UNIVERSITY OF THE AEGEAN
Department of Shipping Trade and Transport

&

UNIVERSITY OF WEST ATTICA
Department of Industrial Design and Production Engineering

JOINT
MASTER DEGREE PROGRAM IN
«NEW TECHNOLOGIES IN SHIPPING AND TRANSPORT»

THESIS:

LOCOMOTIVE FAULT PROGNOSIS VIA MACHINE LEARNING

Student Name:

HARRY (CHARALAMPOS) P. GEORGALIS

Supervisor Name:

Dr. NIKOLAOU GREGORY

Place:

PIRAEUS, GREECE

Date:

19 / 02 / 2021

UNIVERSITY OF THE AEGEAN
Department of Shipping Trade and Transport

&

UNIVERSITY OF WEST ATTICA
Department of Industrial Design and Production Engineering

Μέλη Εξεταστικής Επιτροπής

GRIGORIOS NIKOLAOU

Digitally signed by GRIGORIOS NIKOLAOU
Date: 2021.03.09 11:38:24 +02'00'

Νικολάου Γρηγόριος

Παπουτσιδάκης Μιχαήλ

Δρόσος Χρήστος

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

Ο κάτωθι υπογεγραμμένος Γεωργαλής Π. Χάρης του Παναγιώτη, με αριθμό μητρώου 116 φοιτητής του Διϊδρυματικού Προγράμματος Μεταπτυχιακών Σπουδών «Νέες Τεχνολογίες στη Ναυτιλία και τις Μεταφορές» του Τμήματος Μηχανικών Βιομηχανικής Σχεδίασης και Παραγωγής της Σχολής Μηχανικών Πανεπιστημίου Δυτικής Αττικής, δηλώνω υπεύθυνα ότι:
«Είμαι συγγραφέας αυτής της μεταπτυχιακής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος.
Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του διπλώματός μου».

Ο Δηλών

Γεωργαλής Π. Χάρης

PLAGIARISM STATEMENT

I certify that this thesis is my own work, based on my personal study and research and that I have acknowledged all material and sources used in its preparation, whether they are reports, articles, books, lecture notes, and any other kind of document, personal or electronic communication. I also certify that this research has not previously been submitted for assessment in any other unit, except where specific permission has been granted from all unit coordinators involved, or at any other time in this unit, and that I have not copied in part or whole or otherwise plagiarised the work of other students and/or persons.

Harry P. Georgalis

LOCOMOTIVES FAULT PROGNOSIS

Harry P. Georgalis
University of West Attica
University of the Aegean

Author's Notes
Harry P. Georgalis, B.Tech., Electrical Engineer, Rolling Stock Engineer, Technical Office Manager, harisgeor@gmail.com, c.georgalis@trainose.gr.

ABSTRACT

Railway has played a vital role in transportation of both goods and people historically, continuing to hold an important share of the market with great potentials. Locomotives are the main part of a train and as with any mechanism; maintenance and troubleshooting are of critical importance. Except for corrective and preventing maintenance, the new trend in all industries is the fault prognostics, also known as predictive maintenance, whose goal is to detect an upcoming breakdown. Almost every mechanical compartment uses some type of bearings. So, this research focus on two main pillars, the first part is about to review the available technical manual and to count the bearings used in locomotives, while the second part is about to construct a deep machine learning model for bearings fault diagnosis and prognosis based on secondary data.

Keywords: Bearings, BiLSTM, Condition-Based Maintenance, Deep Learning, Fault Prognosis, Genetic Algorithms, K-means, Linear Regression, Locomotives, Modelling, Multi-class Classification, Predictive Maintenance, Prognostics, Regression, Rolling Stock, Signal Processing, Supervised Machine Learning, Vibrations.

ACKNOWLEDGEMENTS

This research project was undertaken as part of the Postgraduate Program of Studies, Master of Science in "New Technologies in Shipping & Transport", which is a joint program of the Department of Shipping trade & Transport and Department of Industrial Design and Production Engineering, collaborated by the University of the Aegean and the University of West Attica respectively.

PREFACE

The initial goal of this thesis was to construct a holistic model via machine learning for locomotives faults prediction. However, this changed because TRAIOSE S.A finally denied access to the SCADA and data loggers. So, a shift forward to different objectives was obligatory. Inside a short period of time, around 6 months, the whole process of the thesis has been included a literature overview, data and information collection, analysis, modeling, evaluation, editing and finally the official presentation. All bearings of Greek fleet locomotives have been tried to be accounted at the first part of the thesis and additionally a high-end method has been designed in order to construct a deep machine learning model for fault predictions based on secondary data for outperforming the older approaches. APA style version 7 is used all the way, compiled with the Universities guidelines.

I would like to thank my supervisor, Dr. Nikolaou Gregory, for his priceless guidance and his excellent coaching. Also, very special thanks to my partner and my mother for their patience and their constant support. Finally, I want to devote this work to my pass-away father without his help I could not have achieved anything.

Harry P. Georgalis

Piraeus, 2021

"Kleos is gained on the battlefield"

Athens, 2021

"Τό κλέος κερδαίνει ἐν τῷ πεδίῳ τῆς μάχης"

Ἐν Ἀθήναις, 2021

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATION

Abbreviation	Meaning
ACC	Accuracy
ADF	Augmented Dickey Fuller
ALCO	American Locomotive Company
AUC	Area Under the Curve
BF	Butterworth Filter
BiLSTM	Bidirectional Long Short-Term Memory
BRL	Basic Rating Life
BS	Bearings Sum
CBM	Condition-Based Maintenance
DNN	Deep Neural Network
DTCWT	Dual-Tree Complex Wavelet Transform
DWT	Discrete Wavelet Transform
FFT	Fast Fourier Transformation
FN	False Negative
FP	False Positive
H	Hypothese
HT	Hilbert Transform
IMS	Intelligent Maintenance Systems
KF	Kalman Filter
LM	Locomotive
RL	Linear Regression
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
ML	Machine Learning
MLW	Montreal Locomotive Works
ND	Normal Distribution
NND	Non-Normal Distribution
PE	Permutation Entropy
PRE	Precision
REC	Recall
RMSE	Root Mean Square Error
ROC	Receiver Operating Characteristics
RQ	Research Question
RUL	Remaining Useful Life
SE	Shannon Entropy
TN	True Negative
TP	True Positive

LOCOMOTIVES FAULT PROGNOSIS

1. INTRODUCTION

The railway has been a major player in transportation of people and products historically (Schwaller, 1997). Even though railway has faced a decline in the market share for the last years, a reverse trend forwards this kind of inland transport happens because of the promoting policy in European Union, mostly driven by its better environmental footprint (European Commission, 2019; European Parliament, 2019).

Locomotive (LM) is the main traction system of a train and the most important part indeed. From an engineering perspective, the possible breakdowns are of highly importance because they play a crucial role in maintenance scheduling and cost overall.

Except for corrective and preventative maintenance (Dhillon, 2006), the state-of-art technique called preventive maintenance that is based on the fault prognosis of spare parts before their breakdown (Mobley, 2002). Machine Learning (ML) and Artificial Intelligence are used for simulation, modeling and predictions of processes in complex systems.

Rotating machinery is commonly used (Li et al., 2019) which usually bearings are installed on. Furthermore, bearings thought to be one of the major reasons for rotating machines breakdowns (Boškoski et al., 2015a).

This research focus on two main pillars, the first part is about to review the available technical manual and to count the bearings used in locomotives, while the second part is about to construct a deep ML model for bearings fault prognosis based on secondary data.

2. BACKGROUND

2.1 Locomotives

A LM is a complex construction, combining plethora of mechanisms which integrate each other. Every mechanism is prone to breakdowns, requiring corrective actions to fix the occurred fault. On the other hand, major mechanisms, as listed below, are scheduled to take maintenance in order to prevent a future failure (Janicki, Reinhard, & Rüffer, 2013).

i. Chassis
ii. Bogies
iii. Diesel engine
iv. Electric motor
v. Transmission gear

Chassis and diesel engine are the most critical parts of LMs which are scheduled for routine maintenance depending mostly on distance (km) and working time (h) respectively. In contrast, in predictive techniques the real condition of the equipment is calculated.

Sensors can be mounted so as that measurements are taken and used as inputs, like (Dhillon, 2006) :

› Vibration
› Temperature
› Tribology
› Ultrasonic
› Acoustics

Moreover, Kostic et al. (2011) outlined extra measurements for electric LMs, such as:

- Voltage
- Current
- Tongue
- Velocity
- Acceleration
- Power
- Traction Force

Any subsystem is prone to breakdown, so every possible fault could be taken as output. Research has been conducted and shown that some of possibly fault can occur in:

- Power traction inverter (Fei et al., 2018)
- Electro-pneumatic brake (Niu et al., 2015)
- Gearbox (Gao et al., 2019)
- Diesel engine (F. Feng et al., 2011)
  - Lubrication system (Gao et al., 2019)
  - Cooling system (Moussa Nahim et al., 2016)
  - Valves (Flett & Bone, 2016)
  - Pump (X. Wang et al., 2014)
  - Bearings (Abdelkrim et al., 2019)
- Electric motor
  - Phase to phase short-circuit (Z. Wang et al., 2016)
  - Bearings (Glowacz et al., 2018)
  - Winding
  - Stator (Glowacz et al., 2018)
  - Rotor (Cheng & Xiong, 2018)

2.2 Machine Learning

Human learns from experience as equal machines learn from data. Vapnik (1998) stated that “The learning process is a process of choosing an appropriate function from a given set of functions”. The basic concept of ML modeling is built on three pillars as illustrated in **Fig. 1**.

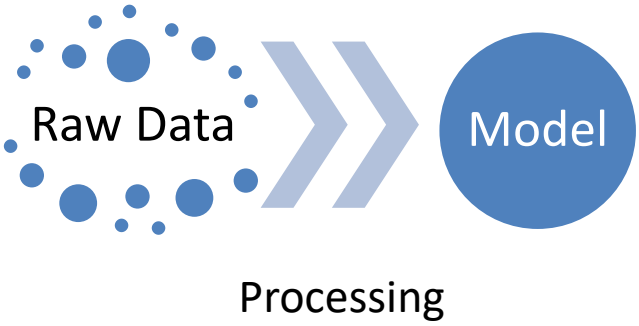


Figure 1 Concept

Data collection, commonly called as data mining, is used by many tools to discover knowledge. Databases and data warehouses are used to save and manage all required information (Han, Kamber & Pei, 2012).

Some of the main ML categories are (Marsland, 2009) :

› Supervised

Learning from examplars is the method that the algorithm trained with a set of examples and the desired responses. After the training period, the algorithm can generalise and find the right responses based on the inputs (Marsland, 2009).

› Unsupervised

In this case, there is not any supervisor. Input regularities and certain patterns are recognised as general forms. Here, clustering is used for density estimation (Alpaydin, 2010).

› Evolutionary

Evolutionary and genetic programming represent this category with Genetic Algorithms (GAs) and Memetic algorithms to be characteristic examples (Eiben & Smith, 2003; Marsland, 2009). In this category, there are methods like ant colony optimasation, with heuristic and meta-heuristic techniques (Eiben & Smith, 2003; Dorigo & Stutzle, 2004) and other bio-inspired systems

such as cellular, neural, developmental, immune, behavioral, collective (Floreano & Mattiussi, 2008) and finally swarm intelligence (Kennedy & Eberhart, 2001).

› Reinforcement

A decision-maker, called agent, try to find a solution to a problem by acting and receiving reward or penalties. As a result, total reward maximisation is achieved via the choice of the best policy (Alpaydin, 2010). Here, the most famous technique is the Markov decision process (Marsland, 2009).

Luger (2009) categorised the ML as :

› Symbol-Based
› Connectionist
› Genetic and emergent
› Probabilistic

Clustering methods also learn from data and they could be applied in parametric approaches, by relaxing the untenable assumptions (Alpaydin, 2010). Clustering is categorised as flat and hierarchical, incremental and probability-based one (Manning, Raghavan & Schütze, 2009, Witten, Frank, & Hall, 2017).

Moreover, two main problems in supervised ML are the use of data for Regression or Classification (Rasmussen & Williams, 2006; Marsland, 2009). In classification, the inputs are assigned, by a classifier, to one of two or more classes. The functions, responsible for separations, are called discriminants and they are able to make prediction based on past data (Alpaydin, 2010). In binary classification, some simple ML algorithms are the mean classifier, naïve bayes, the perceptron nearest neighbours and K-Means (Smola & Vishwanathan, 2008). Nearest neighbours and Kernel can be also used in nonparametric problems (Bishop, 2006).

Additionally, classifiers can be either linear or nonlinear. As linear, they are thought to be the decision hyperplanes, the perceptron algorithms, the least square methods, the mean square estimations, the logistic discriminants and the support vector

machines, based on Kernels. On the other hand, as nonlinear classifiers are thought to be the decision trees, the probabilistic neural networks, the multi-layer perceptrons, the polynomial methods, the backpropagation algorithms e.t.c. (Marsland, 2009; Theodoridis, & Koutroumbas, 2009).

In unsupervised learning, k-means also works well and especially the k-means neural networks. In addition. vector quantisation and self-organising feature maps are very common techniques as well (Kasabov, 2007; Marsland, 2009).

A lot of programs is used in ML, with most popular software to be MatLab and BUGS as well as programming languages such as C, R, Python, PROLOG and LISP (Shawe-Taylor & Cristianini, 2004; Segaran, 2007; Luger, 2009; Marsland, 2009; Thodoridis & Koutroumbas, 2010; Kruschke, 2011; Joshi, 2017).

Neural network algorithms are the utmost edge in the field of ML. MacKay (2005) distinguished three main specifications: the architecture, the activity rules and the learning rules. Furthermore, deep learning architectures are used to train neural networks via multiple layers (Bengio, 2009) and by using backpropagation algorithms (Mitchell, 1997). Applications can be found in linear data analysis and nonlinear pattern recognition (Samarasinghe, 2007).

The main purpose of the majority of the aforementioned methods is to find the best solution to a problem, the so-called optimisation (Segaran, 2007).

2.3 Bearings

According to American Bearing Manufacturers Association (ABMA, n.d.), a bearing is a mechanical component that gives machinery the ability to rotate in a variety of speeds and loads with an easy and effective way. As a vital mechanism, there are lots of types and many categories based on specific characteristics depending on the use, but some major categories are illustrated in **Fig.2** (Al-Waily, 2017).

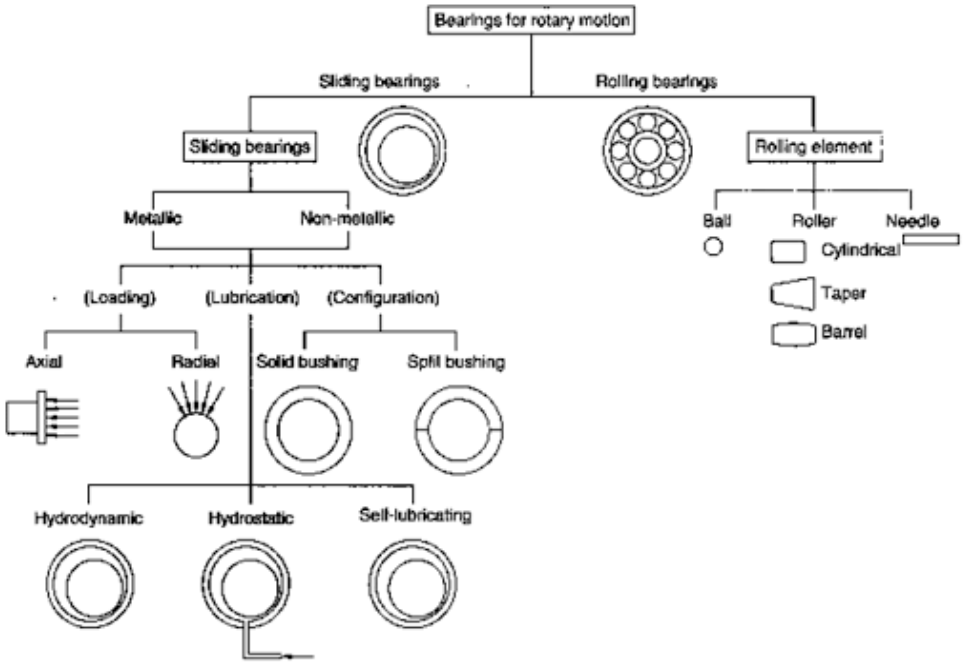
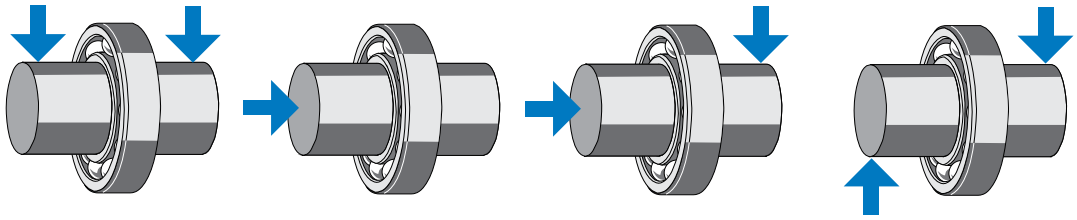


Figure 2 Bearing types

Ball bearings are usually used for shafts with small diameters in contrast with large-diameter shafts which can use taper, spherical cylindrical or toroidal roller bearings. Indeed, load plays a key role for selecting the appropriate bearing, so it should be taken into account the existence of radial, axial, combined or moment loads, as shown in **Fig.3** respectively (SKF, 2018).



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

Figure 3 Bearing Types I

a) Radial load, b) Axial load, c) Combined load and d) Moment load

Additionally, speed is also an important factor for bearings choice. Moreover, speed limit depends on many parameters such as cooling conditions, cage design, internal clearance, temperature and accuracy (SKF, 2018).

In PM and based on DIN ISO 281, life parameters estimations like Basic Rating Life (BRL) for 10^6 revolutions, Fatigue Life Factor and Speed Factor (f_n) are calculated by **Eqs. 1-3** (NSK, 2016; FAG, 1999).

$$BRL = \frac{10^6}{60n}\left(\frac{C}{P}\right)^p = 500{f_h}^p[h] \quad (1)$$

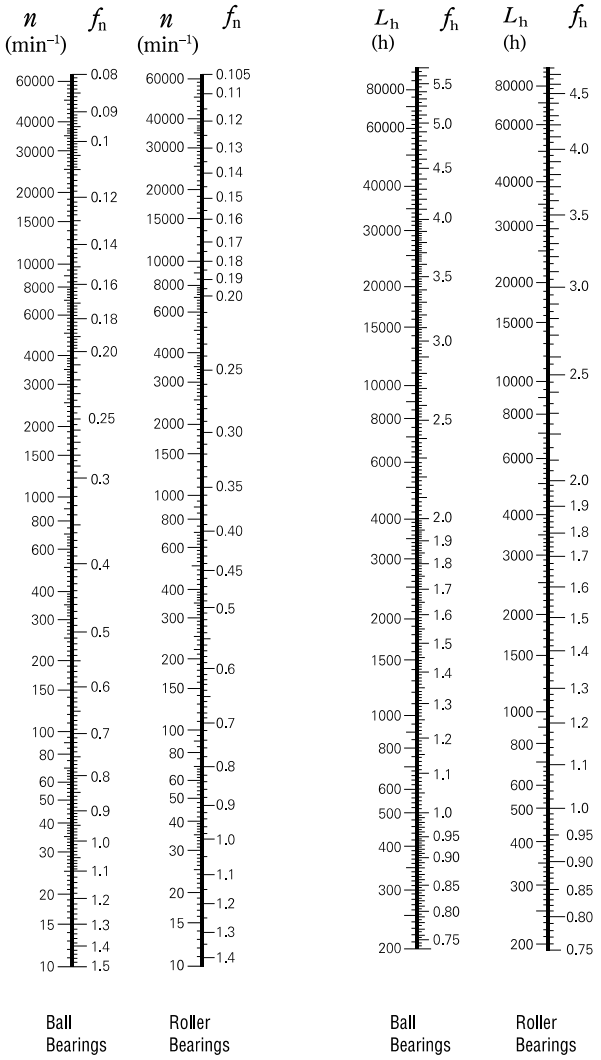
$$f_h = f_n\frac{C}{P} \quad (2)$$

$$f_n = \left(\frac{10^6}{500 \cdot 60n}\right)^{\frac{1}{p}} = 0.03n^{\frac{1}{p}} \quad (3)$$

Where p called life exponent, equal to 3 and 10/3 for Ball and Roller Bearings respectively, illustrated in **Fig. 4** (TIMKEN, n.d.). C stands for load rating, n is the mean rotational speed (rpm) and P is the equivalent load. Furthermore, f_n and f_h are derived from **Fig. 5** (NSK, 2016).

(a) (b)

Figure 4 Bearing Types II
a) open type ball bearing b) spherical roller bearing



(a) (b)

Figure 5 Factors Choice
a) Bearing Speed and Speed Factor
and b) FLF and Fatigue Life

Some usual faults of rolling bearings are corrosion in rolling elements, outer and inner race as well as cage damage and fatigue pitting. A common cause of these is the frictions and the lack of lubricants. The performance degradation produces unique vibration spectra and defect frequency (Saxena et al., 2016), but these signals are more random and cyclostationary than periodic (D. Wang et al., 2018).

2.4 Prognostics and Diagnostics

In prognostics, the real condition of a machinery is quantified by an index called remaining useful life (RUL) that represents the predicted time left before a failure. Indeed, it is crucial to be defined what a failure means and how it is estimated. So, a failure could be either a breakdown or an unsatisfactory performance degradation (D. Wang et al., 2018; Jardine et al., 2006; Tidriri et al., 2019; Xu et al., 2020). However, other distinctive metrics, i.e. extendable useful life, are proposed for use as well (Saxena et al., 2016).

In condition-based maintenance (CBM), there are the physical models and the data-driven models, with the latter to lack the need for complex mathematical modelling and unreasonable assumptions (D. Wang et al., 2018; Kim et al., 2012; Xu et al., 2021).

Data-driven methods consist of three main pillars, data collection, data processing and decision making. Furthermore, data are divided into three categories: value, waveform and multi-dimensional type with last two categories to be processed as signals (Jardine et al., 2006).

Vibration and acoustic signals are the most common waveform data, whose analysis falls into three distinctive categories : frequency-domain, time-domain and time-frequency (D. Wang et al., 2018; Jardine et al., 2006).

Signal can also be transformed into frequency domain, which gives the ability for isolation and identification of the important features. The most popular techniques are the Fast Fourier Transformation (FFT), Hilbert Transform (HT), spectrum and its differentiations such are bispectrum that is found applications in studying bearings (D. Wang et al., 2018; Jardine et al., 2006; Leite et al., 2019).

Time-domain analysis depends on time waveform and features extraction from signals, where descriptive statistics are used (D. Wang et al., 2018). Time synchronous average is a common technique for noise reduction or even remove. Some more

sophisticated techniques are the models of the autoregressive and the autoregressive integrated moving average while there are plethora of other approaches in this category (Jardine et al., 2006; Leite et al., 2019).

Time-frequency analysis overcomes the problem of difficulty in non-stationary waveform signals handling. Short-time Fourier transformation, also known as spectrogram, and Wigner-Vile distribution are widely used techniques, however the latter has got a difficulty in estimation of distribution. Wavelet transform is another approach that offers high time resolution at high frequencies and high frequency resolution at low frequencies and noise reduction. In addition, it can be improved by de-noising the signals via imposing zero frequency filter (Sachan et al., 2020). There also are other more advanced techniques such as wavelet packet transform (WPT), basis pursuit e.t.c. (Boškoski et al., 2014; Jardine et al., 2006).

Value type data analysis combines raw data with features that are processed from raw signals. There are multivariate approaches, e.g. independent and principal component analysis, and regression approaches such as polynomials and ARMA (Jardine et al., 2006).

Data or reliability analysis is another approach that can combines condition monitoring data with extra information such as events. Baseline hazard function, Weibull hazard function, proportional hazards model, potential-functional and installation-potential intervals, Hidden Markov models and EM algorithms are some examples of this analysis (Jardine et al., 2006).

Leite et al. (2018) investigated entropy and divergence in a dataset of bearings run-to-failure test. Moreover, a classification into two states, fault and non-fault, was achieved over all defection types, detecting the bearing life acceptably. Another research is also based on Jensen-Rényi entropy of vibrations and WPT examined bearing faults under a variety of speed and load and shows advantages on no requiring prior knowledge and no limits on statistical limits in required signals (Boškoski et al., 2015a).

In high speeds, e.g. 350 km/h and above, acoustic emissions are more suitable than vibration signals because of larger outliers, more stable Kurtosis and having a more Gaussian distribution of waveform patterns (Xu et al., 2021).

Classification and multi-class classification have been applied in bearings fault diagnosis through many proposed techniques with the results of some methods to reach outstanding and some perfect scores in accuracy, but not tested by other metrics (Spoerre, 1997; Yuan, & Tang, 2011; Gryllias, & Antoniadis, 2012; Ben Ali et al., 2015; Jia, Tahir, Khan, Iqbal, & Hussain, 2017; Carlo, Perkins, & Caputo, 2021).

Ball bearings have been studied as a case study of fault prognosis depending on estimations of health state probabilities. For dimensionality reduction and over-fitting avoidance, the features are extracted by using distance evaluation. After real health state being estimated from vibrations via SVM and the RUL estimations are very close to the real values (Kim et al., 2012).

Shao et al. (2018) proposed a novel method for bearing fault detection. After extracting some time-domain statistical features jointed in an index through local linear embedding (LLE), a continuous deep belief network, tuned by genetic algorithm shown a superiority over other sophisticated regressions, i.e. MAE = .24 and RMSE=.1.

Haidong et al. (2020) constructed a superior and very accurate method for early bearing fault prognosis. A sophisticated gate recurrent unit is combined with a modified training algorithm are fed with a complex variant entropy of vibration signals resulting an extraordinary performance (Haidong et al., 2020).

For small amount of data, missing values and poor information as a result, there are techniques that resolve this problem. The metabolism grey forecasting model combined with particle filter is thought to be a useful tool for knowledge extraction. Applying this technique in bearings temperature, good predictions happen with robustness and effectiveness outperforming other methods (Li et al., 2019). Additionally, incomplete dynamic Bayesian networks with gaussian mixture gives an

early fault alert and estimates the RUL with a stable way even when missing date increases (Zhang et al., 2018).

To sum up, a gap is found in collecting and listing the bearings used in LMs and specifically for the greek fleet. In addition, the majority of research in bearings fault prognostics focus on the time-series regression based on time steps.

3. METHODOLOGY

The research philosophy is followed is positivism such as the researcher is thought to be independent, the reality is observed as objective and the phenomena can be simplified into smaller parts by following quantitative research methods. Moreover, secondary data is observed and processed through quantitative techniques (Zukauskas et al., 2018)

A case study based on the greek locomotive fleet is conducted based on literature overview of the available technical manuals for counting the used bearings. Secondly, an applied research is made on the construction of a deep ML algorithm for bearings fault predictions.

3.1 Research Questions

This study aims to summarise all bearing attached on locomotives and to construct an algorithm for optimal fault predictions through ML. Moreover, the purpose is divided into research objectives which form the following Research Questions (RQs).

RQ1:Can it be counted the amount of bearings mounted on the mechanisms of each locomotive type?

RQ2:Could algorithms be designed that outperform the existing techniques on bearings fault diagnosis and prognosis?

3.2 Hypotheses

The hypothetico-deductive method is used for hypotheses construction and their testing. The RQ1 is analysed in below hypotheses (Hs) as follows, with Hx.0 denotes the null hypothesis and Hx.1 denotes the alternative (Sekaran et al., 2009).

H1.0: ADtranz LM does not contain more bearings than MLW LM (BS3<BS1).

H1.1: ADtranz LM contains more bearings than MLW LM (BS3>BS1).

H2.0: ADtranz LM does not contain more bearings than Siemens (BS3<BS2).

H2.1: ADtranz LM contains more bearings than Siemens LM (BS3>BS2).

In the same way, the RQ2 is divided in below Hs.

H3.0: The proposed multi-class classification model does not perform in bearing fault diagnosis perfectly (AUC ≠ 1 and ACC ≠ 1).

H3.1:The proposed multi-class classification model performs in bearing fault diagnosis perfectly (AUC = 1 and ACC = 1).

H4.0:The proposed linear regression does not predict adequately the RUL (ACC < .7).

H4.1:The proposed linear regression predicts adequately the RUL (ACC > .7).

H5.0:The proposed regression model of RUL does not outperform its rivals (MAE > .24 and RMSE > .1).

H5.1:The proposed regression model of RUL outperforms its rivals (MAE < .24 and RMSE < .1).

The answers to RQs and Hs fulfill the posed objectives.

4. ANALYSIS

4.1 Case Study

The greek rolling stock, e.g. LMs, will be the examined in present case study. Historically, Greek State Railways was found in 1920, whose successor Hellenic Railways Organisation, known as OSE was found in 1970. A subsidiary of the Group was found in 2008, called TRAINOSE SA with main operations in utilization. In 2013, the rolling stock maintenance industry was separated from OSE SA and was transferred to the new corporation EESSTY SA. In 2017, the Italian State Railways, called FS Group, under its subsidiary Ferrovie Dello Stato Italiane S,p.A. acquired the TRAINOSE SA, following a second acquisition of EESSTY S.A. in 2019.

Depending on technical drawings, manuals and instructions, a collection of given information is tried to list all possible bearings which are mounted on the Greek fleet of LMs alone.

4.1.1 Locomotives

The operational fleet of LMs counts 96 units in total, whose all specifications are illustrated in **Table 1**.

Table 1
Locomotives Information

LM Type	**ADtranz (DE 2000)**	**Siemens (Hellas Sprinter)**	**MLW 500**	**MLW 450**
Description	Diesel - Electric	Electric	Diesel - Electric	
Gauge length	1,435 mm			
Year	1998 – 2004	1999 – 2006	1973 (2009[b])	1974 (2004[b])
Number	220.001 – 220.036	120.001 – 120.030	A 451 – A 470	A 501 – A 510
Country - Origin	Germany – Bombardier Transportation	Germany – Siemens – KRAUS MAFFEI	Canada – MLW	
Total Units	36	30	20	10
Tare (ton)	–	–	114	117.6

Weight (ton)	80 ±2%	80	120	124
Diesel Tank capacity (lt)	3,500	–	4,000	
Axes type	Bo Bo		Co Co	
Tires diameter (max)	1,100 mm	1,250 mm	1,016	1,017
Max. Speed	160 km/h	200 km/h	149	
Min. Speed	25 km/h	20 km/h	22	26
Power Source	2 x MTU 12V 396TC13	25 kV – 50Hz (GTO –thyristor technik)	ALCO 251F/V12	ALCO 251F/V16
Net Power (kW)	2,100	5,000	1,985	2,908
Max. Traction Power (kN)	260	300	273	290
Max. Braking Power (kN)	160	160	–	–
Max. Power Supply	400 kVA (1,500V – 50Hz)	–	–	–
Transmission	Electric			
Braking System – Manufacturer	Electropnematic KNORR BREMSE + Electrodynamic		Electropneumatic - Westinghouse 26L - Electrodynamic	
Nu. Bogies	2	2	2	
Nu. axes/bogie	2	2	3	
Total length (mm)	19,400	19,580	17,755	19,392
Total width (mm)	2,950	3,000	–	–
Total height (mm)	4,260	4,300	–	–
Biaxes distance (mm)	2,650 mm	3,000	1,702 / 1,702	
Bogies centre distance (mm)	11,400	9,900	16,578	18,212
UIC[a] coding	505-1	505-2	–	–

[a] UIC is the International Union of Railway
[2] Refurbishment was made

4.1.1.1 MLW (MX-627)

Montreal Locomotive Works (MLW) was a Canadian LMs manufacturer that was merged via acquisition by American Locomotive Company (ALCO). By code name MX-627, 450 and 500 series are diesel-electric LMs operates on freight transport for many years in Greece.

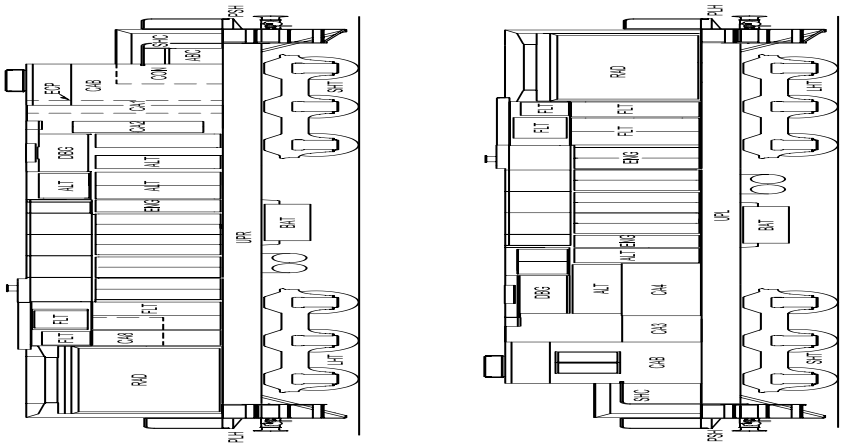


Figure 6 MLW LM

In **Table 2**, the main compartments are described as illustrated in **Fig. 6** (ALCO, 2003). Furthermore, some and all needed bearings are listed in **Table 3 – 4.**

Table 2
MLW Compartments

Code[α]	*DESCRIPTION*
ABC	Air Brake Compartment
ALT	Alternator Compartment
BAT	Battery Compartment
CA1	Control Area #1
CA2	Control Area #2
CA3	Control Area #3
CA4	Control Area #4
CA8	Control Area #8
CON	Control Console
DBG	Dynamic Brake Compartment
ECP	Engine Control Panel
ENG	Engine and Engine Compartment
FLT	Filter Compartment
LHT	Long Hood Truck
PLH	Platform at Long Hood end
PSH	Platform at Short Hood end
RAD	Radiator Compartment
SHC	Short Hood Compartment
SHT	Short Hood Truck
UPL	Underneath Platform on Left side
UPR	Underneath Platform on Right side

Note. Source: ALCO, 2003.
[α] ALCO abbreviations

Table 3
MLW Bearings I

Quantity	*DESCRIPTION*
1	Bearing, ball, drive shaft
1	Bearing, thrust
1	Bearing, centering
2	Bearing, needle
1	Bearing, thrust, rotating bushing

Table 4
MLW Bearings II

Quant.[a]	*Type*	*Manufactures*	*Description*	*Dimensions*	*Comp.[b]*
1	6317 ZZ C3 E AS2S	NSK	Deep groove ball bearings with two shields	85x180x41mm	
1	6203 2Z	FAG / Standard program 41500/2 DA051978	Deep groove ball bearings with two shields	17x40x12 mm	
1	6222 ZZ C3 E AS2S	NSK	Deep groove ball bearings with two shields	110x200x38m m	
1	6309 Z	FAG / Standard program 41500/2 DA051978	Deep groove ball bearings with single shield	45x100x25mm	
1	6204-2RS		Deep groove ball bearings with two shields and two O-rings	20x47x14 mm	Heater
1	6209 2RS1 C3	DIN 625	Deep groove ball bearings with single shield and O-ring	45x85x19 mm	Air compressor motor
1	6309 2Z	SKF General catalogue 3200 / IE 121985	Deep groove ball bearings with two shields	45x100x25mm	Generator fan motor
1	6309 2RSR	FAG catalogue FAGWL 41510 GR	Ball bearing with single shield	45x100x25mm	Air compressor motor
6	NJ320EMC4 + HJ320E		Cylindrical bearings with ring		Traction electric motor
6	NU 330 E M C4 NU330E/B/M2 /C4/ZS/SV 1.52	STEYR			Traction electric motor
36			BS1		

[a]*Quantity.* ; [b]*Compartment*

4.1.1.2 Siemens (Hellas Sprinter)

They are the only electric LM in Greece, **Fig. 7** (Siemens, 2004). In 1999, the first batch of six units was delivered, following a second batch. Initially, the first six were numbered with different coding as show in **Table 5**. In **Table 6**, they are listed the bearings of **Fig. 8-10** (Siemens, 2004).

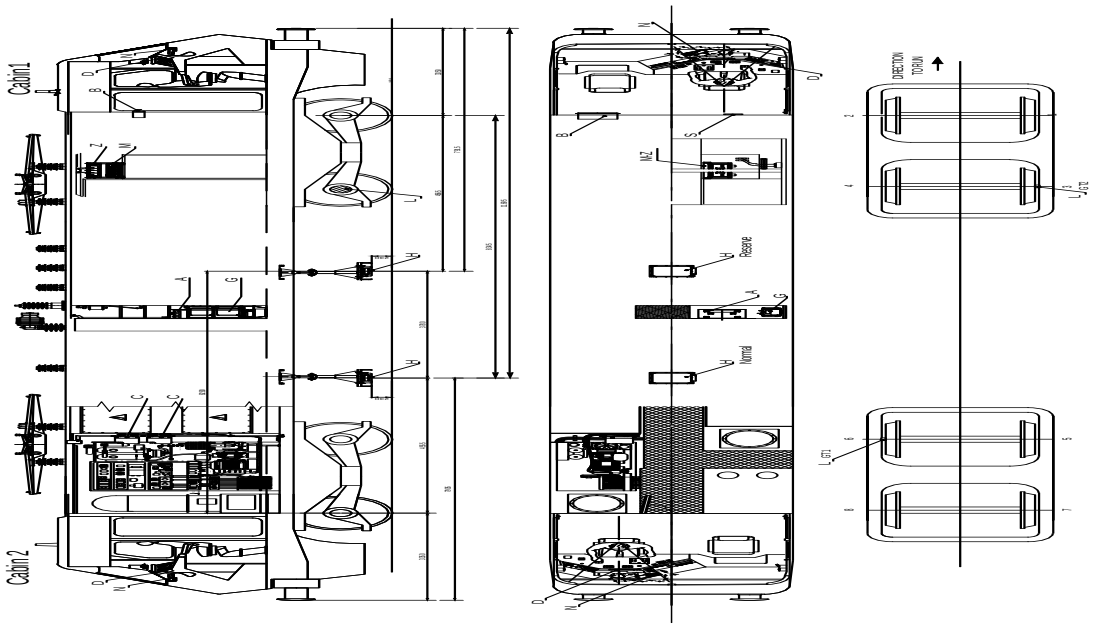


Figure 7 Siemens Hellas Sprinter

Table 5
Siemens 1st Series

Initial No.[a]	*Present Numbering*
H561	120.001
H562	120.002
H563	120.003
H564	120.004
H565	120.005
H566	120.006

[a] Number

Table 6

Siemens Bearings

No[a]	*Dr.*[b]	*DESCRIPTION*	*Type*	*Qnt.*[c]	*Comp.*[d]
3.20	Fig. 8	Cylindrical rolling bearing for electric motor D-END	FAG N326E.M1.R265.290.F 1. DIN 43283-N-326 ECM	4	Electric traction motor
4.21	Fig. 9	Rear tapered rolling bearing set		2	Traction subsystem
4.24	Fig. 9	Front tapered rolling bearing set		2	Traction subsystem
5.30	Fig. 7	Cylindrical rolling bearing for electric motor N-END with ceramic coating	SKF BC1B 322652 A. DIN 43283	4	Electric traction motor
		Deep groove ball bearings with two shields and two O-rings	6209 2RS1 C3 DIN 625 45x85x19mm	2	Air compressor
BS2				14	

[a] Number. [b] Drawing. [c] Quantity. [d] Compartment

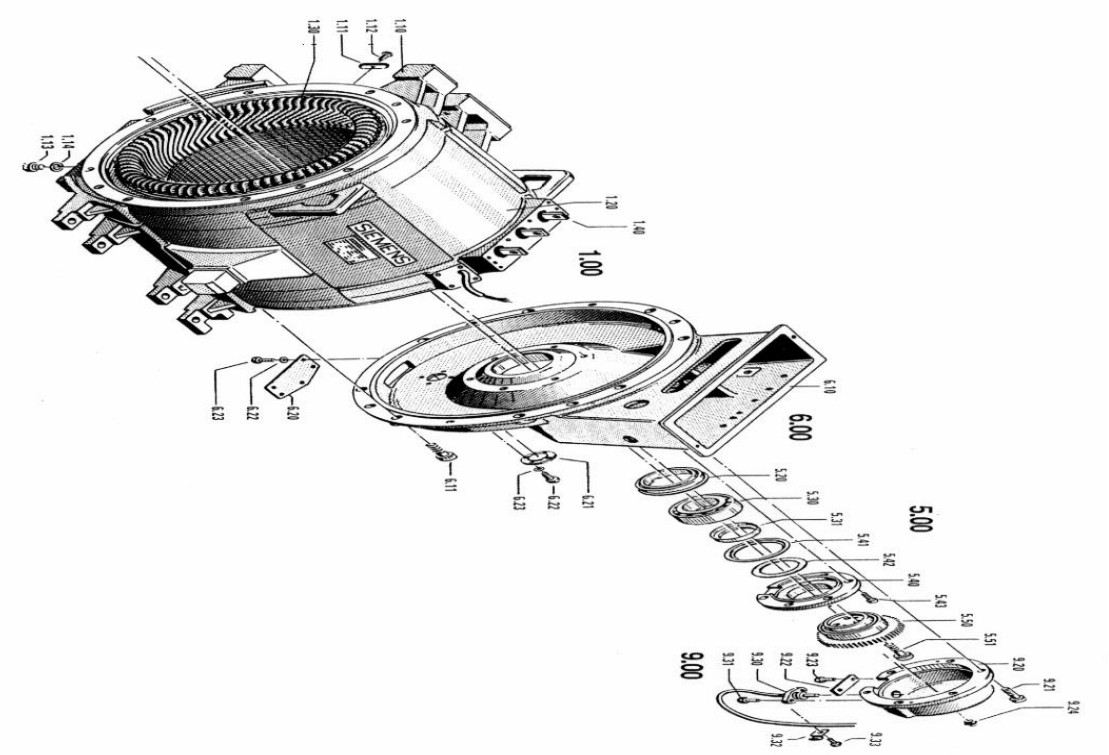


Figure 8 The electric motor stator

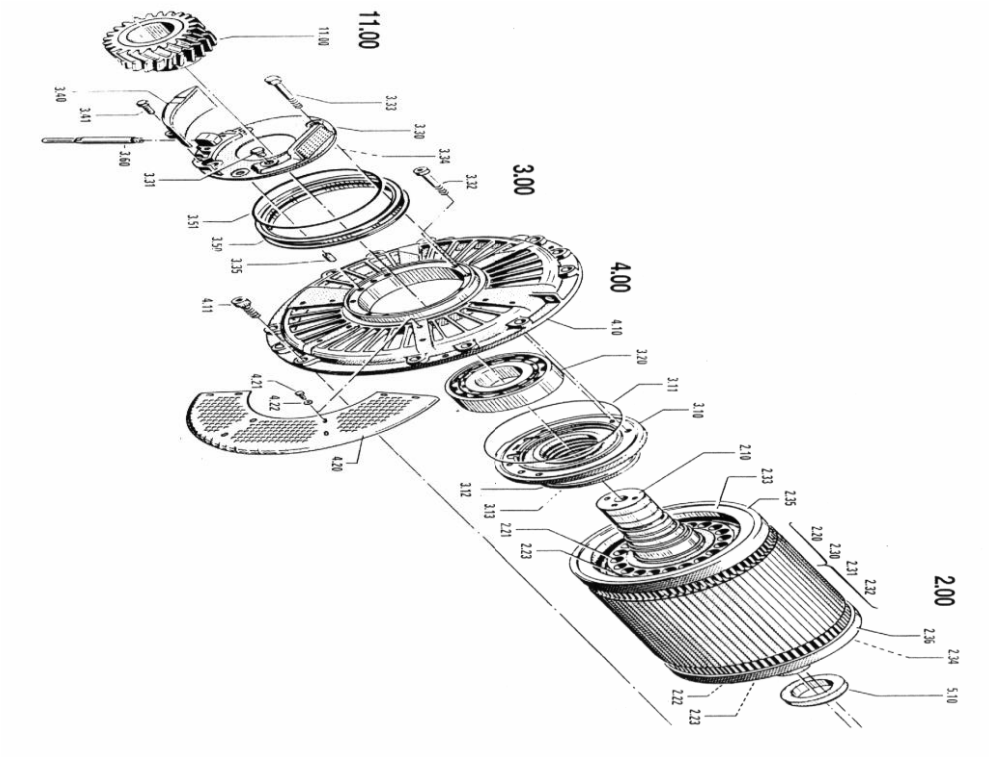


Figure 9 The electric motor rotor

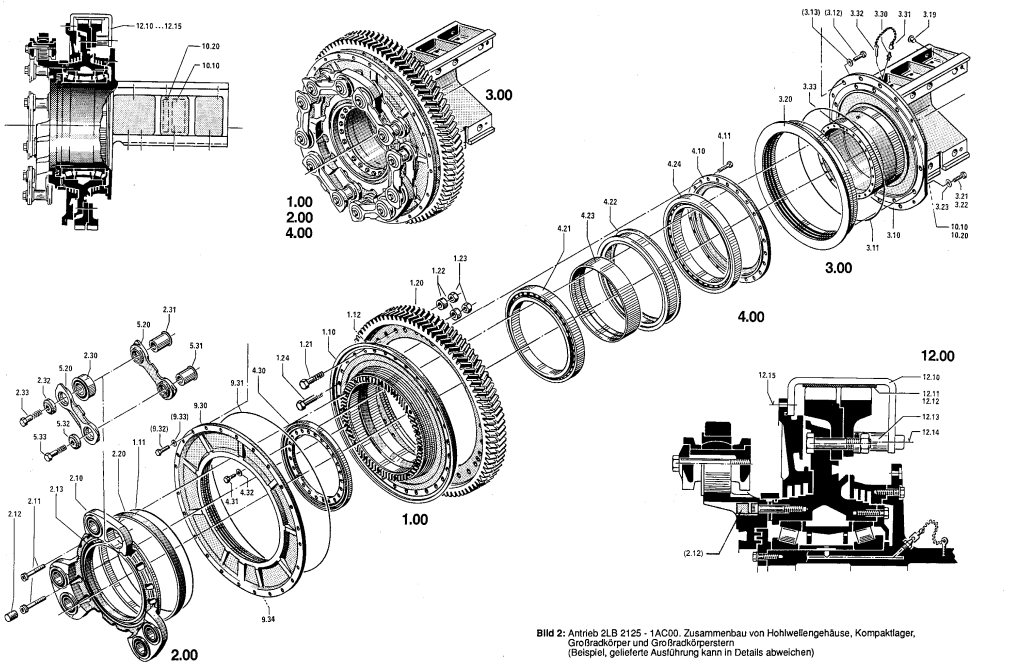


Figure 10 Traction subsystem

4.1.1.3 ADtranz DE 2000

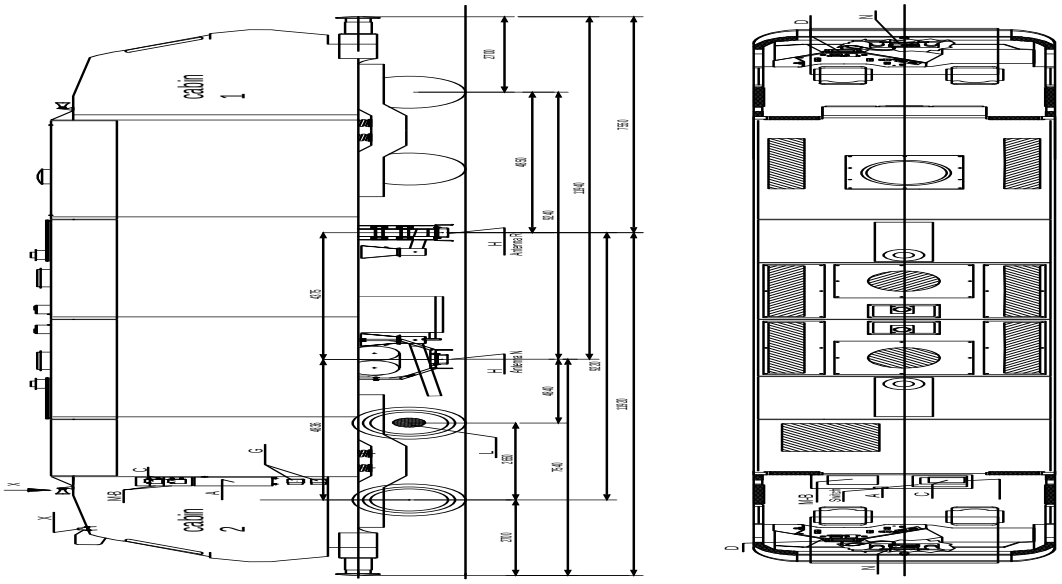


Figure 11 ADtranz DE 2000

ADtranz DE 2000, as shown in **Fig. 11-12**, is a diesel-electric LM made in Kassel and assembled in Oerlikon in Switzerland. Initially, they were numbered A.471 – A.496 as listed in **Table 7**. The bearings are listed in **Table 8** (Bombardier, 2003).

