

Data mining for diagnosis, monitoring, and prediction in wind power plants

by

Mostafa SAEIDI

THESIS PRESENTED TO ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
IN PARTIAL FULFILLMENT FOR A MASTER'S DEGREE
WITH THESIS IN ELECTRICAL ENGINEERING
M.A.Sc.

MONTREAL, 14TH AUGUST 2020

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

© Copyright 2020, Mostafa Saeidi, All right reserved

© Copyright reserved

It is forbidden to reproduce, save or share the content of this document either in whole or in parts. The reader who wishes to print or save this document on any media must first get the permission of the author.

BOARD OF EXAMINERS (THESIS M.Sc.A.)
THIS THESIS HAS BEEN EVALUATED
BY THE FOLLOWING BOARD OF EXAMINERS

Mr. Ambrish Chandra, Thesis Supervisor
Département de génie électrique at École de technologie supérieure

Mr. Zhaoheng Liu, President of the Board of Examiners
Département de Génie Mécanique at École de technologie supérieure

Mr. Pierre Jean Lagacé, Member of the jury
Département de génie électrique at École de technologie supérieure

Mr. Philippe Cambron, External Evaluator
Power Factors

THIS THESIS WAS PRESENTED AND DEFENDED
IN THE PRESENCE OF A BOARD OF EXAMINERS AND PUBLIC
<DEFENCE DATE OF THE THESIS>
AT ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

ACKNOWLEDGMENT

It is a pleasure to thank the many people who have helped and inspired me during my M.A.Sc. study. First of all, I would like to express my deep and sincere gratitude to my supervisor, Ambrish Chandra. His understanding, encouraging and personal guidance have provided a fundamental basis for the present thesis. His patience, guidance, encouragement, and useful critiques through this research work helped me enormously. Many thanks also to Dariush Faghani, Phippe Cambron, and Francis Pelletier from Power Factors. Their detailed and constructive advices helped me in the development of my team working and professionalism skills and this work. I would like to thank the members of my examining committee: Pierre Jean Lagacé, Philippe Cambron, and Zhaoheng Liu, who have agreed in evaluating this thesis and provided me valuable comments. Thanks also to my many colleagues in the GREPCI (le Groupe de recherche en électronique de puissance et commande industrielle) who provided a stimulating and friendly environment to learn and grow. Also, I would like to give my special thanks to my entire family and parents, Mohammadreza and Farahnaz for providing me with unfailing support and continuous encouragement throughout my years of study. Most of all, thanks to my dearest friend, Pardis for her understanding, endless love, and encouragement which were fundamental for me during my study. They have always supported, inspired, and loved me. To all of them I dedicate this thesis. I also thank National Sciences and Engineering Research Council NSERC of Canada, and École de Technologie Supérieure for their financial support. This work would not have been possible without their support.

Extraction de données pour le diagnostic, la surveillance et la prévision dans les centrales éoliennes

Mostafa SAEIDI

RÉSUMÉ

De nos jours, les sources d'énergie renouvelables jouent un rôle clé dans l'ingénierie des systèmes électriques. Parmi la multiplicité des sources d'énergie, l'énergie éolienne est plus populaire en raison de ses avantages substantiels tels que le bas coût de l'énergie produite, sa technologie bien établie et le fait qu'elle soit plus accessible. Cependant, ce type de technologies a des coûts considérables qui peuvent réduire les bénéfices des investisseurs et des consommateurs. Parmi les différents coûts, les coûts d'exploitation et de maintenance jouent un rôle clé, car de nombreuses éoliennes installées vieillissent. Pour réduire les coûts d'exploitation et d'entretien, il est essentiel d'actualiser les stratégies définies pour la surveillance de l'état dans les parcs éoliens. L'outil de surveillance de l'état approprié doit être développé après avoir effectué une étude complète des techniques de surveillances de l'état existantes. Pour passer d'une surveillance de l'état manuelle à une approche automatique et intelligente, les concepts d'apprentissage automatique et d'exploration de données seront utilisés dans cette application. La nécessité d'une stratégie moderne de surveillance de l'état coïncide avec l'émergence de l'apprentissage en profondeur au cours de la dernière décennie. Une nouvelle surveillance de l'état est proposée dans cette thèse, afin de surveiller la température des enroulements des générateurs dans un parc éolien, en utilisant le modèle de mémoire à court terme qui est une méthode de réseau de neurones récurrents. Ce modèle d'apprentissage en profondeur est un choix pratique et unique pour la prédiction en raison de sa capacité à prendre en compte les dépendances à long terme entre les caractéristiques séquentielles d'entrée qui, dans ce cas, sont l'ensemble des données chronologiques fournies par le système SCADA. Les données chronologiques de SCADA sont fournies par Power Factors qui dispose d'une vaste base de données d'informations sur les performances des différentes centrales électriques, qui a été rassemblée par un système SCADA à une fréquence de 1 Hz. Le nouveau modèle proposé dans cette thèse est évalué à partir d'un parc éolien dans la province de Québec (Canada), où les scénarios de cas seront des conditions de fonctionnement sain et anormal. De plus, le signal d'erreur est calculé entre les signaux réels et prédits pour évaluer la bonne performance du modèle. Enfin, ce modèle est comparé avec le modèle de régression linéaire multiple pour montrer l'efficacité et la grande précision du modèle proposé.

Mots-clés: surveillance de l'état, mémoire à court terme, SCADA, apprentissage en profondeur, température des enroulements du générateur.

Data mining for diagnosis, monitoring, and prediction in wind power plants

Mostafa SAEIDI

ABSTRACT

Nowadays, renewable energy sources play the key role in power system engineering. Among multiplicity of energy sources, wind energy is more popular due to its substantial advantages such as lower prices of produced electricity, their established technology, and being more accessible. However, these technologies have considerable costs that can mitigate profit for the investors and consumers. Among the different costs, the operation and maintenance (O&M) costs have the key role since many installed wind turbines are getting older. To mitigate the O&M costs it is crucial to update the defined strategies for condition monitoring (CM) in wind farms. The proper CM tool should be developed after exerting complete survey of existing CM techniques. To shift from a manual CM toward an automatic and smart approach, the machine learning and data mining concepts will be utilized in this application. The need for a modern condition monitoring strategy coincides with the emergence of Deep Learning (DL) in the recent decade. In this thesis, a novel condition monitoring is proposed to monitor the temperature of generator windings in a wind farm using Long Short-Term Memory (LSTM) model which is a Recurrent Neural Network (RNN) method. This DL model is a practical and unique choice for prediction due to its ability for considering long-term dependencies among input sequential features, which in this case are the provided time-series datasets by SCADA system. The time-series SCADA datasets are provided by Power Factors which has a vast database of information about performance of different power plants which is gathered by SCADA system at 1Hz frequency. The novel proposed model in this thesis is evaluated in healthy and abnormal operation modes of a wind farm in Quebec Province, Canada as case scenarios. Moreover, the error signal is calculated between real and predicted signals to evaluate the proper performance of the model. Finally, this model is compared with Multiple Linear Regression model to show the effectiveness and high precision of the proposed model.

Keywords: condition monitoring, long short-term memory, SCADA, deep learning, generator windings temperature

TABLE OF CONTENTS

	Page
INTRODUCTION	16
CHAPTER 1 Problem Description	19
1.1 Supervisory Control and Data Acquisition (SCADA) systems	20
1.2 Condition monitoring challenges of wind power plants	21
1.3 Objectives of project	23
1.4 Research Contribution	24
1.5 Methodology	25
1.5.1 A thorough literature review	25
1.5.2 Refining raw data	26
1.5.3 Pre-interpretation of data based on grid codes	26
1.5.4 ML usage for interpretation of data	27
CHAPTER 2 Literature Review	28
2.1 Common failures in main equipment	28
2.1.1 Blades	29
2.1.2 Gearbox	29
2.1.3 Main Shaft	29
2.1.4 Power Converters	30
2.1.5 Electric Machine	30
2.2 Spectral Analysis	31
2.3 Signal Trending	35
2.4 Physical Model	36
2.5 ML and signal processing techniques	38
CHAPTER 3 Proposed Data Preprocessing Technique and Regression Model	41
3.1 Correlation Analysis	41
3.2 Multiple Linear Regression (MLR)	42
3.3 Artificial Neural Network (ANN)	44
3.3.1 Activation Function	45
3.3.1.1 Binary Step Function	45
3.3.1.2 Linear Activation Function	46
3.3.1.3 Sigmoid Function	47
3.3.1.4 Hyperbolic Tangent Function	48
3.3.2 Optimizer	49
3.3.3 Loss Function	50
3.3.4 Learning Rate	52
3.3.5 Dropout	52
3.4 Recurrent Neural Network (RNN)	54
3.4.1 Vanishing Gradient Challenge:	55
3.5 Long-Short Term Memory (LSTM)	56

CHAPTER 4	Results & Discussions.....	61
4.1	Stacked LSTM	61
4.2	Bidirectional LSTM.....	67
4.3	Multivariate LSTM.....	69
4.4	Performance Comparison.....	77
CONCLUSION.....		80
RECOMMANDATIONS		81

LIST OF TABLES

	Page
Table 0.1	Number of wind turbine and installed wind power : Canada & Quebec ...21
Table 0.2	O&M costs in Canada.....33
Table 2.1	Carrier frequencies (CF) and their side-bands33
Table 4.1	The candidate parameters for inputs and output70
Table 4.2	Correlation values of candidate parameters71
Table 4.3	Tuned hyperparameters for developed LSTM models78
Table 4.4	MSE and R-squared values for different models78

LIST OF FIGURES

	Page
Figure 0.1	Installed wind power plant capacity in Canada16
Figure 1.1	Structure of a SCADA system76
Figure 1.2	Time trace of substation voltage23
Figure 1.3	The flowchart of project methodology25
Figure 2.1	Failure rate in wind power plant components.....28
Figure 2.2	Common failure rate in power converters of wind power plants.....30
Figure 2.3	The stator current spectrum32
Figure 2.4	The stator current spectrum in different operational condition34
Figure 2.5	The spectrum of active power in different operation condition.....34
Figure 2.6	Drive train temperature35
Figure 2.7	The gathered raw data by SCADA system for 750 kW wind turbine36
Figure 2.8	Pre-processed data for a 750-kW wind turbine.....37
Figure 2.9	Generator winding fault detection38
Figure 3.1	Correlation scenarios between two variables.....42
Figure 3.2	The topology of a one-layer ANN44
Figure 3.3	A graphical representation of neuron structure.....45
Figure 3.4	Binary step activation function representation46
Figure 3.5	Linear activation function representation46
Figure 3.6	Sigmoid activation function representation47
Figure 3.7	Hyperbolic tangent activation function.....48
Figure 3.8	Training cost of MNIST MLP with dropout.....50

Figure 3.9 Learning rate effect on finding the optimal point52

Figure 3.10 Dropout implementation on an ANN.....53

Figure 3.11 RNN topology over time.....54

Figure 3.12 Simple LSTM structure.....56

Figure 3.13 LSTM structure.....57

Figure 4.1 Generator windings actual temperature: WT 762

Figure 4.2 Generator windings actual temperature: WT 67.....62

Figure 4.3. Training dataset: Generator Winding Temperature WT 764

Figure 4.4. Testing dataset: Generator Winding Temperature WT 7.....65

Figure 4.5 Predicted generator windings temperature by Stacked LSTM: WT 766

Figure 4.6 Predicted generator windings temperature by Stacked LSTM: WT 6766

Figure 4.7 Bidirectional LSTM structure67

Figure 4.8 Predicted target variable by Bidirectional LSTM: WT 7.....68

Figure 4.9 Predicted target variable by Bidirectional LSTM: WT 67.....69

Figure 4.10 Scatter plots of candidate parameters71

Figure 4.11 Generator cooling system temperature: WT 772

Figure 4.12 Produced active power: WT 7.....73

Figure 4.13 Generator cooling system temperature: WT 6773

Figure 4.14 Produced active power: WT 67.....74

Figure 4.15 Predicted target variable by Multivariate LSTM: WT 7.....75

Figure 4.16 Predicted target variable by Multivariate LSTM: WT 67.....75

Figure 4.17 Histogram of model residuals75

Figure 4.18 Density plot of model residuals75

Figure 4.19 The graphical representation of models performance.....79

LIST OF ABBREVIATIONS

RES	Renewable Energy Source
O&M	Operation and Maintenance
AC	Alternating Current
DC	Direct Current
SCADA	Supervisory Control and Data Acquisition
RTU	Remote Terminal Units
PLC	Programmable Logic Control
LSTM	Long Short-Term Memory
PMSG	Permanent Magnetic Synchronous Generator
MPPT	Maximum Power Point Tracking
DFIG	Doubly-Fed Induction Generator
RNN	Recurrent Neural Network
MLR	Multiple Linear Regression
ANN	Artificial Neural Network
NLP	Natural Language Processing
MLP	Multi-Layer Perception
SGD	Stochastic Gradient Descent
GD	Gradient Descent
ADAM	Adaptive Moment Estimation
MSE	Mean Square Error
MAE	Mean Absolute Error
MBE	Mean Bias Error

