

---

MACHINE LEARNING-BASED  
PREDICTIVE MAINTENANCE FOR  
OPTICAL NETWORKS

---

**Dissertation**  
zur Erlangung des akademischen Grades  
Doktor der Ingenieurwissenschaften  
(Dr.-Ing.)  
der Christian-Albrechts-Universität zu Kiel

vorgelegt von  
Khouloud Abdelli  
aus  
El Hamma

Jahr  
2022

---

Vorsitz: Prof. Dr.-Ing. Michael Höft

1. Gutachter: Prof. Dr.-Ing. Dipl.-Wirt. Ing. Stephan Pachnicke

2. Gutachter: Prof. Dr.-Ing. Bernhard Schmauß

Tag der mündlichen Prüfung: 16. März 2023

## Abstract

Optical networks provide the backbone of modern telecommunications by connecting the world faster than ever before. However, such networks are susceptible to several failures (e.g., optical fiber cuts, malfunctioning optical devices), which might result in degradation in the network operation, massive data loss, and network disruption. It is challenging to accurately and quickly detect and localize such failures due to the complexity of such networks, the time required to identify the fault and pinpoint it using conventional approaches, and the lack of proactive efficient fault management mechanisms. Therefore, it is highly beneficial to perform fault management in optical communication systems in order to reduce the mean time to repair, to meet service level agreements more easily, and to enhance the network reliability.

In this thesis, the aforementioned challenges and needs are tackled by investigating the use of machine learning (ML) techniques for implementing efficient proactive fault detection, diagnosis, and localization schemes for optical communication systems to enhance the availability and reliability of such systems. In particular, the adoption of ML methods for solving the following problems is explored:

- Degradation prediction of semiconductor lasers (one of the most important optical components in optical communication systems),
- Lifetime (mean time to failure) prediction of semiconductor lasers,
- Remaining useful life (the length of time a machine is likely to operate before it requires repair or replacement) prediction of semiconductor lasers,
- Optical fiber fault detection, localization, characterization, and identification for different optical network architectures,
- Anomaly detection in optical fiber monitoring.

The different proposed ML-based approaches for solving the above-mentioned problems are validated using either synthetic or experimental data. Given the obtained results, such approaches outperform the conventionally employed methods for all the investigated use cases by achieving better prediction accuracy and earlier prediction or detection capability. To the best of our knowledge, the work in this thesis is one of the first works to investigate the use of machine learning for solving the aforementioned problems.

---

## Zusammenfassung

Optische Netze bilden das Rückgrat modernen Telekommunikationsnetze und verbinden die Welt schneller als je zuvor. Solche Netze sind jedoch anfällig für verschiedene Ausfälle (z. B. Unterbrechung von Glasfasern, fehlerhafte optische Geräte), die zu einer Beeinträchtigung des Netzbetriebs, massiven Datenverlusten und Netzünterbrechungen führen können. Die genaue und schnelle Erkennung und Lokalisierung der Fehler ist aufgrund der Komplexität solcher Netze, des Zeitaufwands für die Identifizierung und Lokalisierung des Fehlers mit herkömmlichen Ansätzen und des Fehlens proaktiver, effizienter Fehlermanagementmechanismen eine Herausforderung. Daher ist ein Fehlermanagement in optischen Kommunikationssystemen von großem Nutzen, um die mittlere Reparaturzeit zu verkürzen, Service Level Agreements leichter einzuhalten und die Zuverlässigkeit des Netzes zu erhöhen.

In dieser Arbeit werden die oben genannten Herausforderungen und Bedürfnisse angegangen, indem der Einsatz von Techniken des maschinellen Lernens (ML) für die Implementierung effizienter proaktiver Fehlererkennungs-, Diagnose- und Lokalisierungsverfahren für optische Kommunikationssysteme untersucht wird, um die Verfügbarkeit und Zuverlässigkeit solcher Systeme zu verbessern. Insbesondere wird der Einsatz von ML-Methoden zur Lösung der folgenden Probleme erforscht:

- Degradationsvorhersage von Halbleiterlasern (eine der wichtigsten optischen Komponenten in optischen Kommunikationssystemen),
- Lebensdauervorhersage (mittlere Zeit bis zum Ausfall) von Halbleiterlasern,
- Vorhersage der Restnutzungsdauer von Halbleiterlasern (die Zeitspanne, in der eine Maschine wahrscheinlich noch funktioniert, bevor sie repariert oder ersetzt werden muss),
- Erkennung, Lokalisierung, Charakterisierung und Identifizierung von Fehlern in optischen Fasern für verschiedene optische Netzwerkarchitekturen,
- Erkennung von Anomalien bei der Überwachung von optischen Fasern.

Die verschiedenen vorgeschlagenen ML-basierten Ansätze zur Lösung der oben genannten Probleme werden entweder anhand synthetischer oder experimenteller Daten validiert. Die erzielten Ergebnisse zeigen, dass diese Ansätze die konventionellen Methoden für alle untersuchten Anwendungsfälle übertreffen, durch eine bessere Vorhersagegenauigkeit und eine frühere Vorhersage oder Erkennungsfähigkeit. Die in dieser Dissertation vorgestellten Ergebnisse zeigen erstmals umfassend den Einsatz von maschinellem Lernen für die Lösung der oben genannten Probleme.

## Acknowledgement

"Feeling gratitude and not expressing it is like wrapping a present and not giving it."  
William Arthur Ward

That's why, I would like to take this opportunity to thank everyone who assisted me in completing my PhD.

First and foremost, I would like to thank God Almighty for giving me the strength, knowledge, and patience to successfully complete my PhD. Without his blessing, this accomplishment would not have been possible.

My deepest gratitude goes to my supervisor, Prof. Dr. Stephan Pachnicke for his expert guidance, unwavering support, and encouragement throughout my PhD study. I am also very thankful to him for introducing and helping me understand the art of scientific research, critical analysis, strategic decision making and peculiarities of academic writing. I am very grateful and fortunate enough to have had Prof. Dr. Stephan Pachnicke as my PhD supervisor.

My sincere thanks also go to my supervisor at ADVA, Dr. Helmut Griebner for advising and supporting me technically and professionally during my PhD journey. He was generous enough to devote numerous hours in reviewing my conference and journal paper submissions, which helped me improve the quality of my research.

I would like to extend my gratitude to the members of my dissertation committee, Prof. Dr.-Ing. Bernhard Schmauß and Prof. Stephan Pachnicke. I would also like to thank the committee chairman, Prof. Dr.-Ing. Michael Höft.

I am also very grateful to my colleagues Mr. Carsten Tropschug, Dr. Joo Cho, and Mrs. Gabriele Hermann for their continuous support, positive spirited attitude, and sharing expert knowledge.

I would like also to thank the laser manufacturers Vertilas and II-VI Incorporated for data sharing and support.

Finally, I would like to express my deepest gratitude to my parents, my brother, and my sisters for their immense support and encouragement throughout my life in general.

---

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation . . . . .	3
1.2	Research Questions and Contributions . . . . .	3
1.3	Thesis Methodology and Outline . . . . .	6
<b>2</b>	<b>Optical Fiber Communications</b>	<b>7</b>
2.1	Optical Transmitters . . . . .	9
2.1.1	Semiconductor Lasers . . . . .	9
2.1.2	Optical Modulator . . . . .	19
2.2	Optical Fiber . . . . .	20
2.2.1	Basic Structure . . . . .	21
2.2.2	Fiber Types . . . . .	22
2.2.3	Fiber Attenuation . . . . .	22
2.2.4	Dispersion . . . . .	26
2.2.5	Nonlinear Effects . . . . .	27
2.3	Optical Amplifiers . . . . .	28
2.3.1	General Applications of Optical Amplifiers . . . . .	29
2.3.2	Optical Amplifier Parameters . . . . .	29
2.3.3	Optical Amplifier Types . . . . .	31
2.4	Passive Optical Components . . . . .	32
2.4.1	Fiber Joints . . . . .	32
2.4.2	Fiber Couplers . . . . .	34
2.4.3	Optical Attenuator . . . . .	36
2.5	Optical Receiver . . . . .	36
2.5.1	Working Principle of Photodiode . . . . .	37
2.5.2	Characteristics of Photodiode . . . . .	37
2.6	System Architectures . . . . .	39
2.6.1	Point-to-Point Links . . . . .	39
2.6.2	Point-to-Multipoint Links . . . . .	39
2.6.3	Meshed Networks . . . . .	40
2.7	System Monitoring . . . . .	41

---

<b>3</b>	<b>Predictive Maintenance</b>	<b>45</b>
3.1	Maintenance Strategies . . . . .	45
3.1.1	Run-to-Failure Maintenance . . . . .	45
3.1.2	Preventive Maintenance . . . . .	46
3.1.3	Predictive Maintenance . . . . .	47
3.1.4	Proactive Maintenance . . . . .	48
3.2	Evolution of Maintenance Strategies . . . . .	48
3.3	Benefits . . . . .	49
3.4	Enabling Technologies . . . . .	50
3.5	Machine Learning . . . . .	51
3.5.1	What is machine learning? . . . . .	51
3.5.2	Standard Learning Tasks . . . . .	52
3.5.3	Learning Stages . . . . .	53
3.5.4	Machine Learning Techniques . . . . .	53
3.5.5	Performance Metrics . . . . .	72
3.6	Use Cases . . . . .	74
3.6.1	Equipment Failure Prediction . . . . .	74
3.6.2	Equipment Remaining Useful Life Prediction . . . . .	74
3.6.3	Anomaly Detection in Equipment Behavior . . . . .	75
3.6.4	Equipment Failure Diagnosis . . . . .	75
3.7	Challenges . . . . .	75
<b>4</b>	<b>Predictive Maintenance for Semiconductor Lasers</b>	<b>77</b>
4.1	Degradation Prediction Use Case . . . . .	78
4.1.1	Model Architecture . . . . .	78
4.1.2	Anomaly Detection . . . . .	79
4.1.3	Experimental Data . . . . .	80
4.1.4	Experimental Results . . . . .	81
4.2	Degradation Type Identification Use Case . . . . .	86
4.2.1	Model Architecture . . . . .	86
4.2.2	Synthetic Data Generation . . . . .	87
4.2.3	Results . . . . .	88
4.3	Laser Lifetime Prediction Use Case . . . . .	90
4.3.1	Model Architecture . . . . .	90
4.3.2	Synthetic Data Generation . . . . .	91
4.3.3	Conventional Lifetime Prediction Method . . . . .	93
4.3.4	Results . . . . .	93

---

4.4	Remaining Useful Life Prediction Use Case . . . . .	95
4.4.1	Model Architecture . . . . .	95
4.4.2	Experimental Data . . . . .	95
4.4.3	Results . . . . .	96
4.4.4	Discussion . . . . .	98
4.5	Predictive Maintenance Framework for Semiconductor Lasers . . . . .	100
4.5.1	Proposed Framework . . . . .	100
4.5.2	Methodology . . . . .	104
4.5.3	Experimental data . . . . .	105
4.5.4	Synthetic data . . . . .	106
4.5.5	Validation results of the performance prediction model . . . . .	108
4.6	Conclusion . . . . .	112
<b>5</b>	<b>Fault Management for Optical Fiber Links</b>	<b>113</b>
5.1	Point-to-Point Optical Links . . . . .	113
5.1.1	Reflective Fault Detection, Localization and Characterization . .	114
5.1.2	Reflective and Non-Reflective Fault Diagnosis and Localization .	124
5.1.3	Combined Approach . . . . .	128
5.2	Point-to-multipoint Optical Link Predictive Monitoring . . . . .	135
5.2.1	Optical Fault Identification, Localization, and Characterization	137
5.2.2	Anomaly Detection in Optical Fiber Monitoring . . . . .	141
5.3	Conclusion . . . . .	148
<b>6</b>	<b>Conclusions and Future Work</b>	<b>149</b>
	<b>Bibliography</b>	<b>154</b>
	<b>Erklärung</b>	<b>I</b>

## Acronyms

**Adam** Adaptive moment estimation.

**AE** Autoencoder.

**AI** Artificial intelligence.

**ANN** Artificial neural network.

**APC** Angled physical contact.

**AUC** Area under curve.

**BiGRU** Bidirectional gated recurrent unit.

**BiLSTM** Bidirectional long short-term memory.

**CNN** Convolutional neural network.

**CVAE** Conditional variational autoencoder.

**DAE** Denoising autoencoder.

**DCAE** Denoising convolutional autoencoder.

**DFB** Distributed feedback laser.

**EDFA** Erbium-doped fiber amplifier.

**ELU** Exponential linear unit.

**FL** Federated learning.

**FPA** Fabry–Pérot amplifier.

**FWM** Four-wave mixing.

**GAN** Generative adversarial network.

- 
- GRU** Gated recurrent unit.
- IoT** Internet of things.
- KNN** K-nearest neighbors.
- LOF** Local outlier factor.
- LR** Linear regression.
- LSTM** Long short-term memory.
- MAE** Mean absolute error.
- MAPE** Mean absolute percentage error.
- ML** Machine learning.
- MLP** Multilayer perceptron.
- MLT** Multi-task learning.
- MSE** Mean squared error.
- MTTF** Mean-time-to-failure.
- MTTR** Mean-time-to-repair.
- MZM** Mach-Zehnder modulator.
- OCSVM** One-Class support vector machine.
- ODN** Optical distribution network.
- OLT** Optical line terminal.
- ONU** Optical network unit.
- ORL** Optical return loss.
- OTDR** Optical time-domain reflectometer.
- P2M** Point-to-multi-point.
- P2P** Point-to-point.

- 
- PC** Physical contact.
- PON** Passive optical network.
- RF** Random forest.
- RLU** Rectified linear unit.
- RMSE** Root mean squared error.
- ROC** Receiver operating curve.
- RUL** Remaining useful life.
- SBS** Stimulated Brillouin scattering.
- SDN** Software defined networking.
- SNR** Signal-to-noise ratio.
- SOA** Semiconductor optical amplifier.
- SPM** Self-phase modulation.
- SRS** Stimulated Raman scattering.
- SVR** Support vector regression.
- UPC** Ultra-physical contact.
- VAE** Variational autoencoder.
- VCSEL** Vertical-cavity surface-emitting laser.
- WDM** Wavelength division multiplexing.
- XPM** Cross-phase modulation.

# 1 Introduction

Optical fiber systems are the backbone of modern telecommunications. Optical fibers allow information to be transmitted from ultrashort cables in data centers to transoceanic distances around the globe efficiently at enormous speeds, providing connectivity to billions of users worldwide, and thus connecting the global world faster while ensuring lower latency and higher bandwidth than ever before [1]. The impact of fiber-optics on communication, various industries (e.g. healthcare, government etc.), and our daily lives cannot be overstated [1]. The ever-increasing demands for higher speed and higher capacity in optical fiber communication systems to achieve higher data rate transmission have made it even more important to ensure the reliability, resilience, and survivability of such systems.

Optical communication networks are prone to a wide range of failures, including fiber cuts, malfunctions of optical devices, soft-failures (e.g. gradual signal degradation), unintentional failures caused by natural disasters, wear out or human errors, etc [2]. Such failures may cause network disruption, resulting in massive financial and data losses, or they may gradually degrade the network operations. For example, the fiber cuts in the Mediterranean Sea in 2008 caused loss of 70% of Egypt's connection to the outside world and between 50% and 60% of India's network outbound connectivity on the westbound route [3]. As a result, it is critical to conduct fault detection, localization (locating the failed element (e.g., the node or link responsible for the failure), identification (determining the cause/source of the failure), and recovery as efficiently as possible. Early and accurately predicting faults and pinpointing them to help optical network operators to plan and schedule their maintenance activities, saving CAPEX/OPEX and reducing mean time to repair (MTTR). This makes it easier to meet service level agreements and improve customer satisfaction by reducing downtime and improving network quality.[4].

Traditionally, it is often challenging to quickly localize and identify failures in optical fiber communication systems due to their complexity, and the fact that fault management in such systems is inherently cross-layer (both physical-layer and network-layer have to be considered for performing the fault analysis) [3]. The discovery of incidents in

such complex systems requires considerable expert knowledge and long probing time until a fault (e.g., broken fiber) is identified and pinpointed. Therefore, a long time is typically required for repairing or replacing the defective components. For instance, the MTTR of a fiber cut is 12 hours, according to Telcordia [5]. Furthermore, proactive failure management mechanisms for early detection of potential anomalies before resulting in system disruption are not yet implemented in most systems. The conventional approaches based on predefined fixed thresholds and/or "if-then" rules are not suitable for modern optical networks featured by high dynamics as they cannot adapt to changing network conditions [3].

In this thesis, the aforementioned issues are addressed by investigating the use of several machine learning (ML) techniques for developing efficient proactive fault management schemes for optical communication systems to enhance the availability and reliability of such systems. In particular, the adoption of ML methods for solving the following problems is investigated:

- Degradation prediction of semiconductor lasers (one of the most important optical components in optical communication systems),
- Lifetime (mean time to failure) prediction of semiconductor lasers,
- Remaining useful life (the length of time a machine is likely to operate before it requires repair or replacement) prediction of semiconductor lasers,
- Optical fiber fault detection, localization, characterization, and identification for different optical network architectures,
- Anomaly detection in optical fiber monitoring.

The implemented ML-based approaches for addressing the aforementioned problems are evaluated using either synthetic or experimental data (a meaningful representative data set for each investigated task). These ML techniques outperform the conventionally employed methods by achieving better prediction accuracy and earlier prediction or detection capability.

In the following sections, the motivation behind this thesis as well as the research questions and contributions are discussed.

## 1.1 Motivation

The following factors (key enablers) are the primary drivers paving the way for effective ML applications for fault management in optical networks:

- **Data availability:** Modern optical equipment (transceivers, reconfigurable optical add-drop multiplexers, amplifiers etc.) is now deployed in optical communication systems with built-in monitoring capabilities [3]. Such equipment is able to generate a large amount of monitoring data, which can be fed into ML algorithms to extract insights from that data for solving fault management problems,
- **Network telemetry:** The monitored data retrieved from monitors or equipment installed in optical communication systems can be collected and elaborated in centralized locations thanks to new advanced control/management solutions such as software defined networking (SDN),
- **Breakthroughs in ML algorithm development:** the advances in ML algorithm development (e.g. improved architecture, better optimization algorithms, etc.) help to improve the performance of ML methods,
- **Advancements in computational power:** The continued advances in computational power help to speed up the training and the inference of the ML algorithms for solving large-scale problems.

## 1.2 Research Questions and Contributions

The main research question investigated in this thesis is:

**Could machine learning methods improve the reliability and the survivability of the optical networks and thus help to realize a self-healing optical communication system?**

The main research question is broken down into the following two sub-questions in order to enhance the reliability of optical communication systems in a systematic fashion.

- **[RQ1] Could machine learning techniques early and accurately predict the degradation and the lifetime of optical components deployed in an optical communication system?**

Improving the reliability of optical components and particularly optical transmitters is critical for ensuring a reliable optical communication system. According to [6], laser problems account for more than 92% of optical transmitter failures. Therefore, it is of great importance to improve the reliability of lasers, which is governed by degradation mechanisms. Such degradation mechanisms are very complex due to the diversity of the factors inducing the degradation, and the complexity of the laser structures [7]. Therefore, it is very challenging to fully understand the physical processes governing degradation. That is why, using physics-based models that use explicit mathematical equations to model the degradation behavior of the system is difficult. Conventionally, the laser lifetime is estimated by extrapolating a mathematical fit of the laser current or output power over time. However, such a reliability extrapolation is inaccurate and can result in a significant overestimation or underestimation of the laser's actual lifetime [8]. The laser is considered degraded, if the value crosses the threshold, which is determined based on the laser design and specifications. However, the threshold approach is imprecise and leads to a high false alarm rate.

To address these challenges, this thesis proposes various ML-based approaches that do not require system specific knowledge or physical models to perform prognostics and health monitoring of laser devices. Such approaches learn system behavior by leveraging monitored data, and thus predict laser degradation and remaining useful life. They are validated using either synthetic or experimental laser reliability data, which reflect both normal operation behavior and the degradation process under different operating conditions. The various proposed approaches outperform conventional methods in terms of prediction accuracy. Contributions to this research question are:

- Develop an ML-based model for early predicting the degradation of lasers deployed in optical networks,
- Propose an ML method for identifying the failure mode (e.g. sudden degradation, gradual degradation) that a laser device might exhibit during operation,
- Develop an ML model for remaining useful prediction of lasers,
- Propose an ML model for lifetime (MTTF) prediction of laser devices,
- Propose an predictive maintenance framework (combined approach whereby different ML models for solving different predictive maintenance tasks (e.g. performance degradation prediction, anomaly detection, remaining useful

prediction) are involved) for lasers deployed in optical communication systems

- Perform comparative analysis of the proposed ML approaches and the conventional methods as well as other ML algorithms.

All these contributions are presented in Chapter 4.

- **[RQ2] Could machine learning techniques quickly and accurately detect, localize, and identify faults in optical fiber links?**

Fiber-optic monitoring has primarily been accomplished through the use of optical time domain reflectometry (OTDR), a technique based on Rayleigh backscattering that is widely used for measuring fiber characteristics as well as detecting and localizing fiber faults [9]. However, OTDR traces can be noisy and hard to interpret even for trained and experienced field engineers. This is primarily due to noise overwhelming the signals. Analyzing OTDR traces using conventional methods can take a long time because a lot of averaging of OTDR measurements is required to remove the noise and achieve a good event detection and localization accuracy.

Various ML techniques for accurately and quickly detecting, localizing, characterizing, and identifying optical fiber faults given noisy OTDR traces are presented in this thesis. The proposed methods achieve good detection and diagnosis capabilities and high localization accuracy even for low SNR values, and outperform conventionally employed techniques. Contributions to this research questions include:

- Propose different ML models for fault detection, localization, and characterization given noisy OTDR signals for point-to-point optical links,
- Develop data-driven approaches for fault management in point-to-multi-point optical links,
- Present an ML-based framework for anomaly detection in optical fiber monitoring,
- Propose an ML framework combining an ML-based model for noise removal of OTDR signals and another ML method for fault detection, localization, and diagnosis with the denoised signal as input, in order to improve the event analysis accuracy under very low SNR conditions,
- Perform comparative analysis of the proposed ML approaches and the conventional methods as well as other ML algorithms.

All these contributions are presented in Chapter 5.

### 1.3 Thesis Methodology and Outline

A generic methodology is adopted in the different phases of this thesis work. It is composed of three main building blocks:

- Data preparation: generating a meaningful data set (a good representative for the problem to be solved using ML models) using analytical models, simulations, or experiments,
- Model development: training the ML model given the prepared data set and optimizing the model to achieve a good performance,
- Assessment: testing and evaluating the performance of the developed ML model (the generalizability and the robustness) using unseen or new data, comparing the performance of the ML model with other ML algorithms and/or conventional methods.

The rest of this thesis is structured as follows:

In Chapter 2, background knowledge of optical communication systems including optical components (e.g. lasers), optical fibers, system monitoring, is presented.

In Chapter 3, the different types of maintenance strategies, the key enablers for predictive maintenance, several ML algorithms, and use cases are presented.

In Chapter 4, different ML approaches for solving several uses cases for laser predictive maintenance are discussed.

In Chapter 5, several ML models for fault management for different optical communication system architectures are investigated.

In Chapter 6, the main conclusions, research limitations, and insights to future work are given.

## 2 Optical Fiber Communications

Fiber-optical communication technology has been evolving over the last few decades so rapidly that it has become the backbone of the global telecommunication networks, one of the main pillars of the “information age”, and is widely used in a wide range of industries, including medical, industrial and government.

The birth of the optical communication era happened thanks to the breakthrough of the laser in 1960, the pioneering discovery of Charles Kao and George Hockham in 1966 about fiber light-loss properties [10], and Corning’s invention of the first low-loss optical fiber in 1970 [11]. The first commercial fiber-optic communication systems were successfully deployed in the 1980s [12]. Since then, a tremendous series of technological advances related to optical fibers, semiconductor lasers and other related optical devices, such as dispersion-compensating optical fibers, tunable laser diodes, the invention of the Erbium-doped fiber amplifier (EDFA) in 1986 [13], and wavelength division multiplexing (WDM) [14], have led to a big improvement in the information-carrying capacity that significantly increased a millionfold from the tens of Mb/s in the 1970s to many tens of Tb/s in recent years [15]. Today more than 80% of the world’s long-distance voice and data traffic is carried over hundreds of millions of kilometers of optical fiber cables, creating an intricate web of connectivity forming the foundation of the World Wide Web [15], and thus connecting the global world faster than ever before.

An optical communication system is a particular type of telecommunication system that uses optical fibers for carrying information from one place to another, whether they are separated by few kilometers or transoceanic distances. Adopting optical fiber as a transmission medium offers a plethora of metrics and extremely appealing features, e.g., low transmission loss (as low as 0.15 dB/km [16]), massive bandwidth ( $\sim 100$  THz, small size and low weight (hair-sized dimensions), immunity to interference (free from electromagnetic interference [16]), signal security (high degree of data security), low cost etc [17]. Therefore, fiber-optic communication systems have enabled amounts of information to be exchanged over long distances, rapidly replacing older conventional telecommunication systems such as copper wire or microwave radio transmission systems in most applications. As shown in Fig. 2.1, a generic optical communication system consists of an optical transmitter, an optical-fiber transmission medium (i.e.,

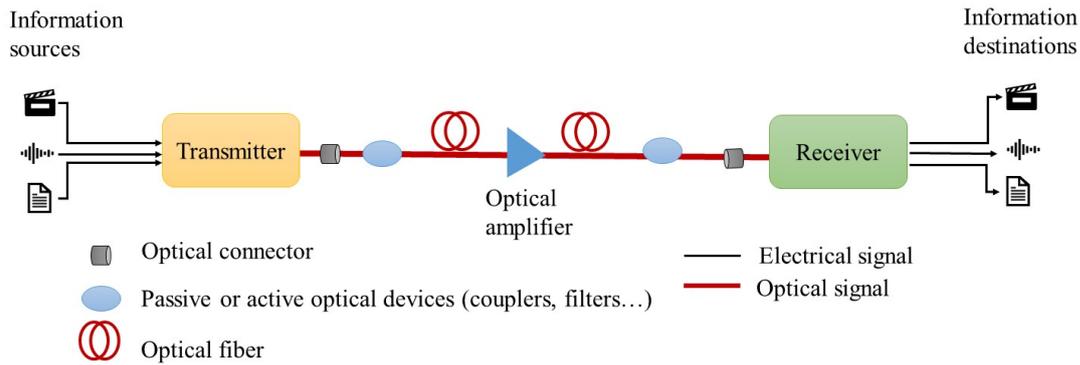


Figure 2.1: Main elements of an optical fiber communication link. The basic components are the transmitter, the optical fiber, and the receiver. Additional components include optical amplifier, connectors, and other passive and active optical devices [18].

communication channel), and a receiver. The information (video, audio, data etc.) to be conveyed enters to the transmitter, where it is converted into an optical form. The resulting light signal is then transmitted through an optical fiber to the receiver, where the optical information signal is converted back to the electrical signal. In addition to the aforementioned main basic building blocks, an optical-fiber communication system may include additional components such as optical amplifiers, which are used to amplify the optical signals when they are attenuated after long distances, connectors, couplers, and other passive components that do not require external power, and further active photonic devices.

The different elements of an optical fiber transmission link are discussed in more details in the following sections. The main two components of an optical transmitter namely the optical source such as a semiconductor laser and the optical modulator are covered in Section 2.1. Section 2.2 gives details on the optical fiber properties including attenuation behavior and signal dispersion characteristics. Section 2.3 is devoted to the study of optical amplifiers. The passive optical components such as optical connectors, couplers, and optical attenuators are addressed in Section 2.4. Section 2.5 deals with the optical receivers and mainly the photodiodes. The different optical network system architectures like point-to-point links, passive optical networks, and meshed networks, are described in Section 2.6. Section 2.7 discusses performance measurement and monitoring including test instruments for fiber link characterization and fault detection and localization.

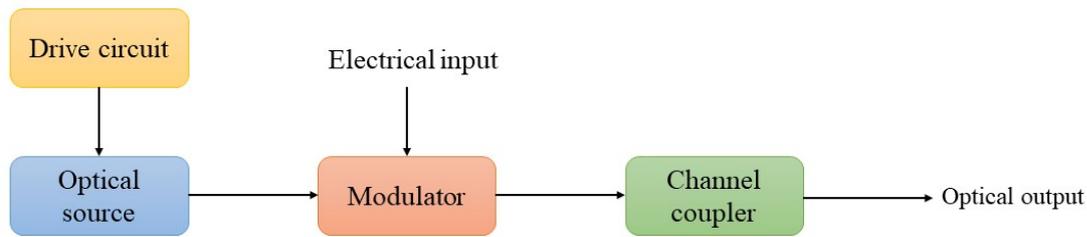


Figure 2.2: Components of an optical transmitter [19].

## 2.1 Optical Transmitters

The role of an optical transmitter is to convert the electrical signal in optical form and then to launch it to the optical-fiber transmission medium. As shown in Fig. 2.2, an optical transmitter is composed of a drive circuit, an optical source, a modulator, and a channel coupler. The optical sources are considered as the heart and the soul of optical transmitters. Optical fiber communication systems typically use semiconductor optical sources, primarily semiconductor lasers or light-emitting diodes, due to the numerous advantages they provide, such as high efficiency, compact size, the ability to directly modulate the light level with an information signal at high frequencies, high reliability, low cost, etc [19]. For low-cost, low-rate, short-distance, and grey (i.e. a single wavelength without stabilization) systems, optical sources can be used to perform direct signal modulation over a wide bandwidth by varying the injection current, eliminating the need for an external modulator inside the optical transmitter and thus simplifying and lowering the cost of the transmitter design. High rate, long distance, and high spectral efficiency systems, on the other hand, necessitate external modulation and WDM with wavelength stabilization. ; The channel coupler, which couples the modulated signal into the optical fiber, is a microlens that efficiently focuses the optical output signal into the entrance plane of an optical fiber [19].

In the next Subsections 2.1.1 and 2.1.2, we leap into the technical details of semiconductor lasers and modulators.

### 2.1.1 Semiconductor Lasers

The first demonstration of a semiconductor laser dates back to 1962 [20]. The early lasers were Gallium-Arsenid (GaAs) homojunction devices operating at low temperature, and with poor performance. In 1970, the achievement of continuous wave operation at room temperature [21] paved the way for the development of semiconductor lasers thanks to the use of heterostructure design [22]. Since then, tremendous improvements in semiconductor laser technology in terms of novel physical concepts, new

materials, and new crystal growth techniques for their fabrications have been accomplished, enhancing the performance, functionality, and the productivity of semiconductor lasers. Today, semiconductor lasers have been widely used in various applications in different fields of science and industry such as fiber-optic communications as primary light sources, medicine for diagnostics and for surgery, metrology, 3D sensing, spectroscopy, etc [23].

In this subsection, the basic concepts such as simulated emission are first discussed. Afterwards, the operating principle of semiconductor lasers is addressed. Then, some types of semiconductor lasers namely distributed feedback lasers (DFBs) [24, 25, 26] and vertical-cavity surface-emitting lasers (VCSELs) [27, 28, 29] are covered. Finally, the reliability of semiconductor lasers is considered.

### ***Fundamentals of Laser***

The laser working principle is a result of three key transition processes, namely, absorption, spontaneous emission, and stimulated emission [30]. Figure 2.3 illustrates these three processes, where  $E_1$  and  $E_2$  denote the ground state energy (i.e., the state of lowest energy), and the energy at the excited state (i.e., the state having more energy than the ground state), respectively. The transition between these two states leads to either the absorption or the emission of a photon with an energy  $h\nu = (E_2 - E_1)$  [19]. If an incident photon of energy  $h\nu$  may be excited from a lower energy state into a higher energy state, the process is referred to as absorption or sometimes stimulated absorption. If an electron is at the excited state  $E_2$ , which is an unstable state, it returns to the lower energy state  $E_1$  by emitting a photon of energy  $h\nu$ . If the emission process occurs in an entirely random manner without any external stimulation as shown in Fig. 2.3 (b), it is called spontaneous emission. Whereas if the emission process happens due to an external stimulation initiated by an existing photon of energy  $h\nu$  causing the electron at the excited state to return back to the ground-state level with the creation of a second photon in phase with the incident photon as illustrated in Fig. 2.3 (c), this process is referred to as simulated emission [31].

For systems operating in thermal equilibrium at room temperature, the spontaneous emission dominates over the simulated emission, and thus the emitted radiation in this regime is random and incoherent. Therefore, the lasers must operate at non-thermal equilibrium state. This can be achieved, if the stimulated emission can be the dominant process. The stimulated emission can exceed the other processes only, if the atomic density in the excited state is greater than the atomic density in the ground state [19]. This condition is known as population inversion because the population is the inverse of thermal equilibrium. The population inversion is achieved by pumping lasers with

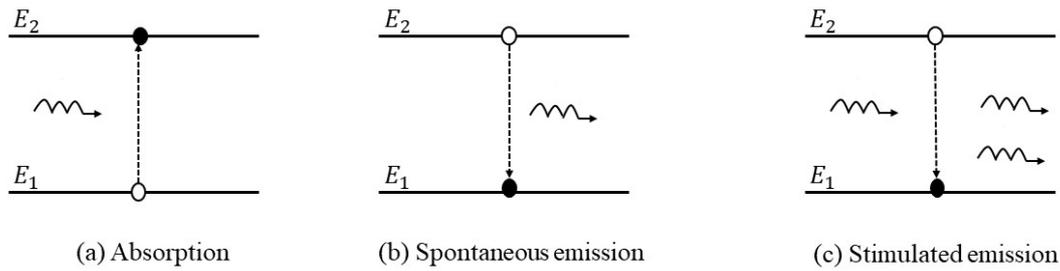


Figure 2.3: Energy level diagrams showing the transition processes; the white dot representing the initial state of the electron and the black dot denoting the final state after the transition took place [19].

an external energy source in order to excite atoms into the upper energy level  $E_2$  and hence obtain a nonequilibrium distribution. To create a population inversion, at least a three-level system [32, 33] is required. Fig. 2.4 illustrates the energy-level diagrams for two systems with three and four energy levels respectively, that can produce population inversion. As shown in Fig. 2.4 (a), the three-level system is composed of a ground level  $E_0$ , a metastable level  $E_1$ , and an excited level lying above the metastable state  $E_2$ . The pumping excites the electrons from the ground state to the higher level  $E_2$ . Given that  $E_2$  is a short-lived state, the electrons will quickly decay either to  $E_1$  or to  $E_0$  [34]. Since the metastable state  $E_1$  exhibits a much longer lifetime than  $E_2$  ( $\sim 1000$  times longer), over a period of time, the density of accumulated atoms at  $E_1$  increases above those in the ground state and thus a population inversion between the levels  $E_1$  and  $E_0$  is achieved. Consequently, the lasing can take place between  $E_1$  and  $E_0$ . A downside with the three-level systems is the very high pumping requirements to produce the population inversion. By contrast, the four-level systems, illustrated in Fig. 2.4 (b), require much lower pumping powers. For such systems, the electrons are excited to the fourth level energy  $E_3$  and then they decay rapidly to the third level  $E_3$ . Since the population at the second level  $E_2$  is relatively low, a population inversion between  $E_2$  and  $E_3$  is obtained and thereby the lasing occurs, creating radiative electron transitions between these two levels.

The basic working principle of a laser is shown in Fig. 2.5. A laser system consists of three main parts: (i) an amplifying or gain medium, (ii) a pump source exciting the gain medium, and (iii) an optical cavity or resonator. The gain medium is the material (solid, liquid, gas, or semiconductor) in which the spontaneous and stimulated emission processes take place resulting in the phenomenon of optical gain or amplification, and it determines the properties of emitted light such as lasing wavelength (or frequency) [35]. The pump source provides the energy required to create the population inversion and thereby satisfying the conditions of light amplification. The optical cavity is typ-

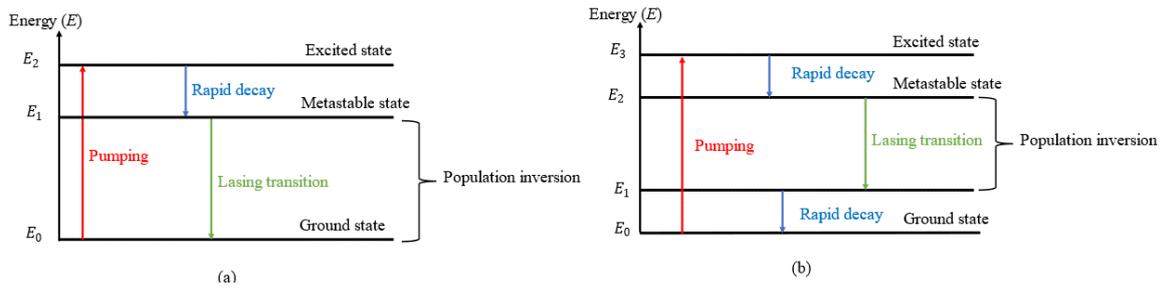


Figure 2.4: Energy-level diagrams showing the population inversion achieved at: (a) three-level and (b) four-level systems [34].

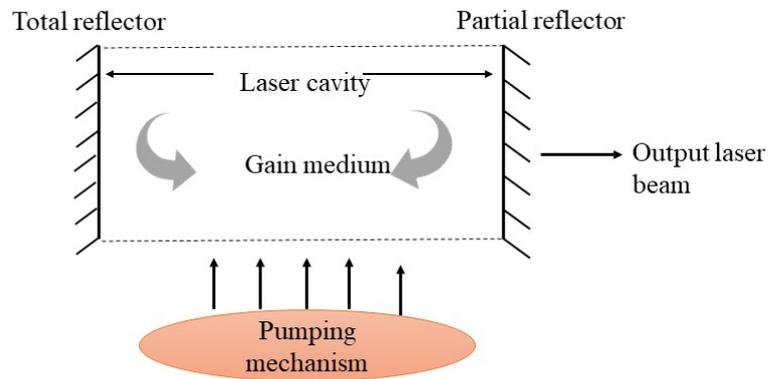


Figure 2.5: Basic laser system. It consists of three major parts: (1) optical cavity, (2) gain medium for light amplification, and (3) pumping mechanism to maintain the gain medium in population inversion [18, 36].

ically formed using two reflective optical elements (e.g., mirrors) surrounding the gain medium in order to force the light to flow back and forth through the medium [35], and thus providing positive feedback of light to sustain the laser action. The light leaking out of the optical cavity forms the output beam of laser.

### ***Properties of Semiconductors***

Semiconductors are materials with a moderate electrical conductivity lying in between that of conductors and insulators [37]. Semiconductors, conductors, and insulators can be distinguished on the basis of the forbidden energy gap or band gap  $E_g$  defined as the difference of energy levels of conduction band and valence band [38], as illustrated in Fig. 2.6. Insulators have a large band gap ( $E_g > 3$  eV), whereas semiconductors possess lower energy gaps, which lie roughly in the range of 0.1 to 3 eV [39]. Conductors have very small or no band gap. For a semiconductor in thermal equilibrium, the occupation probability for an electron to occupy an energy level  $E$  is described by the Fermi-Dirac

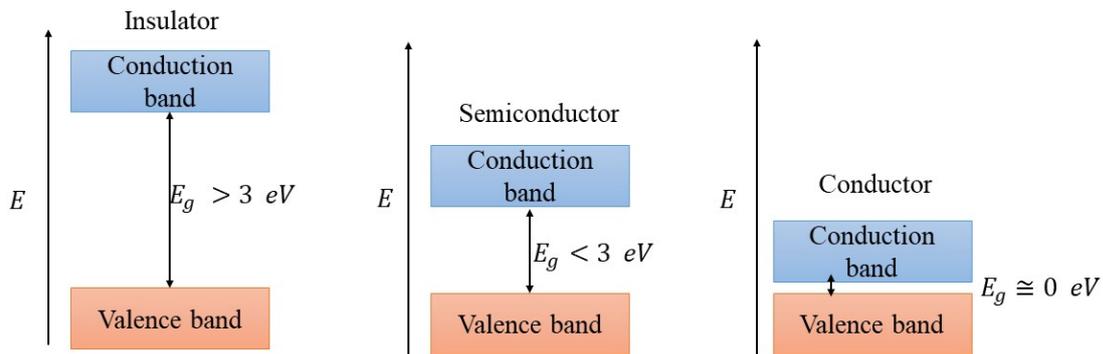


Figure 2.6: Energy-level diagrams showing the difference among insulator, semiconductor, and conductor [39].

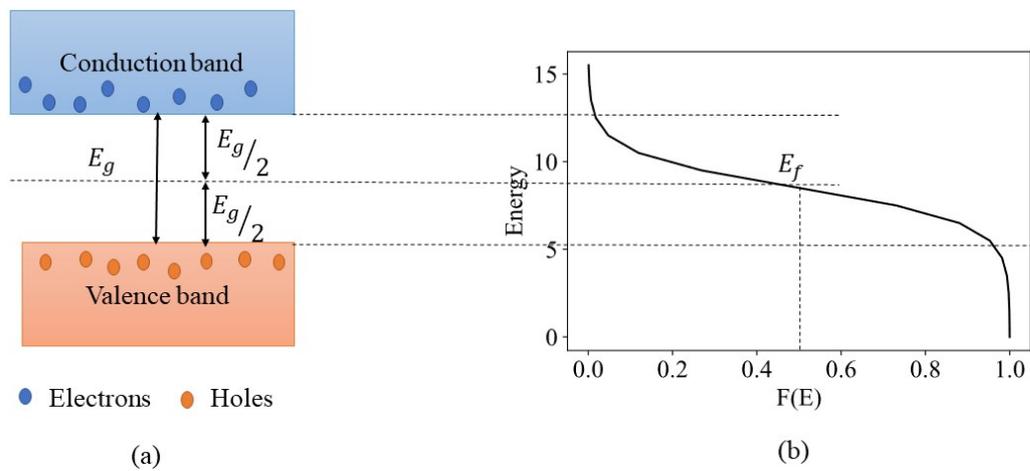


Figure 2.7: a) The energy band structure of an intrinsic semiconductor at a temperature above absolute zero. (b) The Fermi–Dirac probability distribution [16].

distribution [40]:

$$F_E = \frac{1}{1 + \exp\left(\frac{E - E_f}{k_B T}\right)} \quad (2.1)$$

where  $E_f$  denotes the Fermi energy or Fermi level.

For an intrinsic semiconductor (i.e., undoped semiconductor) defined as a perfect semiconductor crystal containing no impurities [16], the energy band structure at a temperature above absolute zero shows an equal number of electrons and holes in the conduction band and the valence band and the symmetry of the density of states around the center of the band gap as illustrated in Fig. 2.7 (a), and the Fermi level lies in the middle of the energy gap [41] as shown in Fig. 2.7 (b). The conductivity of an intrinsic semiconductor is very low at a room temperature. Therefore, to increase the conductivity, doping by adding impurities to an intrinsic semiconductor is commonly adopted. There are two types of dopants, n-type (electron donors), and p-type (electron

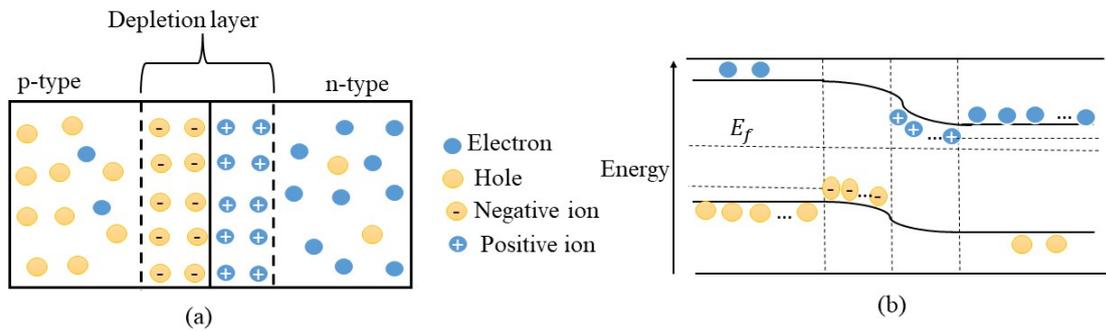


Figure 2.8: a) p-n junction of a semiconductor. (b) The energy band diagram corresponding to (a) [45].

acceptors) dopants [42]. By bringing n-type and p-type semiconductors into contact, a p-n junction, a fundamental building block of semiconductor devices, is formed, as illustrated in Fig. 2.8 (a). The electrons in the n-type semiconductor diffuse to the p-side and combine with holes, which similarly spread from the p-side to the n-side. The free electrons crossing the junction from n-side to fill the holes in p-side atoms create negative ions at p-side, whereas the free electrons leaving the parent atoms at n-side to fill the holes in p-side atom induce positive ions at n-side [43]. As a result, a region near the junction, called a depletion region, is created. This region acts like a barrier preventing further flow of majority charge carriers. When the junction is in an equilibrium state, the Fermi level becomes flat across the n-type and p-type semiconductors as illustrated in Fig. 2.8 (b). The width of the depletion layer depends on the dopant concentration of n- and p-type regions and any applied voltage [44]. When the p-n junction is under forward bias (applying an external positive voltage to the p-type side with respect to the n-type), the depletion region is reduced or even removed, and thereby the current can flow readily through the junction.

### ***Principle of Semiconductor Lasers***

Semiconductor lasers are specifically a p-n junction, emitting coherent light when it is forward biased [46]. The electrons from the n-type side and the holes from the p-type region recombine with each other at junction to emit photons. The released photons then stimulate the electrons and the holes to emit radiation. Consequently, stimulated emission takes place, and thereby producing lasing. The population inversion is achieved by forward biasing of a p-n junction. The laser oscillation (i.e optical feedback) is attained by cleaving the end faces of the p-type and n-type semiconductors [31]. The active medium of a semiconductor laser is a p-n junction. If the p-n junction is made of single type of semiconductor material, the semiconductor laser is called a

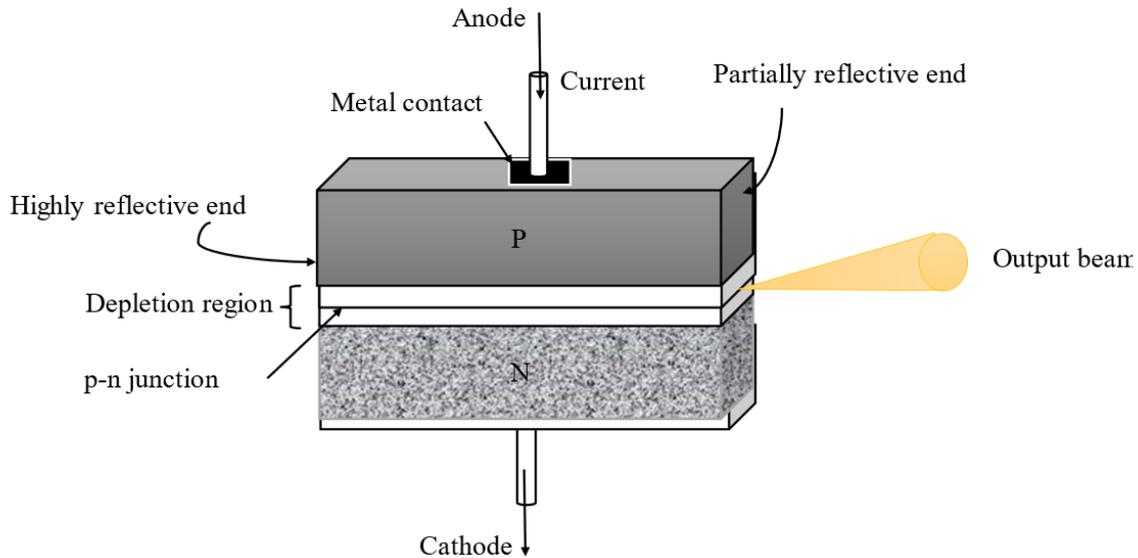


Figure 2.9: Basic structure of a semiconductor laser [48].

homojunction [47]. Whereas if the medium is composed of different types of semiconductor materials, the semiconductor is known as heterojunction laser [47]. Figure 2.9 illustrates the simplest structure of a semiconductor laser consisting of a thin active layer (light emission layer) sandwiched between p-type and n-type cladding layers.

### **Types of Semiconductor Lasers**

The most commonly used semiconductor lasers in optical communication systems are the following:

- Distributed Feedback (DFB) Lasers [49] are a type of lasers that achieves the optical feedback using Bragg reflection gratings (corrugated reflectors) formed by periodically changing the refractive index, rather than the usual cleaved mirrors [50]. The grating is etched onto one of the cladding layers surrounding the active region [50] as shown in Fig. 2.10. DFB lasers oscillate at a single longitudinal mode with a narrow linewidth thanks to the Bragg grating inducing the reflection at certain wavelength.
- Tunable Semiconductor Lasers [51, 52], a kind of lasers allowing the tunability (i.e., adjustment) of the output wavelength, are required by many lightwave systems such as wavelength-division-multiplexed optical communication, and coherent optical communications, which demand wavelength stabilization to ensure that the laser frequency is usually  $\pm 2.5\text{GHz}$  within the nominal channel frequency. They are composed mainly of an active medium with a laser gain across

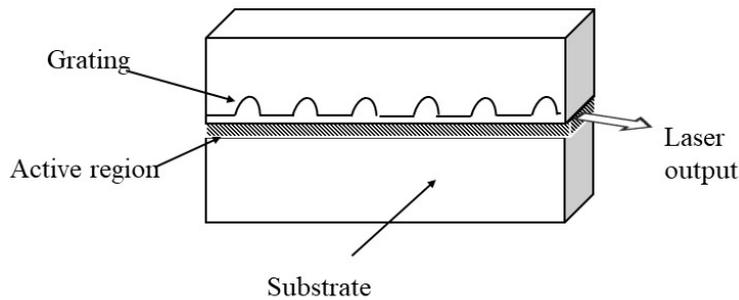


Figure 2.10: Structure of a distributed feedback (DFB) laser [39].

an adequate range of wavelengths and a laser resonator limiting oscillation to a narrow band of wavelengths. There are several tuning mechanisms such as changing the refractive index with temperature, adjusting the injection current and/or the temperature, changing the length of the cavity by moving an external mirror back and forth and thus modifying the resonant wavelength, etc [53].

- Vertical-Cavity Surface-Emitting Lasers (VCSELs) [54] are a type of semiconductor lasers emitting coherent light from the substrate surface vertically as opposed to the conventional edge emitting semiconductor lasers which radiate a beam from surfaces [55]. VCSELs have emerged as the key light source for optical communications, particularly widely deployed in high-speed local area networks such as Gigabit Ethernet or short-reach optical interconnects for datacom links [56]. Figure 2.11 illustrates the structure of a typical VCSEL which includes several layers. The active region (p-n junction) made of multiple quantum wells to improve the light gain, lies between Bragg reflectors with alternating high and low refractive index material, which act as reflective mirrors providing positive feedback. VCSELs offer a plethora of advantages such as low power consumption, high-speed modulation with low driving current ( $<100 \mu\text{A}$ ), low cost and small packaging capability, single longitudinal mode operation with extremely small resonant cavity ( $\sim 1 \mu\text{m}$ ), narrow circular beam for direct fiber coupling, etc [57, 19].

### ***Light-Current Characteristics***

The Light-current curve, also called the power-current (P-I) curve, is considered as a key performance characteristic of semiconductor lasers characterizing the emission properties as it illustrates not only the threshold level but also the driving current required to obtain a certain amount of power. Figure 2.12 shows the P-I curves of a

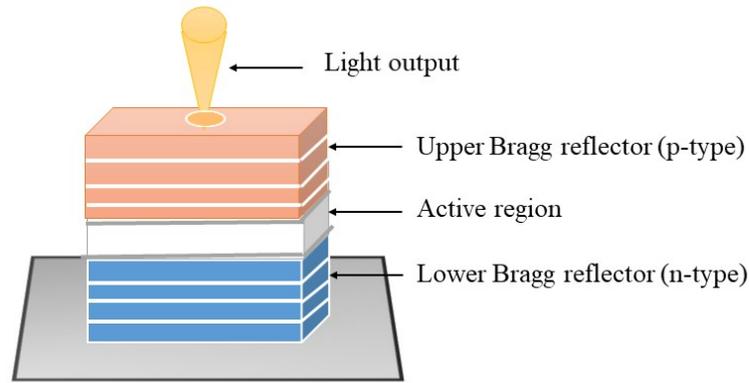


Figure 2.11: Structure of a vertical-cavity surface-emitting laser (VCSEL) [58].

VCSEL laser at temperatures in the ranges of 50-90 °C. The threshold current ( $I_{th}$ ) tends to increase exponentially with the temperature, i.e [19],

$$I_{th} = I_0 \exp\left(\frac{T}{T_0}\right) \quad (2.2)$$

where  $I_0$  is a constant,  $T$  denotes the device absolute temperature, and  $T_0$  is a characteristic temperature describing the quality of the material, used to express the temperature sensitivity of the threshold current.

The slope of the P-I curve for  $I > I_{th}$  is called the slope efficiency, also referred as differential (quantum) efficiency,  $S = \Delta P / \Delta I$ . This characteristic describes the efficiency of a laser in converting electrical current in optical power. For the laser-oscillating region ( $I > I_{th}$ ), the curve is generally close to linear, thus the output optical power  $P$  can be expressed as [59]:

$$P = S (I - I_{th}) \quad (2.3)$$

### **Reliability**

Semiconductor lasers should operate reliably over a long period of time ( $10^5$  hours) in order to ensure reliable data transmission in optical communication networks. The reliability of a semiconductor laser is governed by degradation mechanisms determining its lifetime. Figure 2.13 shows the different failure modes that can be roughly categorized as rapid, sudden, and gradual degradation. If a laser is operated at a constant current, the output power is decreased during laser degradation. Whereas at a constant optical output power operation mode, the current is increased as the device performance degrades in order to maintain constant output power. Rapid degradation is normally

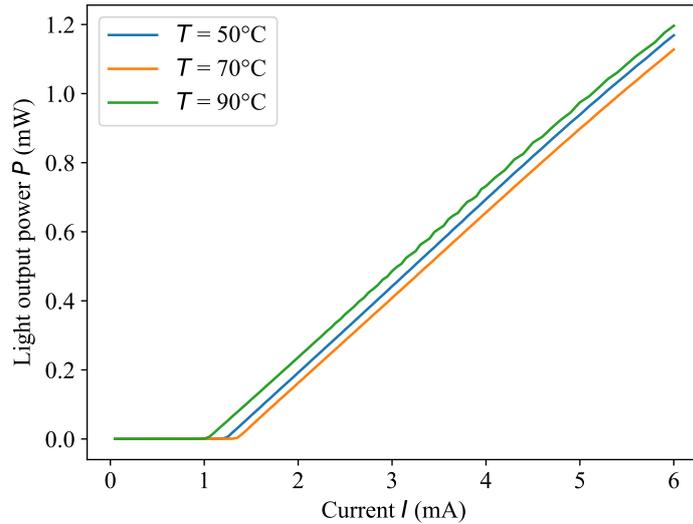


Figure 2.12: P-I curves at several temperatures for a VCSEL laser [18].

observed in the first hundred hours of operation, and manifests itself as a quick decrease in the optical output power or a quick increase in the current. Gradual degradation, the usual failure mode of a device operating over its service life, occurs over a long period (several thousand hours), and determines ultimately the lifetime of the laser. It is characterized by gradual decrease in the quantum efficiency [60]. A device can also degrade suddenly after a regular operation life of the laser, which represents the end of life for the device. Sudden or catastrophic degradation usually occurs at laser mirror facets or in the bulk of the cavity at certain defects [7]. It is very challenging to predict the devices which are likely to degrade catastrophically. The commonly adopted approach to screen out the lasers that might be prone to sudden degradation is to perform burn-in or accelerated aging tests conducted under high stress conditions such as high temperatures or high drive current levels in order to speed up the degradation of the devices. Under highly stressed conditions, fragile devices will exhibit sudden or rapid degradation and thereby failing early before the end of the aging tests. The degradation rate is used to determine the semiconductor laser lifetime and the mean time to failure (MTTF) at high temperature  $T_1$ . The MTTF at the normal operating temperature  $T_2$  is calculated by extrapolation using Arrhenius model [61] as follows:

$$MTTF_{T_2} = MTTF_{T_1} \exp\left(\frac{E_a}{k_B} \left(\frac{1}{T_2} - \frac{1}{T_1}\right)\right) \quad (2.4)$$

where  $E_a$  denotes the activation function.

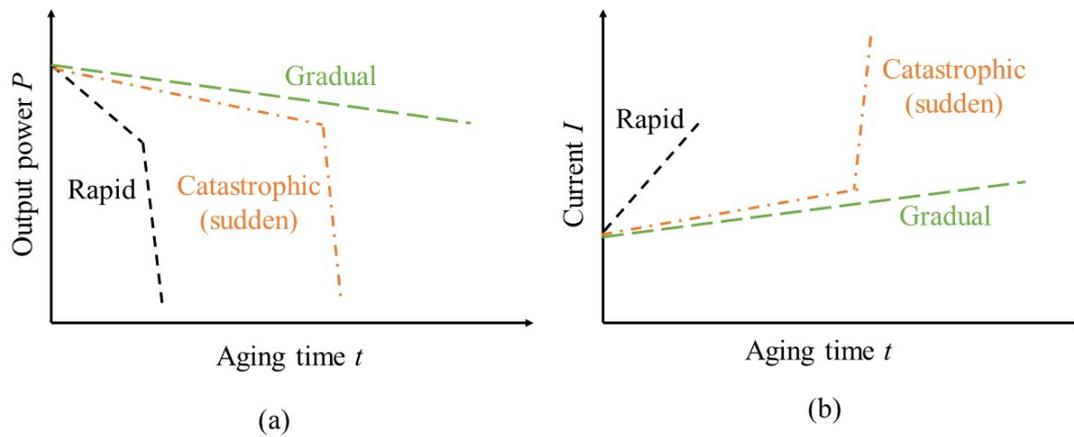


Figure 2.13: Degradation modes: (a) in constant current, and (b) in constant power modes.

### 2.1.2 Optical Modulator

An optical modulator [62] is used to modulate the property of the light such as intensity, phase, frequency, or polarization (direction) with the information signal [63]. The choice of the light property to be moderated is determined by the characteristics of the optical fiber, the available optical sources, and detectors as well as the overall system considerations. The modulation can be applied by the direct modulation of semiconductor lasers [64, 65] or by adopting external modulation of a continuous wave (CW) optical source [66, 67] as shown in Fig. 2.14. The direct modulation of lasers is realized by varying the injection current. Although such modulation method is simple and robust, it is limited by many physical phenomena including chirping, which produces a widening of the laser linewidth, turn on delay, laser resonance frequency, saturation, clipping, and laser non-linear effects [68]. The aforementioned phenomena produce nonlinear distortions, and cause bandwidth and transmission limitations in optical fibers arising from dispersion effects. Therefore, it is undesirable to adopt such a modulation scheme for operation at data rates greater than 2.5 Gb/s [68, 18]. For higher data rates, and more complex and higher performance systems, an external modulator is typically used to temporally vary the steady optical output power level emitted by the laser [18]. It can be physically integrated into the same package with the optical source, or it can be a separate device. The external modulation offers multiple modulation performance advantages such as much wider bandwidth ( $\sim 60$  GHz), larger on/off extinction ratios, and superior modulation spectral purity, however, it is more complex and expensive. The main types of external modulators widely used in optical communication systems are the electro-optical (EO) phase modulator (also

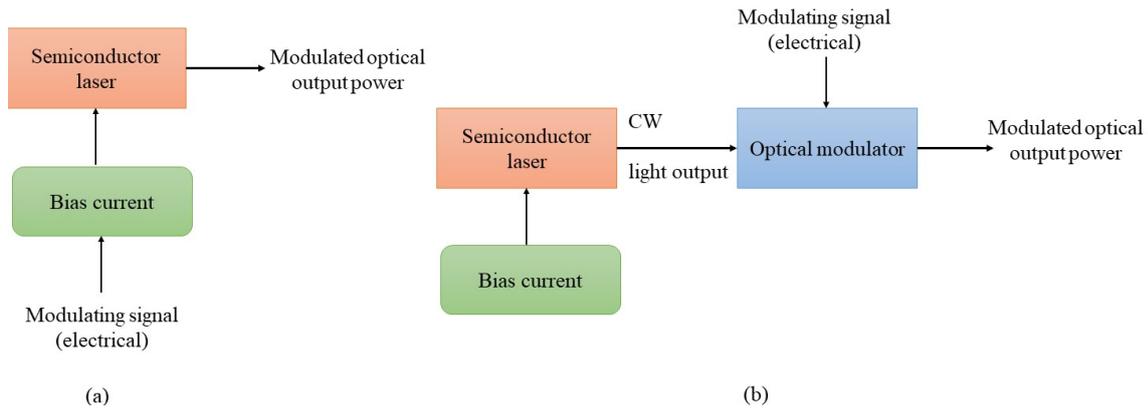


Figure 2.14: (a) Schematic diagram of direct modulation, (b) schematic diagram of external modulation [69].

called Mach-Zehnder Modulator or MZM) [70] and the electro-absorption (EA) modulator [71]. The EO modulator is an optical device used for modulating the power, phase, or polarization of light with an electrical control signal. Its operating principle is based on the Pockels effect (also known as linear electro-optical effect), which is the change in the refractive index induced by an electric field in proportion to the field strength [72]. Figure 2.15 shows the typical structure of an MZM. The input light is split in two separate paths. The applied voltages  $V_1$  and  $V_2$  change the phase of the light signals in both paths. Then, the phase modulated signals are recombined either constructively or destructively [18]. The MZMs can offer pure amplitude modulation (i.e., zero chirp modulation) and thus they can be adopted for several types of modulation formats requiring precise control of the phase and amplitude. The EA modulator is a semiconductor-based optical modulator used to vary and control the intensity of a laser beam (i.e., the output power) via an electric voltage [73]. Its basic principle is based on the Franz–Keldysh effect (i.e., the change in absorption spectrum by applying an electric field (reverse bias), and thereby varying the bandgap)[74]. Such effect can be improved by adopting nano structures such as quantum wells to induce another effect called quantum confined Stark effect. The EA modulators bring several benefits such as zero biasing voltage, high bandwidth, low chirp, high-speed intensity modulation, etc.

## 2.2 Optical Fiber

This section focuses on the optical fibers as the communication channel in optical communication systems. In Subsection 2.2.1, the structure of an optical fiber is discussed. Subsection 2.2.2 covers the different types of optical fibers. The fiber attenuation and

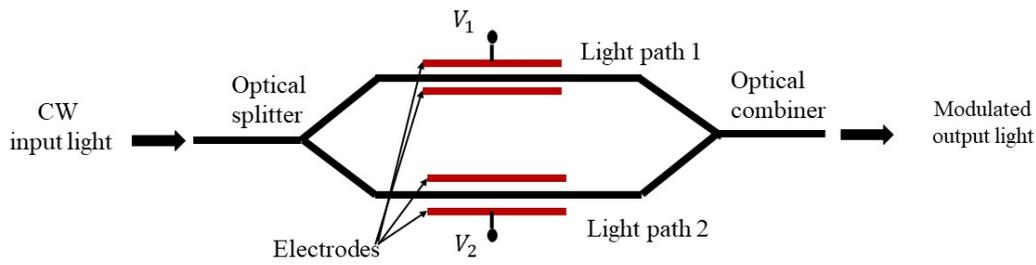


Figure 2.15: Schematic diagram of a Mach-Zehnder Modulator [18].

optical absorption are considered in Subsections 2.2.3 and 2.2.4, respectively. The loss mechanisms in optical fibers including scattering and bending losses are discussed in Subsections 2.2.5 and 2.2.6. The origin of fiber dispersion is considered in Subsection 2.2.7. Subsection 2.2.8 is devoted to nonlinear effects.

