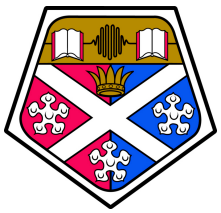
Unsupervised Data-Driven Analysis of Ultrasonic Inspection Data



Panagiotis Zacharis

Department of Electronic and Electrical Engineering

University of Strathclyde

A thesis submitted for the degree of

Doctor of Philosophy

2022

Copyright

Acknowledgements

First and foremost, I would like to express my gratitude to my primary supervisor, Dr Graeme West, for his immense support, invaluable guidance and endless patience throughout all these years.

I would also like to thank Dr Gordon Dobie, Prof Anthony Gachagan, Dr Timothy Lardner, and Dr Christopher Wallace, for helping me make sense of the nuclear and ultrasonic world by sharing their knowledge and expertise with me. My thanks extend to the researchers and academic staff of the Advanced Nuclear Research Centre for all the useful discussions and debugging sessions.

I would like to thank Duncan Hawthorne for generously funding my PhD. My gratitude extends to Bruce Power for supporting this PhD with their expertise and valuable data.

I am deeply grateful to my parents and sister, for their support and understanding through my seemingly never ending education.

Finally, special credit goes to Essi, for her moral support and understanding, without whom I could not have completed this thesis.

Abstract

Ultrasonic inspection is a key part of condition-based maintenance in the nuclear power industry, and it is widely used for flaw detection and characterisation in critical components. The analysis of ultrasonic inspection datasets is a time-consuming and complex task commonly carried by expert analysts. The need for robust and efficient interpretation of inspection data is especially important now that advancements in ultrasonic hardware enable the capture of high-resolution data in far greater rates and volumes than in the recent past, driving research efforts to create automated procedures for signal classification and flaw detection. This thesis provides new data-driven approaches for analysing large volumes of ultrasonic inspection data in an unsupervised manner, without requiring individually labelled ultrasonic signals. The first method utilises the DBSCAN clustering algorithm at its core, and along with the proposed subsampling method, two-stage clustering procedure, and automated parameter estimation procedure, it provides efficient flaw detection without a pre-defined state of normality. The analysis is then extended to large-scale ultrasonic inspection datasets that offer challenges both in terms of size, but also heterogeneity across the different neighbourhoods of the inspected surface. The proposed method utilises a neighbourhood-based transformation of the signals' variability measures as a means of creating a more homogenous feature space that allows for distance metric that is relevant across the surface, offering significant computational benefits. The application of the proposed methods is focused on pressure tubes, a critical reactor asset of CANDU reactors, and is tested across a large number of real-world datasets showing satisfactory detection rates and efficient performance.

Contents

List of Figures

List of Tables

LIST OF TABLES

List of Acronyms

ADAPT	Automated Data Analysis for Pressure Tubes
AGR	Advanced Gas-cooled Reactor
ANDE	Advanced NDE
ANN	Artificial Neural Network
APC	Axial Pitch-Catch Probe
CANDU	CANada Deuterium Uranium
CBM	Condition-Based Maintenance
CIGAR	Channel Inspection Gauging Apparatus for Reactors
CM	Condition Monitoring
CNN	Convolutional Neural Network
CPC	Circumferential Pitch-Catch Probe
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DHC	Delayed Hydride Cracking
DL	Deep Learning
DNN	Deep Neural Network

LIST OF ACRONYMS

DWT	Discrete Wavelet Transform
FT	Fourier Transform
GH	General Helical
IQR	Interquartile Range
KADS	Knowledge Acquisition and Documentation Structuring
LSTM	Long Short-Term Memory
ML	Machine Learning
NB	Normal Incidence Probe
NB10	Normal Incidence Probe at 10 MHz
NB20	Normal Incidence Probe at 20 MHz
NDE	Non-Destructive Evaluation
NPP	Nuclear Power Plant
O&M	Operations and Maintenance.
PHWR	Pressurised Heavy Water Reactor
SIGKDD	Special Interest Group on Knowledge Discovery and Data Mining
SSCs	Systems, Structures, and Components
STFT	Short-Time Fourier Transform
SVM	Support Vector Machine

TBM	Time-Based maintenance

Chapter 1

Introduction

1.1 Motivation for the Work

Ultrasonic non-destructive testing is widely used for flaw detection and characterisation in critical components within industries such as nuclear power, oil & gas, and aerospace, with key applications including examination of reactor pressure vessels [1], offshore oil pipeline inspection [2], and health monitoring of aircraft wings [3].

Ultrasonic inspection is a broad category of methods which enable the examination of materials with the use of high-frequency sound signals. Commonly, the ultrasonic pulses are transmitted into the object of interest and their echoes are captured for further analysis which can reveal potential flaws in the material [4]. However this can be a complex procedure because a received signal is comprised of all the reflections that occurred within the material, resulting in complex signals that carry both important structural information and noise [5, 6].

Besides the extensive research on noise suppression techniques for ultrasonic flaw detection signals [7, 8, 9, 10], traditionally analysing ultrasonic data for flaw detection is a task carried manually by expert analysts due to its complexity. This is a time-consuming approach and studies have shown that although manual ultrasonic inspection can be accurate, it can also produce variability , depending on the inspection skills, training and emotional status or fatigue of inspectors [11, 12].

Furthermore, modern ultrasonic testing equipment is increasingly capable of capturing larger volumes of high-resolution data, which introduces additional challenges, and further accentuates the need for assisting the manual analysis of ultrasonic data with reliable and efficient automated methods.

Several researchers have approached the problem of automated ultrasonic analysis through model-based finite element methods [13, 14, 15, 16, 17, 18] that compare simulated signal responses with the received signals and evaluate the discrepancies. However, modelling the physical and mechanical characteristics of a system requires many assumptions and approximations [4], especially in an environment where the behaviours of surface corrosion and foreign material follow a random behaviour [19].

An alternative approach for analysing ultrasonic inspection datasets is data-driven methods which algorithmically identify patterns within the available data and allow for the health evaluation of the inspected system [20, 21]. This thesis explores data-driven ultrasonic analysis methods, in order to identify limitations and challenges, and to develop methods for the automated assessment of ultra-

sonic datasets. The thesis was inspired by, and uses, existing real-world inspection datasets of core components of CANDU (CANada Deuterium Uranium) reactors called pressure tubes. These are critical components as they contain the nuclear fuel and the coolant. Manufacturing flaws, as well as flaws developed during in-service operation, facilitate Delayed Hydride Cracking (DHC) which can lead to coolant leakage and can potentially damage the reactor [22].

Current inspection processes utilise multiple ultrasonic probes that inspect subsets of pressure tubes during planned statutory plant shutdowns. The datasets are then manually assessed by expert analysts who report the position, size, and type of any detected flaws. This is a laborious work that depends upon expert judgement and a level of subjectivity, due to large volumes of data and complex decision-making required by the diverse surface conditions. As the ultrasonic health assessment lies on the critical path to restarting the reactor, it constitutes a crucial activity with significant financial effects following the principal safety objectives. Providing a means of automatically supporting this assessment in a reliable, repeatable, and rapid manner would be of significant benefit.

Existing expert-system research focuses on the development of a decision support tool for the analysis of ultrasonic inspection data for the CANDU reactors through a knowledge-based approach [23]. Although that work has presented analysis results comparable to human analysts, the significant recent increases in data resolution and volumes, as well as the difficulty of replicating human expert operation in highly complex cases, has generated the need for an alternative parallel procedure to support this analysis process.

The data-driven methods presented in this thesis are tested as a means of supporting the health assessment of pressure tubes. Typically, data-driven approaches to ultrasonic health assessment require individually labelled signals, as they utilise supervised machine learning algorithms. However, the existing pressure tube inspection process does not directly generate individually labelled signals that can be used in a supervised machine learning setting. Although bounding boxes of rectangular areas known to contain flaws are provided, the individual signals are not accompanied by the information of whether their location is within or outside the often irregularly shaped flawed area. Furthermore, available verified flaw reports do not always reflect the precise state of the available data, as they can be influenced by resources unavailable to this research, i.e. older inspection datasets and historical results.

This thesis therefore presents unsupervised methods that explore the structures and patterns within the ultrasonic inspection datasets while maintaining computational times and generalisation ability at levels that enable their applicability to real-world datasets. These methods involve integrated processes focused on transforming and reducing the large amount of available information to levels that facilitate the discovery of potential underlying meaningful groups; i.e. groups of healthy and abnormal areas.

Reducing the information space is important for computational reasons but also for lessening the effects of the 'curse of dimensionality' that appears in high-dimensional spaces [24] and allowing easier interpretation. The techniques proposed in this thesis attempt to accommodate this through feature selection aimed to serve the purpose of each of the analyses, a subsampling/remapping method

for reducing the search space without adversely affecting subsequent analyses, and clustering procedures that group the transformed inspection data by operating without strictly predefined states of normality. This presented subsampling method is specific to the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) clustering algorithm [25], which is utilised to offer an efficient automated clustering approach that locates surface flaws within high resolution ultrasonic inspection datasets captured by different inspection probes.

Furthermore, this thesis considers the problem of efficiently detecting anomalous areas within large-scale inspection datasets where the application of computationally demanding algorithms is challenging. As it is shown, within the scope of the entire population of features of a large-scale inspected area there can be various levels of heterogeneity across different parts of the surface as it is an active environment affected by imperfect geometry and surface conditions. Effectively this means that two healthy areas could appear different, when compared to each other, or, a healthy signal obtained from a noisy area could appear similar to a signal obtained by a noise-free defective area. This could occur as signals of different sections could be affected by different neighbourhood-specific causes, such as poorly focused signals due to probe misalignment caused by sagging [26]. The proposed approach is seeking to create a feature transformation process that enables intuitive inferences that are relevant across the surface. This feature transformation is based on the idea of tracking changes of the feature values locally, instead of investigating the relationship of each feature value within the entire feature space. This can be approached by representing each signal by some distance function of its feature values and its immediate physical neighbours' fea-

ture values only, thus reducing drastically the size of the feature space that each sample has access to and creating a metric that is relevant across the large-scale surface.

1.2 Contributions to Knowledge

- *Subsampling method that reduces the clustering search space of the DBSCAN algorithm.*

  The proposed subsampling method selects a representative input subset which participates in the clustering procedure and re-maps the generated clustering labels back to the original input data. This reduces the search space of the DBSCAN clustering algorithm while ensuring that the results are not adversely affected. The method is effective for large volumes of low-cardinality input data containing multiple repeating unique sets of features values. The method acts as a wrapper for the DBSCAN algorithm, and to the best of the author's knowledge, this direction has not been previously explored.

- *Flaw detection through a two-stage clustering method of ultrasonic inspection data.*

  This method compartmentalises the clustering of the feature space from the location-based clustering offering computational benefits and allows for groups of flawed signals that are spatially disassociated. Following the feature-based clustering, only the location properties of the minority groups

are analysed. This distinction is not based on strictly pre-defined normality conditions, but only on the assumption that the datasets are intrinsically imbalanced, thus the group of healthy samples should be the most populous.

- *DBSCAN parameter estimation method for ultrasonic inspection datasets.*

  This approach provides an estimation of the effect that the DBSCAN parameter of local radius has on the cluster formation datasets captured by different ultrasonic inspection probes, without utilising explicit knowledge of the probes characteristics.

- *Feature transformation method for normalising large-scale heterogeneous ultrasonic inspection data.*

  The transformation method acts across ultrasonic inspection data of large-scale environments which present heterogeneity due to neighbourhood-specific surface conditions and distortion. This generates a more homogenous feature space that allows for a distance metric that is relevant across the inspected surface, offering a computationally effective method for detecting areas of abnormal behaviour.

1.3 List of Publications

- P. Zacharis, G. M. West, G. Dobie, T. Lardner, and A. Gachagan, "Data-driven analysis of ultrasonic pressure tube inspection data.," in *10th International Topical Meeting on Nuclear Plant Instrumentation, Control, and Human-Machine Interface Technologies (NPIC&HMIT)*, (San Francisco, California, United States), June 2017.

- P. Zacharis, G. West, G. Dobie, T. Lardner, and A. Gachagan, "Data-Driven Analysis of Ultrasonic Inspection Data of Pressure Tubes," *Nuclear Technology*, vol. 202, pp. 153–160, June 2018. Publisher: Taylor & Francis eprint: https://doi.org/10.1080/00295450.2017.1421803.

- P. Zacharis, G. West, C. Wallace, G. Dobie, and A. Gachagan, "Automated pressure tube defect analysis.,", Poster Session at *2019 Innovation Showcase: Nuclear Asset Management and Industrial Informatics*, (Toronto, Canada), May 2019.

- C. Wallace, G. West, P. Zacharis, G. Dobie, and A. Gachagan, "Experience, testing and future development of an ultrasonic inspection analysis defect decision support tool for CANDU reactors," in *11th Nuclear Plant Instrumentation, Control and Human-Machine Interface Technologies (NPIC&HMIT)*, (Orlando, Florida, United States), Feb. 2019.

1.4 Structure of Thesis

The remainder of the thesis is organised as follows:

- **Chapter 2** provides the background necessary for the understanding of the inspection datasets and setup that inspired this thesis. The chapter provides an overview of nuclear plant maintenance strategies, and focuses on the pressure tube inspection process, the challenges involved in the assessment of the inspection datasets, and current research towards automating this process. It is argued that the challenges of this process generate a need for automated data-driven inspection analysis processes that can provide support in a rapid and robust manner.

- **Chapter 3** reviews existing research efforts concerned with the automated analysis of ultrasonic data for health assessment purposes. Having recognised that a data-driven approach would be more appropriate for the purposes of this thesis, this chapter focused mainly on supervised and unsupervised data-driven methods and explores their applicability in datasets containing a diverse set of flaws and surface conditions.

- **Chapter 4** proposes a novel unsupervised data-driven approach to automatically analyse collections of ultrasonic signals by individually characterising each signal and producing 2D maps that flag potentially flawed areas.

The analysis consists of multiple steps that consecutively reduce and transform the search space. The chapter introduces a subsampling method that allows for more efficient density clustering, and a two-stage clustering procedure designed to efficiently utilise the DBSCAN clustering algorithm for grouping and noise reduction. Furthermore, a DBSCAN parameter estimation method is presented, which can be applied across datasets captured by different ultrasonic inspection probes without utilising explicit knowledge of the probes characteristics. The proposed procedure is tested by utilising an automated evaluation procedure across a large number of real-world datasets captured from sections of pressure tubes containing flaws, showing satisfactory results.

- **Chapter 5** extends the unsupervised ultrasonic analysis to large-scale ultrasonic inspection datasets which are problematic for current memory intensive algorithms. The chapter presents an anomaly detection method targeting datasets that also exhibit various levels of heterogeneity across different neighbourhoods of the inspected surface. The proposed method utilises a neighbourhood-based transformation of the signals' variability measures as a means of creating a more homogenous feature space that allows for distance metric that is relevant across the surface, offering significant computational benefits.

  Finally, the chapter presents the outcome of the method application to real-world large-scale datasets that shows promising results, as the transformation allows isolation of surface anomalies, including threatening flaws.

- **Chapter 6** concludes the thesis by summarising and discussing the key outcomes of this work and suggesting possible directions for future research.

Chapter 2

Background

2.1 Introduction

Nuclear power is currently the second largest source of low-carbon energy worldwide, after hydroelectricity. As of 2021, there are 443 operational reactors globally [27], and 51 under construction, primarily in China, India, and Korea [28].

A key driver for the nuclear industry is the long-term supply of electricity in a safe, reliable, clean and cost-effective manner. With safety being the highest priority, the situational awareness of the state of nuclear power plant (NPP) systems, structures, and components (SSCs) is crucial, especially considering the unique characteristics of the harsh service conditions that they are exposed to.

Thus, effective maintenance, through the assessment, management, and mitigation of asset degradation is a necessary requirement of both the operators and the regulatory bodies, as it ensures the safe and continuous operation of NPPs while enabling the assessment of NPPs' life extension applications. Condition monitoring of the key components is an essential activity of plant ageing management [29], ensuring safe continued operation and minimising unplanned outages [30].

In the case of Canada Uranium Deuterium (CANDU) reactor, ultrasonic non-destructive examination has been utilised routinely during planned outages for the health assessment of pressure tubes, a crucial component of the reactor as it contains the nuclear fuel and the pressurised coolant.

This chapter provides an overview of nuclear plant maintenance strategies, and focuses on the pressure tube inspection process, the challenges involved in the assessment of the inspection datasets, and current research towards automating this process.

2.2 Maintenance Strategies in the Nuclear Power Industry

The earliest and simplest approach to general machinery maintenance involved corrective or replacement interventions during breakdowns. This maintenance approach (referred to as run-to-failure, or run-to-maintenance) is a source of unplanned downtimes and is suitable only for simple non-critical components of which the consequences, risks, and costs of failure are within acceptable margins. However, the uniquely low risk tolerances of nuclear industry result in performance standards that are essentially the same for both critical and non-critical SSCs. This generates expectations of no failures during all SSCs' operational life and rendering the run-to-failure approach non-viable.

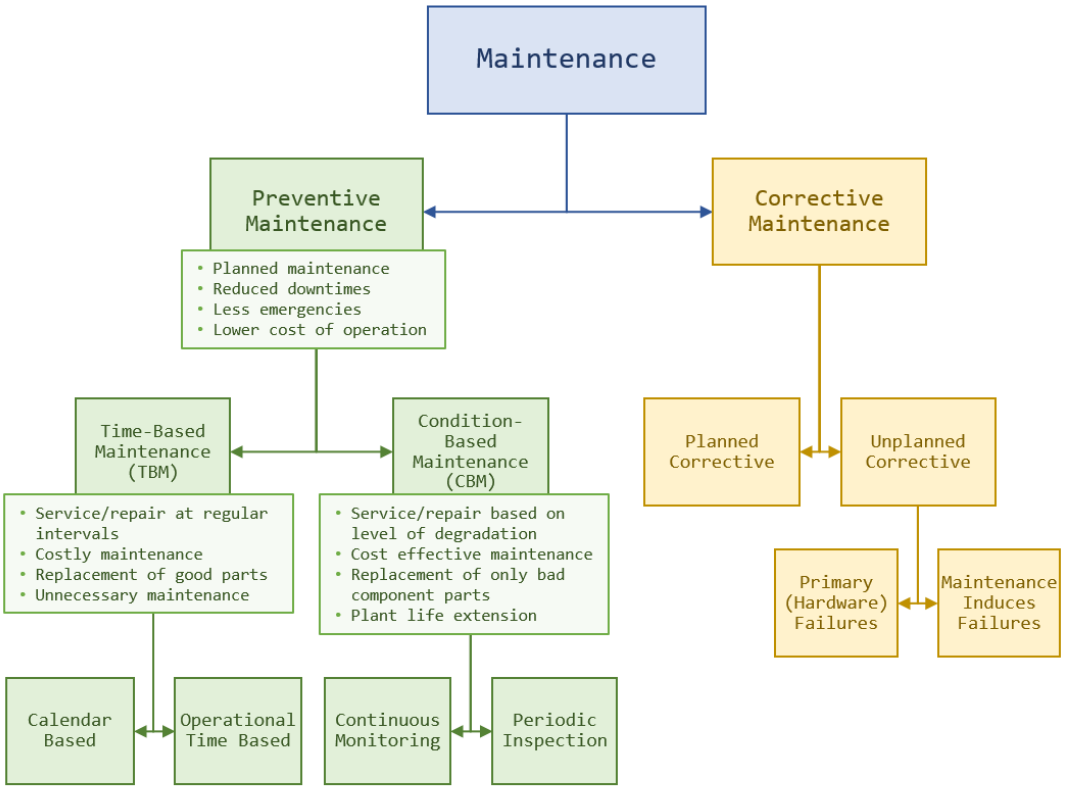


Figure 2.1: Maintenance strategies.

Contemporary maintenance approaches follow mainly proactive preventive strategies that aim to mitigate the shortcomings of corrective approaches. The concept of preventive maintenance was introduced by time-based maintenance (TBM) which involve periodic maintenance based on estimated component failure cycles [31]. These cycles can be either Calendar Based, where the degradation intervals are determined regardless of the operational overhead of the SSC (e.g. lubricant change and replacement of contact sealings), or Operational Time Based, where the operational period is considered to be a significant factor. Typically, the history of failures or incidences of unsatisfactory component behaviour is utilised, and the calculated frequency is determined through the expected mean period between failures.

Compared to corrective maintenance, TBM can reduce unplanned maintenance activities and downtimes, while establishing improved SSC reliability and increased safety-related confidence. However, due to the high uncertainty involved in the estimated failure frequencies [32] and as the majority of failure modes is independent of equipment working age [33], TBM can significantly increase the cost of operation due to unnecessary maintenance activities and replacement of well functioning equipment, while potentially initiating early failures through regular disassembly and reassembly of equipment.

Enabled by the advancement of technology, and driven by the need to reduce the significant maintenance costs that can arise from TBM approaches, the focus across industries has been shifted towards the development of more cost-effective approaches to preventive maintenance, such as condition-based maintenance (CBM) [34, 35] (Figure 2.1).

2.2.1 Condition-Based Maintenance

CBM is a maintenance strategy that determines preventive maintenance actions through the analysis of collected data related to the condition or performance of the SSCs, rather than through scheduled preventive maintenance intervals. This approach allows the operators to significantly reduce the number of unnecessary maintenance operations and enforces better situational awareness of the health of the assets.

The importance of efficient maintenance is especially evident in NPPs, as operations and maintenance (O&M) costs contribute about 60-70% of the cost of generation [30], thus directly affecting the offered price of electricity. At present, CBM is playing a key role in the nuclear industry as well as industries such as aerospace, chemical and military.

A CBM implementation strategy is comprised of acquisition, integration and analysis of diagnostic and performance data to support timely decisions regarding maintenance regimes of critical assets [36]. At its core, CBM is concerned with monitoring and diagnosing the condition of the assets, which can be performed either in a continuous online manner, by utilising real-time surveillance data, or off-line during regularly performed non-destructive inspections.

2.2.2 Condition Monitoring

Condition monitoring (CM) activities aim to enhance the knowledge of the state of a system by measuring and analysing parameters related to that system. CM covers a wide range of activities which can be performed either manually or automatically through the use of intelligent systems. Examples of manual condition monitoring within the nuclear sector include direct visual inspection of AGR (Advanced Gas-cooled Reactor) [37], the extraction of thin slivers from CANDU pressure tubes to determine their deuterium and hydrogen content [38], and manual analysis of ultrasonic inspection data of pressure tubes [39]. Manual inspection can provide direct and detailed information regarding the condition of an asset, but can be a time-consuming and laborious work that depends upon expert judgement and a level of subjectivity. As this process often lies on the critical path of resuming operation [23], it constitutes an activity with significant financial impact.

Intelligent systems can support condition monitoring activities by automating the complex decision-making involved in analysing inspection data in a reliable, repeatable, and rapid manner. Strategies for automated condition monitoring vary depending on the availability of data, the complexity of the acting degradation mechanisms, and the extent and quality of the available knowledge. The following section describes different approaches to creating intelligent systems which include expert systems, model-based, data-driven, and hybrid systems.

2.2.2.1 Expert Systems

Expert systems is a branch of Artificial Intelligence (AI) designed to simulate the reasoning and judgement of qualified specialists within a particular field [40]. Expert systems rely on the assumption that experts generate single pieces of knowledge which can be combined in an appropriate sequence to derive a solution to a problem. By formalising both the domain knowledge and the problem solving strategy they can be represented and incorporated into an expert system that emulates the decision making process of an expert [41]. A differentiating factor between expert systems and conventional programs is their ability to provide an explanation of their behaviour and the clear separation of knowledge and methods which enables expert systems to provide easily amendable dynamic knowledge bases.

A critical step in implementing an expert system, which often is the most difficult [42], is the knowledge acquisition process. This can involve knowledge extraction from sources such as reports and databases, but often the primary source is domain experts [42, 43]. The elicitation, interpretation, and organisation of the information are realised by practitioners, called knowledge engineers, in conjunction, to some level, with a knowledge acquisition module that aims to automate the process [44]. The knowledge engineer interacts with the expert through observation and a series of interviews where expected problems are analysed and discussed. The product of these interviews is formalised protocols and problem

solving mechanisms that can be derived either by the introspection of the expert on his problem solving steps, or by the analysis of the transcribed interview records by the knowledge engineer [42].

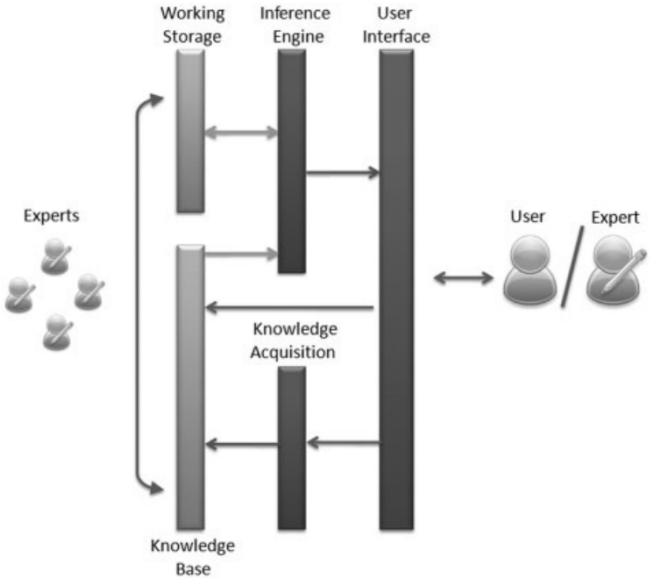


Figure 2.2: Typical expert system components [45].

Typically an expert system (Figure 2.2) consists of the following components:

- *Knowledge base*: The container of the rules resulted from the formalisation of the domain expertise.
- *Working storage*: The container of information related to the specific instance of a problem.
- *Inference engine*: The processing core of an expert system which directs the problem solving decision process that controls the flow of information and acts on the current knowledge base and working storage to derive problem-specific recommendations.

- *Knowledge acquisition module*: An important role of the knowledge acquisition module is to facilitate the interpretation and formalisation process of the domain expertise into usable rules. Furthermore, it enables the transfer of the discovered knowledge into the system through the population, update, and expansion of knowledge bases.

- *User interface*: It provides the means for a user-friendly interaction between the end-users and the expert system through graphical interfaces, menus etc. A user interface is responsible for accommodating the user to present the specifics of a problem to the expert system, as well as enabling the system to present the underlying rules and reasoning in human-understandable way and convey the outcome of the analysis to the end-user.

A major benefit of the knowledge-based approach is that its outputs are auditable as the reasoning behind them is transparent. Moreover, future re-evaluations of the inspection and condition monitoring processes can be incorporated into the system through updating of the knowledge base. These advantages enabled knowledge-based expert systems to be widely applicable in the industrial setting [46]. Garcia et al. [47] presented a failure mode and effects analysis of nuclear safety systems utilising the opinion of five experts to identify potential failure modes and to formalise them into a set of logical inferences. Murray et al. [48] developed a visual inspection processing software that significantly reduces the in-core inspection times of AGR nuclear power plants. This approach uses domain expert knowledge that reduces the complexity and computational effort and currently being tested to assess the feasibility of replacing the existing manual image processing method by a leading energy company [49].

West et al. [50] developed a rule-based expert system that extracts information related to the health of AGR cores through the analysis of data related to the frictional interface between the fuel assembly and the fuel channel (Fuel Grab Load Trace data). This system has been deployed within the nuclear operator in the UK and is used to support the analysis of refuelling events of AGR stations.

2.2.2.2 Model-Based

Closely related to knowledge-based expert systems, are the model-based systems which incorporate available expert knowledge into explicit mathematical models that emulate the physical behaviour of the system of interest. Often researchers approach this problem through finite element methods [13, 14, 15, 16, 17, 18] that compare received inputs and outputs of a system with simulated responses and evaluate the discrepancies.

The development and calibration of such models is a sophisticated procedure that aims to closely match the behaviour of the numerical model to the behaviour of reference states of a real asset [51]. The application of a model that is well calibrated to healthy reference states to new test cases can reveal legitimate discrepancies that indicate the presence of damage. Solving the inverse problem of identifying the model parameter values that generate the observed responses, can potentially inform the location of a flaw and quantify the damage levels.

Verification and validation of the numerical models are crucial processes that build credibility in the models and enable them to produce predictions with quantifiable confidence. The definitions given by the Los Alamos National Laboratory

are: "Verification is the process of determining that a model implementation accurately represents the developer's conceptual description of the model and its solution. Validation is the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model.". Therefore, verification is primarily concerned with the mathematical aspect of assessing the accuracy between the model outputs and known solutions. On the other hand validation is concerned with whether the simulated physical representation of the model is sufficient to represent the real world physical behaviour of the model [52].

Model-based methods have demonstrated their applicability to nuclear reactor condition monitoring through control rod vibration analysis, loose part localisation, identification of core barrel material degradation [53], and three-dimension sizing of stress corrosion cracking of stainless steel used in recirculating pipe and in shroud in Boiling Water Reactor nuclear plants [54]. Commonly, monitoring procedures relying on physical models of an asset, operate in a hybrid mode where an advanced statistical or signal processing algorithm is required as a pre.processing step to transform the raw data into the appropriate input [55, 54].

Although a model-based approach can provide a complete condition monitoring system that is based on first-principles, developing such a complex approach can be challenging as the availability of detailed knowledge and representative data required for robust parameter estimation can be scarce [56]. Therefore, in the case of complex systems modelling the physical and mechanical characteristics of a system may require assumptions and approximations of the physical relations [57].

2.2.2.3 Data-Driven

An alternative to knowledge-driven expert systems and model-based systems is the use of data-driven approaches which categorise inspection data through the identification of learned patterns within features extracted from the data. In a condition monitoring setting operational and inspection measurements can be analysed by a model and infer the condition of an asset by either classifying its state as one of the known healthy or faulty states, or by determining whether the observed measurements suggest a state that is sufficiently similar, or dissimilar, to the expected "healthy" state.

The machine learning (ML) field emerged out of traditional statistics, data mining, and artificial intelligence communities and offers a large pool of algorithms that are currently widely used in condition monitoring applications across multiple industries [58, 59]. ML algorithms rely heavily on the analysis of historical data to identify patterns that can be later used to analyse new data.

In the case where the historical data are accompanied by labels that describe the class of each instance of observations, the algorithm can be automatically trained to learn useful patterns through the process that is called supervised learning. A common training approach performs gradual self-improvement of a model by iteratively updating the model's parameter values guided by the comparison of the model's predictions to the ground truth labels. This training approach is prevalent in artificial neural network models which consist of multiple layers containing neuron-inspired nodes that act as a mapping between the inputs and the target class labels.

If the class labels are unknown, the data can be analysed using unsupervised learning algorithms which aim to extract meaningful patterns from the data without guidance from a ground truth. An example of unsupervised learning subcategories is clustering analysis, where the data is partitioned into groups such that individual observations in the same group are more similar to each other. A hybrid approach, commonly used when there is an abundance of unlabelled data but only limited related training data, is semi-supervised learning which combines principles from both unsupervised and supervised learning methods [60] and utilises both categories of data to improve performance.

Data-driven methods have recently gained widespread popularity due to displaying significant capabilities in analysing sequential data and performing computer vision tasks. Furthermore, advances in sensor technology are enabling the generation of large amounts of data, while enhanced computational efficiency and computing power are enabling the rapid digestion of data from the algorithms.

A challenge of effectively training an ML model is that although training procedures automatically handle the tuning of the algorithm's internal weight values, e.g. through backpropagation [61], the tuning of a model's hyperparameters which explicitly direct the structure of a model relies on heuristic approaches. Selecting the appropriate hyperparameters can have a significant effect on the performance and behaviour of the model, and it is generally aimed towards maintaining a balance between overfitting the training data and underfitting the data. Overfitting happens to models that are high capacity relative to the size and noisiness of the training data [62]. A consequence of this can be that a model learns the input data, and not its abstraction, thus performing well in the training data, while

underperforming in new unseen data due to lack of generalisation capabilities. This is especially prevalent in large machine learning models when the number of trainable model parameters significantly outnumbers the number of available training examples. Underfitting is the opposite effect of overfitting which happens when a model is too simple, or restricted, compared to the complexity of the training data, resulting in poor predictive performance both in training and new data. Therefore, a heuristic approach towards the optimal hyperparameter setup is to identify the hyperparameter values that maintain reasonably good model performance in known training data, but also to maximise the performance for data previously unseen by the model.

An additional challenge of data-driven methods is the limited explicability of the results due to the black-box nature of complex models, i.e. the inability to provide an interpretable reasoning behind a decision. Nevertheless, data-driven methods offer a unique flexibility that allows them to be used in a wide array of condition monitoring applications.

Earlier work has demonstrated the feasibility of both supervised [63, 64] and unsupervised [65] data-driven monitoring and fault diagnosis applications in nuclear energy sector. More recently, data-driven methods have become widely used for the health management of nuclear power plants across a variety of applications and types of inspection data [66, 67].

Chen et al. [68] implemented a crack detection method applied on inspection videos of underwater metallic surfaces of nuclear power plants. The method is based on Convolutional Neural Networks (CNNs) and showcased high hit rate

against low false positives per frame, however its applicability is determined by the availability of large amounts of training data as well as computational power. Chae et al. [69] utilised Long Short-Term Memory (LSTM) networks for the detection of pipe thinning due to flow accelerated corrosion. The analysis was performed on vibration data and it was shown that LSTMs outperformed other machine learning algorithms, such as SVMs (Support Vector Machines) and CNNs, as LSTMs are more well-suited towards the analysis of sequences of data. Berry [60] expanded on previous work on the detection of cracked graphite bricks in AGRs [70, 50, 37] by applying semi-supervised learning techniques. This allowed for the combination of infrequent labelled data, derived from detailed inspections of the core, with abundant unlabeled monitoring measurements taken during refueling operations, which improved the classification of graphite brick health states. Wang [71] used a combination of the unsupervised and supervised approaches kernel principal component analysis, similarity clustering and SVM for fault diagnosis in the reactor coolant system of pressurised water reactor. This enabled a highly accurate fault diagnosis process that also allows for visual illustration of the results leading to enhanced interpretability.