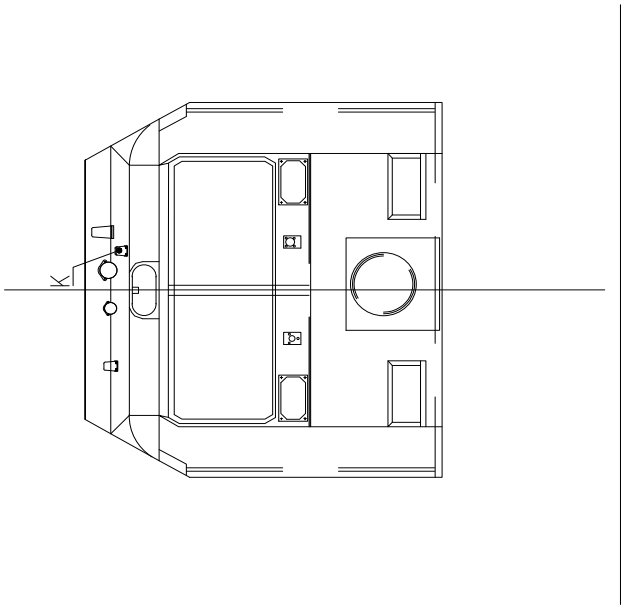


Figure 12 ADtranz DE 2000 front without bogies

Table 7
Code Numbers

Initial No.[a]	*Present Numbering*
A.471	220.001
A.472	220.002
A.473	220.003
A.474	220.004
A.475	220.005
A.476	220.006
A.477	220.007
A.478	220.008
A.479	220.009
A.480	220.010
A.481	220.011
A.482	220.012
A.483	220.013
A.484	220.014
A.485	220.015
A.486	220.016
A.487	220.017
A.488	220.018
A.489	220.019
A.490	220.020
A.491	220.021
A.492	220.022
A.493	220.023
A.494	220.024
A.495	220.025
A.496	220.026

[a] Number

Table 8
ADtranz Bearings

Quant [a]	*Type*	*Manufactures*	*Description*	*Dimensions*	*Compartment*
2	6206 2Z C3		Deep groove ball bearings with two shields		Fan electric motor
1	6208 2Z		Deep groove ball bearings with two shields	8x22x7mm	Air compressor
1	NUP226 ECM P63 VA379 NUP226EBM 2P63 SV1.52.7'	STEYR	Cylindrical Single-row bearing	130x230x4mm	Electric Generator
9	Undefined				Oil pump motors
2	NU222 ECMR/ P64VA309	SKF DIN 43283	Cylindrical Single-row bearing		Traction electric motor
15			BS3		

[a] Quant.: Quantity

4.2 Applied Research

Applied Research is undertaken so as that new knowledge is acquired focused on practical goals (OECD, 2012). In the present research, secondary data are used combined with multiple techniques in order a better outcome is achieved.

4.2.1 Dataset

The IMS Bearing Data has been collected by the Center for Intelligent Maintenance Systems (IMS), which includes data from three run-to-failure experiments. On a shaft, four bearings are mounted rotated by an alternative current motor at 2,000 rpm and applied a radial force of 6,000 lbs. Furthermore, two high sensitivity accelerometers were placed on each bearing in order to measure X and Y axes vibrations. In addition, the sampling rates was set at 20kHz by resulting the collection 20,480 points every second (Qiu, Lee & Lin, 2006).

In No. 1 set, 2,156 samples were taken every ten minutes, except from the initial 43 recordings that were logged every five minutes. The experiment took place from October 22, 2003 12:06:24 to November 25, 2003 23:39:56 and ended with the failure of the inner race in bearing 3 and the roller element in bearing 4 (Qiu et al., 2006).

In No. 2 set, 984 samples were logged every ten minutes. The experiment took place from February 12, 2004 10:32:39 to February 19, 2004 06:22:39 and ended with the failure of the outer race in bearing 1. However, only single-axis accelerometers were installed in that experiment (Qiu et al., 2006).

Finally, in No. 3 set, 4448 samples were collected every ten minutes. The experiment took place from March 4, 2004 09:27:46 to April 4, 2004 19:01:57 and ended with the defect of the outer race in bearing 3 (Qiu et al., 2006).

The programming language, which is being used, is Python on Macintosh environment. Both built-in and off-the-shelf libraries are used inside in the code in order the purposes to be reached.

4.2.2 Theory

Plethora of techniques are used and tested in the present study, so a short and comprehensive overview of their theory is presented as follows.

4.2.2.1 Complex Numbers

Fortunately, there are not missing values to handle cause of the data derived from an experiment. So, there is no need for using some traditional techniques. For a complex number with the form of **Eq. 4**, its magnitude is calculated by **Eq. 5**. (Κατωπόδης, Μακρυγιάννης & Σάσσαλος, 1995) and in this way all measurements are transformed into positive values.

$$\vec{z} = x + jy \tag{4}$$

$$|\vec{z}| = \sqrt{x^2 + y^2} \tag{5}$$

4.2.2.2 Preprocessing and Feature Engineering

Data preprocessing is an essential step and prepares the data for a ML model. Cleaning and tranformations are used to remove outliers and rescale the data in order to be more compatible for ML models (Chakrabarty, Mannan & Cagin, 2016).

The numeric representation of raw data is called features. Moreover, the process of formulating the most valuable features is named Feature Engineering (Zheng & Casari, 2018).

4.2.2.2.1 Signal Processing

The analysis, modification and synthesis of discrete-time signals so as to derive useful information is called signal processing (Oppenheim, & Schafer, 1999).

4.2.2.2.1.1 Transforms

4.2.2.2.1.1.1 Hilbert Transform

For a real signal s(t), its Hilbert Transform is defined as shown in **Eq. 6**, while the analytic signal is defined in **Eq. 7** (Yaguo, 2017).

$$\hat{x}(t) = \mathcal{H}\{s(t)\} = \frac{\int_{-\infty}^{\infty} \frac{s(t)}{t-\tau} d\tau}{\pi} = x(t) \cdot \frac{1}{\pi \cdot t} \quad (6)$$

$$z(t) = s(t) + j\hat{s}(t) \quad (7)$$

The instantaneous amplitude and phase are shown in **Eqs. 8-9** respectively (Yaguo, 2017).

$$a(t) = \sqrt{s(t)^2 + \hat{s}(t)^2} \quad (8)$$

$$\theta(t) = tan^{-1}\left(\frac{\hat{x}(t)}{x(t)}\right) \cdot \quad (9)$$

4.2.2.2.1.1.2 Fast Fourier Transform

In 1965, Fast Fourier Transform was developed in order to calculate the discrete Fourier transform (Smith, 2002). Generally, the Fourier Transform is given by the integral as formed in **Eq. 10**, with ω to be the continuous variable (Fischer-Cripps, 2002).

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}dt \tag{10}$$

An approximation of **Eq. 11** is the finite sum as defined in **Eq. 11**, where N denotes the equally distanced data points, iΔt is the time interval of data value $y_i(i\Delta t)$ during the i-th time space. So, the discrete samples are defined as shown in **Eq. 12** (Fischer-Cripps, 2002).

$$F(\omega) = \sum_{i=0}^{N-1} y_i(i\Delta t)e^{-j\omega(i\Delta t)}\Delta t \tag{11}$$

$$\omega_k = \frac{2k\pi}{N\Delta t} \tag{12}$$

$\forall\, k\epsilon(0, N-1)$, the actual amplitude spectrum is defined as follows in **Eq. 13** (Fischer-Cripps, 2002).

$$\frac{F(\omega_k)}{\Delta t} = \sum_{i=0}^{N-1} i\Delta t \cdot y_i \cdot e^{-\frac{j2k\cdot\pi\cdot i}{N}} \tag{13}$$

In terms of cosine-sine form, **Eq. 13** is transformed into **Eq. 14**, where C(k) is a complex number and is also known as discrete Fourier transform (Fischer-Cripps, 2002).

$$C(k) = \sum_{i=0}^{N-1} i\Delta t \cdot y_i \left[cos\left(\frac{2k\cdot\pi\cdot i}{N}\right) - jsin\left(\frac{2k\cdot\pi\cdot i}{N}\right)\right] \tag{14}$$

4.2.2.2.1.1.3 Discrete Wavelet Transform

In late 1980, the discrete wavelet transform (DWT) was constructed as formula (Percival & Mondal, 2012). In Hilbert space denoted L^2I, the $\psi(t)\varepsilon L^2I$ is called wavelet only if **Eq. 15** is satisfied (Ouadfeul et al., 2012).

$$\int_{-\infty}^{\infty} \psi(t)dt = 0 \tag{15}$$

Then, **Eq. 16** is called wavelet transform (Ouadfeul et al., 2012).

$$\psi_\alpha(t) = f(t) \cdot \psi_\alpha(t) \tag{16}$$

Where $\psi_\alpha(t)$ is the dilation of $\psi(t)$ and is defined in **Eq. 17** (Ouadfeul et al., 2012).

$$\psi_\alpha(t) = \frac{1}{a} \cdot \psi_\alpha\left(\frac{t}{a}\right) \tag{17}$$

In practice, a is separated in a binary form. So, when $a=2^j (j\epsilon Z)$, the wavelet becomes as defined in **Eq. 18** (Ouadfeul et al., 2012).

$$\psi_{2^j}(t) = \frac{\psi_{2^j}\left(\frac{t}{2^j}\right)}{2^j} \tag{18}$$

Its wavelet transform defined as shown in **Eq. 19** (Ouadfeul et al., 2012).

$$W_{2^j} f(t) = f(t)\psi_{2^j} = \frac{1}{2^j} f(t)\psi\left(\frac{t}{2^j}\right) \tag{19}$$

On the contrary, the reverse transform is defined as illustrated in **Eq. 20** when x(t) satisfied the criterion of **Eq. 21** (Ouadfeul et al., 2012).

$$f(t) = \sum_{-\infty}^{\infty} W_{2^j} \cdot f(t) \cdot x(t) \quad (20)$$

$$\sum_{-\infty}^{\infty} \hat{\psi}\left(2^j \omega\right) \cdot x\left(2^j \omega\right) \quad (21)$$

Dispersed in time domain, DWT is obtained. Moreover, an effective and fast representation is approximately calculated by **Eqs. 22-23**, where **Eq. 23** finds the DWT coefficients of $f(t)$ based on 2^j scale (Ouadfeul et al., 2012).

$$S_{2^j} f = S_{2^{j-1}} f \cdot H_{j-1} \quad (22)$$

$$W_{2^j} f = S_{2^{j-1}} f \cdot G_{j-1} \quad (23)$$

The H_j and G_j denote the gains of discrete filters when $(2^j\text{-}1)$ zeros are inserted every second sample of H and G, whose relationship is defined in **Eq. 24** (Ouadfeul et al., 2012).

$$g_k = \bar{h}_{1-k}(-1)^{k-1} \quad (24)$$

4.2.2.2.1.1.4 Dual-tree Complex Wavelet Transform

The evolution of DWT is the Dual-Tree Complex Wavelet Transform (DTCWT), having the advantages of being directionally selective (helpful in higher than 1D dimensions). In this technique, reconstruction and decomposition run parallelly through high-pass and low-pass filters for each scale as illustrated in **Fig. 13** (Wang et al., 2010).

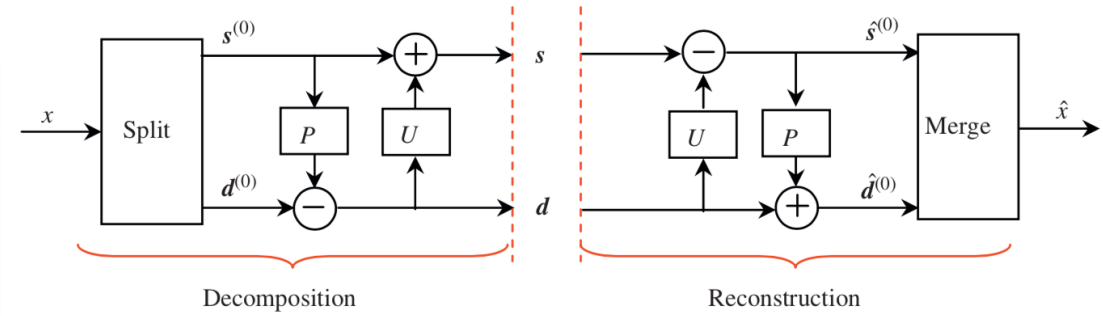


Figure 13 The decomposition and composition transform based on lifting scheme

For $\psi_h(t)$ and $\psi_g(t)$ being the real-value wavelet and its transform, $\psi^C(t)$ can be described as seen in **Eq. 25** (Wang et al., 2010).

$$\psi^C(t) = \psi_h(t) + j\psi_g(t) \tag{25}$$

The wavelet and scaling coefficients of the upper tree are $d_i^{Re}(k)$ and $c_j^{Re}(k)$ and they are defined in **Eqs. 26-27** respectively (Wang et al., 2010).

$$d_i^{Re}(k) = \sqrt{2}\int_{-\infty}^{\infty} \psi_h\left(2^i t - k\right)x(t)dt, \qquad for\ \ i\epsilon(1,j) \tag{26}$$

$$c_j^{Re}(k) = \sqrt{2}\int_{-\infty}^{\infty} \varphi_h\left(2^j t - k\right)x(t)dt \tag{27}$$

Where i and j denote the scale factor and the maximum scale respectively. In contrast, the $d_i^{Im}(k)$ and $c_j^{Im}(k)$ are the lower tree coefficients and if replaced by $\psi_h(t)$ and $\psi_g(t)$ respectively, the final DTCWT coefficients are calculated by **Eqs. 28-29** (Y. Wang et al., 2010).

$$d_i^C(k) = d_i^{Re}(k) + jd_i^{Im}(k),\ \ for\ \ i\epsilon(1,j) \tag{28}$$

$$c_j^C(k) = c_j^{Re}(k) + jc_j^{Im}(k) \tag{29}$$

4.2.2.2.1.2 Filters

4.2.2.2.1.2.1 Butterworth Filter

Butterworth Filter (BF) with N order works as a low-pass filter and achieves a maximum flat response in the passband. The poles have got equal distances and places around a circle with radius that is equal to the cutoff frequency of the filter (Lobontiu, 2010), **Fig. 14**. Dey et al. (2019) claims that a cutoff frequency between 0.1 and 15Hz is sufficient enough for noise reduction. The N poles and the frequency response are given by **Eqs. 30-31** (Thompson, 2014).

$$-sin\frac{\pi(2m-1)}{2N}+jcos\frac{\pi(2m-1)}{2N}m=1,2,3,\dots N \quad (30)$$

$$\left|H(j\omega)=\frac{1}{\sqrt{1+\ \omega^{2N}}}\right| \quad (31)$$

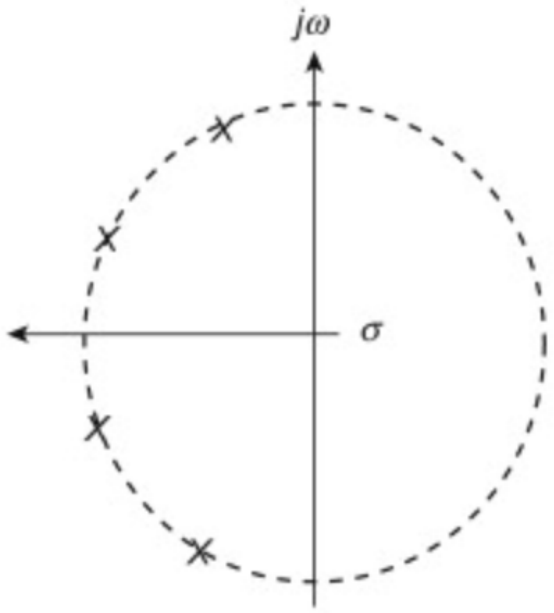


Figure 14 Poles

4.2.2.2.1.2.2 Kalman Filter

In the early 60s, Kalman Filter (KF) started to be implemented. Formulated for prediction is shown in **Eqs. 32-35**, where K denotes the gain, P the state error covariance matrix, R is the noise covariance matrix of noise, x and y are the independent and dependent variables respectively (Suthar et al., 2018).

$$K=\frac{C^T\cdot P_{old}}{R+C^T\cdot CP_{old}} \quad (32)$$

$$x_{new}=(y-x_{old}C)K+x_{old} \quad (33)$$

$$P_{new}=P_{old}-CP_{old}K \quad (34)$$

$$y_{est}=x_{new}C \quad (35)$$

4.2.2.2.1.3 Entropies

4.2.2.2.1.3.1 Shannon Entropy

Shannon entropy is a non-parametric technique, but thought to be inefficient when noise exists. Based on the probability mass function p(k), with k=1,…,N and a time series s(t), **Eq. 36** forms the SE (Boškoski et al., 2015b; Leite et al., 2019).

$$S(p) = -\sum_{k=1}^{N} log\big(p(k)\big) \cdot p(k) = -\sum_{k=1}^{N} ln\big(p(k)\big) \tag{36}$$

4.2.2.2.1.3.2 Permutation Entropy

When order d ≥2 via Shannon entropy, the Permutation Entropy (PE) is defined as follows in **Eq. 37**, with π denotes d! (Leite et al., 2019).

$$H(p) = -\sum_{k=1}^{n} log\big(p(\pi)\big) \cdot p(\pi) \tag{37}$$

4.2.2.2.2 Tests

4.2.2.2.2.1 *Normal Distribution*

The distribution of the data, if normal distribution (ND) or non-normal distribution (NND), is tested by the **Eq. 38** (D'Agostino, 1971; D'Agostiono & Pearson, 1973).

$$z_k^2 + z_s^2 \sim x^2(2) \tag{38}$$

Where $x^2(2)$ denotes the chi-square distribution, taking 2 degrees of freedom and for sample size n larger than twenty (n>20). The Kurtosis and Skewness tests are computed in **Eqs. 39-45**. (D'Agostino, 1971; D'Agostiono et al., 1973).

$$z_k = \frac{Kurtosis}{s.e.} \tag{39}$$

$$z_s = \frac{Skewness}{s.e.} \tag{40}$$

$$s.e. = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}} \tag{41}$$

$$Kurtosis = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma}\right)^4 \tag{42}$$

$$Skewness = \sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma}\right)^3 \tag{43}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}} \tag{44}$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n}(x_i) \tag{45}$$

4.2.2.2.2.2 *Stationarity*

Stationarity means constant variance and mean. Augmented Dickey_Fuller test (ADF) is a statistic test for examine if a unit root exists in time series. If the DF_t of **Eq. 49** is negative together with a set low value of marginal significance p-value, the null hypothesis of a unit root presence is rejected and the series is stationary (Hamilton, 2000). However, there are plethora of criteria for testing the stationarity such as Anderson, Von Neuman e.t.c. (Αλεξανδρόπουλος, Κατωπόδης, Παλιάτσος, & Πρεζεράκος, 1994).

Random walk is the simplest version of ADF and it is difined by **Eqs. 46-48** (Greene, 2002).

$$y_t = \gamma y_{t-1} + \varepsilon_t \tag{46}$$

$$\varepsilon_t \sim N[0, \sigma^2] \tag{47}$$

$$Cov[\varepsilon_t, \varepsilon_s] = 0. \ \ \forall\, t \neq s \tag{48}$$

For the null hypothesis, it is γ=1. On the other hand, for conventional ratio t, it is taken the function of **Eq. 49** (Greene, 2002). Where σ_{est} denotes the standard error of estimate in **Eq. 50 (**McHugh, 2008), γ is the actual value and $\hat{\gamma}$ is the predicted value.

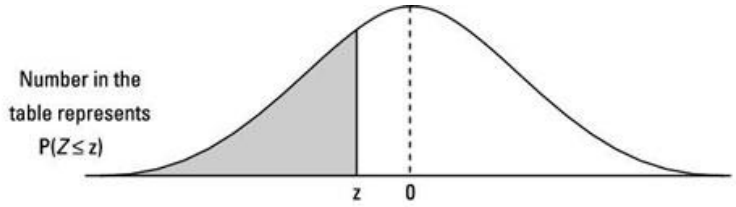
$$DF_t = \frac{\hat{\gamma} - 1}{\sigma_{est}(\hat{\gamma})} \tag{49}$$

$$(49) \Rightarrow \sigma_{est}(\hat{\gamma}) = \sqrt{\frac{\sum_{i=1}^{N} (\gamma - \hat{\gamma})^2}{N}} \tag{50}$$

Finally, p-value is calculated by z statistics by computed z from **Eq. 51** (Αλεξανδρόπουλος et al., 1994) and then by using the **Tables 9-10** (Rumsey, 2016).

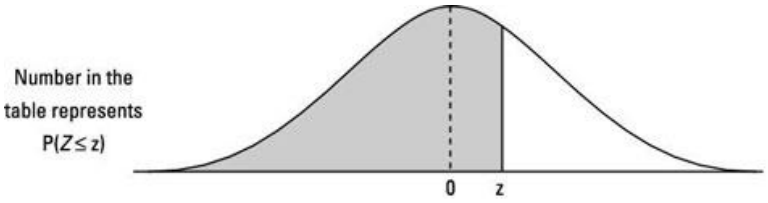
$$z = \frac{|\bar{x} - \mu_0|}{\sigma / \sqrt{n}} \tag{51}$$

Table 9
Probabilities I



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
–3.6	.0002	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
–3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
–3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
–3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
–3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
–3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
–3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
–2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
–2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
–2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
–2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
–2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
–2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
–2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
–2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
–2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
–2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
–1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
–1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
–1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
–1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
–1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
–1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
–1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
–1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
–1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
–1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
–0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
–0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
–0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
–0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
–0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
–0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
–0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
–0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
–0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
–0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Table 10
Probabilities II



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999

4.2.2.2.2.3 *Correlation*

4.2.2.2.2.3.1 Pearson

The Pearson Product Moment Correlation is a measurement of how much two parametric variables x-y are related to each other. Given a set significance, the metrics of the correlation is calculated by coefficient rho, as illustrated in **Eq. 52**. The more rho is reaching the 1, the more the correlation is, otherwise the 0 means no correlation at all (Chen, & Popovich, 2002).

$$r(x,y) = \frac{\sum_{i=1}^{N} \frac{(x_i-\bar{x})(y_i-\bar{y})}{n}}{\sqrt{\frac{\sum_{i=1}^{N}(x_i-\bar{x})^2 \cdot \sum_{i=1}^{N}(y_i-\bar{y})^2}{n}}} \text{ , } \forall\, r\epsilon[0,1] \tag{52}$$

4.2.2.2.2.3.2 Kendall

In nonparametric variables, Kendall's tau can be used instead of Spearman's rho. If P denotes the number of concordant pairs and Q denotes the number of discordant pairs, given a set significance the coefficient tau is computed by **Eq. 53**. Similarly, to Pearsons'r, the more the Kendall's tau is reaching 1, the more the correlation exists and the more it is reaching 0, the less the correlation exists (Chen et al., 2002).

$$t(x,y) = \frac{2(P-Q)}{n(n-1)} \text{ , } \forall\, t\epsilon[0,1] \tag{53}$$

4.2.2.2.2.3.3 Autocorrelation

The Autocorrelation Function is statistic test as formulated by **Eq. 54** (Tsay, 2012).

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^{T}(x_t-\bar{x})(x_{t-k}-\bar{x})}{\sum_{t=1}^{T}(x_t-\bar{x})^2} \tag{54}$$

4.2.2.2.3 Feature Scaling

Feature scaling or rescaling is the preprocessing step through data are transformed in order that a better model is constructed. Normalisation and standardisation are among the most popular techniques.

4.2.2.2.3.1 *Normalisation*

4.2.2.2.3.1.1 Logarithmic

Feng et al. (2014) claim that the log-normal transformation is the most popular techniques for skewed data, but they outline that the method performs poorly on skewed data handling, linear modeling and hypothesis testing. Furthermore, the log-normalisation is thought to happen either the data has got a log shape or the distribution is close to parametric. **Eq. 55** defines the log transform (Feng, Wang, Lu, & Tu, 2013).

$$x_{i,new} = log(x_{i,old}) \tag{55}$$

4.2.2.2.3.1.2 Min-max

Although this method is not used in the study, it is presented. As similar to z-score, its formula is defined by **Eq. 56** (Ozdemir, & Susarla, 2018).

$$x_{i,new} = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{56}$$

4.2.2.2.3.2 *Standardisation*

Data standardisation is the process of transforming the raw data into a target structural form (Loshin, 2009). Z-score is a simple method for outlier detection and defined by **Eq. 57** (Loshin, 2009). However, in bibliography there is a contradiction

in which standardization method is more effective (Mohamad & Usman, 2013; Steinley, 2004).

$$z_i = \frac{x_i - \bar{x}}{\sigma} \tag{57}$$

4.2.2.3 Machine Learning

4.2.2.3.1 Classification

There are two major types of supervised learning categories, regression and classification. In classification, it is desired the prediction of at least two labeled classes. When there are only two classes, then the classification is called binary in contrast to the case of three or more classes exist and then it is named multi-class classification (Müller et al., 2017).

The cross-entropy loss function for discrete distributions p and q is calculated for binary and multiclass classification by **Eq. 58-59** respectively (Zhu, He, Zhang, & Cui, 2020).

$$H(p,q) = -\big(ylog(\hat{y}) - (1-y)log(1-\hat{y})\big) \tag{58}$$

$$H(p,q) = -\sum_{1} p_i log q_i \tag{59}$$

4.2.2.3.2 Regression

In regressions, a mathematical formula, as formed in **Eq. 60**, is constructed in order to simulate the data as accurate as possible. Linear Regression (LR) is the most widely used method and in use for many decades (Müller & Guido, 2017).

$$\hat{y} = b + \sum_{i=0}^{n} w_i x_i \tag{60}$$

4.2.2.3.3 K-means

K-means is a technique that finds the cluster centers of specific regions of a dataset. Initially, it assigns each data point and then computes the mean of the data as long as no change in clusters occurs (Müller & Guido, 2017).

If protype vectors denote $\mu_1,\ldots,\mu_n$ and an indicator vector r_{ij} is equal to 1 if, and only if cluster j and x_i are assigned. By minimising the distortion measure of **Eq. 61**, the distances between each data point is also minimised (Smola & Vishwanathan, 2008).

$$J(r,\mu) \coloneqq \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n} r_{ij}\|x_i - \mu_j\|^2 \tag{61}$$

Where the Euclidean square norm is denoted by $\|\cdot\|^2$, $\mu = \{\mu_j\}$ and $r = \{r_{ij}\}$. For J minimizing, a two-stage strategy is adopted.

Stage 1

By keeping the μ fixed and determining r, for i-th data solution of x_i data point is calculated by **Eqs. 62-65** (Smola et al., 2008).

If :

$$r_{ij} = 1 \tag{62}$$

Then:

$$j = argmin\|x_i - \mu_{j'}\|^2 \tag{63}$$

Else if :

$$r_{ij} = 0 \tag{64}$$

Then:

$$j = 0 \tag{65}$$

Stage 2

By keeping the r fixed and determining μ, J is formed as a quadratic function of μ. So, deriving by keeping μj=0 for all j, as shown in **Eq. 66** (Smola et al., 2008).

$$\sum_{i=1}^{m} r_{ij}(x_i - \mu_j) = 0 \tag{66}$$

$$(66) \Leftrightarrow \mu_j = \frac{\sum_i r_{ij} x_i}{\sum_i x_i} \tag{67}$$

The all process cancels when the assignments of the cluster stayed almost unchanged at some point in calculations (Smola et al., 2008).

4.2.2.3.4 Long Short-Term Memory

Sherstinsky (2020) suggests the standardisation of the data before the Long Short-Term Memory (LSTM) network is fed and trained. LSTM is a variant of recurrent neural networks and passes information from the past outputs to current ones via storage elements (Elsheikh et al., 2019).

Fig. 15 illustrates the structure of LSTM. At a specific time step t, LSTM holds a hidden memory $\tilde{C}_t$ and 3 gate units: the input gate i_t the output gate o_t and the forget gate f_t, which are calculated by **Eqs. 68-71** (Cui et al., 2020).

$$f_t = \sigma_g(W_f \cdot x_t + U_f \cdot h_{t-1} + b_f) \tag{68}$$

$$i_t = \sigma_g(W_i \cdot x_t + U_i \cdot h_{t-1} + b_i) \quad (69)$$

$$o_t = \sigma_g(W_o \cdot x_t + U_o \cdot h_{t-1} + b_o) \quad (70)$$

$$\tilde{C}_t = tanh(W_C \cdot x_t + U_C \cdot h_{t-1} + b_C) \quad (71)$$

Where the Ws detone the weight matrices mapping, the Us denote the weight metrices and bs denote bias vectors. Furthermore, $\sigma_g(\cdot)$ is the sigmoid function of the gate and $tanh(\cdot)$ denotes the hyperbolic tangent function. Moreover, the layer output h_t and the cell output C_t are given by **Eqs. 72-73** (Cui et al., 2020).

$$h_t = o_t \odot tanh(C_t) \quad (72)$$

$$C_t = o_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (73)$$

Where $\odot$ is the operator of matrix/vector multiplication (Cui et al., 2020).

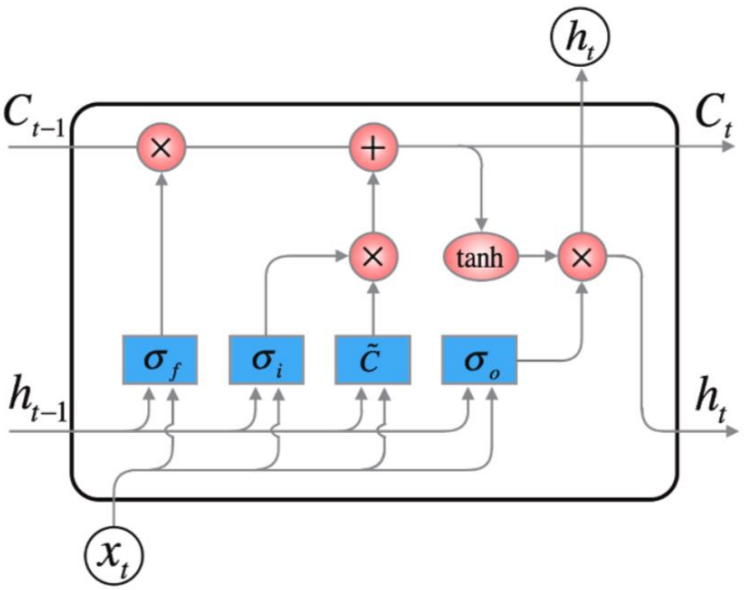


Figure 15 LSTM

Total loss $\mathcal{L}$ during training at each iteration is given by **Eq. 74**.

$$\mathcal{L} = loss(\hat{x}_{T+1} - x_{T+1}) = loss(h_T - x_{T+1}) \quad (74)$$

Where the loss function is symbolized as $loss(\cdot)$ (Cui et al., 2020).

The Bidirectional LSTM (BiLSTM) means the calculation of output for both forward and backward directions. So, $\vec{h}_t$ denotes the forward output depending on the forward order of inputs and masks such as [$x_1, x_2, \ldots, x_T$] and [$m_1, m_2, \ldots, m_T$] and following the reverse order for backward direction and its output $\overleftarrow{h}_t$ (Cui et al., 2020). Lample et al. (2016) illustrates the BiLSTM as shown in **Fig. 16**.

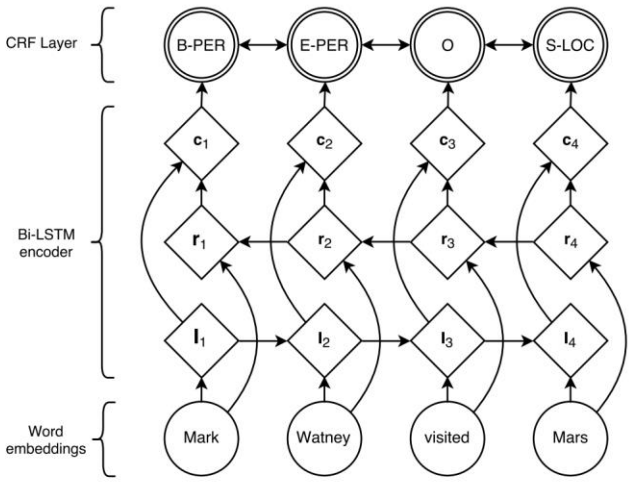


Figure 16 BiLSTM

Before the computation of the loss, activation outputs should be chosen. There is the choice of either the Sigmoid activation function or the Softmax. Loss.

The Sigmoid function, also called logistic function, of an element x_i is given by **Eq. 75** (Witten et. al., 2017).

$$f(x_i) = \frac{1}{1 + e^{-x_i}} \tag{75}$$

The Softmax Loss is computed by **Eq. 76**, where x_j the scores for each class C (Witten et. al., 2017).

$$f(x_i) = \frac{e^{x_i}}{\sum_j^C e^{x_i}} \tag{76}$$

Rectified Linear Units (ReLU) can also be used as a loss function, as shown in **Eq. 77** (Zhang et al. 2017).

$$\sigma(z) = \max(0, z) \tag{77}$$

Eq. 78 illustrates first layer inputs of DNN. The **Eq. 79** shows the number of the last layer outputs. The x denotes the input data, y the outputs and α^M, while W and b are random parametres (Zhang et al. 2017).

$$\alpha^1 = x \tag{78}$$

$$a_j^m = \sigma\left(\sum_j w_{ij}\alpha_i^{m-1} + b_j^{m-1}\right), j > 1 \tag{79}$$

$$y = \alpha^M \tag{80}$$

4.2.2.3.5 Validation

Cross-validation, **Fig. 17**, is thought to be more accurate than simple train-test split technique for the evaluation of a model. K-fold cross-validation splits the dataset into k equal parts, called folds. Each fold is treated as testing set and the remaining folds are used for training each model. In this way, k models are constructed and the means of their performance is thought to be a very representative metrics of the model (Müller et al., 2017).

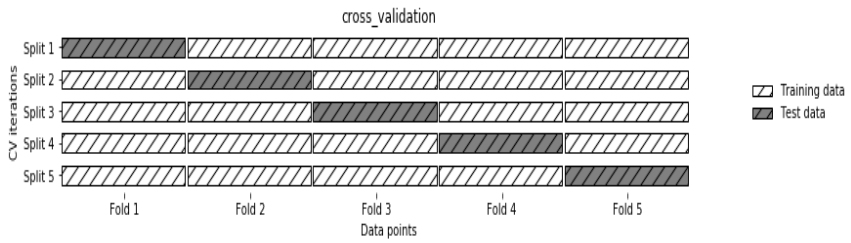


Figure 17 Cross-Validation

Although cross-validation is a very strong tool, it is too heavy and demanding for processing sources, which is not practical for evaluating the also heavy DNNs. So, the data is split into train and test parts and afterwards a small part of these sets is used for validation.

4.2.2.3.6 Tuning

Reimers and Gurevych (2017) summarised some hyperparameters optimasation techniques for fine tuning, which are adagrad, adadelta, RMSProp, Adam and Adam with Nesterov momentum, called Nadam. They also test their performance and find that Adam and its variant outperform in sequence labeling tasks.

The other parameters are chosen by GA. The selected technique is the Population-Based Incremental Learning in which there is a representation of the population of individuals by a single genetic (Baluja, & Caruna, 1995).

Like natural genes dispose a length, the same happens in artificial genes. So, the length of genes is a custom paramtre should be set. Population and generations sizes are also important paramtres for setting (Floreano et al., 2008).

Cicirello and Smith (2000) proposes a uniform variation of partially matched crossover (Goldberg, 1989) taking into account the population, the crossover rate, the probability of any allele, the mutation rate and the halting tolerance T as given by **Eq. 81**.

$$\frac{Fitness(MostFit) - Fitness(LeastFit)}{FitnessMost(Fit)} < T \tag{81}$$

Beyer and Schwefel (2002) suggests a modern approach for mutation calculation according to an extended log normal rule as defined by **Eq. 82**. For the (10, 100) evolution strategy is used c=1.

$$\sigma_t = e^{\frac{c}{\sqrt{2n}}N_0(0,1)}\left[\sigma_{t-1}e^{\frac{c}{\sqrt{2\sqrt{n}}}N_{1,1}(0,1)}, \dots \ , \sigma_{t-1}e^{\frac{c}{\sqrt{2\sqrt{n}}}N_{1,n}(0,1)}\right] \tag{82}$$

As a selection method it is chosen the elitism and more specifically the non-dominated sorting GA (NSGA-II) as proposed by Deb et al. (2002). Since of limited processing sources population size is set at 20, the number of generations is also set at 20. Additionally, the crossover probability is chosen 0.8 and the mutation probability 1/n, where n is the number of variables.

4.2.2.3.7 Model Evaluation

4.2.2.3.7.1 *Classification Metrics*

There are plenty of metrics for ML model evaluation and especially for classification. **Eqs. 78-80** express the accuracy (ACC), the recall (REC), the precision (PRE) (Leonard, 2017) the Area Under the Curve (AUC) (Wang, Zeng, & Zhu, 2010).

$$ACC = \frac{TP + TN}{TN + TP + FN + FP} \tag{78}$$

$$REC = \frac{TP}{FN + TP} \tag{79}$$

$$PRE = \frac{TP}{FP + TP} \tag{80}$$

Where TP denotes the true positives, TN are the true negatives, FP are the false positives and FN are the false negatives. An AUC of 0.5 is thought unacceptable, until 0.69 is moderate, 0.7 to 0.79 is good, 0.8 to 0.89 is excellent and above 0.9 is outstanding (Hosmer, & Lemeshow, 2013).

4.2.2.3.7.2 *Linear Regression Metrics*

R-Squares calculates the fitness of a LR and it is described by the formula of. the **Eq. 81** where y is the real value, $\hat{y}_i$ is the estimated value and $\bar{y}_i$is the means of the values for n samples (Seber, 1977). Indeed, there are some important metrics such as the Mean Absolute Error (MAE) (Willmott, & Matsura, 2005), the Root Mean Squared Error (RMSE) **Eqs 82-83** (Neill, &Hashemi, 2018; Haidong et al., 2019).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{81}$$

$$MAE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n} \tag{82}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{83}$$

4.2.3 Method

A bibliographic research is conducted on available technical manuals and other sources in order to summarise all possible bearings. On the other hand, by the combination of well-known techniques with other state-of-art methods, a very modern method is designed to improve the bearing fault prognosis.

4.2.3.1 The Proposed Approaches

As illustrated in **Fig. 18**, a multi-level method is designed so as that better results are achieved. Firstly, the raw signals are inserted in the programming language and their time intervals are calculated as well. Afterwards, the data are processed through filters and transforms in order noise is removed. Entropies compose the information into a single value for each sample of the datasets. By statistics, custom indices are constructed to improve the overall performance. Then clustering is deployed for concentrating the information further and then categorical encoding forms an array easily explainable and interpretable. Before modelling, data is firstly fluffed and split into train and validation sets. Finally, modelling is made via ML and validation is used for performance testing.

Since limited processing power and memory, some techniques should be rejected based on the processing demand. Assigning of the elapsed time needed for execution of code lines, as illustrated in **Table 15**, the lighter demanding techniques are chosen. In addition, the number of parameters need each technique is taken into consideration under the scope of keeping the model as simple as possible. Last but not least, the effectiveness in transforming the data into a stationary and normally distributed form is thought as a vital criterion for the selection among the tested methods.

The dataset is split into four time-intervals. In this way, this classification is thought to support the algorithm to identify the RUL via segmentation. Except of One-hot encoding, categorical encoding is chosen and made by a custom code for better epxlainablity and interpretability. Tuning is chosen to be made by GAs. Cause of limited computational power, the most demanding and the least efficient techniques are rejected.

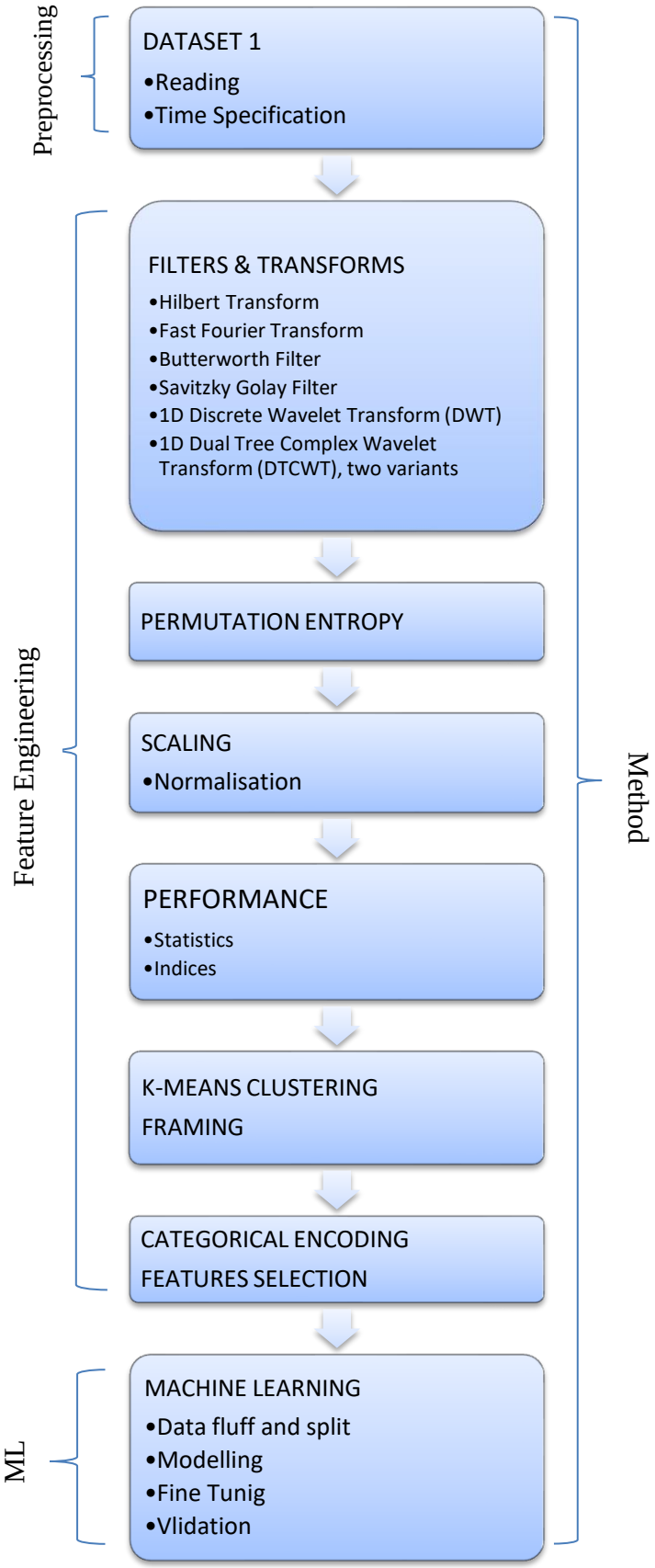


Figure 18 Proposed ML Method

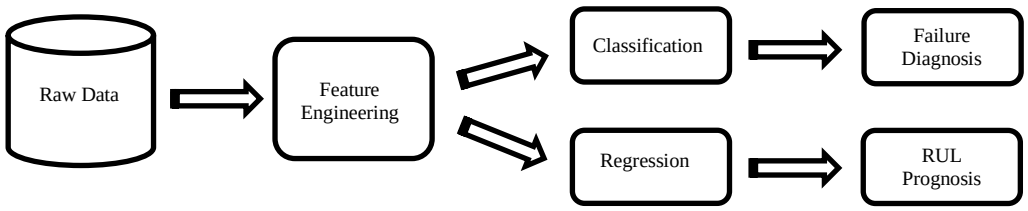


Figure 19 Proposed ML approaches

As illustrated in **Fig. 19**, the processed signal is used as input to feed the BiLSTM and the three operation states are inserted as outputs. Then, multi-class classification is employed to construct the ML model. On the other hand, RUL is estimated by a regression ML model with looking back time steps.

4.2.3.2 Code

During the whole process of programming, many parts should be divided into small tasks for whom custom code is written. In other cases, already existed libraries are used as off-the-shelf solutions.

