



# **Data-Driven Modelling of Gas Turbine Engines**

Hamid Asgari

Master's Thesis  
Master of Engineering - Big Data Analytics  
2021

MASTER'S THESIS	
Arcada University of Applied Sciences	
Degree Programme:	Master of Engineering - Big Data Analytics
Identification number:	URN:NBN:fi:amk-2021060815126
Author:	Hamid Asgari
Title:	Data-Driven Modelling of Gas Turbine Engines
Supervisor (Arcada):	Amin Majd
Commissioned by:	
<p><b>Abstract:</b></p> <p>This study investigates and compares linear and nonlinear data-driven models of a gas turbine engine. These linear models consist of <i>Ridge</i>, <i>Lasso</i>, and <i>Multi-Task Elastic-Net</i> models, which are set up based on linear regressions. The nonlinear model explored in this study is a <i>recurrent neural network (RNN)</i> model. A comprehensive code is written in <i>Python</i> programming language and run in the <i>Jupyter Notebook</i>; an open-source web application for creating and sharing documents.</p> <p>The data employed for this study are open-source experimental time-series datasets of a single-shaft gas turbine. These datasets are composed of hourly average sensor measurements of eleven variables. However, in this research, only nine variables, which are directly related to the technical parameters of the gas turbine are considered, as the main objective is to model the system dynamics. Five of the features are considered as the system inputs and the four remaining features are used as the system outputs. The five datasets were collected over five years from 2011 to 2015, totally including 36732 records. Three of these datasets are employed for the training process and the remains are used for the validation purpose.</p> <p>The results of this study demonstrate that the prediction of the dynamic behavior of the gas turbine for <i>Ridge</i>, <i>Lasso</i>, and <i>Multi-Task Elastic-Net</i> models are almost similar and satisfactory for three out of four gas turbine output parameters. However, the <i>RNN</i> model predicts the system dynamics with higher accuracy compared to other models and that the results are satisfactory for all of the output parameters. The conclusion demonstrates that the nonlinear <i>RNN</i> model has a superior capability in capturing the complex nonlinear dynamics of the gas turbine compared to the three linear models.</p>	
Keywords:	Gas Turbine, Data-Driven, Black-Box, Modelling, Simulation, Recurrent Neural Network ( <i>RNN</i> ), Machine Learning, Linear, Nonlinear
Number of pages:	80
Language:	English
Date of acceptance:	2021

# CONTENTS

<b>1</b>	<b>INTRODUCTION .....</b>	<b>12</b>
1.1	Operation principles of gas turbines .....	13
1.2	Significance to the field .....	16
1.3	Research needs and questions .....	17
1.4	Statement of the problem .....	18
1.5	Purpose of the study.....	19
1.6	Thesis outline .....	19
1.7	Summary .....	20
<b>2</b>	<b>LITERATURE REVIEW .....</b>	<b>21</b>
2.1	Overview of data-driven modelling of gas turbine engines .....	21
2.2	The outcome of the literature review .....	30
2.3	Summary .....	31
<b>3</b>	<b>METHODOLOGY .....</b>	<b>32</b>
3.1	Data collection and preparation.....	32
3.1.1	<i>Data structure of ambient temperature (AT)</i> .....	35
3.1.2	<i>Data structure of ambient pressure (AP)</i> .....	36
3.1.3	<i>Data structure of ambient humidity (AH)</i> .....	37
3.1.4	<i>Data structure of turbine energy yield (TEY)</i> .....	38
3.1.5	<i>Data structure of air filter difference pressure (AFDP)</i> .....	39
3.1.6	<i>Data structure of gas turbine exhaust pressure (GTEP)</i> .....	40
3.1.7	<i>Data structure of gas turbine inlet temperature (TIT)</i> .....	41
3.1.8	<i>Data structure of gas turbine AFTER temperature (TAT)</i> .....	42
3.1.9	<i>Data structure of compressor discharge pressure (CDP)</i> .....	43
3.2	Training process and setting up the data-driven models .....	44
3.2.1	<i>Linear models</i> .....	44
3.2.1.1	<i>Ridge and Lasso regressions</i> .....	45
3.2.1.2	<i>Multi-Task Elastic-Net regression</i> .....	45
3.2.2	<i>Nonlinear model</i> .....	46
3.3	Selection of optimal data-driven models .....	47
3.4	The coefficient of determination ( $R^2$ ).....	47
3.5	Summary .....	48
<b>4</b>	<b>RESULTS.....</b>	<b>49</b>
4.1	Tuning alpha parameters .....	49
4.2	Training and Validation Results.....	49
4.3	Results for the Ridge model .....	51

4.3.1	<i>Ridge model: results for gas turbine exhaust pressure (GTEP)</i> .....	51
4.3.2	<i>Ridge model: results for turbine inlet temperature (TIT)</i> .....	52
4.3.3	<i>Ridge model: results for turbine after temperature (TAT)</i> .....	53
4.3.4	<i>Ridge model: results for compressor discharge pressure (CDP)</i> .....	54
4.4	Results for the Lasso model.....	55
4.4.1	<i>Lasso model: results for gas turbine exhaust pressure (GTEP)</i> .....	55
4.4.2	<i>Lasso model: results for turbine inlet temperature (TIT)</i> .....	56
4.4.3	<i>Lasso model: results for turbine after temperature (TAT)</i> .....	57
4.4.4	<i>Lasso model: results for compressor discharge pressure (CDP)</i> .....	58
4.5	Results for the Multi-Task Elastic-Net (MTEN) model.....	59
4.5.1	<i>MTEN model: results for gas turbine exhaust pressure (GTEP)</i> .....	59
4.5.2	<i>MTEN model: results for turbine inlet temperature (TIT)</i> .....	60
4.5.3	<i>MTEN model: results for turbine inlet temperature (TAT)</i> .....	61
4.5.4	<i>MTEN model: results for compressor discharge pressure (CDP)</i> .....	62
4.6	Results for the RNN model.....	63
4.6.1	<i>RNN model: results for gas turbine exhaust pressure (GTEP)</i> .....	63
4.6.2	<i>RNN model: results for turbine inlet temperature (TIT)</i> .....	64
4.6.3	<i>RNN model: results for turbine after temperature (TAT)</i> .....	65
4.6.4	<i>RNN model: results for compressor discharge pressure (CDP)</i> .....	66
4.7	Comparison of the models .....	67
4.8	Features' importance.....	68
4.9	Summary .....	69
<b>5</b>	<b>DISCUSSION .....</b>	<b>70</b>
5.1	Significant observations .....	70
5.2	Limitations .....	71
5.3	Recommendations for future research.....	72
5.4	Conclusion.....	73
	<b>REFERENCES .....</b>	<b>74</b>

## FIGURES

Figure 1. The schematic of a Single-Shaft Gas Turbine .....	14
Figure 2. Brayton Cycle in P-V and T-S Frames .....	14
Figure 3. Typical Single-Shaft Stationery and Aero-derivative Gas Turbine Engines .....	15
Figure 4. The schematic of the ANN Model and the Interpretation of the Outputs .....	24
Figure 5. Block Diagram of an ANN-Based Model of an ADGT .....	28
Figure 6. The Sensor Locations/Parameter Sources of a Single-Shaft Gas Turbine .....	33
Figure 7. Data Structure of Ambient Temperature (AT) for Training Process .....	35
Figure 8. Data Structure of Ambient Temperature (AT) for Validation Process .....	35
Figure 9. Data Structure of Ambient Pressure (AP) for Training Process .....	36
Figure 10. Data Structure of Ambient Pressure (AP) for Validation Process .....	36
Figure 11. Data Structure of Ambient Humidity (AH) for Training Process.....	37
Figure 12. Data Structure of Ambient Humidity (AH) for Validation Process.....	37
Figure 13. Data Structure of Turbine Energu Yield (TEY) for Training Process .....	38
Figure 14. Data Structure of Turbine Energu Yield (TEY) for Validation Process .....	38
Figure 15. Data Structure of Air Filter Difference Pressure (AFDP) for Training Process .....	39
Figure 16. Data Structure of Air Filter Difference Pressure (AFDP) for Validation Process .....	39
Figure 17. Data Structure of Gas Turbine Exhaust Pressure (GTEP) for Training Process.....	40
Figure 18. Data Structure of Gas Turbine Exhaust Pressure (GTEP) for Validation Process .....	40
Figure 19. Data Structure of Turbine Inlet Temperature (TIT) for Training Process .....	41
Figure 20. Data Structure of Turbine Inlet Temperature (TIT) for Validation Process .....	41
Figure 21. Data Structure of Turbine After Temperature (TAT) for Training Process.....	42
Figure 22. Data Structure of Turbine After Temperature (TAT) for Validation Process .....	42
Figure 23. Data Structure of Compressor Discharge Pressure (CDP) for Training Process .....	43
Figure 24. Data Structure of Compressor Discharge Pressure (CDP) for Validation Process .....	43
Figure 25. A Typical Recurrent Neural Network .....	46

Figure 26. Variations of $R^2$ for Ridge, Lasso, and MTEN Models with the Alpha Parameter for the Training and Test Datasets .....	50
Figure 27. A Comparison Between Outputs of GT Engine and Ridge Model for GT Exhaust Pressure (GTEP) for Training Datasets (Accuracy: 94%) .....	51
Figure 28. A Comparison Between Outputs of GT Engine and Ridge Model for GT Exhaust Pressure (GTEP) for Test Datasets (Accuracy: 96%) .....	51
Figure 29. A Comparison Between Outputs of GT Engine and Ridge Model for Turbine Inlet Temperature (TIT) for Training Datasets (Accuracy: 91%) .....	52
Figure 30. A Comparison Between Outputs of GT Engine and Ridge Model for Turbine Inlet Temperature (TIT) for Test Datasets (Accuracy: 89%) .....	52
Figure 31. A Comparison Between Outputs of GT Engine and Ridge Model for Turbine After Temperature (TAT) for Training Datasets (Accuracy: 55%).....	53
Figure 32. A Comparison Between Outputs of GT Engine and Ridge Model for Turbine After Temperature (TAT) for Test Datasets (Accuracy: 46%).....	53
Figure 33. A Comparison Between Outputs of GT Engine and Ridge Model for Compressor Discharge Pressure (CDP) for Training Datasets (Accuracy: 99%) .....	54
Figure 34. A Comparison Between Outputs of GT Engine and Ridge Model for Compressor Discharge Pressure (CDP) for Test Datasets (Accuracy: 99%) .....	54
Figure 35. A Comparison Between Outputs of GT Engine and Lasso Model for Gas Turbine Exhaust Pressure (GTEP) for Training Datasets (Accuracy: 94%).....	55
Figure 36. A Comparison Between Outputs of GT Engine and Lasso Model for Gas Turbine Exhaust Pressure (GTEP) for Test Datasets (Accuracy: 96%) .....	55
Figure 37. A Comparison Between Outputs of GT Engine and Lasso Model for Turbine Inlet Temperature (TIT) for Training Datasets (Accuracy: 91%) .....	56
Figure 38. A Comparison Between Outputs of GT Engine and Lasso Model for Turbine Inlet Temperature (TIT) for Test Datasets (Accuracy: 89%) .....	56
Figure 39. A Comparison Between Outputs of GT Engine and Lasso Model for Turbine After Temperature (TAT) for Training Datasets (Accuracy: 55%).....	57
Figure 40. A Comparison Between Outputs of GT Engine and Lasso Model for Turbine After Temperature (TAT) for Test Datasets (Accuracy: 46%).....	57

Figure 41. A Comparison Between Outputs of GT Engine and Lasso Model for Compressor Discharge Pressure (CDP) for Training Datasets (Accuracy: 99%) .....	58
Figure 42. A Comparison Between Outputs of GT Engine and Lasso Model for Compressor Discharge Pressure (CDP) for Test Datasets (Accuracy: 99%) .....	58
Figure 43. A Comparison Between Outputs of GT Engine and MTEN Model for GT Exhaust Pressure (GTEP) for Training Datasets (Accuracy: 94%) .....	59
Figure 44. A Comparison Between Outputs of GT Engine and MTEN Model for Gas Turbine Exhaust Pressure (GTEP) for Validation Datasets (Accuracy: 96%) .....	59
Figure 45. A Comparison Between Outputs of GT Engine and MTEN Model for Turbine Inlet Temperature (TIT) for Training Datasets (Accuracy: 91%) .....	60
Figure 46. A Comparison Between Outputs of GT Engine and MTEN Model for Turbine Inlet Temperature (TIT) for Validation Datasets (Accuracy: 89%) .....	60
Figure 47. A Comparison Between Outputs of GT Engine and MTEN Model for Turbine After Temperature (TAT) for Training Datasets (Accuracy: 54%).....	61
Figure 48. A Comparison Between Outputs of GT Engine and MTEN Model for Turbine After Temperature (TAT) for Validation Datasets (Accuracy: 45%).....	61
Figure 49. A Comparison Between Outputs of GT Engine and MTEN Model for Compressor Discharge Pressure (CDP) for Training Datasets (Accuracy: 99%) .....	62
Figure 50. A Comparison Between Outputs of GT Engine and MTEN Model for Compressor Discharge Pressure (CDP) for Validation Datasets (Accuracy: 99%) .....	62
Figure 51. A Comparison Between Outputs of GT Engine and RNN Model for GT Exhaust Pressure (GTEP) for Training Datasets (Accuracy: 98%) .....	63
Figure 52. A Comparison Between Outputs of GT Engine and RNN Model for GT Exhaust Pressure (GTEP) for Test Datasets (Accuracy: 97%) .....	63
Figure 53. A Comparison Between Outputs of GT Engine and RNN Model for Turbine Inlet Temperature (TIT) for Training Datasets (Accuracy: 97%) .....	64
Figure 54. A Comparison Between Outputs of GT Engine and RNN Model for Turbine Inlet Temperature (TIT) for Test Datasets (Accuracy: 94%) .....	64
Figure 55. A Comparison Between Outputs of GT Engine and RNN Model for Turbine After Temperature (TAT) for Training Datasets (Accuracy: 90%).....	65

<b>Figure 56.</b> A Comparison Between Outputs of GT Engine and RNN Model for Turbine After Temperature (TAT) for Test Datasets (Accuracy: 81%).....	65
<b>Figure 57.</b> A Comparison Between Outputs of GT Engine and RNN Model for Compressor Discharge Pressure (CDP) for Training Datasets (Accuracy: 99%) .....	66
<b>Figure 58.</b> A Comparison Between Outputs of GT Engine and RNN Model for Compressor Discharge Pressure (CDP) for Test Datasets (Accuracy: 99%) .....	66
<b>Figure 59.</b> The Comparison Chart for Accuracy Scores ( $R^2$ ) of Different Models for GTE Output Parameters .....	67
<b>Figure 60.</b> The Tree Model of Features's Importance for the Gas Turbine Engine .....	68

## TABLES

<b>Table 1.</b> Classification of Gas Turbine Engines Based on Their Capacities .....	16
<b>Table 2.</b> Gas Turbine Engine Variables .....	34
<b>Table 3.</b> Operational Time-Series Datasets .....	34
<b>Table 4.</b> Comparison of Four Different Models in Terms of Accuracy Scores ( $R^2$ ) for GTE Output Parameters .....	67



## ABBREVIATIONS

Adaptive Network-Based Fuzzy Inference System	<i>ANFIS</i>	Industrial Gas Turbine	<i>IGT</i>
Aero-Derivative Gas Turbine	<i>ADGT</i>	Inlet Guide Vane	<i>IGV</i>
Air Filter Difference Pressure	<i>AFDP</i>	Internal Combustion Engine	<i>ICE</i>
Ambient Humidity	<i>AH</i>	Lasso Regression	<i>LR</i>
Ambient Pressure	<i>AP</i>	Low-Pressure Turbine	<i>LPT</i>
Ambient Temperature	<i>AT</i>	Micro Gas Turbine	<i>MGT</i>
Artificial Intelligence	<i>AI</i>	Multi-Layer Perceptron	<i>MLP</i>
Artificial Neural Network	<i>ANN</i>	Multi-Task Elastic-Net	<i>MTEN</i>
Backpropagation Neural Networks	<i>BPNN</i>	Multiple-Input and Multiple-Output	<i>MIMO</i>
B-Spline Neural Network	<i>BSNN</i>	Nitrogen Oxide	<i>NO<sub>x</sub></i>
Carbon Monoxide	<i>CO</i>	Nonlinear Auto-Regressive with Moving Average Exogenous Inputs	<i>NARMAX</i>
Compressor Discharge Pressure	<i>CDP</i>	Nonlinear Auto-Regressive with Exogenous Inputs	<i>NARX</i>
Compressor Turbine	<i>CT</i>	Powe Turbine	<i>PT</i>
Cumulative Sum	<i>CUSUM</i>	Pressure-Volume	<i>PV</i>
Dry Low Emission	<i>DLE</i>	Radial Basis Function	<i>RBF</i>
Extreme Learning Machine	<i>ELM</i>	Recurrent Neural Network	<i>RNN</i>
Fault Detection and Isolation	<i>FDI</i>	Ridge Regression	<i>RR</i>
Feedforward Neural Network	<i>FFNN</i>	Shaft Dynamic-Based	<i>SD</i>
Fuel Control Unit	<i>FCU</i>	Single-Input and Single-Output	<i>SISO</i>
Gas Turbine	<i>GT</i>	Support Vector Machine	<i>SVM</i>
Gas Turbine Engine	<i>GTE</i>	Temperature-Entropy	<i>TS</i>
Gas Turbine Exhaust Pressure	<i>GTEP</i>	Turbine After Temperature	<i>TAT</i>
Genetic Algorithm	<i>GA</i>	Turbine Energy Yield	<i>TEY</i>
Heavy-Duty Gas Turbine	<i>HDGT</i>	Turbine Inlet Temperature	<i>TIT</i>
High-Pressure Turbine	<i>HPT</i>	Wavelet Neural Networks	<i>WNN</i>

## FOREWORD

This thesis presents the outcome of my research in the area of data-driven modelling of gas turbine engines; submitted as a partial fulfillment for the requirements of Master's Degree in Big Data Analytics at Arcada University of Applied Sciences. During this exciting and fruitful journey, I had the opportunity to expand my knowledge, and to develop my skills according to the cutting-edge research activities in the field of data analytics.

There is no doubt that without the help, and support of my teachers, colleagues, family members, and friends, it would not be possible for me to complete this journey. Among them, first of all, I would like to sincerely express my special appreciation to Dr. Magnus Westerlund; the Head of the Master's Programme, for providing such a fantastic opportunity for me, and also for his excellent lectures, help, and support during the study period at Arcada. I would also like to express my heartfelt thanks and gratitude to my Supervisor Dr. Amin Majd; Senior Researcher & Lecturer at Arcada, for his close collaboration and supervision, support, advice, and also for his very useful lectures during my studies. My sincere appreciation goes to Dr. Leonardo Espinosa Leal; Senior Researcher & Lecturer in Big Data Analytics at Arcada, who has greatly assisted me throughout the whole journey by giving great invaluable lectures and guidance.

I would also like to warmly thank Dr. Kaj-Mikael Björk (the Principal Lecturer in Analytics and Logistics at Arcada), Dr. Sandra Becker (the Lecturer of Visual Analytics course), Andrej Scherbakov-Parland (Senior Lecturer in Information Technology at Arcada), Juha Nyholm (the Software Specialist at CSC), Peter Li (Instructor and AI Scientist at Silo AI), Dusan Sovilj (Data Scientist at Arctic Wolf), Teppo Kuisma (VP of Business Development Department at Silo AI), Roberto Zicari (Professor of Database and Information Systems at Goethe University Frankfurt), Emmanuel Raj (Senior Machine Learning Engineer at TietoEVRy), and Dr. Thomas Forss (Founder and CEO at StageZero Technologies) for their precious efforts and invaluable lectures, presentations, and supports. I also take this opportunity to appreciate Dr. Jenna Kanerva, Dr. Samuel Rönqvist, and Dr. Filip Ginter; the members of the Natural Language Processing (NLP) Team at the University of Turku, who greatly helped me to develop my skills in this interesting area by giving excellent lectures and technical advice.

I also heartily thank the staff in Admission, Administration, Enrolment, Student Services, Library, ICT Services, and all other educational and technical sections in the Arcada University for their endless efforts and for being very supportive to me and my colleagues, especially in the current difficult situation due to COVID-19 pandemic. I am also sincerely grateful to my fellows, classmates, and friends, with I, had the pleasure of working and cooperating during my studies at Arcada.

Finally, I heartily thank my family members; my father, mother, brothers, and sisters for being supportive to me in ups and downs throughout my entire life, by giving endless love and inspiration. Without having them, I would not be able to stand in the place where I am today.

Best regards,

Hamid Asgari

# 1 INTRODUCTION

This chapter explains the research background and provides required information about the structure of this thesis. It also presents the problem, the purpose of the study, research needs and questions, the significance to the field, necessary definitions, advantages, and limitations.

Development of the first practical gas turbine (*GT*) as a jet aircraft engine was carried out by Frank Whittle in Britain in 1930s [1]. It was after World War II that applications of *GTs* significantly increased due to the enhancement in a variety of scientific fields such as material science, aerodynamics, and cooling systems, which in their turn led to remarkable improvements in engine efficiency. In the past 50 years, characteristics such as compactness, low weight, low initial cost, availability, adaptability, reliability, fast start capability, multiple fuel applications, and short delivery time have increased the usage of gas turbines in industrial platforms and offshore utilities [2, 3]. Nowadays, gas turbine engines (*GTE*) are widely used in oil field platforms, power plants, refineries, petrochemical plants, and gas stations for power generation. They also act as jet engines in aeronautical industries. Gas turbines are classified as internal combustion engines, capable of converting chemical energy (stored in fuel) to mechanical energy [4]. These engines are employed as drivers for large generators, compressors, and pumps.

Modelling of gas turbines may be used for the prediction of their dynamics, as well as for design and/or performance optimization. They may also be used for system identification, performance prediction, sensor validation, condition monitoring, fault detection, evaluation of emissions, or control system design [5]. *GT* models can be classified as white-box and black-box models [6]. In a white-box model, the system dynamics and mathematical equations are available for setting up the model using different methods, techniques, and software tools such as *MATHEMATICA* and *Simulink/MATLAB*. These equations for most industrial systems are time-dependents and usually have coupled and nonlinear nature [6]. They are needed to be simplified by making different assumptions to be possible to get solved. If access to the system dynamics and equations is impossible or limited, black-box (data-driven) models are employed. In this case, values of significant features of the system during operation in a specific period of time are

collected and divided into inputs and outputs of the system. Here again, a variety of methodologies and software may be used to map the inputs to the outputs to disclose the relationships among the system variables and to build up the desired model(s). One of the fast-growing and popular approaches in black-box modelling during the recent decades is using a subset of artificial intelligence (*AI*), which is called artificial neural network (*ANN*). *ANN* simply resembles the function of the human brain in solving complicated problems. It can be used in different scientific fields for a variety of applications such as data processing, system identification, modelling, condition monitoring, fault diagnostics, and control of highly nonlinear dynamic systems. In a dynamic system, the values of the system variables change with time, based on earlier applied signals, while in a static system, there are instantaneous direct links among the system variables [7]. Some researchers try to improve empirical models by employing the theoretical structure of the system. This approach is called the gray-box method. A gray-box model is a combination of physical and statistical models [8]. In this hybrid model, experimental data and mathematical modelling techniques are combined to improve the accuracy of the resulting model [6].

## 1.1 Operation principles of gas turbines

Figure 1 presents the main components of a single-shaft gas turbine including compressor, combustor, and turbine; all together so-called gas generator (*GG*). These components are installed on a single central shaft and rotate together. The operation of *GTEs* is based on the Brayton cycle; which is shown in Figure 2 in both temperature-entropy (*T-S*) and pressure-volume (*P-V*) frames [9]. As these figures illustrate, air enters the compressor at point 1 and passes through the compressor. The hot compressed air enters the combustion chamber (combustor) at point 2 and is mixed with fuel. The mixture is ignited to produce the hot gases. The produced gasses enter the turbine at point 3 and pass through it; forcing the turbine, and consequently, the connected shaft and compressor to rotate. The output of the shaft could be connected to a large generator, pump, or compressor to shape a turbo-generator, turbo-pump, or turbo-compressor respectively. Although the actual processes in the compressor (1-2) and turbine (3-4) are irreversible and non-isentropic, they are ideally assumed as isentropic processes. The ideal processes in the combustor (2-3) and environment (4-1) are also assumed to be isobaric, although there are considerable pressure losses in the combustor and air filters [10].

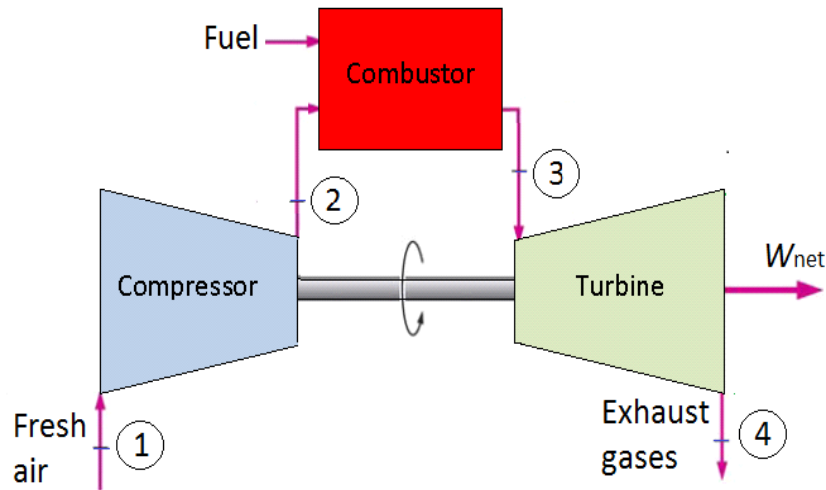


Figure 1. The schematic of a Single-Shaft Gas Turbine

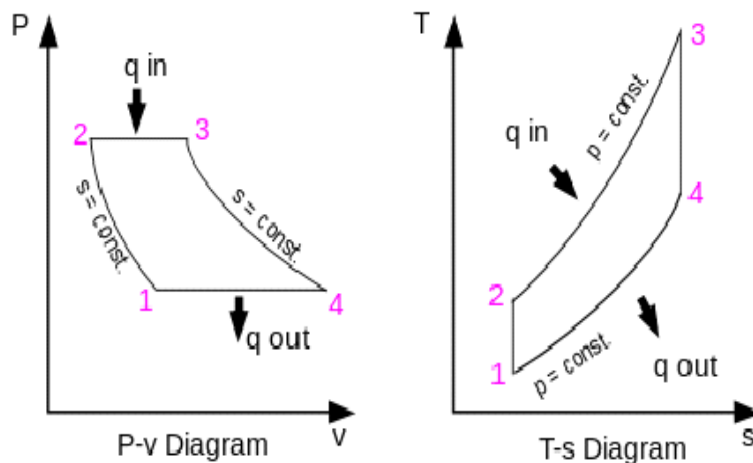
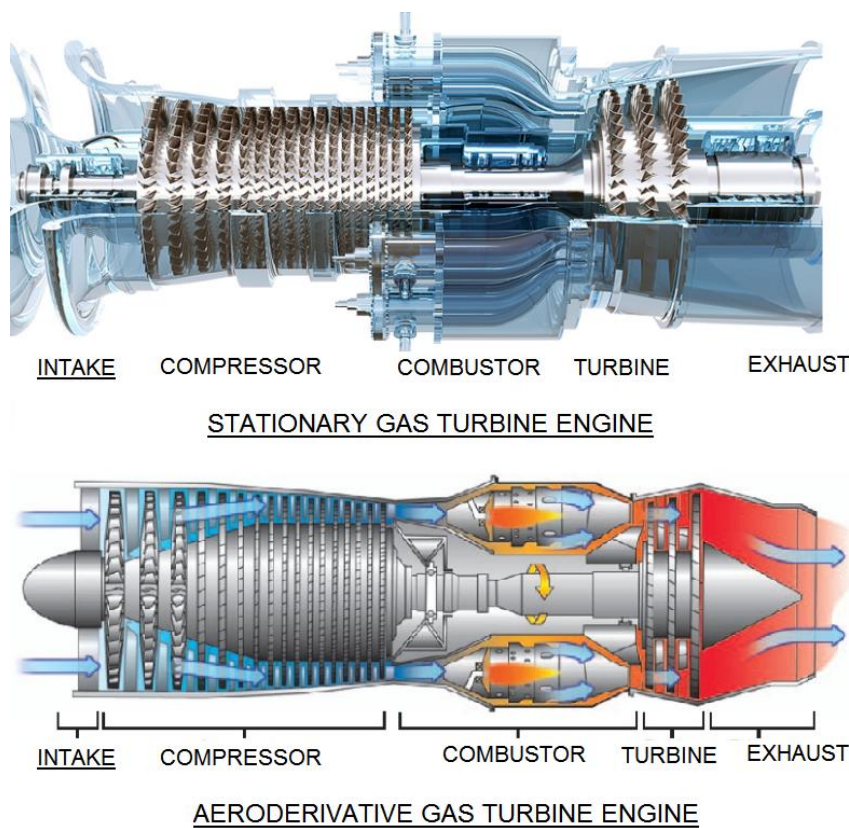


Figure 2. Brayton Cycle in P-V and T-S Frames [9]

Gas turbines can be classified differently based on the type, capacity, size, and application. There are two main types of *GTs* including stationary and aero-derivative gas turbines. The engine acts as the main driver for turbo-generators, turbo-pumps, or turbo-compressors in stationary gas turbines. A turbo-generator is used for producing electrical power. Aero-derivative *GTs* such as turbo-jets, turbo-props, and turbo-fans are employed as propulsion systems to produce the required force (thrust) based on Newton's third law (so-called action and reaction law), for moving an airplane through the air. Gas turbines may also be grouped as single-shaft and split-shaft (twin-shaft, triple-shaft). All main

components in a single-shaft *GT* are installed on a central shaft and rotate together. However, there could be several shafts and turbines in a split-shaft gas turbine. In a twin-shaft *GT*, the first turbine which is connected to the compressor may be called high-pressure turbine (*HPT*), compressor turbine (*CT*), or gas generator (*GG*) turbine. Compressor and its accessories employ the power produced by *CT* for rotation. The second turbine, named power turbine (*PT*) or low-pressure turbine (*LPT*), is mechanically disconnected from the gas generator. *PT* produces the output power of the engine. [Figure 3](#) shows and compares typical single-shaft stationary and aero-derivative gas turbine engines [\[11\]](#).



