A FRAMEWORK FOR DATA-DRIVEN FAULT DETECTION AND IDENTIFICATION WITH MULTI-SCALE KERNEL FISHER DISCRIMINANT ANALYSIS IN CHEMICAL PROCESS SYSTEMS

NORAZWAN BIN MD NOR

FACULTY OF ENGINEERING
UNIVERSITY OF MALAYA
KUALA LUMPUR

2018

A FRAMEWORK FOR DATA-DRIVEN FAULT DETECTION AND IDENTIFICATION WITH MULTI-SCALE KERNEL FISHER DISCRIMINANT ANALYSIS IN CHEMICAL PROCESS SYSTEMS

NORAZWAN BIN MD NOR

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF ENGINEERING
UNIVERSITY OF MALAYA
KUALA LUMPUR

2018

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Norazwan bin Md Nor

Matric No: KHA120099

Name of Degree: Doctor of Philosophy

Title of ~~Project Paper/Research Report/Dissertation~~/Thesis (“this Work”):

A FRAMEWORK FOR DATA-DRIVEN FAULT DETECTION AND IDENTIFICATION WITH MULTI-SCALE KERNEL FISHER DISCRIMINANT ANALYSIS IN CHEMICAL PROCESS SYSTEMS

Field of Study: Health, Safety and Environment (Chemical Process)

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya (“UM”), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate’s Signature Date:

Subscribed and solemnly declared before,

Witness’s Signature Date:

Name:

Designation:

A FRAMEWORK FOR DATA-DRIVEN FAULT DETECTION AND IDENTIFICATION WITH MULTI-SCALE KERNEL FISHER DISCRIMINANT ANALYSIS IN CHEMICAL PROCESS SYSTEMS

ABSTRACT

Fault detection and identification (FDI) framework plays an important role to ensure consistent and reliable operation of chemical process systems. The FDI framework has two main tasks, namely to detect the presence of a fault and to classify the location and type of the fault. In most cases, it is impractical to develop precise model from first principles as it requires the involvement of process complex physics and the interactions among the different components creating the process. Therefore, data-driven FDI methods, which can make use of process data to capture their trends and dynamics, provide an attractive alternative for the quick development and deployment of FDI solutions. One of the main objectives of this thesis is to develop a hybrid framework for data-driven FDI in chemical process systems. This framework integrated a novel multi-scale dimensional reduction method for pre-processing step and an improved data-driven FDI framework. This thesis focuses on proposing a dimensionality reduction method based on multi-scale kernel Fisher discriminant analysis (multi-scale KFDA), in which discrete wavelet transform (DWT) was combined with kernel Fisher discriminant analysis (KFDA) method. Initially, DWT was applied to extract the dynamics of the process at different scales. The wavelet coefficients obtained during the analysis were reconstructed using the inverse discrete wavelet transform (IDWT) method and then, they were fed into the KFDA to produce discriminant vectors. Finally, the discriminant vectors were used as inputs for the classification task in fault identification step. Apart from that, complete fault identification procedures based on adaptive neuro-fuzzy inference system (ANFIS), support vector machine (SVM), Gaussian mixture model (GMM), and k-nearest neighbor (*k*NN) were developed to investigate the parameters that could be optimised for better fault identification. Furthermore, this thesis extended the proposed multi-scale KFDA-based fault identification methods to a hybrid data-driven FDI framework. In the hybrid FDI framework, all classification methods used previously were combined into a single classification framework. Hence, a complete data-driven hybridisation FDI framework for chemical process systems was proposed and analysed. The proposed FDI frameworks were applied in three different chemical processes: the simulation of Tennessee Eastman process, the fed-batch penicillin fermentation process, and a real industrial data set of semiconductor etch process. Notably, the fault detection and classification results demonstrated the effectiveness of the proposed methods.

Keywords: fault detection and identification, data-driven, multi-scale KFDA, chemical process systems

SATU KERJA KERANGKA UNTUK PENGESAN KESALAHAN DAN PENGENALAN DIPACU DATA DENGAN ANALISIS DISKRIMINASI FISHER KERNEL BERBILANG SKALA DALAM SISTEM PROSES KIMIA

ABSTRAK

Kerja kerangka pengesan dan pengenalan kesalahan (FDI) memainkan peranan penting untuk memastikan operasi sistem proses kimia adalah konsisten dan dipercayai. Terdapat dua tugas utama di dalam kerja kerangka FDI; untuk mengesan kewujudan sesuatu kesalahan dan mengelas lokasi dan jenis kesalahan tersebut. Dalam kebanyakkan kes, adalah tidak praktikal untuk membangunkan model yang tepat berasaskan prinsip pertama kerana ia memerlukan penglibatan fizik proses yang kompleks dan juga interaksi di antara komponen berlainan yang membentuk proses tersebut. Oleh itu, kaedah FDI dipacu data mampu menggunakan data proses itu untuk penangkapan trend dan dinamikanya, menyediakan alternatif menarik untuk pembangunan dan penggunaan pantas untuk penyelesaian FDI. Objektif utama tesis ini adalah untuk membangunkan suatu kerja kerangka FDI hibrid untuk sistem proses kimia, menyepadukan kaedah baharu pengurangan dimensi untuk langkah pra-prosesan dan penambahbaikan kerja kerangka FDI dipacu data. Tesis ini berfokus kepada cadangan kaedah pengurangan dimensi berasaskan analisis diskriminasi Fisher kernel berbilang skala (multi-scale KFDA), dimana perubahan gelombang kecil diskrit (DWT) digabung bersama kaedah analisis diskriminan Fisher kernel (KFDA). Pada mulanya, DWT digunakan untuk mengekstrak dinamika proses tersebut pada skala yang berbeza. Pekali gelombang kecil yang diperolehi semasa analisis telah dibina semula menggunakan kaedah perubahan gelombang kecil songsang (IDWT), dimana kemudiannya ia disuap kepada KFDA untuk menghasilkan vektor diskriminan. Akhirnya, vektor diskriminan tersebut digunakan sebagai masukkan untuk tugas pengelasan dalam langkah pengenalan kesalahan. Dalam tesis ini juga, suatu prosedur lengkap bagi pengenalan kesalahan telah dibangunkan berasaskan kepada sistem inferens neuro-kabur ubah suai (ANFIS), mesin vektor sokongan (SVM), model campuran Gaussian (GMM) dan k-kejiranan terdekat (*k*NN) untuk menyiasat pelbagai parameter yang boleh dioptimakan untuk pengenalan kesalahan yang lebih baik. Tesis ini melanjutkan cadangan kaedah pengenalan kesalahan berasaskan KFDA berbilang skala untuk suatu kerja kerangka FDI dipacu data hibrid. Bagi kerja kerangka FDI hibrid, setiap kaedah pengelasan yang telah digunakan sebelum ini digabungkan menjadi pengelasan tunggal hibrid. Maka, suatu kerja kerangka FDI dipacu data hibrid untuk sistem proses kimia telah dicadang dan dikaji. Kerja kerangka FDI yang dicadangkan telah digunakan dalam tiga proses kimia yang berbeza; simulasi proses Tennessee Eastman dan proses fermentasi penisilin berkelompok suapan, dan satu set data industri sebenar. Keputusan penegesanan dan pengelasan kesalahan menunjukkan keberkesanan kaedah yang telah dicadangkan.

Kata kunci: pengesan dan pengenalan kesalahan, dipacu data, KFDA berbilang skala, sistem proses kimia

ACKNOWLEDGEMENTS

Subhanallah. Alhamdulillah. Allahuakbar. I would like to express my sincere appreciation and gratitude to my supervisors Prof. Dr Mohd Azlan Hussain and Assoc. Prof. Dr Che Rosmani Che Hassan for their invaluable recommendations in my research work and their constant amount of encouragement throughout my pursuit of a deeper level of knowledge. I am humbled to have worked with a very considerate professor that truly values the spirit of learning in research and further endeavours. I am grateful to my colleagues at the University of Malaya for their friendliness and support throughout my time in graduate school. Thanks to the Universiti Sains Malaysia (USM), especially School of Chemical Engineering, and the Ministry of Higher Education for their support in terms of SLAB-RLKA fellowship for this study. Finally, I would like to thank my wife, Nazira Anisa and both of our parents; Md Nor Md Ali, Rohani Saari, Rahim Jaafar, and Norshadah Ismail, for their love, supports, and encouragement. Last but not least, my adorable sons and daughter, Umar Al Faruq, Sumayyah, and Abdurrahman for their understanding and motivation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ANFIS	:	Adaptive neuro-fuzzy inference system
ANN	:	Artificial neural network
BN	:	Bayesian network
CWT	:	Continuous wavelet transform
DPLS	:	Discriminant partial least squares
DWPT	:	Discrete wavelet packet transform
DWT	:	Discrete wavelet transform
FDA	:	Fisher discriminant analysis
FDI	:	Fault detection and identification
GA	:	Genetic algorithm
GMM	:	Gaussian mixture model
ICA	:	Independent component analysis
IDWT	:	Inverse discrete wavelet transform
KFDA	:	Kernel Fisher discriminant analysis
KICA	:	Kernel independent component analysis
*k*NN	:	k-nearest neighbor
KPCA	:	Kernel principal component analysis
KPLS	:	Kernel partial least squares
LNGMM	:	Local and Nonlocal Gaussian Mixture Model
MLP	:	Multi-layer perceptron
MRA	:	Multi-resolution analysis
MSKFDA	:	Multi-scale kernel Fisher discriminant analysis
OWA	:	Ordered weighted averaging
PCA	:	Principal component analysis

PLS	:	Partial least squares
RBF	:	Radial basis function
SVDD	:	Support vector data description
SVM	:	Support vector machine
TEP	:	Tennessee Eastman process

CHAPTER 1: INTRODUCTION

1.1 Background

Effective fault detection and identification (FDI) of chemical processes is important to ensure consistency and high product quality, as well as the safety of these processes. Any abnormal process operation should be detected in the early stage to reduce the risk of public damage and large economic loss. Accordingly, the root cause of the process fault can be identified so that corrective action may be taken to move the plant back to its normal operating condition. Therefore, the process monitoring and FDI tasks are crucial as most of the modern chemical process systems are large-scale in nature, highly complicated, and complex with large numbers of variables (Ayoubi & Isermann, 1997). An appropriate action by FDI system could save up to billions of dollars in the process industries (Hong, Tian-You, Jin-Liang, & Martin, 2009).

Undoubtedly, various modelling methodologies have been developed for achieving efficient fault monitoring, detection, and identification in chemical process systems. These methodologies are broadly divided into three types: quantitative model-based methods, qualitative model-based methods, and process history-based methods (Venkatasubramanian, Rengaswamy, & Kavuri, 2003; Venkatasubramanian, Rengaswamy, Kavuri, & Yin, 2003; Venkatasubramanian, Rengaswamy, Yin, & Kavuri, 2003). Quantitative model-based methods such as the observer-based method and parameter estimation method utilise mathematical models constructed from the first principles for process monitoring. However, the effectiveness of this approach depends on the precision of the mathematical models constructed.

Meanwhile, qualitative model-based methods such as signed digraphs and fault tree analysis method employ cause-effect reasoning to describe the system behaviour.

However, this approach is restricted to systems with a relatively small number of variables or states because the creation of the knowledge base can be time-consuming.

On the other hand, process history-based methods, also known as data-driven methods, such as artificial neural network (ANN) and support vector machine (SVM) find patterns or compute meaningful statistics from the process historical data. This approach eliminates the use of detailed models for large-scale systems which can be expensive and difficult to develop. The availability of vast amounts of process history data has encouraged researchers to develop and improve the data-driven FDI methods, where one of the concerns is the transformation of massive amounts of data into a particular form of knowledge representation that will enable proper detection and identification of faults.

In this thesis, terminologies with the definition suggested by the Technical Committee SAFEPROCESS in Isermann and Ballé (1997) are used. Several terminologies that are frequently used in this thesis are defined below:

- Fault: Unpermitted deviation of at least one characteristic property or variable of the system.
- Fault detection: Determination of faults present in a system and time of detection.
- Fault identification: Identifying the variables that are most significant to the fault.
- Fault classification: A fault classification approach that classifies test data and determines the fault class of particular data.
- Monitoring: A continuous real-time task of determining the conditions of a physical system by recording information for recognising and indicating anomalies of the behaviour.

Therefore, the FDI system is a system that is used to determine and classify any undesirable deviation of a characteristic property of the process system.

1.2 Problem Statements

In general, faults in chemical process systems are common occurrences and may result in serious consequences. However, an effective FDI system can help in improving the system's reliability. For this reason, many opportunities exist for the theoretical and application contributions in the FDI research area, especially in chemical process systems. Therefore, in this thesis, various problems of FDI in chemical process systems based on the literature survey are addressed.

Firstly, feature extraction is a fundamental aspect to the pattern recognition and fault classification of FDI framework. In classification applications where raw data are collected from scientific experiments or simulations, the resulting data may have been exposed to noise, excited by external dynamics, or affected by uncertainties which are not accounted for during the data acquisition process. Furthermore, the data can also have high dimensionality, which can hinder the accuracy of decision algorithms if additional dimensions of the data make the patterns inseparable. Apart from that, they could also add the computational complexity burden to the FDI system to train the algorithms. Thus, the feature extraction method is needed to enhance the separability of patterns in the raw data to improve the accuracy of classification and to reduce the computational complexity in training and implementing the classifiers.

Secondly, almost all of the actual chemical processes are nonlinear with dynamic characteristics. Besides, the process data always have other characteristics such as non-Gaussian distribution, dynamic processes, time-varying processes, or multimode behaviours. In the literature, most of the state-of-the-art data-driven methods focused on dealing with these types of data characteristics. Generally, the independent component analysis (ICA)-based, Gaussian mixture model (GMM)-based, and support vector data description (SVDD)-based methods are suggested for non-Gaussian distribution problems, whereas self-organizing map (SOM)-based method is recommended for the nonlinear processes. On the contrary, Fisher discriminant analysis (FDA)-based method is recommended for time-varying processes, whereas the multi-way approach is suggested for batch processes. However, for systems with multiple process characteristics, the method that can be employed in the FDI framework could be complicated. Therefore, it is essential to have a general guideline in the method selection for the FDI system development based on the known process characteristics.

Thirdly, most processes are multi-scale in nature due to the occurrences of faults at different locations and localisations, both in terms of time and frequency. The solution for this non-stationary signal has been suggested through reducing the dimensionality of the data, where they are decomposed into various time scales to extract the appropriate information from the faulty process data. This approach called the multi-scale method was introduced with the intention to improve the performance of traditional single scale-based nonlinear dynamic methods for fault identification purpose. Therefore, general methods of pattern classification based on multi-scale local basis design need to be addressed, and the improvements obtained through the modification of the multi-scale approach should also be explored.

The reconstruction process usually will produce a large data set of featured attributes. Even though the improved wavelet analysis like discrete wavelet transform (DWT) has been proposed to replace the continuous wavelet transform (CWT) method, it is still difficult to accurately find the fault conditions with vision or classifier. This is because DWT is associated with a scaling function, where it creates a multi-resolution analysis (MRA) of the input database. Thus, the Parseval's theorem is proposed to improve the multi-scale analysis for the fault classification. The Parseval's theorem relates the energy of the faulty signal to the energy of the respective components and their wavelet coefficient. Hence, the Parseval's theorem could be efficient in extracting the significant feature at different resolution levels from the input signal based on the energy distribution features, and it can be divided into the approximate and detailed coefficients.

Nonetheless, problem could arise from the fact that most of the FDI method performs very well for one fault, but very poorly for another. Additionally, a lot of these methods may be sensitive to numerous faults, but the false alarm can also be very high. Moreover, all available fault classification methods have their drawbacks, although a number of methods may perform consistently well for all or nearly all types of faults. In addition to the drawbacks related to the classification techniques, most of the existing monitoring systems are application-specific. Thus, a systematic hybrid framework that is capable of automatically selecting the optimal combination of the detection and classification methods needs to be developed.

1.3 Research Objectives

The overall objective of this study is to investigate the integration applicability of data-driven methods to the developed multi-scale discrimination analysis in the FDI applications, especially in chemical process systems. In order to address this objective, multivariate statistical monitoring technique called kernel Fisher discriminant analysis (KFDA) was integrated with DWT analysis to produce multi-scale kernel Fisher discriminant analysis (MSKFDA) method. This method was proposed for the dimensional reduction and fault detection framework.

Several different classification methods such as adaptive neuro-fuzzy inference system (ANFIS), support vector machine (SVM), k-nearest neighbor (*k*NN), and GMM methods were also investigated and implemented to the multi-scale discriminant analysis method. The guidelines for the selection of these methods had been summarised based on the process characteristics. Finally, a data-driven FDI framework that is hybrid-based was proposed to achieve better performance compared to the established methods. In order to fill the gaps in the FDI research area, these four objectives were outlined:

i. To develop a novel multi-scale dimensional reduction method through the integration of KFDA with the DWT analysis.
ii. To evaluate the integration of novel multi-scale KFDA application for feature extraction with data-driven classifiers such as ANFIS, SVM, GMM, and *k*NN.
iii. To develop a guideline for method selection for data-driven FDI framework system, with respect to the process characteristics.
iv. To propose a hybrid data-driven FDI framework, combining multiple data-driven methods such as ANFIS, SVM, GMM, and *k*NN with multi-scale KFDA method for chemical process systems.

1.4 Research Significance and Scope of Work

This study provides an effective feature to detect faults that occur in chemical process systems by applying wavelet transformations, KFDA, and data-driven-based classification methods. It explores the state-of-the-art classification methods such as ANFIS, GMM, *k*NN, and SVM in order to complete the task of FDI.

The contributions of this study include the development of an intelligent multi-scale discrimination analysis, in which multi-scale KFDA method was applied to perform dimensional reduction and process monitoring for chemical process systems with complex and highly nonlinear structure. The contributions also include the development of a new and complete framework based on data-driven methods and the novel feature extraction method to perform the FDI framework. An integrated FDI framework was proposed by synthesising the above methods with other types of classification and pattern recognition methods.

A novel multi-scale KFDA-ANFIS FDI framework that is designed for classifying the faults had also been developed in this study. The contribution of this work is evident by providing a framework based on a complete structure of wavelet analysis and ANFIS for data-driven FDI. The proposed method is applicable to a wide range of chemical process systems as both simulated and real-world data had been utilised to test the performance of the framework designed. Furthermore, a new FDI framework based on GMM, *k*NN, and SVM had also been proposed and successfully applied to detect and diagnose faults in chemical process systems. The performance of these methods had been demonstrated on the Tennessee Eastman plant and other processes, as well as compared to other conventional methods.

Another aspect that is widely explored in this research area is the application of hybrid methods of computational intelligence techniques to solve a single problem. They are often used to solve complex problems, in particular, one technique is typically used to fix the weaknesses of the other. Thus, the study also developed a new hybrid-based framework of multi-scale data-driven methods.

1.5 Thesis Structure

Chapter 1 presents the introduction of the FDI methods with their background study. The problem statement, research objectives, the significance of the study, and the thesis structure are also presented. In Chapter 2, the characteristics and problems in FDI of chemical process systems are reviewed. The selected data-driven FDI methods, namely ANFIS, SVM, GMM, and *k*NN, are also reviewed with discussion on their applications and current issues. In Chapter 3, multi-scale KFDA for feature extraction is proposed and the approach combines the DWT with the KFDA. Meanwhile in Chapter 4, multiple new FDI schemes are proposed based on the combination of multi-scale KFDA and data-driven classification methods such as ANFIS, SVM, *k*NN, and GMM. In Chapter 5, the hybridisation of data-driven classification methods with multi-scale KFDA method is proposed. The effectiveness of the proposed methods is illustrated by three case studies: the fed-batch penicillin fermentation process, the etch process, and the Tennessee Eastman process. Finally, Chapter 6 concludes the thesis with few recommendations for future works.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

In this chapter, a comprehensive review of data-driven fault detection and identification methods in chemical process systems is presented. The review starts with the introduction of general characteristics of chemical processes, followed by problems commonly occurred during the detection and identification of these chemical process systems. After that, this chapter reviews the application of multi-scale feature extraction methods, with focus on chemical process systems, followed by the introduction of data-driven FDI methods, such as ANFIS, SVM, GMM, and kNN. Finally, a summary of the hybrid data-driven methods in FDI systems is presented, with attention to their applications in chemical process systems.

2.2 Chemical process systems

2.2.1 The characteristics of chemical process systems

Two different types of process systems, batch and continuous processes always have their own characteristics and difficulties, especially for process monitoring and fault identification. Continuous process, the most common type of industrial process, has been widely applied in chemical, petrochemical, and metallurgical industries. By operating in a continuous way, the process should be around the optimal state for most of the time to produce constant outputs. The process data need to be collected under the stationary condition, where it runs under a stable operating condition. In contrast, batch process has a finite operation duration with tight process specifications. The set-point of the batch process also changes frequently to accommodate different process conditions, so that various grades of products can be produced in a single batch process.

Batch processes have many different characteristics and variable correlations, coming from multiple process phases. It is more difficult to establish a monitoring model for batch processes. A three-dimensional data structure is a significant characteristics of batch processes. Multivariate statistical techniques, such as multi-way PCA and multi-way PLS have been widely applied in batch process analysis and monitoring to handle these problems. But these methods cannot reveal the changes of process correlations along the time direction because they take the entire batch data as a single object (Zhang, Zhao, Wang, & Wang, 2017). In general, there are three main issues need to be addressed for process monitoring in batch processes; multiple phases, transition characteristics from phase to phase, and uneven-length problems.

Therefore, many methods have been established for multiphase batch processes, i.e. multi-block or phase-separated methods. Multi-block methods use a single model with data grouped in several blocks to extract the relations between complex process variables and reflect local behaviours of a process. Although multi-block methods have been successfully applied in many industrial areas, most of these traditional multi-block methods adopt the concept that prior process knowledge is known, even though it is unavailable in many industrial processes. This has contributed to the limitation of the multi-block method applications. Besides, many batch processes may have significant transition behaviours from phase to phase, while most of the multi-block methods have not considered well the transitions between any two adjacent phases.

Phase-based methods construct a separate statistical model for each phase of a batch process. Since the prior process knowledge is hard to obtain in many processes, an important issue for phase-separated methods is how to divide the batch process into different phases. According to the fact that the changes of the process correlations may relate to the phase shift in multiphase batch processes, many clustering-based phase division methods have been well developed. However, most of these methods assume that each batch has the same and equivalent duration and the sampling intervals.

In fact, uneven-length problems widely exist in batch processes. The length and the intervals may vary from batch to batch in terms of practical operational conditions (Zhang et al., 2017). To solve this problem, many methods have been developed in the past years. The GMM is a common method which can tackle the uneven problem in batch process monitoring. However, the duration time of transitional phases is much shorter than the stable time, the information contained in transitional phases may be overwhelmed when all the historical data are used to train a GMM. Furthermore, GMM may not efficiently capture the local features of transitions and do not consider the phase information.

The nonlinear relationships among different process variables are also very common in the process industry, as well as between the process and quality variables. While the linear relationship of the data can be easily captured by the traditional multivariate statistical analysis method, the data non-linearity is difficult to model. For instance, PCA-based method which assumes linear relationships between the Gaussian latent variables can be inefficient when dealing with highly nonlinear process that have non-Gaussian underlying variables (Slišković et al. 2012). Moreover, different processes may have quite different nonlinear relationships among process variables. Therefore, different types of nonlinear models have been developed to handle this problem, such as kernel-based methods and neural network-based methods.

In real practice, not all of the process variables are Gaussian-distributed, where they may follow different types of non-Gaussian distributions due to the non-Gaussian noise, feedback control systems, and different data transformations. For Gaussian processes, the latent variables have been assumed to follow the Gaussian-distribution, where traditional PCA- and PLS-based methods can be applied to develop the T^2 and SPE statistics together with their control limits. Otherwise, the control limits may be inaccurate, and the model unable to represent the boundary of normal operation region of the process, leading to false alarms. Therefore, for non-Gaussian processes, several types of methods have been proposed to improve the process monitoring and FDI performance, such as ICA, GMM and SVDD.

What distinguished ICA from PCA is that it looks for components that are both statistically independent and non-Gaussian (Fan et al. 2014). While PCA can only impose independence up to second-order statistics information, ICA involves higher-order statistics of the data. Therefore, ICA may reveal more meaningful information in the non-Gaussian data than PCA (Slišković et al., 2012). Although ICA has demonstrated its effectiveness in non-Gaussian process monitoring, it is a linear statistical method, requiring the assumption that the process data have linear structure. However, in industrial environments, the collected process data are usually nonlinear. Therefore, ICA may fail to conduct effective and adequate feature extraction from nonlinear process data, which may lead to unsatisfactory monitoring performance (Cai, Tian, & Chen, 2017).

Kernel ICA (KICA) is introduced to tackle the nonlinear process monitoring problem, which is better than that of ICA. Essentially, KICA is an integration of KPCA with ICA. It can be seen that KICA has been utilized as an effective means for nonlinear and non-Gaussian process monitoring. However, the current KICA-based monitoring methods seldom investigate the significance of different KICs to process monitoring (Cai et al., 2017). In addition, other methods that are also able to handle the non-Gaussian data information have been developed, such as support vector machine (Hsu, Chen, & Chen, 2010) and multi-scale statistical modelling methods, such as multi-scale

PCA (Lau et al., 2013), multi-scale PLS (Lee et al., 2009), and multi-scale FDA (Deypir, Boostani, & Zoughi, 2012).

In the scenario of big data, algorithm efficiency will be more important. For the case of non-Gaussian distribution, the volume of data set will increase the accuracy but not too much due to the over-fitting effect. The processing speed will decrease a lot for KDE and SVDD methods, because they are both nonparametric models, which makes the computation load proportional to the number of training samples. However, the performance of linear algorithm and parametric model will be enhanced a lot and the speed will not be lowered too much (Li & Qin, 2016).

The real process may also be varying from time to time, or switched from one operation to another. For example, the fluctuations in raw materials, set-point changes, aging of equipment, and seasoning effect, the operation conditions of industrial processes change frequently in practice. In order to keep the industrial process under control, the process monitoring and FDI method should be updated according to the change of the operation conditions. On contrary, when the process consists of several different operation conditions, the multi-mode process monitoring method should be applied. Thus, different methods have been implemented, such as adaptive learning techniques and multi-model methods. For processes that are slow-varying or have multiple operating conditions, it is challenging to apply the traditional statistical methods, since they are based on the assumption that the process has only one stable operating region. Therefore, problems will arise when those techniques are applied to varying processes. When the process is slow-varying, many adaptive and recursive monitoring approaches have been developed, such as recursive PLS (Li et al., 2005) and recursive PCA (Zhang, Li, & Teng, 2012). Then, moving-window approaches like moving-window PCA were developed to improve the monitoring efficiency of time-varying processes. Furthermore, to monitor nonlinear time-varying processes, a moving window approach was developed when the kernel PCA and kernel ICA model were used (Jiang & Yan, 2013b; Liu et al., 2009).

The dynamical characteristic also presented in process variables due to the influence of random noise and process disturbance. Due to the dynamic data behaviour, the data sample acquired at the present time may be correlated with those samples before and after present time. For different processes and operation conditions, the dynamic steps may be different from each other. Thus, changes of dynamic relationships among process variables cannot be efficiently featured, which may cause a severe fault to the process. Because of the dynamic relationships, the fault may also be biased by random noises and other disturbances. In order to improve the monitoring performance for dynamic processes, dynamic multivariate statistical analysis methods such as dynamic PCA-based and dynamic ICA-based methods have been proposed. Besides, the process data may also show multi-scale characteristic, inconsistent sample rates, measurement delay, etc.

2.2.2 The problems of fault detection and identification

The first problem is regarding a high-dimensional feature vector. Modern chemical processes always consist of various components, parts, or operation equipment, and each of these parts may have significant numbers of input and measurement variables (Zhao, Liu, Dong, Sun, & Ji, 2017). As a result, the entire process may generate a large number of high-dimensional data samples. To handle these high-dimensional process data is a challenge for the data-driven approaches (Ge, Song, & Gao, 2013). Thus, in order to achieve optimal results, the irrelevant and redundant information hidden in the feature values have to be filtered out, often been done by feature selection or feature extraction. Furthermore, the dimensionality reduction of the feature extraction can be a key factor in reducing the misclassification rate when a pattern classification system is applied to fault identification. Moreover, it is important especially when the dimensionality of the observation space is large while the numbers of observations in the classes are relatively small.

Many multivariate statistical methods have been proposed by researchers for dimensional reduction, such as PCA, PLS, ICA, and FDA-based methods. For instance, the commonly used feature extraction tools are the PCA and its nonlinear version known as kernel PCA (KPCA), which represent the current state-of-the-arts in extracting linear and nonlinear features, respectively and they have been successfully applied to industrial process monitoring and fault detection (Deng, Tian, Chen, & Harris, 2017a). Meanwhile, the filter-based feature selection methods could be developed from F-statistics, information gain, and mutual information, whereas the wrapper-based feature selection includes genetic algorithm and ANN methods. For example, PCA is employed for feature extraction, where the visualisation of high-dimensional feature vectors is mostly done with the first two values of the PCs of the original feature vector. However, the PCA methods, including nonlinear PCA and ICA, extract latent variables from process measurements under feedback control and monitor changes in the process condition, actuators, sensors, and disturbances, which could lead to a poor quality product. Different to the traditional PCA and KPCA methods, which only monitor one layer of linear or nonlinear features, Deng et al. (2017a) proposed a method that integrates multiple layer-wise linear and nonlinear features.

Meanwhile, PLS including nonlinear PLS uses quality data to guide the decomposition of the process data and extract latent variables, which are most relevant to the product quality. On the other hand, if categorical quality data for multiple fault cases are available, the discriminant partial least squares (DPLS) method provide an alternative to diagnose quality-relevant faults. Based on the reduced variable space, process monitoring can be carried out more easily (Ghosh, Ramteke, & Srinivasan, 2014). With deep model structure, raw data are transformed into low level of data features, which are further used to compute higher level of data representations. Deep learning has turned out to be very helpful to discover the intricate information (Deng et al., 2017a). Typical deep learning methods include convolutional neural network (CNN), deep belief network (DBN), and deep auto-encoder network (DEN).

In addition, multi-scale monitoring usually applies the signal pre-processing method in conjunction with PCA and PLS method to further extract features in terms of time scales so that each time scale can be sensitive to certain faults for fault detection and identification. The multiple scale analysis provides the ability to de-noise or filter the data so that uninterested time scale variations can be ignored. With the development of signal processing techniques, many features can be extracted to find their feature patterns and connections to faults, which will provide useful information for fault detection and identification (Dai & Gao, 2013). The features extracted from the output signals can be in time and/or frequency domains, such as signal mean, variance, trends, or spectra in a frequency band. Various signal analysis techniques have been introduced, i.e. fast Fourier transform (FFT), Short Time Fourier Transform (STFT), spectral estimation, wavelet transform, and sequence analysis (Agrawal, Panigrahi, & Subbarao, 2014). For instance, continuous wavelet transform (CWT) method allows the acquisition of multi-scale resolutions, while discrete wavelet transform (DWT) has computation efficiency and ability to reduce noise in raw signals. Discrete wavelet packet transform (DWPT) method was also used to enhance the power and flexibility of the DWT, with various adaptive methods developed for the selection of optimal basis wavelets. Although these three signal-based methods are able to work individually for the FDI system, recently there have been many works that combine these methods. For example, DWT analysis is combined with intelligent classification techniques such as artificial neural network and neuro-fuzzy system, to identify the faulty signals.

For the nonlinearity behaviour of the processes, most of the nonlinear methods for process monitoring have resulted in better performance compared to the linear methods. A related category of nonlinear data-driven methods suitable for process monitoring is from the machine learning methods, including ANN and SVM. Although the linear relationship of the data can be easily captured by the traditional multivariate statistical methods, the nonlinear data is difficult to model. For instance, a PCA-based method which assumes linear relationships between variables and Gaussian latent variables, thus, it can be inefficient when dealing with highly nonlinear and have non-Gaussian underlying variables (Slišković, Grbić, & Hocenski, 2012). Therefore, kernel learning method has been combined with some traditional methods such as PCA and PLS for nonlinear process monitoring. For example, kernel PLS (KPLS) can efficiently compute regression coefficients in high-dimensional Gaussian feature spaces using nonlinear kernel functions. It can also handle a wide Gaussian range of nonlinearities due to its ability to use different kinds of kernels (Zhang, Teng, & Zhang, 2010). Another example is a kernel FDA (KFDA), which is used instead of the traditional FDA to eliminate the weakness towards the nonlinear processes (Ferreira & Trierweiler, 2009).