2.3 Pressure Tubes Health Assessment

2.3.1 CANDU Reactor

Table 2.1: CANDU and CANDU-Derived reactors globally [72]

Facility	Units	Net Capacity (MWE)
Bruce Power (Canada)	8 CANDU reactors	6,288
Darlington (Canada)	4 CANDU reactors	3,512
Pickering (Canada)	6 CANDU reactors	3,094
Point Lepreau (Canada)	1 CANDU reactor	660
Wolsong (South Korea)	4 CANDU reactors	2,576
Qinshan (China)	2 CANDU reactors	1,354
Cernavoda (Romania)	2 CANDU reactors	1,300
Emblase (Argentina)	1 CANDU reactor	600
Karachi (Pakistan)	1 CANDU reactor	90
Tarapur (India)	2 CANDU-derived reactors	980
Rajasthan (India)	1 CANDU reactor 4 CANDU-derived reactors	187 808
Kaiga (India)	4 CANDU-derived reactors	808
Kakrapar (India)	2 CANDU-derived reactors	404
Madras (India)	2 CANDU-derived reactors	410
Narora (India)	2 CANDU-derived reactors	404

CANDU reactor is a type of pressurised heavy water reactor (PHWR) that utilises natural uranium to generate electric power. CANDU reactors were first designed in Canada and currently there are 30 operable reactors globally in Canada, Argentina, China, India, Pakistan, Romania, and South Korea. Furthermore, the majority of nuclear reactors in India are PHWR-type based on CANDU reactors. Table 2.1 lists the currently operational CANDU-type reactors, which constitute about 11% of the nuclear power reactors in use today globally [72].

Similar to other nuclear power plant types, the reactor is the heat source that enables the generation of electricity. To facilitate that, the CANDU nuclear power plant consists of interconnected units that control the flow of energy (Figure 2.3); from the generation and regulation of thermal heat, to its transformation into mechanical energy and finally electrical energy.

The key units of a CANDU nuclear plant include [73]:

- The nuclear reactor generates the heat through nuclear fission.
- The primary heat transport system is responsible for transferring the heat from the reactor to the steam cycle.
- The steam generator is part of the secondary heat transport system, where light water is converted to steam.
- The steam turbine converts the pressurised steam to mechanical energy.
- The electrical generator converts the mechanical energy into electrical energy.

The design of CANDU reactors enables high neutron economy, as this is critical for achieving fission through a sustained chain reaction, given that natural uranium has low fissile content. Some of the unique features of the CANDU reactors' design that enable this process are [73, 75, 76]:

- Use of heavy water (deuterium oxide D_2O) as both coolant and moderator, which has a reduced tendency to absorb neutrons, compared to light water.

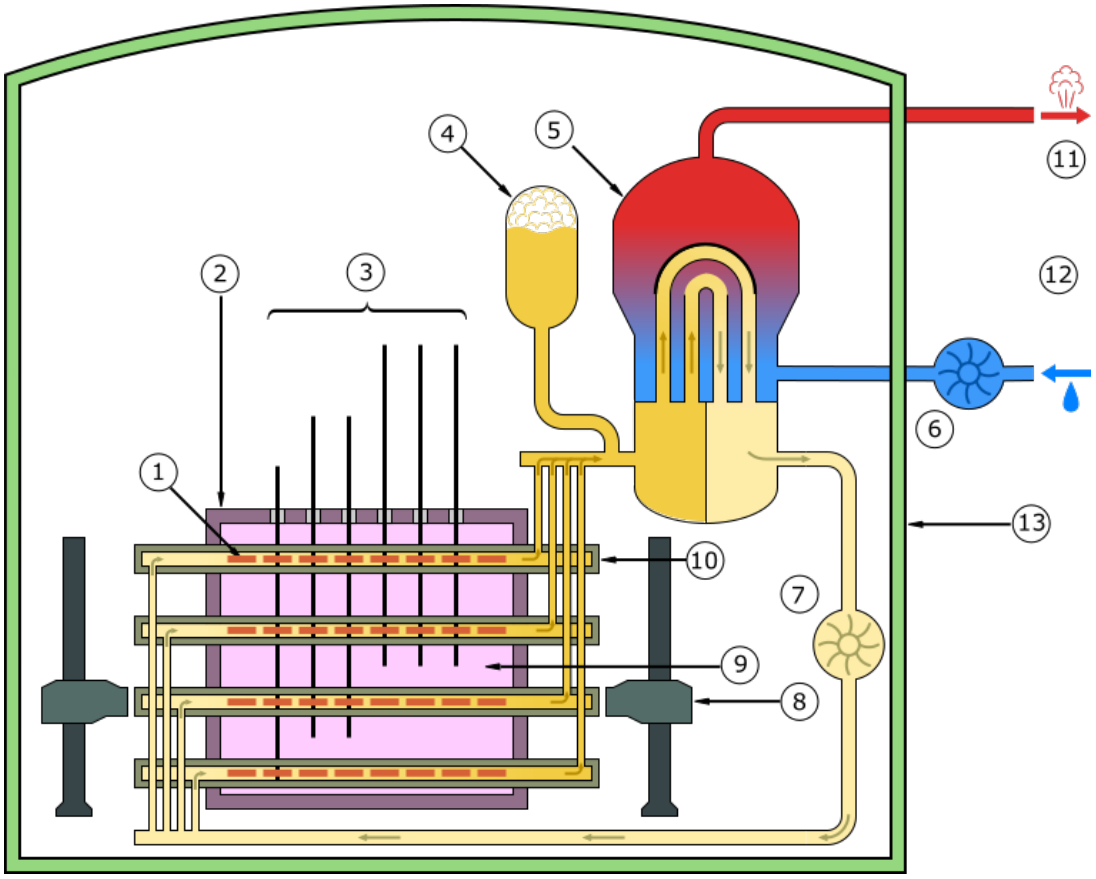


Figure 2.3: Schematic of a CANDU plant (Adapted from [74]). Primary heavy-water loop: Hot, Cold. Secondary light-water loop: Hot, Cold. 1) Fuel bundle, 2) Calandria (reactor core), 3) Adjuster rods, 4) Heavy water pressure reservoir, 5) Steam generator, 6) Light water pump, 7) Heavy water pump, 8) Fueling machines, 9) Heavy water moderator, 10) Fuel channel, 11) Steam going to steam turbine, 12) Cold water returning from turbine, 13) Reinforced concrete containment.

- The core structural components consist of low neutron absorbing materials (zirconium alloys).
- Online refuelling. This minimises the neutron absorption by the control materials and results into low core excess reactivity, eliminating the need for burnable neutron absorbers.
- The reactor design facilitates lower resonance absorption rates and better neutron thermalisation.

The reactor consists of a low-pressure stainless steel tank, called calandria, which contains the moderator. Around 480 fuel channels run across the length of the calandria tank. Each fuel channel (Figures 2.4, 2.5) consists of a thin zircaloy tube (calandria tube) that contains another, thicker, zircaloy tube called pressure tube. The two tubes are separated by the annulus garter-spring spacers that create a gap allowing the flow of an insulating gas called annulus gas. The fuel bundles are located inside the pressure tubes and they are cooled by the hot (≈300°C) pressurized (≈10MPa) heavy water coolant that runs through the pressure tubes. This configuration enables the insulation of the cool (≈70°C) unpressurised moderator from the hot pressurized coolant.

2.3.2 Pressure Tubes

Pressure tubes are one of the critical components of pressurised heavy water reactors as they contain the nuclear fuel bundles and the pressurised coolant. During their operating life they are required to operate reliably in an extremely

harsh environment of high pressure, temperature and neutron flux. Although the extensive quality controls and monitoring of the pressure tubes have ensured a good overall performance [38], pressure tubes are subjected to some degradation mechanisms facilitated by the severe conditions, which could lead to their failure.

One of the principal degradation mechanisms that pressure tubes experience is concentration of hydrogen/deuterium at areas of large tensile stresses. This can precipitate hydrides that initiate a process called Delayed Hydride Cracking (DHC) that can cause coolant leakage potentially damaging the reactor. Therefore there is a need to closely monitor any pressure tube tensile stress concentrations, such as flaws and defects created during tube manufacture, installation, commissioning, or operation. Examples of such flaws include refuelling scratches, fuel fretting, crevice corrosion, and debris fretting.

2.3.3 Inspection Setup

The current inspection process is performed during planned outages on data obtained by tools employing multiple ultrasonic inspection probes [39]. The available ultrasonic inspection tools that facilitate the scanning process are known as CIGAR (Channel Inspection Gauging Apparatus for Reactors) and ANDE (Advanced NDE).

The more recently introduced system ANDE offers advanced hardware capabilities and productivity rates over the CIGAR system. However, the basic operating principles of these systems are similar, allowing them both to perform ultrasonic inspection of pressure tubes and deliver data compatible to some degree with

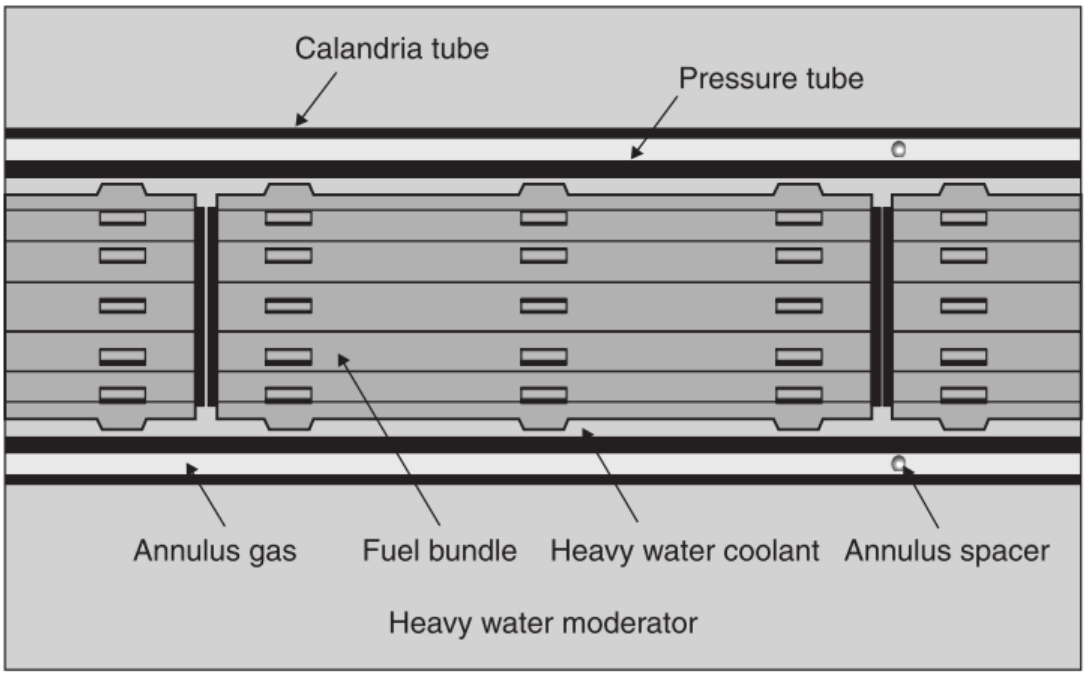


Figure 2.4: CANDU reactor fuel channel arrangement [75].

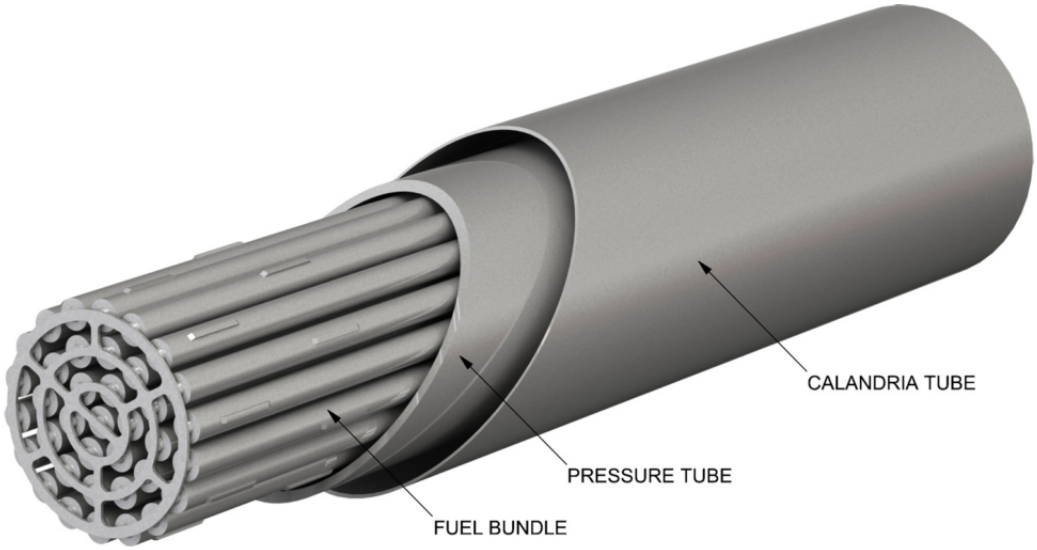


Figure 2.5: CANDU reactor fuel channel arrangement 3D [77].

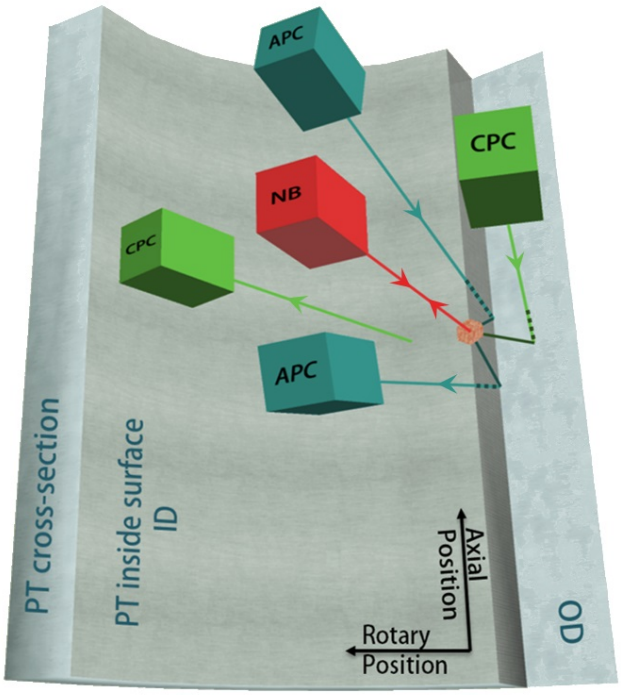


Figure 2.6: Ultrasonic inspection probe orientations.

existing software. The tools consist of a combination of normal incidence probes (NB), operating in pulse-echo mode at 10 MHz (NB10) and 20 MHz (NB20), and pairs of axial pitch-catch probes (APCs) and circumferential pitch-catch probes (CPCs) operating at 10 MHz via full skip propagation. The orientation of the probes is such that any point of the pressure tube is inspected from multiple directions (Figure 2.6).

The data acquisition process of CIGAR is carried out in two stages. During the first stage a general low resolution NB and shear wave general helical (GH) scan is performed across the two halves of the tube. This provides indications of potential flaws based on gated amplitude responses, and serves as the main flaw sizing path. At the second stage, the scanning of the flagged abnormal areas by

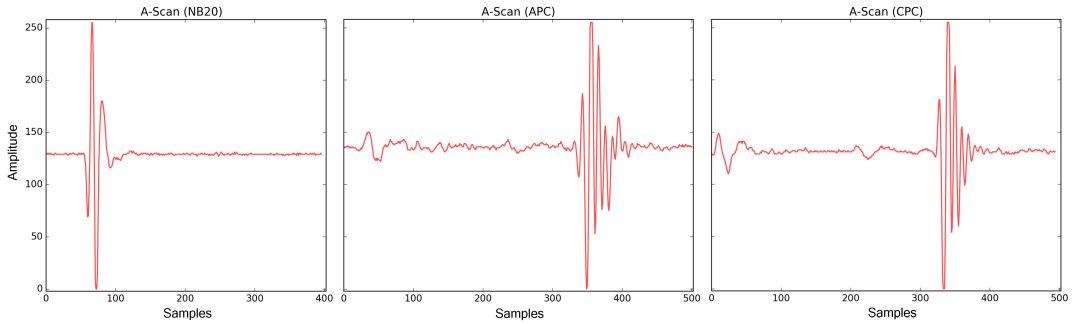


Figure 2.7: Typical examples of ultrasonic signals captured by different probes of the CIGAR tool. The signals are from the same, randomly selected, position of a pressure tube.

the two NB probes, along with the APC and CPC pairs, provides more detailed 3D datasets that serve primarily as a depth assessment source. As all four beams meet at the same position of the pressure tube internal diameter, any potential flaw can be examined from four different directions. Examples of high-resolution ultrasonic signals obtained at the same position of a flagged area by the NB20, APC, and CPC probes can be seen in Figure 2.7. Given the sub-millimetre scanning increment, the resulting dataset for a region or interest is a collection of thousands of signals per cm^2. These individual signals are known as A-Scans and each signal is the captured time-amplitude waveform of one transmitted pulse.

Further possible representations of the captured signals include the B-Scans and C-Scans, which provide different functionalities. B-Scans are stacked A-Scans along a certain axis, and therefore assist the task of flaw depth estimation as they provide a 2-dimensional cross-section view. On the other hand, C-Scans can assist the localisation of potential flaws as they provide a top-down 2-dimensional overview of the scanned surface through single value representation (e.g. maximum amplitude) of each signal.

This assessment process is also followed for datasets acquired by the ANDE system, however the improved capabilities of ANDE enable the acquisition of detailed 3D data from multiple (12) probes across the full length of the pressure tube during the initial scanning, eliminating the need for the tool to revisit areas of interest. This, in combination with the increased resolution of the acquired signals, result in datasets around 1000 times larger than those of CIGAR, posing a great overhead for any existing analysis process.

2.3.4 Existing Analysis Process

The extracted ultrasonic inspection datasets are first streamed from the inspection tools to a local data storage at the equipment site before being transferred to a dedicated analysis centre where trained analysts inspect the data and characterise any flaws detected.

The analysis is performed manually with the use of a proprietary software which is called FLAW. The first stage of the analysis consists of an automated flagging process, offered by the FLAW software, which indicates regions that potentially contain flaws and is based on preset signal amplitude thresholds [78].

Two analysts are provided with a detailed set of information on the history of the pressure tube in question, and are then tasked with independently reviewing the data based on the flagged regions. The analysts complete their review of the available C-Scans (2-dimensional top-down view of the entire channel) and

B-Scans (high resolution slices of stacked A-Scans extracted from the flagged regions) and summarise their findings in a common format, reporting measured locations, sizes, depths and types of flaw.

A third analyst then acts to consolidate these results and performs a resolution of any conflicting interpretations. The length of time required for the complete analysis process of a pressure tube depends on the number and complexity of the flaws identified. When a finalized set of flaws have been agreed, they are compared with a set of criteria developed to determine if any single flaw poses significant structural risk to the pressure tube. The key characteristic in determining this is the depth and root radius of the flaw, which are modelled using a separate analysis software.

In occasions where specific flaw criteria exceed permissible limits, a polymer replica is recorded of the flaw in order to make a direct measurement of the geometry of the flaw, rather than the inferred ultrasonic measurement. This process is complex and time consuming and often requires multiple attempts in order to capture a complete replica. At the end of the analysis process, when all flaws have been analysed and classified, a judgement is made of the integrity [79] of the pressure tube and the ability of the tube to return to service. This process is repeated across all inspected tubes and forms part of the critical path during outage in returning the reactor to power.

2.3.4.1 Existing Process Issues

The large volume of data, intensive analysis process and time pressures of the inspection regime require a number of human analysts to support each outage, often analysing only sections of particular pressure tubes as part of the process described earlier. Despite the use of a multi-pass analysis and a resolution analyst, the process introduces a level of subjectivity and variability in the results. An increased volume of inspection data generated by new tools, combined with ageing plant and a potential increase in the number and complexity of flaws, is likely to drive a requirement to increase the efficiency of analysis. Currently only a subset of the highest resolution inspection data is utilised, for specific regions of tube, but if the complete set of data were to be analysed in the future, new analysis processes would be required [80].

2.3.5 Expert-System Analysis of Pressure Tube Ultrasonic Inspection Data

The challenges of the existing manual analysis process of ultrasonic data gathered from the CANDU pressure tubes led to the development of an automated expert system aiming to save time on the critical path, reduce human subjectivity within the process, and increase the repeatability of measurements [23]. The prototype software system, known as ADAPT (Automated Data Analysis for Pressure Tubes), has been implemented in MATLAB and provides end-to-

end defect detection and characterisation. Testing analysis on historical data demonstrated the viability of a decision support tool which can aid the existing analysis process [80].

ADAPT is based on the principles of expert systems and it integrates rule bases and analysis algorithms derived from experts in the analysis of ultrasonic inspection data of CANDU pressure tubes to replicate and automate the existing analysis process, as shown in Figure 2.8. The process followed by experts is reproduced by adopting the CommonKADS approach to knowledge representation [6]. KADS (Knowledge Acquisition and Documentation Structuring) defines a structured methodology for knowledge-based systems design and this methodology matured into the CommonKADS approach [81].

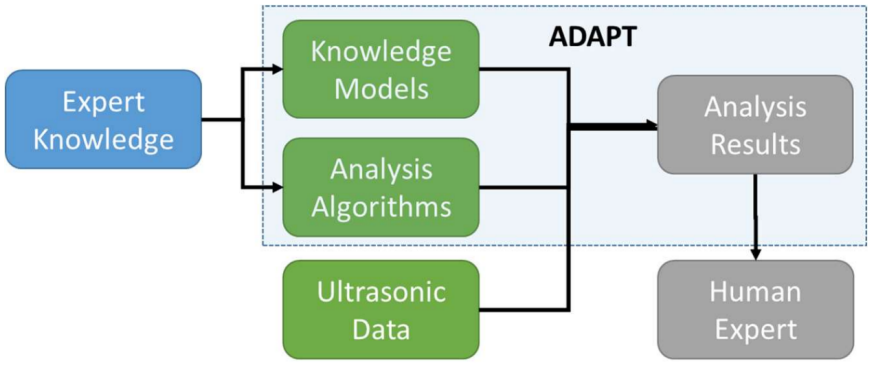


Figure 2.8: A conceptual overview of ADAPT [80].

One benefit of the knowledge-based approach is the ability for the analysis results and the decisions made by the system at each stage to be fully explicable and auditable. Moreover, future extensions or changes of the current understanding of the inspection analysis process can be incorporated into the system through updating of the knowledge base.

The system prototype has been developed to provide decision support for different stages of the analysis process. Figure 2.9 depicts the major modules that enable the 'end-to-end' processing of the ultrasonic inspection data. The input to the system consists of the same ultrasonic inspection datasets that a human analyst would receive, and the output is a set of reports in format readable by the software currently in use by the analysts, FLAW.

The individual analysis steps are examined below.

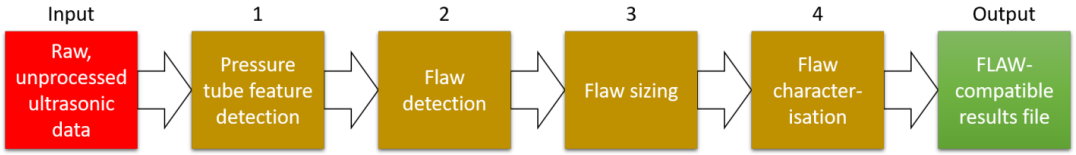


Figure 2.9: ADAPT: A high-level flowchart of the major modules comprising the automated analysis process [23].

2.3.5.1 Pressure Tube Feature Detection

The first stage of the analysis (Step 1 of Figure 2.9) involves the localisation of three mechanical features that are located at each end region of a pressure tube. These features are the burnish mark, three rolled joints, and the end of pressure tube. The burnish mark and the rolled joints (depicted in Figure 2.10) result from the process of mechanically fitting the pressure tubes to the reactor core.

These features are normally located by analysts by identifying bands of equal amplitude along the radial dimension of the 20MHz normal incidence datasets. This procedure is replicated in ADAPT's feature detection module by calculating

the mean value of the amplitudes across the axial positions and identifying the peaks and troughs that comply with the expected characteristic properties and distances.

Identifying these features enables the determination of the boundaries of the critical area, where the presence of DHC is more common due to the non-uniform stresses present in this area [39].

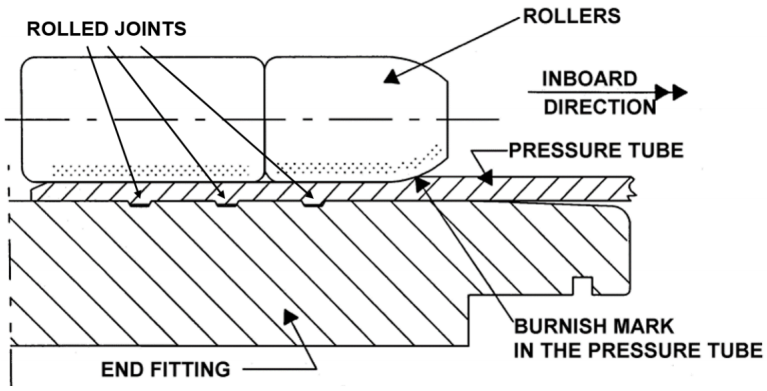


Figure 2.10: An annotated outline [82] of the end region of the pressure tube. The localisation of the three rolled joints and burnish mark is the first analytical step of the assessment process.

2.3.5.2 Flaw Detection

Following the pressure tube feature detection ADAPT processes the available GH datasets to locate any potential flaws (Step 2 of the flowchart in Figure 2.9). The locations of potential flaws are identified by applying different amplitude drop thresholds depending on the area of pressure tube and the inspection tool used.

The flagged sections are examined to determine whether they should be considered as noise, based on processes defined within the inspection specification documentation. However, at this stage only a basic noise classification algorithm is applied, as although relative operating procedures exist, there is still scope for subjectivity regarding what would constitute noise. Therefore the noise classification is not finalised until further characteristics of the flagged areas are also taken into consideration, such as length, width, and B-scan response. Subsequent system modules accommodate that by providing the ability to classify flagged areas as noise if their properties satisfy the noise conditions.

2.3.5.3 Flaw Sizing

The third step of the automated process (Figure 2.9) is concerned with the sizing and depth of the flagged areas.

The goal of the sizing stage is to measure the length and width of a flagged region and is accomplished through the examination of the 10MHz normal incidence probe datasets.

For this process a threshold amplitude value is calculated based on the expert analysts‘ operating procedure, which is related to the neighbourhood of the flagged section. This value serves as the threshold that the amplitudes of the normal incidence responses must exceed in order to be characterised as non-defective. The areas that fall below that threshold, and intersect with the corresponding flag, are incorporated by the flag and the dimensions of this area along the axial and radial axes represent the length and width of the flaw.

The depth assessment of the potential flaws within flagged areas requires detailed examination of B-scan datasets obtained from the corresponding pressure tube region. The knowledge model for this process is a combination of information extracted from inspection specification documentation and formal domain knowledge capture sessions with expert analysts. The algorithm operates on A-Scans obtained from the 20MHz normal incidence probe, as well as the two pairs of pitch-catch probes. During the first processing stage each A-Scan within the anomalous area is examined to identify signal features that can potentially act as depth measurement points. Each signal feature is assigned with a confidence measure that is adjusted by applying a set of domain rules. Following the confidence assignment, the feature with the highest confidence within each signal is selected to represent the depth measurement of the corresponding position. This procedure generates a depth-map of the area that allows for the variations of depth to be examined. This enables the identification of steep depth gradients that pose a risk for hydride build-up and DHC due to the increased mechanical stress at the tip.

2.3.5.4 Flaw Classification

The final step of the automated process (Step 4 of Figure 2.9) is responsible for classifying the detected flaws based on the characteristics measured at the earlier steps, i.e. location, length, width, and depth. Currently the available knowledge base of the system differentiates only between the most common classes of flaws, such as debris frets, scrapes, and fuel bundle bearing pad frets.

2.3.5.5 Capabilities and Limitations

The ADAPT prototype has been tested against a large volume of historical data and the satisfying results demonstrate the viability of decision support tool that can assist the current analysis process [80]. However, it has been highlighted that certain complex tasks within the existing assessment procedures cannot be replicated exactly through formalised rules and partly rely on expert human judgement and intuitive visual interpretation.

An example of such case is the examination of the B-Scan datasets which apart from enabling the depth measurement of flaws, also constitute sources for the detection of potential flaws. While the amplitude response flagging of the GH-produced C-Scans is responsible for the detection of the bulk of flaws, expert analysts also review areas of the high-resolution B-Scans to detect anomalous areas potentially undetected by the initial detection process. This requires significantly more subjective visual examination of a variety of data in either B-Scan or A-Scan formats obtained from a variety of sources (NB, APC, CPC). Furthermore, the recent deployment of the ANDE tool introduced additional inspection probes that deliver greater volumes of higher resolution signals across the entire pressure tube.

2.4 Discussion

This chapter has provided an overview of the importance of pressure tube health assessment through ultrasonic non-destructive evaluation and the challenges of performing this task in a robust and timely manner. Although current research efforts have demonstrated the viability of an automated system that can support this process, there is still a need for tackling the problems and procedures that cannot be formalised into exact knowledge and strategy required by expert systems to emulate the subjective actions of an expert.

Introducing analysis methods that are driven by the structure of the underlying ultrasonic inspection data should enable the highly subjective task of detecting and measuring flaws through the examination of high-resolution B-Scan datasets. Moreover, it is a step towards providing a more thorough assessment of the entirety of a pressure tube through the analysis of newly introduced multi-probe full-body high-resolution scans and understanding the challenges that such a large-scale environment poses. To support the discussion about the feasibility of such methods, the following chapters provide an overview of available data-driven techniques, as well as a review of their application to relevant problems.

Chapter 3

Overview of Automated Ultrasonic Analysis Methods

3.1 Ultrasonic Non-Destructive Evaluation

Ultrasonic non-destructive evaluation has become a key technology in advanced diagnostics and monitoring [83]. The earliest patented practical application of ultrasonic evaluation, following important foundational work of Sokolov and Muhlhauser [84], can be considered to be Dr Firestone's (1898–1986) 1942 'Flaw Detecting Device and Measuring Instrument' – a single quartz transducer used in what became known as pulse-echo technique [83, 85], which was later applied to the inspection of air force bomber landing gears. The technology was rapidly employed and developed, and further advancements in the understanding of the interactions between ultrasonics and materials led to more advanced inspection systems that can now provide health assessment support across a wide range of industries in their inspection tasks.

Ultrasonic inspection uses high-frequency (greater than 20kHz) acoustic waves for the detection of surface and subsurface anomalies in materials. The sound waves are fed into the material and by assessing the changes in their propagation characteristics it is possible to infer information about possible flaws [4, 86]. The changes that can occur in a propagating wave can be summarised as [87]:

- Reflections at material interfaces/boundaries or material discontinuities
- Time of transit discontinuities
- Attenuation by absorption and scattering
- Changes in the spectral response

As the signal is shaped by the properties of the material, it can contain echoes not only from a defect but also from insignificant imperfections in the structure of the material (e.g. porosity and inhomogeneity). These echoes can be characterised as noise, as they can potentially mask flaw echoes in the received signal, i.e. they reduce the signal-to-noise (SNR) ratio [5].

3.2 Automated Ultrasonic Evaluation

The need for robust interpretation of large volumes of inspection data and minimising errors and variability due to human factors [11, 12] is the drive behind the extensive research efforts to create automated procedures for signal classification and automatic detection of defects.

Intelligent systems can support the automated evaluation of ultrasonic inspection signals by automating the complex decision-making involved in identifying and sizing material defects. As ultrasonic nondestructive evaluation is a subset of condition monitoring, the implementation of ultrasonic analysis intelligent systems can be broadly approached through any of the strategies described in Section 2.2.2, i.e. knowledge-based expert systems, model-based, and data-driven.

Although both knowledge-based expert systems (Section 2.3.5) and model-based approaches [13, 14, 15, 16, 17, 18] can offer valuable automated analyses given sufficient expert specialised knowledge, the scope of this thesis is to provide a more system-agnostic approach that relies on data-driven methods.

3.2.1 Data-Driven

3.2.1.1 Supervised Approaches

Similarly to various condition monitoring applications, data-driven methods have also gained popularity as effective means of analysing collections of ultrasonic inspection data, as their flexibility allows them to be applied across multiple environments and for different tasks. Supervised methods are a subset of data-driven methods and are based on machine learning techniques that algorithmically explore datasets of signals and infer knowledge through gradual self-improvement of their performance (training). One of the most extensively used machine learning technique is the Artificial Neural Networks (ANNs).

ANN is a machine learning paradigm inspired by the biological nervous systems, consisting of several highly interconnected processing units (neurons) (Figure 3.1) working in unison to solve specific problems [88]. ANNs infer knowledge through a learning process that adjusts the inter-neuron connection weights, and their success derives from their ability to solve complex non-linear problems [89].

Their learning process initially requires the quantification of the difference between the target and predicted output values (loss function). Subsequently, the gradient of the loss function with respect to the network's weights guides the update of the weights, aiming towards reaching the minimum of the loss function. Specifically, the computation of the gradients is commonly performed by an efficient implementation of the chain-rule, called back-propagation algorithm [90], whereas an optimisation algorithm (e.g. Stochastic Gradient Descent [91, 92], or Adam [93]) is numerically reaching the minimum loss, through the informed update of the model weights, which is guided by the calculated gradients [94]. This iterative process repeats for either a specified number of cycles, or until a stopping criterion is satisfied, e.g. the loss is sufficiently low.

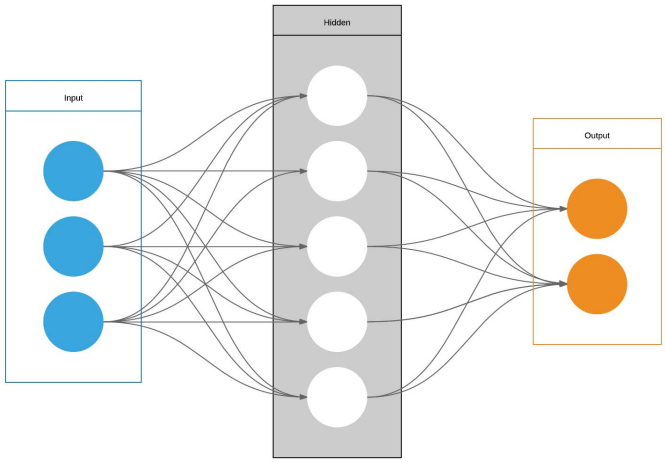


Figure 3.1: A typical architecture of an Artificial neural network (ANN) with a single hidden layer [95]. The interconnected groups of artificial neurons are updated during the training phase, guided by the gradient of the loss function which quantifies the difference between the outputs of the last layer and the expected outputs.

ANNs have been extensively applied to automated ultrasonic evaluation applications. Margrave et. al [96] presented an evaluation of various types and configurations of neural networks in regards to supporting ultrasonic flaw detection and classification in steel plates. The results showed promising classification rates on real or simulated defects with minimum pre-processing. D'Orazio et. al [97] further demonstrated the capabilities of ANNs by applying them into the analysis of ultrasonic inspection data of composite materials. Following a simple normalisation procedure, and no advanced signal processing method for signal denoising and feature extraction, their model demonstrated the ability to perform effective classification of signals and defects in materials of varying thickness and textures.