[Figure 3](#). Typical Single-Shaft Stationary and Aero-derivative Gas Turbine Engines [\[11\]](#)

Before setting up the *GT* models, some considerations such as type, size, capacity, configuration, and even manufacturer of the engine should be taken into account. Besides, the modelling techniques, strategies, and objectives should be carefully investigated at the beginning of the modelling process [\[12\]](#). According to [Table 1](#), gas turbine engines may also be categorized into five main groups based on their sizes and capacities (output power) [\[2\]](#).

Table 1. Classification of Gas Turbine Engines Based on Their Capacities

Type of GT Engine		Symbol	Capacity (MW)	Efficiency (%)	Application
1	Micro gas turbine	MGT	0.02 – 0.35	15 - 30	an alternative to internal combustion engines (ICE)
2	Small gas turbine	-	0.5 – 2.5	15 - 25	simple cycles
3	Aero-derivative gas turbine	ADGT	2.5 - 50	35 - 45	Aeronautical Industry
4	Heavy-duty gas turbine	HDGT	3 - 480	30 - 46	large power generation units
5	Industrial gas turbine	IGT	2.5 - 15	30 - 39	extensive use in petrochemical plants

## 1.2 Significance to the field

Identification, modelling, simulation, and control of gas turbine engines have been the subjects of many studies in recent decades. It is especially because of these engines' significant role and applications in aeronautical, power generation, marine propulsion, and oil & gas industries. Modelling of *GTs* has improved and developed technical and economical strategies for design, manufacturing, operation, performance optimization, and maintenance planning of these engines. The *GT* models have also shown great capabilities for condition monitoring and fault detection. So far, many efforts have been carried out to develop generations of gas turbines with the capability of overcoming available challenging engineering problems from both technical and economical perspectives [13, 14]. Offline modelling of gas turbines is a very good approach for the design, optimization, performance prediction, and diagnostics of *GTEs*. It can assist both users and manufacturers in different ways. Offline models may be employed by manufacturers to evaluate the performance of the engines during the design and manufacturing phases, while online models can be utilized on industrial sites for performance prediction, condition monitoring, fault diagnosis, sensor validation, or troubleshooting.



### 1.3 Research needs and questions

Although remarkable analytical and experimental models of *GTEs* have been investigated and developed so far to deeply understand the complexity of the nonlinear nature of these engines, there are still strong motivations for the students, scientists, manufacturers, and other professionals to continue their research activities in this fascinating area. Moreover, the electricity market has shown a high demand for the electricity produced by employing turbo-generators, especially for industrial applications. It has encouraged the power producers and other professionals in the field to look for novel methodologies for optimization of design and performance, and improving the manufacturing process of stationary gas turbines. It is especially because of the complexity of these systems and the need for developing models for a variety of objectives and applications with higher accuracies and reliabilities. It would also be of great importance to researchers to explore a variety of approaches, techniques, and methodologies in the field with their own benefits and limitations. Both transient and steady-state operations of gas turbines under different environmental conditions, load fluctuation, and system disturbances are still needed further investigations. Besides, a variety of linear and nonlinear models may be developed and compared for different types of the engines, under different operating conditions, and from different perspectives.

A linear model of a system is defined as a model, which is set up based on the totally linear equations of the system constraints and objective functions. If each of these equations is nonlinear, the model is considered a nonlinear model. Because of the complex nonlinear dynamics of many industrial types of equipment, solving nonlinear systems of equations for setting up a model is usually impossible without making some assumptions, and going through a simplification process to pave the way for linear analysis. Although there is a variety of techniques and methods for linearization of nonlinear systems, keeping the sensitive and effective nonlinear aspects of the system for setting up an accurate model should be critically discussed and considered. Otherwise, the resulting model may fail to predict the system dynamic with acceptable accuracy and reliability.

## 1.4 Statement of the problem

Even a quick and short survey in scientific sources shows that there have been remarkable research efforts regarding both physics-based and data-driven modelling of *GTEs*. However, it can be also observed that despite all those efforts, many problems still exist in the field during design, manufacturing, operation, and maintenance phases, which need extensive attention and effort for finding appropriate solutions. Some of these problems can be highlighted as follows:

- White-box models of gas turbines rely on dynamic, thermodynamic, and energy balance equations with a high degree of nonlinearity. To solve these complex and coupled equations, they should be simplified by making some assumptions and employing different linearization techniques, which consequently resulted in setting up models lacking the capability to precisely capture the whole dynamics of the system. Therefore, getting attention to the alternative solutions, which are independent of the system physics has been considerably increased. Among the different techniques, data-driven models have shown a high capability in capturing the nonlinear dynamics of *GTEs*.
- Industrial equipment gradually deteriorates; losing efficiency and operability. Therefore, after years of service in the industry, the original dynamic and thermodynamic equations and the corresponding *GT* models need to be revised through theoretical and experimental processes, which are not easy in most cases. Besides, replacement of the old engines requires huge financial sources and may not be economically possible. Fortunately, data-driven models can act as good alternatives in such situations, because of the flexibility to adapt to new conditions.
- Although setting up linear models are much easier and require a shorter period of time compared to nonlinear models, it is not as accurate and reliable as nonlinear models. However, further investigations are still needed to evaluate and compare a variety of linear and nonlinear models of gas turbines. Data-driven models, such as *RNN*, have shown high capabilities in capturing complex nonlinear dynamics of industrial equipment, and may also be utilized to optimize the processes of design, manufacturing, operation and maintenance of gas turbine engines eventually led to saving a remarkable amount of money and time.

## 1.5 Purpose of the study

Considering the concepts already discussed in this chapter, the major objective of the study can be outlined as follows:

- Novel data-driven models of an industrial gas turbine engine are investigated and the results are compared for different models in terms of accuracy and reliability.
- Three linear models including *Ridge Regression (RR)*, and *Lasso Regression (LR)*, and *Multi-Task Elastic-Net (MTEN)* are trained and verified by using five extensive operational time-series datasets for normal operations of the engine.
- A nonlinear model of the system is set up and validated by employing recurrent neural networks (*RNN*). A combination of different training and transfer functions, different number of neurons and different time-delays for the recurrent connections are employed to explore the best solution.
- To evaluate the capability of the resulting models in capturing the system dynamics, outputs of the resulting models are compared to the system outputs (targets) including values of the corresponding operational data.
- To train and validate the models, a comprehensive programming code is written in the *Jupyter Notebook* environment by using *Python* programming language and the relevant libraries and modules such as *pyrenn*.
- It is shown that the resulting *RNN* model can be applied reliably for performance prediction of the engine by following changes in the system inputs.

## 1.6 Thesis outline

This thesis investigates novel linear and nonlinear data-driven modelling of *GTEs*, according to the following outline:

[Chapter 1 \(Introduction\)](#) is presenting an introduction to the subject of this study. Operation principles of gas turbines were already explained and the significance of the study to the field was described. In the next steps, research needs and questions, the statement of the problem and the purpose of the study were briefly presented and discussed.

**Chapter 2 (Literature Review)** presents an overview of the literature in the area of data-driven (black-box) modelling and simulation of gas turbine engines. The most relevant investigations for different stationary and aero-derivative gas turbine engine are briefly reviewed in this chapter.

**Chapter 3 (Methodology)** elucidates the methodology of this research. In this chapter, the process of data acquisition and preparation is explained and the structure of datasets for significant variables of the engine is presented. Then, the training process and setting up the data-driven models including *RR*, *LR*, *MTEN* and *RNN* models are fully described. At the end of the chapter, the coefficient of determination criteria for selecting the optimal model is presented and a summary of the chapter is provided.

**Chapter 4 (Results)** illustrates the results of this research for the final data-driven models of the engine investigated in this study. The results for the linear and nonlinear techniques including *RR*, *LR*, *MTEN*, and *RNN* are figured for four output parameters of the system consisting of gas turbine exhaust pressure (*GTEP*), turbine inlet temperature (*TIT*), turbine after temperature (*TAT*), and compressor discharge pressure (*CDP*). Then, the R-squared ( $R^2$ ) scores are calculated for the training and validation processes, and the resulting models are evaluated and compared for the selection of the optimal model. Finally, the features of the gas turbine are assessed in terms of the role and importance in the modelling process and a summary of the chapter is provided.

**Chapter 5 (Discussion)** presents the overall results of this study, and provides discussions and concluding statements. Significant observations, limitations, and recommendations for the upcoming research activities and possible improvements in the field of modelling, and simulation of gas turbine engines are also briefly discussed in this chapter.

## **1.7 Summary**

This chapter explained the background and main motives for this study. It also presented the operation principles of gas turbine engines and a short description of the significance of the study to the field. Besides, research needs and questions, the statement of the problem, the purpose of the study, and the thesis outline were briefly discussed.

## 2 LITERATURE REVIEW

One of the best strategies to manufacture gas turbine engines with higher efficiency, durability, and reliability is to employ modelling and simulation techniques. These techniques could also be employed for performance optimization, condition monitoring, fault detection, sensor validation, and maintenance planning. The benefits of using modelling methods have been created a strong motivation for professionals to keep studying and working in this area. A survey in the literature shows that many sources are available in the area of data-driven modelling and simulation of gas turbines, and significant research efforts have been carried out in the field. This chapter presents an overview, discussion, and concluding remarks regarding the most important parts of these efforts in recent decades.

### 2.1 Overview of data-driven modelling of gas turbine engines

So far, a variety of data-driven (black box) models of gas turbine engines have been developed for different types of *GTs* from different perspectives and for different purposes [15, 16, 17]. This variety indicates that building up a generic model of gas turbines, as some researchers like Visser et al. [18] desired and tried to do, may not cover all details of the dynamics of a specific engine. That is the main reason for setting up a variety of models based on different methodologies and approaches. Data-driven models have shown a high capability to predict the dynamic behavior of industrial equipment when access to the system dynamics is restricted or limited.

To disclose the nonlinear dynamics of *GT* engines, a variety of black-box models have already been investigated and developed. These studies cover a wide range of approaches including (but not limited to) multi-layer perceptron (*MLP*), feedforward neural network (*FFNN*), B-spline neural networks (*BSNN*), backpropagation neural networks (*BPNN*), nonlinear auto-regressive with exogenous inputs (*NARX*), nonlinear auto-regressive moving average with exogenous inputs (*NARMAX*), Elman neural network (*ENN*), and radial basis function (*RBF*).

Ibrahim et al. [16] developed two data-driven static and dynamic (*NARX*) models of a triple-shaft *ADGT* engine. They investigated different configurations of activation

functions, training algorithms, and the number of neurons. They concluded that although the accuracies of both models were acceptable, the performance of the static model was slightly better than the dynamic one. However, the *NARX* model showed a higher capability in generalization and capturing the system dynamics with a longer time for the training process because of the feedback connections.

Nikpey et al. [19] investigated an *ANN*-based model of an *MGT* located in a combined cycle power plant for condition monitoring purposes using the modified *MGT* data. They performed a four-step sensitivity analysis to explore the relevance of the system inputs and outputs, and to evaluate the influence of input parameters on the prediction accuracy of outputs. According to the results of sensitivity analysis, three variables including power, compressor inlet temperature, and compressor inlet pressure were the most effective system inputs. The results also demonstrated that the model accuracy was significantly affected by the compressor inlet temperature and pressure. The resulting *ANN*-based model could successfully predict the engine performance with high accuracy and reliability.

Vatani et al. [20] presented two *ANN*-based methods for health monitoring of a gas turbine in order to predict the system degradation under the effect of *GT* measurable parameters such as critical temperatures. They employed measured datasets, defined lower and upper threshold bounds, and used different scenarios to set up the *RNN* and *NARX* models, and to demonstrate the capability of the models in the prediction of the engine's dynamics. The results demonstrated that the required data for training the *NARX* model was less than the *RNN* network. Besides, the *NARX* model had fewer prediction errors compared to the *RNN* architecture.

Tarik et al. [21] developed a *NARX* model for identification of an *HDGT* in dry low emission (*DLE*) mode in order to reduce nitrogen oxide (*NO<sub>x</sub>*) emission. They used operational datasets to train, and validate the model. Nine variables were selected for the modelling process. Power demand, thermocouple temperature setpoint, and the opening of the metering valve, pilot valve, and guide vane were considered as five inputs of the model, while the compressor pressure, gas fuel flow, thermocouple average temperature, and output power were assigned as four model outputs. According to the results, the

*NARX* model could predict the system dynamics in *DLE* mode with a good generalization and high accuracy.

Bartolini et al. [22] explored applications of an adaptive network-based fuzzy inference system (*ANFIS*) and *ANN* to micro gas turbines. The aim of the study was to disclose unavailable experimental data for completing the engine performance diagrams. They showed that the engine performance could reliably be predicted by the *ANN* model. They also studied the effects of load changes and variations of ambient temperature, pressure, and humidity on the output power, and realized that the engine was more affected by load than ambient conditions. The results also demonstrated that the output power was more influenced by the variations of ambient temperature than pressure and humidity changes.

To explore the impacts of *MGTs* on the stability of distribution systems, a *NARX* model of an *MGT* and its relevant distribution system dynamics was developed by Jurado [23]. Testing the model under electrical disturbances and different operating conditions indicates that the resulting *NARX* model was capable of modelling the micro gas turbine dynamics in the non-isolated mode for high and low amplitude dynamics of the system.

Lazzaretto and Toffolo [24] applied *ANN* to a single-shaft *GTE* for off-design modelling purposes. To predict the engine performance, they employed *FFNN* and the analytical method. They also studied the application of scaling techniques to the generalized maps of the turbine and compressor in order to set up new maps for the engine. They validated the resulting maps by utilizing the operational datasets obtained from the real plants. The system of equations of the model was solved by using a commercial simulator. However, the researchers did not include the effect of changes in internal parameters in the analytical model. The simulator results were utilized to train the *FFNN*. The resulting model showed a high level of accuracy and reliability for capturing the correlations among significant thermodynamic parameters of the engine and prediction of the system dynamics.

Amozegar et al. [25] developed a novel methodology by employing ensembles of dynamic neural networks for system identification, fault detection and isolation (*FDI*), and health monitoring of a single-spool engine. Three different ensembles and a group of neural networks including *MLP*, *NARX*, and *RBF* and a dynamic support vector machine

(SVM) were utilized to identify the system dynamics. Five structures for each of the *RBF-NARX*, *MLP-NARX* and *SVM-NARX* models were trained individually to estimate the values of five outputs of the system. Among these individual learning models, the best performance belonged to the *RBF-NARX* model.

Ogaji et al. [26] investigated fault detection of a twin-shaft stationary *GT* by using simulated data of a pre-developed nonlinear aero-thermodynamic model and employing three different *ANN* architectures respectively for partitioning the system measurements into no-faults and faults groups, classifying the faults into a component or a sensor fault, and isolating dual or single faulty sensors, in order to determine the values of the faults by measuring the differences between the corresponding inputs and outputs of the network. The results demonstrated the capability of *ANN* as a high-speed tool for real-time control problems.

Arriagada et al. [27] explored an *ANN*-based model of a single-shaft *GTE* for fault diagnosis. Ten faulty datasets and one healthy dataset of the system were used to train a reliable feedforward *MLP* neural network for identification of the faults, and generation of early-stage warnings. Figure 4 illustrates a schematic of the *ANN*, named 14-H-28 according to its structure, with fourteen measured *GT* parameters as inputs [27]. The desired outputs were unique combinations of twenty-eight binary numbers, arranged in a graphical display.

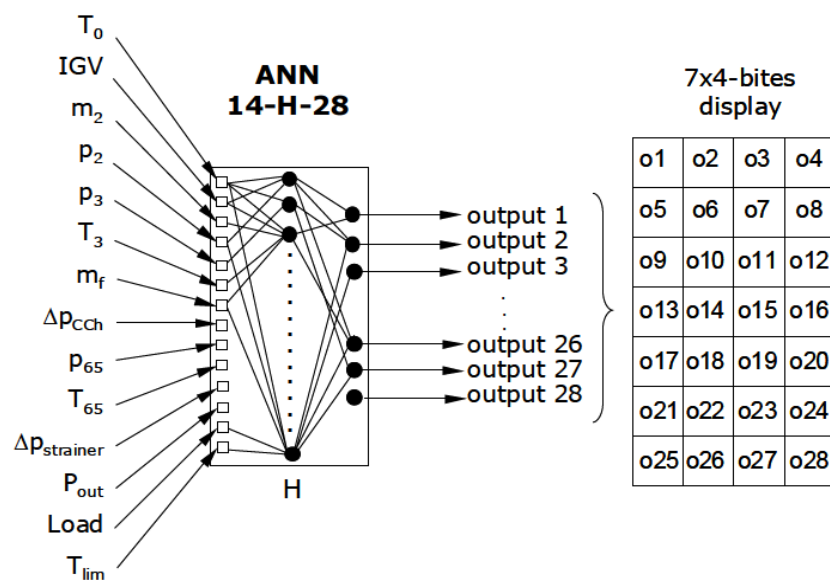


Figure 4. The schematic of the *ANN* Model and the Interpretation of the Outputs [27]



Mehrpanahi et al. [28] presented a dynamic model of a single-shaft *GTE* in start-up and loading modes by using operational datasets. Different methodologies and scenarios were employed to identify the system dynamics, and especially the shaft speed characteristics during the start-up operation. These methods included using *ANNs*, shaft dynamic-based (*SD*) function, linear regression, *NARX* networks, and time-delayed transfer functions. The position of inlet guide vanes (*IGV*) and fuel discharge rate were considered as inputs of the system, while the shaft speed was assigned as the engine's output. The results showed the relative superiority of this methodology in the estimation of the shaft speed in the start-up mode.

Bai et al. [29] developed a novel method for *GTs* anomaly detection by using normal datasets, and employing the *NARX* model and prior knowledge fusion. They also carried out an autocorrelation analysis to depict the engine's dynamic behavior and claimed that it was the first time that normal pattern extraction of gas turbines had been studied. They investigated the healthy engine under different operational conditions and extracted the unchanged features in order to perform robust and sensitive anomaly detection. The results showed that the *NARX* model had superiorities in recognition of normal patterns, and identification of *GT* dynamics compared to other machine learning-based methodologies such as *MLP*, *ENN*, and extreme learning machine (*ELM*).

To identify the dynamics of a power plant *HDGT*, *NARX* models of the engine were explored by Basso et al. [30]. They investigated isolated and non-isolated operational modes and employed black-box identification techniques in order to set up accurate reduced-order nonlinear models. They also used the *Gram-Schmidt* procedure to estimate the parameter of the *NARX* models. To establish the best structure for the nonlinear models, both step-wise and forward regressions were studied, and different input signals were examined.

To set up a neural network (*NN*) model of a single-shaft gas turbine with appropriate robustness, accuracy, and computational time, Bettocchi et al. [31] developed a feedforward *MLP* model, by using the datasets covering the whole operational range, generated by a cycle program already calibrated on the system. The resulting model included fifteen inputs, six outputs, and one hidden layer with sixty neurons. They concluded that the resulting model was capable of the real-time simulation of *GT* engines.

In another study, Bettocchi et al. [32] examined a multiple-input and multiple-output (*MIMO*) neural network for fault diagnosis of the engine. Simani and Patton [33] employed a data-driven model for output estimation, and fault detection and isolation (*FDI*) of a single-shaft *GT* prototype. They used an identified linear model to avoid the complexity of nonlinear dynamics of the engine, and claimed that the presented *FDI* strategy could provide robustly solutions for minimization of the impacts of modelling noise and errors, in response to maximization of fault sensitivity. For verification of the robustness of the proposed approach, the researchers applied the proposed approach to the *GT* simulated datasets obtained in presence of modelling and measurement errors.

Chiras et al. [34] investigated a *NARMAX* model of an *ADGT* engine to determine the nonlinearity nature and order of the system. For this purpose, they employed identification techniques and nonparametric analysis in frequency and time domains. In another study, Chiras et al. [35] utilized a forward-regression orthogonal estimation algorithm to set up a *NARMAX* model for an aero-derivative twin-shaft *GT*. They also studied the nonlinear relation between the shaft rotational speed and the fuel flow rate. To validate the model performance, they examined the system performance for small and large signal tests and observed that the results were satisfactorily close to the results of the model previously developed for the engine. Chiras et al. [36] also employed *FFDN* to model the relationship between fuel rate and shaft rotational speed of an *ADGT*. They realized that using a nonlinear model was necessary for modelling the high-amplitude dynamics of *GTEs*. Chiras et al. [37] presented a global *NARMAX* model of a twin-shaft aero gas turbine. They explored both linear and nonlinear models of the engine and proposed a simple *NARX*-based system identification method with satisfactory performance for small and high amplitude tests.

Condition-based maintenance of *GT* engines was explored by Fast et al. [38], by applying *ANN* techniques to simulation data. Fast et al. [39] also set up a simple *ANN*-based model of a single-shaft gas turbine with very high prediction accuracy, by employing the measured data obtained from the engine when working under full load. Besides, they combined the cumulative sum (*CUSUM*) technique with *ANN* to monitor the system conditions, and to detect the engine performance anomalies [40]. In another study, Fast et al. [41] investigated application of *ANNs* to diagnosis and condition monitoring of a

combined cycle power plant. Fast [42] employed *ANN*-based approaches for sensor validation, diagnosis, and condition monitoring of *GT* engines.

Torella et al. [43] simulated the effects of an air system on the performance of a large turbo-fan engine by using two different configurations of *BPNN*. To improve the model accuracy for the first configuration, a variety of training algorithms and the different number of hidden layers were examined. A computer programming code was derived for the second configuration in order to build up the *BPNN* for modelling the *GT* air system in the presence and also in the absence of faults. Although the proposed approach could be applied to the air system for diagnostics and troubleshooting purposes, multiple faults and the effect of sensor noise could not be covered by the *BPNN* model.

*NARX* models of the fuel control unit (*FCU*) of a turbo-shaft *GTE* was investigated by Salehi et al. [44]. They used *NARX* models because of the nonlinear nature of the *FCU* and not having enough information about the system dynamics and equations. The structure of the *NARX* model included both linear (feedforward) and nonlinear (*NARX*) parts. The inputs of the nonlinear block were the system input/output regressors. The resulting model showed satisfactory accuracy to represent the real system in a wide operating range.

*MLP* models for an *IGT* were explored by Nozari et al. [45]. They applied, tested, and validated a nonlinear method for system identification, in order to detect and isolate the engine faults. Besides, a comparative study was performed by using other relevant works to show the benefits of the applied method. In another effort, Nozari et al. [46] applied linear *neuro-fuzzy* and *MLP* methods for *FDI* purposes.

Ruano et al. [47] applied *NARX* and *ANN*-based models to an aircraft engine for nonlinear identification of the shaft speed dynamics under normal operation. Among the different *ANN*-based structures including *MLP*, *RBF*, and *BSNN*, the latter one showed the best performance. The researchers also used a genetic programming tool to determine the structures of *BSNN* and *NARMAX* models.

An *ANN*-based adaptive technique was developed for system optimization by Dodd and Martin [48] for modelling and control of an *ADGT* engine. An *FFNN* network was

employed for the system modelling. The resulting model could maximize thrust at a constant desired level of fuel flow and could decrease the turbine blades' temperature, which consequently could increase the *GT* life. Figure 5 illustrates the block diagram of the *ANN*-based model with its inputs and output [48].



Figure 5. Block Diagram of an *ANN*-Based Model of an *ADGT* [48].

A genetic algorithm (*GA*) approach was applied by Breikin et al. [49] for an *ADGT* dynamic modelling during the engine cruise operation. They used operational data for parameters estimation of the reduced-order model. Comparing the results of *GA* to the traditional techniques demonstrated that *GA* provided good flexibility in choosing the *GT* performance metrics.

Rahmoune et al. [50] investigated fault detection of a twin-shaft engine by employing a *NARX* model. The real data obtained from the vibrations of bearings of the engine were employed to train and validate the resulting model. It was shown that *NARX* models could effectively be used for fault detection and optimal condition monitoring of gas turbines. Rahmoune et al. [51], and El Ashmawi et al. [52, 53] explored novel *ANN*-based models for monitoring and fault diagnosis of an *IGT* engine.

A feedforward *MLP* neural network was employed by Badihi et al. [54] for estimation of the fuel flow rate of a jet engine. *Simulink-MATLAB* was employed to set up a mathematical model for the generation of the required data for the network training. The resulting model showed a high capability in the prediction of the engine performance parameters.

Yoru et al. [55] applied *ANN* to exergetic analysis of the *GTEs* providing power and heat in a power plant cogeneration system. They compared the values obtained from the

exergetic analysis with the model outlet values, and showed that much closer exergetic results could be achieved by employing the *ANN*.

Loboda et al. [56] developed *MLP* and *RBF* models for fault detection of an *ADGT* engine. According to the results, the computational time for training the *RBF* was much more than *MLP*, and its accuracy was a little more than the accuracy of *MLP*.

Yu et al. [57] proposed a new methodology for modelling aircraft engines by using *NARX* models and wavelet neural networks (*WNN*). They employed time-series datasets obtained from three aircraft engines to set up and validate the models.

A set of single-input and single-output (*SISO*) neural networks were applied to an aircraft engine by Tayarani-Bathaie et al. [58] for fault detection purposes. They performed different simulations to evaluate the accuracy of the proposed approach.