Stochastic faults, also known as random faults are triggered with the random variation in the process operating conditions. The process operating state changes with the state is sensitive to the changes of input materials, process fouling, catalyst activity changes, production of different product quality grades and changes in external environment, for example, the disturbances and variability in raw material and operation. This malfunction can be abrupt or gradual degradation of performance, with different rapidity and severity of the fault. Due to the dynamic data behaviour, the data sample obtained at the present time may be correlated with those samples before and after the present time. For different processes and operation conditions, the dynamic steps may be different from each other. In order to improve the monitoring performance for dynamic processes, dynamic multivariate statistical analysis methods such as dynamic PCA-based and dynamic ICA-based methods have been suggested to be applied in the FDI system. A Bayesian network (BN) is also an excellent tool to characterize processes with stochastic uncertainty using conditional probability-based state transitions (Gharahbagheri, Imtiaz, & Khan, 2017a). Researchers have also used Bayesian network for improving process fault identification in different ways. Due to the stochastic nature of process variation, a false alarm may be generated in monitoring a system while the process is operating in normal condition.

Also, it is usual in practice that the normal process data experience slow but normal drift. However, this drift will cause false alarm if not adapted to normal models. For those processes that are slow-varying or have multiple operating conditions, it is difficult to apply the traditional statistical methods, since they are based on the assumption that the process has only one stable operating region. Therefore, problems will arise when those techniques are applied to varying processes. When the process is slow-varying, many adaptive and recursive monitoring approaches have been developed, such as recursive PLS (Li et al., 2005) and recursive PCA (Zhang, Li, & Teng, 2012). Then, moving-window approaches like moving-window PCA, moving-window KPCA, and moving-window KICA were developed to improve the monitoring efficiency of time-varying processes (Jiang & Yan, 2013b; Liu et al., 2009).

For instrumental faults such as sticking valve in reactor cooling water valve, it is more difficult to diagnose. For this fault, the transient values of the processes must be taken into account, where the sensitive threshold for the adaptive PCA method is suggested for fault detection. The variance sensitive adaptive threshold is suggested to overcome false alarm which occurs in the transient states according to changing process condition (Alkaya & Eker, 2011). Meanwhile, incipient faults describe the wear and ageing of system components and thus relatively difficult to handle due to the slowly developing nature of these faults. Most of the machine learning methods, including neuro-fuzzy, SVM, and GMM methods have successfully been applied for the cases of incipient fault identification. In addition, the signal processing approaches such as wavelet transform method is also been suggested to improve the FDI system, with the detection of incipient faults in chemical process systems.

Another major problem in chemical processes is the high level of complexity. The complexity of a plant could be understood as the difficulty to model its global emergent behaviour using single modelling techniques. A complex system can be represented by a system whose global behaviour emerges from the interactions between its large numbers of basic components and is difficult to represent analytically. Therefore, a distributed fault identification methodologies have been suggested, where the monitored system was divided into subsystems, having a reasonable complexity level, and applying the state-of-the-art FDI methods, on each one of the subsystems based on their problems. By implementing this strategy, the FDI method retains their ability to treat the local nonlinearities, noise, and uncertainty.

2.3 Feature extraction methods

Data-driven methods typically need pre-processing steps to help in improving its ability to classify the patterns in the process data. These pre-processing steps, including signal conditioning, input normalization, and feature extraction, are needed for denoising, to make sure the input values are within the range of the capabilities of the algorithms, and for enhancement of the separability classes, respectively. Undoubtedly, an efficient extraction of the most discriminant information elements, followed by a suitable combination of this information would lead to a successful pattern recognition scheme.

2.3.1 Multivariate statistical feature extraction methods

Multivariate statistical methods such as PCA, PLS, FDA and ICA have been extensively researched for extracting process information from massive data in various field. For example, PCA-based feature extraction method, which is combined with T^2 and Q monitoring charts together with the contribution plots has been developed and widely used in many processes. The PCA method determines the lower dimensional representation of data by capturing the data directions that have the most variance. However, the conventional PCA is linear transformation method in nature, while most of industrial processes database have nonlinear characteristics and dynamical properties. Therefore, the conventional PCA may lead to false-alarm when applying to process monitoring (Zhang, Li, & Hu, 2012).

Contrary to the prominent advances in the PCA algorithms, the PLS algorithm has received surprisingly little attention in the field of process monitoring despite the fact that the PLS method may be a better choice for the monitoring of industrial processes especially when the correlation between the process and quality variables should be emphasized. As a consequence of nonlinear processes, a kernel PCA (KPCA) has used kernel functions to complete nonlinear transformation for nonlinear process monitoring, where it showed better performance in most industrial applications. Moreover, kernel PLS (KPLS) provides good monitoring performance by finding those latent variables that present a nonlinear correlation with the response variables and at the same time improves model understanding. The two scaled matrices in KPLS are also decomposed in feature space, respectively (Zhang & Ma, 2011).

Meanwhile, kernel FDA (KFDA) can extract a nonlinear relationship among variables, while it can take the classification information into account. With proper parameters, KFDA may project the variables along some discriminative directions, and thus KFDA has the potential advantage for better feature extraction (Liu et al., 2010). KFDA is a more generalized FDA, which uses the kernel trick in the same manner as SVM in order to solve a linearly separable problem in a nonlinearly transformed higher dimensional space (Liu et al., 2010).

Although nonlinear methods have been successfully applied, there are usually restricted to analyse the events for only a single static scale corresponding to the sampling frequency. Data from almost all practical processes are inherently dynamical multi-scale due to events occurring at different locations and with different localization in time and frequency. Thus, the data need to be represented at several different dynamical scales.

2.3.2 Multi-scale feature extraction methods

Wavelet analysis technique has grown significantly, having advanced with the progress of digital signal-processing algorithms and technology. It provides a multi-resolution representation of signals with different time-frequency scales. Using dyadic wavelet function into the basis, the wavelet transform connects with the sub-band filter bank by multi-resolution analysis (MRA), and then, using the filter will carry out DWT. The original signal can be decomposed into several signals with different scales or resolutions and can reconstruct the signals using IDWT. The MRA can examine the signal at different time windows and frequency bands, and presented it as two important foundations of the algorithm, i.e. scaling function and wavelet function. Moreover, these decomposed signal will not change their information in the time-domain.

To consider this multi-scale nature in process monitoring, researchers have employed wavelet analysis to transform time-domain signals into the time-frequency domain, such as in Bakshi (1998) and Misra, Yue, Qin, & Ling (2002). Among the wavelet transformation analysis used are wavelet transform (Zhang & Ma, 2011), stationary wavelet transform (Liu et al., 2010), discrete wavelet transform (Choi, Martin, Morris, & Lee, 2007; Lau, Ghosh, Hussain, & Che Hassan, 2013; Wang & Romagnoli, 2005), and discrete fast wavelet transform (Fourie, De Vaal, & Vaal, 2000).

The advantage of the multi-scale methods is the dynamical data will be analysed at different scales, where the effects of the noise in the original data can be eliminated. However, conventional basis selection based on compression- and approximation-based criteria are not the best approaches for classification and segmentation purposes in some cases. Therefore, a number of approaches where multivariate statistical analysis method is combined with wavelet analysis have been proposed, aimed at enhancing process performance monitoring capabilities.

There are many different wavelet-based methods reported in the literature for feature extraction such as short-time Fourier transform, time-average analysis, Wigner-Ville Transform, resonance demodulation technique, CWT, DWT and discrete fast wavelet transform (Härdle, 2011). These analytic methods usually extract the vibration, sound emission, or other signal components of different phenomena as input signals for fault detection (Wu, Hsu, & Wu, 2009). In above mentioned methods, the DWT analysis provides an efficient method for extracting the time-frequency features. Nevertheless, the CWT needs high computational time and amount of data. Contrary, the DWT does not have these flaws. The DWT allows a systematic decomposition of a signal into its sub-band levels which can be performed with minimum bias of the signal even for stationary signal analysis.

The CWT techniques is using different scales to analyse the signal. The signal must be calculated to obtain the wavelet coefficients with every possible scale. For this reason, the signal will generate a large quantity of computation wavelet coefficients. Therefore, the dyadic scale and translation is adopted to reduce computation wavelet coefficients and computation time. The dyadic method called DWT, where it can be represented with filters concept as a complementary filters. The complementary filters contain of a high-pass filter and a low-pass filter, which is obtained high frequency and low-frequency wavelet coefficients, respectively. In the study by Wu and Hsu (2009), a DWT technique is used as an extraction method of the fault feature, while the technique of intelligent classification is based on fuzzy logic inference.

To further diminish the dimensionality of the extracted feature vectors and enhance the model performance, the statistics over the set of the wavelet coefficients were used. For instance, the PCA is combined with wavelet analysis for multi-scale process performance monitoring. Generally, it was called multi-scale PCA (MSPCA) method, where several variants of MSPCA-based monitoring have also been suggested in the literature. Among the works that are based on these wavelet analyses together with multivariate statistical methods are, multi-scale PCA (MSPCA) (Fourie & De Vaal, 2000; Lau et al., 2013; Misra et al., 2002; Wang & Romagnoli, 2005), robust multi-scale PCA (RMSPCA) (Wang & Romagnoli, 2005), multi-scale KPCA (MSKPCA) (Choi et al., 2007; Choi, Morris, & Lee, 2008; Maulud, Wang, & Romagnoli, 2006), multi-scale PLS (MSPLS) (Lee, Lee, & Park, 2009), multi-scale KPLS (MSKPLS) (Zhang & Hu, 2011; Zhang & Ma, 2011), and multi-scale KFDA (Md Nor, Hussain, & Che Hassan, 2015).

Moreover, as a tool for fault identification, the contribution of measured variables to monitoring statistics in MSKPCA have also been derived by calculating the derivative of a monitoring statistic with respect to each variable (Choi et al., 2008). However, the MSPCA and MSPLS have a linearity assumption, thus, limiting their application (Zhang & Ma, 2011). Therefore, to solve the nonlinear problem of observed data, MSKPCA and MSKPLS approaches are proposed. For example, Zhang and Ma (2011) proposed MSKPCA and MSKPLS approaches together with an algorithm of variable contributions to monitoring statistics for fault identification in nonlinear multi-scale processes, while the multi-scale nonlinear PCA proposed by Fourie et al. (2000) is based on multilevel wavelet decomposition and nonlinear PCA via an input-training neural network. The latter method combines the ability of nonlinear PCA with wavelet analysis to extract deterministic features and approximately decorrelating auto-correlated measurements. Another variant, called SFM-MSKPCA proposed by Zhang, Li, & Hu (2012) is based on the combination of the Kronecker production to build the dynamic model, the wavelet decomposition technique, the sliding median filter technique to remove the disturbances and noises, and KPCA, where it gives nonlinear dynamic interpretation.

Additionally, multi-block approaches can also reduce the complexity of process analysis and monitor processes in a hierarchical manner. In general, the low frequency components corresponds to background and the high frequency components correspond to noise, whereas the middle frequency components correspond to useful information. Thus, by incorporating the wavelet decomposition technique and modified KPLS into the framework of the multi-block PLS, the MSKPLS algorithm, as proposed by Zhang & Hu (2011) could be effectively used for the process monitoring. On the contrary, Alawi & Julian Morris (2007) compared a multi-scale multi-block modelling approach with conventional multi-way PCA approach for batch process of penicillin fermentation simulation.

On the other hand, several nonlinear parameter estimation techniques were also applied and combined with multi-scale methods to extract fault-related features hidden in the measured signals. For example, the methods based on approximate entropy and multi-scale entropy (MSE) were used for fault identification. Although these entropy-based methods are simple and require much less computation time, they have very good performance in fault identification. Other works in the literature explore extensions or alternatives of wavelet analysis, such as empirical mode decomposition (EMD) (Liu, Cao, Chen, He, & Shen, 2013; Zamanian & Ohadi, 2011), ensemble EMD (EEMD) (Xue & Zhou, 2016), multi-scale entropy (MSE) (Zhang, Xiong, Liu, Zou, & Guo, 2010), and multi-scale permutation entropy (MPE) (Tiwari, Gupta, & Kankar, 2015; Wu, Wu, Wu, Ding, & Wang, 2012) to calculate entropy over multiple scales for fault identification.

However, the corresponding computation time would increase rapidly if more levels, such as five or more, are to be considered. This would inevitably result in an exponential increase in the number of possible scale combinations. Therefore, a more efficient algorithm is needed to select desirable scales or scale combination and to choose optimal parameters for the classifiers with or without dimension reduction (Liu et al., 2010). Another concerning issues about the multi-scale classification is its generalization ability, where the current method has limitations for systems with time-dependent characteristics. Despite these issues, these wavelet analyses were also combined with other constructed classifiers such neural networks (Dash, Nayak, Senapati, & Lee, 2007), SVM (Md Nor, Hussain, & Che Hassan, 2017a), Bayes classifier (Liu et al., 2010), GMM (Md Nor, Hussain, & Che Hassan, 2017b), and ANFIS networks (Lau et al., 2013). Therefore, further study such as constructing discriminative scale features and discriminant analysis is required for a better classification (Reis and Bauer, 2009).

2.4 Data-driven FDI methods

This section tries to provide an overview of recently efforts in the development and application of data-driven based fault detection and identification system, specifically for chemical processes.

2.4.1 Adaptive Neuro-Fuzzy Inference Systems (ANFIS)

The ANFIS is a Takagi-Sukeno fuzzy model of integration where the final fuzzy inference system is optimized via the ANNs training. It maps input through input membership functions and associated parameters, and then through output membership functions to outputs. The initial membership functions and rules for the fuzzy inference system can be designed by employing human expertise about the target system to be modelled. Then ANFIS can refine the fuzzy if-then rules and membership functions to describe the input/output behaviour of a complex system.

Commonly used identification approach like neural network could achieve good performance in certain situation because of its strong self-learning ability. However, the poor interpretability for “black box” modelling and knowledge completeness limit the usage of this method (Zhao et al., 2017). Therefore, adaptive neuro-fuzzy inference system (ANFIS) is utilized for the decision making method. This system has the merits of fuzzy logic, capable of mimicking the way human beings reason, and artificial neural networks with the adaptive learning ability. Thus, the results are more efficient than the results when either neural network or fuzzy logic method is applied alone.

Neuro-fuzzy-based fault detection method was utilized by Evsukoff & Gentil (2005), Mok & Chan (2008), and Ruiz, Nougues, & Puigjaner (2001) in their research on fluidized bed coal gasifier, simulated nuclear reactor, and two-tank water level control system. It has also applied by Blázquez et al., (2006), Lau, Heng, Hussain, & Mohamad Nor (2010), and Sabura Banu & Uma (2011).

Furthermore, the ANFIS-based fault detection method was also combined with wavelet analysis (Chen, Wang, & McGreavy, 1998; Li, Mei, Zhou, Tang, & Huang, 2006), wavelet analysis with PCA method (Lau et al., 2013) and neuro-fuzzy observer (Uppal, Patton, & Witczak, 2006) to gain more advantages during the monitoring and fault detection process. The ANFIS-based FDI method was also been utilized with the combination of different types of fault detection methods like Kalman filter optimized with genetic algorithm (Khalid & Akram, 2011; Khoukhi, Khalid, Doraiswami, & Cheded, 2012) and fuzzy inference system (FIS) with multi-scale PCA (Kiong, Rosmani, & Hassan, 2010).

Unlike its application for modelling, when used for pattern classification, ANFIS takes superior features that contain rich faulty information as its inputs instead of those who can mostly represent the system. Since it is reliable and robust in real-time fault classification, it is very suitable for nonlinear systems which may have noisy data measurements, multi-faults and incomplete human expertise. When appropriate features corresponding to objective faults have been found, ANFIS can also be trained and validated for fault classification, where it also could be an alternative promising technique for future fault identification of chemical process system.

2.4.2 Support Vector Machine (SVM)

The SVM is a set of related supervised learning methods that analyze data and recognize patterns, and it is used for the classification and regression analyses. The original SVM algorithm was invented by Vapnik and the current standard softmargin was proposed by Cortes and Vapnik, where the basic SVM deals with only the binary classification. The SVM uses two main ideas. First, kernel functions are used to transform the problem from the original input space into a highly dimensional one, called the feature space, where linear separation of training samples belonging to different classes is possible. Second, to find the best separating hyperplane, the concept of maximum margin is introduced. Finally, the optimization problem which defines the SVM is convex and quadratic, and therefore it can be solved efficiently.

In recent years several different and modified SVM methods have been proposed to be used, since the fault identification is usually a multi-class classification problems. Among them, a combination method named one-against-one SVM and one-against-all SVM have been used by Jing and Hou, and Yin and Hou for fault identification in a chemical process system. In their recent work, Yin and Hou also shared that the combination of SVM and other approaches usually performs better than SVM alone. Among the SVM-based fault detection work that has been applied are Jing & Hou (2015); Kulkarni, Jayaraman, & Kulkarni (2005); Xiao, Wang, Zhang, & Xu

(2014); Yélamos, Escudero, & Graells (2006, 2007); Yélamos, Escudero, Graells, & Puigjaner (2009); and Yelamos, Graells, & Puigjaner (2007) in their respective research areas.

Furthermore, SVM-based fault detection method have also been combined by a few methods of pre-processing and dimensional reduction like nonlinear kernel function (Yin, Gao, Karimi, & Zhu, 2014), genetic algorithm (Han, Gu, Kang, & Li, 2011), genetic algorithms with contribution charts together with proximal SVM (Chiang, Kotanchek, & Kordon, 2004), wavelet analysis (Liu et al., 2010), ICA (Bo, Qiao, Zhang, Bai, & Zhang, 2010; Hsu, Chen, & Chen, 2010), time structure ICA (TSICA) (Cai, Tian, & Zhang, 2015), RFE (Mahadevan & Shah, 2009), correlation analysis (Chen, Chen, Chen, & Lee, 2011), contribution chart (Tafazzoli & Saif, 2009), KPCA with KICA (Zhang, 2009) and non-negative matrix factorization (Peng, Chen, Ye, & Jiao, 2014).

The modified version of SVM algorithms; called SVC-based framework also was utilized in fault detection system (Souza, Granzotto, Almeida, & Oliveira-lopes, 2014). Moreover, the SVM-based classification method was also modified, such as SVC and combined with optimization methods like grid-search method, genetic algorithm and PSO (Yang, Zhou, Wang, & Pan, 2015). A few applications on SVM-based FDI method were published with different types of detection and pre-processing methods like fuzzy rules with recursive feature elimination (RFE) (Yong, Zheng, Zheng, Youxian, & Zheng, 2007) and GMM with BIC and ICA combination (Monroy, Benitez, Escurado, & Graells, 2010). Recently, SVDD has been considered as a promising process monitoring and FDI method for non-Gaussian processes. Compared to the ICA-based method, which incorporated a kernel density estimation procedure, the SVDD-based method is more computationally efficient, relies on a quadratic programming cost function to handle the non-Gaussian process fault (Chen et al., 2016).

In summary, SVM embodies many important principles. It has superior generalization ability where it provides efficient means of trading the training error and could handle the classification problems with small samples. It also solves the classification problem directly without trying to solve the much harder problem of estimating the distribution of data samples. Furthermore, even in the nonlinear case, the very central minimization task is stated as a convex optimization problem for which efficient numerical methods of finding the global optimum solution exist.

2.4.3 Gaussian Mixture Model (GMM)

Another non-Gaussian monitoring approach is based on the Gaussian mixture model (GMM), in which the process data are assumed to follow one of the multiple Gaussian distributions with different means and covariances at a fixed prior probability (a density estimate based method). In this way, the globally non-Gaussian process data can be attributed into one of the multiple clusters and then the corresponding local Gaussian models can be selected to detect the process fault in each operation region (Ferreira and Trierweiler 2009). Furthermore, GMM method also has more flexibility to express the data in the form of multiple Gaussian distributions. GMM has also shown to outperformed ANN in a number of non-linear classification problems (Marwala, Mahola, & Nelwamondo, 2006). For the mixture of multiple Gaussian distributions, GMM was proposed to describe the whole distribution with fewer Gaussian distributions based on Bayesian criterion. Measured data is expected to follow a mixture of multiple Gaussian distributions. Therefore, it has been widely applied in monitoring of various processes.

The GMM provides a probabilistic approach to estimate the pdf of the nominal process data and therefore enables more accurate calculation of the confidence bounds. GMM is a promising approach to maintaining a low rate of both false alarms and missing errors in process performance monitoring. Chen and Zhang (2010) extended the GMM technique for the modelling and online performance monitoring of batch manufacturing processes. Furthermore, a missing-value-based contribution analysis method is proposed to facilitate the identification of the detected process fault.

As GMM does not reduce the dimensionality of measurements, a mixture of factor analysis model was proposed for process monitoring inspired by GMM. Firstly, in order to reduce the dimensionality of the process data, the GMM model is sometimes combined with the traditional PCA model, which leads to a mixture form of the PCA model. GMM-based fault detection framework was utilized by adding pre-processing methods; PCA (Chen, Morris, & Martin, 2006; Choi, Park, & Lee, 2004) and multi-way PCA (Chen & Zhang, 2010) to improve its performance. GMM-based FDI method was also utilized with the combination of detection methods like KFDA (Zhu & Song, 2011). Choi et al. (2004) proposed PCA-GMM based fault detection and DA-GMM based fault isolation based on the assumption that the dataset are linear and Gaussian, whereas Li, Tan, and Li (2012) applied improved KFDA-GMM where GMM is applied for fault isolation and identification on the KFDA subspace. The variable-weighting vector and feature vector selection were applied to improve the KFDA discriminant performance. Multi-scale FDA based on CWT, GMM and *k*NN has been applied, where GMM and *k*NN have been applied individually as the classifiers (Gharavian, Almas Ganj, Ohadi, & Heidari Bafroui, 2013).

A Bayesian inference-based GMM multimode fault detection scheme has been developed and the operation mode identification problem has been discussed. GMM can characterise operation modes well, but using GMM to represent faulty data is not appropriate because the data are usually non-Gaussian distributed. (Jiang, Huang, & Yan, 2016) developed GMM and optimal principal components (OPCs)-based Bayesian identification system for fault identification in multimode process, where GMM and Bayesian inference is employed for operating mode identification, while PCA and GA-based optimal PC selection method is performed to generate lower dimension and more efficient evidence.

GMM-based weighted KICA (WKICA) has been proposed to monitor nonlinear and non-Gaussian processes (Cai et al., 2017). GMM employed to estimate the probability of each obtained KIC for measuring the individual KIC's importance. This enable the assignment of different weight values to the extracted KICs according to their measured importance for highlighting the important process information when online fault detection is implemented. A contribution plots method for fault identification is developed based on the idea of sensitivity analysis.

Monroy, Benitez, Escudero, and Graells (2010) applied ICA as feature extraction technique, GMM-BIC technique as a clustering algorithm, and ML-SVM approach as learning algorithm. GMM-BIC is an unsupervised classification algorithm based on the probabilistic modelling of the feature space by means of a mixture of Gaussian distributions. In order to provide a final classification, this approach takes into account both the likelihood of the data with respect to the model and the complexity of the model itself, thus avoiding overfitting. BIC index could be replaced by the Akaike Information Criterion (AIC), an alternative measure of the goodness of fit to the probabilistic model.

A global and distributed GMM has been proposed for modelling and monitoring multi-model plant wide processes (Zhu, Ge, & Song, 2016). Specifically, a global GMM is first built for capturing the global operating mode of the current process, while the sub-block GMMs are extracted from the global model so as to monitor the local behaviour of each process theme. In contrast to traditional GMM, the distributed GMM can be more feasible for reflecting local process changes. The author applied the Bayesian co-clustering approach to extract topic information simultaneously from both sample-wise and variable-wise directions due to the fact that process knowledge is not always available. Once the Bayesian co-clustering has been implemented, the block division and mode recognition have also been simultaneously accomplished.

Yu (2016) proposed application of GMM with manifold regularization, called local and nonlocal manifold regularization-based GMM (LNGMM) for estimation of the distribution of process observation. A failure probability-based indicator LGP that fuses both of local and global information extracted from the baseline. It is capable to handle complicated data distributions with nonlinearity and multimodal features by using a mixture of multiple Gaussian components with manifold regularization technique. Furthermore, a probabilistic indicator integrating local information (Mahalanobis distance) and global information (negative log likelihood probability) from a baseline GMM constructed by normal process to monitor process states.

Recently, Cai et al. (2017) proposed GMM-based weighted KICA method. In weighted KICA (WKICA), GMM is used to estimate the probabilities of the KICs extracted by KICA. Then, the significant KICs are discriminated based on the estimated probabilities and assigned with larger weights to capture the significant information during online fault detection. After a fault is detected, a nonlinear contribution plots is developed based on the idea of a sensitivity analysis to identify the fault variables. The

GMM has also been applied to estimate complicated data distributions in some processes by a limited set of Gaussian components, and GMM-based statistics were developed for multimode or nonlinear process monitoring. Some indicators based on Mahalanobis distance and log likelihood probability are extracted from GMM constructed by normal data for fault detection in processes with complicated data distribution.

2.4.4 K-Nearest Neighbor (*k*NN)

K-nearest neighbor method (*k*NN) is a widely-used nonparametric decision rule, where there are several types of *k*NN methods with different distances and decision rules used for classification (He & Wang, 2007). It has a good performance without requiring knowledge of the probability distribution function of the data. The number of neighbor k, the only free parameter in kNN, can be optimized with respect to the leave-one-out error. The *k*NN-based FDI method was utilized with the combination of pre-processing and detection methods like KFDA (Zhu & Song, 2011).

However, for uneven batch processes, a variable moving window-k nearest neighbor (VMW-*k*NN) based local modelling with irregular phase division and monitoring strategy was proposed by Zhang et al. (2017). For each sample, a pseudo time-slice was constructed by searching samples that are closely similar to the concerned sample. VMW strategy is applied to vary the searching range, while the *k*NN rule is used to find the similar samples. After that, automatic sequential phase division procedure is applied in similarity evaluation tool to get different irregular phases and ensure their time sequence. The affiliation of each new sample is evaluated to determine the proper phase model and fault status can be distinguished from phase shift event. *k*NN has also been used as part of an hybrid ensemble, combined with neural networks, decision trees, and quadratic Bayes classifiers in a single ensemble (Woods, 1997). In Todorovski (2003) it was combined with two algorithms for learning decision trees, a rule learning algorithm, and a naïve Bayes algorithm.

To summarize, the most important ability of *k*NN is the ability of handling data which is not linearly separable. Furthermore, it tries to locally approximate a complicated function using the nearest stored samples. Therefore, this property of *k*NN can be utilized to make within-class data in a neighbourhood become distant when mapped to the new low-dimensional space.

2.5 Hybrid data-driven methods in FDI system

The research area of fault detection and identification in chemical processes is still very active, as summarized in Table 2.1, where each data-driven FDI method that has been applied is based on specific process characteristic, and researchers aim to tackle some of the drawbacks of these methods. Thus, the main motivation for developing hybrid FDI frameworks is that no single method is able to satisfy all the requirements of an accurate FDI system.

For example, pattern recognition methods have the ability to incorporate expert knowledge, which is positive if such information is available, but also require expert knowledge in their construction, which is negative if such information is not available. On the other hand, the future of process monitoring and FDI systems also need to meet a variety of needs including reliability, ability to handle uncertainty, and ability to utilise large quantities of data. Therefore, the solution for handling these demands is to develop hybrid FDI methods, where it is a promising idea to handle those data characteristics simultaneously by combining different FDI approaches together.

Recently, it is the trend to apply hybrid observer approaches in FDI systems due to the rapid development of software that allows the easy combination of observer algorithms (Mohd Ali et al. 2015). In the following, a review of hybrid FDI approaches is carried out.

Each FDI method works under its assumption, which means a method works well in one condition may not provide satisfactory performance in another condition. Thus, a hybrid FDI framework system is proposed, which is consists of multiple methods to perform the collective problem-solving in one single FDI framework, which is also referred as a multi-agents method (Natarajan & Srinivasan 2014). Each of these methods offers their inherent advantages respectively, and a combination of these would certainly be appropriate to overcome limitations of individual methods to improve the performance. For example, the hybridization of SVM and *k*NN would produce high robustness, while the combination C4.5 method with *k*NN would give higher accuracy (Askarian et al. 2016).

Therefore, multiple classifiers or identification methods have been combined to perform collective process fault identification and a scheduler to regulate the decision-making of these identification methods. Early hybrid FDI system where multiple identification methods have been applied is proposed by Mylaraswamy and Venkatasubramanian (1997), where a blackboard-based framework called Dkit have been developed based on this concept. This framework developed based on input-output interface to collect the process and diagnosis results, multiple methods for fault diagnosis (i.e. signed digraph (SDG) method, probability density function classifier, and qualitative trend analysis (QTA) method), and a scheduler as a conflict decision-maker. The framework integrates the SDG method with the other two methods to narrow down the scope of search rapidly for identification and diagnosis task.

Table 2.1: Classification of data-driven FDI literatures based on process characteristics

Characteristics	Methods
Batch process	Multivariate statistical techniques: multi-way PCA, multi-way PLS (Zhang, Zhao, Wang, & Wang, 2017), GMM (Zhang et al., 2017)
Non-Gaussian process	ICA (Cai et al., 2017; Fan & Wang, 2014; Slišković et al., 2012), KICA (Cai et al., 2017), GMM (Hsu, Chen, & Chen, 2010) Multi-scale PCA (Lau et al., 2013), multi-scale PLS (Lee et al., 2009), multi-scale FDA (Deypir, Boostani, & Zoughi, 2012)
Slow varying process	Adaptive and recursive monitoring approaches: Recursive PLS (Li et al., 2005), recursive PCA (Zhang, Li, & Teng, 2012) Moving-windows approaches: Moving-window PCA, moving-window KPCA, moving-window KICA (Jiang & Yan, 2013b; Liu et al., 2009)
Dynamical	Dynamic multivariate statistical analysis methods: Dynamic PCA, dynamic ICA Bayesian network (Gharahbagheri et al., 2017a)
High dimensional data	Multivariate statistical analysis: PCA, PLS, ICA, FDA-based KPCA (Deng et al., 2017a) KPLS (Zhang & Ma, 2011) DPLS (Ghosh et al., 2014) KFDA (Liu et al., 2010) Deep learning methods (Deng et al., 2017a) Multiple scale analysis (Agrawal et al., 2014; Härdle, 2011):
Nonlinearity	Machine learning methods: ANN, SVM Kernel learning methods:

Characteristics	Methods
	KPCA (Zhang, Teng, & Zhang, 2010), KPLS, KFDA (Ferreira & Trierweiler, 2009)
Instrumental fault	Sensitive adaptive threshold (Alkaya & Eker, 2011)
Incipient fault	Machine learning methods: Neuro-fuzzy, SVM, GMM Signal processing methods: Wavelet transform method Multi-scale PCA (Fourie & De Vaal, 2000; Lau et al., 2013; Misra et al., 2002; Wang & Romagnoli, 2005) Multi-scale KPCA (Choi et al., 2007, 2008; Maulud et al., 2006) Multi-scale PLS (Lee, Lee, & Park, 2009) Multi-scale KPLS (Zhang & Hu, 2011; Zhang & Ma, 2011)
Complexity	Multi-block approach Multi-block PLS (Zhang & Hu, 2011) Multi-block multi-scale (Alawi & Morris, 2007)
Signal processing	Nonlinear parameter estimation techniques: Multi-scale entropy (Zhang, Xiong, Liu, Zou, & Guo, 2010) Multi-scale permutation entropy (Tiwari, Gupta, & Kankar, 2015; Wu, Wu, Wu, Ding, & Wang, 2012) Empirical mode decomposition, EMD (Liu, Cao, Chen, He, & Shen, 2013; Zamanian & Ohadi, 2011), EEMD (Xue & Zhou, 2016)

However, unlike the previous approach, Maurya et al. (1999) has proposed to use SDG first as a filter to reduce the set of possible faults, where the measured response pattern and the predicted response is compared, and the QTA to identify the actual fault by assigning ranks to the faults from level 1. This framework improve the performance, where it addressed the problem of inaccurate identification due to noise using denoising techniques and uncertainties safety-critical faults within the set of fault candidates by overriding the SDG-based candidates fault set when the confidence index is low. This hybrid framework only focused on incipient faults in the industrial benchmark of TEP, where the diagnostic resolution is improved while the computational complexity is reduced.

