QATAR UNIVERSITY

COLLEGE OF ENGINEERING

MACHINE LEARNING TECHNIQUES FOR MOTOR GEAR FAULT DETECTION AND DIAGNOSIS

BY

ULA MUTASIM HIJAWI

A Thesis Submitted to

the College of Engineering

in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Electrical Engineering

January 2022

COMMITTEE PAGE

The members of the Committee approve the Thesis of
Ula Mutasim Hijawi defended on 21/11/2021.

Dr. Mustafa Serkan Keranas
Thesis Supervisor

Dr. Ridha Hamila
Thesis Co-Supervisor

Dr. Faouzi Alaya Cheikh
Committee Member

Dr. Devrim Unal
Committee Member

Approved:

Khalid Kamal Naji, Dean, College of Engineering

ABSTRACT

HIJAWI, ULA, Masters : January : 2022, Masters of Science in Electrical Engineering

Title: Machine Learning Techniques for Motor Gear Fault Detection and Diagnosis

Supervisor of Thesis: Prof. Mustafa, Serkan, Kiranyaz.

Many industrial facilities, amongst others, are very sensitive to any sudden hazards that can be expensive and resource-costly; therefore, they should be monitored over the day. In this work, new Machine Learning (ML)-based solutions are explored, with an emphasis on compact neural networks methods, for monitoring sensitive equipment using different data measurement types collected from condition monitoring sensors. Consequently, the proposed thesis work tackles a challenging topic that solves problems of high interest to the academic and industrial engineering community. ML algorithms have been thoroughly employed to build numerous structural damage detection systems and Structural Health Maintenance (SHM). ML-based techniques for facility condition monitoring are investigated and presented in a comparative analysis.

Using convolutional layers, network parameters are greatly decreased in Convolutional Neural Networks (CNNs) by local connectivity and exchanging weights, which comprise a group of kernels with a limited receptive area, unlike a typical neural network with a complete connection through each node.

Training of conventional, deep 2-Dimensional (2D) CNNs requires an intensive process in order to obtain an appropriate generalization capacity. This typically involves large-scale datasets, which in turn raises the computational issues considerably. One-Dimensional (1D) CNNs, on the other hand, have recently been proposed in many 1D signal processing applications, including gear condition

monitoring. They have recently been designed to overcome these disadvantages by working directly and more effectively on 1D signals.

Nevertheless, CNNs involve homogenous configurations that entirely rely on the linear neuron model. It is evident that, in many complex problems, the required learning performance can only be achieved by deep CNNs with immense complexity. Self-Organized Operational Neural Networks (Self-ONNs), however, are recently introduced to overcome the drawback of the convolutional neurons targeting a greatly complex and nonlinear solution space. With minimum level of complexity of the network and least data to train, Self-ONNs can model multi-modal and sophisticated functions and boost diversity by involving heterogeneity with a flexible set of operators that can be optimized.

In this study, the design and implementation of compact 1D CNNs, as well as Self-ONNs targeting gear cracking fault detection and diagnosis are explored for three different sensor types: acoustic, current, and vibration. It is shown that the presented contribution can detect cracking fault occurrences as well as diagnose the fault severity condition. The proposed approach can effectively deal with limited training data acquired in a physical motor setup designed for this study. The performance of the system is continuously evaluated twice; over 1D CNNs and Self-ONNs, in terms of predetermined metrics for validation on real current, vibration, and acoustic signals collected at a lab in Qatar University. The results indicate to the outstanding performance of Self-ONNs in contrast to 1D CNNs in challenging problems among the three signal types for gear fault detection and level percentage diagnosis. Moreover, further analysis indicates the fastest detection of the first-time occurrence of a faulty frame that can be achieved.

The proposed thesis topic presents a social, health, economic, and environmental impact, in addition to scientific and academic dissemination. It is also aligned with different priority themes of Qatar's national priority research themes.

DEDICATION

To my beloved parents and sisters,

The strongest and kindest woman I know, my mother,

My dear husband Abdullah, for his everlasting support,

My dearest friend, Qamar,

And to my teachers and professors.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude and special thanks to Professor Serkan Kiranyaz and Professor Ridha Hamila for their continuous support and guidance in the past years that helped sharpening my knowledge and nurturing my research potential.

My genuine thanks and appreciation extend to Mr. Anas Tahir for his technical assistance, to Mr. Omran Abdallah for assisting me in recording the data, and to Professor Sadok Sassi for facilitating the means for data collection.

I would also like to sincerely thank Professor Ahmed Massoud for his continuous support throughout my master's study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Protection and efficiency are essential considerations for manufacturing activities. Rotating machines are critical instruments employed in many sectors, yet they are vulnerable to damage due to difficult working environments and lengthy running durations. Failures of spinning machines can be observed as quickly as possible in order to avoid critical injury and abrupt stoppage in machine activity [1]. Failures contribute to disruptions in operations and, consequently, major economic losses. Examples of rotating machine components include pumps, bearing, gears, blades, motors, shafts, and engines.

Mechanical devices, e.g., wind turbines, aero-engines, aircraft, vehicles, and mining machines, utilize gearboxes widely in the transfer of speed and electricity [2]. However, when operating constantly in a hard-working setting, the gearboxes are highly vulnerable to numerous failures. When a fault occurs in the main machinery during service, odd features or device instability may lead to long downtimes, improved preservation losses, economic losses, or even loss of life. For example, transmission loss is one of the most important causes of wind turbine damage, and the efficiency of transmission networks is a significant concern for the wind industry. The error-diagnostic technique in the transmission method has been a popular field of study in recent years. While the gearbox is intact, false alarms may occur due to considerable noise or the complicated nature of the gearbox. It is therefore critical that the diagnostic device fails to efficiently minimize noise and abstract error.

Overall, the earlier servicing is maintained, the more effective and less costly it will be. In the study of engines, specialists determine whether a divergence from the usual state can contribute to restoration, or even replacement, as depicted from Figure 1 [3].

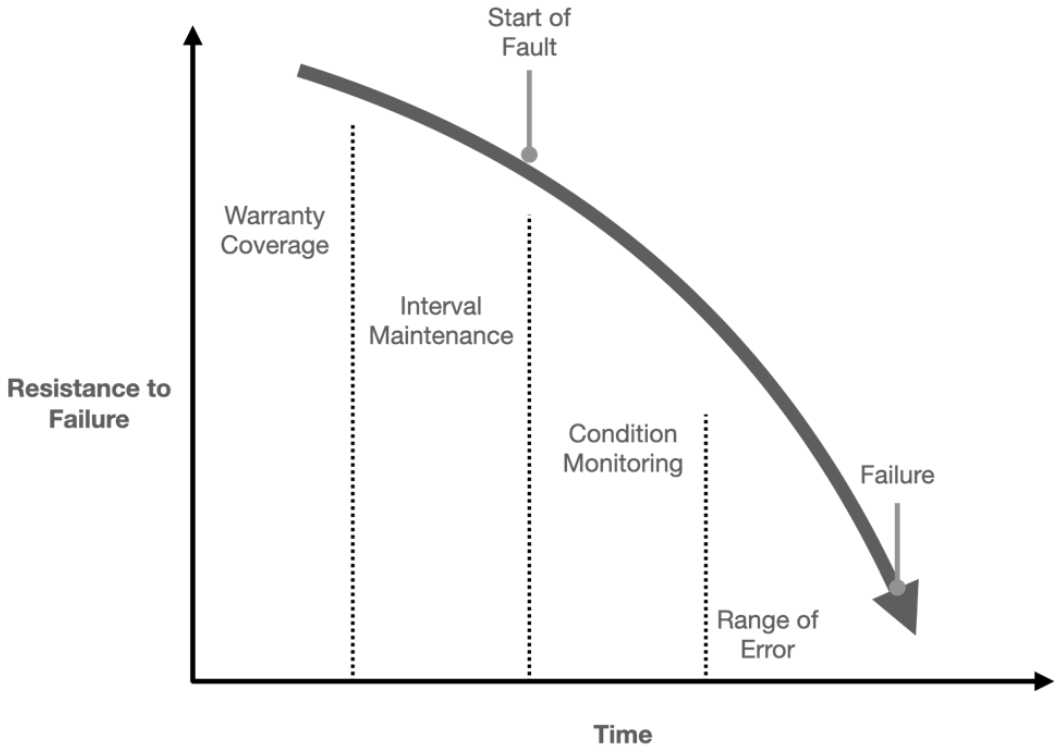

Figure 1. Typical machinery lifecycle. Adapted from [3].

Thus, initially, an unused machine is considered at the guaranteed phase, where failure cannot be completely excluded, although it is reasonably uncommon and may normally be tracked back to manufacturing faults. Targeted measures by adequately qualified service members can only begin in the following process of interval maintenance. This involves regular repairs done regardless of the state of the equipment at regular intervals. If the age of the system gets mature, the state monitoring process is reached. Faults can be predicted from this point forward. Figure 1 indicates the illustrated six changes: (1) beginning with adjustments in the ultrasonic spectrum; (2) accompanied by vibrations; (3) by means of a lubricant examination; (4) a minor rise in temperature; (5) the first symptoms of an outstanding malfunction may be observed until the real failure happens in the form of perceivable noise; and

(6) the generation of heat.

1.1 Background

Through the exponential advancement of information systems, Artificial Intelligence (AI) approaches are prevailingly utilized in a multitude of systems to support data analysis. However, current model-based experiments on electrical rolling devices and rotating machinery encounter problems in the evaluation of wellbeing due to noise-ridden, dynamic environments [4].

Recently, the most popular machine learning systems employed in fault detection and diagnosis approaches include Deep Neural Networks (DNN), Convolutional Neural Networks (CNN), the Boltzmann Restricted Machines (RBM), Stacked Autoencoders (SAE), and the Deep Belief Networks (DBN). Condition control of the machinery elements, such as gears, may be viewed as a pattern identification challenge that has been effectively resolved by intelligent diagnostic methods. According to current literature, condition monitoring, in general, consists of four steps: data collection, extraction, selection, and classification of features.

CNNs are widely used active learning models with applications in computer vision, voice recognition, and fault detection [5].

Conventional two-dimensional (2D) CNNs were successfully used for recognizing 2D signals, e.g., images and video frames [6], growing to be the de-facto method for numerous applications over vast data repositories. Although, 2D CNNs, particularly those with deep architectures, exhibit, first, high computational complexity. Thus, 2D CNNs cannot be ideal for error detection when handling one-dimensional (1D) signals in the absence of specialized hardware. In addition, the training of deep CNNs demands a large supply of training data in order to obtain an appropriate generalization capacity [6]. When deep 2D networks are trained over limited datasets, overfitting can easily occur, which significantly reduces the

generalization capability of the network. Compact 1D CNN models are designed in order to overcome those disadvantages to effectively execute on 1D signals [7], showing computationally light and reliable results in several real-time monitoring applications.

It is worth noticing that, on the other hand, traditional neural networks are based on a homogenous structure with a neuron that can only perform linear transformations (e.g., linear weighted sum or linear convolution). When highly complex, nonlinear problems are encountered, such traditional network architectures manifest varying or unsatisfactory performance levels, and therefore, the desired diversity and learning performance can only be achieved with the deep CNNs given high depth and complexity [8]. Recently introduced as an innovative network model, Self-organized Operational Neural Networks (Self-ONNs) are devised in order to overcome drawbacks of conventional CNNs by enabling neurons to use any desired nonlinear operator, expanding diversity and learning capacity for greatly sophisticated, multi-modal functions with minimum computational complexity and data needed for training [8].

In this work, a detailed investigation is conducted using compact 1D CNNs as well as Self-ONNs, not only on the gear fault detection but also diagnosis aiming at the classification of different fault severity levels. Particularly, the core objective of the presented work is the evaluation of the system's performance over Self-ONNs against 1D CNNs, and to find out the most (and the least) suitable signal type for gear fault diagnosis among current, vibration, and acoustic signal types.

1.2 Thesis Objectives

In order to carry out the investigation on the different signal types for gear fault diagnosis over 1D CNN and Self-ONNs, the core objectives of the presented work are summarized as follows:

1. Introducing a novel benchmark dataset encapsulating three types of raw sensed signal types: current, vibration, and acoustic signals.
2. Introducing a comprehensive investigation over different signal types for gear fault detection and severity diagnosis, which ranks the signal types with respect to the diagnosis accuracy and inference time for detection.
3. Proposing a solution that is not restricted to only detecting real-time gear fault occurrences, but also classifies the Cracking Level Percentage (CLP) to predict the fault severity.
4. Investigating the system's performance over compact 1D CNNs, as well as on Self-ONNs, for the three signal types via several diagnosis performance metrics.
5. Determining the signal type yielding the highest diagnosis performance and the fastest diagnosis time.

1.3 Thesis Outline

The rest of this thesis is organized as follows: Chapter 2 reviews recent ML techniques for machinery fault detection, with a focus on gear faults. Chapter 3 presents the methodology of the proposed gear fault diagnosis system over 1D CNNs and Self-ONNs, as well as the datasets employed. Chapter 4 reports experimental results with a discussion on the limitations and best practices. Chapter 5 concludes the thesis report with final remarks.

CHAPTER 2: MACHINE LEARNING TECHNIQUES FOR MACHINERY FAULT DIAGNOSIS

This section summarizes recent work that involves machine learning techniques, particularly compact models for machinery fault detection, with a focus on gear fault diagnosis (GFD).

2.1. Machine Learning for Machinery Fault Detection

In the research study proposed in [9], past implementations of ML for mechanical system failure have focused on historical data, restricting the use of components with an extended service duration. Besides, reported failure evidence is seldom valid for the particular conditions. The work directly tackles these problems for roller bearings that has race faults by producing data based on high-fidelity roller simulations of bearing dynamics, which are employed to train and test ML models. Following, a number of ML methodologies are compared and contrasted. From there, the authors devise the Nearest-Neighbor Dynamic Time Warping (NNDTW) for fault identification as a parameter-free method for detecting faults. Two new implementations of ML algorithms for error recognition are proposed in this work. They do not depend on statistical features; rather CNNs and NNDTW applied to angle synchronous averaging are seen to equal or outperform statistical feature-based classifiers. The novel implementation of NNDTW is of particular importance because new data can be used without re-training the algorithm and can provide better insight into the fault mechanism when it finds the most comparable simulation. The suggested technique can be extended to any (pseudo) cyclo-stationary defect in a spinning machine, and, on a broader scale, simulation evidence can enhance ML in industrial condition monitoring applications.

In [10], a novel Deep Learning (DL)-based method, named Deep Output Kernel

Learning, is devised to perform a collective diagnostic of multiple fault forms. By following Multi-Layer Extreme Learning Machine (ML-ELM), the work uses autoencoders for adaptive feature extraction, and then employs the selected deep features to create an output kernel regularizer function. Following optimization, the matrix of the output kernel is attained, which is used to construct the final diagnostic model via combining the multiple outputs of the proposed classifier. The experimental findings of the Case Western Reserve University (CWRU) and Intelligent Maintenance Systems (IMS) datasets indicate that, relative to a popular approach and eight standard ML-driven diagnostic approaches, the system is proven to effectively boost fault diagnosis precision at a reasonable latency.

Moreover, in [11], the authors pursued wavelet-based extraction for ball bearings' fault diagnosis using spalls that are localized on independent bearing elements. Hence, a number of requirements needed for both algorithmic training and testing, namely statistical, are determined through the Minimum Shannon Entropy Criterion wavelet technique. Following, the authors selected seven separate base wavelets for the analysis, where the complex Morlet wavelets achieved the best results from the raw vibration wavelet coefficients. Also, three AI methods are used for the detection of faults, two of them are supervised, i.e., vector assistance, vector quantization, while the third is unsupervised, namely self-organizing maps. Results of fault classification indicate that Support Vector Machine (SVM) achieved relatively superior performance compared to the other techniques. The findings demonstrate the possible use of these AI approaches to establish successful maintenance methods to deter catastrophic failures and reduce operational costs.