In addition to the above-mentioned efforts, one can also refer to other studies focused on *ANN*-based *GT* models, which indicate the strength of *ANN* in the identification of complex dynamics of gas turbine engines. For instance, Asgari et al. [59, 60, 61, 62, 63] investigated data-driven modelling, simulation, and control of gas turbines using different methodologies. Mu and Rees [64] developed a nonlinear model or control of an aircraft engine. They employed neural networks and *NARMAX* for the identification of the system dynamics under a variety of operating conditions. A model of an *HDGT* for the prediction of transient behavior during the startup process was presented by Kim et al. [65, 66]. Tavakoli et al. [67] extracted the parameters of an *HDGT* by providing an educational modelling approach. To minimize sensor failures and their consequences including shutdowns, and to decrease the necessary number of sensor calibrations, Palmé et al. [68] presented an *ANN*-based technique for sensor validation of *GTEs*. Spina and Venturini [69] applied *ANN* to model a single-shaft *GTE* and its diagnostic system. They used measured datasets and employed different patterns to train the network with a low computational time. And finally, Mohammadi et al. [70] applied *MLP* with dynamic processing units to a twin-shaft *ADGT* for fault detection. The capability of the resulting model was verified by conducting varieties of simulations.

## 2.2 The outcome of the literature review

As the literature survey indicates, remarkable studies have been done so far in the area of data-driven modelling of gas turbine engines, each with its own advantages and limitations. The outcome of these activities has had significant impacts on optimization and cost-cuts of design and manufacturing processes, and improvements in the condition monitoring, operation, fault diagnosis, and maintenance planning of gas turbines. However, to approach the optimal models as closely as possible, and to design and manufacture *GTEs* with high reliability, good performance, and cost-effectiveness, further efforts still need to be done in order to unfold the complex dynamics of these systems. Because of the nonlinear nature of *GTE* dynamics, and their different sizes, capacities, and applications, study in this area is still a challenging issue. Some of the subjects that indicate the necessity of further research in the area of data-driven modelling of *GTEs* already explained in the previous chapter. The following issues may also be added to those subjects:

- *GT* models, which are usually set up by using simplified dynamics and linearized equations, may not be accurate enough to precisely capture the complex dynamics of nonlinear systems. Hence, any simplification or linearization methodology, and the related effects and consequences on the system dynamics should be precisely taken into account and analyzed at the beginning of the modelling process [71, 72]. However, data-driven models have the potential and power to satisfactorily capture the complex nonlinear dynamics of *GTEs* because of their independence of the physics of the system.
- There are a variety of *ANN* architectures in the field of data-driven modelling of gas turbines. This variety is based on the network topology, the number and type of inputs and outputs, data flow, training algorithms, activation functions, the number of layers and neurons, the number of feedback connections, and the related time-step delays, etc.
- Building up new *GT* models by employing novel methodologies could greatly help in a further and deeper understanding of the *GT* dynamics. For instance, a recurrent neural network, as a nonlinear model, could effectively capture the nonlinear aspects of *GT* dynamics with high accuracy. Besides, such models could

save money and time during design, manufacturing, operation, and maintenance periods.

- The literature lacks enough sources regarding comparisons of linear and nonlinear models in terms of deviations from the real systems. It could be of great significance and also interesting to see, understand and evaluate the application, benefits and limitations of different linear and nonlinear methods in the area of *GT* modelling.
- *GTEs* are complicated systems with high degrees of nonlinearity and with complex, nonlinear, and coupled dynamic equations. Modelling and simulation of these systems cover a wide range of studies and research activities. Even in the area of data-driven modelling, a variety of modelling approaches, and techniques with static, dynamic, linear, or nonlinear structures may be employed for different types of *GTEs*. Among these methodologies, some may need further investigations and developments, and some may even have not been explored so far. Hence, research in data-driven modelling still needs to be continued for different types of gas turbine engines.

## 2.3 Summary

This chapter briefly reviewed and discussed the data-driven models of gas turbine engines including significant and the most relevant studies and investigations for different types of gas turbines. According to the outcome of the literature survey, although many remarkable studies have been done so far in this area, further developments are still required in order to resolve the current challenges in design, manufacturing, commissioning, operation, maintenance planning, fault diagnosis, sensor validation, troubleshooting, control, etc. This study is to provide solutions to some of the current problems in this area.

### 3 METHODOLOGY

The application of suitable procedures and techniques for modelling industrial systems plays a significant role in the reliable and accurate prediction of the dynamic behavior of these systems. Coupled and nonlinear dynamic and thermodynamic equations have made the modelling of gas turbines very complicated. Despite many efforts in the field, it is still a challenging issue to set up novel models of gas turbines by considering the effect of all variables and by making a trade-off correlation among generalization, accuracy, reliability, and robustness. Besides, there are varieties of gas turbines produced by different manufactures in terms of type, size, structure, application, and capacity, which have led to diversity in modelling methods.

In general, there are two main approaches in modelling industrial systems including white-box and black-box (data-driven) models. A white-box model is usually employed when the mathematical equations of the system are available. In this case, the complex equations should be simplified enough to be solved and modelled, which eventually decreases the accuracy of the resulting model(s). Researchers may need to use different linearization techniques during the modelling process and to employ engineering software as powerful simulation and modelling tools.

A black-box model is usually employed when access to the mathematical equations or in general, the physics of the system is impossible or limited. In this case, the system data including values of different variables are collected and used to unfold the hidden relationships among the system variables. Data-driven modelling is a very good alternative to white-box modelling and has been investigated in a variety of scientific fields. It consists of different approaches such as supervised learning, unsupervised learning, and reinforcement learning. In this study, different supervised data-driven approaches for a gas turbine engine will be employed and the resulting models will be compared in terms of their accuracies and reliabilities.

#### 3.1 Data collection and preparation

The data employed for this study are open-source datasets collected from a single-shaft *GTE* located in Turkey. A model of this engine was already employed for the prediction



of carbon monoxide ( $CO$ ), and nitrogen oxide ( $NO_x$ ) emissions from the engine [73]. These datasets include the hourly average values of eleven variables collected from the related  $GT$  sensors. However, in this research, nine out of eleven variables, which are directly related to the technical parameters of the  $GTE$  are considered, as the purpose of this study is to set up dynamic models of the system. Figure 6 shows the  $GTE$  and the sensor locations for picking up the values of the system parameters. These parameters are shown in dashed red rectangles [73].

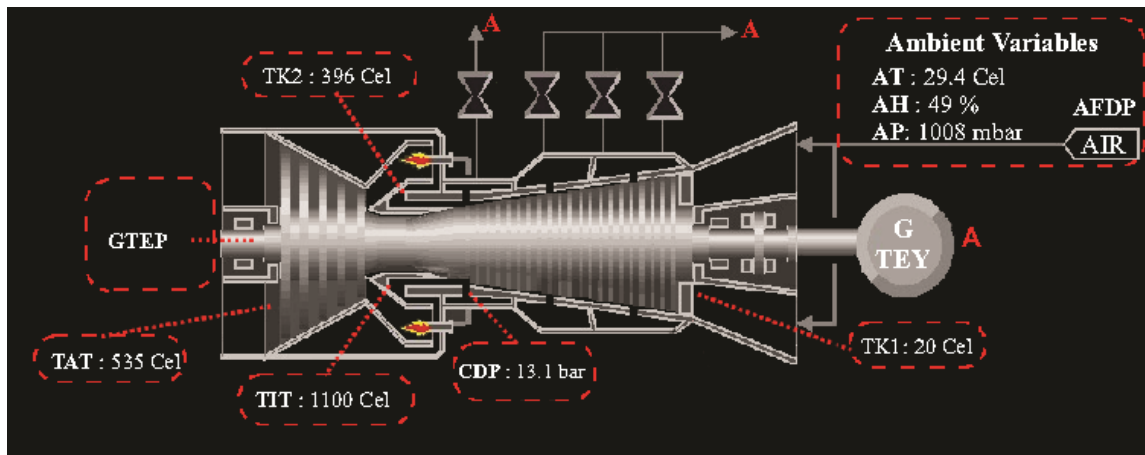


Figure 6. The Sensor Locations/Parameter Sources of a Single-Shaft Gas Turbine [73]

For the modelling process, five variables including ambient pressure ( $AP$ ), ambient temperature ( $AT$ ), ambient humidity ( $AH$ ), turbine energy yield ( $TEY$ ), and air filter difference pressure ( $AFDP$ ) are considered as inputs. The other four parameters consisting of gas turbine exhaust pressure ( $GTEP$ ), turbine after temperature ( $TAT$ ), turbine inlet temperature ( $TIT$ ), and compressor discharge pressure ( $CDP$ ) are appointed as target variables. Table 2 explains the engine variables and their operational ranges. The data include five datasets collected over five years from 2011 to 2015, with a total of 36732 records. Table 3 presents details of the datasets. In order to remove incorrect, incomplete, improperly formatted, irrelevant, or duplicated data, all data are checked, cleaned, and filtered before the training process. Among the five datasets, three datasets including 21947 records are considered for the training process, which cover about 60% of the whole available experimental data. The remaining data consisting of two datasets with a total of 14785 records (40% of the whole data) are assigned for the test (validation) process. Different combinations of the datasets may be considered for training and validation purposes. Running the programming code for a variety of combinations of

datasets showed that the best results were achieved when the combination of the datasets was assigned according to [Table 3](#). [Figures 7 to 24](#) represent the data structures (values) for different variables of the engine for the training and validation processes.

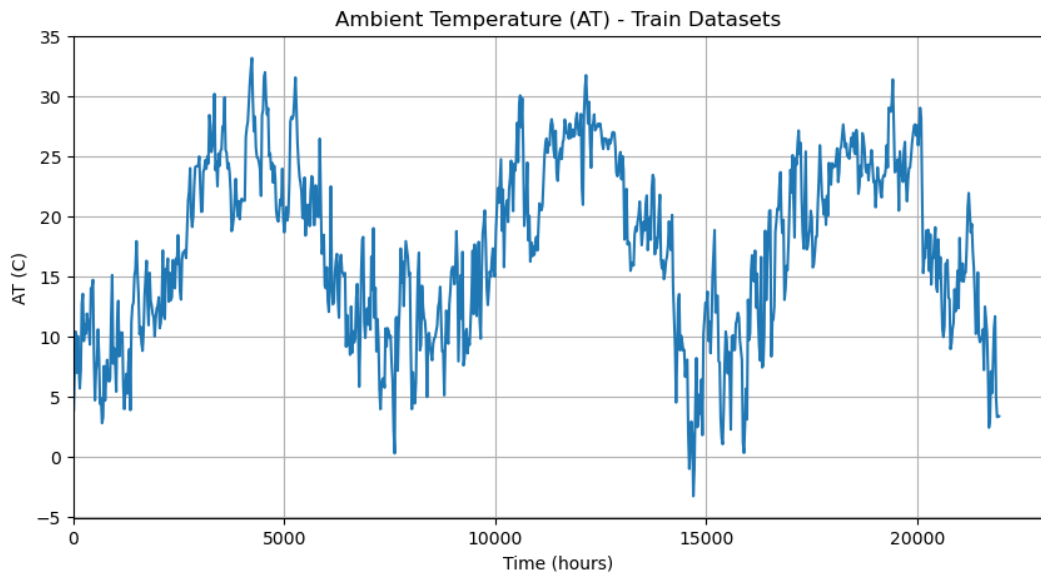
*Table 2. Gas Turbine Engine Variables*

	<b>Variable</b>	<b>Symbol</b>	<b>Operational Range</b>	<b>Unit</b>	<b>Input / Target</b>
1	Ambient temperature	AT	6.23 - 37.10	C°	Input
2	Ambient pressure	AP	985.85 - 1036.56	mbar	Input
3	Ambient humidity	AH	24.08 - 100.20	%	Input
4	Turbine energy yield	TEY	100.02 179.50	MWH	Input
5	Air filter difference pressure	AFDP	2.09 - 7.61	mbar	Input
6	Gas turbine exhaust pressure	GTEP	17.70 - 40.72	mbar	Target
7	Turbine inlet temperature	TIT	1000.85 - 1100.89	C°	Target
8	Turbine after temperature	TAT	511.04 - 550.61	C°	Target
9	Compressor discharge pressure	CDP	9.85 - 15.16	mbar	Target

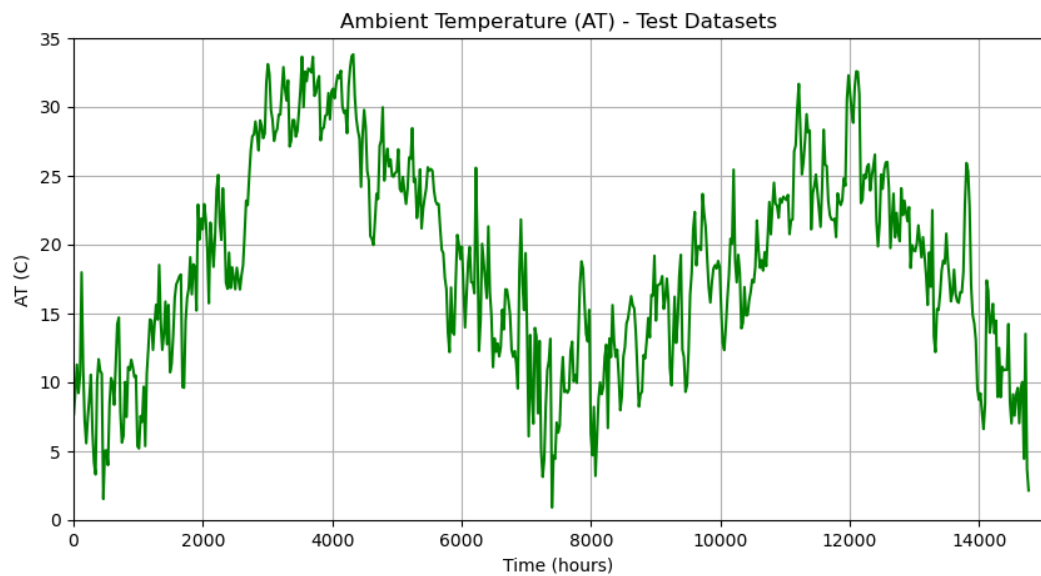
*Table 3. Operational Time-Series Datasets*

	<b>Dataset</b>	<b>Number of Data</b>	<b>Sampling Time (hours)</b>	<b>Training / Validation</b>
1	GT_2011	7411	1	Training
2	GT_2012	7628	1	Validation
3	GT_2013	7152	1	Training
4	GT_2014	7158	1	Validation
5	GT_2015	7384	1	Training
Total Training Datasets		21947	1	Training (60%)
Total Validation Datasets		14785	1	Validation (40%)
Total Datasets		36732	1	Training/Validation

### 3.1.1 Data structure of ambient temperature (AT)



*Figure 7. Data Structure of Ambient Temperature (AT) for Training Process*



*Figure 8. Data Structure of Ambient Temperature (AT) for Validation Process*

### 3.1.2 Data structure of ambient pressure (AP)

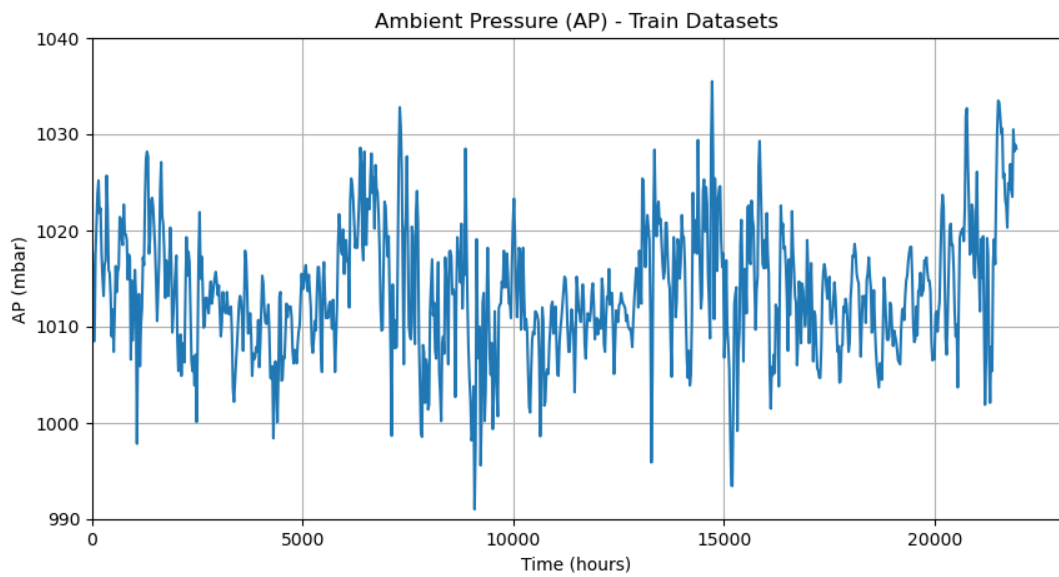


Figure 9. Data Structure of Ambient Pressure (AP) for Training Process

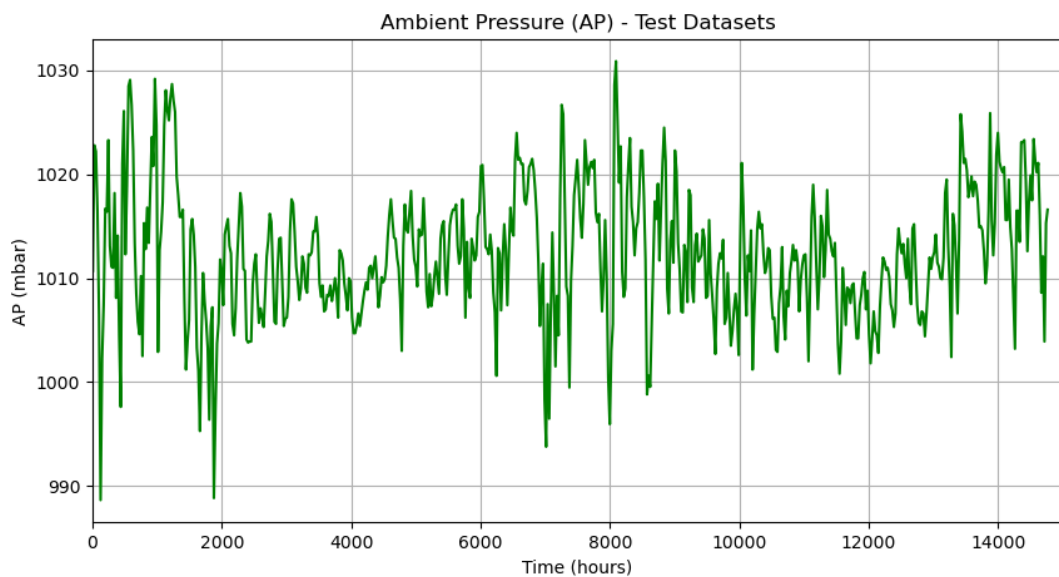


Figure 10. Data Structure of Ambient Pressure (AP) for Validation Process

### 3.1.3 Data structure of ambient humidity (AH)

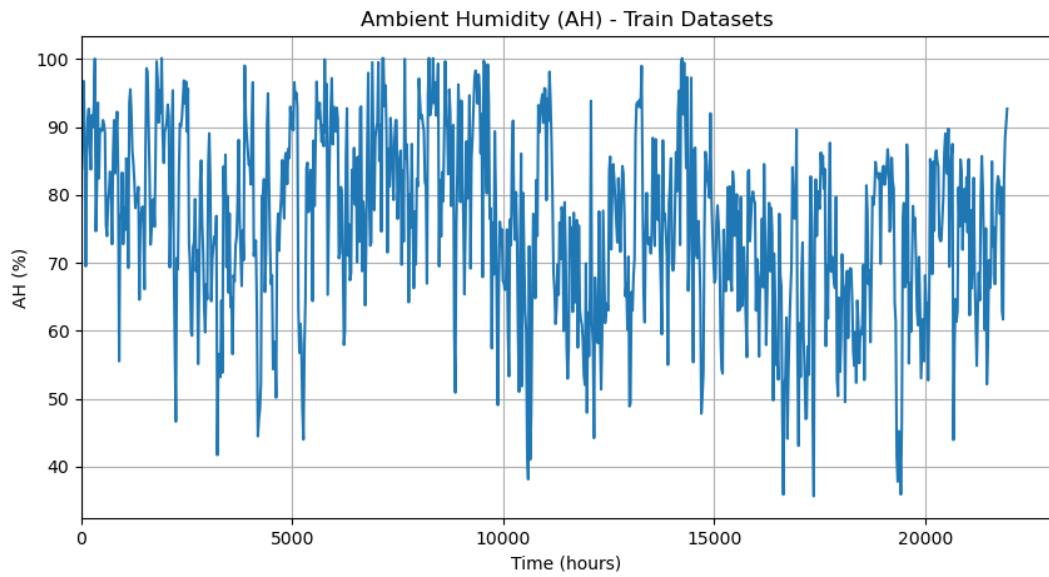


Figure 11. Data Structure of Ambient Humidity (AH) for Training Process

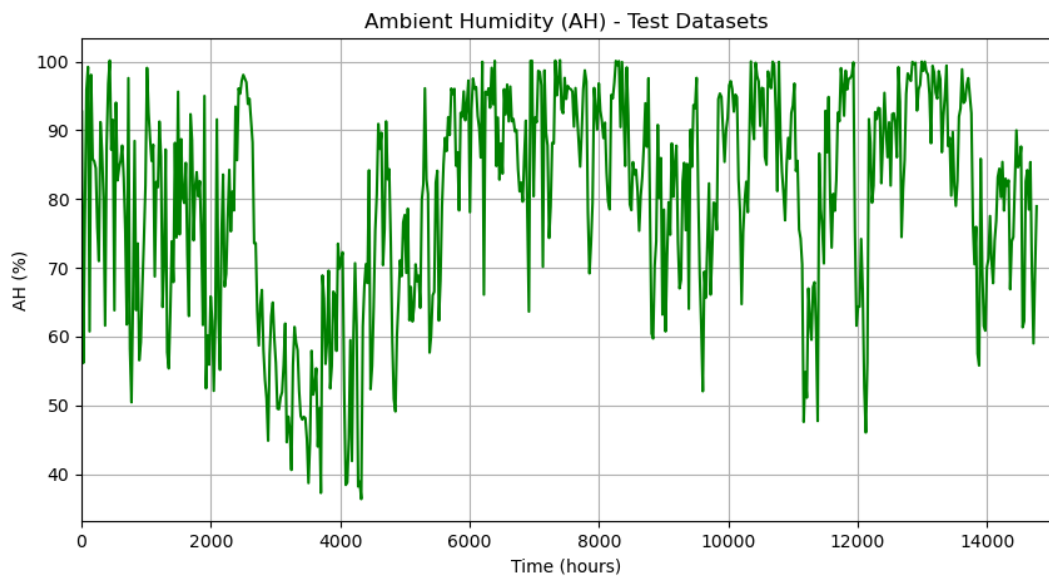


Figure 12. Data Structure of Ambient Humidity (AH) for Validation Process

### 3.1.4 Data structure of turbine energy yield (TEY)

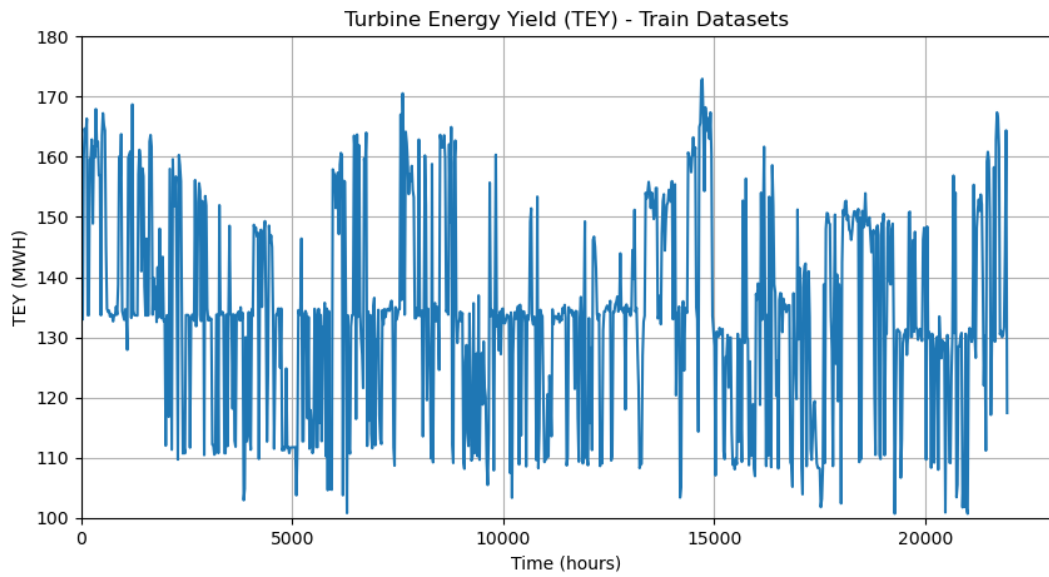


Figure 13. Data Structure of Turbine Energy Yield (TEY) for Training Process

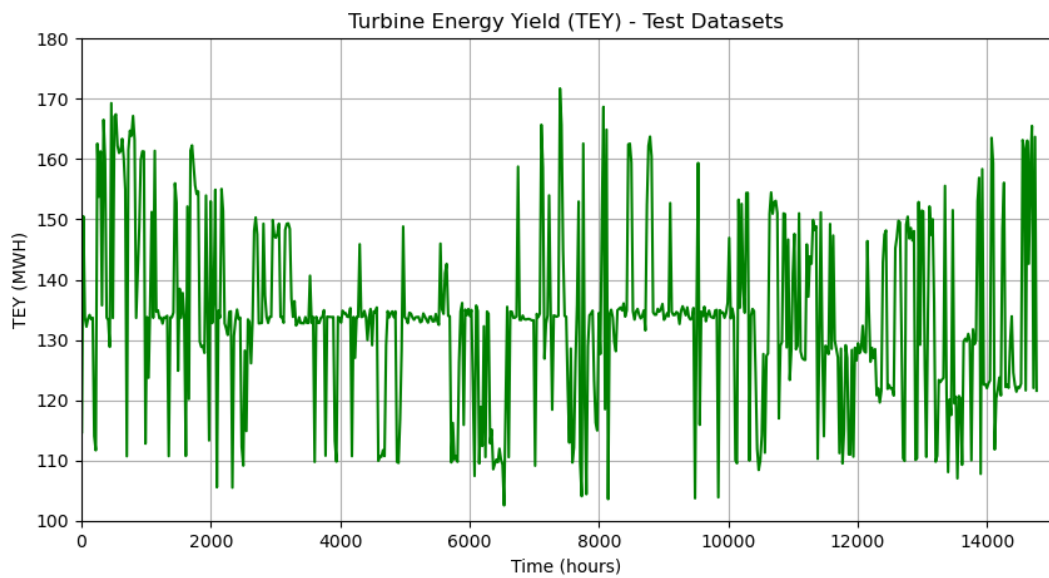


Figure 14. Data Structure of Turbine Energy Yield (TEY) for Validation Process

### 3.1.5 Data structure of air filter difference pressure (AFDP)

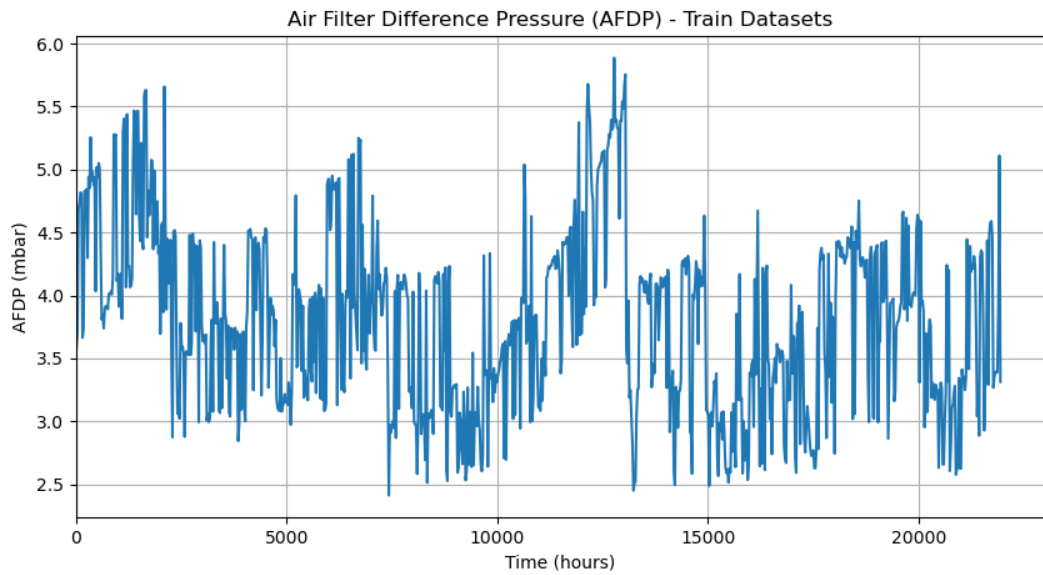


Figure 15. Data Structure of Air Filter Difference Pressure (AFDP) for Training Process