Statistical methods such as PCA and PLS have also been combined with supervised methods like ANN or with the time-frequency analysis to extract statistical features, while others have focused on combining the statistical methods with the analytical redundancy method for powerful isolation capabilities. One of the recent hybrid methods has been proposed by Jiang (2015), where the CVA-FDA method was implemented to identify and diagnose the presence of overlap data. When the FDA cannot provide good separation on highly serial correlation data, they have applied CVA to handle the serial correlations first before the FDA performed fault identification and identification. This approach has reduced the misclassification rate by approximately 40% compared to using FDA alone. Another hybridization, KPCA method has been combined with KICA (Zhang, 2009). This method has combined the advantage of KPCA related to Gaussian part with the advantage of KICA for the non-Gaussian part to develop a nonlinear dynamic approach. These hybrid fault detection methods later were combined with another method for classifying faults, such as SVM as proposed by Zhang (2009).

However, the implementation of hybrid fault identification strategy in FDI system means that there are two or more diagnostic methods with multiple outputs from them, with a possibility in conflicting results (Chen, Ding, Luo, & Zhang 2017). Thus, there is a need for decision-making method to regulate the results and final output. Few types of decision-making methods, which combine these methods have been employed for monitoring, control and identification system, such as averaging, majority voting, modified majority voting, Bayesian probability, and Dempster-Shafer method (Karimi and Jazayeri-Rad 2014; Seng Ng and Srinivasan 2010; Uraikul et al. 2007; Zhang and Ge 2015; Zhang 2006). Other methods, such as ordered weighted averaging (OWA) were also used as the multi-attributed data is hybridised into aggregated values of a single attribute (Salahshoor et al. 2010).

Another examples, Chetouani (2014) has combined ANN with Bayesian classification method to estimate the PDFs of the process to solve the fault detection problem. The Bayes theorem was also combined with neural adaptive black-box identification, called NARX (Nonlinear Auto Regressive with eXogenous input) model for fault detection system. Meanwhile, Ge et al. (2010) has combined several linear subspace methods with Bayesian inference, whereas Salahshoor et al. (2010) has integrated multiple classifiers such as SVM and ANFIS into the FDI system to enhance the fault detection and diagnostic tasks.

Recently, the hybridization of fault discriminant to enhance KPCA method is proposed by Deng and coworkers (Deng et al. 2017). Since the KPCA-based fault detection method ignores available prior fault information, it may unable to provide the best fault detection performance for nonlinear process monitoring. Therefore, an enhanced KPCA is proposed, known as a fault discriminant enhanced KPCA (FDKPCA), where KPCA and kernel local-nonlocal preserving discriminant analysis (KLNPDA) were used to produce kernel principal components (KPCs) and fault discriminant components (FDCs) based on normal operating data and prior fault data simultaneously. Monitoring statistics are then constructed for both the KPCA and KLNPDA sub-models. Then, the Bayesian inference was applied to transform these monitoring statistics into overall fault probabilities by weighting the results of the two-sub-models. The proposed KLNPDA scheme provides more effective online fault detection capability.

Gharahbagheri, Imtiaz, and Khan (2017a, 2017b) integrated identification information from various identification methods such as KPCA and sensor validation module, combined them with process knowledge using Bayesian network to complete a process of FDI framework. Bayesian network has been applied to diagnose internal state faults and disturbance faults, whereas sensor module separate the sensor faults from process faults. However, the research in this area still inadequate, such as process in dynamic condition and time-varying variables. Moreover, chemical processes are also prone to unknown disturbances.

Jiang and Huang (2016) proposed distributed monitoring based on multivariate statistical analysis and Bayesian method for large-scale plant-wide processes. Within the statistical framework based on PCA, they proposed a stochastic optimization algorithm based performance-driven process decomposition method, while the basic Bayesian identification system is used as the decision making system. The process decomposition aspect has been highlighted in order to achieve the best possible monitoring performance. Through the obtained sub-blocks, local monitors are established to characterise local process behaviours, before a Bayesian fault identification system is established to identify the process status. However, some issues like missing data, multiple sampling rate, and communication delay should be further addressed to improve the framework.

For large systems without knowledge rules or with complex models leading to costly and time-consuming process, a data-driven approach can be used to decrease high-dimensional data into the interested information of lower dimensional-space. After that, a classification technique or expert rule in knowledge-based domains can be used for modelling instead of an analytical-based method. For instance, fuzzy logic, ANN or SVM can map the relations between inputs and outputs in terms of non-linear models. Recently, researchers have developed algorithms that combined two or more methods in what is called hybrid systems. These algorithms are applied as estimators to overcome the limitations of the single algorithm and to further increase the estimator's performances. For example, ANN will only allow reasoning from input to outputs and this can be overcome by using the adaptive neuro-fuzzy inference systems (ANFIS) (Yetilmezsoy et al., 2011). Normally, the combination utilised the advantages of each of the algorithms such as the hybrid neural network (HNN), ANFIS, fuzzy neural network (FNN) and expert system neural network (ES-NN) (Sivan, Filo, & Siegelmann, 2007).

A new FDI based on kernel entropy component analysis (KECA) and multi-scale PCA (MSPCA) was proposed for nonlinear chemical process by Zhang, Qi, Wang, Gao & Wang (2017). The KECA-based method have an angle-based statistic, designed to express the distinct angular structure that KECA reveals, which is able to measure the similarity between probability density functions. Each KECA classifier is dedicated to a specific fault, which provides an expendable framework for incorporating new faults identified in the process. As to the fault features are submerged because of multi-scale property of data, an enhanced KECA method for fault detection and identification is developed, by adding multi-scale PCA (MSPCA) for feature extraction to improve the classification effect of KECA.

Multi-class SVM based process supervision and fault identification scheme is proposed by Gao and Hou (2016) to predict the status of the Tennessee Eastman process. PCA is firstly used to reduce the feature dimension. Then, to increase prediction accuracy and reduce computation load, the optimization of SVM parameters is accomplished with the grid search (GS) method which generates comparable classification accuracy to genetic algorithm (GA) and particle swarm optimization (PSO) while being more efficient than the latter two algorithms. They proposed SVM integrated GS-PCA fault identification approach. Meanwhile, in Yang and Hou (2016) work, PCA and recursive feature elimination are combined with SVM for fault detection and identification from the simulation of fed-batch fermentation penicillin process. Another efficient multimode fault identification, proposed by Jiang et al. (2016) has utilised the GMM and Bayesian inference method for identification of the operating mode, before local PCA is implemented in each mode for extraction of distinct fault features. Then, the principal components are selected using genetic algorithm. Finally, a Bayesian probability-based diagnosis system is developed to identify the process fault statuses.

Gas turbine engine FDI is proposed by developing an ensemble of dynamic neural network identifiers, by constructing three separate or individual dynamic neural network architectures. Specifically, a dynamic MLP, a dynamic RBF neural network, and a dynamic SVM are trained to individually identify and represent the gas turbine engine dynamics. Next, three ensemble-based techniques are developed to represent the gas turbine engine dynamics, namely, two heterogeneous ensemble models and one homogenous ensemble model. The best selected stand-alone model is the dynamic RBF and the best selected ensemble architecture is the heterogeneous ensemble in terms of their performances in achieving an accurate system identification are then selected for accomplishing the FDI task and objective. The required residual signals are generated by using both a single mode-based solution and an ensemble-based solution under various gas turbine engine health conditions, as illustrated through detailed quantitative confusion matrix analysis and comparative studies.

Finally, in Askarian et al. (2016), an exploiting classifier and combined methods were assessed in TEP, for which diverse incomplete observations were produced. The use of several indicators revealed that the trade-off between performances of the different schemes. SVM and C4.5, combined with *k*NN, produce the highest robustness and accuracy, respectively. Bayesian network and centroid appears as inappropriate options in terms of accuracy, while Gaussian NB is sensitive to imputation values. In addition, feature selection was explored for further performance enhancement, and the proposed contribution index showed promising results.

2.6 Summary

In this chapter, the state-of-the-art methods and challenges of data-driven based FDI frameworks in chemical process systesms are briefly reviewed. The high numbers of implementations of these methods are due to their simple formulation without needing perfect knowledge of the system model, their ability to reduce time and cost for models development, their easy implementation, and their adaptability. Moreover, these approaches can handle high-dimensional and correlated process variables, especially in complex and large-scale systems. Although a number of research has been accomplished, and a wide range of methods is available in the area of process monitoring and fault identification in chemical process systems, most of them can perform well in the laboratory or testing settings, but some are not suitable for field implementation. Furthermore, most of these studies mainly focus on solving problems of data characteristics such as nonlinearity, non-Gaussian distribution, and data autocorrelation, where it is still hard for any single FDI method to perform effectively for all possible faults a process could have. Therefore, it seems that hybridisation-based framework is required for developing a complete and robust fault detection and identification tool.

CHAPTER 3: MULTI-SCALE KFDA FEATURE EXTRACTION METHOD

3.1 Introduction

In general, the proposed multi-scale KFDA feature extraction method consists of three different steps: the data acquisition and normalization, wavelet transformation, and multi-scale KFDA discriminant vector. In addition, the wavelet transformation step will have another three steps; which are the DWT decomposition, threshold determination, and inverse DWT reconstruction step. The general methodology of the proposed multi-scale KFDA method is highlighted by the flowchart of Figure 3.1.

The strategy consists of two major steps after data acquisition and normalisation, as seen in the flowchart, which are the feature extraction step and the fault classification step. Firstly after data acquisition from the process system, the input database, consisting of the normal and faulty data is pre-processed by the normalisation method. Normalisation of the data was done with each of the variables linearly scaled in the range of [0, 1] to avoid the domination between the larger and smaller numerical range in each of the variables.

Then, the normalized data was fed to the multi-scale KFDA feature extraction step. In this step, the discrete wavelet transform (DWT) method is applied with its multi-scale feature decomposition to give the distinguished characteristic features of the input data. DWT decomposed each of the variables in the input dataset individually, before the significant wavelet coefficients were retained. The retained wavelet coefficients will have higher values compared to the threshold value when a significant event occurred.

After extracting the input features, the DWT reconstruction method called inverse discrete wavelet transform (IDWT) was applied to the retained coefficients.

Through these procedures, the dimensions of the input patterns can be reduced and useful information can be extracted. After that, the reduced dimensional database was further distinguished and separated by the KFDA method into the discriminative feature space, called multi-scale KFDA discriminant vector. The detail steps of the proposed multi-scale KFDA FDI framework is further described in the following sub-chapters.

3.2 Data acquisition and normalization

Data acquisition was applied to the FDI system to obtain process data, where it was measured from the simulated process under different faulty conditions. The acquired process data is divided into two subsets; the training dataset and the testing dataset. The training dataset was used to develop the FDI models, while the testing dataset was reserved to evaluate the performance of the developed classification models.

The normalisation technique was applied to the acquired input database to linearly scale the collected variables in the range of [0,1] using a mix-max normalisation equation, as expressed by Eq. 3.1. The normalization scale is important before applying the multivariable methods to avoid the domination of greater numerical range over the smaller numerical range.

$$v'_{ij} = \frac{v_{ij} - min_i}{max_i - min_i} \tag{3.1}$$

In this equation, max_i is the maximum and min_i is the minimum of the ith attribute values, and v_{ij} is the value of ith attribute of jth object and v'_{ij} is the normalized value. This initial step is crucial to improve data quality as well as to improve the accuracy and efficiency of the statistical and computational process.

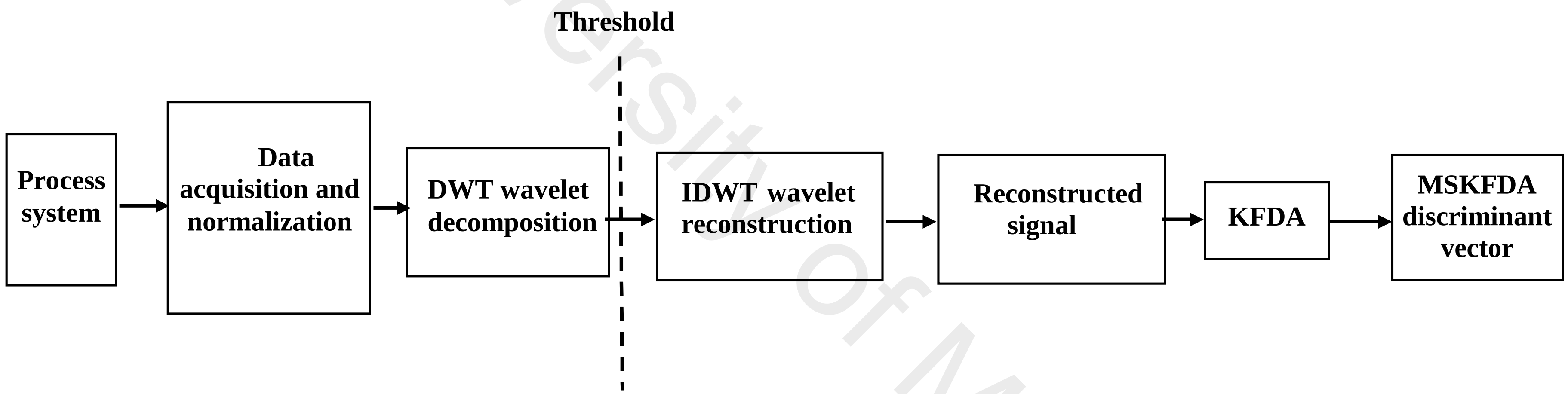

Figure 3.1: Flowchart of multi-scale KFDA feature extraction method

3.3 Wavelet Transformation

Wavelet transformation involved the implementation of discrete wavelet transform (DWT) with Daubechies (dB) wavelet as the mother wavelet. After the threshold has been determined and the wavelet coefficients were retained, the inverse discrete wavelet transform (IDWT) has been applied to reconstruct the selected coefficients.

3.3.1 DWT decomposition

Each of the variables in the input data was individually decomposed by applying the DWT approach with Daubechies wavelet as the mother wavelet. Briefly, the multi-scale KFDA approach contained, for example, input matrix x with m variables and n samples, which can be considered as an $n \times m$ data matrix. Each of the m columns was individually decomposed, where each of the m variables was applied with the same level of decomposition, and labelled as L. The scaling and wavelet function values were calculated for each iteration and presented as coefficients a_l and d_l, as given by Eqs. 3.2 and 3.3, respectively:

$$a_l = \square_0 x_{2l} + \square_1 x_{2l+1} + \square_2 x_{2l+2} + \cdots + \square_{L-1} x_{2L-1} \tag{3.2}$$

$$d_l = g_0 x_{2l} + g_1 x_{2l+1} + g_2 x_{2l+2} + \cdots + g_{L-1} x_{2L-1} \tag{3.3}$$

Then, the wavelet approximation coefficients, a_l and detail coefficients, d_l from each of the variable decompositions, were collected and constructed in their respective matrix, approximation coefficients matrices and detail coefficient matrices. The matrix size of $m \times \frac{n}{2^L}$ depends on the number of decomposed variables, m, number of observations, n, and the level of decomposition, L as shown in Eq. 3.4.

$$\begin{bmatrix} \vdots \\ a_i \\ a_{i+1} \\ a_{i+2} \\ \vdots \end{bmatrix} = \begin{bmatrix} \ddots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \iddots \\ \cdots & h_0 & h_1 & h_2 & h_3 & 0 & 0 & 0 & 0 & \cdots \\ \cdots & 0 & 0 & h_0 & h_1 & h_2 & h_3 & 0 & 0 & \cdots \\ \cdots & 0 & 0 & 0 & 0 & h_0 & h_1 & h_2 & h_3 & \cdots \\ \iddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \bullet \begin{bmatrix} \vdots \\ x_{2i} \\ x_{2i+1} \\ x_{2i+2} \\ \vdots \end{bmatrix}$$

$$\begin{bmatrix} \vdots \\ d_i \\ d_{i+1} \\ d_{i+2} \\ \vdots \end{bmatrix} = \begin{bmatrix} \ddots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \iddots \\ \cdots & g_0 & g_1 & g_2 & g_3 & 0 & 0 & 0 & 0 & \cdots \\ \cdots & 0 & 0 & g_0 & g_1 & g_2 & g_3 & 0 & 0 & \cdots \\ \cdots & 0 & 0 & 0 & 0 & g_0 & g_1 & g_2 & g_3 & \cdots \\ \iddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \bullet \begin{bmatrix} \vdots \\ x_{2i} \\ x_{2i+1} \\ x_{2i+2} \\ \vdots \end{bmatrix} \tag{3.4}$$

As a result, a total of $L + 1$ matrices were formed, with each representing the approximation coefficients and detail coefficients at different level of the decomposition scale.

3.3.2 Threshold determination

In the next step, threshold determination was initiated, with the retained wavelet coefficient larger than the threshold parameter. The wavelet coefficients usually correspond to a significant action in the process, where the aim of the threshold determination, or scale selection would be to find the most discriminating features. Stein's unbiased likelihood estimate method was applied in the soft threshold method based on the nonlinear transform, as shown by Eq. 3.5 and 3.6:

$$S(x) = \begin{cases} sign(x)(|x| - T), & |x| \geq T \\ 0, & |x| < T \end{cases}, \tag{3.5}$$

$$T = \sqrt{2\sigma^2 \log n \ /n}, \tag{3.6}$$

where T is a threshold, n is the length of the input vector and σ^2 is the estimated variance of the input data. In this work, the soft thresholding value for normalized data with range of [0,1] is 0.4. The reconstruction method was applied after the threshold has been calculated.

3.3.3 IDWT reconstruction

Then, the multi-scale model was produced by restructuring important scales from the previous decomposition and threshold step. The inverse discrete wavelet transformation (IDWT) was applied for restructuring the deterministic components in the variables from the retained wavelet coefficient following the threshold calculation. For example, all the coefficients through the filter ▯$_0$ and ▯$_1$ were recombined to be reconstructed in the time-domain space, as shown in Figure 3.2. The reconstruction process consists of level 2 up-sampling, as denoted by ↑2 symbol.

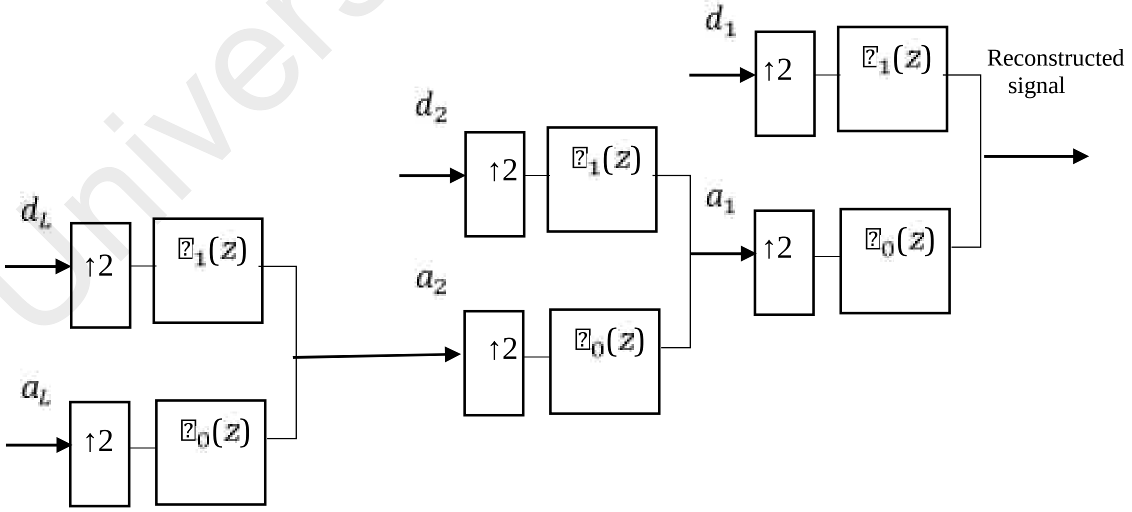

Figure 3.2: Inverse Discrete Wavelet Transform for signal reconstruction

A reconstruction method was applied by constructing $L + 1$ single-scale classifiers including L detail classifications at all levels and one approximation classifier at the coarsest level. Based on the results of both cross-validation and testing data validation, the scales were added one by one to the present scales until the optimal overall classification accuracy was obtained.

3.4 Multi-scale KFDA discriminant vector

During the final step of multi-scale KFDA feature extraction method, the KFDA approach was applied to each of the reconstructed matrices, with the objective of extracting the discriminative attributes across the multi-scale data. The key idea behind the KFDA algorithm was to enable homogenous data to come close within each other, and the heterogeneous data to be distant from each other by projecting data to the high dimensional space using kernel mapping.

The KFDA discriminant vector can be obtained using the within-class-scatter matrix, S_W and between-class-scatter matrix, S_B by maximizing the criterion in Eq. 3.7:

$$\max_{v_k \neq 0} \frac{v_k^T S_B v_k}{v_k^T S_W v_k}, \tag{3.7}$$

where v_k is the KFDA vector, and can be expressed as shown by Eq. 3.8:

$$v_k = \sum_{i=1}^{m} \alpha_i \phi(x_i) \tag{3.8}$$

with coefficient α_i, i = 1, …, *m*.

Based on the KFDA discriminant score vectors for multiple types of pattern data, the first-two dimension feature vectors of the score vectors became the inputs for the data-driven methods for the fault classification task.

3.5 Improved Multi-scale KFDA using Parseval's Theorem

An improved multi-scale feature extraction method is proposed using discrete wavelet transform (DWT) with Parseval's theorem and kernel Fisher discriminant analysis (KFDA). The DWT decomposition analysis performance is applied with Daubechies mother wavelet for the decomposition. The DWT decomposed the time-domain database into the wavelet coefficients with time- and frequency-domain by the Parseval's theorem, before the features were further reduced by KFDA. The KFDA method is used to determine the between-class-scatter and the within-class-scatter matrices. The output of KFDA feature space is reconstructed and used as input to feed the classifiers.

In general, the structure of the proposed system is shown in the Fig. 3.3, with the steps of the improved multi-scale KFDA framework summarized as below:

1. Each of the variables in the original signals are decomposed into both time- and frequency-domain using discrete wavelet transform (DWT).
2. In each level of decomposition for each of this variable, the DWT method with the Parseval's theorem was applied to produce the wavelet decomposition coefficients.
3. The transformed wavelet coefficients larger than the threshold limit were retained for the reconstruction step.
4. Inverse discrete wavelet transform (IDWT) was applied for reconstructing the retained wavelet coefficients.
5. Kernel FDA was applied to the retained wavelet coefficients to find the optimal feature vector for the distinguished input features.
6. A new lower dimensional input features from KFDA was fed for the classification and identification step.

The DWT is derived from CWT, and its advantage is it does not shift and scale continuously, and can only be operated in discrete steps. Although the DWT can improve the drawbacks of CWT, but it is still hard to find the fault conditions accuracy by vision or classifier. Because of the reconstruction process, it will produce a large data quantity of the signal features. For this reason, the energy signal is derived based on Parseval’s energy theorem (Wu et al., 2009). From the scaling and wavelet functions where an orthonormal basis is formed, the Parseval’s theorem relates the energy of the signal distortion to the energy of component expansion and their wavelet coefficient, as given in Eq. 3.9 below.

$$E_{signal} = \int |y(t)|^2 dt = \sum_{i=-\infty}^{\infty} |a(i)|^2 + \sum_{j=j0}^{\infty} \sum_{i=-\infty}^{\infty} |d_j(i)|^2 \tag{3.9}$$

Thus, the energy of the signal at different decomposition levels can be computed as

$$E_{signal} = \left[\|a_{od}\|, \|d_{od}\|, \ldots, \|d_{J-1\,d}\| \right], \tag{3.10}$$

$$\|a_{od}\| = \sqrt{\sum_{i=-\infty}^{\infty} |a_{od}(i)|^2} \tag{3.11}$$

$$\|d_{jd}\| = \sqrt{\sum_{i=-\infty}^{\infty} |d_{jd}(i)|^2} \tag{3.12}$$

where $\| \|$ is the $L^2(R)$ normal and J represents the total number of resolution levels, $a_{od}(i)$ is the approximation DWT coefficient, and $d_{jd}(i)$ is the detail DWT coefficients for the disturbance signal at level j, respectively. A similar calculation is made on a pure signal to attain E_{pure}. The feature vector x_0 is generated by substracting E_{pure} from E_{signal} as

$$x_0 = \Delta E = E_{signal} - E_{pure} \tag{3.13}$$

Then, the wavelet approximation coefficients, $a_{od}(i)$ and detail coefficients, $d_{jd}(i)$ from each of the variable decompositions are collected and constructed in their respective matrix, with the matrix size of $m \times \frac{n}{2^L}$,which is depended on number of decomposed variables, m, number of observations, n, and the level of decomposition, *L* as previously shown in Eq. 3.4.

As a result, a total of *L+1* matrices are formed, each being represented by the approximation coefficients and detail coefficients at a different level of decomposition scale. After that, the multi-scale model was produced by restructuring the important scales from the previous decomposition step, where the inverse discrete wavelet transformation (IDWT) was applied for restructuring the deterministic components in the variables from the retained wavelet coefficient.

Finally, the KFDA approach is applied to the reconstructed matrix, with the objective to extract the distinguished features across the multi-scale data. The key idea of the proposed KFDA algorithm is to enable homogenous data to come close to each other and the heterogeneous data to be distant from each other, by projecting data to the high dimensional space by kernel mapping.

The KFDA discriminant vector can be obtained using the within-class-scatter matrix, S_w and between-class-scatter matrix, S_B by maximizing this criterion:

$$\max_{v_k \neq 0} \frac{v_k^T S_B v_k}{v_k^T S_W v_k}, \tag{3.14}$$

where v_k is the KFDA vector, and can be expressed as

$$v_k = \sum_{i=1}^{m} \alpha_i \phi\ x_i \tag{3.15}$$

with coefficient α_i, i = 1, …, *m*.

Based on the KFDA discriminant score vectors for the multiple types of pattern data, the first-two dimension feature vectors of the score vectors will be the inputs to the ANFIS models for pattern classification. For classifying the fault of the test engine, the coefficient cA of approximated version and coefficient cD of the detailed version at each resolution level will be used to extract the features of the fault sound emission signals. The energy distribution of approximated version P_a and detailed version P_d can be computed as follows:

$$P_a = \frac{1}{N_j} \sum_{h} |cA|^2 = \frac{\|cA\|^2}{N_j} \tag{3.16}$$

$$P_d = \frac{1}{N_j} \sum_{h} |cD_{j,h}|^2 = \frac{\|cD_j\|^2}{N_j} \tag{3.17}$$

where $\|cA\|$ and $\|cD_j\|$ are the norm of the expansion coefficients cA and cD_j. For classifying the different faults, the energy distribution value P_a and P_d will be adopted in fault classification.

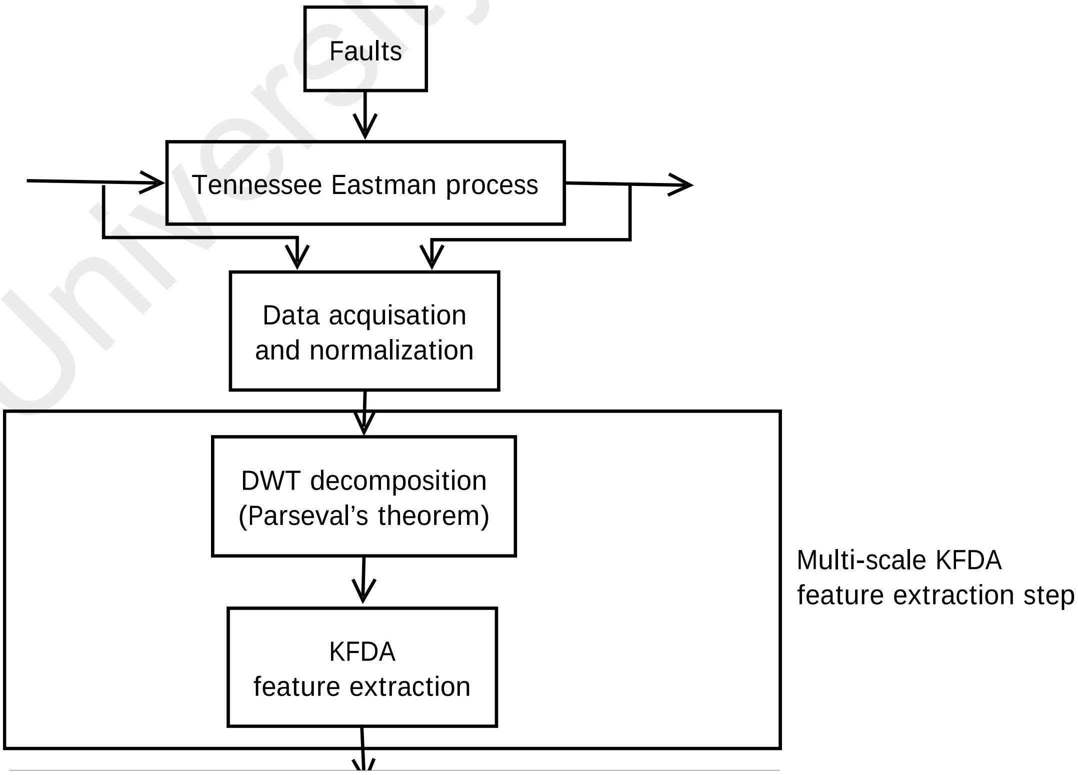

Figure 3.3: Block diagram of improved multi-scale KFDA with Parseval's theorem for feature extraction

3.6 Case Studies

The realistic simulation of the Tennessee Eastman process is chosen as the case study for the proposed multi-scale KFDA-data-driven FDI methods. This simulation of the industrial process is used to evaluate the proposed methods, where the details of the process and its performance evaluation criteria were described in the following sections.

3.6.1 Tennessee Eastman process simulation

The realistic simulation of the Tennessee Eastman process (TEP) was chosen as a case study for the proposed FDI framework. This simulation of the industrial process was introduced by Downs and Vogel (Downs & Vogel, 1993) for evaluating and benchmarking various process control and fault monitoring methods. The details can be found in Chiang, Russell, and Braatz (2001). The Tennessee Eastman process, as shown in Figure 3.4, was developed by the Eastman Chemical Company with the intention of providing a realistic simulation of an industrial process. The control strategy implemented in the process has been described in Lyman and Georgakis (1995).

The process involves four irreversible exothermic gas reactions. These reactions rates depend on the temperature and the concentration of the reactants. The process has five major units: reactor, product condenser, vapor-liquid separator, recycle compressor and product stripper. Four reactants are the inputs of the process that produce two products and two by-products named alphabetically from A to H. The heat of the reactions is removed by the cooling water in heat exchanger. The products and unconverted reactants leave the reactor, as vapor which is partly converted to a liquid in the condenser as shown in the Eq. 3.18 to Eq. 3.21.

$$A(g) + C(g) + D(g) \rightarrow G(liq) \tag{3.18}$$

$$A(g) + C(g) + E(g) \rightarrow H\ liq \tag{3.19}$$

$$A(g) + E(g) \rightarrow F(liq) \tag{3.20}$$

$$3D(g) \rightarrow 2F(liq) \tag{3.21}$$

The process consists of 52 variables in total as summarized in Table 3.1 containing 41 measured and 11 manipulated variables. The input variables are XMEAS(1) until XMEAS(36) and XMV(1) until XMV(11), where XMEAS(1) to XMEAS(36) are the process measurements, and XMV(1) to XMV(11) are the manipulated variables; the output variables are XMEAS(37) until XMEAS(41), where XMEAS(37) to XMEAS(41) are the quality measurements.