NEMA National Electrical Manufacturers Association
CNN Convolutional Neural Network

LIST OF SYMBOLS

k	Harmonic index
f_s	Supplying frequency of induction machine
s	Slip
f_{brb}	Broken rotor bar harmonic frequency
f_{st}	Stator windings harmonic frequency
f_{FRU}	Faulty frequency related to unbalanced rotor in induction machines
p	Number of pole pairs
l	Supply time harmonic constant
f	Supplying frequency
x	Input feature of models
y	Dependent variable or target variable
\bar{x}	Mean value of input feature
\bar{y}	Mean value of dependent variable
β	Regression coefficient
ε	Error of model
r_{xy}	Correlation between x and y variables
σ	Sigmoid activation function
Tanh	Tangent hyperbolic activation function
$E(k)$	Prediction error vector at time step k
c_t	Cell state or memory state
h_t	Output of LSTM model
W	Vector of weights

W_{in}	Vector of weights for input features
W_{rec}	Vector of weights for previous LSTM cell output
α	Learning rate of model

INTRODUCTION

In recent decades, renewable energy sources (RESs) have been the focal point of research among scholars due to their considerable advantages over conventional and fossil fuel-based power plants. Penetration of RESs has been increased substantially in power network in recent decade to provide both consumers and network operators with technical and economic benefits. Technological development of power electronics has also boosted RESs utilization in both transmission and distribution levels of power network (Garcia-Vera, Dufo-Lopez & Bernal-Agustin, 2019). In recent decade, Canadian government has increased the investment in RES. Currently, Canada is ranked 9th in global wind power generation by producing 2% of global wind power generation and 4% of Canada's electricity generation (Canadian Wind Energy Association, 2020). In Figure 0.1, the increase in installed capacity of wind power plants in Canada is shown.

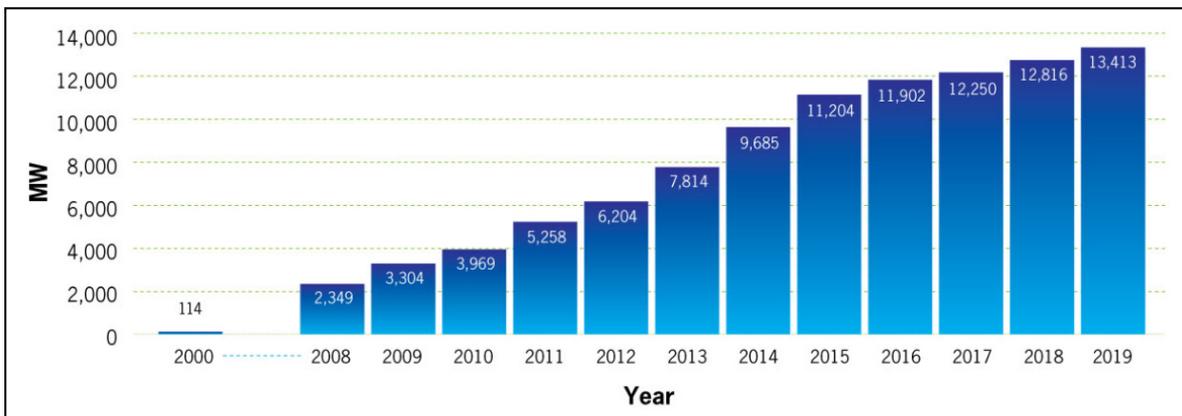


Figure 0.1 Installed onshore wind power plant capacity in Canada
Taken from CanWEA (2020)

It is obvious from Figure 0.1 that by the end of 2019, more than 13GW wind power is installed all over Canada and wind power plants have a high share of power production in the network. Besides, the number of installed turbines has been also increased in Canada and Quebec. In Table 0.1, the numbers of turbines and installed wind power capacity in Canada and Quebec are presented (CanWEA, 2020).

Table 0.1 Number of wind turbine and installed wind power: Canada and Quebec

Region	No. of Wind Turbine	Total Installed Capacity (MW)
Canada (Dec. 2019)	6771	13413
Quebec (Dec. 2019)	1990	3882

Many of these turbines are installed more than 10 years ago and precise condition monitoring strategies should be implemented for them to prevent any failure in the system. It is worth mentioning that the O&M expenses in wind power industry are increasing considerably. Table 0.2 shows the increasing trend of O&M costs in Canada (Canadian wind Energy Association, 2017).

Table 0.2 O&M costs in Canada

Year	O&M (M\$)
2017	290
2020	450

According to Table 0.2, it is anticipated that the O&M costs of wind power plants in Canada will increase to 450M\$ by the end of 2020 (CanWEA, 2017). Hence, scholars in collaboration with wind industry are trying to find innovative ways to decrease O&M costs in recent years. The scope of this research is to develop a powerful tool for fault diagnosis, condition monitoring, and fault prediction for wind power plants (specifically for generator windings) in collaboration with Power Factors and Natural Sciences and Engineering Research Council (NSERC). The rest of this report is organized as follows:

- Chapter 1 defines the problem and motivation for solving it. Objectives of project and the expected contributions are discussed in this chapter.

- Chapter 2 expounds the previous works on condition monitoring in wind power plants. Furthermore, an overview on Machine Learning (ML) and Deep Learning (DL) usage for processing SCADA data is provided in this chapter.
- Chapter 3 explains detailed methodology of project for achieving the defined goals and elaborates the concepts, advantages, and disadvantages of Multiple Linear Regression, Artificial Neural Networks, Recurrent Neural Network, Long Short-Term Memory. Moreover, the concept of Correlation Analysis for feature selection is explained.
- Chapter 4 defines two case scenarios which represents the healthy and faulty operation modes of a wind farm in the North of Quebec Province, Canada. The provided datasets by Power Factors are preprocessed and after implementing Correlation Analysis, the chosen input features are fed to different types of LSTM models in order to predict the generator winding temperature in both case scenarios. Finally, the obtained results are compared to indicate the performance of model. In the final section, a conclusion from this thesis is provided and some possible ideas for future works are discussed.

CHAPTER 1

PROBLEM DESCRIPTION

In last decade, electricity usage in urban and rural areas is increased substantially. Considering the electricity history, AC generators gained much more attention based on their abilities over DC generators. Most of the primitive power plants were diesel and steam generators; however, steam and hydro power plants widely added to power networks after a while. The increasing penetration of fossil fuel-based power plants in power grid have raised many environmental concerns which coerces many to see RES as a unique solution for meeting the electricity demand while considerably decreases Co₂ emission (Østergaard et al., 2020).

Wind power plants among different RESs have more established technology which made them more popular. Moreover, the energy source is free in these power plants which make them one of the affordable RESs. Besides all the merits that wind power plants have, it should be noted that their presence in power grid creates challenges for grid operators and owners of wind power plants. The intermittent production of power (due to unpredictable nature of wind speed) needs specific control algorithms and considerations that are not taken into account for fossil fuel-based power plants.

Another equally noteworthy challenge in operation of wind power plants is their considerable maintenance expenses. For several years, RESs have been advertised as cutting-edge technologies; however, after passing 20 years from their operation in power network, they cannot be considered as new technologies anymore. More equipment in wind power plants is reaching the end of their life span (Papatzimos et al., 2019). As a result, the O&M costs are increasing substantially in recent years. It has been anticipated that by the end of 2020, over 450 M\$ is needed for O&M costs of wind industry in Canada (CanWEA, 2017). Hence, it is highly important for governments and wind power plants owners to keep their competitiveness with fossil fuel-based power plants. Considering high share of O&M costs in total expenses of

power plants, finding solutions for reducing O&M costs of wind power plants will be the best way to increase the competitiveness of these power plants.

1.1 Supervisory Control and Data Acquisition (SCADA) systems

It is worth mentioning that Supervisory Control and Data Acquisition (SCADA) systems are widely used in wind power plants to gather data from equipment and send them to command center for processing and operation. The gathered data by SCADA systems can be used for monitoring the system operation. In the Fig. 1.1, it has been illustrated how SCADA system works (Inductive Automation, 2018).

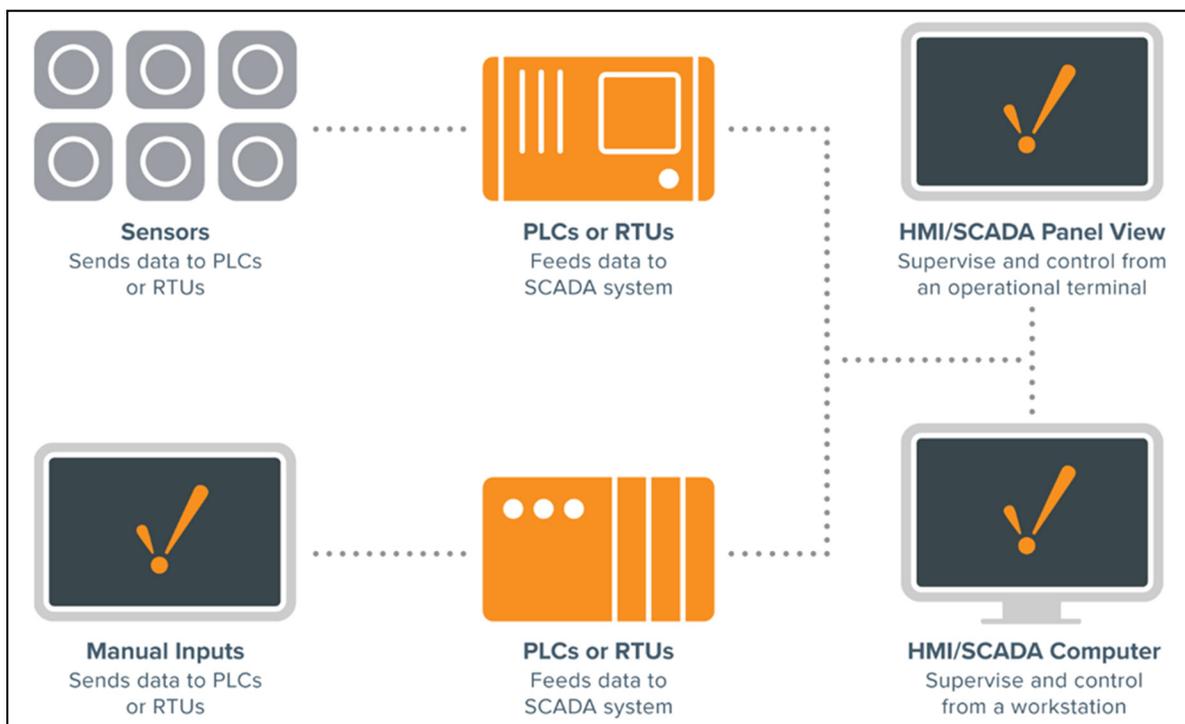


Figure 1.1 Structure of a SCADA system
Taken from Inductive Automation (2018)

The installed sensors, all over the wind power plant, gather data and send them to programmable logic controllers (PLCs) or remote terminal units (RTUs). Data can be sent to PLCs and RTUs manually, as well. PLCs and RTUs convey the gathered data to SCADA system for further actions. In the next step, SCADA systems prepare data for supervisory and

operational actions (Setiawan et al., 2018). Usually SCADA systems gather the data from different equipment by 10-mins average. However, SCADA systems are capable of gathering data for higher frequency up to 1 Hz. It is worth mentioning that 10-mins average gathering of data make analyzing system parameters more complex, especially for electrical parameters which usually have faster variations than mechanical ones.

1.2 Condition monitoring challenges of wind power plants

Considering the fact that studying horizon (which can be over 10 years) is much bigger than the gathering sample rate (in our case it is 10-mins average), the gathered data will be enormous over the years and as a result, their analysis will be a time-consuming process. The raw data which are gathered by SCADA systems provide the operators with several challenges that should be considered to guarantee proper operational analysis of wind power plants. These challenges can be explained as follows:

- Many of the gathered data may be out of rational operational boundaries. These data can affect proper condition monitoring analysis and moreover, the time required for analysis will be increased significantly. Hence, the raw data should be pre-processed based on the elimination these data (Yang & Shen, 2020).
- Any failure in measurement systems can affect the gathered data. It is highly important to know that the data have been gathered and processed under healthy operation of sensors, PLCs, and RTUs. By knowing the failure report of measurement devices, the faulty data can be deleted from our dataset (Jiang & Srivastava, 2020).
- Condition monitoring analysis is significantly related to the system operation point. Considering the fact that operation point of wind power plants is more variable than fossil fuel-based power plants (due to their dependence on wind speed), it can be deduced that condition monitoring in wind power plants are more complicated.