### 2.2.1 Basic Structure

An optical fiber is a thin, flexible, transparent, cylindrical dielectric waveguide consisting of a solid glass cylinder called the core surrounded by a dielectric cladding as shown in Fig. 2.16. The core serves as the medium for light propagation. The typical core sizes are 9 and 50  $\mu\text{m}$  [18]. The core and the cladding are produced from the same material of silica glass ( $\text{SiO}_2$ ), however, they have different optical material properties, i.e, different refractive indexes denoted by  $n_1$  and  $n_2$ , respectively, in such a way that  $n_2$  is less than  $n_1$  in order to achieve light guiding in the optical fiber. The difference in refractive indexes can be realized by doping the silica by different dopants. Please recall that the refractive index denotes the optical property of the material describing, how the material affects the speed of the light travelling through it, expressed as:

$$v = \frac{c}{n} \quad (2.5)$$

where  $v$  denotes the velocity of light inside a material having a refractive index  $n$ , and  $c$  is the speed of light in vacuum. The cladding material is pointedly chosen to ensure the efficiency of the light propagation along the core of the optical fiber. Furthermore, the cladding minimizes the scattering loss resulting from the dielectric discontinuities at the core surface, and it prevents the core of the fiber from absorbing surface contaminants. The standard cladding diameter is 125  $\mu\text{m}$  for most optical fiber types [18]. The core and the cladding layers are surrounded by an abrasion-resistant plastic buffer coating with a nominal diameter of 250  $\mu\text{m}$ , to protect the fiber from mechanical stress and environmental effects [18]. Lastly, an outer protective polymer

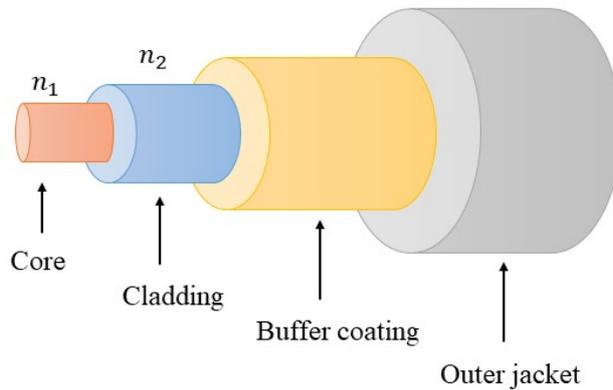


Figure 2.16: Schematic diagram of a generic optical fiber structure [18].

jacket encapsulate the optical fiber.

### 2.2.2 Fiber Types

Based on the refractive index profile, optical fibers can be classified into two types: step-index and graded-index fiber as depicted in Fig. 2.17 [18, 75]. For the step-index fiber category, the refractive index of the core is constant throughout it and encounters an abrupt change (or step) at the core-cladding interface or boundary. Whereas for the graded-index fiber type, the refractive index of the core changes gradually as a function of the radial distance from the fiber center [18]. The step-index and graded-index fibers can be further categorized into single mode and multimode types. Single mode fibers allow only one mode of propagation to be travelling through the fiber, and are generally step-index fibers. They usually have a very small core diameter [76]. They are widely used for long haul communication systems because of high bandwidth [76]. The multimode fibers allow many propagation modes. They can be either step-index or graded-index fiber. They have larger core diameter compared to the single mode fibers. They make it easier to launch optical power into the fiber as well as the connecting together of similar fibers. They are widely used for short fiber links due to the limited bandwidth restricted by intermodal dispersion.

### 2.2.3 Fiber Attenuation

Fiber attenuation (also known as signal attenuation, fiber loss, or power level reduction) [77] is considered as one of the most important properties of an optical fiber that limits the performance of fiber-optic communication systems as it reduces the average power reaching the receiver. It determines the maximum unamplified or repeaterless distance between the transmitter and the receiver or an in-line amplifier [78]. The attenuation

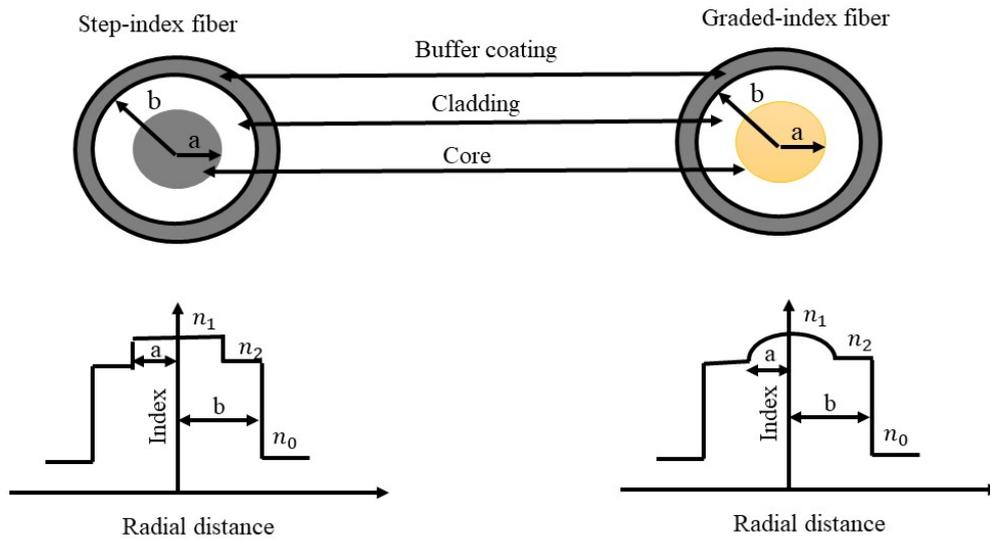


Figure 2.17: Refractive index profiles in step-index fiber and graded-index profile [19].

in an optical fiber is caused by several mechanisms including absorption, scattering, and bending losses [79]. The absorption is associated with the material composition and fabrication process of fiber [80], whereas the scattering is related to both the fiber material and to waveguide imperfections, and the bending losses are induced by perturbations (both microscopic and macroscopic) of the fiber geometry [81, 18]. Fiber attenuation can be expressed by the general relation [19]:

$$\frac{dP}{dz} = -\alpha P \quad (2.6)$$

where  $\alpha$  denotes the attenuation coefficient,  $P$  is the optical power, and  $z$  represents the propagated distance along the optical fiber.

If  $P_{in}$  is the power launched at the input of the fiber, the output power  $P_{out}$ , the remaining power after propagating a distance  $z$  along the fiber length, is given by [19]:

$$P_{out} = P_{in} \exp(-\alpha z) \quad (2.7)$$

$\alpha$  is often expressed in units of dB/km by using the relation [19]:

$$\alpha \text{ (dB/km)} = \frac{-10}{z} \log_{10} \left( \frac{P_{out}}{P_{in}} \right) = 4.343 \alpha \quad (2.8)$$

and refers to it as the fiber loss.

The fiber attenuation depends on the wavelength of transmitted light.

### ***Optical Absorption***

Absorption in an optical fiber is considered as the major cause of fiber optic loss during transmission [82]. It is defined as the portion of fiber attenuation originating from the conversion of the optical power in another energy form such as heat [83]. It is the result of three mechanisms: atomic defects (i.e., imperfections in the atom structure of the fiber material), intrinsic absorption by basic fiber-material properties, and extrinsic absorption by impurity atoms in fiber material [84, 85]. The atomic defects cause absorption due to the missing of oxygen molecules, the presence of high-density clusters of atom groups, or oxygen defects in the fiber material structure [18]. The absorption losses originating from these defects are usually negligible compared to the losses due to the intrinsic and extrinsic absorption factors [84]. The intrinsic absorption arises from electronic absorption bands in the ultraviolet region, and it is mainly caused by the characteristic vibration frequency atomic bands in the near-infrared region. For an absolutely pure fiber material with no impurities or imperfections, the only existing absorption mechanism is intrinsic. Therefore, the intrinsic absorption sets the fundamental lower limit on absorption [18]. The extrinsic absorption is induced due to the presence of minute quantities of impurities in the fiber material including the water ions dissolved in the glass and transition metal ions such as iron, copper, chromium, and vanadium [18].

### ***Scattering Losses***

Scattering losses in optical fibers are caused by microscopic variations in the material density, compositional fluctuations, and structural inhomogeneity or defects occurring during fiber manufacturing [18]. Such losses fall into two categories: linear scattering and non-linear scattering mechanisms. The linear scattering effects result in linearly transferring a portion of the total optical power within one directional mode into a different mode, that can be a leaky or radiative mode, which does not propagate within the fiber core, but it radiates out from the fiber, and thereby causing the attenuation of the transmitted signal within the optical fiber [86]. They are characterized by having no change in frequency for the scattered light. The linear scattering mechanisms can be further classified into Rayleigh scattering and Mie scattering [87]. Rayleigh scattering is elastic scattering of particles (e.g., atoms, molecules) much smaller than the optical wavelength  $\lambda$  (typically  $\lambda / 10$ ). During the propagation of the light inside the core of the optical fiber, the elastic collisions between the light wave and the silica molecules give rise to Rayleigh scattering mechanism as shown in Fig. 2.18. Rayleigh scattering accounts for 96% of the attenuation in optical fibers [88]. The amount of the scattering

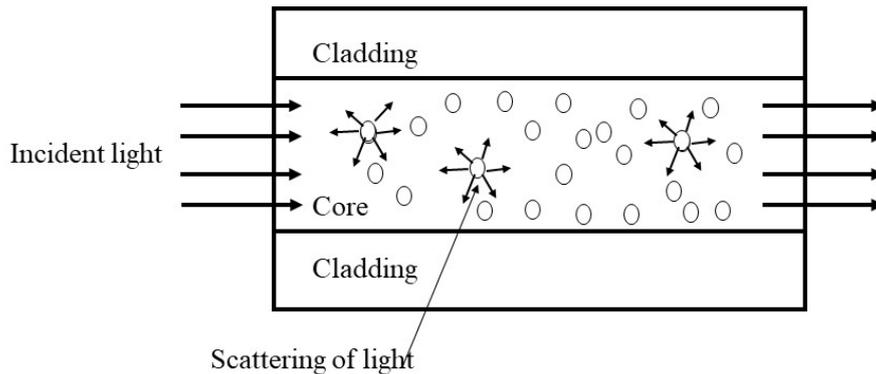


Figure 2.18: Schematic diagram of Rayleigh scattering in optical fibers [91].

of light is inversely proportional to the fourth power of wavelength [83], so it decreases dramatically with the increasing wavelength as depicted in Fig. 2.19. The scattering loss at a wavelength  $\lambda$  resulting from Rayleigh scattering can be approximated by [89, 90]:

$$\gamma_R = \frac{8\pi^3}{3\lambda^4} n^8 \rho^2 \beta_T k_B T_f \quad (2.9)$$

where  $\gamma_R$  is the Rayleigh scattering coefficient,  $\rho$  denotes the photoelastic coefficient,  $\beta_T$  is the isothermal compressibility of the material, and  $T_f$  represents the fictive temperature (i.e., the temperature at which the density fluctuations are frozen into the glass as it solidifies [18]).

Mie scattering mechanism [92] describes the scattering of propagating electromagnetic waves by spherical particles, which have a diameter similar to or larger ( $> \lambda / 10$ ) than the wavelength of the incident light. It occurs due to inhomogeneities in the composition of silica, irregularities in the core-cladding interface, or difference in core cladding refractive index, etc [85]. The nonlinear scattering effects in optical fibers arise from the inelastic scattering of a photon to a lower energy photon [93]. There are two dominant nonlinear scattering phenomena namely stimulated Raman scattering (SRS) and stimulated Brillouin scattering (SBS) [94]. The fundamental difference between the two is that optical phonons participate in SRS, whereas acoustic phonons take part in SBS [93]. Consequently, SBS occurs only in the backward direction while SRS can occur in both forward and backward directions.

### **Fiber Bending Losses**

Optical fibers suffer from bending losses which come into two forms: macrobending and microbending [95, 96]. Macrobending refers to the attenuation associated with wrapping or bending the optical fiber. Macrobends, having larger radii than that of

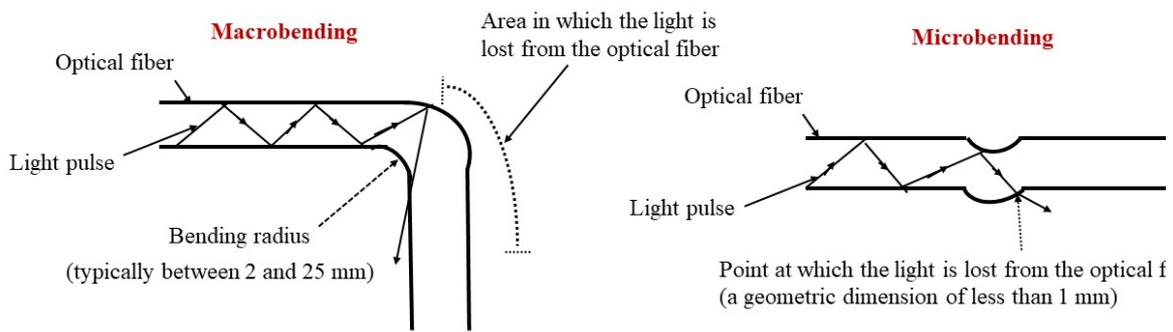


Figure 2.19: Bending losses in optical fibers [99].

the fiber diameter [18], can be seen by the human eye, and occur when a fiber cable turns around a mandrel or a corner as illustrated in Fig. 2.19. When the fiber is bent, light can leak out of the fiber. As the bend becomes more sharp, more light leaks out. Microbends [97, 98] (bends too small to be seen with the naked eye) are random microscopic bends of the fiber core arising from non-uniformities in the manufacturing of the fiber or by non-uniform lateral pressures created during the cabling of the fiber.

#### 2.2.4 Dispersion

Optical dispersion, a major factor limiting the quality of transmitted signal in optical links, refers to the process of broadening the light pulses during the propagation along the fibers, due to material properties and imperfections [18, 100]. Dispersion can be broadly classified into two categories: intermodal dispersion (also called as intermodal delay) and intramodal dispersion (also known as chromatic dispersion) effects. The intermodal dispersion [101] is a distortion mechanism appearing only in multimode fibers. It arises from the different group velocity (i.e., the speed at which energy in a particular mode propagates through the fiber) of all propagation modes which induce the light to spread out temporally as it travels along the fiber. Chromatic dispersion [102, 103] is a phenomenon of pulse spreading over time that takes place within a single mode. It stems from the finite spectral emission of the optical source (defined as the band of wavelengths over which the optical source emits light): as the optical sources are not monochromatic, the input pulse includes several wavelength components, traveling at different speeds, causing the pulse to spread. Intermodal dispersion includes material dispersion and waveguide dispersions [104]. Material dispersion arises due to the wavelength dependence of the refractive index on the fiber core material. Waveguide dispersion occurs due to the change in propagation velocity for the different spectral components induced by the difference in core-cladding spatial power distribu-

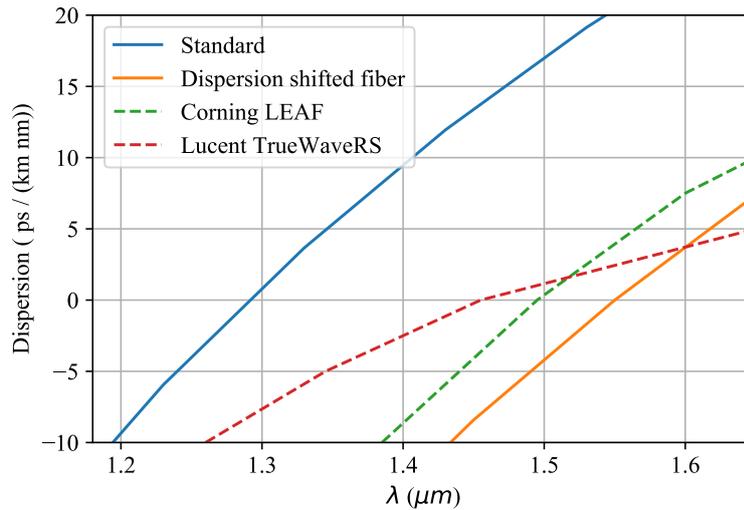


Figure 2.20: Typical wavelength dependence of the dispersion for standard, dispersion shifted, and dispersion flattened single mode fibers [18, 19].

tions, along with the speed variations of the various wavelengths [105, 18]. The resultant chromatic dispersion is expressed by [18]:

$$D(\lambda) = \frac{1}{L} \frac{d\tau}{d\lambda} \quad (2.10)$$

where  $\tau$  denotes the group delay, and  $L$  represents the length of fiber. The dispersion is typically expressed in ps/(km.nm).

The spreading  $\sigma$  of an optical pulse transmitted over an optical fiber of length  $L$  is given by [18]:

$$\sigma = D(\lambda)L\sigma_\lambda \quad (2.11)$$

where  $\sigma_\lambda$  denotes the the half-power spectral width of the optical source [18].

The dispersion behavior varies with wavelength and fiber type [18]. Figure 2.20 illustrates typical examples of the wavelength dependence of the dispersion for different types of fibers namely standard (conventional), dispersion shifted, and non-zero dispersion-shifted single mode fibers. As examples of non-zero dispersion-shifted fiber type, Figure. 2.20 depicts two different fibers produced by different manufacturers, including Corning LEAF and Lucent TrueWaveRs.

### 2.2.5 Nonlinear Effects

The nonlinear effects (i.e., intensity-dependent phenomena) [106, 107] in optical fibers arise from either the variation of the refractive index of the medium with optical intensity or the inelastic scattering mechanisms. As shown in Fig. 2.21, the nonlinear effects

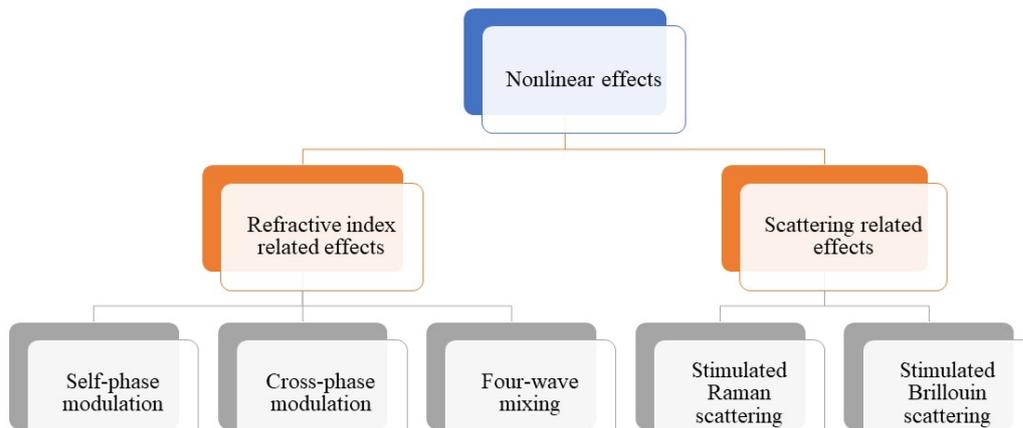


Figure 2.21: Block schematic showing the nonlinear effects in optical fibers [107].

can be broadly classified based on their characteristics into two types: nonlinear refractive index related effects referred as Kerr nonlinearities [108], and inelastic scattering effects [18]. The Kerr effects manifest in three different forms: self-phase modulation (SPM), cross-phase modulation (XPM), and four-wave mixing (FWM) processes [109, 110]. SPM [111] effect causes the phase alteration of a pulse by its own intensity and thereby inducing its spectral broadening which produces dispersion-like effects limiting transmission rates in some long-haul optical communication systems. Similar to the SPM effect, XPM [112] can alter and spread out the spectrum of a pulse induced by the intensities of all other pulses, possessing different wavelengths or polarizations, propagating in the fiber. FWM [113] is a third order nonlinear optical effect in which two or three photons at different frequencies interact to generate one or two additional photons with sum and difference frequencies during the propagation in the waveguide. The scattering effects include SRS and SBS mechanisms [18] briefly discussed in Subsection 2.2.3.2.

## 2.3 Optical Amplifiers

To compensate for the fiber losses and dispersion effects limiting the transmission distance of a fiber-optic communication system, optical amplifiers which amplify (boost) the optical signal directly without interconversion of photons to electrons, are widely adopted thanks to the lower complexity and cost advantages compared to traditional optoelectronic repeaters in which the optical signal is first converted into an electronic current and then regenerated using a transmitter [19, 16]. The optical amplifiers are considered to be lasers without optical cavity or with suppressed feedback from the cavity [114]: an optical amplifier can amplify the input optical signal, but it cannot

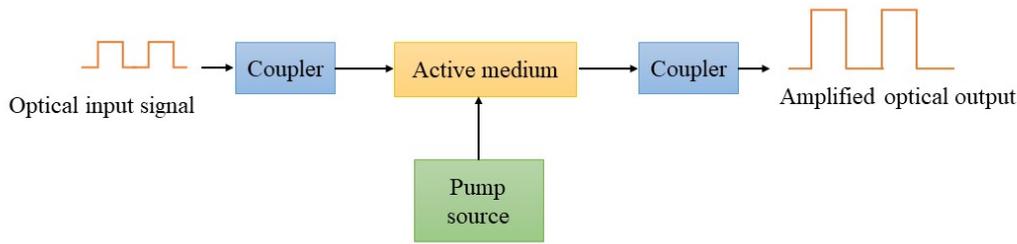


Figure 2.22: Basic operation of a generic optical amplifier [18].

produce a coherent optical output by itself [18]. The working principle of a generic optical amplifier is illustrated in Fig. 2.22. Here, the active medium is provided with energy from an external source called the pump and thus raising the electrons to higher energy levels to produce a population inversion. The incident signal photon triggers these excited electrons, which lose energy in form of photons and thus drop to lower levels through a stimulated-emission process. Those stimulated photons have the same wavelength as the incoming optical signal, thereby amplifying the optical signal [18].

In following subsections, the applications of optical amplifiers in optical communication systems as well as the different types of optical amplifiers are addressed.

### 2.3.1 General Applications of Optical Amplifiers

There are three common applications of optical amplifiers: power boosters (of transmitters), in-line amplifiers, and optical pre-amplifiers [17]. The power amplifiers are placed immediately after the optical transmitter to boost the transmitted power before launching into the fiber link [18] as illustrated in Fig. 2.23 a. The in-line amplifiers are set at intermediate points in the transmission link to compensate for transmission loss [18] as shown in Fig. 2.23 b. As shown in Fig. 2.23 c, the optical pre-amplifiers are located just before the optical receiver to amplify the optical signal before the photodetection process in order to overcome the signal-to-noise degradation caused by the thermal noise caused by the receiver electronics and thus improving the receiver sensitivity [18].

### 2.3.2 Optical Amplifier Parameters

**Amplifier Gain:** The amplifier gain  $G$  is considered as one of the most important parameters of an optical amplifier. It is defined as [18]:

$$G = \frac{P_{out}}{P_{in}} \quad (2.12)$$

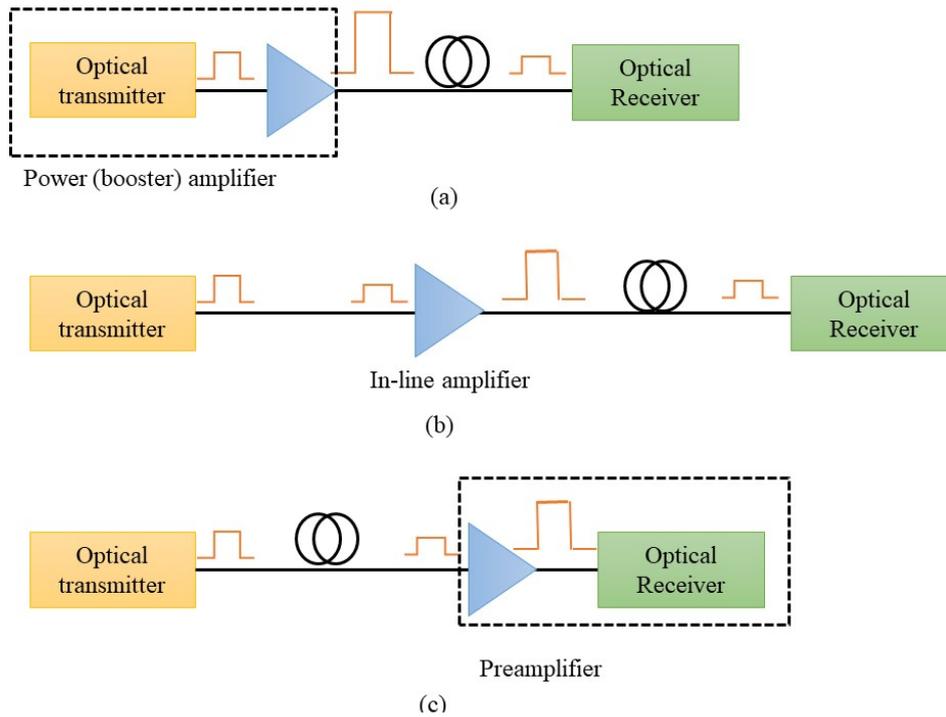


Figure 2.23: Applications of optical amplifiers: (a) booster of transmitted power, (b) in-line amplifier to increase transmission distance, and (c) preamplifier to enhance receiver sensitivity [18].

where  $P_{out}$  and  $P_{in}$  denote the input and the output powers respectively of the optical signal being amplified.

**Noise Figure:** The noise figure of an optical amplifier  $F_n$  is a measure quantifying the signal-to-noise ratio (SNR) degradation experienced by an optical signal after passing through the amplifier mainly due to the amplified spontaneous emission (ASE) noise generated during the amplification process.  $F_n$  is defined as the SNR at the input of an amplifier divided by the SNR at the amplifier output, and is expressed as [19]:

$$F_n = \frac{(SNR)_{in}}{(SNR)_{out}} = \frac{1 + 2 n_{sp} (G - 1)}{G} \quad (2.13)$$

where  $n_{sp}$  denotes the spontaneous-emission or population-inversion factor.

For large values of  $G$ ,  $F_n$  becomes equal to  $2 n_{sp}$ . For an ideal amplifier ( $n_{sp} = 1$ ), the SNR of the amplified signal is degraded by a factor of 2 (or 3 dB) [19]. Generally, for the most used amplifiers,  $F_n$  exceeds 3 dB and can be as large as 6-8 dB [19, 115].

**Gain Bandwidth:** The gain bandwidth (also called amplification bandwidth) represents the width of the optical frequency range in which significant gain is available from

an optical amplifier [116], and typically is 5 THz (100 channels with 50GHz) wide.

### 2.3.3 Optical Amplifier Types

Based upon the physical mechanism used to amplify the optical signal, the optical amplifiers can be categorized into three main types of optical amplifiers: semiconductor optical amplifiers (SOAs) [117, 118], Erbium-doped-fiber amplifiers (EDFAs) [119, 120], and Raman amplifiers [121]. An SOA, known as a laser amplifier, is an optical amplifier based on a semiconductor gain medium, whereby an electrical pump is used to achieve population inversion. SOAs can provide high internal gain (up to 35 dB) with relatively low power consumption [16]. They have proven to be particularly suitable for use with a single-mode fiber thanks to their single-mode waveguide structure [16]. They are of small size and thus can be integrated to produce subsystems. SOAs can be classified into two types: Fabry–Pérot amplifiers (FPAs) [122] and traveling-wave amplifiers (TWAs) [123, 124]. The primary difference between the two types is the facet reflectivity. FPAs have a reflectivity around 30%, whereas TWAs have a reflectivity of around 0.01% [125]. EDFAs are one of the most commonly used optical amplifiers for fiber loss compensation in long-haul optical communication systems [126]. An EDFA uses the Erbium-doped fiber as an optical gain medium, whereby an optical pump is adopted to produce population inversion. It operates in the C-band (1,530 to 1,565 nm) and L-band (1,560 to 1,605 nm) [126]. It can amplify multiple optical signals at different wavelength simultaneously, making it eminently suitable for WDM system applications [127]. EDFAs offer a plethora of advantages including high gain over wide spectral bandwidths ( $> 30$  dB with gain flatness  $< \pm 0.8$  dB and  $< \pm 0.5$  dB in C- and L-band, respectively), low noise figure (4.5 to 6 dB), direct and simultaneous amplification of wavelength division multiplexed signals, high pump power utilization ( $> 50\%$ ), easy deployment, etc [126, 127]. However, it is hard to integrate EDFAs with other components as their size is not small (typically ten meters long) [127]. The dynamic control of EDFAs is required as dropping channels can give rise to errors in surviving channels [127]. Raman amplifiers make use of SRS effect within the fiber [18, 128]. There are typically two types of Raman amplifiers: distributed Raman amplifier which uses the transmission fiber itself as gain medium into which a backward pump is injected, and lumped Raman amplifier which adopts the dispersion compensation fiber or highly nonlinear fiber as gain medium [23]. The low noise feature (3-5 dB) and wide gain bandwidth (up to 10 nm) make Raman amplifiers very attractive for long distance optical communication systems [129]. However, Raman amplifiers require high pump power and provide low gain (10 dB) [130].

## 2.4 Passive Optical Components

Passive optical components, which require no input power to function, are important devices in optical communication systems, actively deployed not only in FTTH (Fiber to the Home) networks but also in metro/core networks. In the following subsections, typical examples of optical passive components including fiber joints, fiber couplers, and optical attenuators are discussed.

### 2.4.1 Fiber Joints

#### *Joint Loss, Return Loss, and Reflectance*

A crucial aspect of the fiber joint is the optical loss encountered at the connection, which significantly depends on the alignment of the two jointed fibers. An attenuation (insertion loss) at the fiber joint can be induced due to Fresnel reflection phenomena causing a small proportion of the light to be reflected back into the transmitting fiber, due to the change in refractive index at the jointed interface (i.e. glass–air–glass) [16]. The loss (in decibels) due to Fresnel reflection at a fiber–fiber connection is expressed as follows [16]:

$$LOSS_{Fresnel} = -10 \log_{10} \left( 1 - \left( \frac{n_1 - n}{n_1 + n} \right)^2 \right) \quad (2.14)$$

where  $n_1$  denotes the refractive index of the fiber core, and  $n$  is the refractive index of the medium between the two jointed fibers.

Another source of optical loss at a fiber–fiber connection is arising from misalignment of the two jointed fibers. The misalignment can be induced due to several factors associated with deviations in the geometrical and optical parameters of the jointed optical fibers, comprising a mismatch in the core and/or cladding diameters, a mismatch in numerical apertures, and differing refractive index profiles [16]. Figure 2.24 illustrates the possible types of optical fiber misalignments that might occur when jointing two coupled compatible optical fibers: longitudinal misalignment (or endface separation), lateral misalignment, and angular misalignment. The aforementioned types of misalignments result in different optical losses at the joint depending on the type of the fiber, the core diameter, and the distribution of the power propagating between the different modes.

Optical return loss (ORL) [18] is defined as the ratio (in dB) of the optical power ( $P_{in}$ ) launched at a system interface and the optical power reflected back from into the input path due to by light reflections off the interface of the polished end surface of

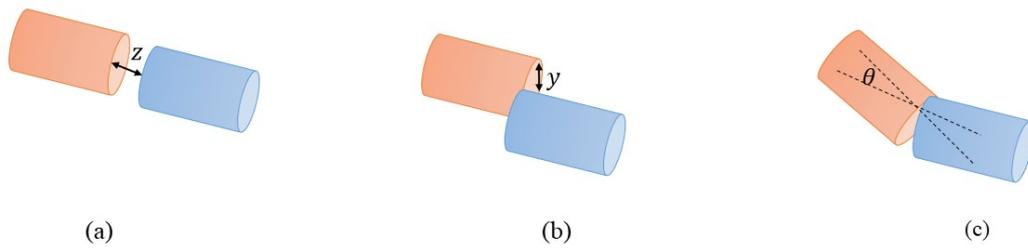


Figure 2.24: Types of optical fiber misalignments: (a) longitudinal misalignment, (b) lateral misalignment, (c) angular misalignment [16].

the connectors and air. It is expressed as follows [131]:

$$\text{ORL (dB)} = -10 \log_{10} \left( \frac{P_{in}}{P_{ref}} \right) \quad (2.15)$$

Reflectance (called also "back reflection" or Fresnel reflection) of a connection [132] is the amount of light that is reflected back up the fiber toward the source. It is used to describe back reflection at a connector pair. It is calculated as follows:

$$\text{Reflectance (dB)} = -10 \log_{10} \left( \frac{P_{ref}}{P_{in}} \right) \quad (2.16)$$

The reflectance is sometimes referred to as the inverse of ORL, having the opposite sign, e.g., -40 dB reflectance is 40 dB return loss. Please note that a high return loss or low reflectance indicates a good quality fiber joint and usually results in low insertion loss [132].

### **Joint Types**

Fiber joints can be classified into two major categories: splices (i.e., permanent, or semipermanent joints achieved by bonding or thermal fusion), and connectors (i.e., removable joints permitting easy, fast, and manual coupling and uncoupling of optical fibers) [16]. Depending upon the used splicing method, fiber splices may be categorized into two broad types: fusion and mechanical splicing. Fusion splicing is the process of permanently fusing or welding two optical fibers using an electric arc or heat [133]. Mechanical splicing is the act of precisely holding two optical fibers in alignment by utilizing different mechanical means, enabling the light to pass from one fiber to the other seamlessly [134]. It may be performed using various techniques including the utilization of tubes around the fiber ends (tube splices) or V-grooves into which the butted fibers are located (groove splices) [16]. Optical connectors are harder to perform than the fiber splices as they must ensure similar tolerance requirements to splices

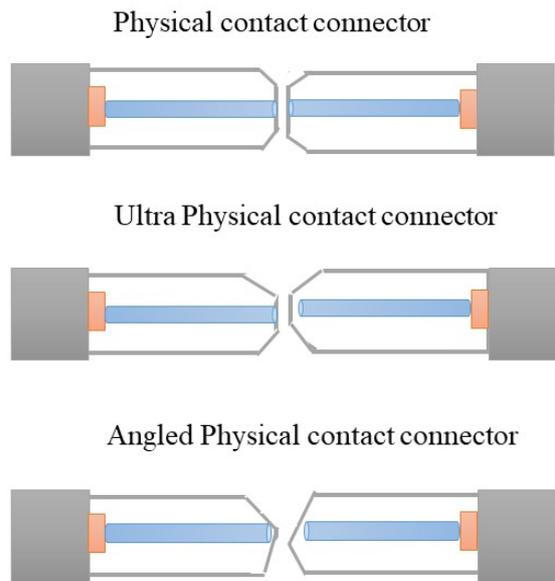


Figure 2.25: Types of optical fiber connectors [135].

in order to enable an efficient connect/disconnect capability for joining optical fibers. They are used as terminations of optical fibers. Fiber optic connectors can be classified into three main types based on the shape of the end faces of connector ferrules (i.e. the housing for the exposed end of a fiber): physical contact (PC) connector, ultra-physical contact (UPC) connector, and angled physical contact (APC) connector [135], as illustrated in Fig. 2.25. PC connectors are polished with a slightly cylindrical cone head designed to reduce the air gap, and thus to increase the optical return loss. The typical reflectance values achieved by the PC connectors range from -30 dB to -40 dB, which are much better than the ones yielded by the original flat connectors (-14 dB or roughly 4%) [135]. An UPC connector is an improvement of PC connector with an extended polishing for a better surface finish, and with a better return loss (typically 50 dB) than the PC connector [135]. The APC connector is designed with the aim of reducing the back reflection more efficiently. Its ferrule end face radius is polished at an angle of  $8^\circ$  [135]. The typical optical return loss of APC connectors is 60 dB or higher, better than the other connector types [135].

### 2.4.2 Fiber Couplers

An optical fiber coupler [136, 137] is a branching device distributing the light from a main fiber into one or more branch fibers [16]. Fiber couplers are widely adopted in optical fiber information distribution systems such as local area networks (LANs),

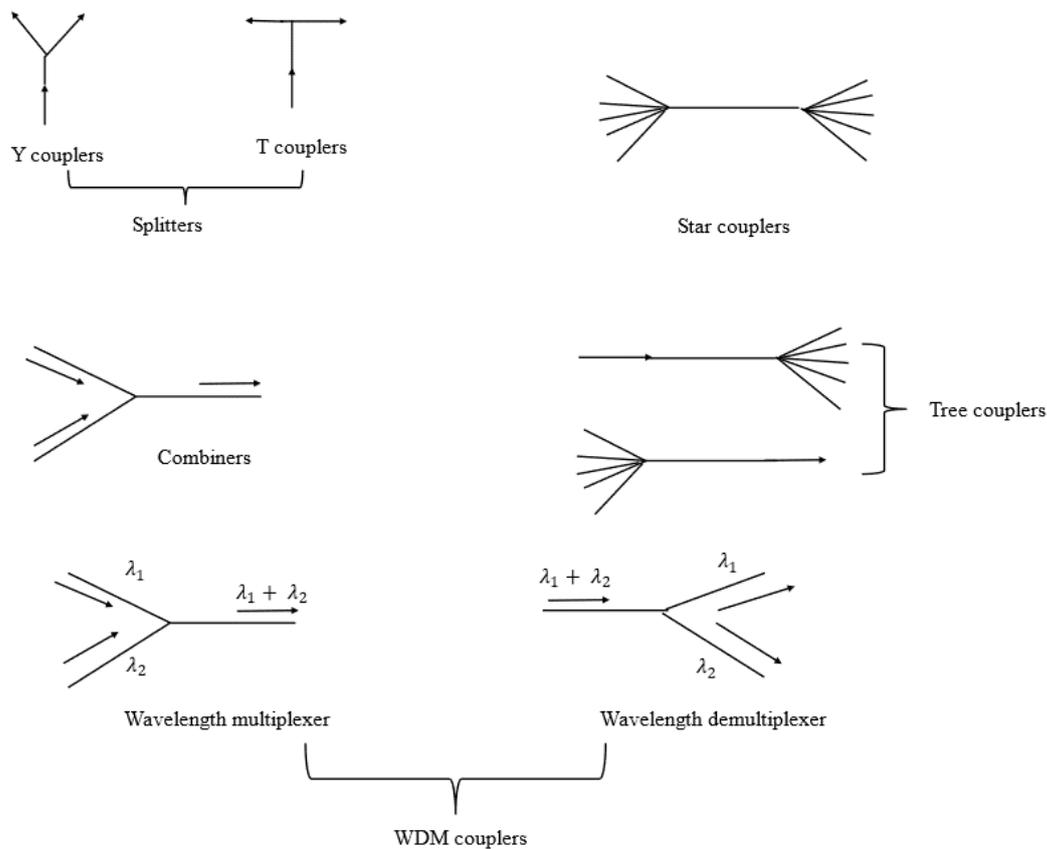


Figure 2.26: Different types of optical fiber couplers [16].

WDM systems, data buses, etc. As illustrated in Fig. 2.26, optical fiber couplers can be classified into different types including splitters, combiners, star couplers, tree couplers and WDM couplers, mostly based on the number of input and output ports. The splitters split optical signals into multiple paths. They can be further categorized into T couplers and Y couplers (also called tap couplers), whereby the former having equal power distribution and latter an uneven power distribution [138]. The combiners combine two or more optical signal inputs from different paths into one output. The star couplers have multiple inputs and outputs, and are used to distribute the power from all of the inputs to all outputs [139]. The tree couplers have either multiple inputs with a single output or a single input with multiple outputs [139]. WDM couplers distribute optical signals according to their wavelengths by permitting them to be transmitted in parallel on a single fiber. They either combine the different peak wavelength optical signals onto the fiber (i.e. multiplex) or separate the different wavelength optical signals output from the fiber (i.e. demultiplex) [16].

### 2.4.3 Optical Attenuator

An optical attenuator [45] is a fiber-coupled device used to reduce the power level of an optical signal in a controlled way. Fiber-optic attenuators are widely used in optical communication systems for different applications including, testing and evaluation of components or modules, laboratory experiments, equalizing the power of different channels in the WDM system, etc. Generally, optical attenuators can be broadly classified as fixed optical attenuators (FOAs) and variable optical attenuators (VOAs) [140]. FOAs are designed to have an unvarying level of attenuation in optical fiber, expressed in dB, typically between 1 dB and 30 dB, such as 1 dB, or 5 dB, etc [141, 142]. Whereas VOAs provide an adjustable degree of attenuation. A VOA generally adopts a variable neutral density filter. It offers multiple benefits including stability, wavelength and mode insensitivity features, large dynamic range, etc. The VOAs can be further categorized into two types: stepwise variable attenuator and continuously variable attenuator [142]. A stepwise variable attenuator can vary the attenuation of the optical signal in known steps such as 0.5 dB or 1 dB [142]. Continuously variable optical attenuators provide a precise level of attenuation with flexible adjustment. Optical attenuators are available for both fiber types: single and multimode fibers. There are multiple methods used to fabricate attenuators comprising absorbing materials, partially reflective mirrors, taped fibers with longitudinal gap, etc [45].

## 2.5 Optical Receiver

The role of an optical receiver [143, 144] is to convert the optical signal received at the output end of the optical fiber into the original electrical signal [145]. As shown in Fig. 2.27, an optical receiver is composed of a coupler, a photodetector, and a demodulator. The coupler focuses the received optical signal onto the photodetector [146]. The photodetector is the heart of the optical receiver. It is used to convert the light into electricity by adopting the photoelectric effect. A photodetector in an optical communication system is a semiconductor photodiode. The design of the demodulator depends on the modulation scheme adopted by the fiber optic communication system [78].

In Subsection 2.5.1, the working principle of a photodiode is explained. Subsection 2.5.2 is devoted to a discussion about the characteristics of a photodiode.

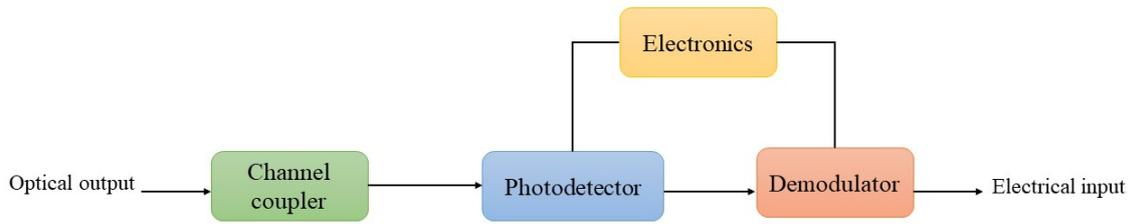


Figure 2.27: Components of an optical receiver. [19].

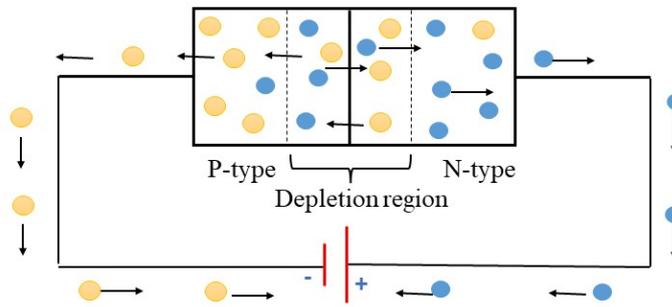


Figure 2.28: Schematic diagram of a photodiode [150, 45].

### 2.5.1 Working Principle of Photodiode

A photodiode (also called a photodetector, a light detector, and photo-sensor) [147, 148] is a semiconductor device with a P-N junction converting photons (light) into electrical current generated by photon absorption mechanism [149]. It is electrically reverse-biased. As illustrated in Fig. 2.28, the photons with the proper energy (wavelength) are annihilated to create electron-hole pairs in the depletion region by raising an electron from the valence band to the conduction band, leaving a hole behind. The bias voltage induces these current carriers to move quickly away from the junction region. Thus, the holes move to the anode and the electrons move to the cathode to generate a photocurrent proportional to the light.

### 2.5.2 Characteristics of Photodiode

**Quantum Efficiency** The quantum efficiency ( $\eta$ ) [18, 16] is defined as the probability that a photon incident on the photodiode generates a photocarrier contributing to the photocurrent. It is the fraction of the number of electron-hole pairs generated to the number of incident photons, and it is calculated as follows [39]:

$$\eta = (1 - R) \epsilon (1 - \exp(-\alpha W)) \quad (2.17)$$

where  $R$  denotes the power reflection coefficient at the air-semiconductor surface,  $\epsilon$  represents the fraction of photocarriers contributing to the measured photocurrent,  $\alpha$  is the absorption coefficient, and  $W$  refers to the thickness of the depletion or active region of the photodiode.

**Responsivity** The responsivity or photoresponse (sometimes also called sensitivity) ( $\rho$ ) [39] is a measure of the effectiveness of a photodiode in converting optical power into an electrical current or voltage. It depends on the wavelength of the light, the type of active material in the detector, and the structure and operating conditions of the photodiode [39]. It is calculated as follows [39]:

$$\rho (A/W) = \eta \frac{\lambda_0 (\mu m)}{1.24} \quad (2.18)$$

The responsivity is proportional to the quantum efficiency and the free-space wavelength  $\lambda_0$ .

**Dark Current** Dark current is the current generated in the photodiode in the absence of light or an incident optical power [151]. The dark current arises from the generation of an electron-hole pair due to thermal radiation or stray light.

**Speed or Response Time** The speed of response or the bandwidth (i.e. the frequency at which the output signal has dropped by 3 dB (50%) below the power at a low frequency) of a photodetector depend on the following three factors [152]:

- The transit time ( $\tau_t$ ) of the photogenerated carriers through the depletion region is calculated as [39]

$$\tau_t = \frac{W}{v}, \quad (2.19)$$

where  $v$  denotes the speed of the carrier.

- The slower diffusion of carriers occurring outside the depletion region
- The RC time constant ( $\tau_{RC}$ ), which depends on the diode's capacitance  $C$  and resistance  $R$  and it given by [39]:

$$\tau_{RC} = R C, \quad (2.20)$$

The total response time ( $\tau_{tot}$ ) can be calculated as the root mean square of the RC time constant and the transit time [39].

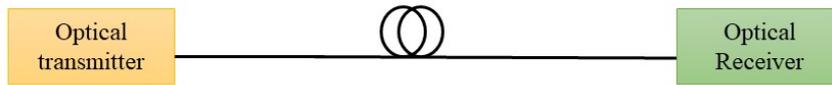


Figure 2.29: Block diagram of a simple point-to-point link.

$$\tau_{tot} = \sqrt{(\tau_t^2 + \tau_{RC}^2)} \quad (2.21)$$

## 2.6 System Architectures

The architectures of the fiber-optic communication systems can be broadly classified as point-to-point links, point-to-multipoint links, and meshed networks. In the following subsections, each aforementioned system architecture is described.

### 2.6.1 Point-to-Point Links

The point-to-point link or connection (also known as P2P) is the simplest setup in an optical fiber communication system, connecting two nodes, which communicate back and forth, directly together with a common link [153] as illustrated in Fig. 2.29. It establishes a direct connection for transmitting data with a dedicated bandwidth. Examples of P2P communication include connecting computers and terminals within the same building or between two buildings with a relatively short transmission distance ( $< 10$  km) [153]. A P2P communication scheme provides multiple benefits compared to other optical communication topologies such as highest bandwidth as the entire bandwidth of the link is shared by only two nodes, low latency, simplicity, ease of maintenance and handling, etc. However, such topology is used only for small areas and if one of the nodes stops working or the link is broken, the entire network becomes dead.

### 2.6.2 Point-to-Multipoint Links

The point-to-multipoint link (also known as P2MP) or passive optical networks (PONs) [154, 155] refers to a one-to-many connection from a single location to multiple locations by enabling multiple nodes or devices to share a single link. A PON system is based on using multiple wavelengths and bidirectional transmission over a single fiber [18]. PON has become a widely used broadband fiber access network solution in the world today thanks to a plethora of benefits offered such as wide bandwidth, cost effectiveness, scalability, flexibility, etc. PONs are renowned as the predominant technology

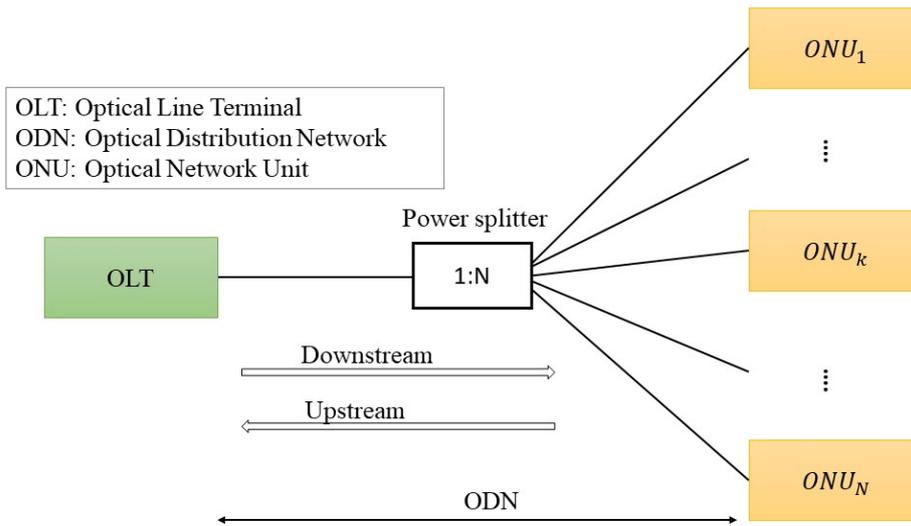


Figure 2.30: Architecture of a standard passive optical network [159].

for fiber-to-the-home (FTTH) deployment, providing various communication and multimedia services including high quality triple play service capabilities for data, voice and video, and high-speed internet access, in a cost-effective manner due to the passive components deployed in the transmission line and the optimal fiber infrastructure [156]. As shown in Fig. 2.30, a PON system is composed of an optical line terminal (OLT) installed in the central control station, optical distribution network (ODN) containing passive optical devices such as splitters, and several optical network units (ONUs) installed at the user's premises providing services such as data, IPTV (i.e. interactive network television), and voice (using IAD, i.e. Integrated Access Device), truly realizing "triple-play" applications [157]. The OLT sends data to an ONU and allocates bandwidth between the ONUs. The ODN provides an optical transmission channel between the OLT and the ONU. PONs today are typically implemented using two prominent technologies: Ethernet (EPON) and Gigabit PON (GPON) [18]. PON systems lower the operation-and-maintenance expenses (OPEX) due to the fact that no power requirements or active electronic components, which are more prone to failures in the PON outside plant, are required [158].

### 2.6.3 Meshed Networks

An optical mesh network is a type of optical telecommunication system employing optical fiber, synchronous digital hierarchy, and a ring architecture. Optical mesh networks are widely deployed in metropolitan, regional, national, or international areas thanks to their efficiency and reliability features [160]. They are faster and less error prone than other network architectures, and they support backup and recovery plans

for established networks in case of any disaster, damage, or failure [160].

## 2.7 System Monitoring

Optical communication networks have to be monitored continuously to ensure a high degree of reliability and availability. Various measurement or monitoring techniques have been developed to test the operational behavior of the communication system elements including fibers and optoelectronic devices, as well as characterize them in terms of fundamental parameters such as optical power, polarization, and spectral content, in order to ensure that the network functions properly [18]. Operational fixed parameters of many optical components can be extracted from vendor data sheets. Such fixed parameters include fiber parameters such as core and cladding diameters, refractive-index profile, mode-field diameter, and cut-off wavelength [18]. There is no need to remeasure them. However, variable parameters of communication system elements may change with operational conditions and thus need to be measured when the link is deployed [18]. Particularly, it is paramount to perform accurate measurements of the optical fiber, specifically measuring the dispersion and the attenuation parameters. A commonly used technique for measuring the total fiber attenuation per unit length is the cut-back (measuring the optical power transmitted through a long and a short length of the same fiber using identical input couplings) or insertion-loss method [18]. When an optical fiber link is deployed, operational parameters of interest including bit-error rate (BER), timing jitter, and SNR have to be measured as well in order to be used for maintenance purpose and monitoring functions to check the status of the devices or to localize faults in optical fibers [18]. Figure 2.31 shows some of the relevant test parameters of interest and at what points in an optical link they are of importance [18]. The monitoring techniques can be broadly classified as internal or external monitoring methods. The internal monitoring function can be integrated into the device and remotely controlled or configured. It enables to collect periodically the operational performance parameters of a deployed device in an optical network, such as measured output power, temperature, current, or gain etc. The performance monitoring instrument can be located externally (outside the optical communication devices or elements). The external optical-test instruments include optical spectrum analyzers (measuring optical power as a function of wavelength), attenuators, a visual fault locator (a handheld pen-sized instrument that uses a visible laser light source to locate events such as fiber breaks [18]), eye diagram (a traditional technique for quickly and intuitively assessing the quality of a received signal [18]), and optical time-domain reflectometer (OTDR), etc.

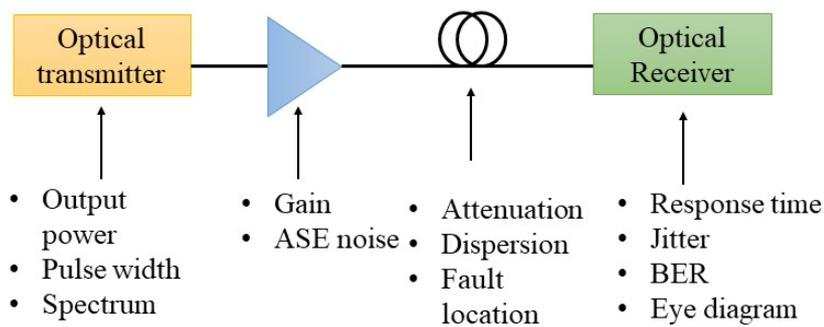


Figure 2.31: Some performance measurement parameters of interest for an optical fiber link [18].

Given that Chapter 5 discusses the application of ML techniques for fault management, it is important to provide an overview of how an OTDR works, its characteristics, and what types of faults can be seen in an OTDR trace data.

**Optical Time-Domain Reflectometer** An optical time-domain reflectometer (OTDR) [161], a technique based on Rayleigh backscattering, is a versatile portable instrument widely applied for measuring the characteristics of an installed optical fiber link such as fiber attenuation, length, and connector and splice losses, and for detecting and localizing fiber fault or anomalies within a link. OTDR operates like an optical radar. As shown in Fig. 2.32, an OTDR typically consists of a light source (laser), a receiver (photodetector), and a data acquisition and processing module. An OTDR operates by sending a series of optical pulses into one of the ends of the fiber under test using a directional coupler. Part of these pulses are reflected or scattered back towards the source as a result of Rayleigh scattering. The amplitude of this reflected signal is recorded as a function of propagation time of the light pulse, which can be converted into the position on the optical fiber. Thus, a recorded OTDR trace (or waveform) illustrating the positions of faults or events along the fiber, is generated, and used for event analysis. Figure 2.33 shows a typical OTDR trace whereby the vertical axis denotes the measured returning (back-reflected) signal in decibels, and the horizontal axis represents the distance between the OTDR device and the measurement point in the optical fiber. As illustrated in Fig. 2.33, a typical waveform shows firstly a large initial pulse resulting from Fresnel reflection at the input end of the fiber, then a long decaying tail arising from the Rayleigh scattering effect, the events induced due to connectors or splices or physical changes in the fiber, and lastly a large spike caused by the Fresnel reflection at the end of the fiber. The events shown in an OTDR trace can be classified

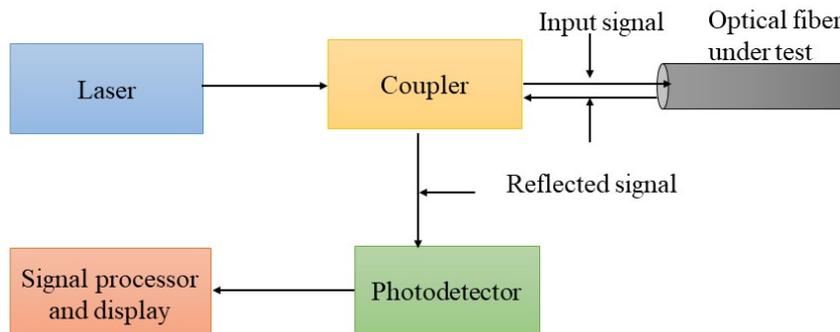


Figure 2.32: Operational principle of an optical time-domain reflectometer (OTDR) [18].

as reflective or non-reflective events. The reflective events are Fresnel peaks arising from sudden changes in density of the material [162]. Examples of reflective events include PC connectors and mechanical splices [162]. The non-reflective events are induced due to MFD (modal field diameter) variations as a result of geometric changes or differences in the glass fiber [162]. Examples of non-reflective events are APC connectors or fusion splices. The accuracy of the event detection and measurement depends on the SNR achieved by an OTDR at that point, and which is defined as the ratio between the back-reflected signal and the noise level. The SNR depends on multiple factors including the OTDR pulse width, how often the OTDR samples the signal, and the distance to the measurement point. An important performance parameter is the dynamic range, which is defined as the difference between the initial backscatter power level at the front connector and the noise level peak at the far end of the fiber [18]. Dynamic range indicates the maximum fiber loss that can be measured and denotes the time required to measure such fiber loss. To increase the OTDR's dynamic range, the backscatter level has to be increased and/or the noise level must be reduced. To increment the backscatter level, the OTDR's pulse width or laser power has to be increased [161]. Reducing the receiver's bandwidth, increasing the amount of the averaging of OTDR waveforms, or filtering OTDR traces could help to decrease the noise level [161]. Larger pulse widths increase dynamic range but reduce the OTDR's ability to measure closely spaced events, limiting the OTDR's two event resolution. Decreasing the receiver's bandwidth may reduce noise, but it makes the OTDR instrument slow to respond to changes in signal, increasing the minimum distance between two events. Although filtering helps to reduce noise, it introduces distance-measurement errors, which may impair event detection and localization accuracy. Excessive laser power could cause nonlinear effects in the fiber. The averaging method is commonly used to reduce noise and thus provides dynamic range gain. However, averaging many OTDR

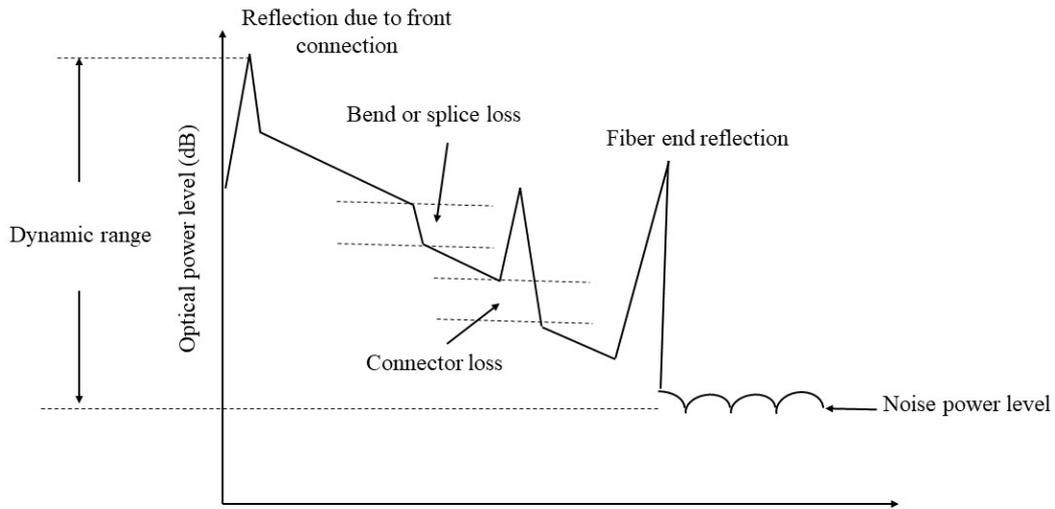


Figure 2.33: Representative of a typical OTDR trace showing the events along an optical fiber link [18].

traces to remove noise takes time, especially for long fibers. Another important OTDR performance parameter to mention is the measurement range, which indicates how far an OTDR can identify events in an optical fiber link. The maximum measurement range  $R_{max}$  in km can be calculated as follows [18]:

$$R_{max} = \frac{D_{OTDR}}{\alpha} \quad (2.22)$$

where  $D_{OTDR}$  denotes the dynamic range expressed in dB, and  $\alpha$  represents the attenuation given in dB/km.

## 3 Predictive Maintenance

Predictive maintenance has emerged as an efficient and advanced maintenance strategy for the manufacturing industry to increase operational efficiency and productivity, and to reduce maintenance costs. The global predictive maintenance market is expected to reach \$13 billion by 2026 [163, 164].

This chapter focuses on the predictive maintenance strategy, its benefits, its key enabling technologies including ML, and the challenges of adopting such a maintenance strategy in industries.

### 3.1 Maintenance Strategies

The maintenance strategies can be usually classified into four different categories: breakdown or run to failure maintenance, preventive or time-based maintenance, predictive or condition-based maintenance, and proactive or prevention maintenance [165]. Each maintenance philosophy has its own benefits and shortcomings to be discussed in the following subsections.

#### 3.1.1 Run-to-Failure Maintenance

The run-to-failure maintenance (also called failure-driven maintenance) is a reactive (corrective) maintenance management technique that deliberately allows the machinery to break down and only repairs or replaces damaged components when the equipment comes to a full stop [165]. It is based on the principle “fix it when it fails”. This approach works well for non-critical components, whose breakdowns do not jeopardize business operations or for assets with a low risk of failure [166]. Such a maintenance strategy requires minimal planning (no maintenance schedule is planned before the failure of the equipment), allows the extended use of equipment (an asset is replaced only after its breakdown), and it is simple and easy to implement with minimal resources [167]. However, it is considered as the most inefficient and expensive maintenance strategy [165]. As most of equipment’ failures are highly unpredicted, it is challenging to anticipate when labor manpower and spare parts will be required for repairs or maintenance activities and to efficiently plan for resources to react immediately. The major

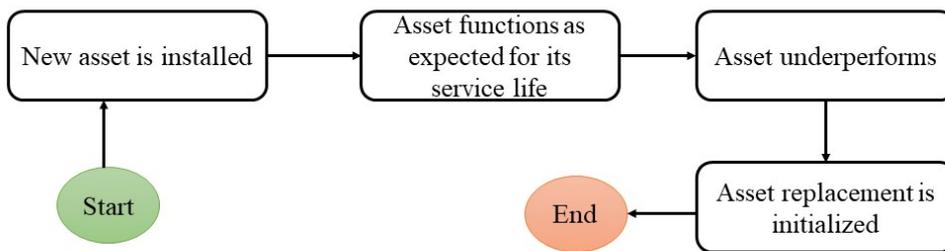


Figure 3.1: Workflow of run-to-failure maintenance strategy [167].

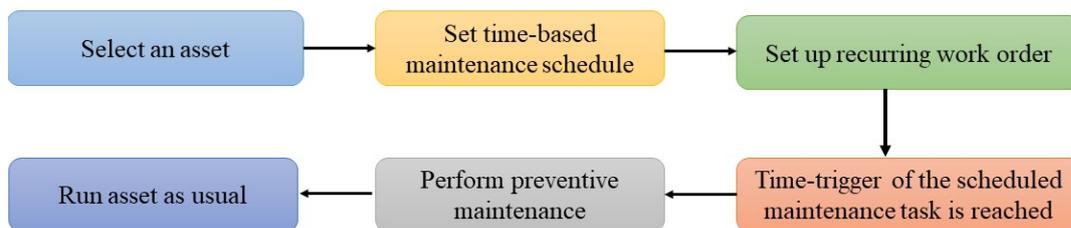


Figure 3.2: Workflow of preventive maintenance strategy [168].

expenses associated with this type of maintenance philosophy include high spare parts inventory costs, high breakdown costs, high labor costs associated with performing the maintenance, and low production availability [165].