There are 21 process faults (Fault 1 to Fault 21) simulated in the Tennessee Eastman process as summarized in Table 3.2. These 21 faults represent several types of process faults, such as step disturbances, random variations, slow kinetic drift, valve sticking, and some unknown conditions, where all of these faults were utilized for the training and testing of the developed classification system.

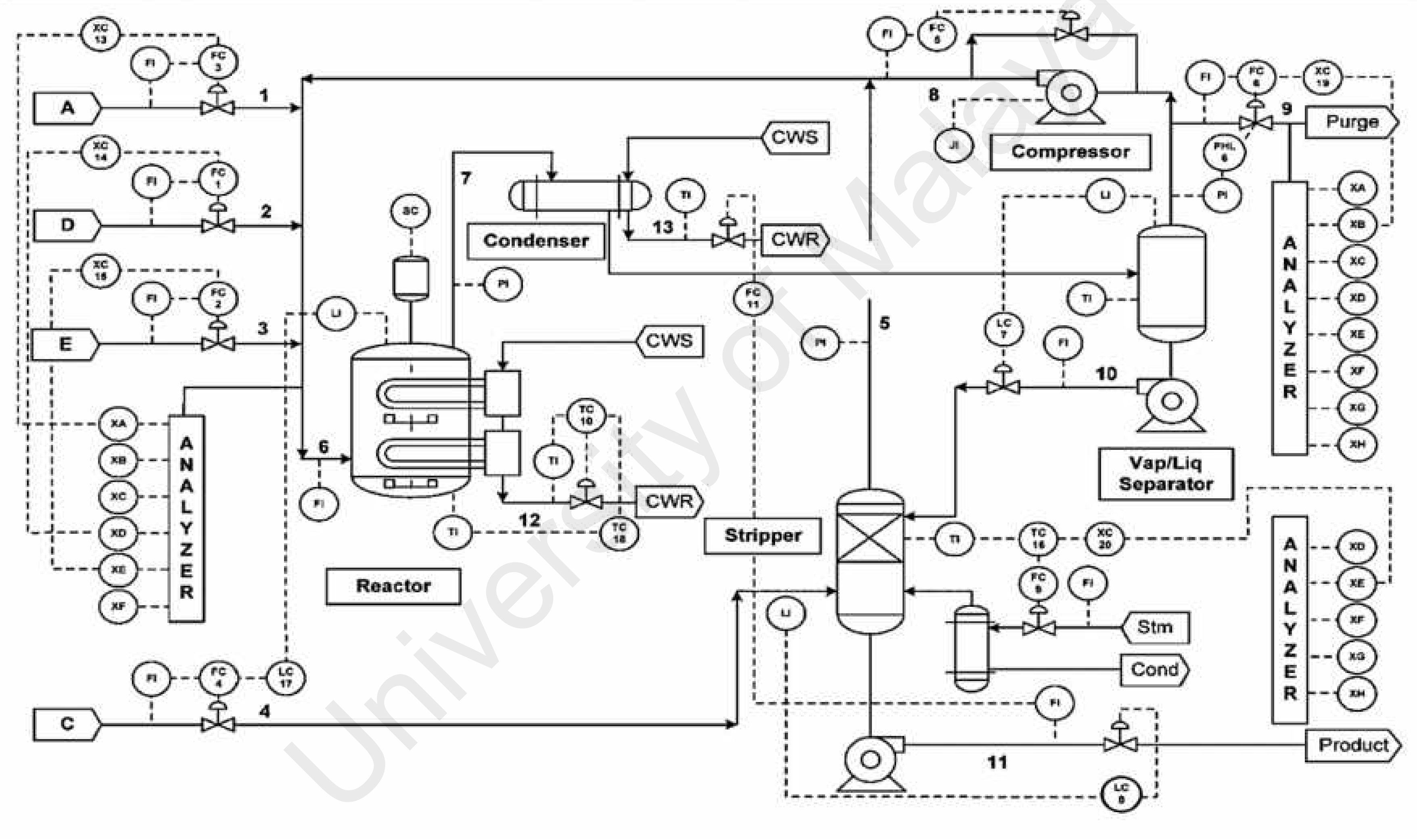

Figure 3.4: The Tennessee Eastman process plant

Table 3.1: Measured and Manipulated Variables of the Tennessee Eastman process

Identification	Description	Identification	Description
XMEAS(1)	A feed Stream 1	**XMEAS(27)**	Reactor feed component E
XMEAS(2)	D feed Stream 2	**XMEAS(28)**	Reactor feed component F
XMEAS(3)	E feed Stream 3	**XMEAS(29)**	Purge component A
XMEAS(4)	Total feed Stream 4	**XMEAS(30)**	Purge component B
XMEAS(5)	Recycle flow	**XMEAS(31)**	Purge component C
XMEAS(6)	Reactor feed rate	**XMEAS(32)**	Purge component D
XMEAS(7)	Reactor pressure	**XMEAS(33)**	Purge component E
XMEAS(8)	Reactor level	**XMEAS(34)**	Purge component F
XMEAS(9)	Reactor temperature	**XMEAS(35)**	Purge component G
XMEAS(10)	Purge rate	**XMEAS(36)**	Purge component H
XMEAS(11)	Separator temperature	**XMEAS(37)**	Product component D
XMEAS(12)	Separator level	**XMEAS(38)**	Product component E
XMEAS(13)	Separator pressure	**XMEAS(39)**	Product component F
XMEAS(14)	Separator underflow	**XMEAS(40)**	Product component G
XMEAS(15)	Stripper level	**XMEAS(41)**	Product component H
XMEAS(16)	Stripper pressure	**XMV(1)**	D feed flow Stream 2
XMEAS(17)	Stripper underflow	**XMV(2)**	E feed flow Stream 3
XMEAS(18)	Stripper temperature	**XMV(3)**	A feed flow Stream 1
XMEAS(19)	Stripper steam flow	**XMV(4)**	Total feed flow Stream 4
XMEAS(20)	Compressor work	**XMV(5)**	Compressor recycle valve
XMEAS(21)	Reactor cooling water outlet temp.	**XMV(6)**	Purge valve
XMEAS(22)	Separator cooling water outlet temp.	**XMV(7)**	Separator product liquid flow
XMEAS(23)	Reactor feed component A	**XMV(8)**	Stripper product liquid flow
XMEAS(24)	Reactor feed component B	**XMV(9)**	Stripper steam valve
XMEAS(25)	Reactor feed component C	**XMV(10)**	Reactor cooling water flow
XMEAS(26)	Reactor feed component D	**XMV(11)**	Condenser cooling water flow

Table 3.2: Faults defined in the Tennessee Eastman process

Fault ID	Description	Type
DV1	A/C feed ratio, B composition constant (stream 4)	Step
DV2	B composition, A/C ratio constant (stream 4)	Step
DV3	D feed temperature (stream 2)	Step
DV4	Reactor cooling water inlet temperature	Step
DV5	Condenser cooling water inlet temperature	Step
DV6	A feed loss (stream 1)	Step
DV7	C header pressure loss-reduced availability (stream 4)	Step
DV8	A, B, C feed composition (stream 4)	Random variation
DV9	D feed temperature (stream 2)	Random variation
DV10	C feed temperature (stream 4)	Random variation
DV11	Reactor cooling water inlet temperature	Random variation
DV12	Condenser cooling water inlet temperature	Random variation
DV13	Reaction kinetics	Slow drift
DV14	Reactor cooling water valve	Sticking valve
DV15	Condenser cooling water valve	Sticking valve
DV16-DV20	Unknown	Unknown
DV21	Valve for stream 4 fixed at the steady-state position	Constant position

In this work, all of the faults in the TEP were tested and analysed by the proposed framework. Data acquisition was established for 21 of the designated faults introduced at different range of operating conditions. Data acquisition for each fault has been carried out to include the 52 types of variables, based on three-minute interval time sampling. Meanwhile, faults were designated to occur after one hour of running the process with the training data and after eight hours of running the process with the testing data, with the specification for each designated process fault tabulated in Table 3.2. The proposed frameworks, previously discussed in sections 3.1 until 3.4 were then applied for these datasets, with the classification performance evaluated based on the testing datasets.

3.6.2 Performance evaluation

For the classification performance, the separability and classification between these fault databases can be seen through the classification projection diagram. Additionally, the identification performance for the proposed multi-scale KFDA-data-driven methods were also can be illustrated using the confusion matrix, as shown in Figure 3.5 (Dogantekin, Dogantekin, & Avci, 2011).

From the confusion matrix, the total number of samples was considered for each of the designated faults. Data points that were correctly classified were represented by the diagonal elements and highlighted in green squares, whereas the others were misclassified data points, coloured in red squares; either as the correct or incorrect responses. The accuracy per output class was given by the grey squares, while the blue squares presented the overall accuracy of the classifier. For example, the classification efficiency was 100% when there was no misclassification.

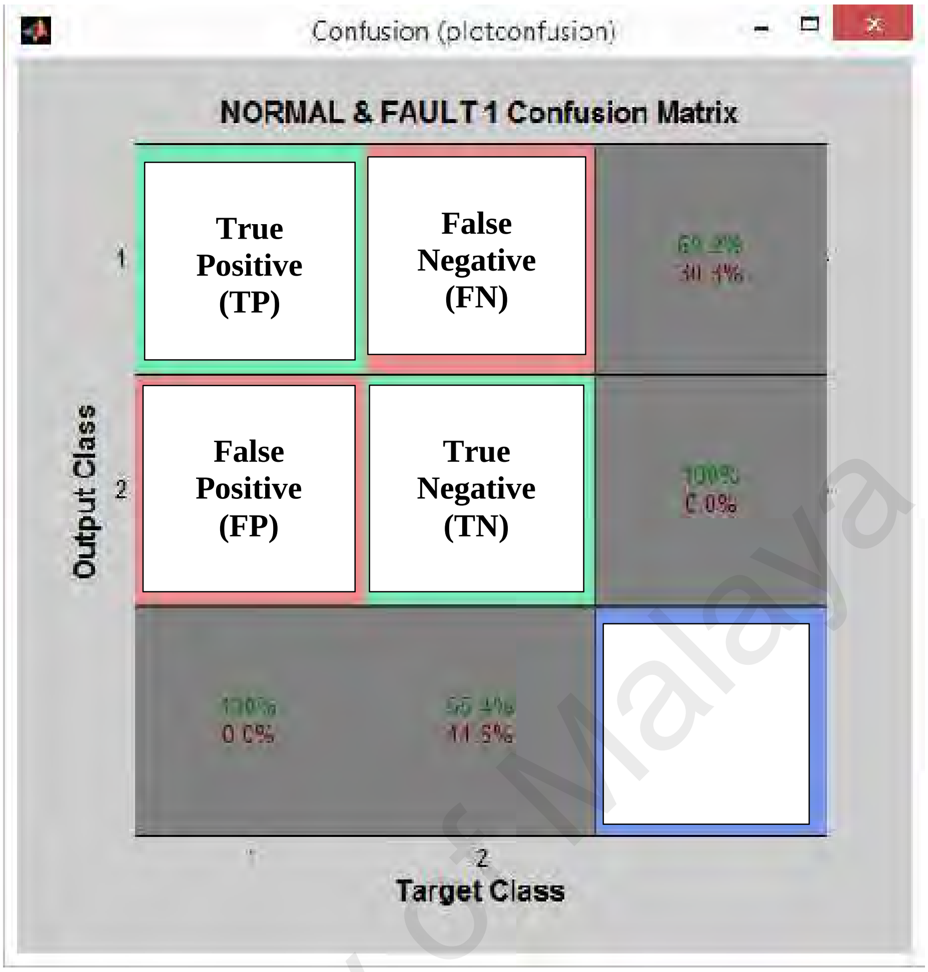

Figure 3.5: Confusion matrix definition

Furthermore, the overall performance evaluation of the proposed methods can be measured using the tool defined as the accuracy rate, given by Eq. 3.22 below:

$$Accuracy\ Rate \ = \ 1 - \frac{FP+FN}{TP+TN+FP+FN} \times 100\%, \tag{3.22}$$

where True Positive (TP) is a fault indication on faulty operating condition, False Positive (FP) is a fault indication on normal operating condition, True Negative (TN) is a normal indication of normal operating condition, and False Negative (FN) is a normal indication of faulty operating condition.

3.7 Results and Discussion

The proposed multi-scale KFDA-data-driven methods were implemented to detect and diagnose the faults of the TEP database, with the data for training step were used to develop data-driven FDI models, while the data for testing step were used to evaluate these classification results through the accuracy rate, as shown in Eq. 3.22. An evaluation of the multi-scale KFDA-based feature extraction work was designed to investigate the efficiency of the wavelet decomposition in the proposed approach, using the normalized data of Fault 4 from the reactor cooling water flow variable (XMV10), as shown in Figure 3.6.

From Fig. 3.6(b), the approximation coefficient for Fault 4 data of the transformed signal clearly shows a significant difference in amplitude of the plot compared to the normal data. This features show that the disturbance or fault event has occurred in this data. The detailed coefficient for level 5 decomposition in Fig. 3.6(e) also shows some distinctive characteristics to differentiate the normal and fault condition in the database. As shown and proven by Fig. 3.6, discrete wavelet analysis transformation decomposed the faulty data into the significant information related to the process variables. After that, the KFDA was applied on all of these multi-scaled data for fault classification. As examples, Fig. 3.7 and 3.8 show the classification of Fault 4, Fault 9 and Fault 11 based on the FDA projection and the proposed multi-scale KFDA, respectively. Clearly, from Fig. 3.7, the FDA is able to classify the Fault 4 and Fault 9 data but failed to separate and classify Fault 11. The FDA method has been unable to distinguish the Fault 11 because of the separation between-classes is not large enough while the distribution within-classes also was quite large. This is due to the fact that all variables in the data sets could not be separated without proper elimination of the insignificant information, which shared common characteristics.

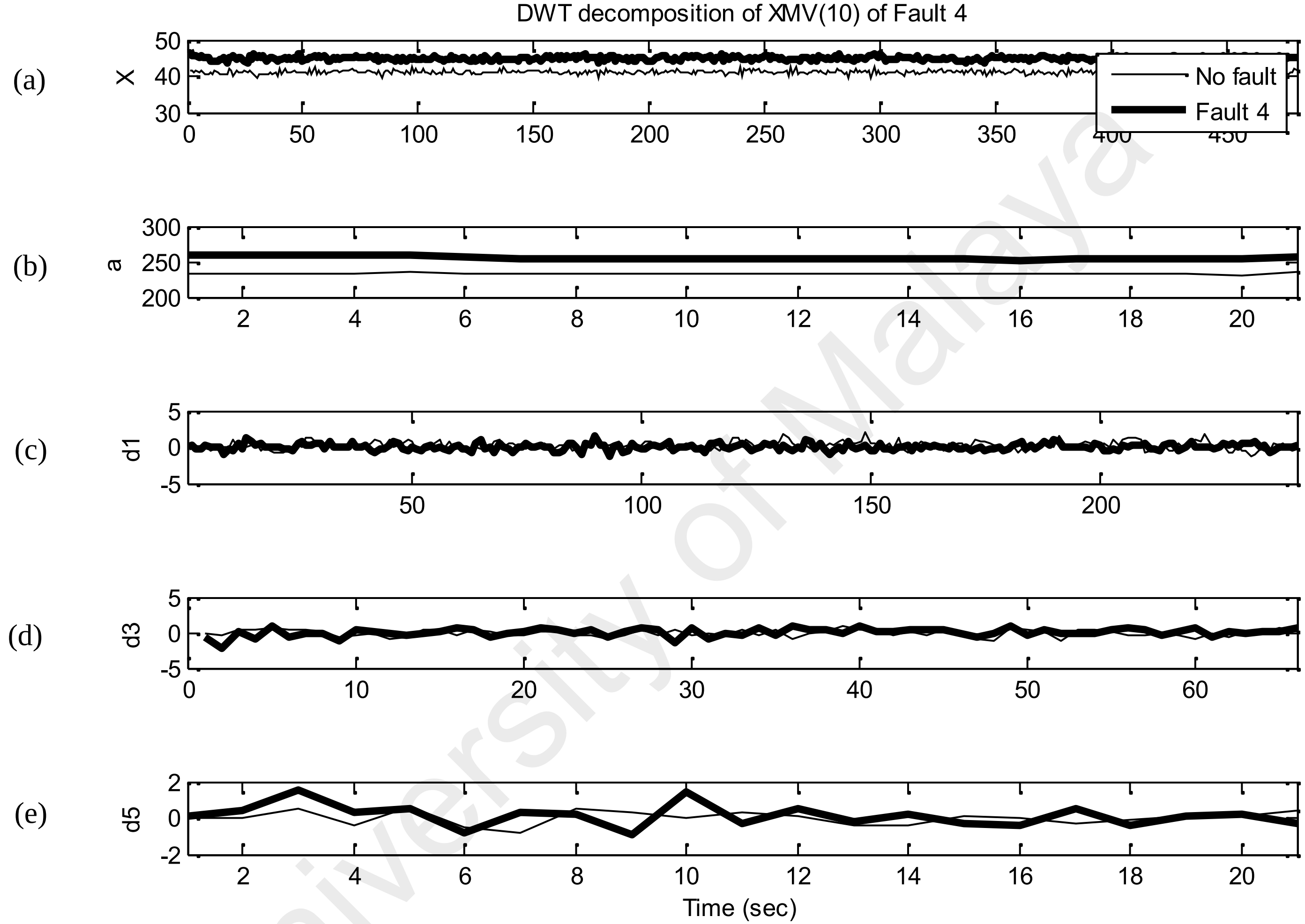

Figure 3.6: Variable XMV10 with normal and Fault 4 (a) original data; (b) fifth-level approximate coefficient decomposition, (c) first-level detail coefficients decomposition, (d) third-level detail coefficients decomposition, and (e) fifth-level detail coefficients decomposition.

In contrast to the FDA approach, the proposed multi-scale KFDA method has improved the power of discrimination and classification, especially in multiple time and frequency domains, as shown in Fig. 3.8, which shows the projection data onto the first two multi-scale KFDA vectors. From the figure, it shows that there is a large separation in-between-class distribution while the scattering of within-classes distance is shortened compared to the FDA. This classification has proved that the proposed multi-scale KFDA has better discriminative power than the normal FDA method for all the faults 4, 9, and 11.

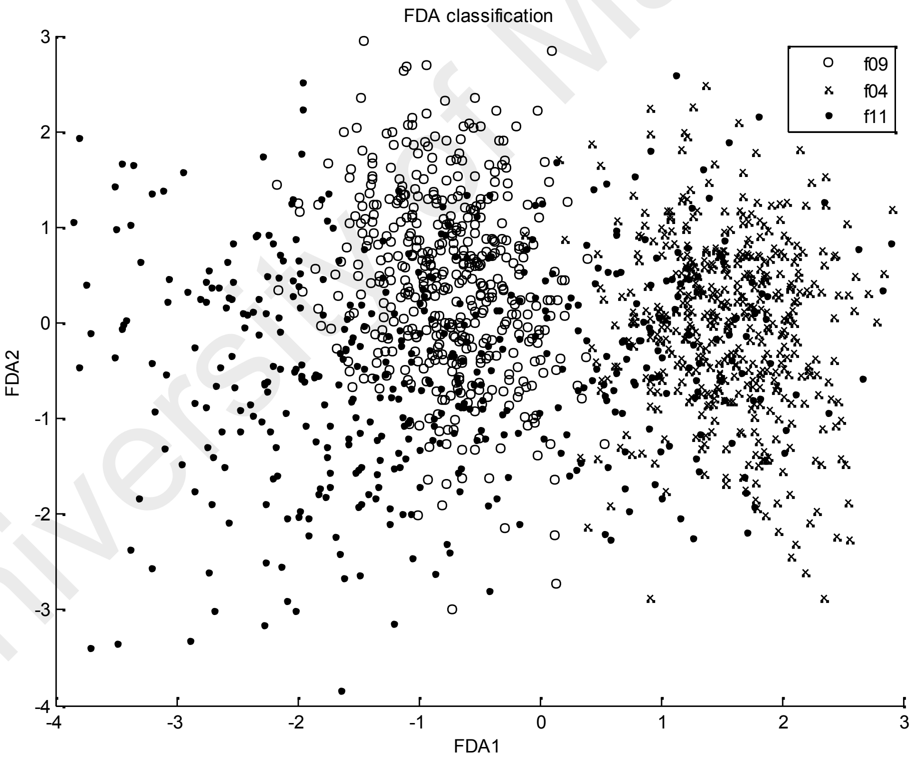

Figure 3.7: Fault 4, Fault 9 and Fault 11 classification projection using normal FDA

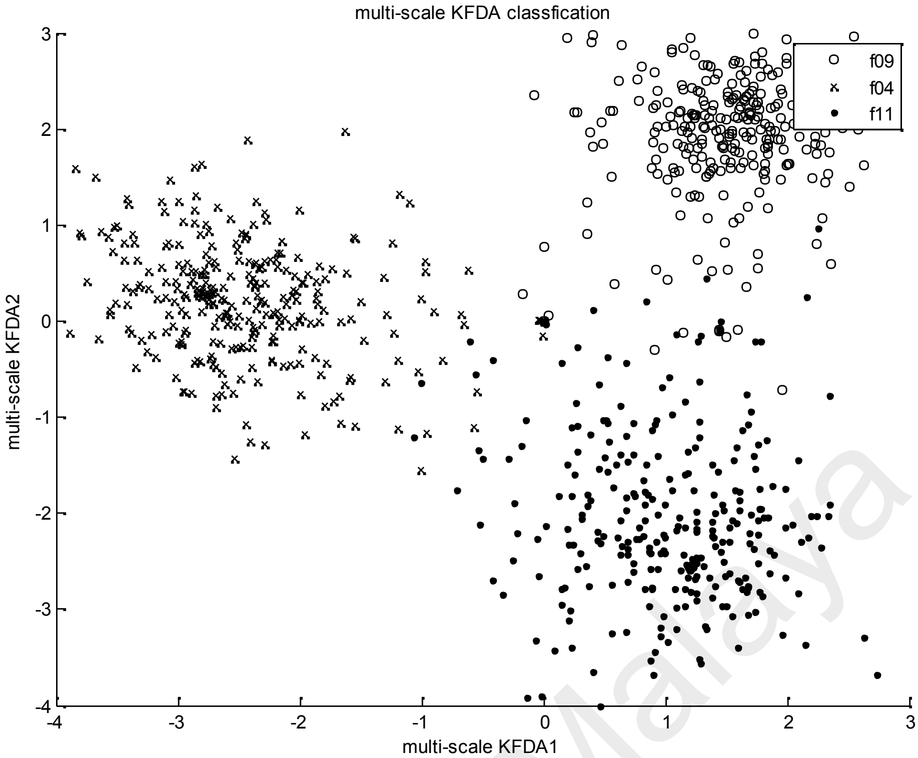

Figure 3.8: Fault 4, Fault 9, and Fault 11 classification projection using multi-scale KFDA

3.8 Summary

In this chapter, multi-scale KFDA-based feature extraction for the Tennessee Eastman process is presented. Data discrimination based on the proposed multi-scale KFDA methodology enhanced the extraction by taking into consideration the multi-scale information compared to other methods that considered only single scale nature. It also can provide a better separation of the features and improve the extraction of features that are relevant to a faulty situation from both time and frequency domain. The application of the proposed multi-scale KFDA also shows better fault identification performance than KFDA and FDA. The high misclassification rate for FDA also shows the advantage of nonlinear technique when the fault data are highly overlapped.

CHAPTER 4: MULTI-SCALE KERNEL FISHER DISCRIMINANT ANALYSIS DATA-DRIVEN FAULT DETECTION AND IDENTIFICATION METHODS

4.1 Introduction

Four different types of data-driven methods were combined with multi-scale KFDA feature extraction method for FDI tasks in this chapter. The data-driven methods, namely ANFIS, SVM, GMM, and *k*NN were implemented individually, to form multi-scale KFDA-ANFIS, multi-scale KFDA-SVM, multi-scale KFDA-GMM, and multi-scale KFDA-*k*NN methods, respectively.

As seen in the flowchart of Figure 4.1, the strategy consisted of two major steps after data acquisition and normalisation, which were the feature extraction step and the fault classification step. Firstly, after data acquisition from the process system, the input database consisting of the normal and faulty data was pre-processed using the normalisation method. Normalisation of the data was done by ensuring each of the variables was linearly scaled in the range of [0, 1] to avoid domination of the larger numerical range over the smaller numerical range.

Then, the normalised data were fed to the multi-scale KFDA feature extraction step. In this step, the DWT method was applied with its multi-scale feature decomposition to give the distinguished characteristic features of the input data. DWT decomposed each of the variables in the input data set individually, before the significant wavelet coefficients were retained. The retained wavelet coefficients will have higher values compared to the threshold value when a significant event occurs.

After extracting the input features, the DWT reconstruction method called IDWT was applied to the retained coefficients. Through these procedures, the dimensions of the input patterns can be reduced and useful information can be extracted.

Consecutively, the reduced dimensional database was separated by the KFDA method into the discriminative feature space, before it was fed to the pattern classifier to perform the next step, the fault classification step. In fault classification step, the data-driven methods were applied to the framework to find the patterns in the extracted multi-scale KFDA subspaces. If any faults were detected, the classifier will diagnose them by assigning the features to the corresponding detected fault classes. If the classifier detected normal condition, the data used by the classifiers will be stored into the database as normal data.

Each of the proposed multi-scale KFDA-data-driven-based method was applied with designated faults of the Tennessee Eastman process, namely Fault 1 until Fault 21 (as discussed in the previous chapter). The proposed methods were implemented to detect and diagnose the faults. The training data were used to develop data-driven classification models, while the testing data were used to evaluate the classification performance. The details of the steps and performance of the proposed multi-scale KFDA-based classification framework are further explained in the following sections.

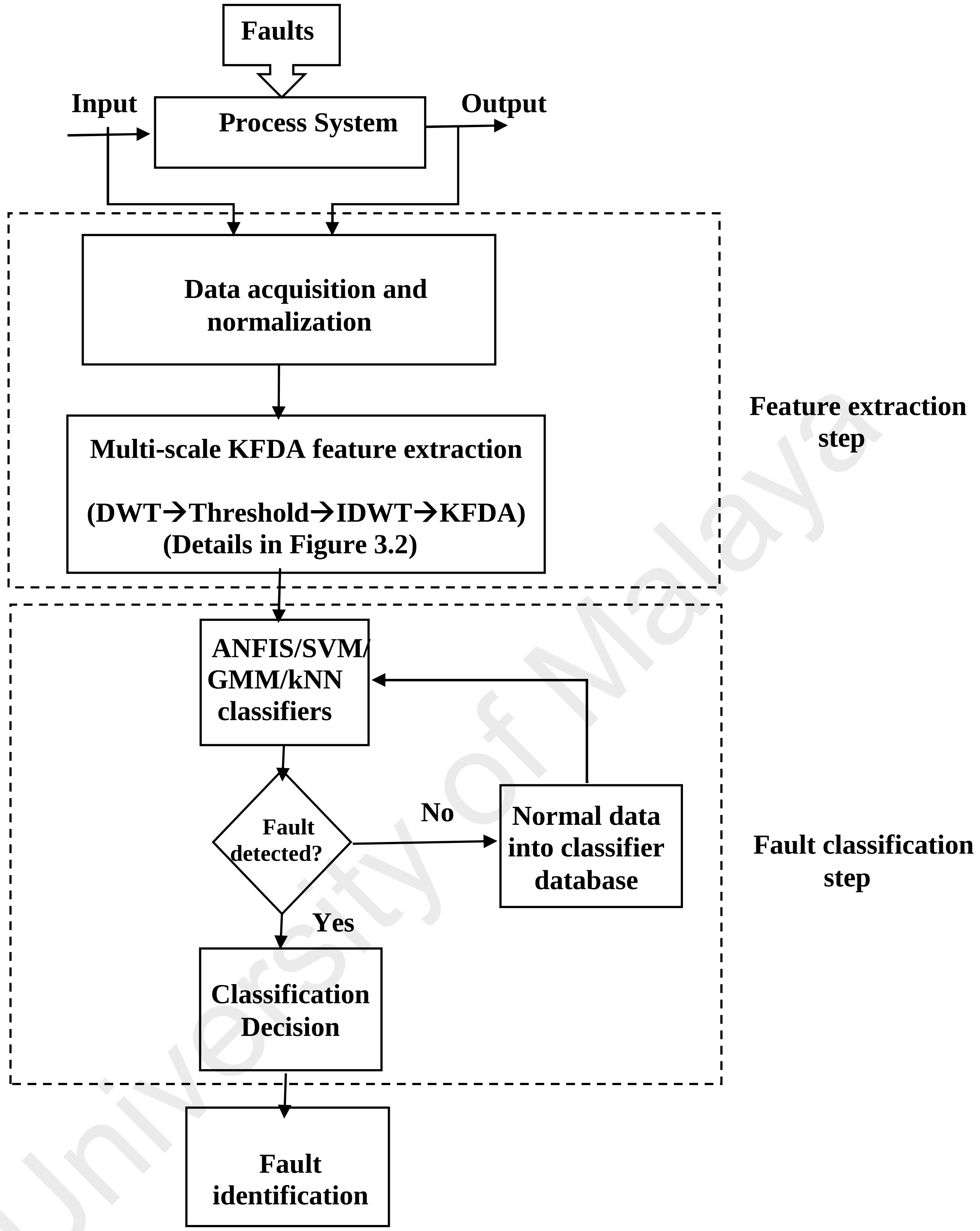

Figure 4.1: General flowchart of the proposed multi-scale KFDA-based data-driven FDI method

4.2 Multi-scale KFDA-ANFIS Method

For the proposed multi-scale KFDA-ANFIS FDI method, the ANFIS-based fault classification method was selected. Generally, in ANFIS-based fault classification, the first step was to train the ANFIS classifiers. During the training phase, ANFIS learned and extracted the fault-symptom relationship from the data by changing its parameters based on a hybrid learning algorithm. The training targets were set at 0.1 and 0.9 for the normal and faulty conditions, respectively. The learning algorithm applied was a combination of least-squares and gradient descent methods, involving forward and backward passes. In forward pass, the least-squares estimation method identified the consequent parameters when the node outputs advance to the output membership function (MF) layer. Meanwhile, the gradient descent method updated the premise parameters via back-propagation error signals.

As opposed to the traditional ANFIS classifier which is developed to detect all faults in one network, the proposed ANFIS fault classification framework applied one ANFIS classifier to diagnose one specific fault, as shown in Figure 4.2. Hence, the number of faults occurring in the database determined the number of ANFIS classifiers constructed for the system, with each of the classifier dedicated to one specific fault. Only the specific ANFIS classifier will respond to and be triggered by the respective fault, as the classifier output exceeded the threshold value, while other classifiers remained under the threshold limit. The ANFIS model which exceeded the threshold value will be identified as showing the right class of the fault. The multiple ANFIS classifiers proposed in this work helped to reduce the computation load, as complex chemical processes usually have high number of variables. Furthermore, they also increased the efficiency of the fault classification, as well as reducing the time for network training.