- The wind power plants experience shut-down periods due to different reasons such as grid operators command, curtailment, high wind speed and etc. The shut-down periods produce blank data in the database which affect the continuity of condition monitoring analysis.
- Another challenge in analyzing SCADA data for wind power plants is mutual interactions between mechanical and electrical parameters. This correlation fails any successful individual investigation for electrical or mechanical data analysis. Hence, any specific parameter monitoring in these systems needs more complicated control algorithms to consider the correlation between mechanical and electrical parameters (Moeini et al., 2019).
- The other challenge for condition monitoring in wind power plants is the geographical position of turbines. It is very important to consider the turbines' location in the condition monitoring, because the wind speed, shading effects, ambient temperature, altitude, and etc. are several parameters that related to the geographical position of turbines which affect the generated power (Aziz et al., 2019), (Shen et al., 2019), (Patrizi, Ciani, Guidi & Bartolini, 2019).

These challenges should be considered in developing any condition monitoring tool for wind power plants. To show the complexities of these challenges in the time trace of real data, the voltage of a wind power plant in high voltage side of its substation is shown in Fig. 1.2.

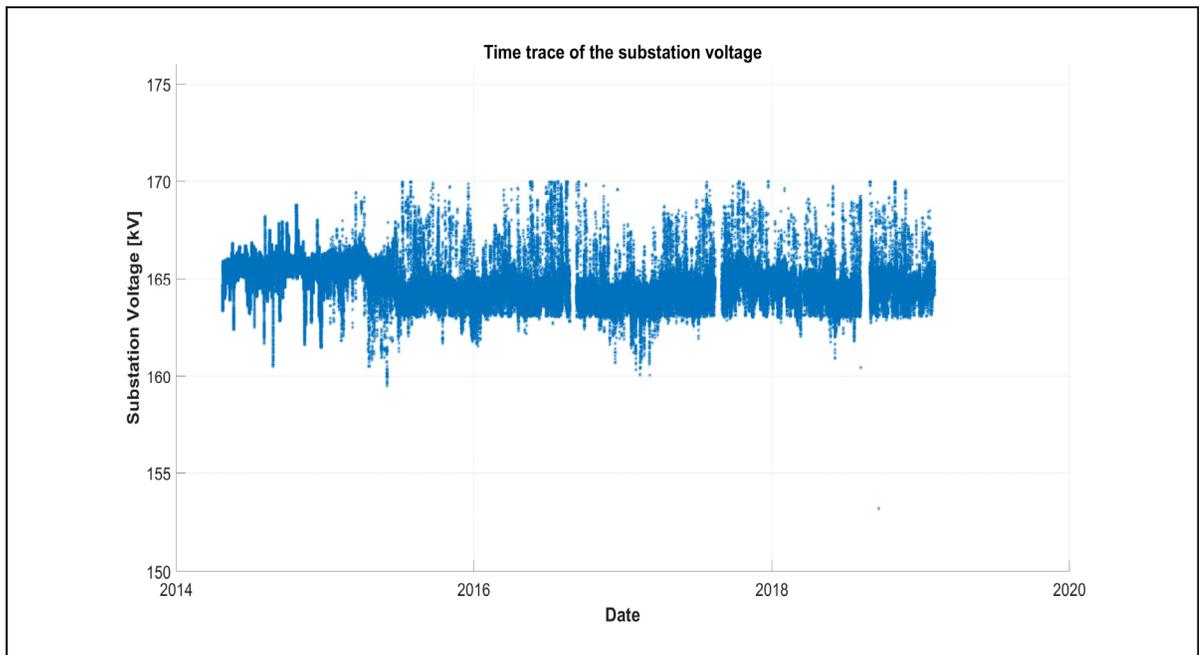


Figure 1.2 Time trace of substation voltage

This wind power plant is located in Quebec Province, Canada and the data is provided in collaboration with Power Factors; however, the name of this power plant is not mentioned here due to confidentiality reasons. The nominal voltage at high voltage side of substation is 161 kV. From Fig. 1.2, it can be understood that most of gathered data deviate around 161 kV, which is the nominal voltage at the high voltage side of substation. However, there are many data which is below 140 kV and many periods have zero voltage. The root cause of these considerable deviations can be any of the mentioned challenges.

1.3 Objectives of project

In order to reduce the O&M costs of wind power plants, there is a need for an accurate condition monitoring tool to overcome mentioned challenges and also reduce operational costs

of wind power plants. This goal cannot be achieved without exploiting ML techniques. As a result, the main objective of this project is to develop a precise condition monitoring tool for monitoring the temperature of generators' windings in wind power plants by using the gathered data from SCADA systems and using data mining and ML techniques for processing them.

1.4 Research Contribution

As it was mentioned earlier, few works have focused on electrical side of condition monitoring in wind power plants and mechanical equipment such as gearbox, blades, bearings and etc. were the focal point of research in recent decades. Hence, any endeavor for monitoring electrical equipment can be a contribution. Furthermore, since this project is in collaboration with an industrial compartment (Power Factors), it is important to know that there are several practical contributions for this project beside the research contributions.

In this project a diagnosis/monitoring/predictive tool is developed by DL approach for integrating to Power Factors's DIAGNOSTIX tool. Therefore, the development of this type of tool will help the electric power industry, in general and help Power Factors to keep its competitive edge at national and global market and benefit several industries in renewable energy sources areas, as well.

In the following, several research contributions of this project are explained:

- In this project, mutual investigation of electrical and mechanical parameters will be the focal research point. Moreover, the correlation analysis will be performed between mechanical and electrical parameters of system to extract exact relations between different parameters.
- Another contribution of this project is anomaly detection in the provided parameters by early prediction of faults and malfunction in the system. This will give the owners of wind power plants a unique opportunity to prevent any damage to their equipment.

- The main contribution of this project is implementing ML techniques for monitoring the temperature of generator windings based on the gathered real data from SCADA systems. Since the provided temperature signal is time-series data, Long Short-Term Memory (LSTM) technique is the best method for developing a monitoring model. Besides, a comprehensive literature review on ML applications on wind farm maintenance will be provided.

1.5 Methodology

To reach the mentioned goal in the previous chapter, it is important to propose a well-organized and detailed methodology. In the following, the methodology of this research is demonstrated in a flowchart.

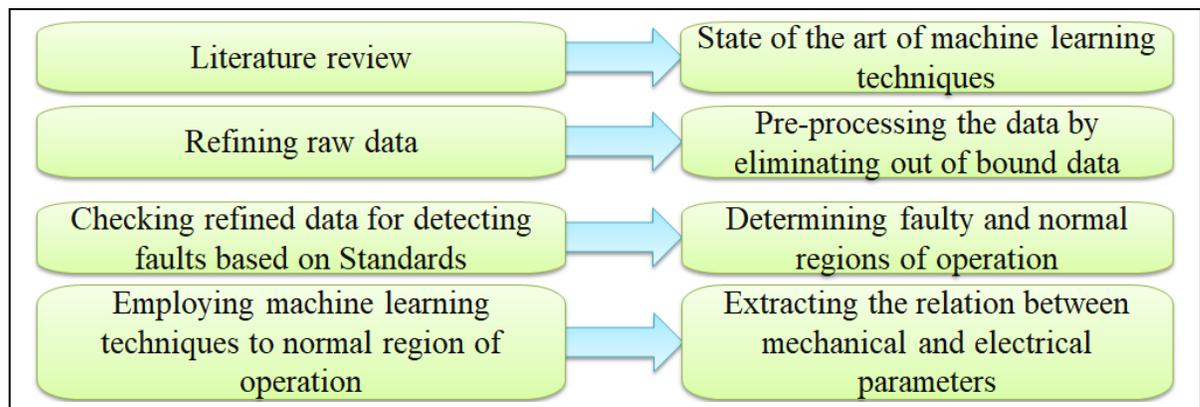


Figure 1.3 The flowchart of project methodology

The methodology has been proposed in several steps as follows:

1.5.1 A thorough literature review

Similar to every other project, it is necessary to perform a complete literature survey about the problem in order to gather required knowledge. The main focus on this project will be on two topics as follows:

- It is necessary in the first step to review the works that have been done on condition monitoring for equipment of wind power plants. Although most of the works has only focused only individual equipment for condition monitoring, these papers give good insight about the most common failures in them.
- Another equally noteworthy topic for literature review is the methods that have been used by papers for processing the gathered data by SCADA systems. The main focus in this part will be on the ML and data mining techniques which are used for processing the data and early stage prediction of faults in the system.

1.5.2 Refining raw data

As it was mentioned earlier, the gathered data by SCADA system can be logically false due to various reasons such as sensor failures, shut-down periods, and etc. It is crucial to exert a refining operation on these data before using them for further condition monitoring analysis and refining operation is called “pre-processing” the data. The main focus on preprocessing the data is to delete those data that are logically false. As an example, the amplitude of depicted substation voltage in Fig. 1.2. cannot be negative; hence, any negative voltage in the database should be eliminated to guarantee proper further analysis.

1.5.3 Pre-interpretation of data based on grid codes

It should be noted that each of the equipment that have been used in wind power plants can operate within ranges for normal operation or transient operation (when a fault occurs in the main grid or other parts of power plant). Moreover, the grid operators define various regulations in order to satisfy different indices to guarantee proper delivery of electrical energy to costumer. These regulations are known as grid codes which in this project are defined by Hydro-Quebec as the grid operator in Quebec Province. The grid codes will be scrutinized in literature review section.

After refining the gathered data in previous part, it is important to exert a pre-processing analysis based on the existing grid codes and defined operational ranges for each of the parameters. This will give us a good sight about the faulty and healthy operation regions. Moreover, it can be seen how close the system is operating to defined boundaries.

1.5.4 ML usage for interpretation of data

After determining the healthy operation regions, the dataset will be ready for being processed by ML methods in order to move from manual condition monitoring, which was based on data checking with grid codes and normal operation ranges, to a smart condition monitoring method which is not only shows the healthy and faulty operation regions, but also determines the correlations between different mechanical and electrical parameters to extract the root cause of failures in the system.

Knowing the correlations of mechanical and electrical parameters and also, the healthy and faulty states of system are two handy tools for interpretation of existing operational behaviors of system. The developed condition monitoring tool for analyzing the system behaviors in previous sections can monitor the system behaviors and detect the under-performance and faults in the system. However, it is crucial for wind power plants owners to prevent any damage to the equipment. Hence, the predictive approach will be added to the developed model in order to follow the trends of equipment operational behaviors. The predictive approach also can help us to estimate the life-span of equipment and moreover, redesign the maintenance schedule.

CHAPTER 2

LITERATURE REVIEW

2.1 Common failures in main equipment

As it was mentioned earlier, wind power plants are using SCADA system to execute their supervisory control and to gather data from equipment. It is worth noting that SCADA systems are already installed in most wind power plants; hence, their exploitation for condition monitoring purposes needs no additional tools or expenses which makes them a suitable choice for condition monitoring (Ye & Zhou, 2013).

Wind power plants are consisted from different electrical and mechanical equipment which make them vulnerable to various faults. It has been reported in literature that converters, control system, and rotors are among top three main faulty equipment in wind power plants (Hahn, Durstewitz & Rohrig, 2006), (Santos & González, 2019). In Fig. 2.1, the failure rate of wind power plant components is shown (Hahn et al., 2006).

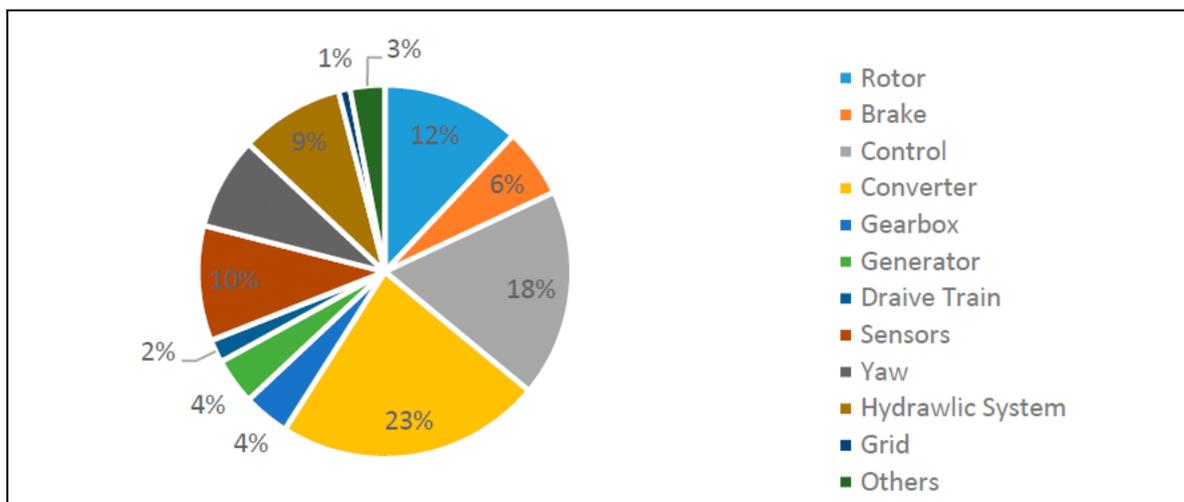


Figure 2.1 Failure rate in wind power plant components
Taken from Hahn et al. (2006)

In the following paragraphs, the papers that have focused on a specific component will be explained as follows:

2.1.1 Blades

Blades are one of the mechanical components of wind power plants which should be designed very carefully in order to guarantee maximum energy extraction from wind power. Fatigue, surface crack, material aging, deformation, and false design of pitch angle are some of the possible root causes of failures in blades (Finnegan, Flanagan & Goggins, 2020), (Mustafa, Barabadi, & Markeset, 2019). Moreover, the environmental factors such as weather condition such as icing can have considerable impact on underperformance of blade (Zeng et al., 2013). Mostly acoustic emission and vibration sensors are used in order to detect failures in the blades (Helander et al., 2017).