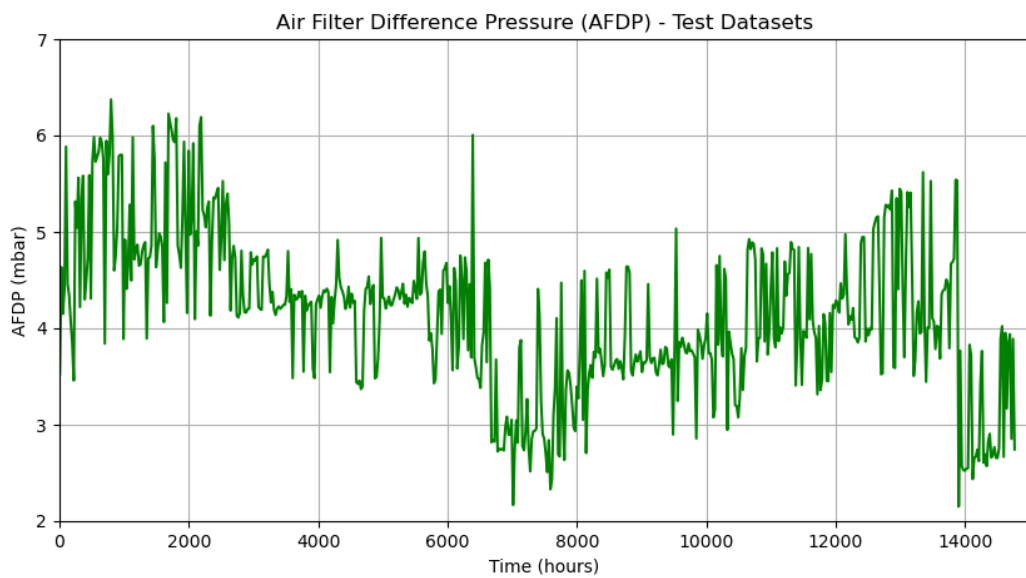


Figure 16. Data Structure of Air Filter Difference Pressure (AFDP) for Validation Process

### 3.1.6 Data structure of gas turbine exhaust pressure (GTEP)

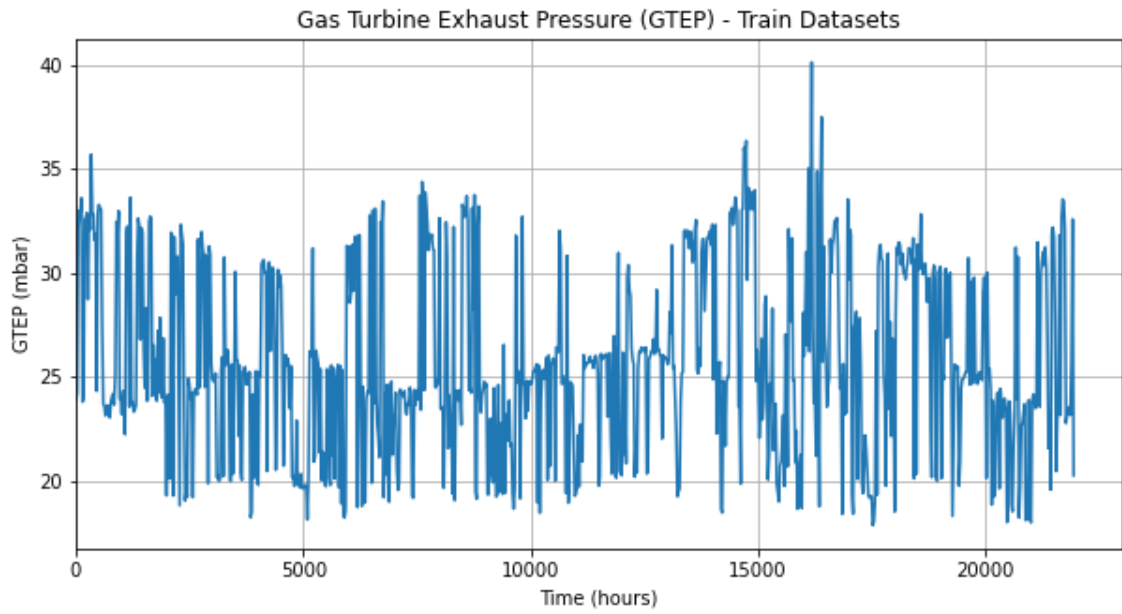


Figure 17. Data Structure of Gas Turbine Exhaust Pressure (GTEP) for Training Process

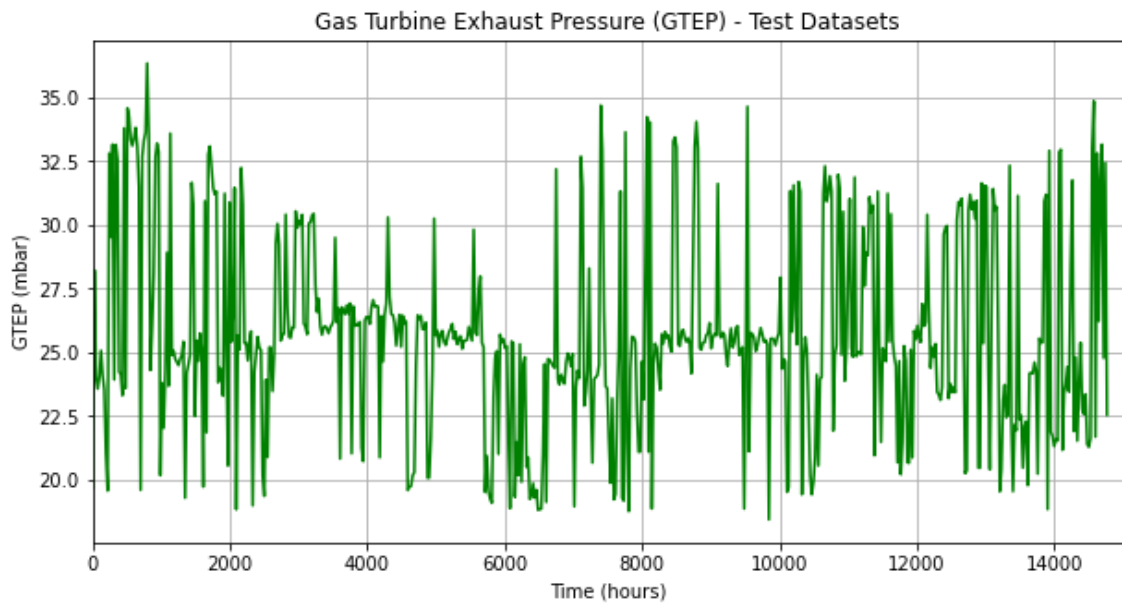


Figure 18. Data Structure of Gas Turbine Exhaust Pressure (GTEP) for Validation Process



### 3.1.7 Data structure of gas turbine inlet temperature (TIT)

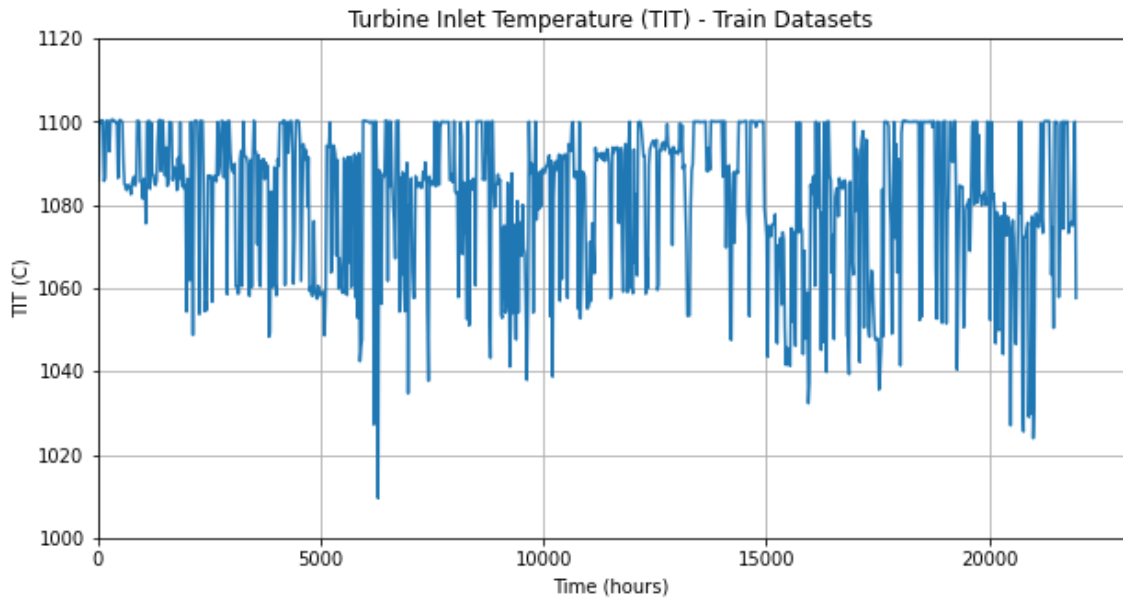


Figure 19. Data Structure of Turbine Inlet Temperature (TIT) for Training Process

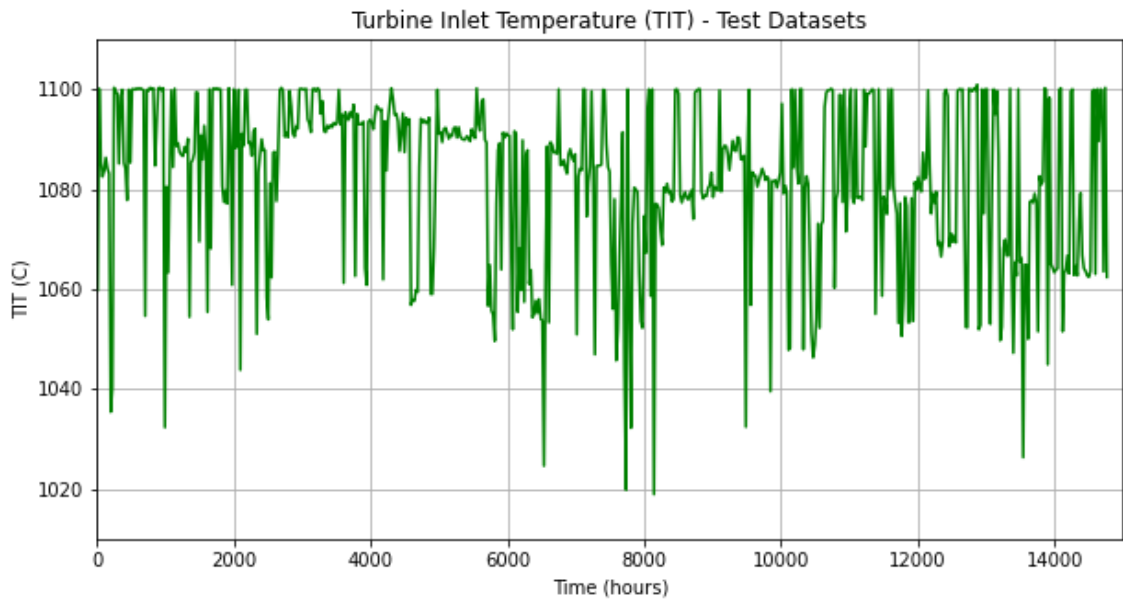


Figure 20. Data Structure of Turbine Inlet Temperature (TIT) for Validation Process

### 3.1.8 Data structure of gas turbine after temperature (TAT)

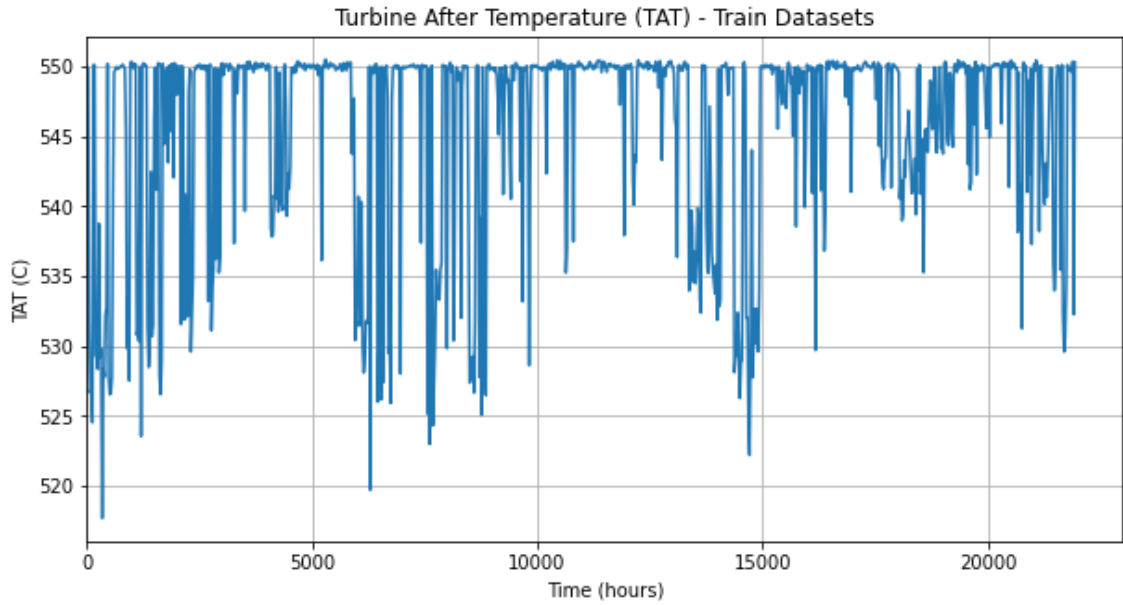


Figure 21. Data Structure of Turbine After Temperature (TAT) for Training Process

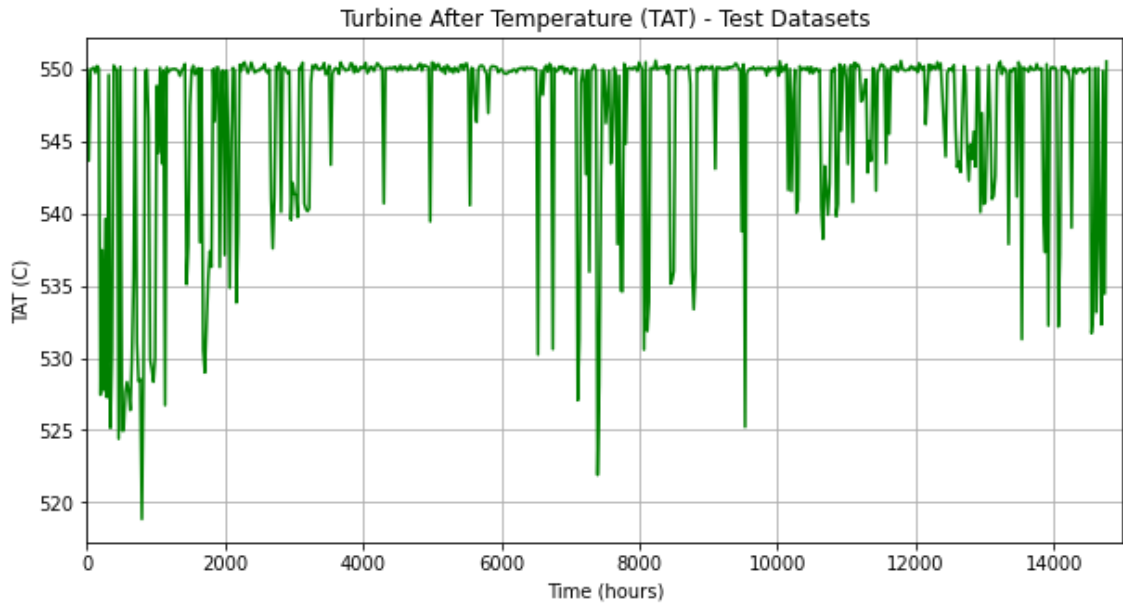


Figure 22. Data Structure of Turbine After Temperature (TAT) for Validation Process

### 3.1.9 Data structure of compressor discharge pressure (CDP)

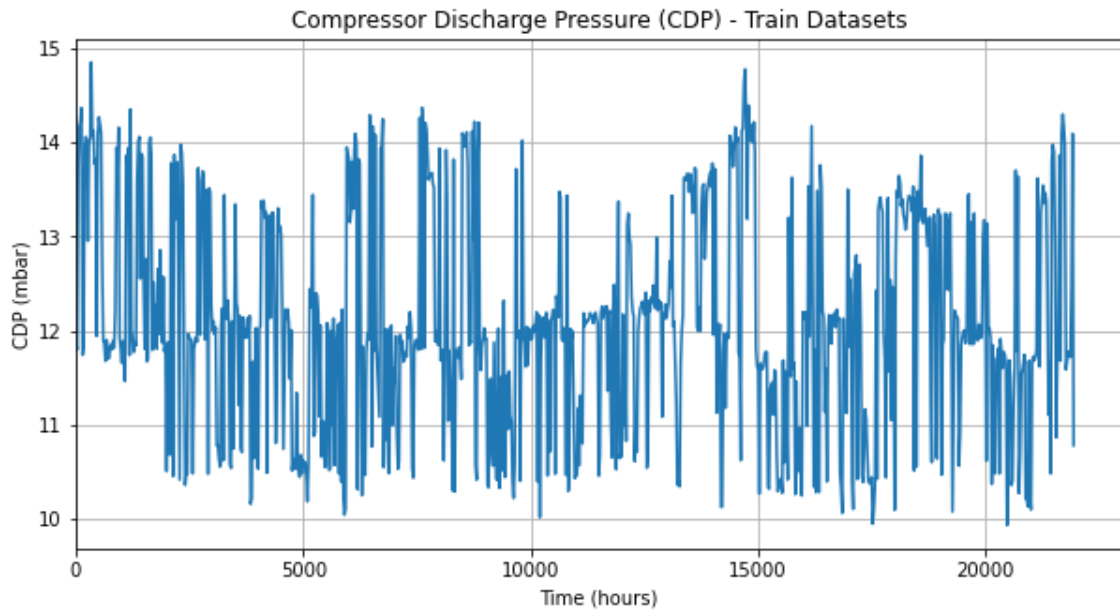


Figure 23. Data Structure of Compressor Discharge Pressure (CDP) for Training Process

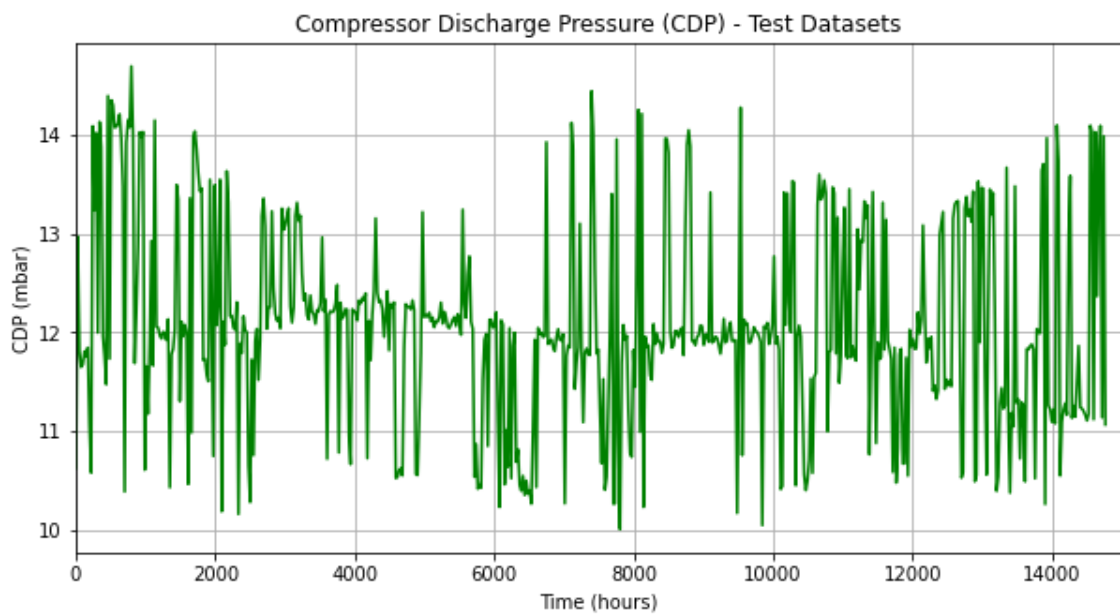


Figure 24. Data Structure of Compressor Discharge Pressure (CDP) for Validation Process

## 3.2 Training process and setting up the data-driven models

As already explained, the gas turbine variables are divided into five inputs and four targets. The objective is to set up an optimal model, which maps inputs to targets as accurately as possible. The generalization and reliability of the resulting model are also very important and should be considered for the training process and selection of the optimal model. To achieve this goal, different data-driven approaches are explored and the resulting models are compared in terms of accuracy, reliability, and generalization. For this purpose, different data-driven techniques are employed. Considering a large number of data, three out of five datasets are used for the training phase, and the other two datasets are employed for the validation process. The data-driven modelling approaches that are considered for this study include *RR*, *LR*, *MTEN*, and *RNN*. For this purpose, a comprehensive code is written in the *Jupyter Notebook* environment by using *Python* programming language.

There are different web-based open-source applications for creating and sharing documents. One of the best of these applications is the *Jupyter Notebook*. It contains live code, visualizations, equations, and narrative text. The *Jupyter Notebook* can be utilized for data cleaning and transformation, data visualization, statistical modelling, numerical simulation, machine learning, etc. [74]. *Python* is a high-level and general-purpose programming language, which has been developing rapidly in recent years. It is widely used because of being user-friendly, interpreted, and powerful in many scientific applications. The design philosophy of *Python* was based on code readability with its remarkable use of significant indentation.

### 3.2.1 Linear models

Linear regression, as a statistical technique, is widely used for predictive analysis by modelling the relationship between two sets of variables as a linear regression equation. In this approach, a continuous variable (target) is modelled as a linear function of other variables (inputs), so that the regression model can predict the system output when the input variables are available. In this study, three linear models including *Ridge*, *Lasso*, and *Multi-Task Elastic-Net* are employed. These techniques are accessible and can

be used through the machine learning module (*scikit-learn*) integrated into *Python*. Both the *Jupyter Notebook* and *Python* are used as programming language environments [75].

### 3.2.1.1 Ridge and Lasso regressions

*Ridge* and *Lasso* regressions, also respectively called *L1* and *L2* regularizations, are data-driven based techniques, which are employed to make linear models with lower complexity. Despite the simple linear regression technique, they can prevent over-fitting. *Ridge* regression reduces the model complexity and multi-collinearity, but cannot be considered as a good method for feature reduction. *Lasso* regression also helps in feature selection, but it has some limitations for dealing with specific types of data. If the number of observations ( $n$ ) is less than the number of predictors ( $p$ ), maximum  $n$  non-zero predictors will be picked up by *Lasso*, even if all of the predictors are relevant. Besides, if the number of highly collinear variables is two or more than two, just one of them will be randomly selected by *Lasso* regression, which negatively affects the data interpretation. The process of reducing (deleting) features is called dimensionality reduction.

### 3.2.1.2 Multi-Task Elastic-Net regression

As already mentioned, in *Lasso* regression, many features are eliminated, and overfitting is reduced, while in *Ridge* regression, the effects of nonsignificant features in the prediction of output values are reduced. *Multi-Task Elastic-Net* regression (*MTEN*), as a regularized regression technique, linearly combines *L1* regularization of *Ridge* with *L2* regularization of *Lasso* methods. In *MTEN* regression, feature coefficient reduction characteristic from the *Ridge* is combined with feature elimination characteristic from *Lasso*, which consequently improves the prediction capability of the model. The capability of *MTEN* in fitting multiple regression tasks can equalize the selected features for all the regression problems [75].

### 3.2.2 Nonlinear model

In dynamic systems, the quantities of variables, or in general, the system behavior changes over time. Modelling of these systems is more difficult compared to statistical systems, which are independent of time. One of the best methodologies for modelling *GTEs*, as dynamic systems with coupled nonlinear equations, is using recurrent (feedback) neural networks, which is a class of artificial neural networks. *RNN* is employed for modelling of time-domain behavior of dynamic systems. In addition to the current inputs, the outputs of *RNN* also depends on previous inputs. Such as other classes of *ANNs*, the structure of *RNN* includes input(s), output(s), and hidden layer(s). Details of *RNN* should be determined before the training process. The details consist of the number of inputs, number of outputs, number of hidden layers, number of neurons in the hidden layer(s), the type of training algorithms and transfer functions, and time-step delays of the recurrent connections. Although it is sometimes necessary to increase the number of neurons to catch the nonlinear dynamics of the system, this will make *RNN* more complex and does not necessarily lead to a more accurate model. Adjusting an appropriate number of neurons would result in a better converging network. Figure 25 shows a typical *RNN* with seven inputs, three outputs, and the hidden layer with ten neurons [76]. In this study, the *RNN* model is set up by using *Pyrenn* module in the *Python* programming environment. *Pyrenn* is a powerful tool in making recurrent neural networks.

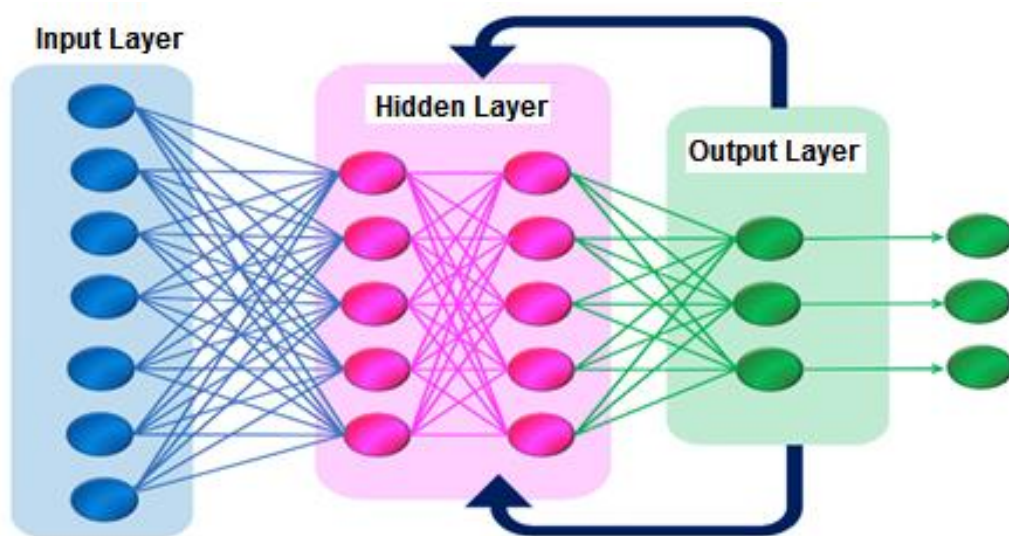


Figure 25. A Typical Recurrent Neural Network [76]

Equation 1 indicates the mathematical expression of the *NARX* model with its feedback connections [77].

$$y(t) = f [u(t-1), u(t-2), \dots, u(t-n_u), y(t-1), y(t-2), \dots, y(t-n_y)] \quad (\text{Equation 1})$$

In this equation,  $y$  and  $u$  respectively indicate output and externally determined variables.  $y(t)$  shows the next value of the dependent output signal, which is calculated on the base of the regression on previous values of the independent input and output signals. The dynamic architecture of *RNN* has made it suitable for supervised data-driven modelling for time-series predictions of nonlinear dynamic systems [77].