In [12], a systematic research is performed considering different diagnostic approaches by using ML for the identification of small bearing defects (hole and

scratch). This research seeks to explain the disparity between the diagnostic approach and its usefulness in the identification of IM-bearing defects, which is done using Fast Fourier Transform (FFT) and corresponding derived features are employed for training. Depending on the required application and use case, the diagnostic tool can be chosen from the sample.

In [13], the authors proposed a novel fault diagnostic technique for Rolling-Element Bearing (REB) based upon a specialized fuzzy sliding mode observer. Next, the Laguerre based AutoRegressive with eXogenous input (ARX-Laguerre) algorithm is proposed to model bearings in noisy, unpredictable environments. Comparatively, the ARX-Laguerre technique uses a fuzzy algorithm to improve the precision of the device modeling. Following, a traditional sliding mode observer is used for addressing issues of fault prediction for devices with high instability, e.g., the spinning devices. Consequently, the employed observer adaptively boosts reliability, versatility, and precision of approximating roller bearing faults. Finally, the machine learning method, named Decision Tree (DT), adapts to the threshold values used in this analysis for problems of fault detection and fault recognition, where the feasibility of the proposed approach is verified through the freely accessible CWRU Vibration Dataset.

In [14], a statistical hybrid motor-current bearing fault detection technique is proposed, where the Genetic Algorithm (GA) hand in hand with a DL model are employed. In the first position, the mathematical characteristics are derived from current signals. Following, features are reduced using the GA, resulting in three separate classification algorithms, i.e., K-Nearest Neighbors (KNN), DTs, and Random Forests (RF), which are employed to detect bearing faults. Using a fusion of classifiers results in precision enhancement and lighter computational load. Results indicate that the accuracy of the three classifiers is more than 97%, where the feasibility of the

devised model is compared against other previously introduced strategies.

With a general intelligent diagnostic approach consisting of the four steps: data collection, feature extraction, feature selection, and feature classification, Z. Lei et al. [15], for instance, have proposed the Domain Adaptation-Multiscale Mixed Domain Feature (DA-MMDF) method to target bearings under Polytropic Working Conditions (PWC). This method achieves complete, fault-sensitive, and adaptive feather extraction with high computational efficiency unlike other methods, in which either only one domain or several domains of a single scale, is employed for feature extraction. The proposed method outstands others in that, despite their adaptive feature extraction ability, are incapable of interpreting features. In particular, it employs the GS_XGBoost module to interpret the significance and the sensitivity of the extracted features, at the same time, and sort them. This is achieved with the simultaneous diagnosis of cross-domain fault by the domain adaptive module.

Similarly, Z. Huang et al. [16] introduced the Multi-source Dense Adaptation Adversarial Network (MDAAN) for fault diagnoses under PWC considering multi-source information fusion and facilitating sufficient utilization of inherent labels throughout the classification process. In their proposal, frequency spectra of the vibration signals are leveraged for the best utilization of data information using convolutional deep feature extraction and fusion.

Further, in the work of L. C. Brito et al. [17], an isolation forest based unsupervised model is employed for bearing fault detection. The technique utilizes vibration signals as input. Also, dimensionality reduction has been applied to increase model compactness as well performance.

A time-frequency based bearing fault detection and diagnosis system is introduced by W.-B. Zoungrana et al. [18]. The described method analyzes sensor

vibration signals and detect abnormalities as well as recognize fault level (e.g., ordinary, degraded, and faulty).

The details of individual features and drawbacks of the aforementioned related work of ML-based techniques for machinery fault detection are summarized in Table 1.

Table 1. Summary of Related Work for Machine Learning-Based Techniques for Machinery Fault Detection.

Work	ML Method	Signal Type	Application	Features	Drawbacks
[9]	Nearest-Neighbor Dynamic Time Warping (NNDTW)	Vibration	Bearing fault diagnosis	Parameter-free model.	Model developed on simulations.
[10]	Deep Output Kernel Learning	Vibration	Bearing fault diagnosis	Uses autoencoders for adaptive feature extraction.	Does not classify fault severity.
[11]	Wavelet-based extraction	Vibration	Bearing fault diagnosis	Uses localized spalls on independent bearing elements. Uses multiple algorithms for fault detection.	Does not classify fault severity.

Work	ML Method	Signal Type	Application	Features	Drawbacks
[12]	Fast Fourier Transform	Current	Bearing fault detection of induction motors	Uses multiple algorithms for fault detection.	Focused on detecting only minor faults and requires big data for high performance.
[13]	Advanced machine learning observer	Vibration	Rolling element bearing	Uses multiple algorithms for fault detection, including fuzzy sliding observer. Detection under variable conditions.	Minimal information on fault severity classification.
[14]	Hybrid genetic algorithm DL model	Current	Bearing fault detection of induction motors	Employs classifier fusion.	Does not classify fault severity.
[15]	Domain Adaptation-Multiscale Mixed Domain Feature (DA-MMDF)	Vibration	Bearing fault diagnosis for Polytropic Working Conditions (PWC)	Relatively high computational efficiency. Tested under variable speed and load conditions.	Employs manual parameter tuning. Does not classify fault severity.
[16]	Multi-source Dense Adaptation Adversarial Network (MDAAN)	Vibration	Bearing fault diagnosis for Polytropic Working Conditions (PWC)	Employs transfer learning	Does not classify fault severity.

Work	ML Method	Signal Type	Application	Features	Drawbacks
[17]	Isolation forest	Vibration	Bearing fault diagnosis	Employs dimensionality reduction	Does not classify fault severity.
[18]	Time-frequency	Vibration	Bearing fault diagnosis	Detects the operation modes: normal, degraded, and faulty.	Does not classify fault severity.

As has been preceded in Section 2.1 and Section 1.1, and as will be discussed in the following section, AI approaches have been commonly utilized in many systems to support fault detection of rotating machinery maintenance, which has largely grabbed the attention of recent research. As illustrated in Figure 2, a broad range of popular ML systems employed in fault detection and diagnosis systems has been reviewed in the literature, including the Boltzmann Restricted Machine (RBM), the Deep belief Network (DBN), and the Deep Neural Network (DNN), Stacked Autoencoder (SAE), Support Vector Machine (SVM), Convolutional Neural Network (CNN), Random Forests (RF), K-Nearest Neighbors (KNN), etc.

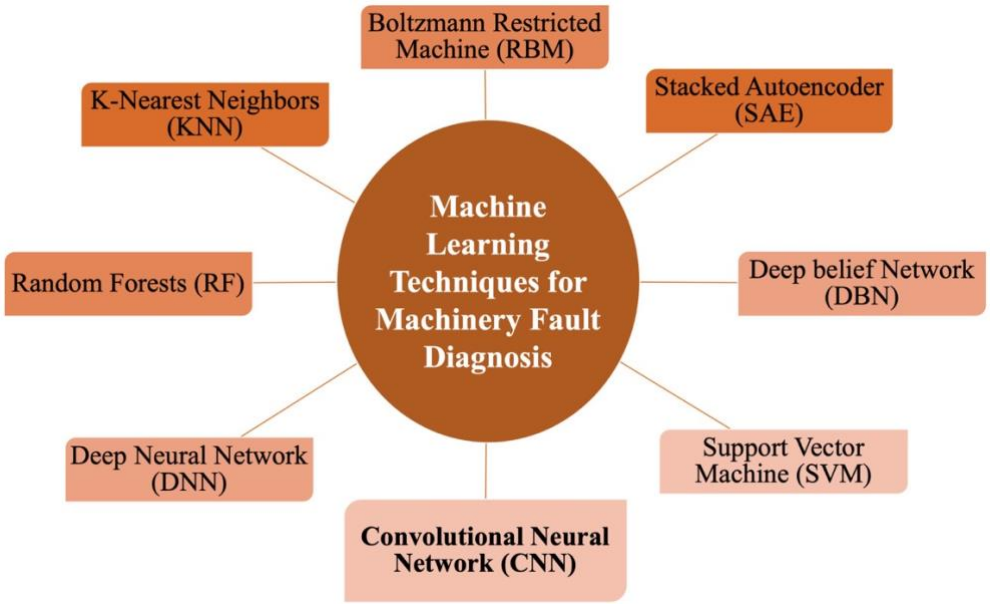

Figure 2. Some of the reviewed Machine Learning techniques for Machinery Fault Detection.

As will be noticed in the following section, and amongst of the reviewed techniques in the literature, it is evident that CNNs are widely used as active learning models with applications in fault detection [5]. Several researchers have recently attempted to use deep 2D CNNs to diagnose faults in bearings and gears. It is noted, however, that 2D CNNs, particularly those with deep architectures or of large-scale datasets, exhibit high computational complexity [6]. Thus, 2D CNNs cannot be ideal for error detection when handling 1D signals without special hardware or requiring large training data to obtain an appropriate generalization capacity of the network [6]. Moreover, when deep 2D networks are trained over limited datasets, overfitting can easily occur. On the other hand, compact 1D CNN models have been designed to overcome these disadvantages to run directly and more effectively on 1D signals. They showed high-speed and reliable results in several real-time monitoring applications.

2.2 ANN-Based Methods for Gear Fault Detection

Considering machine fault diagnoses, numerous signal processing techniques are extensively focused on the extraction of fault characteristics [19]. Widely employed Artificial Neural Networks (ANNs) techniques of signal processing aim to suppress the noise for the detection of fault characteristics. Compared to commonly utilized signal processing techniques, Stochastic Resonance (SR) can take advantage of signal noise in order to recover characteristics, and hence faults. Consequently, extensive usage has been made of fault-specific extraction and system fault identification.

Rapid recognition of faults of gears has become a major difficulty, since transmission faults exist mainly at the microstructure or also at the material stage, yet such results can solely be observed indirectly at the machine. So, the gear fault diagnostic device's performance is highly dependent on the classifier's feature [20]. While a lot of research attention has been devoted to recent approaches focusing on DNNs with feature extractions that are adaptive, they typically involve a significant collection of data for training.

Accordingly, failures in planetary gearboxes may contribute to injuries, downtime, not to mention expensive, regular maintenance requirements. To tackle such issues, Motor Current Signal Analysis (MCSA) are introduced as non-intrusive means for identifying defects in rotating machinery. Various DNN-based diagnostic approaches have been published, offering efficient methods to analyze vast volumes of data and generate reliable diagnostic reports, with near real-time performance. However, most modern techniques focus on detecting faults through vibration-based signals. The authors in [21] performed research on the diagnostic faults in planetary gears through pre-processing existing signals as well as utilizing DNN-based systems. Compared with related techniques, the merits of the approach are shown by the usage of laboratory results, which include significant measuring signals representing various

health problems under varying loading configurations.

In several industrial applications, gears are considered as one of the most critical elements of a mechanical piece of machinery, especially transmission devices. Also, gear pitting faults may trigger failure, which may result in safety and financial catastrophes, and therefore, gears should be regularly diagnosed, which is commonly carried out using raw vibration signal [22]. Accordingly, the work of [23] suggested a novel approach called Augmented Deep Sparse Auto-encoder (ADSAE), which is employed for gear pitting fault diagnosis with minimum data. Such approach creatively incorporates a deep, sparse autoencoder algorithm, where the efficacy is validated by examining six variants of gear pitting environments. Results show promising rise in both the generalization and robustness of the network with very high precision.

The authors in [24] analyzed gear states using linear and nonlinear techniques. The linear methods are Pareto maps, biplots, key angles, while nonlinear techniques include curvilinear component analysis, assuming that the manifolds of the given class are well-differentiated and compact enough. Therefore, a shallow neural network is utilized, rather than a deep one. Preliminary results support the authors' claim and compare it with the other published approaches.

In the work of [25], a method of diagnosis of gear loss was proposed based on the extraction of a numerical low-cost hybrid handcrafted feature collection, namely the Mel-Frequency Cepstral Coefficient (MFCC) and the Gamma Tone Cepstral Coefficient (GTCC), which are momentarily derived from vibration data with errors of a temporal type. Thus, a Long Short-Term Memory (LSTM) classifier is employed due to its merits in processing time-series data. A ten-fold cross-validation is applied to two separate data sets, with findings demonstrating that successful implementation of the classifier for the identification of gear defects. Nevertheless, only one 1D signal type is

addressed in their work for gear fault detection. Moreover, the datasets used in this study are of restricted sizes.

Furthermore, a new approach for detecting faults in gear pitting was introduced in [26]. The presented system was built upon a deep, sparse autoencoder, where the approach incorporates dictionary learning through a sparsely encoded network of auto-encoders. Sparse dictionary coding is used as an extraction tool for system malfunction diagnosis. The provided method utilizes a stacked auto-encoder network to conduct dictionary learning for raw vibration data automatic feature extraction, which are applied to locate gear pitting faults. Results reported on vibration signals are obtained from a number of tests on gearboxes' gears and compared to the current DL techniques.

Moreover, X. Wang et al. [27] proposed another bearing fault diagnosis that is dependent on fusing vibration and acoustic signals and extracting the features from the resulting fused data using 1D CNNs. This is done to achieve higher robustness with respect to other methods based on a single-modal sensor.

On the other hand, X. Li et al. [28] have targeted early gear pitting faults in mixed operating conditions using the vibration signal at different speeds. In the proposed method, the number of network parameters is reduced by leveraging channel convolution with point-by-point convolution. To achieve this, the method implements a separable convolution with a residual connection network identifying the study's significance in contrast to other related work.

Also, Yao et al. [29] presented a DL-based gear fault diagnostic approach focused on sound signal analysis for gear faults. By building up a CNN, namely with end-to-end architecture, both raw time and frequency domain signals can be used as input. In addition, a multi-channel audio sensing may be combined with the CNN without the need for an additional fusion technique. The obtained experimental findings

indicate that the proposed approach has much better efficiency in the diagnosis of gear loss relative to other conventional gear fault diagnostic approaches involving function engineering. An open-access sound signal database for the diagnostic of gear fault is also published. However, both studies [26] and [29] address only one signal type for gear fault detection.

There are certain limitations and drawbacks of the aforementioned studies. All of them were proposed on a single type of signal. Some of them exhibit high computational complexity; a challenge that confronts real-time applications, and others involve datasets that are usually limited in size leading to unreliable methods. Finally, most of them are only fault detectors and cannot grade the severity of the fault. The details of individual features and drawbacks of the prior art are encapsulated in Table 2.

Table 2. Summary Of Prior Art For ANN-based Gear Fault Detection.

Work	ML Method	Signal Type	Application	Features	Drawbacks
[20]	Transfer learning and Deep CNN	Vibration	Gearbox fault diagnosis	Requires minimal training. No pre-processing required. Fault severity analysis.	High computational complexity and requires special hardware.
[21]	DNN	Gear rotational speed	Planetary gear fault diagnosis	Verified using seven large datasets from the planetary gearboxes. Designed for industrial applications.	Requires extensive training time and high computational complexity.

Work	ML Method	Signal Type	Application	Features	Drawbacks
[22]	CNN	Acoustic emission	Aerospace gear pitting fault analysis	No frequency analysis required. Includes sparse autoencoder.	Not optimized for complex gear pitting faults.
[23]	Enhanced deep sparse autoencoder (ADSAE)	Vibration	Aerospace gear pitting fault analysis	Employs data augmentation.	High computational complexity.
[24]	Linear and nonlinear techniques	Vibration, torque, acoustic pressure and electrical current.	Induction Machine gear fault detection	Extensive comparison between deep and shallow methods. Uses data from different sensors	Does not classify fault severity. Not optimized for complex gear faults.
[25]	Gamma Tone Cepstral Coefficient (GTCC) and Mel-Frequency Cepstral Coefficient (MFCC)	Vibration	Gearbox fault detection	Long Short-Term Memory (LSTM) is employed.	Does not classify fault severity. Does not address variant signal types. Not optimized for complex gear faults.
[26]	Deep sparse autoencoder	Vibration	Gearbox fault detection	Does not require supervised parameter tuning.	Does not classify fault severity. Does not address variant signal types.
[27]	1D-CNN data fusion	Acoustic and Vibration	Bearing fault diagnosis	Multi-sensor data fusion with visualization analysis.	Does not classify fault severity.

