

GRAPH-BASED MACHINE LEARNING
FRAMEWORKS FOR PROCESS MONITORING AND
OPTIMIZATION

UMANG GOSWAMI



DEPARTMENT OF CHEMICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY DELHI

JANUARY 2025

Graph-Based Machine Learning Frameworks for Process Monitoring and Optimization

by

Umang Goswami

Department of Chemical Engineering

Submitted

in fulfilment of the requirements of the degree of

Doctor of Philosophy

to the



INDIAN INSTITUTE OF TECHNOLOGY DELHI

January 2025

THESIS CERTIFICATE

This is to certify that the thesis titled **Graph-Based Machine Learning Frameworks for Process Monitoring and Optimization**, submitted by **Umang Goswami (2020CHZ8402)**, to the Indian Institute of Technology Delhi, for the award of the degree of **Doctor of Philosophy**, is a bonafide record of the research work done by him under my supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Prof. Hariprasad Kodamana

Research Supervisor

Dept. of Chemical Engineering &
Yardi School of Artificial In-
telligence, Indian Institute of
Technology-Delhi, Hauz khas,
Delhi-110016

Date: January 07, 2025

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my supervisor, Prof. Hariprasad Kodamana, whose unwavering support, encouragement, and vast knowledge have been instrumental throughout this journey. His guidance was the beacon that led me through many challenging moments, and his technical expertise consistently pushed me to explore every dimension of the problems I faced. His dedication and professionalism inspired me to push my own limits, and his empathy—an exceptional quality—set him apart and made this journey more manageable.

I am also grateful to Prof. Manojkumar Ramteke, Prof. Sudip Pattanayek, and Prof. Subhasish Datta for their insightful feedback and for monitoring my progress throughout each semester. Their perspectives helped me stay on track and grow as a researcher. I also had the privilege of collaborating with Prof. Manojkumar Ramteke, Prof. Sandeep Kumar and Prof. Agam Gupta on various projects, and their clear and concise input played a crucial role in shaping my research direction.

My sincere thanks also go to my labmates, who have been constant companions during this journey, whether it was debugging a tricky piece of code or discussing broad research ideas. Special thanks to my seniors — Tanuja, Reena, Avan, Nikita, and Jyoti—for their invaluable guidance. I would also like to thank my batch mates Nasre, Deepak, and Ahtesham for their support and for helping to keep my spirits high. To my juniors—Arjun, Shobhita, Anirudh, Amal, Robin, Vinayak, and Divyam—thank you for your assistance and camaraderie.

I would also like to extend my appreciation to the officials of the Department of Chemical Engineering and IIT Delhi for their help with administrative matters, which ensured that my work proceeded smoothly.

They say friends are the family we choose, and I have been fortunate to have a wonderful circle of friends who walked this path with me. I am especially thankful to Anusha for keeping me grounded and motivated during the toughest days and to Raghav and Rituraj

for their constant companionship and support.

Finally, I would like to thank the almighty, and I would also like to express my heartfelt thanks to my sister, Poorvi, whose humour lightened the most difficult days, and to my parents, whose unwavering love and strength have been the foundation of my journey. Without their care and prayers, none of this would have been possible. I would also like to remember my grandparents, whose encouragement and belief in my abilities have always been a source of inspiration.

ABSTRACT

The integration of Industry 4.0 technologies has transformed process industries by embedding real-time analytics, automation, and data-driven decision-making into core operations. Central to this transformation is the adoption of advanced fault detection and diagnosis (FDD) systems and change point detection (CPD) methodologies, which are critical for ensuring the safety, reliability, and efficiency of industrial processes. In complex process units, timely detection and isolation of faults are paramount to prevent undesirable events that could lead to operational downtime, financial losses, or catastrophic failures.

Traditional modeling approaches, such as First-Principles models and knowledge-based techniques, often struggle to adapt to the dynamic and complex nature of modern industrial processes. The variability and real-time updates inherent in these systems make it challenging for such models to provide accurate and timely insights. Data-driven approaches, empowered by machine learning (ML) and deep learning (DL) algorithms, have emerged as powerful tools to address these challenges by learning directly from process data. However, conventional ML and DL methods may not fully capture the structural dependencies and relational information present in industrial processes.

This thesis presents innovative solutions leveraging graph-based machine learning to enhance fault detection, change point detection, and multi-objective optimization in industrial hydrocracking operations. Graph machine learning models offer significant advantages over traditional methods by effectively handling structural data, capturing data dependencies, enabling enhanced representation learning through node embeddings, and improving generalization to new processes without extensive retraining. By representing process units as nodes and their interactions as edges, graph-based models naturally encapsulate the complex relationships within industrial systems.

Fault detection and diagnosis are addressed by developing a Spectral Weighted Graph Auto-Encoder (SWGAE) framework that models multivariate time-series data as complete graphs with weighted adjacency matrices. Training the model on normal operational data

allows for the establishment of thresholds to detect anomalies. This approach not only identifies faults but also isolates the contributing variables by calculating their percentage contributions, thereby pinpointing the source of anomalies within the process unit. Comparative studies demonstrate that the SWGAE method outperforms traditional techniques such as Auto-Encoders (AE), Long Short-Term Memory networks (LSTM), and Dynamic Principal Component Analysis (DPCA) in terms of accuracy and reliability.