### 3.3 Selection of optimal data-driven models

To select the optimal model among the four trained models including *RR*, *LR*, *MTEN*, and *RNN*, the outputs of each model for the four output parameters are compared to the corresponding experimental validation data, and the average coefficient of determination ( $R^2$ ) for each of the four output parameters is calculated. The model with the minimum error (highest value of  $R^2$  accuracy score) will be selected as the optimal model among the four investigated models.

### 3.4 The coefficient of determination ( $R^2$ )

The coefficient of determination, so-called  $R^2$  or *R*-squared, is a statistical criterion (measurement) for evaluation of the accuracy on the base of differences between the real and the corresponding predicted values of data.  $R^2$ , like linear regression, is used to determine how well a line fits an experimental dataset, especially when comparing models. It shows the fraction of the total output variations captured by a model, and an indication of how well does a line follows the variations within a set of data. Equation 2 represents the mathematical expression of  $R^2$  [78].

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum_i^n ((y_i - \hat{y}_i)^2)}{\sum_i^n ((y_i - \bar{y}_i)^2)} \quad (\text{Equation 2})$$

In Equation 2,  $n$  is the number of data,  $y_i$  is the available data,  $\hat{y}_i$  is the model prediction, and  $\bar{y}_i$  is the average (mean value) of the available data.  $SS_{res}$  indicates the residual sum of squared errors of the regression model, while  $SS_{tot}$  shows the total sum of squared errors.  $R^2$  is a value between 0 and 1. If the modeled values exactly match the observed

values,  $SS_{res} = 0$  and  $R^2 = 1$ . In a baseline model, which always predicts  $\bar{y}$ ,  $R^2 = 0$ . The model capability in prediction of the dependent variable becomes better, as the value of  $R^2$  becomes closer to 1 [78].

### 3.5 Summary

This chapter explained the methodology of this study. It provided information about the necessity and benefits of data-driven modelling of industrial systems with complex dynamics and nonlinear behaviors. The process of data collection and data preparation, and the structure of the datasets were presented in detail by using illustrative tables and figures. And finally, the procedures for writing the programming code, setting up four different linear and nonlinear models, and selecting the optimal model were described. According to the objective of this research, different data-driven methodologies are employed to model a single-shaft gas turbine located in Turkey. Five datasets collected over five years from 2011 to 2015, totally including 36732 records are used for the modelling and validation processes. For this purpose, a comprehensive code is written in the *Jupyter Notebook* and *Python* programming language environments. *Ridge* regression, *Lasso* regression, *Multi-Task Elastic-Net* regression and *RNN* are applied to the datasets and the results are compared in terms of accuracy, reliability, and generalization. Finally, the optimal model is chosen on the basis of the results obtained from different modelling approaches.



## 4 RESULTS

As already explained in the previous chapter, the datasets of the *GTE* were split into training and validation sets. A comprehensive code was written in the *Jupyter Notebook* environment by using *Python* programming language for training and setting up four different models including *Ridge*, *Lasso*, *Multi-Task Elastic-Net*, and *RNN*. The resulting models were validated against validation datasets and coefficients of determination ( $R^2$ ) for four output parameters of the engine were calculated for each of the models. Finally, the results for each of the models were figured and compared.

### 4.1 Tuning alpha parameters

To obtain the optimal model of the engine models, the significant training parameter *Alpha* was tuned for the linear models through the integration of the required code in the programming environment. [Figure 26](#) illustrates the results of the tuning for the *Ridge*, *Lasso*, and *MTEN* models respectively. As we can see from these figures, the accuracy score  $R^2$  increases as the *Alpha* parameter approaches zero. Therefore, the *Alpha* parameter for training the three linear models was tuned quite close to zero.

### 4.2 Training and Validation Results

Four different linear and nonlinear models were trained, validated, and evaluated (scored) using the measured time-series datasets according to the procedure already explained in the previous chapter (Methodology). [Figures 27-34](#), [35-42](#), [43-50](#), and [51-58](#) respectively show the results of the simulations for the *Ridge*, *Lasso*, *MTEN*, and *RNN* models, for both training and validation processes. The results will be evaluated, compared, and discussed in the next chapter (Discussion).

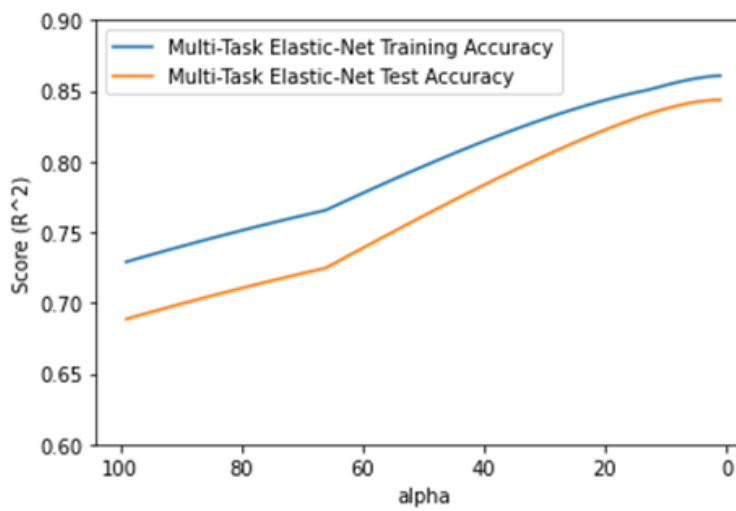
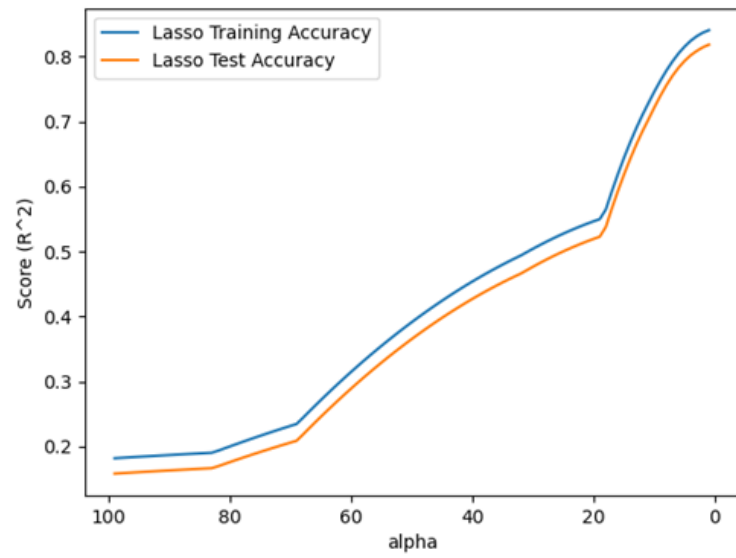
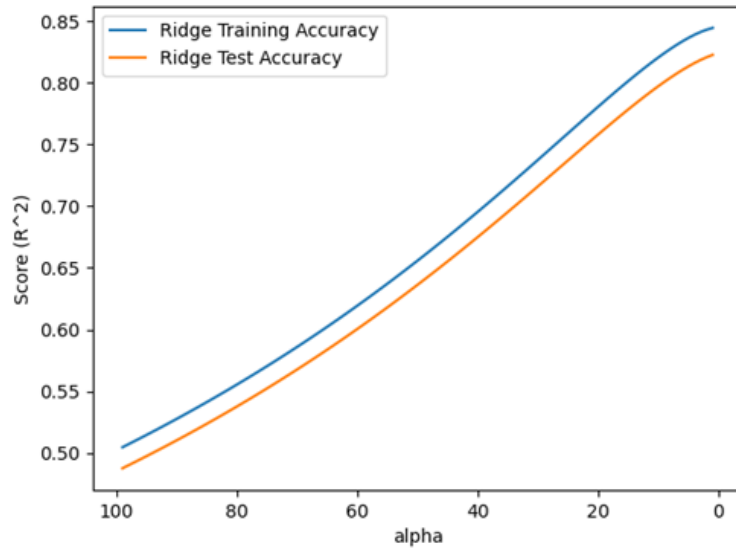


Figure 26. Variations of  $R^2$  for Ridge, Lasso, and MTEN Models with the Alpha Parameter for the Training and Test Datasets

## 4.3 Results for the Ridge model

### 4.3.1 Ridge model: results for gas turbine exhaust pressure (GTEP)

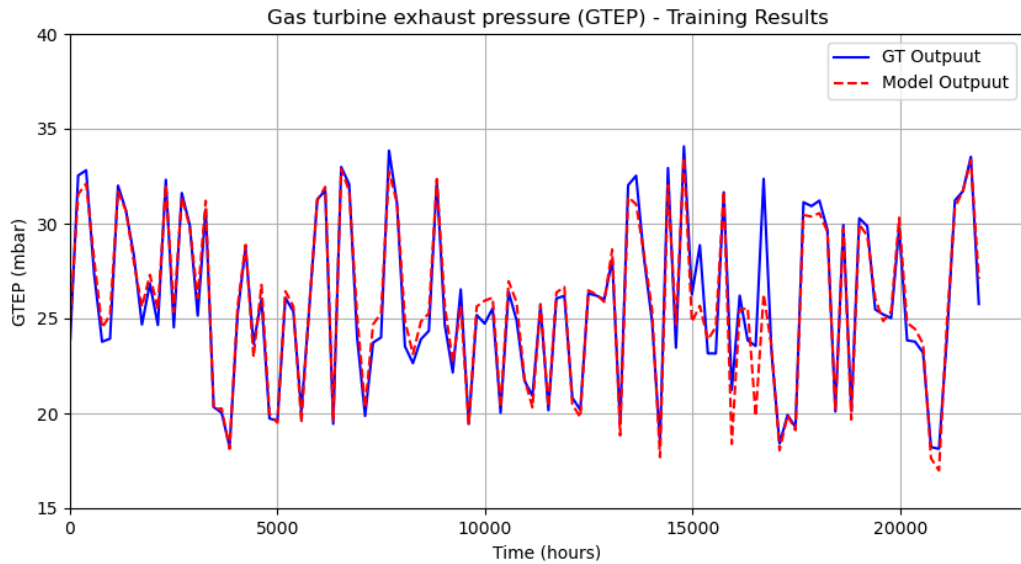


Figure 27. A Comparison Between Outputs of GT Engine and Ridge Model for GT Exhaust Pressure (GTEP) for Training Datasets (Accuracy: 94%)

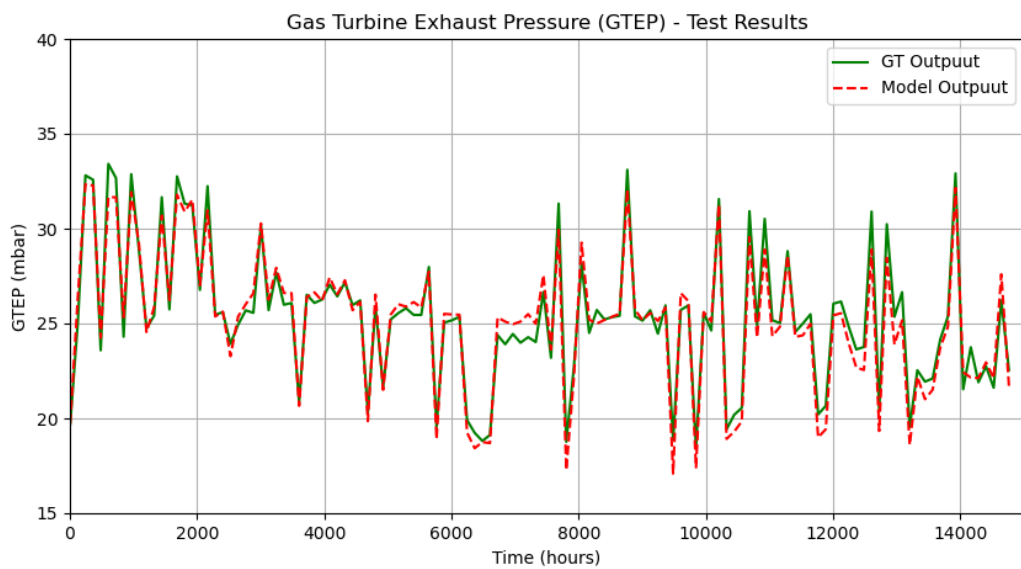


Figure 28. A Comparison Between Outputs of GT Engine and Ridge Model for GT Exhaust Pressure (GTEP) for Test Datasets (Accuracy: 96%)

### 4.3.2 Ridge model: results for turbine inlet temperature (TIT)

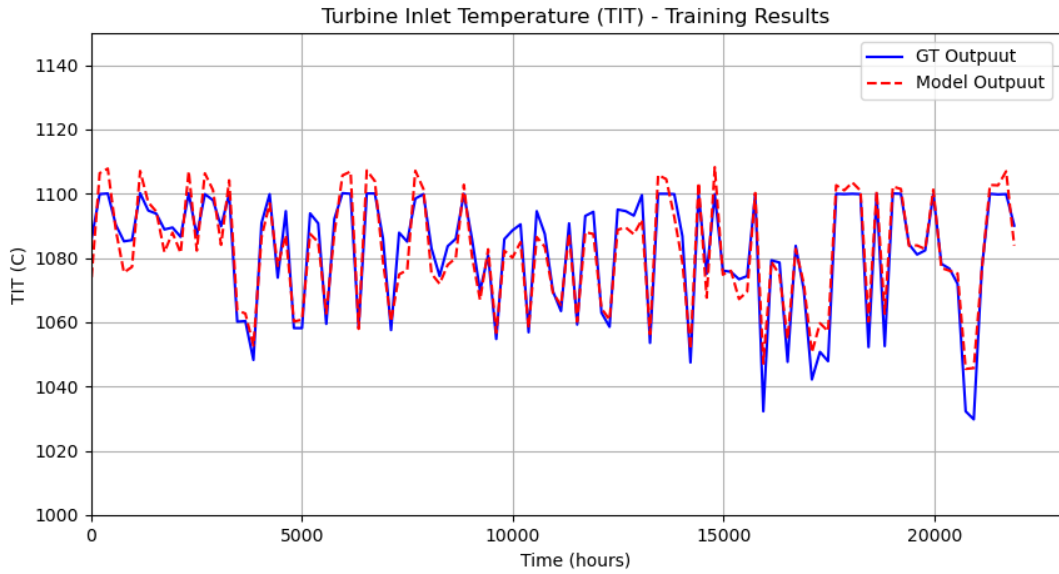


Figure 29. A Comparison Between Outputs of GT Engine and Ridge Model for Turbine Inlet Temperature (TIT) for Training Datasets (Accuracy: 91%)

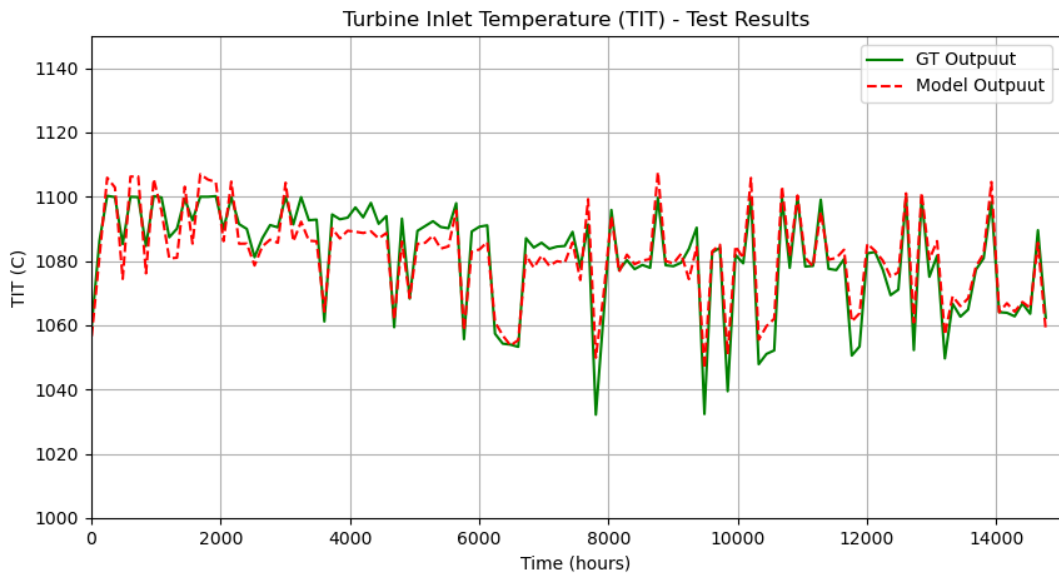


Figure 30. A Comparison Between Outputs of GT Engine and Ridge Model for Turbine Inlet Temperature (TIT) for Test Datasets (Accuracy: 89%)

### 4.3.3 Ridge model: results for turbine after temperature (TAT)

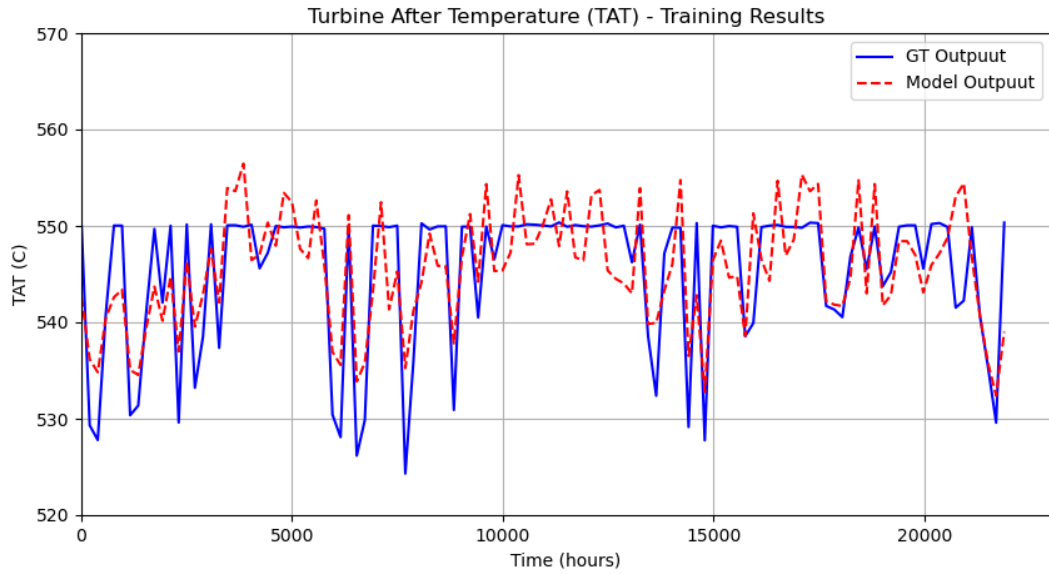


Figure 31. A Comparison Between Outputs of GT Engine and Ridge Model for Turbine After Temperature (TAT) for Training Datasets (Accuracy: 55%)

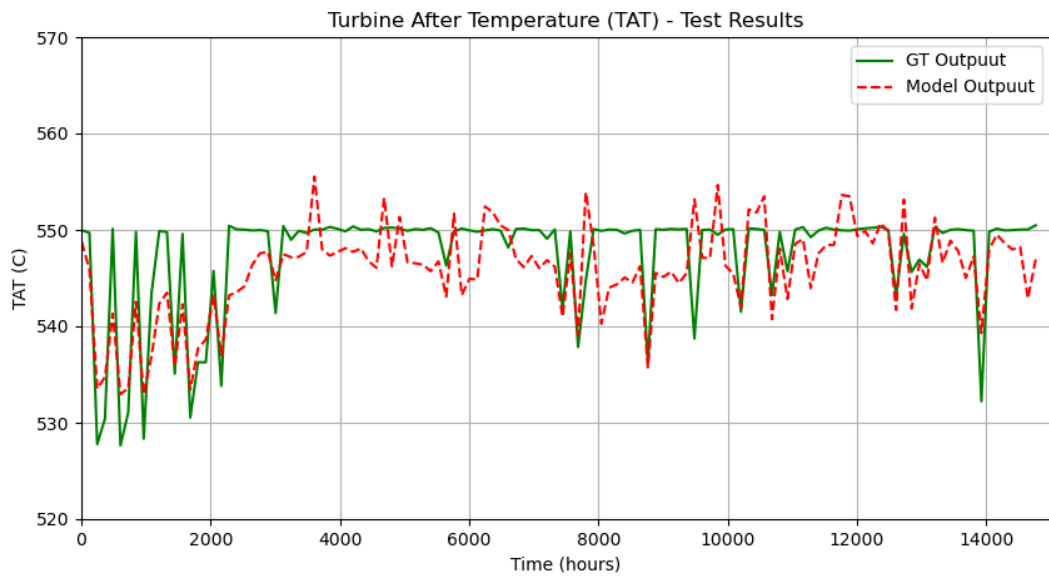


Figure 32. A Comparison Between Outputs of GT Engine and Ridge Model for Turbine After Temperature (TAT) for Test Datasets (Accuracy: 46%)

### 4.3.4 Ridge model: results for compressor discharge pressure (CDP)

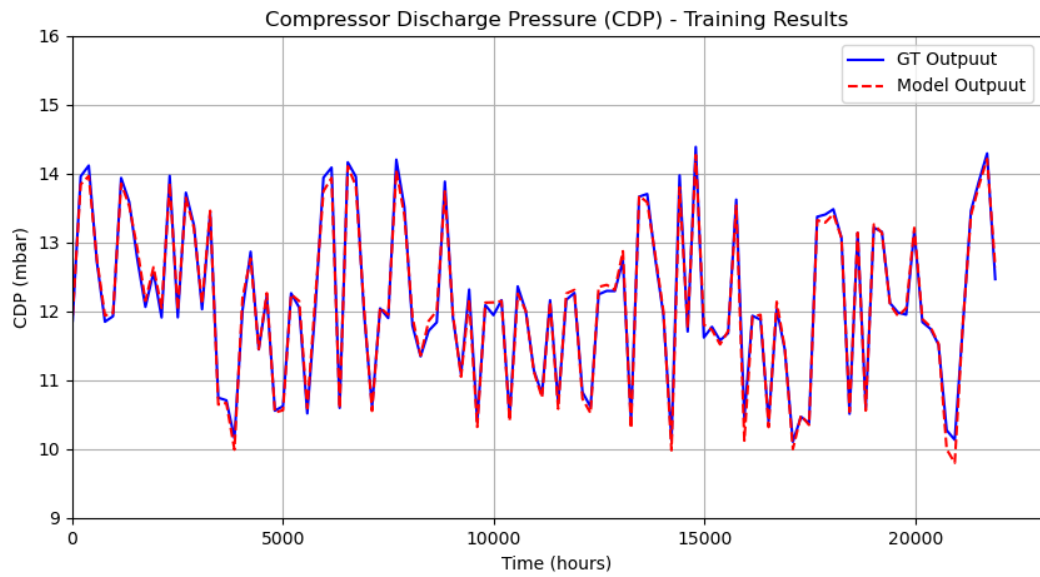


Figure 33. A Comparison Between Outputs of GT Engine and Ridge Model for Compressor Discharge Pressure (CDP) for Training Datasets (Accuracy: 99%)

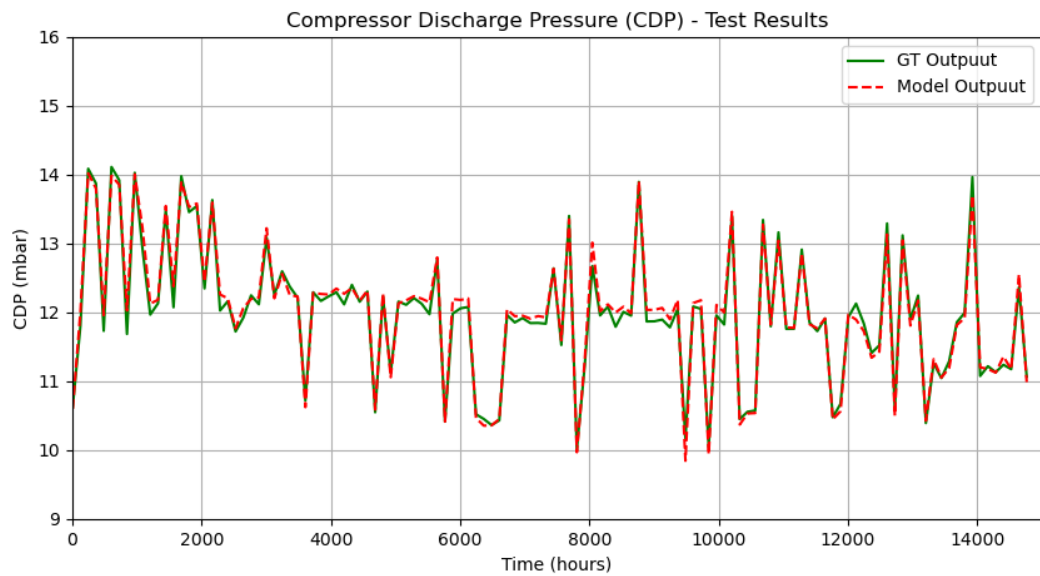


Figure 34. A Comparison Between Outputs of GT Engine and Ridge Model for Compressor Discharge Pressure (CDP) for Test Datasets (Accuracy: 99%)

## 4.4 Results for the Lasso model

### 4.4.1 Lasso model: results for gas turbine exhaust pressure (GTEP)

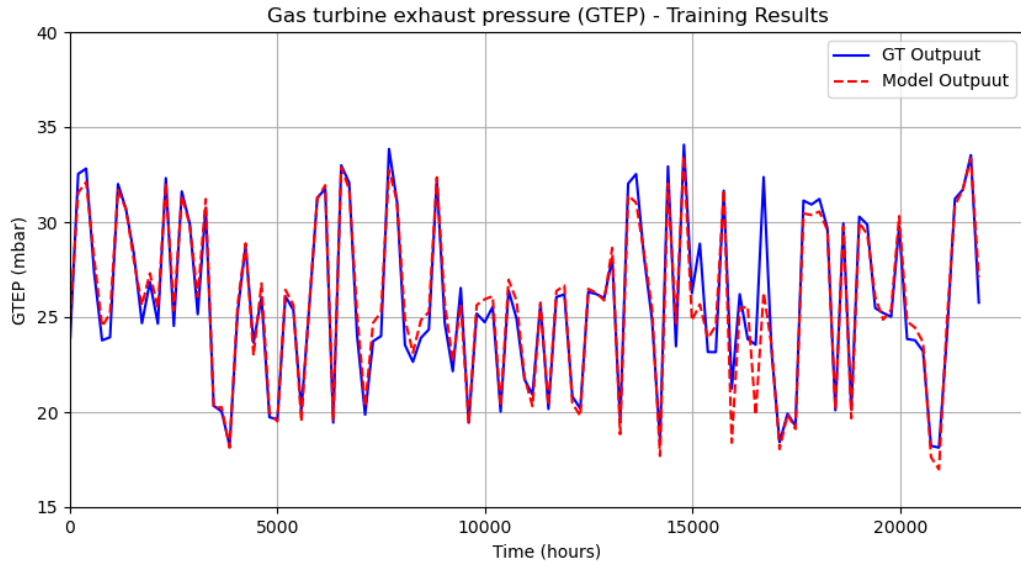


Figure 35. A Comparison Between Outputs of GT Engine and Lasso Model for Gas Turbine Exhaust Pressure (GTEP) for Training Datasets (Accuracy: 94%)