Work	ML Method	Signal Type	Application	Features	Drawbacks
[28]	1D separable convolution	Vibration	Early gear pitting faults	Multi gear speed analysis in varying conditions. Optimized network parameters using point-by-point convolution.	Does not classify fault severity.
[29]	CNN	Acoustic	Gearbox fault detection	Can process data without manual configuration. Accompanied by sound signal dataset.	Does not classify fault severity. Does not address variant signal types. Requires additional 1D→2D conversions.

2.3 Self-Organized Operational Neural Networks (Self-ONNs)

Operational Neural Networks (ONNs), recently pioneered as an advancement to conventional CNNs, especially in highly complex and deep networks, where a generalized model is needed to elevate the classification performance. Heterogeneous by nature, ONNs can learn quite complex models using the least network complexity and data [8].

In order to attain maximum heterogeneity and boost computational efficiency, Self-Organized ONNs (Self-ONNs) have recently been proposed [30]. Self-ONNs are equipped with generative neurons to optimize operators during training, with experimental results signifying its prime performance in terms of training and computational efficiency, compared with conventional ONNs and CNNs.

A viable application of ONNs is medical image analysis. Devecioglu et al.

proposed a compact Self-ONN-based glaucoma detector from digital early fundus images [31]. Compared with the performance of CNNs on three datasets, the proposed technique excelled in terms of accuracy and computational efficiency. It is also worth noting that Self-ONNs relied on limited training to create the classification model, achieving a performance climb of approximately 8-12% F1-score compared to the conventional and deep CNN counterparts.

Moreover, electrocardiogram (ECG) classification is a topic of popular discussion, and novel methods have been put into practice to boost performance beyond the state-of-the-art. Henceforth, a Self-ONN-based ECG classifier is proposed to overcome existing challenges [32]. With its self-organization nature, Self-ONNs excel over conventional ONNs and 1D CNNs through its novel operator search function to optimize for the most suitable operator. Using the least amount of training data, the proposed classifier produced superior results compared to conventional techniques. Moreover, Kiranyaz et al. proposed a robust signal peak detection on Holter ECG with the use of Self-ONNs, achieving a higher computation efficiency and superior detection results on the China Physiological Signal Challenge-2020 (CPSC) dataset compared against deep CNN implementations [33].

On a similar note to this report's work, an early bearing fault detection and diagnosis system is developed based on 1D Self-ONNs [34]. By enhancing the learning performance of its CNN counterpart, the authors' solution provided a real-time computational performance. The proposed technique is also tested on the NSF/IMS bearing vibration dataset and benchmarked against conventional diagnosis techniques.

2.4 Core Concepts

From a biological standpoint, linear first-order models of living neurons inspired artificial neurons used in traditional ANNs [35]. Biological learning is primarily conducted at the neuronal stage of the human nervous system. Every neuron

can handle electric signals as per three operations: receiving signals by synaptic connections from other neurons; the pooling of the processed output at the soma of the nucleus; and the activation of the final signal at the Axon hillock.

2.4.1 Two-Dimensional CNNs

Conventional 2D deep CNNs have gradually been the predominant technique used in various Computer Vision (CV) and DL applications. They are highly complex and data-hungry ML paradigms. Figure 3 overviews the structure of 2D CNNs.

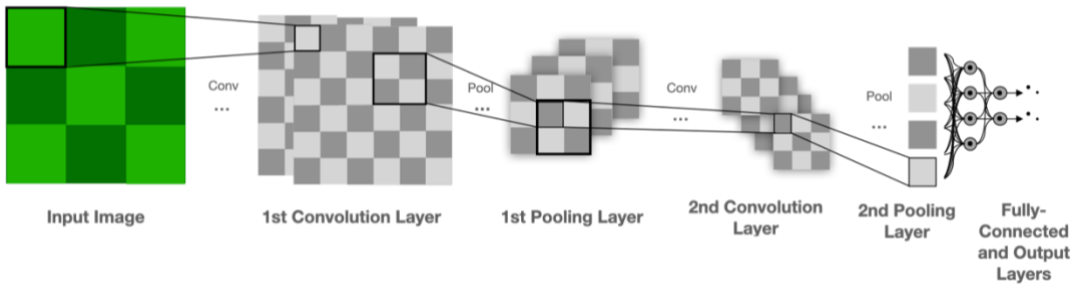

Figure 3. Overview of 2D CNN.

Unlike a typical neural network with a complete connection through each node, a CNN greatly decreases the number of the parameters in the network via local connectivity and exchanging weights using convolutional layers, which comprises a group of kernels with a limited receptive area. Each kernel passes through the input volume, executing the convolution process.

2.4.2 One-Dimensional CNNs

Alternatively, the 1D variant of CNNs, namely 1D CNNs, has recently been proposed for several applications over 1D signal repositories [35]-[39]. 1D CNNs can be desirable to their 2D equivalents due to their lower computational complexity. Compact 1D CNNs generally require fewer parameters, do not require high computational resources (can use CPU effectively), and can achieve real-time results

in low-cost applications [35]. In the meanwhile, kernel parameters stay constant to manage the free parameters amount. Hence, for a convolutional layer in the l^{th} layer, the 1D forward propagation computation is depicted as follows [35]:

$$x_k^l = b_k^l + \sum_{i=1}^{N_{l-1}} conv1D\left(w_{ki}^{l-1}, s_i^{l-1}\right) \tag{1}$$

where x_k^l is denoted as the input b_k^l, refers to as the bias of the k^{th} neuron at layer l, s_i^{l-1} is the output of the i^{th} neuron at layer $l-1$, w_{ki}^{l-1} is the kernel from the i^{th} neuron at layer $l-1$ to the k^{th} neuron at layer l. $conv1D$ is utilized to perform 1D convolution.

The algorithm for back-propagation (BP) is expressed as follows: The back propagating error begins from the output Multi-Layer Perceptron (MLP) layer. Let us assume $l = 1$ for the input layer and assume $l = L$ for the output layer. Let N_L be the number of classes in the dataset and so, for an input vector p, and its target and output vectors, $\boldsymbol{t^p}$ and $\left[y_1^L, \cdots, y_{N_L}^L\right]'$ respectively. Henceforth, in the output layer for the input p; the Mean-Squared Error (MSE), E_p, can be defined as:

$$E_p = \text{MSE}\left(\boldsymbol{t^p}, \left[y_1^L, \cdots, y_{N_L}^L\right]'\right) = \sum_{i=1}^{N_L} \left(y_i^L - t_i^p\right)^2 \tag{2}$$

To differentiate E_p by every parameter in the network, the delta error, $\Delta_k^l = \frac{\partial E}{\partial x_k^l}$ should be found, we can use the chain-rule:

$$\frac{\partial E}{\partial w_{ik}^{l-1}} = \Delta_k^l y_i^{l-1} \qquad \text{and} \qquad \frac{\partial E}{\partial b_k^l} = \Delta_k^l \tag{3}$$

Thus, from the 1st MLP layer to the last CNN layer, a process of scalar BP is carried out as:

$$\frac{\partial E}{\partial s_k^l} = \Delta s_k^l = \sum_{i=1}^{N_{l+1}} \frac{\partial E}{\partial x_i^{l+1}} \frac{\partial x_i^{l+1}}{\partial s_k^l} = \sum_{i=1}^{N_{l+1}} \Delta_i^{l+1} w_{ki}^l \tag{4}$$

Once the 1st BP is done from the subsequent layer, l+1, to the current layer, l, then the BP can be executed for the input delta of the CNN layer l, Δ_i^l. Let a map of zero-order up-sampled be: $\mathrm{us}_k^l = \mathrm{up}(s_k^l)$, then the delta error can be computed as:

$$\Delta_k^l = \frac{\partial E}{\partial y_k^l} \frac{\partial y_k^l}{\partial x_k^l} = \frac{\partial E}{\partial \mathrm{us}_k^l} \frac{\partial \mathrm{us}_k^l}{\partial y_k^l} f'(x_k^l) = \mathrm{up}(\Delta s_k^l) \beta\, f'(x_k^l) \tag{5}$$

where $\beta = (ss)^{-1}$. So, the BP of the delta error $\left(\Delta s_k^l \overset{\Sigma}{\leftarrow} \Delta_i^{l+1}\right)$ can be defined as

$$\Delta s_k^l = \sum_{i=1}^{N_{l+1}} conv1Dz\left(\Delta_i^{l+1}, rev(w_{ki}^l)\right) \tag{6}$$

where $rev(\,)$ is employed to reverse the array and $conv1Dz(\,)$ is used to carry out a zero-padding 1D convolution process. Hence, the weight and bias sensitivities are be defined as:

$$\frac{\partial E}{\partial w_{ik}^l} = conv1D(s_k^l, \Delta_i^{l+1}) \quad \text{and} \quad \frac{\partial E}{\partial b_k^l} = \sum_n \Delta_k^l(n) \tag{7}$$

As weight and bias sensitivities are calculated, biases and weights with the learning factor, ε, can be updated as:

$$w_{ik}^{l-1}(t+1) = w_{ik}^{l-1}(t) - \varepsilon \frac{\partial E}{\partial w_{ik}^{l-1}} \quad \text{and} \quad b_k^l(t+1) = b_k^l(t) - \varepsilon \frac{\partial E}{\partial b_k^l} \tag{8}$$

Figure 4 depicts the structure of 1D CNNs. In this work, compared to the reviewed literature, a technical investigation is carried out using compact 1D CNNs, not only for gear fault detection, but also gear fault diagnosis with the classification of different severity levels on current signals, vibration signals, and acoustic signals.

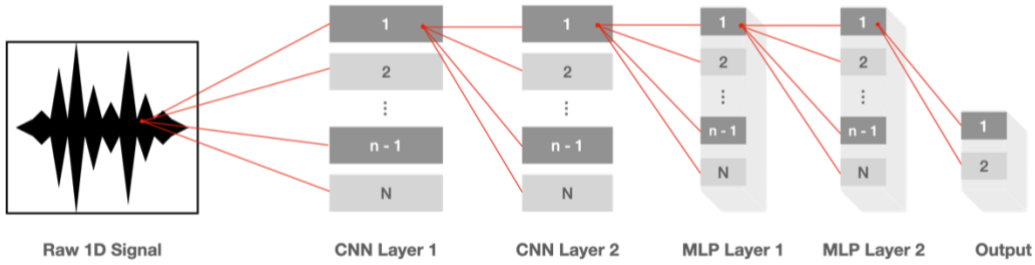

Figure 4. Overview of 1D CNN.

2.4.3 Self-Organized ONNs (Self-ONNs)

CNNs, as a derivative of Multi-Layer Perceptrons (MLPs), correspond to one specific operator set model: multiplication for the nodal operator, summation for the pooling operator, and sigmoid for the activation function. As illustrated in Figure 5 (left), this transformation is repeated in all neurons in all layers in the network, making the network homogenous and linear, whereas the networks in biological neural systems are heterogenous; highly diverse, and almost all the time nonlinear. Therefore, a homogeneous network architecture is considered as a major disadvantage encountered in CNNs, as they impose the exact same linear neuron model in all kernel connections. This, however, does not represent the heterogeneity property of the actual biological neural systems, which constitutes variant types of neurons. Such immature models can be adequate for a relatively straightforward linear solution space, nevertheless, their performance degrades with a nonlinear solution space of high complexity [8], [30].

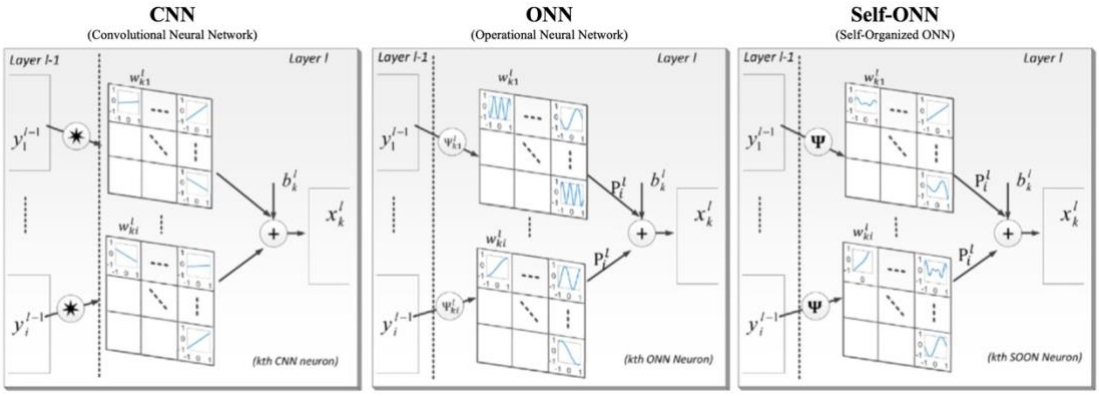

Figure 5. An illustration of the nodal operations in the kernels of the k^{th} CNN (left), ONN (middle), and Self-ONN (right) neurons at layer l [30].

Although there are several studies in the literature that attempt to overcome this drawback, none address the core problem. To model the true models of biological neurons, ONNs [8] have been recently proposed. They apply the same operator set to all neurons within the same hidden layer, but not necessarity summation for the pooling operator or multiplication for the nodal operator, making it a heterogeneous and non-linear network model that outstands CNNs in many challenging learning problems.

Adopting the core concept of Generalized Operational Perceptrons (GOPs), ONNs have been developed to present a heterogeneous network model that incorporates diversity found in biological neural networks and outperforms conventional MLPs failing in many challenging problems. Furthermore, ONNs can be established by determining the most suitable operator sets to be used in the entire set of neurons of the hidden layers. If the optimal operator set in an ONN is found to be multiplication for the nodal operator, summation for the pooling operator, and sigmoid for the activation function, then the homogenous ONN model is identical to that of a CNN model. Therefore, ONNs do not cancel CNNs, but they extend the exclusive usage of convolutions in a hidden layer. ONNs, consequently, are considered as a superset of

CNNs sharing the core concept of the heterogeneous nature of GOPs and learning complex and multi-modal spaces by facilitating higher diversity with training data.

Although an ONN outstands a CNN in maximizing the learning performance, it is still limited by two certain drawbacks. First, as shown in Figure 5 (middle), it still exhibits limited heterogeneity due to the usage of a single operator set for all neurons in a hidden layer. Second, it is also limited by a fixed set of operators (the nodal operators) that are specified in a predefined operator set library containing standard functions [30].

To address this limitation, further optimization is facilitated by the recently proposed Self-organized ONNs (Self-ONNs) with generative neurons [30]. Owing to their generative neurons, Self-ONNs do not need an in-advance fixed library set of operators, as they self-organize the network without searching for the optimal nodal operator during training. As illustrated in Figure 5 (right), first, Self-ONNs can approximate the optimal nonlinear or linear nodal operator, such that every neuron can create any suitable composition of operators including linear, sinusoids, exponential, hyperbolic, etc. Therefore, a nodal operator can be a composite function, and not necessarity a standard function, that is of a certain order refered to as the Q^{th} order, so they don't need to search in the operator set library in advance. Second, this is done for every kernel element and for every kernel connection, in contrary to applying the same nodal operator for all neurons within the same hidden layer as in ONNs.