### 3.1.2 Preventive Maintenance

Preventive maintenance is a time-driven approach whereby the maintenance activities are scheduled at predetermined time intervals based on elapsed time or hours of operation [165]. Figure 3.2 illustrates the workflow of such maintenance philosophy.

In preventive maintenance management, the repairs or replacements of assets are planned or adjusted based on MTTF statistic or a bathtub curve [169, 170] representing the failure rate of an asset over time. Figure 3.3 shows a typical bathtub curve, which can be divided in three regions: early failure (or break-in) or “infant mortality” stage characterized by a decreasing failure rate after removing the early failures due to manufacturing or installation problems, a normal life regime known as the intrinsic failure period characterized by a constant failure rate, and a wear-out stage characterized by an increasing failure rate after a long period of use due to the degradation of the equipment [171]. The main drawback of adopting the preventive maintenance approach is that scheduled maintenance can result in carrying out maintenance tasks too early leading to unnecessary repairs or too late, resulting in high costs due to un-

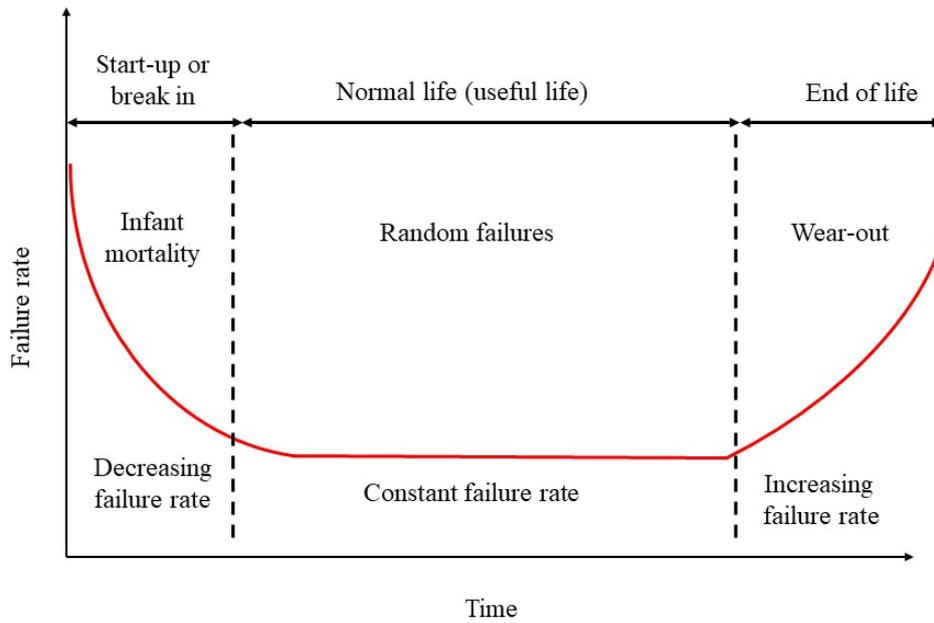


Figure 3.3: Typical bathtub curve [170, 172].

expected catastrophic failures [165]. It is possible that some components are rebuilt without showing any sign of functional failure when there is still some residual life left in them. Hence, reducing the production and wasting the labor and material used to perform the unnecessary repairs. Sometimes, machines may fail before the scheduled maintenance activities and thereby they are repaired in such cases using run-to-time methods leading to high maintenance costs [165]. In many cases, the performance of assets may diminish due to incorrect repair or rebuild methods.

### 3.1.3 Predictive Maintenance

Predictive maintenance is a philosophy or attitude consisting of scheduling maintenance activities only when a functional failure is detected [173]. It uses data monitoring and analysis tools and methods to monitor the performance or the condition of an equipment during operation, and to detect deterioration signs or anomalies, and thereby fixing them before they result in failure and thus preventing catastrophic breakdowns [165]. The predictive maintenance strategy is considered as a condition-driven preventive maintenance approach which instead of relying on MTTF statistics to schedule maintenance activities, makes full use of the monitoring data of the operating condition of an equipment or system efficiency to estimate the actual MTTF and thus, to plan maintenance work on an "as-needed" basis [165]. Predictive maintenance helps to reduce the maintenance costs, to increase the production capacity, to improve the productivity and profitability of manufacturing or production plant, and to minimize

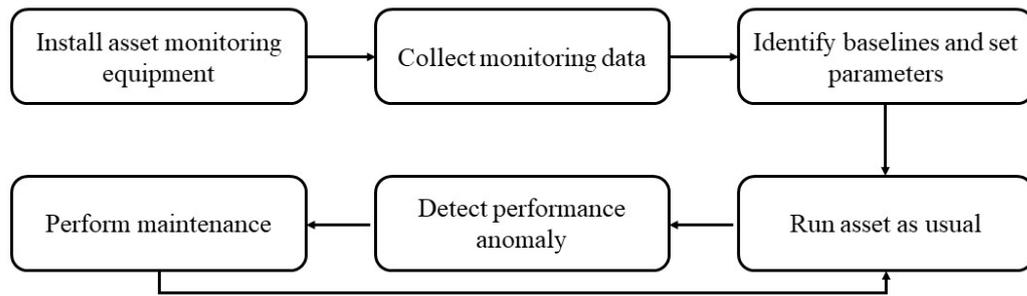


Figure 3.4: Workflow of predictive maintenance strategy [175]

the number of unscheduled outages [174]. However, it might be costly to implement the required smart technologies including sensors or monitoring equipment to collect the operational monitoring data, and to provide training to personnel (or hire skilled personnel).

### 3.1.4 Proactive Maintenance

Proactive maintenance [173, 165] is a maintenance strategy that lays emphasis on identifying and fixing the root causes of equipment failures and thus preventing breakdowns caused by underlying equipment conditions [176]. Proactive measures, after analyzing failures, are carried out in order to avoid recurrence of such failures. The proactive maintenance approach uses all of the predictive/preventive maintenance methods discussed above along with root cause failure analysis detecting and pinpointing the problems leading to breakdowns. It is to be noted that redesign or modification of equipment might be required to prevent failures from occurring repeatedly. The maintenance activities are scheduled in an orderly fashion like in the predictive maintenance strategy. The proactive maintenance approach leads to a substantial increase in production capacity, and helps to avoid downtime and to increase equipment reliability and availability [165]. However, one disadvantage of adopting such maintenance strategy is the requirement for procurement of specialized equipment and hiring extremely knowledgeable personnel to perform the maintenance tasks [173].

## 3.2 Evolution of Maintenance Strategies

Maintenance strategies have evolved in the industry over the years as illustrated in Fig. 3.6. A run-to-failure maintenance approach was adopted in the early days of production technology and was a reactive method whereby equipment repairs or rebuilds were carried out only after a functional failure occurred [173]. Given that breakdown

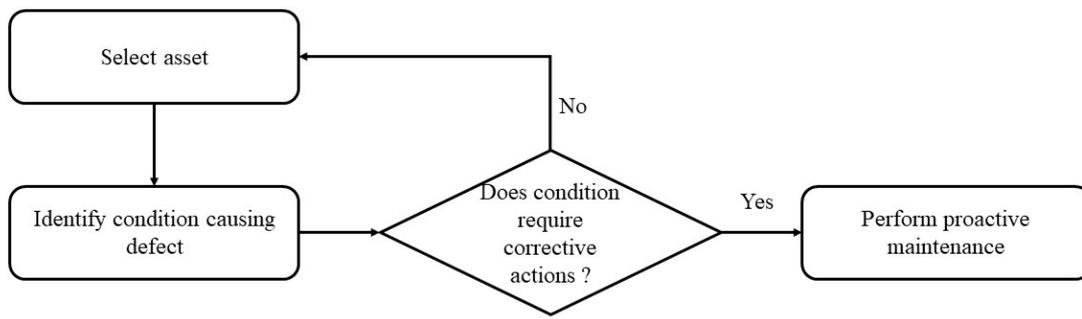


Figure 3.5: Workflow of proactive maintenance strategy [177]

maintenance was inefficient and led to high maintenance costs, a time-based or preventive maintenance approach was developed to improve the reliability and reduce the costs [173]. Equipment maintenance activities were performed regularly after a certain number of running hours, even if there was no evidence of a functional failure [173]. The disadvantage of this strategy was that assets were being replaced even when there was still some functional lifetime left in them. This strategy unfortunately could not help to reduce maintenance costs. Therefore, there was a need for scheduling the maintenance or overhaul of equipment based on the condition of the equipment. This led to the evolution of the predictive maintenance strategy [173]. For such an approach, the equipment is continuously monitored, and maintenance work is planned, only if a defect or anomaly are detected. Today, adopting predictive maintenance in industry has proven tangible benefits including reducing maintenance costs, minimizing downtime, and improving productivity. After early detecting equipment problems or failures, there was a tendency to perform the root cause analysis in order to eliminate the occurrence of such problems, which led to an evolved kind of maintenance, the proactive maintenance.

### 3.3 Benefits

Adopting a predictive maintenance program offers a plethora of benefits that can be enumerated in the following way [165]:

- Reduce maintenance costs: The implementation of a predictive maintenance program results in an average reduction in the costs associated with maintenance operations (e.g., costs of labor and overhead of the maintenance department, the materials cost of repair parts, tools, etc.) by more than 50%.
- Improve repair time: Leveraging predictive maintenance strategy helps to reduce the actual time required to repair or rebuild equipment by an average of 60% since

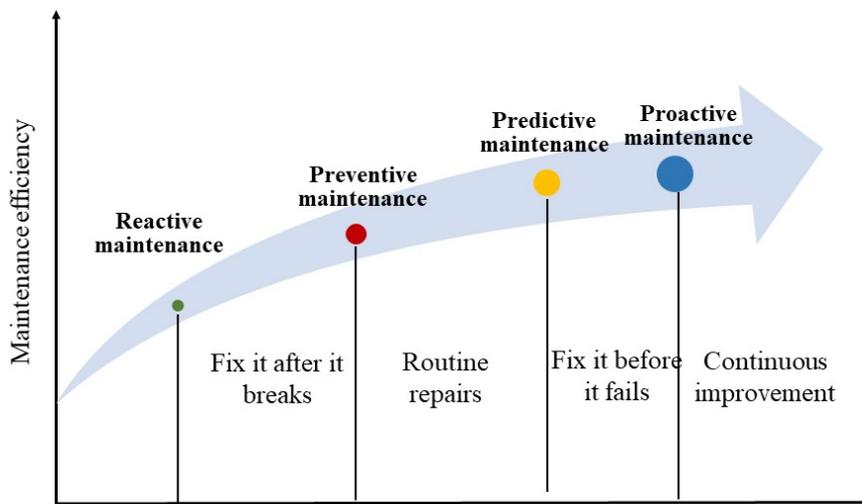


Figure 3.6: Evolution of maintenance approaches [178].

the early detection of equipment problems makes the maintenance activities carry out faster and smoother.

- Increase equipment lifespan: The early detection of equipment defects and elimination of sudden or catastrophic failures extends the operational useful lifetime of an asset by an average of 30%.
- Minimize spare parts inventories: Prediction of equipment failures reduces spare parts inventories by more than 30%.
- Lessen unexpected equipment failure: The continuous monitoring of the actual equipment condition helps to early detect impending asset problems that could bring the equipment or system to a grinding halt and thus implement the required measures to fix them before their occurrence and thereby preventing catastrophic or unexpected failures. Implementing a predictive maintenance approach reduces the number of catastrophic, unexpected equipment failures by an average of 55% [165, 174].

### 3.4 Enabling Technologies

Several emerging technologies including big data (data warehouses), cloud computing, artificial intelligence (AI) or specifically machine learning (ML), edge computing, and the internet of things (IoT) play a pivotal role in enabling predictive maintenance and accelerating its evolution and deployment in industry [179]. Figure 3.7 illustrates the involvement or the interaction between the different aforementioned technologies or

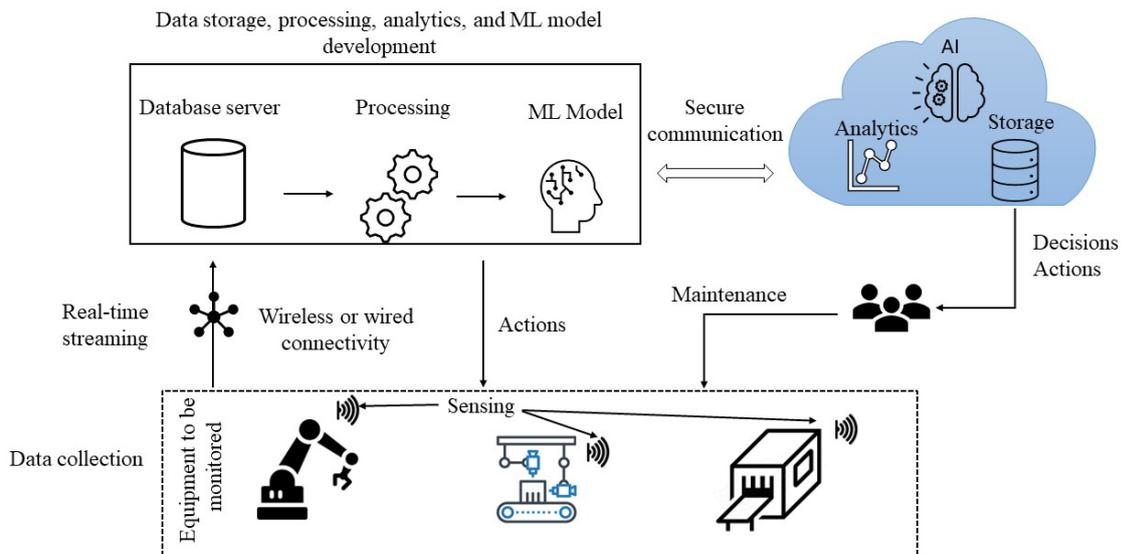


Figure 3.7: General architecture of a predictive maintenance system [180].

mega-trends to build a data-driven intelligent predictive maintenance system. Machines deployed in a manufacturing site are continuously monitored using health monitoring sensors to withstand a variety of operating conditions such as temperature, pressure, vibration, and so on. Monitoring data is transmitted via wired or wireless connectivity to an edge or cloud server. The data is then preprocessed and stored on a centralized database server. Afterwards, the data is analyzed using ML techniques to identify any anomalies. If an anomaly is discovered, a maintenance action is taken.

## 3.5 Machine Learning

### 3.5.1 What is machine learning?

Machine learning (ML, known as predictive analytics or statistical learning) [181] is a subfield of artificial intelligence, combining fundamental concepts in computer science with ideas from statistics, probability, and optimization, which provides computers with the ability to learn without being explicitly programmed [182]. It enables systems to learn from experience (i.e., past information available to the learner, typically an electronic data collected and preprocessed for analysis) by extracting knowledge from the input data (i.e., “experience”), in order to improve performance or to make accurate predictions [183]. The quality and the size of the data impacts the performance achieved by the ML algorithms trained with such data. The function of an ML system can be descriptive (extracting insights from the data to explain what happened), or predictive (predicting what will happen given the input data) or prescriptive (recom-

mending a course of actions for a given situation) [184]. ML methods can be used to tackle or solve complex tasks routinely performed by humans (e.g., driving, speech recognition, image understanding, etc.) or tasks that go beyond human capabilities requiring the analysis of very large and complex data sets (e.g., astronomical data, genomic data, etc.) including weather prediction, search engines, and electronic commerce [185]. ML admits a very broad set of practical applications including computer vision applications (e.g., object recognition, face detection, content-based image retrieval), natural language processing, speech processing applications including speech recognition, speech synthesis, speaker identification, and language modeling, and many other problems such as network intrusion [183].

### 3.5.2 Standard Learning Tasks

Some standard learning tasks [183] that are widely investigated are discussed in the following:

- *Classification*: this is the task of predicting a class or category (i.e., discrete output variable) for a given observation or item. For example, a bank transaction can be classified as belonging to one of two classes: “normal“ and “fraudulent“. The number of categories or classes assigned in such tasks can be two (a two-class or binary classification problem) or more (a multi-class classification problem) [186].
- *Regression*: This is the task of predicting a continuous outcome (real value such as an integer or floating-point value) for each observation or item. Examples of regression tasks include predicting house prices, predicting future stock prices, predicting sales of a product, etc. The skill of a regression predictive model is estimated by reporting an error in made predictions (the difference between the true and predicted values) such as root mean squared error [187].
- *Clustering*: this is the task of dividing a set of items or data into homogeneous subsets (clusters). The goal of cluster analysis is to find groups or subsets such that the observations within each cluster are quite similar to each other. Cluster analysis is widely used in different applications including retail marketing (e.g., retail customer segmentation), streaming services (identify viewers who have similar behavior), or health insurance (to identify “clusters” of consumers that use their health insurance in specific ways) [188].
- *Ranking*: this is the task of learning to order items according to some criteria.

### 3.5.3 Learning Stages

A list of definitions and terminology commonly used in ML is briefly explained in the following [183]:

- *Features*: the set of attributes representing the input of an ML model.
- *Labels*: the output of an ML model. For classification tasks, the labels are the classes or categories assigned. Whereas for regression problems, the labels are real values.
- *Training sample*: Items or instances of data used to train an ML algorithm.
- *Validation sample*: Instances of data used to tune or select appropriate values for the ML algorithm's free parameters (hyperparameters)
- *Test sample*: Instances of data used to evaluate the performance of a ML model after training it.
- *Loss function*: A function that measures the difference, or loss, between a predicted label and a true label.

To solve a learning task using the ML method, the learning workflow, illustrated in Fig. 3.8, is usually adopted. Firstly, a meaningful data representative of the problem to be tackled is collected. Then, the collected data is preprocessed and randomly split into training sample, validation sample, and testing data set. The training data set is used to train the ML model and the validation sample is adopted to select the hyperparameters leading to best performance. The ML algorithm is trained by minimizing the loss function. Once the ML model is trained, its performance is evaluated using the unseen test sample.

### 3.5.4 Machine Learning Techniques

The ML techniques can be broadly classified into three different categories: supervised learning, unsupervised learning, and reinforcement learning [183, 190]. As illustrated in Fig. 3.9, these categories differ in the types of input data available to the learner, the method by which the ML model is built, and the type of output. In a **supervised learning** scenario [191, 192], the ML algorithm receives a set of labeled data (data with clearly defined outputs or labels) and learns a mapping function on input data and its labels or targets in a supervised process by matching the predictions. The supervised learning category can be further classified into classification and regression problems. In **unsupervised learning** [191, 192], the ML method exclusively receives unlabeled

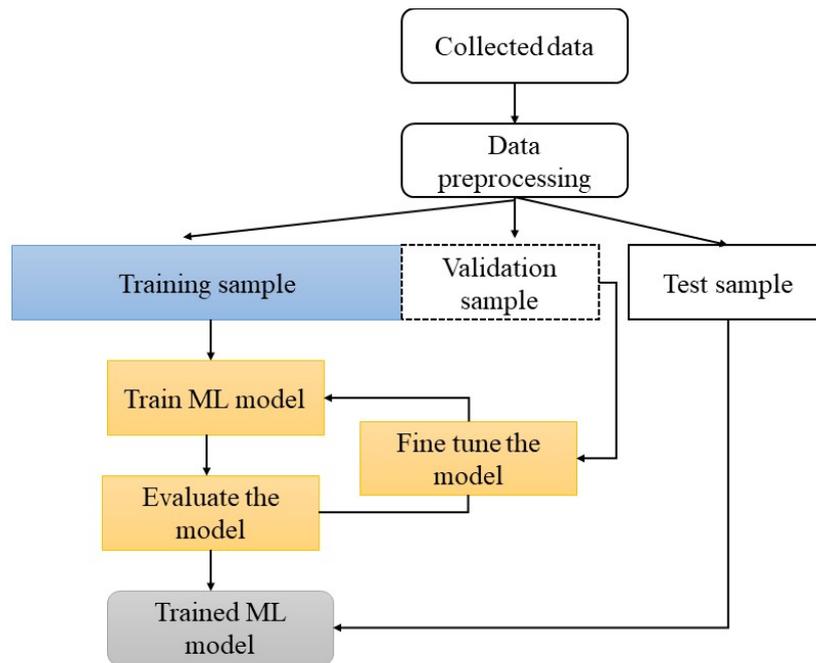


Figure 3.8: A generic machine learning workflow illustrating the typical stages of a learning process [189].

data (data without labels), trains without any guidance to find similar features in input data and to classify it into classes. Association and clustering are examples of unsupervised learning problems. In **reinforcement learning** [193], the ML algorithm (the agent) is trained independently through a trial and error approach to learn the optimal behavior (taking the best course of actions) in an interactive environment (a kind of simulation scenario that the agent has to explore) to obtain the maximum reward [194]. Figure 3.10 illustrates the working principle of a reinforcement learning method. While interacting with an environment, the agent evaluates the current situation (state), and makes an action. Based on the executed action, immediate feedback which can be either positive (a reward) or negative (a punishment for making a mistake) is received. The goal of the agent is to maximize rewards or minimize punishments. Exploration [195] or exploitation are types of reinforcement learning problems. Reinforcement learning methods can be used in different applications including self-driving cars, gaming, and healthcare [196, 197].

#### 3.5.4.1 Supervised Learning Algorithms

**Support Vector Machine** A support vector machine (SVM) [200] is a type of supervised ML technique used for solving both regression and classification problems. The core idea of SVM is to find an optimal hyperplane (line) that best separates the dataset into two categories or classes [201] as shown in Fig. 3.11. The objective of an

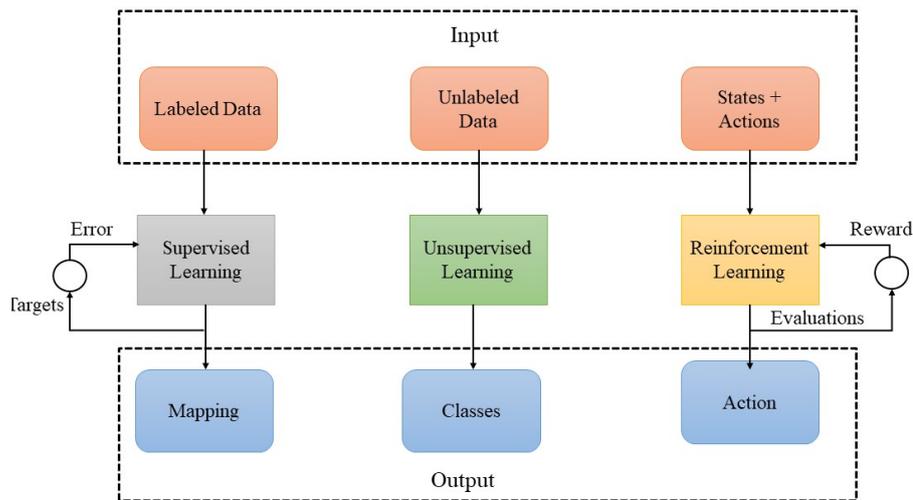


Figure 3.9: Different types of learning scenarios including supervised learning, unsupervised learning, and reinforcement learning [198].

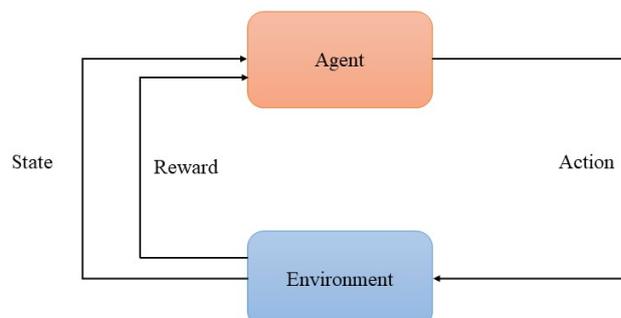


Figure 3.10: Illustration of the working principle of a reinforcement learning method [199].

SVM algorithm is to maximize the margin (i.e., the distance between the hyperplane and the nearest data point from either set (support vectors)). SVMs are widely used in different applications such as text classification tasks (e.g., category assignment, sentiment analysis, etc.), image recognition, and handwritten digit recognition [202]. SVMs achieve high accuracy and work well for small datasets. However, they are less efficient for larger and noisier datasets [203].

**Decision Tree** A decision tree [205] is a supervised learning technique employed for both classification and regression tasks, but mostly it is preferred for solving classification problems. The objective of a decision tree is to build a model that can be used to predict the class or the target variable by learning simple decision rules inferred from prior training data [206, 207]. For predicting the output or the target, the working principle of a decision tree algorithm starts from the root node and then compares

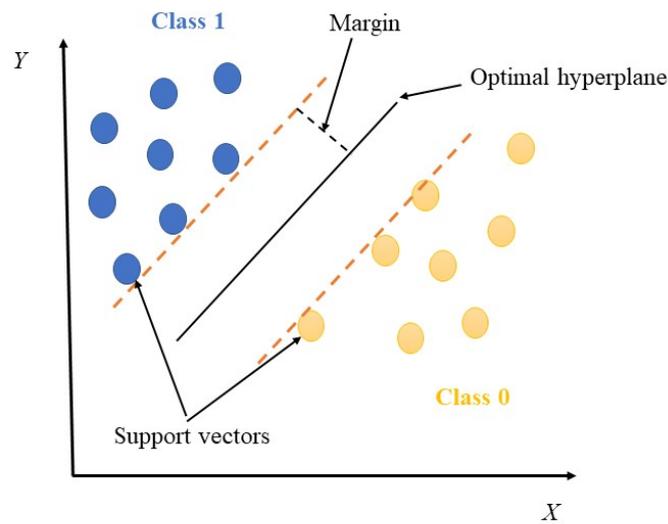


Figure 3.11: Principle of the support vector machine method for classification analysis [204].

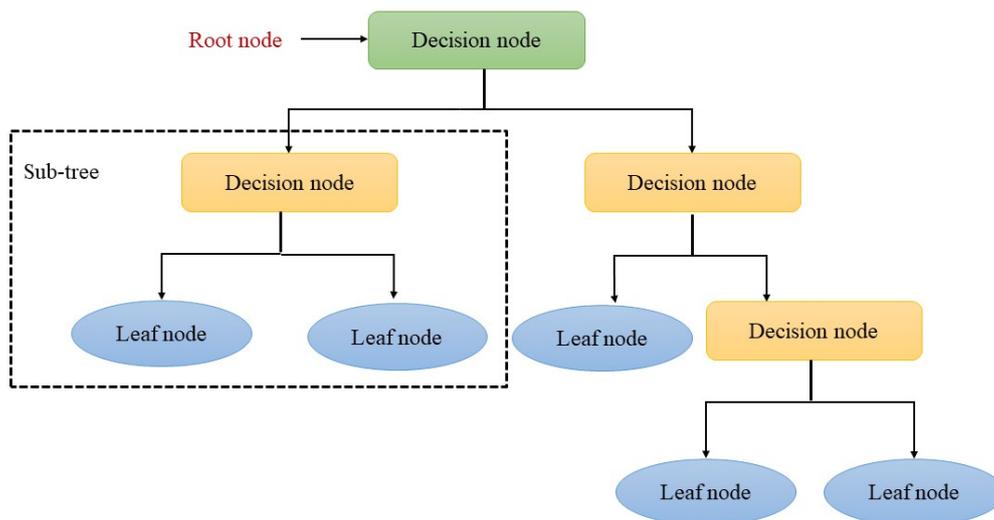


Figure 3.12: Working principle of a decision tree method [208].

the attribute of that node with the dataset attribute [207]. Based on the outcome of the comparison, it is divided into different branches, called decision nodes, which are further split into other nodes, called leaf nodes, defining the final output as illustrated in Fig. 3.12. Based on the type of the target or the output, decision tree algorithms can be classified into two categories: categorical variable decision trees (the type of the output is categorical), and continuous variable decision trees (the target is a continuous variable) [206].

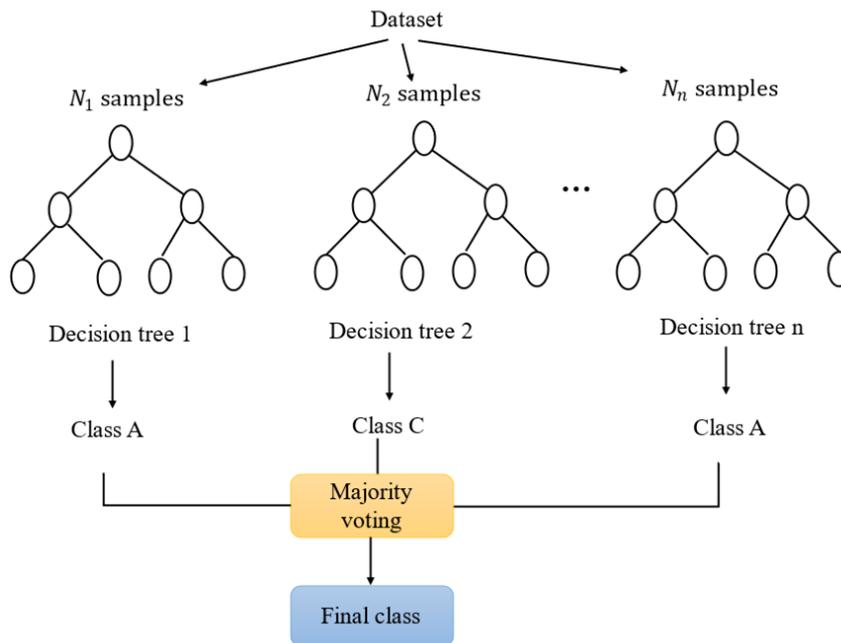


Figure 3.13: Principle of a random forest method employed for classification [211].

**Random Forest** A random forest [209] is one of the most commonly used supervised learning techniques widely employed for solving both classification and regression problems. It is an ensemble of decision trees trained using the “bagging” (known as bootstrap aggregation) approach (combining the output of multiple decision tree models to generate the final output) [210]. Figure 3.13 illustrates the working principle of a random forest method for classification analysis. Firstly, random samples are extracted from the training dataset. Then, a decision tree model is built for each sample. Afterwards, each decision tree generates an output. Lastly, the final output is calculated based on majority voting or averaging for classification and regression tasks, respectively [210].

**K-Nearest Neighbors** K-nearest neighbors (KNN) [212] is a simple and easy-to-implement supervised learning technique used for both classification and regression tasks. It determines the output or the target of a data point by majority voting or averaging of the  $k$  closest points to that data point, for classification and regression tasks, respectively [213]. The working principle of an KNN algorithm is illustrated in Fig. 3.14. The distances between the data point and all other points are first calculated. Then, the  $k$  points closest to the data point are selected. In case of a classification problem, the class of the data point is calculated by voting for the most frequent label or class given the classes of the  $k$  closest points (nearest neighbors) [214]. Whereas for regression, the output of the data point is computed as the mean or the average of the

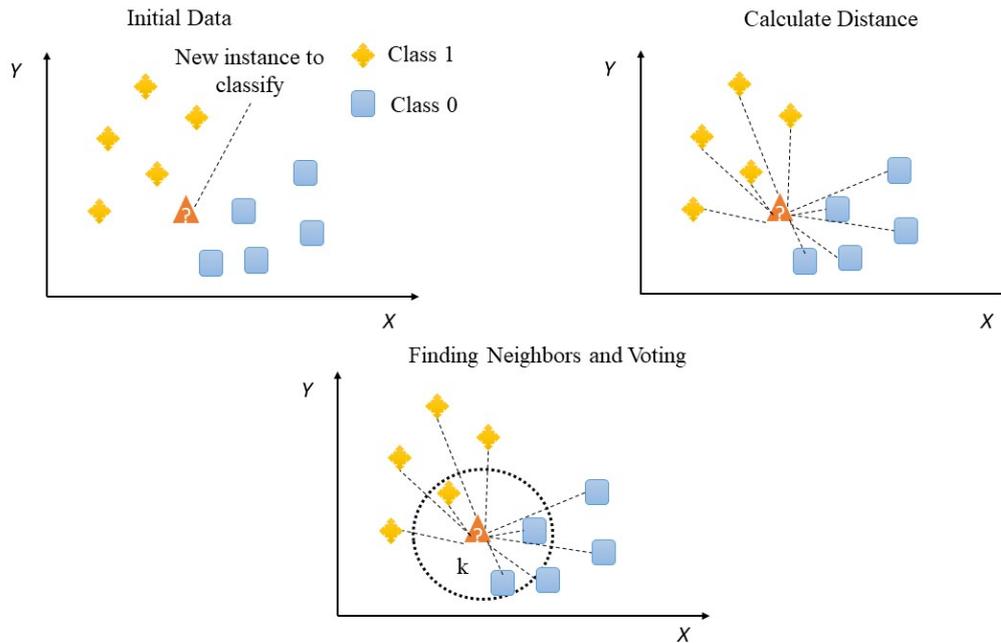


Figure 3.14: Principle of K-nearest neighbors algorithm [214].

outputs of the  $k$  nearest neighbors [214].

**Logistic Regression** Logistic regression [215] is a supervised learning method used for solving classification problems, and most commonly for tackling binary classification tasks (the outcome or target is binary or dichotomous). It is employed to calculate or predict the probability of a certain class or event. Logistic regression methods can be classified into three types: binary logistic regression (two possible outcomes), multinomial logistic regression (multiple unordered outcomes), and ordinal logistic regression (ordered outcomes). Logistic Regression (sometimes called the logistic model or logit model) is the statistical fitting of an s-curve logistic or logit function (called the sigmoid function) to data in order to calculate the probability of the occurrence of a specific categorical event based on one or more predictor variables [216] as illustrated in Fig. 3.15. The logistic function is generally expressed as follows [216]:

$$P(x) = \frac{1}{1 + e^{-x}}, \quad (3.1)$$

where  $e$  denotes Euler's number.

**Neural networks** Neural networks [217, 218, 219] are computational models inspired by the human brain structure. They have proven to excel in solving complex problems. They are universal approximation algorithms capable of learning any approximation

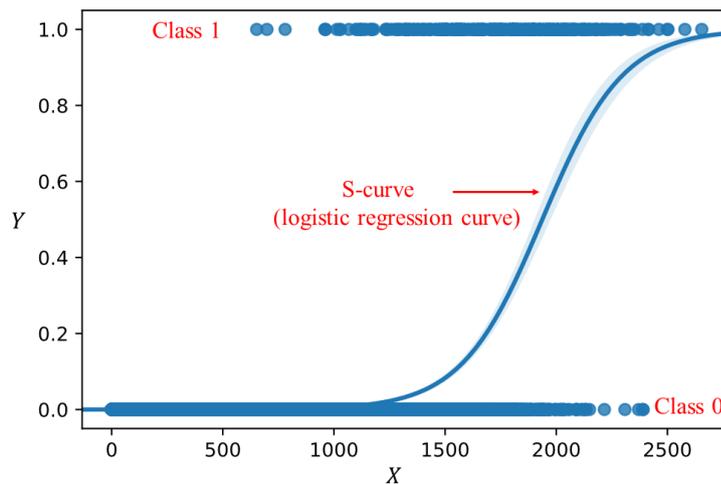


Figure 3.15: Principle of logistic regression [216].

function mapping inputs to outputs. The building block of a neural network is an artificial neuron, a simple computational unit, which takes inputs, aggregates them, and then produces an output based on an activation function (i.e., a simple mapping of summed weighted input to the output of the neuron) [220] as shown in Fig. 3.16. Conventionally, nonlinear activation functions are widely used as they are useful to better capture complex patterns in the data. Nonlinear functions like the logistic function, also called the sigmoid function, normalize the output of the neuron to a range between 0 and 1 with an s-shaped distribution [220]. They give a smooth gradient while converging, and they are useful for output neurons of binary classification-aimed neural networks [221]. However, they are computationally "expensive" functions due to the computation of an exponent, which makes the convergence of the network slower [221]. They are prone to the vanishing gradient problem inducing poor learning of neural networks. Another frequently used activation function to be mentioned is the the hyperbolic tangent function also called tanh which normalizes the output of the a neuron to a a range between -1 and 1. Unlike sigmoid, tanh is a zero-centric function, which eases the optimization of the loss function [221]. However, like sigmoid, tanh suffers from the vanishing gradient issue and is highly compute-intensive [221]. Rectifier activation function (ReLU) has been shown to provide better results and it is the most used activation function [220]. ReLU is a computationally inexpensive function so that the neural network converges very quickly. ReLU can deal with the vanishing gradient issue for positive values (saturation and vanishing gradient only occur for negative values as it gives zero value as inactive in the negative axis) [221]. To solve the issue of saturation for the negative values, leaky ReLU has been proposed by introducing a small slope (having the negative values scaled by small factor such as 0.01 instead

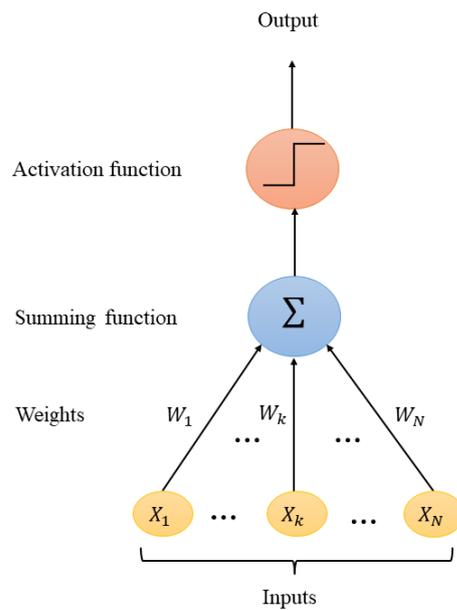


Figure 3.16: Model of a simple artificial neuron. [220].

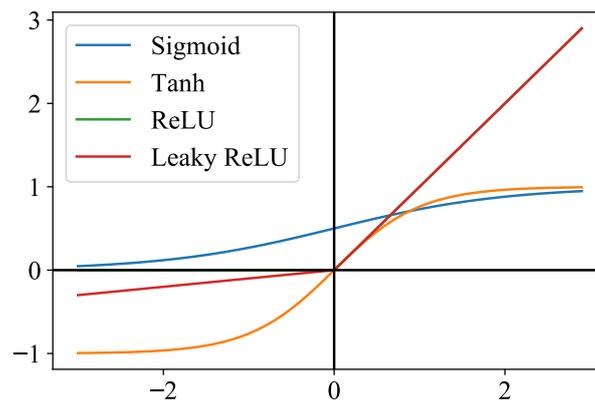


Figure 3.17: Commonly used activation functions for neural networks [222].

of 0 as of ReLU in order to enable their corresponding neurons to “stay alive”) [221]. Figure 3.17 illustrates the outputs of the aforementioned activation layer functions. Neural networks are networks of neurons. A row of neurons is called layer. A neural network is composed of multiple layers including an input layer or visible layer passing the input to the next layer, one or more hidden layers, and an output layer responsible for outputting a value or vector of values dependent on the type of problem (regression or classification) as shown in Fig. 3.18. If a neural network has more than one hidden layer, it is often called deep neural network.

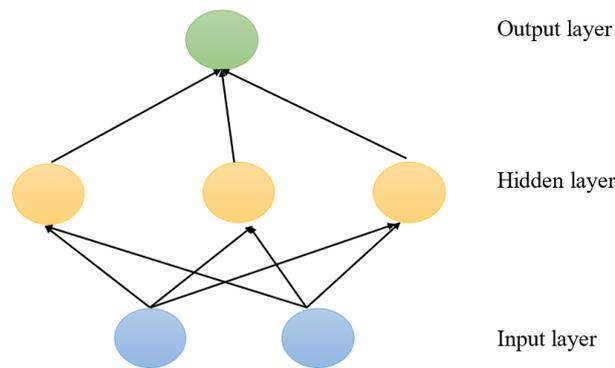


Figure 3.18: Architecture of a simple neural network [220].

**Convolutional Neural Networks** Convolutional neural networks (CNNs) [223, 224, 225] are a type of artificial neural networks, biologically inspired by the modular structure of the human visual cortex. They have been widely used in computer vision and have become the state of the art for several object recognition tasks such as handwritten digit recognition [226, 227]. They are powerful at extracting features. The architecture of a CNN comprises different hidden layers namely the convolutional, the pooling and the fully connected layers, made by stacking them on top of each other in sequence as shown in Fig. 3.19. The convolutional layers, the core building block of CNN, are composed of filters and feature maps. A filter can be described as a neuron of a layer that has weighted inputs and produces an output similar to an artificial neuron. A feature map is the outcome of applying a filter to the preceding layer, which has been used to extract features. A patch or receptive field is a fixed-size square that serves as the input size. A given filter is drawn across the entire previous layer and shifted one pixel at a time resulting in an activation of the neuron and the output is stored in the feature map [220]. The stride is represented by the distance that the filter passes over the input from the previous layer with each activation. By consolidating the features learned and expressed in the previous layer's feature map, the pooling layers are used to down-sample the previous layers' feature maps. They are used to compress feature representations and, in general, to reduce model overfitting. Pooling layers create feature maps based on the average or maximum value for the input value. The fully connected layer is the normal flat feedforward neural network layer used at the end of the network to create final nonlinear combinations of features and for making predictions.

**Recurrent Neural Networks** A recurrent neural network [229, 230] is a special type of artificial neural network perfectly suited for solving problems that involve sequential

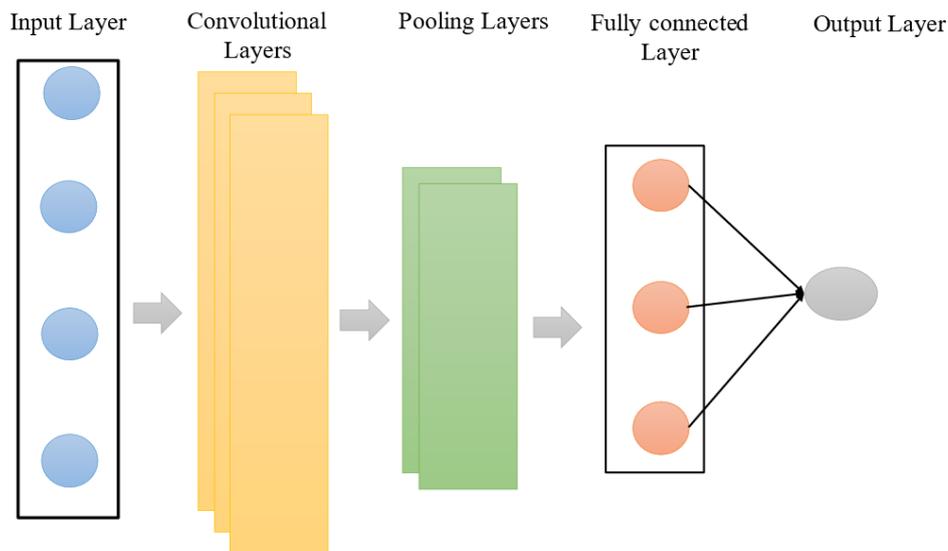


Figure 3.19: Architecture of a standard convolutional neural network [228].

data (e.g., time series, speech, text, financial data, etc.) such as language translation, natural language processing, speech recognition, etc. Recurrent neural networks are distinguished by their “memory” (feedback loops), which remembers all the information about what has been calculated (previous outputs or past inputs), whereby the output is computed by considering the current input and previously received inputs as shown in Fig. 3.20. Based on the length of inputs and outputs, recurrent neural networks can be classified into four different types: one-to-one (a single input and output), one-to-many (a single input and multiple outputs), many-to-one (taking a sequence of multiple inputs and predicting a single output), and many-to-many (taking multiple inputs and outputs) [231], as illustrated in Fig. 3.21. Although recurrent neural networks are good at processing variable length sequential data, they are unable to capture long term dependencies due to the vanishing gradient problem (the gradient becomes so small that updating parameters becomes insignificant, and thereby the algorithm stops learning) and exploding gradient issue (the gradient becomes too large, which makes the model unstable and leads to longer training times as well as poor model performance) arising from the recurrent connections. Such issues can be solved by adopting improved or advanced recurrent neural network architectures such as long short-term memory or gated recurrent unit algorithms to be discussed in the following subsections.

**Long short-Term Memory** Long short-term memory (LSTM) [234] is a specific kind of recurrent neural network used to process sequential data and to capture long-term sequential dependencies [235]. The core computational unit of LSTM is called

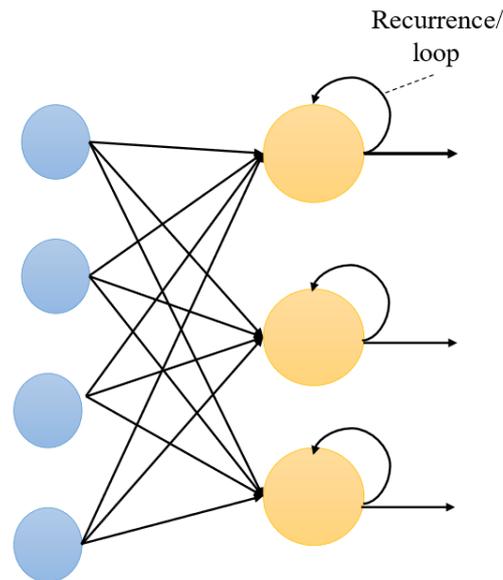


Figure 3.20: Architecture of a recurrent neural network [232].

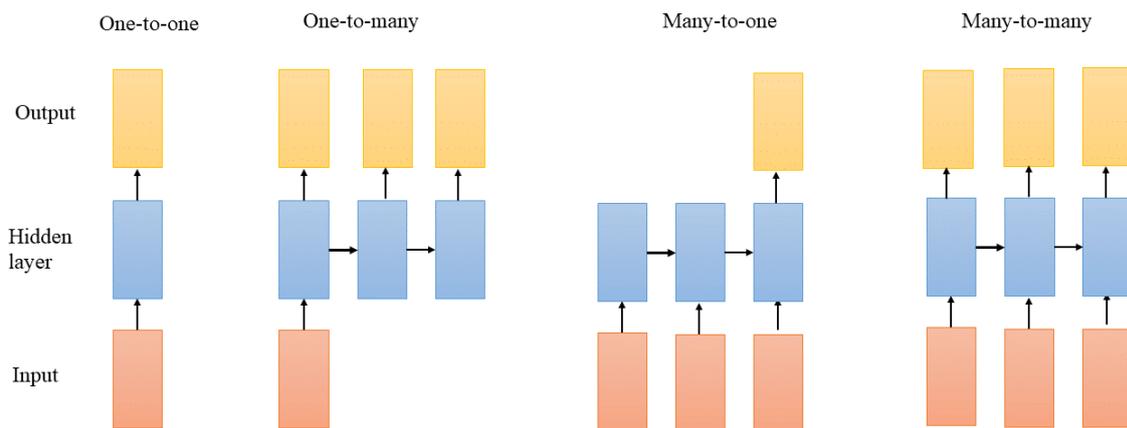


Figure 3.21: Different types of recurrent neural networks [233].

memory cell or block memory, containing weights and three gates, controlling the flow of information to the cell state [236]. The forget gate decides what information to throw away from the cell state [235]. The input gate determines what new information to store in the cell state, and the output gate decides what to output [235]. As shown in Fig. 3.22, the previous cell state  $c_{t-1}$  interacts with the previous cell output  $h_{t-1}$  and the present input  $x_t$  to determine, which elements of the internal state vector should be updated, kept, or discarded [235]. The LSTM cell is updated by applying the following equations [235]:

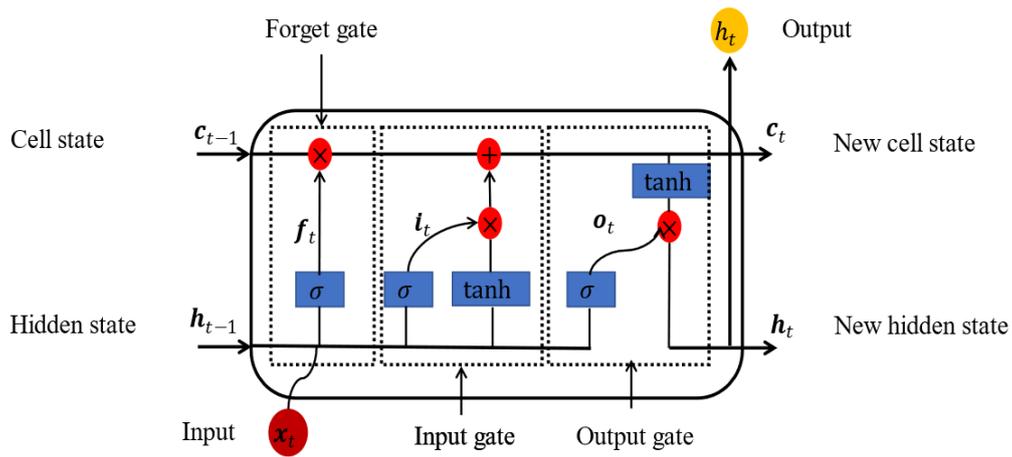


Figure 3.22: Structure of the Long Short-Term Memory (LSTM) cell [235].

$$f_t = \sigma(W_{xf} \cdot x_t + W_{hf} \cdot h_{(t-1)} + b_f) \quad (3.2)$$

$$i_t = \sigma(W_{xi} \cdot x_t + W_{hi} \cdot h_{(t-1)} + b_i) \quad (3.3)$$

$$\hat{c}_t = \tanh(W_{xc} \cdot x_t + W_{hc} \cdot h_{(t-1)} + b_c) \quad (3.4)$$

$$c_t = f_t \circ c_{(t-1)} + i_t \circ \hat{c}_t \quad (3.5)$$

$$o_t = \sigma(W_{xo} \cdot x_t + W_{ho} \cdot h_{(t-1)} + b_o) \quad (3.6)$$

$$h_t = o_t \circ \tanh(c_t) \quad (3.7)$$

where  $\sigma$  is the logistic sigmoid function, and  $f$ ,  $i$ ,  $c$  and  $o$  denote the forget gate, input gate, cell activation and output gate vectors, respectively [235]. 'o' represents the Hadamard (element-wise) product operator, all  $b$  are learned bias vectors, all  $W$  are trainable weight matrices, and  $c_t$  is a candidate cell value [235].

**Bidirectional Long Short-Term Memory** Bidirectional long short-term memory (BiLSTM) [237] is an extension of LSTM that helps to improve the performance of the model. As shown in Fig. 3.23, BiLSTM consists of two LSTMs: one forward LSTM model that takes the input in a forward direction beginning from the start of the input sequence, and one backward LSTM model that moves backward, beginning from the end of the input sequence. The output  $y_t$  of the model is generated by combining the forward output  $\vec{h}_t$  and the backward output  $\overleftarrow{h}_t$  as described by the following equations [235]:

$$\vec{h}_t = LSTM(x_t, \overrightarrow{\mathbf{h}_{t-1}}) \quad (3.8)$$

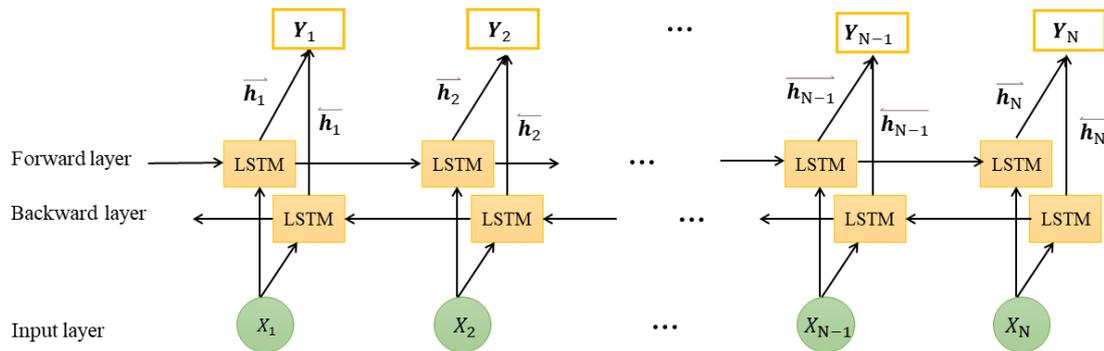


Figure 3.23: Structure of the bidirectional long short-term memory (BiLSTM) [235].

$$\overleftarrow{h}_t = LSTM(x_t, \overleftarrow{h}_{t-1}) \quad (3.9)$$

$$y_t = \overrightarrow{h}_t \oplus \overleftarrow{h}_t \quad (3.10)$$

where  $\oplus$  denotes an element-wise sum.

**Gated Recurrent Unit** The gated recurrent unit (GRU) [238], proposed by Cho et al. in 2014 to tackle the gradient vanishing problem, is an improved version of standard recurrent neural networks, used to process sequential data and to capture long-term dependencies [239]. The typical structure of a GRU, shown in Fig. 3.24, contains two gates, a reset and an update gate, controlling the flow of the information [239]. The update gate regulates the information that flows into the memory, while the reset gate controls the information flowing out the memory [239]. The GRU cell is updated at each time step  $t$  by applying the following equations [239]:

$$z_t = \sigma(W_z \cdot x_t + W_z \cdot h_{(t-1)} + b_z) \quad (3.11)$$

$$r_t = \sigma(W_r \cdot x_t + W_r \cdot h_{(t-1)} + b_r) \quad (3.12)$$

$$\hat{h}_t = \tanh(W_h \cdot x_t + W_h \cdot (r_t \circ h_{(t-1)}) + b_h) \quad (3.13)$$

$$h_t = z_t \circ h_{(t-1)} + (1 - z_t) \circ \hat{h}_t \quad (3.14)$$

where  $z_t$  denotes the update gate,  $r_t$  represents the reset gate,  $x_t$  is the input vector,  $h_t$  is the output vector,  $W$  and  $b$  represent the weight matrix and the bias vector, respectively [239].  $\sigma$  is the gate activation function. ' $\circ$ ' is the element-wise product operator. Equation (3.11) represents the update gate, Eq. (3.12) is the reset gate, Eq. (3.13) computes a candidate state for the current time step using the parts of the previous hidden state, and Eq. (3.14) shows how the output  $h_t$  is calculated [239].

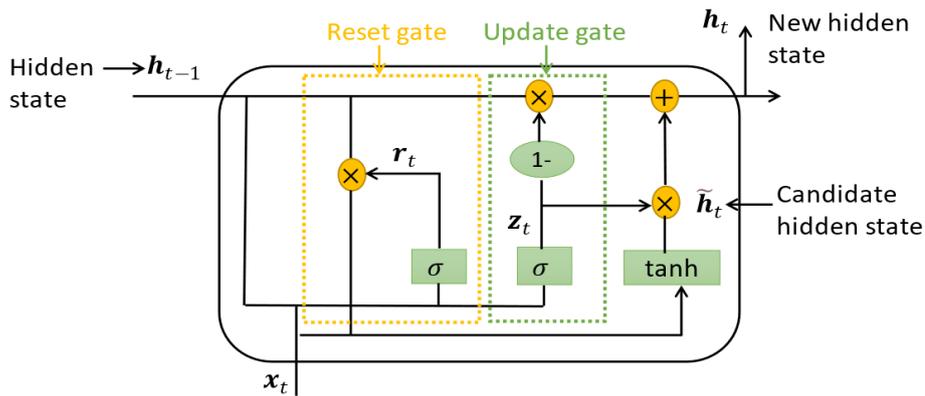


Figure 3.24: Structure of the gated recurrent unit (GRU) cell [239].

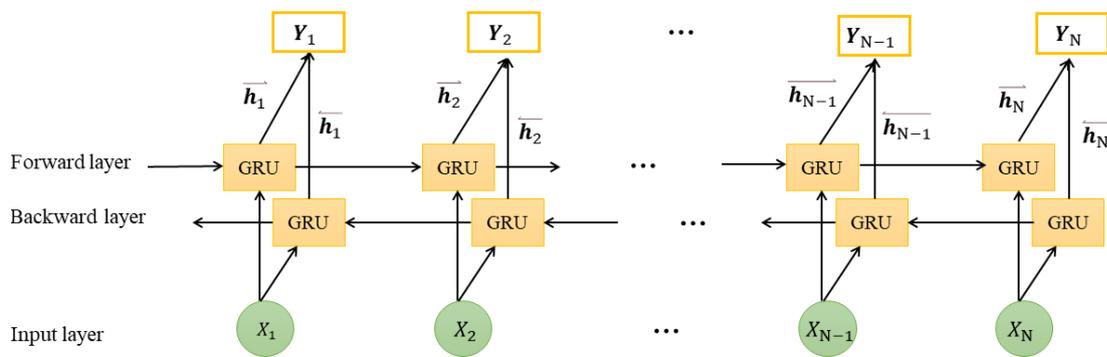


Figure 3.25: Structure of the bidirectional gated recurrent unit (BiGRU) [240].

**Bidirectional Gated Recurrent Unit** Bidirectional GRU (BiGRU) [240] is an extension of GRU that helps to improve the performance of the model. As illustrated in Fig. 3.25, BiGRU consists of two GRUs: one is a forward GRU model that takes the input in a forward direction, beginning from the start of the input sequence, and the other is a backward GRU model that moves backward, beginning from the end of the input sequence [240]. The output  $y_t$  of the model is generated by combining the forward output  $\vec{h}_t$  and the backward output  $\overleftarrow{h}_t$  as described by the following equations [240]:

$$\vec{h}_t = GRU(x_t, \vec{h}_{t-1}) \quad (3.15)$$

$$\overleftarrow{h}_t = GRU(x_t, \overleftarrow{h}_{t-1}) \quad (3.16)$$

$$y_t = \vec{h}_t \oplus \overleftarrow{h}_t \quad (3.17)$$

where  $\oplus$  denotes an element-wise sum.

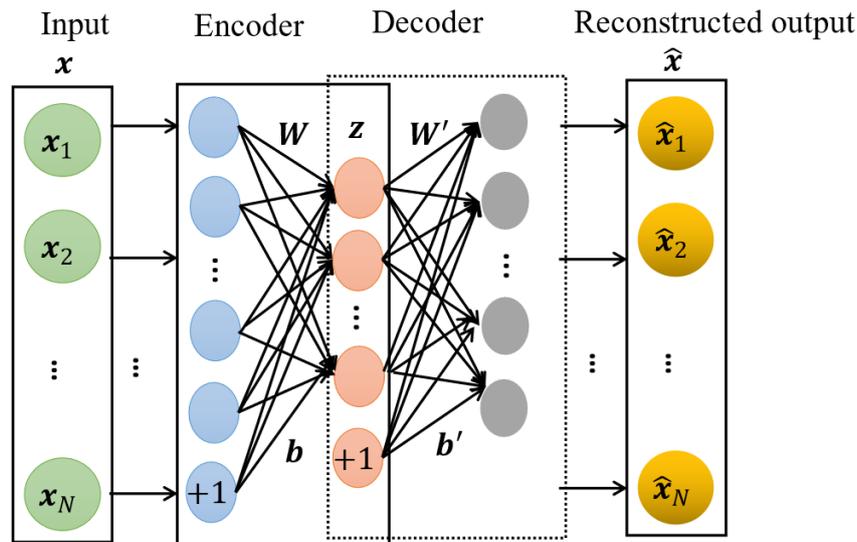


Figure 3.26: Structure of a standard autoencoder: the training objective is to minimize the reconstruction error between the output  $\hat{x}$  and the input  $x$  [235].

### 3.5.4.2 Unsupervised Learning Algorithms

**Autoencoder** An autoencoder (AE) [241] is a type of artificial neural network seeking to learn a compressed representation of an input in an unsupervised manner [235]. Fig. 3.26 shows the standard architecture of the AE. An AE consists of two sub-models namely the encoder and the decoder, whereby the former compresses an input  $x$  into lower-dimensional latent-space representation  $z$ , and the latter reconstructs the output  $\hat{x}$  given the compressed representation  $z$  as follows [235]:

$$z = f(Wx + b), \quad (3.18)$$

$$\hat{x} = g(W'z + b'), \quad (3.19)$$

where  $f$  and  $g$  represent the activation functions of the encoder and the decoder respectively. The weight matrix  $W$  (resp.  $W'$ ) and bias vector  $b$  (resp.  $b'$ ) are the learnable parameters for the encoder (resp. decoder) [235].

The training objective of the AE is to minimize the reconstruction error between the output  $\hat{x}$  and the input  $x$ , referred to as the loss function, typically the mean square error (MSE), defined as [235]:

$$L(\theta) = \sum \|x - \hat{x}\|^2 \quad (3.20)$$

where  $\theta = \{W, b, W', b'\}$  denotes the set of parameters to be optimized.

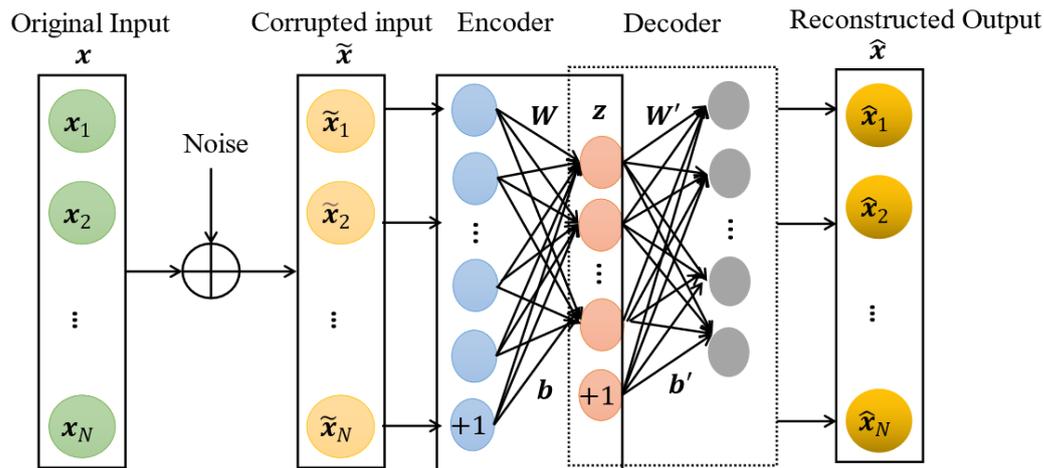


Figure 3.27: Structure of a standard denoising autoencoder: the training objective is to minimize the reconstruction error between the output  $\hat{x}$  and the original input  $x$  (uncorrupted input) [235].

Autoencoders can be useful as non-linear feature extraction and dimensionality reduction techniques, which are used to remove noise or redundant or irrelevant information from high dimensional input data (leading to model overfitting) while providing a useful lower dimensional representation (encoding) of the input. The encoded features or compressed data representation can then be used to train various ML methods (e.g., artificial neural networks, support vector regression, etc.) for solving various learning tasks (e.g. classification, regression, etc.) [242]. Anomaly detection is another application for autoencoders. In general, autoencoders are trained using only data representing normal behavior. A well-trained autoencoder should be able to reconstruct normal instances very well with small errors, but it should fail to reproduce anomalous observations (not seen during the training phase) with much larger reconstruction errors or losses [242].

**Denoising Autoencoder** A denoising autoencoder (DAE) [243, 244] is an extension and a stochastic version of the autoencoder [235]. It lowers the likelihood of learning the identity function by randomly corrupting the input (e.g., adding noise) and attempting to reconstruct the original, uncorrupted input. Figure 3.27 shows a standard architecture of the DAE. The input  $x$  is corrupted by adding some noise to get  $\tilde{x}$ . The encoder then converts the noisy input  $\tilde{x}$  to a low dimensional representation  $z$ , which the decoder uses to reconstruct the output  $\hat{x}$ .