Iyer et. al [98] presented a comprehensive ANN-based signal classification method of signals obtained from a concrete pipe that uses a different signal feature extraction method. The overall approach consisted of three major steps: signal preprocessing, feature extraction and ANN classification. During the preprocessing stage the signals were processed by the discrete wavelet transform (DWT), passing them through a series of filters to decompose them into their detail and approximate signals. DWT is an analysis technique able to construct time-frequency representations of a signal [99, 100], a property that is not provided by Fourier transform (FT) [101], but it can also be obtained through Short-Time Fourier Transform (STFT) [102]. Then a set of coefficients of representative signals from the entire spectrum of defect classes were clustered to identify the frequency bands that were considered to contain the most information. These clusters were then converted into feature vectors by taking the energy of each cluster as a single feature. The feature vectors were then fed into the NN model for training it into classifying signals to their appropriate defect classes. Furthermore, the ANNs were compared to a statistical classifier (Linear Discriminant Analysis) revealing ANNs' ability for significantly more accurate classifications.

Although ANNs have proven their effectiveness throughout an extensive range of applications, research has presented examples of Support Vector Machines (SVMs) (Figure 3.2) providing better classification results than ANNs [103, 104]. This should cause caution when the classification problem involves a small sample and high dimensional feature spaces [105] which can lead to ANNs overfitting the training data, resulting in poor generalisation capabilities (Section 2.2.2.3).

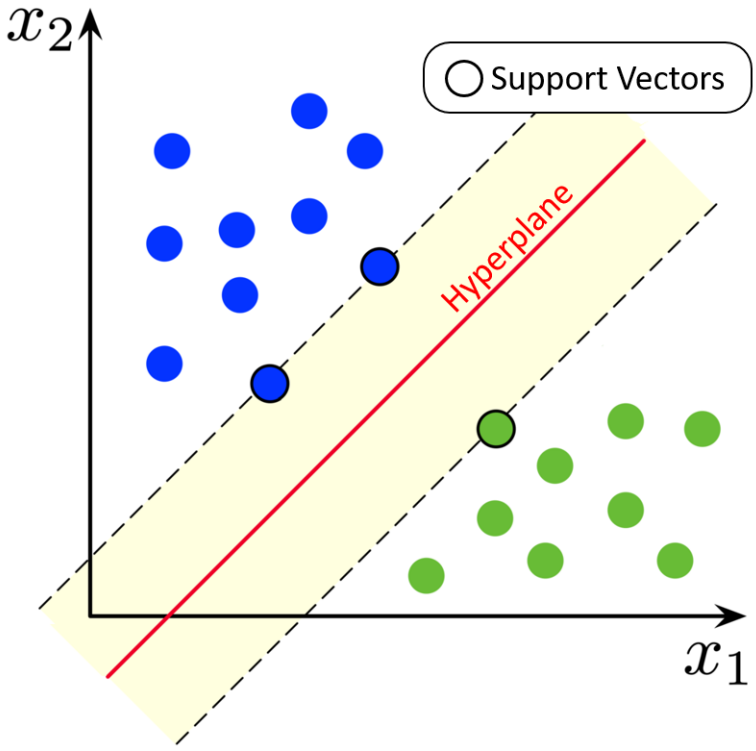


Figure 3.2: Support vector machine (SVM) [106]: Find the hyperplane (red) that maximises its distance from the nearest data point on each side. The data points that define these distances are called 'support vectors'.

SVMs iteratively try to find the hyperplane within the labelled feature space that best separates the different classes. The best dividing plane is considered to be the plane that maximises the distance of group points that are closest to points from different groups [107, 108]. Given the SVMs' performance in classification accuracy they are a popular choice as a means to automatically classify ultrasonic signals. However, they have been found to perform poorly in the presence of noisy signals [109], therefore available applications use pre-processing techniques for denoising purposes. Lee et al. [110] performed SVM automatic classification of ultrasonic shaft inspection data showing that its performance was superior when

used DWT as denoising and feature extraction tool, than when used Fast Fourier Transform. Hassan et al. [109] further showed that applying Kalman filter as a preprocessing technique provides better SVM classification accuracy, than DWT.

Recent advances in computing power, GPU-accelerated back-propagation [111, 112] and regularisation techniques to decrease overfitting, such as dropout [113], have allowed for deeper learning networks able to outperform traditional machine learning algorithms in various tasks, including time series data [114] and computer vision tasks [115]. Deep neural networks (DNNs) rely on multiple stacked layers that perform consecutive non-linear transformations to their inputs driven by the goal of minimising the loss between the predicted labels and the actual labels. This deep hierarchical learning architecture enables the construction of high-level feature representations that utilise the lower-level feature representations of the earlier layers [94].

Munir et al. [116] trained a single-layer ANN and a multi-layer DNN on datasets containing A-Scans captured from different artificial flaws within weldments. The comparison of these models showed no significant improvement in performance of the DNN over the ANN for a dataset containing single-frequency samples. However, the DNN model outperformed the ANN when used on a mixed-frequency dataset. More recently, Munir et al. [117] augmented their original mixed-frequency dataset, which contained a small number of mainly noise-free signals, by introducing Gaussian noise and shifting the A-Scans in the time domain. These increased the size of the dataset by generating new signal instances containing different levels of signal to noise ratio (SNR) and different simulated defect

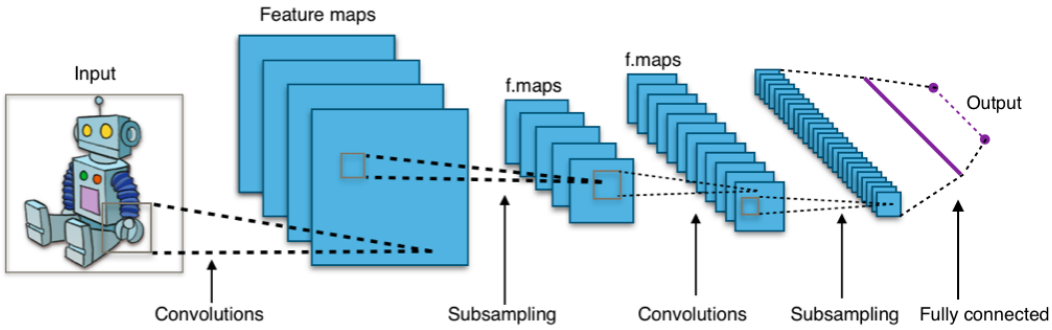


Figure 3.3: Convolutional neural network (CNN): Typical architecture [121].

depths. A DNN model, analogous to the [116], was then evaluated against a Convolutional Neural Network (CNN), showing that the CNN model performed significantly better than the DNN model.

CNNs (Figure 3.3) are well-suited for structured data and are frequently used in domains such as computer vision [118, 119] where multi-dimensional (2 face dimensions and variable depth) filters are convoluted with the input image to generate high-level convolved features that preserve local pixel relations. The convolution layers are the core building blocks of CNNs and are usually periodically complimented by pooling layers that serve as down-sampling layers that help reduce the computational footprint and reduce overfitting. Finally, the output layer of a CNN architecture consists of a classifier layer that generates the predictions that serve the loss function that drives the back-propagating training process [120].

Munir et al. [117] utilised 1D filters convolved with the augmented A-Scans to train a CNN, connected to a fully connected neural network. The CNN outperformed the DNN and proved to be more robust towards low SNR instances.

Meng et al. [122] have also demonstrated an automatic ultrasonic signal classification system utilising CNNs. Here, instead of directly feeding the raw signals to the CNN, the signals were first individually transformed into a 2D representation consisted of their wavelet transform coefficients. Furthermore, a linear SVM was utilised as a top layer to facilitate the signal classification task. The experimental results demonstrated improved performance over ANN and SVM classifiers with manually generated features.

Medak et al. [123] applied more complex CNN-based architectures that specialise in image object detection, to automatic defect detection in B-Scan images. Specifically they used popular one-stage detector architectures that rely on predefined anchored rectangles, such as EfficientDet [124], YOLOv3 [125] and RetinaNet [126]. Several variants of each model were applied to ultrasonic inspection data of six stainless steel blocks containing artificially created defects. The results showed that the base model of EfficientDet (EfficientDet-D0) outperformed the more complex variants (EfficientDet-D1 to EfficientDet-D7), as well as all the variants of YOLOv3 and RetinaNet.

Furthermore, [123] highlighted that the performance of the best performing model (EfficientDet-D0) varies based on several factors. First, the generalisation error of the model can be reduced by following the commonly [127] used practice of augmenting the available data in order to force the model to learn invariant feature embeddings [128]. Second, it was shown that larger input image resolution (512×512) increased the model's performance, compared to image resolution of 384×384. The third configuration that had a positive impact was initialising

the model weights using the weights of a model pre-trained on a large dataset of common objects (COCO dataset [129]), instead of randomly initialising the weights.

The last important factor is relevant to single-shot object detection models and is concerned with the selection of informed priors for the pre-defined dense grid of bounding boxes (called anchors) that is used by the models to locate objects. As the shape and scale of these anchors are directly affecting a model's efficiency and accuracy, it is important to further investigate these hyperparameters, instead of using the publicly available default values. This holds especially true for domains such that ultrasonic inspection, where the typical objects of interest (e.g. defects) have more extreme aspect ratios than typical objects found in large publicly available datasets used in deep learning research. Redmon and Farhadi [130] introduced the idea of identifying good anchor priors by using clustering on the training, and it was further shown by [123] that this approach yields better defect detection performance on ultrasonic inspection images, compared with utilising default anchors.

More recently, Posilović el al. [131] focused on deep learning-based anomaly detection methods for the analysis of ultrasonic inspection B-Scans. The followed approach is categorised as semi-supervised as the detection models were trained using only images that do not contain defects. This enables the models to 'learn' the expected state of normality, allowing them to flag outliers during inference as potential defects.

3.2.1.2 Unsupervised Approaches

A significant drawback of the previous data-driven ultrasonic signal analysis methods for defect detection is the requirement of labelled data, i.e. class definition for each individual signal for the training stage. There is notably little research in the area of unsupervised signal analysis for the detection of flaws where the obtained signals have no accompanying label.

The majority of existing research has focused on the analysis of low frequency guided-waves and outlier analysis to determine the presence of a defect on large structures. Zhu et al. [132] utilised guided-waves and outlier analysis to diagnose an artificial defect in a truss. The waves were processed to generate a set of relevant damage sensitive features used to construct a uni- or multi- dimensional damage index. The damage index was then fed to an outlier analysis based algorithm aiming to detecting anomalous conditions. However, the experimental data were collected in a controlled environment and a selected number of waves were selected to represent the baseline. The damage indices were assumed to follow a Gaussian distribution and a threshold of 3σ was set to indicate the outlier threshold. It was found that several combinations of features outperform other combinations, depending on the setup of actuator-sensor pairs relative to the defect. Therefore, the results associated with the data from the experiment cannot necessarily be extended to other situations where damage is located in another position or the truss type is different.

You et al. [133] utilized a clustering algorithm (K-Means) and illustrated the feasibility of the method on polymer specimen containing artificial defects. However, three crucial issues make this method unsuitable for our purpose. The first two are related to the K-Means clustering method (Figure 3.4).

K-Means is a popular clustering algorithm [134] that operates by randomly initialising centres and iteratively assigns each data point to its nearest centre until the recomputed centres do not change positions between two consecutive rounds [135].

This process requires a pre-defined number of expected clusters; a requirement that cannot be satisfied by our task which is concerned with analysing datasets that generally will contain unknown number of flaws. Furthermore, the objective of K-Means, which is the minimisation of the sum of Euclidean distances of data points from its cluster centres, serves well only (hyper)spherical shapes of clusters [136]; again, an assumption that cannot be guaranteed by the multi-dimensional noisy datasets.

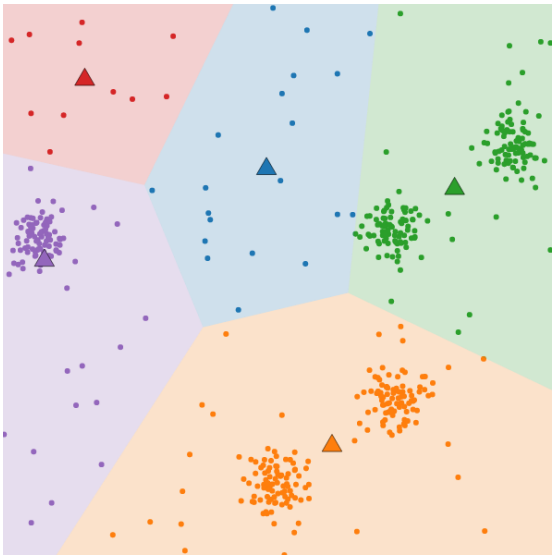


Figure 3.4: K-Means clustering algorithm: Partition dataset into a pre-defined number of clusters (here: 5), such that the within-cluster variance (within cluster sum of squares) is minimised.

Secondly, the assumption of point-to-point consistency between the signals, a requirement for the applied pre-processing step, cannot be guaranteed by the current pressure tube inspection process. This is due to the time shift that can occur in the received echoes due to slight inspection probe misalignment caused by small deformities and sagging of the pressure tubes.

3.3 Discussion

This chapter presented an overview of existing research efforts concerned with the automated analysis of ultrasonic data for health assessment purposes. Having recognised that a data-driven approach would be more appropriate for the

purposes of this thesis, this chapter focused mainly on supervised and unsupervised data-driven methods. As it has been discussed current research utilises assumptions and methods that are not appropriate for the datasets of this work.

Specifically, the major issue limiting the applicability of supervised methods to pressure tube ultrasonic inspection datasets is the requirement for individually labelled signals, i.e. information regarding whether each signal (A-Scan) is part of a flaw or not. Although currently there is available information in form of flaw locations and bounding boxes, using this to make claims about each individual A-Scan would introduce label uncertainty issues because:

- Flaws are irregularly shaped and the available information is in form of rectangular bounding boxes containing both healthy and flawed areas.
- Some datasets suffer from misalignment issues causing bounding box shifting which leads in some flaws extending outside the flagged area.

This label uncertainty will generate noise in the training dataset as several healthy signals will be labelled as abnormal and vice versa.

Regarding the unsupervised approaches it is observed that they mainly utilise small samples that enable the determination of well-defined states of normality. In the case of guided waves, the lower frequencies of these signals have long-distance propagation that allows the binary assessment of the structural integrity of large structures, but it differs from our datasets and goal of precisely locating and sizing multiple flaws within the same inspected area. Furthermore, current research in-

vestigating higher frequencies pulse-echo and pitch-catch methods are performed on small-scale controlled environments that also enable the pre-definition of the normal state.

These reasons generate the need for new approaches that allow for flaw location and pixel-wise segmentation of complex-shapes. These approaches should be able to operate across large-scale real-world noisy datasets captured by probes from various surface conditions and geometric imperfections. This exploration of the structures and patterns within the ultrasonic inspection datasets should also maintain computational times that enable their applicability to real-world datasets, as well as interpretability

One potential direction, that has not yet been adopted by the ultrasonic analysis research, could be based on current deep learning image segmentation state-of-the-art methods that utilise Vision Transformers (ViT) [137]. Specifically, the models STEGO [138] and Leopart [139] currently lead the unsupervised semantic segmentation tasks related to cityscapes [140] and object recognition [141] respectively.

However, modern Deep Learning methods are, at the current stage, incompatible with the main goal of this thesis, which is to minimise the friction in the adoption of ultrasonic analysis data-driven decision support methods in real-world industrial settings. Firstly, DL methods are computationally intensive and require GPUs for optimal performance [142]. This is a requirement that cannot be guaranteed by the computers commonly used for the analysis of pressure tube ultrasonic inspection data. Secondly, as this thesis is aiming to aid expert an-

alysts who are responsible for the health assessment of critical assets through the interpretation of inspection data, it is important to favour interpretable and transparent methods, both for the feature extraction and modelling, over more black-box approaches.

The new methods should involve interpretable processes focused on transforming and reducing the large amount of available information to levels that facilitate the discovery of potentially anomalous areas that could contain flaws. Finally, as the most important issue encountered in observed complex real-world ultrasonic data is the lack of a precisely defined state of normality that is universally true across large scale assets such as pressure tubes, the proposed methods should be able to generalise across a diverse set of inspected areas which can potentially differ in the characteristics of their healthy/unhealthy subregions.

Overall, by introducing approaches that respect the presented requirements, it should facilitate the introduction of data-driven approaches to real live testing, concurrently running along with manual assessment. This would be an especially important step, as extensive testing of data-driven methods in existing industrial environments should reveal potential inefficiencies, establish potential benefits, and eventually gain the trust as a useful decision support tool.

Chapter 4

Automated Ultrasonic Cluster Analysis for Flaw Detection

4.1 Introduction

This chapter proposes a novel unsupervised data-driven approach to automatically analyse collections of A-Scans by individually characterising each A-Scan and producing 2D maps (C-Scans) that flag potentially flawed areas. This approach operates across unlabelled signals and requires no explicit predefinition of normal state. As established in Chapter 3, the thesis will follow unsupervised approaches that do not require specialised computer equipment, aiming to provide methods that can be directly used with on-premises computers.

The analysis consists of multiple steps that consecutively reduce and transform the search space, focusing on cluster analysis methods. The main contributions of this chapter consist of a subsampling method that allows for more efficient density clustering, and a two-stage clustering procedure designed to efficiently utilise the DBSCAN clustering algorithm for grouping and noise reduction. Fur-

thermore, a DBSCAN parameter estimation method is presented, which can be applied across datasets captured by different ultrasonic inspection probes without utilising explicit knowledge of the probes characteristics.

The proposed procedure was tested across a large number of real-world datasets captured from sections of pressure tubes containing flaws utilising an automated evaluation procedure showing satisfactory results.

4.2 Ultrasonic Inspection Data

The datasets involved in the following analysis contain the inspection ultrasonic waveform data obtained by the 20MHz NB, the APC and the CPC probes of the CIGAR tool from sections of the inside surface of several pressure tubes. The tool scans the area of interest in axial and rotary increments effectively providing a set of three waveforms for each of the individual positions. The recorded waveforms consist of sampled echo amplitudes in the time domain and are presented in the form of A-Scans.

Each dataset is obtained from separate areas of interest that are known to contain surface flaws of debris fretting type. However, the available A-Scans are unlabelled, i.e. the individual signals are not accompanied by the information of whether their location is within or outside the irregularly shaped flawed areas.

The locations of the flaws, as well as the dimensions of their bounding boxes, have been verified manually by independent analysts, and provide the ground truth necessary for the performance assessment of the proposed data-driven approach.

4.3 Feature Selection

As the aim of this chapter is to identify and isolate groups of signals, it is useful to condense the information contained in each signal (average of 490 data points of amplitude-time pairs) into small subsets of representative features which act as the input to the learning process. Driven by the requirement to provide a general system that is able to adapt to the diverse surface pressure tube conditions and the irregularly shaped flaws, the features are selected based on a set of characteristics that aims to facilitate the formation of a non-system-specific procedure.

The basis of the proposed selection is that the features should be descriptive characteristics of the individual signals that do not require explicit threshold-based assumptions about the signal, nor are parameter-sensitive. Therefore, descriptive statistical measures that quantitatively summarise the entire signal were chosen instead of methods based on the detection of explicitly defined signal characteristics. Furthermore, the signals are not filtered through signal transformation procedures, such as Fourier Transform (FT), Short-Time Fourier Transform (STFT) or Discrete Wavelet Transform (DWT). These techniques are known to be effective [143], however they would increase the computational effort, as well as the complexity of the system, as they require careful parameter selection [144, 145]. Therefore, since the main scope of this thesis is the introduction of system agnostic analysis methods, the feature selection process has been focused on features that are commonly used by the experts due to their effectiveness [143], easy to interpret, and require no parameter tuning. Nevertheless, the general methods presented on this thesis can be seamlessly applied to any extended set of features.

The extracted features from each N-sample discrete signal x are:

- Sample variance:

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2, \tag{4.1}$$

  where $\bar{x}$ is the sample mean:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i, \tag{4.2}$$

- Maximum amplitude:

$$peak(x) = \max(x), \tag{4.3}$$

- Minimum amplitude:

$$trough(x) = \min(x), \tag{4.4}$$

- Index of maximum amplitude:

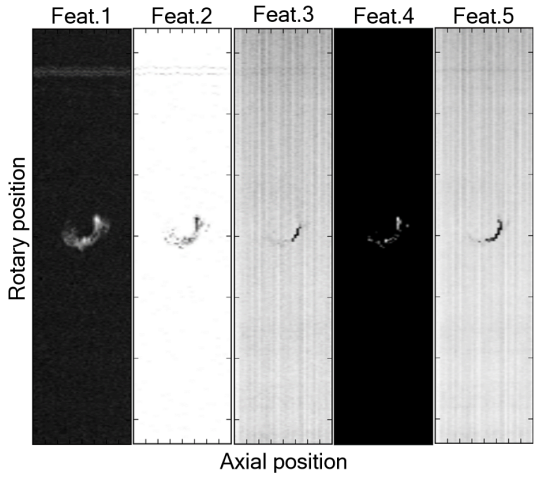
$$i_{max}(x) = argmax(x), \tag{4.5}$$

- Index of minimum amplitude:

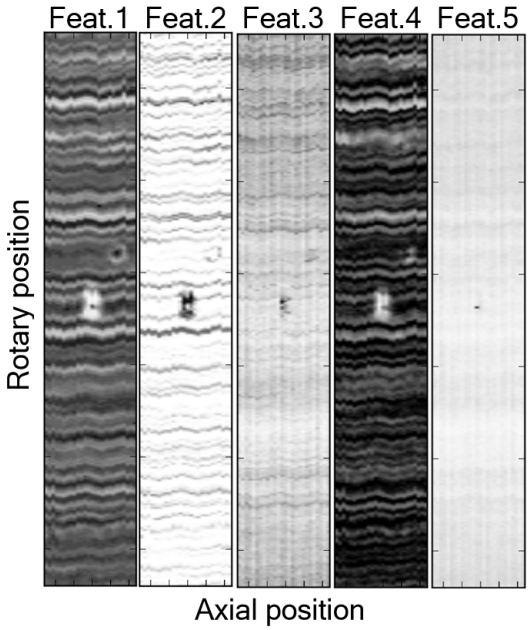
$$i_{min}(x) = argmin(x), \tag{4.6}$$

These features, although simple, are showing promising capabilities of carrying information about the condition of the pressure tube surface and potential flaws. A way of illustrating this information is a two-dimensional top view presentation (C-Scan) of the dataset where each pixel represents the physical position of a signal, and its intensity represents the value of the corresponding signal's feature (e.g. Figure 4.1).

Figure 4.1a illustrates an example where the condition of the inside surface of the pressure tube enables each feature to produce a clear view of the location and shape of the flaw. However, this is not the general case as the presence of various factors, such as foreign material or/and accumulated oxide on the surface of pressure tubes and within the flaw, can cause additional signal scattering and reflections creating noisy patterns that are more challenging to be addressed by the features (e.g. Figure 4.1b). The random behaviour of these factors poses a problem for modelling and predicting the degree to which they affect the performance of the features on each unique pressure tube environment. Therefore, the proposed methodology utilises the full set of equally weighted features to represent the signals in the unsupervised learning process.



(a) Example flagged area 1: The clean condition of the surface allows for a clearer view of the flaw.



(b) Example flagged area 2: The presence of foreign material, or/and oxide, on the surface, interferes with the signals, resulting in noisy patterns that make the distinguishing between flaw and surface more difficult.

Figure 4.1: Examples of C-Scans generated for two areas containing flaws, utilising the values of individual features.

4.4 Unsupervised Signal Analysis

4.4.1 Selection of Clustering Algorithm

The problem of grouping together unlabelled ultrasonic signals is being approached through unsupervised clustering, which acts as the backbone of the multi-step analysis presented in this chapter. Specifically, this chapter focuses on the utilisation of the machine learning algorithm called DBSCAN (Density-Based Spatial Clustering of Applications with Noise) [25]. DBSCAN clustering has proven to work in practice and has been awarded the 2014 Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD) test of time award.

In addition to DBSCAN's core advantages which are presented below, DBSCAN was selected because it is well suited for the novel subsampling/remapping process introduced on Section 4.4.3, which enables further computational improvements.

DBSCAN utilises a density-based notion of clusters where a cluster of data points is considered to be a high-density region which is separated from the other clusters by low-density regions.

This approach allows DBSCAN to detect arbitrarily shaped clusters [146], and differs from other popular clustering algorithms, such as K-Means [147] and Mean Shift [148], where the clusters are defined by their centres and causes them to underperform when processing non-spherical shaped clusters.

Another important advantage of DBSCAN within the application area of this thesis is that it operates without requiring a predefined number of expected clusters. On the other hand, algorithms such as K-Means and Gaussian Mixture Models [149] require to the number of clusters as a parameter. Although it is possible to heuristically estimate the optimal number of clusters in a dataset [150], it is not computationally efficient for the large number of datasets used in this thesis as a separate analysis would be required for each dataset. Therefore, since the proposed application is aimed at handling inspection signals, the utilised algorithm should be able to determine the different (*a priori* unknown) number of groups and not forcing the data to fit the assumptions about the dataset.

DBSCAN requires two input parameters (Eps: Maximum local radius of the neighbourhood of a point, $MinPts$: Minimum number of points in the Eps-neighbourhood of a point) that operate as a measure of minimum density level estimation and operates by iteratively distinguishing each data point as one of the following types (Figure 4.2):

- Core point: a point with a dense neighbourhood
- Border point: a point that belongs to a cluster but its neighbourhood is not dense
- Noise point: a point that does not belong to any cluster

DBSCAN calculates the density around the user-specified *Eps*-neighbourhood of each data point, and applies the $MinPts$ threshold to assign a type (i.e. core, border, noise) to that data point. Then, core points with overlapping *Eps*-neighbourhoods (called density reachable points) are joined a cluster, followed by the assignment of border points to their corresponding clusters.

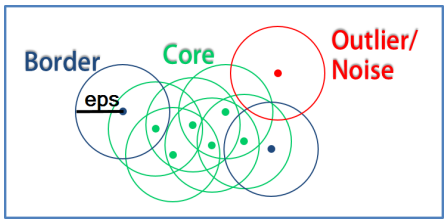


Figure 4.2: Example of DBSCAN clustering with parameters $MinPts = 3$ and $Eps = radius$

An illustration of its capabilities is presented on Figure 4.3 which shows how different popular clustering algorithms handle arbitrarily shaped clusters, as well as the 'null' dataset that represents one homogeneous cluster, with the corresponding computational efficiencies summarised on Figure 4.4. The comparison additionally includes the algorithms BIRCH [151], OPTICS [152], Agglomerative Clustering [153], Ward hierarchical clustering [154], Spectral Clustering [155, 156], and Affinity Propagation [157].

As it is presented on Figure 4.4, the computational efficiency of DBSCAN is competitive when the datasets under consideration are of moderate size ($n \leq 5\,000$). However, as the sample size scales up, DBSCAN experiences a decline in performance relative to the top performing algorithms. To illustrate the computational performances on large sample sizes, an additional comparison test was conducted on a single-cluster artificial dataset of sample size $n_{large} = 150\,000$. As evidenced

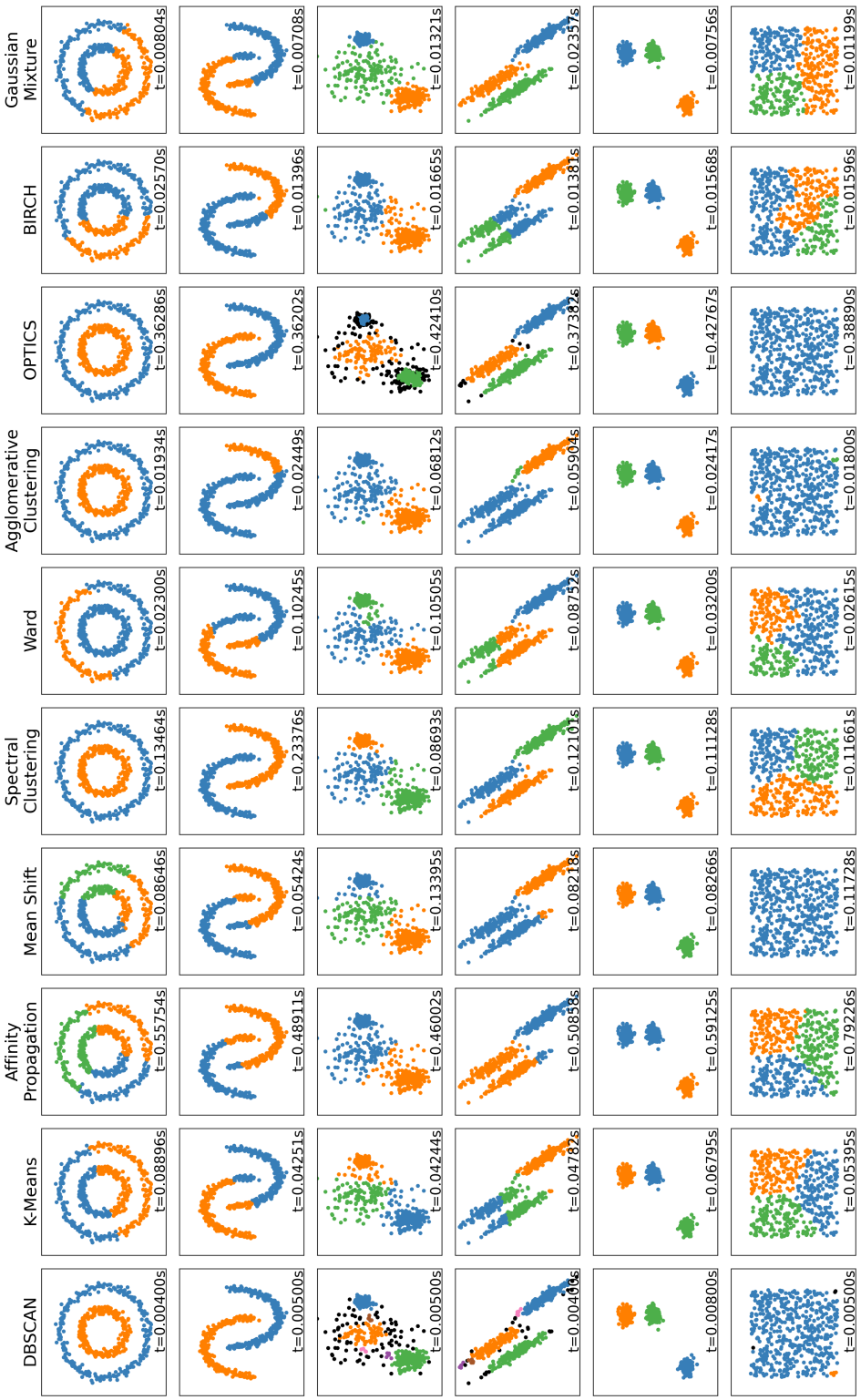
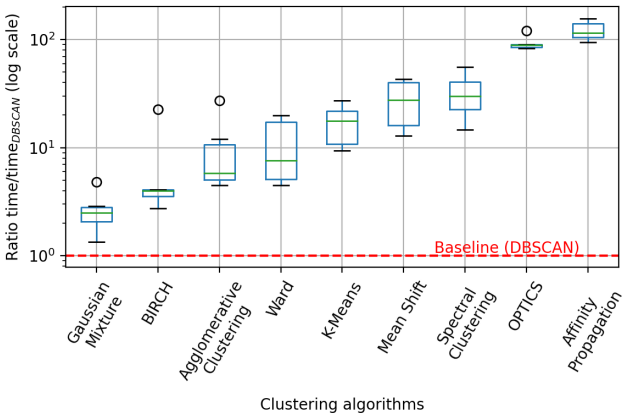
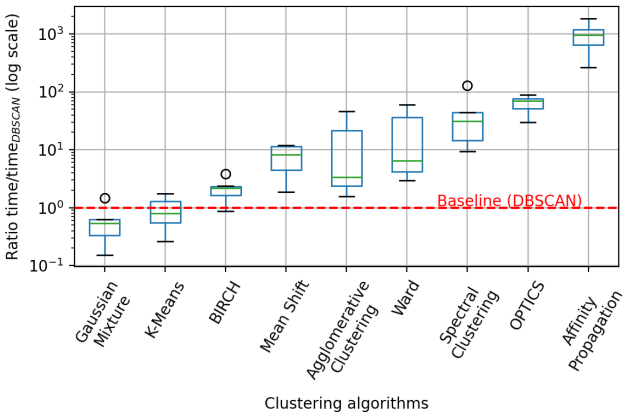


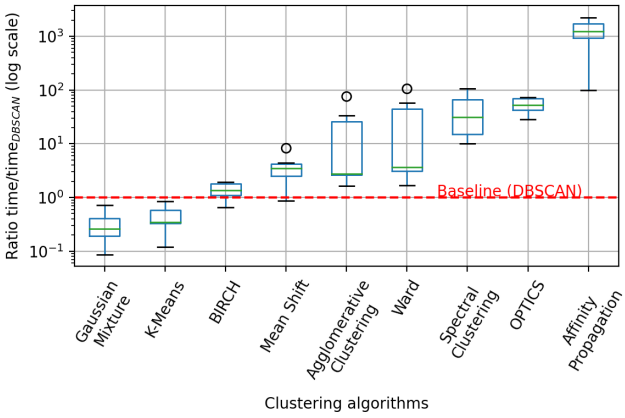
Figure 4.3: Qualitative comparison of clustering performance for popular clustering algorithms on publicly available artificial datasets [158].
The computational time comparison for this experiment is presented on Figure 4.4.