Algorithm 1 Time Reading Pseudo-Code

for *j={1, 2,…,2156}* **do**
  if *j<2* **then**
  vector$_j$*(1, 1)* ← $read\_name(file\_name_j)$
    x_j ← vector$_j$*(1, 1)*
  end if
  if *j*> 2 **then**
    vector$_j$*(1, 1)* ← $read\_csv(file\_name_j)$
    x_j ← vector$_j$*(1, 20480)*
    matrix(j, 1) ← $x_j \cup x_{j-1}$
    tr← *matrix(j, 1)*
  end if
end for

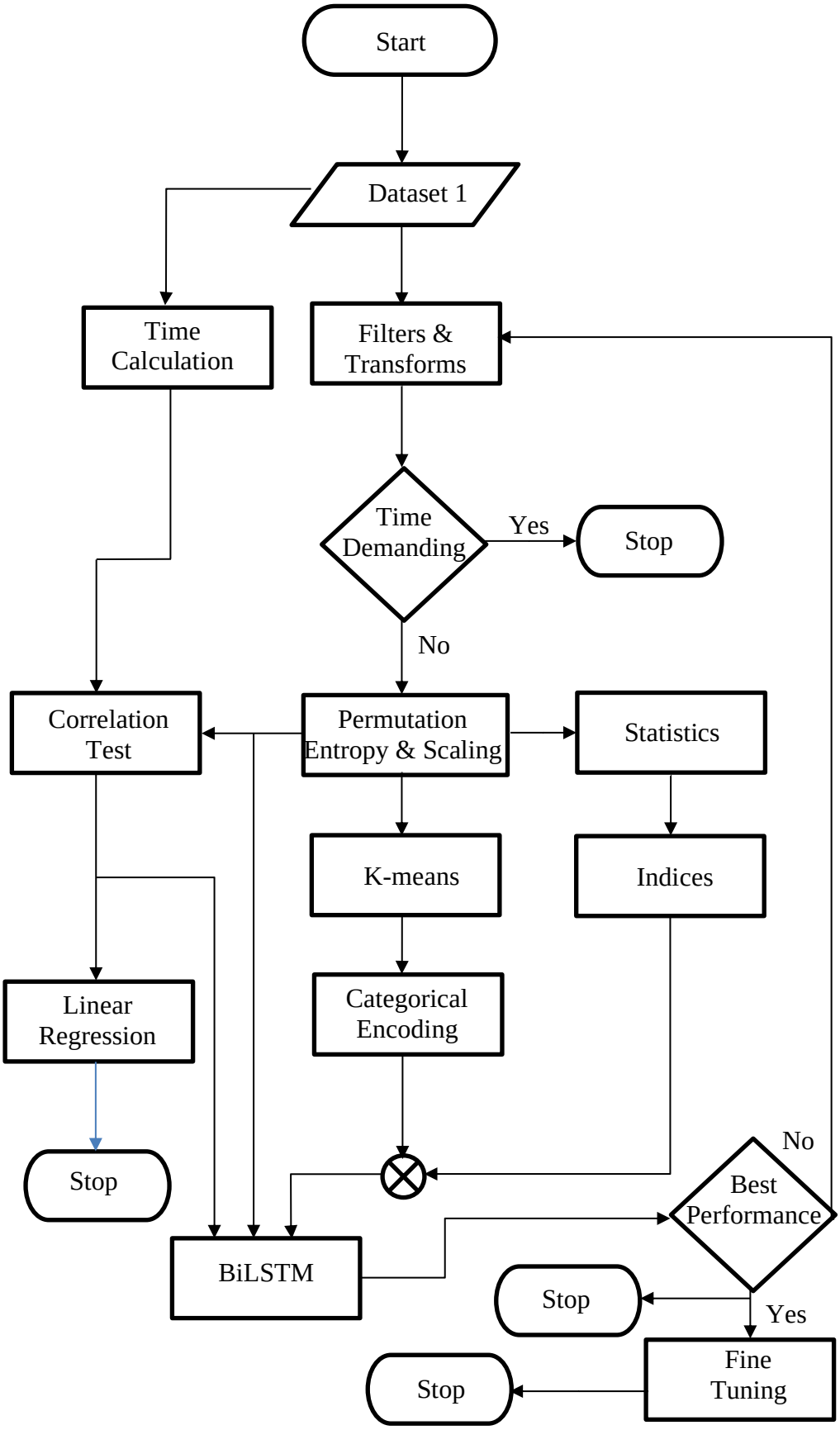


Figure 20 Algorithms Flowchart

Algorithm 2 Time Difference Calculation Pseudo-Code

for $j=\{1, 2, \ldots, 2156\}$ **do**
  if $j<2$ **then**
    split() **where** *separator(.)*
      $vector_j(1, 2) \leftarrow split(file\_name_j)$
      $y_i \leftarrow vector_j(1, 2)$
      $vector_1(1, 2) \leftarrow \{0, 0\}$
      $matrix(j, 2) \leftarrow vector_1(1, 2)$
  end if
  if $j>2$ **then**
    split() **where** *separator(.)*
      $vector_j(1, 2) \leftarrow split(file\_name_j)$
      $y_i \leftarrow vector_j(1, 2)$
      $matrix(j, 2) \leftarrow y_j \cup (y_j - y_{j-1})$
      $td \leftarrow matrix(j, 2)$
  end if
end for

Algorithm 3 RUL Pseudo-Code

for $j=\{1, 2, \ldots, 2156\}$ **do**
  if $j<2$ **then**
    $s_j \leftarrow (td_{j,1} \cdot 60) + td_{j,2}$
  end if
  if $j>2$ **then**
    $s_j \leftarrow (td_{j,1} \cdot 60) + td_{j,2}$
    $vector(j, 1) \leftarrow s_j + s_{j-1}$
    $tp \leftarrow vector(j, 1)$
  end if
end for
for $j=\{1, 2, \ldots, 2156\}$ **do**
  $vector(j, 1) \leftarrow s_j / s_{2156}$
  $RUL \leftarrow vector(j, 1)$
  $vector(j, 1) \leftarrow (s_j / s_{2156}) \cdot 100$
  $RUL\% \leftarrow vector(j, 1)$
end for

Algorithm 4 Transformations Pseudo-Code

for *j={1, 2,…,2156}* **do**
  if *j<2* **then**
  vector$_j$*(1, 20480)* ← $read\_name(file\_name_j)$
    x_j ← vector$_j$*(1, 20480)*
  end if
  if *j*> 2 **then**
    vector$_j$*(1, 20480)* ← $read\_csv(file\_name_j)$
    x_j ← vector$_j$*(1, 20480)*
    matrix(j, 204800) ← $x_j \cup x_{j-1}$
    values← *matrix(j, 204800)*
  end if
end for
matrix(2156, 204800) ← *transform(matrix(2156, 204800))*
matrix(2156, 1) ← *entropy(matrix(2156, 204800))*
trfd_values ← *matrix(2156, 1)*
matrix(8, 1) ← *statistics(trfd_values)*
stat ← *matrix(8, 1)*

Note. Calculations for each bearing

In **Fig. 20**, all steps of the programming are shown by an algorithm flowchart. At the beginning, the raw data, which are encoded in CSV format, are read and inserted into a variable as matrix by **Algorithm 1**. Then, labels of the samples are treated by **Algorithms 2** and **3** so as that RUL is calculated.

Time is chosen to be divided into four-time intervals with 254 samples each. These four-time segments are used instead of RUL in order to support the modeling by saving processing sources.

Overall the process, the heavier and ineffective techniques are rejected so as to reach an optimal solution for the problem. So, after denoising the vibration signals via filters or transforms, as shown in **Algorithm 4,** PE is employed to compress the information. Afterwards, statistics help to form supportive indicines.

Clustering by K-means is used to compose the information more while the custom categorical encoding forms a matrix of 413,952 rows and 38 columns in order to be the final dataset for testing. The custom encoding is used for a better understanding of the whole process. Finally, after testing the enough effective methods, the best approach is chosen to construct DNNs by BiLSTM via GAs.

For RUL estimation, the Pearson correlation is used to identify the most closely correlated value and then a LR is constructed. Finally, a regression model based on looking back time steps is constructed to be tested the prediction of the time series.

5. RESULTS

In this section, only some important results are presented so as to avoid information overload. For more insights, it is suggested to take a look at the Appendix.

5.1 Bearings Lists

In this section, some comprehensive lists of bearing are presented with some extra information about technical specifications, manufactures, models, use or compartment if information was accessible.

Tables 11-14 enlist all accessible information about the bearings mounted on each type of LM (MLW, n.d.; Bombardier, 2003, Siemens, 2004).

5.1.1 MLW LM

Table 11

MLW LM Bearings I

Quant.[a]	Type	Manufactures	Description	Dimensions	Compartment
1	6317 ZZ C3 E AS2S	NSK	Deep groove ball bearings with two shields	85x180x41mm	
1	6203 2Z	FAG / Standard program 41500/2 DA051978	Deep groove ball bearings with two shields	17x40x12mm	
1	6222 ZZ C3 E AS2S	NSK	Deep groove ball bearings with two shields	110x200x38mm	
1	6309 Z	FAG / Standard program 41500/2 DA051978	Deep groove ball bearings with single shield	45x100x25mm	
1	6204-2RS	SKF/Koyo	Deep groove ball bearings with two shields and two O-rings	20x47x14mm	Heater
1	6209 2RS1 C3	DIN 625	Deep groove ball bearings with single shield and O-ring	45x85x19mm	Air Compressor
1	6309 2Z	SKF General catalogue 3200 / IE 121985	Deep groove ball bearings with two shields	45x100x25mm	
1	6309 2RSR	FAG catalogue FAGWL 41510 GR	Ball bearing with single shield	45x100x25mm	
6	NJ320EMC4 + HJ320E	NSK	Cylindrical bearings with ring	100x215x47mm	Traction electric motor
6	NU 330 E M C4 NU330E/B/M2 /C4/ZS/SV 1.52	STEYR/FAG	Cylindrical Roller singe row	150x320x65mm	Traction electric motor

[a] Quantity

The BS1 totals 26 bearings based on **Tables 11-12**.

Table 12

MLW LM Bearings II

Quantity	DESCRIPTION
1	Bearing, ball, drive shaft
1	Bearing, thrust
1	Bearing, centering
2	Bearing, needle
1	Bearing, thrust, rotating bushing

Note: without identification

5.1.2 Siemens LM

Table 13

Siemens LM Bearings

No[a]	Dr.[b]	DESCRIPTION	Type	Qnt.[c]	Compartment
3.20	Fig. 8	Cylindrical rolling bearing for electric motor D-END	FAG N326E.M1.R265.290.F1 DIN 43283-N-326 ECM	4	Electric traction motor
4.21	Fig. 9	Rear tapered rolling bearing set	Z-534052.TR1 FAG	4	Traction subsystem
4.24	Fig. 9	Front tapered rolling bearing set	Z-534052.TR1 FAG	4	Traction subsystem
5.30	Fig. 7	Cylindrical rolling bearing for electric motor N-END with ceramic coating	SKF BC1B 322652 A DIN 43283	4	Electric traction motor
-	-	Deep groove ball bearings with two shields and two O-rings	6209 2RS1 C3 DIN 625 45x85x19mm	2	Air compressor

[a] drawing. [b] quantity

The BS2 totals 18 bearings based on **Tables 13**.

5.1.3 ADtranz LM

Table 14

ADtranz LM Bearings

Quant.		*Manufactures*	*Description*	*Dimensions*	*Compartment*
2	6206 2Z C3	SKF	Deep groove ball bearings with two shields	40x80x18mm	Electric Motor Fan
1	6208 2Z	SKF	Deep groove ball bearings with two shields	8x22x7mm	Air compressor
1	NUP226 ECM P63 VA379 NUP226EBM 2P63 SV1.52.7		Cylindrical Single-row bearing	130x230x4mm	Electric Generator
8	NJ 1880 MP + HJ 1880	FAG/SKF	Cylindrical Single-row bearing	400x500x46mm	Gearbox
4	NP 273081	Timken	Double Cup Conical bearing	101.6x63.5x8mm	Gearbox
8	BC2-0098	SKF	Cylindrical Roller Double Row bearing	160x270x170mm	Bogies
1	RNU 1940 E.M	FAG	Cylindrical Roller bearing	130x280x58mm	Water Pump
2	NU222 ECMR/ P64VA309	SKF DIN 43283	Cylindrical Single-row bearing	110x200x38mm	Traction electric motor

Note. Bombardier (2003)

The BS3 totals 27 bearings based on **Table 14**.

5.2 Bearings Fault Prognosis

In this section, different metrics are used to examine each method about effectiveness and efficiency.

5.2.1 Processing Demand

The elapsed time for the execution of the same task is taken into account in order to be chosen which techniques are heavy enough and rejected, **Table 15.**

Table 15
Techniques Comparison

Method	Elapsed Time (s)	Parameters
KF	51.7154870033264 16	n_iter=5
BF_lp	0.08866715431213379	T=5.0 Fs = 1000 Cutoff_freq=30 Order=4
1D SGF	0.007277011871337891	Win_lenght=5 Polyorder=2
HT_m	0.0949089527130127	
FFT_m	0.037882089614868164	
DWT	0.0004799365997314453	Wavelet=biort1.1 Samples jump=14
DTWT_b_lp	0.00024890899658203125	Level=3
DTWT_b_hp_m	0.0002980232238769531	Level=3
DTWT_q_lp	0.0002467632293701172	Level=3
DTWT_q_hp_m	0.000308990478515625	Level=3
PE	0.007987022399902344	Log Normalisation Order=2
MPE	0.05527925491333008	Order=2 Delay Time=1 Scale=1
SE	0.8500769138336182	

Notes. Elapsed time varies on each program execution; Task includes from Dataset 1, 1st bearing, X axis, 1st sample, 20,480 points; _b denotes biort; _q denotes qshift; _lp denotes low pass; _m denotes magnitude; _hp denotes high pass.

5.2.2 Exploratory Analysis

In this section, after having applied signal processing some calculation and statistics are used in order to find out some initial results that will support the ML process.

5.2.2.1 Permutation Entropy

Even tough scatter plots do not show some specific patterns, **Fig. 36-53,** statistics help to identify some hidden characteristics inside data. In **Tables 19-40**, it has been found that standard deviation among bearings follows a stable pattern that distinguish the investigating bearings and their axes. Moreover, for the case of DTCWT low-pass variants the standard deviation σ_1 of the bearing 1 at the Y axis has got a lower value compared with σ_3 and σ_4 of bearings 3 and 4 at the X axis. In contrast, the reverse happens for the case of the other DTCWT high-pass variant but for X axis.

5.2.2.2 Test of Normality

Having transformed the data by aforementioned techniques before and after the implementation of normalisation, the effectiveness of each technique is calculated by Pearson rho metrics (D'Agostino et al., 1973).

In **table 16**, it is shown that before normalisation the data is already almost normal distributed (ND). Furthermore, the HT and FFT achieve perfectly transform the data form non-normal distribution (NND) into ND. After PE transform, the FFT and the DTCWT outperform and reach a perfect percentage, as illustrated in **Table 17**. Both results are presented in the graph of **Fig. 21**.

Table 16

Techniques Comparison I

Method	ND	NND	ND ratio (%)
Raw_Data	17195	53	99.69271799628943
BF_lp	16365	883	94.88056586270872
1D SGF	15005	2243	86.99559369202227
HT_m	17248	0	100
FFT_m	17248	0	100
DTWT_b_lp[a]	12424	22072	36.01576994434137
DTWT_q_lp[a]	12424	22072	36.01576994434137

Notes. Task includes 4 bearings, X and Y axes, 2156 samples, 20,480 points; Before Permutation Entropy Tranform and log normalization; _b denotes biort; _q denotes qshift; _lp denotes low pass; _m denotes magnitude

[a]Including 3 and 4 levels

Table 17

Techniques Comparison II

Method	ND	NND	ND ratio (%)
Raw_Data	22	2	91.66666666666666
BF_lp	15	9	62.5
1D SGF	22	2	62.5
HT_m_pe_d	23	1	95.83333333333334
FFT_m_pe_d	24	0	100
DTCWT_b_lp_pe_d[a]	32	0	100
DTCWT_b_hp_pe_d[a]	32	0	100
DTCWT_q_lp_pe_d[a]	32	0	100

Notes. Task includes 4 bearings, X and Y axes, 2156 samples, 20,480 points; b denotes biort; q denotes qshift; pe denotes the Permutation Entropy Transform; d denotes log normalization; hp denotes high pass; lp denotes low pass; m denotes magnitude.

[a]Including 3 and 4 levels

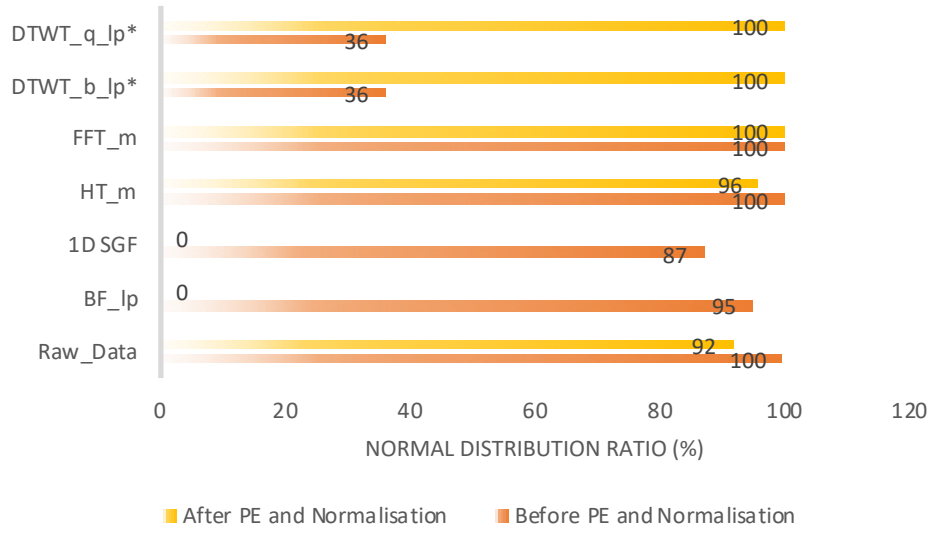


Figure 21 Test of Normality (%)

5.2.2.3 Test of Stationarity

The **Table 18** and **Fig. 22** show that DTCWT low-pass variants are able to transform all data into a stationary form.

Table 18
ADF test

Method	Stationary	Non-stationary	Stationarity ratio (%)
Raw_data_pe_d	6***	18***	25
BF_lp_pe_d	12***	12***	50
1D SGF_pe_d	8***	16***	33.33333333333333
HT_m_pe_d	8***	16***	33.33333333333333
FFT_m_pe_d	15***	9***	62.5
DTCWT_b_lp_pe_d[a]	32***	0***	100
DTCWT_b_hp_pe_d[a]	28***	4***	87.5
DTWT_q_lp_pe_d*	32***	0***	100

Notes. Task includes 4 bearings, X and Y axes, 2156 samples, 20,480 points; _b denotes biort; _q denotes qshift; _pe denotes the Permutation Entropy Transform; _d denotes log normalization; _hp denotes high pass; _lp denotes low pass; m denotes magnitude
***p<.001

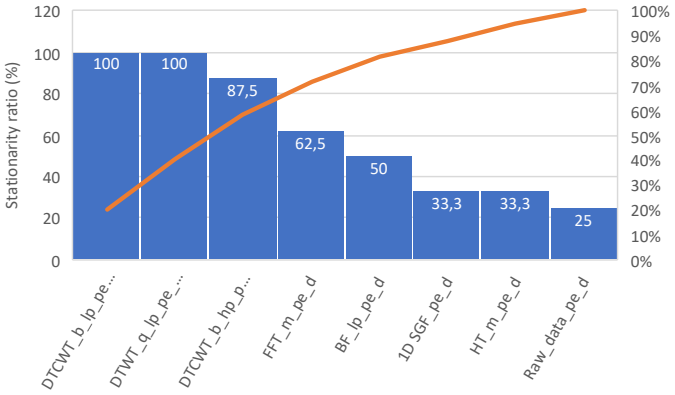


Figure 22 Test of Stationarity

5.2.2.4 Correlation

The correlation inside every technique is examined by the heatmaps of the **Fig. 56-60**, where it is found out that DTCWT variants are not able to distinguish the correlation among the bearings. However, the other techniques outline that there is moderate to strong correlation between X and Y axis for each bearing, which is thought to be normal. It is worth-mention that there is a strong enough correlation between the bearing 3 Y-axis and the bearing 2 for both X and Y axis.

As for autocorrelation, DTCWT, the FFT shows form of white noise, in contrast with the HT, the SGF and the raw signal after PE with level=3 that have strong positive values, as show in **Fig. 66-73**. Moreover, the BF shows a positive autocorrelation, but weak. The RUL is highly autocorrelated as illustrated in **Fig. 23**.

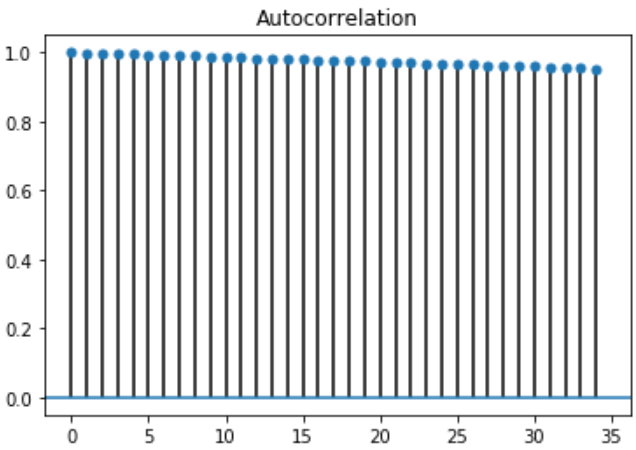


Figure 23 Autocorrelation I: RUL

5.2.3 Machine Learning Models

In this section, costumed indices are developed in order to support the ML modeling. Furthermore, K-means are deployed to simplify the information before the construction of the model. Then, BiLSTM builds plethora of Deep Neural Networks (DNNs), optimising the hyperparameters with NADAM, and finally fine tuning is approached by GAs.

5.2.3.1 Indices

The test of indices is made by sequential intervals as illustrated in **Fig. 24**. So, the standard deviation is calculated each time as same as the four indices.

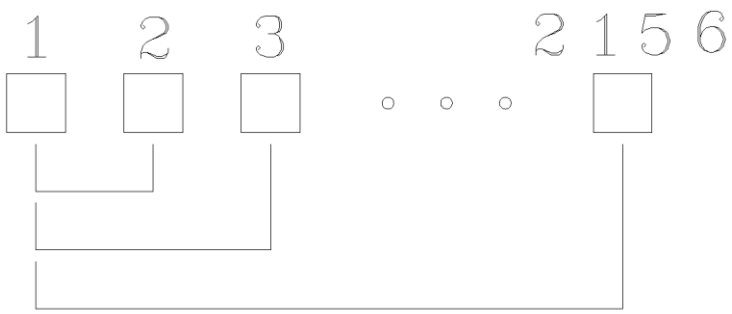


Figure 24 Index Sampling

After examining the statistics of all possible cases, **Tables 19-40**, four indices are constructed based on the comparison between the bearing 1, used as a benchmark, and the goal bearings 3 and 4. The indices are formed as defined by **Eq. 84-85**. Notice should be given in the fact that only after the second sample standard is deviation possible to be calculated. For this reason, first two values are replaced by 1 in order to be capable to examine all 2,156 samples.

For $A = \{1,2\}$ and $x \in A$,

$$h_x = \begin{cases} \sigma_x < \sigma_1,\ h_x = 1 \\ \sigma_x > \sigma_1,\ h_x = 0 \end{cases} \quad (84)$$

For $B = \{3,4\}$ and $y \in B$,

$$h_y = \begin{cases} \sigma_y > \sigma_1,\ h_y = 1 \\ \sigma_y < \sigma_1,\ h_y = 0 \end{cases} \quad (85)$$

The **Tables 41-52** show the results of the indices h_x calculations. **Fig. 25-26** show that the case of DTCWT biort low-pass variant gives the most reliable results for the indices h_x and the biort high-pass variant performs the worst in h_x indices, i.e. it performs the best in h_y indices.

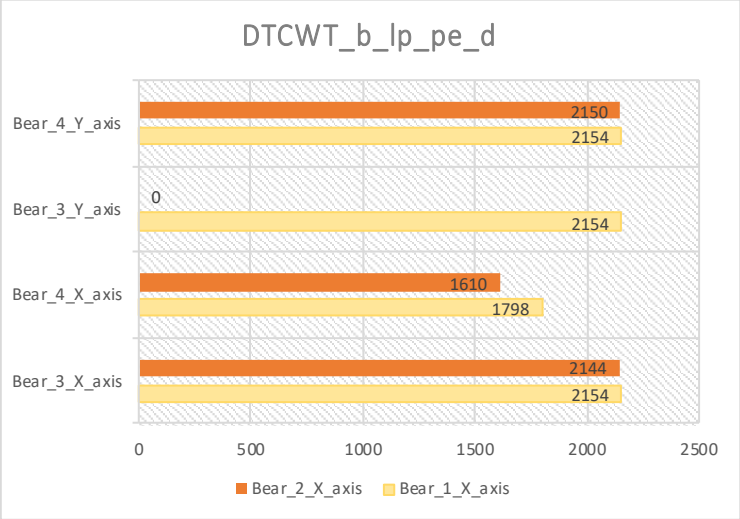


Figure 25 Index h_x

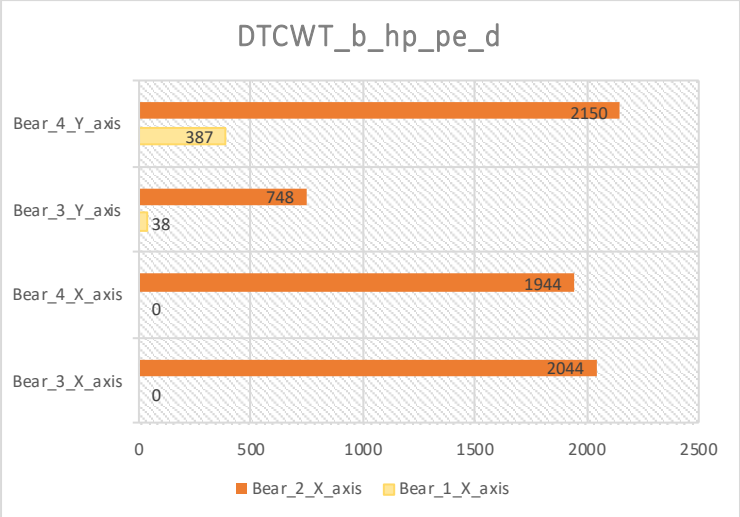


Figure 26 Index h_y

5.2.3.2 Time Segmentation

Since the choice of approaching the problem with the classification, time is ought to be taken as discrete intervals. These intervals would be in a range between two and the maximum number of instances. However, it is desirable to be chosen a relatively small number of time intervals in order to support the ML modeling and save processing sources, resulting a lighter program and shorter execution time. For the choice of the right number of the time intervals, it is used the maximum intercept of the even division of the three datasets instances as given by the algorithm 4.

Algorithm 4 Time Segmentation

```
l = 2156
k = 984
j = 4448
for i in range(1, (j)):
    if (((j/i)-(j//i))==0) & (((k/i)-(k//i))==(((j/i)-(j//i)))) & (((l/i)-(
        l//i))==(((j/i)-(j//i)))):
        print(i)
```

5.2.3.3 K-means Clustering

The K-means clustering is used in order to compose more the information. The cases of two and three clusters are examined. **Fig. 74-89** illustrate all tested techniques for two clusters, but no sound conclusions can be made because of no clear patterns in data, i.e. all data shows non-uniform shapes in their values.

5.2.3.4 Modeling

In this section, the raw data and all aforementioned techniques are used for feeding the BiLSTM network to construct ML models. Although the models perform poorly initially, gradually by the use of more complex techniques the performance evolves.

Initially, all unclustered signals feed the DNN to construct a binary classification model that is able to distinguish the working state from failure. The parametres are deliberately constant for all cases so as that the results are comparable. Moreover, they are chosen 100 neurons, no hidden layers, the sigmoid activation, binary cross-entropy for loss function, 20% for validation, Nadam for hyperparametres optimization, batch

size equals to 100, 20% dropout and stable pseudo-randomly shuffling. The results show an unacceptable poor performance for all techniques, as illustrated in **Table 53** and **Fig. 27**.

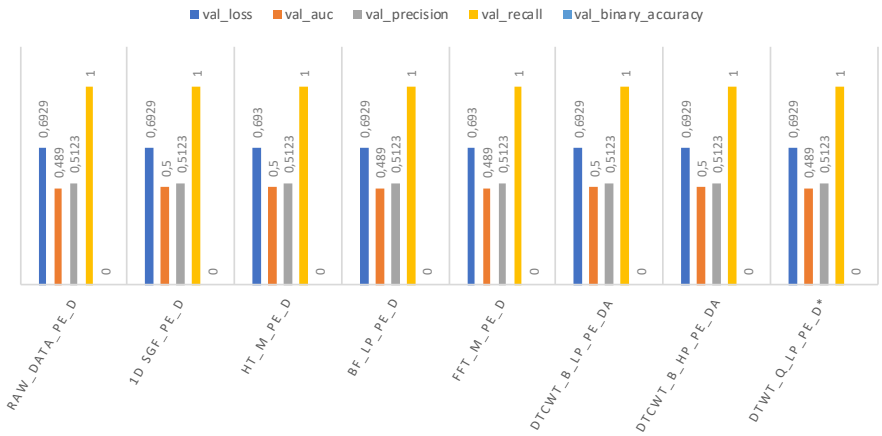


Figure 27 Binary Classification

By clustering via K-means, the multi-class classification of three distinctive operational states, i.e. working, failure 1 and failure 2, is improved. However, the performance of all method is not acceptable yet, **Table 54** and **Fig. 28**. The parametres are set as: 100 neurons, no hidden layers, the softmax activation, categorical cross-entropy for loss function, 20% for validation, Nadam for hyperparametres optimisation, batch size equals to 100, 20% dropout and stable pseudo-randomly shuffling. Furthermore, the instances are 413,952 in a categorical form.

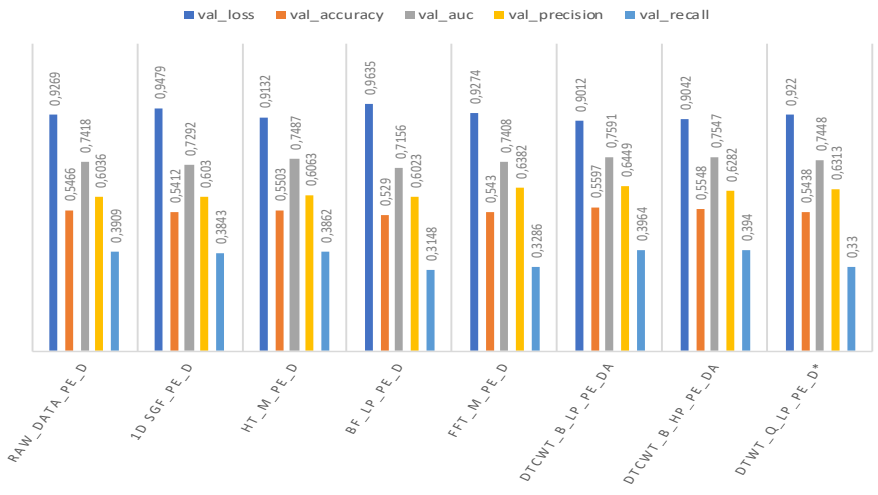
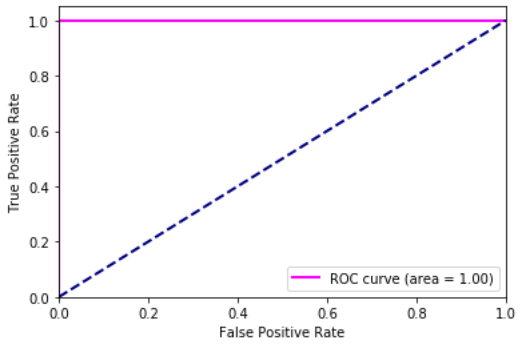
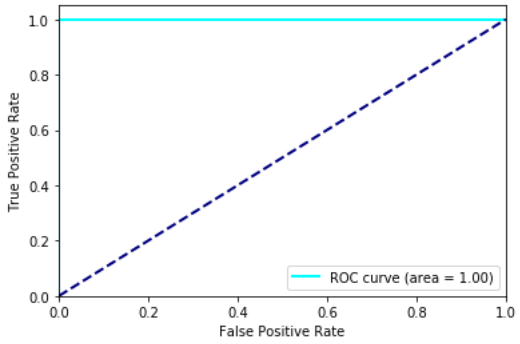


Figure 28 Multi-class classification without custom indexing

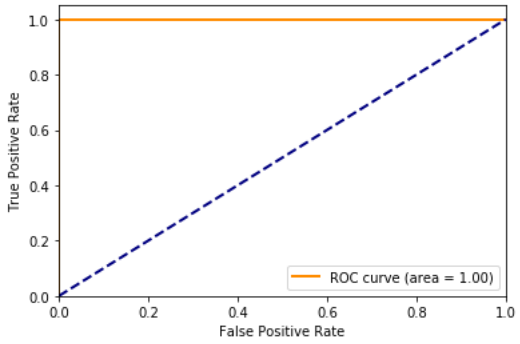
Inserting the custom indices, the multi-class classification performs perfectly in terms of accuracy, auc, precision, entropy e.t.c. as illustrated in **Fig. 29**.



(a)



(b)



(c)

Figure 29 ROC for
a) working state, b) failure 1 and c) failure 2

By inserting the four time-segments for reaching fault prognosis and with the same parametres, the models perform highly enough as illustrated in **Table 55** and **Fig. 30**, which it also shown DTCWT biort low-pass variant outperforms the other techniques in.

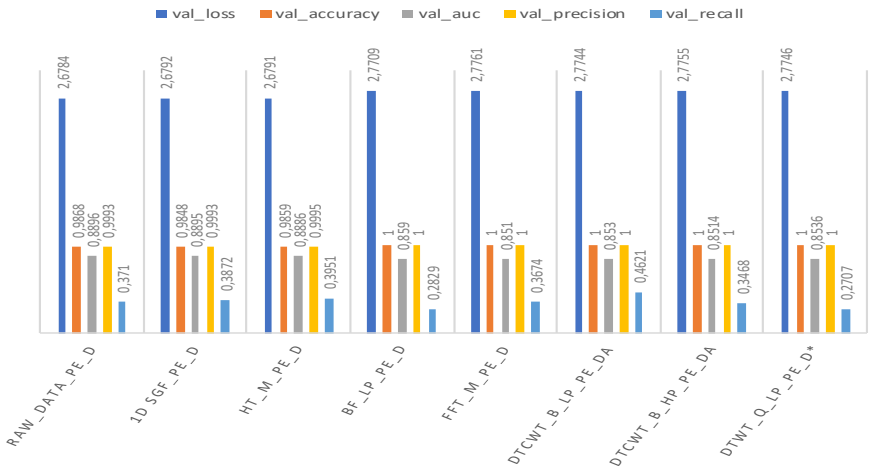


Figure 30 Multi-label classification

Afterwards, DTCWT biort low-pass variant is trained multiple times through GA to find out the optimal hyperparametres: BiLSTM units and epochs. Moreover, taking into account the processing demand, it is chosen the population size equals to 20, the number of generations also equals to 20, the batch size equals to 100 and the genes length equals to 13. Optimasing based on the fitness loss, i.e. the categorical crossentropy, it is found that units should be equal to 100 and the epochs equals to 63 throughout evolution process as illustrated in **Fig. 23**. The performance of the fine-tuned BiLSTM of the DTCWT biort low-pass variant is shown in the **Fig. 32** and **Tables 56**.

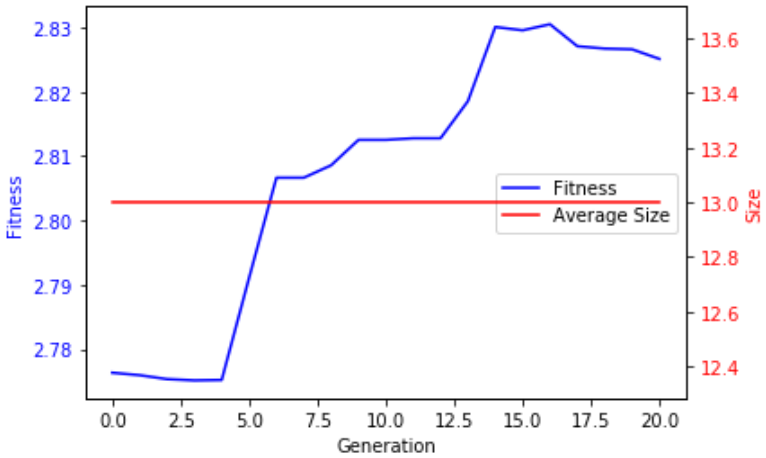


Figure 31 Evolution Process of Optimasing the BiLSTM

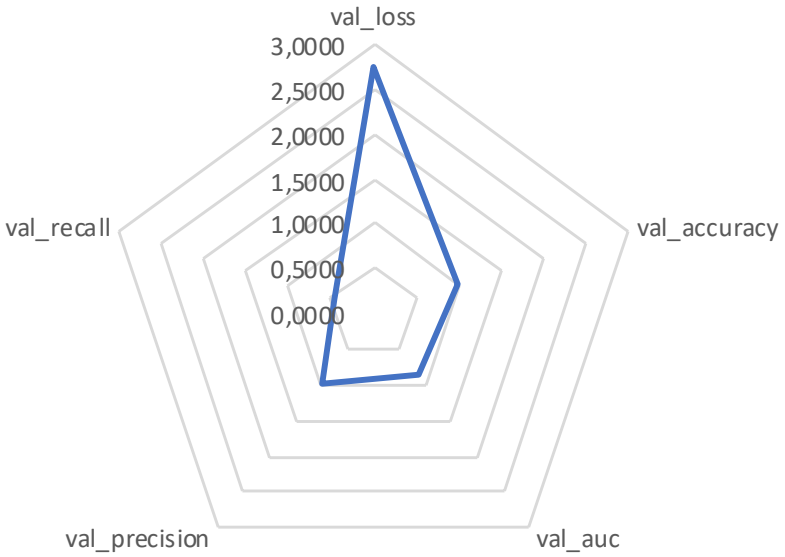


Figure 32 Spider Chart: The performance of the fine-tuned BiLSTM

However, having a close look in the performance of the proposed method, it is found that the time segments prediction again performs poorly as shown in **Fig. 33**, i.e. the classes 3 to 6 have got approximately AUC = .5.

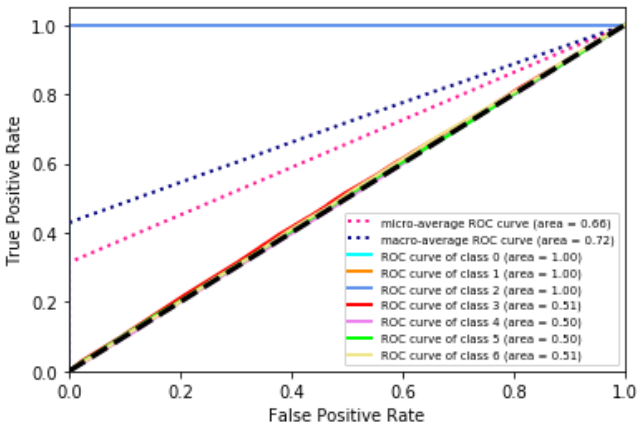


Figure 33 ROC curve

Taking the values of HT for Y axis, time is estimated by LR with pseudo-randomly shuffling and it is achieved an acceptable performance, i.e. $R^2 \approx .73$, RMSE≈223.43 and MAE≈.73 as plotted in **Fig. 34**, where it is also deserved the simple and mean confidence bands for p=.5.

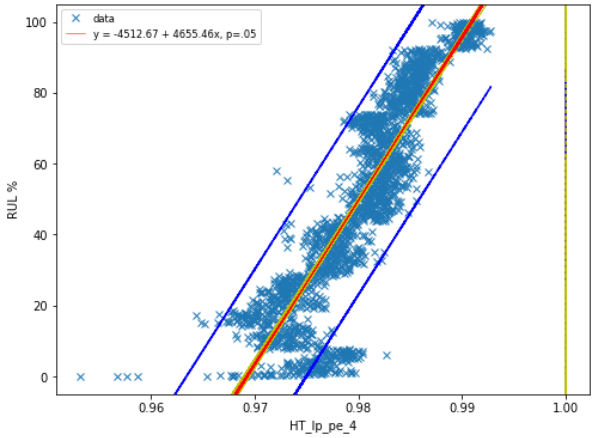


Figure 34 Linear Regression of RUL

A BiLSTM model is constructed for RUL estimation by dividing the time into 359 instances of around 20 minutes each and setting a looking back time steps equal to 4. Furthermore, for biort variant of the DTCWT for level=4 after PE, log normalisation 100 training epochs, ReLU activation and MAE for loss function. As shown in **Fig. 35**, the model for a 20% of validation shows as minimum values RMSE≈.087, MAE≈.007, MAE before validation loss≈.007. The regression is illustrated in **Fig. 36**.

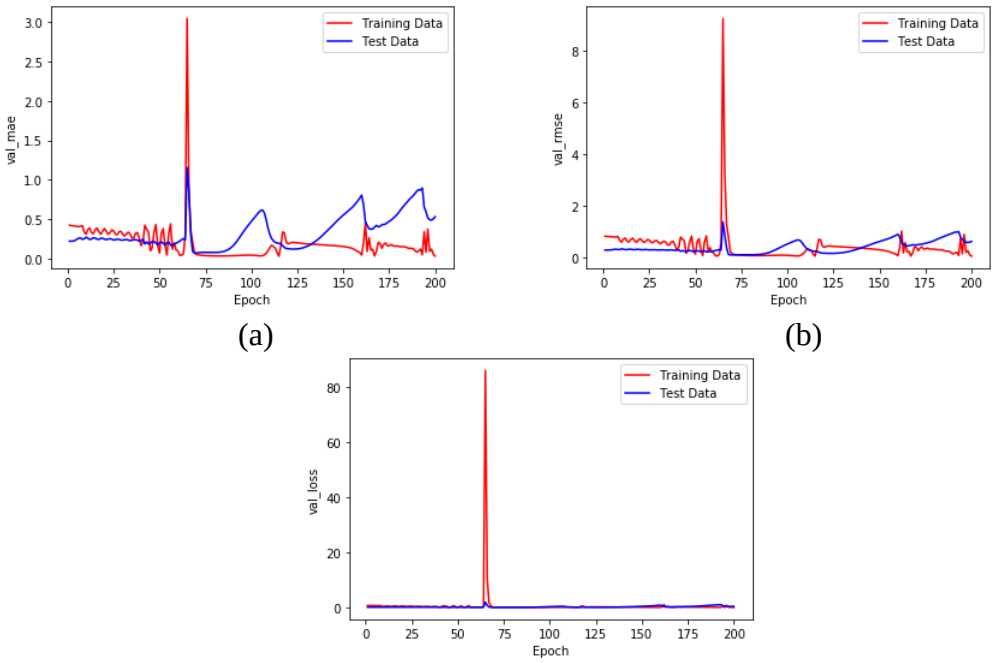


Figure 35 BiLSTM regression validated performance
a) MAE, b) RMSE and c) Loss

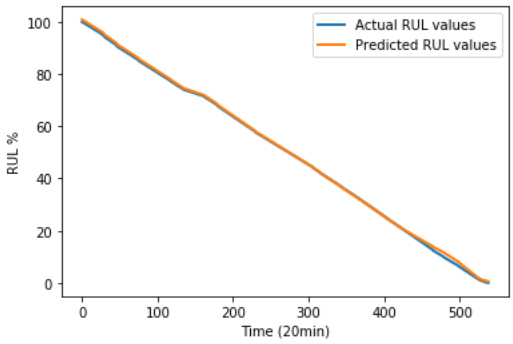


Figure 36 Regression of RUL

Finally, an array is constructed with 17,248 instances. Afterwards, four features are selected, i.e. the processed signal of DTCWT biort variant with level=4 and form=2 is taken as input and the operating states, failure 1, failure 2 and working, are taken as outputs. Then, time step is set to be equal to 3, the same as the units. A BiLSTM with 4 hidden layers is fed with softmax activation function, loss is chosen the categorical cross-entropy and the optimiser is selected the Nadam. For 20 epochs, the batch size to be equal to 1,533, 20% dropout, 20% validation and without shuffling the constructed model shows a perfect performance in terms of validated metrics such as AUC, PRE and ACC.

6. CONCLUSION

At the beginning, a literature review through technical manual was conducted in order all bearings used in the Greek fleet of LMs are listed. Enough types of bearings are encountered and the results show that the MLW LM has got more bearings than the other two types. Furthermore, the ADtranz LM has got more bearings that the Siemens LM.

The most common type is the deep grove ball bearings with two shields in all three LMs besides other types e.g. cylindrical, front/rear taped bearings and ball bearings with single shield. The electric motors require bearings in contrast with the other mechanisms.

For the bearing fault prognosis, it is used the first dataset of the IMS database. During all the process, many techniques are used and tested. A multi-level method is proposed with the addition of custom indices for constructing a multi-label classification ML mode.

Initially, signal is processed by plenty of transforms and filters for denoising. Afterwards, PE transforms all information into a single value for each instance. Then, all values are clustered by K-means resulting to constructing clusters that help the BiLSTM algorithm in estimations. However, the application of K-mean shew no strong effect in the performance evaluation.

The crucial point is the construction of four indices based on the descriptive statistics which boosts the overall performance. By inserting these custom indices, the method can classify each three distinct working states, i.e. working, failure 1 and failure 2. The proposed model performs perfectly in terms of categorical accuracy, auc, precision and recall, showing superiority over the previous methods.

Time segmentation is proved that is not a rigorous technique for RUL estimation. By using the BiLSTM algorithm tuned by GA, the overall performance remains highly enough in all metrics. However, the final ML model predicts only the working state, but not the RUL segments.

Afterwards, a LR is constructed based on the most correlated value for RUL estimation with an acceptable performance. Then, a regression model for RUL based on time steps is constructed showing superiority over last techniques.

The prediction of the operation states is also made by a multi-class classification ML model using the processed signal. The model outperforms its rivals by reaching perfect scores in all three metrics as happened also in the other approach.

All null hypotheses are rejected and the hypotheses H.1.1, H2.0, H3.1, H4.1 and H5.1 are accepted. So, the research questions are answered with yes. Finally, the future research could include the three datasets for modelling and online real-time successors of the present algorithms for products to scale.

7. BIBLIOGRAPHY

Αλεξανδρόπουλος, Α., Κατωπόδης, Ε., Παλιάτσος, Α., & Πρεζεράκος, Ν. (1994). *ΣΤΑΤΙΣΤΙΚΗ*. Αθήνα, Ελλάδα: Σύγχρονη Εκδοτική.

Abdelkrim, C., Meridjet, M. S., Boutasseta, N., & Boulanouar, L. (2019). Detection and classification of bearing faults in industrial geared motors using temporal features and adaptive neuro-fuzzy inference system. *Heliyon*, *5*(8), e02046. https://doi.org/10.1016/j.heliyon.2019.e02046

According to American Bearing Manufacturers Association (n.d.). *ABMA 20*. https://www.americanbearings.org/

ALCO (2003). *Drawings. OSE RAILWAY IN GREECE MX-627 UPGRADE.*

Alpaydin, E. (2010). *Introduction to Machine Learning* (2nd ed.). Massachusetts, USA The MIT Press. ISBN 978-0-262-01243-0

Al-Waily, M. (2017). *Bearing Design*. https://doi.org/10.13140/RG.2.2.15963.36643

Baluja, S., & Caruana, R. (1995). Removing the genetics from the standard genetic algorithm. *Proceedings of the Twelfth International Conference on Machine Learning*, pages 38–46. Morgan Kaufmann, San Mateo, CA.

Beyer, H.-G., & Schwefel, H.-P. (2002). Evolution strategies – A comprehensive introduction. *Natural Computing*, *1*(1), 3–52. https://doi.org/10.1023/A:1015059928466

Ben Ali, J., Saidi, L., Mouelhi, A., Chebel-Morello, B., & Fnaiech, F. (2015). Linear feature selection and classification using PNN and SFAM neural networks for a nearly online diagnosis of bearing naturally progressing degradations. *Engineering Applications of Artificial Intelligence*, *42*, 67–81. https://doi.org/10.1016/j.engappai.2015.03.013

Bengio, Y. (2009). *Learning Deep Architectures Vol.2.* Montréal, Canada: Foundation and Trends in Machine Learning, NOW the essence of knowledge.