**Denosing Convolutional Autoencoder** A denosing convolutional autoencoder [245] has the same standard DAE structure with convolutional encoding and decoding layers instead of fully connected layers [235]. Each convolution layer consists of multiple kernels used to extract features or so-called feature maps [235]. The latent representation  $z_k$  of feature map  $k$  is represented by [245]:

$$z_k = W_k \cdot x + b_k, \quad (3.21)$$

The output of the decoder is expressed as [245]:

$$\hat{x} = g \left( \sum_{k \in H} W'_k \cdot x + b'_k \right) \quad (3.22)$$

where  $H$  denotes the group of latent feature maps.

**Variational Autoencoder** A variational autoencoder (VAE) [246, 247, 248] is a specific type of deep generative models, having the same architecture as traditional autoencoders. The VAE forces the model to learn the input distribution by constraining the latent-space representation  $z$  to be distributed according to a prior distribution  $p_\theta(z)$ , typically a multivariate unit Gaussian distribution, in contrast to the conventional autoencoder, which deterministically encodes the input into the latent representation and then outputs the reconstructed input.  $x$  is sampled from  $p_\theta(x|z)$ , which can be obtained using a decoder neural network (generative model) with parameters  $\theta$ , to decode  $z$  into a distribution over the observation  $x$ . As the true posterior  $p_\theta(z|x)$  is intractable for a continuous latent space  $z$ , variational inference techniques are frequently used to find an approximate posterior  $q_\phi(z|x)$ , which can be represented by an encoder neural network (recognition network) with parameters  $\phi$ . Typically, this posterior is supposed to be  $N(\mu_\phi(x), \sigma_\phi^2(x))$ , with the neural network optimizing  $\mu_\phi(x)$  and  $\sigma_\phi(x)$  parameters. As shown in Fig. 3.28, a VAE is composed of an encoder, a latent distribution, and a decoder. The input  $x$  is fed into the encoder (i.e.,  $q_\phi(z|x)$ ) to produce the parameters of the latent distribution (e.g., the mean  $\mu$ , the standard deviation  $\sigma$ ). Then, the latent vector  $z$  is produced by performing random sampling from the distribution. The decoder (i.e.,  $p_\theta(z|x)$ ) then adopts the latent vector to reconstruct the original input  $x$ . The encoder and the decoder networks are trained simultaneously by maximizing the objective function which can be expressed as follows [248, 246]:

$$\mathcal{L}_{VAE}(\theta, \phi; x) = \mathbb{E}_{q_\phi(z|x)} [\log(p_\theta(x|z))] - D_{KL}[q_\phi(z|x) || p_\theta(z)], \quad (3.23)$$

where  $p_\theta(x|z)$  denotes the likelihood of  $x$  given  $z$ , and  $\theta$  and  $\phi$  represent the decoder

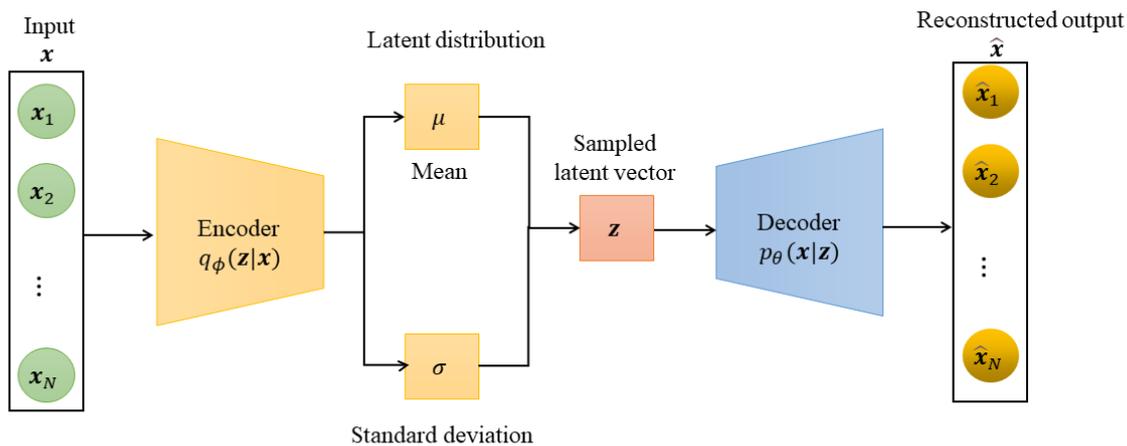


Figure 3.28: Structure of a variational autoencoder (VAE) [242].

and the encoder parameters to be optimized.

The first term of Eq. (3.23) is the reconstruction likelihood, which denotes how well the decoder learns the training data. The second term of Eq. (3.23) represents the Kullback-Leibler divergence ( $D_{KL}$ ) between the probability distribution learnt by the encoder and the prior distribution of the latent representation  $z$ , which acts as a regularization term to make the posterior distribution resemble the prior distribution. The purpose of training a VAE is to jointly determine the optimal  $\theta$  and  $\phi$  parameters by maximizing the objective function (Eq. (3.23)) with respect to both parameters. The maximization of the objective function with respect to  $\theta$  maximizes the reconstruction likelihood (i.e. reducing the reconstruction error between the input and its reconstruction), whereas maximizing Eq. (3.23) with respect to  $\phi$  minimizes the ( $D_{KL}$ ) (i.e. maximizing the similarity between the true posterior  $p_\theta(z|x)$  and its approximation  $q_\phi(z|x)$ ).

**Conditional Variational Autoencoder** A conditional variational autoencoder (CVAE) [249, 250] is an extension of the standard VAE by conditioning the encoder and the decoder with auxiliary covariate information (i.e., additional property of the data such as label or a class) as shown in Fig. 3.29. The encoder is conditioned on the input  $x$  and the covariates  $c$  whereas the decoder is conditioned on the latent representation  $z$  and  $c$ . Hence, the objective function of CVAE is expressed as follows [251]:

$$\mathcal{L}_{CVAE}(\theta, \phi; x, c) = \mathbb{E}_{q_\phi(z|x,c)} [\log(p_\theta(x|z,c))] - D_{KL}[q_\phi(z|x,c) || p_\theta(z|c)] \quad (3.24)$$

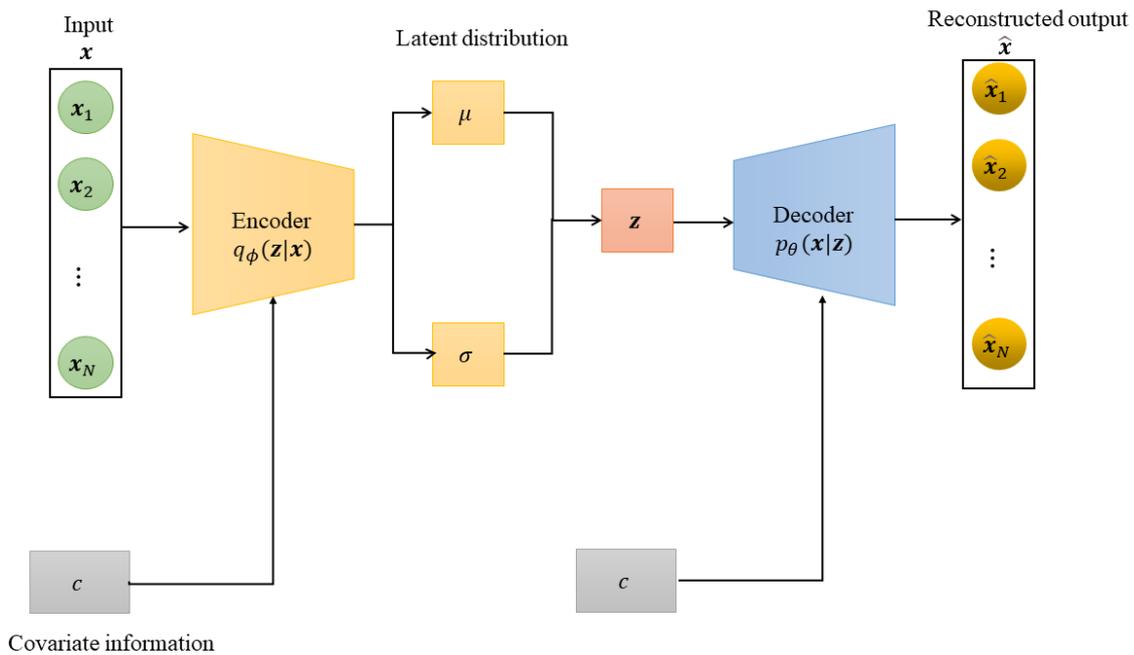


Figure 3.29: Architecture of a conditional variational autoencoder [242].

**Generative Adversarial Network** Generative adversarial networks (GANs) [252, 253, 254] are a type of generative models able to create new content such as an image, text, or audio. The GAN model architecture consists of two sub-models, namely the generator  $G$  and the discriminator  $D$ , which are trained and optimized simultaneously while competing with each other.  $G$  is trained to produce realistic samples from a random noise input  $z$ , to fool  $D$ , which is trained to distinguish the real samples  $x$  from the fake ones made by  $G$ . The objective function of a GAN model ( $\mathcal{L}_{GAN}$ ), whereby  $G$  tries to minimize it and  $D$  tries to maximize it, can be formulated as follows [255]:

$$\mathcal{L}_{GAN} = \min_G \max_D \left[ \mathbb{E}_{x \sim p_{data}(x)} [\log(D(x))] + \mathbb{E}_{z \sim p_z(z)} [\log((1 - D(G(z))))], \quad (3.25)$$

**Multitask Learning** Multi-task learning (MTL) [256] is a learning paradigm in ML that aims to improve the generalization performance of multiple tasks by learning them jointly while sharing knowledge across them. It has been widely used in a variety of fields, including natural language processing and computer vision [4]. The MTL approaches can be divided into two types: hard parameter sharing and soft parameter sharing [256]. The hard parameter sharing method involves sharing the hidden layers with the various tasks (completely sharing the weights and parameters between all tasks) while preserving task-specific output layers learnt independently by each task [4]. Whereas for the soft parameter sharing approach, a model with its own parameters is learnt for each task, and the distance between the parameters of the model is then

regularized to encourage similarities among related parameters [4].

### 3.5.5 Performance Metrics

To evaluate the performance of an ML model, several metrics are adopted based on the type of the problem or task to be solved (e.g., classification or regression problems).

#### 3.5.5.1 Classification Evaluation Metrics

To assess the performance of a classification model (“a classifier”), the following metrics are commonly used:

- *Confusion matrix*: it is a performance measurement method summarizing the prediction results on a classification problem, whereby the number of correct and incorrect predictions made by the ML method are summarized with count values and broken down by each class. Figure 3.30 illustrates the confusion matrix for a binary classification problem. The different elements of a confusion matrix can be defined as follows:
  - True positive (TP) denotes the number of “positive” instances correctly classified (the predicted class and actual class both are positive).
  - True negative (TN) represents the number of “negative” instances correctly classified (the predicted class and actual class both are negative)
  - False negative (FN) denotes the number of “positive” instances misclassified as “negative” (the predicted class is negative whereas the actual class is positive)
  - False positive (FP) denotes the number of “negative” instances misclassified as “positive” (the predictive class is positive whereas the actual class is negative)
- *Precision* (P) quantifies the relevance of the predictions made by the ML model. It is expressed as:

$$P = \frac{TP}{TP + FP} \quad (3.26)$$

- *Recall* (R) provides the total relevant results correctly classified by the ML model. It is formulated as:

$$R = \frac{TP}{TP + FN} \quad (3.27)$$

		True class	
		Positive	Negative
Predicted class	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Figure 3.30: Confusion matrix of a classifier [257].

- *F1 score* score is the harmonic mean of the precision and recall, calculated as:

$$F1 = \frac{2PR}{P + R} \quad (3.28)$$

- *Accuracy* (A) can be defined as the total number of correctly classified instances divided by the total number of test instances. It is calculated as follows:

$$A = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.29)$$

- *Receiver operating curve* (ROC) is a graphical plot illustrating the performance of a classification model at all classification thresholds. It plots the true positive rate (TPR) (the proportion of observations that are correctly predicted to be positive out of all positive observations) against the false positive rate (FPR) (the proportion of observations that are incorrectly predicted to be positive out of all negative observations).
- *Area under curve* (AUC) measures the degree of separability between the classes (the ability of a classifier to distinguish between classes). The higher the AUC, the better the performance of the ML model at discriminating between the classes. An excellent or perfect model achieves an AUC near to 1.

### 3.5.5.2 Regression Evaluation Metrics

The most commonly used metrics for evaluating the performance of a regression model are the following:

- *Mean Squared Error* (MSE) is computed as the mean or average of the squared differences between predicted and actual value. It is given by:

$$MSE = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N} \quad (3.30)$$

where  $y_i$  denotes the  $i^{th}$  expected value, and  $\hat{y}_i$  represents the  $i^{th}$  predicted value.

- *Root Mean Squared Error* (RMSE) is calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad (3.31)$$

- *Mean Absolute Error* (MAE) is calculated as the average of the absolute error values.

$$MAE = \frac{\sum_{i=1}^N |y_i - \hat{y}_i|}{N} \quad (3.32)$$

- *Mean absolute percentage error* (MAPE) measures the error between actual and predicted values as a percentage. It is calculated as follows:

$$MAPE = \frac{\sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{|y_i|}}{N} \quad (3.33)$$

## 3.6 Use Cases

### 3.6.1 Equipment Failure Prediction

For this use case, the goal is to predict, if an equipment will fail in a given period (e.g., next week, next month) or not. In machine learning terminology, this is referred to as a binary classification problem (class 1: “equipment will fail”, class 0: “equipment will not fail”). A labeled dataset including both normal and failure instances/observations has to be fed to the ML model for the training and learning phase.

### 3.6.2 Equipment Remaining Useful Life Prediction

In this use case, the purpose is to predict the remaining useful life (RUL) of an equipment, which is defined as the time remaining for a component to perform its intended function before failing. In machine learning terms, this is considered as a regression problem (the target/the output is to predict the value of the RUL). A dataset including the operational behavior of an equipment at different stages of an asset life cycle (e.g., normal operation, early degradation phase, 1% remaining life etc.) has to be available for solving such problem.

### 3.6.3 Anomaly Detection in Equipment Behavior

In this use case, the objective is to detect unusual or abnormal behaviors or patterns (e.g., asset degradation, power issues, failures, etc.) during the equipment operation, and to trigger an alert as soon as an anomaly is spotted. This use case can be considered either as a supervised learning problem or most commonly as a unsupervised learning task. To solve such a use case as a supervised learning problem (binary classification), labeled data composed of “normal” and “abnormal” instances has to be provided.

### 3.6.4 Equipment Failure Diagnosis

The goal of this use case is to identify the type of equipment failures or anomalies. This use case can be considered as a multi-classification problem. A labeled dataset including instances for each type of failures to be investigated has to be given for the ML model in order to learn the patterns characterizing each failure type.

## 3.7 Challenges

Implementing a predictive maintenance program is not trouble-free. Several challenges may be encountered by deploying such a strategy in industry. Among them, we can list the following [258, 259]:

- *Lack of failure data:* As equipment failures do not occur frequently, collecting enough failure data can be a time-consuming and challenging task. Therefore, developing accurate and reliable ML models to solve predictive maintenance use cases can be very difficult due to the lack of sufficient training data.
- *Integration:* To implement a fully functional predictive maintenance program, there is a need to establish an IoT network or platform to collect, process and store the data, and to integrate it with other existing programs such as an ERP system (Enterprise Resource Planning), an MES (Manufacturing Execution System) or other supervisory and process control systems. The different systems have to be connected through newly developed API's (Application Programming Interfaces).
- *Security and privacy concerns:* As many data of the production or manufacturing plant is being collected, processed, and stored somewhere in a database that can be located in the cloud, there are a lot of concerns about the security of the IoT network and the privacy of the data.

- *Costs*: Implementing a predictive maintenance strategy can be expensive. The costs include the costs for buying and building new smart sensors, training the staff to work with the new technologies, building of the IoT platform etc.

## 4 Predictive Maintenance for Semiconductor Lasers

Since its inception in 1962, the semiconductor laser has rapidly evolved to meet the demands of optical networks in terms of high launch power, low linewidth, moderate electrical power consumption, and wavelength stability [260, 261]. Furthermore, these applications have stringent requirements for semiconductor laser reliability.

Degradation mechanisms govern the reliability of such lasers. A variety of internal and external conditions, such as contamination, crystal flaws, ambient temperature, and so on, can have a negative impact on laser performance[7]. Many of these factors are uncontrollable, and their interaction can result in complex degradation mechanisms that are difficult to model [262]. Furthermore, the complexity of the laser structures, as well as the variety of factors causing degradation, make it difficult to fully comprehend the physical processes governing degradation. Consequently, it is challenging to adopt physics-based models that make use of explicit mathematical equations to model the system's degradation behavior.

Data-driven prognostic approaches, which do not require system specific knowledge or physical models to conduct the prognostics and health monitoring, have recently gained popularity. They learn the system behavior by extracting insights from the gathered data, and thereby predict the future state of degradation, the type of failure modes, and the remaining useful life. The availability of run-to-failure data sets that indicate both the behavior for normal operation and the degradation process under various operating conditions is necessary for the development of such prognostic algorithms.

In this chapter, several ML-based data-driven diagnostic and prognostic approaches for different semiconductor laser predictive maintenance use cases are presented and discussed. The validation of the proposed ML models is carried out using either synthetic data or experimental data obtained from accelerated life tests conducted under high stress conditions (for example, high temperatures) to speed up the degradation and thereby shorten the time to failure of the device as it takes a long time (many years) to collect meaningful reliability field data. Section 4.1 discusses the application of an unsupervised ML approach, specifically the conditional variational autoencoder, to the use case of early semiconductor laser degradation prediction. Section 4.2 describes the

use of a supervised learning technique (LSTM) to identify the different types of laser failure modes. The application of an artificial neural network model to predict the MTTF of a semiconductor laser is discussed in Section 4.3. Adopting ML techniques to tackle the semiconductor laser RUL prediction use case is covered in Section 4.4. Section 4.5 describes a combined ML approach comprised of different ML models for addressing various semiconductor laser predictive maintenance use cases.

## 4.1 Degradation Prediction Use Case

The goal of this use case is to predict the degradation behavior that semiconductor laser devices deployed in optical networks may exhibit during operation, allowing for the scheduling of maintenance plans to be performed as early and as efficiently as possible. This section's content has already been published in [242].

### 4.1.1 Model Architecture

The proposed approach, referred to as semi-CVAE (SCVAE) in the following, is a variant of CVAE in which only the decoder is conditioned on both the latent variable and the operating conditions (auxiliary covariate information), while the encoding is unaffected by the operating conditions. Figure 4.1 shows the proposed approach for semiconductor laser degradation prediction. The proposed model is composed of two sub-models, the encoder, and the decoder. The encoder compresses the input  $x$ , composed of a sequence of output power measurements  $[x_0, x_1, \dots, x_5]$ , and outputs the parameters of the latent space distribution, namely the mean vector  $\mu$  and the variance vector  $\sigma$ . After that, the latent vector  $z$  is sampled from the distribution  $N(\mu, \sigma)$ . The operating conditions ( $oc$ ) namely the temperature  $T$  and the laser current  $I$ , which influence the degradation trend, are combined with  $z$  and fed to the decoder to produce a reconstructed input  $\hat{x}$ . The objective function of the SCVAE model is expressed as follows:

$$\mathcal{L}_{VAE}(\theta, \phi; x, oc) = \mathbb{E}_{q_\phi(z|x)} [\log(p_\theta(x|z, oc))] - D_{KL}[q_\phi(z|x) || p_\theta(z|oc)], \quad (4.1)$$

The encoder's structure includes two GRU layers with 40 and 20 cells, respectively, that are used to extract the features of the sequential input  $x$ . The decoder is inversely symmetric to the encoder model. Adaptive moment estimation (Adam) is selected as an optimizer. ReLU is adopted as an activation function for the hidden layers of the model.

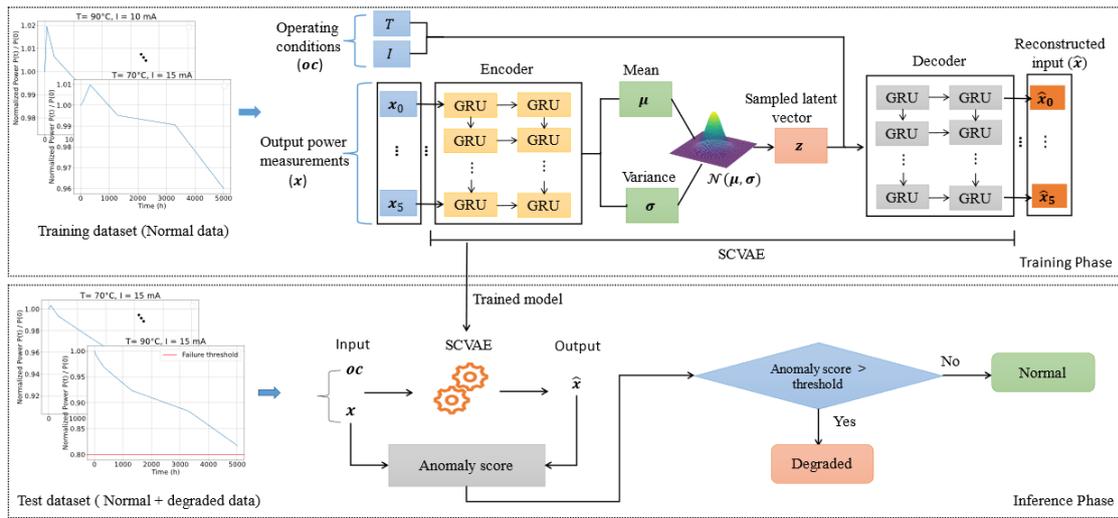


Figure 4.1: Proposed GRU-based SCVAE architecture for semiconductor laser degradation prediction [242].

#### 4.1.2 Anomaly Detection

The SCVAE model is trained with data modeling the normal operating behavior only in order to learn the normal pattern. The reconstruction error (i.e., square  $L_2$  norm) is adopted as anomaly score during the inference phase. Reconstructed input data with a high anomaly score is considered to be abnormal (e.g., sudden degradation, gradual degradation, etc.) because it is expected that the fully trained SCVAE reconstructs normal data by yielding very low reconstruction errors, while failing to produce anomalous data that was not encountered during the training phase. The process of the classification of an observation as abnormal or degraded (i.e., abnormal) is outlined in Algorithm 1.

The instance is categorized as "degraded", if the calculated anomaly score is higher than a threshold " $\alpha$ ," else it is regarded as "normal". " $\alpha$ " is a parameter to be optimized. Once an anomaly is detected, the type of degradation  $D_{type}$  (e.g., sudden or gradual failure modes) can be determined by establishing a two-threshold system optimized based on the reconstruction errors achieved by the SCVAE for each degradation type, as described below.

$$D_{type} = \begin{cases} gradual, & \text{if } \alpha < e_r < \beta \\ sudden, & \text{if } e_r > \beta \end{cases} \quad (4.2)$$

where  $\beta$  is a parameter to be optimized. The SCVAE model is expected to reconstruct "sudden degradation" samples with higher reconstruction errors than "gradual degradation" samples because the gradual failure mode trend is slower and looks more like

**Algorithm 1** SCVAE based anomaly detection

---

**Input** Normal dataset ( $x, oc$ ), degraded dataset  $(\mathbf{x}^{(i)}, \mathbf{oc}^{(i)})$ ,  $i=1, \dots, N$ , threshold  $\alpha$

**Output** Reconstruction error

$\phi, \theta \leftarrow$  train an SCVAE given the normal data ( $x, oc$ )  $\mu_z(\mathbf{i}), \sigma_z(\mathbf{i}) = f_\theta(z|\mathbf{x}^{(i)})$ .

$z_{(i)} \sim N(\mu_z(\mathbf{i}), \sigma_z(\mathbf{i}))$

$e_r(\mathbf{i}) = \|\mathbf{x}^{(i)} - \hat{\mathbf{x}}^{(i)}\|$

**for**  $i = 1, \dots, N$  **do**

**if**  $e_r(\mathbf{i}) > \alpha$  **then**

$\mathbf{x}^{(i)}$  is degraded.

**else**

$\mathbf{x}^{(i)}$  is normal.

**end if**

**end for**

---

normal patterns than the sudden degradation trend, which is faster and looks different.

### 4.1.3 Experimental Data

Experimental data derived from accelerated aging tests performed on VCSEL devices operating under various operating conditions is adopted to validate the proposed approach. The aging tests are performed at high temperatures (70°C or 90°C) to significantly accelerate laser degradation and thus device failure. Under constant current operation, the output power (i.e., degradation parameter) is monitored for 15,000 hours. The device's failure criterion is defined as a decrease in output power of more than 20% from its initial value. Figure 4.2 depicts aging test results for several semiconductor lasers performed under various operating conditions. In total, a dataset of 6,786 samples composed of sequences of output power measurements monitored for 5,000 hours, combined with the operating conditions  $T$  and  $I$ , is constructed. The state of the device (normal or degraded) based on the aforementioned failure criterion is assigned to each sample. The data is then normalized and randomly divided into training data (composed of 80% of the samples) and test dataset (the remaining 20%). The training dataset contains only samples of normal devices, whereas the test dataset comprises samples of both, normal (25%) and degraded devices (75%). In order to test the effectiveness of the ML model in predicting degradation early, only degraded lasers that failed after 5,000 hours are taken into consideration.

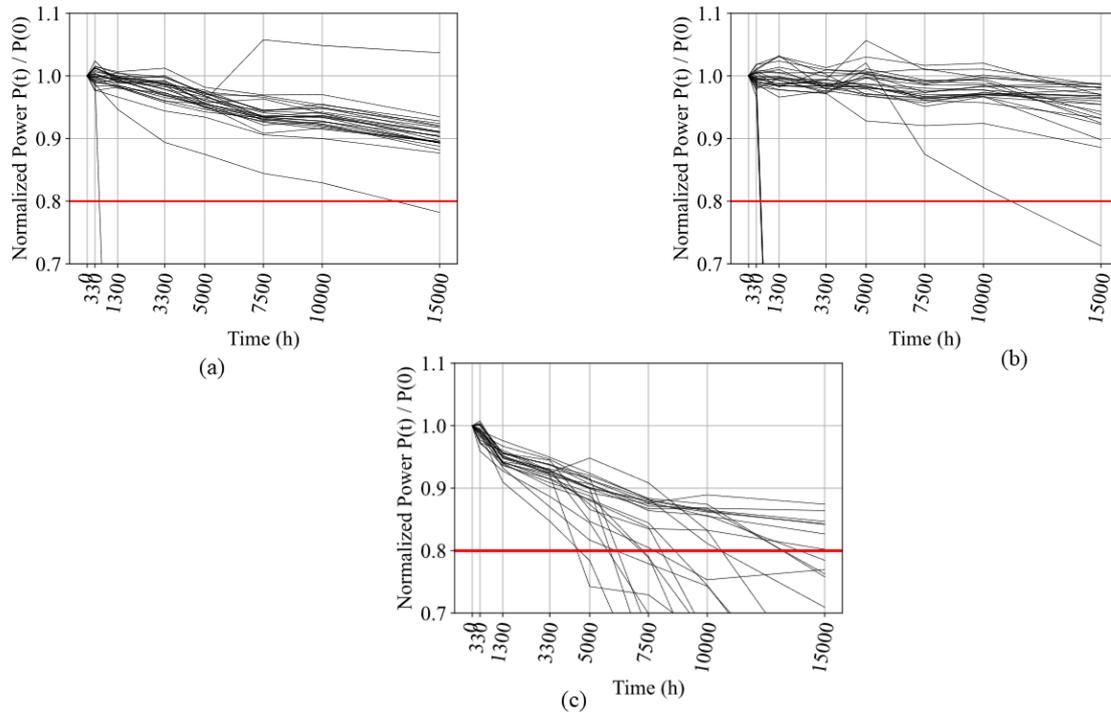


Figure 4.2: Experimental aging test data of VCSELs conducted under different operating conditions: (a)  $T = 70^{\circ}\text{C}$ ,  $I = 15\text{ mA}$ , (b)  $T = 90^{\circ}\text{C}$ ,  $I = 10\text{ mA}$ , (c)  $T = 90^{\circ}\text{C}$ ,  $I = 15\text{ mA}$  [242].

#### 4.1.4 Experimental Results

##### 4.1.4.1 Performance Evaluation

The SCVAE model's degradation prediction capability is improved by selecting an optimal threshold  $\alpha$ . Figure 4.3 shows the precision, the recall (i.e. sensitivity), and the F1 score curves as a function of  $\alpha$ . If the threshold is set too low, many degraded laser devices are categorized as normal, resulting in a higher false positive ratio. However, if the threshold is too high, many healthy devices are classified as degraded, increasing the false negative ratio. As a result, the optimal threshold that provides the best precision and recall tradeoff (i.e., maximizing the F1 score) is chosen. For the optimal chosen threshold of 0.013, the precision, the recall, and the F1 score are 98.6%, 92.2%, and 95.3%, respectively. The normal and abnormal reconstruction score distributions appear well separated, as shown in Fig. 4.4, indicating that the SCVAE has efficiently learned the relevant features characterizing the normal behavior or pattern. The ROC curve shows that the SCVAE model can discriminate between normal and degraded classes very well, with an AUC (i.e. degree of separability between classes) of 0.96, and that the optimally chosen threshold generalizes well for the test dataset.

An unseen test dataset consisting of measurement sequences with 5,000 hours of ran-

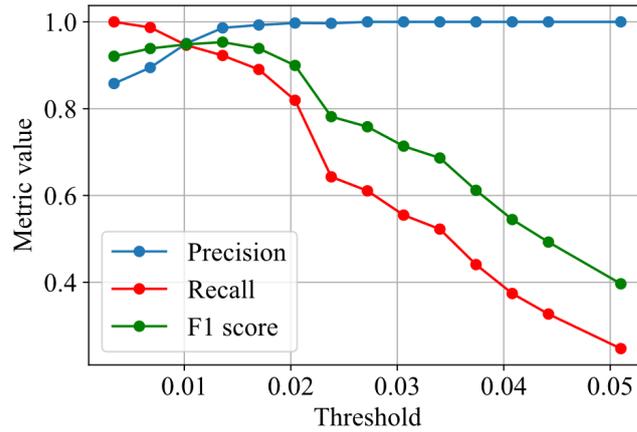


Figure 4.3: The optimal threshold selection based on the precision, recall and F1 scores yielded by SCVAE [242].

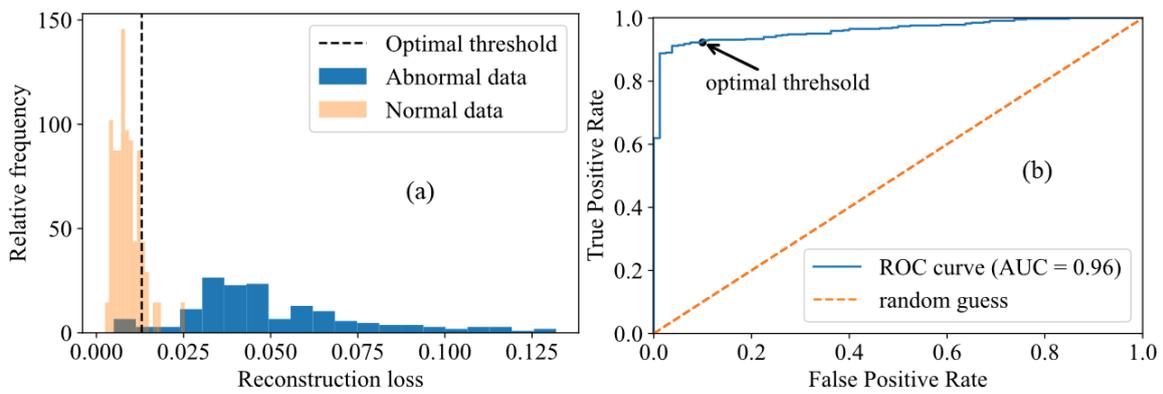


Figure 4.4: (a) Representation of the anomaly score distributions, (b) Receiver operating characteristic (ROC) curve achieved by the SCVAE [242].

domly selected normal and degraded VCSELs that failed after 5,000 hours is taken into consideration to assess the early degradation prediction power of the SCVAE model. As shown in Fig. 4.5, the proposed approach can accurately and early predict the degraded or failed devices before reaching the failure criterion or the end of the aging test (i.e., 15,000 hours), demonstrating the usefulness of the SCVAE in early predicting the failed devices and thus reducing the time and costs of conducting the aging tests over a long period of time (the aging time can be shortened by 66.6% while achieving high detection accuracy). Please note that conventional approaches based on a threshold system for laser failure detection fail to predict degraded devices at 5,000 hours because the decrease in output power for the considered devices at that time does not meet the failure criterion, whereas the ML model can predict failed devices earlier. Consequently, it is beneficial to use an ML model to predict degradation early.

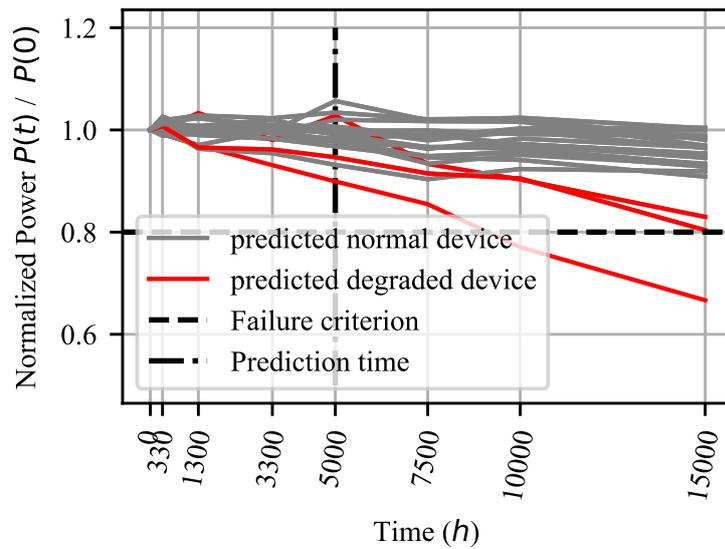
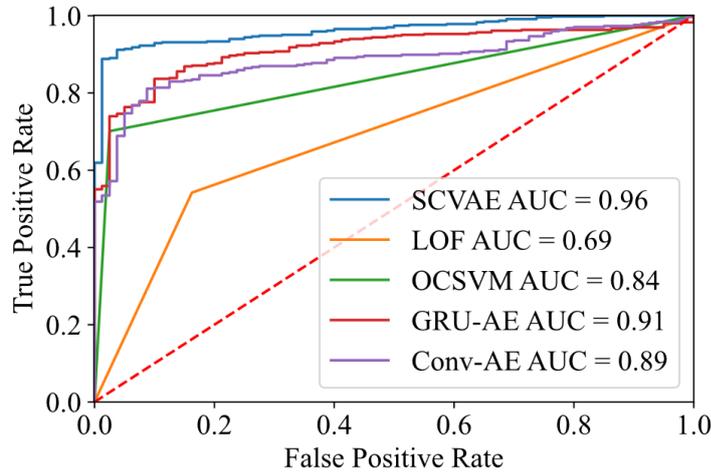


Figure 4.5: Assessment of early degradation prediction capability of the SCVAE model [242].

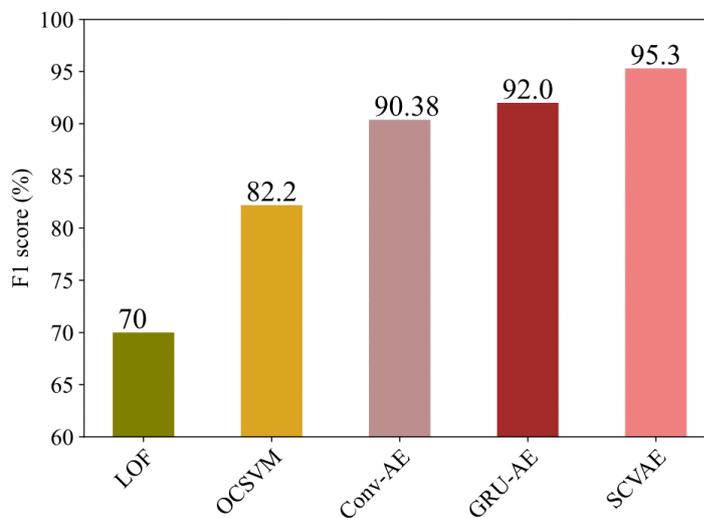
#### 4.1.4.2 Comparison of SCVAE with other ML models

The proposed SCVAE model is compared with the following anomaly detection-based ML algorithms: One-Class Support Vector machine (OCSVM) [263], Local Outlier Factor (LOF) [264], a standard GRU-based autoencoder (GRU-AE), and a convolutional based autoencoder (Conv-AE). The hyperparameters of the various ML models mentioned above are optimized. OCSVM has a linear kernel with a coefficient of 0.1 and a parameter of 0.6 that represents an upper bound on the fraction of training errors and a lower bound on the fraction of support vectors. The tuned hyperparameters of LOF are a neighborhood size of 100 (which defines the neighborhood for the computation of local density) and a contamination parameter of 0.5 (which specifies the proportion of points to be labeled as anomalies). The standard AE is made up of an encoder and a decoder sub-model with four layers, where the encoder has two GRU layers with 40 and 20 cells, respectively, and the decoder's architecture is inversely symmetric to the encoder's structure. The convolutional autoencoder architecture consists of a 7-layer encoder and decoder sub-model, with the encoder composed of a series of 3 convolutional layers containing 32, 16, 32 filters of size 31 with a stride of 2, 1, 1, respectively. The results depicted in Fig. 4.6 prove that the proposed model outperforms the other ML approaches by achieving the highest F1 and AUC values. In comparison, SCVAE achieves significant improvements in AUC and F1 scores of more than 5% and 3.3%, respectively. In comparison to the considered baseline (conv-AE, a recently presented model for laser anomaly detection [265]), the proposed approach improves prediction

capability by 7% and 4.98% in AUC and F1 score metrics, respectively. The comparison results of the deterministic autoencoders GRU-AE and Conv-AE show that GRU-AE outperforms Conv-AE in terms of AUC and F1 score, owing to the fact that GRU is better at processing sequential data and capturing relevant features than the convolutional layers.



(a)



(b)

Figure 4.6: Comparison of different ML methods in terms of: (a) area under curve (AUC) score, and (b) F1 score [242].

#### 4.1.4.3 Degradation Type Classification

The capability of SCVAE in categorizing the type of the degradation is optimized by choosing an optimal threshold  $\beta$ . The precision, recall and F1 score curves as function

of  $\beta$  are depicted in Fig. 4.7. If  $\beta$  is set too high, many sudden degradation samples will be classified as gradual degradation, resulting in a low recall score. If  $\beta$  is too low, several gradual degradation failures will be labeled as sudden degradation failure modes, leading to a low precision score. As a result, the threshold  $\beta$  that maximizes the F1 score is chosen as the optimal threshold. The precision, recall, and F1 score for the optimally chosen threshold of 0.1 are 96.9%, 94.2%, and 95.5%, respectively. Figure 4.8 illustrates that the type of degradation (i.e., sudden or gradual) can be determined based on the anomaly score (i.e., the reconstruction error) obtained by the SCVAE. Based on the optimally selected threshold  $\beta$  of 0.1, the anomalous samples (those with anomaly scores higher than  $\alpha$ ) can be categorized as "gradual" or "sudden".

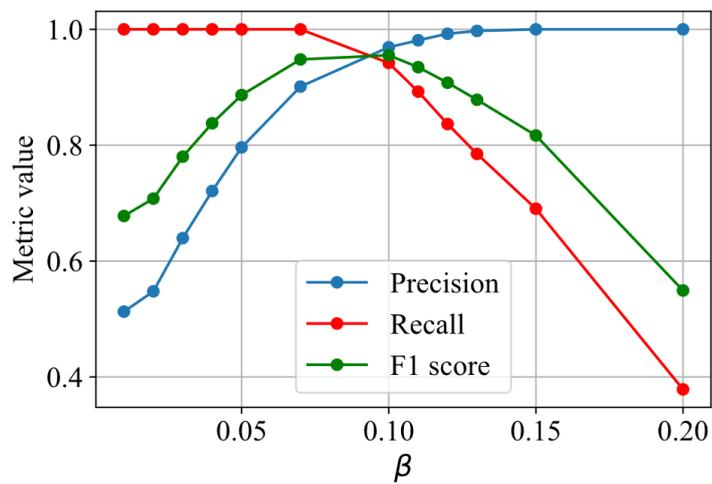


Figure 4.7: The optimal threshold selection for degradation type classification based on the precision, recall and F1 scores yielded by SCVAE [242].

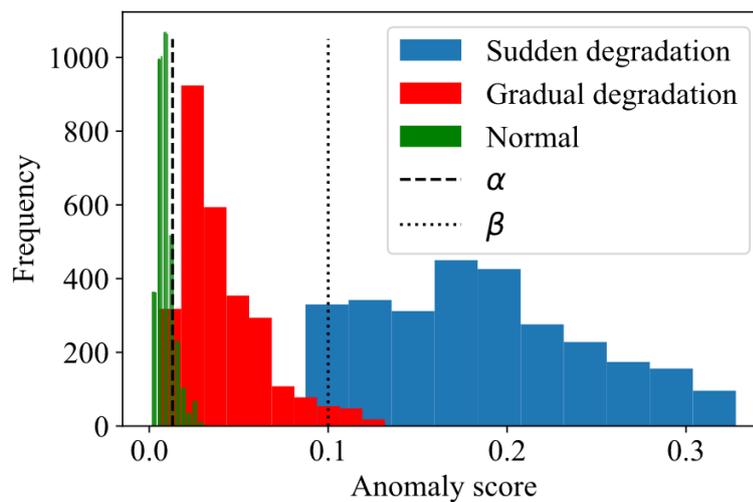


Figure 4.8: Degradation type classification according to the anomaly score [242].

## 4.2 Degradation Type Identification Use Case

The goal of this use case is to identify the type of degradation (i.e. failure mode) that lasers deployed in optical networks may exhibit during operation as early as possible. The previous section demonstrated how to identify the type of degradation by using a two-threshold system optimized based on the reconstruction errors obtained by the autoencoder model (unsupervised learning method). Please note that the optimally chosen thresholds  $\alpha$  and  $\beta$  must be optimized for each of the operating conditions (e.g. normal operating conditions) that differ from the ones used to select the thresholds (stressed conditions), using representative data that includes device failures observed under those operating conditions, because the patterns of the different degradation types under such operating conditions may differ slightly from the patterns of the failure models under the operating conditions used to select  $\alpha$  and  $\beta$ . However, optimizing thresholds for each of the operational circumstances can be a time-consuming and inefficient process. That is why, in this section, a different approach (a supervised learning method) is presented to solving the learning task "early identification of laser degradation types" under different operating conditions. A meaningful data set with an equal and sufficient number of samples or examples for each type of degradation observed under various operational conditions must be available to train such an approach. The content of this section has already been published in [266].

### 4.2.1 Model Architecture

The laser failure mode identification task is formulated as a multi-classification problem, with LSTM as the algorithm of choice because it is well-suited for processing sequential data (time series data) and capturing relevant features.

Figure 4.9 shows the architecture of the proposed approach for early identifying the type of degradation. The input of the ML model is composed of the historical current measurements ( $[I_1 \dots I_{100}]$ ), combined with the laser parameters influencing the degradation, namely the current threshold  $I_{th}$ , temperature  $T$ , optical power  $P$ , and the wavelength  $\lambda$ . The input is fed to 2 LSTM layers composed of 32, and 16 cells, respectively. The relevant features extracted by the LSTM layers are then fed to the output layer that outputs the type of the laser degradation (0: normal, 1: gradual degradation, 2: rapid degradation and 3: sudden degradation). As the loss function, a categorical cross-entropy function is used to update the weights of the LSTM model based on the difference between the predicted and desired output.

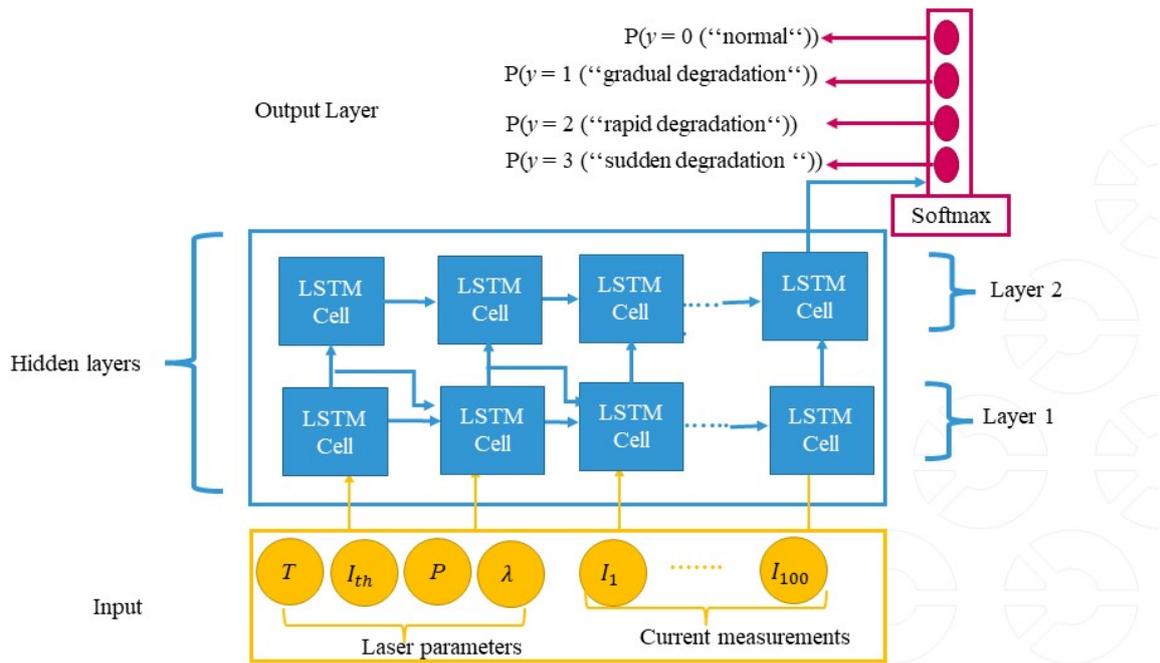


Figure 4.9: Architecture of the proposed ML model for early laser degradation type identification [266].

#### 4.2.2 Synthetic Data Generation

In the absence of real field data to train the ML-algorithm, synthetic data is generated, whereby an analytic model is adopted to simulate the patterns of laser failure modes, including gradual, rapid, and sudden degradation, along with normal laser operation. Equation (4.3) models the variation of the operational current as a function of the time under constant power [267, 266].

$$I(t) = I_0 + \beta \exp\left(P^n \exp\left(\mu_0 - \frac{E_a}{k_B T}\right) t\right) \quad (4.3)$$

where the parameter  $n$  represents the derating exponent,  $E_A$  denotes the activation energy,  $\mu_0$  is the scale parameter,  $\beta$  denotes the non-radiative current, and  $T$  is the temperature.

The input features, which include the optical power  $P$ , the threshold current  $I_0$  and the temperature  $T$ , are extracted from real laser datasheets specifications [268], whereas the underlying coefficients used to create the various degradation patterns are derived using normal distributions.

Approximately 1,500 samples are generated for each type of laser degradation mode, as well as for normal laser behavior. In total, a dataset composed of 6,000 samples modeling the variation of the current as a function of the time under various operating

conditions is built for the development of the laser degradation type identification model. The synthetically generated dataset represents the laser degradation modes at three different time scales: rapid degradation can be observed within the first 100 hours of operation, gradual degradation can last for many hundreds of hours, and catastrophic degradation can occur after many hours of normal operation. The aforementioned data is divided into sequences of length 100. The gradual degradation sequences are averaged to create a compact 100 sequence that represents this failure mode. Figure 4.10 shows the different laser failure modes patterns after preprocessing, where the  $x$ -axis indicates different time indices.

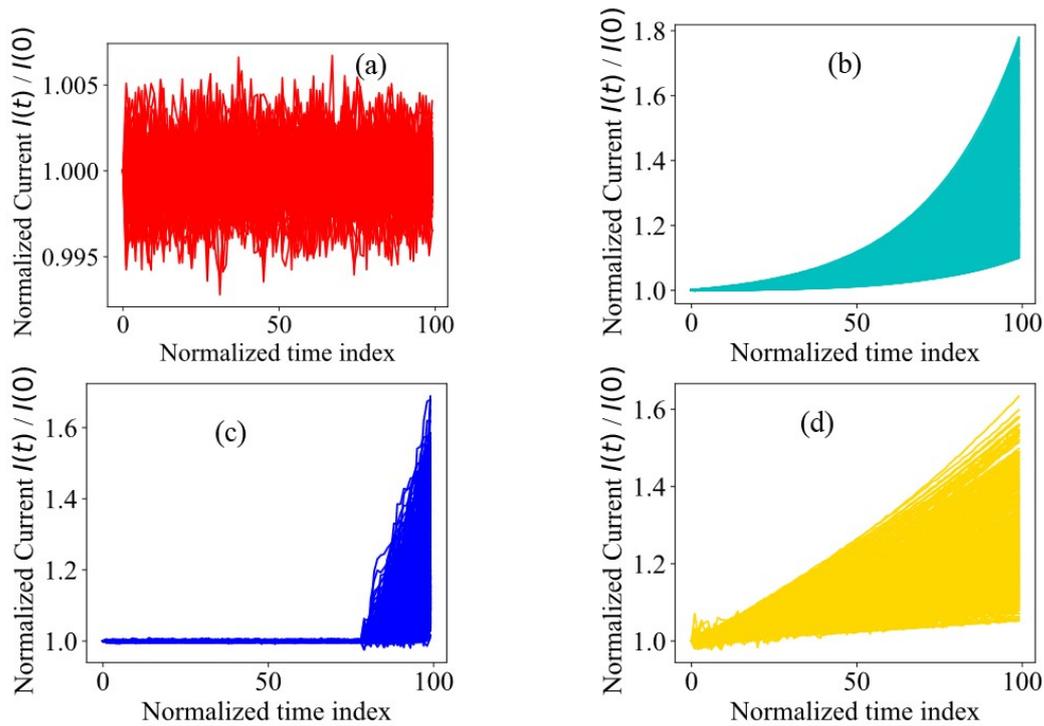


Figure 4.10: Laser failure mode patterns: (a) normal, (b) gradual degradation, (c) sudden degradation, and (d) rapid degradation [266].

After preprocessing, the data is split into a 60% training dataset, a 20% validation dataset and a 20% test dataset. To simulate real field data, the test dataset is altered so that only 20 - 40% of the fault pattern is maintained for each failure mode sample, and prepended with a normal laser operation sequence representing 60% - 80% of the observation.

### 4.2.3 Results

The performance of the implemented LSTM model is compared with several classical ML algorithms namely random forest (RF), KNN and multinomial linear regression

ML Method	Accuracy(%)	Recall (%)	Precision (%)	F1 score (%)
KNN	67.1	67	67	64
LR	89.5	90	91	89
RF	88	88	89	88
LSTM (proposed)	99	99	99	99

Table 4.1: Performance comparison of the different ML models.

(LR), in terms of various metrics including the accuracy, the recall, the precision, and the F1 score. The LSTM model outperforms the other ML methods in terms of all evaluation metrics, as shown in Table 4.1, by achieving the highest values because LSTM is better at processing sequential data and capturing long term dependencies. The confusion matrix shown in Fig. 4.11 (a) demonstrates that the LSTM model classifies the various degradation types and normal behavior with a very high accuracy (more than 98%). The ROC curve shown in Fig. 4.11 (b) proves that the ML model accurately discriminates between the different types of classes by achieving a very high degree of separability (AUC higher than 99%).

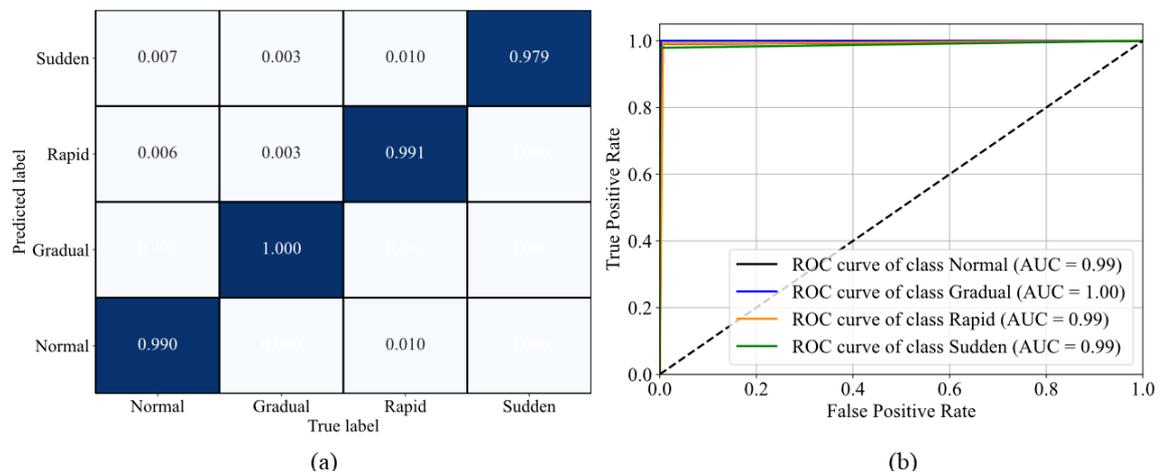


Figure 4.11: Performance evaluation of the ML model for laser degradation identification [266].

The performance of the LSTM-based degradation type identification model is compared with a conventional method, a multi-threshold system that assigns different thresholds ranging from 20% to 80% of the current increase based on failures, EOL (end of life) criteria, and laser specification and design. As shown in Table 4.2, the LSTM model outperforms the threshold system in terms of fault accuracy, which is defined as the accurate detection of the degradation or failure, when tested on two unseen test data: one data containing partial test failure patterns and another data incorporating full patterns. The LSTM model accurately detects the laser failure by achieving a very

Data	ML Method	Conventional method
Partial pattern data	99%	24.41%
Full pattern data	99.6%	79%

Table 4.2: Fault accuracy achieved by the ML model and the conventional method for different unseen test data sets.

high fault accuracy ( $\geq 99\%$ ) for both test data sets. Tested with the partial pattern data, the obtained fault accuracy of the conventional method (24.4%) is very low as the threshold based system could correctly classify only the normal behavior while failing to detect the different degradation types as the increase of the current of the partial patterns incorporating the different failure modes is below the failure criterion threshold.

### 4.3 Laser Lifetime Prediction Use Case

Previous sections have discussed the application of ML techniques to tackle laser diagnostics problems. The following sections go over how to use ML methods to solve laser prognostics tasks. This section will cover the use case, "laser MTTF prediction using ML". The goal of this use case is to predict the MTTF of a laser during operation based on various monitored laser parameters by fully utilizing the power of ML methods in efficiently learning or modeling the dependency between the laser lifetime and the impact of various laser characteristics on its reliability. The content of this section has already been published in [269].

#### 4.3.1 Model Architecture

The architecture of the proposed method for laser MTTF prediction is illustrated in Fig. 4.12. The ML approach based on artificial neural networks (ANNs) takes as input several monitored laser characteristic parameters namely the optical power ( $P_{op}$ ), the current threshold ( $I_{th}$ ), the conversion efficiency ( $\eta$ ), the slope efficiency ( $SE$ ), the voltage ( $V$ ), the wavelength ( $\lambda$ ), and the junction temperature ( $T_j$ ), and outputs the MTTF. The ANN model is made of two hidden layers with 200 and 100 neurons, respectively. A mean square error cross-entropy function is adopted as the loss function to update the weights of the model based on the error between the predicted and the desired output.

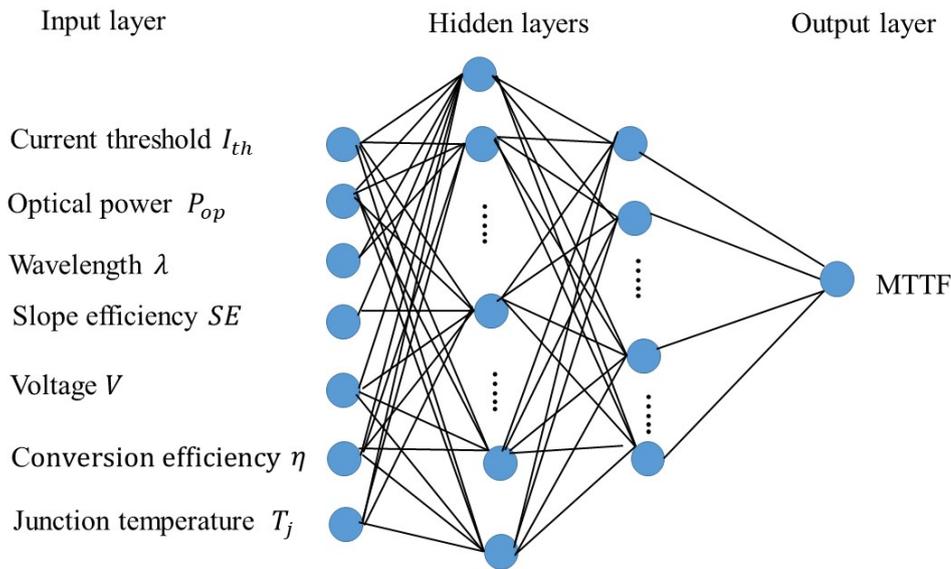


Figure 4.12: Architecture of the proposed ML model for laser lifetime prediction.

### 4.3.2 Synthetic Data Generation

Please note that normally, the ML model would be trained using experimentally derived MTTF values, however, due to a lack of a large dataset, an analytical approach for synthetically generating MTTF values is used.

Synthetic data for low power ( $P_{op} \leq 10$  mW) InGaAsP MQW-DFB lasers operating at a  $\lambda$  range from 1.53 to 1.57  $\mu\text{m}$  with the case temperature  $T_c$  range of  $-40$   $^{\circ}\text{C}$  to  $85$   $^{\circ}\text{C}$  with side mode suppression ratio SMSR of greater than 35 dB is generated. The different laser electro-optical characteristics namely  $I_{th}$ ,  $SE$ ,  $V$  and  $\lambda$  are modeled using a real laser datasheet [270]. Figure 4.13 depicts the performance curves of the modeled parameters.

The different laser parameters namely  $I_{th}$ ,  $SE$ ,  $V$ ,  $\lambda$ ,  $\eta$ , and junction temperature  $T_j$  are computed for given values of  $T_c$  and  $P_{op}$ , randomly chosen from uniform distributions. The  $MTTF_2$  of the produced parameters is determined using a model for two stress terms namely temperature and the forward current. It is expressed as:

$$MTTF_2 = MTTF_1 \left( \frac{I_1}{I_2} \right)^2 \exp \left( \frac{E_a}{k_B} \left( \frac{1}{T_{j,2}} - \frac{1}{T_{j,1}} \right) \right) \quad (4.4)$$

where  $E_a$  denotes the activation energy for the device in units of eV,  $k_B$  is Boltzmann's constant,  $T_{(j,1)}$  and  $T_{(j,2)}$  are different junction temperatures in units of Kelvin,  $I_1$  and  $I_2$  represent the corresponding operating currents at  $T_{(j,1)}$  and  $T_{(j,2)}$ , respectively. The aforementioned parameters are assumed to be fixed, and the experimentally deter-

mined MTTF, denoted by  $MTTF_1$  ( $7 \times 10^5$  hours), based on accelerated aging test under stressed conditions ( $P_{op}=10$  mW,  $T_c = 50$  °C), is used. The process of the data generation described above is illustrated in Fig.4.14.

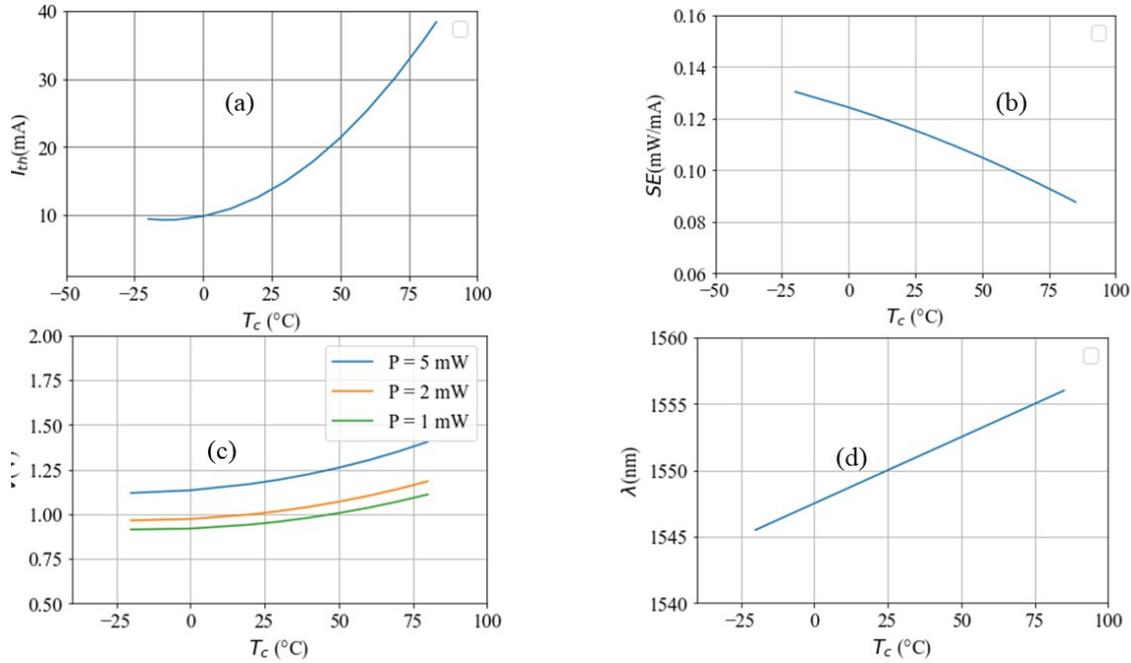


Figure 4.13: a) Current threshold  $I_{th}$  versus case temperature  $T_c$ , b) Slope efficiency  $SE$  versus case temperature  $T_c$ , c) Voltage  $V$  versus temperature  $T_c$ ,  $P_{op}$ , and d) Wavelength  $\lambda$  versus temperature  $T_c$  [269].