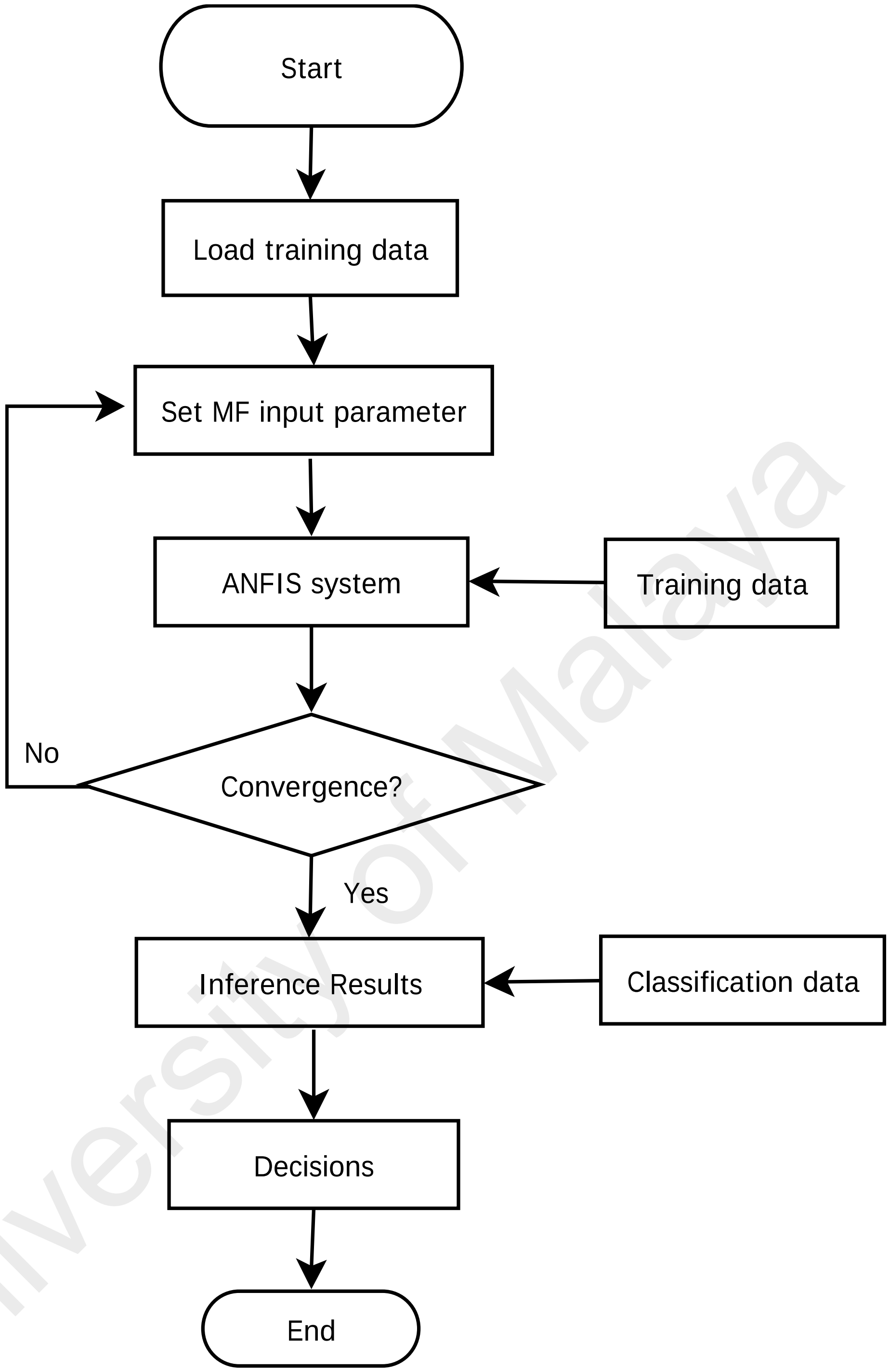

Figure 4.2: Flowchart of ANFIS hybrid learning algorithm

The multiple-input multiple-output ANFIS classification system was developed based on multiple ANFIS network. It was trained and developed for all faults designated from the Tennessee Eastman process to avoid the complexity of the network structure.

This first-order Takagi-Sugeno-type ANFIS with three Gaussian MFs per input was developed and trained using the MATLAB fuzzy logic toolbox. It was trained based on supervised learning and its goal was to be able to approximate unknown functions given by the training data. It should be noted that the classification performance of ANFIS could be significantly enhanced if a first-order system was used with more MF per input. However, the main obstacle is that such system requires massive computational and storage facilities to adapt and save the parameters (Awadallah & Morcos, 2006).

These ANFIS classifiers were trained and obtained for each of the fault designated before the multiple ANFIS scheme was tested on each type of fault, independently. Different types of MFs were tested, namely generalised (‘*gbellmf*’), symmetric Gaussian (‘*gaussmf*’), and triangular (‘*trimf*’) MF. The system parameters and the chosen MFs were automatically tuned during the learning process via a hybrid learning algorithm of gradient descent approach and least-squares estimate to increase the training efficiency. The convergence of the root mean squared (RMS) error was utilised to evaluate the learning process. If the decreasing rate of the RMS error and the performance was not significant, the learning process can be terminated. In other words, the premise parameters of the MFs corresponding to the inputs were changed to eliminate the possibility of local minima trapping based on the training samples.

Once the type and number of MFs were determined, a fuzzy inference system was initially constructed in each ANFIS network using fuzzy clustering of the measurement data. An error was realised from the difference between the desired response and the corresponding FDI system output. This error information was fed back to the relevant FDI system to fine-tune each individual ANFIS network through hybrid and back-propagation learning algorithms.

Through this proposed multi-scale KFDA-ANFIS framework, the FDI can be carried out simultaneously while reducing the training time for the multiple ANFIS classifiers. Apart from that, the computation load and the network structure can also be simplified via this approach. Notably, the proposed fault identification system can be expanded by adding a new ANFIS classifier into the classification system when a new type of fault is identified, making it modular and versatile in nature.

4.2.1 Performance of Multi-scale KFDA-ANFIS

The proposed improved multi-scale KFDA-ANFIS FDI framework was applied to all faults designated to the Tennessee Eastman process, namely Fault 1 to Fault 21. The proposed framework was implemented to detect and diagnose the faults, with the data for training used for framework modelling. Meanwhile, the data for testing were used to evaluate the classification results through the accuracy rate, as shown in Eq. 3.22 earlier. Each ANFIS was uniquely trained to detect and identify its assigned category of process faults. Therefore, only the respective ANFIS classifier will be triggered to indicate the occurrence of the assigned fault. A confusion matrix analysis as in Figure 3.5 was generated to show the accuracy of the proposed method. This confusion matrix contains information about the actual and predicted classifications done by the multi-scale KFDA-ANFIS approach.

From the confusion matrix, an accuracy can be defined as the proportion of the total number of classification that is correct. This was one of the key aspects to determine the success of this proposed approach. For example, Figure 4.3 shows a confusion matrix for Fault 8 classification. From the figure, the accuracy of the classification can be summarised as 92.3%, as indicated by the confusion matrix. Figure

4.4 shows a confusion matrix for Fault 5 classification. From the figure, the accuracy of the classification can be summarised as 97.8%, as indicated by the confusion matrix. Additionally, the occurrence of false fault and missing fault, which are known as type I and type II errors, can also be observed through the confusion matrix. Generally, an increase in the number of local model will increase the type I errors. On the contrary, the type I error will occasionally decrease as the number of local model decreases because the convergence of the model was easier and faster to achieve. However, it can be accompanied by an increase in the occurrence of the type II errors.

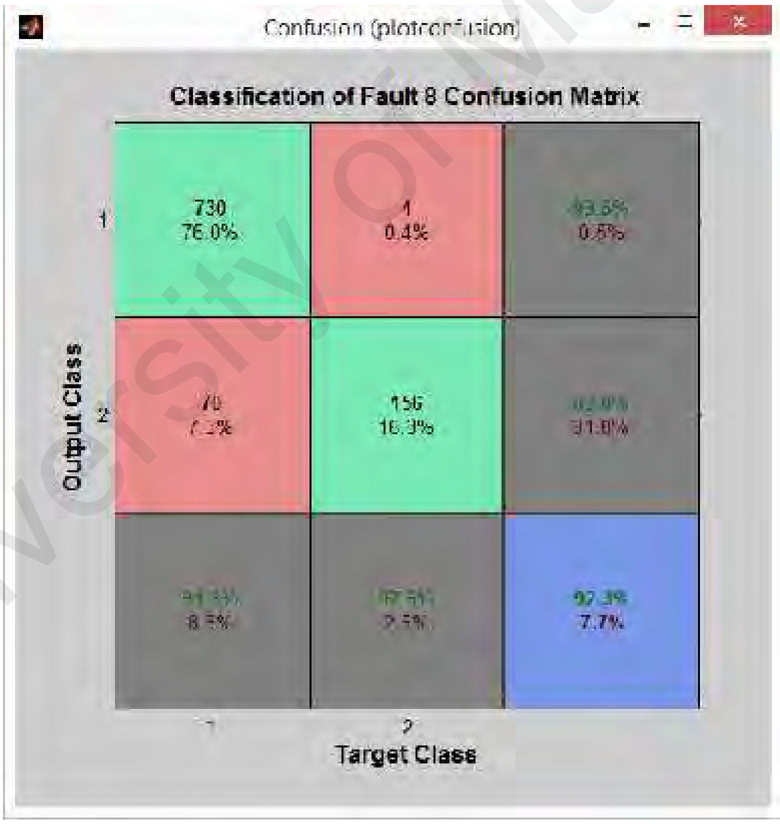

Figure 4.3: Confusion matrix for Fault 8 classification

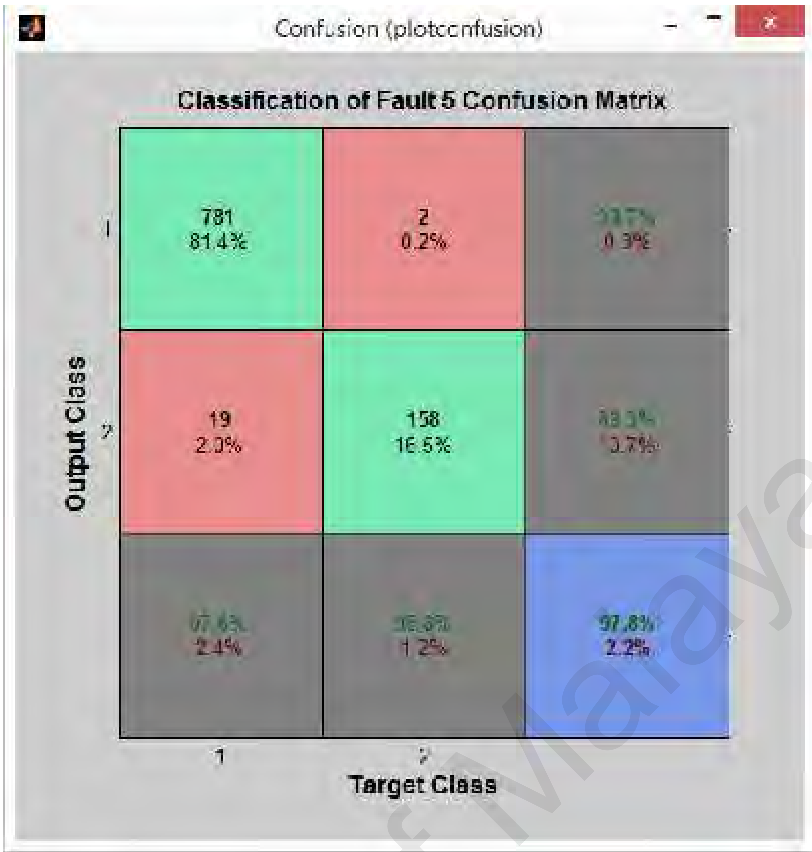

Figure 4.4: Confusion matrix for Fault 5 classification

The aim of classification was to design an input pattern to one of the 21 classes concerned in the TEP represented by the classification labels defined earlier. Table 4.1 presents the classification accuracy rate for the proposed multi-scale KFDA-ANFIS, compared with FDA-ANFIS method and multi-scale PCA-ANFIS method proposed by Lau and coworkers in 2013, for all 21 types of fault classes. As shown in Table 4.1, the results of classification accuracy demonstrated that the detection accuracies were considered to be averagely improved by employing the multi-scale KFDA in the ANFIS-based FDI framework. Even though the fault classification rates for Fault 3, Fault 9, and Fault 15 were quite low for both of the methods, the proposed method had shown promising results as these three faults (Fault 3, Fault 9, and Fault 15) achieved average classification rates of nearly 43.67%. The classification results of multi-scale KFDA-ANFIS were also higher than the multi-scale PCA-ANFIS methods, as recorded for Fault 2, Fault 5, Fault 6, Fault 10, Fault 16, and Fault 21.

Table 4.1: Identification accuracy rate using proposed multi-scale KFDA-ANFIS for all faults of TEP

	Multi-scale KFDA-ANFIS	**FDA-ANFIS**	**Multi-scale PCA-ANFIS** (Lau et al., 2013)
Fault 1	99.12	90.00	96.20
Fault 2	98.64	90.00	93.70
Fault 3	33.72	10.91	4.60
Fault 4	94.39	90.26	97.10
Fault 5	97.85	88.96	93.60
Fault 6	95.26	90.19	97.10
Fault 7	99.78	90.05	97.00
Fault 8	94.77	90.24	96.30
Fault 9	55.06	4.30	7.10
Fault 10	90.12	66.86	71.00
Fault 11	90.88	70.13	90.50
Fault 12	94.36	90.34	96.30
Fault 13	91.20	70.05	92.20
Fault 14	94.42	70.58	97.50
Fault 15	55.27	6.96	12.50
Fault 16	95.88	70.84	79.60
Fault 17	96.54	80.41	95.20
Fault 18	92.85	80.19	84.70
Fault 19	90.66	88.85	95.50
Fault 20	91.24	76.25	89.40
Fault 21	75.41	70.37	69.80
Average	**87.02**	**70.80**	**78.90**

Furthermore, the proposed multi-scale KFDA-ANFIS-based framework combined fault detection and fault identification in one system that can be maintained easily. The ability of ANFIS to adapt and model nonlinear process makes it suitable for efficient process monitoring even if the operating condition of the process has changed.

Moreover, the proposed multi-scale KFDA-ANFIS method classified each fault within an average of 0.2 milliseconds. The computational time for fault classification was limited by the computational ability of the hardware. Thus, it could be enhanced by improving the computer specifications such as the random access memory (RAM) and the processors.

4.3 Multi-scale KFDA-SVM Method

The SVM method is a discriminative classifier which has been widely used in classification tasks. It is often considered as a state-of-the-art classifier since it provides good generalisation, although it may not be the best for every case. SVM is a supervised learning method used for classification and regression based on statistical learning theory. By applying a nonlinear kernel function that maps data points into high-dimensional feature space, SVM can also treat nonlinear classification problem via a large margin hyperplane. Common kernel functions include (i) polynomial, (ii) radial basis function, (iii) linear, and (iv) sigmoid. According to the different classification problems, the different kernel functions can be selected to obtain the optimal classification results.

For example, giving n samples = $\{x_i, y_i\}_{i=1}^{n}$, $y_i \in \{-1, +1\}$, where x_i represents the condition attributes, y_i is the class label, and i is the number of samples. The decision hyperplane of SVM can be defined as (w, b), where w is a weight vector and b is a bias. Let w_0 and b_0 denote the optimal values of the weight vector and bias. Correspondingly, the optimal hyperplane can be written as

$$w_0^T x + b_0 = 0. \tag{4.1}$$

In order to find the optimum values of w and b, the following optimisation problem has to be solved. The C-SVM primal binary optimisation problem for a training data $x_i \in R^n, i = 1, \ldots, l$, and an indicator vector $y \in R^l$ such that $y_i \in \{1, -1\}$ are formulated as

$$\min_{w,b,\xi} \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i \tag{4.2}$$

subject to $y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i$ with $\xi_i \geq 0, i = 1, \ldots, l$, where $\phi(x_i)$ function mapped x_i into a higher-dimensional space, w is the weight vector, b is the bias, ξ is the slack variable, and $C > 0$ is the regularisation parameter. Eq. 4.2 proves that the suitable penalty parameter, C, is subtle to the kernel function types and parameters.

Therefore, the radial basis function (RBF) which is one of most widely used kernel functions in SVM applications, has been employed in this work due to a number of desirable properties. For instance, RBF includes other kernels as special cases and avoids difficulties associated with very large numbers because its values range between 0 and 1. Moreover, it is well known that the SVM classification performance depends on the choice of its parameter settings. For example, through RBF kernel function, only two parameters, C and γ, need to be considered to facilitate the classifier in its classifying task. For this reason, the RBF kernel function for the C-SVM model can be expressed as

$$k(x_i, x_j') = \exp\left(-\gamma \| x_i - x_j' \|^2\right), \gamma > 0. \tag{4.3}$$

These algorithms were implemented in MATLAB, where the C-SVM classifier was developed for each designated fault. Generally, the selection of the regularisation parameter, C, the upper bound on the fraction of support vector, ν, and the RBF kernel parameter, γ, were chosen through a trial-and-error method before the training of the

SVM. Initially, each C and γ values were fixed at 0.1, before they were gradually tuned to increase the SVM classification performance.

The predict class of the testing data could be determined after solving the SVM formulation. For example, let $x1, \ldots, xl$ be the testing data and $f\ x1\ , \ldots, f(xl)$ be the decision values predicted. If the true classes of testing data are known and denoted as $yi, \ldots, yl$, the predicted results could be evaluated by the classification accuracy. The classification accuracy is the percentage of number of correctly predicted data with respect to the total number of testing data. Finally, the proposed multi-scale KFDA-SVM framework was developed based on the combination of the highest classification performance of SVM on each designated fault.

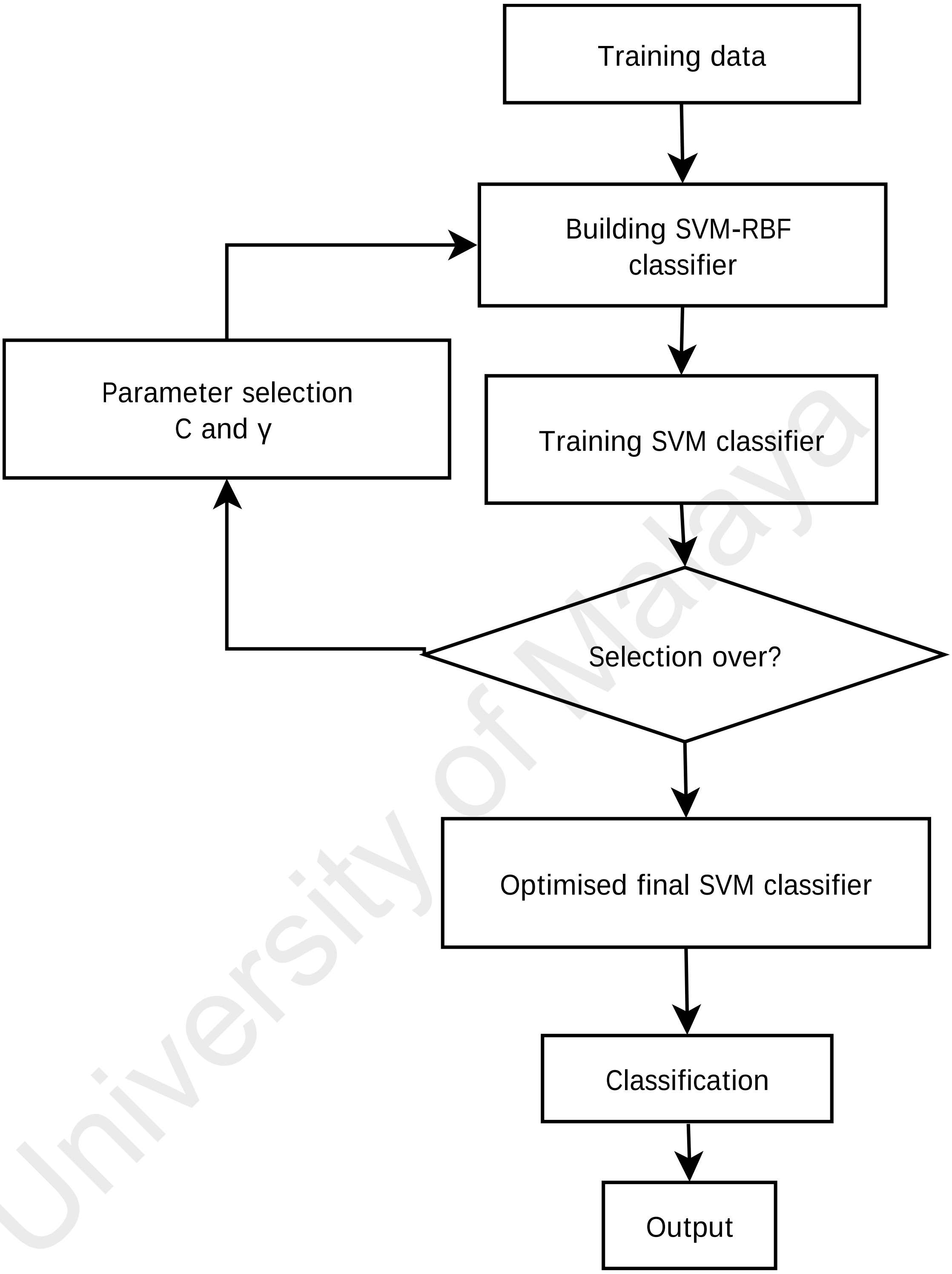

Figure 4.5: Flow-chart of algorithms based on SVM

4.3.1 Performance of Multi-scale KFDA-SVM

On the other hand, the performance of the multi-scale KFDA-SVM method, which combined the proposed multi-scale KFDA feature extraction method with the SVM classifier, was also evaluated and discussed. Figure 4.6 and Figure 4.7 show the final identification of Fault 5 and Fault 16 based on the performance of the multi-scale KFDA-SVM, FDA-SVM, and single SVM fault classification frameworks, respectively. Figure 4.6 presents the confusion matrix of the classification accuracy using the multi-scale KFDA-SVM (Figure 4.6[a]), FDA-SVM (Figure 4.6[b]), and SVM (Figure 4.6[c]) frameworks on Fault 5 of Tennessee Eastman process data. Fault 5 was related to the step change in the condenser cooling water inlet temperature. When this fault occurred, the outlet stream flow rate from the condenser to the vapour-liquid separator also increased. Hence, this increased the separator cooling water outlet temperature and initiated unwanted changes in the process temperature in the system.

Figure 4.6 also demonstrates the comparison of the classification accuracy results for Fault 5 using the proposed multi-scale KFDA-SVM with FDA-SVM and SVM fault detection frameworks. Both the multi-scale KFDA-SVM and FDA-SVM methods can classify the fault data with 99.5% and 91.4% accuracy, respectively. Meanwhile, the SVM classifier without any feature extraction method attained 76.8% accuracy for Fault 5 classification. Notably, the proposed multi-scale framework revealed much higher accuracy in classification compared to the FDA-SVM framework. This was because the combination of wavelet analysis with KFDA has the ability to extract richer faulty information from the measured variables, and to maximise and minimise the between-class and within-class data for better classification performance by these classifiers.

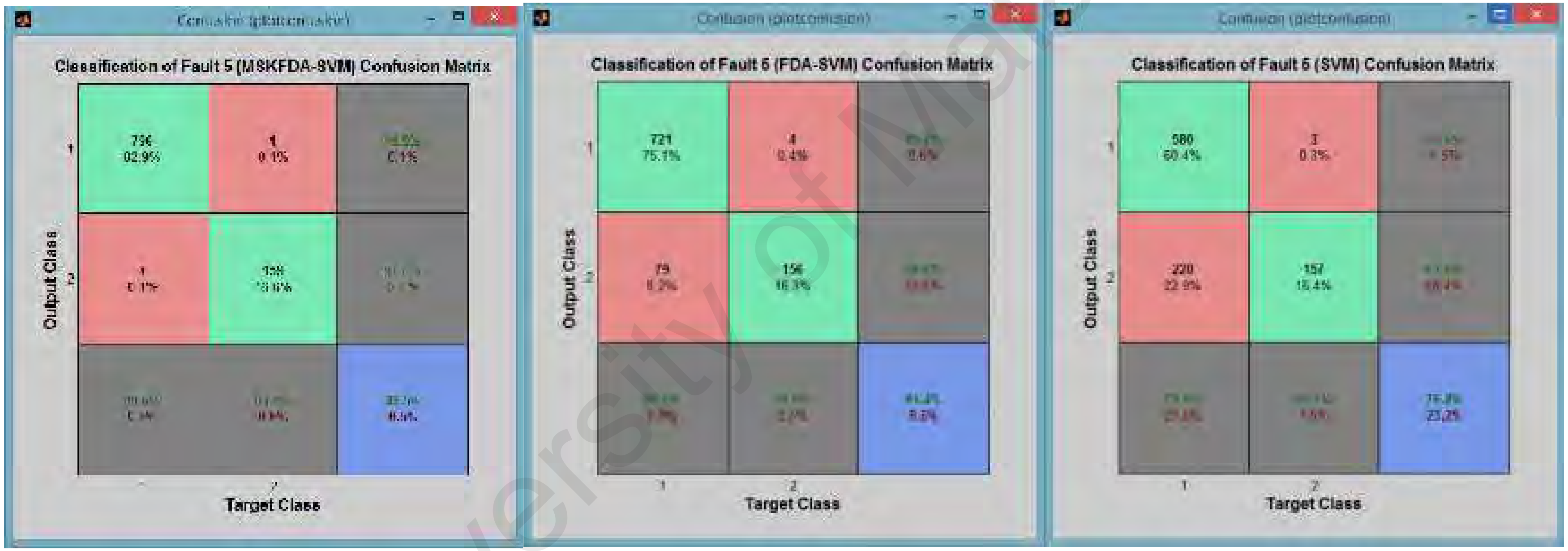

Figure 4.6: Confusion matrices for classification accuracy of Fault 5 for (a) multi-scale KFDA-SVM, (b) FDA-SVM, and (c) SVM framework

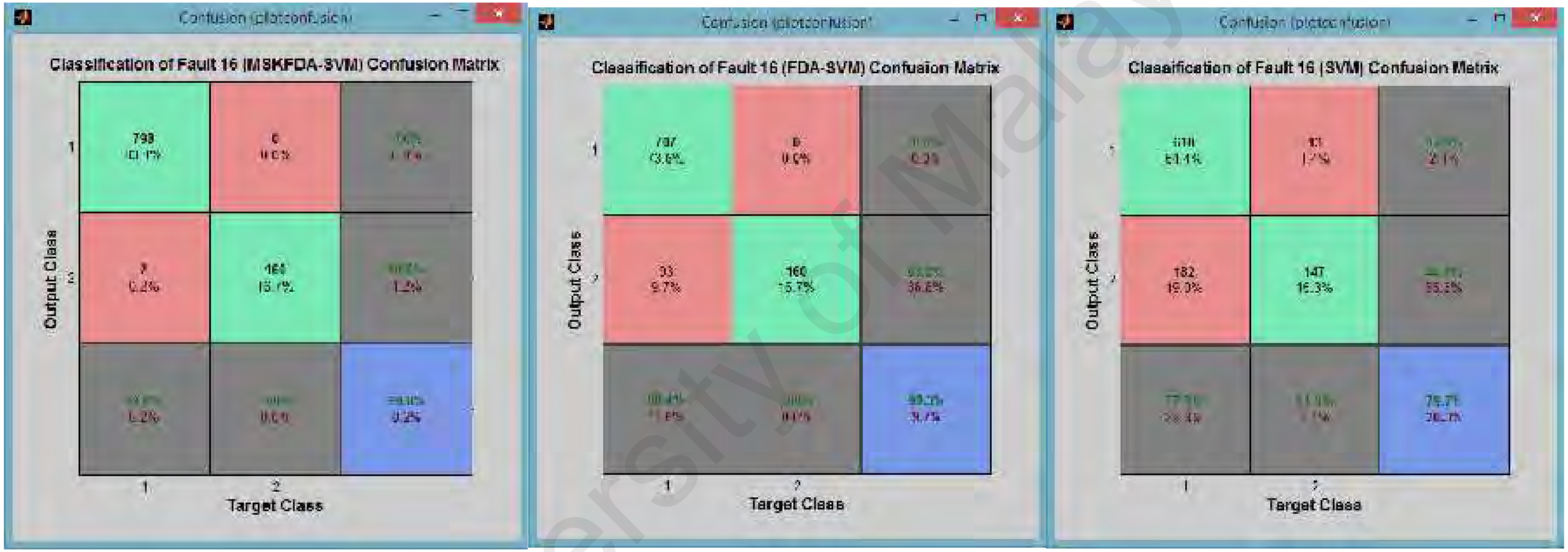

Figure 4.7: Confusion matrices for classification accuracy of Fault 16 for (a) multi-scale KFDA-SVM, (b) FDA-SVM, and (c) SVM framework

The classification results for Fault 16 are shown in Figure 4.7; Fault 16 was an unknown fault in which the root cause was unidentified. This type of fault produces multi-scale deviation in the measured variables, thus, it may lead to difficulties in classification. However, Figure 4.7(a) depicts that the proposed multi-scale KFDA-SVM framework classified the fault with 99.8% accuracy, compared to FDA-SVM and SVM frameworks which only achieved the accuracy of 90.3% and 79.7%, respectively. The higher accuracy achieved by the proposed framework indicated that the multi-scale feature extraction method filtered the data in both time and frequency levels for better classification compared to the other methods.

Table 4.2 presents the classification results of the proposed multi-scale KFDA-SVM and the comparison for all the 21 types of fault classes. It is worthy to note that the multi-scale KFDA-SVM offered higher accuracy rate for TEP classification cases, compared to the other methods. Comparing the identification accuracy percentages presented in the table, it can be observed that the average diagnostic accuracy percentage for the multi-scale KFDA-SVM was significantly higher for most of the faults. Basically, on average, the KFDA-SVM produced higher identification accuracy, even though few unobservable faults such as Fault 9 and Fault 15 had closely similar accuracy. Nonetheless, the detection of faults for Fault 3, Fault 9, Fault 15, and Fault 21 was very difficult as there was no observable change in the means, variance, or the peak time, which makes it very difficult to isolate them from the normality. Because of this, they were excluded in most of the studies by other researchers. The simulation results of the classification accuracy without Fault 3, Fault 9, Fault 15, and Fault 21 are given in Table 4.2.

In effort to further demonstrate the superiority of the proposed multi-scale KFDA feature extraction, a comparison was made to the classification results of the

SVM and ICA-SVM, which were proposed by Mahadevan and Shah (2009), and Hsu and coworkers, as described in Hsu et al. (2013). It is apparent from this table that multi-scale KFDA-SVM had the highest average in classification accuracy (96.79%) compared to ICA-SVM (95.78%) and SVM (83.63%). These present findings also suggested that the proposed multi-scale KFDA-SVM has more capability in diagnosing faults when compared with a traditional combination method like ICA-SVM. Although ICA can extract rich information of the input data, it is still a linear learning technique which may not fit very well for this type of data structure in the TEP process.

Table 4.2: Classification accuracy rate using multi-scale KFDA-SVM for all faults of TEP

	Accuracy rate (%)		
	Multi-scale KFDA-SVM	SVM (Mahadevan & Shah, 2009)	ICA-SVM Hsu et al. (2013)
Fault 1	99.80	99.8	99.6
Fault 2	99.87	98.6	98.1
Fault 3	43.33	-	
Fault 4	95.89	99.6	99.1
Fault 5	99.50	100	99.0
Fault 6	96.02	100	100
Fault 7	99.44	100	99.1
Fault 8	95.81	97.9	97.8
Fault 9	51.56	-	
Fault 10	93.99	87.6	89.4
Fault 11	95.38	69.8	82.7
Fault 12	95.63	99.9	99.5
Fault 13	95.24	95.5	95.5
Fault 14	98.05	100	99.0
Fault 15	64.74	-	
Fault 16	97.80	89.8	93.0
Fault 17	97.32	95.3	95.7
Fault 18	95.15	90.0	91.5
Fault 19	96.87	83.9	96.6
Fault 20	93.63	90.0	92.7
Fault 21	87.64	52.8	
Average	**90.13**	**83.63**	**95.78**

4.4 Multi-scale KFDA-GMM Method

Gaussian mixture model (GMM) is a generative classifier based on a probability density function (*pdf*) distribution. It is composed of a mixture of two Gaussian models widely used for modelling *pdf* in practice. The model offers a good trade-off between the complexity of estimating the parameters of the GMM and the capability of the GMM for accurately modelling underlying density. The main idea of this method is the assumption that the combination of local linear models is capable of complex process description. However, the GMMs require intensive computations and parameter optimisation, as well as have unclear choice of the numbers of mixtures.

Therefore, the GMM is constructed based on expectation-maximisation (EM) method to simultaneously find the optimal parameters, where this set of parameters will iteratively approximate the data distribution and pattern similarity (Zhu & Song, 2011). In this study, the EM algorithm found the optimum model parameters by iteratively refining GMM parameters so that the likelihood of the developed model for the designated process fault feature vectors can be increased.