(a) $n = 500$



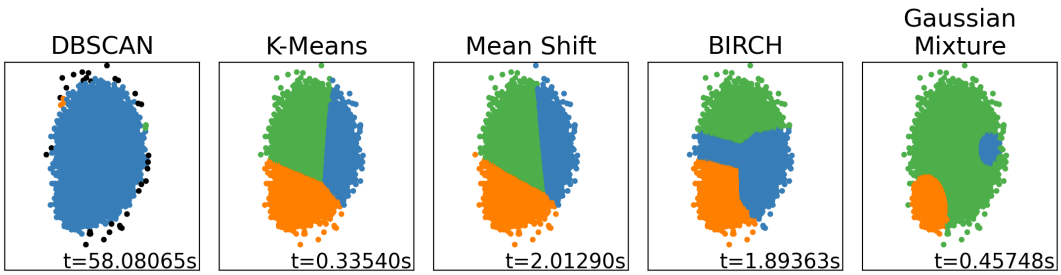
(b) $n = 2500$



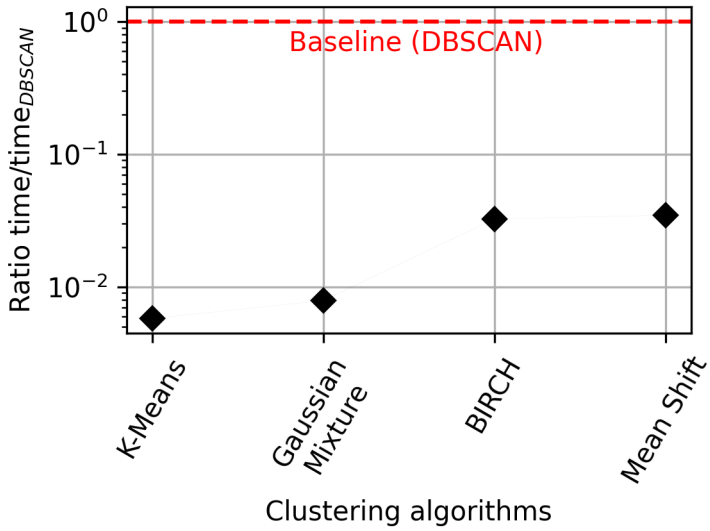
(c) $n = 5000$

Figure 4.4: Computational time comparison of popular clustering algorithms on publicly available artificial datasets [158], for different sample sizes. The computational times are compared to the baseline time of DBSCAN, showing that DBSCAN offers competitive processing times, while providing advantageous clustering characteristics shown on Figure 4.3.

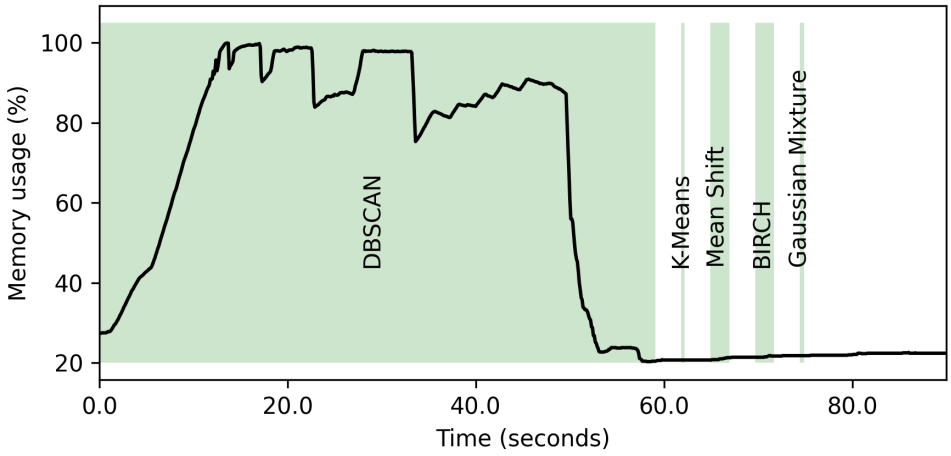
by the plots of Figure 4.5, given a larger sample size, the computational overhead of DBSCAN is significantly higher than that of the rest of the popular clustering algorithms. Therefore, it is important to derive a procedure that preserves the useful clustering characteristics of DBSCAN, while maintaining more efficient time and space complexities.



(a) Generated clusters.



(b) Computational time.



(c) Memory usage.

Figure 4.5: Comparison of popular clustering algorithms on a single-cluster artificial dataset [158], for large sample size $n_{large} = 150\,000$. The computational overhead of the DBSCAN algorithm is significantly higher compared to the other algorithms.

4.4.2 Feature Standardisation

Following the feature extraction step, each signal is represented by its corresponding set of feature values. This enables greater computational efficiency during the subsequent clustering procedure due to the inherent reduction of the search space. Preceding the clustering procedure is a pre-processing step which reformats the features onto the same scale (feature rescaling). This is required by algorithms which use distance metrics (such as the Euclidean distance) as the rescaled values of the features will ensure that features with intrinsically large numeric values will not dominate the computed distances. Two common approaches to rescaling are: *normalisation* and *standardization*.

Normalisation is often approached as a case of 'min-max' rescaling where the feature values are rescaled to fit within fixed intervals such as $[0, 1]$ or $[-1, 1]$. To map each sample feature value $x^{(i)}$ in $[0, 1]$ the following linear transformation (Equation 4.7) can be used:

$$x^{(i)'} = \frac{x^{(i)} - x_{min}}{x_{max} - x_{min}} \tag{4.7}$$

where x_{max} denotes the maximum value of the corresponding feature sample values and x_{min} the minimum value. This transformation can be extended to any interval $[a, b], a, b \in \mathbf{R}$ as follows (Equation 4.8):

$$x^{(i)'} = (b - a)\frac{x^{(i)} - x_{min}}{x_{max} - x_{min}} + a \tag{4.8}$$

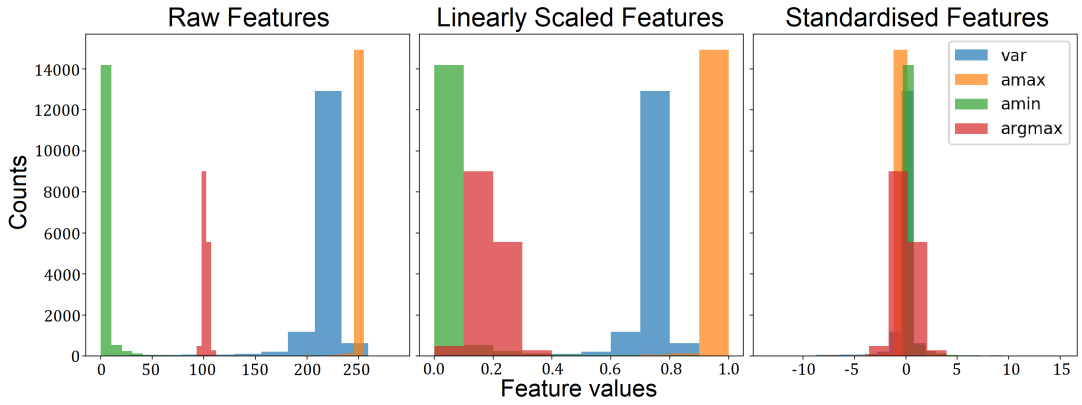
However, this approach is very sensitive to outliers [159] as the presence of a single extreme outlier is sufficient for the remaining sample values to be significantly compressed within a small range of values for the requirement of strict range boundaries to be satisfied [62].

An alternative approach to rescaling is the transformation of the feature values so that they are re-centred around zero with standard deviation 1. This procedure is called standardisation and the calculation of the new standardised value $x_{std}^{(i)}$ of a sample $x^{(i)}$ is expressed by Equation 4.9:

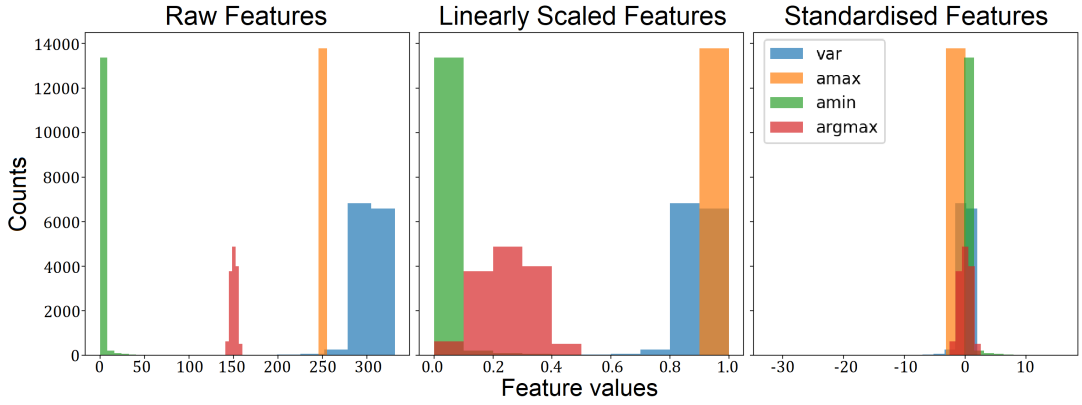
$$x_{std}^{(i)'} = \frac{x^{(i)} - \bar{x}}{\sigma} \tag{4.9}$$

where $\bar{x} = average(x)$ is the sample mean of the particular feature, and σ is the corresponding standard deviation. The effect of normalisation and standardisation have on the distribution of the feature values is illustrated in Figure 4.6.

In contrast to normalisation, standardisation does not map the feature values within a fixed interval, thus being less sensitive to outliers. Naturally, extreme values will still affect the transformed values of the main body of data, as they participate in the calculation of the sample mean and sample standard deviation. However this effect is dependant on sample size and, given the high axial and rotational scanning resolution that enables large number of samples per scanned area, it is expected to be less significant than when performing minmax normalisation where even the most extreme sample values are forced within a fixed interval.



(a) Example dataset 1



(b) Example dataset 2

Figure 4.6: Examples of feature values (variance, maximum, minimum, position of maximum) distributions for different datasets obtained by the Normal Beam probe. The left column presents the distributions of the raw feature values, the middle column shows the effect of linearly rescaling the feature values within a fixed range (here $[0, 1]$), and the right column shows the effect of standardising the feature values to a distribution of sample mean $\bar{x} = 0$ and standard deviation $\sigma = 1$.

The significance of outlier sensitivity of the feature rescaling techniques for this body of work lies on the fact that the detection of samples that are 'abnormally' distant from the main body of samples (assumed to represent the 'healthy' part of the scanned pressure tube surface) constitutes its main objective. As the methods of this research do not assume *a priori* knowledge of the boundaries between 'healthy' and 'unhealthy' subspaces of the feature space, but instead explore the inter-sample Euclidean distances (distance matrix), it is important that the relative expected difference between 'healthy'-to-'healthy' sample distances and 'healthy'-to-'unhealthy' sample distances does not fluctuate significantly between inspection datasets of different pressure tube areas or different pressure tubes. This allows for a narrower range of clustering parameter values that are relevant across different datasets and enable consistent grouping of the main body of 'healthy' samples present within the datasets.

The following comparative test (Figure 4.7) is designed to examine the consistency of clustering results across different datasets depending on the feature rescaling method used as a pre-processing step. This process starts with manually selecting sets of clustering parameters that generate similar results (Table 4.1) on a small sample of inspection datasets for both feature rescaling methods.

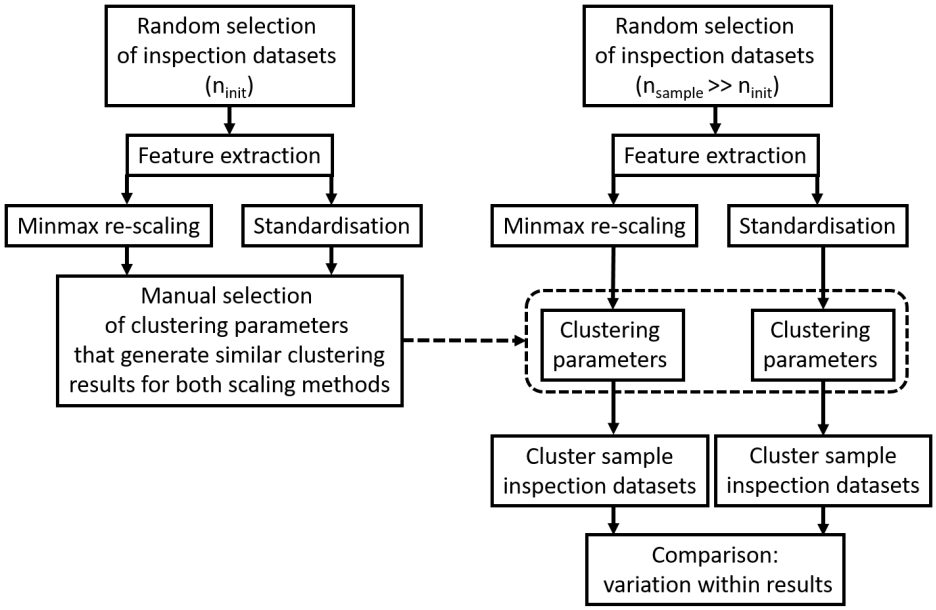


Figure 4.7: Process for generating comparative clustering results from two rescaling methods for variation comparison.

Table 4.1: Clustering parameters used for the two feature rescaling techniques and coverage of the largest clusters.

minimum samples	radius minmax	coverage minmax	radius standardisation	coverage standardisation
3	0.08045	0.984	1	0.983
6	0.055	0.992	0.7	0.991
9	0.04463	0.973	0.6	0.974

The manual selection of clustering parameters was performed by selecting a fixed value for the DBSCAN parameter related to minimum number of neighbours (integer data type) for a sample to be considered as a core point and adjusting

the neighbourhood radius (float data type) parameter until the coverages[1] of the largest clusters generated from normalised and standardised features were matched up to the second decimal.

Each of the fixed (three) sets of clustering parameters was generated using a different randomly selected dataset, and these manually selected sets of clustering parameters were applied to grouping together samples within $n_{sample} = 500$ randomly selected datasets.

Figure 4.8 illustrates the variability of the ratio between the sample population of the largest group identified within each dataset and the total sample population of the corresponding dataset, for different clustering parameter values (DBSCAN parameter expressing the minimum number of neighbours for a sample to be considered as a core point) and the two rescaling methods (minmax for the interval $[0, 1]$, and standardisation by re-centering such as $\bar{x} = 0$ and $\sigma = 1$). It can be observed that standardisation offers transformed values that enable the clustering algorithm to sustain smaller (compared to minmax) variability within its results when applied to different datasets using the same clustering parameter values. This becomes more apparent (Figure 4.9) for larger 'minimum samples' values that utilise smaller radius values that decrease the clustering algorithm's 'reach' resulting in greater variability within the clustering results.

[1]Coverage refers to the ratio of number of samples belonging to the largest cluster, versus the total number of samples.

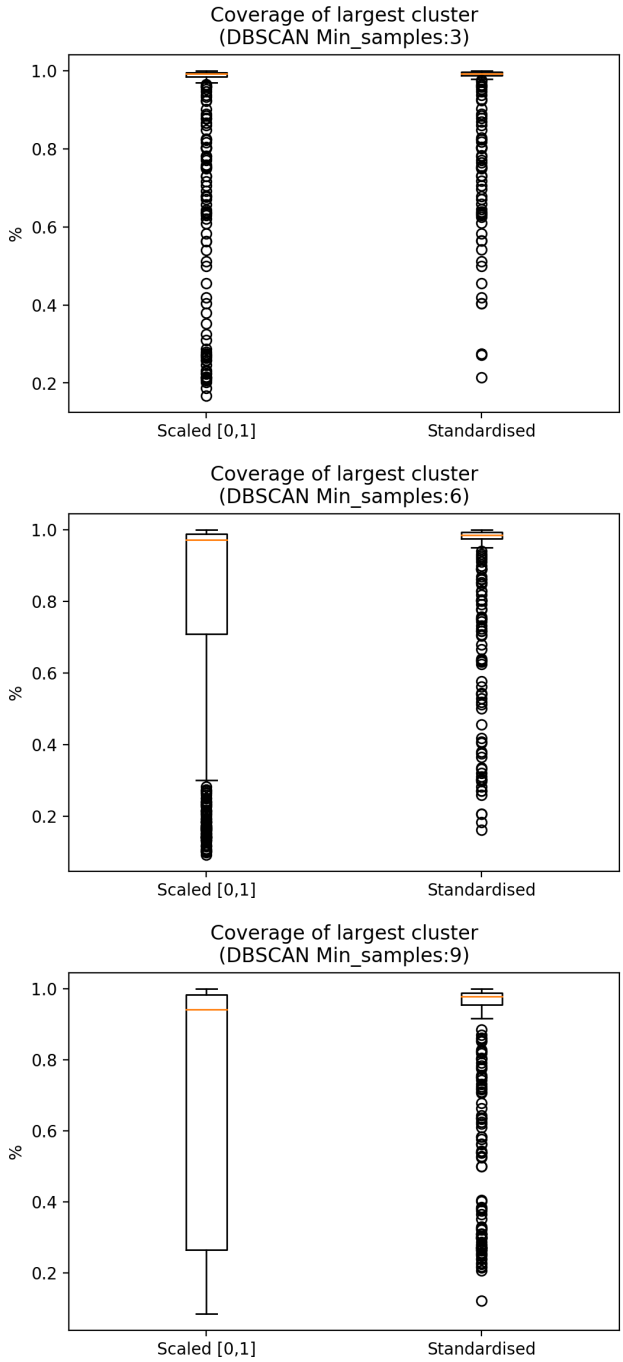


Figure 4.8: Boxplots illustrating the variabilities of the ratio between the sample population of the largest group identified within each dataset and the total sample population of the corresponding dataset, for different clustering values and the two rescaling methods. The plots consist of the following elements: the box illustrates the interquartile range (IQR), the horizontal coloured line represents the median, the whiskers extend up to 1.5∗(IQR) from the corresponding quartile, and the points outside the area defined by whisker plot ends are considered as outliers.

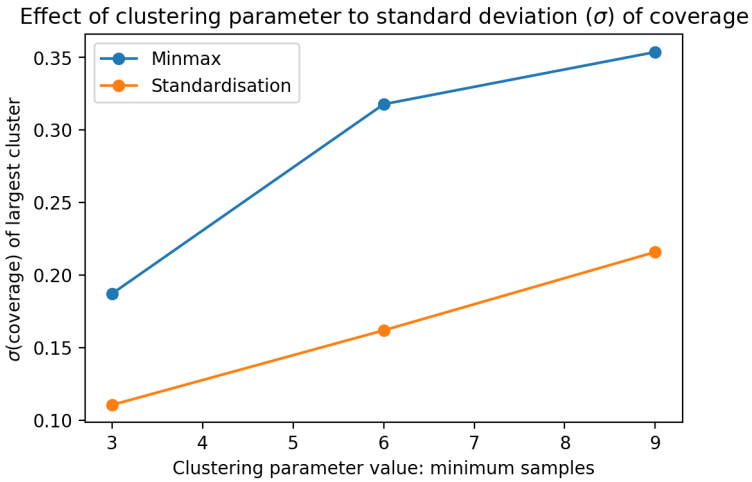


Figure 4.9: For a fixed neighbourhood radius value, by increasing the 'minimum samples' values during the manual parameter selection process, smaller clustering radius values were required to maintain similar initial clustering coverages, reducing the clustering algorithm's 'reach' and generating more variable results.

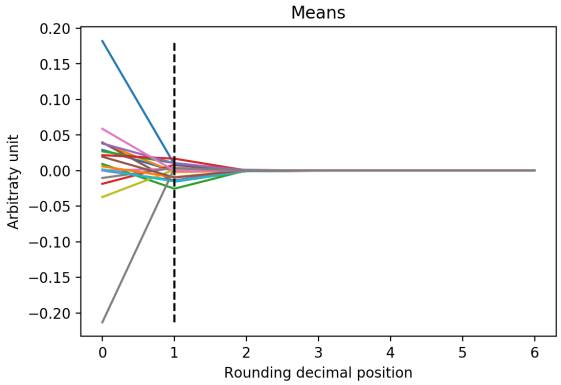
4.4.3 Subsampling DBSCAN Wrapper

Prior to introducing and applying the two-stage clustering procedure, a novel subsampling method, is presented that wraps DBSCAN and reduces the computational effort of the clustering procedure. It is effective for large volumes of low-cardinality input data containing multiple repeating unique sets of feature values. This method subsamples a representative input subset which participates in the clustering procedure. The generated clustering labels can then be re-mapped back to the original input data.

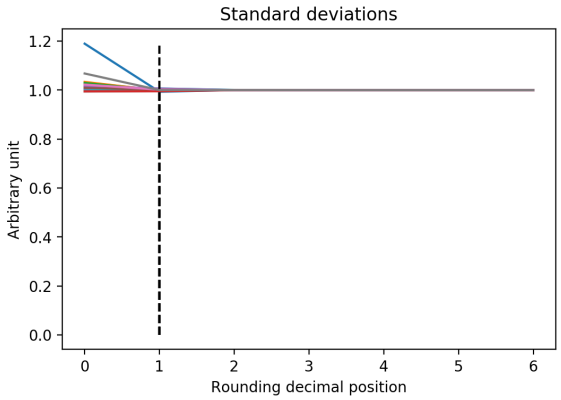
4.4.3.1 Rounding Step

The initial step of this procedure is based on the assumption that the transformation of the feature values during the rescaling step introduced insignificant decimal digits that can be viewed as a source of false precision. Therefore, there should exist decimal positions that the transformed values can be rounded up to without significantly altering the properties of the underlying distribution.

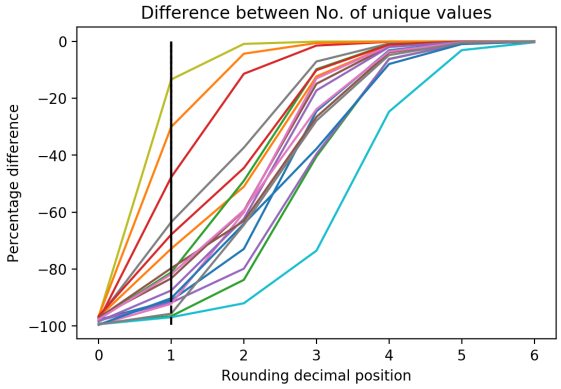
To identify potential appropriate decimal positions, $n = 20$ sample datasets were randomly selected and their standardised feature values were rounded up to different test decimal positions and their new distributions were compared to the original distribution of the scaled feature values ($\bar{x} = 0$, $\sigma = 1$). Figures 4.10a and 4.10b show that the values of the observed means and standard deviations of the test datasets' rounded feature values for the decimal positions 0 (*integer*), 1, 2, 3, 4, 5, and 6, do not deviate significantly from $\bar{x} = 0$, $\sigma = 1$. Furthermore, given the small number of features, by reducing the available decimal points within the feature values the number of unique sets of feature values is also significantly reduced. Figure 4.10c illustrates the percentage reduction between the initial number of unique sets of scaled feature values and the number of unique sets of scaled feature values after the rounding.



(a) Observed means of the test datasets' rounded scaled feature values.



(b) Observed standard deviations of the test datasets' rounded scaled feature values.



(c) By reducing the available decimal points, and given the small number of features, significantly fewer unique sets of feature values are available.

Figure 4.10: Characteristics of the scaled feature values that have been rounded up to different decimal positions.

4.4.3.2 Representative Participating Subsamples

The proposed approach of reducing the computational effort of the clustering procedure is taking advantage of the decreased number of unique sets of values produced by the rounding step and is supported by the fact that one of the requirements for DBSCAN to consider a data point as a core point is that the number of data points within its neighbourhood must be greater or equal to a defined minimum number.

The decreased number of unique sets of values could result in increased counts of duplicated members that are uniquely described by these feature values. Thus, by involving in the computation only enough representative members to satisfy the 'minimum number of samples' parameter of DBSCAN, the clustering results should be sufficiently similar to the clustering results involving all the available samples, while the computational effort could potentially be decreased.

Figure 4.11 illustrates the steps involved in utilising the minimum representatives subsampling method for clustering a feature space with the DBSCAN clustering algorithm. The process splits the rounded feature values into two groups; a non-participating subset and a participating subset, the members of which act as representatives of the non-participating members. The clustering step uses standard DBSCAN Python libraries [158] and requires no alterations to the existing implementation. Following the grouping of the participating feature subset the labels are then mapped to the corresponding non-participating members.

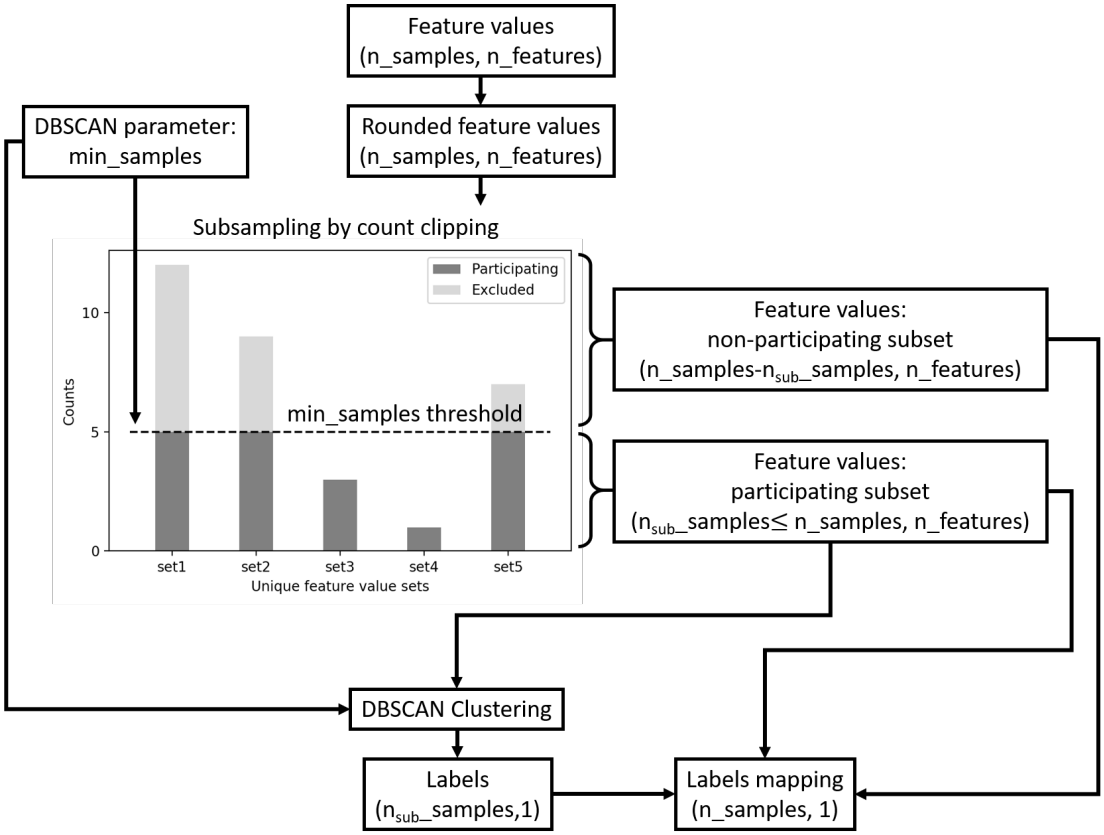


Figure 4.11: Flowchart of subsampling method.

4.4.3.3 Assertions Testing

This section is concerned with demonstrating that the proposed subsampling process is both correct and useful. This is approached by testing two important assertions:

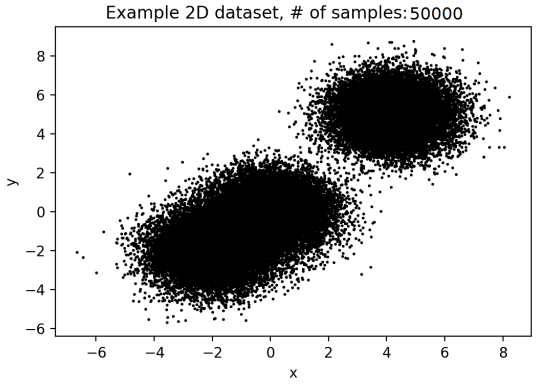
(i) The process of assigning core points by DBSCAN is deterministic and order invariant, i.e. given the same clustering parameters, a cluster will always have the same core points regardless of the order the samples are processed. However, the cluster labelling of a border point approachable by

more than one clusters is non-deterministic and will get assigned the label of the first cluster to approach it. Therefore, if the random state of the involved computational processes is fixed, the groups obtained when applying the subsampling/mapping pre/post-processing step should be identical to the groups generated when DBSCAN is directly applied to the feature space.

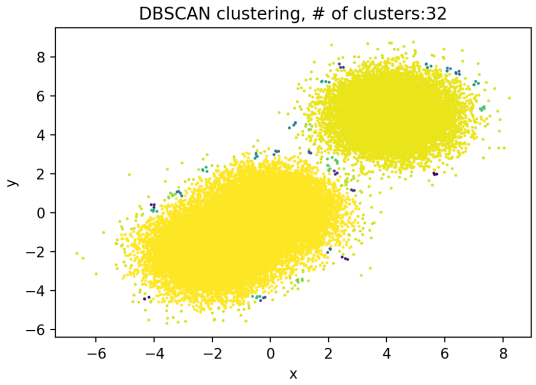
(ii) The computation time of the pre/post-processing step, along with the clustering step, should be less or equal to the clustering time of the original feature space. It is expected that the computational benefit will be greater for larger differences between the size of the set of unique sample values against the sample size of the dataset. This also implies that when every sample of the dataset is unique there will be no computational benefit, i.e. the method will perform exactly as a vanilla DBSCAN.

To compare the clustering operations between the two methods, two-dimensional datasets were generated as test examples. The datasets consist of randomly generated neighbourhoods of points with some level of overlap (illustrated at Figure 4.12a. The datasets were processed with and without subsampling using identical DBSCAN parameter values which were selected such that they allow for more than one cluster.

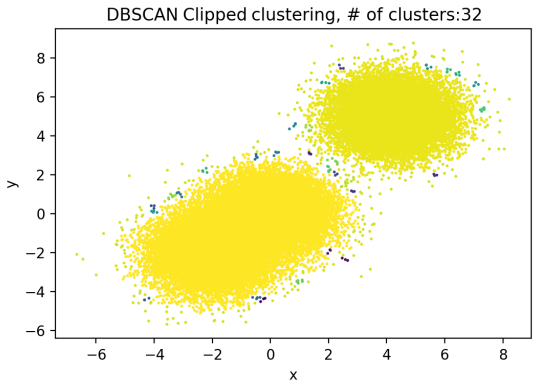
To enable the comparison of the groups, since the order of DBSCAN groups is not deterministic, the resulted groups were sorted according to number of group members. The comparison showed that the groups are exactly identical (Figures 4.12b and 4.12c).



(a) Generated example dataset.



(b) Clustered with DBSCAN.



(c) Clustered with DBSCAN using the proposed subsampling method (DBSCAN Clipped).

Figure 4.12: Scatter plots of randomly generated example dataset. The clustering results of DBSCAN combined with the subsampling method are identical to the original DBSCAN clustering results. The clusters' labelling is reordered by sample size.

To explore potential computational time benefits of subsampling, the two clustering approaches were timed for generated 2D datasets of various sample populations and decimal point accuracies. Figure 4.13 illustrates that the benefits become more evident as the population increases and the decimal point accuracy decreases, i.e. when the ratio of number of unique sets of feature values over the size of population decreases.

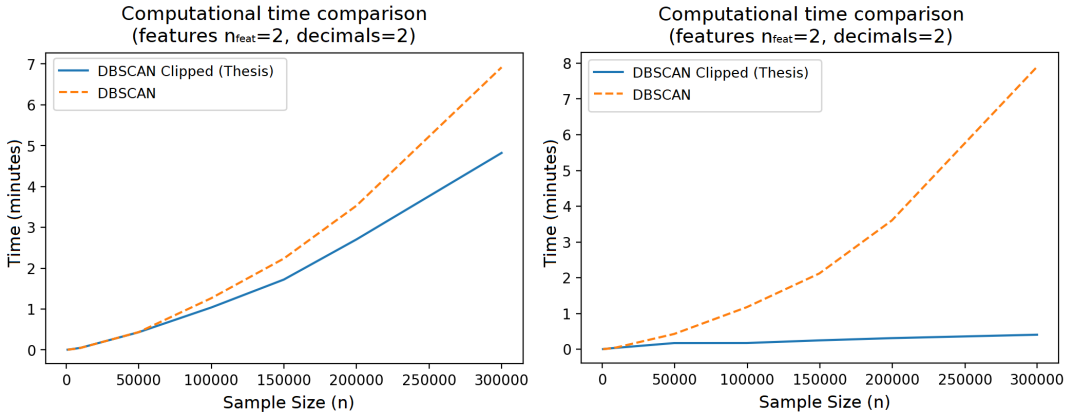
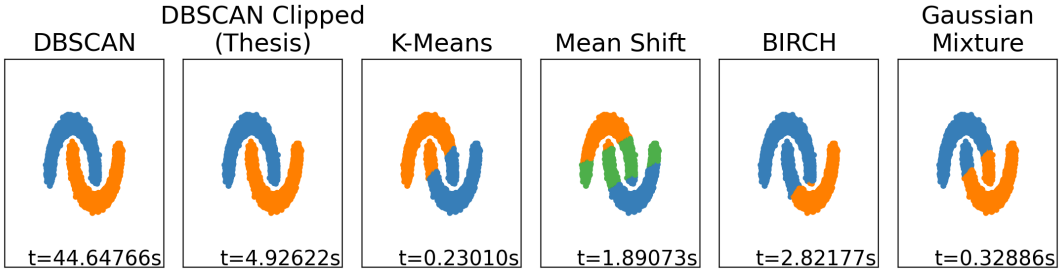
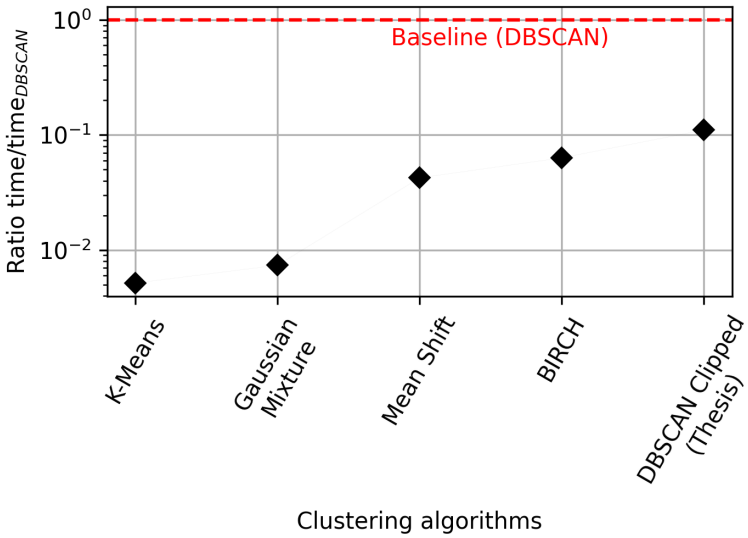


Figure 4.13: Computational time comparison between DBSCAN clustering and DBSCAN clustering wrapped with the subsampling method (labelled here as 'DBSCAN Clipped'). The benefits become more evident as the population increases and by decreasing the decimal point accuracy, i.e. when the ratio of number of unique sets of feature values over the size of population decreases.