Therefore, a generative-neuron can be defined as a neuron having a composite nodal-operator created freely at training. Consequently, a Self-ONN may organize its nodal operators automatically with nodal operator functions optimized to achieve the optimum learning performance. For instance, CNN and ONN neurons include static nodal operators for their corresponding kernels, whereas the generative neuron has the

flexibility to become any desired nodal function at each connection element. Such functionality provides a significant deal of versatility, allowing for the synthesis of any nodal operator function as needed. Also, this training technique propagates errors back via the Self-ONN's operational layers to be able to generate the suitable nodal functions.

Hence, Self-ONNs significantly maximize the generalization capacity and expand diversity of the network. It is noteworthy to mention that Self-ONNs can accomplish similar, and in most cases, greater performance compared with the analogous ONNs with higher computing efficiency, whereas when contrasted with CNNs, the performance difference is even more recognizable when Self-ONNs are employed.

Overall, conventional CNNs, like MLPs, utilize the traditional model of linear neurons; yet impose two structural constraints: (a) kernel restricted connections and (b) weight sharing, and so the linear weighted sum for MLPs is transformed into the convolution process utilized at CNNs. Conversely, ONNs adopt the fundamental concept of GOPs, and therefore, extend the nodal and pool operators' exclusive use of linear convolutions in neurons. The input map of the k^{th} neuron in the current layer, x_k^l, is generated by pooling the final output mappings, y_i^{l-1}, of the previous layer neurons operated with their corresponding kernels, w_{ki}^l, as follows:

$$x_k^l = b_k^l + \sum_{i=1}^{N_{l-1}} oper2D\left(w_{ki}^l, y_i^{l-1}, 'NoZeroPad'\right) \tag{9}$$

$$x_k^l(m,n)\big|_{(0,0)}^{(M-1,N-1)} = b_k^l + \sum_{i=1}^{N_{l-1}} \left(P_k^l \begin{bmatrix} \Psi_k^l\left(w_{ki}^l(0,0), y_i^{l-1}(m,n)\right), \dots, \\ \Psi_k^l\left(w_{ki}^l(r,t), y_i^{l-1}(m+r,n+t), \dots\right), \dots \end{bmatrix} \right)$$

When examining the above equation carefully, it is evident that the pool operator is a summation, i.e., $P_k^l = \Sigma$, and the fact that the nodal operator is linear,

$\Psi_{ki}^{l}\left(y_i^{l-1}(m,n), w_{ki}^{l}(r,t)\right) = w_{ki}^{l}(r,t)y_i^{l-1}(m,n)$, for every neuron, then the produced ONN is both homogeneous and identical to a CNN, verifying the observation that ONNs are considered a superset of CNNs since GOPs are a superset of MLPs.

As previously stated, a generative neuron is produced repeatedly during BP training with no constraints. To enhance the learning performance, each generative neuron in a Self-ONN have self-optimized operators via BP training for each kernel element and connection link. A natural option for generating a composite nodal operator can be the weighted sum, as follows:

$$\Psi(\boldsymbol{W}, y) = w_1 Sin(w_2 y) + w_3 \exp(w_4 y) + \cdots + w_Q y \qquad (10)$$

where w is a Q-dimensional parameter array made of weights as well as internal parameters of each function. However, owing to the various and combined dynamic ranges of the component non-linear functions, such construction of composite functions can have significant stability problems. Furthermore, tuning a plethora of parameters is required, particularly when the operator set library includes a large number of unique nodal operator functions, which conveniently can be formed using other traditional techniques such as Taylor Polynomials or Fourier Series. Due to its less complex nature, Taylor Polynomials is a superior option, which can be depicted for a function, $f(x)$, at a point, $x = a$, as follows:

$$\mathrm{f}(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots \qquad (11)$$

where f', f'' and f''' are the 1st, 2nd, and 3rd derivatives, respectively. Hence, the function of the composite nodal operator can be derived via the Q^{th} order truncated Taylor approximation as follows:

$$\Psi(\boldsymbol{W}, y) = w_0 + w_1(y-a) + w_2(y-a)^2 + \cdots + w_Q(y-a)^Q \qquad (12)$$

where $w_q = \frac{f^{(q)}(a)}{g!}$ is the q^{th} parameter of the Q^{th} order polynomial. Such process is

optimized throughout the training phase to estimate the best-fitting operator for every kernel element of every individual inter-neuron connection. Yet, an obvious problem comes up: this estimate is only viable near $y = a$. So, when the closer the points are to a, the less accurate the approximation becomes. Although, this phenomenon has no effect on Self-ONNs because the nodal operators work over the output of neurons of the preceding layer, where each layer is limited by the activation operator function's generating range. So, in this work, the tangent hyperbolic (tanh) activation function is employed, with a range of [-1, 1]. Hence, $a = 0$ in this instance, and the Q^{th} order Taylor approximation in Eq. (13) is reduced to the Maclaurin series as,

$$\Psi(\boldsymbol{W}, y) = w_0 + w_1 y + w_2 y^2 + \cdots + w_Q y^Q \tag{13}$$

Hence, the bias coefficient, w_0, can be dropped, as the resultant DC bias is factored out by the bias of every neuron. It is crucial to highlight that the existence of generative neurons with the composite nodal operator, a Q^{th} order Maclaurin polynomial, is the major distinction between ONNs and Self-ONNs. Consequently, every kernel element is a Q-dimensional array, and the weight kernels, w_{ik}^l, are 3D matrices that are equal to the $w_{ik}^{l+1}<q>, q = 1, \dots, Q$ array of Q 2D matrices.

In this work, compared to the reviewed literature, a technical investigation is carried out using Self-ONNs, not only for gear fault detection, but also gear fault diagnosis with the classification of different severity levels on current signals, vibration signals, and acoustic signals.

CHAPTER 3: DATASETS AND METHODOLOGY

3.1. System Overview

The workflow of the proposed system involves the design of datasets, in which different raw signal types are extracted to investigate their changing behavior in fault gear detection and diagnosis. This enables the intended facilities to perform safe and cost-effective operations and structural health maintenance. Figure 6 shows an overview of the system's workflow defined in Work Packages (WPs). Each WP is described thoroughly in Chapters 3 and 4.

Figure 6. Proposed Workflow.

In WP1, the raw data of current, vibration, and acoustic signals are recorded from a design test-rig system to diagnose the severity level of their fault condition. In this work, the fault condition is specified as the presence of cracking on the arc of the gear's teeth. Including different signal types facilitates the investigation, presents its outcomes, and determines the best signal type for fault gear analysis.

In WP2, a re-sampling rate for each signal type is determined. The signals are then segmented into non-overlapping frames and then normalized. Before feeding into the 1D CNN and Self-ONN model architectures, all data acquisitions are then prepared to be divided into training, validation, and testing sets over four-Fold Cross-Validation (FCV).

In WP3 and WP4, the model architecture is determined by its hyperparameters and complexity, and the diagnosis is accomplished by classifying the gear faults into five classes based on the cracking percentage levels.

Finally, in WP5, the network's performance is quantized as per five main classification evaluation metrics, which are Accuracy, Precision, Sensitivity, F1-score, and Specificity.

3.2 Benchmark Dataset Creation

The dataset is acquired from a test rig system at Qatar University (QU) as shown in Figure 7. An accelerometer, a microphone, and a current measuring probe were connected to the test rig system and were used to record vibration, acoustic, and current signals, respectively. Gears of different anomalies were attached to the rotor test rig. In the recorded dataset, five classes of faulty gears are considered for detection and diagnosis. The fault severity is represented by the Cracking Level Percentage (CLP), which is a percentage of the cracking with respect to the arc length of a gear's tooth. An example of a healthy gear that does not have any cracking is illustrated in Figure 8 (left) in contrast with a 30% CLP faulty gear (middle), and a 50% CLP faulty gear (right). The five classes of faulty gears are labeled as: 0% CLP (healthy), 10% CLP, 20% CLP, 30% CLP, and 50% CLP (severe).

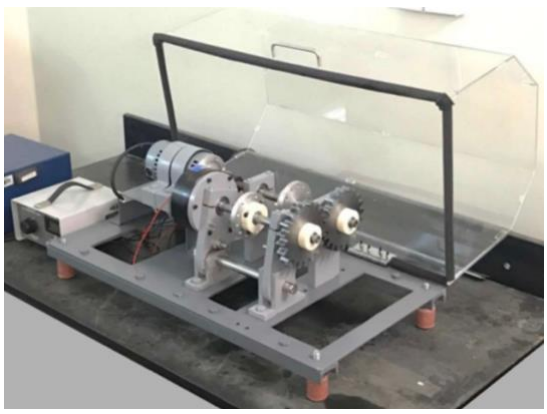

Figure 7. Acquisition test rig system with two gears attached at a QU lab.

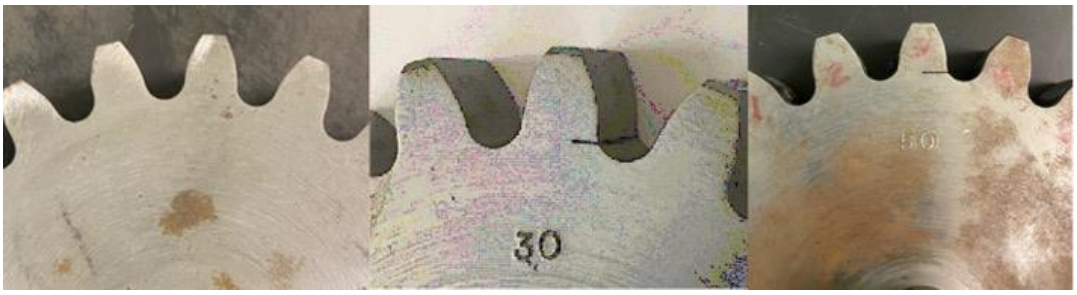

Figure 8. A healthy gear that does not have any cracking (left), a 30% CLP faulty gear (middle), and a 50% CLP faulty gear (right).

At a sampling rate of 51.2 kHz, four different Data Acquisitions (DAs) were taken for 30 seconds. Data streams of all the five labeled classes for each of the different signal types (current, vibration, and acoustic) are recorded in each data acquisition. Therefore, for each DA, 30 x 51,200 = 1,356,000 samples are recorded for each labeled class, hence, a total of 1,356,000 samples x 5 classes = 6,780,000 samples for each signal type. Table 3 summarizes the recordings for each data acquisition.

Table 3. Summary of recorded data acquisitions.

	For Each DA		
Signal type	Current	Vibration	Acoustic
Classes	5 classes 0% CLP, 10% CLP, 20% CLP, 30% CLP, and 50% CLP.	5 classes 0% CLP, 10% CLP, 20% CLP, 30% CLP, and 50% CLP.	5 classes 0% CLP, 10% CLP, 20% CLP, 30% CLP, and 50% CLP.
Sampling rate	51.2 kHz	51.2 kHz	51.2 kHz
No. of samples per class	1356000	1356000	1356000

3.3. Training Phase

In this operation, three independent 1D CNN as well as Self-ONN architectures are trained for each signal type as in Figure 9. Initially, the three datasets required to train each 1D CNN/Self-ONN architecture are down-sampled, segmented into non-overlapping frames, normalized, shuffled, and produced as follows.

1. The first dataset is composed of current signals X_{curr}, pre-processed into N frames, $[C_1, C_2, \ldots, C_N]$.
2. The second dataset is composed of acoustic signals X_{aco}, pre-processed into M frames, $[A_1, A_2, \ldots, A_M]$.
3. The third dataset is composed of vibration signals X_{vibr}, pre-processed into K frames, $[V_1, V_2, \ldots, V_K]$.

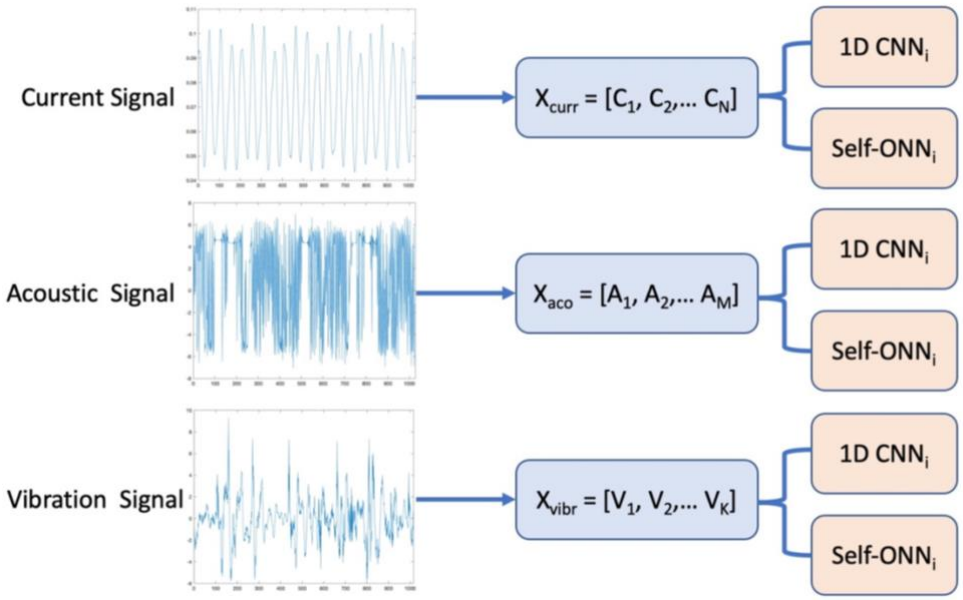

Figure 9. Training of three independent 1D CNN/Self-ONN architectures for different signal types.

It is worth mentioning that, as per the core concept of Self-ONNs, the value of the q^{th} parameter of the Q^{th} order polynomial specifies whether the running model

architecture is 1D CNN or Self-ONN, as 1D CNNs are a special case of the generalized Self-ONNs holding an order of $q = 1$, whereas Self-ONN models can be specified when $q > 1$. In this work, $q = 1$ has been set for evaluating the learning performance of a 1D CNN model architecture, and $q = 3$ and $q = 5$ order values, each at once, have been set for evaluating the learning performance of the Self-ONN model architecture.

Several hyperparameters define the structure of a 1D CNN/Self-ONN, which are usually determined by trial and error [6]:

1. Number of hidden layers
2. Number of hidden fully connected layers
3. Number of neurons in each hidden and fully connected layer
4. Filter kernels size
5. Subsampling factor

In this work, the architecture of the 1D CNN and Self-ONN models has been implemented using Pytorch and Google Colab. Pytorch[1] is a library that consists of several components including 'torch;' a Tensor library with strong Graphics Processing Unit (GPU) support, and 'torch.nn;' a neural networks library that is designed for maximum flexibility and deeply integrated with the 'autograd' library. The computational platform, CUDA, is employed.

Google Colab [40], on the other hand, is an online Python editor from Google Research, allows researchers to write, execute, and collaborate on Python documents (i.e., Jupyter notebooks) on Virtual Machines (VMs). It offers several options for virtual Central Processing Units (vCPUs) with varying performance. Also, it includes different machine types with a selection of vCPUs and memory per vCPU for different

[1] https://pypi.org/project/torch/

applications. Notwithstanding, different GPUs can be employed, where they can be used alongside vCPUs.

The deployed structure of the 1D-CNN and the Self-ONN models is defined in Table 4.

Table 4. Hyperparameters definition of the 1D CNN and Self-ONN architecture models.