In this work, the GMM classifier was developed to describe and classify the faulty process patterns and each fault was represented by one local GMM. However, the faulty pattern, especially in highly complex systems, could be difficult to characterise due to the limitation in a single Gaussian model-based application. In effort to overcome this situation, the number of the Gaussian model needs to be increased to represent every fault. Thus, for the local Gaussian model development, the mean and covariance of each fault data were taken as initial parameters before the parameters were fine-tuned throughout the model training phase. The detailed description of the development of mixture of Gaussian distribution can be referred to in Li and Qin (2016).

For example, an arbitrary probability density of a sample vector, x, $p(x|\theta)$, can be estimated by a mixture of Gaussian density function as follows:

$$p(x|\theta) = \sum_{k=1}^{K} p(x|k)\,P(k) \tag{4.4}$$

where θ is a parameter vector whose entries are model parameters, $p(x|k)$ is the component density, and $P(k)$ is the prior probability of the data point generated from component k of the mixture, satisfying $\sum_k P(k) = 1$ and $0 \leq P(k) \leq 1$.

Then, the posterior probability gives the probability that a data point is assigned to a particular cluster. The posterior probability can be obtained using Bayes' theorem:

$$p(k|x) = \frac{p(x|k)P(k)}{\sum_{i=1}^{K} p(x|i)\,P(i)} = \frac{G_x[\mu_k, \sum_k]P(k)}{\sum_{i=1}^{K} G_x[\mu_i, \sum_i]P(i)} \tag{4.5}$$

where $G_x[\mu_k, \sum_k]$ is a Gaussian density function of x with mean of μ_k and covariance matrix, $\sum_k$. Three kinds of parameters, - μ_k, $\sum_k$, and $P(k)$, must be estimated from the observed data $x^N = \{x_1, x_2, \ldots, x_N\}$. Denoting the responsibility $R_{k,n}$ with $P(k|x_n, \theta)$, the estimated parameters, μ_k, $\sum_k$ and $P(k)$, can be determined from

$$\mu_k^{new} = \frac{\sum_{n=1}^{N} R_{k,n} x_n}{\sum_{n=1}^{N} R_{k,n}} \in \mathbb{R}^m \tag{4.6}$$

$$\sum_k^{new} = \frac{\sum_{n=1}^{N} R_{k,n} x_n x'_n}{\sum_{n=1}^{N} R_{k,n}} - (\mu_k^{new})(\mu_k^{new})^T \in \mathbb{R}^{m \times m} \tag{4.7}$$

$$P(k) = \frac{1}{N}\sum_{n=1}^{N} R_{k,n}, \text{ for a given } R_{k,n}. \tag{4.8}$$

Eq. 4.5 until Eq. 4.8 are the EM algorithms. The expectation step (E-step) of the EM algorithm, shown in Eq. 4.5, involved obtaining the posterior probability, given all the parameters and the prior probability. The maximisation step (M-step), represented by Eq. 4.6 to Eq. 4.8, involved finding all the parameters, given the posterior probability. By repeating the E-step and M-step in an iterative procedure, the calculation converged to a stable solution that represents the maximum likelihood solution of the problem.

The fault classification was then achieved by computing the likelihood of the unknown vibration segment of different fault models. This likelihood was given by

$$\hat{s} = \arg\max_{1\le f\le F} \sum_{k=1}^{K} \log p(x_k|\lambda_f) \tag{4.9}$$

where F represents the number of faults to be diagonalised, $X = \{x_1, x_2, \dots, x_K\}$ is the unknown D-dimension bearing fault vibration segment, and $p(x_k|\lambda_f)$ is the mixture density function given by

$$p(x|\lambda) = \sum_{i=1}^{M} w_i\, p(x) \tag{4.10}$$

with

$$p_i(x_i) = \frac{1}{(2\pi)^{D/2}\sqrt{\Sigma_i}} \exp\{-\frac{1}{2}(x_k - \mu_i)^T (\Sigma_i)^{-1}(x_k - \mu_i)\} \tag{4.11}$$

The mixture weights, w_i, satisfy the constrains,

$$\overset{M}{\underset{i=1}{}} \; w \qquad (4.12)$$

Finally, the process fault classification by the proposed multi-scale KFDA-GMM was completed by calculating the probability of the input feature vector.

4.5 Multi-scale KFDA-*k*NN Method

K-nearest neighbor (*k*NN) is a non-parametric classification method based on the distance between the labelled training samples and the process data. The process data will be assigned to the designated class if the distances of the nearest samples of the data are in the specified class boundary. Its performance can be very sensitive to the influence of outliers and the choice of the number of neighbors to be considered (Arruda, Bandeira, Soares, & Laureano, 2014). For the proposed multi-scale KFDA-*k*NN FDI framework, the *k*NN method was selected for the fault classification task.

The *k*NN classification model was developed based on the estimated probability of the test sample. Then, the distance between the test sample and training sample was computed and sorted. As for final decision, the *k* value that was the nearest to the neighbors was selected for the classification model. Although simple, it was comparable to more sophisticated and complex methods. For example, as P features were picked up from a sample, in the *k*NN classification method, each training record was described in a P-dimensional space based on the value of each of its P input characteristic. The testing sample was then expressed in the same form, and its k nearest neighbors were chosen. Subsequently, the category of each of these k neighbors was counted, and the category with maximum vote was designated as the result of the unknown sample.

The kNNs were generally decided by calculating the Euclidean distance between the testing record and each of the training one. For instance, the Euclidean distance, ED_q, between the testing sample, TE_p, and the qth training sample TR_{qp} is defined as

$$ED_q = \sqrt{\sum_{p=1}^{P} (TE_p - TR_{qp})^2}, \qquad p = 1, 2, \ldots, P, q = 1, 2, \ldots, Q \tag{4.9}$$

where P and Q indicated the figures of input characteristics and training records, respectively. The k elements were called kNN and the algorithm checks which were the classes of k-NN and the most frequent class was assigned to the unknown element class.

The classifier method had a set of classifiers, $\mathbb{F} = \{F_1, F_2, \ldots, F_m\}$, with each classifier performing a mapping of an instance vector, $x \in \mathbb{R}^D$, into the set of labels $Y = \{1, \ldots, k\}$. The design of classifier ensembles must perform two main tasks: constructing the individuals classifiers, F_i, and developing a combination rule that finds a class label for x based on the outputs of the classifiers $\{F_1(x), F_2(x), \ldots, F_m(x),\}$.

4.5.1 Performance of Multi-scale KFDA-GMM and multi-scale KFDA-*k*NN

The multi-class classification problem was further studied with the implementation of multi-scale KFDA with the GMM and *k*NN classifiers. The proposed methods were implemented to detect the faults of the Tennessee Eastman process. Table 4.3 shows the identification accuracy rates for both methods.

Table 4.3: Identification accuracy rate using proposed multi-scale KFDA-GMM and KFDA-kNN for all faults of TEP

	Multi-scale KFDA-GMM	**Multi-scale KFDA-kNN**
Fault 1	98.00	94.00
Fault 2	98.37	95.37
Fault 3	15.75	35.75
Fault 4	62.75	92.75
Fault 5	97.62	97.80
Fault 6	89.00	99.00
Fault 7	95.75	95.75
Fault 8	54.37	92.37
Fault 9	56.50	36.50
Fault 10	23.62	83.62
Fault 11	26.75	89.75
Fault 12	78.12	94.12
Fault 13	73.87	83.87
Fault 14	85.87	89.97
Fault 15	68.75	58.75
Fault 16	93.62	93.62
Fault 17	98.37	88.37
Fault 18	89.0	89.04
Fault 19	62.87	92.87
Fault 20	62.50	82.50
Fault 21	15.12	88.12
Average	**68.88**	**84.44**

A summary of the classification results for FDA, KFDA-Bayesian, PCA-*k*NN, multi-scale KFDA-GMM, and multi-scale KFDA-*k*NN is also tabulated in Table 4.3. The table lists the average identification accuracy rate by utilising the selected faults of the Tennessee Eastman process data set. Comparing the identification accuracy percentages presented in the table, it can be observed that the diagnostic accuracy percentage for multi-scale KFDA-GMM and multi-scale KFDA-*k*NN were significantly higher than the traditional FDA method. On average, the multi-scale KFDA classifications (multi-scale KFDA-GMM and multi-scale KFDA-*k*NN) also produced higher identification accuracy than KFDA-Bayesian and PCA-*k*NN approaches.

Multi-scale KFDA-*k*NN had the highest average of classification accuracy (84.44%) compared with multi-scale KFDA-GMM (68.88%), FDA (62%) (Chiang et al., 2004), KFDA-Bayes (48.05%) (Liu et al., 2010), and PCA-*k*NN (47%) (Liu et al., 2010). The results in Table 4.4 showed that the proposed classification method based on the multi-scale KFDA with GMM and *k*NN had higher proficiency in diagnosing the faults compared to the other methods.

Table 4.4: Identification Accuracy Using Different Approaches for Selected Faults

	MSKFDA-GMM	**MSKFDA-KNN**	**FDA**	**KFDA-Bayes**	**PCA-KNN**
Identification accuracy (%)	68.88	84.44	62	48.05	47.00

4.6 Summary

This chapter was dedicated to evaluate the integration of multi-scale KFDA method for feature extraction with data-driven classifiers, such as ANFIS, SVM, GMM and *k*NN. By comparing the performance and effectiveness of classification for these proposed frameworks in dealing with the case study, the proposed multi-scale KFDA-based data-driven frameworks successfully detected and classified all types of faults with average accuracies higher than 90%. Specifically, the multi-scale KFDA-SVM FDI framework produced the highest classification performance, with accuracy rate of 90.13% compared to the others.

The overall results of this study show that the multi-scale KFDA-based feature extraction method, which was based on the time and frequency levels of filtration of the process data, contributed to the significant performance of fault classification as compared to the single-scale feature extraction method. Moreover, the combination of the proposed multi-scale KFDA-based feature extraction with data-driven methods, namely SVM, ANFIS, GMM, and *k*NN fault classifications, also proved that the multi-scale KFDA-SVM, multi-scale KFDA-ANFIS, multi-scale KFDA-GMM, and multi-scale KFDA-*k*NN frameworks were suitable for implementation in FDI of chemical processes.

CHAPTER 5: DATA-DRIVEN HYBRIDIZATION WITH MULTI-SCALE KFDA FOR FDI FRAMEWORK

5.1 Introduction

In recent years, combining classifiers instead of using a single one for increasing classification and identification accuracy is an active research area. The hybridization of classifiers is a combination of several classifiers whose individual decisions are combined in specific way to classify the testing samples. It is known that hybrid often produces better performance than the single classifiers that compose it. Therefore, in general, the idea of the hybrid FDI framework is to generate several classifiers and group them in such a way as to improve the performance of any single one.

Most of the hybrid FDI frameworks are developed based on a strategy, which it to combine the outputs of the single classifiers that make up the hybrid in such a way that the correct decisions are amplified, and the incorrect ones are cancelled out. Specifically, the classifiers whose decision boundaries are sufficiently different from those of others are much needed for this work. However, the combination of different classifiers might result in conflicted outputs. Thus, the necessary conditions for the hybrid rule also need to be developed for a given classifiers, including the use of decision combination methods to combine the outputs.

Generally, existing decision combination methods can be classified as utility-based and evidence-based method. Utility-based methods, including simple average and voting techniques provide the modest way to combine decisions. Without any prior knowledge from previous estimations, these methods are only based on some accumulating technique which assesses the combined utility function generated from each classifier. On the other hand, evidence-based methods, including the Bayesian method uses a priori information regarding the previous performance of each classifier to combine decisions. However, all combination methods can be applied with any type of outputs, such as abstract, rank, or measurement form (Ng et al., 2010).

For example, the hybrid model used to detect and classify fault condition are such as Fuzzy Min-Max (FMM) neural network-Classification and Regression Tree (CART) (Seera, Lim, Ishak, & Singh, 2013), SVM with an ANFIS classifier (Salahshoor et al., 2010). A hybrid intelligent model that consists of the FMM neural network and the Random Forest (RF) model comprising an ensemble of CART is developed (Seera, Lim, Nahavandi, & Loo, 2014). The majority voting scheme is used to combine the predictions produced by the resulting FMM-RF ensemble (FMM-RFE) members. Multiple classifiers based on several classification algorithms (MLP network, RBF network, *k*NN algorithm) and input features (Hilbert Transform, residual signal, FFT, WPT, and EMD) are combined with genetic algorithm (GA) (Lei, Zuo, He, & Zi, 2010).

The motivation for the development of the hybrid framework lies in the fact that no single diagnostic method adequately addresses all the challenges of complex, industrial-scale diagnostic problems. Furthermore, a hybrid framework in which different classification methods are combined to perform collective problem solving has also been shown to display a lot of promise. Thus, in this thesis, two different decision combination methods have been applied, which are majority voting method and class-specific Bayesian hybrid method. The data-driven classification methods, including SVM, GMM, *k*NN and ANFIS classifiers are also used to obtain the posterior probabilities of class labels. These probabilities are then combined within these proposed hybrid classification frameworks.

5.2 Guideline on data-driven method selection

Various data-driven methods have been considered to be implemented in FDI systems in chemical processes, where each of this method has their strength, limitation, and formulation towards different kinds of process data characteristics (Zhang & Ge, 2015). However, the selection of suitable FDI method for process system is getting difficult due to the variety of these available methods (Li & Cui, 2009). As we are concerned about the fault detection and identification of a chemical process, the guidelines for method selection have been constructed based on the problem-oriented, generally occurred in process systems. Thus, in this section, the general guidelines on how to make a suitable selection of data-driven FDI method to be applied in chemical process system is provided, as shown in Figure 5.1.

The first problem is regarding a high-dimensional feature vector. Modern chemical processes always consist of several components, parts, or operation equipment, and each of these parts may have significant numbers of measured variables. As a result, the whole process may generate a large number of high-dimensional data samples. To handle these high-dimensional process data is a challenge for the data-driven approaches (Ge et al., 2013). Thus, in order to achieve optimal results, the irrelevant and redundant information hidden in the feature values have to be filtered out, often been done by feature selection or feature extraction. Furthermore, the dimensionality reduction of the feature extraction can be a key factor in reducing the misclassification rate when a pattern classification system is applied to fault identification. Moreover, it is important especially when the dimensionality of the observation space is large while the numbers of observations in the classes are relatively small.

Many multivariate statistical methods have been proposed by researchers for dimensional reduction, such as PCA, PLS, ICA, and FDA-based methods. Meanwhile, the filter-based feature selection methods could be developed from F-statistics, information gain, and mutual information, whereas the wrapper-based feature selection includes genetic algorithm and ANN methods. For example, PCA is employed for feature extraction, where the visualisation of high-dimensional feature vectors is mostly done with the first two values of the PCs of the original feature vector. However, the PCA methods, including nonlinear PCA and ICA, extract latent variables from process measurements under feedback control and monitor changes in the process condition, actuators, sensors, and disturbances. This action could lead to a poor quality product. Meanwhile, PLS including nonlinear PLS uses quality data to guide the decomposition of the process data and extract latent variables, which are most relevant to the product quality. On the other hand, if categorical quality data for multiple fault cases are available, the discriminant partial least squares (DPLS) method provide an alternative to diagnose quality-relevant faults. Based on the reduced variable space, process monitoring can be carried out more easily (Ghosh et al., 2014).

In addition, multi-scale monitoring usually applies the signal pre-processing method in conjunction with PCA and PLS method to further extract features in terms of time scales so that each time scale can be sensitive to certain faults for fault detection and identification. The multiple scale analysis provides the ability to de-noise or filter the data so that uninterested time scale variations can be ignored. With the development of signal processing techniques, many features can be extracted to find their feature patterns and connections to faults, which will provide useful information for fault detection and identification (Dai & Gao, 2013). The features extracted from the output signals can be in time and/or frequency domains, such as signal means, variance, trends, or spectra in a frequency band. Various signal analysis techniques have been introduced, i.e. fast Fourier transform (FFT), Short Time Fourier Transform (STFT), spectral estimation, wavelet transform, and sequence analysis (Agrawal et al., 2014). For instance, continuous wavelet transform (CWT) method allows the acquisition of multi-scale resolutions, while discrete wavelet transform (DWT) has computation efficiency and ability to reduce noise in raw signals. Discrete wavelet packet transform (DWPT) method was also used to enhance the power and flexibility of the DWT, with various adaptive methods developed for the selection of optimal basis wavelets. Although these three signal-based methods are able to work individually for the FDI system, recently there have been many works that combine these methods. For example, DWT analysis is combined with intelligent classification techniques such as artificial neural network and neuro-fuzzy system, to identify the faulty signals.

Next, obviously, for the nonlinearity behaviour of the processes, most of the nonlinear methods for process monitoring have resulted in better performance compared to the linear methods. A related category of nonlinear data-driven methods suitable for process monitoring is from the machine learning methods, including ANN and SVM. While the linear relationship of the data can be easily captured by the traditional multivariate statistical methods, the nonlinear data is difficult to model. For instance, a PCA-based method which assumes linear relationships between variables and Gaussian latent variables, thus, it can be inefficient when dealing with highly nonlinear and have non-Gaussian underlying variables (Slišković et al., 2012). Therefore, kernel learning method has been combined with some traditional methods such as PCA and PLS for nonlinear process monitoring. For example, kernel PLS (KPLS) can efficiently compute regression coefficients in high-dimensional Gaussian feature spaces using nonlinear kernel functions. It also can handle a wide Gaussian range of nonlinearities due to its ability to use different kinds of kernels (Zhang, Teng, & Zhang, 2010). Another example is a kernel FDA (KFDA), which is used instead of the traditional FDA to eliminate the weakness towards the nonlinear processes (Ferreira & Trierweiler, 2009).

After that, the FDI method selection is based on different types of faults that likely been occurred in the process systems. The fault is classified into five different groups; deterministic faults, stochastic faults, slow drift, instrumental faults, and incipient faults. The proposed general guideline is using benchmarked Tennessee Eastman process (TEP) as the example, in selecting an appropriate method for the FDI system. The TEP simulation contains 21 pre-programmed faults, where sixteen of these faults are known, and five are unknown, as shown in Table 3.2. Fault 1-7 are associated with a step change in a process variable, e.g., in the cooling water inlet temperature or in feed composition. Fault 8-12 are associated with an increase in the variability of some process variables, while Fault 13 is a slow drift in the reaction kinetics, and fault 14, 15, and 21 are associated with sticking valves.

The deterministic fault is triggered by a step change in a process variable. In TEP simulation, Fault 1 to 7 are associated with a step change during the process. For example, in Fault 1, a step change is induced in the A/C feed ratio in Stream 4, which results in an increase in the C feed and a decrease in the A feed in Stream 4. Since the ration of the reactants A and C changes, the distribution of the variable that related to material balances such as level, pressure, and composition changed correspondingly. Since more than half of the variables deviate significantly from their normal operating behaviour, detecting and diagnosing such a fault should not be a challenging task. Thus, the PCA-based and CVA-based methods are among the suggested FDI methods to be applied with. Therefore, the process monitoring statistics that show poor performance on Fault 1 are likely to perform poorly on the other faults as well.

Stochastic faults, also known as random faults are triggered with the random variation in the process operating conditions. The process operating state changes with the state is sensitive to the changes of input materials, process fouling, catalyst activity changes, production of different product quality grades and changes in external environment. For example, the disturbances and variability in raw material and operation. This malfunction can be abrupt or gradual degradation of performance, with different rapidity and severity of the fault. For example, Fault 11 of TEP occurs when the random variation is induced to the reactor cooling water inlet temperature. The fault prompted large oscillations in the reactor cooling water flow rate, resulting in a fluctuation of reactor temperature. Due to the dynamic data behaviour, the data sample obtained at the present time may be correlated or different from each other. In order to improve the monitoring performance for dynamic processes, dynamic multivariate statistical analysis methods such as dynamic PCA-based and dynamic ICA-based methods have been suggested to be applied in the FDI system.

Also, it is usual in practice that the normal process data experience slow but normal drift. However, this drift will cause false alarm if not adapted to normal models. For those processes that are slow-varying or have multiple operating conditions, it is difficult to apply the traditional statistical methods, since they are based on the assumption that the process has only one stable operating region. Therefore, problems will arise when those techniques are applied to varying processes. When the process is slow-varying, many adaptive and recursive monitoring approaches have been developed, such as recursive PLS (Li et al., 2005) and recursive PCA (Zhang, Li, & Teng, 2012). Then, moving-window approaches like moving-window PCA were developed to improve the monitoring efficiency of time-varying processes. Furthermore, to monitor nonlinear time-varying processes, a moving window approach was developed when the kernel PCA and kernel ICA methods were used (Jiang & Yan, 2013b; Liu et al., 2009).

For instrumental faults such as sticking valve in reactor cooling water valve, also referred as Fault 14, it is more difficult to diagnose. For this fault, the transient values of the processes must be taken into account, where the sensitive threshold for the adaptive PCA method is suggested for fault detection. The variance sensitive adaptive threshold is suggested to overcome false alarm which occurs in the transient states according to changing process condition (Alkaya & Eker, 2011). Meanwhile, incipient faults describe the wear and ageing of system components and thus relatively difficult to handle due to the slowly developing nature of these faults. Most of the machine learning methods, such as neuro-fuzzy, SVM and GMM methods have been applied for the cases of incipient faults, with the fault identification has been successfully achieved. In addition, the signal processing approaches such as wavelet transform method is also been suggested to improve the FDI system, with the detection of incipient faults in chemical process systems.

Another major problem in chemical processes is the high level of complexity. The complexity of a plant could be understood as the impossibility to model its global emergent behaviour using single modelling techniques. A complex system can be represented by a system whose global behaviour emerges from the interactions between its large numbers of basic components and is difficult to represent analytically. Therefore, a distributed fault identification methodologies have been suggested, with the main idea is to partition the monitored system into subsystems, having a reasonable complexity level, and apply state-of-the-art FDI methods, on each one of the subsystems based on their problems. By implementing this strategy, the FDI method retains their ability to treat the local nonlinearities, noise, and uncertainty. Thus, the main task of this complex problem is to select, for each partition region, the available FDI methods that provide best performances.

5.3 Data-driven based hybrid FDI framework

Various decision combination methods, including averaging, majority voting, modified majority voting such as weighted majority voting, Dempster-Shafer, and Bayesian-based fusion have been applied to the fault detection and identification system to provide a correct final decision based on the collection of individual classifier decisions (Ghosh, Ng, & Srinivasan, 2011; Zhang, 2006).

In the averaging method, a simple decision fusion algorithm is applied when the largest n decisions and smallest n decisions are dropped and the average of remaining local decision is compared with a threshold for final decision. For instance, a multi-attribute data is used into aggregated values of a single attribute by OWA (ordered weighted averaging) operators (Salahshoor et al., 2010). The OWA operator provides a general class of parameterized aggregation operators including the min, max, and average. The major thrust behind selecting the OWA operator to aggregate multi-criteria relates to its inherent capability to encompass a range of operators bounded between minimum and maximum.

In a voting-based decision method, an agreement on the type of fault by the majority of the classifiers is required. If none of them agree on the type of faults, the final decision cannot be made. In other words, the winner is the class with the sufficient and highest number of expert votes. For instance, multiple neural networks are developed and their identification are combined through this modified majority voting (Zhang, 2006). This modified majority voting combination scheme intends to provide earlier fault identification. The threshold value of 0.6 is set based on heuristics and it is possible to fine tune this parameter based on a set of training and testing data. The maximum output when the majority of the individual networks give outputs much larger than 0 (which is >0.6), is taken as the final identification result in order to achieve earlier identification. On the other hand, when the majority of the individual networks give outputs which are not much larger than 0, then the median of the outputs is taken as the final identification output in order to provide reliable advance warning.

In weighted voting-based fusion, each expert receives a class-specific weight proportional to its classification accuracy on the training set or a separate validation set and the class with the maximum total weight is considered the winner. The weighted averaging technique is the simplest and most widely used technique combining multiple classifiers, which assigns a nonnegative weight to each individual classifier. By optimizing an objective function, the classifier weights can be estimated using various techniques (Lei et al., 2010). Genetic algorithms (GAs) can be used to search the optimal weights of multiple classifiers. For example, the weighted averaging technique with GAs is employed to combine the outputs of the six ANFISs and come out with the final identification results (Lei, He, Zi, & Hu, 2007).

In a performance-based decision method, more reliable classifiers are prioritized. Each fault discriminant method can adapt using the new faulty data once a identification decision is confirmed. For instance, Bayesian-based fusion uses the Bayes rule to calculate the posterior probability for each class from the confusion matrix and declares the class with maximum posterior probability as the winner (Seng Ng et al., 2010).

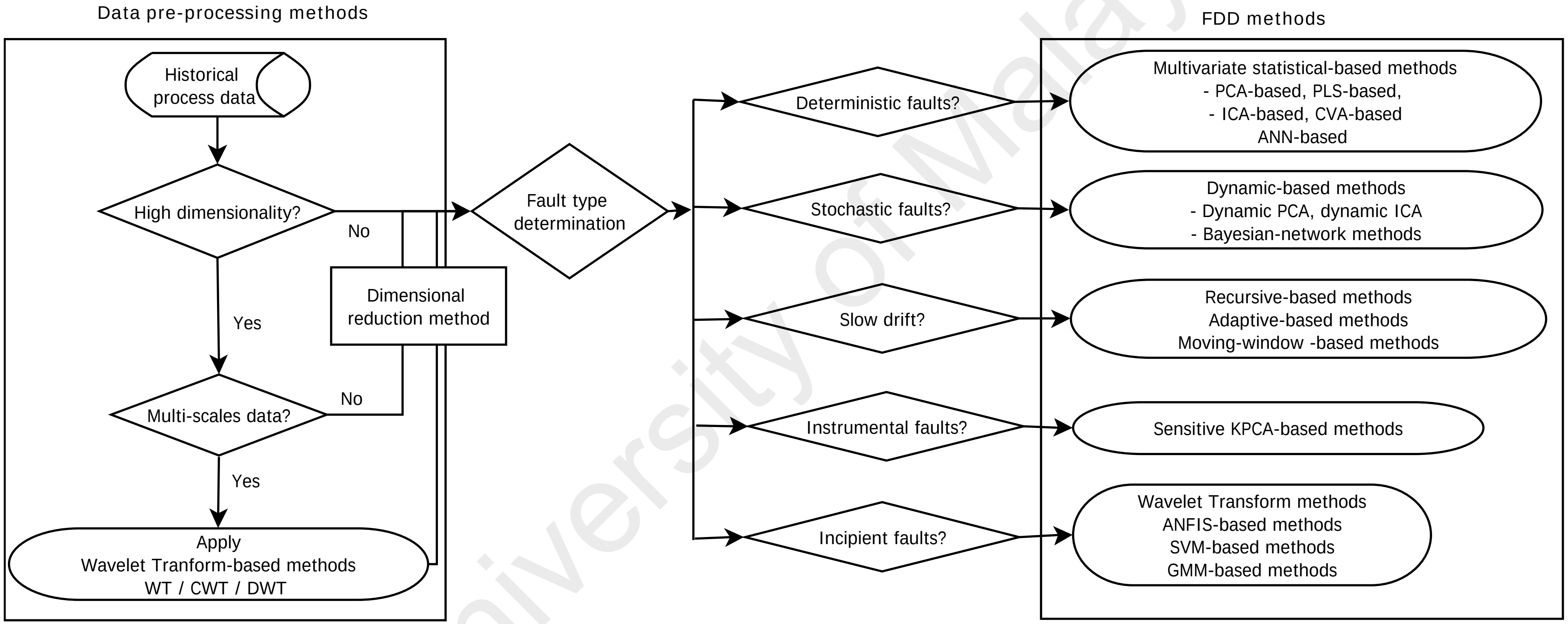

Figure 5.1: General guideline for selecting data-driven FDI methods

5.4 Classifiers hybridization methodology

This section introduces the development of the proposed hybrid classification technique for multi-scale KFDA-data-driven FDI framework. The basic idea is to make use of complementary classification of each individual multi-scale KFDA-data-driven based FDI classifiers. In this framework, each type of classifier is trained over the selected input space from multi-scale KFDA feature extraction and the final decision is then derived through the decision combination methods.

The classifiers are responsible to classify a sample x_t into one of the known fault classes under the presence of fault. The prediction from each classifier is in the measurement form with a degree of similarity, S_j^r, between x_t and fault j.

Suppose there are R diagnostic classifiers $A_r^d, r = [1, R]$. Let the fault database consist of J fault classes $DB = \{C_1, \ldots, C_J\}$. When the process is abnormal, each A_r^d will locate a set of candidates C_j, $j \epsilon [1, J + 1]$, where, class C_{J+1} corresponds to the novel fault.

Decision combination methods (using either majority-voting or Bayesian-based) seeks to identify the most probable fault class, C^{opt} by combining the predictions from all A_r^d, i.e., $\epsilon_1\ x \square C_j\ \square \ldots \square\ \epsilon_R\ x \square C_j\ , j \square\ 1, J + 1$. Similar to the combination of monitoring results, identification results can be combined based on plurality voting. Let T_j^r be the number of votes cast by Agent A_r^d for fault j.

$$\sum_{j=1}^{J+1} T_j^r = 1. \tag{5.1}$$

The combined prediction (fault candidates), C_V^{opt}, is determined based on the votes of all fault classifiers A_r^d as follows:

$$C_V^{opt} = \arg\max_{j \in [1,J+1]} \sum_{r=1}^{R} T_j^r \tag{5.2}$$

In the Bayesian-based decision combination method, we use the classification results produced from the classifiers, A_r^d. Upon detection of an abnormality, i.e., when $G(t) = 1$, the results produced by all A_r^d are collected to form a fault candidate pool.

The posteriori probability of fault j based on the evidence of classifier A_r^d is

$$P^r(x_t \in C_j \mid e_r(x_t) = j) = \frac{n_{ij}^r}{\sum_{i=1}^{M} n_{ij}^r}, \quad i,j \in [1,J] \tag{5.3}$$

A classifier may not always return a single fault candidate. In these cases, following Eq.(5.1), each A_r^d is assigned unit credit, which is distributed equally among all fault candidates FC_r proposed by A_r^d. The highest combination derived from the conditional probabilities of the candidates in candidate pool forms the basis for identifying the optimal candidate C_B^{opt}

$$C_B^{opt} = \arg\max_{j \in CP} \sum_{r=1}^{R} P^r(x_t \in C_j \mid e_r(x_t) = j) = \arg\max_{j \in CP} \sum_{r=1}^{R} \frac{n_{ij}^r / FC^r}{\sum_{i=1}^{M} n_{ij}^r} \tag{5.4}$$

A confirmed identification decision is formed and the performances of the methods are updated based on their classification results. The classifiers that have shown poor performance in identification may adapt by adding the new fault data to their historical datasets and rebuilding the model. All performance estimates and all classification decisions are listed for each classifier for the current fault observation. The classifier with the highest performance estimate is selected and its classification decision is the identification decision. If multiple classifier share the same ranking with the same estimated performance values, depending on the individual decisions either both or all of the classifiers' decisions are considered.

The architecture and algorithms described above have been implemented in Matlab. The classification algorithms by the data-driven methods can be computationally heavy, therefore a parallel implementation of the framework is needed to improve the performance. Four data-driven classifiers are used in this case study for fault classification – ANFIS, SVM, GMM and *k*NN. The predictions from all four classifiers used are in measurement forms and various decision fusion schemes are used to develop the hybrid classification output. The proposed method is tested on three different case studies; the Tennessee Eastman process, the Penicillin fermentation process, and the semiconductor etch process.

5.4.1 Majority voting-based method

In binary classification, majority voting is implemented in the natural way, such as if there are more individual classifiers giving output class 1 instead of class 0, then the aggregated classifier takes the output 1. When a test pattern is presented to the ensemble, final class is decided using the majority voting rule. For example, the class that is predicted the most by the trained classifiers is chosen as the class label for the test pattern.

There are two different voting schemes to determine its label, namely majority voting and weighted-sum voting. Simple majority voting is a decision rule that selects one of many alternatives, based on the predicted classes with the most votes. It is the decision rule used most often in ensemble methods. Weighted majority voting can be made if the decision of each classifier is multiplied by a weight to reflect the individual confidence of these decisions. Simple majority voting is a special case of weighted majority voting, assigning an equal weight of $1/k$ to each classifier where k is the number of classifiers in an ensemble (Arruda et al., 2014).