2.1.2 Gearbox

In some wind turbine topologies, gear box can be eliminated due to the fact that some generators can operate with lower rotational speed such as permanent magnetic synchronous generators (PMSGs). However, gearbox is used in most popular and well-established topologies and it is cardinal to consider them for condition monitoring analysis. The main faults of gearbox are related to faulty design and installation, tooth crack, bearing damage, and torque overload which can increase the temperature of oil and bearing (Salem, Abu-siada, & Islam, 2017), (Jantara & Papaelias, 2020), (Wang et al., 2020).

2.1.3 Main Shaft

Mechanical shaft plays a cardinal role in conveying the torque; hence, any failure or malfunction in it such as misalignment, corrosion, and crack, can decrease the torque. As a result, the characteristic figures of rotational speed and generated power change from their nominal values which can be a sign for failure in the system (Blancke et al., 2016).

2.1.4 Power Converters

Power converters are important components of wind power plants which are responsible for different tasks such as controlling voltage and guaranteeing maximum power point tracking (MPPT). Any failure in them can affect delivery of the generated power to grid considerably. Vibration, humidity, and temperature are three main reasons that can affect the capacitor, PCB, and IGBTs performance (Peyghami, Blaabjerg & Palensky, 2020), (Luo et al., 2019).

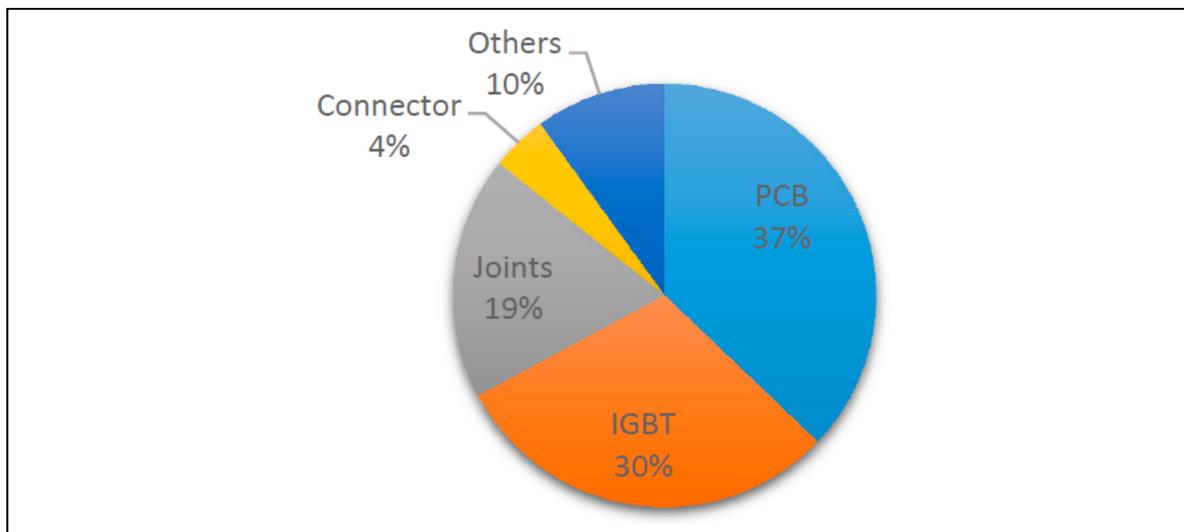


Figure 2.2 Common failure rate in power converters of wind power plants
Taken from Qiao et al. (2015, p. 6546)

It can be seen from Fig. 2.2 that most of failures are related to PCBs and IGBTs (Qiao & Lu, 2015).

2.1.5 Electric Machine

One of the most important components of wind power plants are electric machines which can be considered as the heart of power production in wind farms. Considering the fact that generator deals with both mechanical and electrical sections of system (responsible to transform mechanical energy to electrical one), both mechanical and electrical failures can affect its proper performance. Short-circuit, imbalance voltage and current phases are some of

the main electrical faults in the generators; on the other hand, air gap eccentricity, shaft, and bearing failures are among the most common mechanical failures in the generator (Yucai & Yonggang, 2016), (Artigao et al., 2020), (Zhang et al., 2020). Different tests and detection techniques are used in the literature to diagnosis these faults such as shaft displacement detection, vibration analysis, temperature monitoring, and torque measurement (Qiao & Lu, 2015).

2.2 Spectral Analysis

Another important fault diagnosis technique in generators is performing spectral analysis for electrical parameters. In (Artigao, Honrubia-escribano & Gomez-lazaro, 2018), Artigao suggests that the healthy and faulty operation conditions have their own specific signature in the spectral analysis of electrical parameters. This paper proposes the current spectral analysis for doubly-fed induction generators (DFIGs) for the first time. It is worth mentioning that the advantage of this method is that most of the failures' frequencies are known and based on the current spectrum, the failures and root causes can be detected.

As an example, when the rotor bars break, it is known for a long time that it affects the magnetic and electrical symmetrical structure of rotor; hence, it produces a frequency component in stator current spectrum at the following frequency:

$$f_{brb} = f_s \pm 2sf_s \quad (2.1)$$

Where s is the slip and f_s is the supplying frequency of generator.

Another possible failure in the generators is inter-turn short circuit in the stator winding which produces a frequency component at the following frequency:

$$f_{st} = f_s [k(\frac{1-s}{p}) \pm n] \quad (2.2)$$

Where p is the number of pole pairs, k is integer harmonic order, and n is odd.

The unbalance rotor is another failures root cause in generators which produces stator current harmonics and can be detected via produced frequency component at the following frequency:

$$f_{FRU} = f_s \left| \frac{k}{p} (1 - s) \pm l \right| \quad (2.3)$$

Where l is the supplying time harmonic constant.

This paper also provides other frequency components for various failures such as bearing damage, and air gap eccentricity which can be referred to detect malfunction or underperformance in the generator. The authors gathered their database from a wind farm in Spain for one year (starting September 2015) and extracted the stator current spectrum in different loading condition and also, in both super-synchronous and sub-synchronous speeds. In Fig 2.3., the stator current spectrum is shown by using FFT analysis (Artigao et al., 2018).

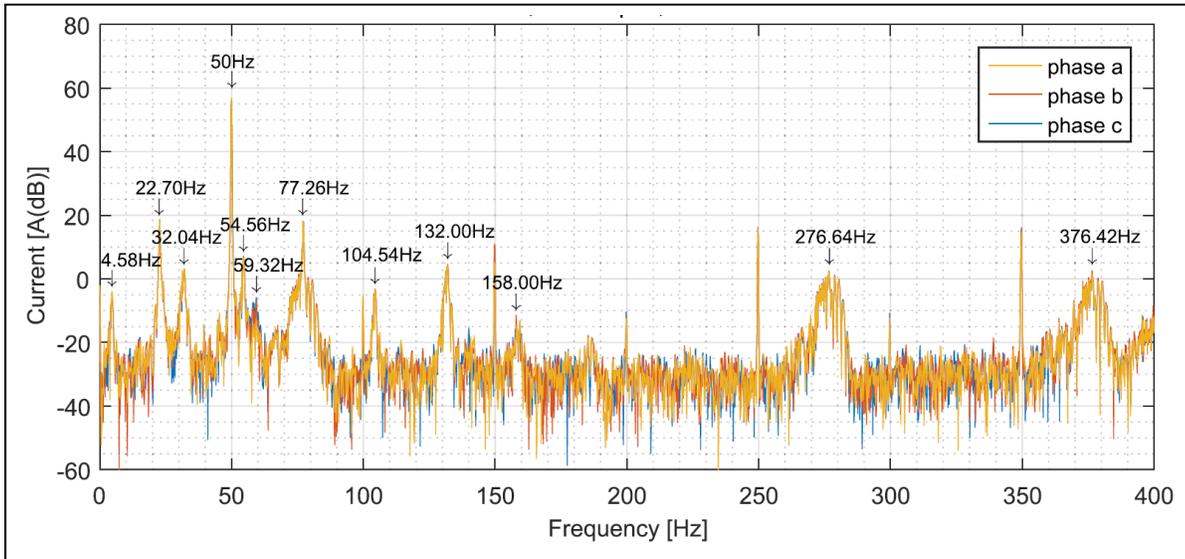


Figure 2.3 The stator current spectrum
Taken from Artigao et al. (2018, p. 5)

As it is demonstrated in Fig 2.3, there are several peak values for current in higher frequencies that might be the sign of failures. The mentioned equations can be used to check the root cause of possible failure and afterwards, rescheduling the maintenance observations.

In (Sarma et al., 2019), Zappala developed the current spectral idea by adding stator active power and rotational speed to the studies for spectral condition monitoring. However, it should be mentioned that this paper is only about generator's rotor electrical failure.

Table 2.1 Carrier frequencies (CF) and their side-bands

Generator Signal	Closed-Form Analytical Expressions	
	Balanced Rotor (CF)	Unbalanced Rotor ($CF \pm 2nsf$)
Stator Current, I_s	$ i \pm 6k(1-s) f$	$ (i \pm 2ns) \pm 6k(1-s) f$
Stator Active Power, Rotational Speed, P_e & N_s	$ (l \pm i) \pm 6k(1-s) f$	$ ((l \pm i) \pm 2ns) \pm 6k(1-s) f$

Table 2.1. indicates the possible increase in current, active power, and rotational speed spectrums due to unbalanced rotor structure. The effect on the mentioned spectrums will be as side bands ($2nsf$). The paper used a DFIG laboratory model with an emulated design and operational data as a DFIG harmonic model to demonstrate the proper performance of suggested frequency carriers. In the following figures, the predicted stator current and active power are shown in both healthy and unbalanced rotor condition in laboratory environment.

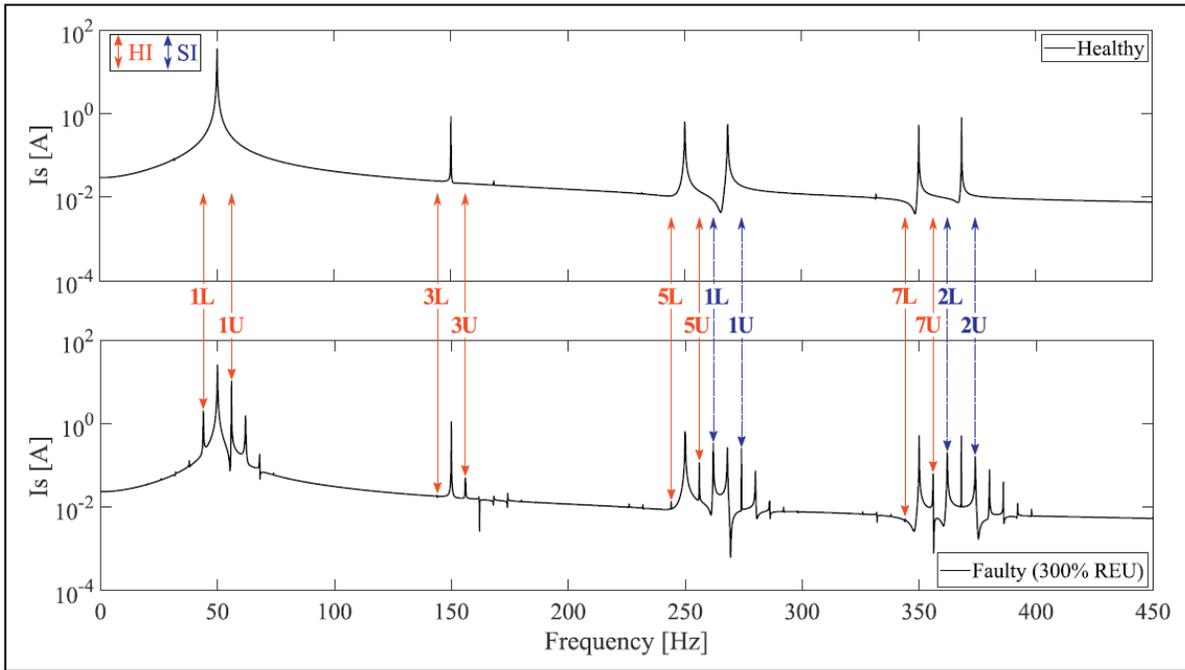


Figure 2.5 The stator current spectrum in different operational condition
 Taken from Sarma et al. (2019, p. 14)