To further enhance fault detection capabilities, a data-agnostic approach using Graph Neural Differential Auto-Encoders (GNDAE) is introduced. This model integrates Graph Convolutional Networks (GCNs) with a Graph Neural Differential Equation (GNDE) solver, providing continuous parameterization of hidden states. The GNDAE model treats the neural network as a differential equation, allowing it to learn values at infinitesimal points and capture both spatial and temporal dependencies in the data. This continuous modeling approach improves the understanding of underlying datasets and leads to highly accurate fault detection results.

Change point detection is addressed by applying the GNDAE model within an unsupervised learning framework. By modeling time-series data as graphs and incorporating contrastive learning techniques, the model effectively identifies significant changes in process parameters, indicating the onset of faults or transitions in operational states. The model's ability to forecast feature values at subsequent time steps enhances its capability to detect magnitude changes, outperforming traditional methods such as CUSUM and LSTM-based approaches in chemical process case studies.

The importance of multi-objective optimization in industrial processes is emphasized, particularly in hydrocracking operations—a critical process in petroleum refining that converts heavy hydrocarbons into lighter, more valuable products like diesel, kerosene, and aviation turbine fuel (ATF). Optimizing the hydrocracking process is essential for maximizing yields of desirable products while minimizing resource consumption, such as hydrogen, thereby enhancing economic efficiency and contributing to sustainable resource utilization.

To address this, the thesis develops an Hydrocarbon Graph Neural Differential Equation (HCGNDE) model that represents the hydrocracker unit by modeling the underlying partial differential equations as graphs. Variables are represented as nodes, and their relationships form the edges, capturing the axial dispersion of components through catalyst beds. The

continuous parameterization inherent in the HCGNDE model allows for precise modeling of concentration changes across the process. The model is then utilized for multi-objective optimization to maximize the yields of ATF and diesel while minimizing hydrogen consumption. Results demonstrate significant improvements, with hydrogen usage reduced by 25.5% and yields of ATF and diesel maintained at high levels. This showcases the practical benefits of applying graph-based machine learning to complex industrial optimization problems.

In summary, this thesis provides comprehensive solutions for enhancing the safety, reliability, and efficiency of industrial process units through advanced graph-based machine learning techniques. By focusing on data-driven approaches aligned with Industry 4.0 objectives, the research addresses critical challenges in fault detection, change point detection, and multi-objective optimization. The proposed methodologies leverage the structural and relational properties inherent in industrial processes, offering robust, accurate, and scalable solutions that outperform traditional methods. The findings have practical implications for improving operational performance, reducing risks, and optimizing resource utilization in critical industrial processes like hydrocracking.

KEYWORDS: Data-Driven Approaches; Industry 4.0; Fault Detection and Diagnosis; Change Point Detection; Process Monitoring; Graph Machine Learning; Machine Learning; Deep Learning; Graph Representation Learning; Multi-Objective Optimization; Hydrocracking Operation; Industrial Processes

सारांश

उद्योग 4.0 प्रौद्योगिकियों के एकीकरण ने वास्तविक समय विश्लेषण, स्वचालन और डेटा-संचालित निर्णय लेने को मुख्य संचालन में एम्बेड करके प्रक्रिया उद्योगों को बदल दिया है। इस परिवर्तन का केंद्र उन्नत दोष पहचान और निदान (FDD) प्रणालियों और परिवर्तन बिंदु पहचान (CPD) पद्धतियों को अपनाना है, जो औद्योगिक प्रक्रियाओं की सुरक्षा, विश्वसनीयता और दक्षता सुनिश्चित करने के लिए महत्वपूर्ण हैं। जटिल प्रक्रिया इकाइयों में, समय पर दोषों का पता लगाना और उन्हें अलग करना अवांछनीय घटनाओं को रोकने के लिए सर्वोपरि है जो परिचालन डाउनटाइम, वित्तीय नुकसान या भयावह विफलताओं का कारण बन सकते हैं।

पारंपरिक मॉडलिंग दृष्टिकोण, जैसे कि प्रथम-सिद्धांत मॉडल और ज्ञान-आधारित तकनीकें, अक्सर आधुनिक औद्योगिक प्रक्रियाओं की गतिशील और जटिल प्रकृति के अनुकूल होने के लिए संघर्ष करती हैं। इन प्रणालियों में निहित परिवर्तनशीलता और वास्तविक समय के अपडेट ऐसे मॉडलों के लिए सटीक और समय पर जानकारी प्रदान करना चुनौतीपूर्ण बनाते हैं। मशीन लर्निंग (ML) और डीप लर्निंग (DL) एल्गोरिदम द्वारा सशक्त डेटा-संचालित दृष्टिकोण, प्रक्रिया डेटा से सीधे सीखकर इन चुनौतियों का समाधान करने के लिए शक्तिशाली उपकरण के रूप में उभरे हैं। हालाँकि, पारंपरिक ML और DL विधियाँ औद्योगिक प्रक्रियाओं में मौजूद संरचनात्मक निर्भरताओं और संबंधपरक जानकारी को पूरी तरह से कैप्चर नहीं कर सकती हैं।