In the identification scheme considered here, since the classifier outputs take continuous values between 0 and 1, majority voting takes the following form:

$$g = \mathrm{median}\left(f_i\left(X(t)\right)\right), i = 1, 2, \ldots, N. \tag{5.5}$$

The median of the individual network outputs is taken as the aggregated neural network output. If the majority of the individual networks have outputs corresponding to a fault close to 1, then the median of the corresponding individual network outputs will also be close to 1.

5.4.2 Class-specific Bayesian based method

The Bayesian fusion approach stores the historical priori class-specific performance of each classifier κ_r, $r = [1, R]$ using conditional probability. The final predictions are then estimated through the Bayes rule of calculating posteriori probability. For a classification problem of J classes on measurement x, if all classes are mutually exclusive (two classes cannot occur concurrently), the Bayesian inference process for evaluating the conditional probability of a class j from κ_r becomes

$$P\left(x \in C_j \mid \kappa_r(x) = j\right) = \frac{P(\kappa_r(x) = j \mid x \in C_j)\, P(x \in C_j)}{\sum_{i=1}^{J} P(\kappa_r(x) = i \mid x \in C_i)\, P(x \in C_i)} \tag{5.6}$$

The conditional probability that implies $x \in C_j$, $j \in [1, J]$, is often estimated from previous performance of κ_r using a confusion matrix (CM). A CM stores class-specific performance of κ_T, and is normally constructed by testing κ_T with some training dataset. The class C_j with the highest P_{E_j} can be selected as the optimal combined prediction

$$E\ x = j, \qquad \text{if } C_j = \underset{j \in \Lambda}{\operatorname{argmax}}(P_{E_j}) \tag{5.7}$$

5.5 Case study and classification performance

The implementation of hybrid multi-scale KFDA-data-driven framework for fault detection and identification has been done on three different case studies; the Tennessee Eastman process, the Penicillin fermentation process, and the semiconductor etch process. The results and its discussion have been summarized in the following sections.

5.5.1 Fed-batch penicillin fermentation process

Penicillin is a group of antibiotics obtained from penicillin fungi discovered by Alexander Flemming in 1928. The process data for fed-batch penicillin fermentation is generated using a mathematical model and PenSim simulator. The process to produce penicillin has nonlinear dynamics, time-varying, and multiphase characteristics. Throughout the entire fermentation process, many factors affect the effectiveness of penicillin fermentation, such as temperature, pH, substrate concentration, dissolved oxygen concentration and so on. Therefore it is imperative to conduct effective process monitoring. PenSim simulator provides a benchmark simulation platform for penicillin fermentation modelling, optimal control and process monitoring. A descriptive flowchart of the process is depicted in Figure 5.2.

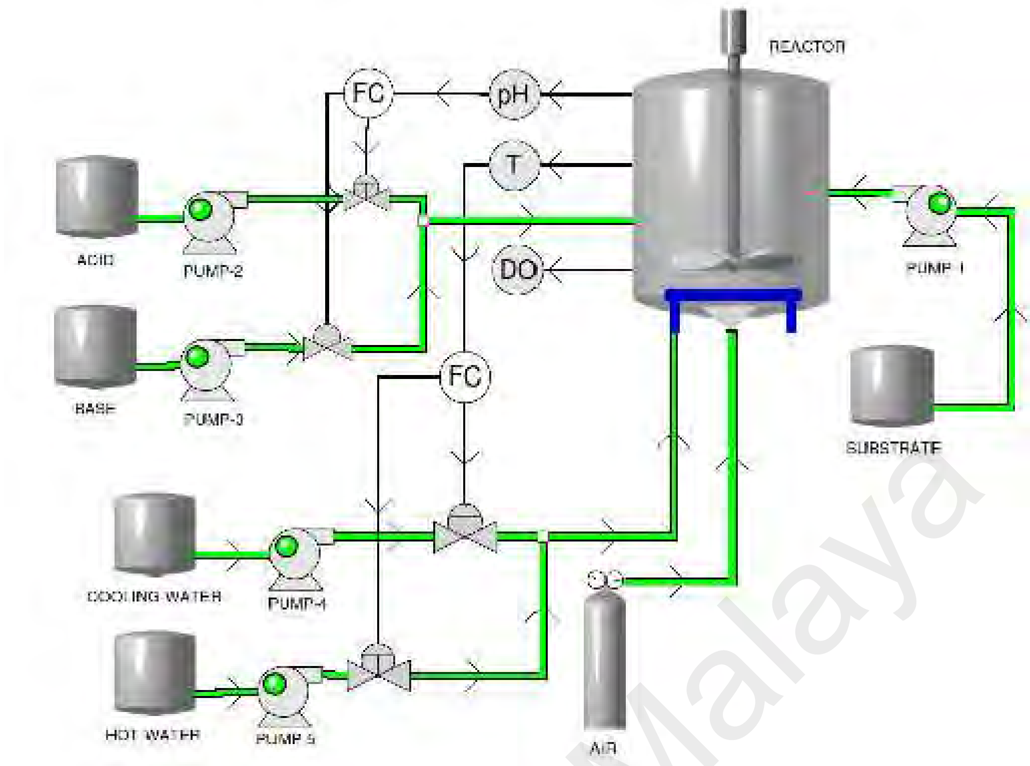

Figure 5.2: Fed-batch penicillin fermentation process

The model has five input variables (1-4 and 14), nine process variables (5-13), and five quality variables, as shown in Table 5.1. Feedback controllers keep pH and temperature near their desired values. These simulations are run under closed-loop control of pH and temperature, while glucose addition is performed open-loop.

Penicillin cultivation process has two operational phases. It has actually four physiological phases: lag, exponential cell growth, stationary, and cell death. The first two phases are conducted as batch operation, while the following two phases are conducted as fed-batch operation. In the first operational phase, fermentation is carried out in batch mode to promote biomass growth resulting in high cell densities. The second phase is a fed-batch operation. When the initial amount of glucose is consumed by the growing cells, additional glucose is fed during the fed-batch operation until the end of the run. In a batch fermentation process that lasts several days, some microorganisms may have different generation times. Slight changes in operating conditions during critical periods may have a significant influence on growth and differentiation of microorganisms, and impact final product quality and yield.

To simulate the physical uncertainty present in each batch due to variable metabolic responses, very small perturbations are introduced into the manipulated variables while generating the batch data set. A reference data set of 60 batches is simulated under nominal conditions with small perturbations using PenSim, resulting in unequal batch lengths. The duration of each batch is 400h, consisting of a pre-culture phase of about 45h and a fed-batch phase of about 355h. The simulation conditions of normal and fault operations are listed in Table 5.1, and the sample number of each data is 400 since the sampling interval is chose as 1h. The data sets consisting of 12 process variables are listed in Table 5.2.

Data are divided into two parts corresponding to batch (phase 1) and fed-batch (phase 2), respectively. Separate indicator variables are selected for each phase. The culture volume (variable 9) decrease is found as a good candidate in phase 1. A derived variables called ‘percent substrate fed’ is calculated from substrate feed rate (variable 3) and used as an indicator variable in phase 2. It is assumed that phase 2 is completed when 25l of substrate is added to the bioreactor.

Table 5.1: Simulations conditions of normal and fault operations

Simulation conditions	**Normal**	**Fault 1**	**Fault 2**	**Fault 3**	**Unit**
Initial conditions					
Substrate concentration	15	15	15	15	g/L
Dissolved oxygen concentration	1.16	1.16	1.16	1.16	g/L
Biomass concentration	0.1	0.1	0.1	0.1	g/L
Penicillin concentration	0	0	0	0	g/L
Culture volume	100	100	100	100	g/L
Carbon dioxide concentration	0.5	0.5	0.5	0.5	g/L
pH	5	5	5	5	-
Fermenter temperature	298	298	298	298	K
Generated heat	0	0	0	0	Kcal
Set points					
Aeration rate	8.6	8.6	8.6	**6**	L/h
Agitator power	29.9	**21**	29.9	29.9	W
Substrate feed flowrate	0.0426	0.0426	**0.027**	0.0426	L/h
Substrate feed temperature	296	296	296	296	K
Temperature set point	298	298	298	298	K
pH set point	5	5	5	5	-
Sampling interval	1	1	1	1	H
Simulation time	400	400	400	400	H

Table 5.2: Variable of FBFP process

Number	**Process variable**	**Unit**
1	Aeration rate	L/h
2	Agitator power	W
3	Substrate feed temperature	K
4	Substrate concentration	g/L
5	Dissolved oxygen concentration	g/L
6	Biomass concentration	g/L
7	Penicillin concentration	g/L
8	Culture volume	L
9	Carbon dioxide concentration	g/L
10	pH	-
11	Fermenter temperature	K
12	Cooling water flow rate	L/h

A fault was imposed to the batch for the second batch. The substrate feed rate was linearly decreased from 0.04 to 0.03l/h due to the fault of a feeding pump until the end of batch operation. The initial time of the fault was 60h. A decrease in its feed resulted in a reduction in penicillin production since glucose is the main carbon source to be fed during the fed-batch fermentation. For the third batch, a 15% step-decrease in the substrate feed rate was introduced at 55h and retained until the end of fermentation. For the fourth batch, a fault was imposed from time 40h to the end of batch. During that time, agitation power was linearly decreased from 30 to 26W. Agitation has a direct influence on the overall oxygen mass transfer coefficient, K_{la}. A decrease in the K_{la} value resulted in a decrease in the dissolved oxygen level in the culture medium, consequently lowering the biomass growth and penicillin concentration (J.-M. Lee, Yoo, & Lee, 2004). Due to the small magnitude of the drift, the effects on the process are not immediately occurring (Ündey, Tatara, & Cįnar, 2003). The monitoring results of hybrid multi-scale KFDA-data-driven method are shown in Table 5.3.

Table 5.3: Fault detection for fed-batch penicillin fermentation process

Fault	Hybrid Multi-scale KFDA-data-driven with Bayesian decision method	Hybrid Multi-scale KFDA-data-driven with majority voting method	Multi-scale KFDA-GMM	Multi-scale KFDA-SVM	Multi-scale KFDA-ANFIS
Fault 1	100	100	77.65	97.75	84.00
Fault 2	100	99.21	88.47	98.20	95.75
Fault 3	100	100	85.12	93.33	95.75
Average	100	99.74	83.75	96.43	91.83

From Table 5.3, both of the proposed hybrid multi-scale KFDA-data-driven methods produced comparable performances. From three types of faults generated during the simulation, the hybrid framework with Bayesian decision method had produced 100% accuracy of fault classification, followed by the classification of 99.74% by the majority voting method. With nearly similar results produced during the testing, the decision method has been a factor that can decide the classification performance.

For instance, for Fault 2 classification, the majority voting methods had been seen made a misclassification of 0.79%. This misclassification has affected its overall performance, even though it average performance can be considered as high as the previous Bayesian decision method. Both Fault 1 and 3 have been totally classified by both of the decision methods, which are also been contributed with the effective individual classification performances. With various individual classifiers have been performed well during the classification, the decision method could have less challenging decision to come out with. Therefore, the improvement achieved through this case study is marginal since the individual classifiers performs reasonably well.

5.5.2 Batch semiconductor etch process

The manufacture of semiconductors is introduced as an example of the online monitoring of batch processes. This study focuses specifically on an Aluminium-stack etch process performed on the commercially available Lam 9600 plasma etch tool (Metal Etcher) (Wise, Gallagher, Butler, White, & Barna, 1999), at the Texas Instruments Inc. the TiN/Al-0.5% Cu/TiN/oxide stack is etched with an inductively coupled BCl_3/Cl_2 plasma. The key parameters of interest are the line width of the etched

Al line, uniformity across the wafer, and the oxide loss. Process conditions must be kept constant in order assure consistent results. This single chemistry etch process mainly consists of six steps, where the first two are for gas flow and pressure stabilization, while step 3 is a brief plasma ignition step. Step 4 is the main etch of the Al layer terminating at the Al endpoint, whereas step 5 acting as the over-etch for the underlying TiN and oxide layers and step 6 vents the chamber. The process chemistry is identical during step 3 through step 5.

In the etch process, it would be ideal to have sensors which directly reflected the state of the wafers in the process. However, with a few exceptions, wafer state sensors are typically unavailable in the original equipment manufacturer (OEM) processing tools. This, the alternative is to select more commonly available process state sensors, with the understanding that wafer state information will have to be inferred. Engineering judgement was also used to select variables that should impact product quality. These variables are listed in Table 6.3.

The machine state sensors, built into the processing tool, collect the available machine data during wafer processing. The machine data consists of 40 process setpoints and measured and controlled variables sampled at 1 second intervals during the etch. A sampling interval of 1s was used in the analysis. A series of three experiments, resulting in three different data groups, were performed where faults were purposely induced by changing specific manipulated variables (TCP power, RF power, pressure, plasma flow rate, and Helium pressure). There are 108 normal operating batches and 20 fault batches. Twenty batches were randomly selected from the normal batches to investigate the effect of false alarms.

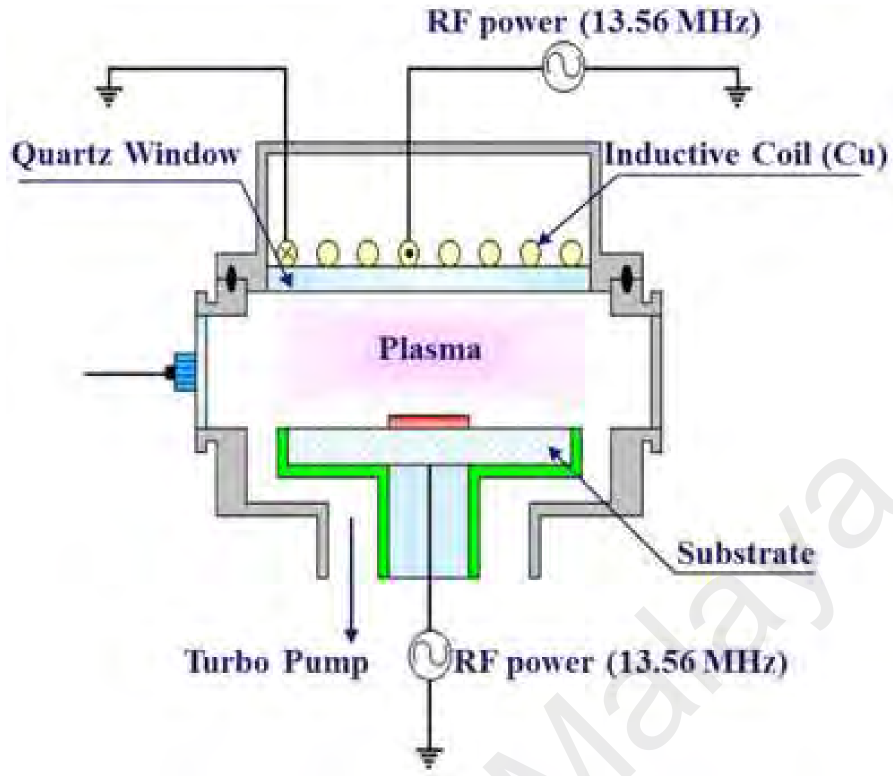

Figure 5.3: Schematic of plasma etching system

Faults were induced in the data by changing the setpoints for controlled variables, such as chamber pressure and plasma power, from the normal recipe. The mean values of the controlled variables were then set back to those of the normal recipe. The effect is a data record where it appears as if a bias had developed in the sensor for the controlled variable.

Three experiments, numbered 29, 31, and 33, were performed to produce the data use here. The experiments were run several weeks apart. Forty-three wafers were processed in each experiment. The three experiments together generated data for 108 normal wafers and 21 wafers with induced faults, though there are several instances where the data records for the RFM and OES are not complete.

Table 5.4: Machine state variables used for process monitoring and FDI

1	BCl_3 flow	11	RF power
2	Cl_2 flow	12	RF impedance
3	RF bottom power	13	TCP tuner
4	RF reflected power	14	TCP phase error
5	Endpoint A detector	15	TCP impedance
6	Helium pressure	16	TCP top power
7	Chamber pressure	17	TCP reflected power
8	RF tuner	18	TCP load
9	RF load	19	Vat valve
10	Phase error		

In this work, only the machine state variables that are collected at a 1s sample interval are considered. This is because, from previous study, it was observed that the machine state variables give the maximum information about the process and are most sensitive to faults in the system. The variables use for comparing the hybrid multi-scale KFDA-data-driven FDI method with others has been reported in Table 5.4. Monitoring on practical industrial data has more difficulties, including serial correlations and large scalability.

Ideally, under normal conditions, a process would be stationary, i.e., retain the same mean and covariance structure over time. Unfortunately, measurements from the etch process are clearly non-stationary. The process set point variables have been removed for the analysis because throughout the process they are maintained constant and do not provide any valuable information. The time and step number variables have also been removed as they do not provide any crucial information about the process.

Changes in the data are primarily due to three sources, aging of the etcher over a clean cycle as residue accumulates on the inside of the chamber, differences in the incoming materials due to changes in the upstream processes, and drift in the process monitoring sensors themselves. In addition, process maintenance can result in sudden shifts in the mean.

The process monitoring and FDI frameworks was tested using etch data from three experiments; experiment number 29, 31 and 33, which is also includes induced faults. A series of specific faults were intentionally induced by changing the TCP power, RF power, pressure, Cl2 or BCl3 flow rate, and Helium pressure, as shown in Table 5.5. These three experiments consists of 128 wafers with 21 faults. In total, there are 107 normal wafers and 21 faulty wafers. The normal samples are split randomly in a 2:1 ratio to form the training and the validation set. This was done randomly so that the validation set has representations from samples belonging to the three different experiments, as shown in Table 5.5.

From Table 5.3, both of the proposed hybrid multi-scale KFDA-data-driven methods produced comparable performances. From 19 types of faults generated during the simulation, the hybrid framework with Bayesian decision method had produced 93.71% accuracy of fault classification, followed by the classification of 87.92% by the majority voting method.

For instance, for Fault 5 classification, the Bayesian methods had been seen made a misclassification of 6.38%, while majority voting-based method had 13.5% of misclassification. This misclassification has affected both of their overall performance, even though it average performance can be considered as high as nearly 88%, whereas the previous Bayesian decision method had produced 93.71% of average accuracy in fault classification. A summarize from 19 fault classes in this case study, various individual classifiers have been performed well during the fault classification, even though the decision method could have some challenging decision to come out with. For instance, the decision in Fault 1, Fault 5, Fault 9, Fault 11, Fault 14 and Fault 19, where there is a large margin between the performances of both decision methods. Therefore, the improvement achieved through this case study is been decided by the right application of hybrid decision methods, as well as the efficient individual classification methods.

5.5.3 Tennessee Eastman process

The details and description of the Tennessee Eastman process was discussed earlier in Section 3.6. The proposed method in this chapter has been applied with the same database, to get a clearer projection of the performance's improvement.

From Table 5.6, both of the proposed hybrid multi-scale KFDA-data-driven methods produced comparable performances. From three types of faults generated during the simulation, the hybrid framework with Bayesian decision method had produced 93.40% accuracy of fault classification, followed by the classification of 90.57% by the majority voting method. With quite similar results produced during the testing, the decision method could be a factor that can decide the classification performance.

For instance, for Fault 5 classification, the majority voting methods had been seen made a misclassification of 0.40%. Fault 5 corresponds to a step change in the condenser cooling water inlet temperature. The fault was introduced at t=460 min as a step change to variable XMV(11) – the condenser cooling water flow. When the fault occurs, the liquid flow rate of the outlet stream from the condenser to the vapor/liquid separator increases. This causes the pressure of the separator and stripper to decrease. For this fault, the control loops are able to compensate the change and all measured variables are returned to set-point. The time it takes to reach steady-state is approximately 10 h.

The same misclassification value has been achieved by the Bayesian decision method. However, the difference in other faults classification, such as in Fault 16 and Fault 19 have been the decider. This misclassification has affected the majority-voting based overall performance, even though it average performance can be considered as high as the previous Bayesian decision method. However, Fault 1, 4, and 7 have been totally 100% classified by both of the decision methods, which are also been contributed with the effective individual classification performances. With various individual classifiers have been performed well during the classification, the decision method could have less challenging decision to come out with. Therefore, the improvement achieved through this case study is marginal since the individual classifiers perform reasonably well.

The summary classification results based on Bayesian decision and majority-voting based strategies are shown in Table 5.6, where all faults are identified successfully by proposed hybrid multi-scale KFDA-data-driven framework. The Bayesian-based decision combination is able to diagnose all of the faults by effectively combining the strength from each method of different classifiers.

Table 5.5: Fault detection for semiconductor etch data

Exp. Run	Fault	Induced fault	Hybrid Multi-scale KFDA-data-driven with Bayesian decision method	Hybrid Multi-scale KFDA-data-driven with majority voting method	MSKFDA-SVM	MSKFDA-ANFIS
Exp. 29	Fault 1	TCP +50	94.12	87.80	79.00	71.88
Exp. 29	Fault 2	RF +10	83.87	79.00	75.75	78.12
Exp. 29	Fault 3	Pr +3	89.97	85.75	74.37	73.87
Exp. 29	Fault 4	TCP +10	98.75	92.37	72.81	85.87
Exp. 29	Fault 5	BCl_3 +5	93.62	86.50	73.62	77.80
Exp. 29	Fault 6	Pr -2	88.37	83.62	81.88	79.00
Exp. 29	Fault 7	Cl_2 -5	89.04	83.62	78.12	79.48
Exp. 29	Fault 8	He Chuck	92.87	91.88	83.87	87.62
Exp. 31	Fault 9	TCP +30	94.12	78.12	70.00	69.00
Exp. 31	Fault 10	Cl_2 +5	83.87	73.87	72.75	71.75
Exp. 31	Fault 11	BCl_3 -3	97.88	85.87	83.62	84.75
Exp. 31	Fault 12	Pr +2	98.21	89.48	88.37	84.12
Exp. 31	Fault 13	TCP -20	98.46	97.62	89.04	83.87
Exp. 33	Fault 14	TCP -15	95.38	89.00	82.87	89.97
Exp. 33	Fault 15	CL_2 -10	95.63	95.15	84.74	84.37
Exp. 33	Fault 16	RF -12	95.24	94.37	93.62	92.81
Exp. 33	Fault 17	BCl_3 +10	98.50	92.81	92.37	92.62
Exp. 33	Fault 18	Pr +1	94.74	93.62	89.00	84.37
Exp. 33	Fault 19	TCP +20	97.80	90.11	82.87	82.81
	Average		93.71	87.92	81.51	81.79

Table 5.6: Fault detection for the Tennessee Eastman process

Fault	Hybrid Multi-scale KFDA-data-driven with Bayesian decision method	Hybrid Multi-scale KFDA-data-driven with majority voting method	MSKFDA-GMM	MSKFDA-SVM	MSKFDA-ANFIS
1	100	100	98.00	99.80	94.00
2	100	99.21	98.37	99.87	95.37
3	75.21	70.26	15.75	43.33	35.75
4	100	100	89.48	95.89	92.75
5	99.60	99.60	97.62	99.50	97.80
6	98.85	100	89.00	96.02	99.00
7	100	100	95.75	99.44	95.75
8	98.19	94	54.37	95.81	92.37
9	82.81	78.56	72.81	51.56	36.50
10	97.60	92.75	23.62	93.99	83.62
11	97.88	96.50	91.88	95.38	89.75
12	98.21	97.0	78.12	95.63	94.12
13	98.46	97.0	73.87	95.24	83.87
14	100	97.0	85.87	98.50	89.97
15	88.85	75.54	68.75	64.74	58.75
16	98.81	76.15	93.62	97.80	93.62
17	99.60	98.81	98.37	97.32	88.37
18	98.19	93.63	89.00	95.15	89.04
19	100	70.11	62.87	96.87	92.87
20	98.96	88.37	62.50	93.63	82.50
21	90.23	77.52	15.12	87.64	88.12
Aver.	93.40	90.57	84.72	90.13	84.44

5.6 Summary

Although most of the studies in the literature focus on solving problems of data characteristics such as nonlinearity, non-Gaussian distribution, and data autocorrelation, it is still hard for any single FDI method to perform effectively for all possible faults a process could have. Therefore, it seems that hybridisation-based framework is required for developing a complete and robust fault detection and identification tool. However, with the difficulties to construct a unified nonlinear model for the process and the emergence of more combinations involving different non-Gaussian data modelling techniques with the support of advanced software tools, it can be expected that the process monitoring and FDI performance such as fault detection speed, fault classification rate, and fault identification accuracy could be further improved. Hence, the proper selection of FDI methods is needed to achieve maximum efficiency in fault detection and identification of the developed system. Thus, the guideline in choosing the suitable data-driven method is provided, aiming to help others in developing the data-driven based FDI in chemical process systems.

Furthermore, the classifiers hybridization methodology has also been discussed in this chapter, where two different hybrid decision methods are used, namely majority voting-based method and class-specific Bayesian based method. To validate the developed hybrid framework for fault classification, three different case studies are implmented; the fed-batch penicillin fermentation process, batch semiconductor etch process, and the Tennessee Eastman process. In summary, the application of Bayesian-based method and majority-voting based decision method are successful. Both of these methods gave marginal performance in overall, for all of case studies in this work. Furthermore, these successful hybrid frameworks have also been contributed by the highly effective individual classification performances. With various individual classifiers have been performed well during the classification, the decision method would have less conflict during the decision determination. Therefore, the improvement achieved through this case study is marginal since the individual classifiers perform reasonably well.

CHAPTER 6: CONCLUSION & RECOMMENDATIONS

6.1 Introduction

This chapter summarises the thesis, discusses its findings and contributions, and outlines directions for future research. This chapter is divided into four sections, starting with Section 6.2 that presents the summary of the thesis. Section 6.3 discusses the contributions of the current work, while Section 6.4 lists the recommendations for future studies. Finally, Section 6.5 provides a conclusion of the entire thesis.

6.2 Summary of the Thesis

This thesis introduced the data-driven FDI methods with their background study. It focused on data-driven techniques in order to explore the opportunity to implement and improve their performance. The main objective of this work is to develop a complete FDI framework based on multi-scale KFDA and hybrid fault classification. This framework can be evaluated on different systems and the proficiency of the proposed method can be compared with other standard approaches.

Chapter 2 reviewed the data-driven FDI methods in chemical process systems comprehensively. The review started with an introduction of the general characteristics of chemical processes, followed by problems that commonly occur during the detection and identification of these chemical process systems. In addition, the application of feature extraction method, mainly based on multivariate statistical analysis and multi-scale analysis, was also included in this chapter. Moreover, the selected data-driven FDI methods, namely ANFIS, SVM, GMM, and *k*NN, were reviewed with discussion on their applications and current issues in chemical process systems. Finally, this chapter provided a summary by reviewing the hybrid data-driven methods in FDI systems and a conclusion of the chapter.

Meanwhile, Chapter 3 introduced the multi-scale KFDA feature extraction method that consists of three different steps, namely the data acquisition and normalisation, DWT, and multi-scale KFDA discriminant vector. The wavelet transformation step had another three steps known as DWT decomposition, threshold determination, and inverse DWT reconstruction. Moreover, this chapter also proposed a multi-scale feature extraction method based on the combination of DWT with Parseval's theorem and KFDA to improve the feature extraction method. Subsequently, the proposed method was tested using the Tennessee Eastman process database, in which the performance was evaluated based on its fault classification accuracy.

In Chapter 4, four types of data-driven classification methods, ANFIS, SVM, GMM, and *k*NN, were proposed to be implemented individually with the multi-scale KFDA method to form multi-scale KFDA-ANFIS, multi-scale KFDA-SVM, multi-scale KFDA-GMM, and multi-scale KFDA-*k*NN, respectively, for fault classification scheme. In the fault classification step, the data-driven methods were applied to the framework to find the patterns in the extracted MSKFDA subspaces. The classifier will diagnose the faults by assigning the features to the corresponding fault classes detected. Each of the proposed MSKFDA-data-driven method was applied with designated faults of the Tennessee Eastman process, from Fault 1 until Fault 21. The proposed methods were developed to detect and classify the faults. The training data were used to develop the classification models, whereas the testing data were utilised to evaluate the classification performance.

Furthermore, Chapter 5 presented the hybridisation of classifiers. In particular, several classifiers were integrated so that their individual decisions were combined in specific way to classify the testing samples. The hybrid FDI framework was developed based on a strategy that amplified the correct combined outputs of the single classifiers, while cancelling out the incorrect ones. This section introduced the development of the proposed hybrid classification technique for multi-scale KFDA-data-driven FDI framework. Basically, the idea was to make use of the complementary classification of each individual multi-scale KFDA-data-driven-based classifier.

In this framework, each type of classifier was trained over the selected input space from multi-scale KFDA feature extraction and the final decision was then derived through the decision combination methods. Two different decision methods were applied in this work, namely majority voting-based method and class-specific Bayesian-based method. The proposed framework was tested on three different case studies: the fed-batch penicillin fermentation process, the semi-conductor etch process, and the Tennessee Eastman process. Finally, this chapter also included a guideline for data-driven method selection for FDI in chemical process systems. The guideline was constructed based on the problems which generally occur in process systems.

6.3 Contributions of the Thesis

In summary, the major contributions of this thesis to the data-driven FDI aspects are as follows:

- A multi-scale dimensional reduction method was developed based on the combination of KFDA and DWT. The proposed method, called multi-scale KFDA was developed to improve the performance of feature extraction task. Multiple scale information, compared to the single scale, was considered for classification of the features in both time and frequency domains. The proposed feature extraction method enhanced the separability of patterns in the database, improved the accuracy of classification, and reduced the computational complexities of the FDI system.

- Data-driven FDI schemes were developed based on the combination of multi-scale KFDA feature extraction method with various types of data-driven classifiers. As most processes are multi-scale in nature, the solution suggested was the multiple scaling of dimensionality reduction of the data. Particularly, it was decomposed into various time scales to extract the appropriate information from the faulty process data. Therefore, the data-driven methods of pattern classification based on multi-scale feature extraction were developed to enhance and improve the classification performance through the modification of the multi-scale approaches.
- A guideline of method selection for data-driven FDI framework was developed to improve the applications of data-driven methods in chemical process systems. The development of the guideline was problem-oriented. Almost all of the actual chemical processes have various types of characteristics such as non-Gaussian distribution, time-varying, and multi-mode behaviours. Therefore, this guideline is useful, especially for systems with multiple process characteristics, as the method that can be employed in the FDI framework could be complicated with the known process characteristics.
- A hybrid data-driven FDI framework was proposed by combining multiple data-driven methods with multi-scale KFDA method for chemical process systems. It is widely known that most of the FDI methods perform very well for one fault, but very poorly for another. Plus, all available fault identification methods have their drawbacks, even though a number of methods may perform consistently well for nearly all types of faults. Thus, there is a need to develop a systematic

hybrid framework that is capable of selecting the optimal combination of the detection and identification methods. Moreover, various case studies were examined in this work to prove the validity of the proposed data-driven agent network.

6.4 Recommendations for Future Work

Based on the present work, many areas of improvement could be investigated in the future. The recommendations are as follows:

- The ultimate goal in FDI development is the successful performance on practical applications. However, several issues have been identified in practical applications, especially from the perspective of the chemical process industries. First is the high requirement of expert knowledge and data samples for training purpose, as the small number of history fault samples is not enough for model training of most data-driven FDI methods. Furthermore, in practical scenarios, critical faults, especially the same kind of faults, do not occur frequently.
- Second is the issue of bad performance in the adaptation and self-learning of the model developed. Real chemical processes can be affected by external environments, different operators, product quality grades, or new kinds of faults during the online operations. Hence, the online FDI methods developed should be able to adapt and have good self-learning abilities to deal with these immeasurable changes.
- This study also suggests several directions for future works related to data-driven-based FDI systems. First, a lot of industrial processes could have multiple faults occurring within the same time windows, thus, the detection and

identification of the multiple faults are considered challenging. Moreover, one important area of active research is applying hybrid methods of computational intelligence techniques to solve multiple problems. They are often used to solve complex real-world problems; one technique is typically used to fix the weaknesses of the other. It may be more effective to apply the hybrid methods of other artificial intelligence algorithms. The effectiveness of the developed hybrid FDI systems depends on the nature of the faults' combination. Therefore, the important aspect in future FDI systems is to be able to make an estimation of faults with incomplete or new abnormal observation during process monitoring.