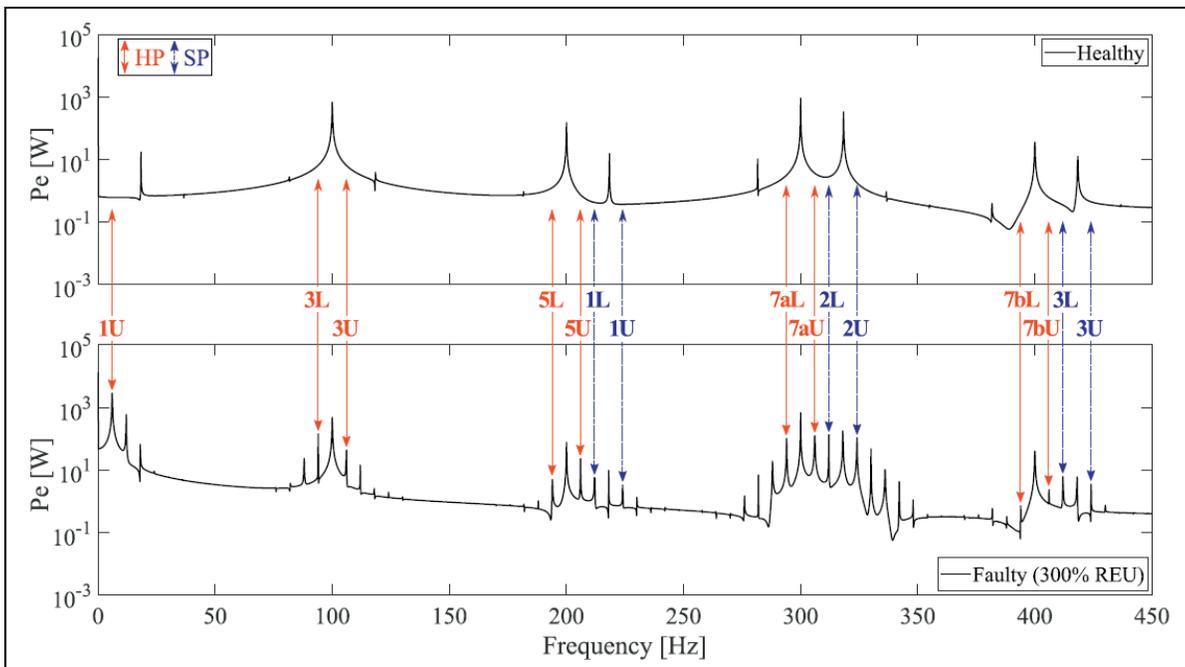


Figure 2.4 The spectrum of active power in different operation condition
 Taken from Sarma et al. (2019, p. 14)

It can be seen from Fig. 2.4 and Fig. 2.5 that the anticipated carrier frequencies for current and active power are correct and side bands appeared in rotor failure condition (Sarma et al., 2019). There are still other components in wind power plants which their failure affects proper performance of the system; however, since the main failures are discussed here, the failures of other components can be neglected. The gathered data by SCADA system can be exploited for condition monitoring purposes via three different approaches: 1) Signal trending, 2) Physical models, and 3) ML and signal processing techniques.

2.3 Signal Trending

This method of analyzing SCADA data is based on comparing the measurements from one turbine with other turbines. As an example, the bearing temperature in one turbine can be compared with other bearings temperatures in other turbines and any deviations will be a sign of failure in this method. One of the techniques that are used in this method is normalizing the quantity of measured parameter in order to make the comparison easier (Orleans, 2013). As an example, the temperature of drive train in a wind farm has been demonstrated in Fig. 2.6.

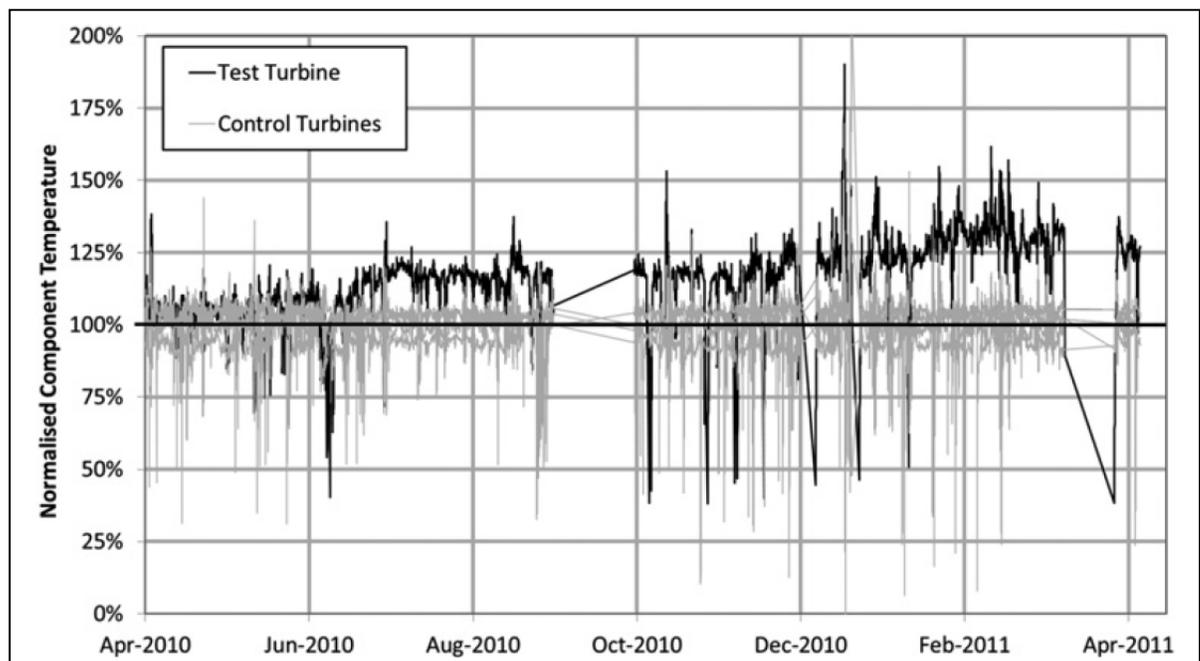


Figure 2.6 Drive train temperature
Taken from Orleans (2013, p. 382)

As it has been shown in Fig. 2.6, the drive train temperature of one turbine is compared with other ones in order to check the deviation. It can be seen after December 2010 the temperature deviates from expected value which can be a sign of failures. It is worth mentioning that this method is based on comparison of different turbines without considering the location of turbines and topology of wind farm. The altitude of turbines from sea level is highly affects the wind speed; hence, this method will have a high rate of false failure alarms though it might lead to detection of some failures in the system.

2.4 Physical Model

The other approach in condition monitoring of wind power plants are based on the physical and mathematical modeling of the system. In (Yang, Court & Jiang, 2013), Yang tried to refine to the data at the first step by sectionalizing the database and showing each section with only

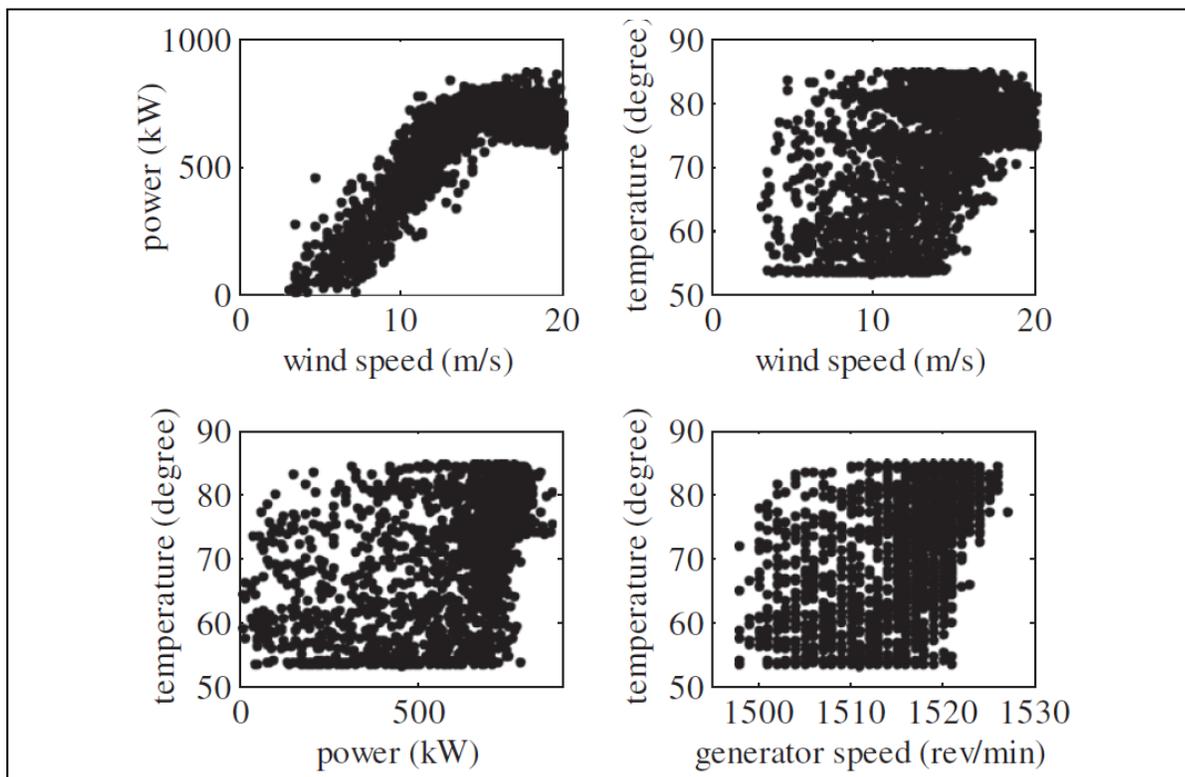


Figure 2.7 The gathered raw data by SCADA system for 750 kW wind turbine
Taken from Yang et al. (2013, p. 365)

one value. In Fig. 2.7., it has been shown how the gathered raw data for a 750-kW wind turbine are pre-processed by sectionalizing.

In Fig. 2.8., the refined data based on sectionalizing method are shown (Yang et al., 2013).

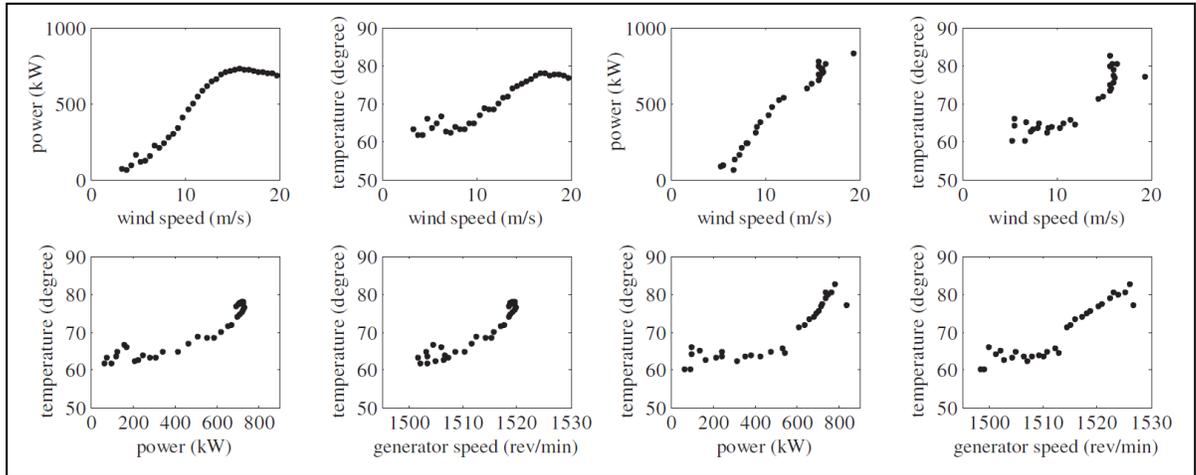


Figure 2.8 Pre-processed data for a 750-kW wind turbine
Taken from Yang et al. (2013, p. 365)

In this paper, a correlation test is also exerted to estimate the future values of parameters based on the variations in other parameters. The following equation the value of estimated value based on the correlation with other parameters.