Bishop, C., M. (2006). *Pattern Recognition and Machine Learning.* Cambridge, UK: Springer. ISBN-13: 978-0387-31073-2

Boškoski, P., Gašperin, M., Petelin, D., & Juričić, Đ. (2015a). Bearing fault prognostics using Rényi entropy based features and Gaussian process models. *Mechanical Systems and Signal Processing*, *52–53*, 327–337. https://doi.org/10.1016/j.ymssp.2014.07.011

Bombardier Transortation GmbH (2004). *Dieselelektrische Lokomotive OSE DE 2000. 3EGH314 264.PBE.de.4. Ordner I-IIX.*

Cheng, J., & Xiong, Y. (2018). Application of Teaching-Learning-Based Optimization Algorithm in Rotor Fault Diagnosis for Asynchronous Motor. *Recent Advancement in Information and Communication Technology:*, *131*, 1275–1281. https://doi.org/10.1016/j.procs.2018.04.341

Chakrabarty, A., Mannan, S., & Cagin, T. (2016). Multiscale Modeling for Process Safety Applications. Oxford, UK: Butterworth-Heinemann, Elsevier. ISBN: 978-0-12-396975.

Chen, P., Y., & Popovich, P., M. (2002). *CORRELATION. Parametric and Nonparametric Measures.* California, USA: Sage Publications. ISBN 0-7619-2228-8 (p).

Cicirello, V., & Smith, S. (2000). Modeling GA Performance for Control Parameter Optimization. In *Proceedings of the Genetic and Evolutionary Computation Conference*.

Greene, W., H. (2002). *Econometric Analysis* (5th ed.). New Jersey, USA: Prentice Hall. ISBN 0-13-066189-9.

Gryllias, K. C., & Antoniadis, I. A. (2012). A Support Vector Machine approach based on physical model training for rolling element bearing fault detection in industrial environments. *Special Section: Local Search Algorithms for Real-World Scheduling and Planning*, *25*(2), 326–344. https://doi.org/10.1016/j.engappai.2011.09.010

Cui, Z., Ke, R., Pu, Z., & Wang, Y. (2020). Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values. *Transportation Research Part C: Emerging Technologies*, *118*, 102674. https://doi.org/10.1016/j.trc.2020.102674

D'Agostino, R., B. (1971). An omnibus test of normality for moderate and large sample size. *Biometrika*, 58, 341-348.

D'Agostino, R., & Pearson, E., S. (1973). Tests for departure from normality. *Biometrika*, 60, 613-622.

Day, N., Ashour, A., S., Fong, S., J., & Borra, S. (Eds.). (2019). *U-Healthcare Monitoring Systems. Volume 1: Design and Applications. A volume in Advances in Ubiquitous Sensing Applications for Healthcare.* Academic Press. Elsevier.

https://doi.org/10.1016/C2017-0-03248-2

Deb, K., Agrawal, S., Pratap, A., & Meyarivan, T. (2000). A Fast Elitist Non-dominated Sorting Genetic Algorithm for Multi-objective Optimization: NSGA-II. In M. Schoenauer, K. Deb, G. Rudolph, X. Yao, E. Lutton, J. J. Merelo, & H.-P. Schwefel (Eds.), *Parallel Problem Solving from Nature PPSN VI* (pp. 849–858). Springer Berlin Heidelberg.

De Carlo, L., Perkins, K., & Caputo, M. C. (2021). Evidence of Preferential Flow Activation in the Vadose Zone via Geophysical Monitoring. *Sensors*, *21*(4). https://doi.org/10.3390/s21041358

Dhillon, B., S. (2006). Maintainability, Maintenance and Reliability for Engineers. *Publisher CRC Press*. ISBN 1420006789.

Dorigo, M. & Stutzle, T. (2004). *Ant Colony Optimization.* Massachusetts, USA: The MIT Press.

Eiben, A., E. & Smith, J., E. (2003). *Introduction to Evolutionary Computing.* Heidelberg, Germany: Springer-Verlag.

Elsheikh, A., Yacout, S., & Ouali, M.-S. (2019). Bidirectional handshaking LSTM for remaining useful life prediction. *Neurocomputing*, *323*, 148–156. https://doi.org/10.1016/j.neucom.2018.09.076

European Parliament (n.d.). *Fact Sheet on the European Union*. Retrieved October 2, 2019, from http://www.europarl.europa.eu/factsheets/en/sheet/123/common-transport-policy-overview

FAG (1999). *FAG Rolling Bearings. Technical Information TI No. WL 43-1190 EA.* www.fagbearing.cc. www.schaeffler.de

Fei, M., Ning, L., Huiyu, M., Yi, P., Haoyuan, S., & Jianyong, Z. (2018). On-line fault diagnosis model for locomotive traction inverter based on wavelet transform and support vector machine. *29th European Symposium on Reliability of Electron Devices, Failure Physics and Analysis (ESREF 2018), 88–90*, 1274–1280. https://doi.org/10.1016/j.microrel.2018.06.069

Feng, C., Wang, H., Lu, N., Chen, T., He, H., Lu, Y., & Tu, X., M. (2014). Log-transformation and its implications for data analysis. *Shanghai Archives of Psychiatry.*

Feng, C., Wang, H., Lu, N., & Tu, X. M. (2013). Log transformation: Application and interpretation in biomedical research. *Statistics in Medicine*, *32*(2), 230–239.

https://doi.org/10.1002/sim.5486

Feng, F., Si, A., & Zhang, H. (2011). Research on Fault Diagnosis of Diesel Engine Based on Bispectrum Analysis and Genetic Neural Network. *CEIS 2011*, *15*, 2454–2458. https://doi.org/10.1016/j.proeng.2011.08.461

Fischer-Cripps A., C. (2002). *Newnes Interfacing Companion.* Newnes. Elsevier. https://doi.org/10.1016/B978-0-7506-5720-4.X5100-6

Flett, J., & Bone, G. M. (2016). Fault detection and diagnosis of diesel engine valve trains. *Mechanical Systems and Signal Processing*, *72–73*, 316–327. https://doi.org/10.1016/j.ymssp.2015.10.024

Floreano, D. & Mattiussi, C. (2008). *Bio-Inspired. Artificial Intelligence. Theories, Methods and Technologies.* Massachusetts, USA: The MIT Press.

Gao, Z., Huo, B., Zhang, J., & Jiang, Z. (2019). Failure investigation of gear teeth fracture of seawater pump in a diesel engine. *Engineering Failure Analysis*, *105*, 1079–1092. https://doi.org/10.1016/j.engfailanal.2019.07.050

Goldberg, E., D. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning.* USA: Addison – Wesley Publishing Company, Inc.
ISBN: 0-201-15767-5.

Glowacz, A., Glowacz, W., Glowacz, Z., & Kozik, J. (2018). Early fault diagnosis of bearing and stator faults of the single-phase induction motor using acoustic signals. *Measurement*, *113*, 1–9. https://doi.org/10.1016/j.measurement.2017.08.036

Haidong, S., Junsheng, C., Hongkai, J., Yu, Y., & Zhantao, W. (2020). Enhanced deep gated recurrent unit and complex wavelet packet energy moment entropy for early fault prognosis of bearing. *Knowledge-Based Systems*, *188*, 105022. https://doi.org/10.1016/j.knosys.2019.105022

Hai, Q., Jay, L. & Jing, L. (2006). Wavelet Filter-based Weak Signature Detection Method and its Application on Roller Bearing Prognostics. *Journal of Sound and Vibration* (289)1066-1090.

Han, J. & Kamber, M. (2012). *Data Mining* (3rd ed.). Waltham, MA, USA: Morgan Kaufmann Publishers, Elsevier.

Hosmer, D., W., Lemeshow, S., & Sturdivant, R., X. (2013). *Applied Logistic Regression* (3rd ed.). NY, USA: Wiley and Sons.
ISBN: 978-0-470-58247-3

Janicki, J., Reinhard, H. & Rüffer, M. (2013). Schienenfahrzaugtechnik. Berlin, Germany: DB-Fachbuch, Bahn Fachverlag GmbH.

Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, *20*(7), 1483–1510. https://doi.org/10.1016/j.ymssp.2005.09.012

Jia, G., Yuan, S., & Tang, C. (2011). Fault Diagnosis of Roller Bearing Based on PCA and Multi-class Support Vector Machine. In D. Li, Y. Liu, & Y. Chen (Eds.), Computer and Computing Technologies in Agriculture IV (pp. 198–205). Springer Berlin Heidelberg.

Joshi, P. (2017). *Artificial Intelligence with Python.* Birmingham, UK: Packt Publishing.

Κατωπόδης, Ε., Μακρυγιάννης, Α., & Σάσσαλος, Σ. (1995). *ΜΑΘΗΜΑΤΙΚΑ Ι. ΑΛΓΕΒΡΑ – ΑΝΑΛΥΤΙΚΗ ΓΕΩΜΕΤΡΙΑ* (Vol. I). Σύγχρονη Εκδοτική.

Kasabov, N. (2007). *Evolving Connectionist Systems. The Knowledge Engineering Approach* (2nd ed.). London, UK: Springer Science + Business Media, Springer-Verlag.

Kennedy, J., & Elberhart, C. (2001). *Swarm Intelligence.* California, USA: Morgan Kaufmann Publishers. ISBN 1-55860-595-9

Kim, H.-E., Tan, A. C. C., Mathew, J., & Choi, B.-K. (2012). Bearing fault prognosis based on health state probability estimation. *Expert Systems with Applications*, *39*(5), 5200–5213. https://doi.org/10.1016/j.eswa.2011.11.019

Kostić, D., Drndarević, V., Marković, P., & Jevtić, N. (2011). Development of methods for acquiring and transfering measurement data in testing the electric locomotives. Transport, 26(4), 367-374. https://doi.org/10.3846/16484142.2011.557217

Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., & Dyer, C. (2016). *Neural architectures for named entity recognition*. arXiv:1603.01360 [cs.CL]

Lee, J., Qiu, H., & Lin, J. (2006). Bearing Data Set. [Data set]. NASA Ames Prognostics Data Repository. http://ti.arc.nasa.gov/project/prognostic-data-repository.

Leite, G. de N. P., Araújo, A. M., Rosas, P. A. C., Stosic, T., & Stosic, B. (2019). Entropy measures for early detection of bearing faults. *Physica A: Statistical*

Mechanics and Its Applications, *514*, 458–472. https://doi.org/10.1016/j.physa.2018.09.052

Leonard, L. C. (2017). Chapter One—Web-Based Behavioral Modeling for Continuous User Authentication (CUA). In A. M. Memon (Ed.), *Advances in Computers* (Vol. 105, pp. 1–44). Elsevier. https://doi.org/10.1016/bs.adcom.2016.12.001

Li, X., Yang, X., Yang, Y., Bennett, I., & Mba, D. (2019). A novel diagnostic and prognostic framework for incipient fault detection and remaining service life prediction with application to industrial rotating machines. *Applied Soft Computing*, *82*, 105564. https://doi.org/10.1016/j.asoc.2019.105564

Lobontiu, N., (2010). *System Dynamics for Engineering Students.* Academic Press. Elsevier. ISBN 978-0-240-81128-4

Loshin, D. (2009). Chapter 10—Data Consolidation and Integration. In D. Loshin (Ed.), *Master Data Management* (pp. 177–199). Morgan Kaufmann. https://doi.org/10.1016/B978-0-12-374225-4.00010-2

Luger, G., F. (2009). *Artificial Intelligence. Structures and Strategies for Complex Problem Solving.* (6th ed.) Boston, USA: Pearson Education, Inc. ISBN-13: 978-0-321-54589-3.

MacKinnon, G., J. (2010, January). *Critical Values for Cointegration Test.* QUEENSU. http://qed.econ.queensu.ca/working_papers/papers/qed_wp_1227.pdf

Manning, C., D., & Raghavan, P., & Schütze, H. (2009c). *An Introduction to Information Retrieval.* Cambridge, UK: Cambridge University Press. ISBN: 0521865719

McHugh, M., L. (2008). Standard error: meaning and interpretation. *Biochemia Medica*, 18(1), 7-13.

Marsland, S. (2009). *Machine Learning: An Algo- rithmic Perspective.* (2nd ed.) Florida, USA: CRC Press /Chapman & Hall. Taylor & Francis Group. ISBN 978-1-4665-8328-3

MLW Industries (n.d.). *Technical Manual DRP.*

Mobley, R, .K. (2002). An introduction to predictive maintenance. *A Volume in Plant Engineering* (2nd ed.). Butterworth-Heinemann. Elsevier Inc. https://doi.org/10.1016/B978-0-7506-7531-4.X5000-3

Mohamad, I., & Usman, D. (2013). Standardization and Its Effects on K-Means Clustering Algorithm. *Research Journal of Applied Sciences, Engineering and Technology*, *6*, 3299–3303. https://doi.org/10.19026/rjaset.6.3638

Moussa Nahim, H., Younes, R., Shraim, H., & Ouladsine, M. (2016). Modeling with Fault Integration of the Cooling and the Lubricating Systems in Marine Diesel Engine: Experimental validation. *8th IFAC Symposium on Advances in Automotive Control AAC 2016*, *49*(11), 570–575. https://doi.org/10.1016/j.ifacol.2016.08.083

Müller, C., A., & Guido, S. (2017). *Introduction to Machine Learning with Python. A Guide for Data Scientists.* California, USA: O' Reilly. ISBN 978-1-449-36941-5

Murphy, K., P. (2012). *Machine Learning: A Probabilistic Perspective.* Massachusetts, USA The MIT Press.
ISBN: 9780262018029

Neil, S., P., & Hashemi, M., R. (2018). *Fundamentals of Ocean Renewable Energy. Generating electricity from the sea* (Vol. A). Elsivier Ltd. p. 336. ISBN 978-0-12-810448-4. https://doi.org/10.1016/C2016-0-00230-9

Niu, G., Zhao, Y., Defoort, M., & Pecht, M. (2015). Fault diagnosis of locomotive electro-pneumatic brake through uncertain bond graph modeling and robust online monitoring. *Mechanical Systems and Signal Processing*, *50–51*, 676–691. https://doi.org/10.1016/j.ymssp.2014.05.020

NSK Company (2016). *NSK Motion & Control. Rolling Bearings for Industrial Machinery. Catalog No. E1103.*
https://www.nsk.com

Oppenheim, A., V., & Schafer, R., W. (1999). *Discrete-Time Signal Processing* (2nd Ed.). New Jersey, USA: Prentice-Hall. ISBN 0-13-754920-2

Ouadfeul, Sid-Ali, Aliouane, L., Hamoudi, M., & Amar, B. (2012). *Wavelet Transforms and Their Applications in Biology and Geoscience.* Researchgate Publications. https://doi.org/10.5772/36747

Ozdemir, S., & Susarla, D. (2018). *Feature Engineering Made Easy.* O'Reilly, Packt Publishing. ISBN 9781787287600

Percival, D. B., & Mondal, D. (2012). 22—A Wavelet Variance Primer. In T. Subba Rao, S. Subba Rao, & C. R. Rao (Eds.), *Handbook of Statistics* (Vol. 30, pp. 623–657). Elsevier. https://doi.org/10.1016/B978-0-444-53858-1.00022-3

Rasmussen, C., E. & Williams, K., I. (2006). Gaussian Processes for Machine Learning. Massachusetts, USA: the MIT Press.

Rumsey, D., J. (2016). *Statistics for Dummies* (2nd ed.) John Wiley & Sons. ISBN-13: 978-1119293521.

Sachan, S., Shukla, S., & Singh, S. K. (2020). Two level de-noising algorithm for early detection of bearing fault using wavelet transform and zero frequency filter. *Tribology International*, *143*, 106088. https://doi.org/10.1016/j.triboint.2019.106088

Saxena, M., Bannett, O. O., & Sharma, V. (2016). Bearing Fault Evaluation for Structural Health Monitoring, Fault Detection, Failure Prevention and Prognosis. *International Conference on Vibration Problems 2015*, *144*, 208–214. https://doi.org/10.1016/j.proeng.2016.05.026

Seber ,G.(1977). *Linear Regression Analysis.* NY, USA: John Wiley & Sons. p. 465

Segaran, T. (2007). Programming Collective Intelligence. Building Smart Web 2.0 Applications.

Sekaran, U. & Bougie, R. (2009). *Research Methods for Business. A Skill-Building Approach* (5th ed.). West Sussex, UK: John Wiley & Sons Ltd.

Shao, H., Jiang, H., Li, X., & Liang, T. (2018). Rolling bearing fault detection using continuous deep belief network with locally linear embedding. Computers in Industry, 96, 27–39. https://doi.org/10.1016/j.compind.2018.01.005

Shawe-Taylor, J. & Cristianini, N. (2004). Kernel Methods for Pattern Analysis. Cambridge, UK: Cambridge University Press.

Siemens (2004). *Lokomotive Hellas Sprinter. MODUL 5. Subsysteme und Geräte E-Teil. PA:#39. Lok.Nr. 120 007 – 120 030. Fahrmotro 1TB2824-1GA02. -1M1, -1M2, -1M3, -1M4. F.0090.90.1. Version 1.0.*

SKF Group (2018). *SKF bearings and mounted products.* www.skf.com/usgencatalog

Sherstinsky, A. (2020). Fundamentals of Recurrent Neural Network (RNN) and Long-Short Memory (LSTM) network. *Physica D: Phenomena, 404, 132306.* https://doi.org/10.1016/j.physd.2019.132306

Smith, W., S. (2002). Digital Signal Processing. *A Practical Guide for Engineers and Scientists.* Newnes. Elsevier. https://doi.org/10.1016/B978-0-7506-7444-7.X5036-5

Smola, A., & Vishwanathan, S., V., N. (2008). *Introduction to Machine Learning.* Cambridge, UK: Cambridge University Press. ISBN 0 521 82583 0

Spoerre, J. K. (1997). Application of the cascade correlation algorithm (CCA) to bearing fault classification problems. Computers in Industry, 32(3), 295–304. https://doi.org/10.1016/S0166-3615(96)00080-2

Ruey S. Tsay. 2012. *An Introduction to Analysis of Financial Data with R* (1st. ed.). Wiley Publishing.

Steinley, D. (2004). Standardizing Variables in K-means Clustering. In D. Banks, F. R. McMorris, P. Arabie, & W. Gaul (Eds.), *Classification, Clustering, and Data Mining Applications* (pp. 53–60). Springer Berlin Heidelberg.

Suthar, K., Shah, D., Wang, J., & Peter He, Q. (2018). Feature-based Virtual Metrology for Semiconductor Manufacturing. In M. R. Eden, M. G. Ierapetritou, & G. P. Towler (Eds.), *Computer Aided Chemical Engineering* (Vol. 44, pp. 2083–2088). Elsevier. https://doi.org/10.1016/B978-0-444-64241-7.50342-6

Tahir, M., M., Khan, A., Q., Iqbal, N., Hussain, A., & Badshah., S. (2017). Enhancing Fault Classification Accuracy of Ball Bearing Using Central Tendency Based Time Domain Features. IEEE Access, 5, 72–83. https://doi.org/10.1109/ACCESS.2016.2608505

Theodoridis, S. & Koutroumbas, K. (2010). An Introduction to Pattern Recognition: A MATLAB Approach (4th ed.). Burlington, USA: Academic Press, Elsevier.

Thompson, M. T., (2014). *Intuitive Analog Circuit Design.* (2nd ed.). Elsevier. https://doi.org/10.1016/C2012-0-03027-X

Tidriri, K., Verron, S., Tiplica, T., & Chatti, N. (2019). A decision fusion based methodology for fault Prognostic and Health Management of complex systems. *Applied Soft Computing*, *83*, 105622. https://doi.org/10.1016/j.asoc.2019.105622

TIMKEN (n.d.) *Products Catalog 3.* www.timken.com

Vapnik, V., N. (1998). Statistical Learning Theory. New York, USA: Wiley-Interscience Publication, John Wiley & Sons, Inc.

Wang, D., Tsui, K., & Miao, Q. (2018). Prognostics and Health Management: A Review of Vibration Based Bearing and Gear Health Indicators. *IEEE Access*, *6*, 665–676

Wang, N., Zeng, N., & Zhu, W. (2010) Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Practical SAS Implementations. *Northeast SAS. Health Care and Life Sciences*, Baltimore, Maryland, 14-17 November 2010, 1-9.

Wang, X., Cai, Y., & Lin, X. (2014). Diesel Engine PT Pump Fault Diagnosis based on the Characteristics of its Fuel Pressure. *International Conference on Applied Computing, Computer Science, and Computer Engineering (ICACC 2013)*, *7*, 84–89. https://doi.org/10.1016/j.ieri.2014.08.014

Wang, Y., He, Z., & Zi, Y. (2010). Enhancement of signal denoising and multiple fault signatures detecting in rotating machinery using dual-tree complex wavelet transform. *Mechanical Systems and Signal Processing*, *24*(1), 119–137. https://doi.org/10.1016/j.ymssp.2009.06.015

Wang, Z., Yang, J., Cao, C., & Gu, Z. (2016). Phase-phase Short Fault Analysis of Permanent Magnet Synchronous Motor in Electric Vehicles. *CUE 2015 - Applied Energy Symposium and Summit 2015: Low Carbon Cities and Urban Energy Systems*, *88*, 915–920. https://doi.org/10.1016/j.egypro.2016.06.112

Willmott, C., J., & Matsura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research.* 30: 79–82. doi:10.3354/cr030079

Witten, I., H., Frank, E. & Hall, M., A. (2017). *DATA MINING. Practical Machine Learning Tools and Techniques* (4th ed.). Burlington, USA: Morgan Kaufmann Publishers, Elsevier.

Xu, G., Hou, D., Qi, H., & Bo, L. (2021). High-speed train wheel set bearing fault diagnosis and prognostics: A new prognostic model based on extendable useful life. *Mechanical Systems and Signal Processing*, *146*, 107050. https://doi.org/10.1016/j.ymssp.2020.107050

Yaguo L. (2017). *Intelligent Fault Diagnosis and Remaining Useful Life Prediction of Rotating Machinery*. Butterworth-Heinemann, https://doi.org/10.1016/C2016-0-00367-4

Zhang, R., Peng, Z., Wu, L., Yao, B., & Guan, Y. (2013). Fault Diagnosis from Raw Sensor Data Using Deep Neural Networks Considering Temporal Coherence. *Physical Sensors, 2017, 17(3), 549*, doi.org/10.3390/s17030549

Zhang, Z., Dong, F., & Xie, L. (2018). Data-Driven Fault Prognosis Based on Incomplete Time Slice Dynamic Bayesian Network1∗This work was supported in

part by the National Natural Science Foundation of China (No. 61374047, No. 61202473) and the Fundamental Research Funds for Central Universities (JUSRP111A49). *10th IFAC Symposium on Advanced Control of Chemical Processes ADCHEM 2018*, *51*(18), 239–244. https://doi.org/10.1016/j.ifacol.2018.09.306

Zheng, A., & Casari, A. (2018). *Feature Engineering for Machine Learning. Principles and Techniques for Data Scientists.* California, USA: O'Reilly Media. ISBN 978-1-491-95324-2.

Zhu, Q., He, Z., Zhang, T., & Cui, W. (2020). Improving Classification Performance of Softmax Loss Function Based on Scalable Batch-Normalization. *Applied Sciences, 2020, 10, 2950, doi:10.3390/app10082950*

Zukauskas, P., Vveinhardt, J., & Andriukaitienė, R. (2018). *Philosophy and Paradigm of Scientific Research*. https://doi.org/10.5772/intechopen.70628

APPENDIX

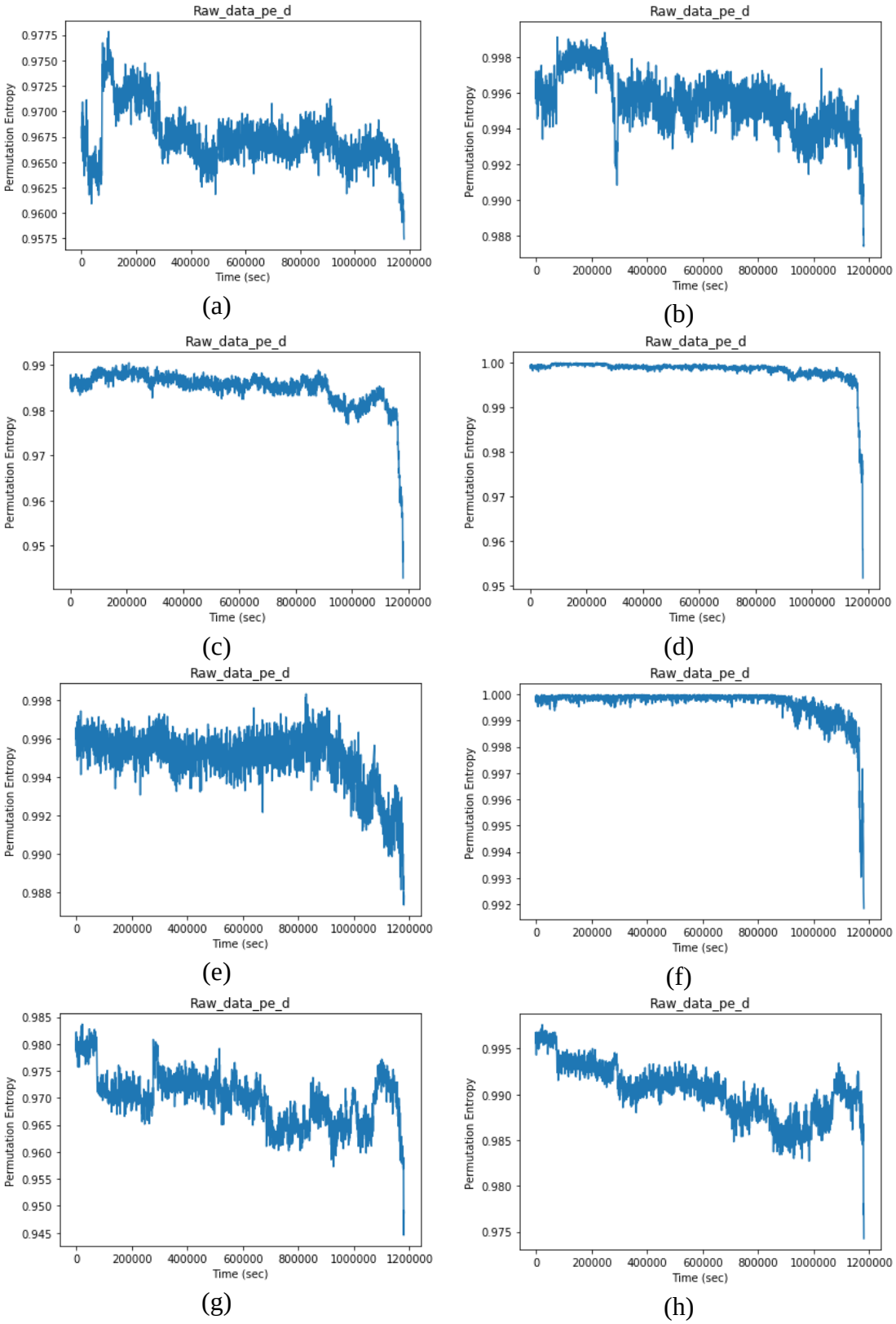


Figure 36 Raw Signal PE level=3

a) Bearing 1 X-axis, b) Bearing 1 Y-axis, c) Bearing 2 X-axis, d) Bearing 2 Y-axis, e) Bearing 3 X-axis,f) Bearing 3 Y-axis, g) Bearing 4 X-axis and h) Bearing 4 Y-axis

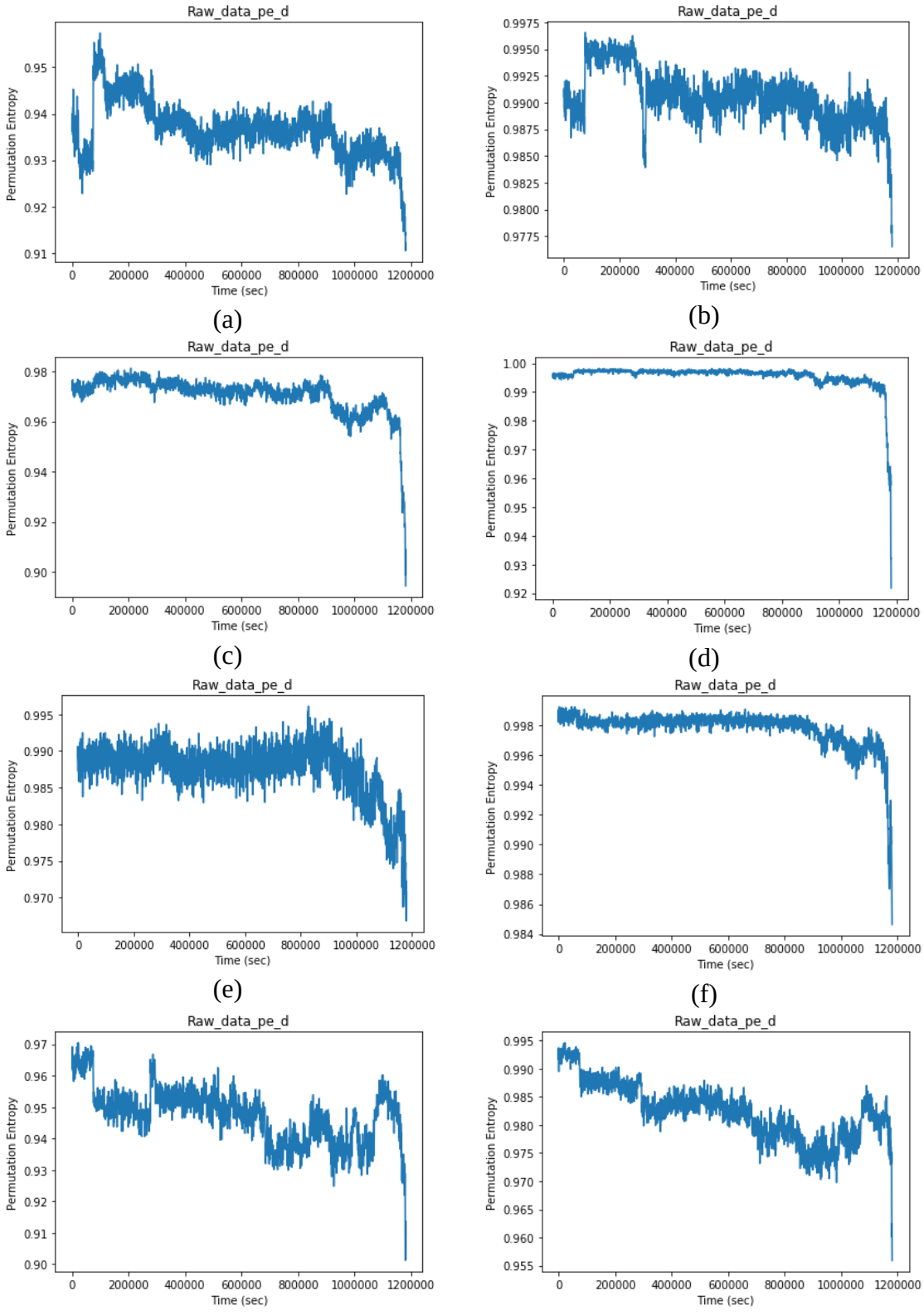


Figure 37 Raw Signal PE level=4
a) Bearing 1 X-axis, b) Bearing 1 Y-axis, c) Bearing 2 X-axis, d) Bearing 2 Y-axis, e) Bearing 3 X-axis,f) Bearing 3 Y-axis, g) Bearing 4 X-axis and h) Bearing 4 Y-axis

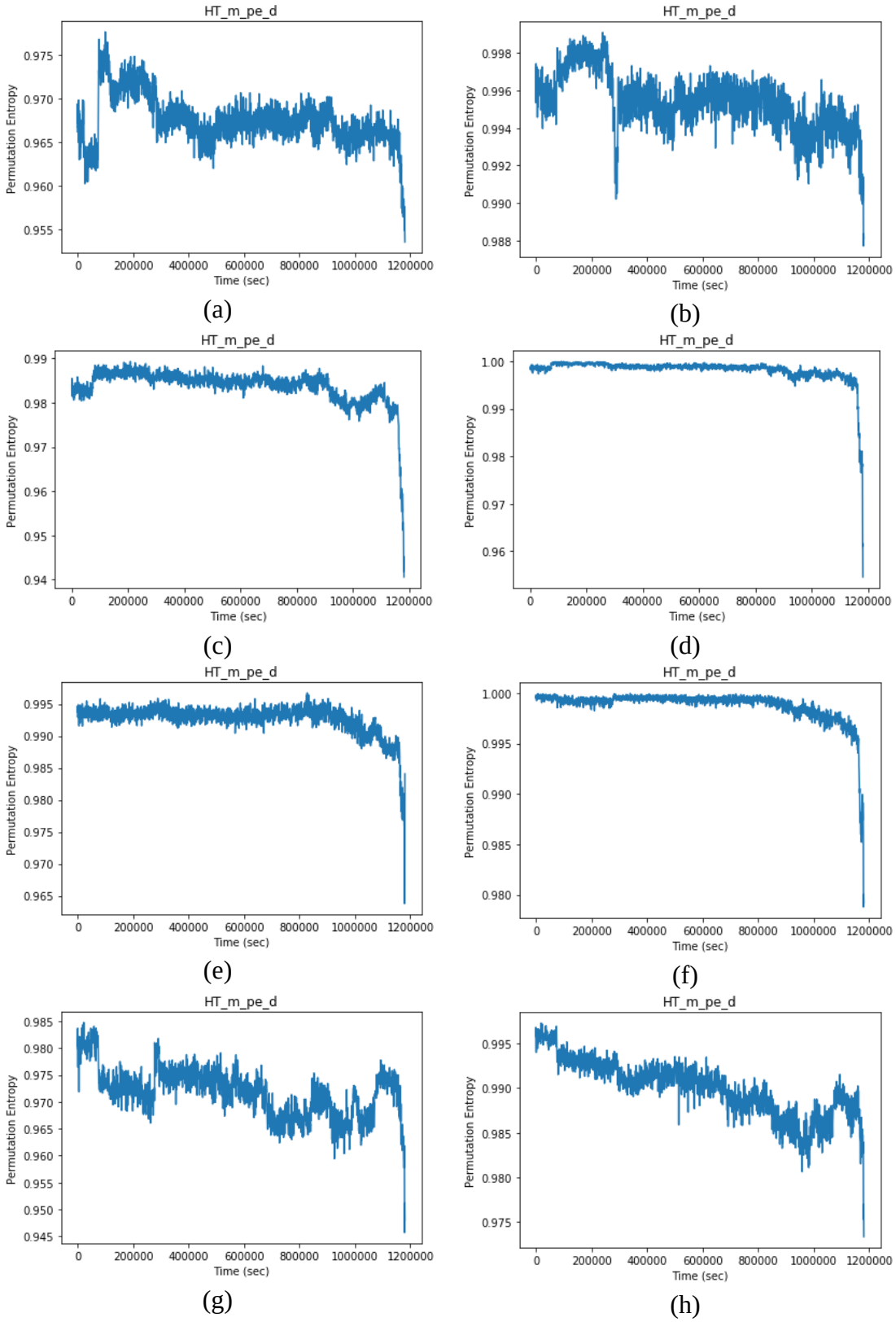


Figure 38 HT lowpass PE level=3

a) Bearing 1 X-axis, b) Bearing 1 Y-axis,c) Bearing 2 X-axis, d) Bearing 2 Y-axis, e) Bearing 3 X-axis,f) Bearing 3 Y-axis, g) Bearing 4 X-axis and h) Bearing 4 Y-axis

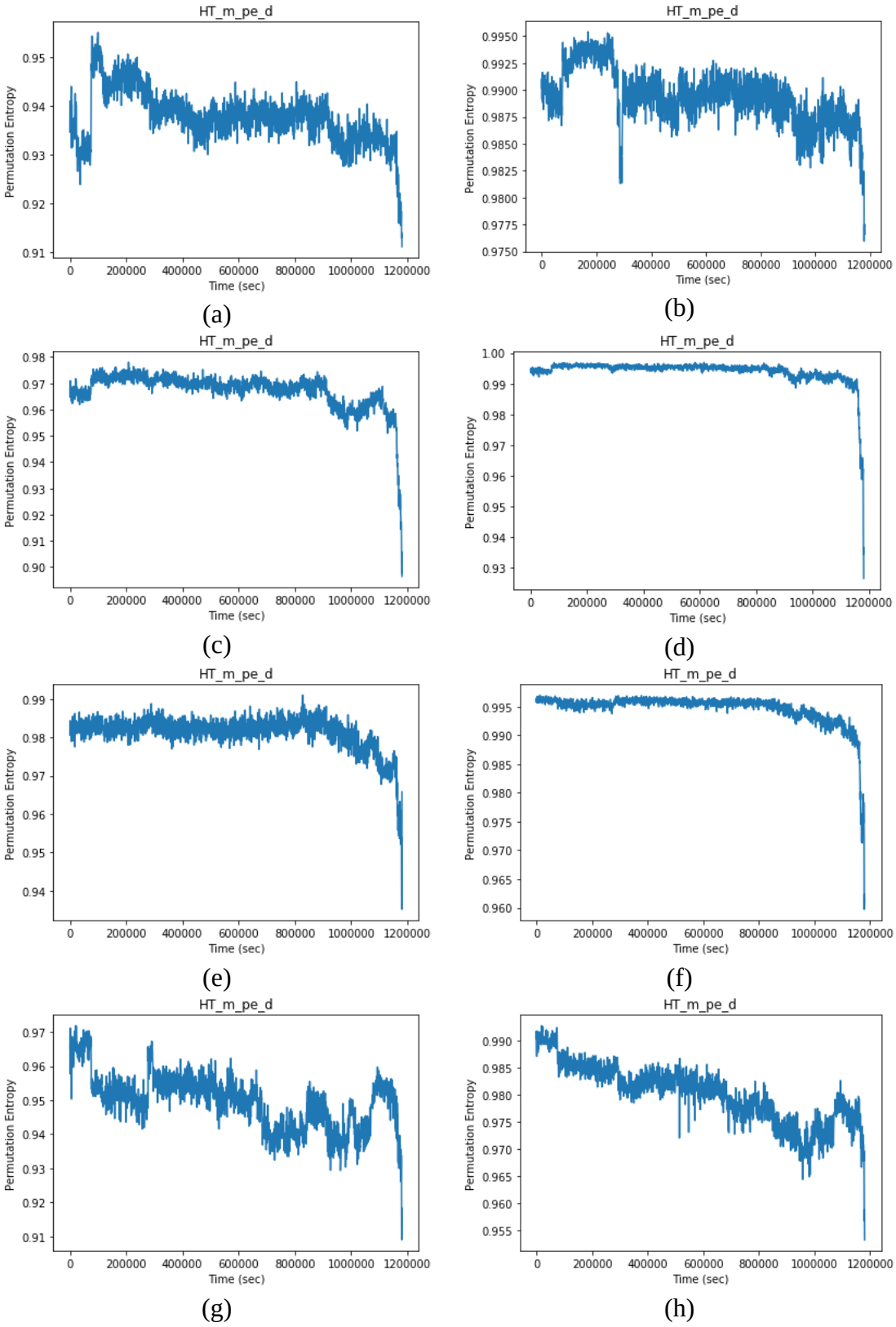


Figure 39 HT lowpass PE level=4

a) Bearing 1 X-axis, b) Bearing 1 Y-axis, c) Bearing 2 X-axis, d) Bearing 2 Y-axis, e) Bearing 3 X-axis,f) Bearing 3 Y-axis, g) Bearing 4 X-axis and h) Bearing 4 Y-axis

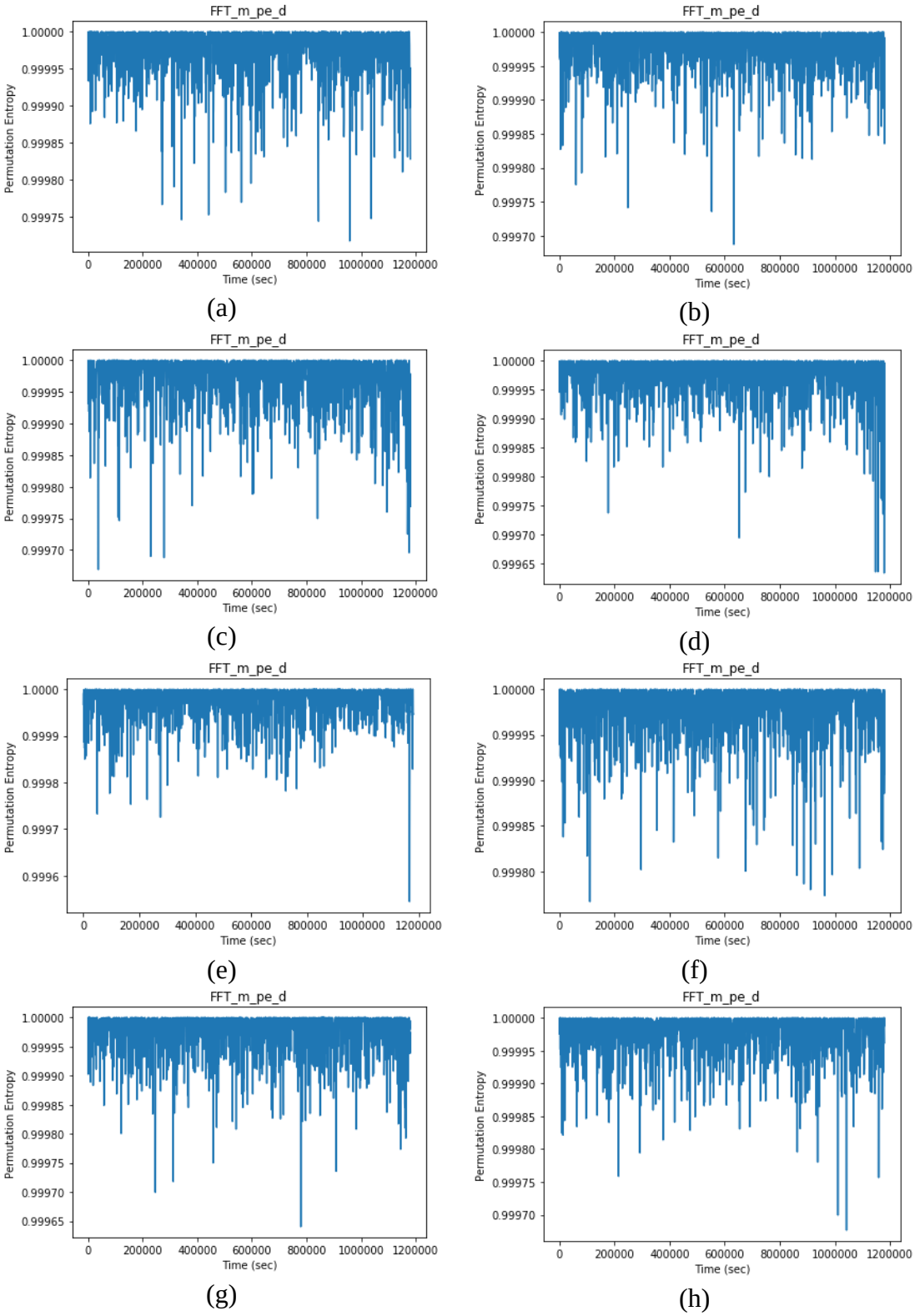


Figure 40 FFT magnitude PE level=3

a) Bearing 1 X-axis, b) Bearing 1 Y-axis, c) Bearing 2 X-axis, d) Bearing 2 Y-axis, e) Bearing 3 X-axis,f) Bearing 3 Y-axis, g) Bearing 4 X-axis and h) Bearing 4 Y-axis

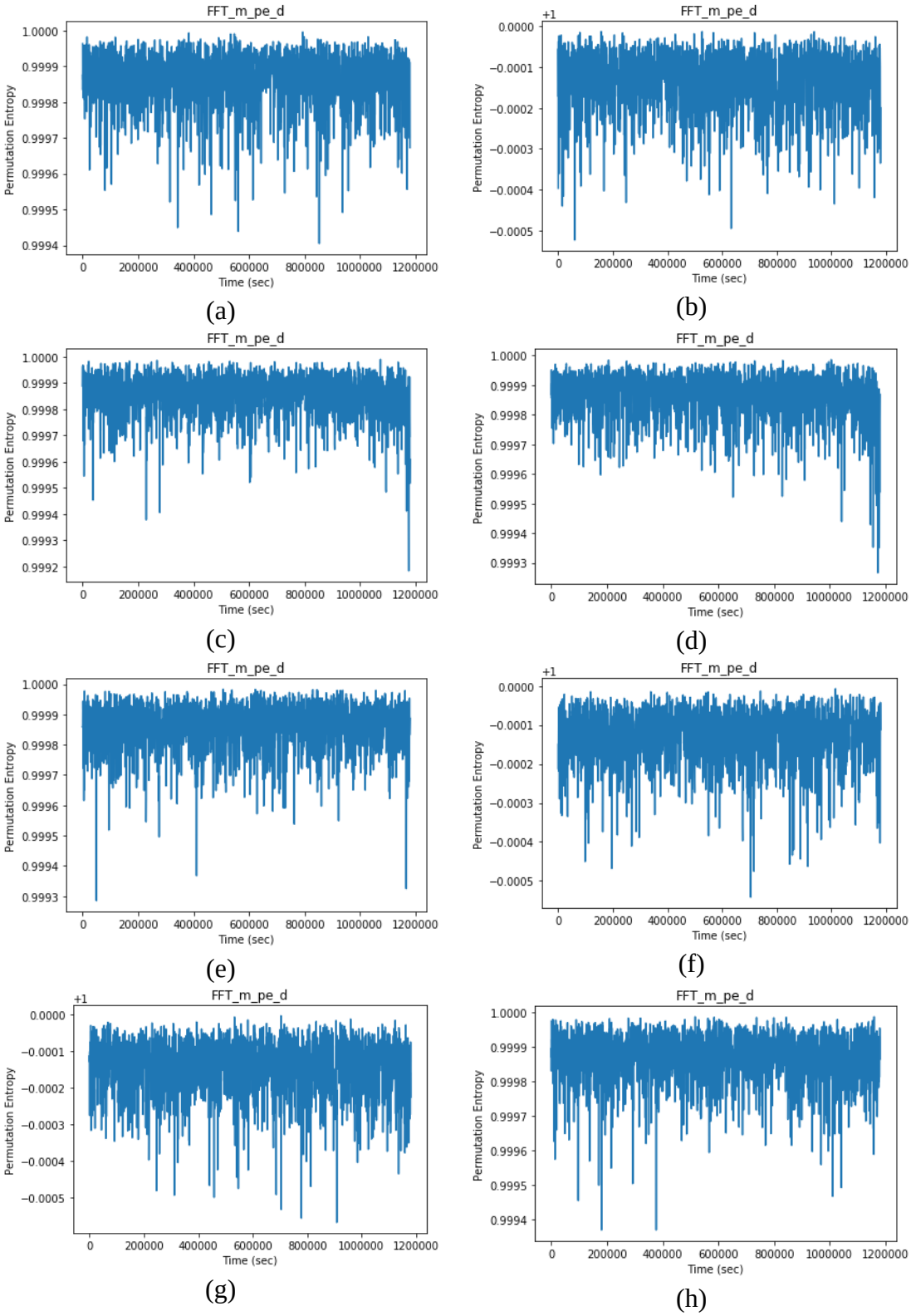


Figure 41 FFT magnitude PE level=4

a) Bearing 1 X-axis, b) Bearing 1 Y-axis, c) Bearing 2 X-axis, d) Bearing 2 Y-axis, e) Bearing 3 X-axis,f) Bearing 3 Y-axis, g) Bearing 4 X-axis and h) Bearing 4 Y-axis