यह थीसिस औद्योगिक हाइड्रोकार्बन संचालन में दोष पहचान, परिवर्तन बिंदु पहचान और बहु-उद्देश्यीय अनुकूलन को बढ़ाने के लिए ग्राफ-आधारित मशीन लर्निंग का लाभ उठाने वाले अभिनव समाधान प्रस्तुत करती है। ग्राफ मशीन लर्निंग मॉडल संरचनात्मक डेटा को प्रभावी ढंग से संभालने, डेटा निर्भरताओं को कैप्चर करने, नोड एम्बेडिंग के माध्यम से उन्नत प्रतिनिधित्व सीखने को सक्षम करने और व्यापक पुनर्प्रशिक्षण के बिना नई प्रक्रियाओं के लिए सामान्यीकरण में सुधार करके पारंपरिक तरीकों पर महत्वपूर्ण लाभ प्रदान करते हैं। प्रक्रिया इकाइयों को नोड्स और उनके इंटरैक्शन को किनारों के रूप में दर्शाकर, ग्राफ-आधारित मॉडल स्वाभाविक रूप से औद्योगिक प्रणालियों के भीतर जटिल संबंधों को समाहित करते हैं।

स्पेक्ट्रल वेटेड ग्राफ ऑटो-एनकोडर (SWGAE) फ्रेमवर्क विकसित करके दोष पहचान और निदान को संबोधित किया जाता है जो बहुभिन्नरूपी समय-श्रृंखला डेटा को भारित आसन्न मैट्रिसेस के साथ पूर्ण ग्राफ़ के रूप में मॉडल करता है। सामान्य परिचालन डेटा पर मॉडल को प्रशिक्षित करने से विसंगतियों का पता लगाने के लिए थ्रेसहोल्ड की स्थापना की अनुमति मिलती है। यह दृष्टिकोण न केवल दोषों की पहचान करता है, बल्कि उनके प्रतिशत योगदान की गणना करके योगदान देने वाले चरों को भी अलग करता है, जिससे प्रक्रिया इकाई के भीतर विसंगतियों के स्रोत का पता चलता है। तुलनात्मक अध्ययनों से पता चलता है कि SWGAE विधि सटीकता और विश्वसनीयता के मामले में ऑटो-एनकोडर (AE), लॉन्ग शॉर्ट-टर्म मेमोरी नेटवर्क (LSTM) और डायनेमिक प्रिंसिपल कंपोनेंट एनालिसिस (DPCA) जैसी पारंपरिक तकनीकों से बेहतर प्रदर्शन करती है।

दोष पहचान क्षमताओं को और बेहतर बनाने के लिए, ग्राफ न्यूरल डिफरेंशियल ऑटो-एनकोडर (GNDAE) का उपयोग करके डेटा-अज्ञेय दृष्टिकोण पेश किया गया है। यह मॉडल ग्राफ कन्वोल्यूशनल नेटवर्क (GCN) को ग्राफ न्यूरल डिफरेंशियल इन्फ्लेक्शन (GNDE) सॉल्वर के साथ एकीकृत करता है, जो

छिपी हुई स्थितियों का निरंतर पैरामीटराइजेशन प्रदान करता है। GNDAE मॉडल न्यूरल नेटवर्क को एक अंतर समीकरण के रूप में मानता है, जिससे यह अनंत बिंदुओं पर मान सीख सकता है और डेटा में स्थानिक और लौकिक दोनों निर्भरताओं को पकड़ सकता है। यह निरंतर मॉडलिंग दृष्टिकोण अंतर्निहित डेटासेट की समझ में सुधार करता है और अत्यधिक सटीक दोष पहचान परिणामों की ओर ले जाता है। परिवर्तन बिंदु पहचान को GNDAE मॉडल को एक अप्रशिक्षित शिक्षण ढांचे के भीतर लागू करके संबोधित किया जाता है। समय-श्रृंखला डेटा को ग्राफ़ के रूप में मॉडलिंग करके और विपरीत शिक्षण तकनीकों को शामिल करके, मॉडल प्रक्रिया मापदंडों में महत्वपूर्ण परिवर्तनों को प्रभावी ढंग से पहचानता है, जो परिचालन स्थितियों में दोषों या संक्रमणों की शुरुआत का संकेत देता है। मॉडल की बाद के समय चरणों में विशेषता मानों का पूर्वानुमान लगाने की क्षमता, परिमाण परिवर्तनों का पता लगाने की इसकी क्षमता को बढ़ाती है, जो रासायनिक प्रक्रिया केस अध्ययनों में CUSUM और LSTM-आधारित दृष्टिकोणों जैसे पारंपरिक तरीकों से बेहतर प्रदर्शन करती है।