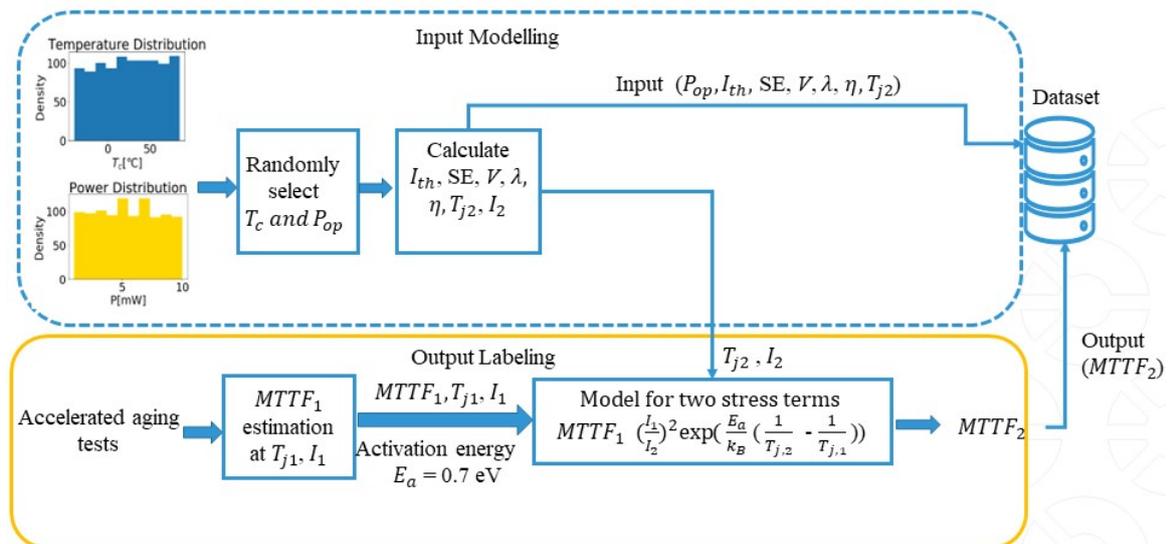


Figure 4.14: Synthetic Laser Reliability Dataset Generation Process [269].

### 4.3.3 Conventional Lifetime Prediction Method

Accelerated aging tests for twenty-five 1.55  $\mu\text{m}$  DFB lasers carried out for 5,000 hours under constant output optical power of 10 mW at 70 °C were performed. The cumulative percent failures versus the failure times of the devices estimated by linear extrapolating the change of  $I_{op}$  up to 50% are plotted on a lognormal probability scale yielding  $1.8 \times 10^5$  hours for median lifetime  $t_m$ , defined as the point of the fitted line where 50% of devices failed the test, and 0.4 for the standard deviation  $\sigma$  computed as  $\ln(t_m/t_1)$  where  $t_1$  denotes the time corresponding to a cumulative failure of 16% [269]. With the obtained  $t_m$  and  $\sigma$ , the  $MTTF_1$  is calculated as:

$$MTTF_1 = t_m \exp\left(\frac{\sigma^2}{2}\right) \quad (4.5)$$

Given a  $MTTF_1$  of  $1.9 \times 10^5$  hours, the  $MTTF_2$  of the same device at a different junction temperature  $T_{j,2}$  can be determined using the Arrhenius model.

### 4.3.4 Results

The accuracy of the model prediction is evaluated by adopting two regression evaluation metrics namely RMSE and the value of scoring function  $S$ , which can be expressed as function of  $h_i$ , representing the difference between the  $i$ th predicted value and  $i$ th true value, as:

$$S = \begin{cases} \sum_{i=1}^N \left( \exp\left(-\frac{h_i}{13}\right) - 1 \right) & \text{for } h_i < 0 \\ \sum_{i=1}^N \left( \exp\left(\frac{h_i}{10}\right) - 1 \right) & \text{for } h_i \geq 0 \end{cases} \quad (4.6)$$

The scoring function assess the estimation error by penalizing overestimated values more than underestimated values, as overestimated values fail to predict device failure, resulting in higher costs.

The proposed model's prediction accuracy is evaluated using previously unseen test data. The ANN model achieves a small RMSE of 0.37 years and a low scoring function value of 19.23. The histogram depicted in Fig. 4.15, which illustrates the distribution of the prediction errors yielded by the ANN model, shows that the proposed approach can overestimate the MTTF by up to 0.9 years and underestimate it by up to 1.1 years. To compare the performance of the ANN model with the conventional laser lifetime prediction method, a test dataset, including different laser characteristics estimated under similar operating conditions like the accelerated aging tests (same optical power  $P_{op}=10$  mW, different temperatures), is produced and adopted. Table 4.3 shows that the ANN model outperforms the conventional approach (CM) by achieving lower RMSE and scoring function values. The prediction error histograms shown in Fig. 4.16

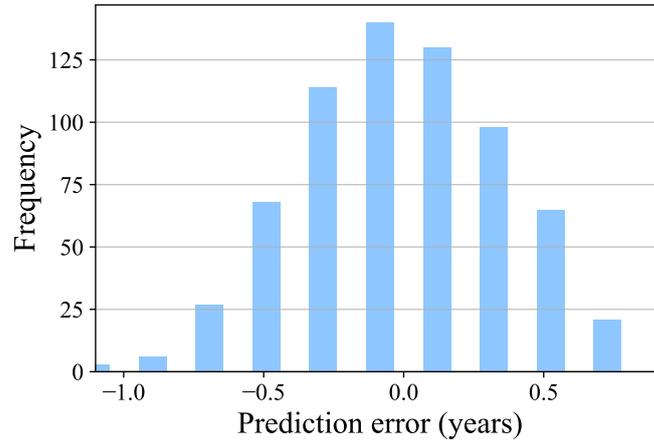


Figure 4.15: Histogram of laser lifetime prediction errors yielded by the proposed approach.

demonstrate that the ANN model accurately predicts the MTTF with small prediction errors of less than 0.8 years, whereas the conventional method overestimates the MTTF (large prediction errors up to 6.7 years).

Evaluation method	ANN	Conventional method
RMSE [years]	0.4	5.1
Scoring function	6.71	137.2

Table 4.3: Comparison of ANN model with conventional method.

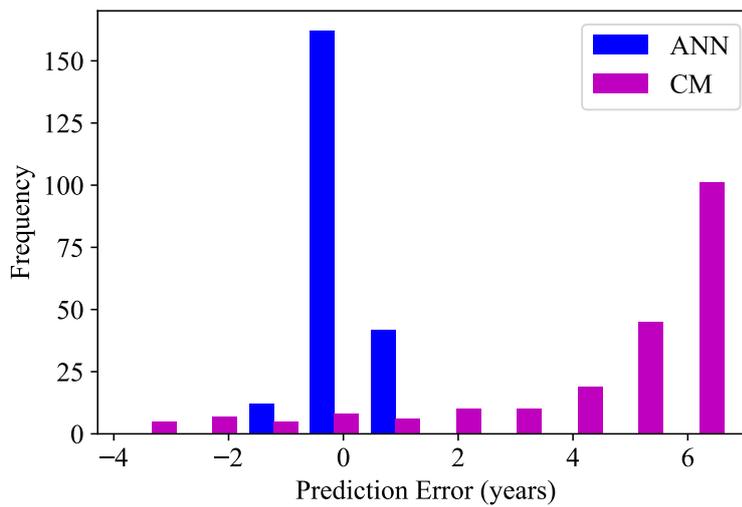


Figure 4.16: Predictions errors achieved by ANN model and conventional method.

## 4.4 Remaining Useful Life Prediction Use Case

In the previous section, an ML-based approach for MTTF prediction using only the monitored laser characteristics, was presented. However, the degradation trend over time, which influences MTTF estimation, is not considered as a feature for the ML model, which may have an adverse effect on the ML model's prediction accuracy. As a result, along with the laser characteristics, information about the variation of the degradation parameters (e.g., current, output power) as a function of time must be provided as an additional input to build a reliable laser lifetime ML model. Therefore, in this section, an ML model for predicting the remaining useful life (RUL) of semiconductor lasers during operation by taking as input the historical output power measurements, modelling the degradation trend over time, as well as operating conditions and laser characteristics, is presented to enhance the prediction accuracy. The content of this section has already published in [271].

### 4.4.1 Model Architecture

As illustrated in Fig. 4.17, the proposed framework is composed of two parallel models (one based on CNN and one on LSTM), which are followed by a fully connected neural network that combines the outputs of each model to perform the RUL estimation regression task. As CNN is good at extracting local spatial features, the operating conditions are fed into the CNN-based model, which consists of three stacked convolution layers with 32, 32, and 16 filters, respectively, followed by max pooling layers. Because LSTM is suitable for capturing dependency in sequential data, the output power measurement sequence is provided to the LSTM-based model, which is composed of three stacked LSTM layers made up of 40, 20, and 10 memory cells, respectively. The CNN model's two-dimensional output is flattened into a one-dimensional vector before being combined with the LSTM model's output. The fused output is then fed into the neural network layer, which consists of 64 neurons and outputs the predicted RUL.

### 4.4.2 Experimental Data

The proposed ML approach is validated using experimental data derived from different accelerated aging tests for VCSEL samples with various oxide aperture sizes performed under several controlled operating conditions, namely the current  $I$ , the junction temperature  $T_j$ , the current density  $J$  and the resistance  $R$ . The aging tests are carried out up to 3,500 hours. The temperature  $T$  is varied from 85°C to 150°C in order to significantly increase laser degradation and thus accelerate device failure. The optical output power is monitored periodically under constant current operation regime. The

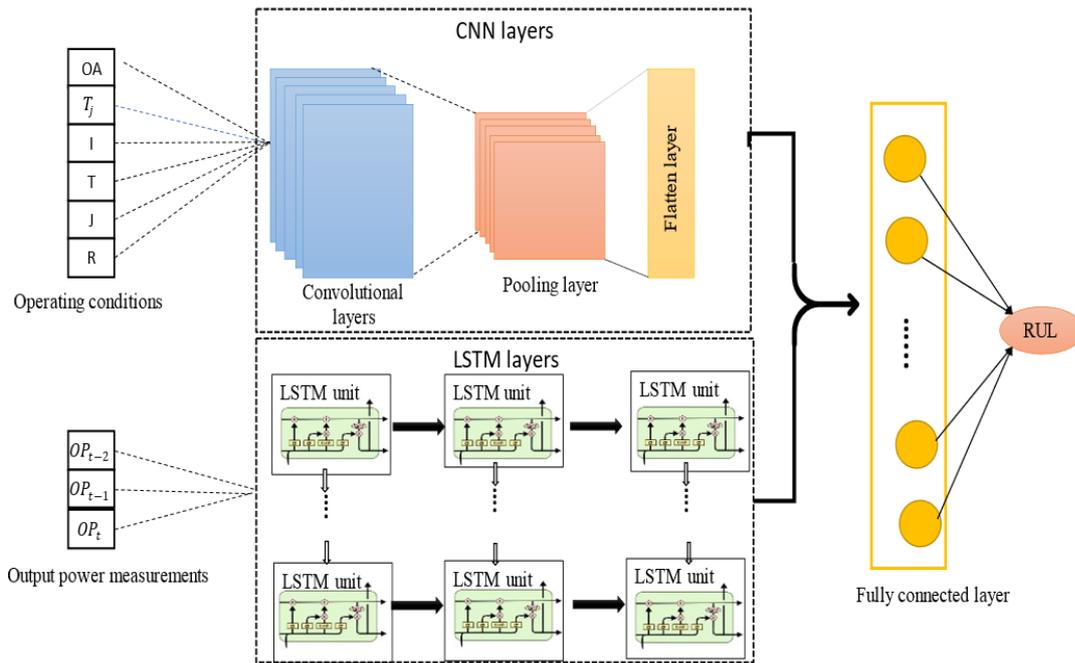


Figure 4.17: Architecture of the proposed model for laser RUL prediction [271].

device's time to failure  $t_f$  is defined as the time at which the output power has dropped by 1 dB (20%) from its initial value. The output power measurements of the various devices are recorded from the start of the aging test until the device reaches  $t_f$  or the end of the test, if the device does not fail during the test. In total, 239 output power time series are produced. The time series are split with a sliding window of size 3 in order to increase the amount of data required to train the machine learning model. The obtained sequences are then combined with the operating conditions including the oxide aperture size ( $OA$ ),  $T_j$ ,  $I$ ,  $T$ ,  $J$  and  $R$ . Each sequence's RUL is calculated as the difference between  $t_f$  and the time  $t$  at which the RUL is predicted. The ML model can only be trained based on the measurement sequences recorded before the failure occurred up to a maximum of 5,000 hours since it is difficult to predict the RUL over a long period of time. In total, a data set of 729 samples is built, normalized, and split into a training (comprising of 80% of the sequences) and a test dataset (the remaining 20%).

#### 4.4.3 Results

The accuracy of the ML model RUL estimation is evaluated by adopting several evaluation metrics namely RMSE, MAE, and the scoring function  $S$ , which can be expressed

as follows:

$$S = \begin{cases} \sum_{i=1}^N \exp\left(-\frac{RUL_{pred(i)} - RUL_i}{a_1} - 1\right) & \text{for } RUL_{pred(i)} < RUL_i \\ \sum_{i=1}^N \exp\left(-\frac{RUL_{pred(i)} - RUL_i}{a_2} - 1\right) & \text{for } RUL_{pred(i)} \geq RUL_i \end{cases} \quad (4.7)$$

where  $N$  represents the total number of samples in the test data set, and  $RUL_i$  and  $RUL_{pred}$  denote the true RUL and the predicted RUL for the test sample  $i$ , respectively. The parameters  $a_1$  and  $a_2$  are user-defined parameters, which control the asymmetric preference for underestimated predictions over overestimated predictions.  $a_1$  and  $a_2$  are set to 250 and 220, respectively. The RMSE and MAE metrics are adopted to assess the closeness between the predicted RULs and the true RULs by equally penalizing underestimated and overestimated predictions. Whereas the scoring metric severely penalizes cases where the predicted RUL is greater than the actual RUL because overestimated values may result in device failure and higher costs. The performance of the proposed approach CNN+LSTM is compared with different other ML algorithms including SVR, RF, MLP, CNN, and LSTM. As depicted in Fig. 4.18, the proposed model outperforms the aforementioned ML methods in terms of all metrics, achieving the lowest RMSE, MAE, and scoring values. It achieves a notable improvement over the different ML approaches particularly SVR and RF, by improving the scoring metric by more than 26.8% and yielding significant improvements in the RMSE and MAE metrics by more than 11.5% and 16%, respectively.

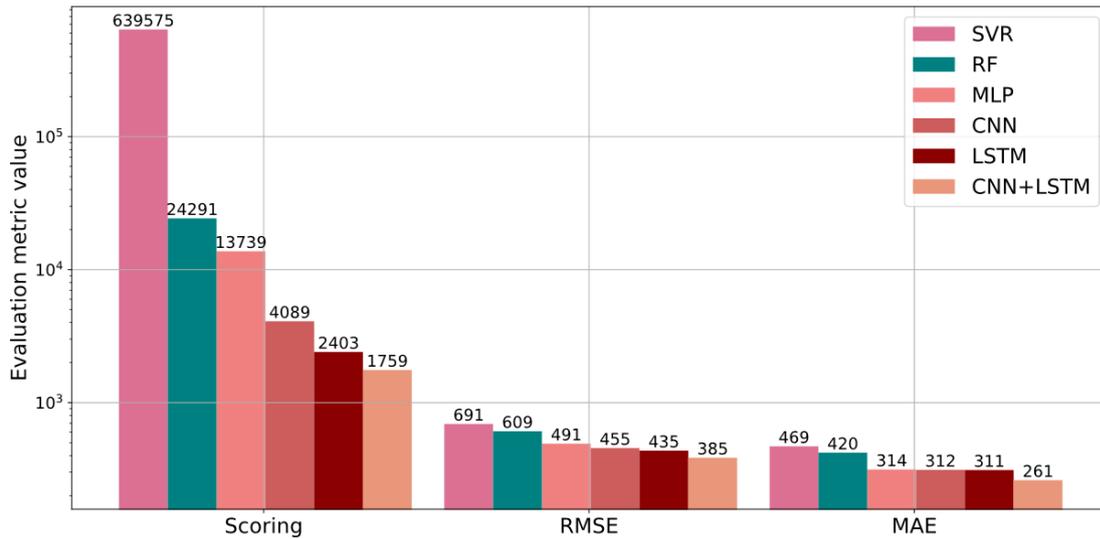


Figure 4.18: Comparison of the results of the proposed model with other methods [271].

The performance of the proposed approach is compared with a conventional laser RUL estimation technique based on a linear least-squares-fitting (LLSF) method in terms of

prediction error. The comparison of the results shown in Fig. 4.19 proves that the proposed ML model significantly outperforms the conventional method by achieving lower prediction errors. The LLSF technique significantly underestimates or overestimates the RUL, whereas CNN+LSTM estimates are more accurate.

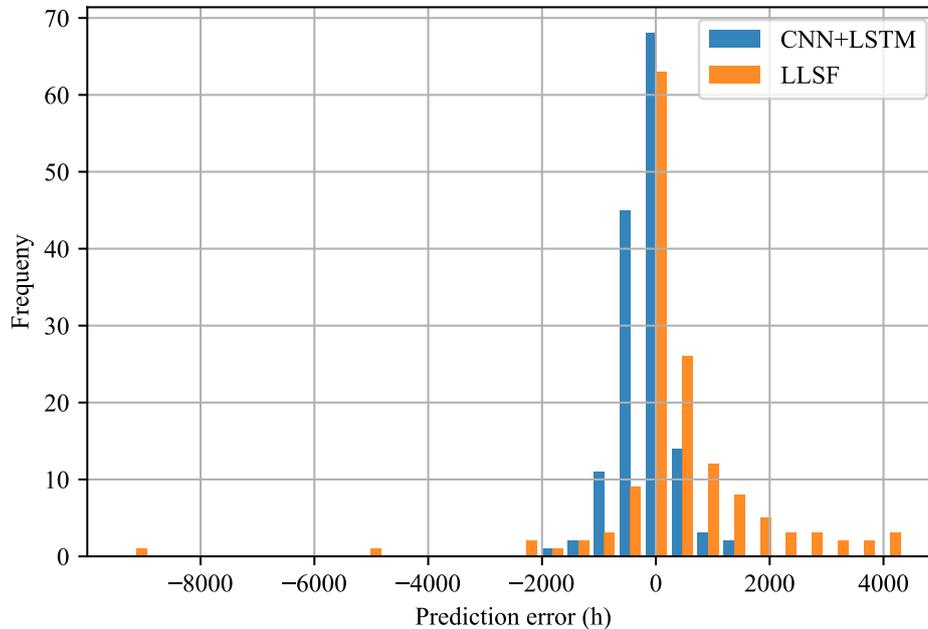


Figure 4.19: Histogram of prediction errors of CNN-LSTM and LLSF [271].

#### 4.4.4 Discussion

The developed ML model is effective at predicting the RUL of semiconductor lasers that degrade gradually. However, predicting the RUL for devices exhibiting sudden degradation is extremely difficult because the degradation mode begins close to the failure time and accelerates rapidly. Therefore, it is beneficial to implement a dedicated ML model for estimating the RUL in case of sudden degradation failures. Hence, two ML models have to be adopted to predict the RUL of lasers dependently on the type of degradation or failure mode predicted (sudden or gradual degradation) as illustrated in Fig. 4.20. The synthetic data presented in Subsection 4.2.2 is adopted to validate the proposed framework shown in Fig. 4.20, which has been published in [272]. The adopted ML model for laser degradation type identification is the LSTM model discussed in Subsection 4.2.1. The architecture of the ML-based RUL prediction models developed for both sudden and gradual degradation is the same, a stacked-LSTM architecture with two hidden layers as shown in Fig. 4.21. Each ML-based RUL prediction model is trained by a dataset containing multiple run-to-failure data sequences

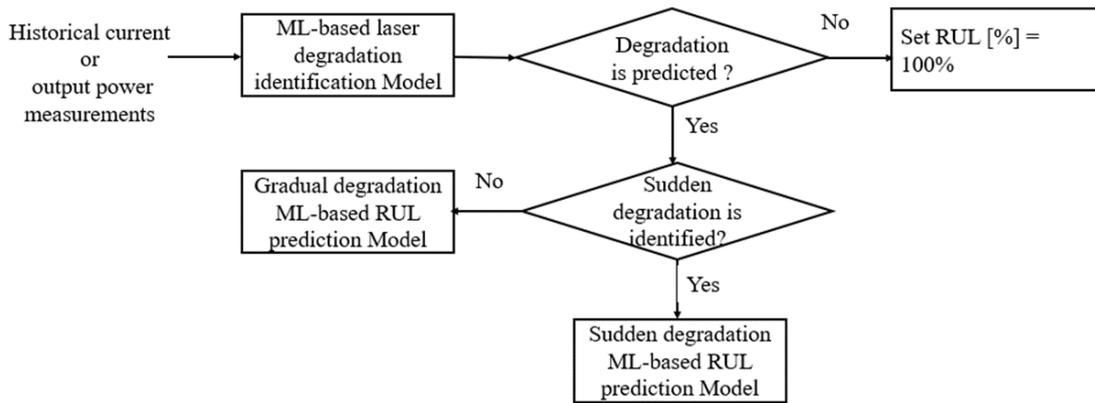


Figure 4.20: Laser RUL prediction approach [272].

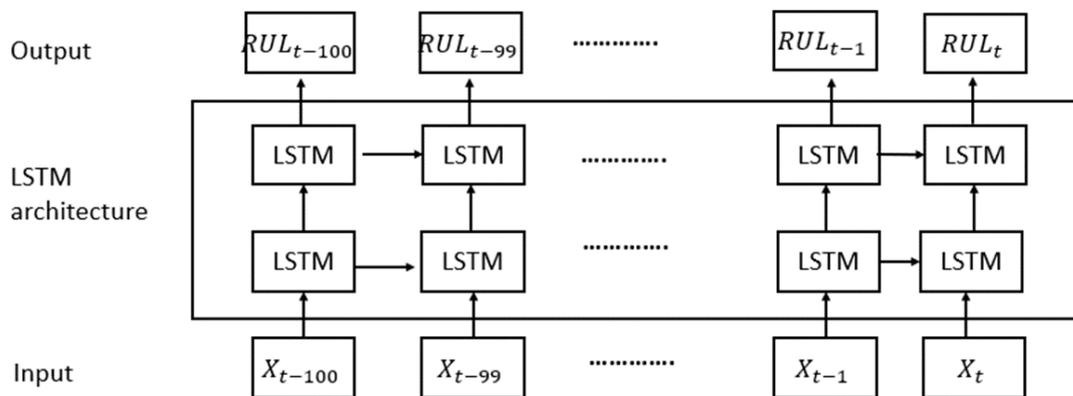


Figure 4.21: LSTM-based RUL prediction architecture [272].

(i.e., univariate time series in form of  $X = [X_{(t-N)}, X_{(t-N-1)}, \dots, X_{(t-1)}, X_{(t)}]$  where each data sample is a vector containing the current value at that sample time as well as the operating conditions). For each data sample, the LSTM model predicts the RUL at that sample time. A piece-wise labelling approach assuming that the laser performance begins to degrade linearly at some point  $\tau$  is adopted to compute the RUL label for each data sample. Eq. (4.10), where  $t_f$  represents the time to failure of the device, provides the formula used for labeling the RUL in form of normalized life percentage.

$$RUL = \begin{cases} 1, & \forall t \leq \tau \\ \frac{t_f - t}{t_f - \tau}, & \forall t > \tau \end{cases} \quad (4.8)$$

The results show that the LSTM-based RUL prediction model for sudden degradation accurately predicts the RUL with an RMSE of up to 36 minutes. The RUL prediction error for the gradual degradation is estimated to reach up to 131 hours. The plots

shown in Fig. 4.22 demonstrate that the predicted RUL tends to be close to the true piecewise RUL, particularly at the end-of-life stage.

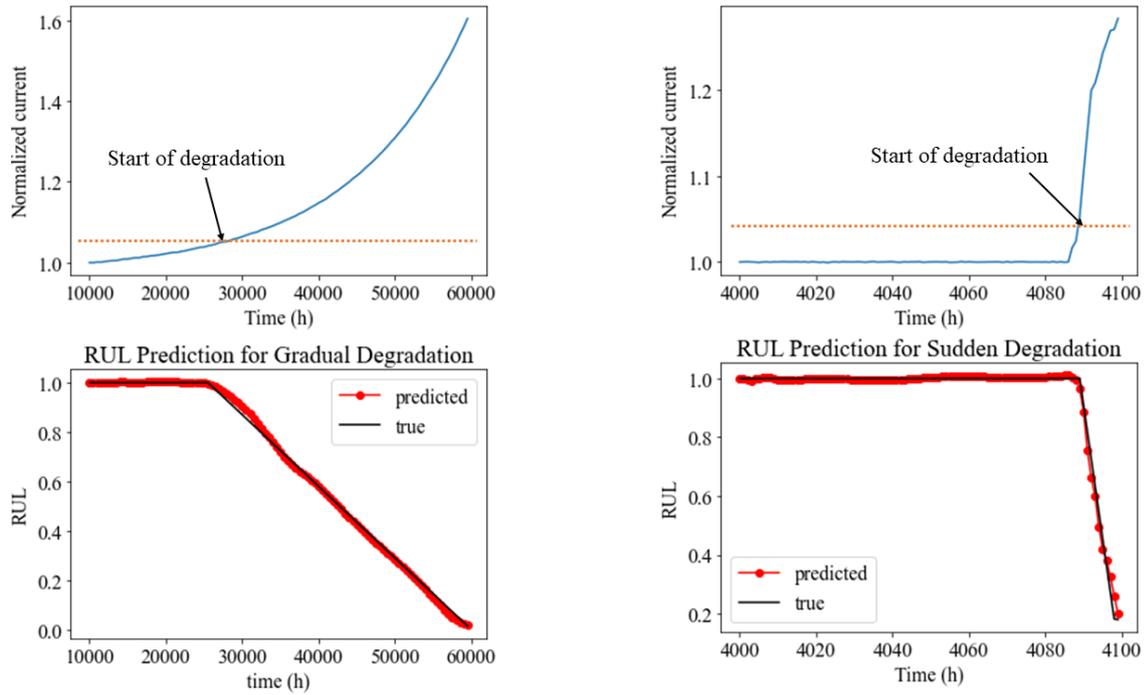


Figure 4.22: The true and the predicted RUL for randomly selected test samples [272].

## 4.5 Predictive Maintenance Framework for Semiconductor Lasers

This section describes a combined framework that incorporates various ML methods for solving the same predictive maintenance tasks (e.g., degradation prediction, RUL prediction, and so on) discussed in previous sections in order to improve laser reliability. The content of this section has already been published in [265].

### 4.5.1 Proposed Framework

The proposed predictive maintenance framework for a semiconductor laser is depicted in Figure 4.23. Following the deployment of the laser device in an optical network, the laser current (i.e. degradation parameter) is monitored on a regular basis under a constant output power operation regime. The laser current measurements that are collected are then saved in a database. Afterwards, the most recent  $k$  monitored current measurements  $I_{(t-k)} \dots I_t$ , are extracted, preprocessed, and fed into an ML model for real-time prediction of the performance degradation trend. The ML model forecasts the next value of the laser current  $I_{(t+1)}$ , which is stored in the database. Finally, the

sequence of current measurements  $I_{(t-k)} \dots I_{(t+1)}$  fed into an anomaly detection model to identify any degradation or abnormal behavior. If a degradation or abnormal behavior is detected, an RUL prediction model is used to determine the RUL of the device, and a notification is sent to the maintenance planning unit for root cause analysis. The maintenance operations are then scheduled based on the predicted RUL.

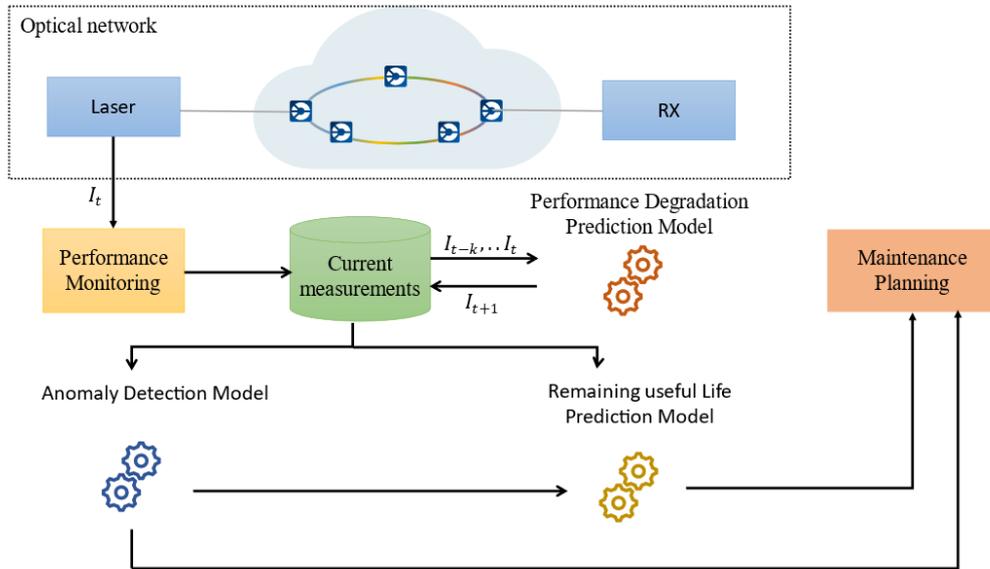


Figure 4.23: Flow chart of the proposed framework for predictive maintenance of semiconductor lasers [265].

#### 4.5.1.1 ML based performance degradation prediction model

The proposed ML model for real-time prediction of performance degradation (i.e., increase in laser current) combines the GRU and the attention mechanism, with the GRU performing a one-step prediction and the attention mechanism assisting the model in focusing more on relevant features to improve prediction accuracy. As shown in Fig. 4.24, the attention based GRU model predicts the next current measurement  $I_{(t+1)}$  given a sequence of 9 historical current measurements  $[I_{(t-8)}, I_{(t-7)}, \dots, I_t]$ . The proposed model has two GRU layers with 64 and 32 cells, respectively, followed by an attention layer, and finally a fully connected layer with no activation function that outputs  $I_{(t+1)}$ . The GRU layers process the sequential input to capture the temporal dependency modeling the degradation trend, and outputs the hidden states  $[h_{(t-8)}, h_{(t-7)}, \dots, h_t]$  (i.e. the learned or extracted features). The attention layer then assigns to each extracted feature  $h_i$  a weight (i.e. attention score)  $\alpha_i$  to generate a context vector  $c_t$  containing the relevant information. The weighted attention features are then fed into a fully connected layer with output  $I_{(t+1)}$ . Using the Adam optimizer, the model

is optimized by minimizing the MSE (i.e. the cost function) between the predicted and true current values.

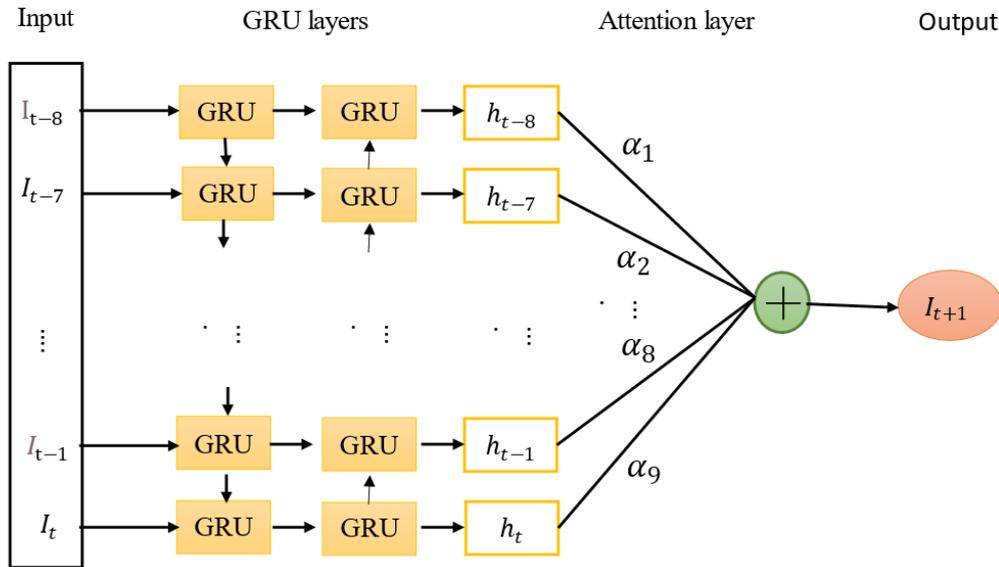


Figure 4.24: Architecture of the proposed attention based GRU model for performance degradation prediction [265].

#### 4.5.1.2 ML based anomaly detection model

The proposed model for detecting laser anomalies is based on a convolutional autoencoder, as shown in Fig. 4.25. The model includes a 7-layer encoder and decoder sub-model. The encoder takes a current measurement sequence of length 10 as input and encodes it into low dimensional features using a series of three convolutional layers each containing 32, 16, 32 filters of size 31 with a stride of 2, 1, 1. respectively. Given the encoder's compressed representation output, the decoder reconstructs the original input. The decoder is made up of four transposed convolutional layers that are used to up-sample the feature maps, with the last transposed convolutional layer using one filter of size 31 and a stride of one to generate the output. The cost function is the MSE, which is adjusted by using Adam optimizer. It should be noted that the model is trained using normal data that represents the laser device's normal state in order to learn the distribution that characterizes normal behavior. Once the model has been trained, the classification of an instance as anomalous/normal is carried out as shown in Fig. 4.26. First, an anomaly score that quantifies the error between the input  $I$  and the reconstructed input  $\hat{I}$  (the output), is computed. MAE is used as an anomaly score in this study. If the calculated anomaly score is greater than a predefined threshold  $\theta$ ,

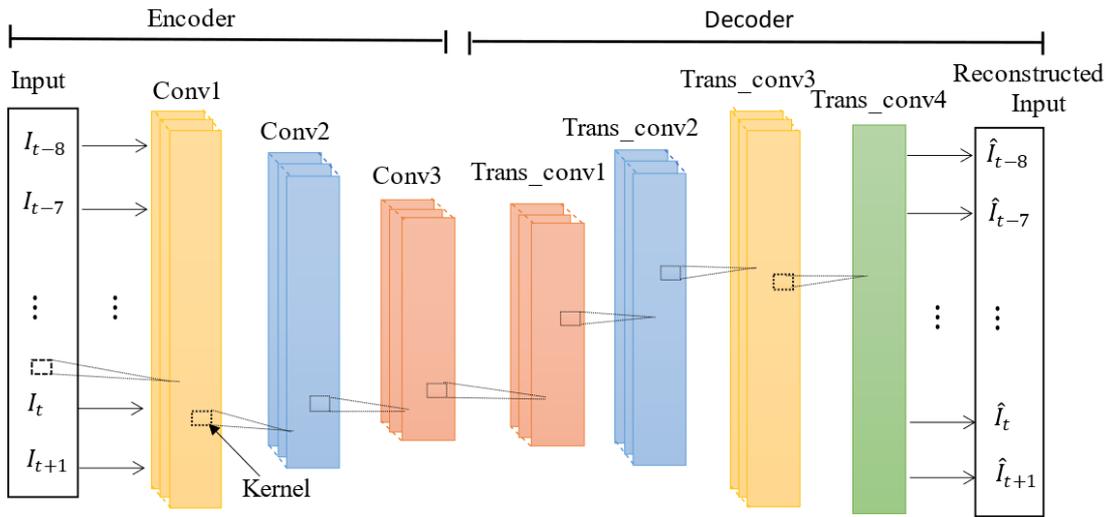


Figure 4.25: Structure of the proposed convolutional autoencoder for laser anomaly detection [265].

the instance is classified as "anomalous," otherwise it is categorized as "normal".  $\theta$  is a hyperparameter that is optimized according to the number of true and false positives.

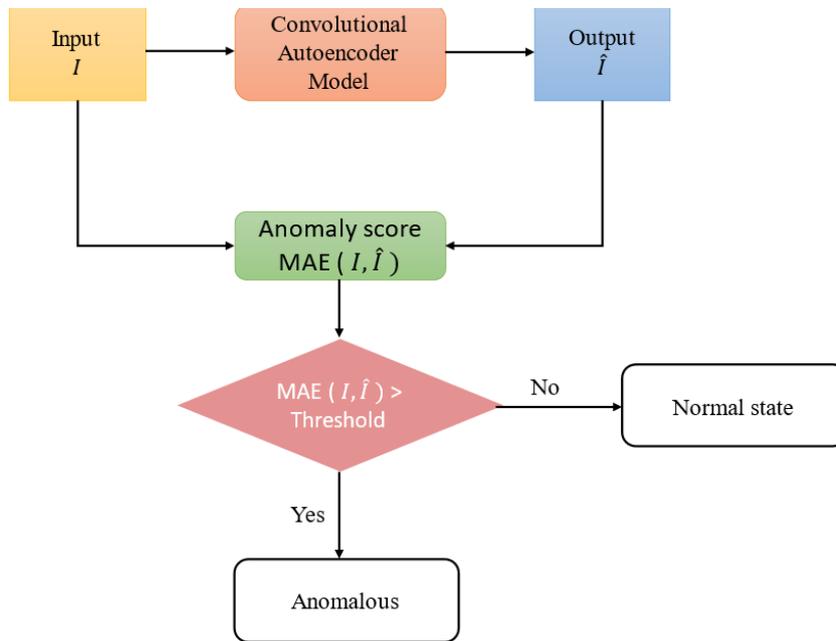


Figure 4.26: Flow chart of the process of instance classification as anomalous or normal [265].

### 4.5.1.3 ML based RUL prediction

The proposed attention-based deep learning model for RUL prediction is depicted in Figure 4.27. In order to improve prediction accuracy, the proposed approach fully utilizes the fusion of the sequential and temporal features learned by the GRU-attention based layers and the statistical features characterizing the degradation trend, namely the root mean square (RMS), kurtosis ( $\beta$ ), and skewness ( $\delta$ ). The sequential input  $[I_{(t-8)}, I_{(t-7)}, \dots, I_{(t+1)}]$  is fed to 2 GRU layers made up of 64 and 32 cells, respectively, whereas the statistical features are given to a fully connected layer composed of 32 neurons. The learned features are then transferred to the attention layer to capture the most important features, which are then merged with the output features of the fully connected layer. The fused features are then fed into a fully connected layer with 32 neurons, followed by a dropout layer to prevent overfitting, and finally the RUL is output. The entire network is simultaneously trained by using an Adam optimizer to minimize the cost function (MSE).

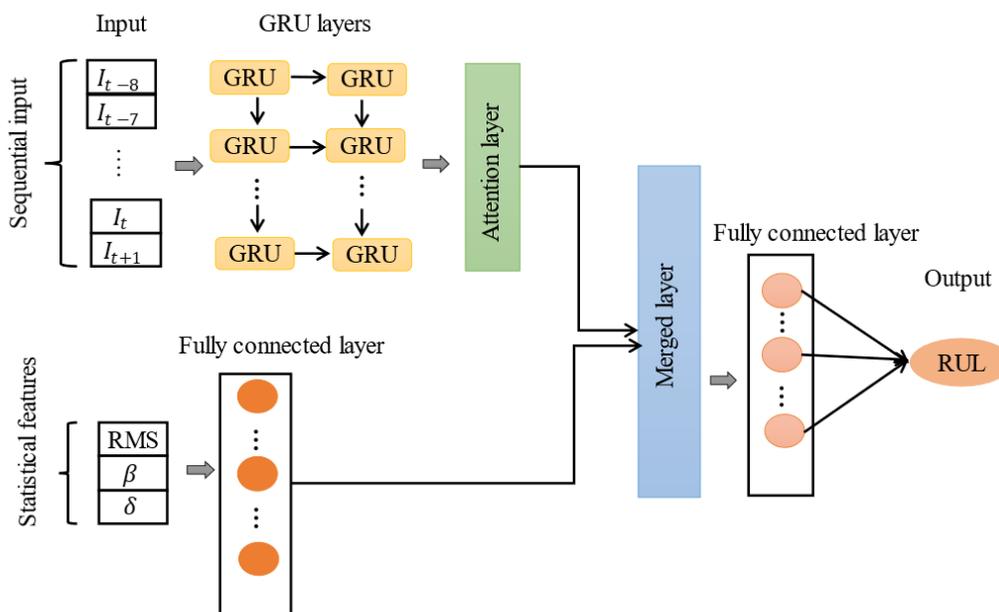


Figure 4.27: Structure of the proposed model for laser RUL prediction [265].

### 4.5.2 Methodology

The methodology depicted in Fig. 4.28 is adopted in order to validate the proposed framework. First, accelerated aging tests are performed under stressed conditions to induce laser degradation and thus accelerate device failure, with the laser current being monitored periodically under constant optical output power. After that, the collected

current measurement data is then normalized and divided into fixed length sequences. The preprocessed data is then fed into a GAN model to train it to synthetically generate realistic data that is similar to the real data. After training the GAN model, the generator is used to generate synthetic data, which is then used to train the various ML models involved in the predictive maintenance framework. Afterwards, each trained ML model is tested with real experimental data to assess its performance.

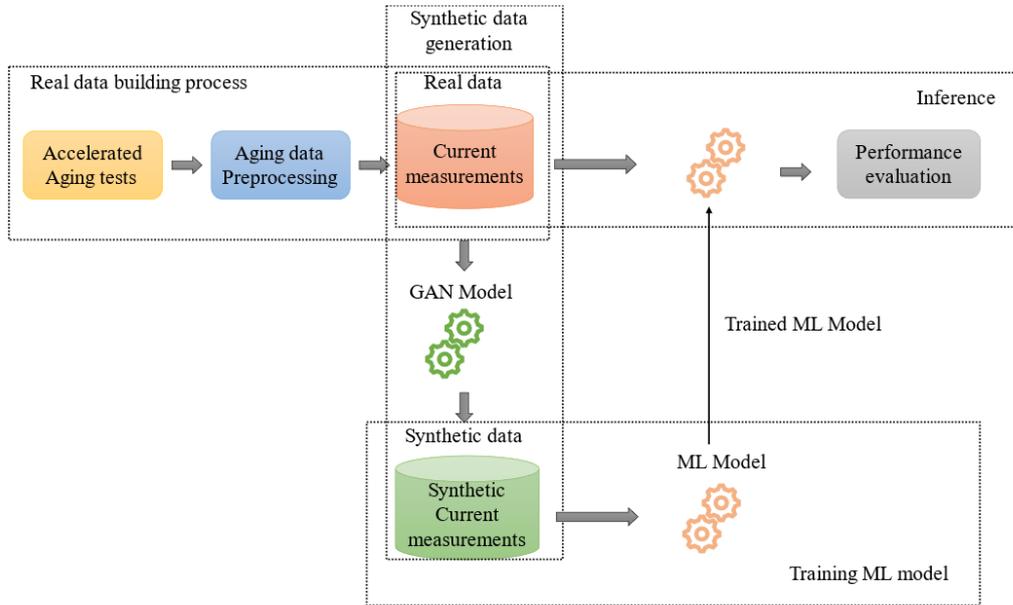


Figure 4.28: Methodology for the validation of the proposed framework [265].

### 4.5.3 Experimental data

Accelerated aging tests are carried out for various tunable laser devices operating at high temperatures of  $90^{\circ}\text{C}$  to significantly increase laser degradation and thus speed up device failure. The current is measured on a regular basis at the following times: 2, 20, 40, 60, 80, 100, 150, 500, 1000, 1500, 2000, and 3000 hours. The device's time to failure  $t_f$  is defined as the time when the current has increased beyond 20% of its initial value. The recorded current measurements of the tested devices are shown in Figure 4.29. A dataset of 384 samples containing the sequences of monitored current measurements of the tested devices (i.e., 384 semiconductor lasers) is generated. The RUL computed as the difference between  $t_f$  and the time  $t$  at which the RUL is predicted, as well as the device state (normal or anomalous/degraded), are assigned to each sample. Since the ML-based anomaly detection model is trained exclusively with normal data, only samples of normal devices are taken into account for training the GAN model. The first ten current values from each considered sample are then extracted, while the remaining

current values are kept for testing the ML-based performance prediction model's short-term and long-term prediction capability. In total, a dataset composed of 278 samples is adopted for training the GAN model.

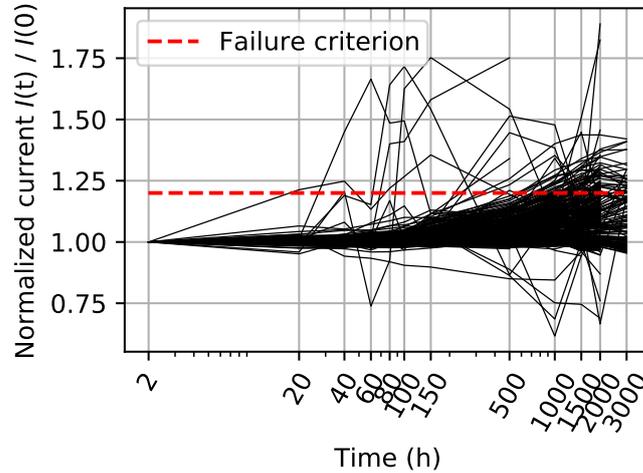


Figure 4.29: Recorded current measurements of different laser devices conducted at 90°C [265].

#### 4.5.4 Synthetic data

Figure 4.30 depicts the process of generating synthetic laser reliability data using GAN. The generator attempts to generate realistic data from a random noise input, whereas the discriminator is trained to differentiate between the generator's fake data and the real data. The generator and discriminator are both updated at the same time. The training procedure is repeated until the generator can generate data samples that the discriminator cannot distinguish from real data. The generator's architecture is made up of one LSTM layer with eight cells, followed by four convolutional layers with 32, 16, 16, 1 filters of size 3×1. The discriminator has 3 convolutional layers with 32 filters of size 3×1. The binary cross entropy serves as the cost function for the GAN model, with the generator attempting to minimize it and the discriminator attempting to maximize it.

After the GAN model has been trained, the generator is used to generate synthetic data. The evaluation metrics percent root mean square difference (PRD), RMSE, and the Fréchet distance (FD) are used to assess the quality of the synthetic data. PRD is used to assess the difference between real and generated data and can be expressed as:

$$PRD = \sqrt{\frac{\sum_{i=1}^N (x_i - \hat{x}_i)^2}{\sum_{i=1}^N (x_i)^2}} \times 100 \quad (4.9)$$

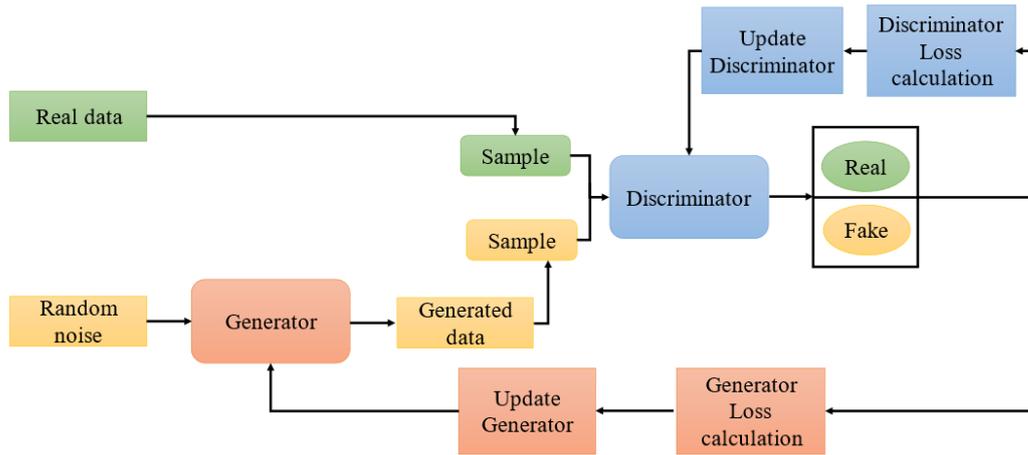


Figure 4.30: Synthetic data generation using a generative adversarial network (GAN) model [265].

where  $x_i$  represents the value of the sampling point  $i$  of the real sequence,  $\hat{x}_i$  denotes the value of the sampling point  $i$  of the generated sequence, and  $N$  is the length of the sequence.

The RMSE measures the consistency between the original and synthetic data.

FD quantifies the similarity between the real data and the generated data curves. Let  $O_R=(u_1, u_2 \dots u_R)$  be the order of points along the segmented real curves, and  $O_S=(v_1, v_2 \dots v_S)$  be the order of points along the segmented synthetic curves. The length  $\|d\|$  of the sequence consisting of couple of points  $(u_{a_1}, v_{b_1}), \dots (u_{a_n}, v_{b_n})$  is calculated as:

$$\|d\| = \max_{i=1..n} d(u_{a_i}, v_{b_i}), \quad (4.10)$$

where  $d$  denotes the Euclidean distance.

FD is computed as:

$$FD(R, S) = \min\{\|d\|\} \quad (4.11)$$

Please keep in mind that a good or optimal synthetic data generation method should have very low or near-zero PRD, RMSE, and FD metrics.

Table. 4.4 demonstrates that the various evaluation metrics are very small, indicating that the synthetic data is very close to the real data.

Evaluation metric	PRD	FD	RMSE
Value	2.87	0.45	0.028

Table 4.4: Assessment of the synthetic data using the metrics PRD, FD and RMSE.

A t-SNE, a technique for visualizing a high-dimensional data into two-dimensional

space (tSNE1 and tSNE2), is used to qualitatively assess how close the distribution of the synthetic data is to the distribution of the real data. Figure 4.31 shows that the distribution of the synthetic data is similar to that of the original experimental data, demonstrating the generator's effectiveness in producing realistic data.

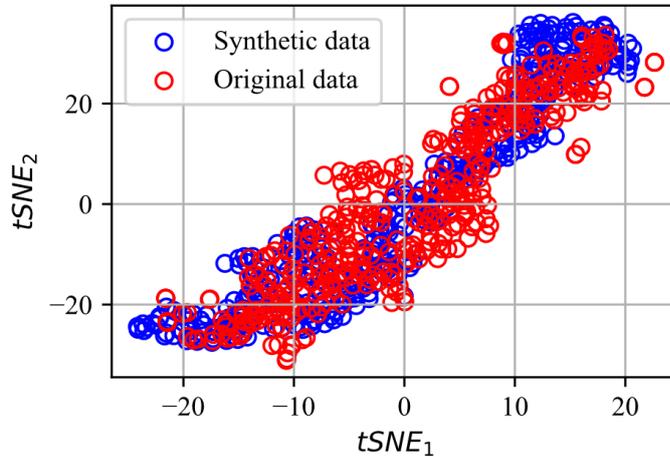


Figure 4.31: t-SNE visualization of the synthetic and real data distributions.

In total, a synthetic dataset of 5,600 samples is produced. Based on the defined failure criteria, the RUL and device state are computed for each sample. The data is then normalized before being fed into the ML models for training.

#### 4.5.5 Validation results of the performance prediction model

The proposed model is compared to other ML techniques, namely multilayer perceptron (MLP), CNN, and RNN by using as evaluation metrics the MAPE and the coefficient of variation of the root mean squared error (CVRMSE), which can be expressed as:

$$CVRMSE = \frac{100}{\bar{x}} \sqrt{\frac{\sum_{i=1}^N (x_i - x'_i)^2}{N}} \quad (4.12)$$

where  $N$  denotes the number of test samples and  $\bar{x}$  represents the average of the true current values. It should be noted that a lower MAPE and CVRMSE value indicates a better prediction capability.

The aforementioned ML models are trained on a synthetic dataset and then tested on a real experimental dataset. The comparison results shown in Fig. 4.32 demonstrate that the proposed model outperforms the other ML algorithms by achieving the lowest MAPE and CVRMSE values, indicating that the proposed method produces better prediction performance.

The ability of the proposed model to predict in the short and long term is evaluated. It is worth noting that the model is trained with current measurements until 1,000

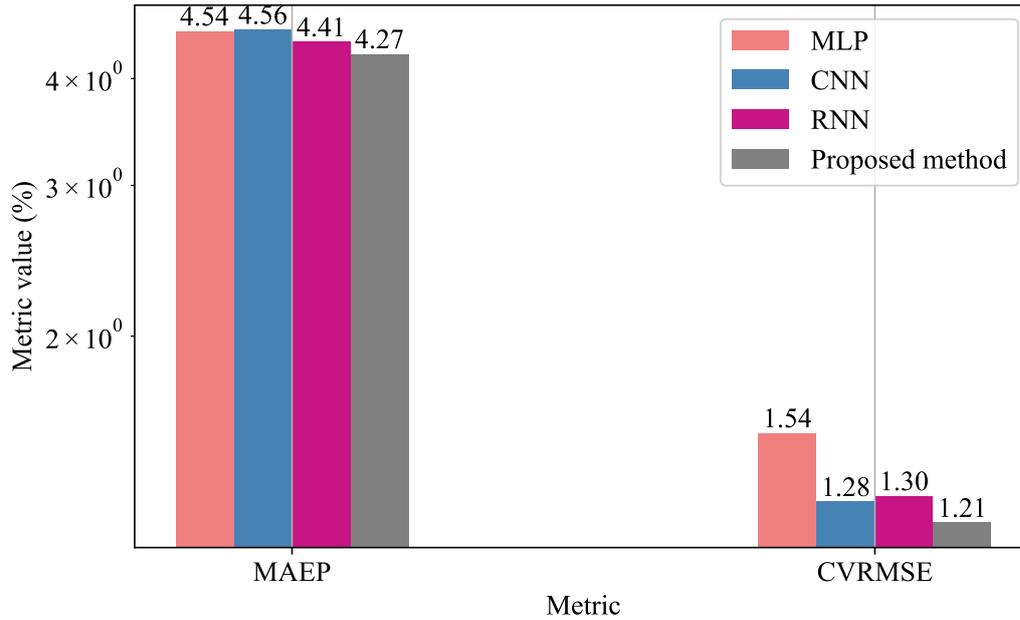


Figure 4.32: Results of the comparison of the proposed method with MLP, CNN and RNN in terms of MAPE and CVMSE [265].

hours and then tested to forecast current values at 1,500 hours, 2,000 hours, and 3,000 hours. The predicted values of two random samples are shown in Figure 4.33. The forecasted values are close to the actual values and follow the same degradation trend as the actual values, demonstrating the effectiveness of the proposed model in predicting current measurements.

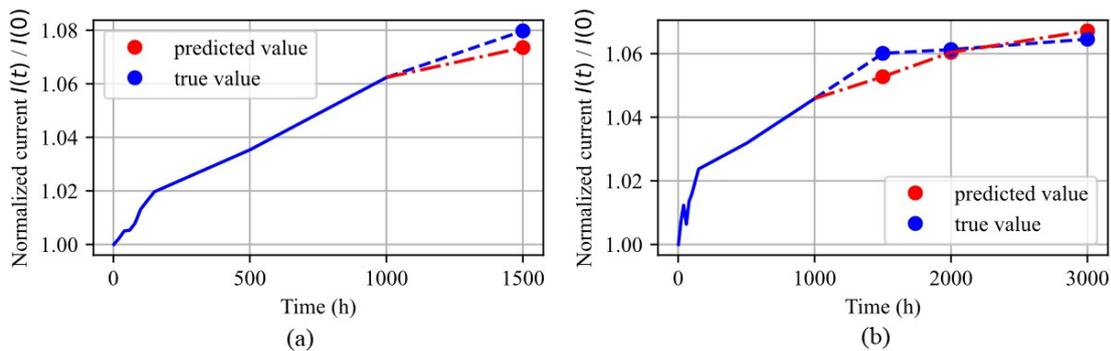


Figure 4.33: Results of laser current prediction for two random test samples: (a) short-term prediction of the current at 1,500 h, (b) long-term prediction of the current up to 3,000 h [265].

**Validation results of the ML model for anomaly detection** The model's detection capability is optimized by selecting the optimal threshold  $\theta$ . Figure 4.34 illustrates the precision, recall and F1 score curves along with  $\theta$ . As discussed in Section 4.1, the

optimal threshold, that provides the best precision and recall tradeoff (i.e. maximizing the F1 score) is selected. For the optimal chosen threshold of 0.019, the precision, the recall, the F1 score and the accuracy are 96.72%, 92%, 94% and 94.24%, respectively.

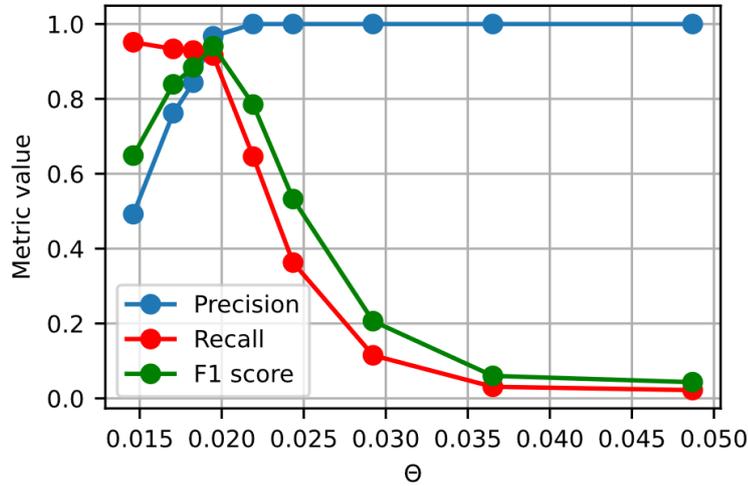


Figure 4.34: The optimal threshold selection based on the precision, recall and F1 score scores yielded by the autoencoder [265].

**Validation results of the ML model for RUL prediction** After training with synthetic data, the proposed model for RUL prediction is tested with experimental data using the RMSE and MAE evaluation metrics. The proposed method's prediction capability with the GRU model without attention mechanism and the attention-based GRU method are compared. Figure 4.35 shows that the proposed model has the lowest RMSE and MAE scores, demonstrating that adding statistical features and the attention mechanism helps to improve RUL estimation capability. The addition of the attention mechanism improves performance by 10.3% and 8% in the RMSE and MAE metrics, respectively. Whereas incorporating both statistical features and an attention mechanism improves prediction capability by 32.6% and 18.5%, respectively, in RMSE and MAE.

The proposed model is compared with other ML techniques, namely RF, SVR, MLP, RNN, LSTM and CNN, using as evaluation metrics the RMSE and the MAE. The results shown in Table 4.5 demonstrate that the proposed method outperforms the other ML models by achieving the smallest scores of RMSE and MAE. The comparison results of the computational time (inference time) of the proposed model and other ML methods in Table 4.8 show that the proposed method consumes more time in testing (inference) due to its deep architecture, and that the shallow ML techniques RF and SVR are much less time consuming in testing. The proposed method also outperforms

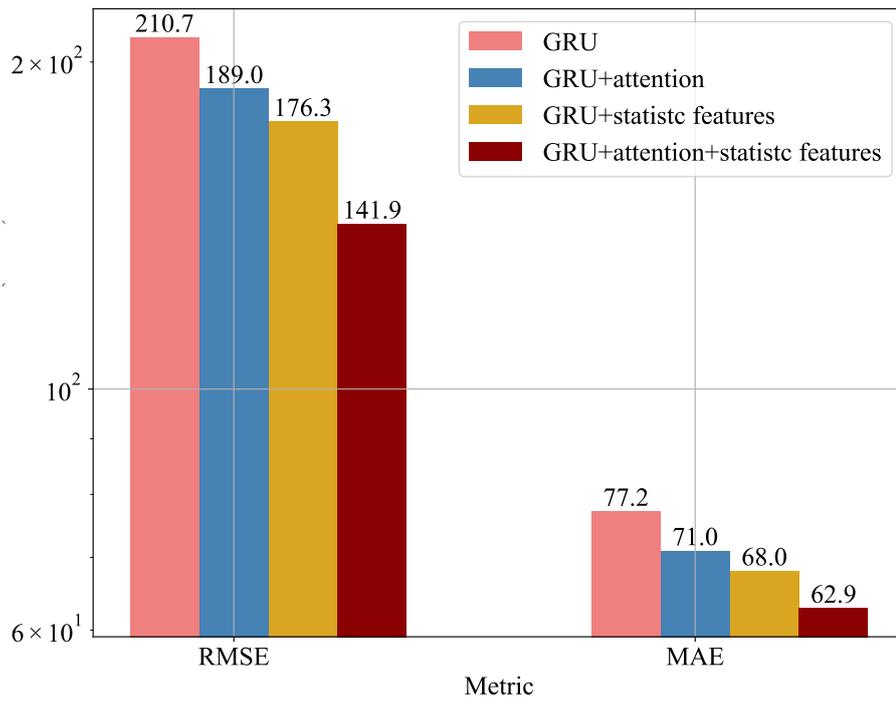


Figure 4.35: Results of the comparison of the proposed model (GRU+ attention+ statistic features) with GRU model, attention based GRU method, and GRU + statistic features model using the RMSE and the MAE metrics [265].

Method	RMSE	MAE	Inference time (55 samples)
RF	1269.37	882.67	0.003 s
SVR	973.8	547.77	0.004 s
MLP	309.38	120.57	0.035 s
RNN	228.12	70.25	0.71 s
CNN	253.26	95.4	0.148 s
LSTM	181.33	93.68	0.49 s
Proposed method	141.9	62.85	0.75 s

Table 4.5: Computational time of proposed model and other methods.

the CNN-LSTM model for laser RUL estimation presented in Section 4.3 (RMSE = 385 hours, MAE = 261 hours) by improving the RMSE and MAE metrics by 63.14% and 75.9%, respectively. The predicted RUL values are compared with the true RULs at various stages of degradation to further evaluate the proposed model's RUL estimation capability. As depicted in Fig. 4.36, the model's RUL values are very close to the true RUL values, demonstrating the proposed model's effectiveness in predicting the RUL of the laser device.

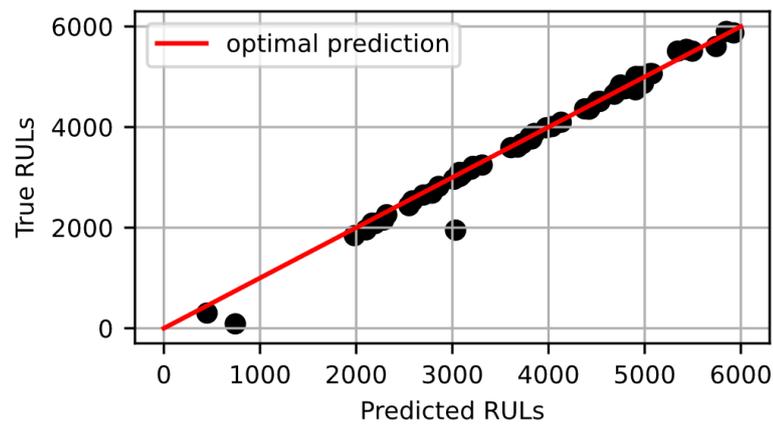


Figure 4.36: Results of Predicted RULs by the proposed model vs. actual RULs [265].

## 4.6 Conclusion

It has been demonstrated that ML can be a promising and efficient tool for improving the reliability of semiconductor lasers and thereby ensuring more reliable optical communication systems. Different ML approaches for solving several predictive maintenance tasks, including lifetime prediction, degradation prediction, RUL prediction, and early degradation type identification, have been investigated and validated using synthetic or experimental laser reliability data comprising either historical current or output power measurements or several monitored electro-optical laser characteristics. The presented approaches achieve good performance and outperform the conventionally employed methods. The same concepts of the proposed ML approaches are readily applicable to other optoelectronic devices such as semiconductor optical amplifiers due to the similarity of their structures.

## 5 Fault Management for Optical Fiber Links

Fiber-optic monitoring has mainly been performed using OTDR in the past. Analyzing OTDR traces, even for experienced field engineers, can be challenging mainly due to noise overwhelming the signal, resulting in inaccurate or unreliable event detection and localization. The noise in OTDR signals can be reduced by taking the average of multiple OTDR measurements. Although averaging can minimize the noise level and thus improving the performance of OTDR event analysis approaches in terms of event detection and fault localization accuracy, it is a time-consuming process. The greater the number of averages, the better the SNR of the OTDR trace becomes, however, the longer it takes to analyze the OTDR data. Given that real-time detection is the industry standard, it is crucial to have a reliable automated diagnostic technique that accurately and quickly locates and identifies faults while processing noisy OTDR data without the need for extensive averaging or assistance from trained personnel, thereby lowering operation and maintenance costs. This can be accomplished by fully utilizing the power of ML methods in solving pattern recognition tasks and extracting insights from OTDR data.