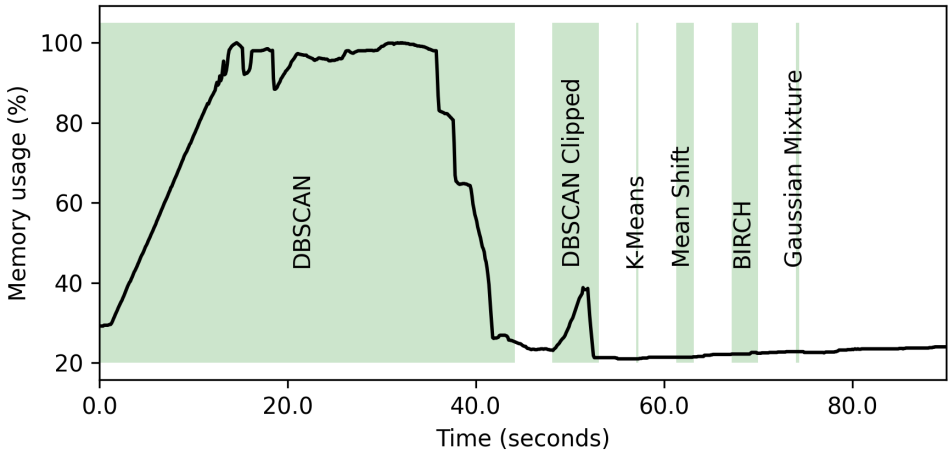
To further demonstrate the efficiency of the proposed method, Figures 4.14 and 4.15 summarise its performance relative to DBSCAN and the algorithms that participated in the large sample size comparison of Figure 4.5. The results indicate that the proposed method exhibits competitive efficiency, while maintaining the useful clustering characteristics of the DBSCAN algorithm.



(a) Generated clusters.

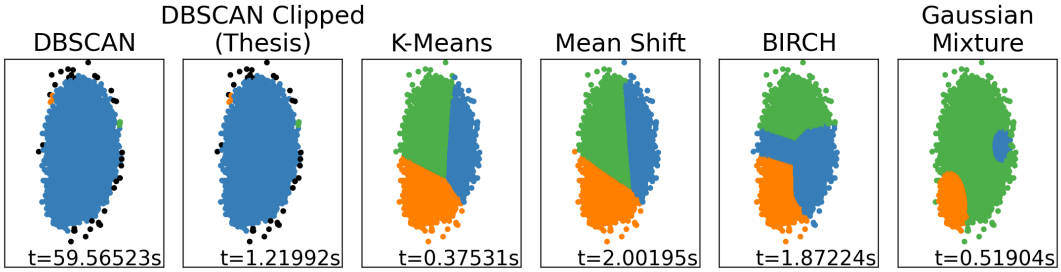


(b) Computational time.

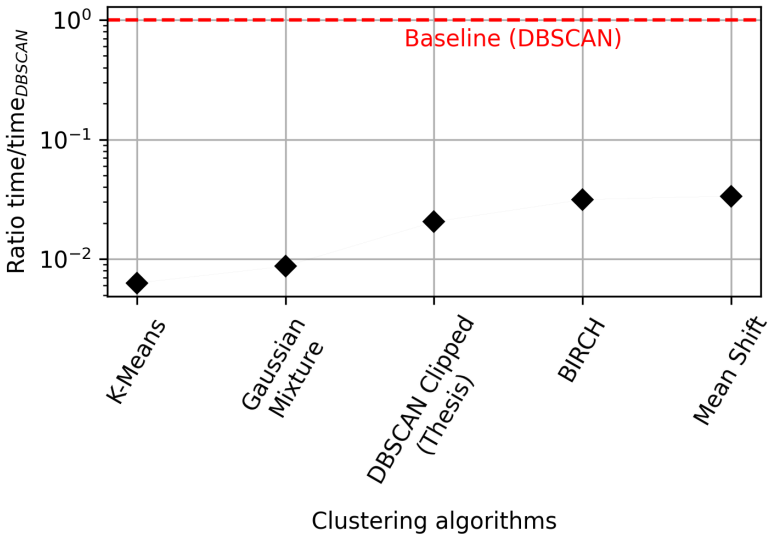


(c) Memory usage.

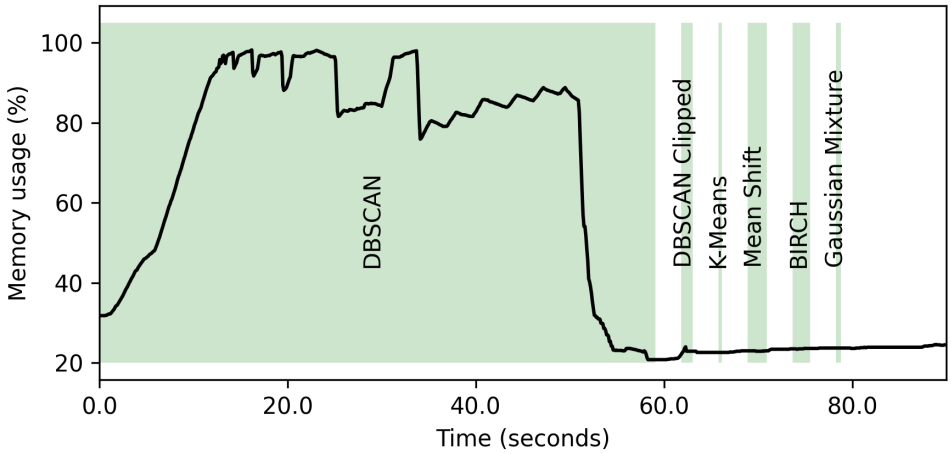
Figure 4.14: Comparison of DBSCAN clustering wrapped with the subsampling method ('DBSCAN Clipped'), DBSCAN and other popular clustering algorithms on a 2-cluster artificial dataset [158], for large sample size of $n_{large} = 150\,000$ rounded at 2 decimal places.



(a) Generated clusters.



(b) Computational time comparison.



(c) Memory usage.

Figure 4.15: Comparison of DBSCAN clustering wrapped with the subsampling method ('DBSCAN Clipped'), DBSCAN and other popular clustering algorithms on a single-cluster artificial dataset [158], for large sample size $n_{large} = 150\,000$, rounded at 1 decimal place.

Finally, the two methods were compared for a sample of B-Scan datasets ($n = 73 \approx 10\%$ of total datasets, each containing data from three probes) to examine the behaviour under more noisy and multi-dimensional datasets. The comparison was repeated for three sets of DBSCAN parameters that were empirically evaluated to provide different levels of 'rigidity'; from 'strict' neighbourhood requirements that result into more clusters and more data points classified as noise, to more meaningful and 'inclusive' neighbourhoods that tend to generate a single major cluster followed by a small number of smaller clusters and noise.

The usage of subsampling consistently decreased the total computational time required by the clustering process and the difference was more pronounced for parameter values that allowed for larger clusters (Table 4.2). The clustering results were identical for all (657 =(73 datasets)×(3 probes)×(3 tested parameters)) the tested cases.

Table 4.2: Comparison of computational time required for clustering sample B-Scan datasets ($n = 73$, each containing data from three probes) with and without subsampling.

Radius DBSCAN	Time (mins) with subsampling	Time (mins) no subsampling	Time difference (%)
0.2	9.583	11.78	-19%
0.6	14.58	18.75	-22%
1.0	12.43	20.35	-39%

4.4.4 Two-Stage Clustering

Having established the pre-processing approach, this section introduces the two-stage clustering process that takes as input the standardised features and outputs the final segmentation of the data. The process separates the feature values from their spatial information and each participates in separated linked clustering stages. Following is a description of the stages and a summary of the benefits that arise from the separation of the clustering process into two stages.

The first stage involves an exploration of the feature space and aims at reducing the size of the sample by eliminating the majority of "healthy" signals, and isolate signals of ambiguous nature. The output of this clustering is a label attached to each unique feature set, indicating the class/group that it belongs. This stage is the most computational intensive of the two stages, as it involves multiple feature values of the entire population of the dataset. However, it utilises the subsampling method which, as is shown in Section 4.4.3, has the potential to decrease the computational time.

Given the unbalanced nature of the datasets, i.e. the healthy section of the tested area is larger than the flawed section, an assumption is made that the group with the larger number of members belongs to the healthy section, thus the remaining groups contain potentially flawed areas or noise. Therefore, only the smaller groups of samples will proceed to the next clustering stage.

The samples that proceed at the second stage of the clustering procedure are being represented only by their location properties, i.e. *axial position* and *rotary position*, and are all considered as potentially "flawed" signals. The dual purpose of this stage is to:

- Eliminate outliers/noise by identifying samples that are not members of the "healthy" group and are located in a sparse neighbourhood
- Group together samples that are not members of the "healthy" group and are located in a dense neighbourhood

This step is essential for formulating the final group comprising of samples which are members of the flaw. It provides the means to group together signals that although obtained from a flawed region, might not be initially grouped together; a consequence of the diverse characteristics that constitute a flaw and their different impact on the reflected signals. Therefore, the location-based clustering is aiming at re-grouping those signals together, utilising the fact they should be located in close proximity.

The output of the location-based clustering is the updated signal labels: Samples that belong to a group are classified as part of the flawed region, and samples that were labelled as noise/outliers are discarded as being members of the healthy region. Figure 4.16 provides a high-level flowchart that summarizes the steps taken by the clustering procedure. This process can be applied to the different datasets obtained by each of the ultrasonic probes (APC, CPC and 20 MHz NB).

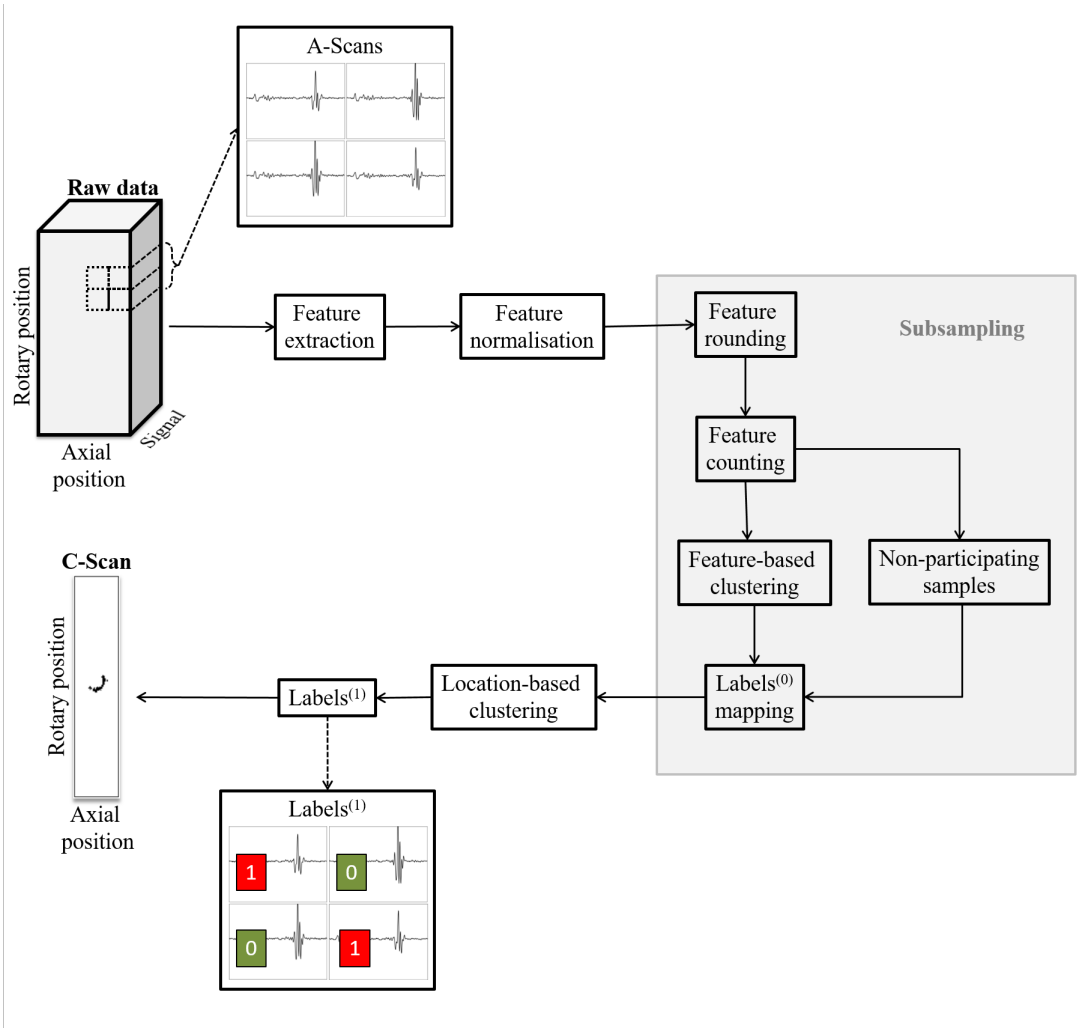


Figure 4.16: Flowchart of the clustering procedure.

The main benefit of not including the location indices within the initial clustering and instead utilising a two-stage clustering approach is decreased computational time due to:

- Decreased search space during the most computationally demanding step as two uniformly distributed feature vectors (location indices) are omitted. Instead, the location indices are participating only in the second clustering stage, where, by design, the number of participating members should be less to the number of the samples contained within the original dataset.
- Facilitation for the existence of multiple data points that share the same rounded feature values. This enables the subsampling method (Section 4.4.3) to offer potentially decreased computational time as some of the repeated data points will be omitted from the clustering procedure. This would never be the case if the location indices are included, as they are strictly not repeating combination of values.

4.4.5 DBSCAN Parameter Selection

As mentioned earlier (Section 4.4.1) the DBSCAN clustering algorithm requires two initial input parameter values, the local cluster radius and the minimum members required to form a cluster, which are referred to as *Eps* and *MinPts* respectively, and both have a significant influence on the clustering results [160].

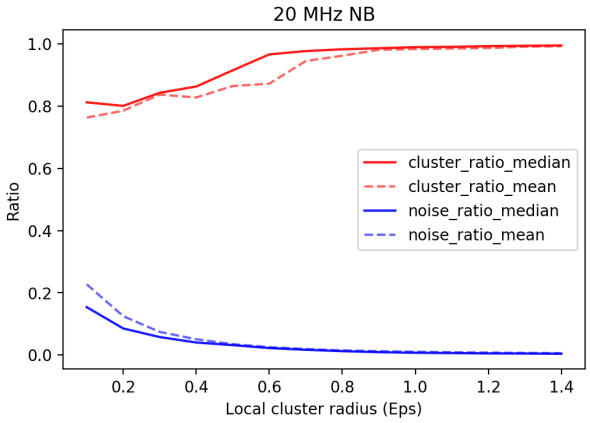
For the purposes of this work, during the first stage of the clustering procedure (feature-based) the minimum members parameter was kept constant at a low $MinPts = 3$ while a range of radius (Eps) values were applied across a sample of datasets. For each clustering outcome two quantities were measured to assist the evaluation of the parameters' suitability:

1. Ratio of sample size classified as noise over the population size of the dataset.

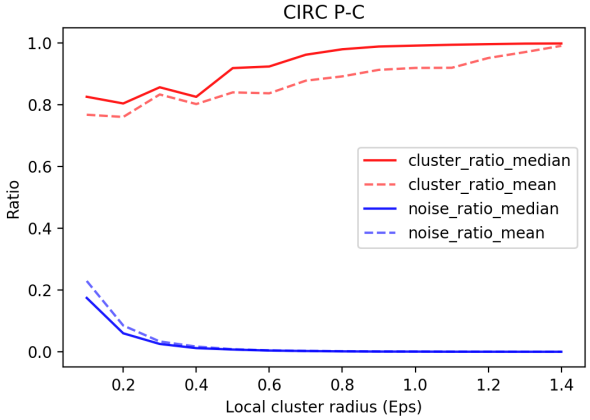
2. Ratio of sample size of the largest cluster over the population size of the dataset.

The purpose of this heuristic approach is to provide an estimation of the effect that the local radius has on the cluster formation on the type of inspection data that this work investigates, without utilising explicit knowledge of the flaw characteristics contained within each dataset.

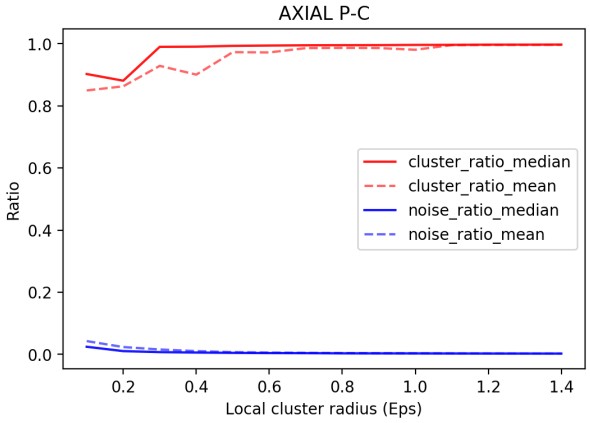
Figures 4.17 illustrate the extracted median and mean of the calculated quantities for each of the sampled ultrasonic probes. These elbow plots are used as guides for estimating minimum local cluster radius values that potentially balance between 'too strict' radius requirements that tend to discard many data points as 'noise', and very 'inclusive' parameters that tend to capture all of the data points into one cluster.



(a) Flaw 1



(b) Flaw 2



(c) Flaw 2

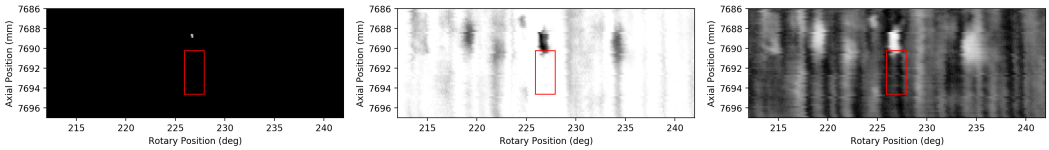
Figure 4.17: Median and mean of the ratios $\frac{n_{\text{noise}}}{n_{\text{population}}}$ and $\frac{n_{\text{largest cluster}}}{n_{\text{population}}}$ of the sample ultrasonic probes. These elbow plots are used as guides for estimating minimum local cluster radius values that potentially balance between 'too strict' radius requirements that tend to discard many data points as 'noise', and very 'inclusive' parameters that tend to capture all of the data points into one cluster.

As currently there is no commonly accepted automated method for *a priori* selection of DBSCAN parameters, this work's selection method is utilising soft assumptions related to the given type of inspection data. These assumptions include that the captured inspected area contains mainly 'healthy' regions and that some of the datasets will not present informative feature values regarding the underlying flawed area as not all types of probes are suitable for all types of flaws.

Based on the aforementioned assumptions and requirements the initial *Eps* values selected for the NB probe (Figure 4.17a) and APC probe (Figure 4.17c) are 0.9 and 0.5 respectively.

However, Figure 4.17b shows a long discrepancy between the mean and the median of the largest cluster ratio, indicating the presence of additional significantly large clusters that persist the merging into the main cluster. Upon visual inspection of the abnormal results it has been identified that in the case of CPC, the feature *argmin*, i.e. the position of the minimum amplitude, can be significantly sensitive to certain surface texture, resulting in the generation of large false positive clusters (an example case is illustrated at Figure 4.18). These resulted in the feature being excluded from the feature space of CPC for any further testing.

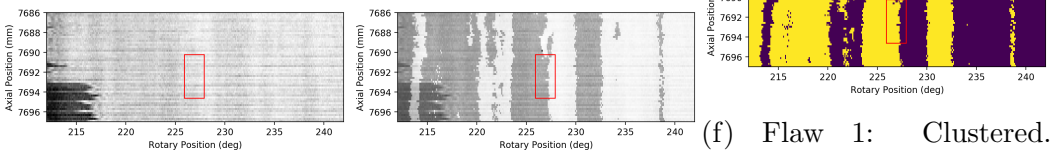
Figure 4.19 plots the mean and median of the largest cluster ratio and noise ratio that were re-generated for the CPC datasets excluding the feature *argmin*, showing faster convergence. The *Eps* value selected for the CPC datasets is 0.6.



(a) Flaw 1: Maximum.

(b) Flaw 1: Minimum.

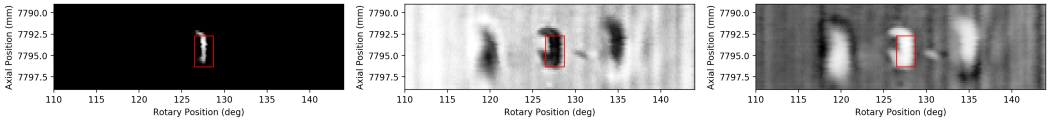
(c) Flaw 1: Variance.



(d) Flaw 1: Position of maximum.

(e) Flaw 1: Position of mininum.

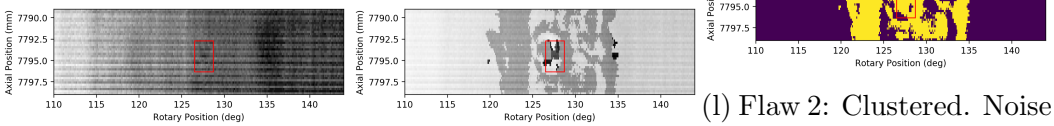
(f) Flaw 1: Clustered. Noise patterns match the patterns showing in the C-Scan of the position of minimum (Figure 4.18e).



(g) Flaw 2: Maximum.

(h) Flaw 2: Minimum.

(i) Flaw 2: Variance.



(j) Flaw 2: Position of maximum.

(k) Flaw 2: Position of minimum.

(l) Flaw 2: Clustered. Noise patterns match the patterns showing in the C-Scan of the position of minimum (Figure 4.18k).

Figure 4.18: C-Scans of features and processed segmented binary labels (bright: anomalous area, dark: healthy area) from two example CPC datasets exhibiting large false positive areas. The patterns indicate that the position of minimum of CPC A-Scans might be inappropriate as a feature as it can be sensitive to surface conditions that are not considered as 'flawed'. The red box shows the expected position and size of the flaws based on analysts' reports.

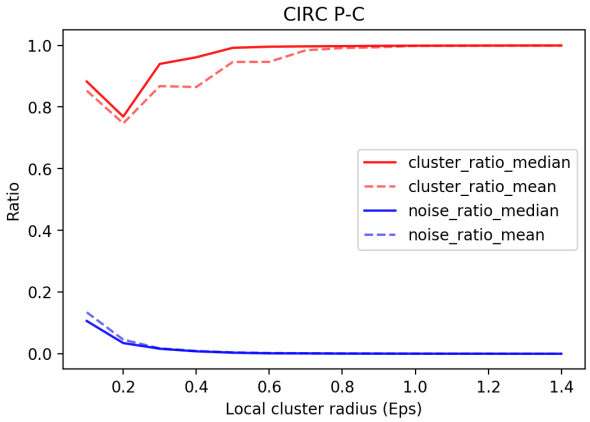
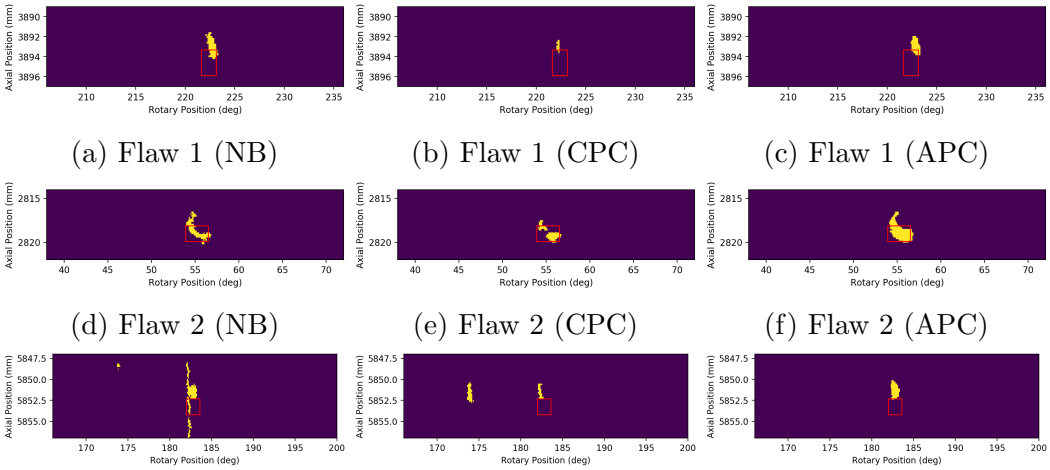


Figure 4.19: The mean and median of the largest cluster ratio and noise ratio that were re-generated for the sample CPC datasets excluding the feature *argmin*, showing faster convergence. The *Eps* value selected for the CPC datasets is 0.6.

4.5 Results

The proposed clustering methodology was applied on 766 datasets, each containing data from 3 ultrasonic probes (NB, CPC and APC) utilising the parameter values determined in Section 4.4.5. The task was completed in 2 hours and 39 minutes, on an Intel(R) Core(TM) i5-8400 CPU 2808 Mhz 6 Cores processor, with 32GB of RAM.

Figure 4.20 presents examples of resulted C-Scans for three different datasets and for each ultrasonic probe. Each pixel represents the label assigned to the corresponding A-Scan: "bright" represents signals classified as "healthy" and the *dark* represents "flawed" signals. The red boxes represent areas where expert analysts have verified the presence of a flaw. Small deviations from the exact



(a) Flaw 1 (NB) (b) Flaw 1 (CPC) (c) Flaw 1 (APC)

(d) Flaw 2 (NB) (e) Flaw 2 (CPC) (f) Flaw 2 (APC)

(g) Flaw 3 (NB) (h) Flaw 3 (CPC) (i) Flaw 3 (APC)

Figure 4.20: Examples of C-Scans resulted from the A-Scan clustering procedure (bright: classified as anomalous areas, dark: classified as healthy areas).

position are expected as there can be misalignment between the primary tool (GH scans) used by analysts to perform the initial flaw analysis and the tool extracting the detailed A-Scans.

Figures 4.20g and 4.20h further include examples of false positives, where locations of the healthy section are indicated as "flawed". This is a result of the different conditions of the pressure tube and the area surrounding the flaw, affecting heavily the propagating signals. However, the diversity of the probe configuration should enable future work to eliminate the noisy clusters by fusing the results of the different probes.

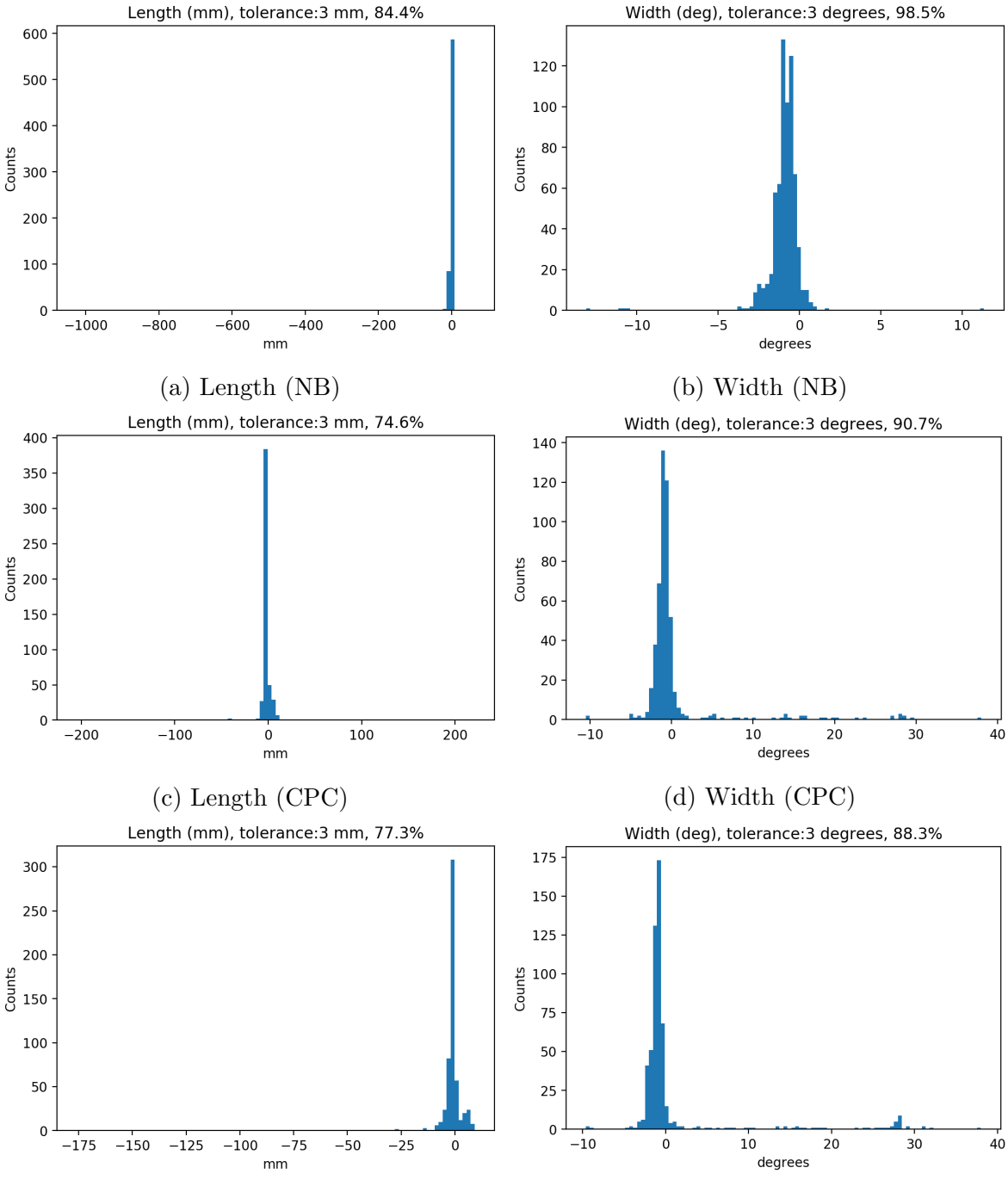
In total, the examined datasets contained 809 flawed areas, of which 93% were detected by the proposed clustering procedure within at least one of the three different probe datasets. Table 4.3 provides an overview of the detection rates achieved by each probe individually. The fact that the individual detection rates

for each probe are lower than the overall detection rate shows that the probes can supplement each other during the flaw detection task. As it was discussed on Section (Inspection Setup) the orientation of the probes is such that any point of the pressure tube's internal diameter can be inspected both directly from above (NB probe) and below through full skip propagation (CPC and APC). The longer travel paths through the material of the pitch-catch beams (APC and CPC) suggest that they are expected to carry echoes with lower SNR, resulting into poorer results, compared to NB, if all flaws are taken into account. Nevertheless, both the experience of the ultrasonic inspection experts, and the results of the data-driven analysis, confirm that pitch-catch beams enhance the overall flaw detection rate as they complement the NB probe on cases where straight normal beam inspection of a flaw is not feasible.

Table 4.3: Detection rates per probe.

Probe	Detected	Percentage
NB	680	84%
CPC	507	63%
APC	564	70%
Total detected: (out of 809)		93%

Figures 4.21 show the distribution of sizing differences between the flagged areas and the sizes measured by expert analysts. The performance regarding the 'length' sizing is lower than the 'width' sizing performance as it involves the sizing of flaw tails which is a heavily subjective task.



(a) Length (NB)

(b) Width (NB)

(c) Length (CPC)

(d) Width (CPC)

(e) Length (APC)

(f) Width (APC)

Figure 4.21: Sizing results (size(flag)-size(verified)).

However, although the 'width' sizing performance is satisfying, the accumulation of outliers near the 30 degrees position of the Figures 4.21d and 4.21f indicate the existence of a factor that contributes towards the flagging of large areas that overlap with the position of the flaw.

To further examine the quality of the produced flags each processed dataset was examined for flags that did not overlap with the verified flaw. Table 4.4 summarises the false positive rates and shows that often additional flagged areas appear alongside the main flag.

Table 4.4: Percentage of datasets containing additional (possible false positives) flags.

Probe	Additional flags
NB	40%
CPC	35%
APC	22%

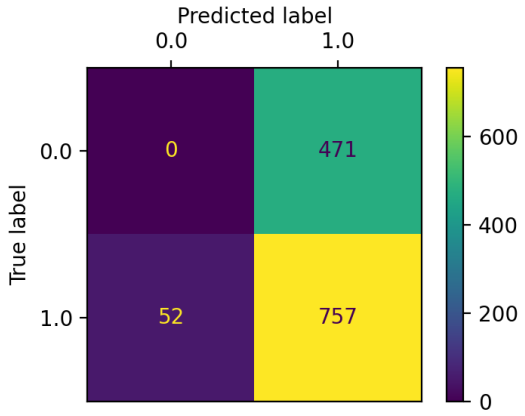


Figure 4.22: Confusion matrix showing the number of True Positives, i.e. detected flaws, False Positives, i.e. additional flags that are possibly not flaws, and False Negatives, i.e. undetected flaws. Due to the nature of the problem, and the approach, there is no notion of True Negatives.

To decrease the effects of false positives, each dataset was re-processed with increased local radius *Eps* values for each probe: *Eps*(NB)= 1.0, *Eps*(CPC)= 0.8 and *Eps*(APC)= 0.8.

Following are the updated results for local radius parameter value. Table 4.5 shows that detection rates have dropped to 90% as a consequence of eliminating some false positive flags (Table 4.6) due to the decreased requirements (i.e. increased local radius parameter) for incorporating samples into the main ('healthy') cluster. Here, the term false positive indicates a large flagged area that happened to overlap with a verified flaw.

Figures 4.24d and 4.24f show that the accumulation of outliers near the 30 degrees position has now been eliminated and only scattered low-count outliers are present.

Table 4.5: Detection rates per probe, after increasing the local radius parameter.

Probe	Detected	Percentage
NB	658	81%
CPC	413	51%
APC	470	58%
Total detected: (out of 809)		90%

Table 4.6: Percentage of datasets containing additional (possible false positives) flags, after increasing the local radius parameter. Decreased values compared to Table 4.4.