$$\hat{y}_i = a_0 + a_1x_i + a_2x_i^2 + \dots + a_kx_i^k \quad (2.4)$$

Which the a_i parameters are calculated in the sensitivity analysis of other parameters as follows:

$$\begin{cases} \frac{\partial(R^2)}{\partial a_0} = 0 \\ \frac{\partial(R^2)}{\partial a_1} = 0 \\ \vdots \\ \frac{\partial(R^2)}{\partial a_2} = 0 \end{cases} \quad (2.5)$$

Where:

$$R^2 = \sum_{i=1}^n [y_i - \hat{y}_i] \quad (2.6)$$

This paper uses these equations to estimate the parameters and predict the system behaviors. A 30-kW wound rotor induction generator is used in laboratory to test the proper performance of this method. The system is tested in presence of slight and severe winding failures. The pre-processing algorithm is employed on the model and the figures have been extracted as follows (Yang et al., 2013):

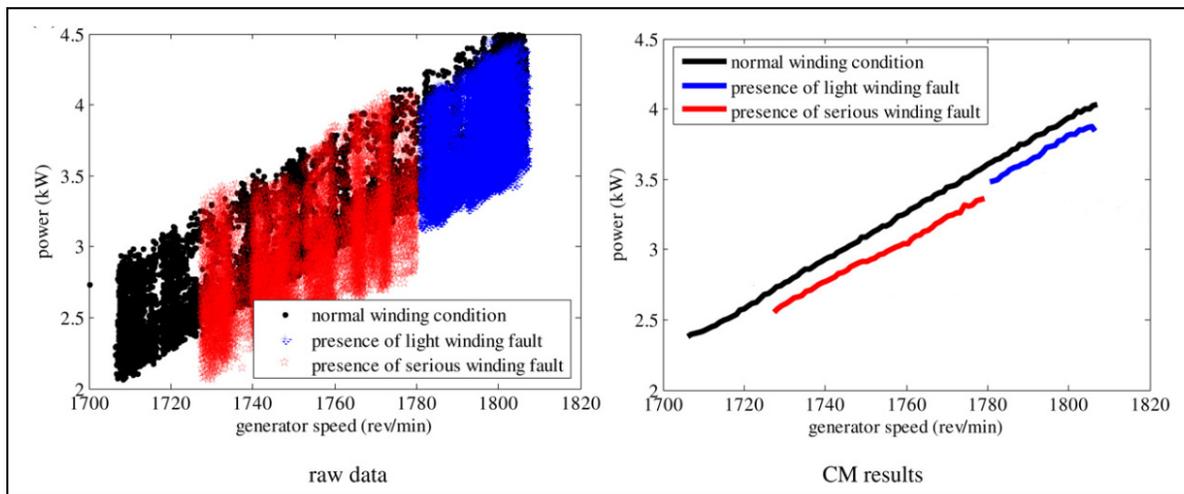


Figure 2.9 Generator winding fault detection
Taken from Yang et al. (2013, p. 365)

It can be seen that pre-processing method made the interpretation easier and also, deviations are clearer.

2.5 ML and signal processing techniques

The gathered data by SCADA system can be interpreted by using different signal processing methods. The Hilbert transform can be used for both time and frequency analysis and demodulate the target signal for feature extraction in order to make the data ready for fault

diagnosis based on the faults' features (Lu et al., 2013), (Gong, Qiao & Member, 2015), (Deng et al., 2019). The other signal processing method is Envelope method which is mainly used for vibration signals and detects the bearing inner and outer faults; however, it needs other signal processing methods to complete condition monitoring procedure, because this analysis is only in time domain (Weller, 2009), (Halpin et al., 2008).

The Statistical Analysis is another well-established method for fault detection wind power plants. In this method, the statistical parameters (mean value and variance) of signal is calculated and extracted for a healthy operation period and it will be recalculated for other periods. If the measured and calculated value deviate from the stored valued, it can be sign of failures in the system. It should be noted that in this method, only the occurrence of faults can be detected and prediction is not available and moreover, presence of noise in the data can highly affect to the interpretation of analysis and lead to a false alarm (Kusiak & Verma, 2012). The signals also can be divided into several different components with their specific scale and frequencies by Wavelet method to provide more data about the operation of system. The wavelet method has been used for monitoring bearings of generator and also gearbox, so far (Zimroz et al., 2012), (Watson et al., 2010), (Gray & Watson, 2010).

Using artificial intelligence is getting wide-spread in the last decade for engineering application due to enhances in computational systems. DL models are widely used in literature for anomaly detection in different systems. Here, we only focus on developed models for time-series data since the purpose of this project is to monitor temperature of generator windings provided by SCADA system with 10-min interval sampling. To handle the sequential data monitoring, Recurrent Neural Networks (RNNs) are the best choice among other developed DL topologies.

Among different RNN models, LSTM shows unique performance for analyzing time-series data with inherent dependencies to previous states of the system (Zhang, Li, Long & Ling, 2018), (Hu & Chen, 2018). In (Lei, Liu & Jiang, 2019), a classification approach is followed using LSTM for anomaly detection in wind farms by setting rotor-related signals. Moreover,

LSTM is used to predict the generated power of a wind farm for enhancing the competitiveness of wind farm in electricity market (Shi et al., 2018). However, usage of LSTMs for monitoring the generator winding is not reported in the literature. The contribution of this project is to utilize LSTM model for anomaly detection in windings of generators in wind farms.

CHAPTER 3

PROPOSED DATA PREPROCESSING TECHNIQUE And REGRESSION MODEL

3.1 Correlation Analysis

In this project, the main object is to predict the temperature of windings of generators in wind farms. For this purpose, considerable datasets of different mechanical and electrical signals are needed to be set as the inputs of ML models. Although these datasets are provided by Power Factors, several preprocessing techniques should be imposed on these raw datasets to be prepared for the models.

To increase the computation speed and decrease the complexity of model, it is crucial to decrease number of inputs to the models. The correlation analysis will be exploited in this project to determine the highly correlated features. The correlation between two datasets can be calculated as follows (Thompson , 2005):

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad (3.1)$$

Where \bar{X} , \bar{y} are the mean values of datasets.

A perfect positive correlation shows that the datasets are highly and linearly correlated and follow the same variation patterns and zero correlation value means that datasets have the least possible relation between each other. Finally, negative correlation value shows that datasets are linearly but negatively related to each other.

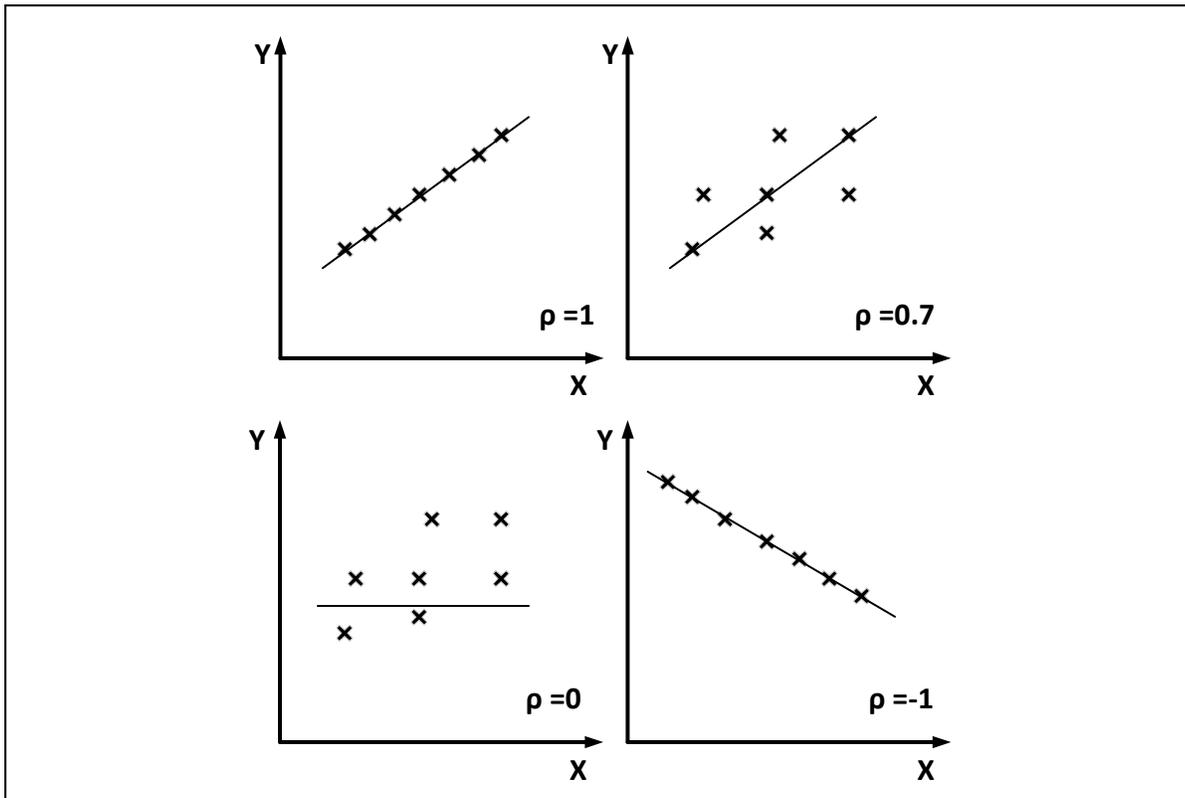


Figure 3.1 Correlation scenarios between two variables

When two variables have a high correlation value, it can be interpreted that their trends of variation are roughly the same. Hence, one of these highly correlated variables can represent the other and decrease the number of features and complexity of the model.

3.2 Multiple Linear Regression (MLR)

A regression approach is used for modeling the relationship between a dependent variable and one (Simple Linear Regression) or several independent variables (MLR). MLRs show excellent performance for predicting a target value when variables are linearly related, though there will be some inevitable errors. In MLR, the linear relationship between features and the dependent variable is obtained as follows (Breiman & Friedman, 1997):

$$y = X\beta + \varepsilon \quad (3.2)$$

Where:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix}; \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}; \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Where y , X , β , and ε are dependent variable, input features, regression coefficients, and error term, respectively. The whole MLR idea relies on the fact that there should be a linear relation between input features and output. The best linear regression is achievable by minimizing the error term in (3.2). The popular approach for minimizing the error term is mostly based on least square modeling which tries to minimize error square which is represented as follows:

$$S(\beta) = \sum_{i=1}^n \varepsilon_i^2 = \varepsilon^T \varepsilon = (y - X\beta)^T (y - X\beta) \quad (3.3)$$

The expanded version (3.3) is shown as follows:

$$\begin{aligned} S(\beta) &= y^T y - \beta^T X^T y - y^T X \beta + \beta^T X^T X \beta \\ &= y^T y - 2\beta^T X^T y + \beta^T X^T X \beta \end{aligned} \quad (3.4)$$

To obtain the optimal regression coefficient, the first derivative of least square equation in (3.4) is calculated as follows:

$$\frac{\delta S}{\delta \beta} = -2X^T y + 2X^T X \beta = 0 \quad (3.5)$$

Hence, the optimal regression coefficient is obtained as follows:

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (3.6)$$

3.3 Artificial Neural Network (ANN)

The brain functionality of humans in obtaining and processing the input signals from environmental experiences of body inspired scholars (McCulloch & Pitts, 1943) to develop an algorithm, named ANN. The developments in neural network modeling is pursued by scientists for reaching a better understanding of brain functionality and more importantly, developing algorithms and computers which have better performance in poorly defined problems such as pattern recognition and Natural Language Processing (NLP). The recent advances in computational systems made ANN-based systems the perfect choice for many different practical applications such as prediction, classification, image processing, and etc. One of the basic and widely popular types of ANNs is multi-layer perception (MLP) which shows a powerful performance in many practical aspects. The MLP structure is consists of three layers: input layer, hidden layer, and output layer. The schematic of a fully inter-connected MLP is shown in Fig. 3.2.