औद्योगिक प्रक्रियाओं में बहु-उद्देश्यीय अनुकूलन के महत्व पर जोर दिया जाता है, विशेष रूप से हाइड्रोक्रैकिंग संचालन में - पेट्रोलियम शोधन में एक महत्वपूर्ण प्रक्रिया जो भारी हाइड्रोकार्बन को डीजल, केरोसिन और विमानन टरबाइन ईंधन (ATF) जैसे हल्के, अधिक मूल्यवान उत्पादों में परिवर्तित करती है। हाइड्रोक्रैकिंग प्रक्रिया का अनुकूलन वांछित उत्पादों की पैदावार को अधिकतम करने के लिए आवश्यक है जबकि संसाधन खपत को कम से कम करना, जैसे कि हाइड्रोजन, जिससे आर्थिक दक्षता में वृद्धि होती है और संधारणीय संसाधन उपयोग में योगदान मिलता है।

इसे संबोधित करने के लिए, थीसिस एक हाइड्रोकार्बन ग्राफ न्यूरल डिफरेंशियल इंकेशन (HCGNDE) मॉडल विकसित करती है जो अंतर्निहित आंशिक अंतर समीकरणों को ग्राफ के रूप में मॉडलिंग करके हाइड्रोक्रैकिंग इकाई का प्रतिनिधित्व करती है। चर को नोड्स के रूप में दर्शाया जाता है, और उनके संबंध किनारों का निर्माण करते हैं, उत्प्रेरक बेड के माध्यम से घटकों के अक्षीय फैलाव को कैप्चर करते हैं। HCGNDE मॉडल में निहित निरंतर पैरामीटराइजेशन प्रक्रिया में सांद्रता परिवर्तनों के सटीक मॉडलिंग की अनुमति देता है। फिर मॉडल का उपयोग बहु-उद्देश्यीय अनुकूलन के लिए किया जाता है ताकि हाइड्रोजन की खपत को कम करते हुए ATF और डीजल की पैदावार को अधिकतम किया जा सके। परिणाम महत्वपूर्ण सुधार प्रदर्शित करते हैं, हाइड्रोजन के उपयोग में 25.5% की कमी आई है और ATF और डीजल की पैदावार उच्च स्तर पर बनी हुई है। यह जटिल औद्योगिक अनुकूलन समस्याओं के लिए ग्राफ-आधारित मशीन लर्निंग को लागू करने के व्यावहारिक लाभों को दर्शाता है।

संक्षेप में, यह थीसिस उन्नत ग्राफ-आधारित मशीन लर्निंग तकनीकों के माध्यम से औद्योगिक प्रक्रिया इकाइयों की सुरक्षा, विश्वसनीयता और दक्षता बढ़ाने के लिए व्यापक समाधान प्रदान करती है। उद्योग 4.0 उद्देश्यों के साथ संरेखित डेटा-संचालित दृष्टिकोणों पर ध्यान केंद्रित करके, अनुसंधान दोष का पता लगाने, परिवर्तन बिंदु का पता लगाने और बहु-उद्देश्यीय अनुकूलन में महत्वपूर्ण चुनौतियों का समाधान करता है। प्रस्तावित कार्यप्रणाली औद्योगिक प्रक्रियाओं में निहित संरचनात्मक और संबंधपरक गुणों का लाभ उठाती है, जो पारंपरिक तरीकों से बेहतर प्रदर्शन करने वाले मजबूत, सटीक और स्केलेबल समाधान प्रदान करती है। निष्कर्षों में परिचालन प्रदर्शन में सुधार, जोखिम को कम करने और हाइड्रोक्रैकिंग जैसी महत्वपूर्ण औद्योगिक प्रक्रियाओं में संसाधन उपयोग को अनुकूलित करने के लिए व्यावहारिक निहितार्थ हैं।

कीवर्ड: डेटा-संचालित दृष्टिकोण; उद्योग 4.0; दोष पहचान और निदान; परिवर्तन बिंदु पहचान; प्रक्रिया निगरानी; ग्राफ मशीन लर्निंग; मशीन लर्निंग; डीप लर्निंग; ग्राफ प्रतिनिधित्व लर्निंग; बहु-उद्देश्य अनुकूलन; हाइड्रोक्रैकिंग ऑपरेशन; औद्योगिक प्रक्रियाएँ।

Contents

ACKNOWLEDGEMENTS	i
ABSTRACT	iii
LIST OF FIGURES	xiv
LIST OF TABLES	xvii
ABBREVIATIONS	xviii
NOTATION	xx
1 INTRODUCTION	1
1.1 Motivation	1
1.2 Research Contributions	6
1.3 Outline of the Thesis	9
2 Literature Review	10
2.1 Traditional Process Monitoring Methodologies	10
2.1.1 Principal Component Analysis (PCA)	11
2.1.2 Autoencoder (AE)	11
2.1.3 Long Short-Term Memory (LSTM)	12
2.1.4 LSTM Autoencoder	13
2.2 Graph representation Learning (GRL)	14
2.2.1 Matrix Factorization Techniques	14
2.2.2 Neural Network-Based Methodologies	17
2.2.3 Large Graph-Based Techniques	30
2.2.4 Random Walk-Based Techniques	32
2.3 Continuous Parameterized Models	34
2.3.1 Neural Ordinary Differential Equations (NODE)	34