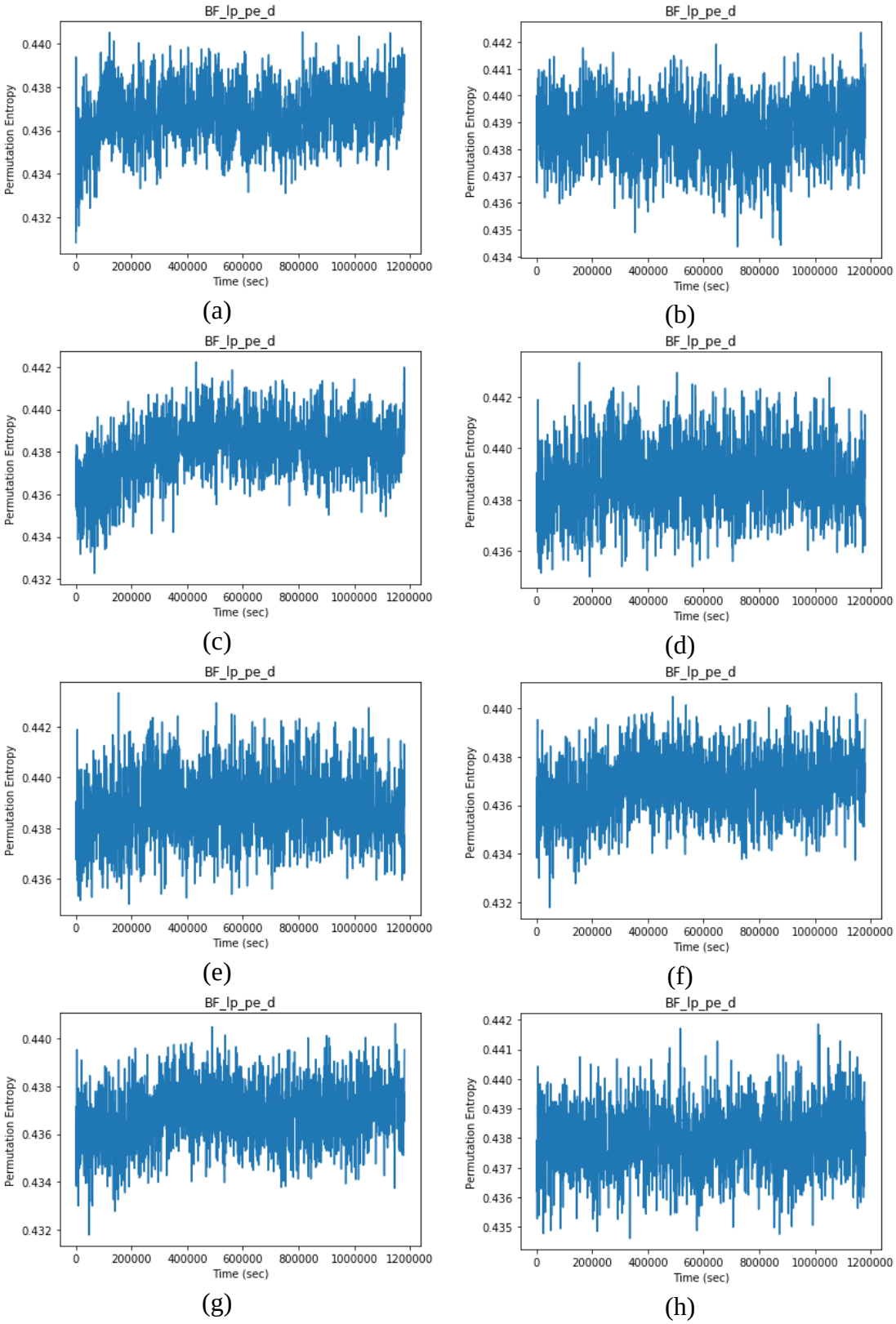


Figure 42 BF low-pass PE level=3

a) Bearing 1 X-axis, b) Bearing 1 Y-axis, c) Bearing 2 X-axis, d) Bearing 2 Y-axis, e) Bearing 3 X-axis,f) Bearing 3 Y-axis, g) Bearing 4 X-axis and h) Bearing 4 Y-axis

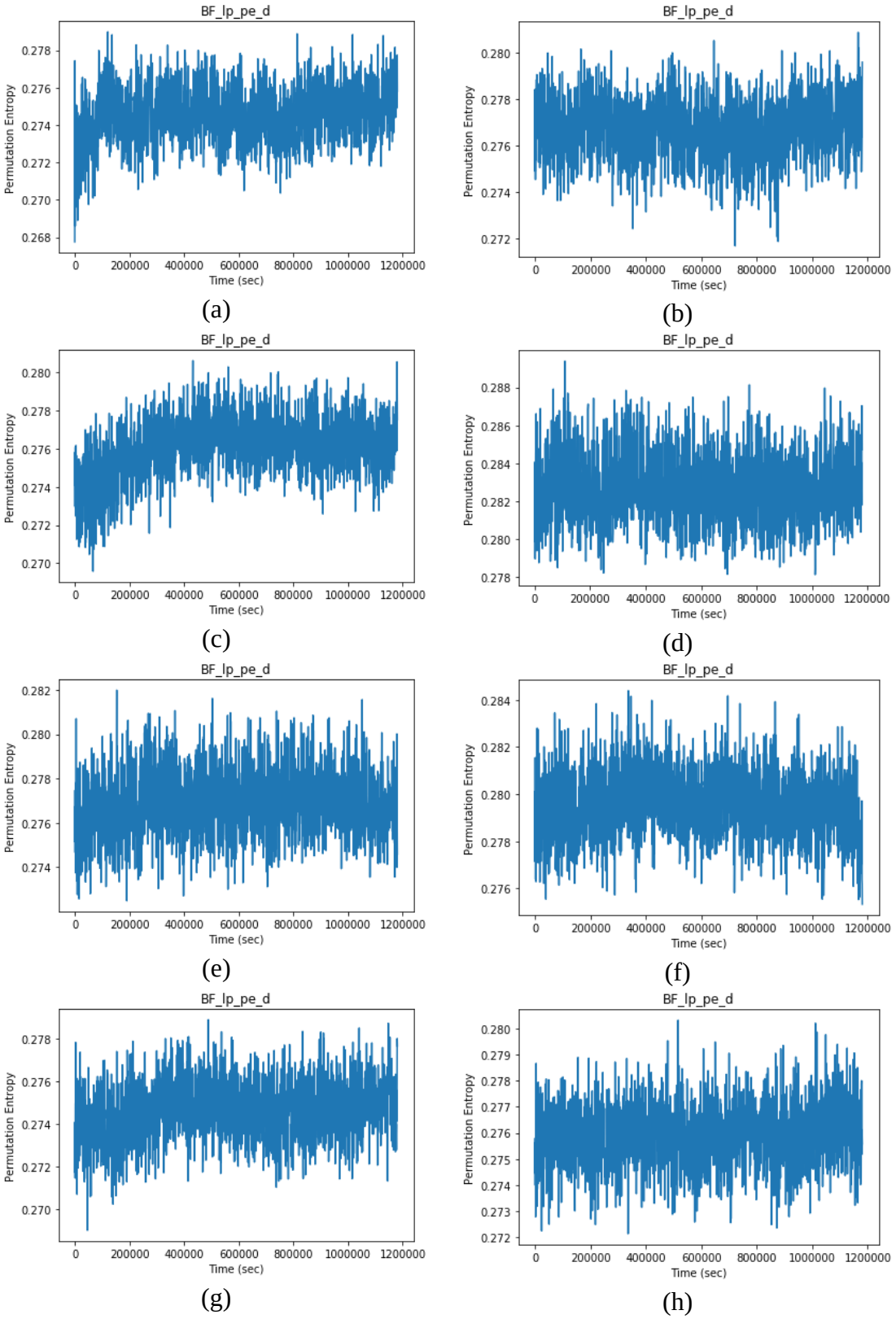


Figure 43 BF low-pass PE level=4

a) Bearing 1 X-axis, b) Bearing 1 Y-axis, c) Bearing 2 X-axis, d) Bearing 2 Y-axis, e) Bearing 3 X-axis,f) Bearing 3 Y-axis, g) Bearing 4 X-axis and h) Bearing 4 Y-axis

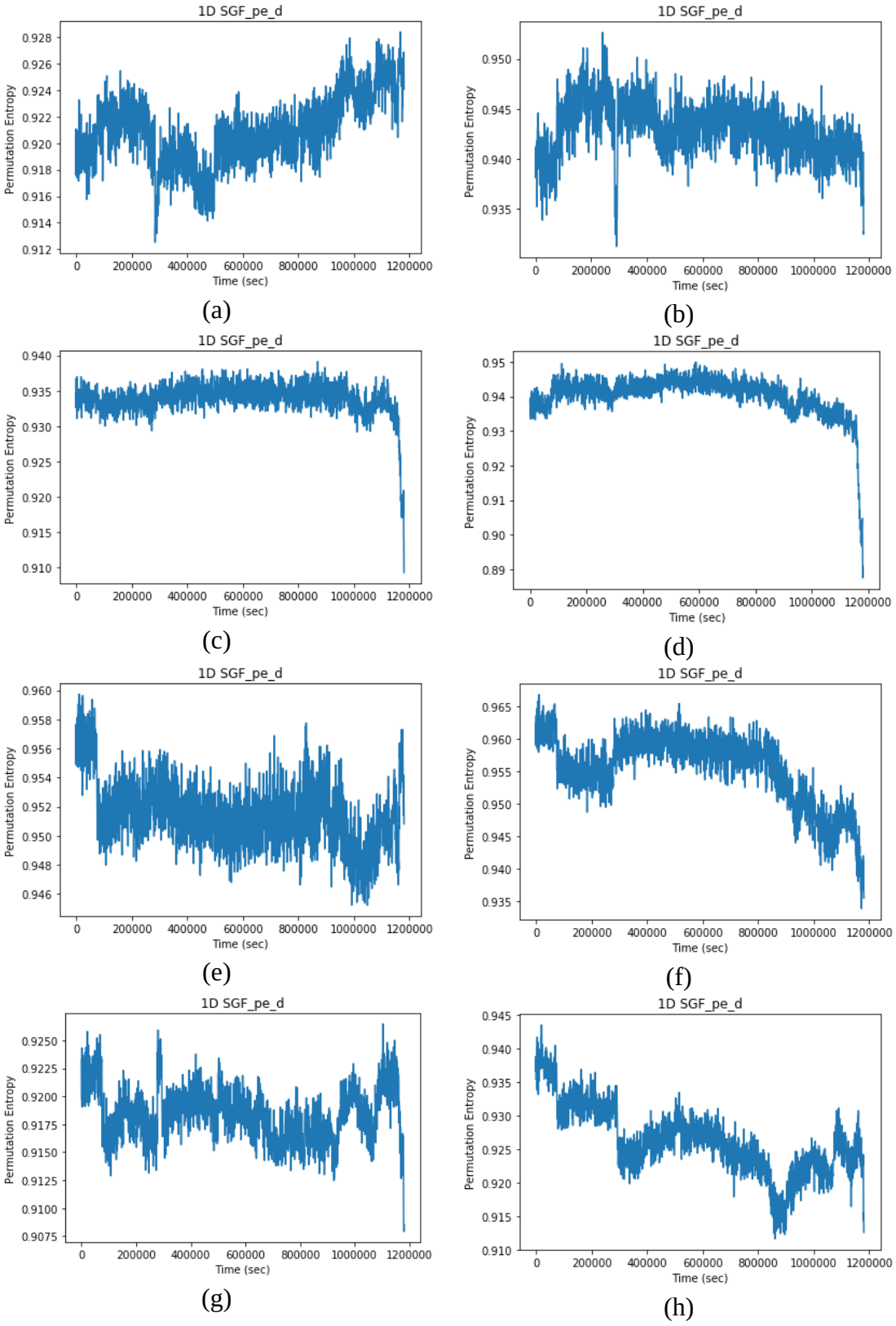


Figure 44 1D SGF PE level=3

a) Bearing 1 X-axis, b) Bearing 1 Y-axis, c) Bearing 2 X-axis, d) Bearing 2 Y-axis, e) Bearing 3 X-axis,f) Bearing 3 Y-axis, g) Bearing 4 X-axis and h) Bearing 4 Y-axis

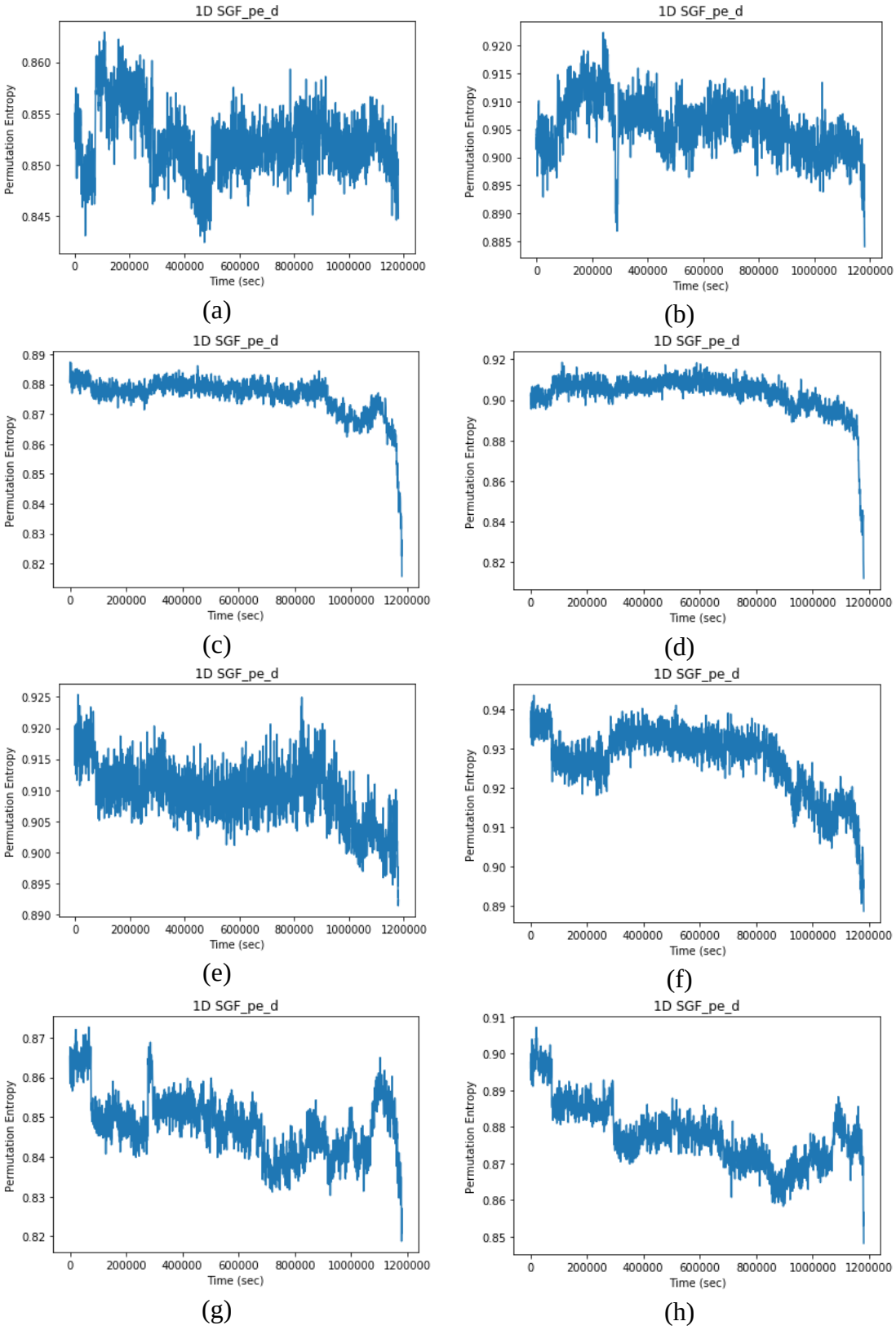


Figure 45 1D SGF PE level=4

a) Bearing 1 X-axis, b) Bearing 1 Y-axis, c) Bearing 2 X-axis, d) Bearing 2 Y-axis, e) Bearing 3 X-axis,f) Bearing 3 Y-axis, g) Bearing 4 X-axis and h) Bearing 4 Y-axis

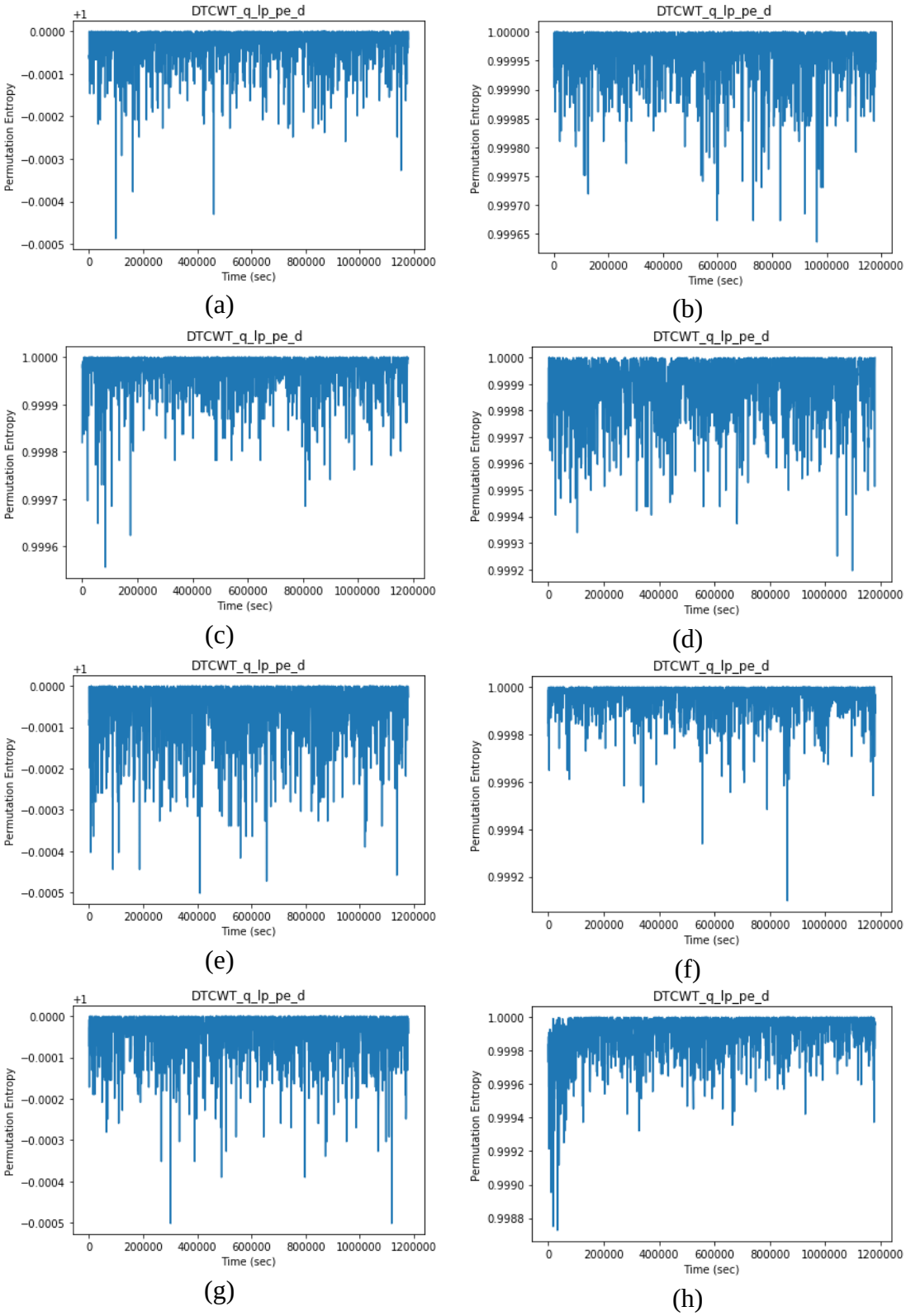


Figure 46 DTCWT qshift level=3 form=1

a) Bearing 1 X-axis, b) Bearing 1 Y-axis, c) Bearing 2 X-axis, d) Bearing 2 Y-axis, e) Bearing 3 X-axis,f) Bearing 3 Y-axis, g) Bearing 4 X-axis and h) Bearing 4 Y-axis

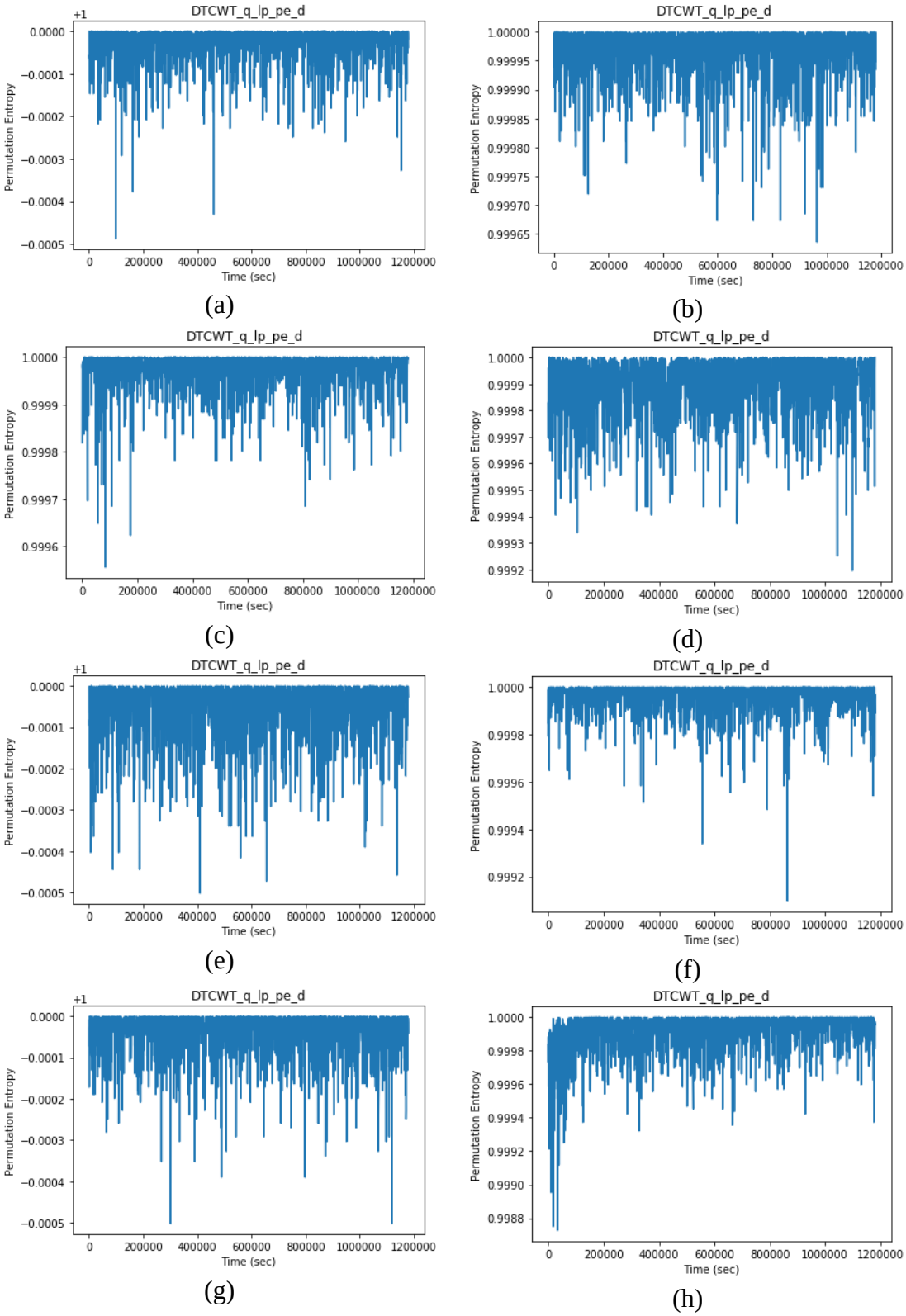


Figure 47 DTCWT qshift lp level=3 form=2

a) Bearing 1 X-axis, b) Bearing 1 Y-axis,c) Bearing 2 X-axis, d) Bearing 2 Y-axis, e) Bearing 3 X-axis,f) Bearing 3 Y-axis, g) Bearing 4 X-axis and h) Bearing 4 Y-axis

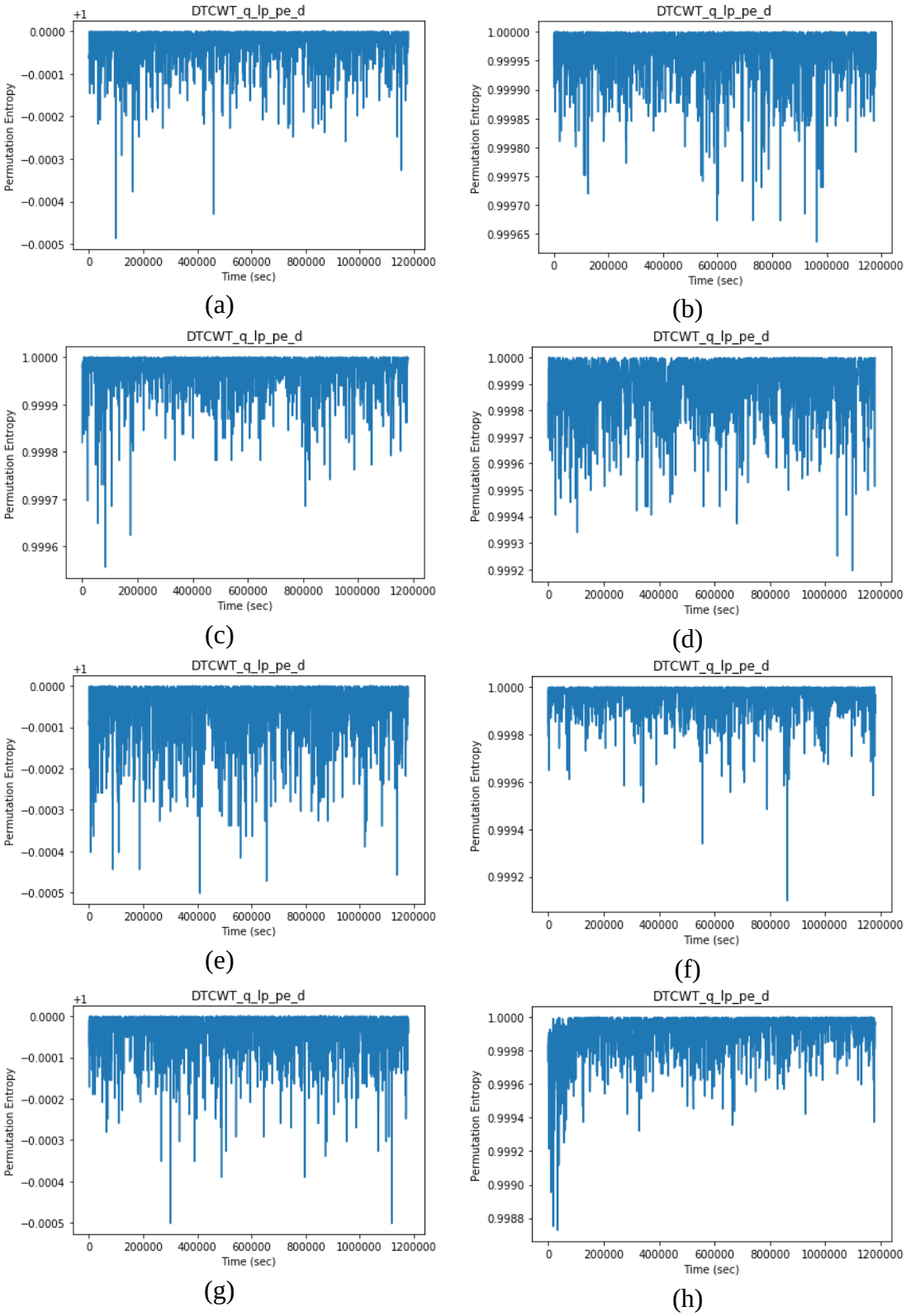


Figure 48 DTCWT qshift lp level=4 form=1

a) Bearing 1 X-axis, b) Bearing 1 Y-axis, c) Bearing 2 X-axis, d) Bearing 2 Y-axis, e) Bearing 3 X-axis,f) Bearing 3 Y-axis, g) Bearing 4 X-axis and h) Bearing 4 Y-axis

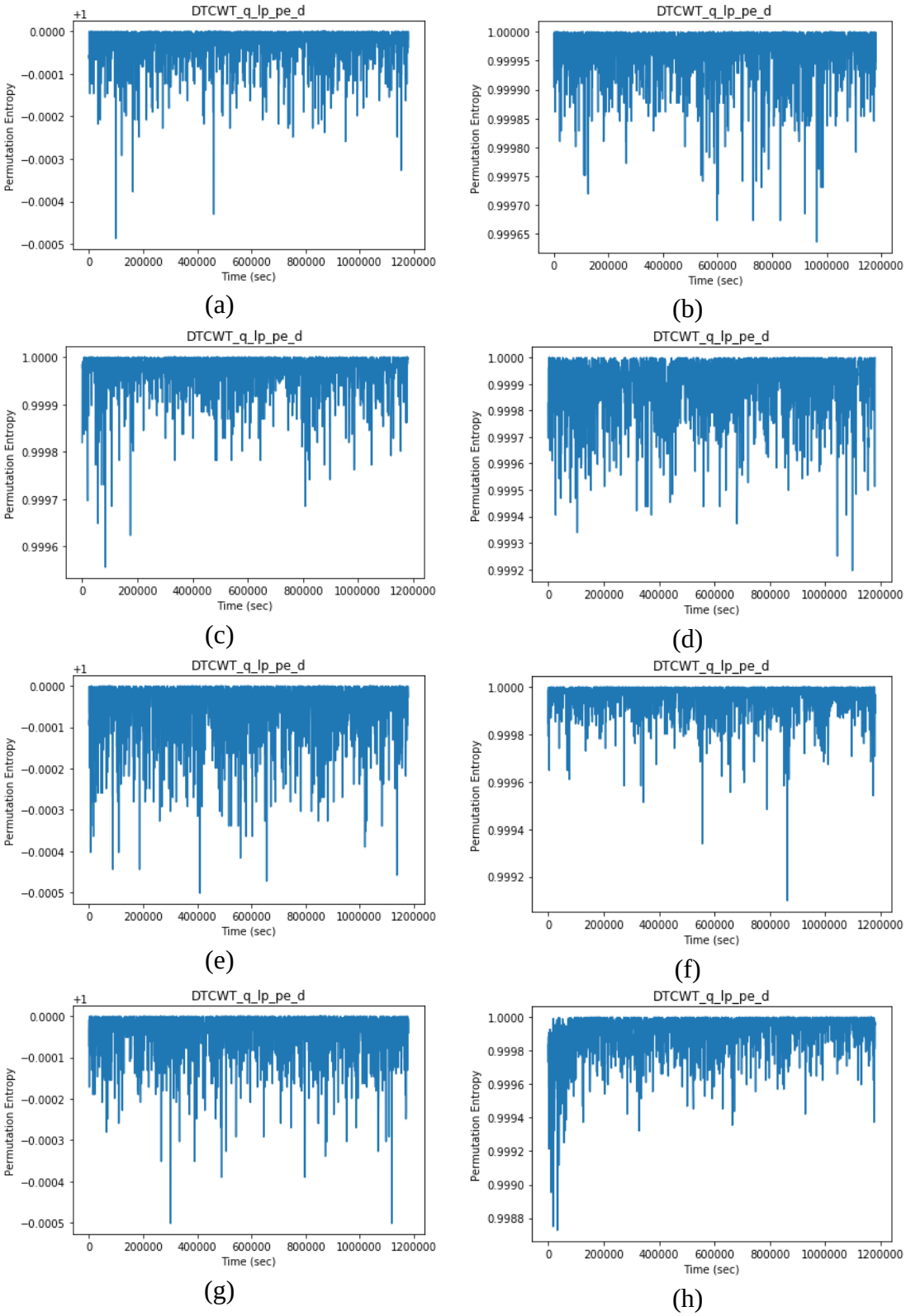


Figure 49 DTCWT qshift lp level=4 form=2

a) Bearing 1 X-axis, b) Bearing 1 Y-axis, c) Bearing 2 X-axis, d) Bearing 2 Y-axis, e) Bearing 3 X-axis,f) Bearing 3 Y-axis, g) Bearing 4 X-axis and h) Bearing 4 Y-axis

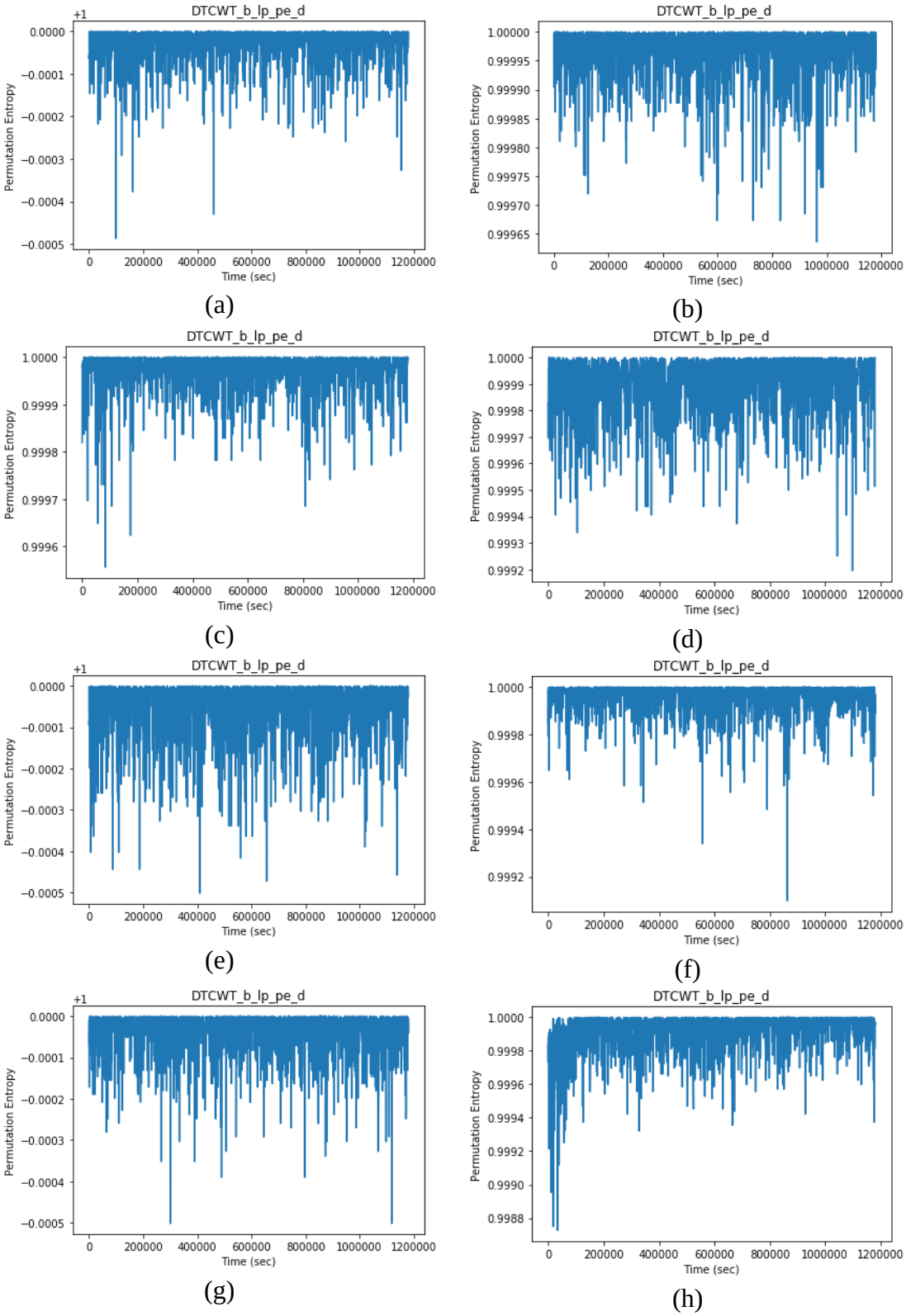


Figure 50 DTCWT biort lp level=3 form=1

a) Bearing 1 X-axis, b) Bearing 1 Y-axis, c) Bearing 2 X-axis, d) Bearing 2 Y-axis, e) Bearing 3 X-axis,f) Bearing 3 Y-axis, g) Bearing 4 X-axis and h) Bearing 4 Y-axis

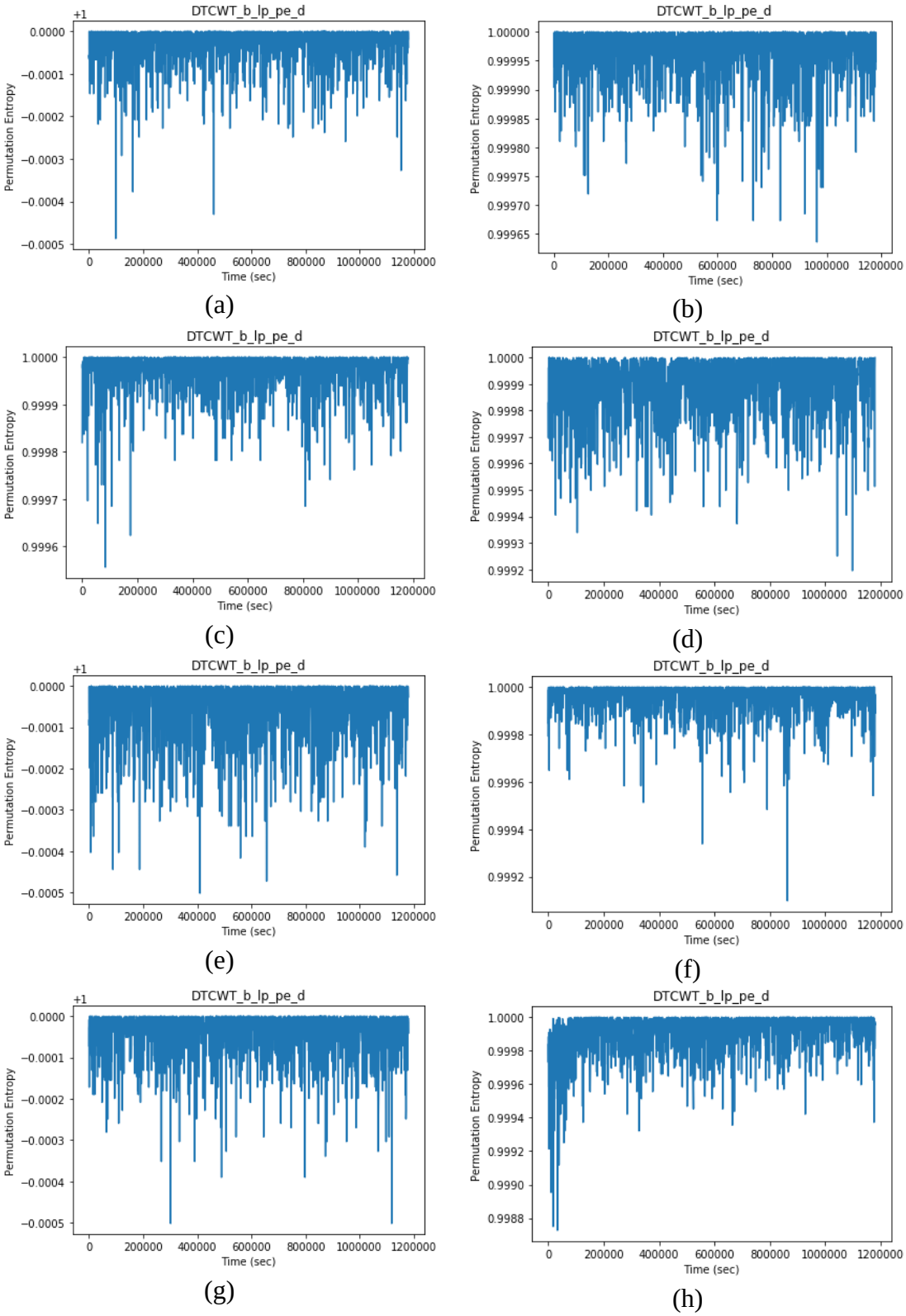


Figure 51 DTCWT biort lp level=3 form=2

a) Bearing 1 X-axis, b) Bearing 1 Y-axis, c) Bearing 2 X-axis, d) Bearing 2 Y-axis, e) Bearing 3 X-axis,f) Bearing 3 Y-axis, g) Bearing 4 X-axis and h) Bearing 4 Y-axis

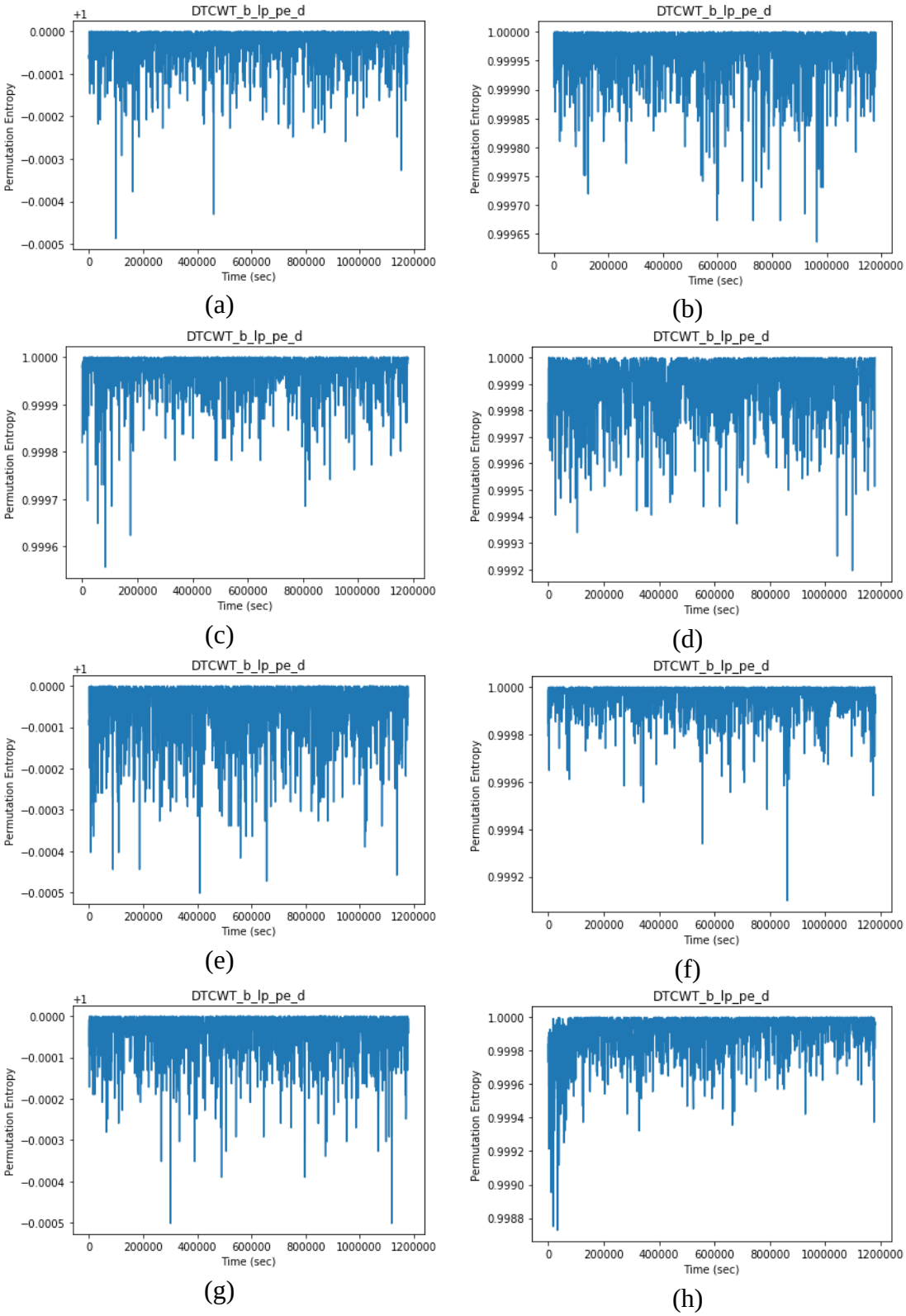


Figure 52 DTCWT biort lp level=4 form=1

a) Bearing 1 X-axis, b) Bearing 1 Y-axis,c) Bearing 2 X-axis, d) Bearing 2 Y-axis, e) Bearing 3 X-axis,f) Bearing 3 Y-axis, g) Bearing 4 X-axis and h) Bearing 4 Y-axis

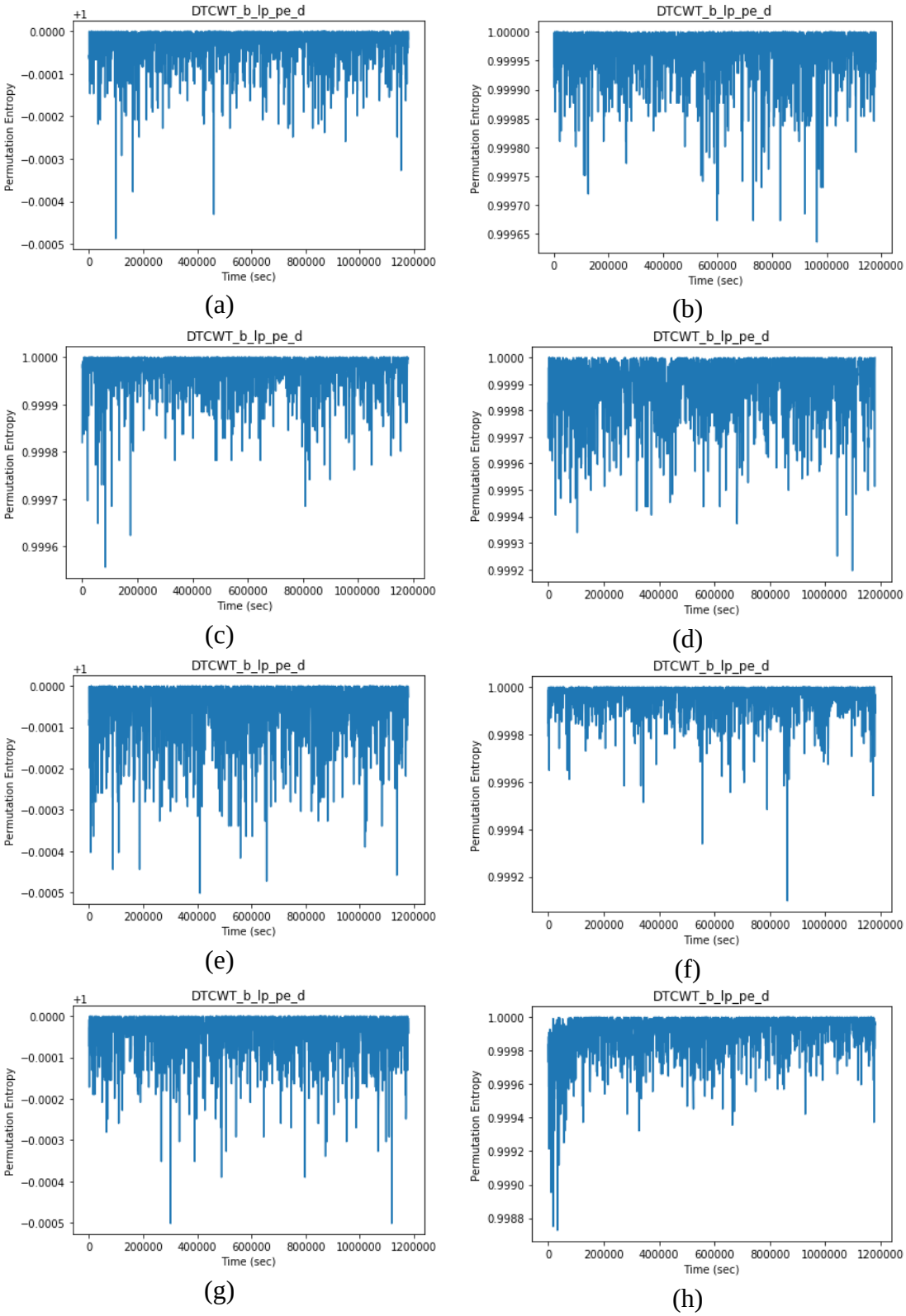


Figure 53 DTCWT biort lp level=4 form=2

a) Bearing 1 X-axis, b) Bearing 1 Y-axis,c) Bearing 2 X-axis, d) Bearing 2 Y-axis, e) Bearing 3 X-axis,f) Bearing 3 Y-axis, g) Bearing 4 X-axis and h) Bearing 4 Y-axis

Table 19
Raw Data - Descriptive Statistics I

	Bearings							
	1,X-axis	1,Y-axis	2,X-axis	2,Y-axis	3,X-axis	3,Y-axis	4,X-axis	4,Y-axis
count	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000
mean	0.967434	0.995503	0.984829	0.998322	0.994962	0.999649	0.969970	0.990361
std	0.002751	0.001555	0.004066	0.003009	0.001426	0.000750	0.004831	0.002995
min	0.957453	0.987393	0.942908	0.951730	0.987342	0.991839	0.944560	0.974247
25%	0.965740	0.994569	0.984115	0.998287	0.994493	0.999728	0.966558	0.988247
50%	0.966989	0.995460	0.985693	0.998921	0.995268	0.999877	0.970294	0.990592
75%	0.968521	0.996409	0.986974	0.999269	0.995858	0.999934	0.972871	0.992332
max	0.977849	0.999378	0.990480	0.999955	0.998314	0.999996	0.983688	0.997618

Note. Level=3; Raw_data_pe_d; Data is unstandardised.