This chapter presents various ML approaches for optical fiber fault detection, localization, and characterization given noisy OTDR data, for different optical network architectures. Section 5.1 discusses the use of ML techniques to improve fault management in simple point-to-point optical links. Section 5.2 covers the application of the ML methods for enhancing optical fault diagnosis and localization in passive optical networks (point-to-multi-point optical links).

### 5.1 Point-to-Point Optical Links

This section investigates various ML-based optical fiber fault management use cases. Subsection 5.1.1 covers the use of ML approaches for detecting, localizing, and characterizing reflective faults. In the Subsection 5.1.2, the identification and the localization of a wide variety of optical fiber faults including not only reflective faults, but also non-reflective and merged events, using ML methods is explored. Subsection 5.1.3 discusses the issue of very noisy OTDR data, which can severely impact the performance of an ML model trained to solve a fiber fault management use case at very low SNR

levels, and proposes a potential solution to improve ML model performance under such conditions.

### 5.1.1 Reflective Fault Detection, Localization and Characterization

The content of the Subsection 5.1.1 has already been published in [4].

Figure 5.1 shows the workflow of the proposed ML based approach for fiber reflective event detection and characterization. The proposed approach enables: (i) faster detection of faulty fiber links monitored by placing a reflector at the end of each fiber; and (ii) more quickly detection, localization, and estimation of the reflectance of the reflective events caused by connectors or mechanical splices in optical fiber links. In operation, an OTDR trace is divided into fixed length sequences, and the ML model is then applied to each derived sequence to detect and characterize the reflective events. Because the position of the reflector is known, only the sequence containing the reflective peak is fed to the ML method to investigate the link's integrity. The integrity of the optical fiber link can be verified by checking the probability of the reflective event and comparing the reflectance value predicted by the model to the initial reflectance value of the reflector. Once a faulty link is identified, the ML model is applied to the remaining OTDR segmented sequences to detect, locate, and predict the reflectance of events.

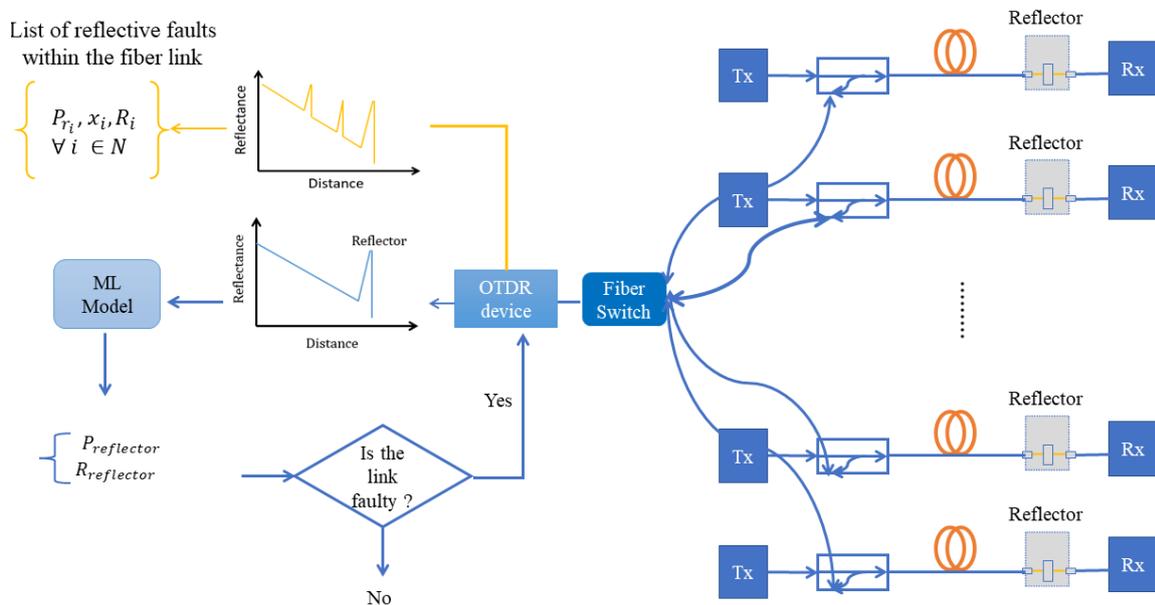


Figure 5.1: Typical operation scenario for fiber monitoring with an OTDR device and receiver end reflectors (right) and estimation flow diagram (left) ( $P_{r_i}$ : probability of the event,  $x_i$ : position,  $R$ : reflectance,  $N$ : number of events) [4].

The following subsections discuss the structure of the adopted ML model, the experimental OTDR data used for training and validating the proposed approach, and the performance results.

### 5.1.1.1 Model Structure

As illustrated in Fig. 5.2, the LSTM-based model's architecture consists of one LSTM hidden layer with 30 neurons and three layers of 15 neurons for each investigated task: the event detection  $T_1$ , the position estimation  $T_2$  and the reflectance prediction  $T_3$ . The ML model receives a 35-length OTDR trace sequence as input and outputs the  $ID_{class}$  (0: no reflective event, 1: reflective event), the position index of the reflective event within the sequence ( $Pos$ ), and the reflectance  $R$ . Fed with input sequences, the LSTM shared layer extracts the useful information underlying the event pattern (i.e., learned knowledge), which is then transferred to the layer of each task. Each task layer improves its generalization performance by combining the information retrieved by itself with the shared knowledge. The overall loss function, which is used to update the weights of the model based on the difference between the predicted and desired output, can be expressed as:

$$L_{total} = \alpha L_{T_1} + \beta L_{T_2} + \delta L_{T_3} \quad (5.1)$$

where  $L_{T_1}$ ,  $L_{T_2}$  and  $L_{T_3}$  represent the loss of  $T_1$ ,  $T_2$  and  $T_3$ , and the first one denote the binary cross-entropy loss whereas the others are regression losses (mean squared errors). The loss weights  $\alpha$ ,  $\beta$  and  $\delta$  are hyperparameters to be tuned.

### 5.1.1.2 Experimental Data

The experimental setup depicted in Fig. 5.3 is used to record OTDR traces that include reflective faults with small and large peaks. The reflective event is caused by placing a reflector at the end of the fibers under test [273]. The reflector is a wavelength-selective reflector with a 95% reflectance that selectively reflects test light pulses sent by the OTDR at 1650 nm while allowing data transmissions at other wavelengths to pass with relatively slight attenuation. The experimental setup parameters chosen (small pulse width, low power, etc.) reduce the dynamic range, weakening the signal in front of the reflective event by increasing noise relative to the reflected pulse. Given that the signal in front of the reflective peak is too low due to the chosen parameters, changing the height of the reflective event would have the same effect as changing the reflectance of the reflector. In order to generate reflective events with small peaks, the height of the peak is varied by placing an internal VOA in front of the photodetector in the

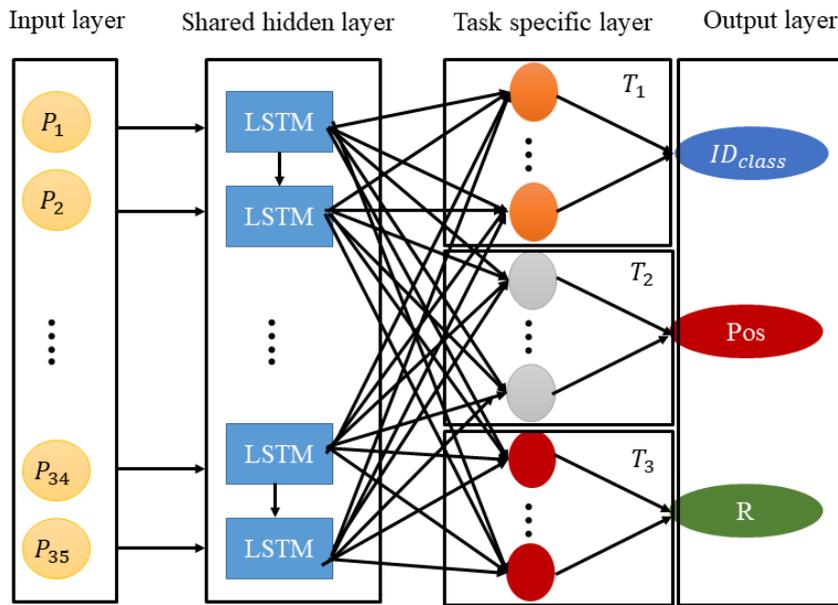


Figure 5.2: Proposed LSTM-based model structure [4].

OTDR device and varying its attenuation settings from 0 to 20 dB. Because OTDR measurement averaging affects the SNR of the OTDR trace, 62 to 64,000 OTDR records are collected and averaged. Furthermore, the laser power is varied from 0 to 12 dBm to influence the SNR. The OTDR configuration parameters namely the pulse width and the sampling period are set to 50 ns and 8 ns, respectively. Figure 5.4 shows an

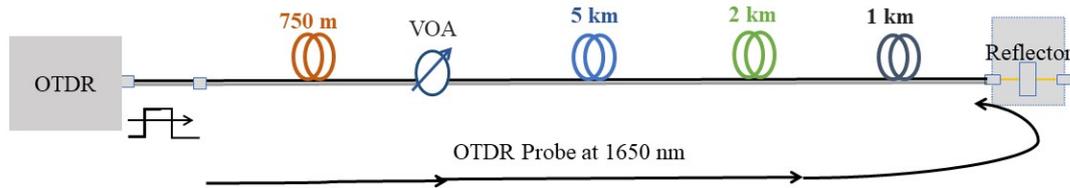


Figure 5.3: Experimental setup for a single reflective event [4].

example of an OTDR trace. The reflective event caused by the reflector is represented by the peak at position 8785 m, which we are looking for (i.e., detecting, locating it and estimating its reflectance). The noise obscures the reflective event, as seen in the trace. This makes it difficult to locate and detect the event. The exponential decay cannot be seen in the trace because the dynamic range is too low due to the parameters chosen (the pulse width, the laser power, and the attenuation value of the received signals). The SNR is defined as [4]:

$$SNR = \frac{a}{\sigma_{noise}} \quad (5.2)$$

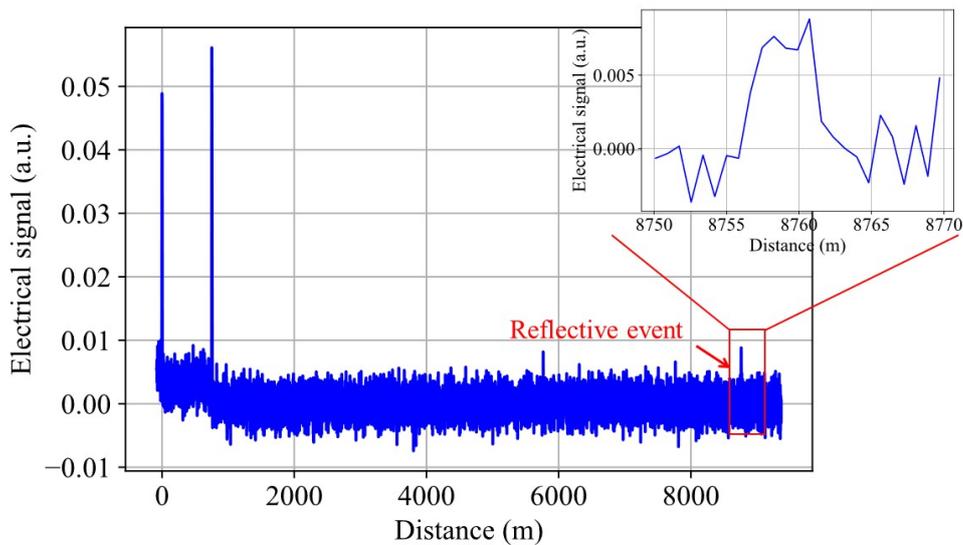


Figure 5.4: OTDR trace including the reflective event [4].

where  $a$  denotes the height of the reflective event and  $\sigma_{noise}$  is the standard deviation of the noise. The event height is estimated as the average of the peak's roof, defined as the average of the two highest values between the event position and the event position plus twice the pulse width length.  $\sigma_{noise}$  can be computed as:

$$\sigma_{noise} = \sqrt{\left(\frac{1}{n} \sum_{N-n+1}^N x_i^2\right) - \left(\frac{1}{n} \sum_{N-n+1}^N x_i\right)^2} \quad (5.3)$$

where  $N$  is the last sampling point of the OTDR trace,  $n$  denotes the number of samples used for the estimation (1000 for this data), and  $x_i$  represents the sampling value.

Up to 6,300 traces containing reflective events with SNR values ranging from 2 to 30 dB and various reflectance values are generated.

### 5.1.1.3 Data Preprocessing

From each OTDR trace, two sequences of length 35 are extracted randomly: one with no events, one with the entire reflective event pattern or just a portion of it if the reflective peak is on the edge of the sequence (to ensure that also partial events can be detected). The  $ID_{class}$  (0: no reflective event, 1: reflective event), the reflective event position index within the sequence, and the reflectance  $R$  are assigned to each sample. The reflectance can be calculated by comparing the peak height to reference measurements for various cleaved fiber ends (see Section VI in [274]). In total, a data set of 12,600 samples is built. The generated data is normalized and divided into a training (60%), a validation (20%) and a test dataset (20%) using random selection.

### 5.1.1.4 Results

**Machine Learning Model Overall Performance** Several metrics, including accuracy, precision, recall, and the F1 score for event detection task evaluation and RMSE and MAE metrics for event position estimation and reflectance prediction task evaluation, are used to assess the performance of the proposed model. The overall performance of the model compared to the single task baselines is depicted in Tab. 5.1. For the sake of comparison, the same architecture as the LSTM-based model is used for the single task models (one LSTM hidden layer with the same number of hidden neurons, one fully connected layer with the same number of hidden neurons). For the event detection task, the improvement in performance ( $\delta$ ) over single task model is computed as the difference between the evaluation metrics of the proposed model and the single task baseline. Whereas for the event position and reflection prediction tasks,  $\delta$  is calculated as follows:

$$\delta(\%) = 1 - \frac{m_{multitask}}{m_{singletask}} \times 100 \quad \forall m \in \{RMSE, MAE\} \quad (5.4)$$

The experimental results shown in Tab. 5.1 demonstrate that the proposed model significantly outperforms the single task baselines.

Metric	Single task	Multitask	$\delta(\%)$
Task 1: Event detection			
Accuracy (%)	92 ± 1.06	93 ± 0.9	+1
Precision (%)	96.5 ± 0.7	96.6 ± 0.7	+0.1
Recall (%)	87.2 ± 1.3	88 ± 1.2	+0.8
F1 score	91.66 ± 1.1	92.1 ± 1.05	+0.45
Task 2: Position estimation			
RMSE (m)	4.4 ± 0.13	2.17 ± 0.07	+50.7
MAE (m)	2.2 ± 0.13	1.15 ± 0.07	+47.7
Task 3: Reflectance prediction			
RMSE (dB)	5.0 ± 0.13	3.65 ± 0.09	+27
MAE (dB)	3.3 ± 0.13	2.5 ± 0.09	+24.2

Table 5.1: ML model performance evaluation.

**Machine Learning Model Performance as Function of SNR** Following that, the performance of the proposed model as a function of SNR is investigated. Figure 5.5 depicts the effects of SNR on the proposed model's accuracy. The Wilson score [275] is used to estimate the confidence intervals. The accuracy increases with SNR, as expected. The accuracy approaches one for SNR values greater than 5 dB. With an SNR of 2 dB, the accuracy is lower because it is difficult to distinguish the event

from the noise, and thus the model misclassified the event class as normal (i.e. noise). Starting at SNR = 3 dB, the model was able to distinguish the event from the noise and detect it with greater than 90% accuracy. The RMSE of event position estimation as

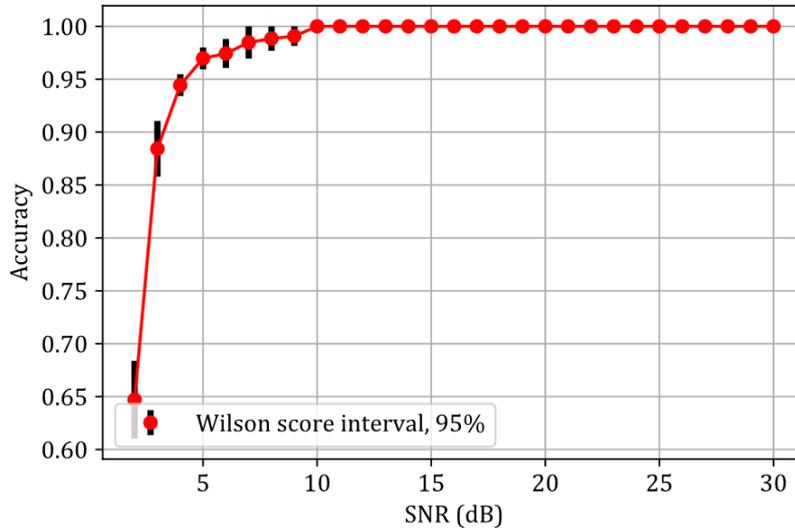


Figure 5.5: Event detection performance evaluation with confidence interval [4].

a function of SNR is shown in Figure 5.6. The RMSE decreases as the SNR increases. The RMSE can be greater than 2.5 m for lower SNR values ( $\text{SNR} \leq 10$  dB), but less than 2 m for SNR values greater than 22 dB, and less than 1 m for SNR values greater than 30 dB. Figure 5.6 also shows that as SNR increases, the RMSE of the reflectance prediction for both cases, mixed and whole peak sequences, decreases. RMSE is greater than 3.5 dB for lower SNR values ( $\text{SNR} \leq 10$  dB) and less than 3 dB for higher SNR values. The upper blue graph depicts the average reflectance prediction error of instances of reflective events with small and large peaks, including partial reflective event sequences. The RMSE could be reduced further, up to 1.1 dB, as shown by the black graph in Fig.5.6, if segmentation could ensure that only whole reflective event patterns occur (lower bound).

**Experimental Settings for Machine Learning Model Performance** The effect of including information about the experimental setup parameters, specifically the laser power ( $P_{laser}$ ), averaging ( $N_{avg}$ ) denoting the number of averaged OTDR traces, attenuation ( $\alpha$ ) of the received OTDR signals, and SNR ( $\gamma$ ), during the training of the ML model to improve performance is investigated. Figure 5.7 depicts the ML model's extended architecture, which includes experimental setup features as an additional input. First, the model described in Subsection 5.1.1.1 is trained using only sequences of signal power levels. To these sequences, the parameters  $P_{laser}$ ,  $N_{avg}$  and  $\alpha$  are then

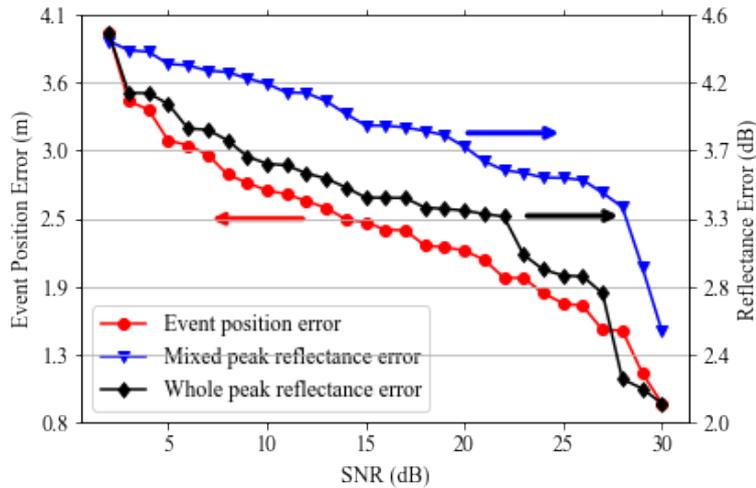


Figure 5.6: Event position and reflectance estimation error (RMSE) for the ML model [4].

added, and the model is retrained again. The same procedure of adding the respective parameters  $\gamma$ , the tuple  $(\gamma, P_{laser}, N_{avg}, \alpha)$  and the tuple  $(\gamma, P_{laser}, N_{avg})$  to the initial training data (i.e. sequences of signal power of length 35), and (re-) training the model is repeated.

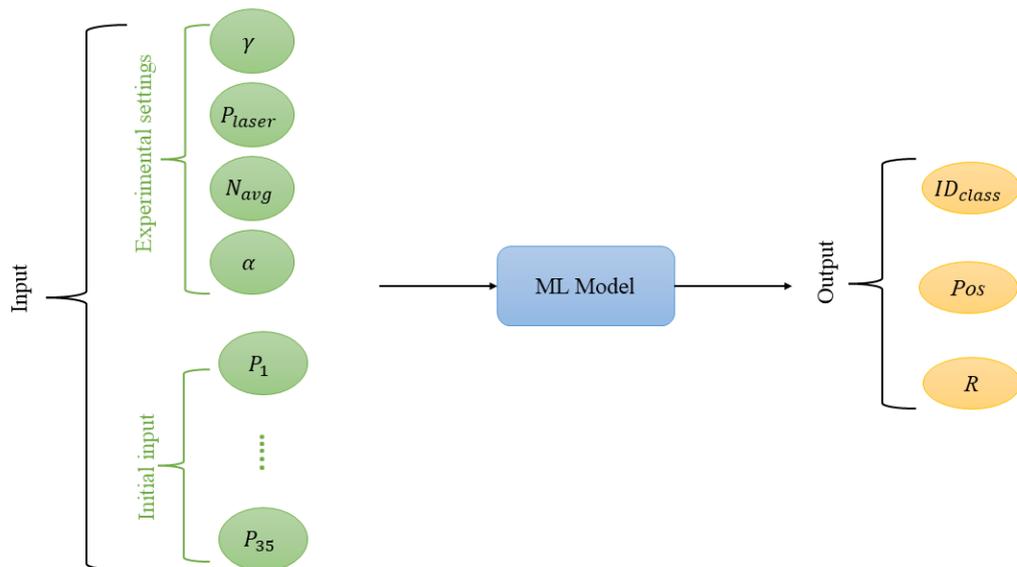


Figure 5.7: Extended ML architecture by adding the experimental setup parameters [4].

The experimental results shown in Fig. 5.8 demonstrate that adding the experimental settings  $P_{laser}$ ,  $N_{avg}$  and  $\alpha$  to the training data increases the ML model's detection capability by 0.75%. Including only  $\gamma$  improves accuracy by more than 1.5%. This

result demonstrates that  $\gamma$  is the most important parameter influencing the detection capability of the ML model, as the other parameters ( $P_{laser}$ ,  $N_{avg}$ ,  $\alpha$ ) only have an impact on the SNR. Training the ML model with all of the experimental setup parameters ( $\gamma$ ,  $P_{laser}$ ,  $N_{avg}$ ,  $\alpha$ ) does not significantly improve ML model performance when compared to only including  $\gamma$  during training. This could be explained by the fact that *gamma* and *alpha* are highly correlated. The accuracy is improved by more than 2% by excluding  $\alpha$  and re-training the ML model with the sequence of signal power combined with the remaining settings, namely  $\gamma$ ,  $P_{laser}$  and  $N_{avg}$ .

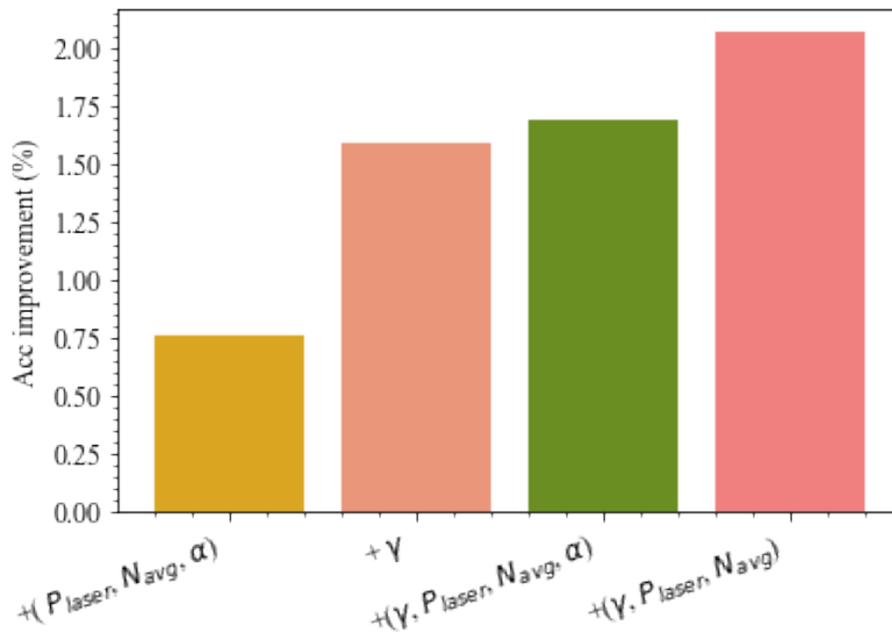


Figure 5.8: Accuracy improvement by adding features (experimental settings) to the initial training data (sequences of signal power levels) [4].

The results shown in Fig. 5.9 demonstrate that including experimental parameters during ML model training improves event localization accuracy. For high SNR values, the models trained with experimental settings combined with the initial sequence of power levels could locate the event with lower RMSE up to reaching the resolution limit of 0.2 m. Because the sampling is 0.8 m, the resolution limit is estimated to be  $\pm 0.4$  m, but as the RMSE of the position error is the average value for each SNR, the resolution limit is derived to be 0.2 m. Training the ML model with the sequence of signal power combined with  $\gamma$  only provides the best localization accuracy for SNR values greater than 8 dB. Including all of the experimental setup parameters ( $\gamma$ ,  $P_{laser}$ ,  $N_{avg}$ ,  $\alpha$ ) during ML model training does not improve event position prediction because the experimental setup parameters are correlated and bring redundant information that hinders the ML model's ability to generalize.

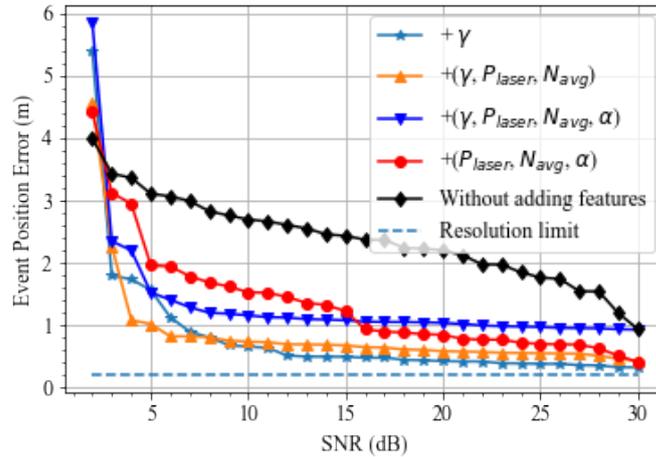


Figure 5.9: Event localization accuracy improvement by adding features (experimental settings) to the initial training data (sequences of signal power) [4].

The study of the effect of including the experimental settings during the training of the ML model on performance demonstrates that training the ML model with  $\gamma$  only combined with the sequence of power signals significantly improves the ML model's performance in terms of event detection capability and event localization accuracy.

**Machine Learning Model versus Conventional Method** It should be noted that the developed ML model was trained using data that included partial and whole peak sequences, as the model will be applied to arbitrarily segmented OTDR sequences. For a fair comparison of the developed ML model with a conventional rank-1 matched subspace detector (R1MSDE) [276], an unseen test dataset, containing only the complete reflective event pattern or no event, is generated. R1MSDE applies the theory of matched subspace detection and associated maximum likelihood estimation techniques to discriminate connection splice events from noise and the Rayleigh component in the OTDR data. As in [276], the reflective event is represented by a rectangular pulse with prior knowledge of the event's duration. The length of the test data window has been optimized to provide the best performance for R1MSDE. As a result, the R1MSDE performance is assessed using a test dataset of 100 sequences (which turned out to be the optimum sequence length). The ML model's performance is evaluated using test sequences of length 35. The detection probability ( $P_d$ ) is used to evaluate the detection capability for fixed false alarm probability ( $P_{FA}$ ).  $P_d$  is the proportion of the total number of correctly detected reflective events, which is expressed as:

$$P_d = \frac{TP}{TP + FN} \quad (5.5)$$

where  $TP$  denotes the number of true reflective event detections, and  $FN$  represents the number of false negative ones.

$P_{FA}$  is calculated as:

$$P_{FA} = \frac{FP}{TN + FP} \quad (5.6)$$

where  $FP$  is the number of false positives, and  $TN$  is the number of true negatives. A comparison of the different detectors in terms of  $P_d$  for  $P_{FA}$  of 0.1, as shown in Fig. 5.10, demonstrates that the ML model outperforms the R1MSDE by achieving higher detection capability particularly for low SNR values. Figure 5.11 depicts the

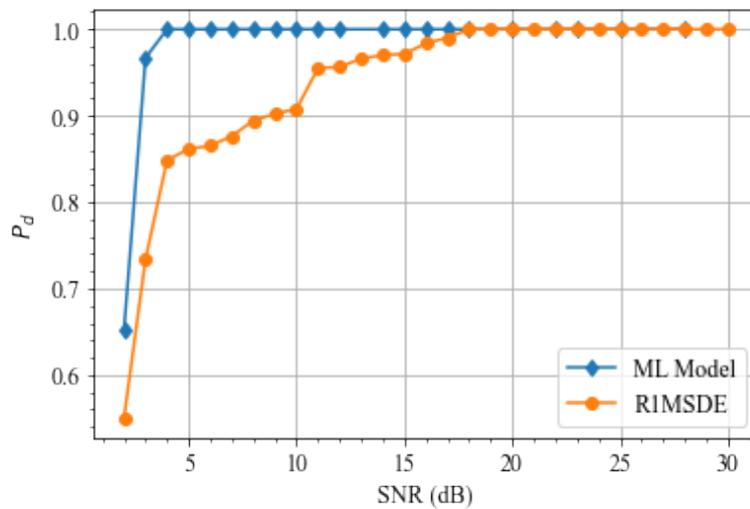


Figure 5.10: Comparison of the detection probability for R1MSDE and ML model for a false alarm probability of 0.1 [4].

results of a comparison of the ML-performance model's with R1MSDE in terms of RMSE of the event position. The ML model outperforms the R1MSDE for lower SNR values by achieving an event position error of less than 4 m. The performance of R1MSDE improves as SNR increases. Figure 5.12 shows the results of the comparison of the ML model's accuracy with R1MSDE as function of the averaging time ( $\tau$ ), which is computed as follows:

$$\tau(s) = \frac{2 N_{avg} L n}{c} \quad (5.7)$$

where  $L$  denotes the length of the fibers under test,  $n$  is the refractive index and  $c$  is the speed of light.

As illustrated in Fig. 5.12, the ML model requires significantly less averaging of OTDR measurements (less measurement time) than the R1MSDE to achieve the same detection capability. This could aid in the detection of events more quickly. The approach requires less than 0.2 s of averaging time to achieve 93% accuracy, whereas the R1MSDE requires more than 12 s.

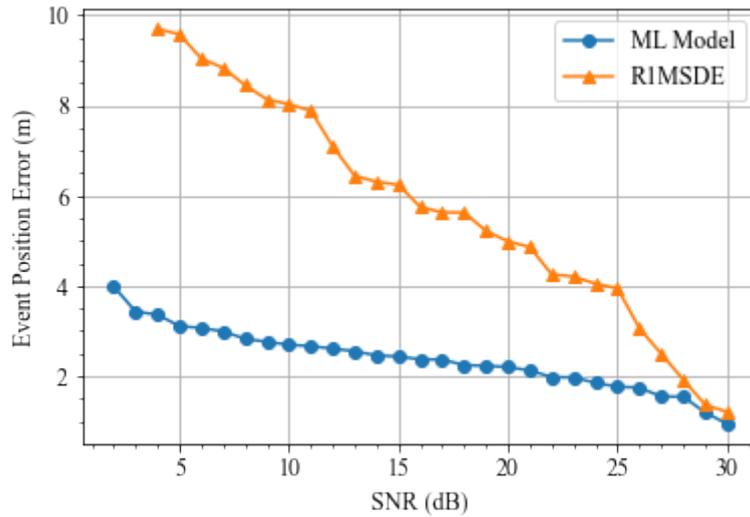


Figure 5.11: Event position estimation error (RMSE) for the ML model vs. R1MSDE [4].

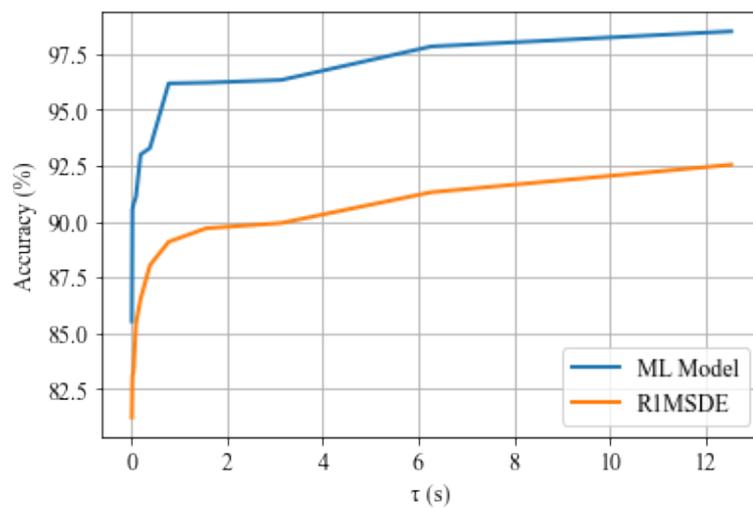


Figure 5.12: Comparison of the accuracy for R1MSDE and ML model as function of the averaging time [4].

### 5.1.2 Reflective and Non-Reflective Fault Diagnosis and Localization

The previous subsection discussed an ML approach for detecting and localizing only reflective events. This subsection investigates the capability of the ML techniques in recognizing more complex types of optical fiber faults such as non-reflective and merged events, when analyzing noisy OTDR data. The content of this subsection has already been published in [277].

### 5.1.2.1 Experimental Data

Both the setup discussed in the previous subsection 5.1.1.2 (referred to as experiment B in the following) and the experimental setup shown in Fig. 5.13 are used to record OTDR traces containing different types of optical fiber faults with varying characteristics (e.g., different reflectances and/or losses). To model specific faults in the fiber link, optical components such as connectors, VOAs, and reflectors are used. To generate non-reflective faults with only attenuation and no reflection, an APC connector is placed at the end of the 1 km fiber of the experiment B. Non-reflective faults are referred to as fiber-bend faults. To produce non-reflective events with abrupt decrease, an APC connector is placed at the end of experiment B (i.e. the end of the 5 km fiber). In that case, the non-reflective faults produced are known as tilted fiber-cut faults. By placing a reflector at the end of the experiment A (after the 1 km fiber), reflective faults (Fresnel reflection) with small peaks are induced. They are known as longitudinal connector misalignment. As discussed in subsection 5.1.1.2, the reflectance of these events varies. Reflective events with a sharp peak and abrupt decrease are caused by inserting a PC connector at the end of experiment B. In such cases, they are referred to as perpendicular fiber-cut. APC connectors are used to make connections between the VOA and the coupler shown in Fig. 5.13. Some of the connectors are deliberately dirty. As a result, merged events (combined), which consist of either two overlapped reflective faults or two overlapped non-reflective and reflective events, are induced. Changing the VOA settings affects the attenuation of non-reflective and combined faults.

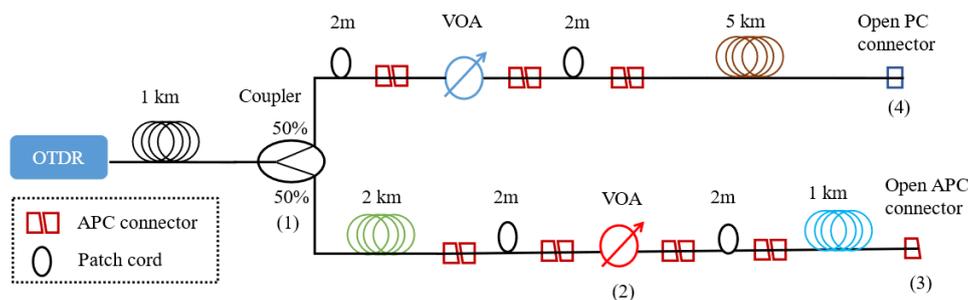


Figure 5.13: Experimental setup for generating non-reflective and reflective faults [235].

Experiment B employs the same OTDR configuration settings (pulse width of 50 ns, wavelength of 1650 nm, sampling rate of 8 ns) as experiment A. Figure 5.14 shows an example of an OTDR trace generated by experiment B, illustrating the different patterns of the investigated event types.

### 5.1.2.2 Data Preprocessing

Five sequences of length 50 are randomly extracted from each OTDR trace generated by experiment B: one containing no event, two containing the merged events, one involving the non-reflective fault, and one containing either reflective or non-reflective events depending on the connector type placed at the end of the setup. In contrast, for each OTDR trace generated by experiment A, two sequences of length 50 (one with no event and one with the reflective event pattern) are chosen randomly. The event type (no event, reflective, non-reflective, merged), the event position index within the sequence, the event loss and/or reflectance, and the event cause are all assigned to each sequence. The "no event" cause is referred to as "no fault" (class 0). The causes of "reflective events" are classified as either longitudinal connector misalignment (class 1) or perpendicular cut (class 2). The "non-reflective" fault causes are classified as fiber bend (class 3) or tilted fiber cut (class 4). Dirty connector and/or fiber bend are the merged event causes (class 5, class 6). In total, a data set of 30,112 samples with SNR values ranging from 0 to 30 dB is built and normalized.

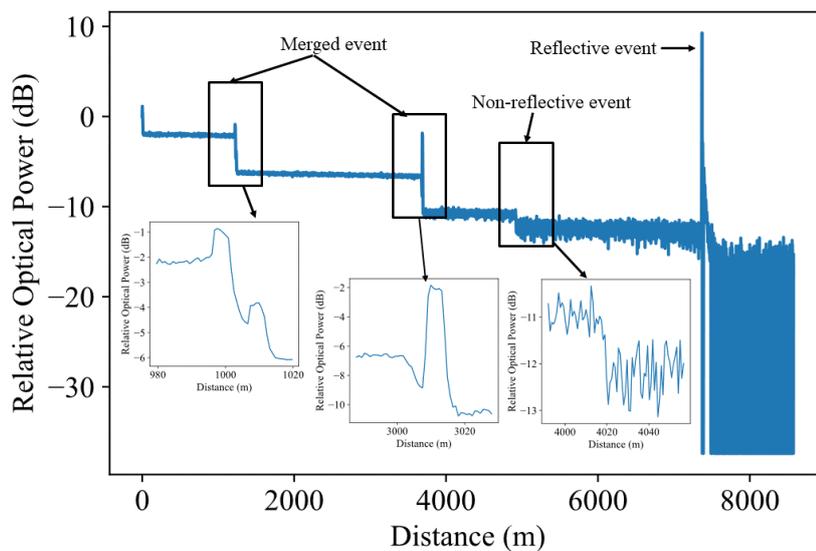


Figure 5.14: Example of OTDR trace [277].

### 5.1.2.3 Model Architecture

The structure of the proposed model is shown in Fig. 5.15. Because BiLSTM is well suited to processing OTDR sequential data and capturing long-term dependency, and CNN has been shown to be good at local feature extraction, a hybrid deep learning approach integrating BiLSTM and CNN is chosen as the model's shared hidden layer

to improve performance by combining their strengths. A shared hidden layer's architecture consists of one BiLSTM layer with 32 cells, followed by CNN layers with primarily one convolutional layer with 32 filters and a max pooling layer, followed by a dropout layer to prevent overfitting. The knowledge acquired by BiLSTM-CNN is then transferred to the four task-specific layers, which are comprised of 16, 20, 32, and 40 neurons, respectively. The total loss of the model is calculated as the weighted sum of the four individual task losses, which are set to 1.5, 0.5, 1.8, and 1, respectively.

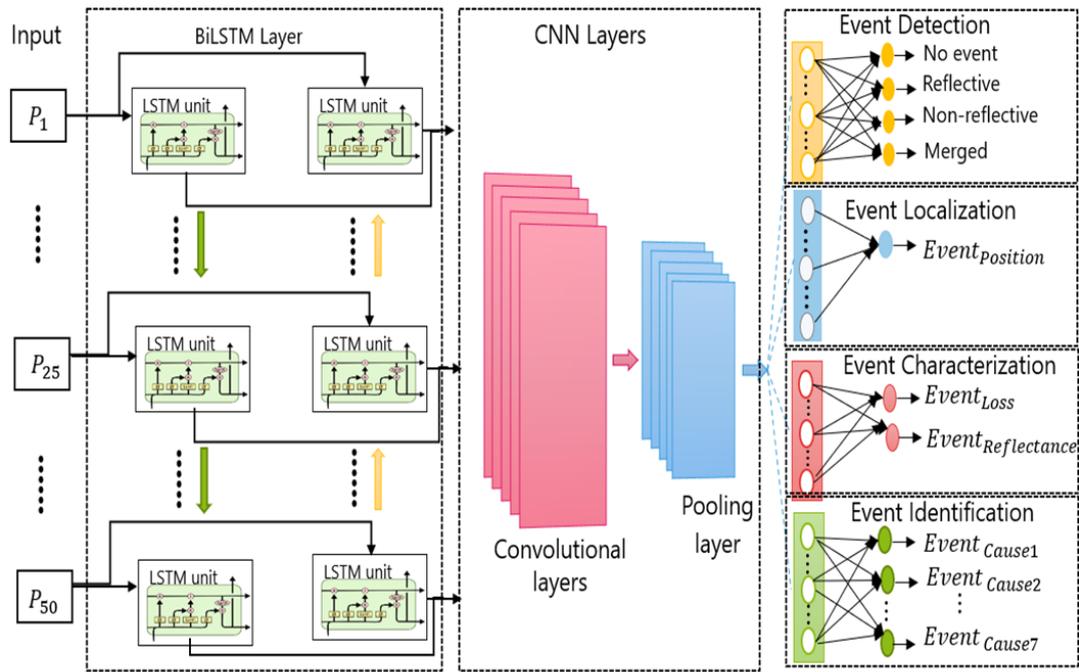


Figure 5.15: Proposed model architecture [277].

#### 5.1.2.4 Results

Several metrics including the detection rate (i.e. detection probability) for the event detection evaluation, RMSE for the event localization and characterization performance assessment, and the ROC curves summarizing the trade-off between the true positive and false positive rates for different threshold settings for the event cause identification evaluation, are adopted. As shown in Fig. 5.16, the detection probability of various events at a false alarm rate (FAR) of 0.01 increases with SNR, and it approaches 1 for SNR values greater than 10 dB. When SNR is more than 20 dB, the RMSE of position estimation is less than 2 m. The RMSE of position estimation decreases with SNR. As SNR rises, the RMSE of the loss ( $RMSE_L$ ) and reflectance ( $RMSE_R$ ) predictions gets smaller. For SNR values greater than 20 dB, the  $RMSE_L$  is less than 0.3 dB,

whereas the  $RMSE_R$  might be as low as 2 dB. The ROC plotted in Fig. 5.16 demonstrates that the model is capable of differentiating between the various event causes. The proposed model is compared to other ML methods, such as CNN, LSTM, and BiLSTM, using various metrics such as accuracy, F1 score, RMSE, MAE, symmetric mean absolute percentage error (SMAPE), and AUC. The results in Tab. 5.2 show that the proposed method outperforms the aforementioned ML techniques in all the metrics. The proposed framework is compared to R1MSDE on an unseen test dataset

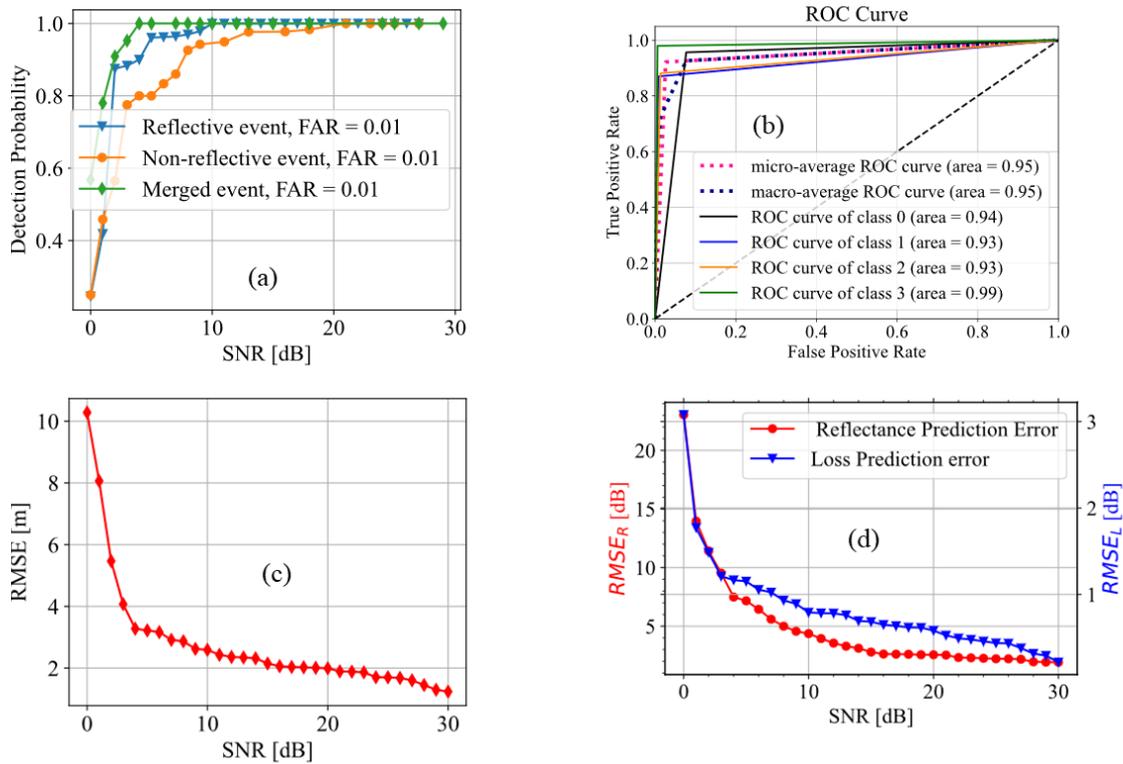


Figure 5.16: Performance of proposed model: a) Fault detection b) Fault localization c) Fault characterization d) Fault cause identification [277].

that contains reflective, non-reflective, and no event sequences. The results shown in Fig. 5.17 demonstrate that the proposed approach outperforms the conventional technique, especially at low SNR values.

### 5.1.3 Combined Approach

The ML-based approaches for OTDR event analysis discussed in the previous subsections have demonstrated significant potential to improve event detection and localization accuracy while also speeding up analysis. When trained on noisy OTDR data with SNR levels ranging from 0 to 30 dB, the technique outperformed conventional techniques in terms of detection and diagnostic capabilities. However, these concepts

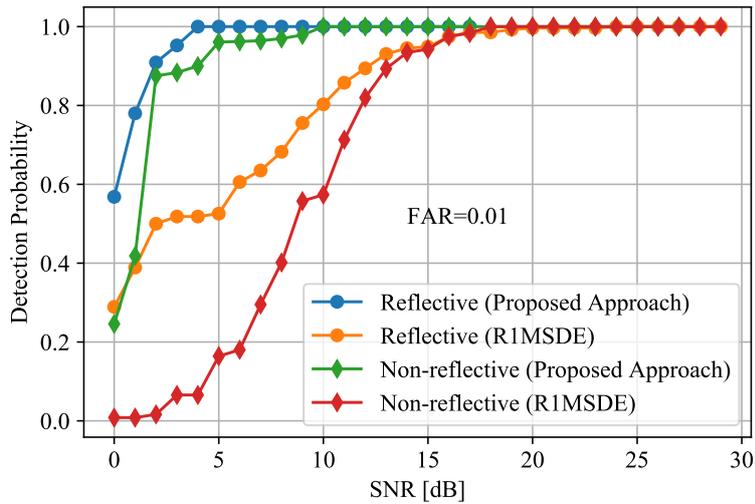


Figure 5.17: ML model vs. conventional method comparison [277].

Metric	CNN	LSTM	BiLSTM	Proposed Method
Task 1: Event detection				
Accuracy (%)	89.5	91.54	91.93	92.32
F1 score	0.89	0.91	0.92	0.92
Task 2: Position estimation				
RMSE (m)	3.31	3.43	2.85	2.73
MAE (m)	2.36	2.46	1.84	1.66
Task 3: Event characterization				
( $RMSE_R$ , $RMSE_L$ )	(10.38, 1.48)	(6.33, 1.03)	(6.15, 1.1)	(4.76, 0.8)
( $SMAPER$ , $SMAPEL$ )	(57.43, 60.94)	(54.59, 55.44)	(55.07, 54)	(53.86, 54)
Task 4: Event identification				
Accuracy (%)	88	89	90	91
AUC	0.92	0.93	0.93	0.94

Table 5.2: Comparison Results of the proposed model with other ML methods

perform poorly at low SNR values (SNR 5 dB), and their generalization and robustness abilities, which underpin the ML model's capability and effectiveness in adapting and reacting to new unseen data, may degrade even further at SNR values lower than 0 dB. Recently, some techniques including ensemble learning [278], structure enhancement (e.g. probabilistic random forest [279]) and Bayesian neural networks [280] have been proposed to improve the learning ability of the ML model in presence of noisy input data. However, these approaches do not completely eliminate overfitting, preventing the model from memorizing the noise patterns. As a result, simply increasing the robustness of the ML model to noise may not significantly improve the detection and diagnostic capability of the ML model under noisy data. As a result, an accurate and reliable event analysis necessitates the removal of noise from the input data while retaining as many significant details of the desired signal as possible in order to improve

the ML model's robust event detection and diagnostic capabilities. The latter requires learning and extracting relevant information and features from a noisy input in order to solve fault detection and diagnosis tasks. To address the aforementioned problem, an approach combining a denoising convolutional autoencoder (DCAE) and BiLSTM is presented in the following, with the former used for noise removal of OTDR signals and the latter used for fault detection, localization, and diagnosis with the denoised signal as input. The method is applied to noisy OTDR signals with input SNR levels ranging from -5 to 15 dB. More details about the architecture of the combined approach and the results as well as the experimental data will be given in the following. The content of this subsection has already been published in [235].

### 5.1.3.1 Proposed Approach

**Denoising Convolutional Autoencoder Model** As shown in Fig. 5.18, the DCAE model for denoising OTDR signals is made up of an encoder and a decoder sub-model with 11 layers each. The encoder receives a noisy OTDR sequence of length 100 as input. It encodes the input into low dimensional features using a series of 5 convolutional layers with 64, 32, 16, 64, and 32 filters of size  $16 \times 1$  with strides of 2, 2, 1, 1, 1. The decoder then reconstructs the output based on the encoder's compressed representation output. It is made up of six transposed convolutional layers. The output is generated by the last transposed convolutional layer with a single filter of size 16 and a stride of 1. The DCAE model's hidden layers have an activation function of exponential linear units (ELU), whereas the output layer has no activation function. Batch normalization is applied to each hidden layer.

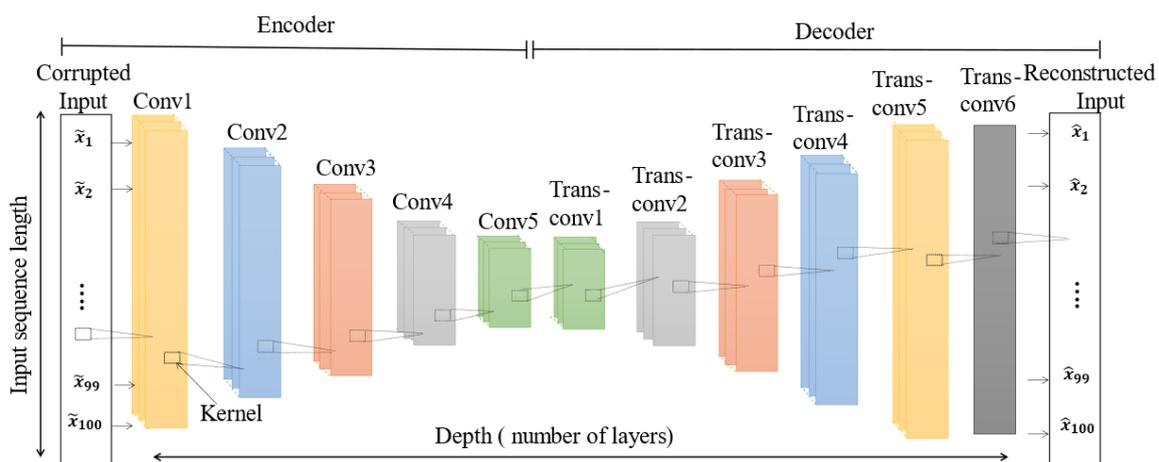


Figure 5.18: Structure of the proposed DCAE model for OTDR signal denoising, Conv: convolutional layer, Trans-conv: transposed convolutional layer [235].

**BiLSTM Model** The model takes an OTDR sequence of length 100 as input and simultaneously outputs the event type ( $E_T$ ), event position ( $E_P$ ), and event cause ( $E_C$ ). The model's architecture, as shown in Fig. 5.19, consists of a shared hidden layer composed of one BiLSTM layer with 32 cells, followed by three task-specific layers made up of 16, 20, and 16 neurons. BiLSTM is chosen as a hidden shared layer because it is well suited to processing OTDR sequential data and capturing long-term dependency.

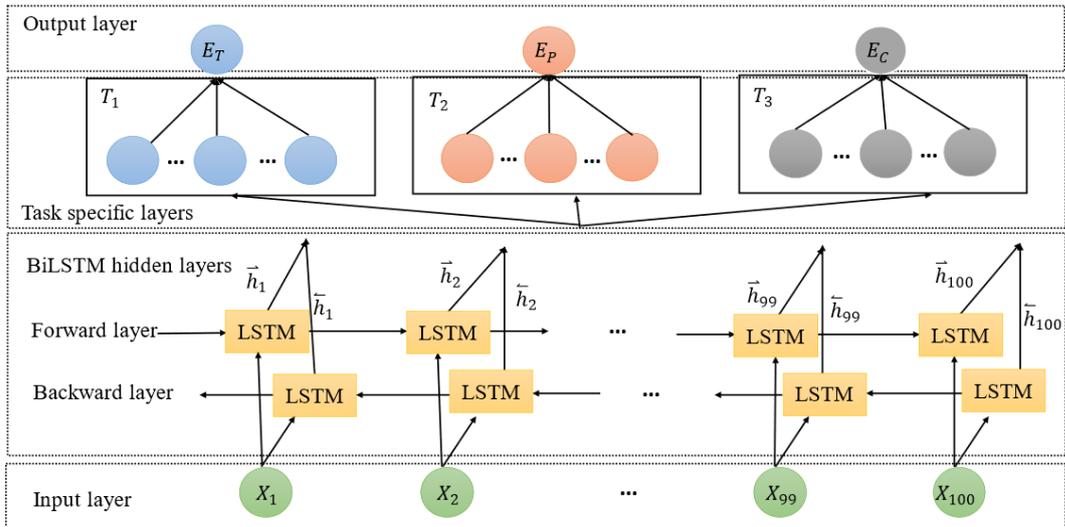


Figure 5.19: Architecture of the proposed multi-task learning based BiLSTM model for fault detection, localization and diagnosis [235].

**Combined Approach** Figure 5.20 depicts the proposed framework, which can be divided into four steps: (1) network monitoring and data collection, (2) dataset creation, (3) DCAE and BiLSTM model training, (4) combining both models, DCAE and BiLSTM, and testing the entire framework using unseen noisy OTDR signals (inference phase). OTDR is used to monitor the fiber optic link by averaging repeated measurements. A noise-free trace is generated for each OTDR setting configuration (pulse width, laser power  $P_{laser}$ ) by first averaging a large number of  $\delta$  OTDR signals to obtain a signal with high SNR and then performing wavelet filtering on the trace, whereas a noisy trace is obtained by averaging a small number of  $\beta$  measurements to obtain a signal with an SNR level ranging from -5 dB to 15 dB. The noisy OTDR traces and their corresponding noise-free traces are split into sequences of length 100 and then normalized. Afterwards, the dataset required for training the DCAE model is produced by combining the noisy sequences (i.e., the input of the DCAE model) and the corresponding noise-free sequences (i.e., the desired output of the DCAE model). For

the training of the BiLSTM model, a dataset containing randomly corrupted noise-free sequences (i.e., the input of the model) and their event characteristics, namely  $E_T, E_P$  and  $E_c$ , labeled as the output of the model, is constructed. The random corruption of the noise-free sequences is carried out by randomly selecting some values of the sequence equal to zero in order to enhance the robustness and improve the generalization capability of the model [280, 235]. The DCAE is then used to denoise the OTDR traces before feeding the denoised signals to the BiLSTM model, which outputs the fault characteristics. To assess its effectiveness, the proposed method is tested on unseen OTDR sequences with SNR values ranging from -5 dB to 15 dB.

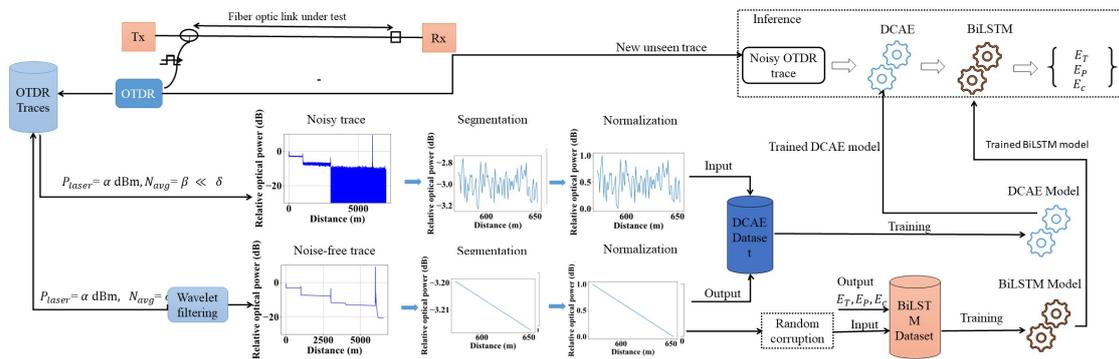


Figure 5.20: Process for training and testing the combination of DCAE and BiLSTM models [235].

### 5.1.3.2 Validation of the proposed Approach

**Experimental Data** The experimental data derived from the experiment setup shown in Fig. 5.13, which has been discussed in the previous subsection, is used To validate the proposed approach. The generated noisy OTDR traces, with SNR values ranging from -5 dB to 15 dB, and their corresponding noise-free traces are split into sequences of length 100 and normalized. For each noise-free sequence, the type  $E_T$  (no event, reflective, non-reflective, merged), the position  $E_P$  (defined as the index within the sequence), and the class  $E_c$  (no event, fiber cut, fiber bend, dirty connector) are assigned. Two datasets: the first dataset containing the noisy and the noise-free sequences, and the second one including the randomly corrupted noise-free sequences and the event characteristics, are built for training the DCAE and BiLSTM models, respectively. Each dataset, which contains 945,172 samples, is divided into three parts: training (60%), validation (20%), and test (20%).

**Validation of the Denoising Capability** The reconstruction error (MSE) and the SNR of the denoised sequence ( $SNR_{out}$ ) are used as evaluation metrics to assess the

performance of the DCAE model. The reconstruction error is defined as the difference between the noise-free sequence and denoised sequences. Figure 5.21 shows that the reconstruction error is very small (less than 0.0035) under all SNR conditions, which demonstrates that DCAE is efficient in noise removal of OTDR signals thanks to its deep architecture. The results shown in Fig. 5.21 prove that the DCAE achieves significant  $SNR_{imp}$  especially for very low SNR levels. To visually assess the performance of

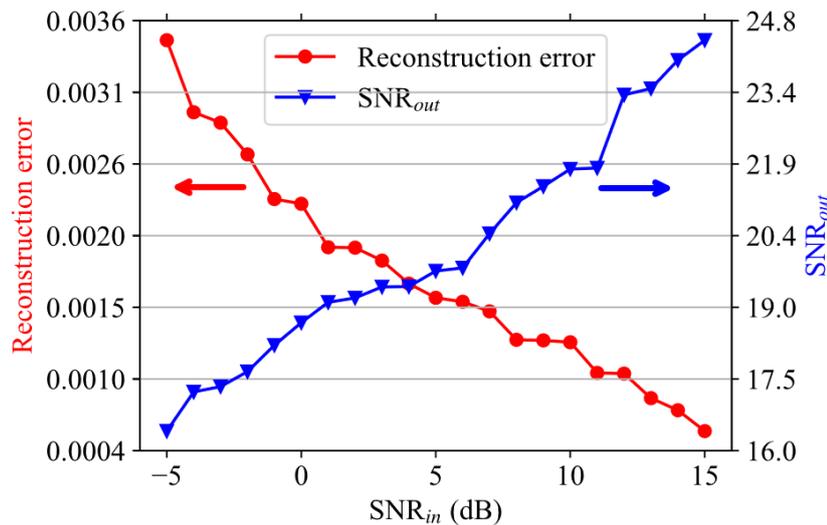


Figure 5.21: Evaluation of DCAE performance using the reconstruction error and the output SNR of the denoised signal as metrics under different input SNR conditions [235].

DCAE, the proposed denoising method is applied to an randomly unseen noisy trace. Figure 5.22 shows that the trace denoised by DCAE is very close to the noise-free trace, which demonstrates the effectiveness of DCAE in noise reduction, which can be very beneficial in improving event analysis. A noisy OTDR sequence incorporating a reflection peak induced by a reflector is considered for denoising using DCAE to investigate the impact of denoising on the spatial resolution measured at the fullwidth half maximum of a reflection event. Figure 5.23 shows that the spatial resolutions of the noisy and denoised sequences are both 5 m, demonstrating that denoising preserves and does not degrade spatial resolution.

**Validation of the Detection and Diagnosis Capability** The proposed combined approach is compared to the BiLSTM model without applying any denoising technique, which is trained with noisy OTDR sequences with SNR values varying from -5 dB to 15 dB. The results shown in Fig. 5.24 demonstrate that (i) the combination of DCAE and BiLSTM achieves better event detection capability compared to the BiLSTM model without denoising, and that (ii) the denoising step conducted by DCAE helps to sig-

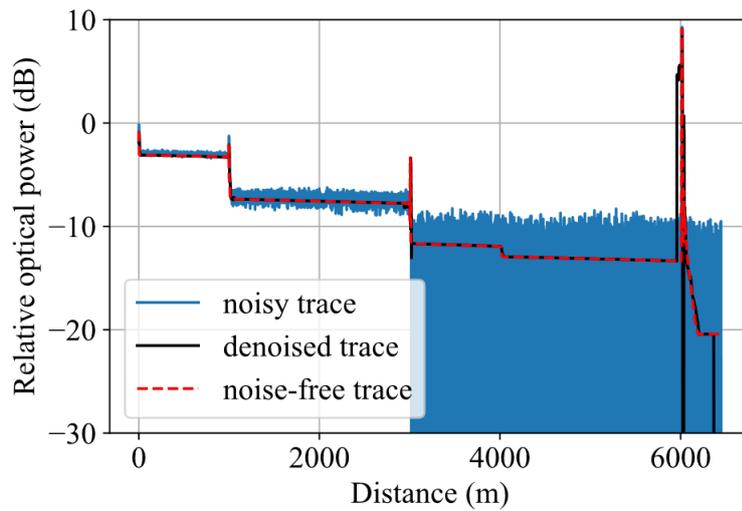


Figure 5.22: Denoising of a random noisy trace using DCAE. The numbers denote the corresponding components shown in Fig. 5.32, inducing the faults [235].