2.3.2	Graph Neural Differential Equations (GNDEs)	34
2.4	Multi-objective Optimization	36
2.4.1	Traditional Multi-objective Optimization Algorithms	36
2.4.2	Neural Network-based Optimization Algorithms	37
2.4.3	Optimization Algorithms in Hydrocracking Case Studies	38
2.5	Research Gaps	39
3	Fault detection and isolation of multi-variate time series data using spectral weighted graph auto-encoders	41
3.1	Methodology	43
3.1.1	Learning of spectral graph	43
3.1.2	Reconstruction-based fault detection using auto-encoders	48
3.1.3	Fault isolation	50
3.2	Numerical experiments	53
3.2.1	Example 1	53
3.2.2	Datasets for experiments	55
3.2.3	Results	57
3.3	Conclusion	63
4	A Graph Embedding Based Fault Detection Framework for Process Systems with Multi-variate Time-series Datasets	65
4.1	Methodology	68
4.1.1	Learning of spectral graph	68
4.1.2	Shallow Embedding Methods	69
4.1.3	Variational Autoencoder for Fault Identification	73
4.2	Experiments conducted with numerical data	75
4.2.1	Datasets for Simulation	76
4.2.2	Results	77
4.3	Conclusion	82
5	Fault Detection Using Graph Neural Differential Auto-encoders (GN-DAE)	87
5.1	Preliminaries	90
5.1.1	Graph Learning	90

5.1.2	Graph Convolution Operation	93
5.1.3	Neural ordinary differential equations	94
5.2	GNDAE Formulation and fault detection framework	95
5.2.1	Fault Detection Framework	98
5.3	Results and Discussion	101
5.3.1	Validation datasets	101
5.3.2	Numerical Experiments	102
5.3.3	Comparison with baseline methods	108
5.4	Conclusion	112
6	Change Point Detection of Processes using Graph Neural Differential Convolutional Network with Contrastive Loss	113
6.1	Methodology	115
6.1.1	Representing Process Systems as Graphs	115
6.1.2	Model Formulation	118
6.1.3	Contrastive Learning Framework	121
6.1.4	Change Point Detection Framework	123
6.2	Case Studies	124
6.2.1	Case Study 1: Interacting Two Tank System	125
6.2.2	Case Study 2: Continuous Stirred Tank Reactor (CSTR)	126
6.2.3	Case Study 3: Carbon Capture and Storage System	127
6.3	Results and Discussion	129
6.3.1	Numerical Experiments	129
6.3.2	Validation of the model	130
6.3.3	Comparison with baseline methods	134
6.4	Conclusions and Future Work	136
7	Multi-objective Optimization of Hydrocracking Processes Using Hydrocracking Graph Neural Differential Equations (HCGNDE) for Enhanced Diesel and Aviation Fuel Yields	137
7.1	Introduction	137
7.2	Hydrocracking Process	140
7.3	Methodology	143

7.3.1	Representing Process Systems as Graphs	143
7.3.2	Neural Ordinary Differential Equations	145
7.3.3	Hydrocracking Graph Neural Differential Equations (HCGNDE) Model	147
7.3.4	Multi-objective Optimization	150
7.4	Results and Discussions	152
7.4.1	GNDE Architecture Details	152
7.4.2	Model Validation	154
7.4.3	Optimization Results	158
7.5	Conclusion	162
8	Conclusion and Future directions	164
8.1	Process Monitoring with Graph-Based Models	164
8.1.1	Spectral Weighted Graph Autoencoders (SWGAE) for Fault Detection and Isolation	164
8.1.2	Graph Embedding-Based Method for Fault Detection	165
8.1.3	Graph Neural Differential Autoencoder (GNDAE) for Process Moni- toring	166
8.1.4	Graph Neural Differential Contrastive Learner (GNDCL) for Change Point Detection	167
8.2	Multi-Objective Optimization with Graph-Based Models	168
8.2.1	Hydrocarbon Graph Neural Differential Equation (HCGNDE) for Multi- Objective Optimization	168
8.3	Future Directions	169
	BIBLIOGRAPHY	171
	APPENDIX	199
A	Fault detection and isolation of multi-variate time series data using spec- tral weighted graph auto-encoders	200
A.1	Fault Detection of other Faults from the Tennessee Eastman data	200
B	Fault Detection Using Graph Neural Differential Auto-encoders (GN- DAE)	208
B.1	Fault Detection for the Tennessee Eastman Process dataset	208
B.2	Fault Detection for the WWTP Process dataset	208

C	Multi-objective Optimization of Hydrocracking Processes Using Hydrocracking Graph Neural Differential Equations (HCGNDE) for Enhanced Diesel and Aviation Fuel Yields	214
C.1	Nomenclature	214
C.2	Equations for simulating the properties of components	214
	LIST OF PUBLICATIONS	224