Table 20
Raw Data - Descriptive Statistics II

	Bearings							
	1,X-axis	1,Y-axis	2,X-axis	2,Y-axis	3,X-axis	3,Y-axis	4,X-axis	4,Y-axis
count	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000
mean	0.936643	0.990464	0.970520	0.995530	0.987542	0.997790	0.947513	0.982435
std	0.005945	0.002492	0.007771	0.004925	0.003534	0.001358	0.008725	0.005154
min	0.910441	0.976529	0.894392	0.921820	0.966779	0.984645	0.901237	0.955909
25%	0.933087	0.989039	0.969154	0.995459	0.986516	0.997746	0.941448	0.978697
50%	0.936360	0.990353	0.972304	0.996628	0.988258	0.998179	0.948319	0.982730
75%	0.939344	0.991717	0.974795	0.997218	0.989627	0.998415	0.952828	0.985803
max	0.957419	0.996552	0.981227	0.998210	0.996109	0.999235	0.970458	0.994626

Note. Level=4; Raw_data_pe_d; Data is unstandardised.

Table 21
BF - Descriptive Statistics I

	Bearings							
	1,X-axis	1,Y-axis	2,X-axis	2,Y-axis	3,X-axis	3,Y-axis	4,X-axis	4,Y-axis
count	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000
mean	0.436758	0.438652	0.438111	0.443913	0.438793	0.441148	0.436772	0.437953
std	0.001283	0.001145	0.001401	0.001533	0.001289	0.001263	0.001211	0.001033
min	0.430824	0.434355	0.432272	0.439994	0.434995	0.437512	0.431784	0.434616
25%	0.435935	0.437887	0.437276	0.442853	0.437930	0.440272	0.435948	0.437274
50%	0.436792	0.438675	0.438237	0.443779	0.438743	0.441179	0.436762	0.437939
75%	0.437627	0.439465	0.439080	0.444897	0.439628	0.441985	0.437606	0.438639
max	0.440537	0.442356	0.442230	0.449971	0.443343	0.445319	0.440613	0.441854

Note. Level=3. BF_lp_pe_d; Data is unstandardised.

Table 22
BF - Descriptive Statistics II

	Bearings							
	1,X-axis	1,Y-axis	2,X-axis	2,Y-axis	3,X-axis	3,Y-axis	4,X-axis	4,Y-axis
count	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000
mean	0.274580	0.276693	0.276101	0.282631	0.276870	0.279504	0.274605	0.275904
std	0.001452	0.001298	0.001580	0.001745	0.001464	0.001434	0.001367	0.001170
min	0.267752	0.271688	0.269579	0.278118	0.272486	0.275330	0.269025	0.272144
25%	0.273641	0.275803	0.275131	0.281433	0.275884	0.278510	0.273692	0.275145
50%	0.274623	0.276697	0.276237	0.282509	0.276846	0.279484	0.274597	0.275917
75%	0.275534	0.277614	0.277191	0.283723	0.277831	0.280473	0.275535	0.276689
max	0.278981	0.280856	0.280620	0.289388	0.281983	0.284394	0.278895	0.280313

Note. Level=4. BF_lp_pe_d; Data is unstandardised.

Table 23
HT - Descriptive Statistics I

	Bearings							
	1,X-axis	1,Y-axis	2,X-axis	2,Y-axis	3,X-axis	3,Y-axis	4,X-axis	4,Y-axis
count	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000
mean	0.967432	0.995333	0.983631	0.998137	0.992623	0.998775	0.971788	0.989950
std	0.002855	0.001516	0.004292	0.002815	0.002533	0.001775	0.004609	0.003259
min	0.953544	0.987706	0.940578	0.954628	0.963766	0.978780	0.945695	0.973372
25%	0.965811	0.994461	0.982697	0.998064	0.992377	0.998787	0.968646	0.987633
50%	0.967133	0.995327	0.984612	0.998683	0.993204	0.999284	0.972113	0.990398
75%	0.968658	0.996215	0.985866	0.999112	0.993855	0.999519	0.974527	0.992200
max	0.977654	0.999108	0.989299	0.999926	0.996718	0.999967	0.984741	0.997277

Note. Level=3, HT_m_pe_d; Data is unstandardised.

Table 24
HT - Descriptive Statistics II

	Bearings							
	1,X-axis	1,Y-axis	2,X-axis	2,Y-axis	3,X-axis	3,Y-axis	4,X-axis	4,Y-axis
count	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000
mean	0.937634	0.989319	0.967040	0.993982	0.981218	0.994649	0.949680	0.980093
std	0.005478	0.002532	0.007405	0.004569	0.004663	0.003127	0.008167	0.005594
min	0.911061	0.975953	0.896515	0.926496	0.935159	0.959738	0.908982	0.953202
25%	0.934340	0.987925	0.965339	0.993738	0.980475	0.994652	0.943974	0.976019
50%	0.937564	0.989257	0.968854	0.995020	0.982156	0.995528	0.950382	0.980835
75%	0.940293	0.990703	0.971106	0.995677	0.983637	0.995986	0.954606	0.983945
max	0.955184	0.995394	0.978032	0.996930	0.990972	0.997015	0.971868	0.992763

Note. Level=4, HT_m_pe_d; Data is unstandardised.

Table 25
FFT - Descriptive Statistics I

	Bearings							
	1,X-axis	1,Y-axis	2,X-axis	2,Y-axis	3,X-axis	3,Y-axis	4,X-axis	4,Y-axis
count	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000
mean	0.999977	0.999978	0.999973	0.999975	0.999976	0.999979	0.999974	0.999977
std	0.000033	0.000031	0.000039	0.000037	0.000036	0.000030	0.000036	0.000032
min	0.999718	0.999687	0.999669	0.999634	0.999544	0.999767	0.999640	0.999676
25%	0.999970	0.999970	0.999965	0.999966	0.999969	0.999971	0.999965	0.999968
50%	0.999990	0.999989	0.999987	0.999989	0.999989	0.999990	0.999987	0.999989
75%	0.999998	0.999998	0.999997	0.999998	0.999998	0.999998	0.999997	0.999998
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

Note. Level=3; FFT_m_pe_d; Data is unstandardised.

Table 26
FFT - Descriptive Statistics II

	Bearings							
	1,X-axis	1,Y-axis	2,X-axis	2,Y-axis	3,X-axis	3,Y-axis	4,X-axis	4,Y-axis
count	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000
mean	0.999862	0.999867	0.999856	0.999861	0.999864	0.999871	0.999858	0.999867
std	0.000076	0.000071	0.000082	0.000078	0.000077	0.000069	0.000075	0.000073
min	0.999405	0.999476	0.999184	0.999266	0.999285	0.999458	0.999432	0.999368
25%	0.999825	0.999831	0.999815	0.999825	0.999827	0.999835	0.999819	0.999832
50%	0.999876	0.999880	0.999873	0.999877	0.999879	0.999883	0.999871	0.999879
75%	0.999915	0.999918	0.999913	0.999915	0.999919	0.999922	0.999911	0.999918
max	0.999996	0.999988	0.999990	0.999986	0.999983	0.999993	0.999998	0.999988

Note. Level=4; FFT_m_pe_d; Data is unstandardised.

Table 27
1D SGF - Descriptive Statistics I

	Bearings							
	1,X-axis	1,Y-axis	2,X-axis	2,Y-axis	3,X-axis	3,Y-axis	4,X-axis	4,Y-axis
count	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000
mean	0.920934	0.942923	0.933920	0.940100	0.951502	0.955373	0.918461	0.926072
std	0.002508	0.002729	0.002440	0.006118	0.002350	0.005332	0.002354	0.005465
min	0.912522	0.931186	0.909248	0.887480	0.945226	0.933913	0.907914	0.911574
25%	0.919272	0.941164	0.933069	0.938187	0.949917	0.952599	0.916747	0.922617
50%	0.920797	0.942975	0.934177	0.941337	0.951267	0.956782	0.918413	0.925725
75%	0.922609	0.944765	0.935226	0.943512	0.952746	0.959130	0.919961	0.929444
max	0.928379	0.952644	0.939155	0.949864	0.959731	0.966733	0.926467	0.943471

Note. Level=3; *1D SGF_pe_d*; Data is unstandardised.

Table 28
1D SGF - Descriptive Statistics II

	Bearings							
	1,X-axis	1,Y-axis	2,X-axis	2,Y-axis	3,X-axis	3,Y-axis	4,X-axis	4,Y-axis
count	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000
mean	0.852322	0.905518	0.876027	0.902890	0.909566	0.927016	0.848054	0.877671
std	0.003203	0.004711	0.006716	0.009923	0.004807	0.008247	0.007650	0.008777
min	0.842393	0.884113	0.815762	0.812126	0.891397	0.888558	0.818797	0.848148
25%	0.850229	0.902461	0.875065	0.900262	0.906756	0.923097	0.842841	0.871489
50%	0.852188	0.905385	0.877755	0.905038	0.909686	0.929305	0.847903	0.876936
75%	0.854092	0.908589	0.879632	0.908219	0.912602	0.932627	0.852311	0.882897
max	0.862953	0.922248	0.887399	0.918511	0.925382	0.943504	0.872613	0.907172

Note. Level=4; *1D SGF_pe_d*; Data is unstandardised.

Table 29
DTCWT - Descriptive Statistics I

	Bearings							
	1,X-axis	1,Y-axis	2,X-axis	2,Y-axis	3,X-axis	3,Y-axis	4,X-axis	4,Y-axis
count	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000
mean	0.999968	0.999967	0.999970	0.999896	0.999946	0.999955	0.999961	0.999903
std	0.000044	0.000046	0.000043	0.000109	0.000072	0.000066	0.000055	0.000129
min	0.999513	0.999636	0.999556	0.999195	0.999498	0.999098	0.999498	0.998727
25%	0.999958	0.999958	0.999962	0.999845	0.999928	0.999939	0.999949	0.999869
50%	0.999985	0.999983	0.999985	0.999928	0.999974	0.999980	0.999983	0.999954
75%	0.999997	0.999997	0.999997	0.999977	0.999994	0.999995	0.999997	0.999990
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

Note. Level=3, form=1; DTCWT_b_lp_pe_d; Data is unstandardised.

Table 30
DTCWT - Descriptive Statistics II

	Bearings							
	1,X-axis	1,Y-axis	2,X-axis	2,Y-axis	3,X-axis	3,Y-axis	4,X-axis	4,Y-axis
count	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000
mean	0.999968	0.999967	0.999970	0.999896	0.999946	0.999955	0.999961	0.999903
std	0.000044	0.000046	0.000043	0.000109	0.000072	0.000066	0.000055	0.000129
min	0.999513	0.999636	0.999556	0.999195	0.999498	0.999098	0.999498	0.998727
25%	0.999958	0.999958	0.999962	0.999845	0.999928	0.999939	0.999949	0.999869
50%	0.999985	0.999983	0.999985	0.999928	0.999974	0.999980	0.999983	0.999954
75%	0.999997	0.999997	0.999997	0.999977	0.999994	0.999995	0.999997	0.999990
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

Note. Level=3, Form=2; DTCWT_b_lp_pe_d; Data is unstandardised.

Table 31
DTCWT - Descriptive Statistics III

	Bearings							
	1,X-axis	1,Y-axis	2,X-axis	2,Y-axis	3,X-axis	3,Y-axis	4,X-axis	4,Y-axis
count	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000
mean	0.999968	0.999967	0.999970	0.999896	0.999946	0.999955	0.999961	0.999903
std	0.000044	0.000046	0.000043	0.000109	0.000072	0.000066	0.000055	0.000129
min	0.999513	0.999636	0.999556	0.999195	0.999498	0.999098	0.999498	0.998727
25%	0.999958	0.999958	0.999962	0.999845	0.999928	0.999939	0.999949	0.999869
50%	0.999985	0.999983	0.999985	0.999928	0.999974	0.999980	0.999983	0.999954
75%	0.999997	0.999997	0.999997	0.999977	0.999994	0.999995	0.999997	0.999990
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

Note. Level=4; Form=1; DTCWT_b_lp_pe_d; Data is unstandardised.

Table 32
DTCWT - Descriptive Statistics IV

	Bearings							
	1,X-axis	1,Y-axis	2,X-axis	2,Y-axis	3,X-axis	3,Y-axis	4,X-axis	4,Y-axis
count	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000
mean	0.999968	0.999967	0.999970	0.999896	0.999946	0.999955	0.999961	0.999903
std	0.000044	0.000046	0.000043	0.000109	0.000072	0.000066	0.000055	0.000129
min	0.999513	0.999636	0.999556	0.999195	0.999498	0.999098	0.999498	0.998727
25%	0.999958	0.999958	0.999962	0.999845	0.999928	0.999939	0.999949	0.999869
50%	0.999985	0.999983	0.999985	0.999928	0.999974	0.999980	0.999983	0.999954
75%	0.999997	0.999997	0.999997	0.999977	0.999994	0.999995	0.999997	0.999990
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

Note. Level=4; Form=2; DTCWT_b_lp_pe_d; Data is unstandardised.

Table 33
DTCWT - Descriptive Statistics V

	Bearings							
	1,X-axis	1,Y-axis	2,X-axis	2,Y-axis	3,X-axis	3,Y-axis	4,X-axis	4,Y-axis
count	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000
mean	0.999968	0.999967	0.999970	0.999896	0.999946	0.999955	0.999961	0.999903
std	0.000044	0.000046	0.000043	0.000109	0.000072	0.000066	0.000055	0.000129
min	0.999513	0.999636	0.999556	0.999195	0.999498	0.999098	0.999498	0.998727
25%	0.999958	0.999958	0.999962	0.999845	0.999928	0.999939	0.999949	0.999869
50%	0.999985	0.999983	0.999985	0.999928	0.999974	0.999980	0.999983	0.999954
75%	0.999997	0.999997	0.999997	0.999977	0.999994	0.999995	0.999997	0.999990
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

Note. Level=3; Form=1; DTCWT_q_lp_pe_d; Data is unstandardised.

Table 34
DTCWT - Descriptive Statistics VI

	Bearings							
	1,X-axis	1,Y-axis	2,X-axis	2,Y-axis	3,X-axis	3,Y-axis	4,X-axis	4,Y-axis
count	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000
mean	0.999968	0.999967	0.999970	0.999896	0.999946	0.999955	0.999961	0.999903
std	0.000044	0.000046	0.000043	0.000109	0.000072	0.000066	0.000055	0.000129
min	0.999513	0.999636	0.999556	0.999195	0.999498	0.999098	0.999498	0.998727
25%	0.999958	0.999958	0.999962	0.999845	0.999928	0.999939	0.999949	0.999869
50%	0.999985	0.999983	0.999985	0.999928	0.999974	0.999980	0.999983	0.999954
75%	0.999997	0.999997	0.999997	0.999977	0.999994	0.999995	0.999997	0.999990
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

Note. Level=3; Form=2; DTCWT_q_lp_pe_d; Data is unstandardised.

Table 35
DTCWT - Descriptive Statistics VII

	Bearings							
	1,X-axis	1,Y-axis	2,X-axis	2,Y-axis	3,X-axis	3,Y-axis	4,X-axis	4,Y-axis
count	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000
mean	0.999968	0.999967	0.999970	0.999896	0.999946	0.999955	0.999961	0.999903
std	0.000044	0.000046	0.000043	0.000109	0.000072	0.000066	0.000055	0.000129
min	0.999513	0.999636	0.999556	0.999195	0.999498	0.999098	0.999498	0.998727
25%	0.999958	0.999958	0.999962	0.999845	0.999928	0.999939	0.999949	0.999869
50%	0.999985	0.999983	0.999985	0.999928	0.999974	0.999980	0.999983	0.999954
75%	0.999997	0.999997	0.999997	0.999977	0.999994	0.999995	0.999997	0.999990
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

Note. Level=4; Form=1; DTCWT_q_lp_pe_d; Data is unstandardised.

Table 36
DTCWT - Descriptive Statistics VIII

	Bearings							
	1,X-axis	1,Y-axis	2,X-axis	2,Y-axis	3,X-axis	3,Y-axis	4,X-axis	4,Y-axis
count	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000
mean	0.999968	0.999967	0.999970	0.999896	0.999946	0.999955	0.999961	0.999903
std	0.000044	0.000046	0.000043	0.000109	0.000072	0.000066	0.000055	0.000129
min	0.999513	0.999636	0.999556	0.999195	0.999498	0.999098	0.999498	0.998727
25%	0.999958	0.999958	0.999962	0.999845	0.999928	0.999939	0.999949	0.999869
50%	0.999985	0.999983	0.999985	0.999928	0.999974	0.999980	0.999983	0.999954
75%	0.999997	0.999997	0.999997	0.999977	0.999994	0.999995	0.999997	0.999990
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

Note. Level=4; Form=2; DTCWT_q_lp_pe_d; Data is unstandardised.

Table 37
DTCWT - Descriptive Statistics IX

	Bearings							
	1,X-axis	1,Y-axis	2,X-axis	2,Y-axis	3,X-axis	3,Y-axis	4,X-axis	4,Y-axis
count	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000
mean	0.999967	0.999973	0.999974	0.999974	0.999972	0.999976	0.999969	0.999973
std	0.000046	0.000038	0.000036	0.000036	0.000043	0.000034	0.000045	0.000037
min	0.999553	0.999626	0.999705	0.999587	0.999360	0.999664	0.999517	0.999645
25%	0.999955	0.999963	0.999965	0.999965	0.999965	0.999967	0.999961	0.999965
50%	0.999983	0.999987	0.999988	0.999987	0.999988	0.999989	0.999986	0.999987
75%	0.999996	0.999997	0.999997	0.999998	0.999997	0.999998	0.999996	0.999998
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

Note. Level=3; Form=1; DTCWT_b_hp_pe_d; Data is unstandardised.

Table 38
DTCWT - Descriptive Statistics X

	Bearings							
	1,X-axis	1,Y-axis	2,X-axis	2,Y-axis	3,X-axis	3,Y-axis	4,X-axis	4,Y-axis
count	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000
mean	0.999967	0.999973	0.999974	0.999974	0.999972	0.999976	0.999969	0.999973
std	0.000046	0.000038	0.000036	0.000036	0.000043	0.000034	0.000045	0.000037
min	0.999553	0.999626	0.999705	0.999587	0.999360	0.999664	0.999517	0.999645
25%	0.999955	0.999963	0.999965	0.999965	0.999965	0.999967	0.999961	0.999965
50%	0.999983	0.999987	0.999988	0.999987	0.999988	0.999989	0.999986	0.999987
75%	0.999996	0.999997	0.999997	0.999998	0.999997	0.999998	0.999996	0.999998
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

Note. Level=3; Form=2; DTCWT_b_hp_pe_d; Data is unstandardised.

Table 39
DTCWT - Descriptive Statistics XI

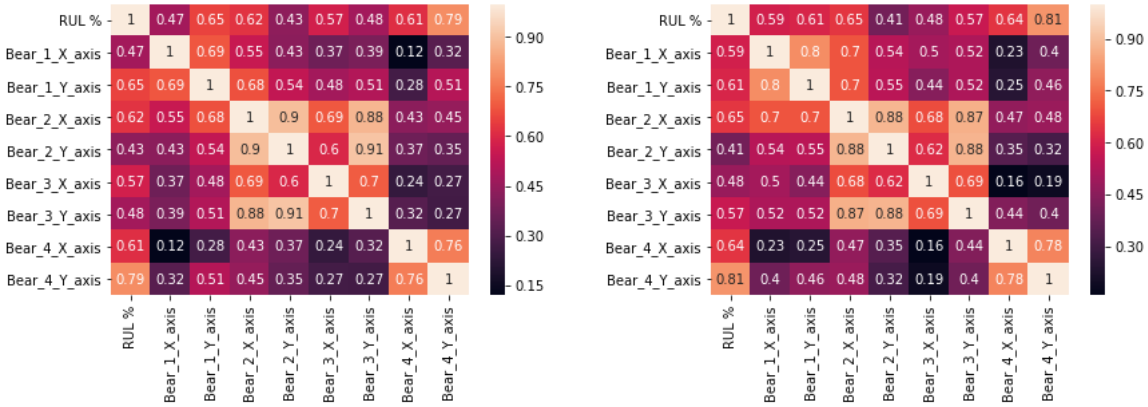
	Bearings							
	1,X-axis	1,Y-axis	2,X-axis	2,Y-axis	3,X-axis	3,Y-axis	4,X-axis	4,Y-axis
count	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000
mean	0.999967	0.999973	0.999974	0.999974	0.999972	0.999976	0.999969	0.999973
std	0.000046	0.000038	0.000036	0.000036	0.000043	0.000034	0.000045	0.000037
min	0.999553	0.999626	0.999705	0.999587	0.999360	0.999664	0.999517	0.999645
25%	0.999955	0.999963	0.999965	0.999965	0.999965	0.999967	0.999961	0.999965
50%	0.999983	0.999987	0.999988	0.999987	0.999988	0.999989	0.999986	0.999987
75%	0.999996	0.999997	0.999997	0.999998	0.999997	0.999998	0.999996	0.999998
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

Note. Level=4; Form=1; DTCWT_b_hp_pe_d; Data is unstandardised.

Table 40
DTCWT - Descriptive Statistics XII

	Bearings							
	1,X-axis	1,Y-axis	2,X-axis	2,Y-axis	3,X-axis	3,Y-axis	4,X-axis	4,Y-axis
count	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000	2156.0000
mean	0.999967	0.999973	0.999974	0.999974	0.999972	0.999976	0.999969	0.999973
std	0.000046	0.000038	0.000036	0.000036	0.000043	0.000034	0.000045	0.000037
min	0.999553	0.999626	0.999705	0.999587	0.999360	0.999664	0.999517	0.999645
25%	0.999955	0.999963	0.999965	0.999965	0.999965	0.999967	0.999961	0.999965
50%	0.999983	0.999987	0.999988	0.999987	0.999988	0.999989	0.999986	0.999987
75%	0.999996	0.999997	0.999997	0.999998	0.999997	0.999998	0.999996	0.999998
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

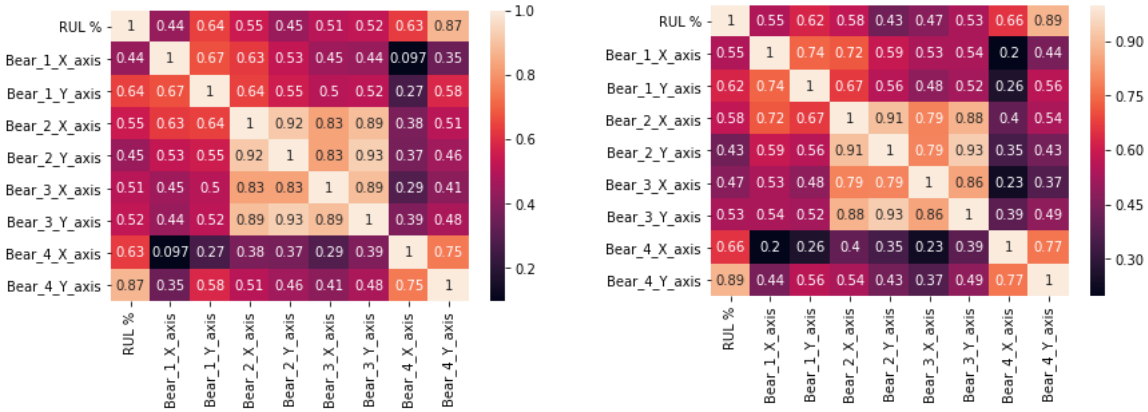
Note. Level=4; Form=2; DTCWT_b_hp_pe_d; Data is unstandardised.



(a) (b)

Figure 54 Pearson Correlation Heatmap I:

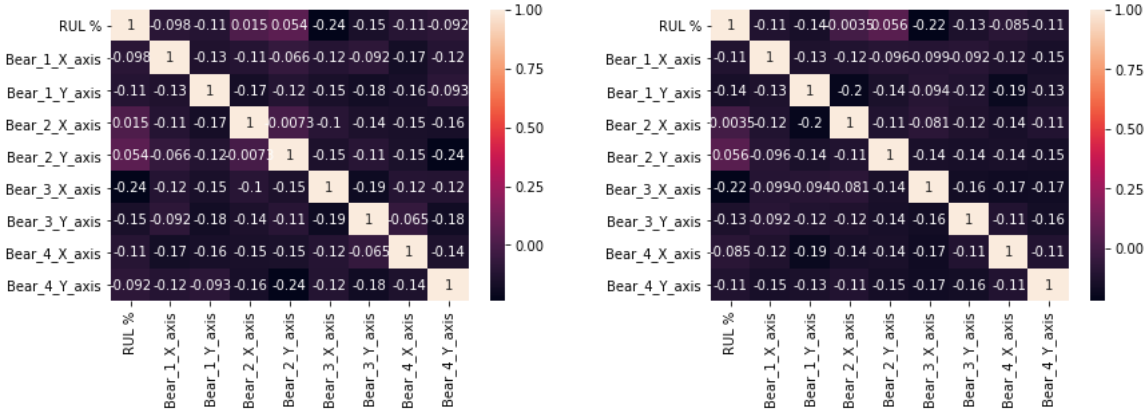
Raw Data after PE and log distribution a) level= 3 and b) level=4



(a) (b)

Figure 55 Pearson Correlation Heatmap II:

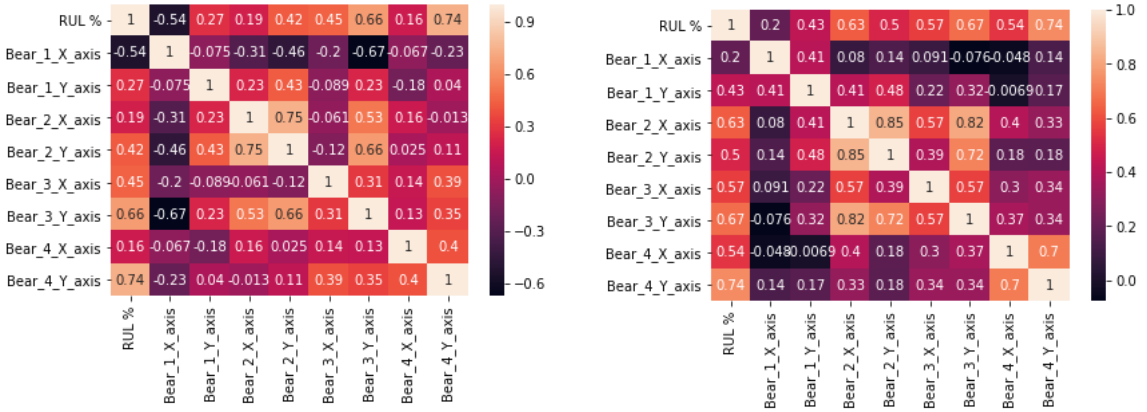
Data after HT, PE and log distribution a) level=3 and b) level=4



(a) (b)

Figure 56 Correlation Heatmap III:

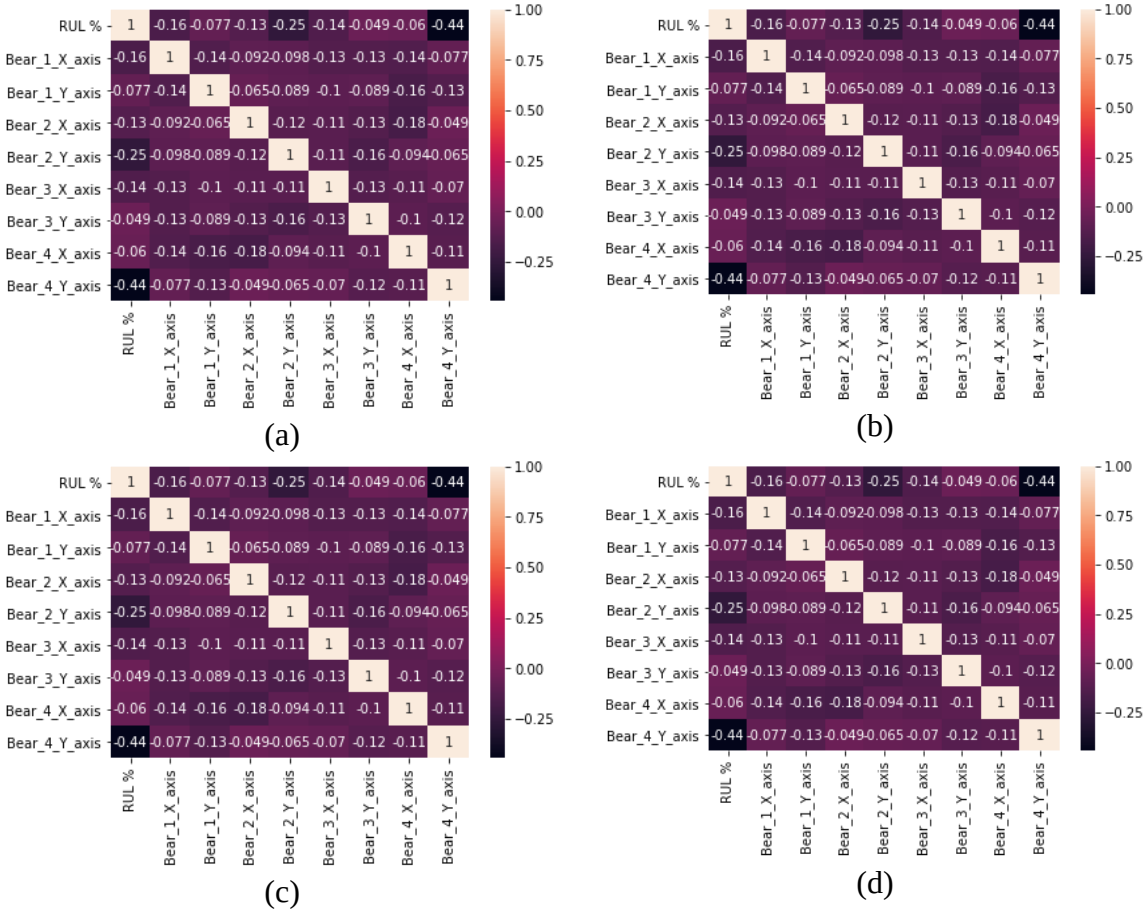
Data after FFT, PE and log distribution a) level=3 and b) level=4



(a) (b)

Figure 57 Pearson Correlation Heatmap IV:

Data after 1D SGF, PE and log distribution a) level=3 and b) level=4



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

Figure 58 Kendall Correlation Heatmap V:

Data after DTCWT_b_hp_pe_d

a) level=3, form=1, b) level=3, form=2,

c) level=4, form=1 and d) level=4, form=2

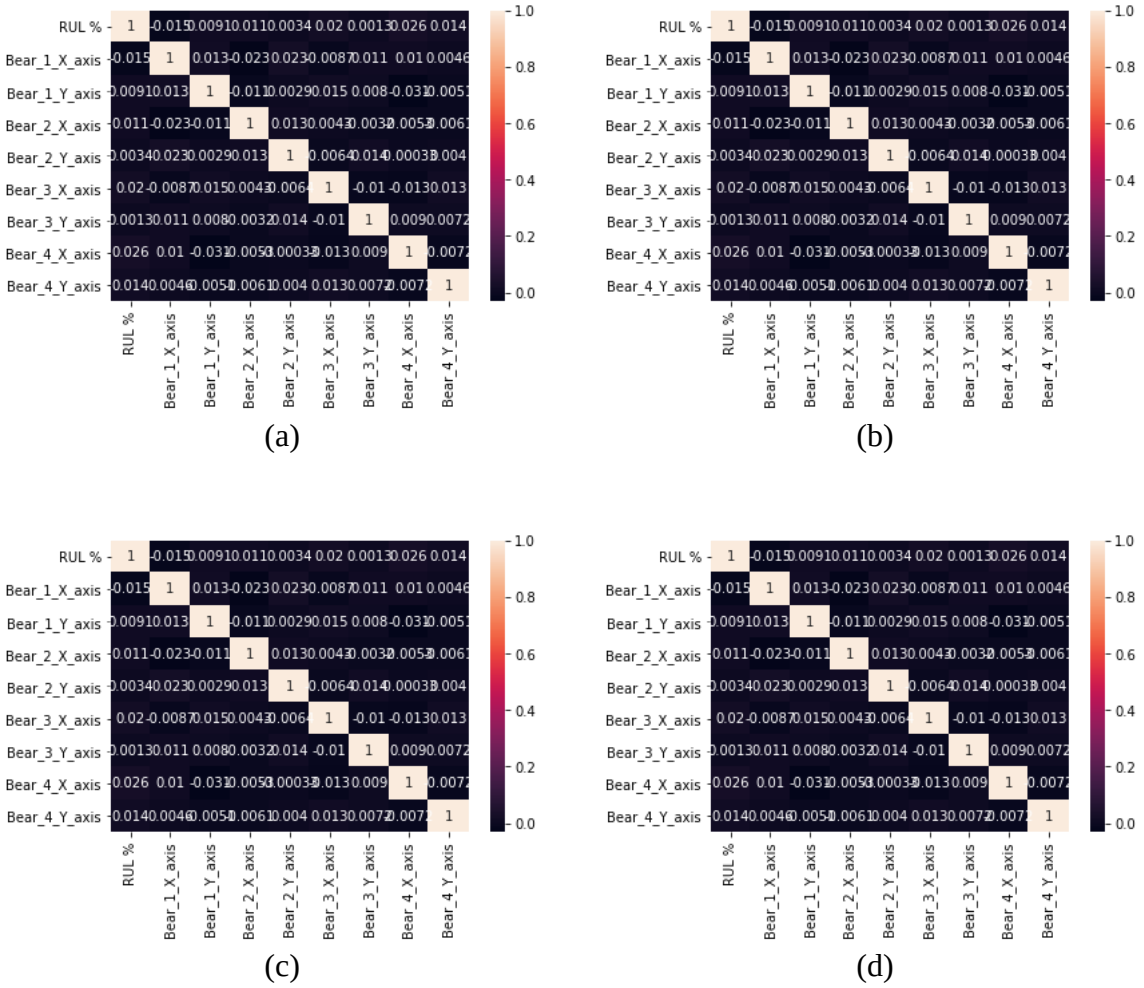
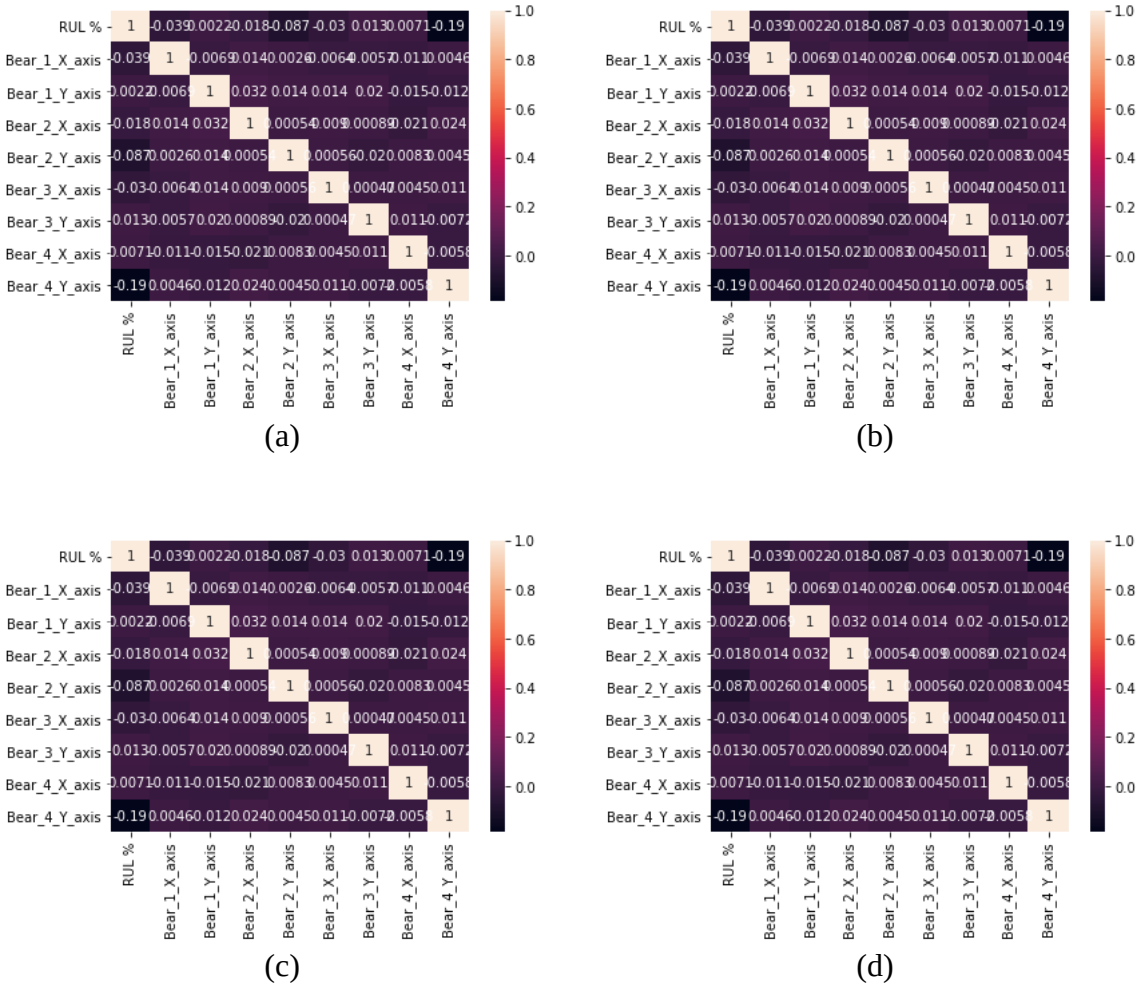


Figure 59 Kendall Correlation Heatmap VI:

Data after DTCWT_q_hp_pe_d

a) level=3, form=1, b) level=3, form=2,

c) level=4, form=1 and d) level=4, form=2



(a) (b)

(c) (d)

Figure 60 Kendall Correlation Heatmap VII:

Data after DTCWT_b_lp_d

a) level=3, form=1, b) level=3, form=2,

c) level=4, form=1 and d) level=4, form=2

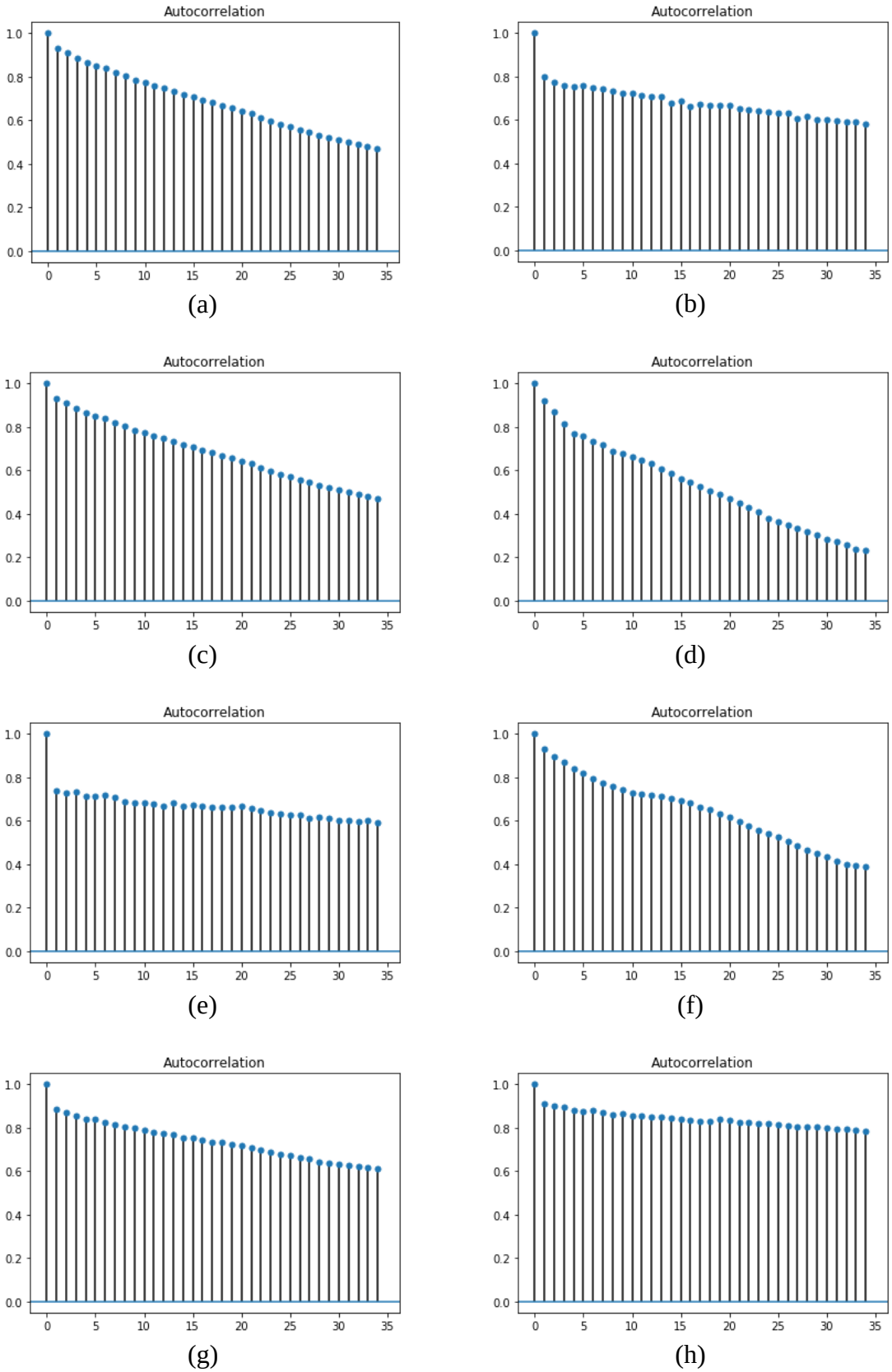


Figure 61 Autocorrelation II: Raw signal after PE level=3, (a,c,e,f) X-axis and (b,d,f,h) Y-axis

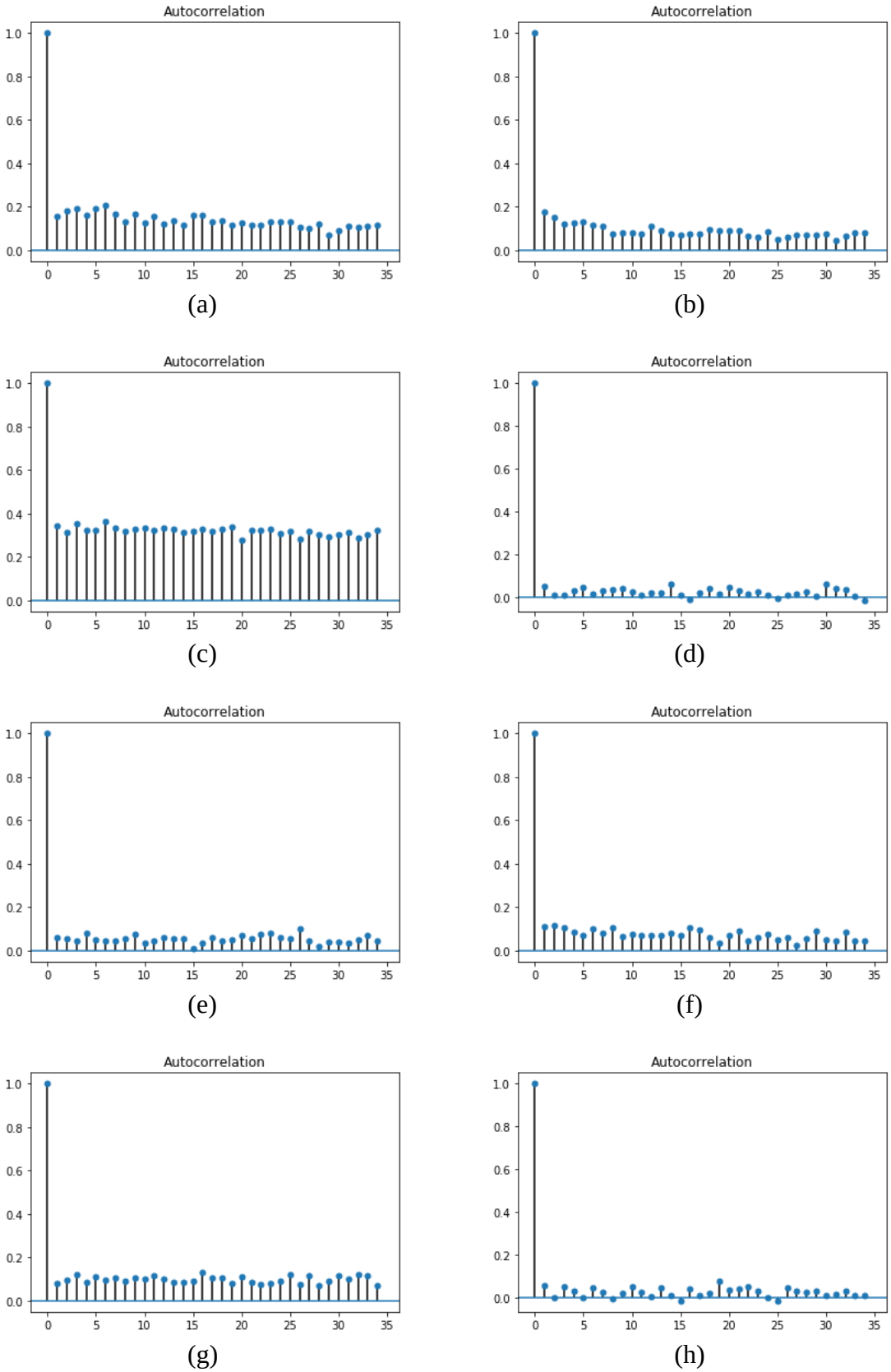


Figure 62 Autocorrelation III. BF low pass band after PE level=3, (a,c,e,f) X-axis and (b,d,f,h) Y-axis