Hyperparameter	Value
Number of hidden learning layers	3
Number of neurons of hidden learning layers	[64, 32, 16]
Subsampling factor in hidden learning layers	[8, 8, 2]
Kernels sizes in hidden learning layers	[41, 11, 11]
Number of hidden MLPs	2
Number of neurons of hidden MLPs	[32, 32]
Activation Function at hidden layers	Tanh
Number of neurons at the output MLP	5

3.4. Network Configuration Parameters

3.4.1. Classification Loss Function

The Mean Squared Error (MSE) function is used as the cost function for the classification networks. By the MSE loss function, a criterion that measures the mean squared error is created between each element in the predicted, y, and the target (truth) t class vectors. This is computed by summing up the squared (pairwise) differences and dividing by the number of such pairs as in Eq. (14).

$$MSE = \frac{\sum_{i=1}^{n}(y_i - t_i)^2}{n} \tag{14}$$

where n is the batch size.

3.4.2. Other Configuration Parameters

Other configuration functions and parameters are summarized in Table 5, which are also determined by trial and error.

Table 5. Configuration parameters used within the network architecture.

Parameter	Value
Learning rate	1e-4
Confidence Interval	0.9
Batch size	8
Optimization function	Adam
Number of epochs	200 for current and vibration signals, and 500 for acoustic signals

CHAPTER 4: EXPERIMENTAL RESULTS AND DISCUSSION

4.1. Experimental Setup

The signal is preprocessed into several stages as follows:

1. Frame down-sampling
2. Frame segmentation
3. Frame normalization
4. Fold cross-validation

The steps are described in detail in the following sub-sections.

4.1.1. Frame Down-Sampling

All signal types are down-sampled by a varying set of factors in all data acquisitions. Down-sampling rates of 5.12, 12.8, and 25.6 kHz are chosen for the current, vibration, and acoustic signals, respectively, based on the observation of the raw signals' domains shown in Figure 10. Nevertheless, this does not impede the consistency of the 1D CNN/Self-ONN models, as the current, vibration, and acoustic signals are independently trained over separate network models with the same aforementioned network configurations and architecture hyperparameters.

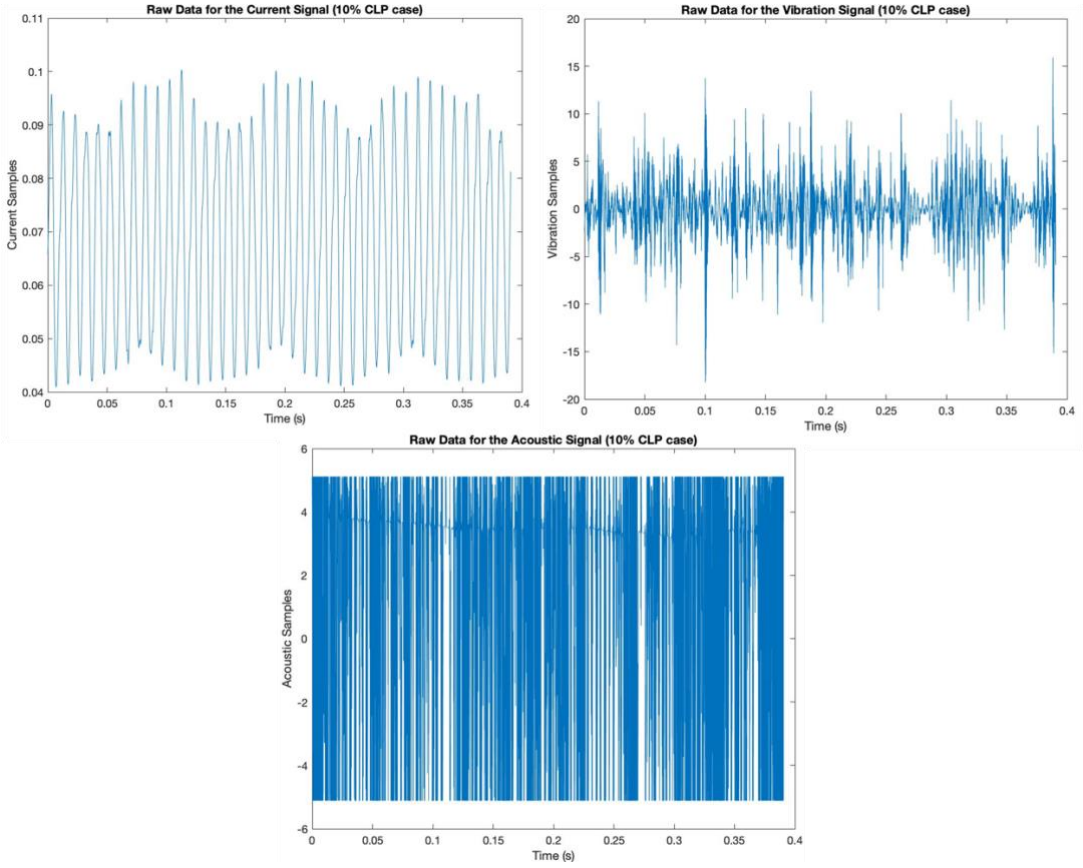

Figure 10. Raw samples of the current (upper left), vibration (upper right), and acoustic (bottom) signals for gear fault detection and diagnosis before the down-sampling process for the 10% CLP class as an example.

4.1.2. Frame Segmentation

In each data acquisition, all signal types are segmented into non-overlapping frames of 1024 samples fixed length, and 200 ms, 80 ms, and 40 ms frame durations for the current, vibration, and acoustic signals, respectively, are resulted.

Table 6 summarizes the down-sampling and segmentation processes, and the produced number of frames for each data acquisition DA.

Table 6. Pre-processing summary for each data acquisition.

For Each DA			
1D Signals	Current	Vibration	Acoustic
Duration	200 ms	80 ms	40 ms
Samples (per class)	1536000	1536000	1536000
Down-sampled rate	5.12 kHz	12.8 kHz	25.6 kHz
Down-sampled number of samples (per class)	153600	384000	768000
Number of Classes	5	5	5
Frame Length (samples)	1024	1024	1024
Number of Frames	150 per class 150 * 5 classes = 750	375 per class 375 * 5 classes = 1875	750 per class 750 * 5 classes = 3750

4.1.3. Frame Normalization

Data normalization is one of the good practices performed before training a neural network in order to obtain a mean close to zero, in addition to yielding faster learning and faster convergence. Each frame is linearly scaled (normalized) within [-1,1] as follows:

$$X_{normalized} = 2\frac{X - X_{minimum}}{(X_{maximum} - X_{minimum})} - 1 \quad (15)$$

where X is a sample amplitude in a frame, and $X_{normalized}$ is its normalized version. The normalized frames afterward are shuffled to be prepared for a four-FCV.

4.1.4. Four-fold Cross-Validation

The three 1D CNN and Self-ONN models are trained using four-Fold Cross-Validation (FCV), with 80% are dedicated to the training set and unseen 20% of data dedicated to the testing set. A 20% portion of the training set is used as a validation set to avoid overfitting.

Table 7 summarizes the data acquisitions, per signal type, used for training, validation, and testing in each fold.

Table 7. A summary of 4-fold cross-validation for each signal type using 4 data acquisitions.

	Fold 1	Fold 2	Fold 3	Fold 4
Training set	DA 1, 2, 3	DA 2, 3, 4	DA 3, 4, 1	DA 4, 1, 2
Validation set	20% of the training set			
Testing set	DA 4	DA 1	DA 2	DA 3

4.2. Classification Evaluation Metrics

The network's performance for each signal type is quantized as per five main classification evaluation metrics, which are Accuracy, Precision, Sensitivity, F1-score, and Specificity.

An overall confusion matrix is produced accumulating all test results per class. The evaluation metrics are computed per class i as follows.

1. Per-class accuracy as in Eq. (16)

$$Accuracy_{class_i} = \frac{TP_{class_i}+TN_{class_i}}{TP_{class_i}+TN_{class_i}+FP_{class_i}+FN_{class_i}} \quad (16)$$

2. Per-class precision as in Eq. (17)

$$Precision_{class_i} = \frac{TP_{class_i}}{TP_{class_i}+FP_{class_i}} \quad (17)$$

3. Per-class sensitivity as in Eq. (18)

$$Sensitivity_{class_i} = \frac{TP_{class_i}}{TP_{class_i}+FN_{class_i}} \quad (18)$$

4. Per-class specificity as in Eq. (19)

$$Specificity_{class_i} = \frac{TN_{class_i}}{TN_{class_i}+FP_{class_i}} \quad (19)$$

5. Per-class F1-score as in Eq. (20)

$$F1-\text{score}_{class_i} = 2 * \frac{Precision_{class_i} * Sensitivity_{class_i}}{Precision_{class_i} + Sensitivity_{class_i}} \quad (20)$$

where True Positives (TP) is the number of frames in the correctly predicted positive class by the model, True Negatives (TN) is the number of frames in the correctly predicted negative class, False Positives (FP) is the number of frames in the incorrectly predicted positive class, and False Negatives (FN) is the number of frames in the incorrectly predicted negative class.

The overall accuracy is calculated as in Eq. (21)

$$Overall\ Accuracy = \frac{\sum_{i=1}^{k} TP_{class_i}}{\sum_{i=1}^{k} S_{class_i}} \tag{21}$$

where k is the total number of classes, and S is the total number of predicated frames for each class i (i.e., $\sum_{i=1}^{k} S_{class_i}$ is the sum of all frame elements in a confusion matrix).

4.3. Gear Fault Detection Experimental Results

The numerical and quantitative evaluation results for the current, vibration and acoustic signals are presented in this section. It is worth mentioning that, for each signal type, the 1D CNN and Self-ONN models at the epoch of the highest validation accuracy are selected for each fold. First, the system runs over the 1D CNN network architecture by setting the order value $q = 1$ for each signal type. The system then runs on the Self-ONN network architecture by adjusting the order value to $q = 3$, and later to $q = 5$ to observe the performance differences. The models' configurations described in Table 4 and Table 5 are set fixed for both 1D CNN and Self-ONN network architectures.

4.3.1. Current Signal Results

In this section, the performance of the trained 1D CNNs and Self-ONNs over 200 BP epochs of the 'current' signal to diagnose the gear fault severity is detailed. The overall test confusion matrices resulted from the 1D CNN as well as the Self-ONN (q=3 and q=5) are summarized in Table 8.

Table 8. Confusion matrix of the current signal over 1D CNN and Self-ONN per class.

Ground Truth	Predicted Classes				
	0% CLP (healthy)	10% CLP	20% CLP	30% CLP	50% CLP (severe)
1D CNN					
0% (healthy)	600	0	0	0	0
10% CLP	0	578	22	0	0
20% CLP	0	71	529	0	0
30% CLP	27	1	5	567	0
50% CLP (severe)	0	0	0	0	600
Self-ONN (q=3)					
0% (healthy)	598	2	0	0	0
10% CLP	0	545	55	0	0
20% CLP	0	82	518	0	0
30% CLP	1	0	9	584	6
50% CLP (severe)	0	0	0	0	600
Self-ONN (q=5)					
0% (healthy)	600	0	0	0	0
10% CLP	0	560	40	0	0
20% CLP	0	76	524	0	0
30% CLP	12	6	18	556	8
50% CLP (severe)	0	0	0	0	600

To observe the network performance in each fold, the training, validation, and testing accuracies and losses are plotted per epoch as in Figure 11 and in Figure 12, respectively, for the system over 1D CNN and Self-ONN.

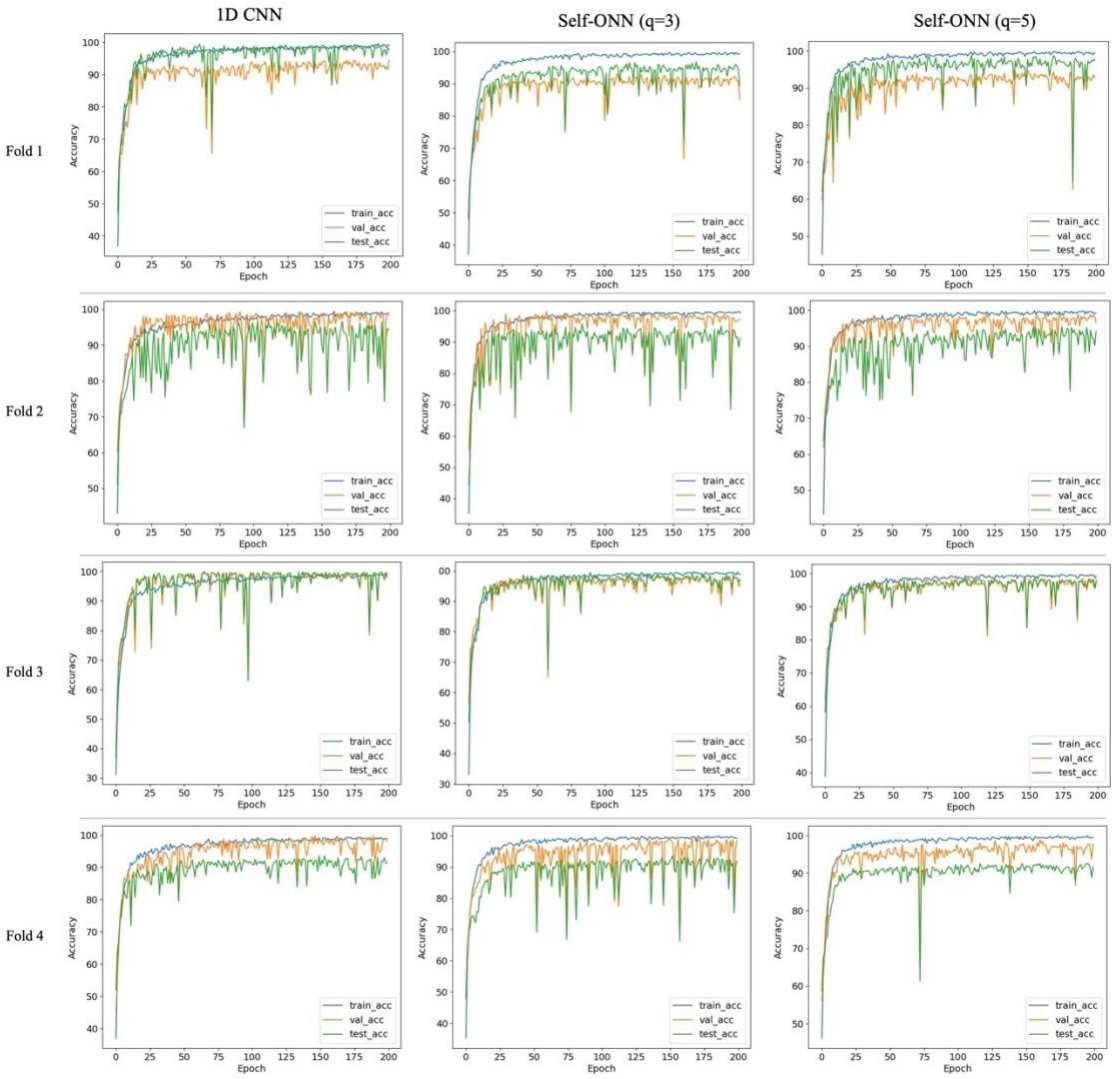

Figure 11. Plots of the training, validation, and testing accuracy per epoch for the current signal over 1D CNN and Self-ONN per fold.