List of Figures

1.1	Overall structure of the thesis.	5
2.1	A schematic of the Auto-encoder Model.	12
2.2	A schematic of the Long Short Term Memory Model.	13
2.3	A schematic of the LSTM-AE Model.	14
2.4	Graph Representation Learning and its classification based on various method-ologies.	15
2.5	Graph Neural Network.	18
2.6	Graph Convolutional Network.	19
2.7	Graph Attention Network.	22
2.8	Graph Isomorphism Network.	24
2.9	Signed Graph Convolutional Network.	26
2.10	A schematic of the Random walk-based approaches comprising of the Node2vec and Deepwalk algorithms.	33
3.1	Schematic of the proposed work.	45
3.2	A step-wise procedure for learning a spectral graph from a time-series dataset.	49
3.3	Auto-encoder framework for detecting the anomaly of edge weights	50
3.4	Step-wise procedure of the Fault identification.	51
3.5	Step-wise Procedure for Fault Isolation.	52
3.6	A simple weighted graph.	54
3.7	Faulty Weighted Adjacency Matrix for Experimental dataset with yellow representing the faulty region and blue representing the normal operating region.	57
3.8	Loss vs Epochs curve for Experimental dataset.	59
3.9	Training reconstruction error with the threshold for Dataset 1 and 2.	60
3.10	Reconstruction error with fault detection when the faults are introduced at various junctures for Dataset 1.	61
3.11	Reconstruction error with fault detection when the faults are introduced at various junctures for Dataset 2.	62

4.1	A diagram illustrating the plan of the proposed work.	68
4.2	Representation of the latent space for mapping hidden features to variables.	74
4.3	VAE framework for detecting the anomalous node embeddings	75
4.4	A systematic sequence of steps outlining the methodology proposed for detecting faults.	76
4.5	Node Embeddings generated by Skip-gram and Matrix factorization Techniques for Npcil dataset for Fault Identification.	79
4.6	Relationship between number of training epochs and loss on NPCIL and TEP dataset.	80
4.7	Trained reconstructed embeddings for NPCIL and TEP dataset.	81
4.9	Node Embeddings generated by Skip-gram and Matrix factorization Techniques for TEP dataset for Fault Identification.	83
4.10	Node Embeddings generated by Skip-gram and Matrix factorization Techniques for TEP dataset for Fault Identification.	84
4.11	Reconstructed Embeddings with Fault Identified for Step (06), Random Variations (08), and Unknown types (18) faults.	85
5.1	Workflow of the proposed study.	90
5.2	Neighboring nodes of the i^{th} node within a window of size w	92
5.3	Proposed Graph Neural Differential Auto-Encoder (GNDAE) model with graph data as input, followed by a couple of GCN layers. The Graph Differential Equation (GDE) block provides continuous depth to the model. The green coloured box represents how the feature information from the nodes at different time steps are processed and transformed.	99
5.4	Sequential procedure for fault identification using the proposed algorithm. The green dotted wrapping visually represents the offline modelling phase, while the online monitoring phase is depicted in blue.	101
5.5	Traing and validation loss curves for both datasets using the proposed GNDAE approach.	103
5.6	Reconstruction profiles of the normal (Fault - 0) and faulty data (Faults 1-21) for the Tennessee Eastman Process dataset by the proposed GNDAE approach.	106
5.7	T^2 profiles for fault identification for TEP dataset with first 160 normal training samples. Blue-coloured points represent the faulty data points. The green colour line represents the onset of faults and red line indicates the fault-tolerant limit.	107
5.8	Identifying Water Data Faults by the proposed GNDAE approach, with Various Fault types introduced at Different Stages.	108

5.9	T^2 profiles for fault identification for WWTP dataset with first 10000 normal training samples along with the faulty data points represented by blue colour. The green line indicates the onset of faults and red line representing the fault tolerant limit.	109
6.1	Workflow of the proposed study.	115
6.2	A sample graph is constructed for a distillation column tray by taking into account its overall material balance, component balance, and heat balance.	117
6.3	Window-based graph extraction.	118
6.4	Proposed Graph Neural Differential Convolutional Network model with input graph along with the feature matrix as an input. The GDE block continuously parameterize the hidden layers and are then passed through a couple of GCN layers for transforming the features of the graph.	121
6.5	Various Graph Data Augmentation Techniques for learning the positive and negative pairs.	122
6.6	Overall procedure for Model Implementation and Change Point Detection.	124
6.7	Details regarding the Two-tank interacting system case study.	125
6.8	Details regarding the CSTR case study.	127
6.9	Details regarding the Carbon capture case study.	130
6.10	Comparison of Loss vs. Epochs for Training and Validation Data across Various Case Studies. The training loss is depicted in blue, while the validation loss is shown in red.	131
6.11	Identification of mean shift points and constant noise mean shift in for interacting tank system.	132
6.12	Identification of mean shift change points and constant noise mean shift by the proposed model.	132
6.13	The validation of the proposed model against the constant noise mean shift, Variance change, outliers with mean shift change points, and Random walk change points.	133
7.1	Workflow of the proposed study.	140
7.2	Two-stage Hydrocracker unit.	143
7.3	A sample graph is constructed for an energy balance equation, which is used for calculating the Temperature gradient along the length of the reactor.	145
7.4	A schematic representation of the proposed HCGNDE model.	150
7.5	Step-wise procedure of Genetic Algorithm implementation.	153
7.6	Two-stage Hydrocracker unit.	155
7.7	Loss vs Epochs Curve for the training of the GNDE model.	156