Probe	Additional flags
NB	34%
CPC	20%
APC	11%

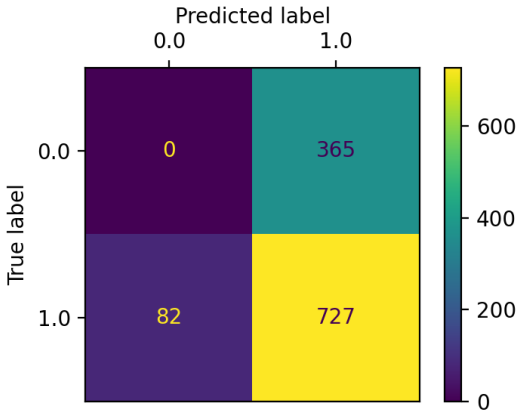


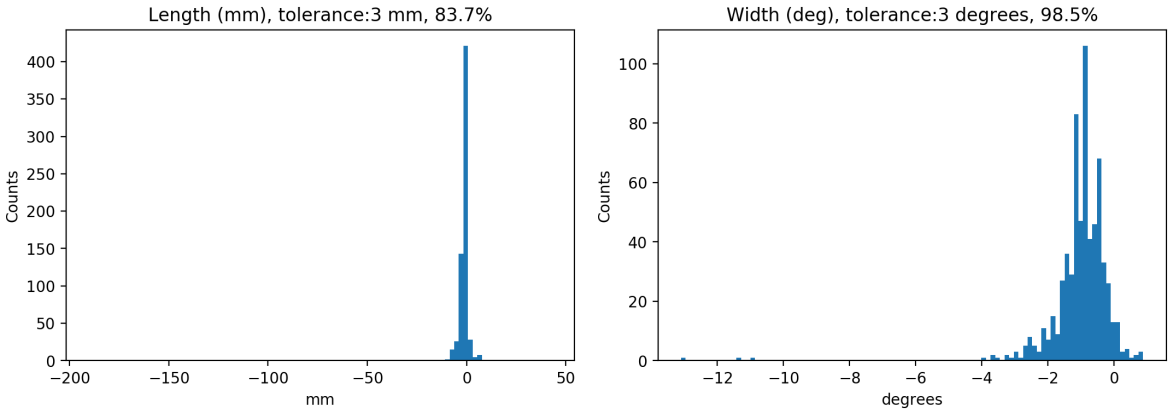
Figure 4.23: Confusion matrix after increasing the local radius parameter. Compared to the smaller local radius results (Figure 4.22), there is a decrease in the false positive counts by n=-106 (-22.5%), however the true positive counts were also reduced by n=-30 (-4%).

Table 4.7: Evaluation metrics for different local radius (*Eps*) values. The increased radius increased the coverage of the 'healthy' cluster, resulting in higher precision, at the expense of recall.

Experiment	Precision score	Recall score	F1 score
Small Local Radius	0.62	0.93	0.74
Large Local Radius	0.67	0.90	0.77

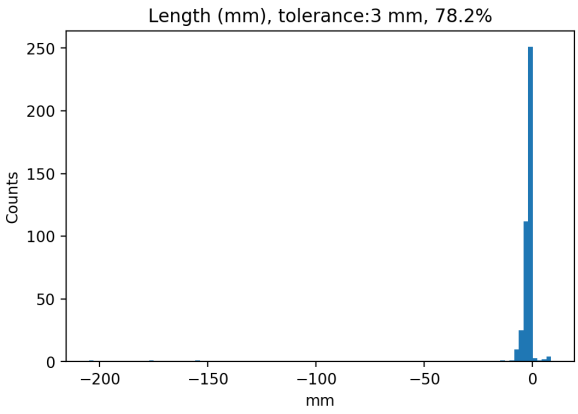
4.5.1 Discussion Of Results

The analysis of the results showed satisfying performance for both detection and sizing. However, this was implemented in an automated way and therefore some errors should be expected due to the 766 datasets and 809 target flaws not being curated individually.

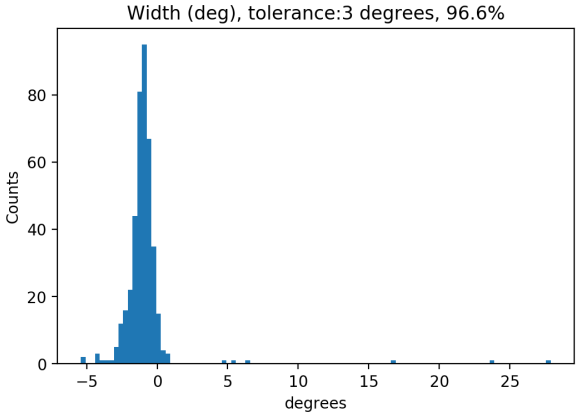


(a) Length (NB)

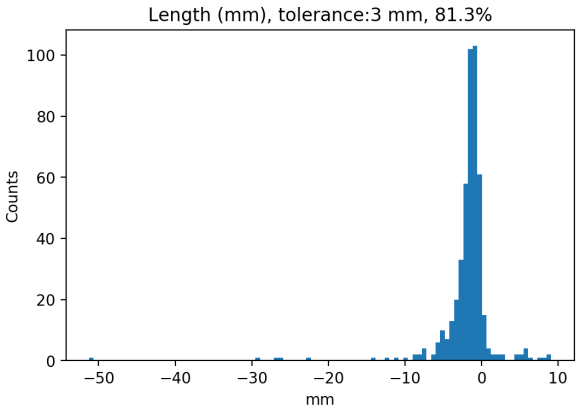
(b) Width (NB)



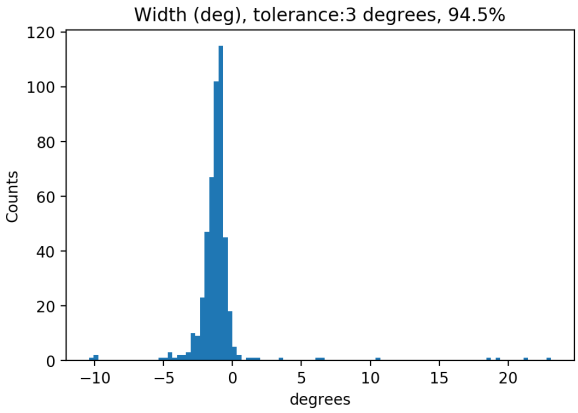
(c) Length (CPC)



(d) Width (CPC)



(e) Length (APC)



(f) Width (APC)

Figure 4.24: Sizing results (size(flag)-size(verified)) with increased local radius parameter value (*Eps*).

Specifically, a small number of 'successful' detections could be due to a large area being incorrectly flagged and overlapping with the flawed area. Nevertheless, the good sizing performance provides additional confidence in the validity of the flaw detection results.

On the other hand, there can be factors that have adverse effect on the reported detection rate, i.e. a situation where it would not be possible from the analysis of a given dataset to detect the associated flaw(s) but due to the automated assessment procedure not identifying this being the case, the flaw is reported as undetected by the algorithm.

To further investigate this potential issue, a manual evaluation of undetected flaws with depth $\geq 0.15mm$ ('dispositionable' flaws) was carried out. This category of flaws contains 302 flaws of which 280 (93%) were reported as 'detected' by the automated assessment. The manual evaluation of the remaining 22 flaws showed that only 6 were wrongfully rejected as noise and therefore it is required to further investigate the causes. The causes of the remaining 18 flaws reported as undetected were due to problems with either the datasets or the automated assessment procedure:

- 5 were duplicated versions of already detected flaws. The automated assessment procedure processes the provided verified analysis documents to extract verified information about the flaws. In these cases it failed to identify these flaws as duplicated due to slight deviations in their descriptions

multiple documents. The produced flags are unique instances that are assigned to overlapping verified Flaws only once, thus one instance of the 'duplicated' flaws is reported as 'missed'.

- 4 belonged to unsuitable datasets as they are obtained from an outlier pressure tube with multiple manufacturing flaws.
- 4 were reported as not detected because of large misalignment between the B-Scans (used by the clustering procedure) and the GH datasets (used by analysts).
- 2 were due to corrupted B-Scans.

Therefore, the examination of this subset of flaws showed that it is expected that some of the flaws were erroneously reported as undetected by the clustering procedure due to miscellaneous causes, such as duplicated reported flaws, misaligned datasets and data quality. An additional known source of potential errors in the assessments of the results is that the automated signal analysis operate only within the given ultrasonic datasets without access to historical results related to prior inspection data which might had an influence on the final flaw size reported by the analysts [80].

4.6 Discussion

This chapter has described the development of a subsampling method and a prototype two-stage clustering procedure that enables the automated analysis of datasets containing ultrasonic inspection data captured by different types of ultrasonic probes.

The pre-processing step showed both qualitative and quantitative approaches for reducing the search space by selecting and transforming discriminative and informative signal features. This enables the usage of the subsampling method, which wraps the DBSCAN clustering algorithm and is able to further reduce the search space. The main idea behind the subsampling method is to identify just-enough representative samples to participate in the computationally demanding clustering task, thus reducing significantly the memory requirements and the computation time. The efficiency of the subsampling method relative to vanilla DBSCAN, is strictly dependent on the number of unique samples of the dataset of interest (i.e. cardinality), relative to the size of the dataset. For datasets that fail to meet this requirement, a level of reduced cardinality can be enforced by exploring various grouping routes, e.g. rounding at a certain decimal place or quantisation. Therefore, the subsampling method introduces an additional analytical aspect that needs to be tuned accordingly. Nevertheless, as demonstrated in the current chapter, it can enable the employment of the useful clustering capabilities of DBSCAN to larger datasets, that would otherwise be highly impractical or even

impossible, achieving computational performance comparable to other prevalent efficient clustering algorithms that exhibit less desirable clustering traits for the objectives of this chapter.

The two-stage clustering procedure takes advantage of the pre-processing step and the subsampling method, and is designed to perform efficient groupings of the signals and noise rejection. This chain of events transforms the original high-volume 3D data to 2D C-Scans that provide information regarding the location, shape, and size of potentially flawed areas.

The proposed procedure was tested across 809 datasets and the produced flags were compared to related historical analysis records showing satisfactory results both for detection and sizing of the flaws. However, although the analysis of errors showed that a subset of errors are due to data quality and the quality of the automated flag assessment procedure, there is still a small percentage of flaws that the proposed procedure was not able to detect.

A further drawback of the proposed procedure is the selection of the several parameters involved throughout the analysis. Although this thesis proposed qualitative heuristic methods to approach the identification of appropriate parameters, there is still room for subjective and manual iterative tuning.

The proposed procedure was designed to offer analytical efficiency and completed the analysis of more than 800 of multi-MB datasets, each containing data from three types of probes, within less than three hours. However, these datasets contain information only for segments of pressure tube surface, and the com-

putationally intensive iterative nature of pre-processing and clustering would be problematic for the analysis of datasets containing data from the entire surface of a pressure tube.

Contributions of this chapter:

- *Subsampling method that reduces the clustering search space of the DBSCAN algorithm.*
- *Flaw detection through a two-stage clustering method of ultrasonic inspection data.*
- *DBSCAN parameter estimation method for ultrasonic inspection datasets.*

Chapter 5

Large-Scale Ultrasonic Anomaly Detection

5.1 Introduction

This chapter presents a novel large-scale anomaly detection method across high-resolution unlabelled and unprocessed ultrasonic scans consisting of collections of A-Scans. The method reduces the size of the data through the extraction and transformation of signal features independently of the various surface neighbourhoods' characteristics, allowing for a normalised and consistent view of the general current state the entire surface that enables the identification of abnormal areas.

This is demonstrated through its application to complete full-length scans of pressure tubes acquired by different probes, aiming to provide a fast and robust automated procedure for supporting the analysts responsible with the health assessment of the pressure tubes during the planned outages.

5.2 Analysis overview

Modern ultrasonic inspection tools, such as ANDE, enable the collection of high-resolution data in volumes that greatly surpass previous generation tools (e.g. CIGAR), due to their more advanced technical capabilities, as well as possessing more probes collecting a wider range of data in parallel.

Although this allows for a more detailed examination of large inspected areas, the computational analysis of large-scale inspection data faces major challenges both in terms of processing time and memory usage. This can render the processing of the entire inspection dataset impractical, especially when the completion of the health assessment is time-sensitive and a requirement for resuming operation.

Chapter 4 addressed the unsupervised computational analysis of smaller sections of high-resolution ultrasonic data captured from surface areas previously identified to potentially contain abnormalities. Although the proposed sub-sampling approach, in conjunction with the dual-stage density-based clustering, enabled the analysis of those datasets (demonstrated using CIGAR datasets), the current chapter is concerned with datasets that cover larger inspection areas, resulting in multi-GB sizes that cannot be similarly processed in memory.

The problem will be approached mainly through an unsupervised data-driven perspective, however, utilising general soft assumptions about signal behaviour during feature pre-processing is an essential step towards a more homogeneous feature space. This section will examine in detail each step of the proposed analysis process depicted in Figure 5.1.

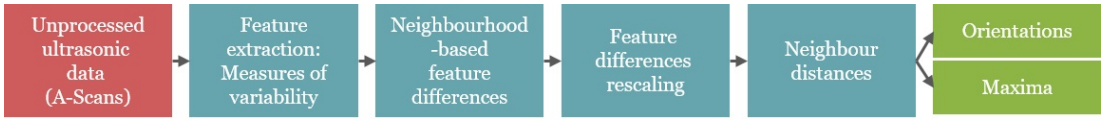
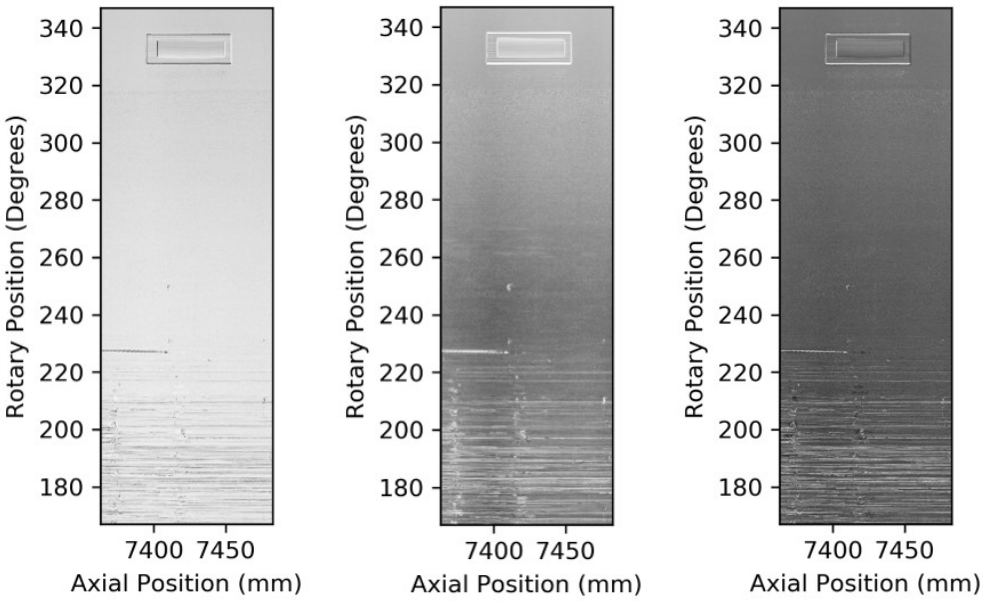


Figure 5.1: High-level overview of the proposed transformation process.

5.2.1 Data description

The datasets of interest within this chapter contain collections of ultrasonic inspection signals obtained by 15 MHz and 20 MHz NB probes of the ANDE tool across the main body of the pressure tube, excluding the end-parts where the tube connects to the reactor core. The length of this area is approximately 6.2 m, covering 95% of the tube's total length. Given the internal tube diameter of approximately 105 mm, and the axial and rotary increments of 0.4 mm and 0.1 degrees respectively, the resulting sample sizes are between 55 and 56 million signals. Depending on the amplitude resolution (8 bit or 14 bit) and the temporal resolution (3600 to 3750 amplitude samples) of the individual A-Scans, the resulting file size of a dataset of a single scan ranges between 180 GB to 400 GB, per probe. This amount of data poses a significant challenge for any automated procedure which aims to perform an assessment within a reasonable amount of time that does not adversely extend the analysis period of the the expert analysts it serves.



(a) Skewness (b) Variance (c) Kurtosis

Figure 5.2: C-Scan of a pressure tube region - Pixel intensities represent the feature value of the corresponding A-Scans.

5.2.2 Feature selection

As the size of the data is restrictively large for single machine analysis, it is required to condense part of the information contained within the datasets. This can be achieved by transforming the individual A-Scans from amplitude-time pairs into small sets of descriptive features that will act as representatives at any subsequent analysis.

The selection of the type and number of features is an open-ended problem. Here, the selection is based on two assumptions: First, that anomalies, such as scratches and flaws, should generally be comprised, at least partly, of areas of increased

surface roughness, compared to the surrounding 'healthier' areas. Rough surface topography affects the reflected signal through incoherent scattering [161, 162] which should result in some change in the shape of the signal.

The second assumption is that the shape alteration should have some effect on the distribution of the sampled discrete intensities of each signal. Therefore statistical measures of variability that describe some aspect of the distribution should form an informative set of features of each signal. Similarly to chapter 4, this chapter favours parameter-free features to participate in the presented analysis, and uses a small set of features. Specifically, variance (Equation 5.1) is a measure of how spread out the distribution is, kurtosis (Equation 5.3) describes the tail shape of the distribution, and skewness (Equation 5.4) is a measure of asymmetry of the distribution.

For N-sample discrete signal x these quantities are calculated as:

$$\sigma^2(x) = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2, \tag{5.1}$$

where $\bar{x}$ is the sample mean:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i, \tag{5.2}$$

$$Kurt(x) = \frac{\frac{1}{N}\sum_{i=1}^{n}(x_i - \bar{x})^4}{(\frac{1}{N}\sum_{i=1}^{n}(x_i - \bar{x})^2)^2} \tag{5.3}$$

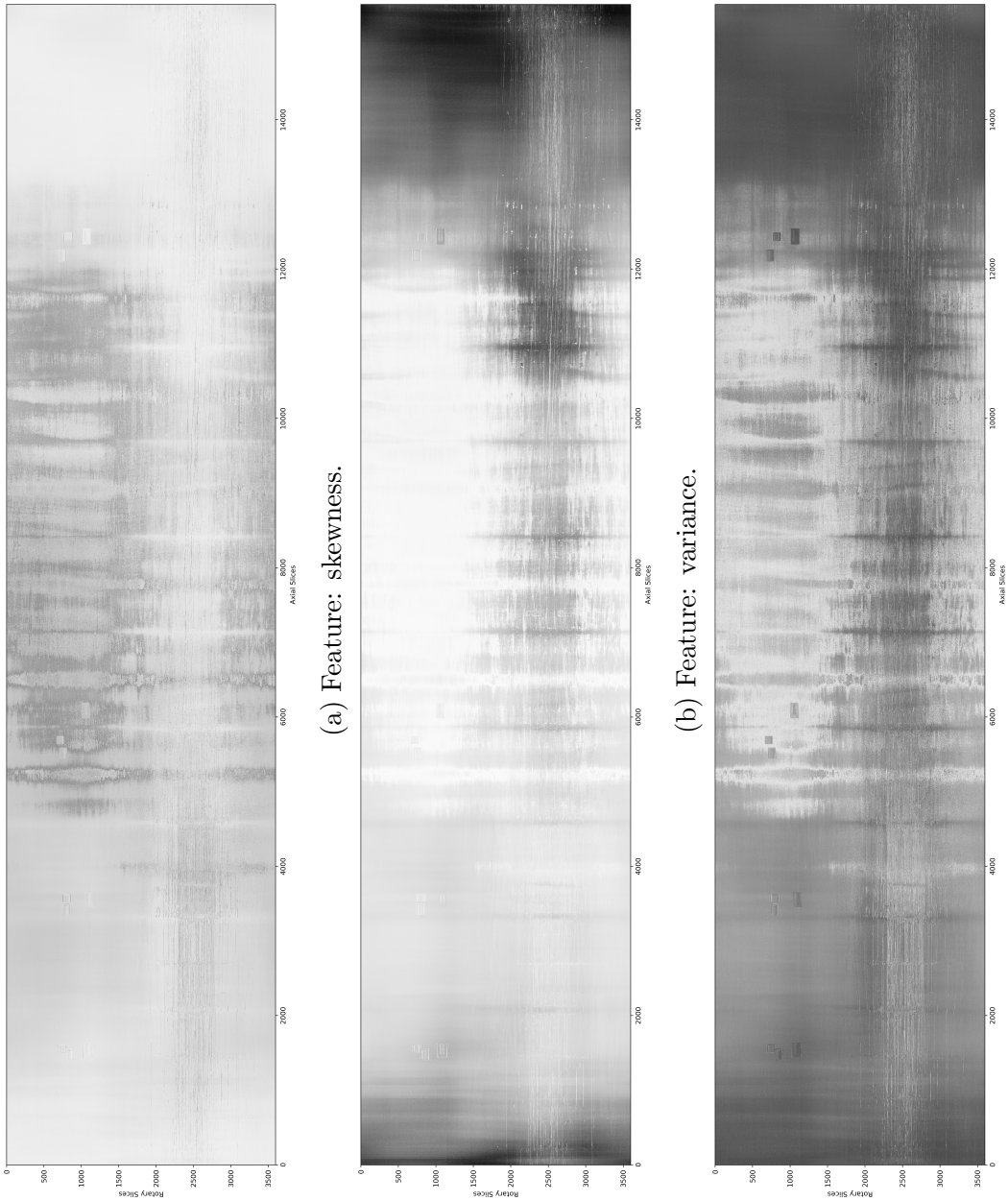
$$Skew(x) = \frac{\frac{1}{N}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\sigma^3(x)} \tag{5.4}$$

Therefore, instead of explicitly attempting to physically model and predict the effects of such a highly randomised event on the signal, the relationships between the signals will be explored through these three descriptive statistics that quantitatively summarise the entire A-Scan.

After sequentially loading each A-Scan and calculating its sample variance, sample kurtosis and sample skewness, the data size has been significantly reduced from the range of 180–400 GB to approximately 1 GB per scan per probe.

Figure 5.2 shows examples of C-Scan sections of a tube created using extracted features from signals obtained by a pulse echo 20 MHz normal incidence probe. C-Scans are two-dimensional top-view presentation where the intensity of each pixel represents the corresponding signal's feature value, and offer a visual way of assessing the information that these features are carrying.

As can be seen in the small-scale C-Scans (Figure 5.2), these features show promising capabilities, as anomalies, such as scratches and machined scrape marks, are distinguishable enough from the surrounding areas. However, within the scope of the entire population of a full-length pressure tube (Figure 5.4b and Figure 5.5) there can be various levels of heterogeneity across different parts of the tube as it is an active large-scale environment affected by imperfect geometry and surface conditions. Effectively this means that two healthy areas could appear different, when compared to each other, or, a healthy signal obtained from a noisy area to appear similar to a signal obtained by a noise-free defective area. This could

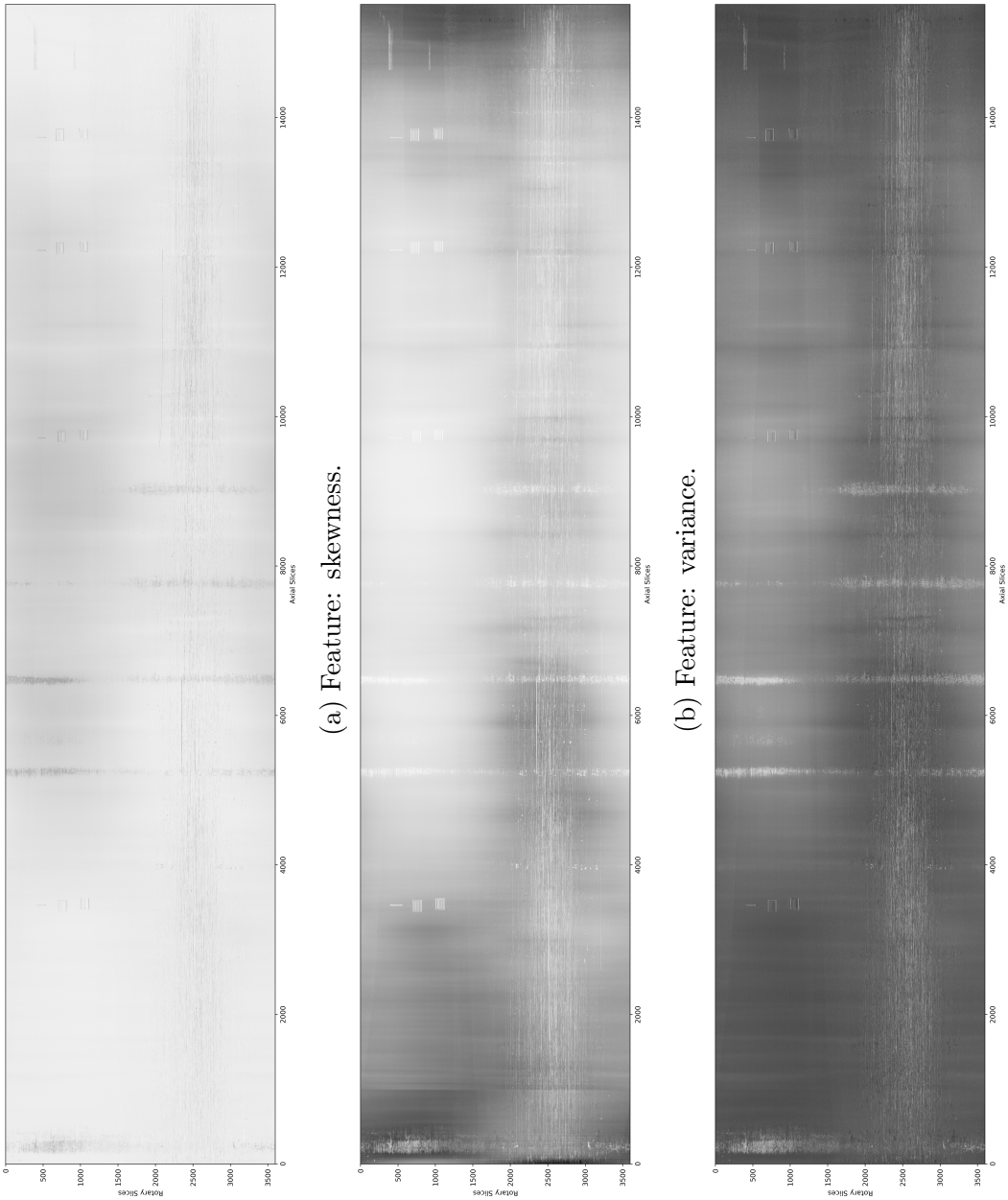


(a) Feature: skewness.

(b) Feature: variance.

(c) Feature: kurtosis.

Figure 5.3: C-Scan across pressure tub #1 (6300 mm by 360 degrees) using a single feature per signal. Neighbourhood-specific surface conditions create a heterogeneous response.

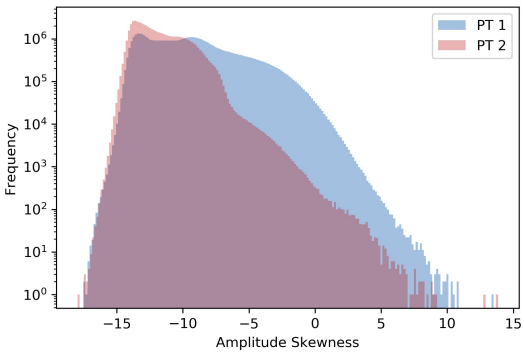


(a) Feature: skewness.

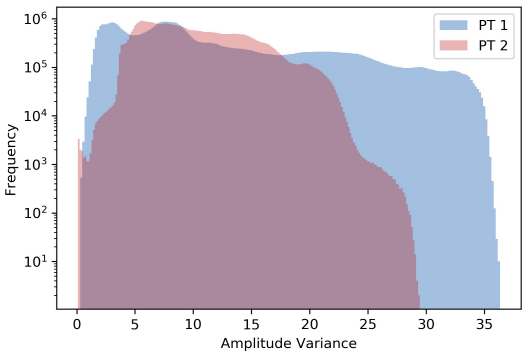
(b) Feature: variance.

(c) Feature: kurtosis.

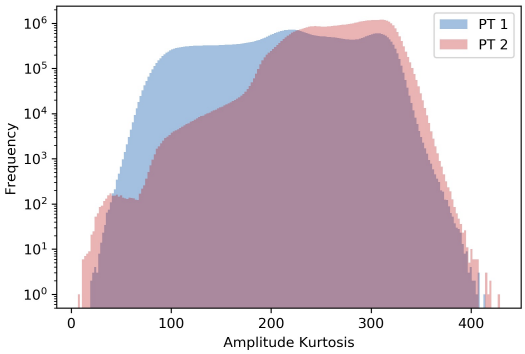
Figure 5.4: C-Scan across pressure tube #2 (6300 mm by 360 degrees) using a single feature per signal. Neighbourhood-specific surface conditions create a heterogeneous response.



(a) Skewness



(b) Variance



(c) Kurtosis

Figure 5.5: Log-histograms of the three features for each pressure tube.

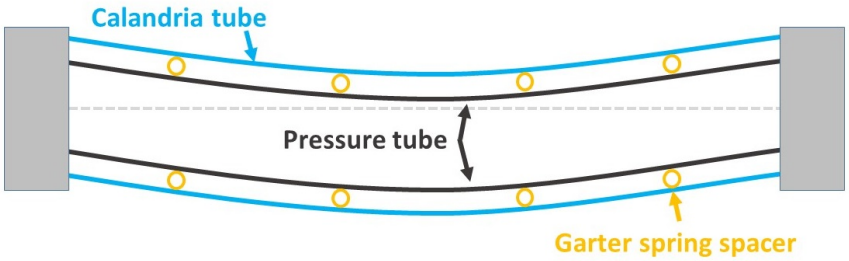


Figure 5.6: Schematic of exaggerated pressure tube sag.

occur as signals of different sections of the tube could be affected by different neighbourhood-specific causes, e.g., poorly focused signals due to probe misalignment [163] caused by sagging of the pressure tube (Figure 5.6), or rough surface area.

5.2.3 Features transformation

The disentanglement of the different behaviours across the tube could be approached through unsupervised machine learning methods, such as density-based clustering, as it has shown promising results in the task of grouping together signals obtained from defective regions [164]. However, these methods are computationally expensive and would require extensive experimentation to mitigate the effects of any possible varying density levels across large and heterogeneous datasets [165].

The proposed approach is seeking to normalise the extracted features, fuse them, and enable intuitive inferences that are relevant across the pressure tube population. This feature transformation is based on the idea of tracking changes of the feature values locally, instead of investigating the relationship of each feature

value within the whole feature space. This can be approached by representing each signal by some distance function of its feature values and its immediate physical neighbours' feature values only, thus reducing drastically the size of the feature space that each sample has access to.

The essential property of a dataset that can support this type of operation is that its members should have strictly defined spatial relationships. The sequentially captured A-Scans of the datasets of interest are captured by a precise helical motion which completes a shifted full revolution every $N = 3600$ samples, enabling the identification of the indices of the 4-connected neighbours for each sample. Naturally, any further computations are not valid for the first and last N samples, as they lack at least one of the four required neighbours. Therefore, they will be excluded from the population and their role will be limited to participating as neighbours, when required.

The indices for the 4-connected neighbours of each of the signals within the dataset $\boldsymbol{S} = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n]$, where $\boldsymbol{s}_i = [a_1, \ldots, a_m]$, n:number of signals, a:amplitude, m:number of amplitude samples per signal, are defined by Eq. 5.5:

$$\boldsymbol{i}_n = (i - N, i - 1, i + 1, i + N), i \in \mathcal{I}, \tag{5.5}$$

where $\mathcal{I}$: index set of $\boldsymbol{S}$. Although the periodicity within the 1D dataset allows for a 2D mapping with indices $\mathcal{I}_x = (\mathcal{I} \mod N)$ representing the rotational index positions, and $\mathcal{I}_y = \lfloor \mathcal{I}/N \rfloor$ representing the axial index positions, it should be noted that 2D reshaping fails to account for the shifting and continuity of the helical capture. Therefore, selecting the x indices of the neighbours $\mathcal{I}_{n_x}^{(2D)}$ based

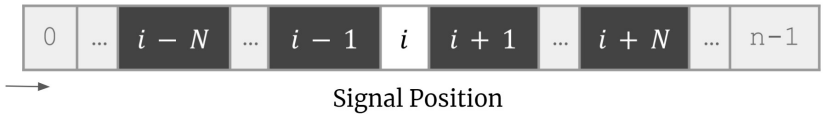
on Manhattan distance of $r = 1$ needs adjustments for the extreme cases of $\grave{\mathfrak{I}}_x = \{0, N-1\} \subset \mathfrak{I}_x$, as shown in Eq. 5.6. The x indices correspond to the rotational index position, and the order matches the order of the neighbours of Eq 5.5, i.e. the order the signals are originally captured and stored during the helical scanning.

$$\boldsymbol{i}_{n_x}^{(2D)} = \begin{cases} (i_x, i_x - 1, i_x + 1, i_x), & \text{for } i_x = \\ & [1, N-2] \\ (0, N-1, 1, 0), & \text{for } i_x = 0 \\ (N-1, N-2, 0, N-1), & \text{for } i_x = N-1 \end{cases} \tag{5.6}$$

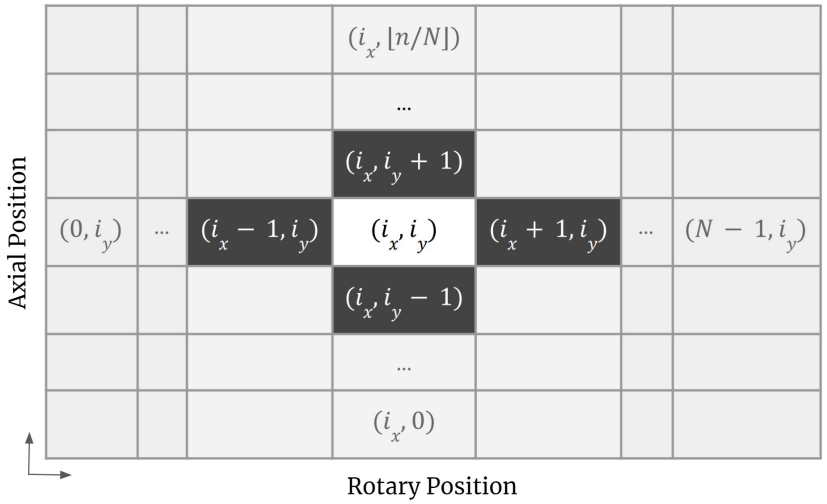
After identifying the appropriate neighbours, the next step is concerned with calculating the Euclidean distances between the variability-based feature vectors $\boldsymbol{f}_i$ and the feature vectors of each of its four neighbours Eq. 5.7:

$$\boldsymbol{d}_i = [\sqrt{\sum_{j=1}^{3}(f_i^{(j)} - f_{i,1}^{(j)})^2}, \ldots, \sqrt{\sum_{j=1}^{3}(f_i^{(j)} - f_{i,4}^{(j)})^2}], \quad i \in I \tag{5.7}$$

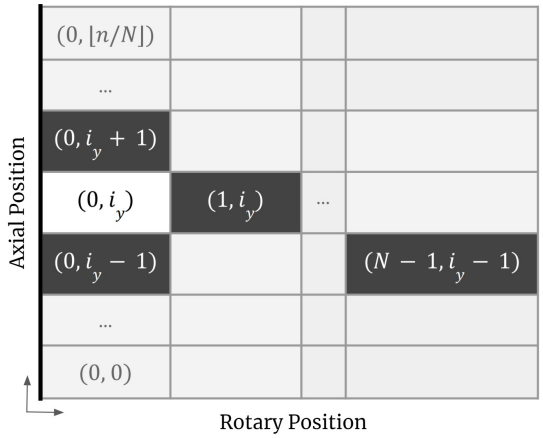
However, prior to the calculation of the distances, it is crucial to reformat the participating data onto the same value range (feature scaling). This is a measure against features with intrinsically large numeric values dominating the results.