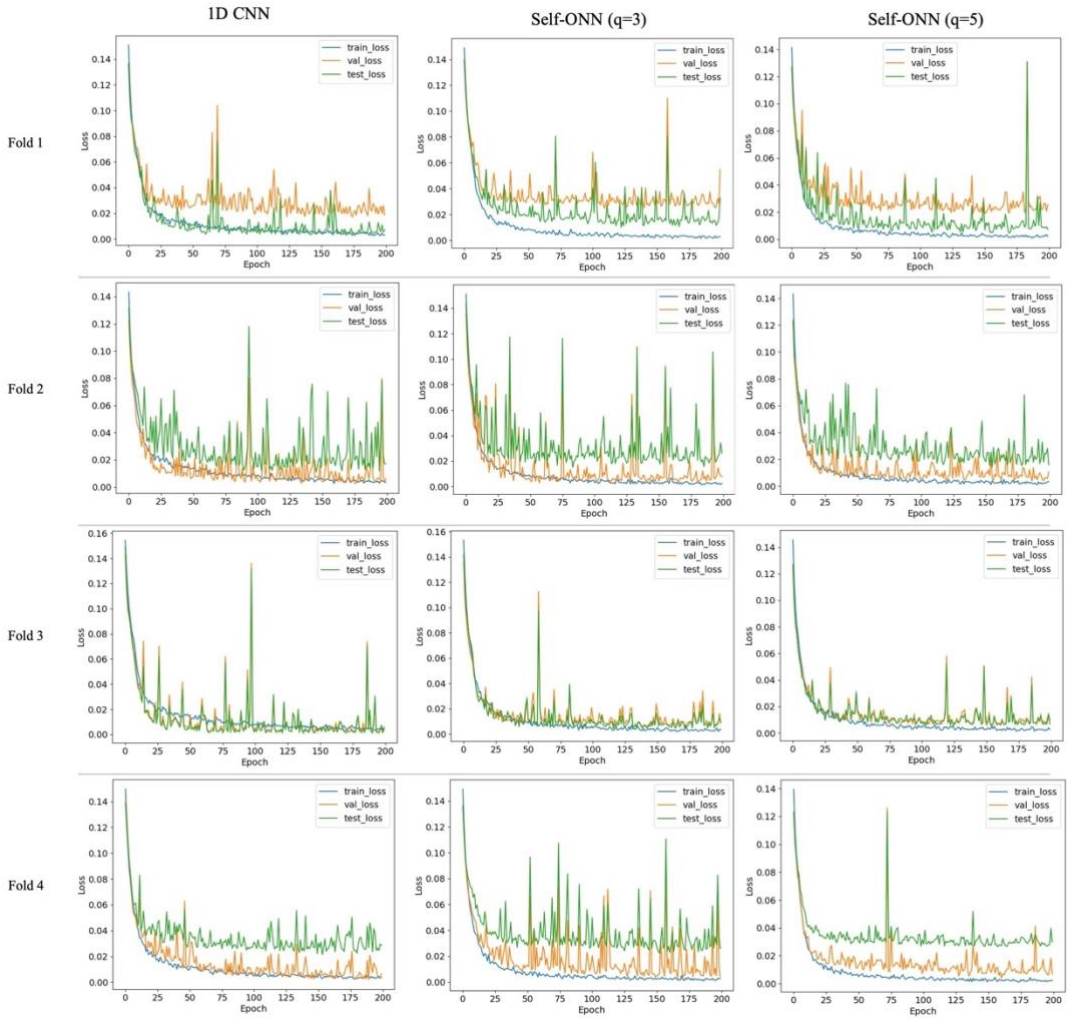

Figure 12. Plots of the training, validation, and testing loss per epoch for the current signal over 1D CNN and Self-ONN per fold.

Moreover, the resulted evaluation metrics after testing per class are summarized in Table 9.

Table 9. Per-class classification evaluation metrics for the current signal over 1D CNN and Self-ONN.

	Accuracy	Precision	Sensitivity	F1-score	Specificity
1D CNN					
0% (healthy)	99.1	95.69	100	97.8	98.87
10% CLP	96.87	88.92	96.33	92.48	97
20% CLP	96.73	95.14	88.17	91.52	98.87
30% CLP	98.9	100	94.5	97.17	100
50% CLP (severe)	100	100	100	100	100
Overall Average	95.8	95.95	95.8	95.79	98.95
Self-ONN (q=3)					
0% (healthy)	99.9	99.83	99.67	99.75	99.96
10% CLP	95.37	86.65	90.83	88.69	96.5
20% CLP	95.13	89	86.33	87.64	97.33
30% CLP	99.47	100	97.33	98.65	100
50% CLP (severe)	99.8	99.01	100	99.5	99.75
Overall Average	94.83	94.9	94.83	94.85	98.71
Self-ONN (q=5)					
0% (healthy)	99.6	98.04	100	99.01	99.5
10% CLP	95.93	87.23	93.33	90.18	96.58
20% CLP	95.53	90.03	87.33	88.66	97.58
30% CLP	98.53	100	92.67	96.2	100
50% CLP (severe)	99.73	98.68	100	99.34	99.67
Overall Average	94.67	94.8	94.67	94.68	98.67

4.3.2. Vibration Signal Results

In this section, the performance of the trained 1D CNN and Self-ONN models by 200 BP epochs over the vibration signal to diagnose the gear fault severity is detailed. The overall test confusion matrices resulting from the 1D CNN as well as the Self-ONN (q=3 and q=5) models are summarized in Table 10.

Table 10. Confusion matrix of the vibration signal over 1D CNN and Self-ONN per class.

Ground Truth	Predicted Classes				
	0% CLP (healthy)	10% CLP	20% CLP	30% CLP	50% CLP (severe)
			1D CNN		
0% (healthy)	1317	24	24	89	46
10% CLP	26	1440	18	13	3
20% CLP	7	14	1469	8	2
30% CLP	69	24	28	1331	48
50% CLP (severe)	30	2	5	25	1438
			Self-ONN (q=3)		
0% (healthy)	1398	13	14	34	41
10% CLP	8	1479	9	2	2
20% CLP	5	16	1477	2	0
30% CLP	92	29	11	1312	56
50% CLP (severe)	20	1	2	5	1472
			Self-ONN (q=5)		
0% (healthy)	1373	18	19	62	28
10% CLP	6	1479	9	4	2
20% CLP	3	6	1488	2	1
30% CLP	50	17	11	1405	17
50% CLP (severe)	11	0	1	11	1477

To observe the network performance in each fold, the training, validation, and testing accuracies and losses are plotted per epoch as in Figure 13 and in Figure 14, respectively .

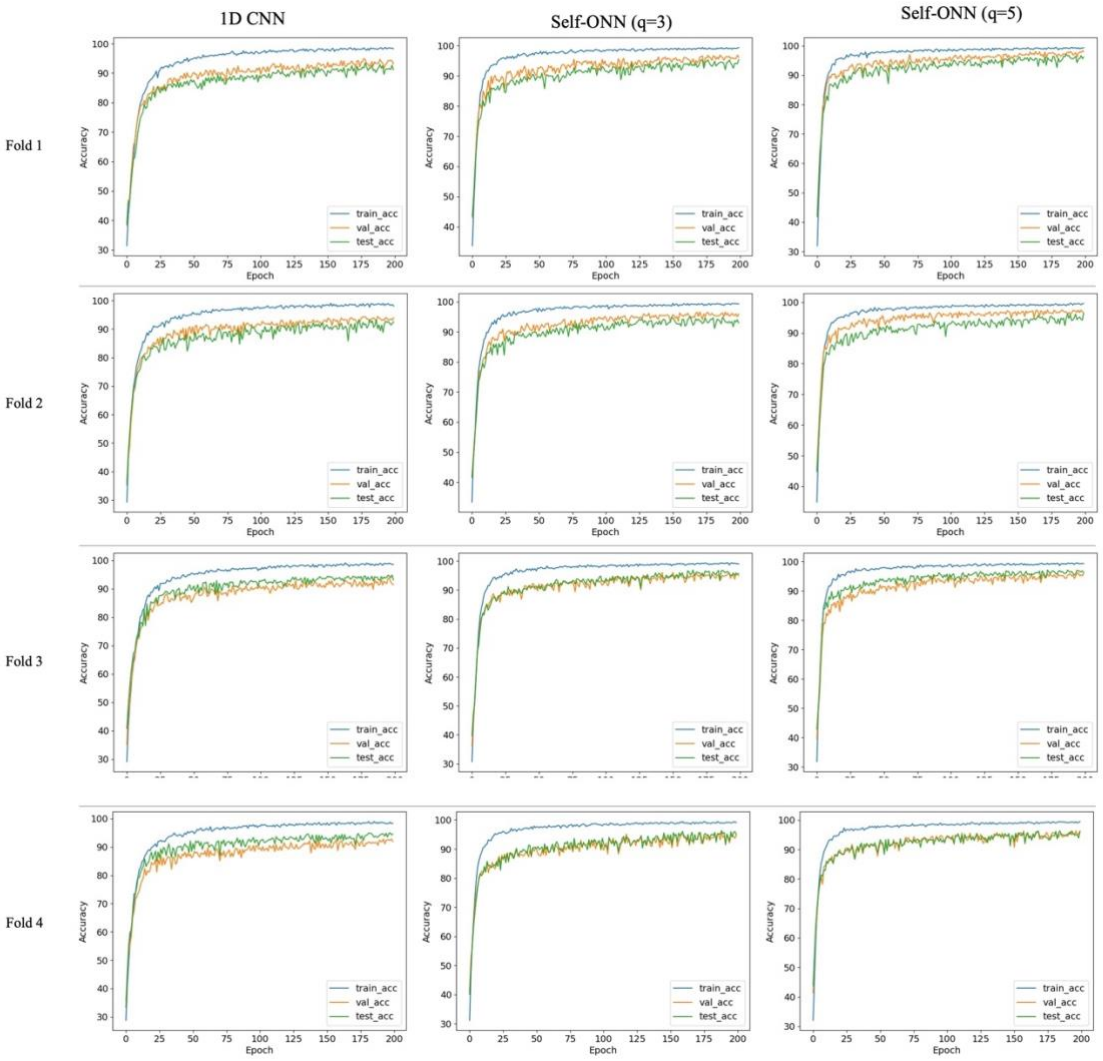

Figure 13. Plots of the training, validation, and testing accuracy per epoch for the vibration signal over 1D CNN and Self-ONN per fold.

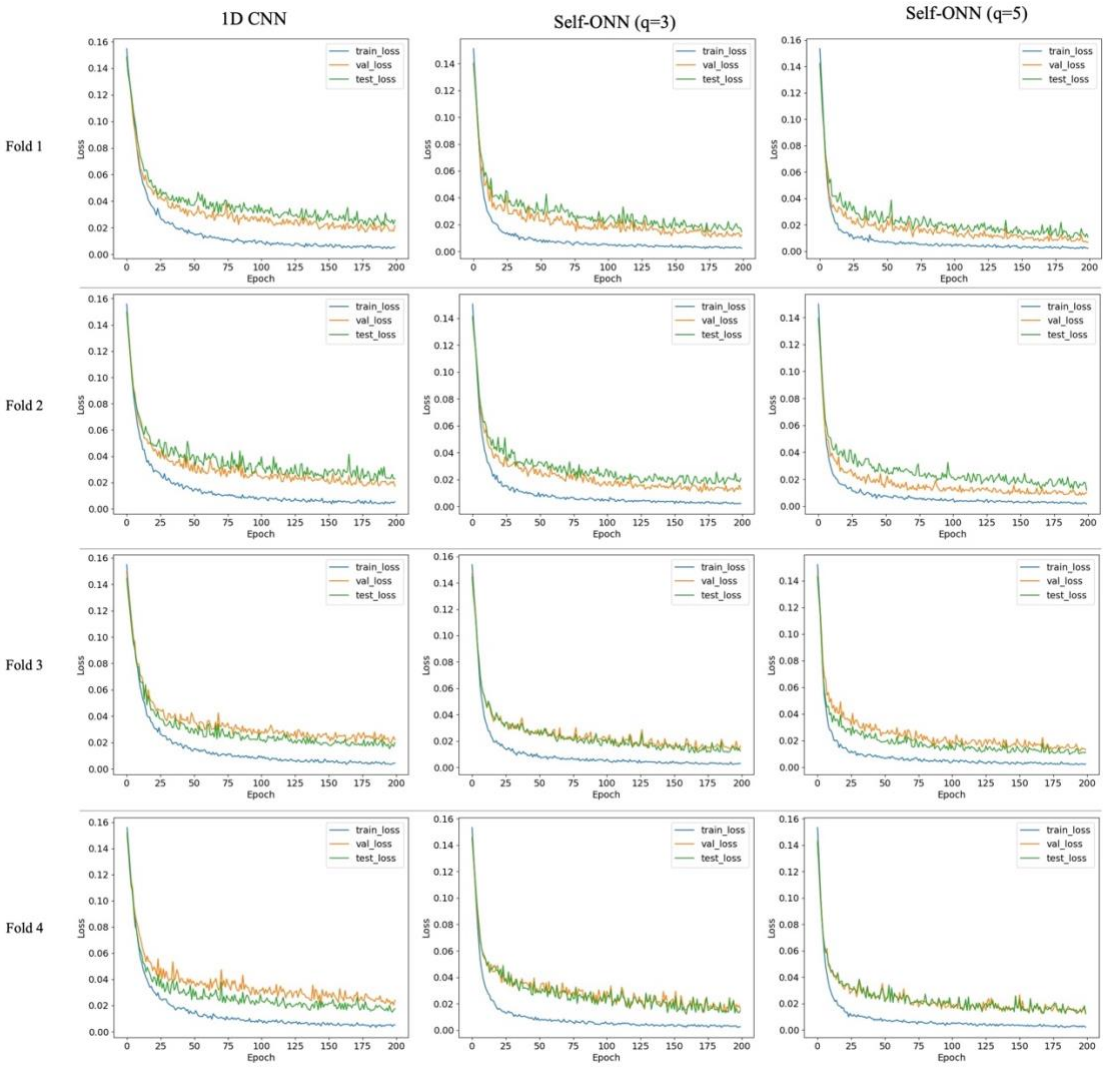

Figure 14. Plots of the training, validation, and testing loss per epoch for the vibration signal over 1D CNN and Self-ONN per fold.

Moreover, the resulted evaluation metrics after testing per class are summarized in Table 11.

Table 11. Per-class classification evaluation metrics for the vibration signal over 1D CNN and Self-ONN.

	Accuracy	Precision	Sensitivity	F1-score	Specificity
1D CNN					
0% (healthy)	95.8	90.89	87.8	89.32	97.8
10% CLP	98.35	95.74	96	95.87	98.93
20% CLP	98.59	95.14	97.93	96.51	98.75
30% CLP	95.95	90.79	88.73	89.75	97.75
50% CLP (severe)	97.85	93.56	95.87	94.7	98.35
Overall Average	93.27	93.22	93.27	93.23	98.32
Self-ONN (q=3)					
0% (healthy)	96.97	91.79	93.2	92.49	97.92
10% CLP	98.93	96.16	98.6	97.36	99.02
20% CLP	99.21	97.62	98.47	98.04	99.4
30% CLP	96.92	96.83	87.47	91.91	99.28
50% CLP (severe)	98.31	93.7	98.13	95.86	98.35
Overall Average	95.17	95.22	95.17	95.13	98.79
Self-ONN (q=5)					
0% (healthy)	97.37	95.15	91.53	93.3	98.83
10% CLP	99.17	97.3	98.6	97.95	99.32
20% CLP	99.31	97.38	99.2	98.28	99.33
30% CLP	97.68	94.68	93.67	94.17	98.68
50% CLP (severe)	99.05	96.85	98.47	97.65	99.2
Overall Average	96.29	96.27	96.29	96.27	99.07

4.3.3. Acoustic Signal Results

In this section, the performance of the trained 1D CNNs and Self-ONNs by 500 BP epochs over the acoustic signal to diagnose the gear fault severity is detailed. Note that, based on the observation of the raw acoustic signal patterns, the number of epochs has been increased from 200 to 500 for this signal type in order to observe the network's performance over a longer time without altering its complexity. The overall test

confusion matrices resulting from the 1D CNN as well as the Self-ONN (q=3 and q=5) models are summarized in Table 12.

Table 12. Confusion matrix of the acoustic signal over 1D CNN and Self-ONN per class.