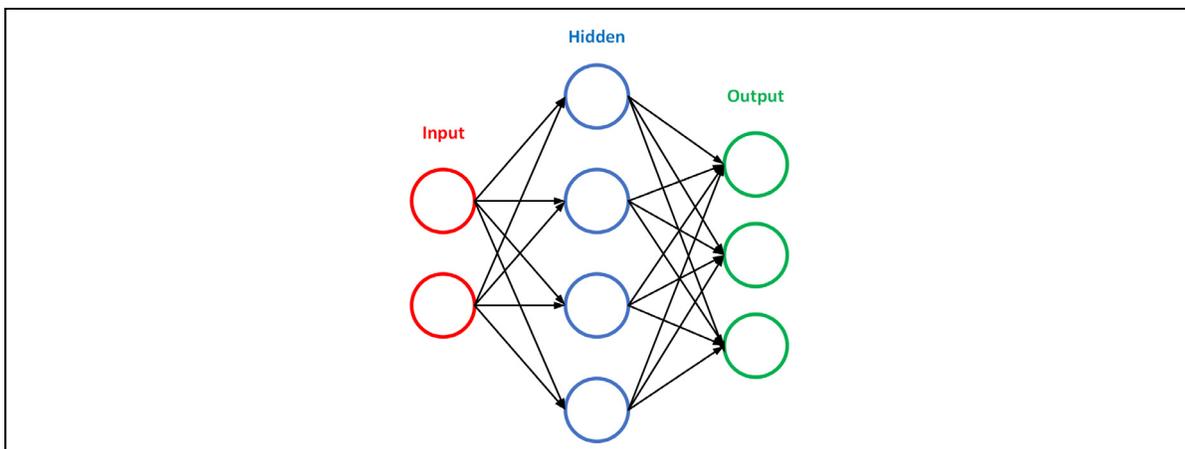


Figure 3.2 The topology of a one-layer ANN

It is worth mentioning that the neurons are the main components of MLP layer which determine the overall performance of system. To have a better understanding about MLP's performance, it is crucial to know exactly how a neuron or node processes the input data. The schematic of a node is presented in Fig. 3.3.

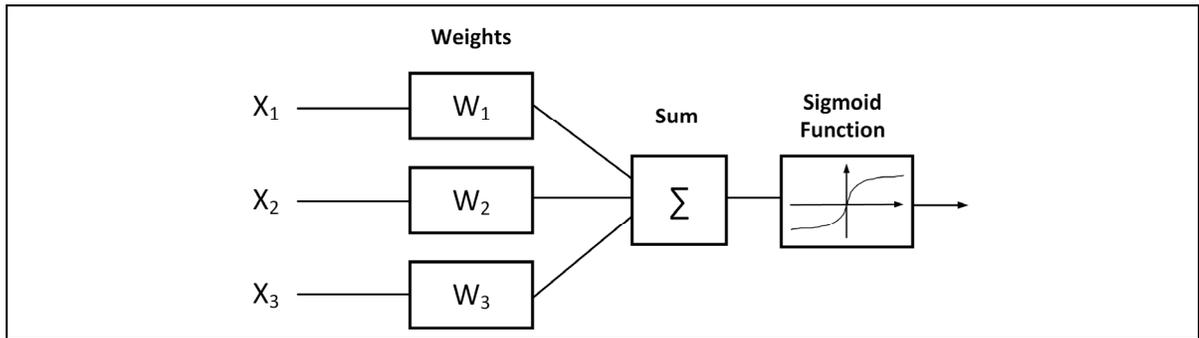


Figure 3.3 A graphical representation of neuron structure

The input data from input layer are fed to neurons in the hidden layer. Each neuron assigns a weight for the incoming data from previous layer to indicate the relative importance of them. In a neuron, the propagation function is exerted to the inputs by calculating the weighted sum of inputs which is added to a bias term. The obtained propagated values of neuron are fed to activation function and the results are passed to the next layer. This structure which relies on one-way flow of information is the basic topology for neural networks which is called feed-forward neural networks.

3.3.1 Activation Function

In terms of biological neural network, the activation function acts as the rate of action potential of firing of cells. Since the activation functions play a cardinal role in obtaining a proper output, dozens of activation functions are developed, which each of them has unique performance for specific problems. Considering the studied problem in this thesis, we only focus on the most relevant activation functions.

3.3.1.1 Binary Step Function

This activation function checks the calculated weighted sum of inputs with a threshold. Once the threshold is passed the neuron is activated and sends the exact same value to the next layer. The graphical representation of Binary Step Function is shown in Fig. 3.4.

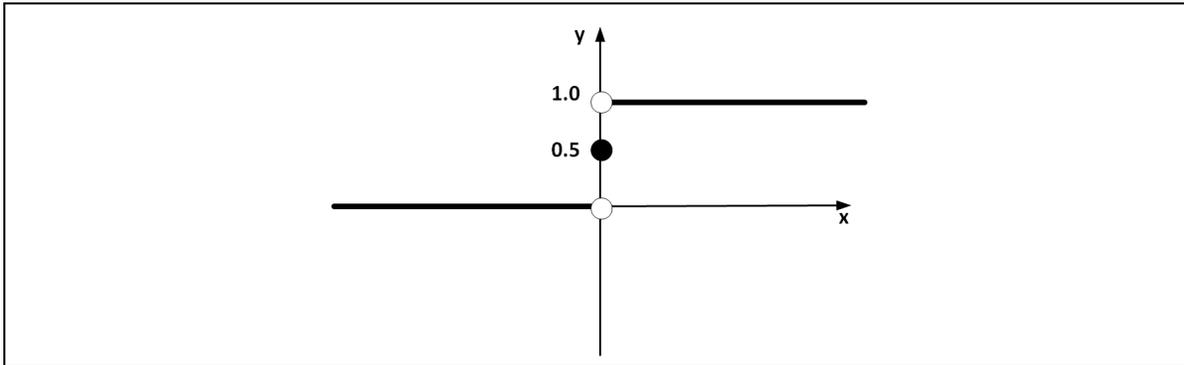


Figure 3.4 Binary step activation function representation

3.3.1.2 Linear Activation Function

The Linear Activation Function produces a proportional of input value and sends it to the next neuron. This activation function provides more freedom by producing multiple outputs instead of binary ones. In gradient descent algorithm, it is essential to have a relation between inputs and the derivative term of activation function and Linear Activation Function derivative term is constant and not related to the inputs; hence, the weights cannot be updated to minimize the error which makes this activation function practically useless. The Linear Activation Function is plotted in Fig. 3.5.

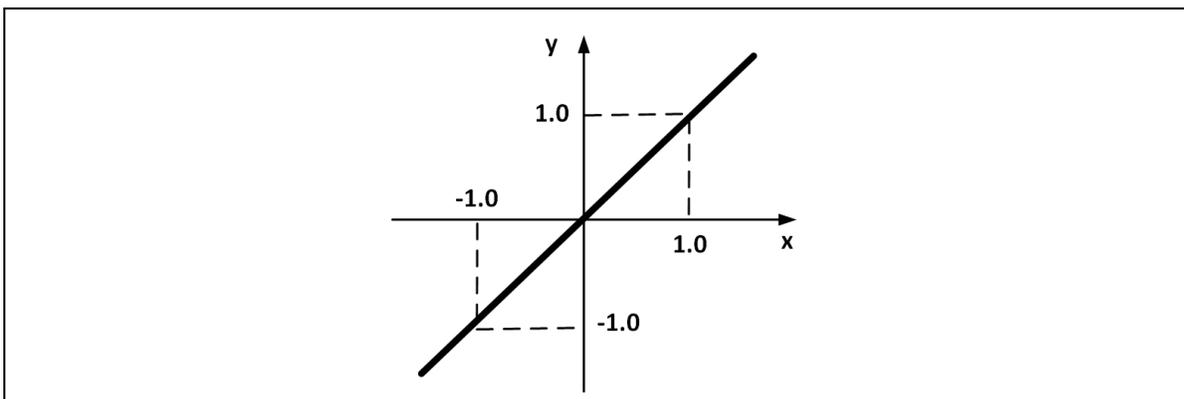


Figure 3.5 Linear activation function representation

3.3.1.3 Sigmoid Function

The Sigmoid Function is a nonlinear activation function which its outputs are between 0 and 1. The Sigmoid curve is shown in Fig. 3.6.

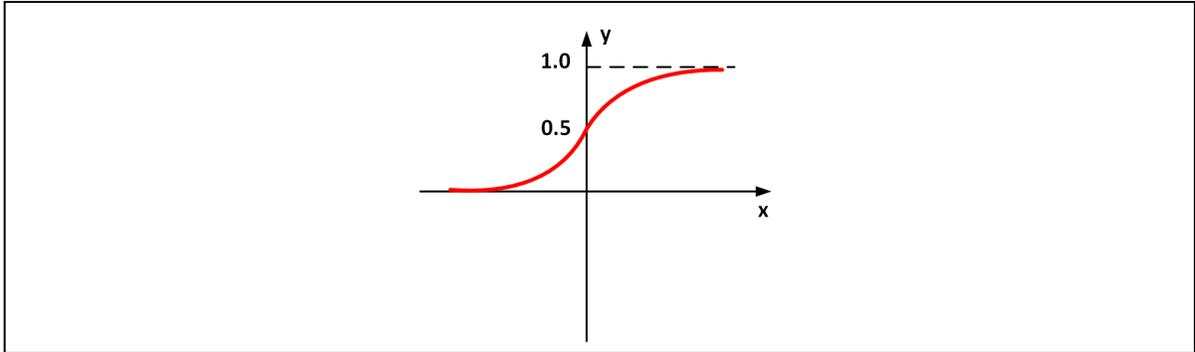


Figure 3.6 Sigmoid activation function representation

The limitation of output between 0 and 1 means that the rate of action potential firing in the cells cannot be faster than a certain rate. The mathematical representation of sigmoid function is shown as follows:

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (3.7)$$

This activation function guarantees that the output will remain between 0 and 1. Sigmoid function provides a fixed output range with continuously differentiable spectrum for different z values. It provides a smooth gradient which prevents any jumps in the outputs of neurons. One of the main disadvantages of this activation function is that it only produces positive output and cannot provide a proper response for problems that need negative output to reduce a variable or forget irrelevant data. Moreover, this activation function suffers from vanishing gradient problem which occurs for very high and very low values of input features. The network refuses to learn or learn very slowly in these ranges due to almost no change in the output of Sigmoid Function in these areas. However, Sigmoid Functions are widely exploited

for regression and classification problems due to their ability in providing a probability range for the output.

3.3.1.4 Hyperbolic Tangent Function

A modified version of sigmoid function is developed to produce both positive and negative output values. The output values for this activation function varies between -1 and 1. The Tanh function is plotted in Fig. 3.7.

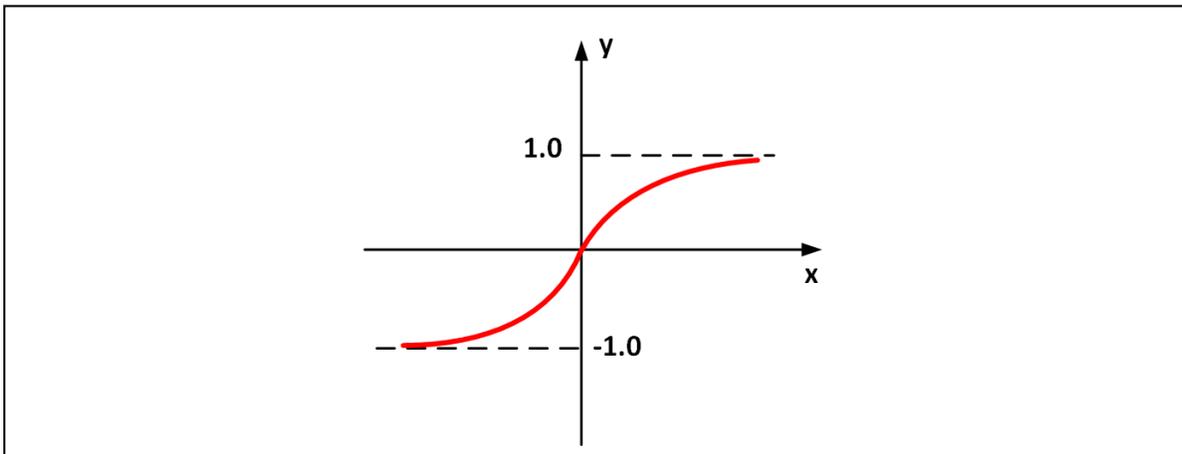


Figure 3.7 Hyperbolic tangent activation function

The mathematical representation of Tanh activation function is represented as follows:

$$\tanh(z) = \frac{2}{1 + e^{-2z}} \quad (3.8)$$

It is worth mentioning that one of the reasons for both sigmoid and Tanh activation functions popularity is their simple derivative forms which makes the backpropagation process for updating weights easier. Besides the sigmoid and Tanh functions difference in producing the negative outputs, Tanh function has a steeper gradient than sigmoid function which should be taken into account based on the problem requirements.

3.3.2 Optimizer

The main focus of an ANN is to minimize the predefined loss function in order to enhance results accuracy. Since it is impossible to calculate the perfect weights in an ANN, finding the acceptable weights is similar to an optimization problem with an algorithm of searching for the optimal weights. The most common algorithm which is used for ANNs is stochastic gradient descent (SGD) that is fortified with backpropagation of error algorithm to update the weights of ANNs. SGD tries to minimize the error by calculating the gradient of dataset and updating the weights values in direction opposite to the gradients in order to find a local minimum value. It is worth mentioning that Gradient Descent (GD) method is uses the same strategy for updating the weights and performs only one update; however, SGD updates the values for each training parameter. Hence, GD is much faster method with less precision and SGD can provide better results.