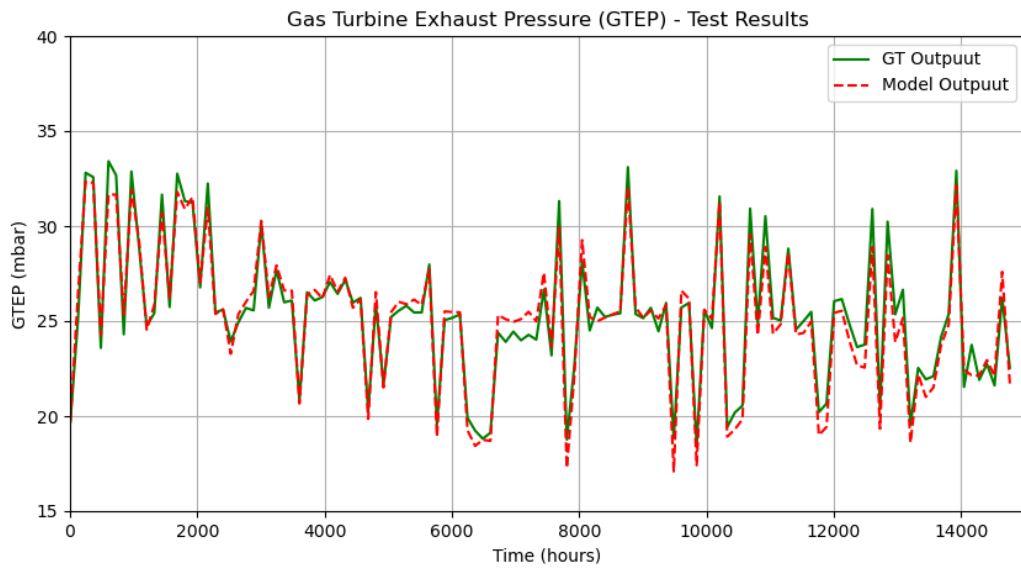


Figure 36. A Comparison Between Outputs of GT Engine and Lasso Model for Gas Turbine Exhaust Pressure (GTEP) for Test Datasets (Accuracy: 96%)

#### 4.4.2 Lasso model: results for turbine inlet temperature (TIT)

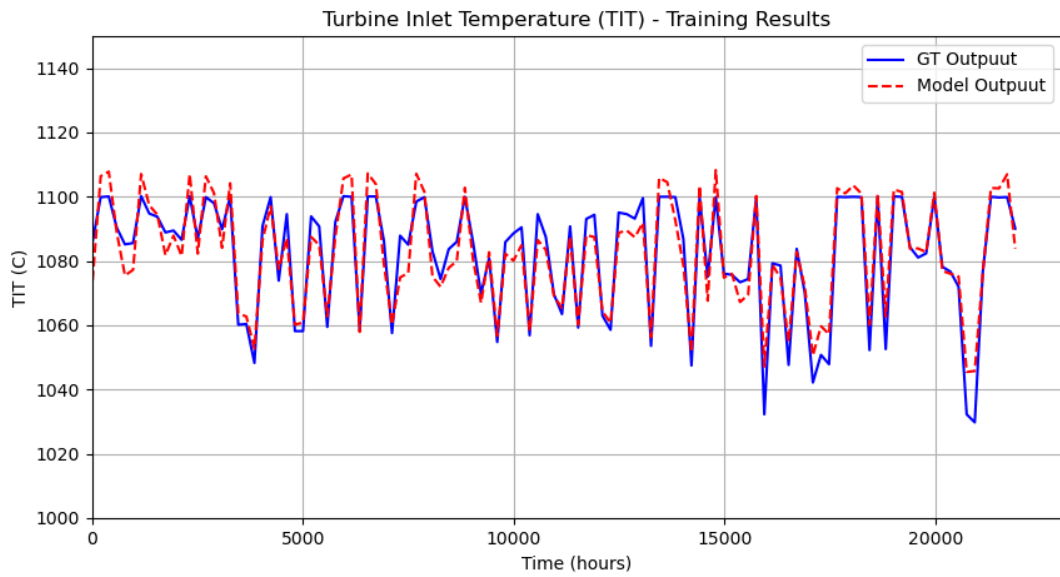


Figure 37. A Comparison Between Outputs of GT Engine and Lasso Model for Turbine Inlet Temperature (TIT) for Training Datasets (Accuracy: 91%)

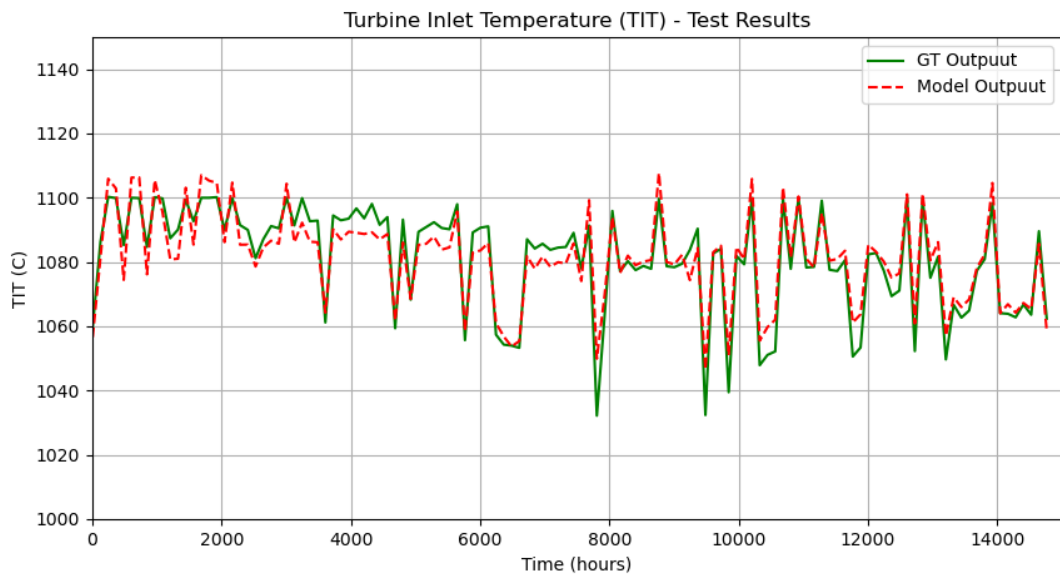


Figure 38. A Comparison Between Outputs of GT Engine and Lasso Model for Turbine Inlet Temperature (TIT) for Test Datasets (Accuracy: 89%)



### 4.4.3 Lasso model: results for turbine after temperature (TAT)

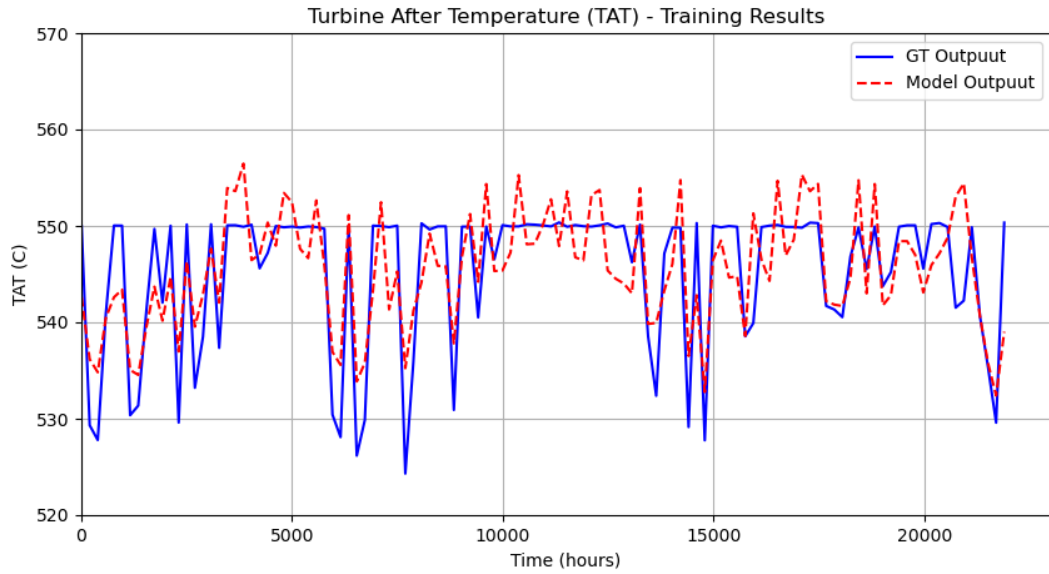


Figure 39. A Comparison Between Outputs of GT Engine and Lasso Model for Turbine After Temperature (TAT) for Training Datasets (Accuracy: 55%)

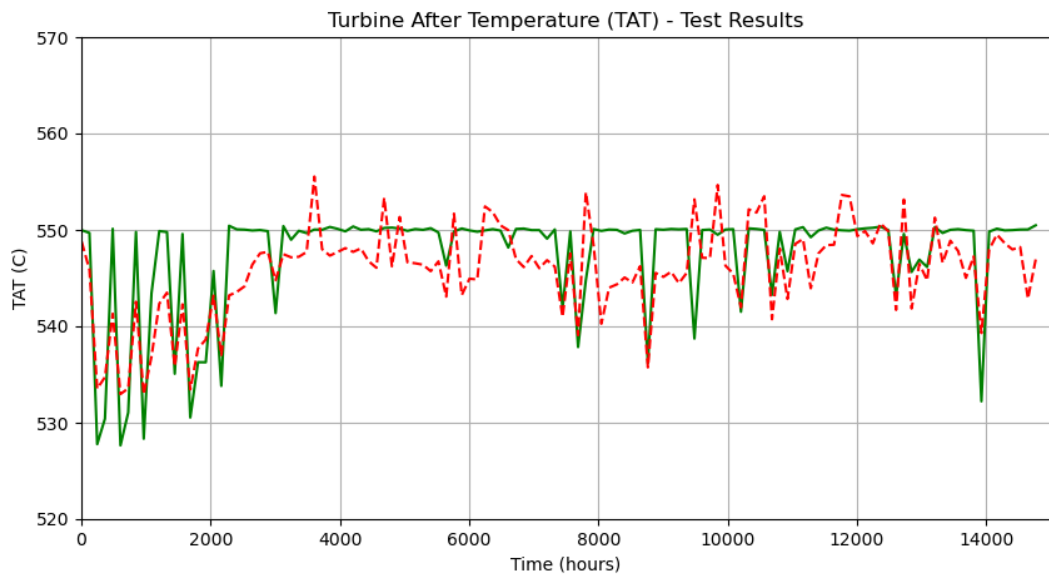


Figure 40. A Comparison Between Outputs of GT Engine and Lasso Model for Turbine After Temperature (TAT) for Test Datasets (Accuracy: 46%)

#### 4.4.4 Lasso model: results for compressor discharge pressure (CDP)

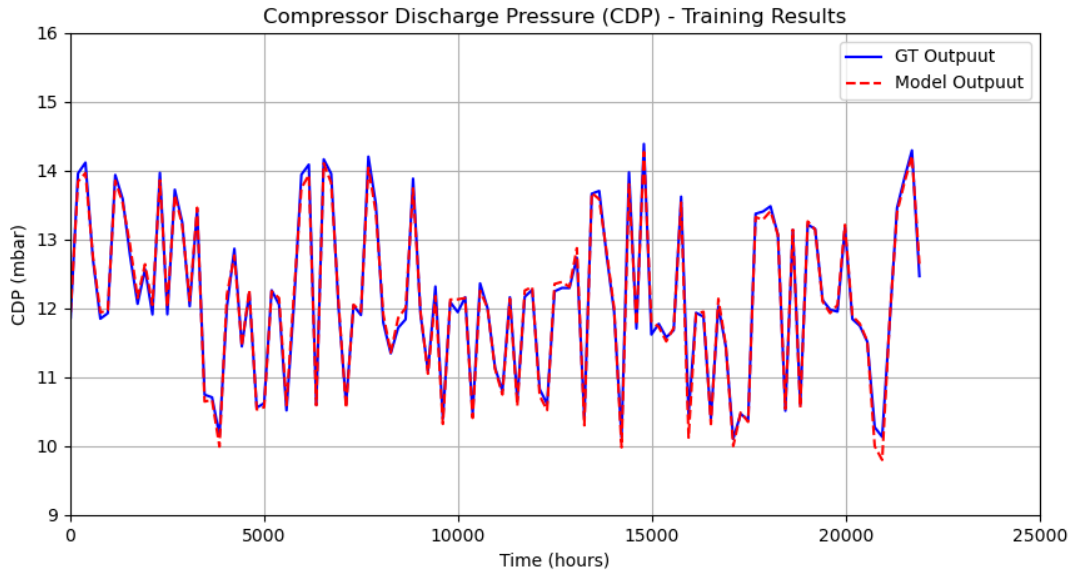


Figure 41. A Comparison Between Outputs of GT Engine and Lasso Model for Compressor Discharge Pressure (CDP) for Training Datasets (Accuracy: 99%)

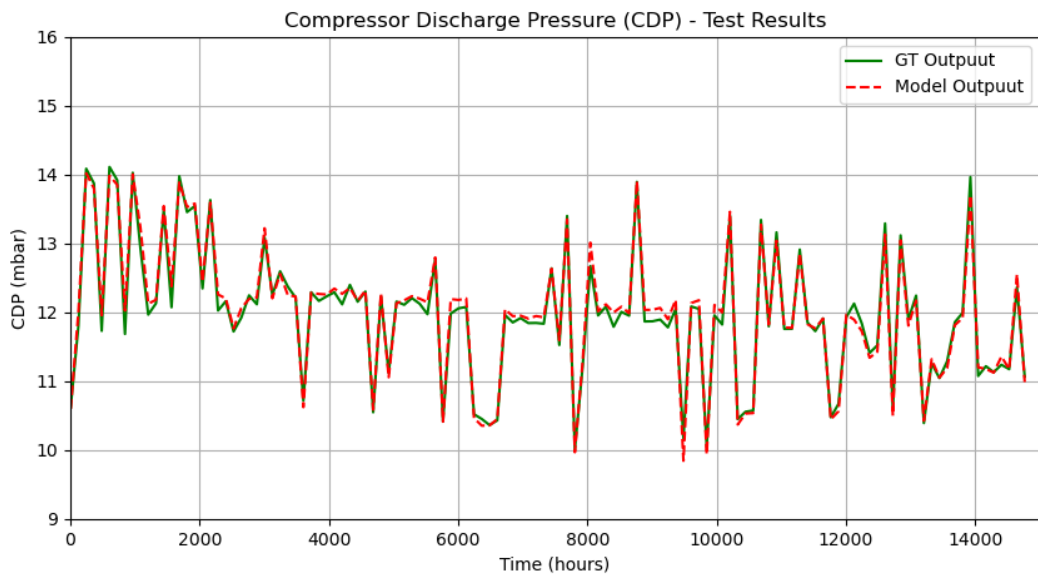


Figure 42. A Comparison Between Outputs of GT Engine and Lasso Model for Compressor Discharge Pressure (CDP) for Test Datasets (Accuracy: 99%)

## 4.5 Results for the Multi-Task Elastic-Net (MTEN) model

### 4.5.1 MTEN model: results for gas turbine exhaust pressure (GTEP)

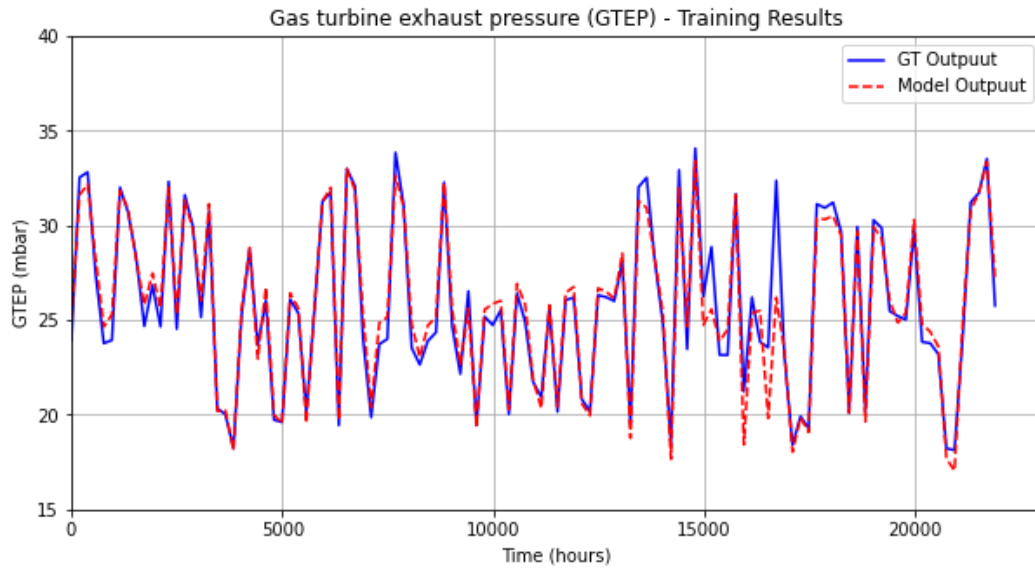


Figure 43. A Comparison Between Outputs of GT Engine and MTEN Model for GT Exhaust Pressure (GTEP) for Training Datasets (Accuracy: 94%)

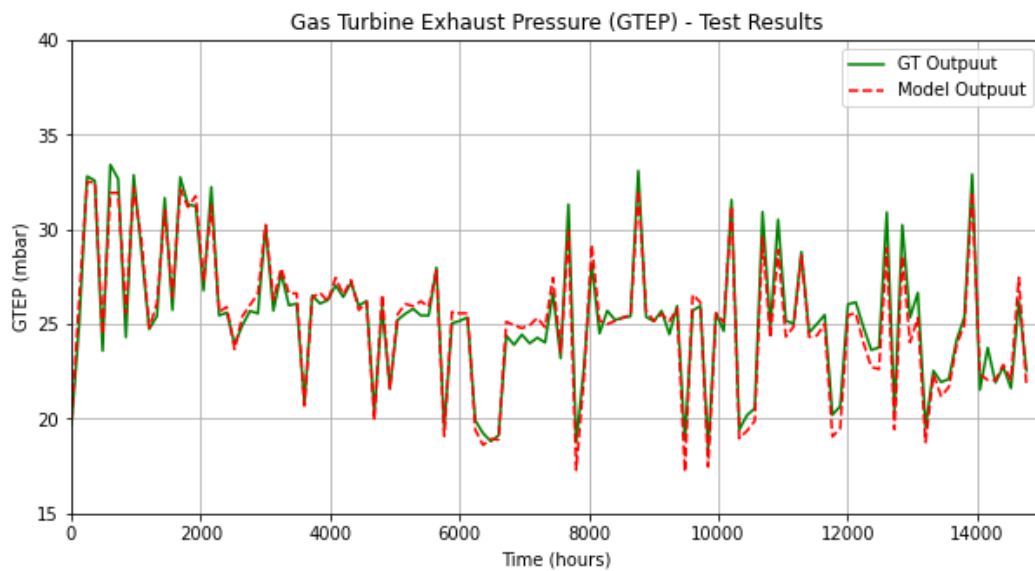


Figure 44. A Comparison Between Outputs of GT Engine and MTEN Model for Gas Turbine Exhaust Pressure (GTEP) for Validation Datasets (Accuracy: 96%)

#### 4.5.2 MTEN model: results for turbine inlet temperature (TIT)

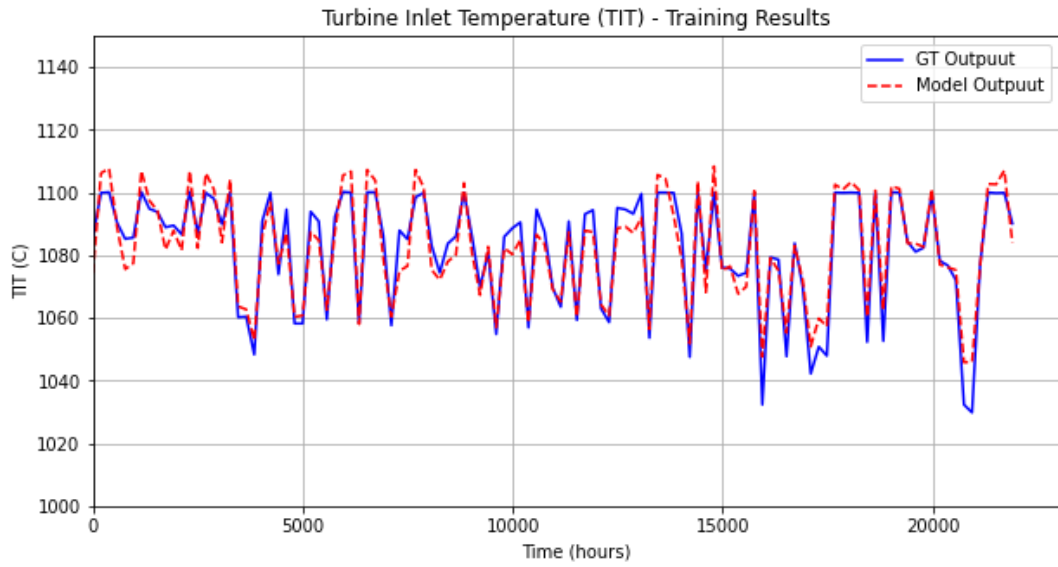


Figure 45. A Comparison Between Outputs of GT Engine and MTEN Model for Turbine Inlet Temperature (TIT) for Training Datasets (Accuracy: 91%)

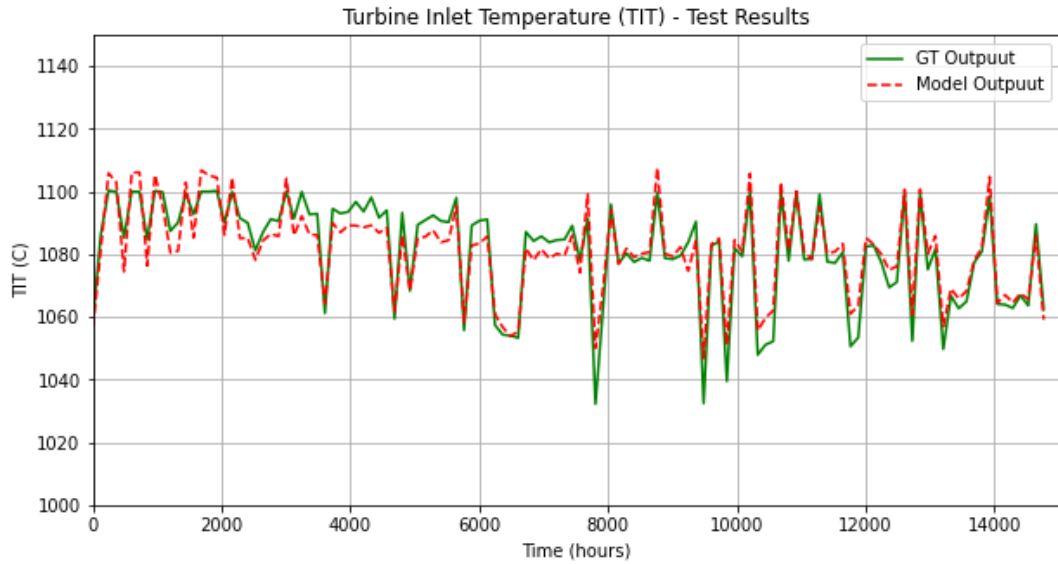


Figure 46. A Comparison Between Outputs of GT Engine and MTEN Model for Turbine Inlet Temperature (TIT) for Validation Datasets (Accuracy: 89%)

### 4.5.3 MTEN model: results for turbine after temperature (TAT)

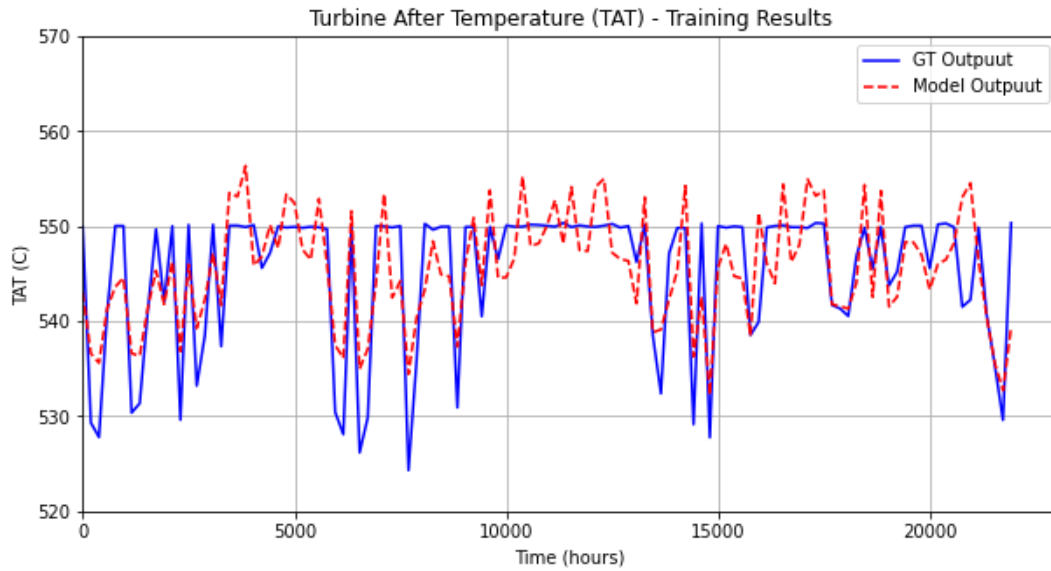


Figure 47. A Comparison Between Outputs of GT Engine and MTEN Model for Turbine After Temperature (TAT) for Training Datasets (Accuracy: 54%)

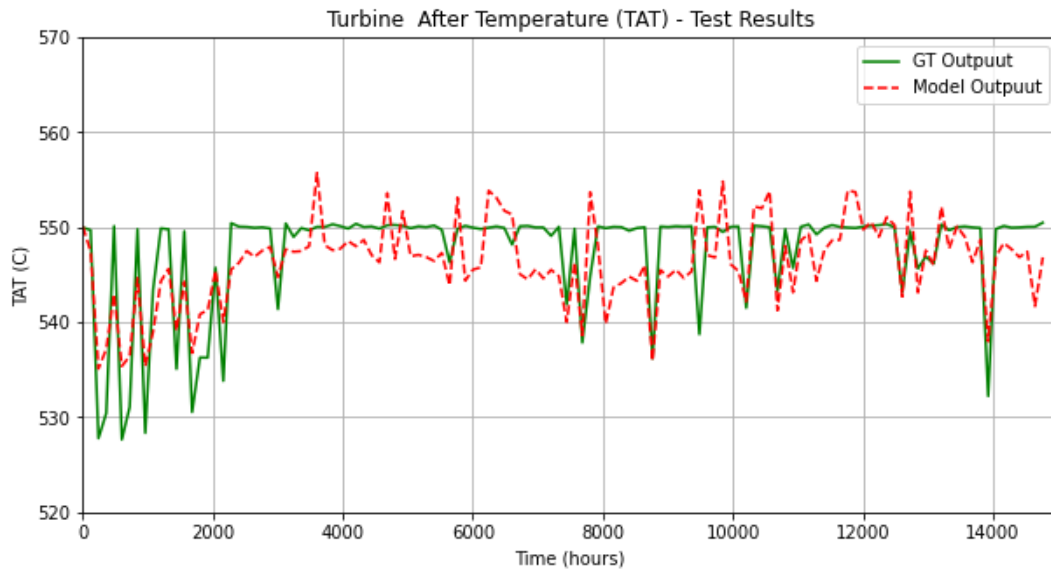


Figure 48. A Comparison Between Outputs of GT Engine and MTEN Model for Turbine After Temperature (TAT) for Validation Datasets (Accuracy: 45%)