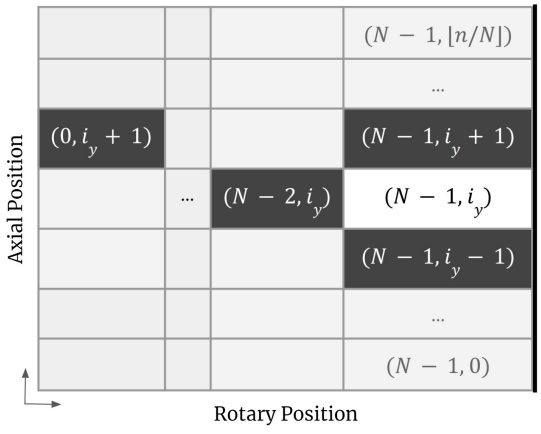
(a) General 1D (stream) setting.



(b) General 2D setting.



(c) Edge case for 2D setting: Start of rotary position.



(d) Edge case for 2D setting: End of rotary position.

Figure 5.7: Indices of the 4-connected neighbours (dark background) for different data layout settings.

For the proposed approach, instead of the original features, the scaling happens on the squared features differences per feature type, per neighbour type: $\boldsymbol{\delta}_n^{(j)} = (\boldsymbol{f}^{(j)} - \boldsymbol{f}_n^{(j)})^2$, before the summation across the features. This is driven by the precondition to identify any possible extreme outliers that might distort the scaled values heavily if they participate in the calculation of the scaling parameters. As discussed above, the neighbourhood-based differences are expected to offer a more homogeneous space, thus facilitating a more intuitive identification of these outliers.

The output of the distance calculation is four distances associated with each sample, one for each neighbour. The second step of the transformation is further reducing the search space by collapsing these four distances into one. A viable option is extracting the element-wise maximum value (Eq. 5.8), thus representing each sample/signal by the maximum change of variability across its 4-connected neighbourhood.

$$\boldsymbol{m} = [\max \boldsymbol{d}_1, \ldots, \max \boldsymbol{d}_n] \tag{5.8}$$

Furthermore, by storing the index of the neighbour that maximises each $\boldsymbol{d}_i$, the direction of change is preserved (Eq. 5.9):

$$\boldsymbol{\kappa}^{(4)} \in \{1, 2, 3, 4\} = [\arg\max \boldsymbol{d}_1, \ldots, \arg\max \boldsymbol{d}_n] \tag{5.9}$$

However, $\boldsymbol{\kappa}^{(4)}$ can be further simplified (Eq. 5.10) by preserving only the orientation of the change: $\boldsymbol{\kappa} = \boldsymbol{\kappa}^{(2)} \in \{-1, 1\}$ with:

$$\kappa(i) = \begin{cases} 1, & \text{for } \kappa^{(4)}(i) = 1, 2 \\ -1, & \text{for } \kappa^{(4)}(i) = 3, 4 \end{cases}, \tag{5.10}$$

thus reducing the cardinality from 4 to 2.

By conventionally choosing $\{-1, 1\}$ to represent horizontal and vertical orientations respectively, the orientation information $\boldsymbol{\kappa}$ can be easily incorporated within the strictly positive $\boldsymbol{m}$ via element-wise multiplication (Eq. 5.11):

$$\hat{\boldsymbol{m}} = \boldsymbol{\kappa} \odot \boldsymbol{m} \tag{5.11}$$

5.3 Application

To assess the potential of the proposed process as a means towards detecting anomalies across large-scale ultrasonic inspection systems, high-resolution datasets from two pressure tubes have been utilised as case examples.

The datasets contain data obtained by a 20 MHz pulse-echo normal-incident probe and a 15 MHz pulse-echo normal-incident probe with stretched focal zone from the use of logarithmic lens. After extracting the required measures of variability from each A-Scan, they were input to the neighbourhood-based process described in Section 5.2.

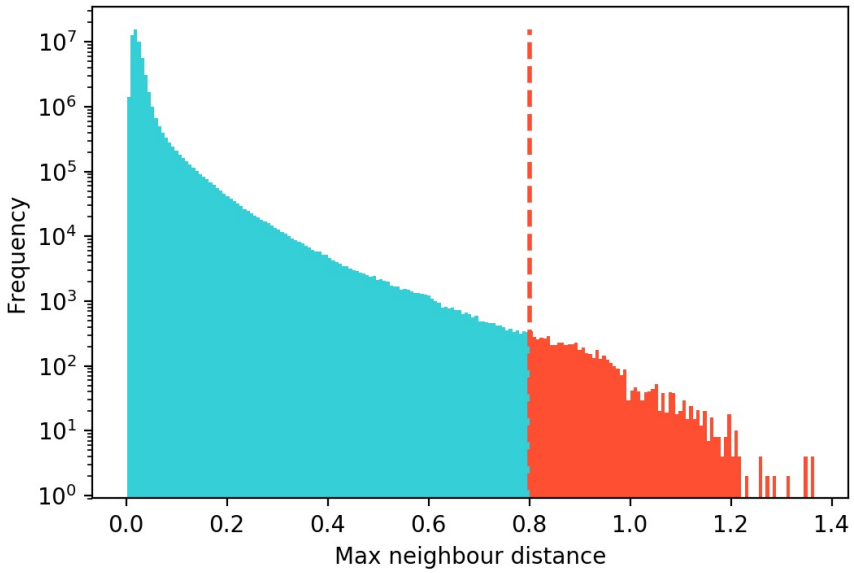
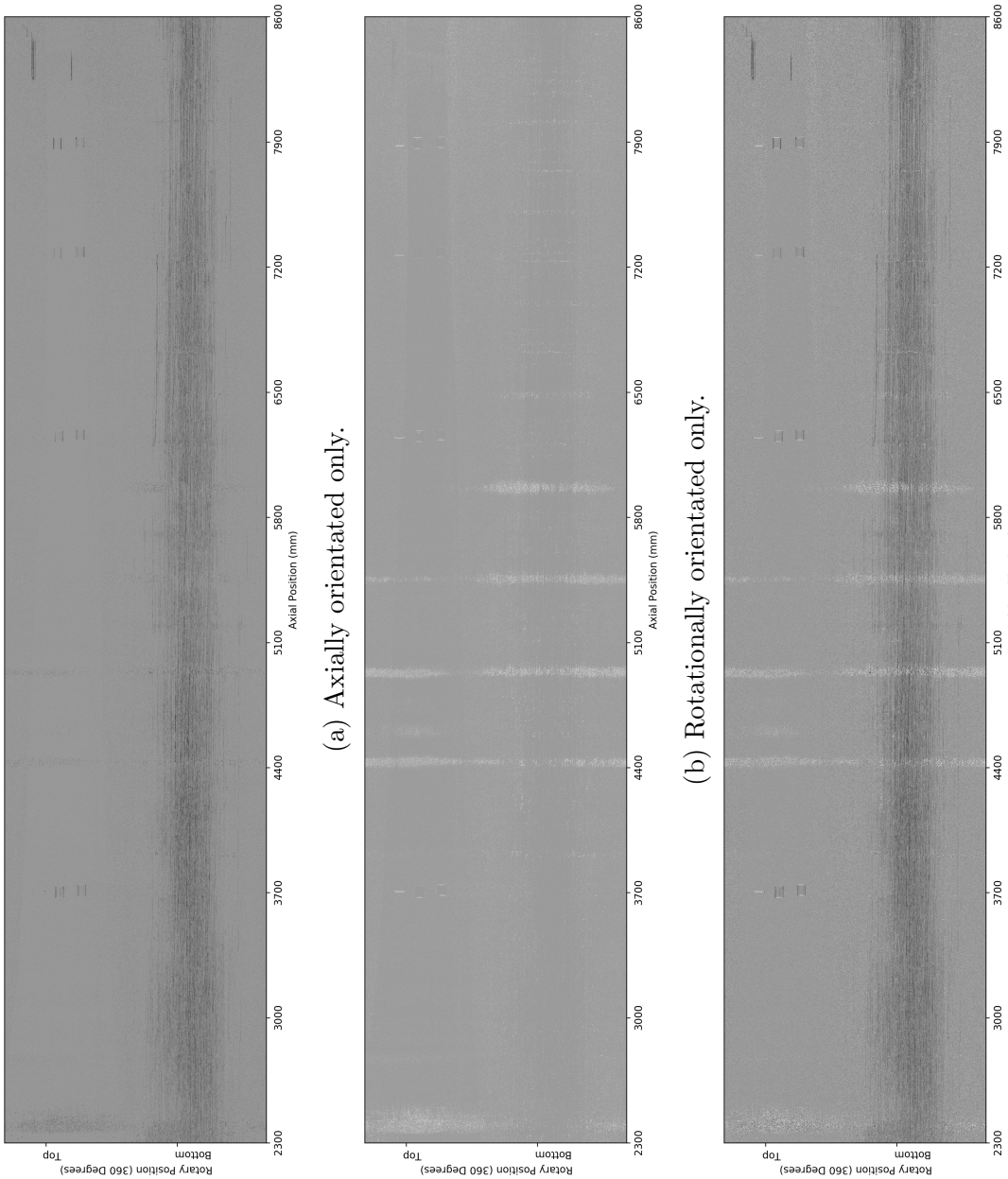


Figure 5.8: Log-histogram of transformed features' distances of a pressure tube.

The resulting maximum changes of variability $\boldsymbol{m}$ (Figure 5.10) offer a more homogeneous space, making it intuitive to assume that signals from healthier regions tend to concentrate towards zero, and anomalies with increased surface roughness diverge monotonically towards $+\infty$.

Furthermore, incorporating the orientation information within $\boldsymbol{m}$ enables the separate investigation of anomalies that manifest along different axes (Figure 5.9).

To further illustrate the normalising effects of the feature transformation method, the MiniBatchKMeans algorithm [166] is used to process the features both before and after their transformation through the proposed method (Figure 5.11). MiniBatchKMeans is a variant of K-Means that is designed to be more computationally efficient, particularly for large datasets. Specifically, instead of calculating

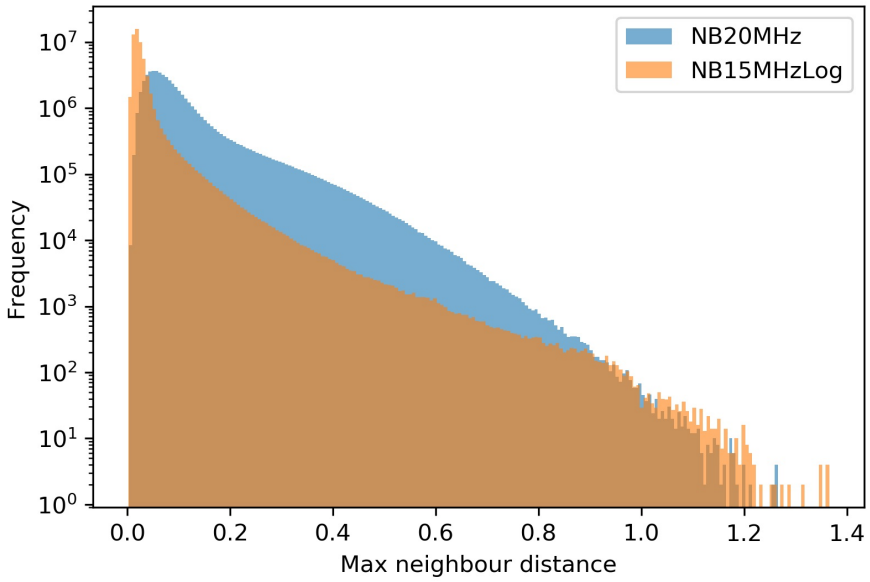


(a) Axially orientated only.

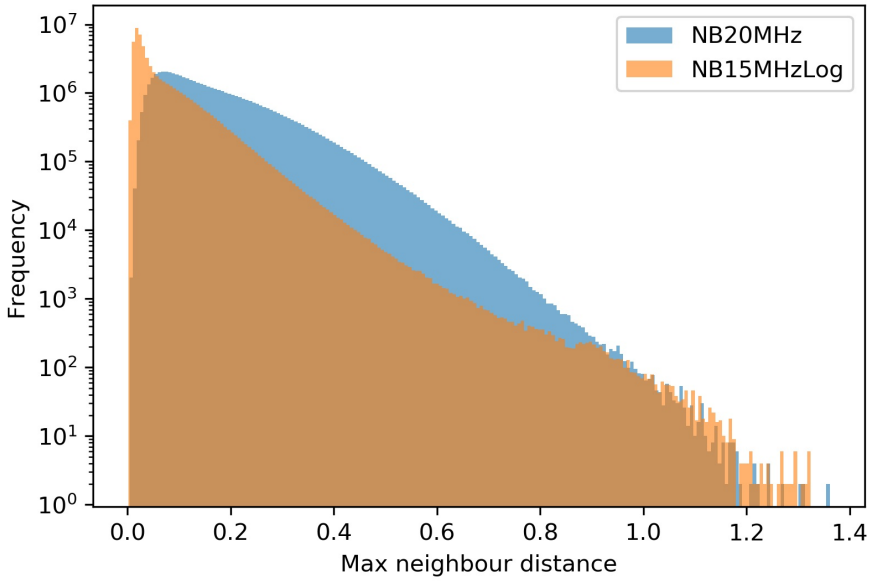
(b) Rotationally orientated only.

(c) Both orientations included.

Figure 5.9: C-Scans across a pressure tube using transformed features' distances.



(a) PT#1



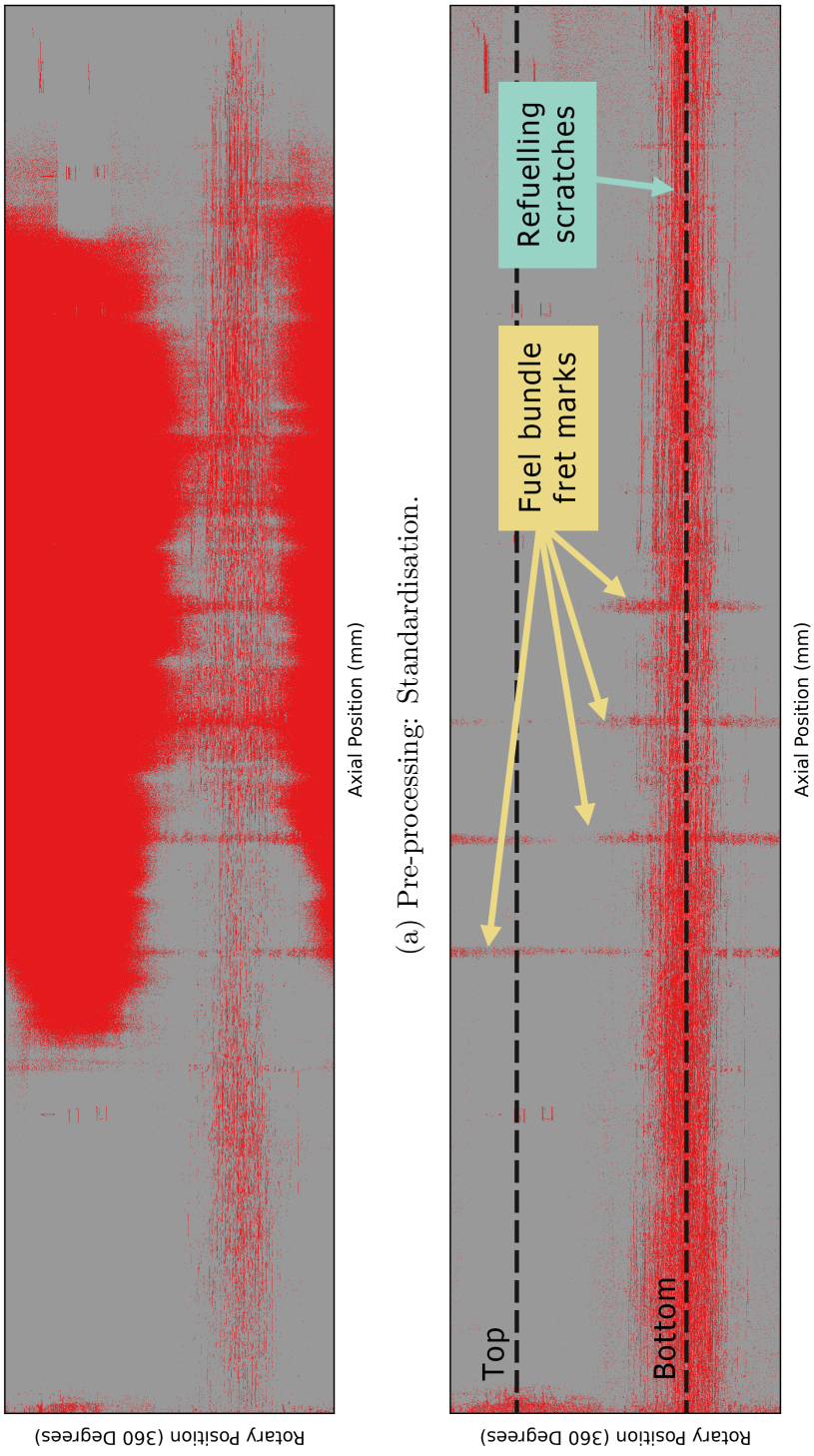
(b) PT#2

Figure 5.10: Log-histograms of the transformed distances for each PT. The less concentrated distributions of (5.10b) indicate that PT#2 contains more rough surface areas than PT#1.

the distance between each data point and each centroid for each iteration, MiniBatchKMeans randomly selects subsets of data, called mini-batches, to update the centroids, instead of using the entire dataset.

Given the *a priori* unknown number of expected clusters the extracted features are processed by instructing MiniBatchKMeans to form two clusters. In the first test setting (Figure 5.11a) the features are rescaled through the common standardisation process (i.e. the features are centred around zero, with a standard deviation of one; presented in Equation 4.9). In this case, MiniBatchKMeans fails to cluster the signals in a meaningful way, as a large proportion of the middle section of the pressure tube is grouped together (red cluster) with the part of the pressure tube that contains service induced scratches. In contrast, the second test setting utilises the proposed neighbourhood-specific feature transformation, which reduces the effects of heterogeneity in the data, resulting in the MiniBatchKMeans producing qualitatively more intuitive results: It groups together (red cluster) the parts of the pressure tube surface that systematically experience service induced damage, such as fuel bundle scratching during refuelling, and fuel bundle bearing pad fretting caused by fuel bundle rocking and vibration [38].

Anomalies found on the surface of a pressure tube include scratches, scrapes, flaws, deposits of foreign material, or erosion/corrosion [19]. Among them, deep sharp flaws are considered of potential concern for the health of the pressure tube and are reportable. Therefore, a critical assessment criterion for the proposed system is its ability to identify areas containing these flaws, while minimising the amount of reported areas containing minor surface anomalies.



(a) Pre-processing: Standardisation.

(b) Pre-processing: Proposed feature transformation.

Figure 5.11: MiniBatchKMeans clustering into 2 groups. The raw feature space consists of the 3 measures of variability discussed on Chapter 5.2.2, excluding any location related information. After clustering the pre-processed feature space, the resulted labels per signal are remapped to their corresponding positions to generate the presented C-Scans.

Using verified measurements from older inspections (CIGAR tool), the locations of the reportable debris flaws (4 flaws for each PT) have been manually identified in the C-Scans produced by the proposed process. Overall, for PT#1, all reportable debris flaws are visually identifiable in both probes' C-Scans. However, for the noisier PT#2, one of the flaws is indistinguishable from the surrounding noisy area when using the resulted C-Scan from the 20 MHz probe.

By gradually increasing the threshold between 'healthy' and 'abnormal' (Figure 5.8), the upper bound where at least a part of all the (visually distinctive) flaws is present can be extracted. Table 5.1 presents the resulted threshold values for each pressure tube - probe pair, along with the percentage of the eliminated samples. The remaining samples appear to be members of verified flaws, scratches and edges of machine-made scrapes. However, since this process acts as a first-pass anomaly filter, without providing any depth information, it is not possible to characterise the threat-level of the detected outliers. For this, the reported areas would act as inputs for depth-specialised algorithms, or be forwarded for further manual analysis.

Another key aspect of the proposed process is the computational time required by the transformation process. The bottleneck of this process is the extraction of features from the 50 to 60 million signals. The required time for this step varies greatly depending on the underlying system. The worst case for this linear batch-process is approximately 4 hours for a single medium capacity computer accessing the data of a network drive. Therefore, this time could be significantly reduced within a multi-machine industrial environment. Following the feature extraction,

the remaining steps of the process are implemented based on vectorised operations [167] enabling the analysis to be performed by a single medium capacity system using high-level interpreted language within few minutes.

The results of the case examples indicate that the proposed process is a promising step towards extracting anomalous areas from large-scale high-resolution ultrasonic inspection data in a robust and efficient way. The assumptions that this process was based upon, as well as the interpretable fusion of the features, allow for a well-informed expectation of its outcomes, limitations, and decision support role within the pressure tube health assessment process. This should allow future investigative analyses to extract general *a priori* parameter estimates that generate results compliant with the respective industry's guidelines. Nevertheless, a conservatively estimated cut-off threshold can provide:

- Visual and statistical information regarding the underlying general current state of the pressure tubes.
- Standardised comparison of current and future scans: Anomaly evolution tracking.
- Heavily decreased search space for the depth-specialised algorithms.

Table 5.1: Upper threshold bounds, with the corresponding percentage of eliminated samples, where at least a part of all the visually distinctive flaws is present.

PT	Probe (NB)	Threshold	Eliminated (%)
1	15MHz L	0.8	99.989
1	20MHz	0.75	99.967
2	15MHz L	0.35	99.279
2	20MHz	0.75*	99.908

* 3 out of 4 flaws

5.4 Discussion

This work has introduced an anomaly detection approach for large-scale ultrasonic inspection datasets achieved through a novel feature transformation. It uses extracted measures of variability from each A-Scan as potential features carrying information about surface roughness around defective areas. Next, their neighbour-based distances are utilised as a means to decrease the effects of neighbourhood-specific deviations that cause heterogeneity across the features. This approach is especially relevant for large ultrasonically inspected structures, such as a full-length pressure tube scan, where the geometric and other neighbourhood-specific surface conditions can affect the captured signals in distinct neighbourhood-specific, creating multimodal feature distributions that are not exclusively related to the actual condition/health of the surface.

Initial application to real-world CANDU pressure tube inspection datasets show that the proposed pre-processing transformation allows for more intuitive clustering results, over pre-processing through the widely used standardisation method, as it allows for clusters that are related to the service-induced damage, in contrast to neighbourhood-related properties.

Furtheremore, due to the resulted reduced homogeneity, the transformed features can be further reduced into a single metric (e.g. through representing each individual signal through the maximum variability distance in its neighbourhood) that is relevant across the large-scale surface, thus providing an intuitive anomaly detection approach that, through thresholding, can isolate areas containing surface wear, including threatening flaws. This provides visual and statistical information regarding the underlying general current state of the pressure tubes, as well as a heavily reduced search space for computationally demanding depth-focused algorithms. As more datasets become available, further exploration should enable pattern recognition techniques to *a priori* separate threatening flaws from low-risk anomalies.

Contribution of this chapter:

- *Feature transformation method for normalising large-scale heterogeneous ultrasonic inspection data.*

Chapter 6

Conclusions

6.1 Summary and Discussion of Contributions

This thesis has presented novel unsupervised data-driven approaches to analysing ultrasonic inspection data for surface flaw detection. The application domain, necessary from understanding the inspection datasets and inspection setup that inspired this thesis, is introduced in Chapter 2. This is provided through an overview of nuclear plant maintenance strategies that leads to the specific maintenance and inspection problem of pressure tubes, a critical asset of CANDU nuclear reactors. Datasets arising from this inspection process offer an application setting for testing the system-agnostic methods proposed by this thesis as they offer a variety of challenges both in terms of complexity and size. The current pressure tube health assessment process, along with recent knowledge-based expert-systems research towards automating this process, are also presented in Chapter 2, and it is argued that a data-driven approach can provide further support in a rapid and robust manner.

A review of existing research efforts concerned with the automated asset health assessment through the analysis of ultrasonic inspection data is presented in Chapter 3. The review is focusing on supervised and unsupervised data-driven methods and explores the applicability of these methods in datasets containing a diverse set of flaws and surface conditions, such as the pressure tube inspection datasets. The key outcome is that existing solutions require assumptions that cannot be guaranteed by the datasets that this thesis is focusing on, which drives the need to further explore data-driven solutions.

A novel automated unsupervised data-driven approach for automatically analysing collections of ultrasonic signals was introduced in Chapter 4. This method utilises the density-based clustering algorithm DBSCAN and individually characterises each signal, producing 2D maps that flag potentially flawed areas. The analysis consists of multiple steps that consecutively reduce and transform the search space. A key step of the proposed method is a novel subsampling method that reduces the search space and enables a more efficient density clustering. This method takes advantage of the way DBSCAN algorithm generates clusters, and selects only just-enough suitable representative samples to participate in the clustering. However, the main drawback of this method is that the computation improvements, relative to DBSCAN, are tied to the requirement of a low cardinality dataset, or the ability to lower the cardinality of the dataset through further grouping that might lower the granularity of the results. Furthermore a two-stage clustering procedure is presented, designed to efficiently utilise the DBSCAN clustering algorithm for grouping and noise reduction. Furthermore, a DBSCAN parameter estimation method is presented, which can be applied

across datasets captured by different ultrasonic inspection probes without utilising explicit knowledge of the probes characteristics. The proposed procedure is tested by utilising an automated evaluation procedure across a large number of real-world datasets captured from sections of pressure tubes containing flaws, showing satisfactory results.

Finally, the unsupervised analysis is extended to large-scale ultrasonic inspection datasets which are problematic for current memory intensive algorithms due to their multimodal nature, as well as their size. An anomaly detection method was introduced in Chapter 5 which targets datasets which, apart from the large size, exhibit various levels of heterogeneity across different neighbourhoods of the inspected surface. The proposed method utilises a neighbourhood-based transformation of the signals' variability measures as a means of creating a more homogenous feature space that allows for distance metric that is relevant across the surface, offering significant computational benefits. Compared to the popular standardisation method on a real-world dataset, the proposed feature transformation method produces a feature space that enables clusters that are less influenced by local geometric and surface properties, and more relevant to service induced surface wear. Furthermore, by reducing the transformed feature space into a single metric that is relevant to the whole of the surface, it is possible to eliminate the vast majority of data, resulting in the isolation of surface anomalies, including threatening flaws. This significant decrease in data size, should be able to provide more sustainable workload to specialised and computationally demanding algorithms responsible to calculating the depth of any flagged areas.

6.2 Future Work

A critical aspect of every data-driven method is the features that represent the population of available input data to the actual analysis. This thesis followed the traditional machine learning approach of manually generating a preselected set of features, which are then fed to a selection process to eliminate potentially uninformative features. This approach is based on the knowledge/biases of the analyst and, although it significantly reduces the problem to a state that allows a traditional machine learning algorithm to be effective, it leaves room for unexplored highly informative hidden features that could advance the performance of the clustering or classification algorithm.

An autoencoder is an unsupervised artificial neural network architecture that enables purely data-driven feature learning from raw input data (Figure 6.1). This architecture is capable of learning lower dimensional representations (codings) of unlabelled input data, while trying to learn to reconstruct the original input data.

By constraining the network through narrow hidden layers, the network is prevented from trivially copying the input to the output and is 'forced' to learn informative representations that can be used as features in a clustering or classification setting. Combined with 1D convolutional filters [168] to make the network more robust against time shifts present in captures ultrasonic signals, and a denoising approach that yields better representations [169], an autoencoder approach could offer a convenient and well-performing feature extraction solution for highly-populous unlabelled datasets similar to those presented in this thesis.

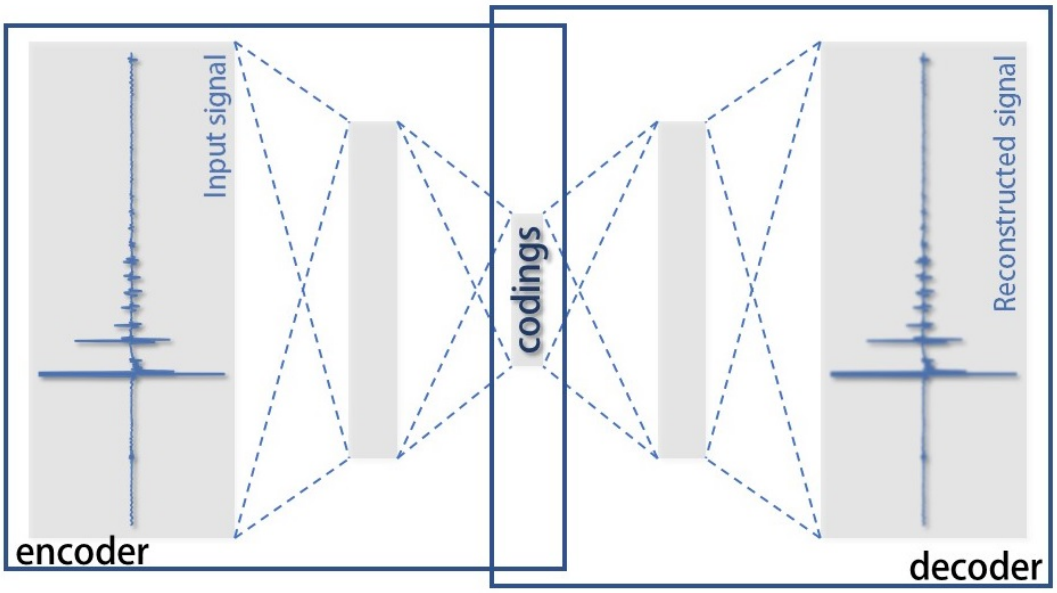


Figure 6.1: Autoencoder: The unsupervised artificial neural network architecture learns a lower dimension representation (codings) of the input data as it tries to reconstruct it. During this process the narrow hidden layer (bottleneck layer) captures discriminative signal features.

A way to approach the unsupervised analysis of ultrasonic inspection data is through the adaptation of deep learning models currently representing the current state-of-the-art in pixel-wise image segmentation tasks. These models are based on recent advances on self-supervised feature learning, a subset of unsupervised learning, where a model learns efficient representations guided by a supervisory signal related to self-defined pseudo labels [139]. Examples of such models are STEGO [138] and Leopart [139] which are capable of learning to segment complex images without utilising labels, and are currently the top performers on unsupervised segmentation tasks [140, 141].

6.3 Conclusion

This thesis presented methods for analysing high-resolution ultrasonic inspection data through unsupervised data-driven approaches, which enable the localisation of abnormal areas. These methods were tested on real-world inspection datasets and it was demonstrated that they can identify surface flaws in an efficient manner across large and heterogeneous datasets. The system-agnostic approach of the methods allowed their application to datasets captured by inspection probes of different specifications, without utilising explicit probe characteristics or a pre-defined a state of normality. This provides a potential step towards automating aspects of ultrasonic assessment that is a critical part of nuclear power plant asset health management. With modularity being in the forefront of current nuclear power plant research, which along with automation aim to increase the economic competitiveness of nuclear energy, independent and flexible automated inspection analysis systems which can handle large rates and volumes of data, will become increasingly important to inspection and maintenance personnel.