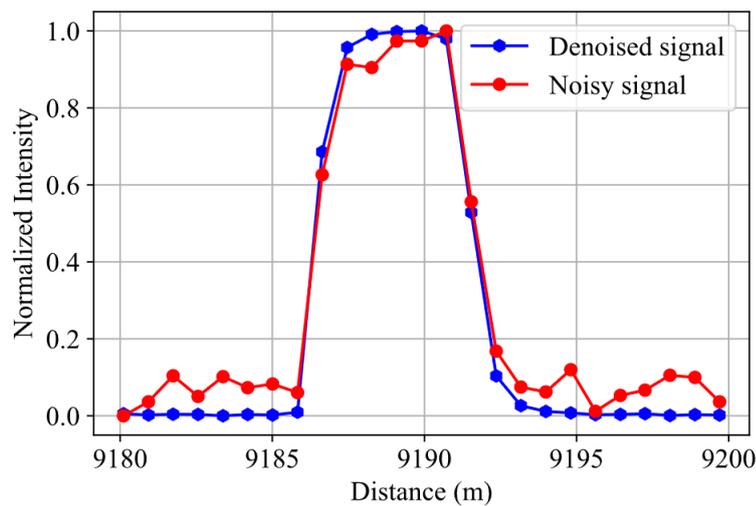


Figure 5.23: Comparison of the spatial resolution without and with denoising [235].

nificantly improve the detection accuracy particularly under very low SNR levels ( $\text{SNR} \leq 0$  dB). For example, the accuracy obtained by the combined method is 91.66% for an input SNR of -1 dB, whereas the accuracy of the BiLSTM model without denoising is only 76.92%, demonstrating that the proposed approach achieves higher event detection performance even at low SNRs.

The comparison of the event localization estimation accuracy of the "DCAE+BiLSTM" combination and BiLSTM without denoising shown in Fig. 5.25 demonstrates that the combined approach achieves smaller event position errors than BiLSTM without denoising. For example, at an SNR level of -4 dB, the combined method produces an event position error of 6 m, whereas BiLSTM without denoising yields an error of 11 m.

The results of the confusion matrices of the event diagnosis for the combined approach and BiLSTM without denoising under an input SNR condition of 0 dB, are illustrated in Fig. 5.26. The overall diagnosis accuracy of the “DCAE+BiLSTM” combination and the BiLSTM model without denoising are 96.37% and 80.13%, respectively. It

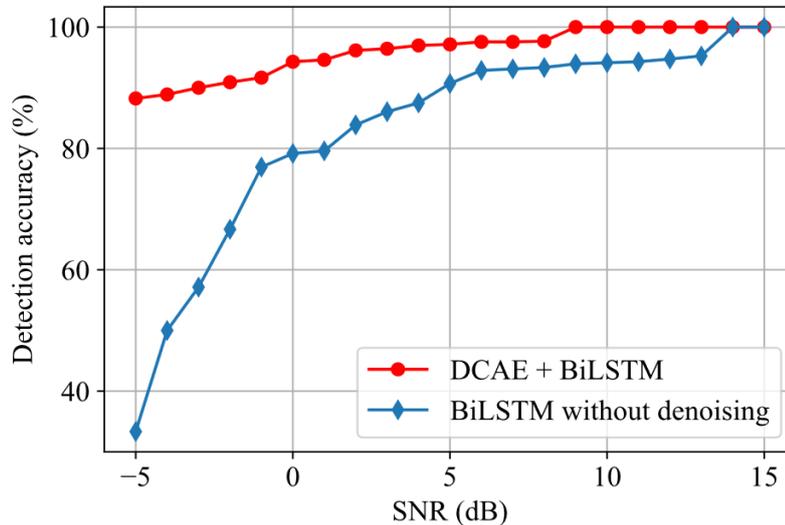


Figure 5.24: Detection accuracy of BiLSTM model with and without denoising [235].

can be seen that the BiLSTM model without denoising misclassifies the fiber bend as a no event because, under low SNR conditions, it is difficult to distinguish the fiber bend’s pattern from the pattern of a normal signal due to noise. Whereas the combined method achieves a higher classification rate of a fiber bend, demonstrating that the DCAE denoising process significantly improves the classification rate of the fault’s cause, particularly for the fiber bend and no event types.

## 5.2 Point-to-multipoint Optical Link Predictive Monitoring

Monitoring PON systems using conventional OTDR can be more challenging than the simple point-to-point links since the backscattered signals from each branch are added together, making it difficult to distinguish between the branches’ backward signals [281]. The event analysis becomes more challenging in the case of equidistant branch terminations because the reflected signals from the same length branches overlap and add up, making it difficult or even impossible to identify the branches. The number of fiber links to be monitored increases as the size of the PON system grows, as does the complexity of the event analysis, leading to a less reliable monitoring [282]. Furthermore, the large splitting loss caused by optical splitters located at the remote node

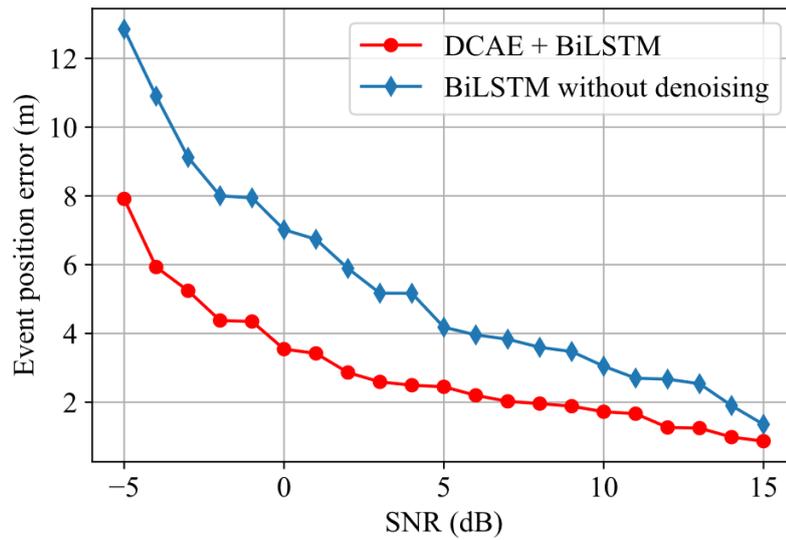


Figure 5.25: Event localization error of BiLSTM model with and without denoising under SNR values varying from -5 dB to 15 dB [235].

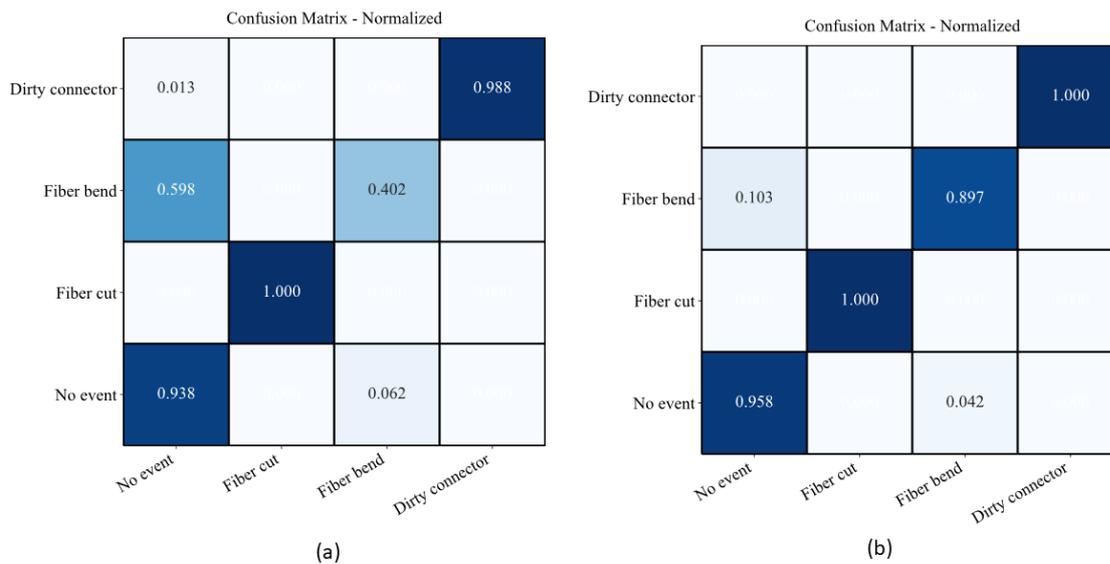


Figure 5.26: Confusion matrices of diagnosis results under SNR condition of 0 dB: (a) the confusion matrix of BiLSTM without denoising, (b) the confusion matrix of DCAE + BiLSTM [235].

results in a significant drop in the backscattered signal (for example, a 1:32 splitter results in a 15 dB loss in measured power [282]). Therefore, a high dynamic range at ODTR and a highly sensitive fault detection method are required for accurate event detection and localization.

This section presents various ML approaches for improving fault monitoring in PON

systems given noisy experimental OTDR signals.

### 5.2.1 Optical Fault Identification, Localization, and Characterization

This subsection discusses a deep learning approach based on GRU autoencoder, referred to as GRU-AE in the following, for identifying, localizing and fully characterizing a wide range of optical fiber faults in PON systems. The content of this subsection has already been published in [283].

#### 5.2.1.1 Experimental Data

The experimental setup shown in Fig. 5.27 is used to record OTDR traces containing various faults in a PON system. Four splitters are used to simulate a real-world PON environment. Multiple faults (i.e. events) including fiber tapping, attenuation splice, dirty connector, fiber bending, fiber break, PC connector and reflector are induced at different locations in the network. Bend radius values of 10 mm, 7.5 mm, 5 mm, and 2.5 mm are used for the fiber bending and tapping events to generate faults with various losses (i.e. different event patterns). To generate various bad splicing fault patterns, different bad attenuation splices with dissimilar losses are performed. Reflectors, PC and open PC connectors generate reflective events with different reflectances. By adjusting the attenuation settings of the VOA from 0 to 30 dB, it is possible to change the height of the reflective peak caused by the reflector. The pulse width, the wavelength, and the sampling time of the OTDR are set to 10 ns, 1650 nm, and 1 ns, respectively.

#### 5.2.1.2 Data Preprocessing

The recorded OTDR traces are divided into sequences of length 30. The event type (no event or one of the aforementioned events), event position index within the sequence, and event reflectance and/or loss are assigned to each sequence. Given that the SNR has a significant impact on the event pattern and the amplitude of the signal is an important characteristic of the reflective event, the SNR ( $\lambda$ ) and a feature  $\delta$  defined as the maximum of the signal amplitude are computed for each sequence. An equal number of samples (instances) for each event are randomly chosen in order to balance the distribution of the event types in the data used to train the ML model. In total, a data set of 125,752 samples with SNR values ranging from 0 to 30 dB is constructed. The data is normalized and split into training, validation, and test datasets in a 60/20/20 ratio.

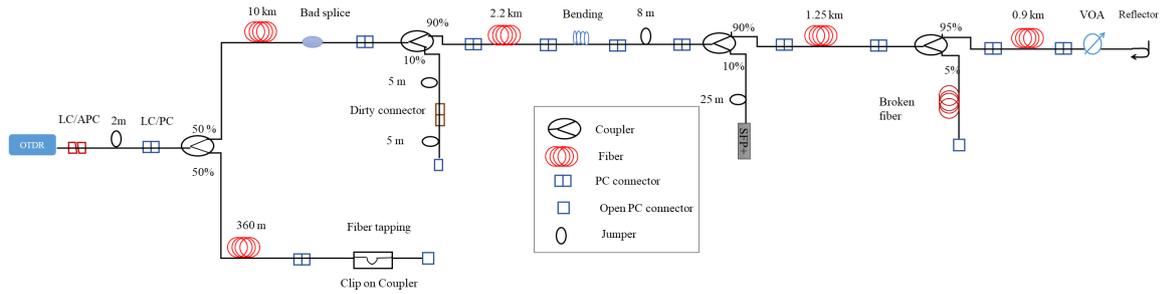


Figure 5.27: Experimental setup for generating OTDR data incorporating several faults at different locations in a PON [283].

### 5.2.1.3 Model Architecture

The architecture of the GRU-AE model is shown in Fig. 5.28. The input of the model is composed of a sequence of power levels (of length 30) combined with the features  $\delta$  and  $\lambda$ . The shared GRU based encoder, which comprises two GRU layers containing 30 and 15 cells, respectively, is used to extract the relevant features modelling the event pattern given the input sequences. The learned representation (extracted features) by the shared encoder is then transferred to several decoders for fault diagnosis, and event localization and characterization. Each decoder is made up of two GRU layers with 15 and 30 cells, respectively, followed by a fully connected layer with 16 neurons. The model's overall loss is calculated as the weighted sum of the four individual decoder losses, which are set to 1, 1.5, 1, and 1, respectively.

### 5.2.1.4 Results and Discussion

Several metrics including the confusion matrix displaying the misclassification rates, the detection probability ( $P_d$ ), and the false alarm rate ( $P_{FA}$ ) are adopted to assess the event detection capability of the GRU-AE model.

The GRU-AE model detects the various faults with high diagnostic accuracy (greater than 93%), as shown in Fig. 5.29. The accuracy for bending and broken fiber events is lower than for other faults because the event patterns of the aforementioned faults are very similar, especially for low SNR sequences, and thus the GRU-AE model misclassified these events. The PC connector is misclassified rarely as either reflector or dirty connector due to the similarity between these reflective events. The GRU-AE model misclassified the reflector sometimes as either dirty connector or open PC connector particularly for high SNR sequences because of the high reflection characterizing these events. For sequences with low SNR values, the height of the reflection is significantly reduced due to the high attenuation set by the VOA, making the pattern of that event look similar to non-reflective events in such cases. As a result, the GRU-AE model

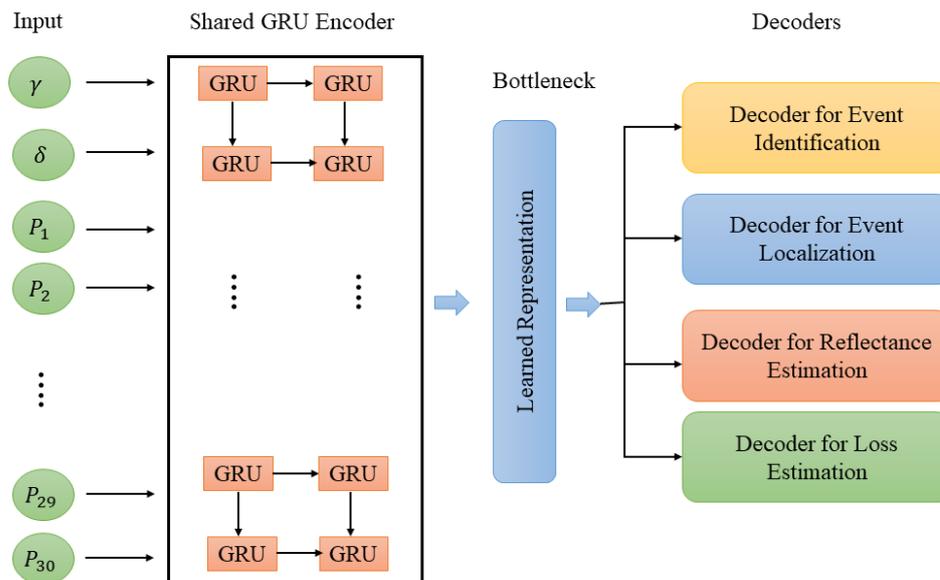


Figure 5.28: Proposed GRU-AE architecture composed of an encoder transferring the learned knowledge to different decoders for fault diagnosis, localization, and characterization [283].

incorrectly classified the reflector as bending, tapping, or bad splice events.

Predicted label	PC connector				0.004		0.006	0.990	
	Reflector		0.002	0.001	0.001	0.022		0.968	0.006
	Fiber break			0.001	0.065		0.932	0.002	
	Dirty connector		0.001			0.983		0.012	0.004
	Bending			0.001	0.931		0.067	0.001	
	Bad splice			0.998	0.001			0.001	
	Tapping	0.004	0.995					0.001	
	No event	0.997	0.002					0.001	
		No event	Tapping	Bad splice	Bending	Dirty connector	Fiber break	Reflector	PC connector
		True label							

Figure 5.29: Confusion matrix of the GRU-AE model [283].

As shown in Fig. 5.30, the detection probability  $P_d$  of the various events increases with SNR, and approaches one for SNR values greater than 10 dB. For lower SNR values ( $\text{SNR} \leq 10$  dB),  $P_d$  is lower owing to noise that influences the event pattern and causes

it to resemble random noise spikes (no event pattern). As a result, distinguishing between the events is difficult for the GRU-AE model. The false alarm rate  $P_{FA}$  for the different faults decreases with SNR and is generally low, as shown in Fig. 5.31. The

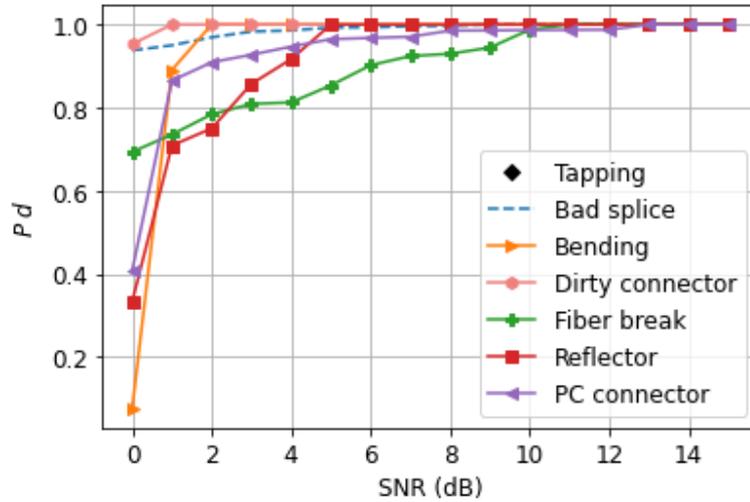


Figure 5.30: Detection probability of the different faults as function of SNR [283].

GRU-AE model's ability to locate the event and characterize it in terms of reflectance and/or loss as a function of SNR is also investigated. The event position error is less than 0.33 m and can be less than 0.17 m for high SNR values, as shown in Fig. 5.32. The loss prediction error is less than 1.4 dB and can reach up to 0.5 dB for high SNR values. The reflectance estimation error is less than 3.5 dB. It can even be reduced to 0.4 dB for high SNR values.

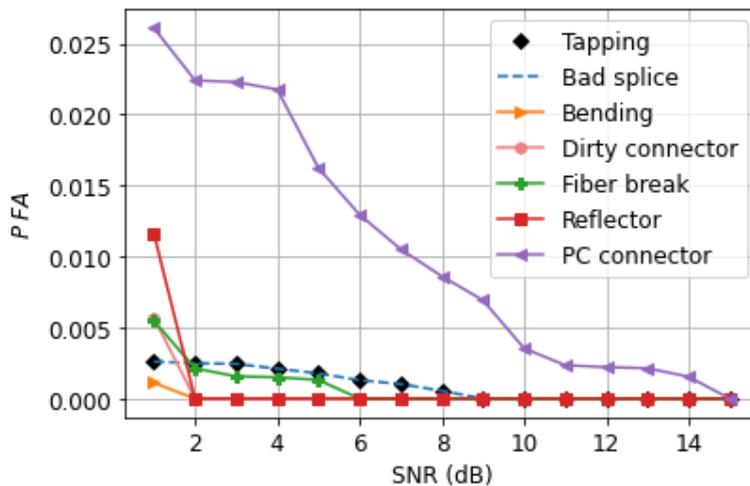


Figure 5.31: False alarm rate of the different faults as function of SNR, false alarm rate of the dirty connector at SNR 0 dB is 0.25 [283].

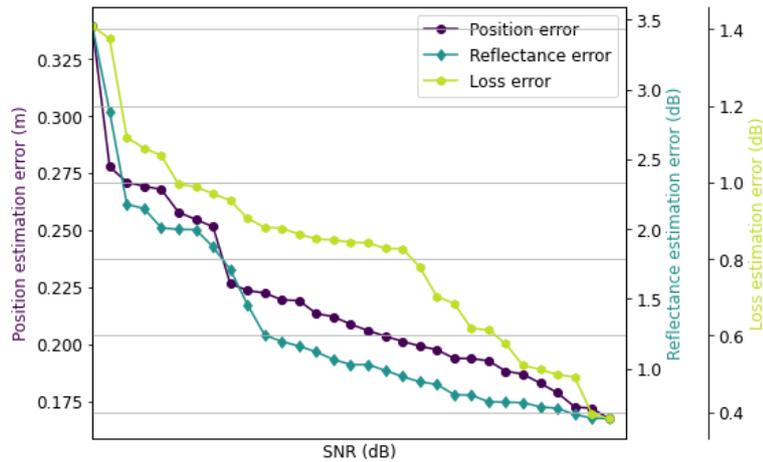


Figure 5.32: Event position, reflectance, and loss estimation error for the GRU-AE model [283].

The GRU-AE approach is compared to a conventional OTDR event analysis method that is based on combining matched templates with typical reflective and non-reflective event patterns, and setting event thresholds, using an unseen test dataset. Please note that the adopted test dataset differs slightly from the training dataset (different laser powers, averaging and SNR values, modified bending radius of the tapping and bending events, different attenuation, slightly changed PON network (one of the splitters is removed) and thus different position of the reflector event). Given that the conventional method detects the event without identifying the type of fault, only the ability of both approaches to detect the fault regardless of the event type is evaluated. The results in Table 5.3 show that the GRU-AE model outperforms the conventional technique in terms of accuracy and RMSE.

Method	Accuracy(%)	RMSE (m)
ML model	99.9	0.8
Conventional method	70	1.43

Table 5.3: Comparison of the GRU-AE approach and conventional method on unseen test dataset using the fault accuracy and the event position error metrics.

### 5.2.2 Anomaly Detection in Optical Fiber Monitoring

An ML model for event recognition, localization, and characterization in PON systems was presented in the previous section. This section discusses a different approach, a combined framework that includes an autoencoder model to detect fiber anomalies and another ML method to recognize and localize the detected fiber faults. The content of this subsection has already been published in [239].

### 5.2.2.1 Proposed Approach

As shown in Fig. 5.33, the proposed framework for fiber monitoring is divided into five major stages: (1) optical fiber network monitoring and data collection, (2) data processing, (3) fiber anomaly detection, (4) fiber fault diagnosis and localization, and (5) mitigation and recovery from fiber failures. OTDR (i.e. fiber monitoring unit) is used to monitor the optical fibers deployed in the network infrastructure on a regular basis. The generated OTDR traces (i.e. monitoring data) are sent periodically to the software-defined networking (SDN) controller managing the infrastructure. The data is then normalized and divided into fixed length sequences. To detect fiber anomalies or faults, the processed data is fed into an ML-based anomaly detection model. If a fiber anomaly is detected, a fault diagnosis and localization ML model is used to diagnose and localize the fault. In order to mitigate the identified fault, a set of recovery rules is applied. When a failure occurs, the SDN controller notifies the network operation center, which notifies the customer about the type of detected fault and its location, as well as the maintenance and repair service in the event of a fiber cut. .

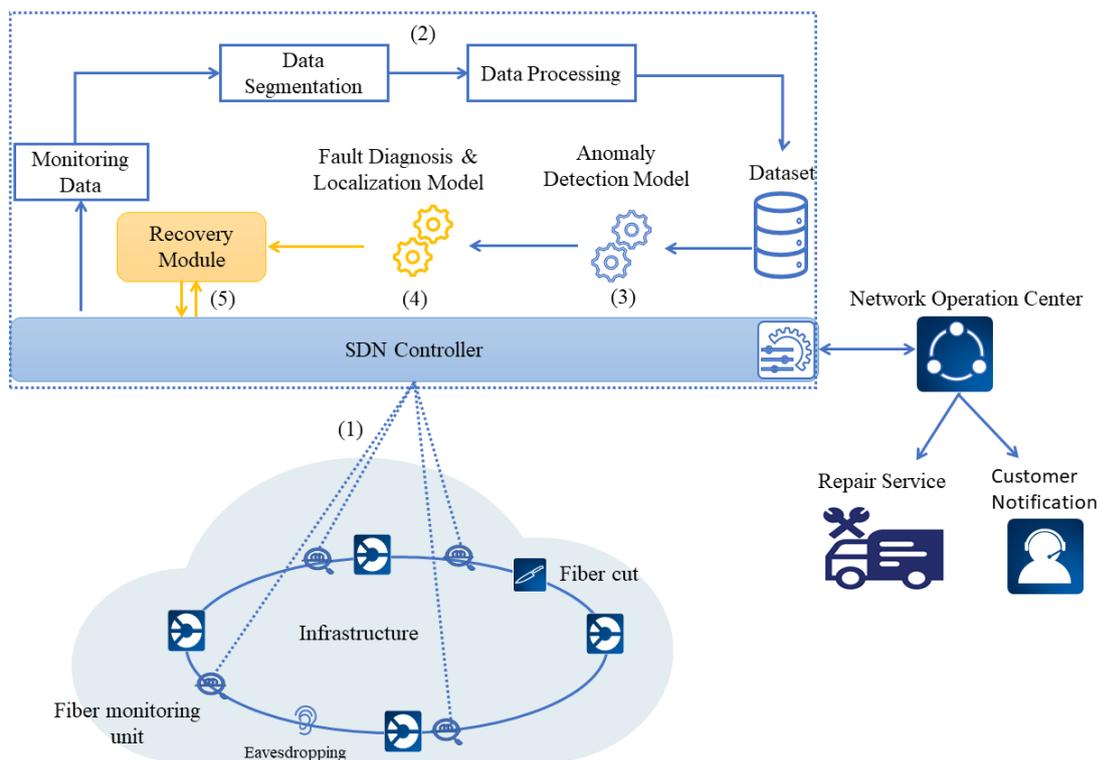


Figure 5.33: Overview of the ML-based fiber monitoring process [239].

**Anomaly Detection Model** The architecture of the proposed autoencoder model for fiber anomaly detection is illustrated in Fig. 5.33. The model is composed of an encoder

and a decoder sub-model with 4 layers. The encoder takes as an input a sequence of length 30 derived from an OTDR trace  $[P_1, P_2, \dots, P_{30}]$ , combined with the SNR ( $\lambda$ ) computed for that sequence. The encoder, which is composed of two GRU layers of 64 and 32 cells, compresses the input into a low-dimensional representation that captures the relevant sequential features modeling the normal state at different SNRs. Afterwards, the decoder reconstructs the output, given the compressed representation. The decoder is inversely symmetric to the encoder part. The cost function is set to MSE, and the Adam optimizer is used to adjust it.

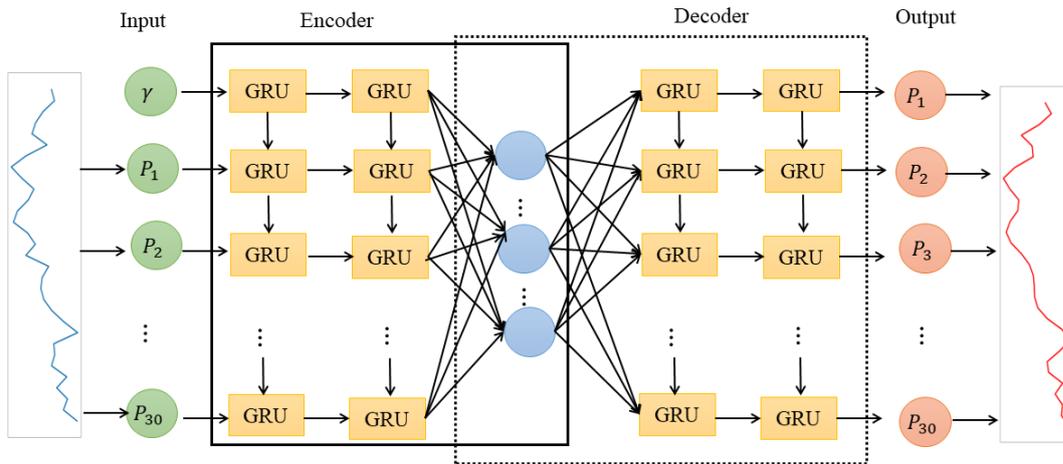


Figure 5.34: Structure of the proposed model for fiber anomaly detection [239].

**Fault Diagnosis and Localization Model** The proposed framework's architecture is made up of shared hidden layers that distribute knowledge across the tasks of fault diagnosis  $T_1$  and the fault position estimation  $T_2$  followed by a specific task layer. The shared hidden layers use a BiGRU network and attention mechanisms. The BiGRU is used to capture the sequential features that characterize each fault's pattern, whereas the attention mechanisms assist the model in focusing more on the relevant features to improve fault diagnosis and localization accuracy. As shown in Fig. 5.35, the input (i.e., the abnormal sample detected by the GRU-based autoencoder) is first fed to two BiGRU layers comprised of 64 and 32 cells, respectively, to capture the relevant sequential features  $[h_1, h_2, \dots, h_{31}]$ . The attention layer assigns to each extracted feature  $h_i$  a weight (i.e., attention score)  $\alpha_i$ . The different computed weights  $\alpha_i$  are aggregated to generate a weighted feature vector  $c$  that captures the relevant information.  $c$  is then transferred to two task-specific layers dedicated to solving the tasks of fault diagnosis ( $T_1$ ) and fault localization ( $T_2$ ), respectively, by leveraging the knowledge extracted by the attention-based BiGRU shared layers. The model is trained by minimizing the loss

function, which is expressed as:

$$L_{total} = \lambda_1 l_{T_1} + \lambda_2 l_{T_2} \quad (5.8)$$

where  $l_{T_1}$  and  $l_{T_2}$  denote the loss of  $T_1$  and  $T_2$ , and the first one is the cross-entropy loss whereas the second one represent the regression loss (MSE). The loss weights  $\lambda_1$  and  $\lambda_2$  denote hyperparameters to be tuned.

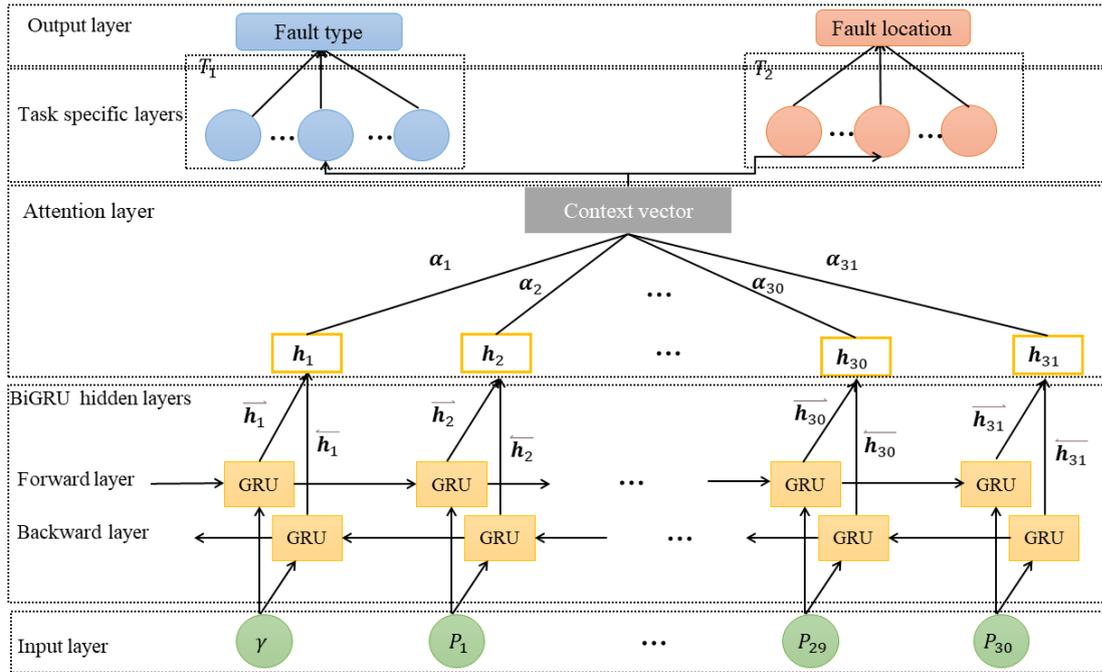


Figure 5.35: Structure of the proposed attention-based bidirectional gated recurrent unit model for fiber fault diagnosis and localization [239].

### 5.2.2.2 Validation of the Proposed Approach

**Data Preprocessing** To validate the proposed framework, the same experimental setup shown in Fig. 5.27 and discussed in Subsection 5.2.1.1 is used to record OTDR signals incorporating the patterns of the different investigated anomalies. The generated OTDR traces are divided into sequences of length 30 and normalized.  $\lambda$  is computed and assigned for each sequence. Only normal state sequences with no fault or normal events induced by optical components are considered for training the GRU-based autoencoder (GRU-AE), whereas normal samples and faulty sequences including an anomaly are used for testing. For training GRU-AE, a dataset of 47,904 samples is created and divided into a training (70%) and a test (30%) dataset. Only faulty sequences are used to train the attention-based BiGRU model. The fault type (fiber

eavesdropping, bad splice, fiber cut, dirty connector) and fault position (defined as the sequence index) are assigned to each sequence. A dataset of 61,849 samples is used for training the fault diagnosis and localization ML model. The said data is split into a training (60%), a validation (20%) and a test dataset (20%).

**Fault Detection Capability** The anomaly detection capability of GRU-AE is optimized by choosing an optimal threshold  $\theta$ . The precision, recall, and F1 score curves as a function of  $\theta$  are depicted in Fig. 5.36. If the threshold is set too low, many faults are classified as normal, resulting in a higher false positive ratio. When the threshold is set too high, many "normal" sequences are classified as "faulty," resulting in a higher false negative ratio. As a result, the optimal threshold is chosen to ensure the best precision/recall tradeoff (i.e., maximizing the F1 score). The precision, recall, and F1 scores for the optimally chosen threshold of 0.008 are 96.9%, 96.86%, and 96.86%, respectively.

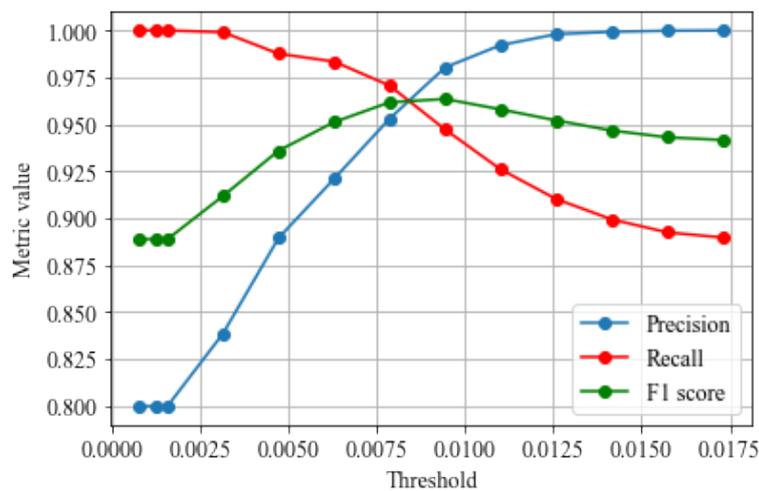


Figure 5.36: The optimal threshold selection based on the precision, recall and F1 scores yielded by GRU-AE [239].

**Comparison of GRU-AE with other ML models** The GRU-AE model is compared to other unsupervised anomaly detection techniques namely isolation forest (IF), LOF, and OCSVM, in terms of F1 score and AUPRC metrics. Like the GRU-AE model, IF, LOF, and OCSVM are trained with normal data and tested with unseen data that includes both normal and abnormal data. The results in Tab. 5.4 show that the GRU-AE model outperforms the other ML methods by yielding the highest F1 and AUPRC scores. In comparison to the tested ML algorithms, GRU-AE improves AUPRC and F1 scores by more than 5.5% and 6.36%, respectively.

Method	F1 score (%)	AUPRC (%)
OCSVM	49.8	94.1
IF	86.9	71.5
LOF	90.5	61.7
GRU-AE	96.86	99.6

Table 5.4: Comparison of different anomaly detection ML methods in terms of area under the precision recall curve (AUPRC) and F1 score.

**Fault Diagnosis Capability** The confusion matrix shown in Fig. 5.37, demonstrates that the attention based BiGRU model (A-BiGRU) diagnoses the different faults with an accuracy higher than 97%. Because the patterns of eavesdropping and bad splice faults in low SNR sequences appear similar, the ML model misclassified these classes slightly. The same applies for dirty connector and fiber cut patterns under low SNR conditions, resulting in low misclassification rates. Figure 5.38 shows the effects of SNR

Predicted label	fiber cut	0.003	0.008	0.988	
	dirty connector	0.001	0.982	0.016	
	bad splice	0.011	0.979	0.009	
	eavesdropping	0.972	0.028	0.001	
		True label			
		eavesdropping	bad splice	dirty connector	fiber cut

Figure 5.37: The confusion matrix of A-BiGRU model [239].

on the diagnosis accuracy of A-BiGRU. For SNR values greater than 10 dB, accuracy approaches 100%. When the SNR is less than 2 dB, the accuracy is worse because it is difficult to distinguish between different faults, owing to noise that disrupts the patterns of the faults, which may appear similar under low SNR conditions. A-BiGRU could discriminate different types of faults with a good accuracy of more than 90% for input sequences with SNR greater than 2dB.

**Fault Localization Capability** To evaluate the A-BiGRU model's fault localization accuracy, two evaluation metrics are used: prediction error (the difference between the

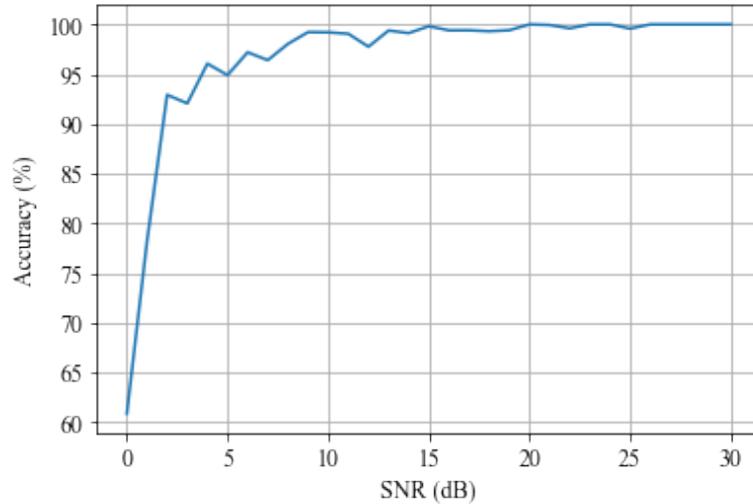


Figure 5.38: The diagnosis accuracy of the A-BiGRU model [239].

predicted and actual position of the fault) and RMSE. Figure 5.39 demonstrates that A-BiGRU achieves very small prediction errors, with a mean of 0.05 m and a standard deviation of 0.2, proving that the ML model accurately localizes the faults. The fault localization estimation performance of A-BiGRU is examined as a function of SNR. A-BiGRU accurately localizes faults with an average root mean square error (RMSE) of 0.19 m, as shown in Fig. 5.40. The RMSE can be higher than 0.35 m for lower SNR values ( $\text{SNR} \leq 10$  dB), but less than 0.1 m for SNR values greater than 27 dB.

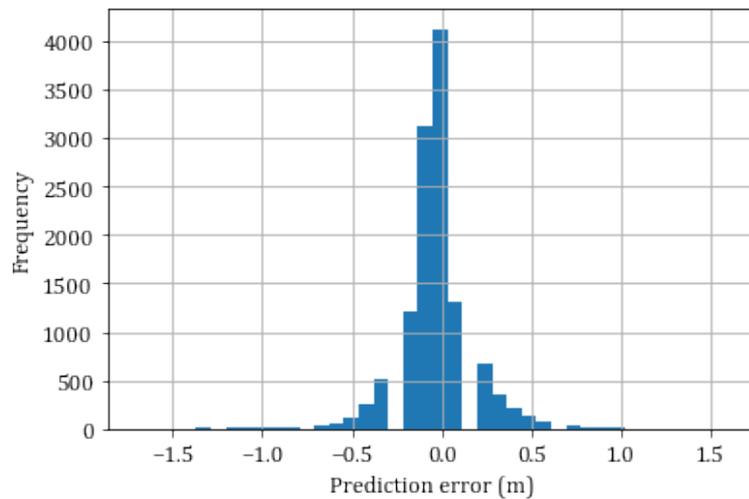


Figure 5.39: Histogram of position prediction errors yielded by A-BiGRU model [239].

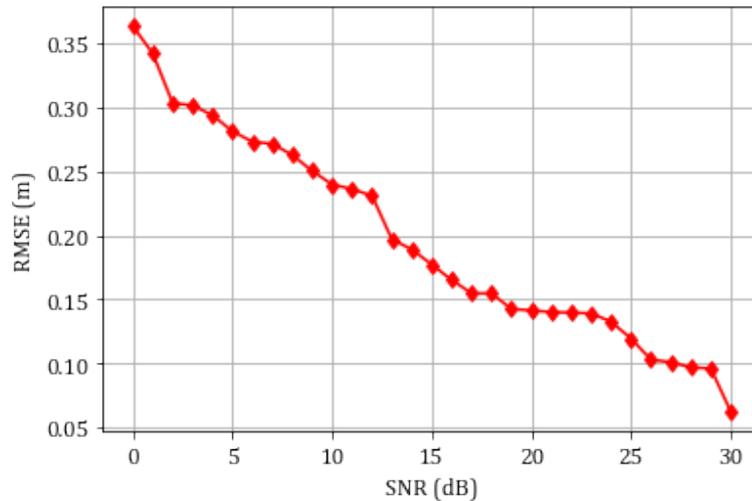


Figure 5.40: Fault position estimation error (RMSE) for the ML model [239].

### 5.3 Conclusion

It has been demonstrated that ML can be an effective tool for optical fault management in optical communication systems. The use of ML for improving fault monitoring in simple point-to-point links and passive optical networks, by leveraging OTDR monitoring data has been investigated. Several ML-based approaches for optical fiber fault detection, localization, characterization, and identification have been proposed and experimentally validated using OTDR data incorporating fault patterns observed at various SNR levels. Each presented ML model takes as input a sequence of an OTDR trace and outputs the fault or event characteristics (i.e., type, location, loss, reflectance, cause, etc.). The proposed approaches achieve good detection and diagnosis capability and high localization accuracy even for low SNR levels, and outperform the conventionally employed techniques. They could help to reduce the measurement time required (less averaging of OTDR measurements) to achieve high detection and localization accuracy. Including the experimental settings, particularly the SNR, as additional input features or performing OTDR signal denoising prior to event analysis could significantly improve the ML model's performance, especially under very low SNR conditions.

## 6 Conclusions and Future Work

In this chapter, research challenges and results of this thesis are concluded. Research limitations are presented to showcase the overall applicability and boundaries of the proposed approaches. Lastly, few potential areas and extensions for future research are given.

### Conclusion

The main focus of this thesis was to develop efficient proactive fault management schemes for optical networks, by making full use of the power of the ML techniques in extracting the useful insights from the monitored data and thus to effectively solve the investigated problems, in order to boost the reliability, the survivability and the resilience of such complex networks. To achieve this goal, several use cases for predictive maintenance in optical communication systems have been investigated with ML methods and validated using either synthetic or experimental data. To this end, it has been worked on the following two questions:

**[RQ1] Could machine learning techniques early and accurately predict the degradation and the lifetime of optical components deployed in an optical communication system?**

**[RQ2] Could machine learning techniques quickly and accurately detect, localize, and identify the faults in optical fiber links?**

It has been shown that ML is a powerful tool to boost the reliability of semiconductor lasers (one of the most critical optical components in an optical communication system) deployed in optical networks. Several ML-based approaches have been presented for solving different predictive maintenance use cases for semiconductor lasers including early degradation prediction, RUL prediction, anomaly detection, and early failure mode identification. The different ML models have been applied to reliability data for various types of semiconductor lasers including VCSELs, DFBs, and tunable edge emitting lasers. Dependently on the investigated use case and the availability of the

reliability data, the input to be fed to the ML model could be historical current (considered degradation parameter) measurements collected under constant output power or output power (regarded as degradation parameter) measurements monitored under constant current operation regime combined with operating conditions or monitored laser characteristics. The presented approaches outperform the conventional methods by achieving better prediction accuracy and earlier detection capability. They enable early and efficient maintenance plan scheduling and thus help to minimize the downtime, and to reduce the maintenance costs.

It has been shown that training the ML model with synthetic data similar to real data, could help to overcome the challenge of lack of reliability data due to the lengthy time required to generate a meaningful data set, which might severely impact the performance of the ML model trained with that small amount of data. To generate synthetic data that has similar properties to real data, it has been demonstrated that the GAN is able to produce laser reliability data that is close to the real experimental data, and that the ML model trained with the synthetic data generated by GAN performs well when tested with experimental real data.

It has been proven that autoencoders could help to deal with the challenge of highly unbalanced laser reliability data (the number of the normal laser behavior samples is higher than the number of the failure modes particularly the number of sudden degradation types). Trained only with normal data modelling the normal laser behavior, the autoencoder model reconstructs the normal instances very well by achieving small errors, while it fails to reproduce the faulty observations (unseen during the training phase) by yielding much greater reconstruction errors. The fault classification threshold to discriminate between the normal laser operation and failure is optimized given the data including both normal and faulty samples. Whereas to distinguish between the different failure modes, another threshold has to be selected given data incorporating sudden and gradual degradation types.

It has been demonstrated that including the operating conditions or specific laser parameters, or some statistical features characterizing the laser degradation trend as additional features combined with the historical current or output power measurements during the training phase could boost the performance of the ML model as such additional features could help the ML model to capture more relevant features underlying the patterns under different conditions.

The presented work in this thesis related to laser predictive maintenance could serve as a great motivation for laser manufactures to adopt ML in their business in order to enhance the reliability of their products and reduce costs. Adopting ML approaches could help them to reduce the time and the costs of carrying out the burn-in or accelerated aging tests for a long time (the aging time can be shortened by 66.6%).

It has been demonstrated that ML methods accurately and quickly detect, diagnose, localize and characterize fiber faults in optical fiber links. Different ML models have been proposed for optical fiber management in different optical communication systems including point-to-point links and passive optical networks. The ML approaches have been validated using experimental noisy OTDR data incorporating different fiber events. The ML models typically take as input a sequence of an OTDR trace combined with SNR and output the type of fiber fault, its location and/or its characteristics in terms of reflectance and/or loss. Such models outperform the state-of-the-art baselines in terms of event detection and localization accuracy particularly under low SNR conditions.

It has been shown that denoising OTDR signals using ML methods before performing event analysis could significantly enhance the performance of the ML model under very low SNR conditions varying from -5 dB to 15 dB. Removing the noise from OTDR traces using an ML method (e.g. convolutional denoising autoencoder) before carrying out event analysis using another ML model could provide an improvement of 13.74% in event detection and diagnosis accuracy compared to the performance of the same ML model trained with noisy OTDR signals, which demonstrates the importance of the denoising step to achieve accurate fault analysis under noisy SNR conditions. It has been proven as well that ML methods are efficient in reducing the noise from OTDR signals without causing loss in the signal information.

It has been demonstrated that adding the experimental settings and particularly the SNR as additional inputs combined with the OTDR sequence during the learning phase could improve the performance of the ML model. For example, including SNR could lead to up 5% increase in the event detection accuracy. Adding event characteristics (e.g., the height of the event) as additional features could help the ML model to discriminate between similar event patterns such as open PC connector, dirty connector, or reflector.

Adopting the proposed ML models for optical fiber fault management given OTDR

signals could help optical network operators to enhance the monitoring of their infrastructures and to reduce the mean time to repair by quickly discovering and pinpointing the link faults, including fiber cuts and thus meeting service level agreements more easily, and improving the customer satisfaction by minimizing the downtimes and enhancing the network quality.

Our main conclusion in this thesis is that machine learning is an efficient tool for enhancing the proactive maintenance in optical networks and could accelerate the path towards a self-healing optical network with high reliability, survivality and resilience.

### **Research Limitations**

This research has some limitations. The experimental data adopted for the validation of the ML approaches for enhancing the semiconductor laser reliability is derived from accelerated aging tests conducted under stressed conditions due to the prohibitively long time required to obtain device failure data under normal operating conditions. However, the degradation trend and the failure patterns in the field under normal conditions might look different to those at extreme conditions, and the performance of the ML models may be adversely affected. Therefore, the ML approaches to be adopted in field during operation have to be re-trained with field monitoring data from the operating devices, incorporating historical measurements of either the current or the output power values.

The experimental OTDR data is generated for specific OTDR configuration settings (e.g. fixed pulse width). Therefore, the proposed ML approaches for fault management in optical links might not perform well, if a test data set produced for different OTDR configuration settings is given, as the patterns under such settings might look different to the those under the considered settings (e.g. the shape and the width of the faults might look different). Hence, the need to re-train the ML models for distinct OTDR configuration settings. Furthermore, as it can be prohibitively expensive and cumbersome to obtain data representing all types of faults or anomalies generated for different scenarios or settings, the ML model for fault identification has to be re-trained by providing new data incorporating the new patterns for the investigated faults or completely new events to be considered. Generally, the proposed ML approaches have to be continually evolved to adapt to changing network conditions by performing a periodic re-training.

## Future Work

Future work could aim to perform the interpretability and the explainability of the proposed ML approaches in this thesis in order to demystify the black-box ML models and try to better understand the outputs of the model as well as the impact of the features or the input on the results.

Further future research work could be to investigate the use of the ML techniques to boost the reliability of other optical components such as optical amplifiers and optical receiver (particularly the photodiode device). The same concepts of the proposed approaches for the different use cases for semiconductor laser predictive maintenance could be readily applicable to the other optoelectronic devices (particularly semiconductor optical amplifiers) due to the similarity of their structures.

Another important research could be to explore the adoption of the ML methods for solving the problem of faulty branch identification in passive optical networks for the cases of equidistant terminations (very challenging scenarios). Another interesting research could aim at automating the OTDR trace analysis for different optical network architectures by making full use of the power of the ML in performing event recognition tasks.

## Bibliography

- [1] A. Ellis and M. Sorokina. *Optical Communication Systems: Limits and Possibilities*. Jenny Stanford Publishing, 2019. ISBN: 9780429650901. URL: <https://books.google.de/books?id=Lb0tDwAAQBAJ>.
- [2] Lena Wosinska et al. “Network Resilience in Future Optical Networks”. In: Apr. 2009, pp. 253–284. ISBN: 978-3-642-01523-6. DOI: 10.1007/978-3-642-01524-3\_10.
- [3] Francesco Musumeci et al. “A Tutorial on Machine Learning for Failure Management in Optical Networks”. In: *Journal of Lightwave Technology* 37.16 (2019), pp. 4125–4139. DOI: 10.1109/JLT.2019.2922586.
- [4] Khouloud Abdelli et al. “Reflective fiber fault detection and characterization using long short-term memory”. In: *Journal of Optical Communications and Networking* 13.10 (2021), E32–E41. DOI: 10.1364/JOCN.423625.
- [5] Myeong-Kyu Song. “Design and Management of Survivable Network: Concepts and Trends”. In: *International Journal of Contents* 5 (June 2009), pp. 43–52. DOI: 10.5392/IJoC.2009.5.2.043.
- [6] Chongjin Xie et al. “Characteristics of Field Operation Data for Optical Transceivers in Hyperscale Data Centers”. In: *Optical Fiber Communication Conference (OFC) 2022*. Optica Publishing Group, 2022, Th2A.1. DOI: 10.1364/OFC.2022.Th2A.1. URL: <http://opg.optica.org/abstract.cfm?URI=OFC-2022-Th2A.1>.
- [7] Juan Jiménez. “Laser diode reliability: crystal defects and degradation modes”. In: *Comptes Rendus Physique* 4.6 (2003). Dossier: Semiconductor lasers, pp. 663–673. ISSN: 1631-0705. DOI: [https://doi.org/10.1016/S1631-0705\(03\)00097-5](https://doi.org/10.1016/S1631-0705(03)00097-5). URL: <https://www.sciencedirect.com/science/article/pii/S1631070503000975>.
- [8] Jia-Sheng Huang. “Temperature and current dependences of reliability degradation of buried heterostructure semiconductor lasers”. In: *IEEE Transactions on Device and Materials Reliability* 5.1 (2005), pp. 150–154. DOI: 10.1109/TDMR.2005.843834.

- [9] Wonkyoung Lee et al. “Identification method of non-reflective faults based on index distribution of optical fibers”. In: *Opt. Express* 22.1 (Jan. 2014), pp. 325–337. DOI: 10.1364/OE.22.000325. URL: <http://opg.optica.org/oe/abstract.cfm?URI=oe-22-1-325>.
- [10] Kwan Chi Kao and G. A. Hockham. “Dielectric-fibre surface waveguides for optical frequencies”. In: 1966.
- [11] Kapron F.P., Keck D.B., and Maurer R.D. “Radiation Losses in Glass Optical Waveguides”. In: *Applied Physics Letters* 17 (1970), pp. 423–425.
- [12] Erik Agrell et al. “Roadmap of optical communications”. In: *Journal of Optics* 18.6 (May 2016), p. 063002. DOI: 10.1088/2040-8978/18/6/063002. URL: <https://doi.org/10.1088/2040-8978/18/6/063002>.
- [13] E. Desurvire. “Erbium-doped fiber amplifiers for new generations of optical communication systems”. In: *Opt. Photon. News* 2.1 (Jan. 1991), p. 6. DOI: 10.1364/OPN.2.1.000006. URL: <http://www.optica-opn.org/abstract.cfm?URI=opn-2-1-6>.
- [14] R. Hui. *Introduction to Fiber-Optic Communications*. Elsevier Science, 2019. ISBN: 9780128053454. URL: <https://books.google.de/books?id=20acDwAAQBAJ>.
- [15] Charles K. Kao. “SAND FROM CENTURIES PAST: SEND FUTURE VOICES FAST”. In: *Nobel Lecture* (Dec. 2009).
- [16] J. M and M.Y. Jamro. *Optical Fiber Communications: Principles and Practice*. Prentice Hall Internacional series in optoelectronics. Financial Times/Prentice Hall, 2009. ISBN: 9780130326812. URL: <https://books.google.de/books?id=7HtqsEOMkXUC>.
- [17] G. Keiser. *Optical Fiber Communications*. McGraw-Hill Companies, 2010.
- [18] G. Keiser. *Fiber Optic Communications*. Springer Nature Singapore, 2021. ISBN: 9789813346659. URL: [https://books.google.de/books?id=5%5C\\_ggEAAAQBAJ](https://books.google.de/books?id=5%5C_ggEAAAQBAJ).
- [19] G.P. Agrawal. *Fiber-Optic Communication Systems*. Wiley Series in Microwave and Optical Engineering. Wiley, 2012. ISBN: 9780470922828. URL: <https://books.google.tn/books?id=yGQ4n1-r2eQC>.
- [20] J J Coleman. In: 27.9 (Aug. 2012), p. 090207. DOI: 10.1088/0268-1242/27/9/090207. URL: <https://doi.org/10.1088/0268-1242/27/9/090207>.
- [21] I Hayashi et al. “Junction lasers which operate continuously at room temperature”. In: *Appl. Phys. Lett.* 17 (1970), p. 109.

- [22] K. Chang. *Handbook of Optical Components and Engineering*. Wiley, 2003. ISBN: 9780471390558. URL: <https://books.google.de/books?id=ChFtQgAACAAJ>.
- [23] C. Headley and Govind Agrawal. *Raman Amplification in Fiber Optical Communication Systems*. Jan. 2005. DOI: 10.1016/B978-0-12-044506-6.X5000-2.
- [24] J.E. Carroll et al. *Distributed Feedback Semiconductor Lasers*. IEE circuits and systems series: Institution of Electrical Engineers. Institution of Electrical Engineers, 1998. ISBN: 9780852969175. URL: <https://books.google.de/books?id=2yHc-485ov8C>.
- [25] G. Morthier and P. Vankwikelberge. *Handbook of Distributed Feedback Laser Diodes, Second Edition*. Optoelectronics. Artech House, 2013. ISBN: 9781608077014. URL: <https://books.google.de/books?id=0RggAgAAQBAJ>.
- [26] J.R. Harwell and W. Zeller. *Distributed Feedback Laser Diodes*. Excelic Press LLC, 2018. ISBN: 9781642241167. URL: <https://books.google.de/books?id=DoXvtQEACAAJ>.
- [27] T.E. Sale. *Vertical Cavity Surface Emitting Lasers*. Electronic & electrical engineering research studies: Optoelectronics series. Research Studies Press, 1995. ISBN: 9780863801747. URL: <https://books.google.de/books?id=2ONRAAAAMAAJ>.
- [28] H. Li and K. Iga. *Vertical-Cavity Surface-Emitting Laser Devices*. Springer Series in Photonics. Springer Berlin Heidelberg, 2002. ISBN: 9783540678519. URL: <https://books.google.de/books?id=3p09QUR13jAC>.
- [29] C.W. Wilmsen, H. Temkin, and L.A. Coldren. *Vertical-Cavity Surface-Emitting Lasers: Design, Fabrication, Characterization, and Applications*. Cambridge Studies in Modern Optics. Cambridge University Press, 2001. ISBN: 9780521006293. URL: <https://books.google.de/books?id=lysu40v0E1YC>.
- [30] *Principles Of Laser Action*. URL: <https://www.worldoflasers.com/laserprinciples.htm> (visited on 2022).
- [31] Subhash Chandra Singh et al. “Lasers: Fundamentals, Types, and Operations”. In: 2012.
- [32] Bahaa Saleh and Malvin Teich. *Fundamentals of Photonics, 1st Edition*. Oct. 1991. ISBN: 9780471839651.
- [33] R. Paschotta. *Field Guide to Lasers*. Field Guide Series. Society of Photo Optical, 2008. ISBN: 9780819469618. URL: <https://books.google.de/books?id=02s6wyYULYAC>.

- [34] Jonas, Jay Newman, and Seyffie Maleki. *Population Inversion can be Achieved in a Three-Level System*. URL: [https://chem.libretexts.org/Courses/Pacific\\_Union\\_College/Quantum\\_Chemistry/15%3A\\_Lasers\\_Laser\\_Spectroscopy\\_and\\_Photochemistry/15.04%3A\\_Population\\_Inversion\\_can\\_be\\_Achieved\\_in\\_a\\_Three-Level\\_System](https://chem.libretexts.org/Courses/Pacific_Union_College/Quantum_Chemistry/15%3A_Lasers_Laser_Spectroscopy_and_Photochemistry/15.04%3A_Population_Inversion_can_be_Achieved_in_a_Three-Level_System) (visited on 2022).
- [35] L. Fichera. *Cognitive Supervision for Robot-Assisted Minimally Invasive Laser Surgery*. Springer Theses. Springer International Publishing, 2016. ISBN: 9783319303307. URL: <https://books.google.de/books?id=AtlCDwAAQBAJ>.
- [36] Ivan Karomi. “Design and Analysis of Quantum Dot Laser (InAsP) for Biophotonic and Mode-Locking Applications”. PhD thesis. May 2018. DOI: 10.13140/RG.2.2.33731.55847.
- [37] *Semiconductor Technology from A to Z Everything about semiconductors and wafer fabrication, Fundamentals: Conductors – Insulators – Semiconductors*. URL: <https://www.halbleiter.org/en/fundamentals/conductors-insulators-semiconductors/> (visited on 2022).
- [38] Electricalvoice. *Difference between Conductor, Semiconductor and Insulator*. URL: <https://electricalvoice.com/difference-between-conductor-semiconductor-and-insulator/> (visited on 2022).
- [39] S. Kumar and M.J. Deen. *Fiber Optic Communications: Fundamentals and Applications*. Wiley, 2014. ISBN: 9780470518670. URL: <https://books.google.de/books?id=yxvnAgAAQBAJ>.
- [40] S. L. Chuang. *Physics of Optoelectronic Devices*. Wiley, New York, 1995.
- [41] *INTRODUCTION TO SOLID STATE PHYSICS, 7TH ED*. Wiley India Pvt. Limited, 2007. ISBN: 9788126510450. URL: [https://books.google.at/books?id=F9Qu5c%5C\\_hUaUC](https://books.google.at/books?id=F9Qu5c%5C_hUaUC).
- [42] *Fundamentals: Doping: n- and p-semiconductors*. URL: <https://www.halbleiter.org/en/fundamentals/doping/> (visited on 2022).
- [43] *PN Junction in Semiconductor Diodes*. URL: <https://www.electronicclinic.com/pn-junction-in-semiconductor-diodes/> (visited on 2022).
- [44] *Width of depletion region*. URL: <https://www.physics-and-radio-electronics.com/electronic-devices-and-circuits/semiconductor-diodes/widthofdepletionregion.html> (visited on 2022).
- [45] C.F. Lin. *Optical Components for Communications: Principles and Applications*. Springer US, 2013. ISBN: 9781475741780. URL: <https://books.google.de/books?id=H4XSBwAAQBAJ>.

- [46] *Semiconductor Diode laser: Principle, Construction, Working, Characteristics, Advantages, Disadvantages and Applications*. URL: [https://www.brainkart.com/article/Semiconductor-Diode-laser--Principle,-Construction,-Working,-Characteristics,-Advantages,-Disadvantages-and-Applications\\_6886/](https://www.brainkart.com/article/Semiconductor-Diode-laser--Principle,-Construction,-Working,-Characteristics,-Advantages,-Disadvantages-and-Applications_6886/) (visited on 2022).
- [47] *Semiconductor laser construction and working*. URL: <https://physicswave.com/semiconductor-laser-construction-and-working/> (visited on 2022).
- [48] *Semiconductor laser*. URL: <https://sites.google.com/site/puenggphysics/home/unit-i/semiconductor-laser> (visited on 2022).
- [49] EUGENE I. GORDON. "Semiconductor lasers—past, present, and future". In: *Conference on Lasers and Electro-Optics*. Optica Publishing Group, 1987, TUN4. URL: <http://opg.optica.org/abstract.cfm?URI=CLEO-1987-TUN4>.
- [50] *SINGLE-MODE SEMICONDUCTOR LASERS*. URL: <https://www.fiberoptics4sale.com/blogs/wave-optics/single-mode-semiconductor-lasers> (visited on 2022).
- [51] F.J. Duarte. *Tunable Lasers Handbook*. Optics and Photonics. Elsevier Science, 1996. ISBN: 9780080519760. URL: <https://books.google.de/books?id=NX1HyVXDHTMC>.
- [52] K. Cheng et al. *Tunable Lasers*. Topics in Applied Physics. Springer Berlin Heidelberg, 2013. ISBN: 9783662106358. URL: <https://books.google.de/books?id=1-7wCAAAQBAJ>.
- [53] *Tuning in to tunable lasers*. URL: <https://www.laserfocusworld.com/lasers-sources/article/16556723/tuning-in-to-tunable-lasers> (visited on 2022).
- [54] T.E. Sale. *Vertical Cavity Surface Emitting Lasers*. Research Studies Press, 1995.
- [55] *VCSEL LASER – THE ADVANTAGES OF VCSEL*. URL: <https://www.fiberoptics4sale.com/blogs/archive-posts/95051078-vcsel-laser-the-advantages-of-vcsel> (visited on 2022).
- [56] Anders Larsson. "Advances in VCSELs for Communication and Sensing". In: *Selected Topics in Quantum Electronics, IEEE Journal of* 17 (Jan. 2012), pp. 1552–1567. DOI: 10.1109/JSTQE.2011.2119469.
- [57] Fumio Koyama. "Recent Advances of VCSEL Photonics". In: *J. Lightwave Technol.* 24.12 (Dec. 2006), pp. 4502–4513. URL: <http://opg.optica.org/jlt/abstract.cfm?URI=jlt-24-12-4502>.