#### 4.5.4 MTEN model: results for compressor discharge pressure (CDP)

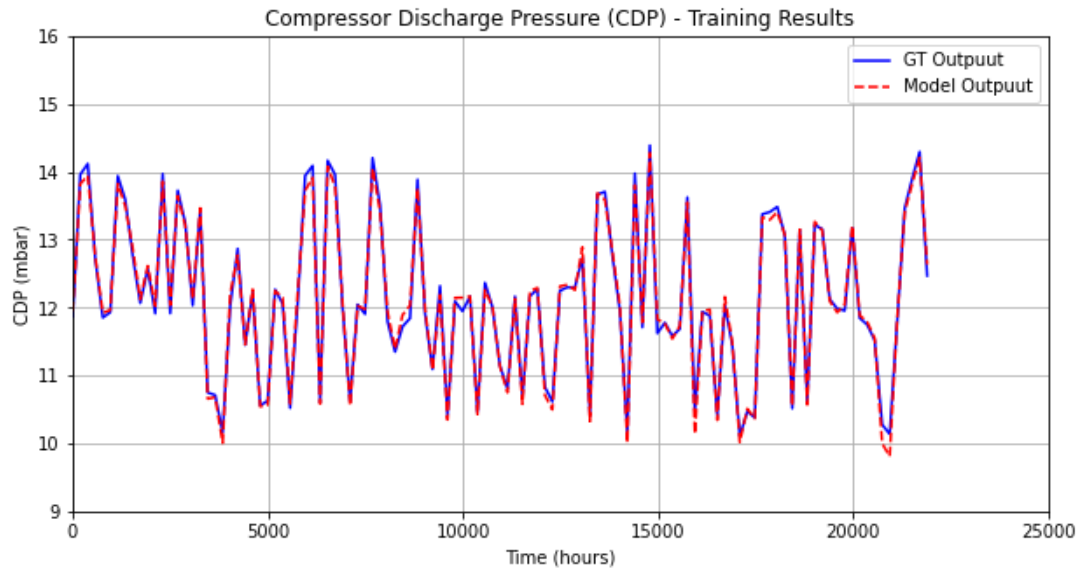


Figure 49 A Comparison Between Outputs of GT Engine and MTEN Model for Compressor Discharge Pressure (CDP) for Training Datasets (Accuracy: 99%)

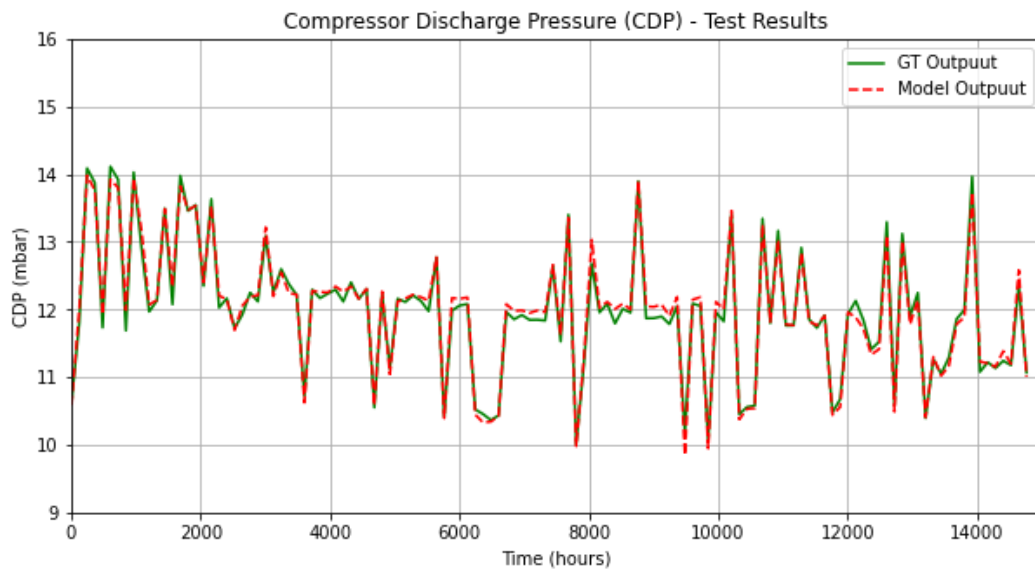


Figure 50. A Comparison Between Outputs of GT Engine and MTEN Model for Compressor Discharge Pressure (CDP) for Validation Datasets (Accuracy: 99%)

## 4.6 Results for the RNN model

### 4.6.1 RNN model: results for gas turbine exhaust pressure (GTEP)

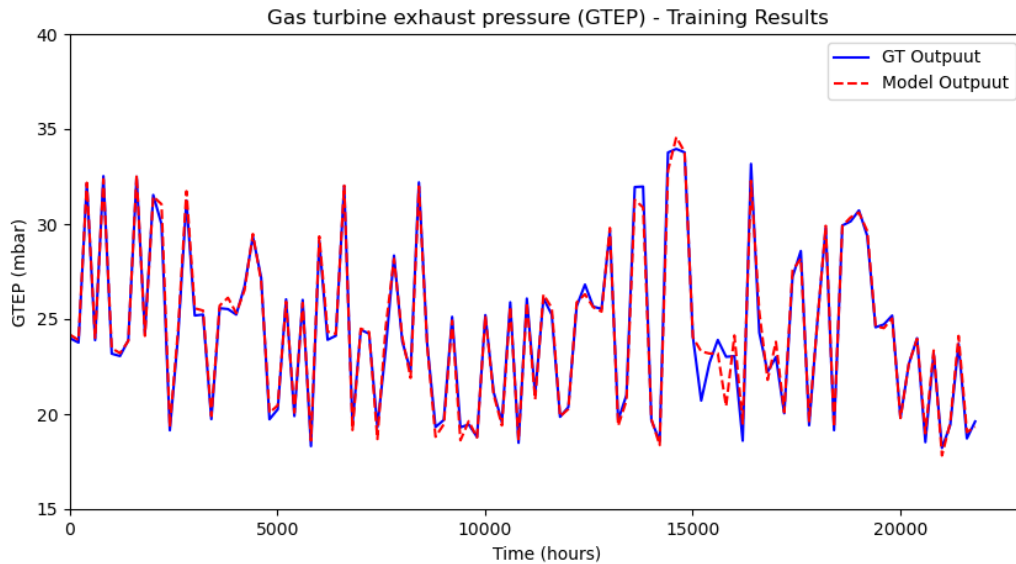


Figure 51. A Comparison Between Outputs of GT Engine and RNN Model for GT Exhaust Pressure (GTEP) for Training Datasets (Accuracy: 98%)

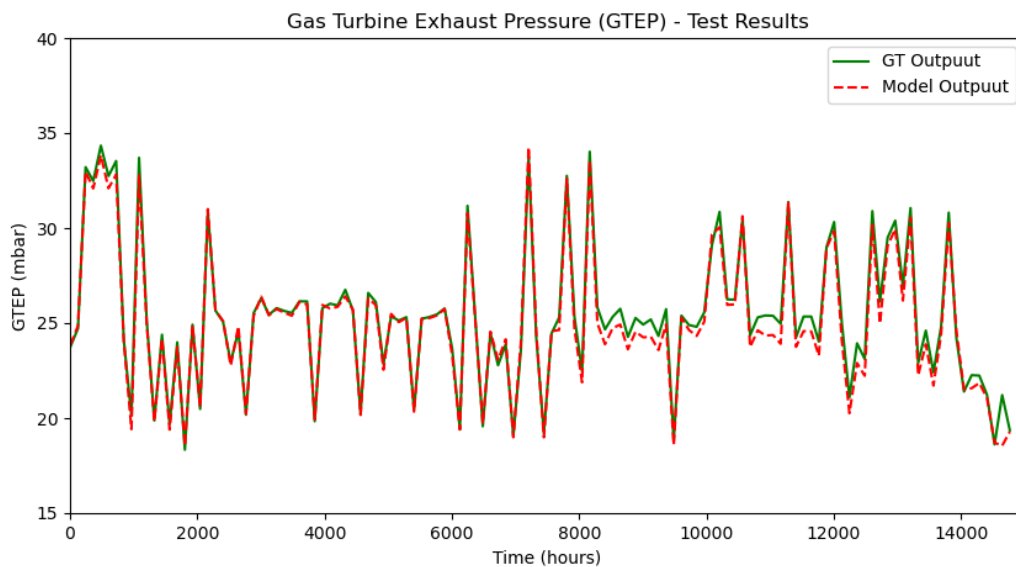


Figure 52. A Comparison Between Outputs of GT Engine and RNN Model for GT Exhaust Pressure (GTEP) for Test Datasets (Accuracy: 97%)

#### 4.6.2 RNN model: results for turbine inlet temperature (TIT)

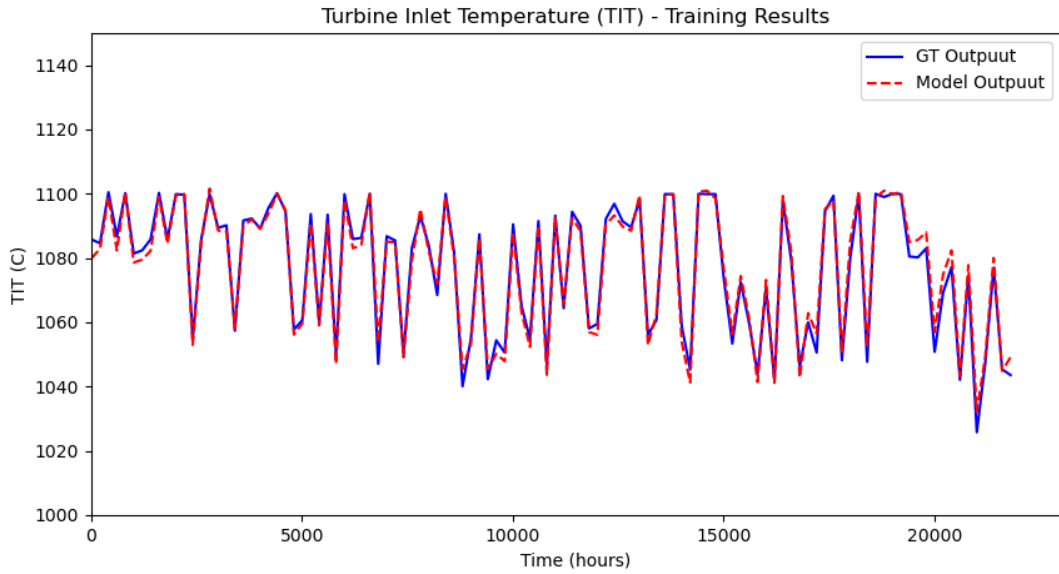


Figure 53. A Comparison Between Outputs of GT Engine and RNN Model for Turbine Inlet Temperature (TIT) for Training Datasets (Accuracy: 97%)

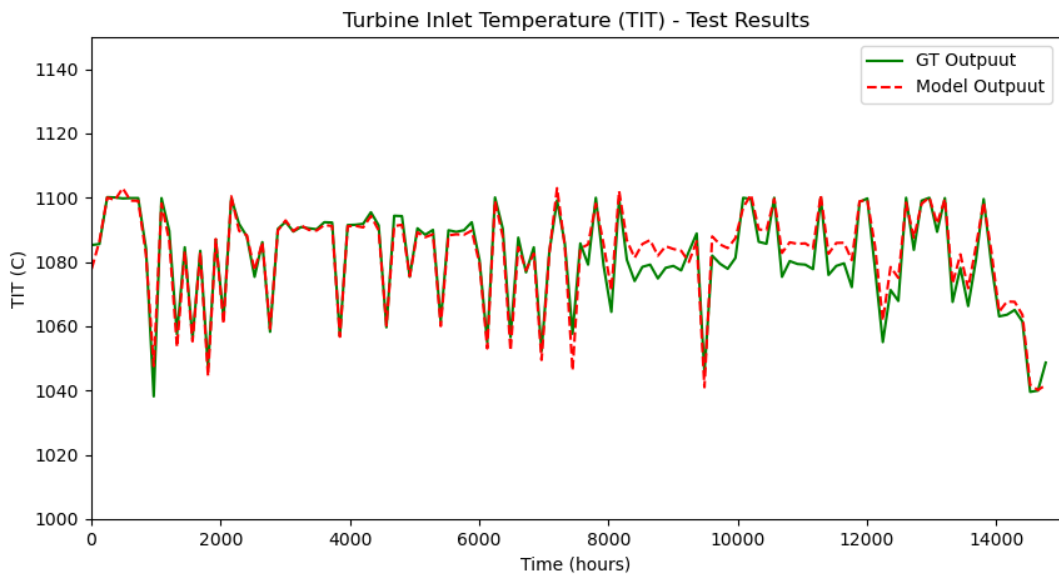


Figure 54. A Comparison Between Outputs of GT Engine and RNN Model for Turbine Inlet Temperature (TIT) for Test Datasets (Accuracy: 94%)



### 4.6.3 RNN model: results for turbine after temperature (TAT)

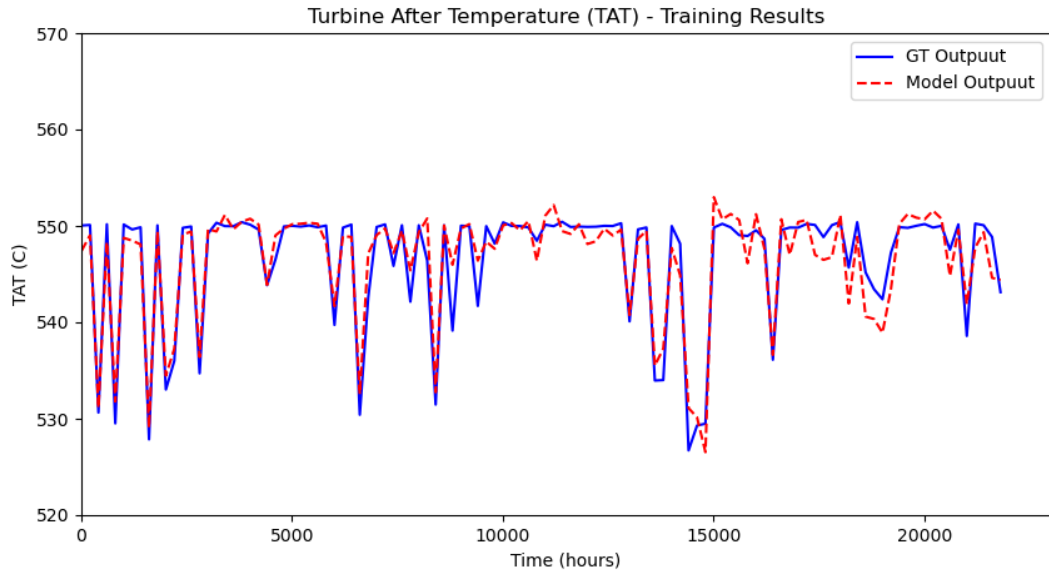


Figure 55. A Comparison Between Outputs of GT Engine and RNN Model for Turbine After Temperature (TAT) for Training Datasets (Accuracy: 90%)

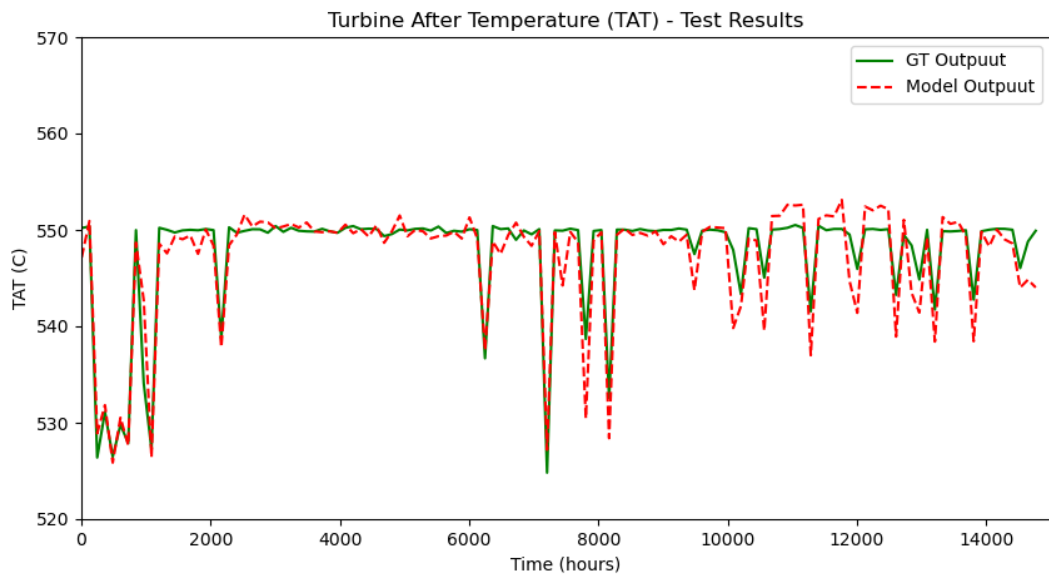


Figure 56. A Comparison Between Outputs of GT Engine and RNN Model for Turbine After Temperature (TAT) for Test Datasets (Accuracy: 81%)

#### 4.6.4 RNN model: results for compressor discharge pressure (CDP)

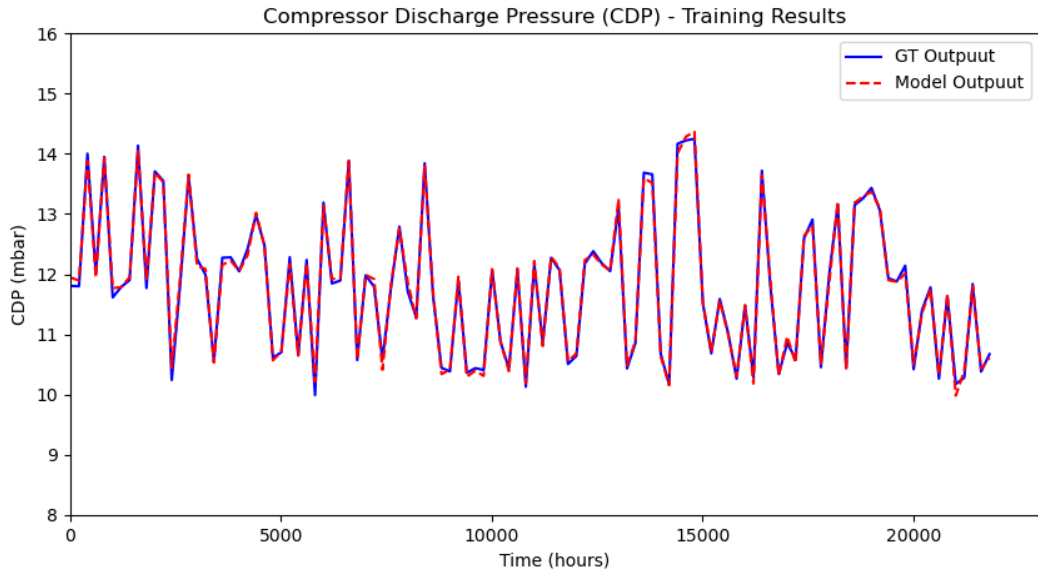


Figure 57. A Comparison Between Outputs of GT Engine and RNN Model for Compressor Discharge Pressure (CDP) for Training Datasets (Accuracy: 99%)

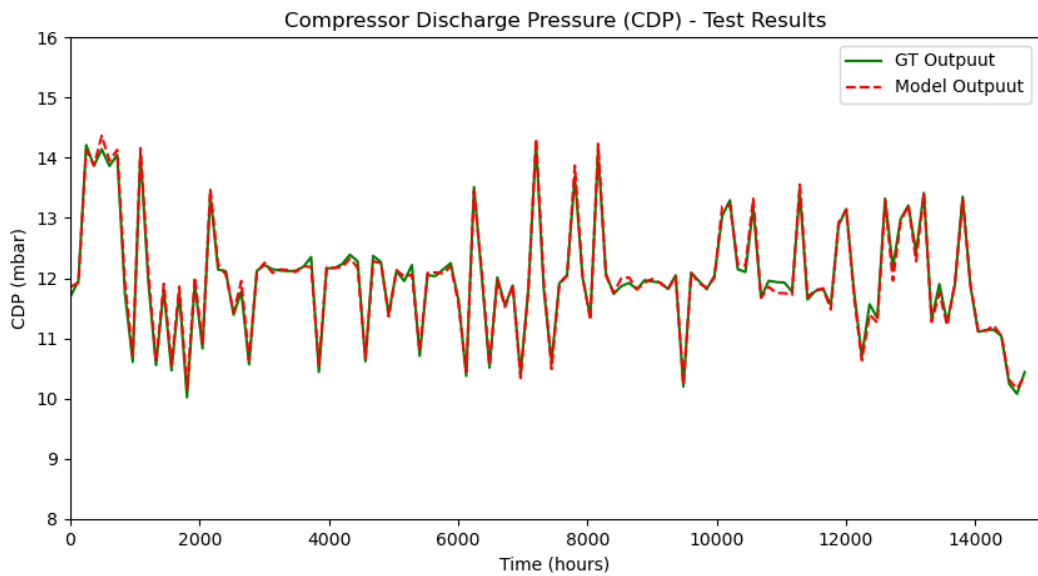


Figure 58. A Comparison Between Outputs of GT Engine and RNN Model for Compressor Discharge Pressure (CDP) for Test Datasets (Accuracy: 99%)

## 4.7 Comparison of the models

Table 4 shows and compares the values of  $R^2$  scores for four outputs of the *Ridge*, *Lasso*, *MTEN*, and *RNN* models. The scores were calculated for the both training and validation processes. The two last columns indicate the average values of the scores for the output *GTE* variables during the training and validation processes. Figure 59 presents a comparison among the four different models in terms of accuracy scores ( $R^2$ ) for the *GTE* output parameters.

Table 4. Comparison of Four Different Models in Terms of Accuracy Scores ( $R^2$ ) for *GTE* Output Parameters

$R^2$	GTEP		TIT		TAT		CDP		Average	
	Training	Validation	Training	Validation	Training	Validation	Training	Validation	Training	Validation
Ridge	0.942	0.958	0.907	0.890	0.552	0.462	0.991	0.988	0.85	0.82
Lasso	0.942	0.958	0.907	0.890	0.552	0.462	0.991	0.989	0.85	0.82
MTEN	0.941	0.962	0.907	0.890	0.539	0.453	0.990	0.988	0.86	0.84
RNN	0.979	0.973	0.973	0.937	0.903	0.818	0.995	0.992	0.963	0.930

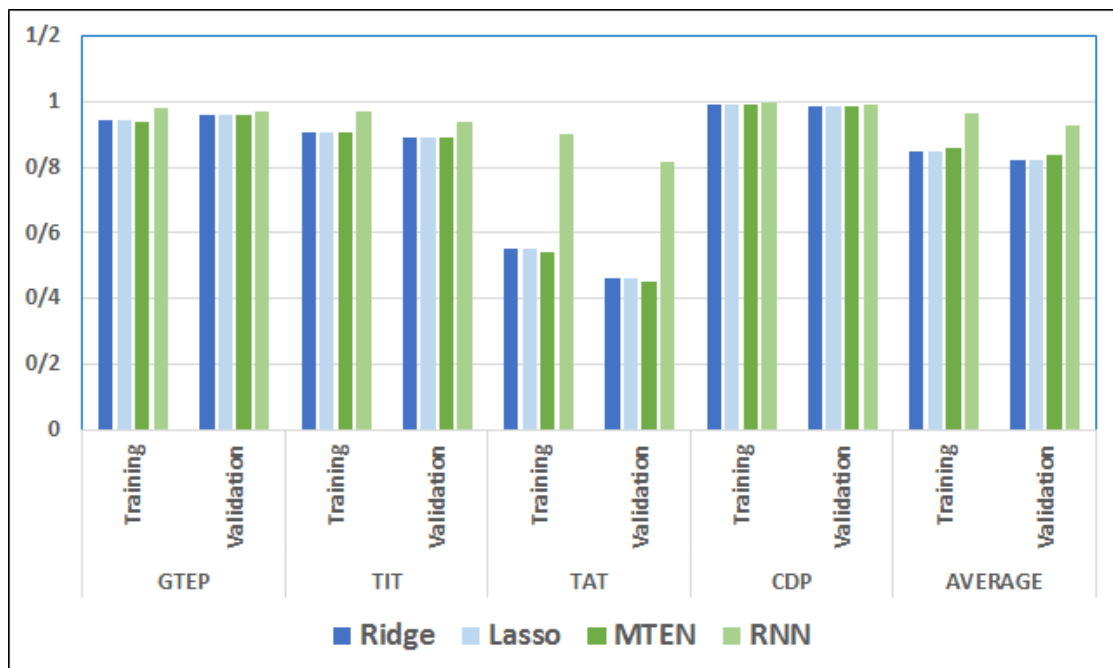


Figure 59. The Comparison Chart for Accuracy Scores ( $R^2$ ) of Different Models for *GTE* Output Parameters

## 4.8 Features' importance

Feature importance is defined as a technique, which assigns scores to the features in a predictive model, and indicates the relative importance of features in the prediction of output(s) [79]. These scores are calculated for cases that include prediction of numerical values (regressions), and also those that include prediction of class labels (classifications). The scores of feature importance can greatly help in predictive modelling problems for a better understanding of the dataset and model. They may also help in the reduction of the number of input features. They can improve the performance of a predictive model and speed up the modelling process by providing insight into the model and datasets, and highlighting the features that are most or least relevant to the target(s). Figure 60 illustrates the features' importance of the *GTE* calculated by using the decision tree model in *Python*. As this figure shows, gas turbine exhaust pressure (*GTEP*) has superior importance compared to other *GTE* parameters. After *GTEP*, ambient temperature (*AT*), and turbine inlet temperature (*TIT*) are the most important features of the engine. Turbine energy yield (*TEY*) has the lowest importance compared to other features.

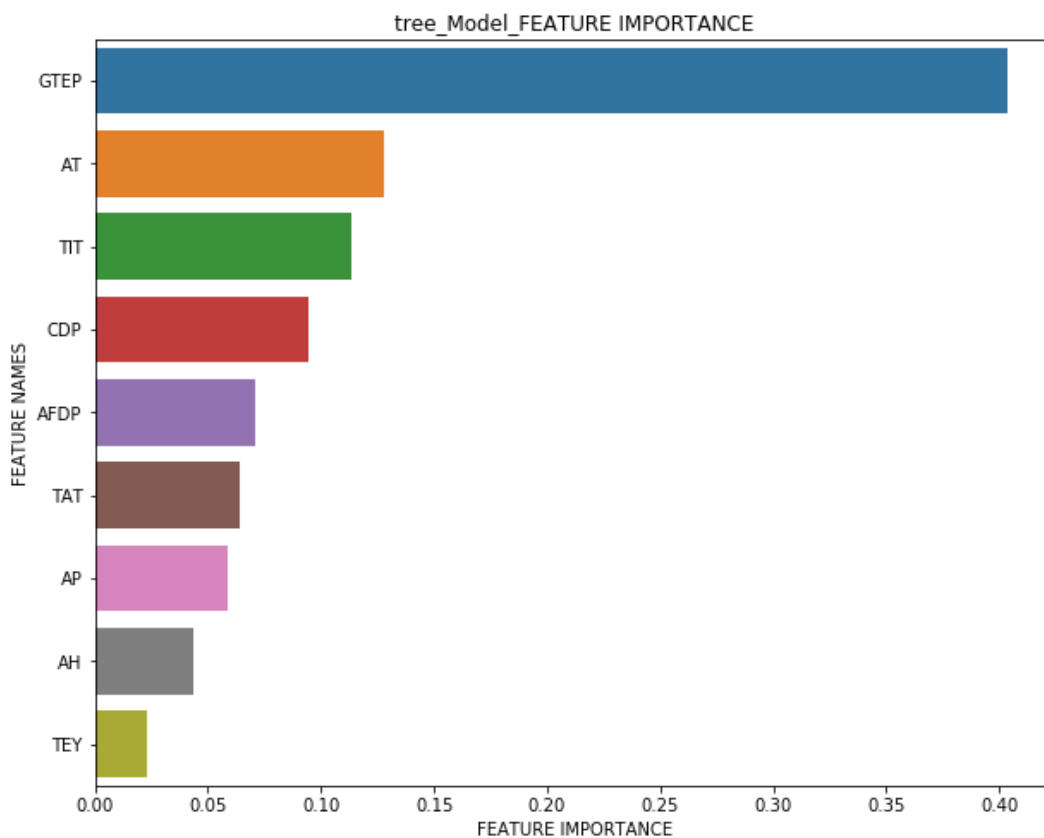


Figure 60. The Tree Model of Features's Importance for the Gas Turbine Engine

## 4.9 Summary

In this chapter, the results of training and validation processes for four linear and nonlinear models including *Ridge*, *Lasso*, *MTEN*, and *RNN* were presented. These models were trained, validated, and evaluated using the measured time-series datasets according to the procedure already discussed in the previous chapter. Before the modelling process started, to obtain the optimal model of the engine for the linear models, the significant training parameter *Alpha* was tuned for the linear models through the integration of the required code in the programming environment, and the results were presented. Finally, the features' importance of the *GTE* parameters was calculated and figured by using a decision tree model in *Python*.