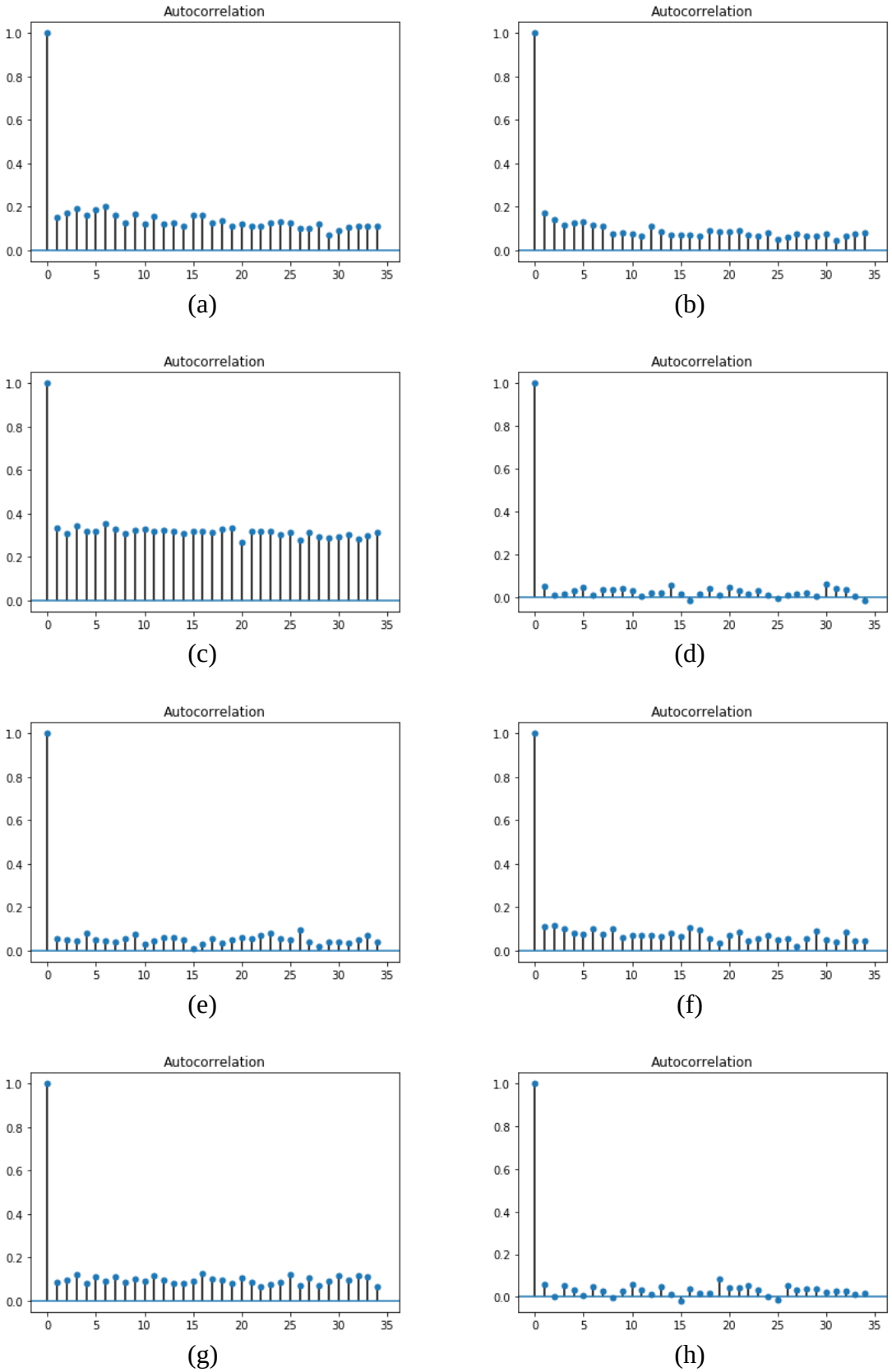


Figure 63 Autocorrelation IV: BF low pass band after PE Level=4, (a,c,e,f) X-axis and (b,d,f,h) Y-axis

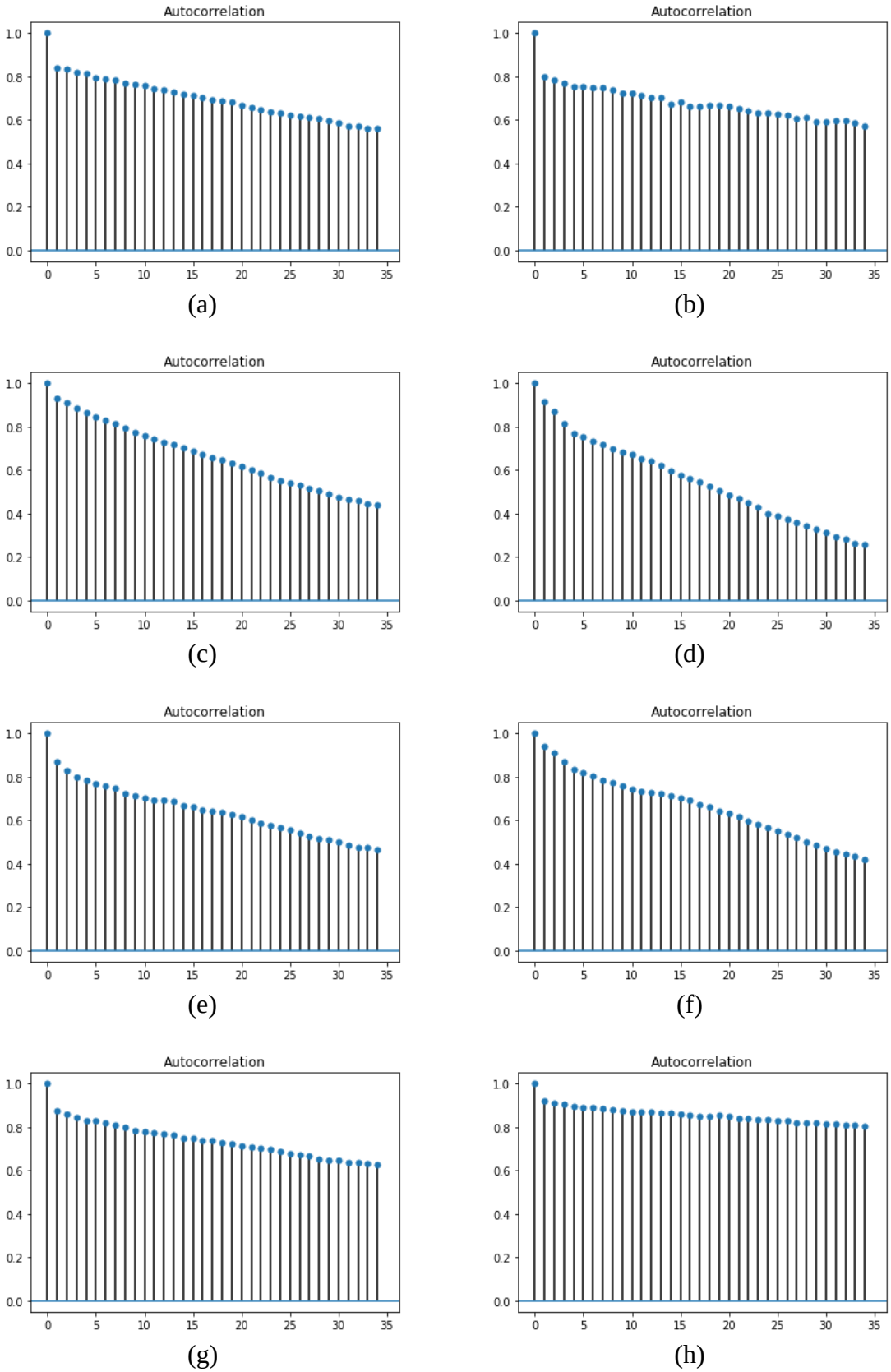


Figure 64 Autocorrelation V: HT magnitude after PE

Level=3, (a,c,e,f) X-axis and (b,d,f,h) Y-axis

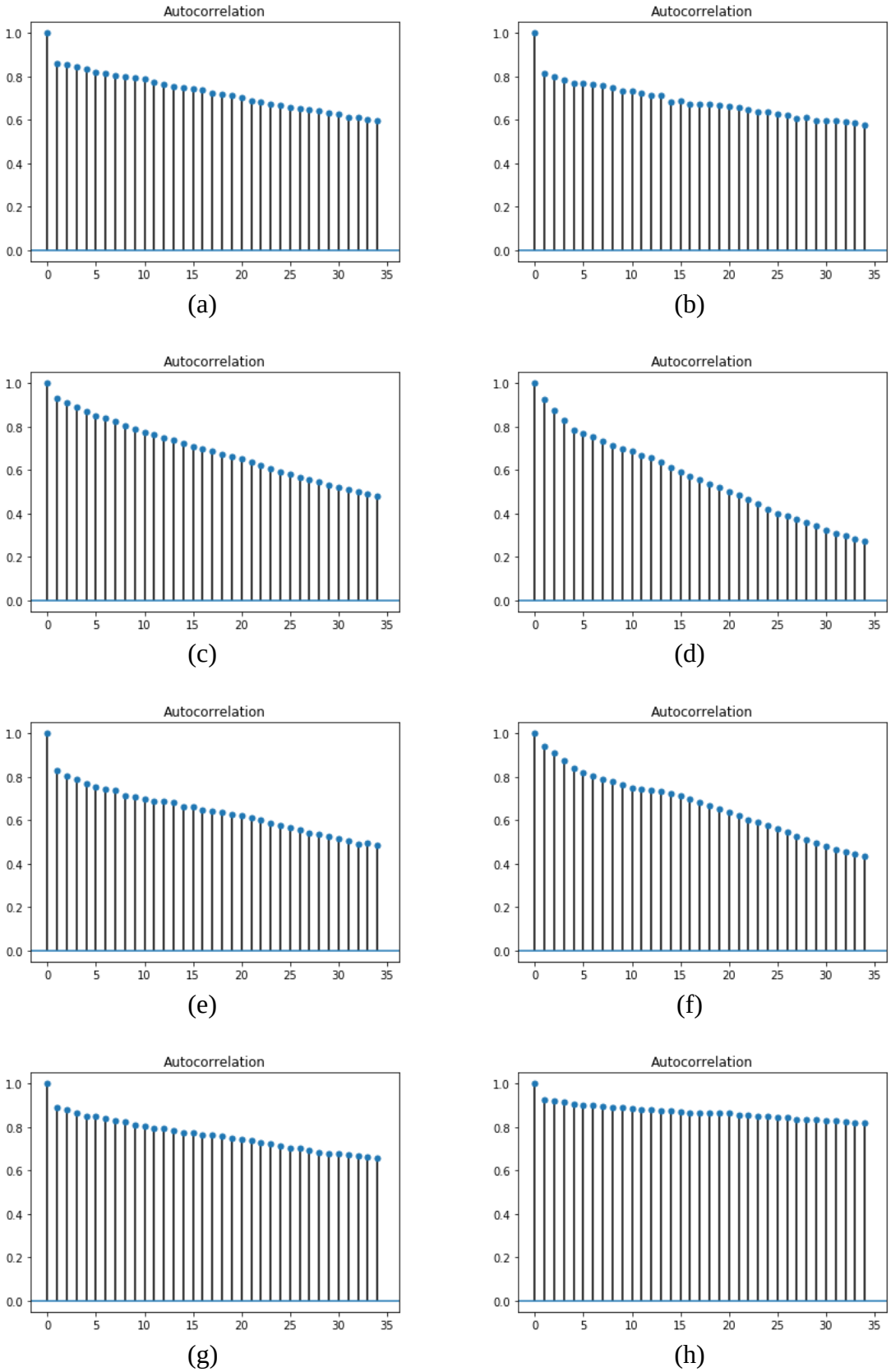


Figure 65 Autocorrelation VI: HT magnitude after PE Level=4, (a,c,e,f) X-axis and (b,d,f,h) Y-axis

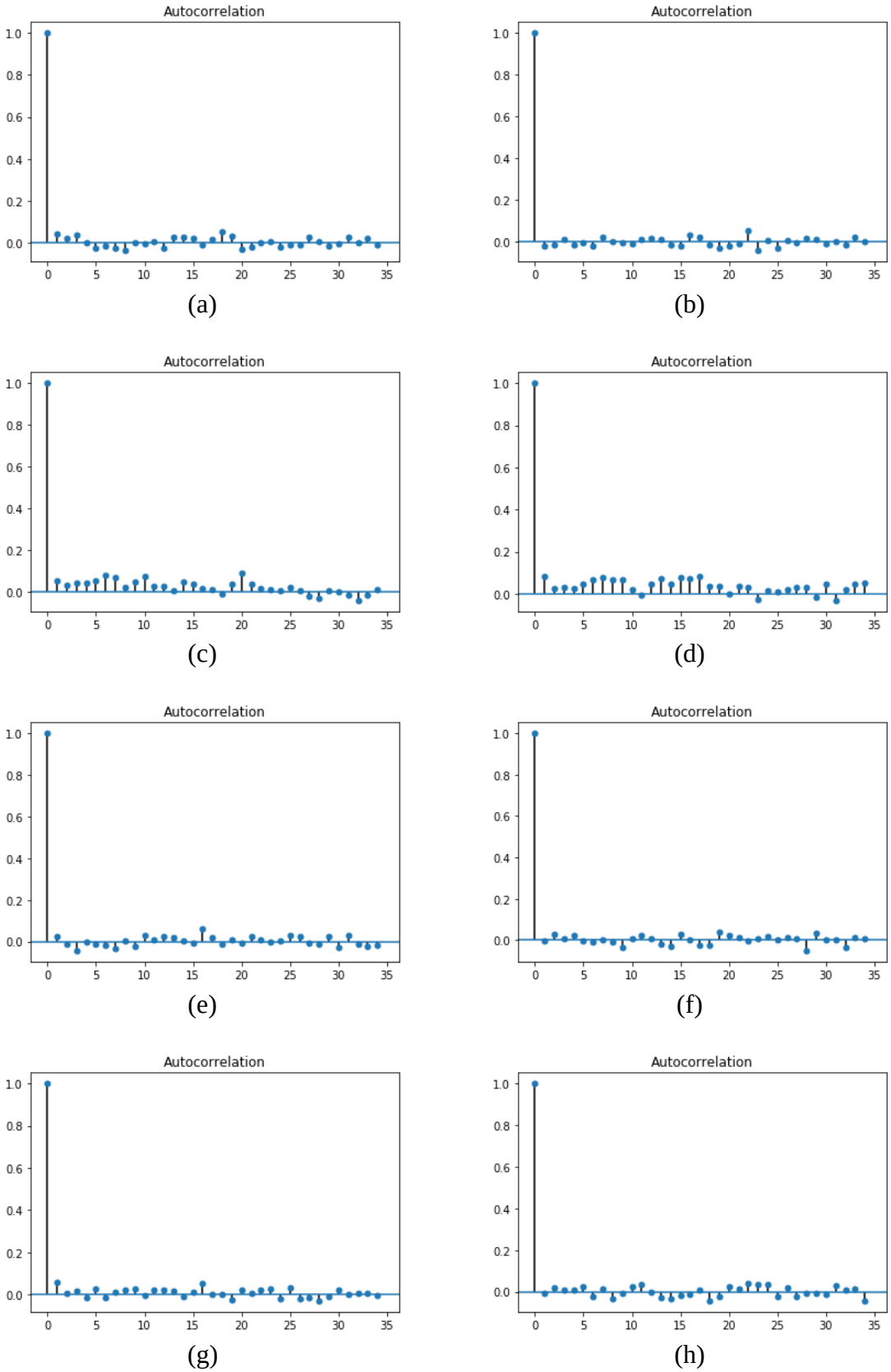


Figure 66 Autocorrelation VII: FFT magnitude after PE

Level=3, (a,c,e,f) X-axis and (b,d,f,h) Y-axis

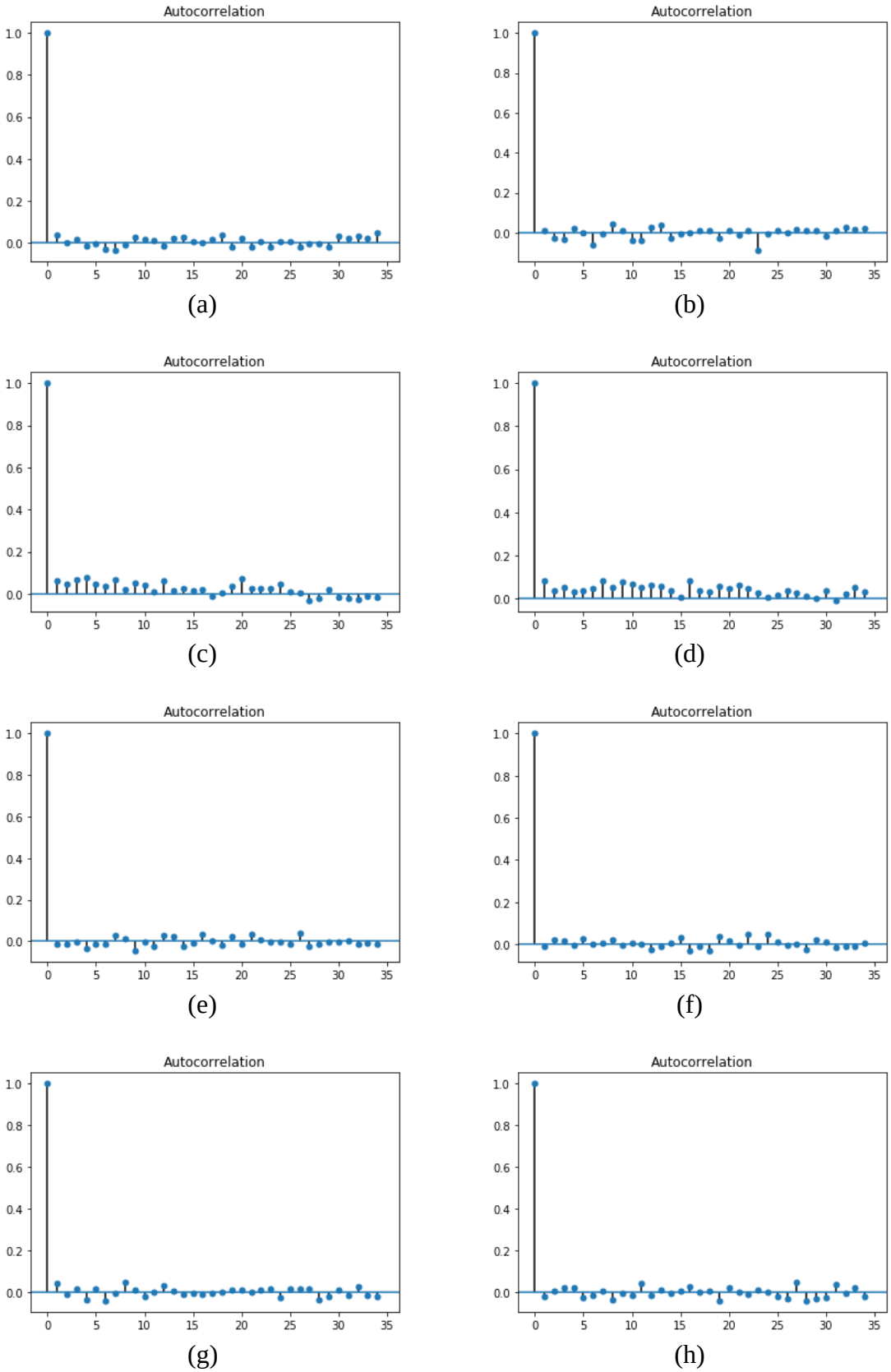


Figure 67 Autocorrelation VIII: FFT magnitude after PE

Level=4, (a,c,e,f) X-axis and (b,d,f,h) Y-axis

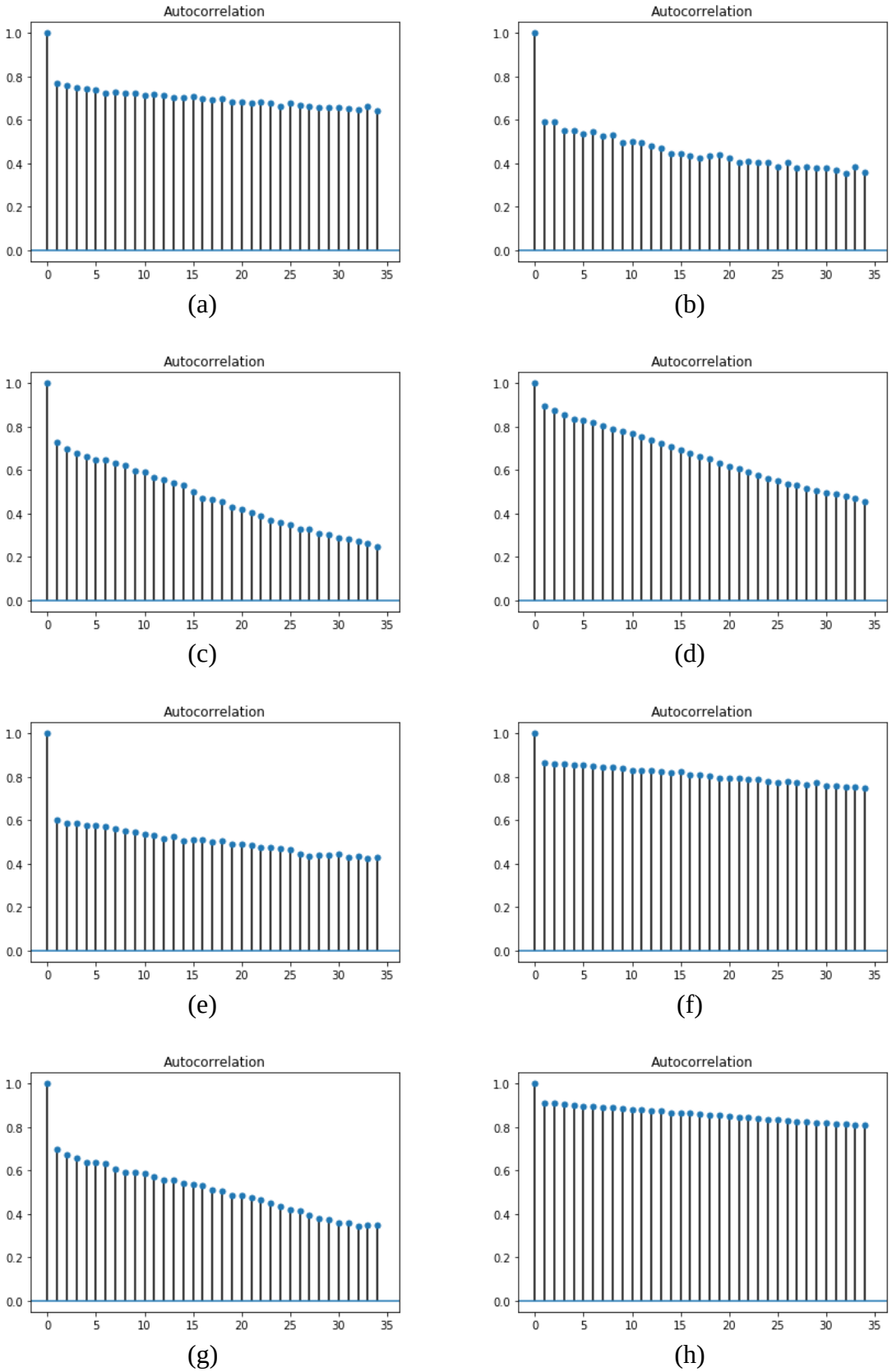


Figure 68 Autocorrelation IX: SGF after PE
Level=3, (a,c,e,f) X-axis and (b,d,f,h) Y-axis

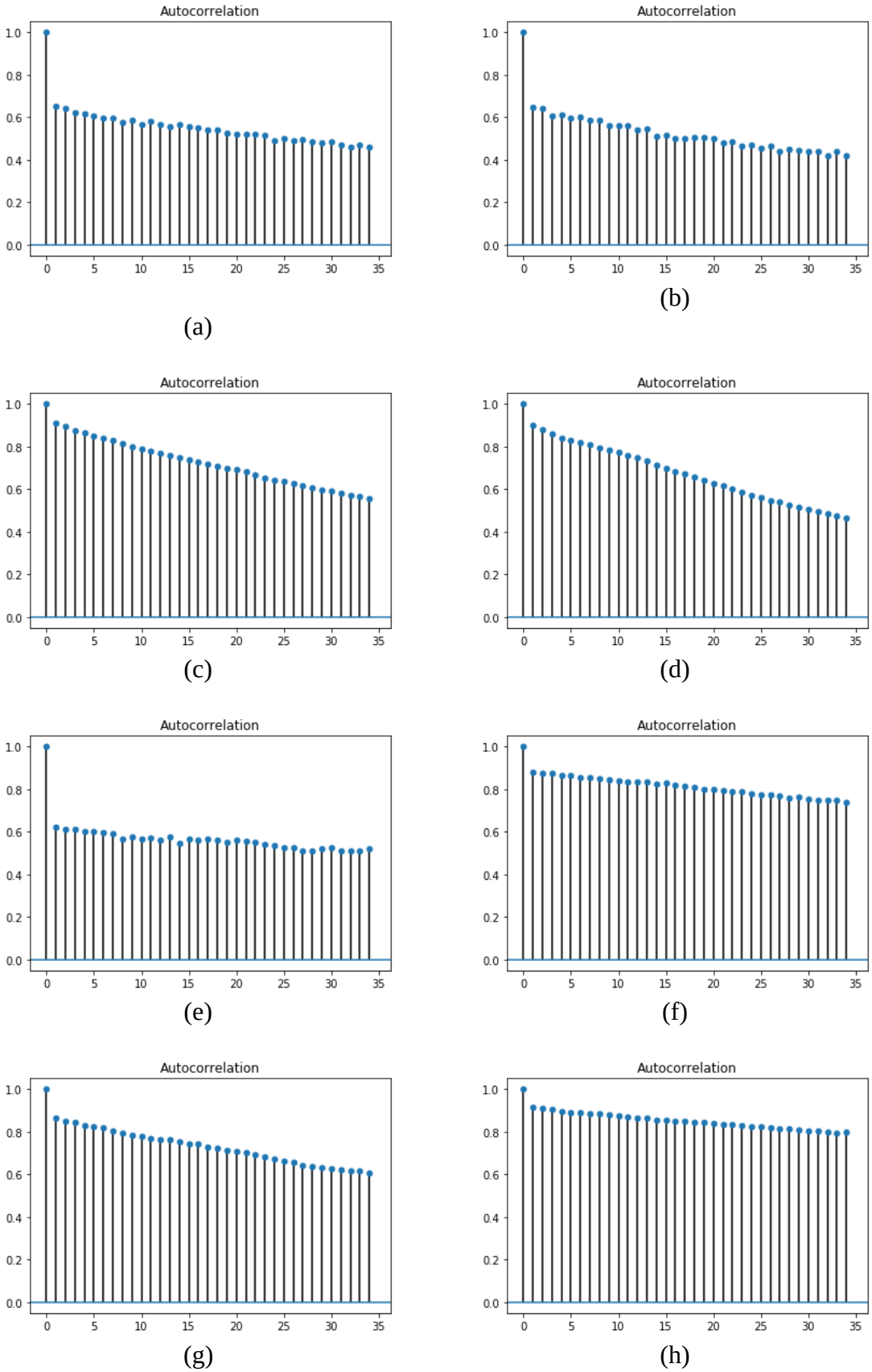


Figure 69 Autocorrelation X: SGF after PE

Level=4, (a,c,e,f) X-axis and (b,d,f,h) Y-axis

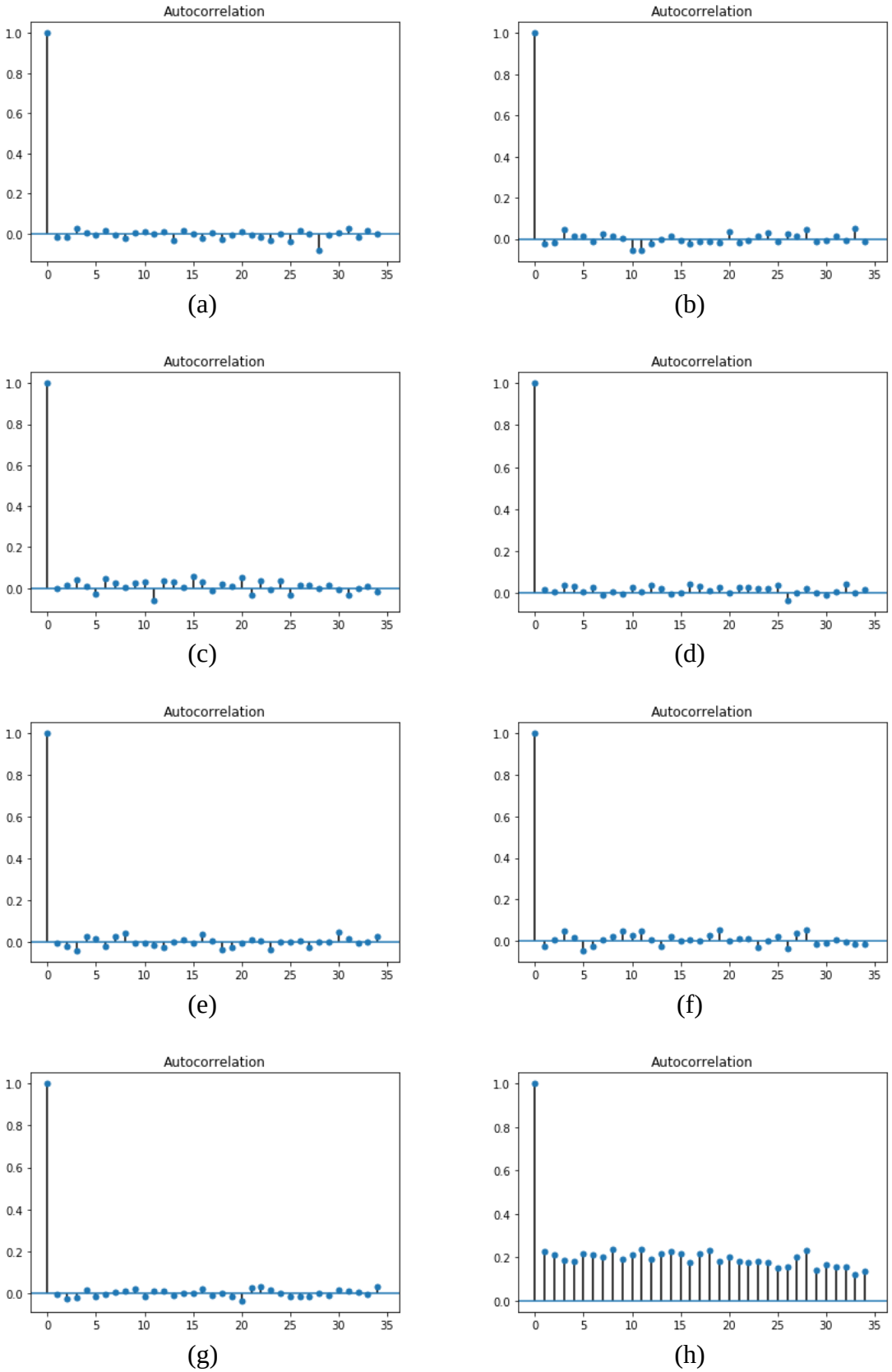


Figure 70 Autocorrelation XI: DTCWT biort low pass after PE

Level=4, Form=2, (a,c,e,f) X-axis and (b,d,f,h) Y-axis

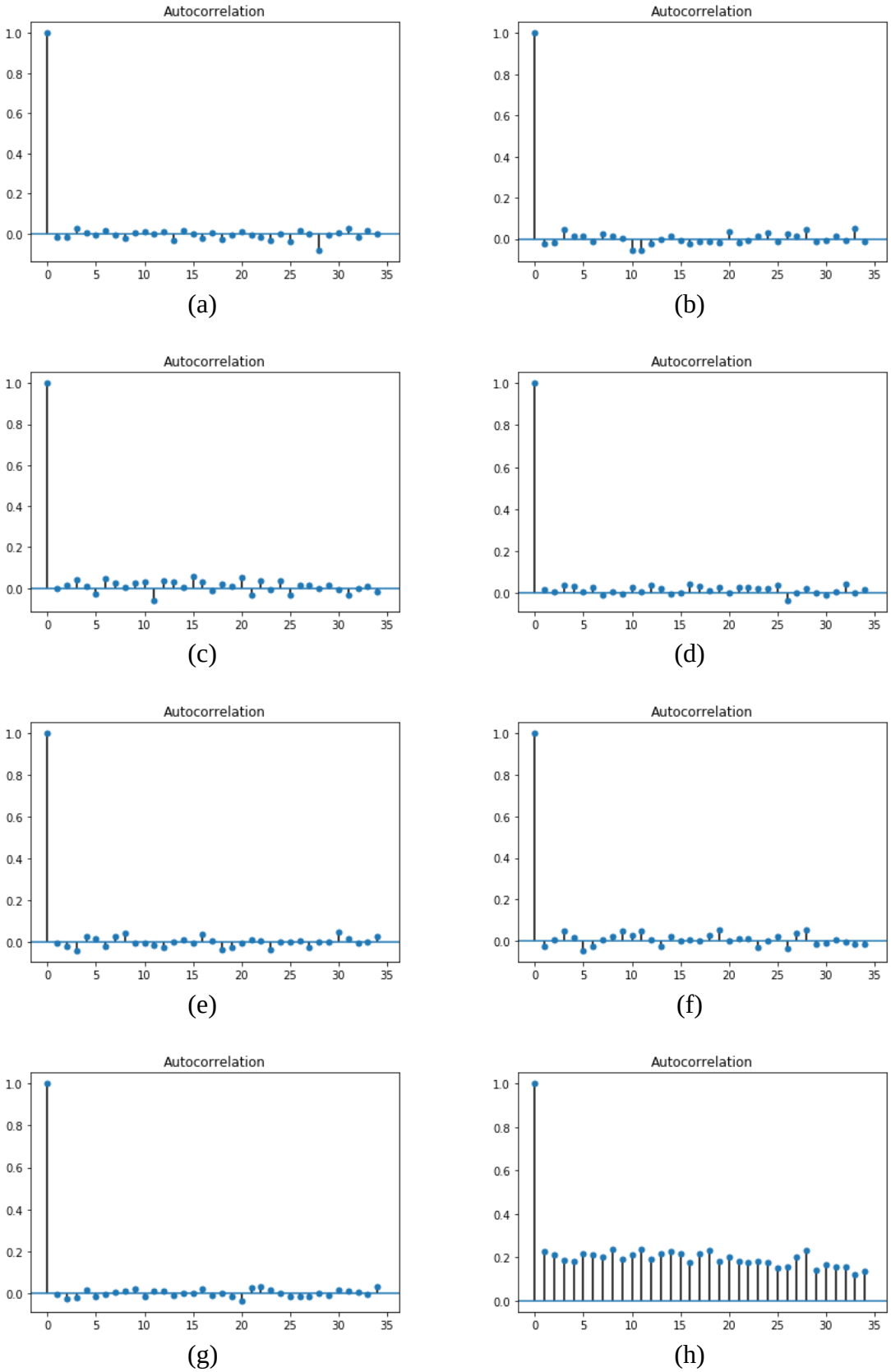


Figure 71 Autocorrelation XII: DTCWT qshift low pass after PE Level=4, Form=2, (a,c,e,f) X-axis and (b,d,f,h) Y-axis

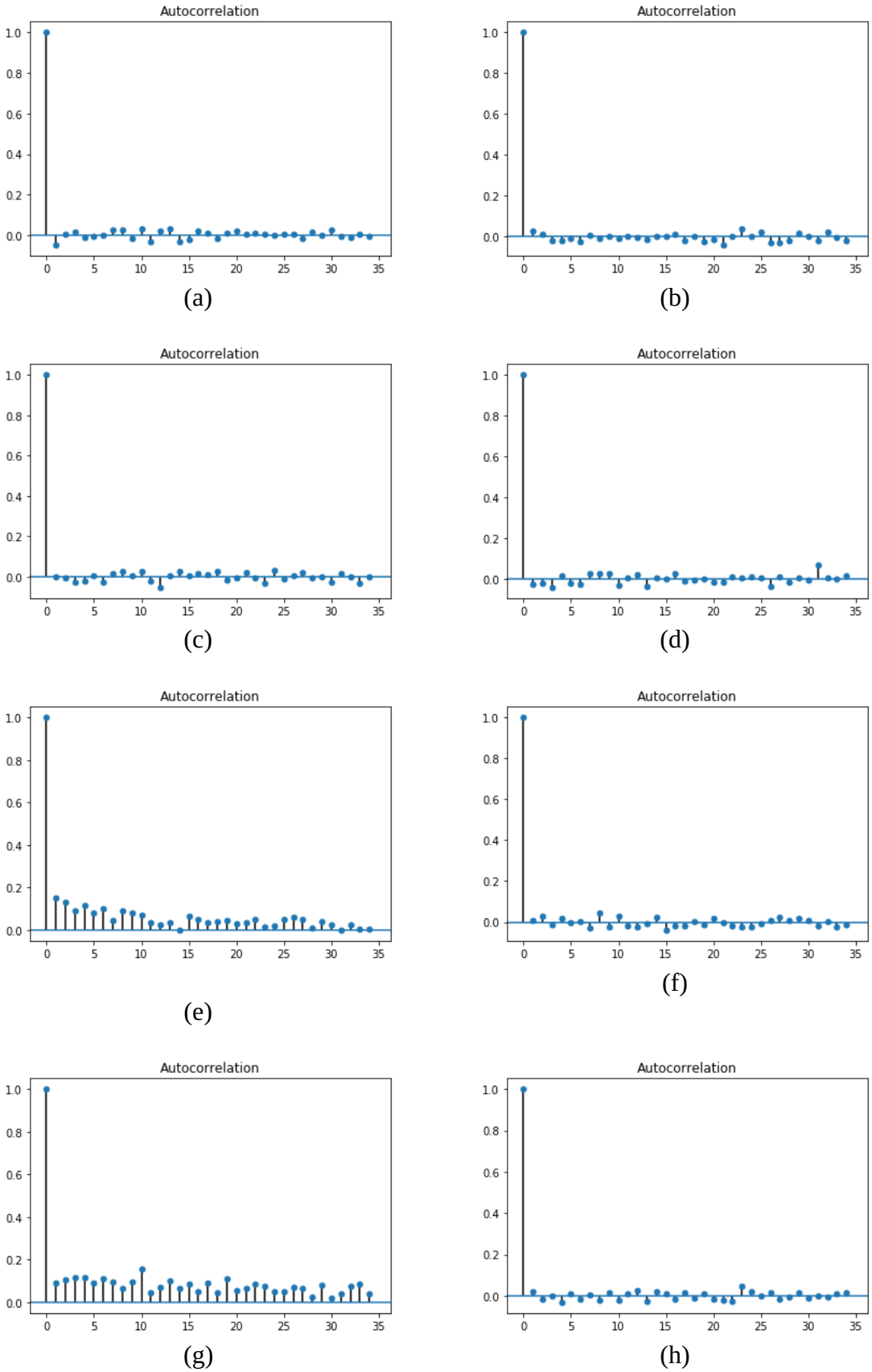


Figure 72 Autocorrelation XIII: DTCWT qshift low pass after PE Level=4, Form=2, (a,c,e,f) X-axis and (b,d,f,h) Y-axis