- Further consideration can be made for the validity of the framework to be implemented in the practical application of the process industry with the incorporation of several external factors. The amount of data that can be generated, collected, and processed for industrial processes has increased significantly and has grown almost exponentially with the development of sensors and computing techniques. However, these large archived data sets which are normally referred to as Big Data, are underutilised to exploit their information for advanced knowledge mining or smart decision-making. Big Data is characterised by four characteristics known as the 4 V's, namely velocity, volume, variety, and veracity, because of their large volume of data, quick rate of collected data (velocity), new features of data (variety), and data veracity. The Big Data also features heterogeneity as new data are increasingly heterogenous in terms of types and time scale. Moreover, the Big Data techniques also highlight efficient data mining methods like deep learning, thus, they expand the usage of data-driven methods. Therefore, future work on the development of Big Data method for FDI system requires an optimal method that can handle rare-event data and data cleaning.

6.5 Conclusion

In this work, multi-scale KFDA-based feature extraction for chemical process systems was presented. Data discrimination based on the proposed multi-scale KFDA methodology enhanced the extraction by taking into consideration the multi-scale information compared to other methods which only considered the single scale nature. It can also provide a better separation of the features and improve the extraction of features that are relevant to a faulty situation from both time and frequency domains.

This work was developed further to evaluate the integration of multi-scale KFDA method for feature extraction with data-driven classifiers such as ANFIS, SVM, GMM, and *k*NN. By comparing the performance and effectiveness of classification for these proposed frameworks when dealing with the Tennessee Eastman process database, the proposed MSKFDA-based data-driven frameworks can successfully detect nearly all types of faults, with a high degree of accuracy. The average accuracy for the performance of these classifications was greater than 90%.

In addition, this work developed a hybrid FDI framework implementing the multi-scale KFDA feature extraction method and multiple types of data-driven classifiers, namely ANFIS, SVM, GMM, and *k*NN methods. The main contributions of this work are that it provided a complete framework for FDI based on multi-scale discriminant analysis for dimensional reduction and feature extraction, and demonstrated the capability of hybrid fault classification methods in identifying faults in chemical processes. A Bayesian-based decision combination strategy was also investigated using three different chemical process systems, namely the benchmark Tennessee Eastman process, the fed-batch penicillin fermentation process, and the semiconductor etch process. The results showed that hybrid classifiers can increase the fault classification performance substantially.

REFERENCES

Agrawal, V., Panigrahi, B. K., & Subbarao, P. M. V. (2014). Review of control and fault diagnosis methods applied to coal mills. *Journal of Process Control*.

Alawi, A., & Julian Morris. (2007). Multiscale Fault Detection and Diagnosis in Fed-Batch Fermentation. In *10th International IFAC Symposium on Computer Applications in Biotechnology* (Vol. 1, pp. 33–38). Cancun, Mexico.

Alkaya, A., & Eker, İ. I. (2011). Variance sensitive adaptive threshold-based PCA method for fault detection with experimental application. *ISA Transactions*, *50*(2), 287–302.

Arruda, P., Bandeira, S., Soares, S., & Laureano, G. T. (2014). Fault Detection in Industrial Plant Using κ -Nearest Neighbors with Random Subspace Method. In *Proceedings on the International Conference on Artificial Intelligence (ICAI)* (pp. 1–6).

Askarian, M., Escudero, G., Graells, M., Zarghami, R., Jalali-Farahani, F., & Mostoufi, N. (2016). Fault diagnosis of chemical processes with incomplete observations: A comparative study. *Computers and Chemical Engineering*, *84*, 104–116.

Awadallah, M. A., & Morcos, M. M. (2006). Automatic diagnosis and location of open-switch fault in brushless dc motor drives using wavelets and neuro-fuzzy systems. *IEEE Transactions on Energy Conversion*, *21*(1), 104–111.

Ayoubi, M., & Isermann, R. (1997). Neuro-fuzzy systems for diagnosis. *Fuzzy Sets and Systems*, *89*(3), 289–307.

Bakshi, B. R. (1998). Multiscale PCA with application to multivariate statistical process monitoring. *AIChE Journal*, *44*(7), 1596–1610.

Blázquez, L. F., Aller, F., De Miguel, L. J., Perán, J. R., L. Felipe Blazquez, Aller, F., … Peran, J. R. (2006). Fault detection by neuro-fuzzy identification in a nonlinear system. *IFAC Proceedings Volumes (IFAC-PapersOnline)*, *6*(PART 1), 228–233.

Bo, C., Qiao, X., Zhang, G., Bai, Y., & Zhang, S. (2010). An integrated method of independent component analysis and support vector machines for industry distillation process monitoring. *Journal of Process Control*, *20*(10), 1133–1140.

Cai, L., Tian, X., & Zhang, H. (2015). Process Fault Detection Method Based on Time Structure Independent Component Analysis and One-Class Support Vector Machine. *IFAC-PapersOnLine*, *48*(21), 1198–1203.

Cai, L., Tian, X., & Chen, S. (2017). Monitoring Nonlinear and Non-Gaussian Processes Using Gaussian Mixture Model-Based Weighted Kernel Independent Component Analysis. *IEEE Transactions on Neural Networks and Learning Systems*, *28*(1), 122–135.

Chen, B. H., Wang, X. Z., & McGreavy, C. (1998). On-line operational support system for faults diagnosis in process plants. *Computers & Chemical Engineering*, *22*, S973–S976.

Chen, K.-Y., Chen, L.-S., Chen, M.-C., & Lee, C.-L. (2011). Using SVM based method for equipment fault detection in a thermal power plant. *Computers in Industry*, *62*(1), 42–50.

Chen, T., Morris, J., & Martin, E. (2006). Probability density estimation via infinite Gaussian mixture model: application to statistical process monitoring. *Journal of the Royal Statistical …*, 1–27.

Chen, T., & Zhang, J. (2010). On-line multivariate statistical monitoring of batch processes using Gaussian mixture model. *Computers and Chemical Engineering*, *34*(4), 500–507.

Chen, Z., Ding, S. X., Luo, H., & Zhang, K. (2017). An alternative data-driven fault detection scheme for dynamic processes with deterministic disturbances. *Journal of the Franklin Institute*, *354*(1), 556–570.

Chetouani, Y. (2014). Model selection and fault detection approach based on Bayes decision theory: Application to changes detection problem in a distillation column. *Process Safety and Environmental Protection*, *92*(3), 215–223.

Chiang, L. H., Russell, E. L., & Braatz, R. D. (2001). *Fault detection and diagnosis in industrial systems*. London: Springer-Verlag.

Chiang, L. H., Kotanchek, M. E., & Kordon, A. K. (2004). Fault diagnosis based on Fisher discriminant analysis and support vector machines. *Computers and Chemical Engineering*, *28*(8), 1389–1401.

Choi, S. W., Park, J. H., & Lee, I.-B. (2004). Process monitoring using a Gaussian mixture model via principal component analysis and discriminant analysis. *Computers & Chemical Engineering*, *28*(8), 1377–1387.

Choi, S. W., Martin, E. B., Morris, J., & Lee, I. B. (2007). Nonlinear multiscale fault detection and identification. *Fault Detection, Supervision and Safety of Technical Processes 2006*, *1*, 120–125.

Choi, S. W., Morris, J., & Lee, I.-B. (2008). Nonlinear multiscale modelling for fault detection and identification. *Chemical Engineering Science*, *63*(8), 2252–2266. Journal Article.

Dai, X., & Gao, Z. (2013). From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis. *IEEE Transactions on Industrial Informatics*, *9*(4).

Dash, P. K., Nayak, M., Senapati, M. R., & Lee, I. W. C. (2007). Mining for similarities in time series data using wavelet-based feature vectors and neural networks. *Engineering Applications of Artificial Intelligence*, *20*(2), 185–201.

Deng, X., Tian, X., Chen, S., & Harris, C. J. (2017a). Deep learning based nonlinear principal component analysis for industrial process fault detection. *2017 International Joint Conference on Neural Networks (IJCNN)*, 1237–1243.

Deng, X., Tian, X., Chen, S., & Harris, C. J. (2017b). Fault discriminant enhanced kernel principal component analysis incorporating prior fault information for monitoring nonlinear processes. *Chemometrics and Intelligent Laboratory Systems*, *162*(October 2016), 21–34.

Dogantekin, E., Dogantekin, A., & Avci, D. (2011). An expert system based on Generalized Discriminant Analysis and Wavelet Support Vector Machine for diagnosis of thyroid diseases. *Expert Systems with Applications*, *38*(1), 146–150.

Downs, J. J., & Vogel, E. F. (1993). A plant-wide industrial process control problem. *Computers & Chemical Engineering*, *17*, 245–255.

Evsukoff, A., & Gentil, S. (2005). Recurrent neuro-fuzzy system for fault detection and isolation in nuclear reactors. *Advanced Engineering Informatics*, *19*(1), 55–66.

Fan, J., & Wang, Y. (2014). Fault detection and diagnosis of non-linear non-Gaussian dynamic processes using kernel dynamic independent component analysis. *Information Sciences*, *259*, 369–379.

Ferreira, L. S., & Trierweiler, J. O. (2009). Modeling and simulation of the polymeric nanocapsule formation process. *IFAC Proceedings Volumes (IFAC-PapersOnline)*, *7*(PART 1), 405–410.

Fourie, S. H., & De Vaal, P. (2000). Advanced process monitoring using an on-line non-linear multiscale principal component analysis methodology. *Computers and Chemical Engineering*, *24*(2–7), 755–760.

Gao, X., & Hou, J. (2016). An improved SVM integrated GS-PCA fault diagnosis approach of Tennessee Eastman process. *Neurocomputing*, *174*, 906–911.

Ge, Z., Zhang, M., & Song, Z. (2010). Nonlinear process monitoring based on linear subspace and Bayesian inference. *Journal of Process Control*, *20*(5), 676–688.

Ge, Z., Song, Z., & Gao, F. (2013). Review of Recent Research on Data-Based Process Monitoring. *American Chemical Society*, (52), 3543–3562.

Gharahbagheri, H., Imtiaz, S. A., & Khan, F. (2017a). Root Cause Diagnosis of Process Fault Using KPCA and Bayesian Network. *Industrial and Engineering Chemistry Research*, *56*(8), 2054–2070.

Gharahbagheri, H., Imtiaz, S. A., & Khan, F. I. (2017b). Application of Bayesian network for root cause diagnosis of chemical process fault. *Industrial and Engineering Chemistry Research*, *56*(8), 2054–2070

Gharavian, M. H., Almas Ganj, F., Ohadi, A. R., & Heidari Bafroui, H. (2013). Comparison of FDA-based and PCA-based features in fault diagnosis of automobile gearboxes. *Neurocomputing*, *121*, 150–159.

Ghosh, K., Ng, Y. S., & Srinivasan, R. (2011). Evaluation of decision fusion strategies for effective collaboration among heterogeneous fault diagnostic methods. *Computers and Chemical Engineering*, *35*(2), 342–355.

Ghosh, K., Ramteke, M., & Srinivasan, R. (2014). Optimal variable selection for effective statistical process monitoring. *Computers and Chemical Engineering*, *60*, 260–276.

Han, H., Gu, B., Kang, J., & Li, Z. R. (2011). Study on a hybrid SVM model for chiller FDD applications. *Applied Thermal Engineering*, *31*(4), 582–592.

Härdle, W. K. (2011). *Handbooks of Computational Statistics*. *Methods*.

Hong, W., Tian-You, C., Jin-Liang, D., & Martin, B. (2009). Data Driven Fault Diagnosis and Fault Tolerant Control: Some Advances and Possible New Directions. *Acta Automatica Sinica*, *35*(6), 739–747.

Hsu, C.-C., Chen, M.-C., & Chen, L.-S. (2010). Integrating independent component analysis and support vector machine for multivariate process monitoring. *Computers & Industrial Engineering*, *59*(1), 145–156.

Isermann, R., & Ballé, P. (1997). Trends in the application of model based fault detection and diagnosis of technical processes. *Control Eng. Practice, 5*(5), 709–719.

Jiang, Q., & Huang, B. (2016). Distributed monitoring for large-scale processes based on multivariate statistical analysis and Bayesian method. *Journal of Process Control*, *46*, 75–83.

Jiang, Q., Huang, B., & Yan, X. (2016). GMM and optimal principal components-based Bayesian method for multimode fault diagnosis. *Computers and Chemical Engineering*, *84*, 338–349.

Jiang, Q., & Yan, X. (2013). Weighted kernel principal component analysis based on probability density estimation and moving window and its application in nonlinear chemical process monitoring. *Chemometrics and Intelligent Laboratory Systems*, *127*, 121–131.

Jing, C., & Hou, J. (2015). SVM and PCA based fault classification approaches for complicated industrial process. *Neurocomputing*, *167*, 636–642.

Karimi, P., & Jazayeri-Rad, H. (2014). Comparing the Fault Diagnosis Performances of Single Neural Networks and Two Ensemble Neural Networks Based on the Boosting Methods. *Journal of Automation and Control*, *2*(1), 21–32.

Khalid, H. M., & Akram, M. (2011). Fault Modeling , Detection and Classification using Fuzzy Logic , Kalman Filter and Genetic Neuro-Fuzzy Systems. *Asian Journal of Engineering, Sciences & Technology*, *1*(2), 45–57.

Khoukhi, A., Khalid, H., Doraiswami, R., & Cheded, L. (2012). Fault Detection and Classification using Kalman Filter and Hybrid Neuro-Fuzzy Systems. *International Journal of Computer Applications*, *45*(22), 7–14.

Kiong, L. C., Rosmani, C., & Hassan, C. (2010). A Two-Step Fault Detection and Diagnosis Framework for Chemical Processes. *Ajche*, *10*(2), 1–9.

Kulkarni, A., Jayaraman, V. K., & Kulkarni, B. D. (2005). Knowledge incorporated support vector machines to detect faults in Tennessee Eastman Process. *Computers and Chemical Engineering*, *29*(10), 2128–2133.

Lau, C. K., Heng, Y. S., Hussain, M. A., & Mohamad Nor, M. I. (2010). Fault diagnosis of the polypropylene production process (UNIPOL PP) using ANFIS. *ISA Transactions*, *49*(4), 559–566.

Lau, C. K., Ghosh, K., Hussain, M. A., & Che Hassan, C. R. (2013). Fault diagnosis of Tennessee Eastman process with multi-scale PCA and ANFIS. *Chemometrics and Intelligent Laboratory Systems*, *120*, 1–14.

Lee, H. W., Lee, M. W., & Park, J. M. (2009). Multi-scale extension of PLS algorithm for advanced on-line process monitoring. *Chemometrics and Intelligent Laboratory Systems*, *98*(2), 201–212.

Lee, J.-M., Yoo, C., & Lee, I.-B. (2004). Fault detection of batch processes using multiway kernel principal component analysis. *Computers & Chemical Engineering*, *28*(9), 1837–1847.

Lei, Y., He, Z., Zi, Y., & Hu, Q. (2007). Fault diagnosis of rotating machinery based on multiple ANFIS combination with GAs. *Mechanical Systems and Signal Processing*, *21*(5), 2280–2294.

Lei, Y., Zuo, M. J., He, Z., & Zi, Y. (2010). A multidimensional hybrid intelligent method for gear fault diagnosis. *Expert Systems with Applications*, *37*(2), 1419–1430.

Li, C., Ye, H., Wang, G., & Zhang, J. (2005). A recursive nonlinear PLS algorithm for adaptive nonlinear process modeling. *Chemical Engineering and Technology*, *28*(2), 141–152.

Li, G., & Qin, S. J. (2016). Comparative study on monitoring schemes for non-Gaussian distributed processes. *Journal of Process Control*.

Li, H., Mei, C., Zhou, N., Tang, Q., & Huang, Y. (2006). Diagnosis of working conditions of an aluminum reduction cell based on wavelet packets and fuzzy neural network. *Chemical Engineering and Processing: Process Intensification*, *45*(12), 1074–1080.

Li, J., & Cui, P. (2009). Improved kernel fisher discriminant analysis for fault diagnosis. *Expert Systems with Applications*, *36*(2 PART 1), 1423–1432.

Li, Z., Tan, G., & Li, Y. (2012). Fault Diagnosis Based on Improved Kernel Fisher Discriminant Analysis. *Journal of Software*, *7*(12), 2657--2662.

Liu, X., Kruger, U., Littler, T., Xie, L., & Wang, S. (2009). Moving window kernel PCA for adaptive monitoring of nonlinear processes. *Chemometrics and Intelligent Laboratory Systems*, *96*(2), 132–143.

Liu, Y., Ye, L., Zheng, P., Shi, X., Hu, B., & Liang, J. (2010). Multiscale classification and its application to process monitoring. *Journal of Zhejiang University*

SCIENCE C (Comput & Electron), *11*(6), 425–434.

Liu, Z., Cao, H., Chen, X., He, Z., & Shen, Z. (2013). Multi-fault classification based on wavelet SVM with PSO algorithm to analyze vibration signals from rolling element bearings. *Neurocomputing*, *99*.

Lyman, P., & Georgakis, C. (1995). Plant-wide control of the tennessee eastman problem. *Computers and Chemical Engineering*, *19*(3), 321–331.

Mahadevan, S., & Shah, S. L. (2009). Fault detection and diagnosis in process data using one-class support vector machines. *Journal of Process Control*, *19*(10), 1627–1639.

Marwala, T., Mahola, U., & Nelwamondo, F. V. (2006). Hidden Markov Models and Gaussian Mixture Models for Bearing Fault Detection Using Fractals. *International Joint Conference on Neural Networks, Vancouver, Canada*, 3237–3242.

Maulud, A., Wang, D., & Romagnoli, J. A. (2006). A multi-scale orthogonal nonlinear strategy for multi-variate statistical process monitoring. *Journal of Process Control*, *16*(7), 671–683.

Md Nor, N., Hussain, M. A., & Che Hassan, C. R. (2015). Process Monitoring and Fault Detection in Non-Linear Chemical Process Based On Multi-Scale Kernel Fisher Discriminant Analysis. *Computer Aided Chemical Engineering*, *37*, 1823–1828.

Md Nor, N., Hussain, M. A., & Che Hassan, C. R. (2017a). Fault diagnosis and classification framework using multi-scale classification based on kernel Fisher discriminant analysis for chemical process system. *Applied Soft Computing*, *61*, 959–972.

Md Nor, N., Hussain, M. A., & Che Hassan, C. R. (2017b). Fault Diagnosis based on Multi-scale Classification using Kernel Fisher Discriminant Analysis and Gaussian Mixture Model and K-Nearest Neighbor Method. *Jurnal Teknologi*, *79*(5–3), 89–96.

Misra, M., Yue, H. H., Qin, S. J., & Ling, C. (2002). Multivariate process monitoring and fault diagnosis by multi-scale PCA. *Computers and Chemical Engineering*, *26*(9), 1281–1293.

Mohd Ali, J., Ha Hoang, N., Hussain, M. A., & Dochain, D. (2015). Review and classification of recent observers applied in chemical process systems. *Computers and Chemical Engineering*, *76*, 27–41.

Mok, H. T., & Chan, C. W. (2008). Online fault detection and isolation of nonlinear systems based on neurofuzzy networks. *Engineering Applications of Artificial Intelligence*, *21*(2), 171–181.

Monroy, I., Benitez, R., Escudero, G., & Graells, M. (2010). A semi-supervised approach to fault diagnosis for chemical processes. *Computers & Chemical Engineering*, *34*(5), 631–642.

Mylaraswamy, D., & Venkatasubramanian, V. (1997). A hybrid framework for large scale process fault diagnosis. *Computers & Chemical Engineering*, *21*, S935–S940.

Natarajan, S., & Srinivasan, R. (2014). Implementation of multi agents based system for process supervision in large-scale chemical plants. *Computers and Chemical Engineering*, *60*, 182–196.

Peng, Y., Chen, X., Ye, Q., & Jiao, J. (2014). Fault detection and classification in chemical processes using NMFSC and structural SVMs. *Canadian Journal of Chemical Engineering*, *92*(6), 1016–1023.

Ruiz, D., Nougues, J. M., & Puigjaner, L. (2001). Fault diagnosis support system for complex chemical plants. *Computers and Chemical Engineering*, *25*, 151–160.

Sabura Banu, U., & Uma, G. (2011). ANFIS based sensor fault detection for continuous stirred tank reactor. *Applied Soft Computing Journal*, *11*(2), 2618–2624.

Salahshoor, K., Kordestani, M., & Khoshro, M. S. (2010). Fault detection and diagnosis of an industrial steam turbine using fusion of SVM (support vector machine) and ANFIS (adaptive neuro-fuzzy inference system) classifiers. *Energy*, *35*(12), 5472–5482.

Seera, M., Lim, C. P., Ishak, D., & Singh, H. (2013). Offline and online fault detection and diagnosis of induction motors using a hybrid soft computing model. *Applied Soft Computing*, *13*(12), 4493–4507.

Seera, M., Lim, C. P., Nahavandi, S., & Loo, C. K. (2014). Condition monitoring of induction motors: A review and an application of an ensemble of hybrid intelligent models. *Expert Systems with Applications*, *41*(10), 4891–4903.

Seng Ng, Y., Srinivasan, R., Ng, Y. S., & Srinivasan, R. (2010). Multi-agent based collaborative fault detection and identification in chemical processes. *Engineering Applications of Artificial Intelligence*, *23*(6), 934–949.

Slišković, D., Grbić, R., & Hocenski, Ž. (2012). Multivariate Statistical Process Monitoring. *Tehnicki Vjesnik*. *41*(10), 4891–4903.

Souza, D. L. De, Granzotto, M. H., Almeida, G. M. De, & Oliveira-lopes, L. C. (2014). Fault Detection and Diagnosis Using Support Vector Machines - A SVC and SVR Comparison. *Journal of Safety Engineering*, *3*(1), 18–29.

Tafazzoli, E., & Saif, M. (2009). Application of combined support vector machines in process fault diagnosis. *Proceedings of the American Control Conference*, 3429–3433.

Tiwari, R., Gupta, V. K., & Kankar, P. K. (2015). Bearing fault diagnosis based on multi-scale permutation entropy and adaptive neuro fuzzy classifier. *Journal of Vibration and Control*, *21*(3), 461–467.

Ündey, C., Tatara, E., & Cįnar, A. (2003). Real-time batch process supervision by integrated knowledge-based systems and multivariate statistical methods. *Engineering Applications of Artificial Intelligence*, *16*(5–6), 555–566.

Uppal, F. J., Patton, R. J., & Witczak, M. (2006). A neuro-fuzzy multiple-model observer approach to robust fault diagnosis based on the DAMADICS benchmark problem. *Control Engineering Practice*, *14*(6 SPEC. ISS.), 699–717.

Uraikul, V., Chan, C. W., & Tontiwachwuthikul, P. (2007). Artificial intelligence for monitoring and supervisory control of process systems. *Engineering Applications of Artificial Intelligence*, *20*(2), 115–131.

Venkatasubramanian, V., Rengaswamy, R., & Kavuri, S. N. (2003a). A review of process fault detection and diagnosisPart II: Qualitative models and search strategies. *Computers and Chemical Engineering*, *27*, 313–326.

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S. N., & Yin, K. (2003b). A review of process fault detection and diagnosisPart III: Process history based methods. *Computers and Chemical Engineering*, *27*, 327–346.

Venkatasubramanian, V., Rengaswamy, R., Yin, K., & Kavuri, S. N. (2003c). A review of process fault detection and diagnosisPart I: Quantiative model-based methods. *Computers and Chemical Engineering*, *27*, 293–311.

Wang, D., & Romagnoli, J. A. (2005). Robust multi-scale principal components analysis with applications to process monitoring. *Journal of Process Control*, *15*(8), 869–882.

Wise, B. M., Gallagher, N. B., Butler, S. W., White, D. D., & Barna, G. G. (1999). A comparison of principal component analysis, multiway principal component analysis, trilinear decomposition and parallel factor analysis for fault detection in a semiconductor etch process. *Journal of Chemometrics*, *13*(3–4), 379–396.

Wu, J.-D., Hsu, C.-C., & Wu, G.-Z. (2009a). Fault gear identification and classification using discrete wavelet transform and adaptive neuro-fuzzy inference. *Expert Systems with Applications*, *36*(3), 6244–6255.

Wu, J. D, & Hsu, C. C. (2009b). Fault gear identification using vibration signal with discrete wavelet transform technique and fuzzy-logic inference. *Expert Systems with Applications*, *36*(2 PART 2), 3785–3794.

Wu, S.-D. De, Wu, P.-H. H., Wu, C.-W. W., Ding, J.-J. J., & Wang, C.-C. C. (2012). Bearing fault diagnosis based on multiscale permutation entropy and support vector machine. *Entropy*, *14*(8), 1343–1356.

Xiao, Y., Wang, H., Zhang, L., & Xu, W. (2014). Two methods of selecting Gaussian kernel parameters for one-class SVM and their application to fault detection. *Knowledge-Based Systems*, *59*, 75–84.

Xue, X., & Zhou, J. (2016). A hybrid fault diagnosis approach based on mixed-domain state features for rotating machinery. *ISA Transactions*.

Yang, C., & Hou, J. (2016). Fed-batch fermentation penicillin process fault diagnosis and detection based on support vector machine. *Neurocomputing*, *190*, 117–123.

Yang, R. A. C., Zhou, Z., Wang, L., & Pan, Y. (2015). Comparison of Different

Optimization Methods with Support Vecto Machine for Blast Furnace Multi-Fault Classification. *IFAC-PapersOnLine*, *48*(21), 1204–1209.

Yélamos, I., Escudero, G., & Graells, M. (2006). Fault diagnosis based on support vector machines and systematic comparison to existing approaches. *16th European Symposium on Computer Aided Process Engineering and 9th International Symposium on Process Systems Engineering*, (2003), 1209–1214.

Yélamos, I., Escudero, G., & Graells, M. (2007). Simultaneous fault diagnosis in chemical plants using Support Vector Machines. *17th European Symposium on Computer Aided Process Engineering – ESCAPE17*, 1–6.

Yelamos, I., Graells, M., & Puigjaner, L. (2007). Simulataneous fault diagnosis in chemical plants using a multilabel approach. *AIChE Journal*, *53*(11).

Yélamos, I., Escudero, G., Graells, M., & Puigjaner, L. (2009). Performance assessment of a novel fault diagnosis system based on support vector machines. *Computers & Chemical Engineering*, *33*(1), 244–255.

Yin, S., Gao, X., Karimi, H. R., & Zhu, X. (2014). Study on support vector machine-based fault detection in Tennessee Eastman process. *Abstract and Applied Analysis*, *2014*.

Yong, M., Zheng, X., Zheng, Y., Youxian, S., & Zheng, W. (2007). Fault diagnosis based on fuzzy support vector machine with parameter tuning and feature selection. *Chinese Journal of Chemical Engineering*, *15*(2), 233–239.

Yu, J. (2016). Process monitoring through manifold regularization-based GMM with global/local information. *Journal of Process Control*, *45*, 84–99.

Zamanian, A. H., & Ohadi, A. (2011). Gear fault diagnosis based on Gaussian correlation of vibrations signals and wavelet coefficients. *Applied Soft Computing Journal*, *11*(8), 4807–4819.

Zhang, F., & Ge, Z. (2015). Decision fusion systems for fault detection and identification in industrial processes. *Journal of Process Control*, *31*, 45–54.

Zhang, J. (2006). Improved on-line process fault diagnosis through information fusion in multiple neural networks. *Computers and Chemical Engineering*, *30*(3), 558–571.

Zhang, L., Xiong, G., Liu, H., Zou, H., & Guo, W. (2010). Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference. *Expert Systems with Applications*, *37*(8), 6077–6085.

Zhang, S., Zhao, C., Wang, S., & Wang, F. (2017). Pseudo Time-Slice Construction Using a Variable Moving Window k Nearest Neighbor Rule for Sequential Uneven Phase Division and Batch Process Monitoring. *Industrial and Engineering Chemistry Research*, *56*(3), 728–740.

Zhang, Y. (2009). Enhanced statistical analysis of nonlinear processes using KPCA, KICA and SVM. *Chemical Engineering Science*, *64*(5), 801–811.

Zhang, Y., Teng, Y., & Zhang, Y. (2010). Complex process quality prediction using modified kernel partial least squares. *Chemical Engineering Science*, *65*(6), 2153–2158.

Zhang, Y., & Hu, Z. (2011). Multivariate process monitoring and analysis based on multi-scale KPLS. *Chemical Engineering Research and Design*, *89*(12), 2667–2678.

Zhang, Y., & Ma, C. (2011). Fault diagnosis of nonlinear processes using multiscale KPCA and multiscale KPLS. *Chemical Engineering Science*, *66*(1), 64–72.

Zhang, Y., Li, S., & Hu, Z. (2012). Improved multi-scale kernel principal component analysis and its application for fault detection. *Chemical Engineering Research and Design*, *90*(9), 1271–1280.

Zhang, Y., Li, S., & Teng, Y. (2012). Dynamic processes monitoring using recursive kernel principal component analysis. *Chemical Engineering Science*, *72*, 78–86.

Zhao, H., Liu, J., Dong, W., Sun, X., & Ji, Y. (2017). An improved case-based reasoning method and its application on fault diagnosis of Tennessee Eastman process. *Neurocomputing*, *249*, 266–276.

Zhu, J., Ge, Z., & Song, Z. (2016). Distributed Gaussian Mixture Model for Monitoring Multimode Plant-wide Process. In *2016 28th Chinese Control and Decision Conference* (pp. 5826–5831).

Zhu, Z.-B. B., & Song, Z.-H. H. (2011). A novel fault diagnosis system using pattern classification on kernel FDA subspace. *Expert Systems with Applications*, *38*(6), 6895–6905.

LIST OF PUBLICATIONS AND PAPERS PRESENTED

Publications

1. Norazwan Md Nor, Mohd Azlan Hussain, Che Rosmani Che Hassan, Process Monitoring and Fault Detection in Non-linear Chemical Process Based on Multi-scale Kernel Fisher Discriminant Analysis, (2015) *Computer-Aided Chemical Engineering*, 37, pp. 1823-1828.

2. Norazwan Md Nor, Mohd Azlan Hussain, Che Rosmani Che Hassan, Fault Diagnosis based on Multi-scale Classification Using Kernel Fisher Discriminant Analysis and Gaussian Mixture Model and K-Nearest Neighbor Method, (2017) *Jurnal Teknologi, 79:5-3, pp. 89-96.*

3. Norazwan Md Nor, Mohd Azlan Hussain, Che Rosmani Che Hassan, Fault Detection and Diagnosis Framework Using Multi-scale Classification Based on Kernel Fisher Discriminant Analysis with SVM Method, (2017) *Applied Soft Computing, 61: 959-972.*

4. Norazwan Md Nor, Mohd Azlan Hussain, Che Rosmani Che Hassan, A Review of Data-driven Fault Detection and Diagnosis Methods: Applications in Chemical Process Systems, *Reviews in Chemical Engineering (In press)*

5. Norazwan Md Nor, Mohd Azlan Hussain, Che Rosmani Che Hassan, Multi-scale Fisher Discriminant Analysis with Adaptive Neuro-Fuzzy Inference System (ANFIS) in Fault Detection and Diagnosis Scheme for Chemical Process Systems, *Neural Computing and Application (Under Revision)*

Proceedings

1. Process Monitoring and Fault Detection in Non-linear Chemical Process Based on Multi-scale Kernel Fisher Discriminant Analysis
12th International Symposium on Process Systems Engineering and 25th European Symposium on Computer Aided Process Engineering, 31 May – 4 June 2015, Copenhagen, Denmark

2. Fault Diagnosis Using Multi-scale Classification Based on Fisher Discriminant Analysis and Support Vector Machine
The 22nd Regional Symposium on Chemical Engineering: RSCE 2015, 24-25 September 2015, Bangkok, Thailand

3. Fault Diagnosis based on Multi-scale Classification Using Kernel Fisher Discriminant Analysis and Gaussian Mixture Model and K-Nearest Neighbor Method
28th Symposium of Malaysian Chemical Engineers (SoMChE 2015), 21-22 October 2015, Putrajaya, Malaysia