## 5 DISCUSSION

This study investigated linear and nonlinear data-driven models of a gas turbine engine. These models consisted of *Ridge*, *Lasso*, and *Multi-Task Elastic-Net* models, which are based on linear regressions, and *RNN* model, which has a nonlinear structure. The data employed for this study were open-source measured time-series datasets of a single-shaft *GTE*. In this research, nine variables, which were directly related to the technical parameters of the *GTE*, were considered to model the system dynamics. Five of the features were considered as the system inputs, and the four remaining features were used as the system outputs. The five datasets were collected over five years from 2011 to 2015, totally including 36732 records. Three of these datasets were employed for training processes and the remains were used for validation purposes as described in the Methodology.

The datasets of the *GTE* were divided into training and validation sets. A comprehensive code was written in *Python* programming language and run in the *Jupyter Notebook* to train and set up three different linear models including *Ridge*, *Lasso*, and *Multi-Task Elastic-Net* models, and a nonlinear *RNN* model. The resulting models were validated against validation datasets. The training and validation results for the models were figured and the coefficients of determination ( $R^2$ ) for four output parameters of the engine were calculated for each of the models. These parameters consisted of gas turbine exhaust pressure (*GTEP*), turbine inlet temperature (*TIT*), turbine after temperature (*TAT*), and compressor discharge pressure (*CDP*).

### 5.1 Significant observations

The following significant observations may be extracted from the resulting models and figures:

- The average accuracy scores for the training of *Ridge*, *Lasso*, and *Multi-Task Elastic-Net* models are respectively 85%, 85%, and 86%, while the corresponding scores for the validation process are respectively 82%, 82%, and 84%. As the results indicate, the prediction of *Ridge*, *Lasso*, and *Multi-Task Elastic-Net* models of the dynamic behavior of the system are almost similar and satisfactory for three of four *GT* output parameters including *GTEP*, *TIT*, and *CDP*. For these

parameters, the outputs of the linear models followed the system outputs with acceptable accuracy from 91% to 99% for the training process, and from 91% to 99% for the validation process. However, the linear models failed to predict the dynamics of *TAT*, which may be called the critical parameter in this study. The accuracy scores for this parameter ranged from 54% to 55% for the training process, and from 45% to 46% for the validation process.

- The nonlinear *RNN*-based model predicted the system dynamics with remarkably higher accuracy compared to *Ridge*, *Lasso*, and *Multi-Task Elastic-Net* models. The average accuracy scores for training and validation of the *RNN* model are 96% and 93% respectively. The most important issue is the fact that the results are satisfactory for all of the output parameters of the engine including the critical parameter *TAT*. The  $R^2$  score of *TAT* for the training and validation processes remarkably increased to 90% and 82% respectively. It demonstrates the power of *RNN* as a nonlinear model for capturing the system dynamics.
- The result of calculations for the features' importance of the *GTE* shows that for the current datasets with a limited number of available parameters, and for the selected models of the engine, *GTEP* has superior importance compared to other available *GTE* parameters. After *GTEP*, ambient temperature (*AT*) and turbine inlet temperature (*TIT*) are the most important features of the engine. Turbine energy yield (*TEY*) has the lowest importance compared to other features.
- The conclusion demonstrates that the nonlinear *RNN* model has superior capability compared to the three linear models, in capturing the dynamics of a complex system like *GTE* with a high degree of nonlinearity.

## 5.2 Limitations

Although the number of datasets and the number of records for each of the datasets were quite enough to train and validate the models in this investigation, the datasets lacked some important features of the gas turbine such as rotational speed of the shaft, and fuel mass flow rate, which both are among the most important variables in gas turbine dynamics. The reason behind this is the fact that the purpose of the prior study [73], for which the data were collected, was to predict and minimize *CO* and *NO<sub>x</sub>* emissions from

the engine, rather than modelling the system dynamics. Lacking these significant features could be at least one of the reasons that the linear models were not capable of capturing *TAT* dynamics. Even, the *RNN* model hardly could achieve a relatively satisfactory prediction for *TAT*. One of the most important requirements for setting up an accurate data-driven model is to have and involve the effective variables of the system dynamics in the training and validation processes. It is also important that the datasets cover the whole operational range of the system, which fortunately was satisfied for this study.

### 5.3 Recommendations for future research

Although remarkable analytical and experimental models of gas turbine engines have been investigated and developed so far in order to deeply understand the complexity of the nonlinear nature of these engines' dynamics, there are still strong motivations for the students, scientists, manufacturers, and other professionals to continue their research activities in this fascinating area. Moreover, the electricity market has shown a high demand for the electricity produced by employing turbo-generators, especially for industrial applications. It has encouraged the power producers and other professionals in the field to look for novel methodologies for optimization of design and performance, and improving the manufacturing process of stationary gas turbines. It is especially because of the complexity of these systems and the need for developing models for a variety of objectives and applications with higher accuracies and reliabilities.

It would also be of great importance to researchers to explore a variety of approaches, techniques, and methodologies in the field with their own benefits and limitations. Both transient and steady-state operations of gas turbines under different environmental conditions, load fluctuation and system disturbances still need further investigations. Besides, other variables of the engine may be calculated by using the dynamic and thermodynamic equations for setting up a gray-box model. Moreover, a variety of linear and nonlinear models may be developed and compared for different types of GTEs under different operating conditions, and from different perspectives.



## 5.4 Conclusion

White-box models of gas turbines rely on dynamic, thermodynamic, and energy balance equations with a high degree of nonlinearity. To solve these complex and coupled equations, they should be simplified by making some assumptions and employing different linearization techniques, which consequently resulted in setting up models lacking the capability to precisely capture the whole dynamics of the system. Therefore, getting attention to the alternative solutions, which are independent of the system physics has been considerably increased. Among the different techniques, data-driven models have shown a high capability in capturing the nonlinear dynamics of GTEs. Besides, industrial equipment gradually deteriorates; losing efficiency and operability. Therefore, after years of service, the original dynamic and thermodynamic equations and the corresponding GT models need to be revised through theoretical and experimental processes, which are not easy in most cases. Moreover, replacement of the old engines requires huge financial sources and may not be economically possible. Fortunately, data-driven models can be very good alternatives in this situation, because of the flexibility and adaptability to the new ambient and operational conditions.

This study investigated four linear and nonlinear data-driven models of a gas turbine engine. These models consist of *Ridge*, *Lasso*, and *Multi-Task Elastic-Net* models, which are based on the linear regressions, and the *RNN* model with a nonlinear structure. As shown in this study, although setting up linear models is much easier and requires a shorter period of time compared to the nonlinear model, it is not as accurate and reliable as the nonlinear model. However, further investigations are still needed to evaluate and compare a variety of linear and nonlinear models of gas turbines. Data-driven models, such as *RNN* have shown a high potential and capability in capturing the nonlinear dynamics of complex systems, and may also be used for design and manufacturing optimization of *GTEs*, eventually led to saving a remarkable amount of money and time.

## REFERENCES

- [1] G. G. Kulikov and H. A. Thompson, Eds., *Dynamic Modelling of Gas Turbines*, Springer London Ltd, 2010, p. 336 pp.
- [2] M. P. Boyce, *Gas Turbine Engineering Handbook*, 4th ed., Oxford: Butterworth-Heinemann Ltd, ELSEVIER Science & Technology, 2011, p. 1000 pp.
- [3] J. W. Sawyer, Ed., *Sawyer's Gas Turbine Engineering Handbook*, Gas Turbine Publications, 1966, p. 552 pp.
- [4] R. T. C. Harman, *Gas Turbine Engineering: Applications, Cycles and Characteristics*, Macmillan, 1981, p. 270 pp.
- [5] A. M. Y. Razak, *Industrial Gas Turbines: Performance and Operability*, First ed., Cambridge, England: Woodhead Publishing Limited, 2007, p. 602 pp.
- [6] M. Jelali and A. Kroll, *Hydraulic Servo-Systems: Modelling, Identification, and Control*, Springer, 2004, p. 355 pp.
- [7] L. Ljung and T. Glad, *Modelling of Dynamic Systems*, Englewood Cliffs, New Jersey: PTR Prentice Hall, Prentice Hall Information and System Sciences Series, 1994, p. 361 pp.
- [8] M. Norgaard, O. Ravn, N. K. Poulsen and L. K. Hansen, *Neural Networks for Modelling and Control of Dynamic Systems: A Practitioner's Handbook*, Springer, 2003, p. 246 pp.
- [9] "Wikimedia Commons Webpage," 2013. [Online]. Available: [http://en.wikipedia.org/wiki/File:Brayton\\_cycle.svg](http://en.wikipedia.org/wiki/File:Brayton_cycle.svg). [Accessed 07 04 2021].
- [10] H. Cohen, G. Rogers and H. I. H. Saravanamut, *Gas Turbine Theory*, 5th ed., Pearson Education, 2001.
- [11] "Savoia Generators," *Advanced Solutions for Power Generation*, [Online]. Available: <http://savoiapower.com/gas-turbines.html>. [Accessed 07 04 2021].
- [12] H. Asgari, X. Q. Chen and R. Sainudiin, "Considerations in Modelling and Control of Gas Turbines - A Review," in *The 2nd International Conference on Control, Instrumentation, and Automation (ICCIA 2011)*, Shiraz, Iran, 2011. doi: 10.1109/ICCIAutom.2011.6356635.
- [13] H. Asgari, X. Q. Chen and R. Sainudiin, "Modelling and Simulation Approaches for Gas Turbine System Optimization," in *Engineering Creative Design in Robotics and Mechatronics*, Hershey, PA: IGI Global, 2013. doi: 10.4018/978-1-4666-4225-6.ch014, pp. 240-264.
- [14] G. V. Beard and M. R. Rollins, "Engineering Economics," in *Power Plant Engineering*, & V. Black, Ed., Springer Science & Business Media Inc., 1996.
- [15] H. Asgari, X. Q. Chen, M. Morini, M. Pinelli, R. Sainudiin, P. R. Spina and M. Venturini, "NARX Models for Simulation of the Start-Up Operation of a Single-Shaft Gas Turbine," *Applied Thermal Engineering*, vol. 93, p. 368–376, 2016. doi: 10.1016/j.applther.
- [16] M. A. Ibrahim, O. Akhrif, H. Moustapha and M. Staniszewski, "Neural Networks Modelling of Aero-derivative Gas Turbine Engine: A Comparison Study," in *Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics (ICINCO)*, Prague, Czech Republic, 2019. doi: 10.5220/0007928907380745.

- [17] O. Mohamed and A. Khalil, "Progress in Modeling and Control of Gas Turbine Power Generation Systems: A Survey," *Energies*, vol. 13, no. 9, p. 2358 (26 pages), 2020. doi: 10.3390/en13092358.
- [18] W. P. J. Visser, O. Kogenhop and M. Oostveen, "A Generic Approach for Gas Turbine Adaptive Modelling," *ASME Journal of Engineering for Gas Turbines and Power*, vol. 128, no. 1, pp. 13-19, 2006.
- [19] H. Nikpey, M. Assadi and P. Breuhaus, "Development of an Optimized Artificial Neural Network Model for Combined Heat and Power Micro Gas Turbines," *Applied Energy*, vol. 108, p. 137–148, 2013. doi: 10.1016/j.apenergy.2013.03.016.
- [20] A. Vatani, K. Khorasani and N. Meskin, "Health Monitoring and Degradation Prognostics in Gas Turbine Engines Using Dynamic Neural Networks," in *Volume 6: Ceramics; Controls, Diagnostics and Instrumentation; Education; Manufacturing Materials and Metallurgy; Honors and Awards*, Montreal, Canada, 2015. doi: 10.1115/gt2015-44101.
- [21] M. H. M. Tarik, M. Omar, M. F. Abdullah and R. Ibrahim, "Modelling of Dry Low Emission Gas Turbine Using Black-Box Approach," in *TENCON 2017 - IEEE Region 10 Conference*, Penang, 2017. doi: 10.1109/tencon.2017.8228169.
- [22] C. Bartolini, F. Caresana, G. Comodi, L. Pelag, M. Renzi and S. Vagni, "Application of Artificial Neural Networks to Micro Gas Turbines," *Energy Conversion and Management*, vol. 52, no. 1, pp. 781-788, 2011. doi: 10.1016/j.enconman.2010.08.003.
- [23] F. Jurado, "Non-Linear Modeling of Micro-Turbines Using NARX Structures on the Distribution Feeder," *Energy Conversion and Management*, vol. 46, no. 3, pp. 385-401, 2005.
- [24] A. Lazzaretto and A. Toffolo, "Analytical and Neural Network Models for Gas Turbine Design and Off-Design Simulation," *International Journal of Applied Thermodynamics*, vol. 4, no. 4, pp. 173-182, 2001.
- [25] M. Amozegar and K. Khorasani, "An Ensemble of Dynamic Neural Network Identifiers for Fault Detection and Isolation of Gas Turbine Engines," *Neural Networks*, vol. 76, p. 106–121, 2016. doi: 10.1016/j.neunet.2016.01.003.
- [26] S. Ogaji, R. Singh and S. D. Probert, "Multiple-Sensor Fault-Diagnoses for a 2-Shaft Stationary Gas Turbine," *Applied Energy*, vol. 71, no. 4, pp. 321-339, 2002. doi: 10.1016/S0306-2619(02)00015-6.
- [27] J. Arriagada, M. Genrup, A. Loberg and M. Assadi, "Fault Diagnosis System for an Industrial Gas Turbine by Means of Neural Networks," in *International Gas Turbine Congress*, Tokyo, 2003.
- [28] A. Mehrpanahi, A. Hamidavi and A. Ghorbanifar, "A Novel Dynamic Modeling of an Industrial Gas Turbine Using Condition Monitoring Data," *Applied Thermal Engineering*, vol. 143, p. 507–520, 2018. doi: 10.1016/j.applthermaleng.2018.07.081.
- [29] M. Bai, J. Liu, J. Chai, X. Zhao and D. Yu, "Anomaly Detection of Gas Turbines Based on Normal Pattern Extraction," *Applied Thermal Engineering*, vol. 166, p. 114664 (18 pages), 2020. doi: 10.1016/j.applthermaleng.2019.114664.

- [30] M. Basso, L. Giarre, S. Groppi and G. Zappa, "NARX Models of an Industrial Power Plant Gas Turbine," *IEEE Transactions on Control Systems Technology*, vol. 13, no. 4, pp. 599 - 604, 2005. doi: 10.1109/TCST.2004.843129.
- [31] R. Bettocchi, M. Pinelli, P. R. Spina and M. Burgio, "Set Up of a Robust Neural Network for Gas Turbine Simulation," in *ASME Turbo Expo 2004*, Vienna, Austria, 2004.
- [32] R. Bettocchi, M. Pinelli, P. Spina and M. Venturini, "Artificial Intelligence for the Diagnostics of Gas Turbines - Part I: Neural Network Approach," *ASME Journal of Engineering for Gas Turbines and Power*, vol. 129, no. 3, pp. 711-719, 2007. doi: 10.1115/1.2431391.
- [33] S. Simani and R. J. Patton, "Fault Diagnosis of an Industrial Gas Turbine Prototype Using a System Identification Approach," *Control Engineering Practice*, vol. 16, no. 7, pp. 769-786, 2008. doi: 10.1016/j.conengprac.2007.08.009.
- [34] N. Chiras, C. Evans and D. Rees, "Nonlinear Gas Turbine Modelling Using NARMAX Structures," *IEEE Transactions on Instrumentation and Measurement*, vol. 50, no. 4, pp. 893-898, 2001. doi: 10.1109/19.948295.
- [35] N. Chiras, C. Evans and D. Rees, "Nonlinear Modelling and Validation of an Aircraft Gas Turbine Engine," in *5th IFAC Symposium on Nonlinear Control Systems*, ST Petersburg, Russia, 2002.
- [36] N. Chiras, C. Evans and D. Rees, *Nonlinear Gas Turbine Modelling Using Feedforward Neural Networks*, Wales: School of Electronics, University of Glamorgan, 2002.
- [37] N. Chiras, C. Evans and D. Rees, "Global Nonlinear Modeling of Gas Turbine Dynamics Using NARMAX Structures," *ASME Journal of Engineering for Gas Turbines and Power*, vol. 124, no. 4, pp. 817-826, 2002. doi: 10.1115/1.1470483.
- [38] M. Fast, M. Assadi and D. Sudipta, "Condition Based Maintenance of Gas Turbines Using Simulation Data and Artificial Neural Network: A Demonstration of Feasibility," in *ASME Turbo Expo*, 2008.
- [39] M. Fast, M. Assadi and S. De, "Development and Multi-Utility of an ANN Model for an Industrial Gas Turbine," *Journal of Applied Energy*, vol. 86, no. 1, pp. 9-17, 2009. doi: 10.1016/j.apenergy.2008.03.018.
- [40] M. Fast, T. Palmé and M. Genrup, "A Novel Approach for Gas Turbines Monitoring Combining CUSUM Technique and Artificial Neural Network," in *ASME Turbo Expo 2009*, Orlando, Florida, USA, 2009.
- [41] M. Fast and T. Palmé, "Application of Artificial Neural Network to the Condition Monitoring and Diagnosis of a Combined Heat and Power Plant," *Energy*, vol. 35, no. 2, pp. 1114-1120, 2010. doi: 10.1016/j.energy.2009.06.005.
- [42] M. Fast, *Artificial Neural Networks for Gas Turbine Monitoring*, Lund: Division of Thermal Power Engineering, Department of Energy Sciences, Faculty of Engineering, Lund University, 2010.
- [43] G. Torella, F. Gamma and G. Palmesano, "Neural Networks for the Study of Gas Turbine Engines Air System," in *International Gas Turbine Congress*, Tokyo, Japan, 2003.
- [44] M. Salehi and A. Montazeri, "Black Box Modeling of a Turboshaft Gas Turbine Engine Fuel Control Unit Based on Neural NARX.," *Journal of Engineering for the Maritime Environment*, vol. 233, no. 3, pp. 949-956, 2018. doi: 10.1177/1475090218797496.

- [45] H. A. Nozari, H. D. Banadaki, M. A. Shoorehdeli and S. Simani, "Model-based Fault Detection and Isolation using Neural Networks: An Industrial Gas Turbine Case Study," in *21st International Conference on Systems Engineering (ICSEng)*, Las Vegas, NV, USA, 2011. doi: 10.1109/ICSEng.2011.13.
- [46] H. A. Nozaria, M. A. Shoorehdeli, S. Simani and H. D. Banadaki, "Model-Based Robust Fault Detection and Isolation of an Industrial Gas Turbine Prototype Using Soft Computing Techniques," *Neurocomputing*, vol. 91, no. 8, p. 29–47, 2012. doi: 10.1016/j.neucom.2012.02.014.
- [47] A. E. Ruano, P. J. Fleming, C. Teixeira, K. Rodriguez-Vazquez and C. M. Fonseca, "Nonlinear Identification of Aircraft Gas-Turbine Dynamics," *NEUROCOMPUTING*, vol. 55, no. 3-4, pp. 551-579, 2003. doi: 10.1016/S0925-2312(03)00393-X.
- [48] N. Dodd and J. Martin, "Using Neural Networks to Optimize Gas Turbine Aero Engines," *Computing & Control Engineering*, vol. 8, no. 3, pp. 129-135, 1997. doi: 10.1049/cce:19970305.
- [49] T. V. Breikin, G. G. Kulikov, P. J. F. Arkov and P. J. Fleming, "Dynamic Modelling for Condition Monitoring of Gas Turbines: Genetic Algorithms Approach," in *16th IFAC World Congress*, Czech Republic, 2005. doi: 10.3182/20050703-6-CZ-1902.00124.
- [50] M. B. Rahmoune, A. Hafaifa, A. Kouzou, M. Guemana and S. Abudura, "Control and Diagnostic of Vibration in Gas Turbine System Using Neural Network Approach," in *8th International Conference on Modelling, Identification and Control (ICMIC)*, Algiers, 2016. doi: 10.1109/icmic.2016.7804177.
- [51] M. B. Rahmoune, A. Hafaifa and G. Mouloud, "Fault Diagnosis in Gas Turbine Based on Neural Networks: Vibrations Speed Application.," *Advances in Acoustics and Vibration*, vol. 5, pp. 1-11, 2017. doi: 10.1007/978-3-319-41459-1\_1.
- [52] W. H. Elashmawi, N. A. Kotp and G. E. Tawel, "Neural Network Monitoring Model for Industrial Gas Turbine," *Asian Journal of Applied Sciences*, vol. 5, no. 3, pp. 608-614, 2017.
- [53] W. H. Elashmawi, "A Novel Proposed Neural Network MAD (Monitoring, Analysis and Diagnose) Model for Industrial Gas Turbine," *International Journal of Soft Computing*, vol. 13, no. 3, pp. 92-101, 2018. doi: 10.36478/ijscmp.2018.92.101.
- [54] H. Badihi, A. Shahriari and A. Naghsh, "Artificial Neural Network Application to Fuel Flow Function for Demanded Jet Engine Performance," in *Proceedings of IEEE Aerospace Conference*, Mont, USA, 2009. doi: 10.1109/AERO.2009.4839656.
- [55] Y. Yoru, T. H. Karakoc and A. Hepbasli, "Application of Artificial Neural Network (ANN) Method to Exergetic Analyses of Gas Turbines," in *International Symposium on Heat Transfer in Gas Turbine Systems*, Antalya, Turkey, 2009. doi: 10.1615/ICHMT.2009.HeatTransfGasTurbSyst.580.
- [56] I. Loboda, Y. Feldshteyn and V. Pon, "Neural Networks for Gas Turbine Fault Identification: Multilayer Perceptron or Radial Basis Network?," *International Journal of Turbo and Jet-Engines*, vol. 29, no. 1, p. 37–48, 2012. doi: 10.1515/tjj-2012-0005.
- [57] B. Yu, W. Shu and C. Cao, "A Novel Modeling Method for Aircraft Engine Using Nonlinear Autoregressive Exogenous (NARX) Models Based on Wavelet Neural Networks," *International Journal of Turbo & Jet-Engines*, vol. 35, p. 161–169, 2018. doi: 10.1515/tjj-2017.

- [58] S. S. Tayarani-Bathaie, Z. N. S. Vanini and k. Khorasani, "Fault Detection of Gas Turbine Engines Using Dynamic Neural Networks," in *25th IEEE Canadian Conference on Electrical & Computer Engineering (CCECE)*, Montreal, QC, 2012. doi: 10.1109/CCECE.2012.6334837.
- [59] H. Asgari, X. Q. Chen and R. Sainudiin, "Analysis of ANN-Based Modelling Approach for Industrial Applications," *International Journal of Innovation, Management and Technology (IJIMT)*, vol. 4, no. 1, pp. 165-169, 2013. doi: 10.7763/IJIMT.2013.V4.383.
- [60] H. Asgari, X. Q. Chen and R. Sainudiin, "Applications of Artificial Neural Networks to Rotating Equipment," in *3rd Conference on Rotating Equipment in Oil and Power Industries*, Tehran, Iran, 2012.
- [61] H. Asgari, X. Q. Chen, M. B. Menhaj and R. Sainudiin, "ANN-Based System Identification, Modelling, and Control of Gas Turbines – A Review," *Manufacturing Science and Technology III, Advanced Materials Research*, Vols. 622-623, pp. 611-617, 2013. doi: 10.4028/www.scientific.net/AMR.622-623.611.
- [62] H. Asgari, X. Q. Chen, M. B. Menhaj and R. Sainudiin, "Artificial Neural Network–Based System Identification for a Single-Shaft Gas Turbine," *ASME Journal of Engineering for Gas Turbines and Power*, vol. 135, no. 9, pp. 092601-7, September 2013.
- [63] H. Asgari, x. Q. Chen, R. Sainudiin, M. Morini, M. Pinelli, P. R. Spina and M. Venturini, "Modeling and Simulation of the Startup Operation of a Heavy-Duty Gas Turbine Using NARX Models," in *Turbo Expo 2014*, Düsseldorf, Germany, 2014.
- [64] J. Mu and D. D. Rees, "Approximate Model Predictive Control for Gas Turbine Engines," in *American Control Conference*, 2004.
- [65] J. H. Kim, T. W. Song, T. Kim and S. T. Ro, "Model Development and Simulation of Transient Behavior of Heavy Duty Gas Turbines," *ASME Journal of Engineering for Gas Turbines and Power*, vol. 123, no. 3, pp. 589-594, 2001. doi: 10.1115/1.1370973.
- [66] J. H. Kim, T. W. Song, T. S. Kim and S. T. Ro, "Dynamic Simulation of Full Startup Procedure of Heavy-Duty Gas Turbines," *ASME Journal of Engineering for Gas Turbines and Power*, vol. 124, no. 3, pp. 510-516, 2002. doi: 10.1115/1.1473150.
- [67] M. R. Bank Tavakoli, B. Vahidi and W. Gawlik, "An Educational Guide to Extract the Parameters of Heavy-Duty Gas Turbines Model in Dynamic Studies Based on Operational Data," *IEEE Transactions on Power Systems*, vol. 24, no. 3, pp. 1366-1374, 2009.
- [68] T. Palmé, M. Fast and M. Thern, "Gas Turbine Sensor Validation Through Classification with Artificial Neural Networks," *Applied Energy*, vol. 88, no. 11, p. 3898–3904, 2011.
- [69] P. Spina and M. Venturini, "Gas Turbine Modelling by Using Neural Networks Trained on Field Operating Data," in *ECOS 2007*, Padova, Italy, 2007.
- [70] R. Mohammadi, E. Naderi, K. Khorasani and S. Hashtrudi-Zad, "Fault Diagnosis of Gas Turbine Engines by Using Dynamic Neural Networks," in *IEEE International Conference on Quality and Reliability (ICQR)*, Bangkok, Thailand, 2011. doi: 10.1109/ICQR.2011.6031675.
- [71] P. Ailer, "Comparison of Linear and Non-linear Mathematical Models Developed for Gas Turbine Control," in *Proceedings of 3rd International Conference on Nonlinear Problems in Aviation and Aerospace (ICNPAA)*, Daytona Beach, Florida, USA, 2000.
- [72] D. Henrion, L. Reberga, J. Bernussou and F. Vary, *Linearization and Identification of Aircraft Turbofan Engine*, LAAS-CNRS and SNECMA Moteurs, France: IFAC, 2004.

- [73] H. Kaya, P. Tufekci and E. Uzun, "Predicting CO and NOx Emissions from Gas Turbines: Novel Data and a Benchmark PEMS," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 27, p. 4783 – 4796, 2019. doi: 10.3906/elk-1807-87.
- [74] "Jupyter," [Online]. Available: <https://jupyter.org>. [Accessed 7 04 2021].
- [75] "Scikit-Learn," [Online]. Available: <https://scikit-learn.org/stable/>. [Accessed 07 04 2021].
- [76] V. Sharma, "Vinod Sharma's Blog," 01 02 2019. [Online]. Available: <https://vinodsblog.com.29.10.2020..> [Accessed 07 04 2021].
- [77] "MathWorks," MathWorks, [Online]. Available: <https://www.mathworks.com/help/deeplearning/ug/design-time-series-narx-feedback-neural-networks.html>. [Accessed 24 04 2021].
- [78] A. Szczepanek, "Omni Calculator," 08 05 2020. [Online]. Available: [omnicalculator.com/statistics/coefficient-of-determination](http://omnicalculator.com/statistics/coefficient-of-determination). [Accessed 24 04 2021].
- [79] J. Brownlee, "Machine Learning Mastery," 30 03 2020. [Online]. Available: <https://machinelearningmastery.com/calculate-feature-importance-with-python/>. [Accessed 07 04 2021].
- [80] M. T. Hagan, H. B. Demuth and O. D. JESÚS, "An Introduction to the Use of Neural Networks in Control Systems," *International Journal of Robust and Nonlinear Control*, vol. 12, no. 11, p. 959–985, 2002. doi: 10.1002/rnc.727.
- [81] M. T. Hagan and H. B. Demuth, "Neural Networks for Control," in *American Control Conference*, San Diego, CA, USA, 1999. doi: 10.1109/ACC.1999.786109.