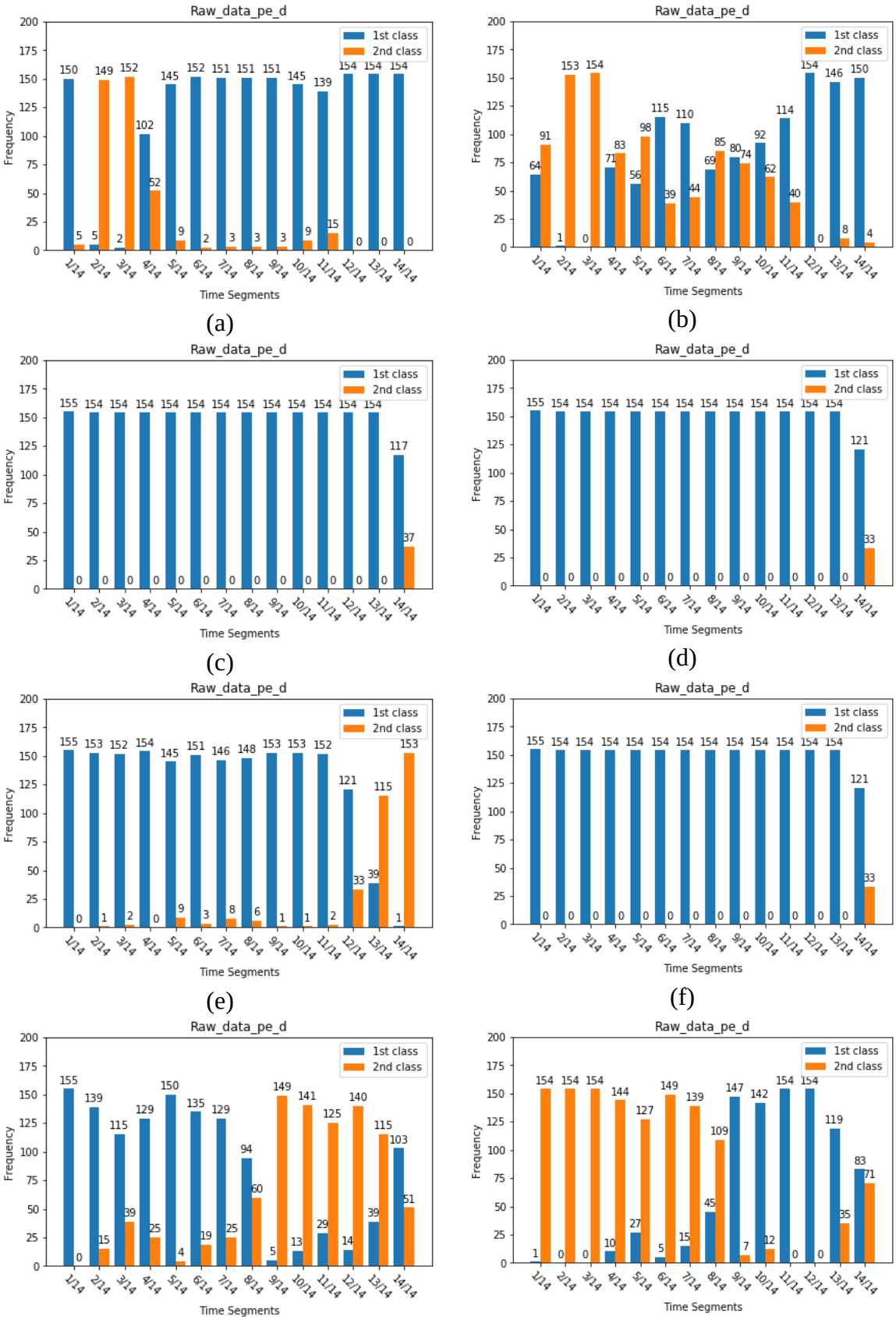


Figure 73 Clustering I Level=3, (a,c,e,f) X-axis and (b,d,f,h) Y-axis

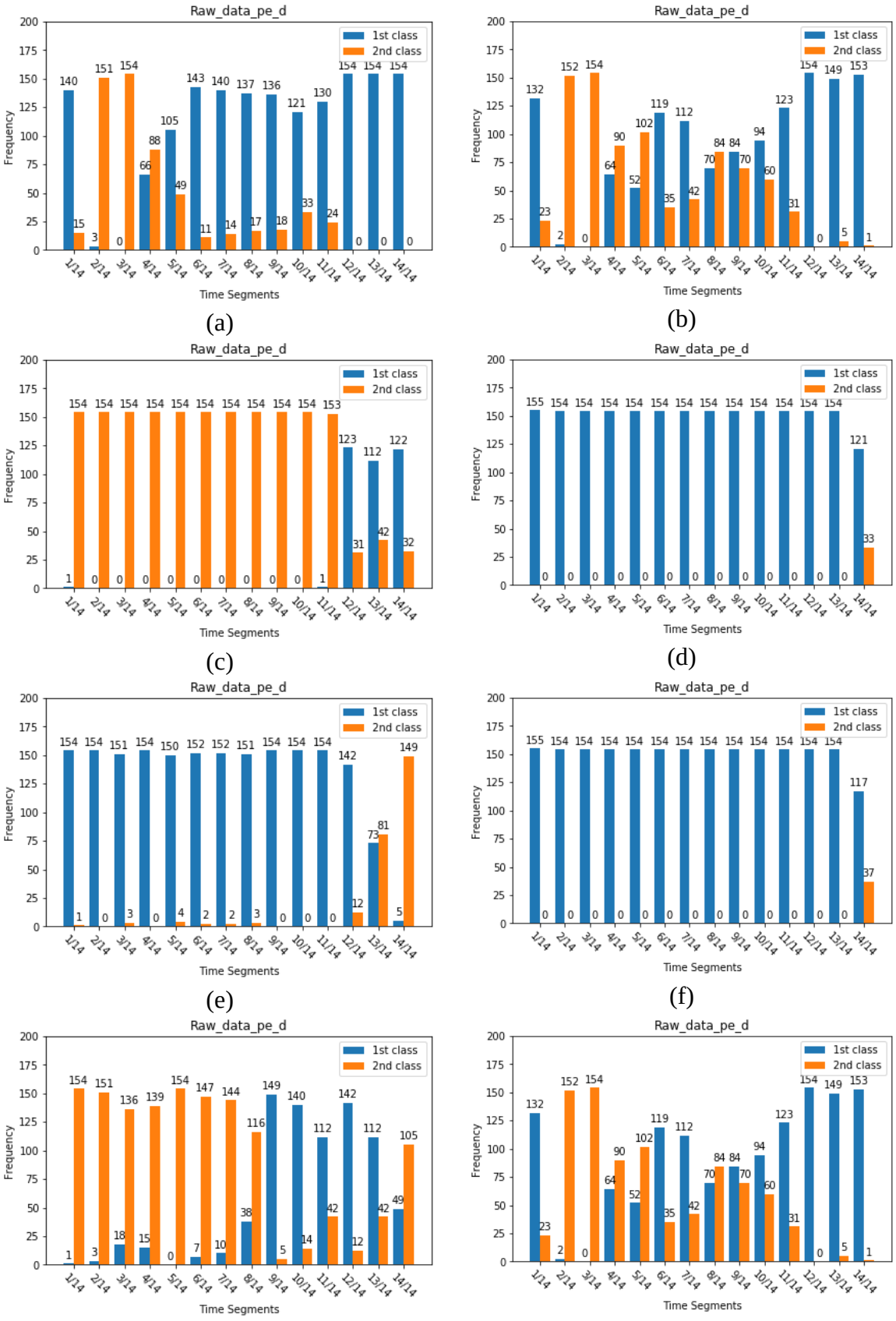


Figure 74 Clustering II Level=4, (a,c,e,f) X-axis and (b,d,f,h) Y-axis

Figure 75 Clustering III Level=3, (a,c,e,f) X-axis and (b,d,f,h) Y-axis

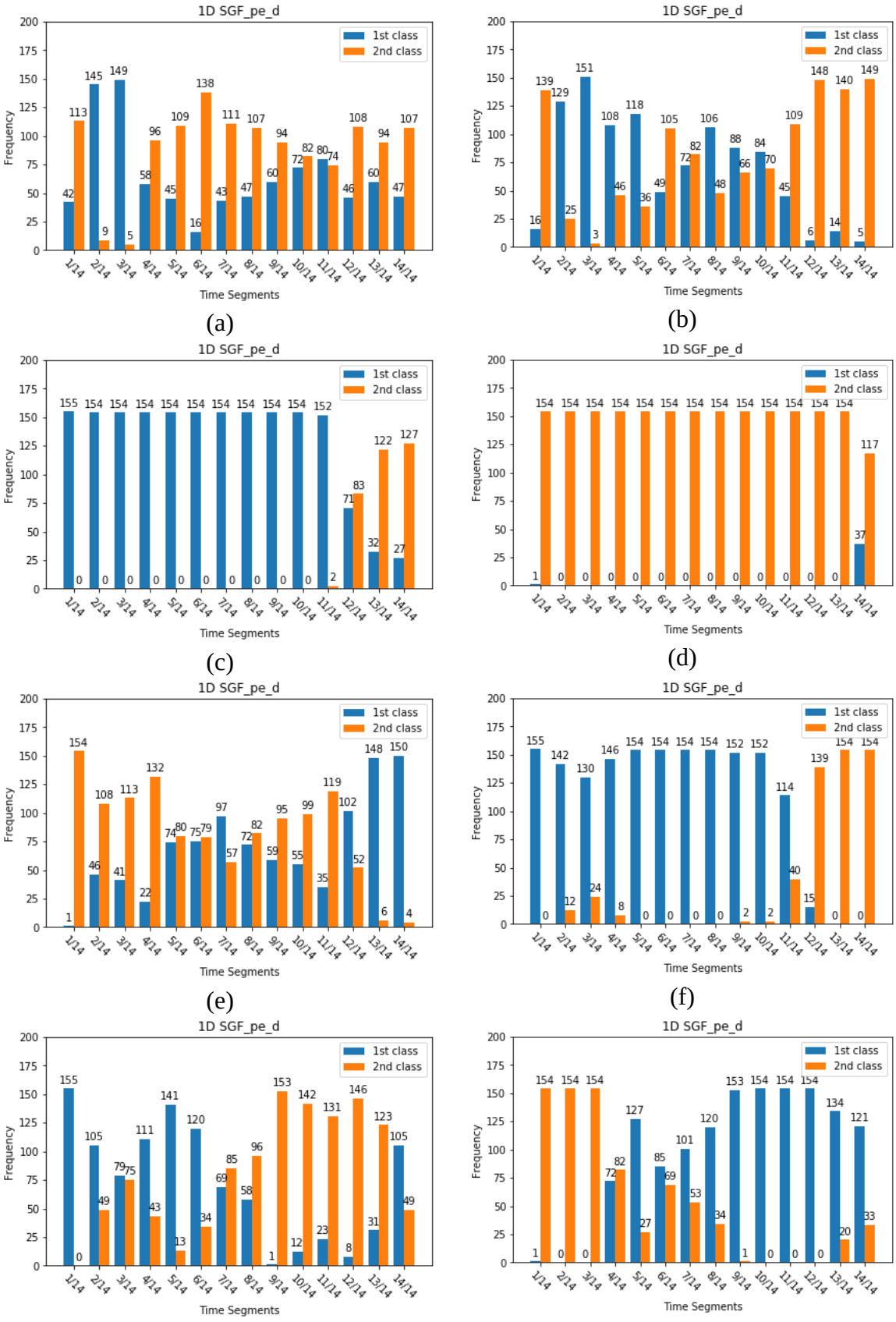


Figure 76 Clustering IV Level=4, (a,c,e,f) X-axis and (b,d,f,h) Y-axis

Figure 77 Clustering V Level=3, (a,c,e,f) X-axis and (b,d,f,h) Y-axis

Figure 78 Clustering VI Level=4, (a,c,e,f) X-axis and (b,d,f,h) Y-axis

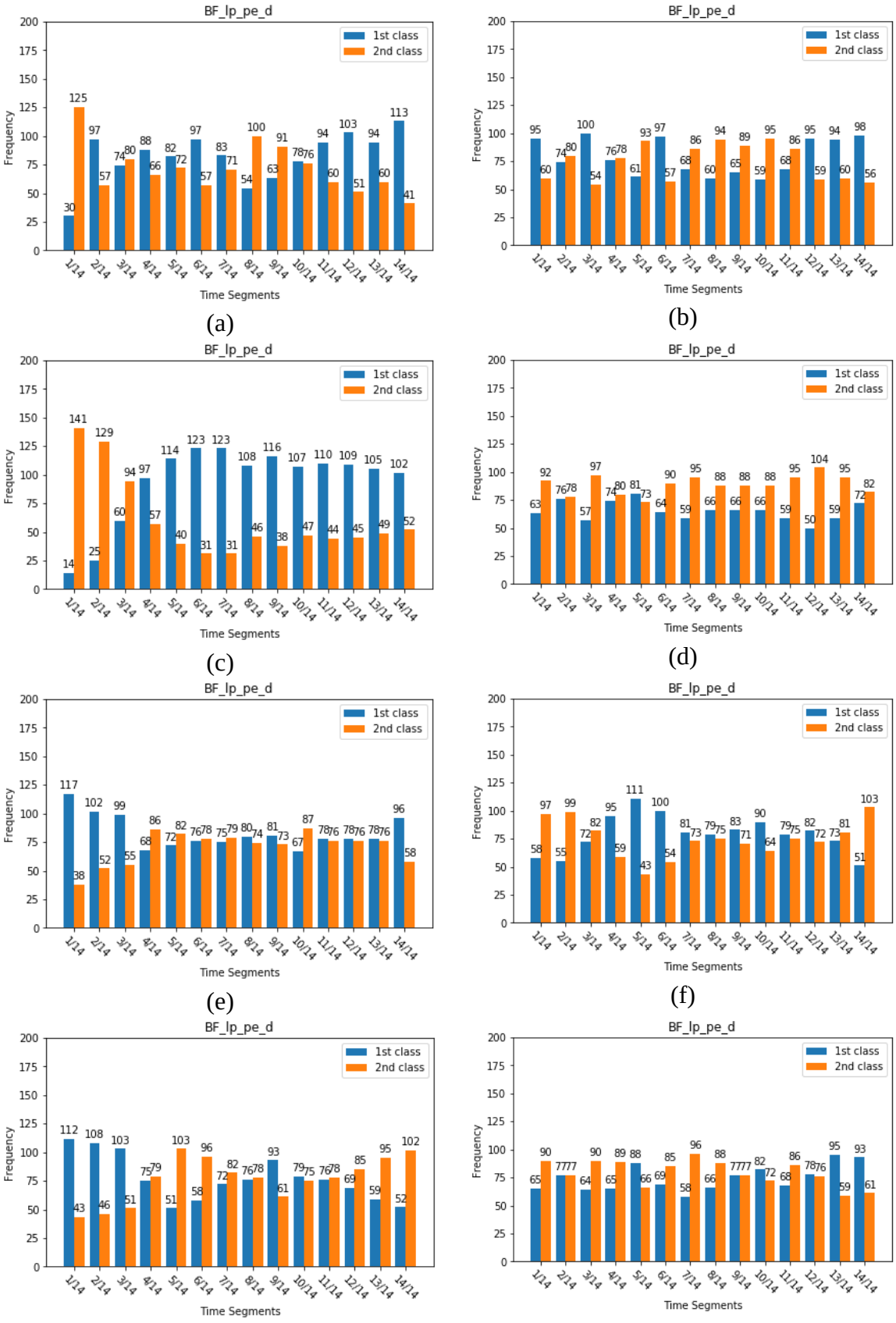


Figure 79 Clustering VII Level=3, (a,c,e,f) X-axis and (b,d,f,h) Y-axis

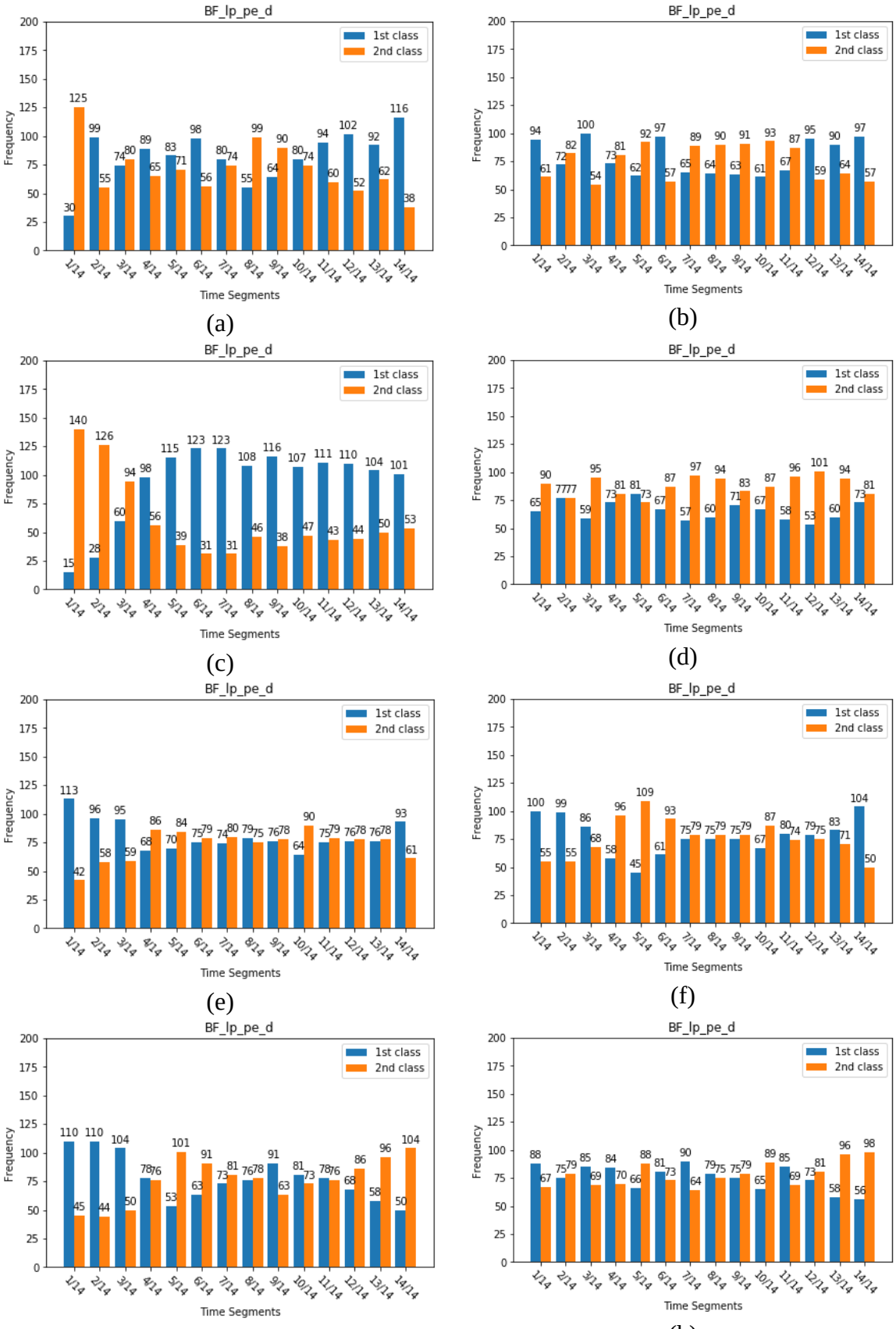


Figure 80 Clustering VIII Level=4, (a,c,e,f) X-axis and (b,d,f,h) Y-axis

Figure 81 Clustering IX Level=3, (a,c,e,f) X-axis and (b,d,f,h) Y-axis

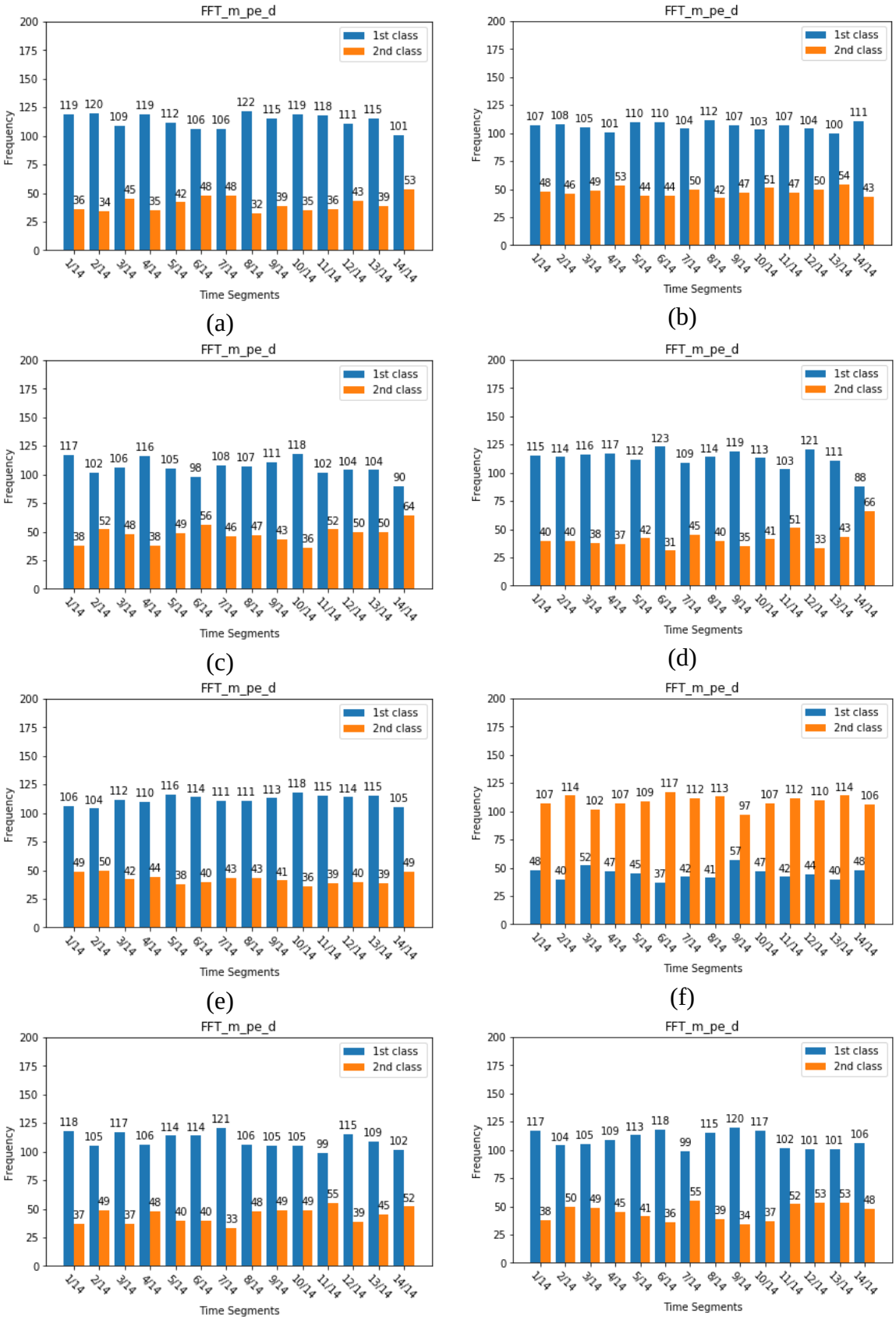


Figure 82 Clustering X Level=4, (a,c,e,f) X-axis and (b,d,f,h) Y-axis

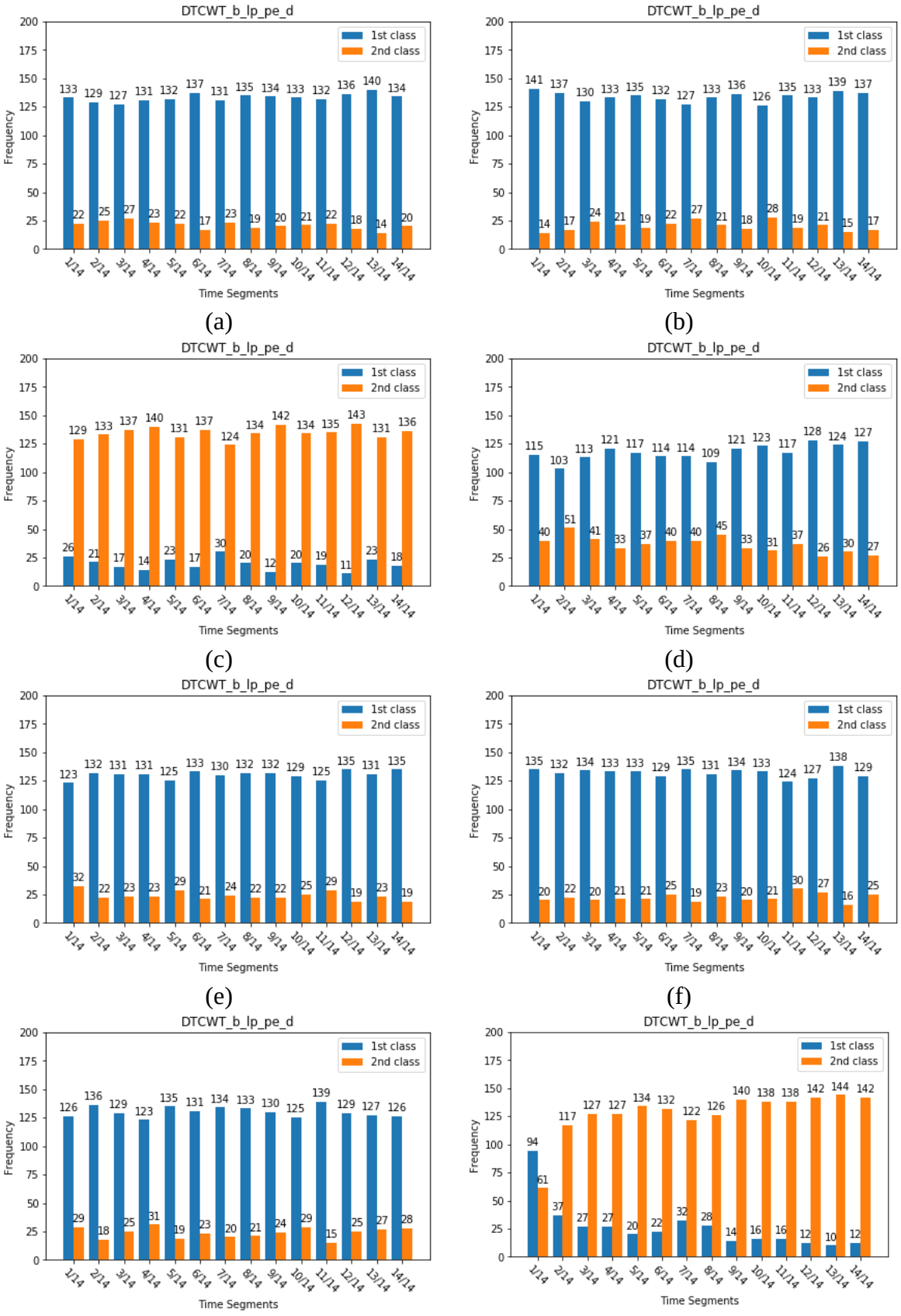


Figure 83 Clustering XI Level=4, form=1, (a,c,e,f) X-axis and (b,d,f,h) Y-axis

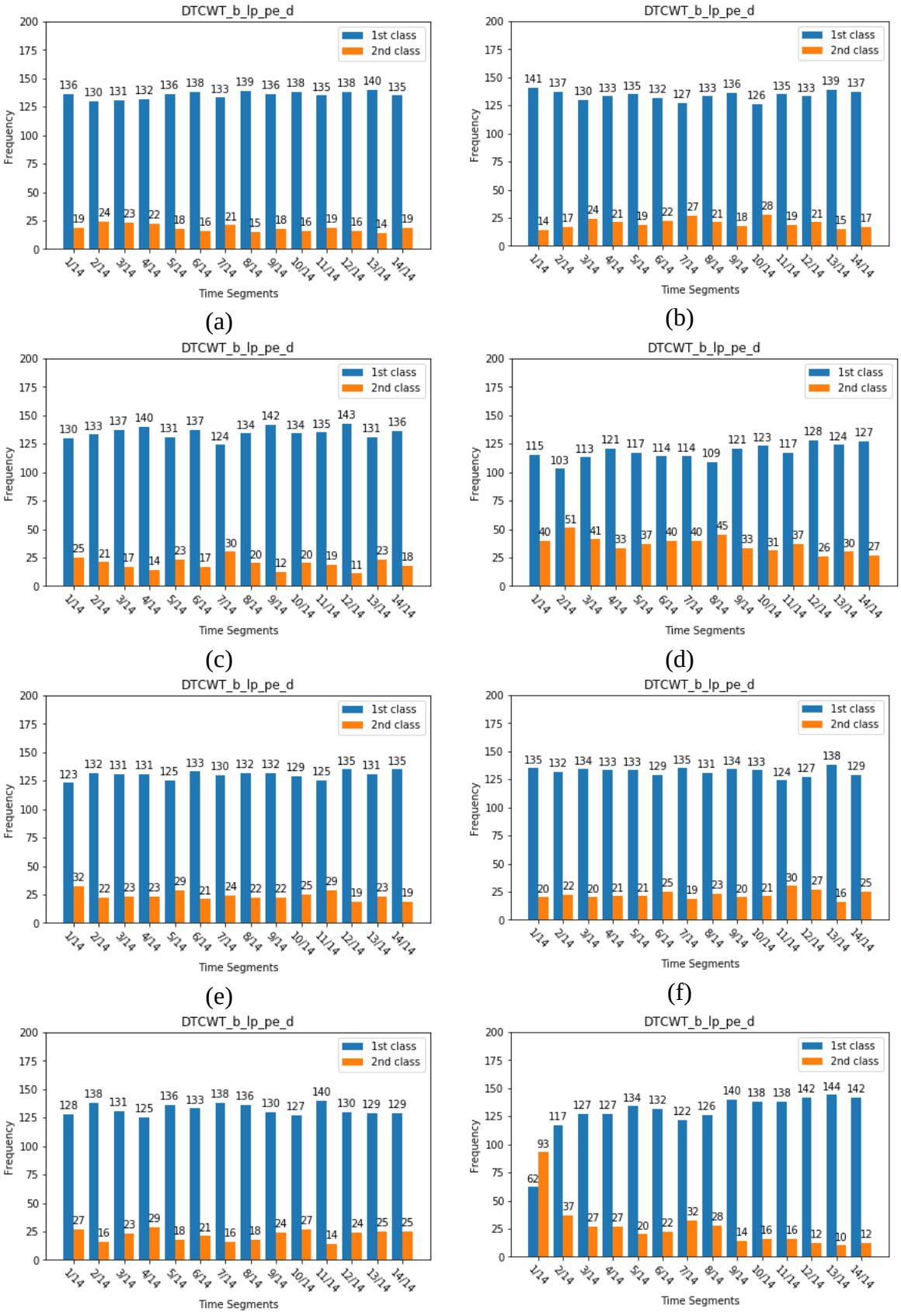


Figure 84 Clustering XII Level=3, form=2, (a,c,e,f) X-axis and (b,d,f,h) Y-axis

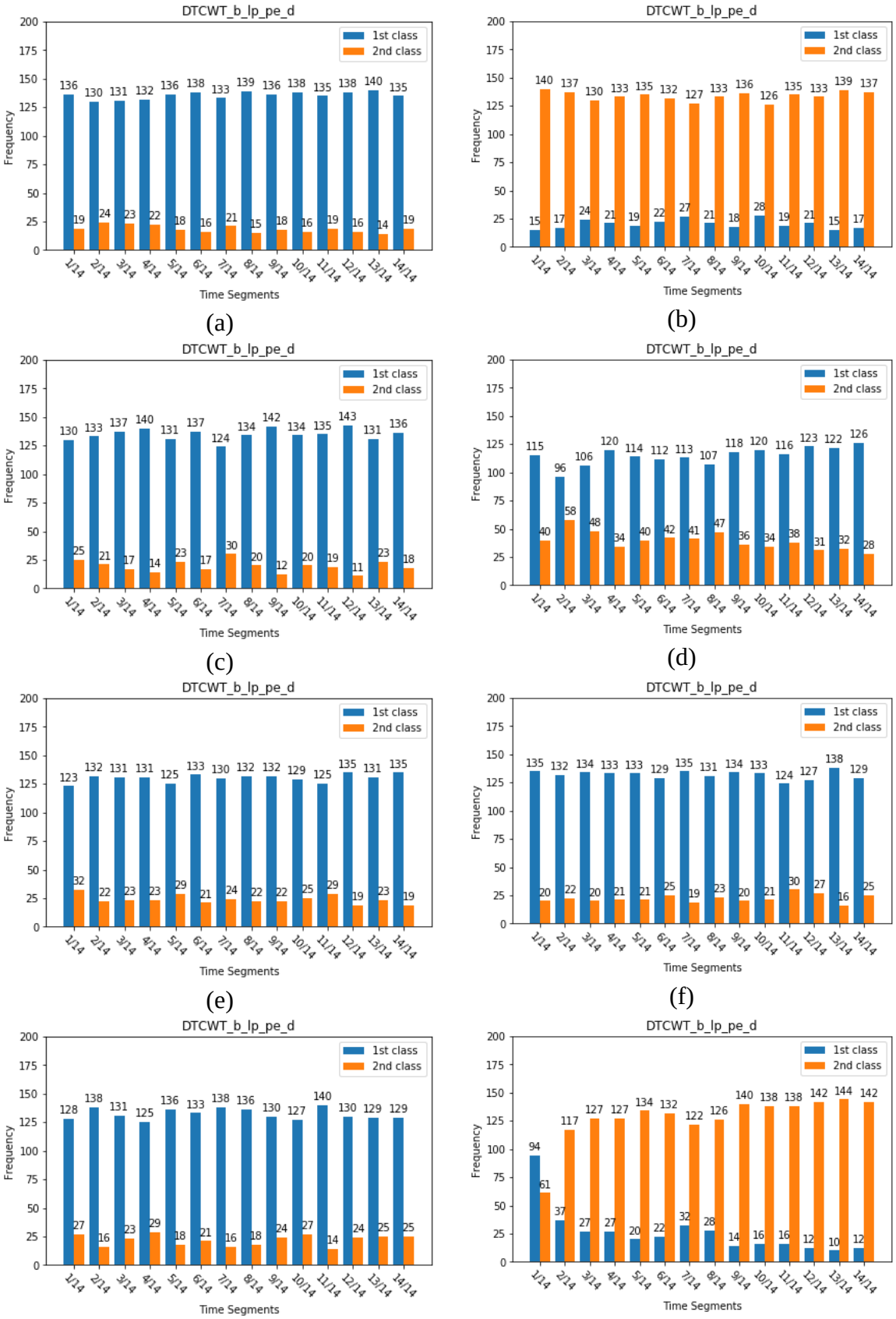


Figure 85 Clustering XIII Level=4, form=1, (a,c,e,f) X-axis and (b,d,f,h) Y-axis

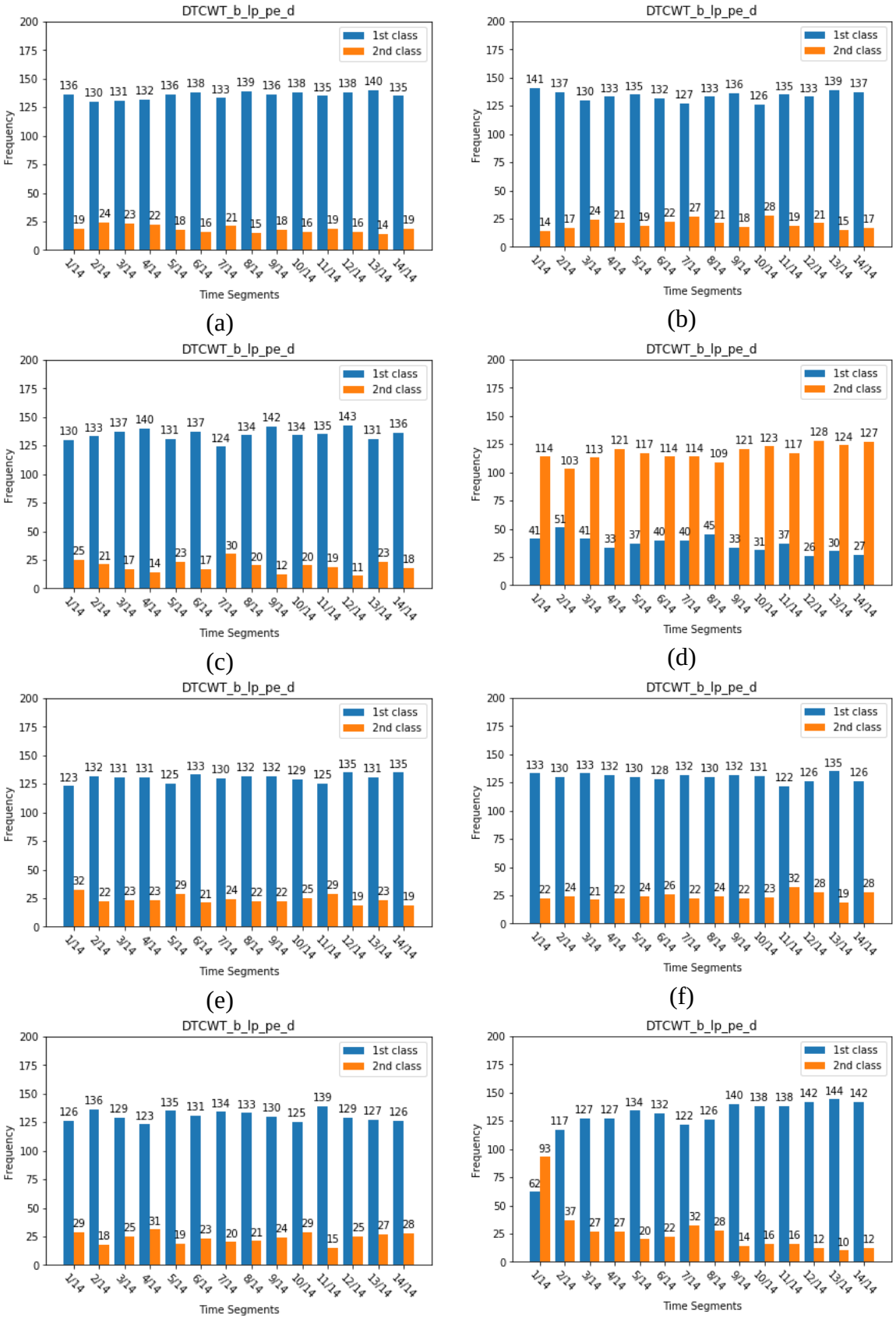


Figure 86 Clustering XIV Level=4, form=2, (a,c,e,f) X-axis and (b,d,f,h) Y-axis

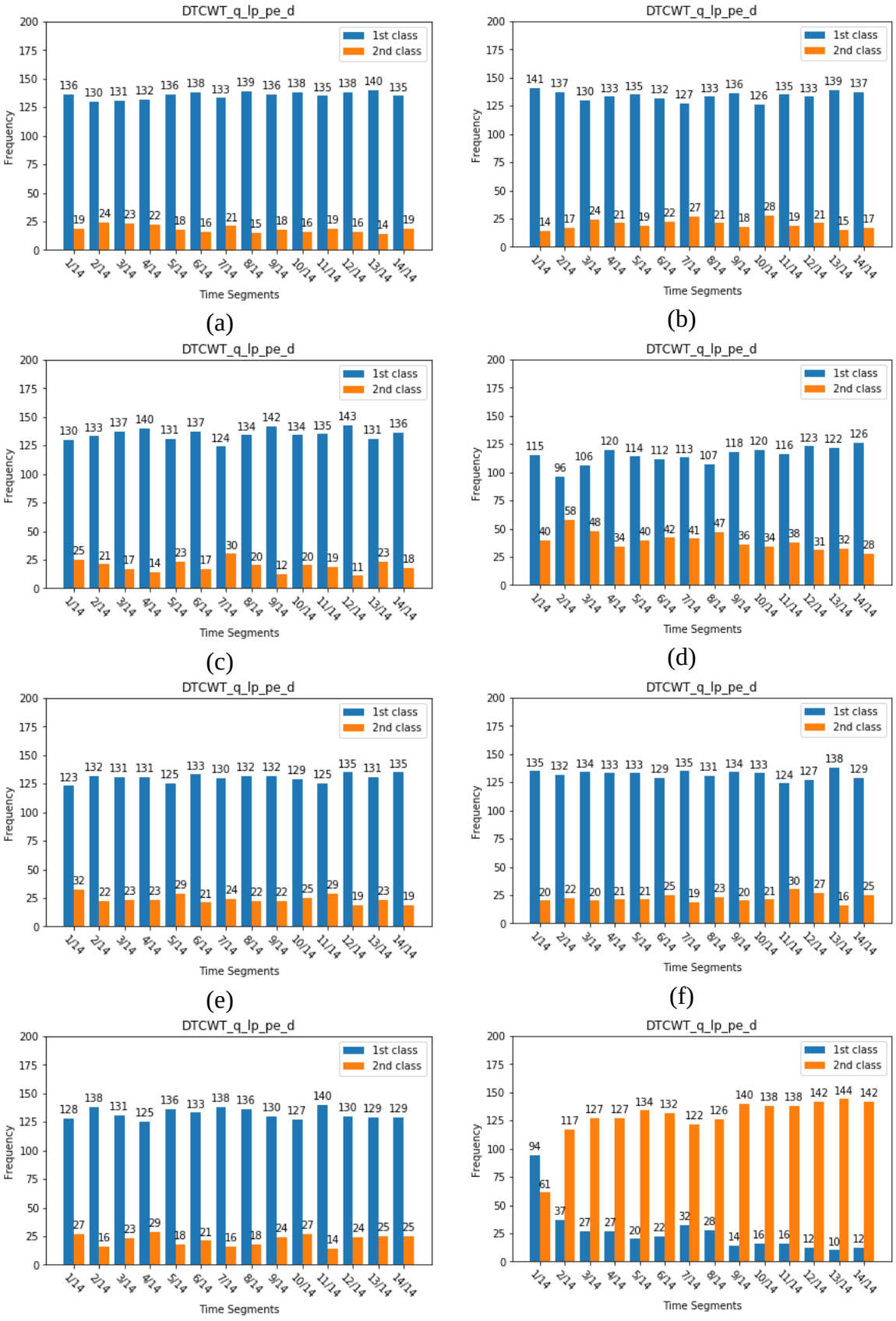


Figure 87 Clustering XV Level=3, form=1, (a,c,e,f) X-axis and (b,d,f,h) Y-axis

Figure 88 Clustering XVI Level=4, form=1, (a,c,e,f) X-axis and (b,d,f,h) Y-axis

Table 41
Raw_data_pe_d I

	Bearings / Axis	
	Bear_1_X_axis	Bear_2_X_axis
Bear_3_X_axis	1	0
Bear_4_X_axis	1,926	2,151
	Bear_1_Y_axis	Bear_2_Y_axis
Bear_3_Y_axis	0	1
Bear_4_Y_axis	2,039	2,156

Note. Level=3

Table 42
Raw_data_pe_d II

	Bearings / Axis	
	Bear_1_X_axis	Bear_2_X_axis
Bear_3_X_axis	3	3
Bear_4_X_axis	1,595	2,151
	Bear_1_Y_axis	Bear_2_Y_axis
Bear_3_Y_axis	0	6
Bear_4_Y_axis	1,817	2,154

Note. Level=4

Table 43
HT_m_pe_d I

	Bearings / Axis	
	Bear_1_X_axis	Bear_2_X_axis
Bear_3_X_axis	0	0
Bear_4_X_axis	1,793	2,149
	Bear_1_Y_axis	Bear_2_Y_axis
Bear_3_Y_axis	9	5
Bear_4_Y_axis	2,053	2,154

Note. Level=3

Table 44
HT_m_pe_d II

	Bearings / Axis	
	Bear_1_X_axis	Bear_2_X_axis
Bear_3_X_axis	0	4
Bear_4_X_axis	1,765	2,151
	Bear_1_Y_axis	Bear_2_Y_axis
Bear_3_Y_axis	16	0
Bear_4_Y_axis	2,062	2154

Note. Level=4

Table 45
FFT_m_pe_d I

	Bearings / Axis	
	Bear_1_X_axis	Bear_2_X_axis
Bear_3_X_axis	2,148	423
Bear_4_X_axis	2,154	32
	Bear_1_Y_axis	Bear_2_Y_axis
Bear_3_Y_axis	117	319
Bear_4_Y_axis	829	597

Note. Level=3

Table 46
FFT_m_pe_d II

	Bearings / Axis	
	Bear_1_X_axis	Bear_2_X_axis
Bear_3_X_axis	1,957	1,283
Bear_4_X_axis	399	3
	Bear_1_Y_axis	Bear_2_Y_axis
Bear_3_Y_axis	0	860
Bear_4_Y_axis	1,809	2,090

Note. Level=4

Table 47
1D SGF_pe_d II

	Bearings / Axis	
	Bear_1_X_axis	Bear_2_X_axis
Bear_3_X_axis	1,982	2,111
Bear_4_X_axis	1,895	2,131
	Bear_1_Y_axis	Bear_2_Y_axis
Bear_3_Y_axis	1,719	1,898
Bear_4_Y_axis	1,807	1,991

Note. Level=3

Table 48
1D SGF_pe_d II

	Bearings / Axis	
	Bear_1_X_axis	Bear_2_X_axis
Bear_3_X_axis	1,607	2,100
Bear_4_X_axis	2,154	2,154
	Bear_1_Y_axis	Bear_2_Y_axis
Bear_3_Y_axis	768	2,002
Bear_4_Y_axis	2,063	2,094

Note. Level=4

Table 49
DTCWT_b_lp_pe_d I

	Bearings / Axis	
	Bear_1_X_axis	Bear_2_X_axis
Bear_3_X_axis	2,154	2,144
Bear_4_X_axis	1,798	1,610
	Bear_1_Y_axis	Bear_2_Y_axis
Bear_3_Y_axis	2,154	0
Bear_4_Y_axis	2,154	2,150

Note. Level=3,4; Form=1,2

Table 50
DTCWT_q_lp_pe_d

	Bearings / Axis	
	Bear_1_X_axis	Bear_2_X_axis
Bear_3_X_axis	2,156	2,146
Bear_4_X_axis	1,791	1,612
	Bear_1_Y_axis	Bear_2_Y_axis
Bear_3_Y_axis	2,154	2
Bear_4_Y_axis	2,154	2,150

Note. Level=3,4; Form=1,2

Table 51
DTCWT_b_hp_pe_d II

	Bearings / Axis	
	Bear_1_X_axis	Bear_2_X_axis
Bear_3_X_axis	0	2,044
Bear_4_X_axis	0	1,944
	Bear_1_Y_axis	Bear_2_Y_axis
Bear_3_Y_axis	38	748
Bear_4_Y_axis	387	2,150

Note. Level=3,4; Form=1,2

Table 52
DTCWT_b_hp_pe_d

	Bearings / Axis	
	Bear_1_X_axis	Bear_2_X_axis
Bear_3_X_axis	0	2,044
Bear_4_X_axis	0	1,944
	Bear_1_Y_axis	Bear_2_Y_axis
Bear_3_Y_axis	38	748
Bear_4_Y_axis	387	2,150

Note. Level=3,4; Form=1,2

Table 53
Binary Classification

	Metrics				
Method	Loss	AUC	PRE	REC	Binary ACC
Raw_data_pe_d	0.6935	0.4913	0.4559	0.4559	0
1D SGF_pe_d	0.6934	0.4994	0.4931	0.4508	0
HT_m_pe_d	0.6934	0.6934	0.4945	0.4945	0
BF_lp_pe_d	0.6934	0.4957	0.4931	0.4351	0
FFT_m_pe_d	0.6934	0.4945	0.4959	0.5069	0
DTCWT_b_lp_pe_d[a]	0.6934	0.4945	0.4934	0.5087	0
DTCWT_b_hp_pe_d[a]	0.6933	0.4991	0.5005	0.4704	0
DTWT_q_lp_pe_d*	0.6934	0.4944	0.4951	0.5064	0
	Validation				
Raw_data_pe_d	0.6929	0.4890	0.5123	1	0
1D SGF_pe_d	0.6929	0.5000	0.5123	1	0
HT_m_pe_d	0.6930	0.5000	0.5123	1	0
BF_lp_pe_d	0.6929	0.4890	0.5123	1	0
FFT_m_pe_d	0.6930	0.4890	0.5123	1	0
DTCWT_b_lp_pe_d[a]	0.6929	0.5000	0.5123	1	0
DTCWT_b_hp_pe_d[a]	0.6929	0.5000	0.5123	1	0
DTWT_q_lp_pe_d*	0.6929	0.4890	0.5123	1	0

Notes. Binary Classification which distinguish the state in working and failure. Task includes 4 bearings, X and Y axes, 2156 instances, 100 neurons, 20 epochs, sigmoid activation, binary cross-entropy and stable pseudo-randomly shuffling; b denotes biort; q denotes qshift; pe denotes the Permutation Entropy Transform; d denotes log normalization; hp denotes high pass; lp denotes low pass; m denotes magnitude.

[a]Including 3 and 4 levels.

Table 54
Multi-class classification without custom indexing

	Metrics				
Method	Loss	Categorical ACC	AUC	PRE	REC
Raw_data_pe_d	0.9340	0.5416	0.7375	0.6031	0.3708
1D SGF_pe_d	0.9547	0.5361	0.7242	0.6004	0.3711
HT_m_pe_d	0.9223	0.5452	0.7438	0.6069	0.3777
BF_lp_pe_d	0.9685	0.5266	0.5921	0.3276	0.9635
FFT_m_pe_d	0.9376	0.5379	0.7343	0.6212	0.3490
DTCWT_b_lp_pe_d[a]	0.9104	0.5547	0.7538	0.6350	0.3929
DTCWT_b_hp_pe_d[a]	0.9160	0.5449	0.7467	0.6225	0.3747
DTWT_q_lp_pe_d*	0.9288	0.5390	0.7407	0.6185	0.3476
	Validation				
Raw_data_pe_d	0.9269	0.5466	0.7418	0.6036	0.3909
1D SGF_pe_d	0.9479	0.5412	0.7292	0.6030	0.3843
HT_m_pe_d	0.9132	0.5503	0.7487	0.6063	0.3862
BF_lp_pe_d	0.9635	0.5290	0.7156	0.6023	0.3148
FFT_m_pe_d	0.9274	0.5430	0.7408	0.6382	0.3286
DTCWT_b_lp_pe_d[a]	0.9012	0.5597	0.7591	0.6449	0.3964
DTCWT_b_hp_pe_d[a]	0.9042	0.5548	0.7547	0.6282	0.3940
DTWT_q_lp_pe_d*	0.9220	0.5438	0.7448	0.6313	0.3300

Notes. Muli-label classification which distinguish the state in working, failure 1 and failure 2. Task includes 4 bearings, X and Y axes, 413,952 instances, 100 neurons, 20 epochs, softmax activation, categorical cross-entropy and stable pseudo-randomly shuffling; b denotes biort; q denotes qshift; pe denotes the Permutation Entropy Transform; d denotes log normalization; hp denotes high pass; lp denotes low pass; m denotes magnitude.

[a]Including 3 and 4 levels.

Table 55
Multi-label classification with custom indexing

	Metrics				
Method	Loss	Categorical ACC	AUC	PRE	REC
Raw_data_pe_d	0.9340	0.5416	0.7375	0.6031	0.3708
1D SGF_pe_d	0.9547	0.5361	0.7242	0.6004	0.3711
HT_m_pe_d	0.9223	0.5452	0.7438	0.6069	0.3777
BF_lp_pe_d	0.9685	0.5266	0.5921	0.3276	0.9635
FFT_m_pe_d	0.9376	0.5379	0.7343	0.6212	0.3490
DTCWT_b_lp_pe_d[a]	0.9104	0.5547	0.7538	0.6350	0.3929
DTCWT_b_hp_pe_d[a]	0.9160	0.5449	0.7467	0.6225	0.3747
DTWT_q_lp_pe_d*	0.9288	0.5390	0.7407	0.6185	0.3476
	Validation				
Raw_data_pe_d	0.9269	0.5466	0.7418	0.6036	0.3909
1D SGF_pe_d	0.9479	0.5412	0.7292	0.6030	0.3843
HT_m_pe_d	0.9132	0.5503	0.7487	0.6063	0.3862
BF_lp_pe_d	0.9635	0.5290	0.7156	0.6023	0.3148
FFT_m_pe_d	0.9274	0.5430	0.7408	0.6382	0.3286
DTCWT_b_lp_pe_d[a]	0.9012	0.5597	0.7591	0.6449	0.3964
DTCWT_b_hp_pe_d[a]	0.9042	0.5548	0.7547	0.6282	0.3940
DTWT_q_lp_pe_d*	0.9220	0.5438	0.7448	0.6313	0.3300

Notes. Multi-label classification which distinguish the state in working, failure 1-2, time segments 1-4. Task includes 4 bearings, X and Y axes, 413,952 instances, 100 neurons, 20 epochs, softmax activation, categorical cross-entropy and stable pseudo-randomly shuffling; b denotes biort; q denotes qshift; pe denotes the Permutation Entropy Transform; d denotes log normalization; hp denotes high pass; lp denotes low pass; m denotes magnitude.

[a]Including 3 and 4 levels.

Table 56
Fine-tuned DTCWT_b_lp_pd_d

Metrics	
Loss	2.7846
ACC	0.9923
AUC	0.8535
PRE	1
REC	0.3026
Validation	
Loss	2.7733
ACC	1
AUC	0.8535
PRE	1
REC	0.4683

Table 57
Used Libraries

Bitstring 3.1.7
Deap 1.3.1
Keras 2.4.3
Mglearn : scikit-learn 0.21.3, pandas 0.25.1, matplotlib 3.1.1, scipy 2.0.0, pyparsing 2.4.2 e.t.c.
Networkx 2.3
Plotly 4.14.3
Pyentrp 0.3.1
Pywt 1.0.3
Scoop 0.7.1.1
Seaborn 0.9.0
Stasmodels 0.12.1

Table 58
System

MacOS Mojave
Version 10.14.6
MacBook Air (Retina, 13-inch, 2019)
Processor 1.6 GHz Intel Core i5
Memory 8 GB 2133 MHz LPDDR3
Graphics Intel UHD Graphics 617 1536 MB