Ground Truth	Predicted Classes				
	0% CLP (healthy)	10% CLP	20% CLP	30% CLP	50% CLP (severe)
			1D CNN		
0% (healthy)	1486	797	258	266	193
10% CLP	658	1764	260	186	132
20% CLP	216	258	2369	30	127
30% CLP	143	115	21	2540	181
50% CLP (severe)	137	91	73	204	2495
			Self-ONN (q=3)		
0% (healthy)	1726	603	176	295	200
10% CLP	658	1865	219	139	119
20% CLP	142	199	2563	16	80
30% CLP	78	73	12	2670	167
50% CLP (severe)	158	57	34	135	2616
			Self-ONN (q=5)		
0% (healthy)	1870	645	154	140	191
10% CLP	631	1972	196	82	119
20% CLP	145	186	2570	9	90
30% CLP	176	96	16	2591	121
50% CLP (severe)	164	69	26	90	2651

To observe the network's performance in each fold, the training, validation, and testing accuracies and losses are plotted per epoch as in Figure 15 and in Figure 16, respectively.

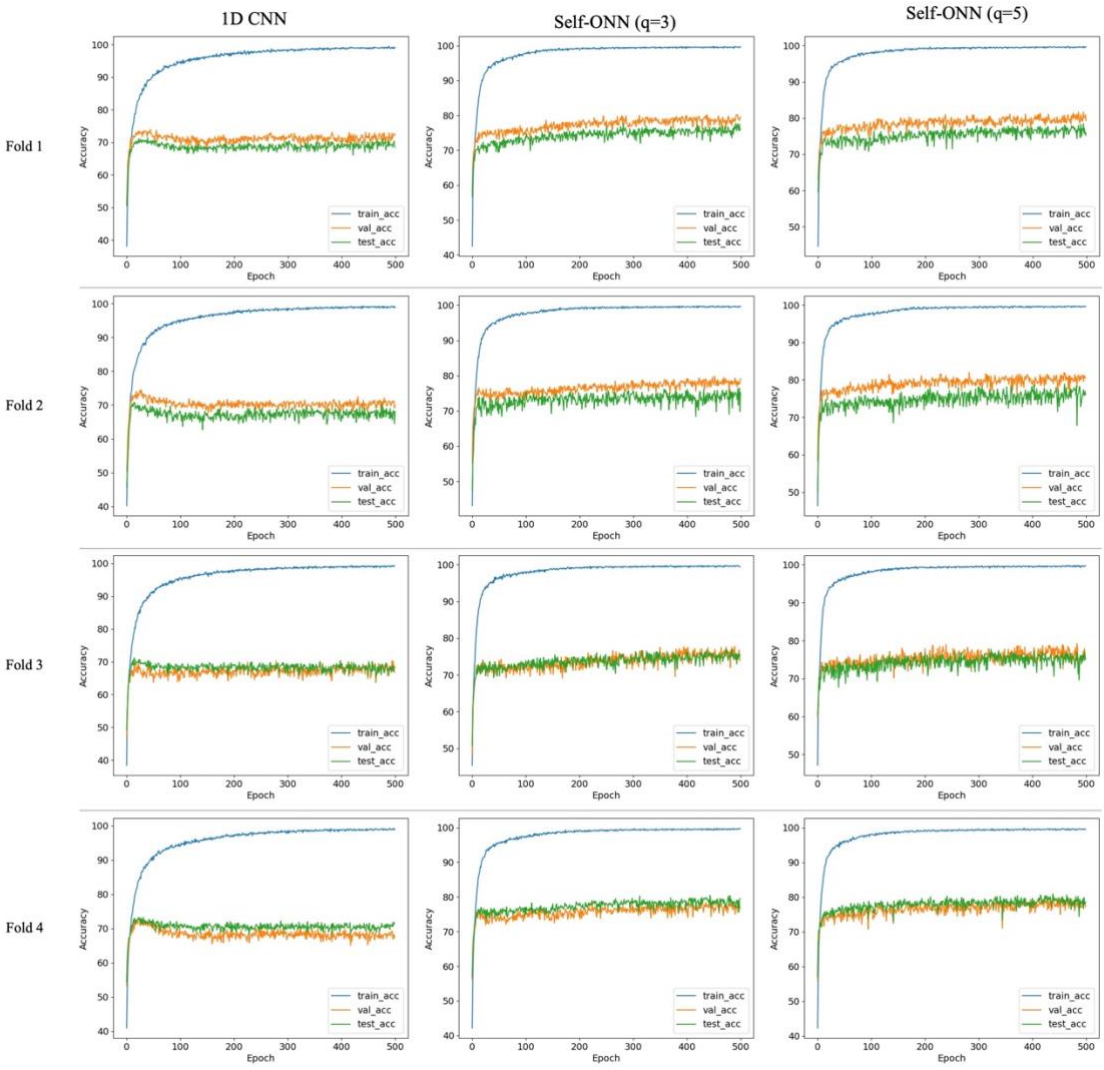

Figure 15. Plots of the training, validation, and testing accuracy per epoch for the acoustic signal over 1D CNN and Self-ONN per fold.

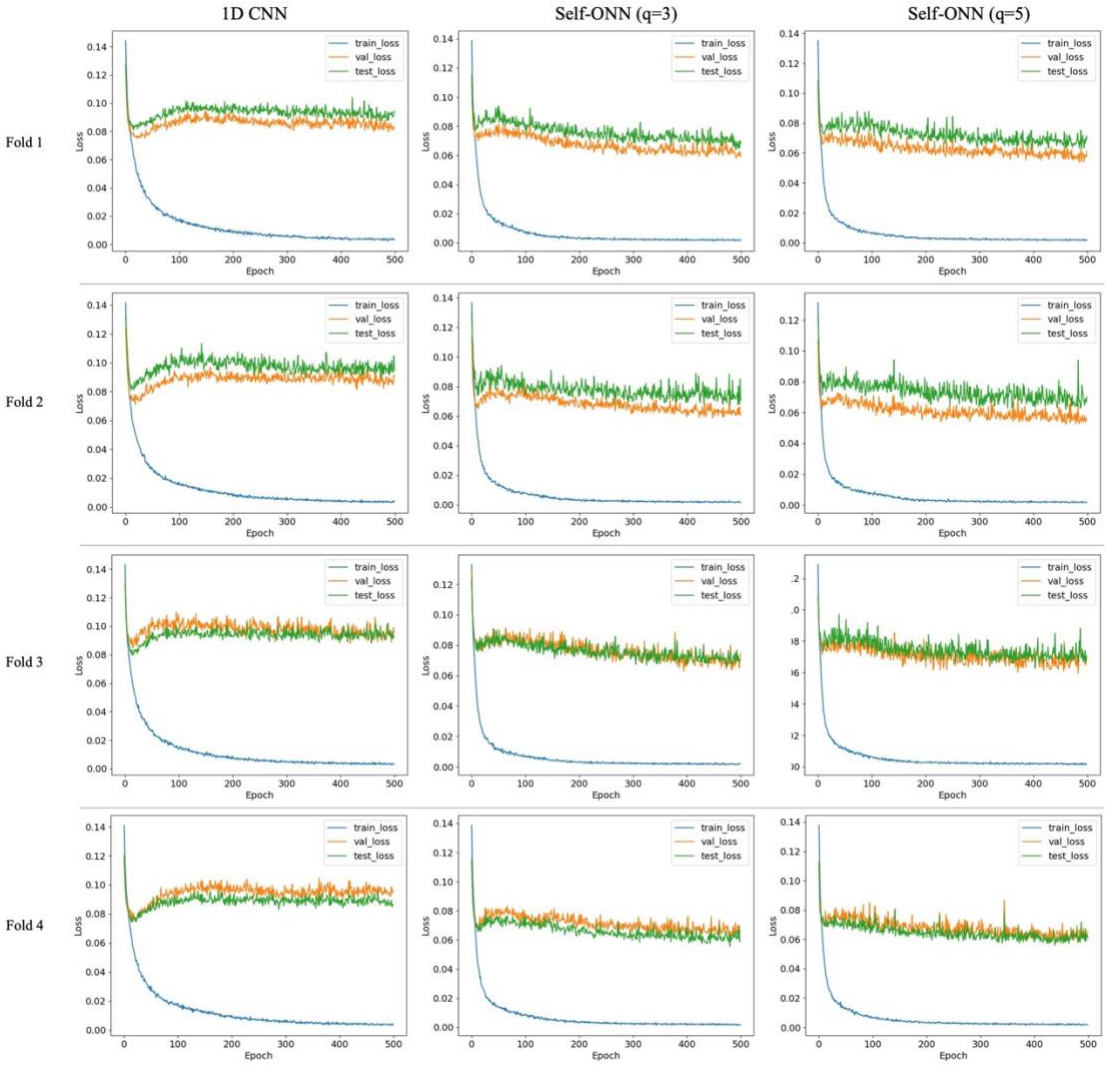

Figure 16. Plots of the training, validation, and testing loss per epoch for the acoustic signal over 1D CNN and Self-ONN per fold.

Moreover, the resulted evaluation metrics after testing per class are summarized in Table 13.

Table 13. Per-class classification evaluation metrics for the acoustic signal over 1D CNN and Self-ONN.

	Accuracy	Precision	Sensitivity	F1-score	Specificity
1D CNN					
0% (healthy)	82.21	56.29	49.53	52.69	90.38
10% CLP	83.35	58.31	58.8	58.55	89.49
20% CLP	91.71	79.47	78.97	79.22	94.9
30% CLP	92.36	78.74	84.67	81.6	94.28
50% CLP (severe)	92.41	79.76	83.17	81.43	94.72
Overall Average	71.03	70.51	71.03	70.7	92.75
Self-ONN (q=3)					
0% (healthy)	84.6	62.49	57.53	59.91	91.37
10% CLP	86.22	66.68	62.17	64.35	92.23
20% CLP	94.15	85.32	85.43	85.37	96.32
30% CLP	93.9	82.03	89	85.37	95.12
50% CLP (severe)	93.67	82.21	87.2	84.63	95.28
Overall Average	76.27	75.75	76.27	75.93	94.06
Self-ONN (q=5)					
0% (healthy)	85.03	62.63	62.33	62.48	90.7
10% CLP	86.51	66.44	65.73	66.08	91.7
20% CLP	94.52	86.77	85.67	86.22	96.73
30% CLP	95.13	88.98	86.37	87.66	97.32
50% CLP (severe)	94.2	83.58	88.37	85.91	95.66
Overall Average	77.69	77.68	77.69	77.67	94.42

4.4 Discussion

The accuracy plots indicate that, with the use of the validation set, the network model with the highest generalization capability was selected for all signal types. Different configurations of the network architecture and hyperparameters were evaluated to achieve the best performance. A slight ‘overfitting’ is apparent for the classification over the acoustic signal; however, it did not affect the overall

performance.

4.4.1. 1D CNN Results

It is observed that the ‘current’ signal type is the best choice among other signal types in terms of diagnosis performance with an overall accuracy of around 96%, a 96% sensitivity, whilst having a precision, F1-score, and specificity above 95%.

So far, what has been covered was only the frame-based classification results, which can give an idea only about the overall classification performance. In practice, further analysis is needed to convert this frame-based classification into a real diagnosis and detection scenario, for which the highest speed and maximum robustness are required. For this purpose, we start by analyzing the speed of detection. For the detection of the gear faults over the ‘current’ signal, 96% frame-based sensitivity is achieved. Thus, that the probability of missing the first faulty frame (regardless of the fault severity) is only 4%. Therefore, the probability of missing the two consecutive frames is around 0.0016, which is negligible. Considering the frame duration for the ‘current’ signal as 200 ms, this means that compact 1D CNNs can detect and diagnose the severity of the gear faults in less than 400 ms. Although it can be inspected that, for the ‘current’ signal type, the network yields the highest confusion when predicting 20% CLP as 10% CLP during classification, there are either no or minor confusions between the other true and predicted classes.

Following the current signal, the second most reliable signal type to consider for gear cracking fault diagnosis is the vibration signal, whereas the least reliable signal type is the acoustic. Similarly, the network models with the highest validation accuracies are selected in each fold. The vibration signal has an overall sensitivity close to 93%. Once again, this is a frame-based sensitivity measure, and for the detection of the gear faults, around a 93% frame-based sensitivity is achieved, which implies that

the probability of missing the first faulty frame (regardless of the fault severity) is 7%. Therefore, the probability of missing the three consecutive frames is around 0.00034, which is negligible. Considering the frame duration for the vibration signal as 80 ms, this means that compact 1D CNNs can detect the gear faults in less than 240 ms. This is a faster detection and early diagnosis capability than the 'current' signal. Although it can be inspected that the network yields the highest confusion when predicting the 0% CLP as 30% CLP during classification, there are either half or much less confusions between the other true and predicted classes for the vibration signal. It is also noticed that the confusions significantly decrease over the Self-ONN model architectures.

Finally, as for the acoustic signal type, the networks at the highest validation accuracies are also selected in each fold to ensure the best generalization capability. Around 71% frame-based sensitivity is achieved. Thus, the probability of missing the first faulty frame (regardless of the fault severity) is 29%, which is the highest among all signal types. Therefore, the probability of missing the six consecutive frames is around 0.0006 which is negligible. However, considering the frame duration for the acoustic signal as 40 ms, this also means that compact 1D CNNs can detect the gear faults using the acoustic signal in less than 240 ms. This is also a faster detection speed than the 'current' signal.

4.4.2. Self-ONN Results

It is observed that, for the 'current' signal type, the Self-ONNs with $q = 3$ and $q = 5$ get a slightly lower performance level than 1D CNNs with a F1-score difference of 1%. Since the detection problem over the 'current' signal type is the easiest due to the periodicity of the signal pattern, the detector does not need a complex/deep network architecture, and hence, yielding a comparable performance level with the Self-ONNs.

Conversely, when dealing with a much more challenging data learning problem

such as over the vibration signal type, a significant improvement on the classification performance is observed by Self-ONNs yielding around a 3% F1-score, sensitivity, and accuracy improvements. The performance on the vibration signal has increased gradually from the 1D CNN case to the Self-ONN with order $q = 3$ to $q = 5$.

Furthermore, the performance gap further widens at the most challenging problem of all three signal types: the acoustic signal. A performance gap of around 7% in the F1-score, sensitivity, and accuracy is observed by the Self-ONNs. Once again, the performance on the acoustic signal has significantly increased from the 1D CNN case to the Self-ONN with order $q = 3$ to $q = 5$.

Moreover, it is observed that the performance generally improves when increasing the Q^{th}order from $q = 3$ to $q = 5$, as a Self-ONN with a $q = 5$ provides a better approximation of the composite nodal operator function during training, as in Eq. (13), in contrast with that of a Self-ONN with a $q = 3$.

Evidently, running Self-ONNs over the two challenging signal types unlocks their potential for gear fault detection and diagnosis, whereas deeper and more complex CNN models are required to accomplish the same performance level.

The diagnosis analysis of the proposed system is presented in Table 14 in terms of the accuracy and sensitivity of the first-time detection of an abnormal frame regardless of the fault severity level. It can be noticed that the system outperforms using the ‘current’ signal measurement for faulty gears in terms of both accuracy and sensitivity with values higher than 99%, followed by the vibration signal type. The system’s diagnosis performance degrades further using the acoustic signal type. It can also be noticed that the detection of the first faulty frame for each signal type improves when running the system on the Self-ONN model. Similarly, there is a slight improvement over the ‘current’ signal when shifting from 1D CNNs to Self-ONNs,

however, more significant improvements are noticed when observing cases of more challenging data learning, as in the vibration and the acoustic signal types, with approximately a 1.6% and 3.7% accuracy and sensitivity increase, respectively, for the vibration signal, and a 2.8% and 12.8% accuracy and sensitivity increase, respectively, for the acoustic signal.