7.8	Comparison of hydrocracking pseudocomponent values predicted by the GNDE model with the simulated values.	158
7.9	Predicted and Simulated concentration values for various Carbon components across the catalyst beds.	159
7.10	Pareto plot for multi-objective optimisation. The three objectives are maximization of ATF and Kerosene with minimization of H_2 in the process stream. The operating points are indicated in blue colour, and projected points on the axes are indicated in green, while the robust operating point is indicated in black.	160
7.11	Plot of Decision Variables for Decision variables with respect to the weight fraction of H_2 for reactors 1 and 2.	161
8.1	Future directions for the proposed graph-based models in process systems engineering, highlighting potential applications, scope, and suggested improvements for the models developed in this thesis.	169
A.1	Reconstruction error with Fault Identification for Step Type Faults for Dataset 2.	202
A.2	Reconstruction error with fault identification for Random, Slow drift and Constant Position Type faults for Dataset 2.	203
A.3	Reconstruction error with fault identification for Random, Sticking and Unknown Type faults for Dataset 2.	204
B.1	The Tennessee Eastman Process: Detailing the Complex Interactions Amongst Reactants, Products, and Process Units	209
B.2	Squared Prediction Error (SPE) analysis for the Normal Training Dataset (Fault 0) and Faults 1-21 in test Datasets. The blue points represent the Normal data and red coloured points represent the faulty data. The threshold is depicted in black.	210
B.3	Wastewater Treatment Process (WWTP) Flowsheet.	212
B.4	Squared Prediction Error (SPE) analysis for the Normal Training Dataset and various types of faults in test Datasets. The blue points represent the Normal data, and the red coloured points represent the faulty data. The threshold is depicted in black.	212

List of Tables

3.1	Nomenclature	44
3.2	Example - Understanding Methodology	54
3.3	Faults considered for the proposed study.	58
3.4	Auto-encoder Network details	58
3.5	Averaged performance metrics compared with baselines for various faults .	63
3.6	Averaged performance metrics for NPCIL Dataset	63
3.7	Averaged performance metrics for Tennessee Eastman Dataset	63
4.1	Parameter Details for Various Methodologies for Case Study 1.	79
4.2	Parameter Details for Various Methodologies for Case Study 2.	82
4.3	Elaboration on the Variational Auto-encoder Network employed for both case studies.	82
4.4	Performance metrics compared against baselines for different types of faults.	86
4.5	Performance metrics for Case Study - 1	86
4.6	Averaged performance metrics for Case Study - 2	86
5.1	Table of Notations	94
5.2	Various types and Explanations of faults in the Nitrous Oxide Dataset of Wastewater Treatment Plants.	103
5.3	Different fault detection methodologies for the Tennessee Eastman Process dataset, evaluated based on their respective MDRs. The MDR values are represented in percentages. The better-performing methodology is highlighted in bold.	110
5.4	Different fault detection methodologies for the WWTP dataset, evaluated based on their respective MDRs. The MDR values are represented in percentages. The values are tabulated for different types of faults. The better-performing methodology is highlighted in bold.	110
5.5	Comparative analysis of the methodologies presented in Chapters 3, 4 and 5.	112
6.1	Effects of Parameter Changes on Two-Tank Interacting System	126
6.2	Effects of Parameter Changes on CSTR System	128

6.3	Key Parameters to Monitor in CCS System, Their Changes, and Effects on the Overall Process	129
6.4	Comparative Analysis of Various Change Point Detection Methodologies Across Multiple Case Studies. This analysis presents values calculated for different types of change points across assorted case studies. Superior performance values are highlighted in bold.	135
6.5	Performance metrics for Case Study 1	135
6.6	Performance metrics for Case Study 2	135
6.7	Performance metrics for Case Study 3	135
7.1	Comaparison of Simulated and Predicted values by the proposed model. . .	157
7.2	Parameters of the Implemented Genetic Algorithm model and their corresponding Specifications.	162
7.3	A comparative analysis of the yields for Diesel, ATF, and the percentage of H_2 required in the process is presented. All values are expressed as percentages.	162
A.1	Faults considered for the proposed study.	201
A.2	Performance Metrics for Tennessee Eastman Dataset for fault 1,	201
A.3	Performance Metrics for Tennessee Eastman Dataset for fault 5,	201
A.4	Performance Metrics for Tennessee Eastman Dataset for Fault 6	201
A.5	Performance Metrics for Tennessee Eastman Dataset for fault 7,	205
A.6	Performance Metrics for Tennessee Eastman Dataset for fault 8	205
A.7	Performance Metrics for Tennessee Eastman Dataset for fault 9	205
A.8	Performance Metrics for Tennessee Eastman Dataset for fault 13	205
A.9	Performance Metrics for Tennessee Eastman Dataset for fault 14	206
A.10	Performance Metrics for Tennessee Eastman Dataset for Fault 15	206
A.11	Performance Metrics for Tennessee Eastman Dataset for Fault 19	206
A.12	Performance Metrics for Tennessee Eastman Dataset for fault 20	206
A.13	Performance Metrics for Tennessee Eastman Dataset for fault 21	207
B.1	Various Fault Types and Their Descriptions in the TEP Dataset.	209
B.2	Different fault detection methodologies for the Tennessee Eastman Process dataset, evaluated based on their respective FARs. The FAR values are represented in percentages. The better-performing methodology is highlighted in bold.	211