References

[1] X. Edelmann, "Ultrasonic examination of austenitic welds at reactor pressure vessels," *Nuclear Engineering and Design*, vol. 129, pp. 341–355, Aug. 1991. 1

[2] Z. Wang, Q. Cao, N. Luan, and L. Zhang, "Development of new pipeline maintenance system for repairing early-built offshore oil pipelines," in *2008 IEEE International Conference on Industrial Technology*, (Chengdu, China), pp. 1–6, IEEE, Apr. 2008. 1

[3] X. Zhao, H. Gao, G. Zhang, B. Ayhan, F. Yan, C. Kwan, and J. L. Rose, "Active health monitoring of an aircraft wing with embedded piezoelectric sensor/actuator network: I. Defect detection, localization and growth monitoring," *Smart Materials and Structures*, vol. 16, pp. 1208–1217, Aug. 2007. 1

[4] L. W. Schmerr and S.-J. Song, *Ultrasonic Nondestructive Evaluation Systems: Models and Measurements*. Boston, MA: Springer US, 2007. 1, 2, 46

[5] G.-M. Zhang and D. M. Harvey, "Contemporary ultrasonic signal processing approaches for nondestructive evaluation of multilayered structures," *Nondestructive Testing and Evaluation*, vol. 27, pp. 1–27, Mar. 2012. 1, 46

[6] T. Lardner, G. West, G. Dobie, and A. Gachagan, "An expert-systems approach to automatically determining flaw depth within candu pressure tubes," in *10th International Topical Meeting on Nuclear Plant Instrumentation, Control, and Human-Machine Interface Technologies (NPIC & HMIT)*, (San Francisco, United States), 2017. 1, 38

[7] A. Abbate, J. Koay, J. Frankel, S. Schroeder, and P. Das, "Signal detection and noise suppression using a wavelet transform signal processor: application to ultrasonic flaw detection," *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 44, pp. 14–26, Jan. 1997. 2

[8] K. Donohue and H. Cheah, "Spectral correlation filters for flaw detection," in *1995 IEEE Ultrasonics Symposium. Proceedings. An International Symposium*, vol. 1, (Seattle, WA, USA), pp. 725–728, IEEE, 1995. 2

[9] Y. Zhu and J. Weight, "Ultrasonic nondestructive evaluation of highly scattering materials using adaptive filtering and detection," *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 41, pp. 26–33, Jan. 1994. 2

[10] M. Kubinyi and R. Smid, "Ultrasonic denoising with a modified wavelet filter," in *Proceedings of the 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems*, (Prague, Czech Republic), pp. 479–482, IEEE, Sept. 2011. 2

[11] P. G. Heasler, T. T. Taylor, J. C. Spanner, S. R. Doctor, and J. D. Deffenbaugh, "Ultrasonic inspection reliability for intergranular stress corrosion cracks," Tech. Rep. NUREG/CR-4908, PNL-6196, 6888871, Nuclear Regulatory Commission, Washington, DC (USA). Div. of Engineering; Pacific Northwest Lab., Richland, WA (USA), July 1990. 2, 46

[12] J. Spanner, R. Badalamente, W. Rankin, and T. Triggs, "Human reliability impact on inservice inspection. Review and analysis of human performance in nondestructive testing (emphasizing ultrasonics). Volume 2," Tech. Rep. NUREG/CR–4436-VOL.2, Battelle Human Affairs Research Center, Seattle, WA (USA); Pacific Northwest Labs., Richland, WA (USA), United States, Mar. 1986. 2, 46

[13] D. Thompson, "EXPERIMENTAL ANALYSIS OF WAVE PROPAGATION IN RAILWAY TRACKS," *Journal of Sound and Vibration*, vol. 203, pp. 867–888, June 1997. 2, 21, 47

[14] R. M. Sanderson, "The Application of Finite Element Modelling to Guided Wave Testing Systems," in *AIP Conference Proceedings*, vol. 657, (Bellingham, Washington (USA)), pp. 256–263, AIP, 2003. ISSN: 0094243X. 2, 21, 47

[15] F. Chen and P. D. Wilcox, "The effect of load on guided wave propagation," *Ultrasonics*, vol. 47, pp. 111–122, Dec. 2007. 2, 21, 47

[16] S. K. Chakrapani, D. Barnard, and V. Dayal, "Finite element simulation of core inspection in helicopter rotor blades using guided waves," *Ultrasonics*, vol. 62, pp. 126–135, Sept. 2015. 2, 21, 47

[17] G. Dib, O. Karpenko, M. Haq, L. Udpa, and S. Udpa, "Advanced Signal Processing Algorithms in Structural Integrity Monitoring," *Procedia Engineering*, vol. 86, pp. 427–439, 2014. 2, 21, 47

[18] P. Rizzo, I. Bartoli, A. Marzani, and F. Lanza di Scalea, "Defect Classification in Pipes by Neural Networks Using Multiple Guided Ultrasonic Wave Features Extracted After Wavelet Processing," *Journal of Pressure Vessel Technology*, vol. 127, pp. 294–303, Aug. 2005. 2, 21, 47

[19] M. Trelinski, "Selected CANDU pressure tube degradation mechanisms and inspection related issues," in *17th World Conference on Nondestructive Testing*, (Shanghai, China), 2008. 2, 132

[20] J. Ye, S. Ito, and N. Toyama, "Computerized Ultrasonic Imaging Inspection: From Shallow to Deep Learning," *Sensors*, vol. 18, p. 3820, Nov. 2018. 2

[21] N. Amiri, G. Farrahi, K. R. Kashyzadeh, and M. Chizari, "Applications of ultrasonic testing and machine learning methods to predict the static & fatigue behavior of spot-welded joints," *Journal of Manufacturing Processes*, vol. 52, pp. 26–34, Apr. 2020. 2

[22] International Atomic Energy Agency, *Delayed hydride cracking in zirconium alloys in pressure tube nuclear reactors: final report of a coordinated research project 1998-2002*. Vienna: IAEA, 2004. OCLC: 1132053464. 3

[23] T. Lardner, G. West, G. Dobie, and A. Gachagan, "Automated sizing and classification of defects in CANDU pressure tubes," *Nuclear Engineering and Design*, vol. 325, pp. 25–32, Dec. 2017. ix, 3, 17, 37, 39

[24] M. Verleysen and D. François, "The Curse of Dimensionality in Data Mining and Time Series Prediction," in *Computational Intelligence and Bioinspired Systems* (D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, J. Cabestany, A. Prieto, and F. Sandoval, eds.), vol. 3512, pp. 758–770, Berlin, Heidelberg: Springer Berlin Heidelberg, 2005. Series Title: Lecture Notes in Computer Science. 4

[25] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, KDD'96, (Portland, Oregon), pp. 226–231, AAAI Press, Aug. 1996. 5, 68

[26] H. Zhao, A. Gachagan, G. Dobie, C. Wallace, and G. West, "Synthetic Aperture Focusing Technique for Correction of Poorly-Focused Ultrasonic Pressure Tube Inspection Data," in *9th European Workshop on Structural Health Monitoring (EWSHM 2018)*, p. 12, July 2018. 5

[27] International Atomic Energy Agency, "Operational & Long-Term Shutdown Reactors," 2021. 12

[28] I. A. E. Agency, "Under Construction Reactors," 2021. 12

[29] I. Aho-Mantila, O. Cronvall, U. Ehrnstén, H. Keinänen, R. Rintamaa, A. Saarenheimo, K. Simola, and E. Vesikari, "Lifetime prediction techniques for nuclear power plant systems," in *Nuclear Corrosion Science and Engineering*, pp. 449–470, Elsevier, 2012. 12

[30] J. Coble, P. Ramuhalli, L. Bond, J. W. Hines, and B. Upadhyaya, "A Review of Prognostics and Health Management Applications in Nuclear Power Plants," *International Journal of Prognostics and Health Management*, vol. 6, no. 3, 2015. 12, 16

[31] R. Ahmad and S. Kamaruddin, "An overview of time-based and condition-based maintenance in industrial application," *Computers & Industrial Engineering*, vol. 63, pp. 135–149, Aug. 2012. 15

[32] I. Wayan Ngarayana, T.-M.-D. Do, K. Murakami, and M. Suzuki, "Nuclear Power Plant Maintenance Optimisation: Models, Methods & Strategies," *Journal of Physics: Conference Series*, vol. 1198, p. 022005, Apr. 2019. 15

[33] L.-j. Huang, H. Jiang, Y. Chen, and S.-j. Chen, "Using Condition-based Maintenance and Reliability-centered Maintenance to Improve Maintenance in Nuclear Power Plants," in *Progress of Nuclear Safety for Symbiosis and Sustainability* (H. Yoshikawa and Z. Zhang, eds.), pp. 177–185, Tokyo: Springer Japan, 2014. 15

[34] R. Ayo-Imoru and A. Cilliers, "A survey of the state of condition-based maintenance (CBM) in the nuclear power industry," *Annals of Nuclear Energy*, vol. 112, pp. 177–188, Feb. 2018. 15

[35] A. K. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems and Signal Processing*, vol. 20, pp. 1483–1510, Oct. 2006. 15

[36] I. A. E. Agency, *Implementation strategies and tools for condition based monitoring at nuclear power plants*. No. 1551 in IAEA-TECDOC, Vienna: IAEA, 2007. OCLC: 1039247056. 16

[37] G. M. West, C. J. Wallace, and S. D. McArthur, "Combining models of behaviour with operational data to provide enhanced condition monitoring of AGR cores," *Nuclear Engineering and Design*, vol. 272, pp. 11–18, June 2014. 17, 26

[38] International Atomic Energy Agency, *Assessment and Management of Ageing of Major Nuclear Power Plant Components Important to Safety: CANDU Pressure Tubes*. No. 1037 in IAEA-TECDOC, Vienna: IAEA, 1998. Publisher: IAEA. 17, 31, 132

[39] M. Trelinski, "Inspection of CANDU reactor pressure tubes using ultrasonics," in *17th World Conference on Nondestructive Testing*, (Shanghai, China), Oct. 2008. 17, 31, 40

[40] P. Jackson, *Introduction to Expert Systems*. International computer science series, USA: Addison-Wesley Longman Publishing Co., Inc., 3rd ed., 1998. 18

[41] F. Puppe, *Systematic Introduction to Expert Systems: Knowledge Representations and Problem-Solving Methods*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1993. OCLC: 851800168. 18

[42] H. P. Hoplin and S. J. Erdman, "Expert systems: an expanded field of vision for business," in *Proceedings of the 1990 ACM SIGBDP conference on Trends and directions in expert systems - SIGBDP '90*, (Orlando, Florida, United States), pp. 1–16, ACM Press, 1990. 18, 19

[43] S. Mittal and C. L. Dym, "Knowledge Acquisition from Multiple Experts," *AI Magazine*, vol. 6, no. 2, pp. 32–36, 1985. 18

[44] M. J. Vlaanderen, *Automated Knowledge Acquisition for Expert Systems: An Overview*. PhD thesis, Faculy of Philosophy, Erasmus University Rotterdam, 1990. 18

[45] M. R. Aniba, S. Siguenza, A. Friedrich, F. Plewniak, O. Poch, A. Marchler-Bauer, and J. D. Thompson, "Knowledge-based expert systems and a proof-of-concept case study for multiple sequence alignment construction and analysis," *Briefings in Bioinformatics*, vol. 10, pp. 11–23, Oct. 2008. viii, 19

[46] V. Venkatasubramanian, R. Rengaswamy, S. N. Ka, and K. Yin, "A review of process fault detection and diagnosis Part III: Process history based methods," *Computers and Chemical Engineering*, vol. 27, no. 3, pp. 327–346, 2003. 20

[47] P. Garcia, R. Schirru, and P. Frutuoso Eˆ Melo, "A fuzzy data envelopment analysis approach for FMEA," *Progress in Nuclear Energy*, vol. 46, pp. 359–373, Jan. 2005. 20

[48] P. Murray, G. West, S. Marshall, and S. McArthur, "Automated in-core image generation from video to aid visual inspection of nuclear power plant cores," *Nuclear Engineering and Design*, vol. 300, pp. 57–66, Apr. 2016. 20

[49] M. G. Devereux, P. Murray, and G. M. West, "A new approach for crack detection and sizing in nuclear reactor cores," *Nuclear Engineering and Design*, vol. 359, p. 110464, Apr. 2020. 20

[50] G. West, S. McArthur, and D. Towle, "Industrial implementation of intelligent system techniques for nuclear power plant condition monitoring," *Expert Systems with Applications*, vol. 39, pp. 7432–7440, June 2012. 21, 26

[51] S. Schommer, V. H. Nguyen, S. Maas, and A. Zürbes, "Model updating for structural health monitoring using static and dynamic measurements," *Procedia Engineering*, vol. 199, pp. 2146–2153, 2017. 21

[52] B. Thacker, S. Doebling, F. Hemez, M. Anderson, J. Pepin, and E. Rodriguez, "Concepts of Model Verification and Validation," Tech. Rep. LA–14167, Los Alamos National Lab, Los Alamos, NM, United States, 2004. 22

[53] International Atomic Energy Agency, *On-line monitoring for improving performance of nuclear power plants. 2: Process and component condition monitoring and diagnostics*. No. NP-T-1.2 in IAEA nuclear energy series Technical reports, IAEA, 2008. 22

[54] F. Kojima, "Structural Health Monitoring of Nuclear Power Plants using Inverse Analysis in Measurements," 2009. 22

[55] A. V. Dyskin, H. Basarir, J. Doherty, M. Elchalakani, G. R. Joldes, A. Karrech, B. Lehane, K. Miller, E. Pasternak, I. Shufrin, and A. Wittek, "Computational monitoring in real time: review of methods and applications," *Geomechanics and Geophysics for Geo-Energy and Geo-Resources*, vol. 4, pp. 235–271, Sept. 2018. 22

[56] E. Zio, "Prognostics and Health Management of Industrial Equipment," in *Diagnostics and Prognostics of Engineering Systems: Methods and Techniques*, Hershey, PA, USA: IGI Global, 2013. 22

[57] J. Liu, R. Seraoui, V. Vitelli, and E. Zio, "Nuclear power plant components condition monitoring by probabilistic support vector machine," *Annals of Nuclear Energy*, vol. 56, pp. 23–33, June 2013. 22

[58] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap," *Mechanical Systems and Signal Processing*, vol. 138, p. 106587, Apr. 2020. 23

[59] A. Stetco, F. Dinmohammadi, X. Zhao, V. Robu, D. Flynn, M. Barnes, J. Keane, and G. Nenadic, "Machine learning methods for wind turbine condition monitoring: A review," *Renewable Energy*, vol. 133, pp. 620–635, Apr. 2019. 23

[60] C. Berry, G. West, S. McArthur, and A. Rudge, "Semi-supervised learning approach for crack detection and identification in advanced gas-cooled reactor graphite bricks," in *10th International Topical Meeting on Nuclear*

Plant Instrumentation, Control, and Human-Machine Interface Technologies, NPIC and HMIT 2017, NPIC and HMIT 2017, (San Francisco, United States), June 2017. 24, 26

[61] P. Werbos, "Backpropagation through time: what it does and how to do it," *Proceedings of the IEEE*, vol. 78, pp. 1550–1560, Oct. 1990. 24

[62] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems*. Beijing [China] ; Sebastopol, CA: O'Reilly Media, Inc, second edition ed., 2019. 24, 76

[63] V. Venkatasubramanian and K. Chan, "A neural network methodology for process fault diagnosis," *AIChE Journal*, vol. 35, pp. 1993–2002, Dec. 1989. 25

[64] V. Venkatasubramanian, R. Vaidyanathan, and Y. Yamamoto, "Process fault detection and diagnosis using neural networks—I. steady-state processes," *Computers & Chemical Engineering*, vol. 14, pp. 699–712, July 1990. 25

[65] X. Han, S. Tian, J. A. Romagnoli, H. Li, and W. Sun, "PCA-SDG Based Process Monitoring and Fault Diagnosis: Application to an Industrial Pyrolysis Furnace," *IFAC-PapersOnLine*, vol. 51, no. 18, pp. 482–487, 2018. 25

[66] G. Hu, T. Zhou, and Q. Liu, "Data-Driven Machine Learning for Fault Detection and Diagnosis in Nuclear Power Plants: A Review," *Frontiers in Energy Research*, vol. 9, p. 663296, May 2021. 25

[67] B. T. Jiang, J. Zhou, and X. B. Huang, "Artificial Neural Networks in Condition Monitoring and Fault Diagnosis of Nuclear Power Plants: A Concise Review," in *Volume 2: Nuclear Policy; Nuclear Safety, Security, and Cyber Security; Operating Plant Experience; Probabilistic Risk Assessments; SMR and Advanced Reactors*, (Virtual, Online), p. V002T08A032, American Society of Mechanical Engineers, Aug. 2020. 25

[68] F.-C. Chen and M. R. Jahanshahi, "NB-CNN: Deep Learning-Based Crack Detection Using Convolutional Neural Network and Naïve Bayes Data Fusion," *IEEE Transactions on Industrial Electronics*, vol. 65, pp. 4392–4400, May 2018. 25

[69] Y. H. Chae, S. G. Kim, H. Kim, J. T. Kim, and P. H. Seong, "A methodology for diagnosing FAC induced pipe thinning using accelerometers and deep learning models," *Annals of Nuclear Energy*, vol. 143, p. 107501, Aug. 2020. 26

[70] G. M. West, S. D. McArthur, and D. Towle, "Knowledge-directed characterization of nuclear power plant reactor core parameters," *Nuclear Engineering and Design*, vol. 241, pp. 4013–4025, Sept. 2011. 26

[71] H. Wang, M.-j. Peng, Y. Yu, H. Saeed, C.-m. Hao, and Y.-k. Liu, "Fault identification and diagnosis based on KPCA and similarity clustering for nuclear power plants," *Annals of Nuclear Energy*, vol. 150, Jan. 2021. 26

[72] Canadian Nuclear Association, "THE CANADIAN NUCLEAR FACTBOOK 2019," 2019. xvi, 27

[73] W. J. Garland, *The Essential CANDU, A Textbook on the CANDU Nuclear Power Plant Technology*. University Network of Excellence in Nuclear Engineering (UNENE), 2014. 28

[74] Inductiveload, "Schematic diagram of the pressurised heavy water cooled version of a CANDU (CANada Deuterium-Uranium) nuclear reactor.," Oct. 2007. [Online]. Wikimedia Foundation. Available from: 'https://commons.wikimedia.org/wiki/File:CANDU_Reactor_Schematic.svg' [Accessed 03 January 2019]. viii, 29

[75] P. Boczar, "CANDU nuclear reactor designs, operation and fuel cycle," in *Nuclear Fuel Cycle Science and Engineering*, pp. 278–299, Elsevier, 2012. viii, 28, 32

[76] B. Rouben, "CANDU Fuel Management Course," Nov. 2003. [Online]. McMaster University, Hamilton, Ontario, Canada. Available from: "https://canteach.candu.org/ content%20library/ 20031101.pdf" [Accessed 05 September 2021]. 28

[77] M. H. A. Piro, F. Wassermann, S. Grundmann, B. W. Leitch, and C. Tropea, "Progress in on-going experimental and computational fluid dynamic investigations within a CANDU fuel channel," *Nuclear Engineering and Design*, vol. 299, pp. 184–200, Apr. 2016. viii, 32

[78] M. Trelinski, "Application of Ultrasonic Testing Methods for Volumetric and Surface Inspection of CANDU Pressure Tubes," in *1st Pan American Conference for Nondestructive Testing (PACNDT '98)*, (Toronto, Canada), Canadian Society for NDT, Sept. 1998. 35

[79] Y.-J. Kim, S.-L. Kwak, J.-S. Lee, and Y.-W. Park, "Integrity evaluation system of CANDU reactor pressure tube," *KSME International Journal*, vol. 17, pp. 947–957, July 2003. 36

[80] C. Wallace, G. West, P. Zacharis, G. Dobie, and A. Gachagan, "Experience, testing and future development of an ultrasonic inspection analysis defect decision support tool for CANDU reactors," in *11th Nuclear Plant Instrumentation, Control and Human-Machine Interface Technologies (NPIC&HMIT)*, (Orlando, Florida, United States), Feb. 2019. viii, 37, 38, 43, 110

[81] G. Schreiber, ed., *Knowledge engineering and management: the CommonKADS methodology*. Cambridge, MA, United States: MIT Press, 2000. 38

[82] W. B. R. Clendening, "Pressure Tube To End Fitting Rolled Joints," Dec. 2002. [Presentation - Online]. Available from: 'https://www.nrc.gov/docs/ML0300/ ML030020213.pdf' [Accessed 06 September 2021]. ix, 40

[83] D. Ensminger and L. J. Bond, *Ultrasonics: fundamentals, technologies, and applications*. Mechanical engineering, Boca Raton, FL, United States: Taylor & Francis, 3rd ed ed., 2012. 45

[84] R. S. Gilmore, "Industrial ultrasonic imaging and microscopy," *Journal of Physics D: Applied Physics*, vol. 29, pp. 1389–1417, June 1996. 45

[85] F. A. Firestone, "The Supersonic Reflectoscope, an Instrument for Inspecting the Interior of Solid Parts by Means of Sound Waves," *The Journal of the Acoustical Society of America*, vol. 17, pp. 287–299, Jan. 1946. 45

[86] L. W. Schmerr, *Fundamentals of Ultrasonic Nondestructive Evaluation: A Modeling Approach*. Springer Series in Measurement Science and Technology, Cham: Springer International Publishing, 2016. 46

[87] P. Burrascano, S. Callegari, A. Montisci, M. Ricci, and M. Versaci, eds., *Ultrasonic Nondestructive Evaluation Systems: Industrial Application Issues*. Cham: Springer International Publishing : Imprint: Springer, 1st ed. 2015 ed., 2015. 46

[88] M. T. Hagan, H. B. Demuth, M. H. Beale, and O. De Jésus, *Neural network design*. s.L: Martin T. Hagan, 2nd edition ed., 2014. 48

[89] S. S. Haykin, *Neural networks: a comprehensive foundation*. Upper Saddle River, NJ, United States: Prentice Hall, 2nd ed., 1999. 48

[90] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, Oct. 1986. 48

[91] H. Robbins and S. Monro, "A Stochastic Approximation Method," *The Annals of Mathematical Statistics*, vol. 22, pp. 400–407, Sept. 1951. 48

[92] J. Kiefer and J. Wolfowitz, "Stochastic Estimation of the Maximum of a Regression Function," *The Annals of Mathematical Statistics*, vol. 23, pp. 462–466, Sept. 1952. 48

[93] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," Jan. 2017. arXiv:1412.6980. 48

[94] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, Nov. 2016. 'http://www.deeplearningbook.org'. 48, 52

[95] LearnDataSci, "Node connections of an artificial neural network," June 2016. [Online]. Wikimedia Foundation. Available from: 'https://commons.wikimedia.org/wiki/ File:Artificial_Neural_Network.jpg' [Accessed 03 September 2021]. ix, 49

[96] F. W. Margrave, K. Rigas, D. A. Bradley, and P. Barrowcliffe, "The use of neural networks in ultrasonic flaw detection," *Measurement*, vol. 25, pp. 143–154, Mar. 1999. 49

[97] T. D'Orazio, M. Leo, A. Distante, C. Guaragnella, V. Pianese, and G. Cavaccini, "Automatic ultrasonic inspection for internal defect detection in composite materials," *NDT & E International*, vol. 41, pp. 145–154, Mar. 2008. 49

[98] S. Iyer, S. K. Sinha, B. R. Tittmann, and M. K. Pedrick, "Ultrasonic signal processing methods for detection of defects in concrete pipes," *Automation in Construction*, vol. 22, pp. 135–148, Mar. 2012. 50

[99] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 674–693, July 1989. 50

[100] M. Vetterli and C. Herley, "Wavelets and filter banks: theory and design," *IEEE Transactions on Signal Processing*, vol. 40, pp. 2207–2232, Sept. 1992. 50

[101] P. Gammell, "Improved ultrasonic detection using the analytic signal magnitude," *Ultrasonics*, vol. 19, pp. 73–76, Mar. 1981. 50

[102] G. K. Sharma, A. Kumar, C. Babu Rao, T. Jayakumar, and B. Raj, "Short time Fourier transform analysis for understanding frequency dependent attenuation in austenitic stainless steel," *NDT & E International*, vol. 53, pp. 1–7, Jan. 2013. 50

[103] M. Ge, R. Du, G. Zhang, and Y. Xu, "Fault diagnosis using support vector machine with an application in sheet metal stamping operations," *Mechanical Systems and Signal Processing*, vol. 18, pp. 143–159, Jan. 2004. 50

[104] F. Kaytez, M. C. Taplamacioglu, E. Cam, and F. Hardalac, "Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines," *International Journal of Electrical Power & Energy Systems*, vol. 67, pp. 431–438, May 2015. 50

[105] P. Yang and Q. Li, "Wavelet transform-based feature extraction for ultrasonic flaw signal classification," *Neural Computing and Applications*, vol. 24, pp. 817–826, Mar. 2014. 50

[106] Larhmam, "SVM Schematic," Oct. 2018. [Online]. Wikimedia Foundation. Available from: 'https://commons.wikimedia.org/wiki/File:SVM_margin.png' [Accessed 03 September 2021]. ix, 51

[107] C. J. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, vol. 2, pp. 121–167, June 1998. 51

[108] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, Sept. 1995. 51

[109] M. Hassan, R. Rajkumar, D. Isa, and R. Arelhi, "Pipeline Defect Classification by Using Non-Destructive Testing and Improved Support Vector Machine Classification," *International Journal of Engineering and Innovative Technology*, vol. 2, no. 7, 2013. 51, 52

[110] K. Lee and V. Estivill-Castro, "Support Vector Machine Classification of Ultrasonic Shaft Inspection Data Using Discrete Wavelet Transform," *Proceedings of the International Conference on Artificial Intelligence, IC-AI'04*, vol. 2, Jan. 2004. 51

[111] D. C. Ciresan, U. Meier, J. Masci, L. M. Gambardella, and J. Schmidhuber, "Flexible, High Performance Convolutional Neural Networks for Image Classification," in *Twenty-Second International Joint Conference on Artificial Intelligence*, IJCAI'11, (Barcelona, Catalonia, Spain), pp. 1237–1242, AAAI Press, July 2011. 52

[112] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *25th International Conference on Neural Information Processing (NIPS'12)*, vol. 1, (Lake Tahoe, Nevada, United States), pp. 1097–1105, Curran Associates, Inc., Dec. 2012. 52

[113] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014. 52

[114] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Deep learning for time series classification: a review," *Data Mining and Knowledge Discovery*, vol. 33, pp. 917–963, July 2019. 52

[115] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing*, vol. 187, pp. 27–48, Apr. 2016. 52

[116] N. Munir, H.-J. Kim, S.-J. Song, and S.-S. Kang, "Investigation of deep neural network with drop out for ultrasonic flaw classification in weldments," *Journal of Mechanical Science and Technology*, vol. 32, pp. 3073–3080, July 2018. 52, 53

[117] N. Munir, H.-J. Kim, J. Park, S.-J. Song, and S.-S. Kang, "Convolutional neural network for ultrasonic weldment flaw classification in noisy conditions," *Ultrasonics*, vol. 94, pp. 74–81, Apr. 2019. 52, 53

[118] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep Learning for Computer Vision: A Brief Review," *Computational Intelligence and Neuroscience*, vol. 2018, Feb. 2018. 53

[119] S. Cantero-Chinchilla, P. D. Wilcox, and A. J. Croxford, "Deep learning in automated ultrasonic NDE–developments, axioms and opportunities," *NDT & E International*, p. 102703, 2022. 53

[120] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 International Conference on Engineering and Technology (ICET)*, pp. 1–6, Aug. 2017. 53

[121] Aphex34, "Typical CNN architecture," Dec. 2015. [Online]. Wikimedia Foundation. Available from: 'https://commons.wikimedia.org/wiki/File:Typical_cnn.png' [Accessed 02 September 2021]. ix, 53

[122] M. Meng, Y. J. Chua, E. Wouterson, and C. P. K. Ong, "Ultrasonic signal classification and imaging system for composite materials via deep convolutional neural networks," *Neurocomputing*, vol. 257, pp. 128–135, Sept. 2017. 54

[123] D. Medak, L. Posilović, M. Subašić, M. Budimir, and S. Lončarić, "Automated defect detection from ultrasonic images using deep learning," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 68, no. 10, pp. 3126–3134, 2021. 54, 55

[124] M. Tan, R. Pang, and Q. V. Le, "Efficientdet: Scalable and efficient object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10781–10790, 2020. 54

[125] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018. 54

[126] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988, 2017. 54

[127] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 6, p. 60, July 2019. 54

[128] R. Atienza, "Improving Model Generalization by Agreement of Learned Representations From Data Augmentation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 372–381, 2022. 54

[129] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common Objects in Context," in *Computer Vision – ECCV 2014* (D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, eds.), Lecture Notes in Computer Science, (Cham), pp. 740–755, Springer International Publishing, 2014. 55

[130] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6517–6525, July 2017. 55

[131] L. Posilović, D. Medak, F. Milković, M. Subašić, M. Budimir, and S. Lončarić, "Deep learning-based anomaly detection from ultrasonic images," *Ultrasonics*, vol. 124, p. 106737, Aug. 2022. 55

[132] X. P. Zhu, P. Rizzo, A. Marzani, and J. Bruck, "Ultrasonic guided waves for nondestructive evaluation/structural health monitoring of trusses," *Measurement Science and Technology*, vol. 21, p. 045701, Mar. 2010. 56

[133] R. You, Y. Yao, J. Shi, K. Zheng, and K.-H. Wang, "Feature-selective clustering for ultrasonic-based automatic defect detection in FRP structures," *Chemometrics and Intelligent Laboratory Systems*, vol. 157, pp. 35–42, Oct. 2016. 57

[134] X. Wu, V. Kumar, J. Ross Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z.-H. Zhou, M. Steinbach, D. J. Hand, and D. Steinberg, "Top 10 algorithms in data mining," *Knowledge and Information Systems*, vol. 14, pp. 1–37, Jan. 2008. 57

[135] B. Bahmani, B. Moseley, A. Vattani, R. Kumar, and S. Vassilvitskii, "Scalable k-means++," *Proceedings of the VLDB Endowment*, vol. 5, pp. 622–633, Mar. 2012. 57

[136] D.-W. Kim, K. H. Lee, and D. Lee, "Detecting clusters of different geometrical shapes in microarray gene expression data," *Bioinformatics*, vol. 21, pp. 1927–1934, May 2005. 57

[137] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in *International Conference on Learning Representations (ICLR) 2021*, Mar. 2022. 60

[138] M. Hamilton, Z. Zhang, B. Hariharan, N. Snavely, and W. T. Freeman, "Unsupervised Semantic Segmentation by Distilling Feature Correspondences," in *International Conference on Learning Representations (ICLR) 2022*, Mar. 2022. 60, 142

[139] A. Ziegler and Y. M. Asano, "Self-Supervised Learning of Object Parts for Semantic Segmentation," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (New Orleans, LA, USA), pp. 14482–14491, IEEE, June 2022. 60, 142

[140] Papers with Code Website, "Cityscapes test Benchmark (Unsupervised Semantic Segmentation)," Dec. 2022. [Online]. 'https://paperswithcode.com/sota/ unsupervised-semantic-segmentation-on' [Accessed 04 December 2022]. 60, 142

[141] Papers with Code Website, "PASCAL VOC 2012 val Benchmark (Unsupervised Semantic Segmentation)," Dec. 2022. [Online]. 'https://paperswithcode.com/sota/unsupervised-semantic-segmentation-on-pascal-1' [Accessed 04 December 2022]. 60, 142

[142] D. Preuveneers, I. Tsingenopoulos, and W. Joosen, "Resource Usage and Performance Trade-offs for Machine Learning Models in Smart Environments," *Sensors (Basel, Switzerland)*, vol. 20, p. 1176, Feb. 2020. 60

[143] A. L. Bowler, M. P. Pound, and N. J. Watson, "A review of ultrasonic sensing and machine learning methods to monitor industrial processes," *Ultrasonics*, vol. 124, p. 106776, Aug. 2022. 64

[144] I. B. Gonçalves, A. Leiria, and M. M. M. Moura, "STFT or CWT for the detection of Doppler ultrasound embolic signals," *International Journal for Numerical Methods in Biomedical Engineering*, vol. 29, no. 9, pp. 964–976, 2013. 64

[145] Z. Zhang, Q. K. Telesford, C. Giusti, K. O. Lim, and D. S. Bassett, "Choosing Wavelet Methods, Filters, and Lengths for Functional Brain Network Construction," *PLoS ONE*, vol. 11, June 2016. 64

[146] J. Sander, M. Ester, H.-P. Kriegel, and X. Xu, "Density-Based Clustering in Spatial Databases: The Algorithm GDBSCAN and Its Applications," *Data Mining and Knowledge Discovery*, vol. 2, pp. 169–194, June 1998. 68

[147] S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, pp. 129–137, Mar. 1982. 68

[148] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, pp. 790–799, Aug. 1995. 68

[149] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker Verification Using Adapted Gaussian Mixture Models," *Digital Signal Processing*, vol. 10, pp. 19–41, Jan. 2000. 69

[150] P. Fränti and S. Sieranoja, "How much can k-means be improved by using better initialization and repeats?," *Pattern Recognition*, vol. 93, pp. 95–112, Sept. 2019. 69

[151] T. Zhang, R. Ramakrishnan, and M. Livny, "BIRCH: an efficient data clustering method for very large databases," *ACM SIGMOD Record*, vol. 25, pp. 103–114, June 1996. 70

[152] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander, "OPTICS: ordering points to identify the clustering structure," *ACM SIGMOD Record*, vol. 28, pp. 49–60, June 1999. 70

[153] K. Chidananda Gowda and G. Krishna, "Agglomerative clustering using the concept of mutual nearest neighbourhood," *Pattern Recognition*, vol. 10, pp. 105–112, Jan. 1978. 70

[154] F. Murtagh and P. Legendre, "Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion?," *Journal of Classification*, vol. 31, pp. 274–295, Oct. 2014. 70

[155] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 888–905, Aug. 2000. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence. 70

[156] A. Ng, M. Jordan, and Y. Weiss, "On Spectral Clustering: Analysis and an algorithm," in *Advances in Neural Information Processing Systems*, vol. 14, MIT Press, 2001. 70

[157] B. J. Frey and D. Dueck, "Clustering by Passing Messages Between Data Points," *Science*, vol. 315, pp. 972–976, Feb. 2007. Publisher: American Association for the Advancement of Science. 70

[158] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and d. Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011. x, xii, 71, 72, 74, 85, 90, 91

[159] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," *Applied Soft Computing*, vol. 97, p. 105524, Dec. 2020. 76

[160] A. Karami and R. Johansson, "Choosing DBSCAN Parameters Automatically using Differential Evolution," *International Journal of Computer Applications*, vol. 91, pp. 1–11, Apr. 2014. 96

[161] P. B. Nagy, L. Adler, and J. H. Rose, "Effects of Acoustic Scattering at Rough Surfaces on the Sensitivity of Ultrasonic Inspection," in *Review of Progress in Quantitative Nondestructive Evaluation: Volumes 12A and 12B* (D. O. Thompson and D. E. Chimenti, eds.), pp. 1775–1782, Boston, MA: Springer US, 1993. 118

[162] W. Choi, F. Shi, M. J. S. Lowe, E. A. Skelton, R. V. Craster, and W. L. Daniels, "Rough surface reconstruction of real surfaces for numerical simulations of ultrasonic wave scattering," *NDT & E International*, vol. 98, pp. 27–36, Sept. 2018. 118

[163] H. Zhao, A. Gachagan, G. Dobie, and T. Lardner, "Wavelet analysis of poorly-focused ultrasonic signal of pressure tube inspection in nuclear industry," *AIP Conference Proceedings*, vol. 1949, p. 170004, Apr. 2018. 123

[164] P. Zacharis, G. West, G. Dobie, T. Lardner, and A. Gachagan, "Data-Driven Analysis of Ultrasonic Inspection Data of Pressure Tubes," *Nuclear Technology*, vol. 202, pp. 153–160, June 2018. 123

[165] R. J. G. B. Campello, D. Moulavi, and J. Sander, "Density-Based Clustering Based on Hierarchical Density Estimates," in *Advances in Knowledge Discovery and Data Mining* (J. Pei, V. S. Tseng, L. Cao, H. Motoda, and G. Xu, eds.), Lecture Notes in Computer Science, (Berlin, Heidelberg), pp. 160–172, Springer, 2013. 123

[166] D. Sculley, "Web-scale k-means clustering," in *Proceedings of the 19th international conference on World wide web*, WWW '10, (New York, NY, USA), pp. 1177–1178, Association for Computing Machinery, Apr. 2010. 129

[167] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, l. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa,

P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020. 135

[168] A. Bowler, M. Pound, and N. Watson, "Convolutional feature extraction for process monitoring using ultrasonic sensors," *Computers & Chemical Engineering*, vol. 155, p. 107508, Dec. 2021. 141

[169] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion," *Journal of Machine Learning Research*, vol. 11, no. 110, pp. 3371–3408, 2010. 141