Table 14. Detection and diagnosis system performance on the detection of the first faulty frame for all signal types.

Signal Type	Model	Accuracy (%)	Sensitivity (%)
Current signal	1D CNN	99.1	99.9
	Self-ONN (q=3)	99.9	99.67
	Self-ONN (q=5)	99.6	99.9
Vibration signal	1D CNN	95.8	87.8
	Self-ONN (q=3)	96.97	93.2
	Self-ONN (q=5)	97.37	91.53
Acoustic signal	1D CNN	82.21	49.53
	Self-ONN (q=3)	84.6	57.53
	Self-ONN (q=5)	85.03	62.33

Moreover, the classification models of the different signal types are compared against each other in terms of computational inference time and the number of trainable parameters. Using a constant number of trainable parameters in all models (33,525), the computational speed performance of the model applied on all signal types does not exceed 1.5 ms for one frame. This inference time, therefore, is considered negligible for the frame duration (200, 80, or 40 ms). Thus, the computational inference time for a frame is around 0.7, 1.8, and 3.4% of the overall system speed for the current, vibration, and acoustic signal types, respectively.

Limitations of the proposed work include testing the network on other cases of gear cracking anomaly levels (i.e., different CLP values), and on other rotor rig systems

to comply with real-world applications as in industrial machinery. These will be the topics for future research.

CHAPTER 5: CONCLUSION

In this study, ML-based techniques for facility condition monitoring are investigated and presented in a comparative analysis. Unlike the traditional fully-connected neural networks, CNNs significantly decrease network parameters through local connectivity and exchanging weights using convolutional layers, which comprises a group of kernels with a limited receptive area. Besides, training deep 2D CNNs requires extensive training in order to obtain an appropriate generalization capacity. This typically involves large-scale datasets, which in turn raises the computational difficulty considerably. Therefore, compact 1D CNNs have been designed to overcome these disadvantages while achieving higher accuracies and computational efficiencies on 1D signals. Nevertheless, the novel Self-ONNs architecture model has been recently proposed as a superset of CNNs, where nodal operators and pooling are generalized to decrease complexity and network depth required by CNNs in many challenging problems.

In this work, the design and implementation of 1D CNNs as well as Self-ONNs for gear fault detection are investigated over three types of signals: current, vibration, and acoustic. Unlike prior works in the literature, this study is not only limited to gear fault detection, but also extends to the diagnosis of gear fault severity.

The performance of the system is continuously evaluated in terms of standard performance metrics for validation on real data recorded from a test rig system at QU. It became evident from the results that, when running the system over 1D CNNs, the ‘current’ signal has been found the most reliable signal type for fault diagnosis, following by vibration and the acoustic signals next in line. Furthermore, the system’s performance over the Self-ONNs was evaluated yielding significant improvements in the vibration and acoustic signals, which were considered as challenging cases in terms

of data learning and classification. Therefore, Self-ONNs unlock the true potential of challenging signal types for gear fault detection and diagnosis.

Moreover, as the diagnosis is based on the frame-based classification accuracy measure, whereas the fault detection is an instantaneous action, the time delay for the detection of the first occurrence of a gear fault was considered. For this purpose, the probability of missing a certain number of faulty frames was computed, and the time delay equivalent to the number of faulty frames with a negligible probability was calculated. The results show that the acoustic signal yields the least time for fault detection due to the shortest frame duration. The computational inference time was also evaluated to be minimal with respect to the overall system's response time considering each of the signal's frame length.

Furthermore, to fulfill a complete investigation on the system's diagnosing performance, the system's performance for detecting the first faulty frames has been evaluated in terms of accuracy and sensitivity. It was observed that, once again, the 'current' signal has the highest sensitivity metric approaching closely to 99%, whereas the vibration and acoustic signals were beyond compare approaching to around 88 and 50%, respectively, in the 1D CNN case. However, when observing the system's performance over Self-ONNs for detecting the first faulty frames, there was a slight improvement on the current signal, and significant improvements on more challenging signal types, approaching a sensitivity of 92 and 62% for the vibration and acoustic signals, respectively. This is analogous to the system's classification performance.

Future work includes targeting an actual health monitoring system, that is, employing a trained system that is used for gear health monitoring on another system and robust to the variations on machines/gears, time and sensor locations, and other environmental parameters.

REFERENCES

[1] S. R. Saufi, Z. A. B. Ahmad, M. S. Leong, and M. H. Lim, “Challenges and Opportunities of Deep Learning Models for Machinery Fault Detection and Diagnosis: A Review,” IEEE Access, vol. 7, pp. 122644–122662, 2019, doi: 10.1109/ACCESS.2019.2938227.

[2] S. Basangar and B. N. Tripathi, “Literature Review on Fault Detection of Equipment using Machine Learning Techniques,” in 2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM), Jan. 2020, pp. 62–67, doi: 10.1109/ICCAKM46823.2020.9051543.

[3] “Demands on Sensors for Future Servicing: Smart Sensors for Condition Monitoring | Analog Devices.” https://www.analog.com/en/technical-articles/a60151-demands-on-sensors-for-future-servicing-smart-sensors-for-condition-monitoring.html# (accessed Mar. 11, 2021).

[4] M. Hamadache, J. H. Jung, J. Park, and B. D. Youn, “A comprehensive review of artificial intelligence-based approaches for rolling element bearing PHM: shallow and deep learning,” JMST Adv., vol. 1, no. 1, pp. 125–151, Jun. 2019, doi: 10.1007/s42791-019-0016-y.

[5] X. Chen, F. Kopsaftopoulos, Q. Wu, H. Ren, and F.-K. Chang, “A Self-Adaptive 1D Convolutional Neural Network for Flight-State Identification,” Sensors, vol. 19, no. 2, Art. no. 2, Jan. 2019, doi: 10.3390/s19020275.

[6] O. Abdeljaber, S. Sassi, O. Avci, S. Kiranyaz, A. A. Ibrahim, and M. Gabbouj, “Fault Detection and Severity Identification of Ball Bearings by Online Condition Monitoring,” IEEE Trans. Ind. Electron., vol. 66, no. 10, pp. 8136–8147, Oct. 2019, doi: 10.1109/TIE.2018.2886789.

[7] O. Abdeljaber, O. Avci, S. Kiranyaz, M. Gabbouj, and D. J. Inman, “Real-time

vibration-based structural damage detection using one-dimensional convolutional neural networks," Journal of Sound and Vibration, vol. 388, pp. 154–170, Feb. 2017, doi: 10.1016/j.jsv.2016.10.043.

[8] S. Kiranyaz, T. Ince, A. Iosifidis, and M. Gabbouj, "Operational neural networks," Neural Comput & Applic, vol. 32, no. 11, pp. 6645–6668, Jun. 2020, doi: 10.1007/s00521-020-04780-3.

[9] C. Sobie, C. Freitas, and M. Nicolai, "Simulation-driven machine learning: Bearing fault classification," Mech. Syst. Signal Process., vol. 99, pp. 403–419, Jan. 2018, doi: 10.1016/j.ymssp.2017.06.025.

[10] W. Mao, W. Feng, and X. Liang, "A novel deep output kernel learning method for bearing fault structural diagnosis," Mech. Syst. Signal Process., vol. 117, pp. 293–318, Feb. 2019, doi: 10.1016/j.ymssp.2018.07.034.

[11] P. K. Kankar, S. C. Sharma, and S. P. Harsha, "Rolling element bearing fault diagnosis using wavelet transform," Neurocomputing, vol. 74, no. 10, pp. 1638–1645, May 2011, doi: 10.1016/j.neucom.2011.01.021.

[12] S. Esakimuthu Pandarakone, Y. Mizuno, and H. Nakamura, "A Comparative Study between Machine Learning Algorithm and Artificial Intelligence Neural Network in Detecting Minor Bearing Fault of Induction Motors," Energies, vol. 12, no. 11, Art. no. 11, Jan. 2019, doi: 10.3390/en12112105.

[13] F. Piltan, A. E. Prosvirin, I. Jeong, K. Im, and J.-M. Kim, "Rolling-Element Bearing Fault Diagnosis Using Advanced Machine Learning-Based Observer," Appl. Sci., vol. 9, no. 24, Art. no. 24, Jan. 2019, doi: 10.3390/app9245404.

[14] R. N. Toma, A. E. Prosvirin, and J.-M. Kim, "Bearing Fault Diagnosis of Induction Motors Using a Genetic Algorithm and Machine Learning Classifiers," Sensors, vol. 20, no. 7, Art. no. 7, Jan. 2020, doi: 10.3390/s20071884.

[15] Z. Lei et al., “An Intelligent Fault Diagnosis Method Based on Domain Adaptation and Its Application for Bearings Under Polytropic Working Conditions,” IEEE Trans. Instrum. Meas., vol. 70, pp. 1–14, 2021, doi: 10.1109/TIM.2020.3041105.

[16] Z. Huang et al., “A multi-source dense adaptation adversarial network for fault diagnosis of machinery,” IEEE Trans. Ind. Electron., pp. 1–1, 2021, doi: 10.1109/TIE.2021.3086707.

[17] L. C. Brito, G. A. Susto, J. N. Brito, and M. A. V. Duarte, “Fault Detection of Bearing: An Unsupervised Machine Learning Approach Exploiting Feature Extraction and Dimensionality Reduction,” Informatics, vol. 8, no. 4, p. 85, Dec. 2021, doi: 10.3390/informatics8040085.

[18] W.-B. Zoungrana, A. Chehri, and A. Zimmermann, “Automatic Classification of Rotating Machinery Defects Using Machine Learning (ML) Algorithms,” in *Human Centred Intelligent Systems*, Singapore, 2021, pp. 193–203. doi: 10.1007/978-981-15-5784-2_16.

[19] Z. Qiao, Y. Lei, and N. Li, “Applications of stochastic resonance to machinery fault detection: A review and tutorial,” Mech. Syst. Signal Process., vol. 122, pp. 502–536, May 2019, doi: 10.1016/j.ymssp.2018.12.032.

[20] P. Cao, S. Zhang, and J. Tang, “Preprocessing-Free Gear Fault Diagnosis Using Small Datasets With Deep Convolutional Neural Network-Based Transfer Learning,” IEEE Access, vol. 6, pp. 26241–26253, 2018, doi: 10.1109/ACCESS.2018.2837621.

[21] F. Li, X. Pang, and Z. Yang, “Motor current signal analysis using deep neural networks for planetary gear fault diagnosis,” Measurement, vol. 145, pp. 45–54, Oct. 2019, doi: 10.1016/j.measurement.2019.05.074.

[22] X. Li, J. Li, D. He, and Y. Qu, “Gear pitting fault diagnosis using raw acoustic

emission signal based on deep learning," Eksploat. Niezawodn., vol. Vol. 21, no. 3, 2019, doi: 10.17531/ein.2019.3.6.

[23] X. Li, J. Li, Y. Qu, and D. He, "Semi-supervised gear fault diagnosis using raw vibration signal based on deep learning," Chin. J. Aeronaut., vol. 33, no. 2, pp. 418–426, Feb. 2020, doi: 10.1016/j.cja.2019.04.018.

[24] G. Cirrincione, R. R. Kumar, A. Mohammadi, S. H. Kia, P. Barbiero, and J. Ferretti, "Shallow Versus Deep Neural Networks in Gear Fault Diagnosis," IEEE Trans. Energy Convers., vol. 35, no. 3, pp. 1338–1347, Sep. 2020, doi: 10.1109/TEC.2020.2978155.

[25] Z. K. Abdul, A. K. Al-Talabani, and D. O. Ramadan, "A Hybrid Temporal Feature for Gear Fault Diagnosis Using the Long Short Term Memory," IEEE Sens. J., vol. 20, no. 23, pp. 14444–14452, Dec. 2020, doi: 10.1109/JSEN.2020.3007262.

[26] Y. Qu, M. He, J. Deutsch, and D. He, "Detection of Pitting in Gears Using a Deep Sparse Autoencoder," Appl. Sci., vol. 7, no. 5, Art. no. 5, May 2017, doi: 10.3390/app7050515.

[27] X. Wang, D. Mao, and X. Li, "Bearing fault diagnosis based on vibro-acoustic data fusion and 1D-CNN network," Measurement, vol. 173, p. 108518, Mar. 2021, doi: 10.1016/j.measurement.2020.108518.

[28] X. Li, J. Li, C. Zhao, Y. Qu, and D. He, "Gear pitting fault diagnosis with mixed operating conditions based on adaptive 1D separable convolution with residual connection," Mechanical Systems and Signal Processing, vol. 142, p. 106740, Aug. 2020, doi: 10.1016/j.ymssp.2020.106740.

[29] Y. Yao et al., "End-To-End Convolutional Neural Network Model for Gear Fault Diagnosis Based on Sound Signals," Appl. Sci., vol. 8, no. 9, Art. no. 9, Sep. 2018, doi: 10.3390/app8091584.

[30] S. Kiranyaz, J. Malik, H. B. Abdallah, T. Ince, A. Iosifidis, and M. Gabbouj, "Self-organized Operational Neural Networks with Generative Neurons," Neural Networks, vol. 140, pp. 294–308, Aug. 2021, doi: 10.1016/j.neunet.2021.02.028.

[31] O. C. Devecioglu, J. Malik, T. Ince, S. Kiranyaz, E. Atalay, and M. Gabbouj, "Real-Time Glaucoma Detection from Digital Fundus Images using Self-ONNs," IEEE Access, pp. 1–1, 2021, doi: 10.1109/ACCESS.2021.3118102.

[32] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks," IEEE Trans. Biomed. Eng., vol. 63, no. 3, pp. 664–675, Mar. 2016, doi: 10.1109/TBME.2015.2468589.

[33] S. Kiranyaz et al., "Robust Peak Detection for Holter ECGs by Self-Organized Operational Neural Networks," arXiv:2110.02381 [cs, eess], Sep. 2021, Accessed: Oct. 14, 2021. [Online]. Available: http://arxiv.org/abs/2110.02381.

[34] T. Ince et al., "Early Bearing Fault Diagnosis of Rotating Machinery by 1D Self-Organized Operational Neural Networks," IEEE Access, pp. 1–1, 2021, doi: 10.1109/ACCESS.2021.3117603.

[35] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," Mech. Syst. Signal Process., vol. 151, p. 107398, Apr. 2021, doi: 10.1016/j.ymssp.2020.107398.

[36] O. Avci, O. Abdeljaber, S. Kiranyaz and D. Inman, "Structural Damage Detection in Real Time: Implementation of 1D Convolutional Neural Networks for SHM Applications", Book: Structural Health Monitoring & Damage Detection, Volume 7 pp 49-54, March 2017. DOI: 10.1007/978-3-319-54109-9_6

[37] T.Ince, S.Kiranyaz , L.Eren , M.Askar and M.Gabbouj, "Real-Time Motor Fault Detection by 1D Convolutional Neural Networks", IEEE Trans. of Industrial Electronics, vol. 63, Issue 11, pp. 7067 – 7075, Nov. 2016.

DOI: 10.1109/TIE.2016.2582729

[38] S. Kiranyaz, A. Gastli, L. Ben-Brahim, N. Alemadi, and M. Gabbouj, "Real-Time Fault Detection and Identification for MMC using 1D Convolutional Neural Networks", IEEE Transactions on Industrial Electronics, May 2, 2018. DOI: 10.1109/TIE.2018.2833045

[39] L. Eren, T. Ince, and S. Kiranyaz, "A Generic Intelligent Bearing Fault Diagnosis System Using Compact Adaptive 1D CNN Classifier", in Journal of Signal Processing Systems (Springer), pp. 1-11, May 2018. https://doi.org/10.1007/s11265-018-1378-3

[40] "Google Colaboratory." https://colab.research.google.com/notebooks/intro.ipynb (accessed Mar. 06, 2021).