- [58] *LASER DIODE AND VCSEL TEST: High-throughput DC production test techniques for laser-diode modules and VCSELs*. URL: <https://www.laserfocusworld.com/test-measurement/test-measurement/article/16556878/laser-diode-and-vcsel-test-highthroughput-dc-production-test-techniques-for-laserdiode-modules-and-vcseles> (visited on 2022).
- [59] *Slope efficiency*. URL: [https://www.rp-photonics.com/slope\\_efficiency.html](https://www.rp-photonics.com/slope_efficiency.html) (visited on 2022).
- [60] H. Kressel. *Semiconductor Devices for Optical Communication*. Topics in Applied Physics. Springer Berlin Heidelberg, 2006. ISBN: 9783540347590. URL: <https://books.google.de/books?id=QoiQDAAAQBAJ>.
- [61] J.L. Spencer. “Calculating laser diode reliability”. In: *IEEE Global Telecommunications Conference and Exhibition. Communications for the Information Age*. 1988, 63–69 vol.1. DOI: 10.1109/GLOCOM.1988.25811.
- [62] L.N. Binh. *Optical Modulation: Advanced Techniques and Applications in Transmission Systems and Networks*. Optics and Photonics. CRC Press, 2017. ISBN: 9781498745246. URL: <https://books.google.de/books?id=IGpQDwAAQBAJ>.
- [63] *Practical Uses and Applications of Electro-Optic Modulators*. URL: <https://www.newport.com/n/practical-uses-and-applications-of-electro-optic-modulators> (visited on 2022).
- [64] Peter J. Delfyett. “Lasers, Semiconductor”. In: *Encyclopedia of Physical Science and Technology (Third Edition)*. Ed. by Robert A. Meyers. Third Edition. New York: Academic Press, 2003, pp. 443–475. ISBN: 978-0-12-227410-7. DOI: <https://doi.org/10.1016/B0-12-227410-5/00371-9>. URL: <https://www.sciencedirect.com/science/article/pii/B0122274105003719>.
- [65] Rongqing Hui. “Chapter 3 - Light sources for optical communications”. In: *Introduction to Fiber-Optic Communications*. Ed. by Rongqing Hui. Academic Press, 2020, pp. 77–124. ISBN: 978-0-12-805345-4. DOI: <https://doi.org/10.1016/B978-0-12-805345-4.00003-2>. URL: <https://www.sciencedirect.com/science/article/pii/B9780128053454000032>.
- [66] Philippe Gallion. “3 - Basics of incoherent and coherent digital optical communications”. In: *Undersea Fiber Communication Systems (Second Edition)*. Ed. by José Chesnoy. Second Edition. Academic Press, 2016, pp. 55–117. ISBN: 978-0-12-804269-4. DOI: <https://doi.org/10.1016/B978-0-12-804269-4.00003-9>. URL: <https://www.sciencedirect.com/science/article/pii/B9780128042694000039>.

- [67] Rongqing Hui. “Chapter 7 - External electro-optic modulators”. In: *Introduction to Fiber-Optic Communications*. Ed. by Rongqing Hui. Academic Press, 2020, pp. 299–335. ISBN: 978-0-12-805345-4. DOI: <https://doi.org/10.1016/B978-0-12-805345-4.00007-X>. URL: <https://www.sciencedirect.com/science/article/pii/B978012805345400007X>.
- [68] *Optical Modulation*. URL: <https://www.lightreading.com/optical-modulation/d/d-id/576926> (visited on 2022).
- [69] Tetsuya Kawanishi. “Direct and External Modulation”. In: *Electro-optic Modulation for Photonic Networks: Precise and high-speed control of lightwaves*. Cham: Springer International Publishing, 2022, pp. 25–32. ISBN: 978-3-030-86720-1. DOI: 10.1007/978-3-030-86720-1\_3. URL: [https://doi.org/10.1007/978-3-030-86720-1\\_3](https://doi.org/10.1007/978-3-030-86720-1_3).
- [70] Mark Jablonski, Yuichi Takushima, and Kazuro Kikuchi. “The Realization of All-Pass Filters for Third-Order Dispersion Compensation in Ultrafast Optical Fiber Transmission Systems”. In: *J. Lightwave Technol.* 19.8 (Aug. 2001), p. 1194.
- [71] “Optical Lattice-Type Add-Drop Multiplexing Filters and Their Use in WDM Networks”. In: *Optical Fiber Technology* 4.1 (1998), pp. 117–134. ISSN: 1068-5200.
- [72] *Electro-optic Modulators*. URL: [https://www.rp-photonics.com/electro\\_optic\\_modulators.html](https://www.rp-photonics.com/electro_optic_modulators.html) (visited on 2022).
- [73] *Electroabsorption Modulators*. URL: [https://www.rp-photonics.com/electroabsorption\\_modulators.html](https://www.rp-photonics.com/electroabsorption_modulators.html) (visited on 2022).
- [74] L. V. Keldysh. “Behavior of Non-metallic Crystals in Strong Electric Fields”. In: *Soviet Journal of Experimental and Theoretical Physics* 6 (Jan. 1958), p. 763.
- [75] Kishore Bhowmik and Gang-Ding Peng. “Polymer Optical Fibers”. In: Mar. 2019, pp. 1–51. DOI: 10.1007/978-981-10-1477-2\_38-1.
- [76] *Fiber Optic Cabling Solutions*. URL: <https://www.cables-solutions.com/whats-difference-between-single-mode-optical-fiber-and-multi-mode-optical-fiber.html> (visited on 2022).
- [77] Davi.J. DiGiovanni, Donal.P. Jablonowski, and Ma.F. Yan. “Chapter 4 - Advances in Fiber Design and Processing”. In: *Optical Fiber Telecommunications IIIA (Third Edition)*. Ed. by Ivan P. Kaminow and Thomas L. Koch. Third Edition. Boston: Academic Press, 1997, pp. 63–91. ISBN: 978-0-08-051316-4. DOI: <https://doi.org/10.1016/B978-0-08-051316-4.50008-4>.

- [78] Govind P. Agrawal. “Optical Communication: Its History and Recent Progress”. In: *Optics in Our Time*. Ed. by Mohammad D. Al-Amri, Mohamed El-Gomati, and M. Suhail Zubairy. Cham: Springer International Publishing, 2016, pp. 177–199. ISBN: 978-3-319-31903-2. DOI: 10.1007/978-3-319-31903-2\_8. URL: [https://doi.org/10.1007/978-3-319-31903-2\\_8](https://doi.org/10.1007/978-3-319-31903-2_8).
- [79] *Attenuation in Fibre*. URL: <https://sites.google.com/site/puenggphysics/home/unit-i/Lecture-16-Attenuation-in-fibre> (visited on 2022).
- [80] *Material absorption losses in silica glass fibers*. URL: [https://www.brainkart.com/article/Material-absorption-losses-in-silica-glass-fibers\\_13608/](https://www.brainkart.com/article/Material-absorption-losses-in-silica-glass-fibers_13608/) (visited on 2022).
- [81] G.C.K.S. Mathur. *Fiber Optics Fundamentals and Advances in Optical Communications*. Blue Rose Publishers, 2021. URL: <https://books.google.de/books?id=hKUbEAAAQBAJ>.
- [82] *Understanding Losses in Optical Fiber*. URL: <https://www.fiber-optic-tutorial.com/understanding-losses-fiber-optic.html> (visited on 2022).
- [83] *Waveguide attenuation*. URL: [https://shopdelta.eu/waveguide-attenuation\\_12\\_aid811.html](https://shopdelta.eu/waveguide-attenuation_12_aid811.html) (visited on 2022).
- [84] Mazin A. Ali. “Transmission Characteristics of Optical Fibers”. In: May 2020.
- [85] *OPTICAL FIBER ATTENUATION*. URL: <https://www.fiberoptics4sale.com/blogs/archive-posts/95052294-optical-fiber-attenuation> (visited on 2022).
- [86] *Linear scattering losses*. URL: [https://www.brainkart.com/article/Linear-scattering-losses\\_13609/](https://www.brainkart.com/article/Linear-scattering-losses_13609/) (visited on 2022).
- [87] M. Born et al. *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*. Cambridge University Press, 1999. ISBN: 9780521642224. URL: <https://books.google.tn/books?id=nUHGpfNsGyUC>.
- [88] *Scattering losses*. URL: [https://www.brainkart.com/article/Scattering-losses\\_11974/](https://www.brainkart.com/article/Scattering-losses_11974/) (visited on 2022).
- [89] Douglas A. Pinnow et al. “Fundamental optical attenuation limits in the liquid and glassy state with application to fiber optical waveguide materials”. In: *Applied Physics Letters* 22 (1973), pp. 527–529.
- [90] R. Olshansky. “Propagation in glass optical waveguides”. In: *Rev. Mod. Phys.* 51 (2 Apr. 1979), pp. 341–367. DOI: 10.1103/RevModPhys.51.341. URL: <https://link.aps.org/doi/10.1103/RevModPhys.51.341>.

- [91] *Scattering losses in optical fiber*. URL: <https://www.ques10.com/p/5152/scattering-losses-in-optical-fiber-1/> (visited on 2022).
- [92] Aleksandr Y. Bekshaev, Konstantin Y. Bliokh, and Franco Nori. “Mie scattering and optical forces from evanescent fields: A complex-angle approach”. In: *Opt. Express* 21.6 (Mar. 2013), pp. 7082–7095. DOI: 10.1364/OE.21.007082. URL: <http://opg.optica.org/oe/abstract.cfm?URI=oe-21-6-7082>.
- [93] Sunil Singh, Ramgopal Gangwar, and Nar Singh. “Nonlinear scattering effects in optical fibers”. In: *Progress in Electromagnetics Research-pier - PROG ELECTROMAGN RES* 74 (Jan. 2007), pp. 379–405. DOI: 10.2528/PIER07051102.
- [94] G. Agrawal. *Nonlinear Fiber Optics*. Optics and Photonics. Elsevier Science, 2013. ISBN: 9780123970237.
- [95] W. Alec Gambling, Hiroshi Matsumura, and Catherine M. Ragdale. “Curvature and microbending losses in single-mode optical fibres”. In: *Optical and Quantum Electronics* 11 (1979), pp. 43–59.
- [96] Tadashi Murao et al. “Design Principle for Realizing Low Bending Losses in All-Solid Photonic Bandgap Fibers”. In: *J. Lightwave Technol.* 29.16 (Aug. 2011), pp. 2428–2435. URL: <http://opg.optica.org/jlt/abstract.cfm?URI=jlt-29-16-2428>.
- [97] W. B. Gardner and D. Gloge. “MICROBENDING LOSS IN COATED AND UNCOATED OPTICAL FIBERS”. In: *Optical Fiber Transmission*. Optica Publishing Group, 1975, WA3.
- [98] J. Sakai and T. Kimura. “Practical microbending loss formula for single-mode optical fibers”. In: *IEEE Journal of Quantum Electronics* 15.6 (1979), pp. 497–500. DOI: 10.1109/JQE.1979.1070029.
- [99] *How to Reduce Various Types of Losses in Optical Fiber?* URL: <https://community.fs.com/blog/understanding-loss-in-fiber-optic.html> (visited on 2022).
- [100] I C Goyal. “Dispersion in Telecommunication Optical Fibres—A Tutorial Review (Invited Paper)”. In: *IETE Journal of Research* 32.4 (1986), pp. 196–205. DOI: 10.1080/03772063.1986.11436598. eprint: <https://doi.org/10.1080/03772063.1986.11436598>. URL: <https://doi.org/10.1080/03772063.1986.11436598>.

- [101] B.K. Tariyal and A.H. Cherin. “Optical Fiber Communications”. In: *Encyclopedia of Physical Science and Technology (Third Edition)*. Ed. by Robert A. Meyers. Third Edition. New York: Academic Press, 2003, pp. 271–294. ISBN: 978-0-12-227410-7. DOI: <https://doi.org/10.1016/B0-12-227410-5/00521-4>. URL: <https://www.sciencedirect.com/science/article/pii/B0122274105005214>.
- [102] J. Bowers and H.F. Chou. “OPTICAL COMMUNICATION SYSTEMS | Wavelength Division Multiplexing”. In: *Encyclopedia of Modern Optics*. Ed. by Robert D. Guenther. Oxford: Elsevier, 2005, pp. 433–440. ISBN: 978-0-12-369395-2. DOI: <https://doi.org/10.1016/B0-12-369395-0/00673-4>. URL: <https://www.sciencedirect.com/science/article/pii/B0123693950006734>.
- [103] Katsunari Okamoto. “Chapter 3 - Optical fibers”. In: *Fundamentals of Optical Waveguides (Second Edition)*. Ed. by Katsunari Okamoto. Second Edition. Burlington: Academic Press, 2006, pp. 57–158. ISBN: 978-0-12-525096-2. DOI: <https://doi.org/10.1016/B978-012525096-2/50004-0>. URL: <https://www.sciencedirect.com/science/article/pii/B9780125250962500040>.
- [104] David Large and James Farmer. “Chapter 4 - Linear Fiber-Optic Signal Transportation”. In: *Broadband Cable Access Networks*. Ed. by David Large and James Farmer. The Morgan Kaufmann Series in Networking. Boston: Morgan Kaufmann, 2009, pp. 81–126. DOI: <https://doi.org/10.1016/B978-0-12-374401-2.00004-8>. URL: <https://www.sciencedirect.com/science/article/pii/B9780123744012000048>.
- [105] Yasuhiro Koike and Roberto Gaudino. “Chapter 10 - Plastic Optical Fibers and Gb/s Data Links”. In: *Optical Fiber Telecommunications (Sixth Edition)*. Ed. by Ivan P. Kaminow, Tingye Li, and Alan E. Willner. Sixth Edition. Optics and Photonics. Boston: Academic Press, 2013, pp. 353–376. DOI: <https://doi.org/10.1016/B978-0-12-396958-3.00010-X>. URL: <https://www.sciencedirect.com/science/article/pii/B978012396958300010X>.
- [106] Fabrizio Forghieri, Robert W. Tkach, and Andrew R. Chraplyvy. “Chapter 8 – Fiber Nonlinearities and Their Impact on Transmission Systems”. In: 1997.
- [107] Sunil Pratap Singh and Nar Singh. “Nonlinear Effects in Optical Fibers: Origin, Management and Applications”. In: *Progress in Electromagnetics Research-pier* 73 (2007), pp. 249–275.

- [108] Sunish Kumar Orappanpara Soman. “A tutorial on fiber Kerr nonlinearity effect and its compensation in optical communication systems”. In: *Journal of Optics* 23.12 (Nov. 2021), p. 123502. DOI: 10.1088/2040-8986/ac362a. URL: <https://doi.org/10.1088/2040-8986/ac362a>.
- [109] G. Millot and P. Tchofo-Dinda. “FIBER AND GUIDED WAVE OPTICS | Non-linear Effects (Basics)”. In: *Encyclopedia of Modern Optics*. Ed. by Robert D. Guenther. Oxford: Elsevier, 2005, pp. 467–471. ISBN: 978-0-12-369395-2. DOI: <https://doi.org/10.1016/B0-12-369395-0/00652-7>. URL: <https://www.sciencedirect.com/science/article/pii/B0123693950006527>.
- [110] Sunish Kumar. “A tutorial on fiber Kerr nonlinearity effect and its compensation in optical communication systems”. In: *Journal of Optics* 23 (Nov. 2021). DOI: 10.1088/2040-8986/ac362a.
- [111] Govind Agrawal. “Self-phase modulation in optical fiber communications: good or bad?” In: May 2007, pp. 1–2. ISBN: 978-1-55752-834-6. DOI: 10.1109/QELS.2007.4431339.
- [112] K. Thyagarajan and Ajoy Ghatak. “CHAPTER 6 - Some Important Nonlinear Effects in Optical Fibers”. In: *Guided Wave Optical Components and Devices*. Ed. by Bishnu P. Pal. Burlington: Academic Press, 2006, pp. 91–100. ISBN: 978-0-12-088481-0. DOI: <https://doi.org/10.1016/B978-012088481-0/50007-3>. URL: <https://www.sciencedirect.com/science/article/pii/B9780120884810500073>.
- [113] Govind P. Agrawal. “Chapter 7 - Fiber-optic communications”. In: *Applications of Nonlinear Fiber Optics (Third Edition)*. Ed. by Govind P. Agrawal. Third Edition. Academic Press, 2021, pp. 309–368. ISBN: 978-0-12-817040-3. DOI: <https://doi.org/10.1016/B978-0-12-817040-3.00013-4>. URL: <https://www.sciencedirect.com/science/article/pii/B9780128170403000134>.
- [114] *Guide to Optical Amplifier*. URL: <https://www.fiber-optical-networking.com/tag/optical-amplifier> (visited on 2022).
- [115] *WHAT IS OPTICAL AMPLIFIER NOISE FIGURE?* URL: <https://www.fiberoptics4sale.com/blogs/archive-posts/95042758-what-is-optical-amplifier-noise-figure> (visited on 2022).
- [116] *Gain Bandwidth*. URL: [https://www.rp-photonics.com/gain\\_bandwidth.html](https://www.rp-photonics.com/gain_bandwidth.html) (visited on 2022).
- [117] Michael Connelly. “Semiconductor Optical Amplifiers”. In: Jan. 2004.

- [118] J. C. SIMON. “Semiconductor laser amplifiers for single-mode fiber communication”. In: *Optical Fiber Communication*. Optica Publishing Group, 1987, THC2. DOI: 10.1364/OFC.1987.TH2. URL: <http://opg.optica.org/abstract.cfm?URI=OFC-1987-TH2>.
- [119] P.C. Becker, N.A. Olsson, and J.R. Simpson. *Erbium-doped Fiber Amplifiers: Fundamentals and Technology*. Electronics & Electrical v. 1. Academic Press, 1999. ISBN: 9780120845903.
- [120] W.J. Miniscalco. “Erbium-doped glasses for fiber amplifiers at 1500 nm”. In: *Journal of Lightwave Technology* 9.2 (1991), pp. 234–250. DOI: 10.1109/50.65882.
- [121] M.N. Islam. *Raman Amplifiers for Telecommunications 2: Sub-Systems and Systems*. Springer Series in Optical Sciences. Springer New York, 2007. ISBN: 9780387215853. URL: <https://books.google.tn/books?id=NIvaBwAAQBAJ>.
- [122] J. Buus and R. Plastow. “A theoretical and experimental investigation of Fabry-Perot semiconductor laser amplifiers”. In: *IEEE Journal of Quantum Electronics* 21.6 (1985), pp. 614–618. DOI: 10.1109/JQE.1985.1072710.
- [123] T. Saitoh and T. Mukai. “Recent progress in semiconductor laser amplifiers”. In: *Journal of Lightwave Technology* 6.11 (1988), pp. 1656–1664. DOI: 10.1109/50.9981.
- [124] Mike J. O’Mahony. “Semiconductor laser optical amplifiers for use in future fiber systems”. In: *Journal of Lightwave Technology* 6 (1988), pp. 531–544.
- [125] *Semiconductor Laser Amplifier*. URL: [https://ebrary.net/206401/engineering/semiconductor\\_laser\\_amplifier](https://ebrary.net/206401/engineering/semiconductor_laser_amplifier) (visited on 2022).
- [126] *Erbium-Doped Fiber Amplifier (EDFA)*. URL: <https://www.fiberlabs.com/glossary/erbium-doped-fiber-amplifier/> (visited on 2022).
- [127] *Optical Amplifier—EDFA (Erbium-doped Fiber Amplifier) for WDM System*. URL: <https://community.fs.com/blog/erbium-doped-fiber-amplifier-edfa.html> (visited on 2022).
- [128] J. Bromage. “Raman amplification for fiber communications systems”. In: *Journal of Lightwave Technology* 22.1 (2004), pp. 79–93. DOI: 10.1109/JLT.2003.822828.
- [129] C.C.K. Chan. *Optical Performance Monitoring: Advanced Techniques for Next-Generation Photonic Networks*. Elsevier Science, 2010. ISBN: 9780080959177. URL: [https://books.google.de/books?id=PvCN%5C\\_biNky0C](https://books.google.de/books?id=PvCN%5C_biNky0C).

- [130] Shu Namiki et al. “Challenges of Raman Amplification”. In: *Proceedings of the IEEE* 94 (June 2006), pp. 1024–1035. DOI: 10.1109/JPROC.2006.873444.
- [131] Frederick C. Allard. “Fiber Optics Handbook for Engineers and Scientists”. In: 1991.
- [132] *Measuring Reflectance or Return Loss*. URL: <https://www.thefoa.org/tech/ref/testing/test/reflectance.html> (visited on 2022).
- [133] *OPTICAL FIBER FUSION SPLICER TYPES (FUSION SPLICING MACHINES) EXPLAINED IN DETAIL*. URL: <https://www.fiberoptics4sale.com/blogs/archive-posts/95047494-optical-fiber-fusion-splicer-types-fusion-splicing-machines-explained-in-detail> (visited on 2022).
- [134] *Fiber Optic Splicing Guide*. URL: [https://tecratools.com/pages/tecalert/splicing\\_guide.html](https://tecratools.com/pages/tecalert/splicing_guide.html) (visited on 2022).
- [135] *PC vs UPC vs APC Connector: Selecting the Right Fiber Connector Type*. URL: <https://community.fs.com/blog/pc-vs-upc-vs-apc-connector-selecting-the-right-fiber-connector-type.html> (visited on 2022).
- [136] Schelto van Doorn. “Fiber Optic Couplers”. In: *Fiber Optic Couplers, Connectors, and Splice Technology II*. Ed. by David W. Stowe. Vol. 0574. International Society for Optics and Photonics. SPIE, 1986, pp. 2–8. DOI: 10.1117/12.950692. URL: <https://doi.org/10.1117/12.950692>.
- [137] Arun K. Agarwal. “Review of optical fiber couplers”. In: *Fiber and Integrated Optics* 6.1 (1987), pp. 27–53. DOI: 10.1080/01468038708219344. eprint: <https://doi.org/10.1080/01468038708219344>. URL: <https://doi.org/10.1080/01468038708219344>.
- [138] *How Do Different Fiber Optic Couplers Work?* URL: <https://www.fiberoptics4sale.com/blogs/archive-posts/95041798-what-are-fiber-optic-attenuators> (visited on 2022).
- [139] *OPTICAL FIBER COUPLERS*. URL: <https://www.fiberoptics4sale.com/blogs/archive-posts/95047750-optical-fiber-couplers> (visited on 2022).
- [140] Sean M. Garner and Steve Caracci. “Variable Optical Attenuator for Large Scale Integration: Design and Optimization”. In: *Integrated Photonics Research*. Optica Publishing Group, 2001, ITuH5. DOI: 10.1364/IPR.2001.ITuH5. URL: <http://opg.optica.org/abstract.cfm?URI=IPR-2001-ITuH5>.

- [141] *WHAT ARE FIBER OPTIC ATTENUATORS?* URL: <https://community.fs.com/blog/how-do-different-fiber-optic-couplers-work.html> (visited on 2022).
- [142] *The Ultimate Guide to Fiber Optic Attenuators*. URL: <https://community.fs.com/blog/basics-of-fiber-optic-attenuator.html> (visited on 2022).
- [143] Tran Van Muoi. "Optical Receivers". In: *Optoelectronic Technology and Light-wave Communications Systems*. Ed. by Chinlon Lin. Dordrecht: Springer Netherlands, 1989, pp. 441–472. ISBN: 978-94-011-7035-2. DOI: 10.1007/978-94-011-7035-2\_16. URL: [https://doi.org/10.1007/978-94-011-7035-2\\_16](https://doi.org/10.1007/978-94-011-7035-2_16).
- [144] Stewart D. Personick. "Optical Detectors and Receivers". In: *J. Lightwave Technol.* 26.9 (May 2008), pp. 1005–1020. URL: <http://opg.optica.org/jlt/abstract.cfm?URI=jlt-26-9-1005>.
- [145] William Shieh and Ivan Djordjevic. "Chapter 3 - Optical Communication Fundamentals". In: *OFDM for Optical Communications*. Ed. by William Shieh and Ivan Djordjevic. Oxford: Academic Press, 2010, pp. 53–118. ISBN: 978-0-12-374879-9. DOI: <https://doi.org/10.1016/B978-0-12-374879-9.00003-4>.
- [146] G. Osborne. *Optical Fibers Telecommunications*. EDTECH, 2018. ISBN: 9781839472381. URL: <https://books.google.de/books?id=7eLEDwAAQBAJ>.
- [147] M.J. Howes and D.V. Morgan. *Optical Fibre Communications: Devices, Circuits, and Systems*. A Wiley-Interscience publication. J. Wiley, 1980. ISBN: 9780471276111. URL: <https://books.google.de/books?id=09IiAAAAMAAJ>.
- [148] H.S. Nalwa. *Photodetectors and Fiber Optics*. Elsevier Science, 2012. ISBN: 9780323137898. URL: <https://books.google.de/books?id=SRAl6wqOBMYC>.
- [149] A. BANDYOPADHYAY and M. JAMAL DEEN. "CHAPTER 5 - Photodetectors for Optical Fiber Communications". In: *Photodetectors and Fiber Optics*. Ed. by Hari Singh Nalwa. Boston: Academic Press, 2001, pp. 307–368. ISBN: 978-0-12-513908-3. DOI: <https://doi.org/10.1016/B978-0-12-513908-3.50011-7>.
- [150] *What is a Photodiode? Working, V-I Characteristics, Applications*. URL: <https://www.electronicshub.org/photodiode-working-characteristics-applications/> (visited on 2022).
- [151] *Dark Current*. URL: [https://www.rp-photonics.com/dark\\_current.html](https://www.rp-photonics.com/dark_current.html) (visited on 2022).

- [152] *PIN PHOTODETECTOR CHARACTERISTICS FOR OPTICAL FIBER COMMUNICATION*. URL: <https://www.fiberoptics4sale.com/blogs/archive-posts/95046662-pin-photodetector-characteristics-for-optical-fiber-communication> (visited on 2022).
- [153] *Architectures of Fiber-optic Communication Systems*. URL: <https://sites.google.com/site/csapgroupc/home/architectures-of-fiber-optic-communication-systems> (visited on 2022).
- [154] G. Keiser. *FTTX Concepts and Applications*. Wiley Series in Telecommunications and Signal Processing. Wiley, 2006. ISBN: 9780471769095. URL: <https://books.google.de/books?id=F9QVc-YfZk8C>.
- [155] C.F. Lam. *Passive Optical Networks: Principles and Practice*. Elsevier Science, 2011. ISBN: 9780080553450. URL: <https://books.google.de/books?id=DS05CVBuhKEC>.
- [156] Huda Abbas and Mark Gregory. “The next generation of passive optical networks: A review”. In: *Journal of Network and Computer Applications* 67 (Mar. 2016). DOI: 10.1016/j.jnca.2016.02.015.
- [157] *ABC of PON: Understanding OLT, ONU, ONT and ODN*. URL: <https://community.fs.com/blog/abc-of-pon-understanding-olt-onu-ont-and-odn.html> (visited on 2022).
- [158] Kivilcim Yüksel et al. “Optical layer monitoring in Passive Optical Networks (PONs): A review (INVITED)”. In: vol. 1. July 2008, pp. 92–98. ISBN: 978-1-4244-2625-6. DOI: 10.1109/ICTON.2008.4598379.
- [159] Deepak Sharma, Payal Arora, and Suresh Kumar. “Q Factor Based Performance Evaluation of Bidirectional TDM PON Network Using Hybrid Amplifier Configurations”. In: *INTERNATIONAL JOURNAL OF COMPUTER SCIENCES AND ENGINEERING* 6 (May 2018), pp. 51–60. DOI: 10.26438/ijcse/v6i4.5160.
- [160] *Optical Mesh Network (OMN)*. URL: <https://www.techopedia.com/definition/25270/optical-mesh-network-omn> (visited on 2022).
- [161] D.R. Anderson, L.M. Johnson, and F.G. Bell. *Troubleshooting Optical Fiber Networks: Understanding and Using Optical Time-Domain Reflectometers*. Elsevier Science, 2004. ISBN: 9780080492360.
- [162] João Moura. “Detection and Characterization of Events with an OTDR”. In: 2016.

- [163] ReportLinker. *Global Predictive Maintenance Industry*. May 2021. URL: [https://www.reportlinker.com/p05799417/Global-Predictive-Maintenance-Industry.html?utm\\_source=GNW](https://www.reportlinker.com/p05799417/Global-Predictive-Maintenance-Industry.html?utm_source=GNW).
- [164] Julien Simon. “Predictive Maintenance with Machine Learning on AWS”. In: *The MEMS & Sensors Technical Congress - MSTC 2021*. 2021.
- [165] R.K. Mobley. *An Introduction to Predictive Maintenance*. Plant Engineering. Elsevier Science, 2002. ISBN: 9780080478692. URL: <https://books.google.tn/books?id=SjqXzxpAzSQC>.
- [166] J.M. Nicod, B. Chebel-Morello, and C. Varnier. *From Prognostics and Health Systems Management to Predictive Maintenance 2: Knowledge, Reliability and Decision*. Wiley, 2017. ISBN: 9781119436850. URL: <https://books.google.de/books?id=iYMtDwAAQBAJ>.
- [167] *What Is Run-to-Failure Maintenance?* URL: <https://www.getmaintainx.com/learning-center/run-to-failure-maintenance/> (visited on 2022).
- [168] *The Ultimate Preventive Maintenance Guide: Nearly Everything You Need to Know*. URL: <https://www.getmaintainx.com/learning-center/the-ultimate-preventive-maintenance-guide-nearly-everything-you-need-to-know/> (visited on 2022).
- [169] “Copyright”. In: *Reliability, Maintainability and Risk (Tenth Edition)*. Ed. by Dr David J Smith. Tenth Edition. Butterworth-Heinemann, 2022, p. iv. ISBN: 978-0-323-91261-7. DOI: <https://doi.org/10.1016/B978-0-323-91261-7.12001-5>. URL: <https://www.sciencedirect.com/science/article/pii/B9780323912617120015>.
- [170] Milton Ohring. “15 - FAILURE AND RELIABILITY OF ELECTRONIC MATERIALS AND DEVICES”. In: *Engineering Materials Science*. Ed. by Milton Ohring. San Diego: Academic Press, 1995, pp. 747–788. ISBN: 978-0-12-524995-9. DOI: <https://doi.org/10.1016/B978-012524995-9/50039-8>. URL: <https://www.sciencedirect.com/science/article/pii/B9780125249959500398>.
- [171] Philip Kosky et al. “Chapter 11 - Industrial Engineering”. In: *Exploring Engineering (Fifth Edition)*. Ed. by Philip Kosky et al. Fifth Edition. Academic Press, 2021, pp. 229–257. ISBN: 978-0-12-815073-3. DOI: <https://doi.org/10.1016/B978-0-12-815073-3.00011-9>. URL: <https://www.sciencedirect.com/science/article/pii/B9780128150733000119>.

- [172] *Dennis J. Wilkins*. URL: [The%20Bathtub%20Curve%20and%20Product%20Failure%20Behavior%20Part%20One%20-%20The%20Bathtub%20Curve,%20Infant%20Mortality%20and%20Burn-in](#) (visited on 2022).
- [173] C. Scheffer and P. Girdhar. *Practical Machinery Vibration Analysis and Predictive Maintenance*. Practical Machinery Vibration Analysis and Predictive Maintenance. Elsevier Science, 2004. ISBN: 9780080480220. URL: <https://books.google.de/books?id=tAvT01t2mwkC>.
- [174] R Keith Mobley. “50 - Predictive Maintenance”. In: *Plant Engineer’s Handbook*. Ed. by R. Keith Mobley. Woburn: Butterworth-Heinemann, 2001, pp. 867–888. ISBN: 978-0-7506-7328-0. DOI: <https://doi.org/10.1016/B978-075067328-0/50052-5>. URL: <https://www.sciencedirect.com/science/article/pii/B9780750673280500525>.
- [175] *What Is Predictive Maintenance?* URL: <https://www.getmaintainx.com/learning-center/predictive-maintenance-quick-guide/> (visited on 2022).
- [176] *What is Proactive Maintenance?* URL: <https://www.accelix.com/what-is-proactive-maintenance/> (visited on 2022).
- [177] *Proactive Maintenance | What is a Proactive Maintenance?* URL: <https://www.upkeep.com/learning/proactive-maintenance> (visited on 2022).
- [178] Jay Lee et al. “Intelligent Maintenance Systems and Predictive Manufacturing”. In: *Journal of Manufacturing Science and Engineering* 142 (July 2020), pp. 1–40. DOI: 10.1115/1.4047856.
- [179] *Predictive maintenance with IoT: The road to real returns*. URL: <https://www.avnet.com/wps/portal/abacus/solutions/markets/industrial/predictive-maintenance-iot/> (visited on 2022).
- [180] Werner Neumann Giuseppina Tomarchio Alessandro Faulisi and Vladimir Janousek. *Jumping into Industry 4.0 with Predictive Maintenance Solutions*. URL: <https://www.eetimes.eu/%EF%BB%BFjumping-into-industry-4-0-with-predictive-maintenance-solutions/> (visited on 2022).
- [181] E. Alpaydin. *Introduction to Machine Learning*. Adaptive computation and machine learning. MIT Press, 2004. ISBN: 9780262012119. URL: [https://books.google.de/books?id=1k0%5C\\_-WroiqEC](https://books.google.de/books?id=1k0%5C_-WroiqEC).
- [182] A. L. Samuel. “Some Studies in Machine Learning Using the Game of Checkers”. In: *IBM Journal of Research and Development* 3.3 (1959), pp. 210–229. DOI: 10.1147/rd.33.0210.

- [183] M. Mohri, A. Rostamizadeh, and A. Talwalkar. *Foundations of Machine Learning*. Adaptive Computation and Machine Learning series. MIT Press, 2012. ISBN: 9780262018258. URL: <https://books.google.de/books?id=maz6AQAAQBAJ>.
- [184] Sara Brown. *Machine learning, explained*. URL: <https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained> (visited on 2022).
- [185] S. Shalev-Shwartz and S. Ben-David. *Understanding Machine Learning: From Theory to Algorithms*. Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press, 2014. ISBN: 9781107057135. URL: <https://books.google.de/books?id=ttJkAwAAQBAJ>.
- [186] Jason Brownlee. *4 Types of Classification Tasks in Machine Learning*. URL: <https://machinelearningmastery.com/types-of-classification-in-machine-learning/> (visited on 2022).
- [187] Jason Brownlee. *Regression Metrics for Machine Learning*. URL: <https://machinelearningmastery.com/regression-metrics-for-machine-learning/> (visited on 2022).
- [188] ZACH. *5 Examples of Cluster Analysis in Real Life*. URL: <https://www.statology.org/cluster-analysis-real-life-examples/> (visited on 2022).
- [189] *Workflow and Data Preparation*. URL: <http://www.theobjects.com/dragonfly/dfhelp/2020-1/Content/Artificial%20Intelligence/Deep%20Learning%20Tool/Workflow%20and%20Data%20Preparation.htm> (visited on 2022).
- [190] Jafar Alzubi, Anand Nayyar, and Akshi Kumar. “Machine Learning from Theory to Algorithms: An Overview”. In: *Journal of Physics: Conference Series* 1142 (Nov. 2018), p. 012012. DOI: 10.1088/1742-6596/1142/1/012012. URL: <https://doi.org/10.1088/1742-6596/1142/1/012012>.
- [191] T. Jo. *Machine Learning Foundations: Supervised, Unsupervised, and Advanced Learning*. Springer International Publishing, 2021. ISBN: 9783030659004. URL: <https://books.google.de/books?id=0egdEAAAQBAJ>.
- [192] K.K. Hiran et al. *Machine Learning: Master Supervised and Unsupervised Learning Algorithms with Real Examples (English Edition)*. BPB Publications, 2021. ISBN: 9789391392352. URL: <https://books.google.de/books?id=4VVDEAAAQBAJ>.
- [193] R.S. Sutton and A.G. Barto. *Reinforcement Learning, second edition: An Introduction*. Adaptive Computation and Machine Learning series. MIT Press, 2018. ISBN: 9780262039246. URL: <https://books.google.de/books?id=sWVODwAAQBAJ>.

- [194] Piyush Verma and Stelios Diamantidis. *What is Reinforcement Learning?* URL: <https://www.synopsys.com/ai/what-is-reinforcement-learning.html> (visited on 2022).
- [195] Pawel Ladosz et al. *Exploration in Deep Reinforcement Learning: A Survey*. May 2022.
- [196] Gary Chan. *Applications of Reinforcement Learning in Real World*. URL: <https://towardsdatascience.com/applications-of-reinforcement-learning-in-real-world-1a94955bcd12> (visited on 2022).
- [197] Muddasar Naeem, Syed Rizvi, and Antonio Coronato. “A Gentle Introduction to Reinforcement Learning and its Application in Different Fields”. In: *IEEE Access* 8 (Jan. 2020), pp. 209320–209344. DOI: 10.1109/ACCESS.2020.3038605.
- [198] *Supervised vs Unsupervised vs Reinforcement Learning – Knowing the differences is a fundamental part of properly understanding machine learning*. URL: <https://starship-knowledge.com/supervised-vs-unsupervised-vs-reinforcement> (visited on 2022).
- [199] *5 Things You Need to Know about Reinforcement Learning*. URL: <https://www.kdnuggets.com/2018/03/5-things-reinforcement-learning.html> (visited on 2022).
- [200] M.A. Hearst et al. “Support vector machines”. In: *IEEE Intelligent Systems and their Applications* 13.4 (1998), pp. 18–28. DOI: 10.1109/5254.708428.
- [201] *Support Vector Machines(SVM) — An Overview*. URL: <https://towardsdatascience.com/https-medium-com-pupalerushikesh-svm-f4b42800e989> (visited on 2022).
- [202] Simon Tong and Daphne Koller. “Support Vector Machine Active Learning With Applications To Text Classification”. In: *The Journal of Machine Learning Research* 2 (Dec. 2001), pp. 45–66. DOI: 10.1162/153244302760185243.
- [203] *Support Vector Machines for dummies; A Simple Explanation*. URL: <https://aylien.com/blog/support-vector-machines-for-dummies-a-simple-explanation> (visited on 2022).
- [204] *Support Vector Machine Algorithm*. URL: <https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm> (visited on 2022).
- [205] Shan Suthaharan. “Decision Tree Learning”. In: *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning*. Boston, MA: Springer US, 2016, pp. 237–269.

- [206] *Decision Tree Algorithm, Explained*. URL: <https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html> (visited on 2022).
- [207] *HOW DECISION TREE ALGORITHM WORKS*. URL: <https://dataaspirant.com/how-decision-tree-algorithm-works/> (visited on 2022).
- [208] *Decision Trees for Dummies*. URL: <https://medium.com/analytics-vidhya/decision-trees-for-dummies-a8e3c00c5e2e> (visited on 2022).
- [209] Leo Breiman Statistics and Leo Breiman. “Random Forests”. In: *Machine Learning*. 2001, pp. 5–32.
- [210] *Bagging and Random Forest Ensemble Algorithms for Machine Learning*. URL: <https://machinelearningmastery.com/bagging-and-random-forest-ensemble-algorithms-for-machine-learning/> (visited on 2022).
- [211] *Random Forest Classifier and its Hyperparameters*. URL: <https://medium.com/analytics-vidhya/random-forest-classifier-and-its-hyperparameters-8467bec755f6> (visited on 2022).
- [212] Oliver Kramer. “K-Nearest Neighbors”. In: *Dimensionality Reduction with Unsupervised Nearest Neighbors*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 13–23.
- [213] *K-Nearest Neighbors (kNN) — Explained*. URL: <https://towardsdatascience.com/k-nearest-neighbors-knn-explained-cbc31849a7e3> (visited on 2022).
- [214] *K-Nearest Neighbor*. URL: <https://medium.com/swlh/k-nearest-neighbor-ca2593d7a3c4> (visited on 2022).
- [215] J. Strickland. *Logistic Regression Inside and Out*. Lulu.com, 2017. ISBN: 9781365819155. URL: <https://books.google.tn/books?id=mvueDgAAQBAJ>.
- [216] Jason Brownlee. *Logistic Regression for Machine Learning*. URL: <https://machinelearningmastery.com/logistic-regression-for-machine-learning/> (visited on 2022).
- [217] M.H. Hassoun and A.P.C.E.M.H. Hassoun. *Fundamentals of Artificial Neural Networks*. A Bradford book. MIT Press, 1995. ISBN: 9780262082396. URL: <https://books.google.tn/books?id=0tk32Y3QkxQC>.
- [218] A.I. Galushkin. *Neural Networks Theory*. Springer Berlin Heidelberg, 2007. ISBN: 9783540481256. URL: <https://books.google.de/books?id=ULds8NuzLtkC>.
- [219] K. Gurney. *An Introduction to Neural Networks*. An Introduction to Neural Networks. Taylor & Francis, 1997. ISBN: 9781857285031. URL: <https://books.google.de/books?id=H0svl1RMMP8C>.

- [220] J. Brownlee and Machine Learning Mastery. *Deep Learning with Python: Develop Deep Learning Models on Theano and TensorFlow Using Keras*. Machine Learning Mastery, 2017. URL: <https://books.google.de/books?id=eJw2nQAACAAJ>.
- [221] *Comprehensive synthesis of the main activation functions pros and cons*. URL: <https://medium.com/analytics-vidhya/comprehensive-synthesis-of-the-main-activation-functions-pros-and-cons-dab105fe4b3b> (visited on 2022).
- [222] *Comparison of Activation Functions for Deep Neural Networks*. URL: <https://towardsdatascience.com/comparison-of-activation-functions-for-deep-neural-networks-706ac4284c8a> (visited on 2022).
- [223] F. Millstein. *Convolutional Neural Networks in Python: Beginner’s Guide to Convolutional Neural Networks in Python*. CreateSpace Independent Publishing Platform, 2018. ISBN: 9781986264020. URL: <https://books.google.tn/books?id=1Wb4swEACAAJ>.
- [224] M. Sewak, M.R. Karim, and P. Pujari. *Practical Convolutional Neural Networks: Implement advanced deep learning models using Python*. Packt Publishing, 2018. ISBN: 9781788394147. URL: <https://books.google.de/books?id=b010DwAAQBAJ>.
- [225] U. Michelucci. *Advanced Applied Deep Learning: Convolutional Neural Networks and Object Detection*. Apress, 2019. ISBN: 9781484249765. URL: <https://books.google.de/books?id=hJyyDwAAQBAJ>.
- [226] Aseem Patil and Milind E Rane. “Convolutional Neural Networks: An Overview and Its Applications in Pattern Recognition”. In: 2020.
- [227] Rikiya Yamashita et al. “Convolutional neural networks: an overview and application in radiology”. In: *Insights into Imaging* 9 (2018), pp. 611–629.
- [228] Francis Jesmar Montalbo and Alvin Alon. “Empirical Analysis of a Fine-Tuned Deep Convolutional Model in Classifying and Detecting Malaria Parasites from Blood Smears”. In: *KSII Transactions on Internet and Information Systems* 15 (Jan. 2021), pp. 147–165. DOI: 10.3837/tiis.2021.01.009.
- [229] F.M. Salem. *Recurrent Neural Networks: From Simple to Gated Architectures*. Springer International Publishing, 2022. ISBN: 9783030899295. URL: <https://books.google.tn/books?id=bJpXEAAAQBAJ>.

- [230] A.K. Tyagi and A. Abraham. *Recurrent Neural Networks: Concepts and Applications*. CRC Press, 2022. ISBN: 9781000626179. URL: <https://books.google.de/books?id=5phxEAAAQBAJ>.
- [231] *What are the types of RNN?* URL: <https://www.educative.io/answers/what-are-the-types-of-rnn> (visited on 2022).
- [232] *A Guide to RNN: Understanding Recurrent Neural Networks and LSTM Networks*. URL: <https://builtin.com/data-science/recurrent-neural-networks-and-lstm> (visited on 2022).
- [233] *The Unreasonable Effectiveness of Recurrent Neural Networks*. URL: <http://karpathy.github.io/2015/05/21/rnn-effectiveness/> (visited on 2022).
- [234] Sepp Hochreiter and Jürgen Schmidhuber. “Long Short-Term Memory”. In: *Neural Computation* 9.8 (Nov. 1997), pp. 1735–1780.
- [235] Khoulood Abdelli et al. “Optical Fiber Fault Detection and Localization in a Noisy OTDR Trace Based on Denoising Convolutional Autoencoder and Bidirectional Long Short-Term Memory”. In: *Journal of Lightwave Technology* 40.8 (2022), pp. 2254–2264. DOI: 10.1109/JLT.2021.3138268.
- [236] J. Brownlee. *Long Short-Term Memory Networks With Python: Develop Sequence Prediction Models with Deep Learning*. Machine Learning Mastery, 2017. URL: <https://books.google.de/books?id=m7SoDwAAQBAJ>.
- [237] Trias Thireou and Martin Reczko. “Bidirectional Long Short-Term Memory Networks for Predicting the Subcellular Localization of Eukaryotic Proteins”. In: *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 4 (2007).
- [238] Kyunghyun Cho et al. “Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation”. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1724–1734. DOI: 10.3115/v1/D14-1179. URL: <https://aclanthology.org/D14-1179>.
- [239] Khoulood Abdelli et al. “Machine Learning-based Anomaly Detection in Optical Fiber Monitoring”. In: *Journal of Optical Communications and Networking* (Mar. 2022). DOI: 10.1364/JOCN.451289.
- [240] Xinyu Liu et al. “Bi-directional gated recurrent unit neural network based non-linear equalizer for coherent optical communication system”. In: *Opt. Express* 29.4 (Feb. 2021), pp. 5923–5933. DOI: 10.1364/OE.416672. URL: <http://opg.optica.org/oe/abstract.cfm?URI=oe-29-4-5923>.

- [241] Mark A. Kramer. “Nonlinear principal component analysis using autoassociative neural networks”. In: *Aiche Journal* 37 (1991), pp. 233–243.
- [242] Khoulood Abdelli et al. “Degradation Prediction of Semiconductor Lasers using Conditional Variational Autoencoder”. In: *Journal of Lightwave Technology* (2022), pp. 1–9. DOI: 10.1109/JLT.2022.3188831.
- [243] Pascal Vincent et al. “Extracting and composing robust features with denoising autoencoders”. In: *ICML '08*. 2008.
- [244] Yoshua Bengio et al. “Deep Generative Stochastic Networks Trainable by Backprop”. In: *ICML*. 2014.
- [245] Lovedeep Gondara. “Medical Image Denoising Using Convolutional Denoising Autoencoders”. In: *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*. 2016, pp. 241–246. DOI: 10.1109/ICDMW.2016.0041.
- [246] Diederik P. Kingma and Max Welling. “Auto-Encoding Variational Bayes”. In: *CoRR* abs/1312.6114 (2014).
- [247] Carl Doersch. “Tutorial on Variational Autoencoders”. In: (June 2016).
- [248] Diederik P. Kingma and Max Welling. “An Introduction to Variational Autoencoders”. In: *ArXiv* abs/1906.02691 (2019).
- [249] Boris Ivanovic et al. “Multimodal Deep Generative Models for Trajectory Prediction: A Conditional Variational Autoencoder Approach”. In: *IEEE Robotics and Automation Letters* 6.2 (2021), pp. 295–302. DOI: 10.1109/LRA.2020.3043163.
- [250] Jaechang Lim et al. “Molecular generative model based on conditional variational autoencoder for de novo molecular design”. In: *Journal of Cheminformatics* 10 (2018).
- [251] Artidoro Pagnoni, Kevin Liu, and Shangyan Li. “Conditional Variational Autoencoder for Neural Machine Translation”. In: *ArXiv* abs/1812.04405 (2018).
- [252] J. Kalin. *Generative Adversarial Networks Cookbook: Over 100 recipes to build generative models using Python, TensorFlow, and Keras*. Packt Publishing, 2018. ISBN: 9781789139587. URL: <https://books.google.tn/books?id=Lx2CDwAAQBAJ>.
- [253] N.K. Manaswi. *Generative Adversarial Networks with Industrial Use Cases: Learning How to Build GAN Applications for Retail, Healthcare, Telecom, Media, Education, and HRTech*. BPB PUNB, 2020. ISBN: 9789389423853. URL: <https://books.google.de/books?id=qyDUDwAAQBAJ>.

- [254] Antonia Creswell et al. “Generative Adversarial Networks: An Overview”. In: *IEEE Signal Processing Magazine* 35 (2018), pp. 53–65.
- [255] Jie Gui et al. “A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications”. In: *IEEE Transactions on Knowledge and Data Engineering* (2021), pp. 1–1. DOI: 10.1109/TKDE.2021.3130191.
- [256] Sebastian Ruder. “An Overview of Multi-Task Learning in Deep Neural Networks”. In: *ArXiv abs/1706.05098* (2017).
- [257] Arda Aras. *Explaining what learned models predict: In which cases can we trust machine learning models and when is caution required ?* Mar. 2020.
- [258] *The Rewards and Challenges of Predictive Maintenance*. URL: <https://www.infoq.com/articles/predictive-maintenance-industrial-iot/> (visited on 2022).
- [259] *Avoid the 8 most common predictive maintenance challenges*. URL: <https://www.sensorfy.ai/blog/8-challenges-that-come-with-implementing-predictive-maintenance/> (visited on 2022).
- [260] Toshio Katsuyama. “Development of semiconductor laser for optical communication”. In: *SEI Technical Review* (Jan. 2009), pp. 13–20.
- [261] Farhan Ujager, S.M.H. Zaidi, and U. Younis. “A review of semiconductor lasers for optical communications”. In: Dec. 2010, pp. 107–111. ISBN: 978-1-4244-9922-9. DOI: 10.1109/HONET.2010.5715754.
- [262] Pisut Koomsap et al. “Condition monitoring and lifetime estimation of a CO2 laser”. In: *Journal of Laser Applications - J LASER APPL* 15 (Nov. 2003). DOI: 10.2351/1.1619994.
- [263] Bernhard Schölkopf et al. “Estimating Support of a High-Dimensional Distribution”. In: *Neural Computation* 13 (July 2001), pp. 1443–1471. DOI: 10.1162/089976601750264965.
- [264] Markus Breunig et al. “LOF: Identifying Density-Based Local Outliers.” In: vol. 29. June 2000, pp. 93–104. DOI: 10.1145/342009.335388.
- [265] Khoulood Abdelli, Helmut Grieger, and Stephan Pachnicke. “A Machine Learning-based Framework for Predictive Maintenance of Semiconductor Laser for Optical Communication”. In: *Journal of Lightwave Technology* (2022), pp. 1–1. DOI: 10.1109/JLT.2022.3163579.

- [266] Khoulood Abdelli, Danish Rafique, and Stephan Pachnicke. “Machine Learning Based Laser Failure Mode Detection”. In: July 2019. DOI: 10.1109/ICTON.2019.8840267.
- [267] Karl Häusler et al. “Degradation model analysis of laser diodes”. In: *Journal of Materials Science: Materials in Electronics* 19 (2008), pp. 160–164.
- [268] *Laser Lab Source :LASER DIODE SOURCE*. URL: <https://www.laserdiodesource.com/laser-diodes-filtered-by-wavelength/1550nm-laser-diodes> (visited on 01/12/2019).
- [269] Khoulood Abdelli et al. “Lifetime Prediction of 1550 nm DFB Laser using Machine Learning Techniques”. In: *2020 Optical Fiber Communications Conference and Exhibition (OFC)*. 2020, pp. 1–3.
- [270] *1500 nm InGaAsP MQW-DFB laser datasheet*. URL: <http://www.cel.com/pdf/datasheets/nd17705.pdf> (visited on 08/25/2019).
- [271] Khoulood Abdelli, Helmut Griesser, and Stephan Pachnicke. “A Hybrid CNN-LSTM Approach for Laser Remaining Useful Life Prediction”. In: July 2021. DOI: 10.1364/OECC.2021.S3D.3.
- [272] Khoulood Abdelli, Helmut Grießer, and Stephan Pachnicke. “Machine Learning Based Data Driven Diagnostic and Prognostic Approach for Laser Reliability Enhancement”. In: *2020 22nd International Conference on Transparent Optical Networks (ICTON)*. 2020, pp. 1–4. DOI: 10.1109/ICTON51198.2020.9203551.
- [273] Hiroyuki Iida et al. “Fresnel Reflection Analysis for Optical Fibre Identification Employing with Three-Wavelength OTDR”. In: *2020 European Conference on Optical Communications (ECOC)*. 2020, pp. 1–4. DOI: 10.1109/ECOC48923.2020.9333328.
- [274] Felix P. Kapron et al. “Fiber-optic reflection measurements using OCWR and OTDR techniques”. In: *Journal of Lightwave Technology* 7 (1989), pp. 1234–1241.
- [275] Edwin B. Wilson. “Probable Inference, the Law of Succession, and Statistical Inference”. In: *Journal of the American Statistical Association* 22.158 (1927), pp. 209–212. DOI: 10.1080/01621459.1927.10502953.
- [276] Fenglei Liu and Christopher Zarowski. “Detection and location of connection splice events in fiber optics given noisy OTDR data. Part II. R1MSDE method”. In: *Instrumentation and Measurement, IEEE Transactions on* 53 (May 2004), pp. 546–556. DOI: 10.1109/TIM.2003.820442.

- [277] Khoulood Abdelli et al. “A BiLSTM-CNN based Multitask Learning Approach for Fiber Fault Diagnosis”. In: *2021 Optical Fiber Communications Conference and Exhibition (OFC)*. 2021, pp. 1–3.
- [278] “A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load”. In: *Mechanical Systems and Signal Processing* 100 (2018), pp. 439–453. ISSN: 0888-3270. DOI: <https://doi.org/10.1016/j.ymssp.2017.06.022>.
- [279] Itamar Reis, Dalya Baron, and Sahar Shahaf. “Probabilistic Random Forest: A machine learning algorithm for noisy datasets”. In: *ArXiv abs/1811.05994* (2018).
- [280] Liu Yang, Xuhui Meng, and George Karniadakis. “B-PINNs: Bayesian physics-informed neural networks for forward and inverse PDE problems with noisy data”. In: *Journal of Computational Physics* 425 (Jan. 2021), p. 109913. DOI: [10.1016/j.jcp.2020.109913](https://doi.org/10.1016/j.jcp.2020.109913).
- [281] Maged Esmail and H. Fathallah. “Physical Layer Monitoring Techniques for TDM-Passive Optical Networks: A Survey”. In: *Communications Surveys Tutorials, IEEE* 15 (Jan. 2013), pp. 943–958. DOI: [10.1109/SURV.2012.060912.00057](https://doi.org/10.1109/SURV.2012.060912.00057).
- [282] Mohammad Rad et al. “Passive Optical Network Monitoring: Challenges and Requirements”. In: *Communications Magazine, IEEE* 49 (Mar. 2011), s45–S52. DOI: [10.1109/MCOM.2011.5706313](https://doi.org/10.1109/MCOM.2011.5706313).
- [283] Khoulood Abdelli et al. “Gated Recurrent Unit based Autoencoder for Optical Link Fault Diagnosis in Passive Optical Networks”. In: *2021 European Conference on Optical Communication (ECOC)*. 2021, pp. 1–4. DOI: [10.1109/ECOC52684.2021.9605969](https://doi.org/10.1109/ECOC52684.2021.9605969).

## Veröffentlichungen des Autors mit Angaben über die Höhe des Eigenanteils bei Publikationen/Manuskripten mit mehreren Autoren

Nachfolgend sind die Veröffentlichungen der Doktorandin aufgelistet, aus denen Teile in der Dissertation verwendet wurden. Da diese Veröffentlichungen Gemeinschaftsarbeiten mehrerer Autoren sind, ist die Höhe des Eigenanteils der Doktorandin in den Tabellen angegeben.

### Beiträge in Fachzeitschriften

Reflective Fiber Faults Detection and Characterization Using Long-Short-Term Memory, IEEE/OSA Journal of Optical Communications and Networking (JOCN), vol. 13, no. 10, pp. E32-E41, October 2021.

Autoren: K. Abdelli, H. Griesser, P. Ehrle, C. Tropschug, S. Pachnicke

Konzeptionierung	Planung	Durchführung	Manuskripterstellung
hoch	hoch	hoch	hoch

Optical Fiber Fault Detection and Localization in a Noisy OTDR Trace Based on Denoising Convolutional Autoencoder and Bidirectional Long Short-Term Memory, Journal of Lightwave Technology (JLT), vol. 40, no. 8, pp. 2254-2264, April 2022.

Autoren: K. Abdelli, H. Griesser, C. Tropschug, S. Pachnicke

Konzeptionierung	Planung	Durchführung	Manuskripterstellung
hoch	hoch	hoch	hoch

Machine Learning-based Anomaly Detection in Optical Fiber Monitoring, IEEE/OSA Journal of Optical Communications and Networking (JOCN), vol. 14, no. 5, pp. 366-375, May 2022.

Autoren: K. Abdelli, J. Y. Cho, F. Azendorf, H. Griesser, C. Tropschug, S. Pachnicke

Konzeptionierung	Planung	Durchführung	Manuskripterstellung
hoch	hoch	hoch	hoch

A Machine Learning-based Framework for Predictive Maintenance of Semiconductor Laser for Optical Communication, Journal of Lightwave Technology (JLT), vol. 40, no. 14, pp. 4698-4708, July, 2022.

Autoren: K. Abdelli, H. Griesser, S. Pachnicke

Konzeptionierung	Planung	Durchführung	Manuskripterstellung
hoch	hoch	hoch	hoch

Degradation Prediction of Semiconductor Lasers using Conditional Variational Autoencoder, Journal of Lightwave Technology (JLT), vol. 40, no. 18, pp. 6213-6221, September, 2022.

Autoren: K. Abdelli, H. Griesser, C. Neumeyr, R. Hohenleitner, S. Pachnicke

Konzeptionierung	Planung	Durchführung	Manuskripterstellung
hoch	hoch	hoch	hoch

## Konferenzbeiträge

Machine Learning based Laser Failure Mode Detection, International Conference on Transparent Optical Networks (ICTON), Angers, France, July 2019.

Autoren: K. Abdelli, D. Rafique, S. Pachnicke

Konzeptionierung	Planung	Durchführung	Manuskripterstellung
hoch	hoch	hoch	hoch

Lifetime Prediction of 1550 nm DFB Laser using Machine learning Techniques, Optical Fiber Communication Conference (OFC 2020), San Diego, USA, March 2020.

Autoren: K. Abdelli, D. Rafique, H. Griesser, S. Pachnicke

Konzeptionierung	Planung	Durchführung	Manuskripterstellung
hoch	hoch	hoch	hoch

Cognitive Laser Reliability Enhancement Using Machine Learning Techniques: A Data-Driven Perspective, International Conference on Transparent Optical Networks (ICTON) Bari, Italy, July 2020.

Autoren: K. Abdelli, H. Griesser, S. Pachnicke

Konzeptionierung	Planung	Durchführung	Manuskripterstellung
hoch	hoch	hoch	hoch

Convolutional Neural Networks for Reflective Event Detection and Characterization in Fiber Optical Links Given Noisy OTDR Signals, ITG Conference "Photonic Networks", Leipzig, Germany, May 2021.

Autoren: K. Abdelli, H. Griesser, S. Pachnicke

Konzeptionierung	Planung	Durchführung	Manuskripterstellung
hoch	hoch	hoch	hoch

A BiLSTM-CNN based Multitask Learning Approach for Fiber Fault Diagnosis, Optical Fiber Communication Conference (OFC 2021), San Francisco, USA, June 2021.

Autoren: K. Abdelli, H. Griesser, C. Tropschug, S. Pachnicke

Konzeptionierung	Planung	Durchführung	Manuskripterstellung
hoch	hoch	hoch	hoch

A Hybrid CNN-LSTM Approach for Laser Remaining Useful Life Prediction, Optoelectronics and Communications Conference (OECC 2021), Hong Kong, China, July 2021.

Autoren: K. Abdelli, H. Griesser, S. Pachnicke

Konzeptionierung	Planung	Durchführung	Manuskripterstellung
hoch	hoch	hoch	hoch

Gated Recurrent Unit based Autoencoder for Optical Link Fault Diagnosis in Passive Optical Networks, European Conference on Optical Communication (ECOC 2021), Bordeaux, France, September 2021.

Autoren: K. Abdelli, F. Azendorf, H. Griesser, C. Tropschug, S. Pachnicke

Konzeptionierung	Planung	Durchführung	Manuskripterstellung
hoch	hoch	hoch	hoch

Secure Collaborative Learning for Predictive Maintenance in Optical Networks, Nordic Conference on Secure IT Systems (Nordsec 2021), Tampere, Finland, November 2021.

Autoren: K. Abdelli, J. Y. Cho, S. Pachnicke

Konzeptionierung	Planung	Durchführung	Manuskripterstellung
hoch	hoch	hoch	hoch

Secure and Robust Federated Learning for Predictive Maintenance in Optical Networks, International Workshop on Trustable, Verifiable and Auditable Federated Learning in Conjunction with AAAI 2022 (FL-AAAI-22), Vancouver, Canada, February 2022.

Autoren: K. Abdelli, J. Y. Cho, S. Pachnicke

Konzeptionierung	Planung	Durchführung	Manuskripterstellung
hoch	hoch	hoch	hoch

Federated Learning Approach for Lifetime Prediction of Semiconductor Lasers, Optical Fiber Communication Conference (OFC 2022), San Diego, USA, March 2022.

Autoren: K. Abdelli, Helmut Griesser. Pachnicke

Konzeptionierung	Planung	Durchführung	Manuskripterstellung
hoch	hoch	hoch	hoch

DeepALM: Holistic Optical Network Monitoring based on Machine Learning, in Optical Fiber Communication Conference (OFC) 2022.

Autoren: J. Y. Cho, J. Pedreno-Manresa, S. Patri, K. Abdelli, C. Tropschug, J. Zou, and P. Rydlichowski

Konzeptionierung	Planung	Durchführung	Manuskripterstellung
mittel	mittel	hoch	niedrig

## Erklärung

Hiermit erkläre ich, dass diese Dissertation mit dem Titel „*Machine Learning-based Predictive Maintenance for Optical Networks*“, abgesehen von der Beratung durch meinen Betreuer Prof. Dr.-Ing. Stephan Pachnicke, nach Inhalt und Form eine eigenständige und nur mit den angegebenen Hilfsmitteln von mir verfasste Arbeit ist.

Die Arbeit hat weder ganz noch teilweise an einer anderen Stelle im Rahmen eines Prüfungsverfahrens vorgelegen und ist bisher weder veröffentlicht noch zur Veröffentlichung eingereicht worden.

Ferner erkläre ich, dass die Arbeit unter Einhaltung der Regeln guter wissenschaftlicher Praxis der Deutschen Forschungsgemeinschaft entstanden ist.

Des Weiteren wurde mir kein akademischer Grad entzogen.

München, der 20. November 2022

---

(KhouLOUD Abdelli)