B.3	Different fault detection methodologies for the WWTP dataset, evaluated based on their respective FARs. The FAR values are represented in percentages. The values are tabulated for different types of faults. The better-performing methodology is highlighted in bold.	213
C.1	Nomenclature	215

ABBREVIATIONS

CPD	Change Point Detection
ML	Machine Learning
DL	Deep Learning
GNDAE	Graph Neural Differential Auto-Encoder
GNDE	Graph Neural Differential Equation
SWGAE	Spectral Weighted Graph Auto-Encoders
GCN	Graph Convolutional Networks
PCA	Principal Component Analysis
DPCA	Dynamic Principal Component Analysis
VAE	Variational Auto-Encoder
LSTM	Long Short-Term Memory
GNDCL	Graph Neural Differential Contrastive Learner
ATF	Aviation Turbine Fuel
HCGNDE	Hydrocracking Graph Neural Differential Equations
ANN	Artificial Neural Networks
F1	F1 Score
SPE	Squared Prediction Error
CUSUM	Cumulative Sum Control Chart
GRL	Graph Representation Learning
GML	Graph Machine Learning
AE	Autoencoder
LSTM-AE	Long Short-Term Memory Autoencoder
GraRep	Graph Representation
HOPE	Higher Order Proximity Embedding
GNN	Graph Neural Network
GAT	Graph Attention Network
GraphSAGE	Graph SAmple and aggreGatE

MPNN	Message Passing Neural Network
GIN	Graph Isomorphism Network
SGCN	Signed Graph Convolutional Network
GAE	Graph Auto-encoder
VGAE	Variational Graph Auto-encoder
SDNE	Structural Deep Network Embedding
LINE	Large Scale Information Network Embedding
GPNN	Graph Partition Neural Networks
LGCL	Learnable Graph Convolutional Layer
NODE	Neural Ordinary Differential Equation
CTGNN	Continuous-Time Graph Neural Networks
NCDE	Neural Controlled Differential Equations
GNSDE	Graph Neural Stochastic Differential Equations
SPDE	Stochastic Partial Differential Equations

NOTATION

$\mathbf{X} \in \mathbb{R}^{n \times m}$	Dataset with n samples and m variables
$\mathbf{X}^T$	Transpose of matrix $\mathbf{X}$
$\mathbf{v}_i$	Eigenvector (principal component)
λ_i	Eigenvalue representing variance captured by principal component
SPE	Squared prediction error
$\mathbf{x}$	Original sample
$\hat{\mathbf{x}}$	Reconstructed sample
$\mathbf{z}$	Latent representation in autoencoder
$f_{\text{enc}}(\mathbf{x})$	Encoder mapping function
$f_{\text{dec}}(\mathbf{z})$	Decoder mapping function
$\mathcal{L}(\mathbf{x}, \hat{\mathbf{x}})$	Reconstruction loss function
$\mathbf{h}_t$	Hidden state in LSTM
$\mathbf{c}_t$	Cell state in LSTM
$\mathbf{f}_t$	Forget gate in LSTM
$\mathbf{i}_t$	Input gate in LSTM
$\mathbf{o}_t$	Output gate in LSTM
$\tilde{\mathbf{c}}_t$	Candidate state in LSTM
$\odot$	Element-wise multiplication
W_f, W_i, W_c, W_o	Weight matrices in LSTM
$\mathbf{b}_f, \mathbf{b}_i, \mathbf{b}_c, \mathbf{b}_o$	Bias terms in LSTM
Y	Low-dimensional embedding in Graph Laplacian Eigen Maps
W_{ij}	Weight of edge between nodes i and j
L	Graph Laplacian matrix
D	Degree matrix
λ	Eigenvalue in Laplacian Eigen Maps
P	Node proximity matrix
U, V	Matrices learned from factorization
A	Adjacency matrix of a graph
$\mathbf{Z}_k$	Embeddings derived from k -step proximity
S	Higher-order proximity matrix in HOPE
$\mathbb{R}^d$	Feature space of dimensionality d
$\mathcal{N}(i)$	Set of neighbors of node i
$h_i^{(l)}$	Feature vector of node i at layer l
$W^{(l)}$	Weight matrix at layer l
$\sigma(\cdot)$	Activation function
$H^{(l)}$	Feature matrix at layer l
$\tilde{A}$	Adjacency matrix with added self-loops
$\tilde{D}$	Degree matrix of $\tilde{A}$
α_{ij}	Attention coefficient in GAT

p, q	Parameters controlling biased random walks in Node2Vec
$P(c_i = x c_{i-1} = v)$	Transition probability in Node2Vec
d_{vx}	Shortest path distance between nodes v and x
$\hat{A}_{ij}$	Reconstructed adjacency matrix in Graph Autoencoder
Z	Latent embedding matrix in Graph Autoencoder
$\frac{dh(t)}{dt}$	Differential equation governing NODEs
$f(h(t), t, \theta)$	Neural network governing the dynamics in NODEs
$\frac{dh_i(t)}{dt}$	Differential equation governing GNDEs