The other popular optimizer which is widely used for ML and DL problems is Adaptive Moment Estimation (Adam) which adapts the learning rate for each weight. Adam relies on the first moment of gradient, which is the mean value, and average of the second moment of gradient, which is the uncentered variance, for adapting the learning rate (Kingma & Ba, 2015). They showed that Adam considerably has lower training cost than other existing popular optimizers for a MLP algorithm on MNIST images dataset. The training costs over iterations for this problem are shown in Fig. 3.8 (Kingma & Ba, 2015).

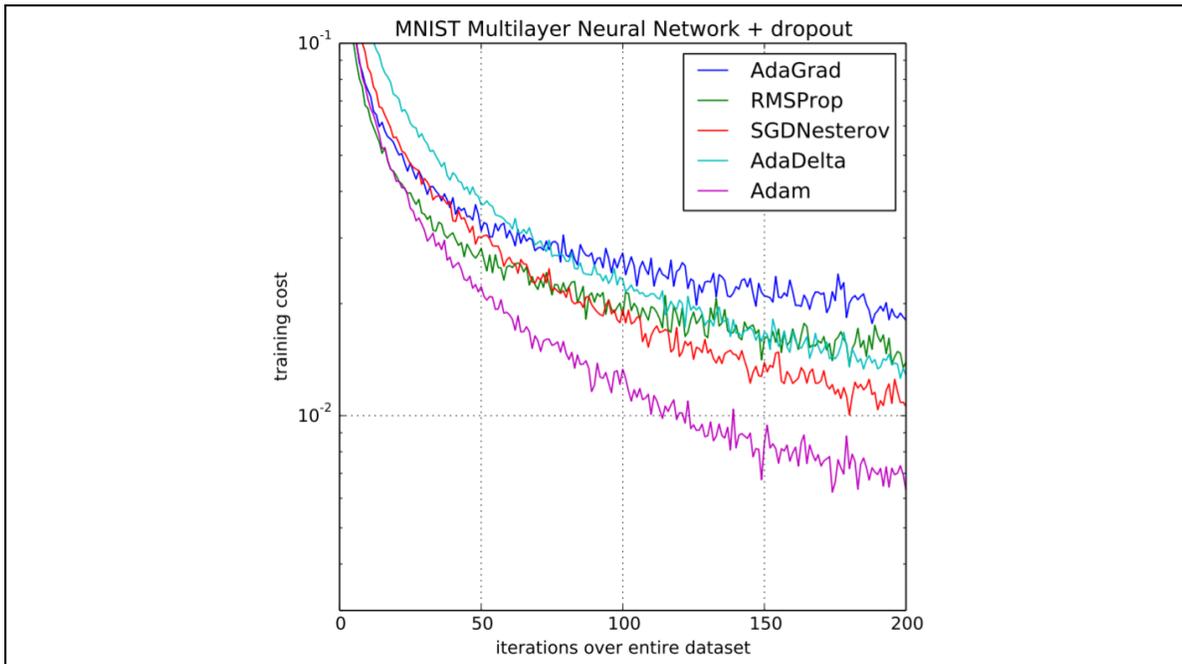


Figure 3.8 Training cost of MNIST MLP with dropout
Taken from Kingma et al. (2015)

It can be understood that Adam has the best performance in terms of training costs among the most popular optimizers. Hence, Adam optimizer is chosen in this thesis to update the weights due its considerable advantages over other existing optimizers.

3.3.3 Loss Function

The obvious definition of loss in ML models is the difference between predicted or obtained output values and the target variables. However, the main challenge is choosing the best method for obtaining the mentioned difference. One of the most popular loss functions for ANNs is Mean Square Error (MSE) which is the average of squared difference between predicted values and target ones. The predictions which are very far away from actual values have considerable effect on MSE value due to squaring. The mathematical representation of MSE is written as follows:

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \quad (3.9)$$

Another equally popular loss function for ML models is Mean Absolute Error (MAE) which is average of sum of errors between predicted and actual target variables. This criterion, same as MSE, neglects the direction of errors by using absolute values of error. However, implementation of MAE is harder than MSE for computing the gradients. On the other hand, MAE has a better performance with outliers since it does not use squared values of error. The MAE concept is formulated as follows:

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (3.10)$$

Mean Bias Error (MBE) is another less popular loss function which is used in ML applications. Considering the direction of error and positive or negative bias is the only merit of this loss function over MAE and MSE.

$$MBE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n} \quad (3.11)$$

Cross Entropy is a well-known loss function which is widely used for classification problems. This loss function utilizes a logarithmic equation of predicted variables to offer a probability value for the output. The deviations of predicted values from the actual ones will increase the Cross Entropy. The mathematical representation of Cross Entropy is written as follows:

$$CrossEntropyLoss = -(y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)) \quad (3.12)$$

It is obvious from (3.12) that when $y_i = 1$ the second term will disappear and $y_i = 0$ will eliminate the first term of equation.

3.3.4 Learning Rate

The step size which indicates the speed and precision of optimizer is called learning rate of neural network. It is crucial to assign a proper learning rate for finding the minimum in order to update the weights of neural network. A high learning rate can lead to neglecting the minimum point in the optimization surface while a low learning rate slows the searching operation or get stuck the operation in a local minimum. The graphical representation of this challenge is shown in Fig. 3.9.

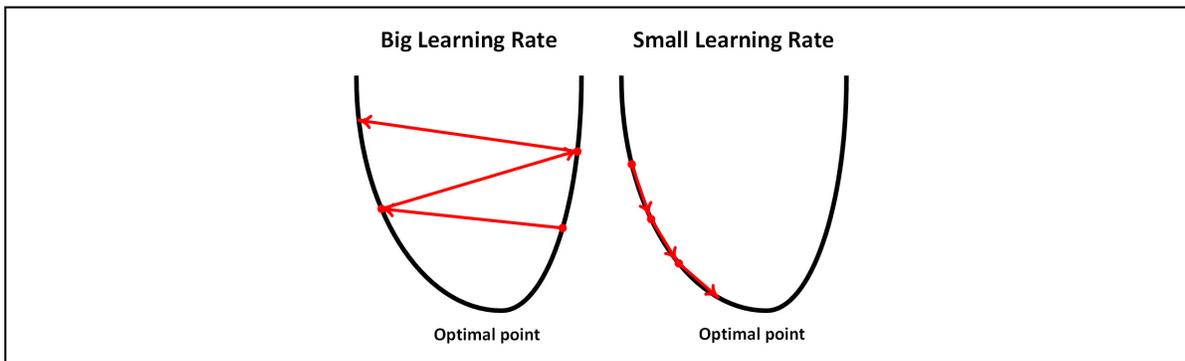


Figure 3.9 Learning rate effect on finding the optimal point

3.3.5 Dropout

The whole concept of learning for a model relies on splitting dataset to training and testing section. The training dataset usually have higher portion of the whole dataset. Training process is exerted by feeding the training datasets to the model in order to find a relation between the target values and training data. Once the model established a reliable relation between them, the testing dataset will be fed to model in order to check the performance of model. Implementation of a precise training model relies on many factors such as the portion of training data to whole dataset, selected input features, the applied filtering range on training data, number of hidden layers, number of neurons, optimizer algorithm, batch size, loss function, number of epochs, time steps, and etc.

Proper adjustment of these parameters, which are called hyperparameters, can vastly increase the efficiency and precision of training process; however, some models will still be highly dependent to the training datasets and perform poorly on any new dataset such as testing data. This phenomenon is called overfitting or overtraining which considerably make the model unreliable. The overfitting problem intensifies when the number of hidden layer increases; hence, DL models usually suffer from overfitting. To overcome this challenge, dropout strategy is widely used by developers.

Dropout strategy, which is a regularization process, randomly ignores the outputs of some neurons in a layer (Hinton et al., 2012). Since this process is random, different neurons of a layer are ignored in each iteration. The graphical representation of dropout is presented in Fig. 3.10.

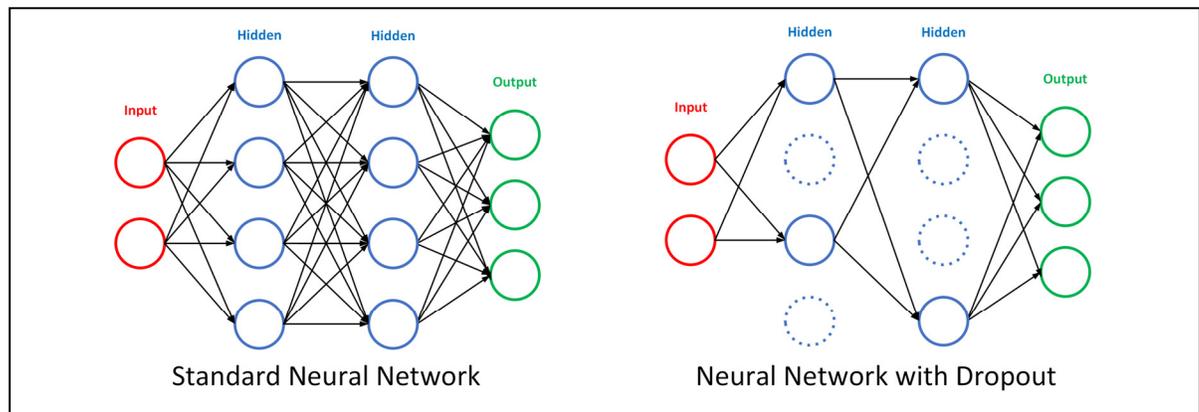


Figure 3.10 Dropout implementation on an ANN

Although this process makes the network performance noisy, it diminishes the adaptation of neurons to the input dataset. Dropout can be applied to input or hidden layers but not the output layer. Implementation of dropout on a network creates a new hyperparameter which can be tuned for achieving highest precision without overfitting. It should be noted that the weights will be larger in a network with dropout strategy since some of the neurons stopped working; hence, it is crucial to scale the weights based on the dropout rate of each layer (Srivastava et al., 2013).

3.4 Recurrent Neural Network (RNN)

RNN is the developed form of ANNs which can perfectly handle time-series and sequential datasets for applications such as machine translation, stock market prediction, text generations and etc. ANNs are usually used for datasets that are independent from each other while RNNs are specifically designed to work with datasets that are dependent to each other (Zaremba, Sutskever & Vinyals, 2014). In RNNs, ANN's backbone is added by a cell memory to look back on a few steps for predicting the future values. The representation of RNN is shown in Fig. 3.11.

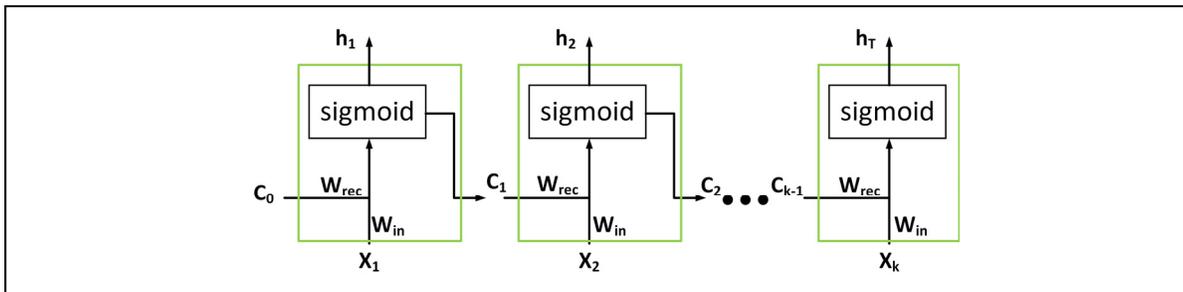


Figure 3.11 RNN topology over time

Where W_{rec} , W_{in} , x , h , and c vectors are the past information weights, input weights, input features, prediction vector, and past information of system, respectively. Fig. 3.11. shows unfolded RNN which means that the neural networks are extended over the chosen sequence. As an example, if the studied problem is Machine Translation and chosen sequence contains 5 words, it means that RNN is unfolded over 5 hidden layers which each layer represents one word. The memory state at time step t is related to the inputs of RNN and previous states of network. RNNs show an acceptable performance for applications where the gap between relevant information and current time is narrow. In the other word, RNNs performance decreases where the gap between previous states and current time step is very wide.

3.4.1 Vanishing Gradient Challenge:

After obtaining $h(k)$, the prediction error vector at time step k , $E(k)$, is calculated. These error values are used for calculating the gradients in Back Propagation Through Time algorithm which is formulated as follows (Hochreiter, 1998):

$$\frac{\partial E}{\partial W} = \sum_{t=1}^T \frac{\partial E_t}{\partial W} \quad (3.13)$$

Due to simplicity, the basic GD algorithm is considered for learning process. The gradient of errors on k time-step is calculated as follows:

$$\frac{\partial E_k}{\partial W} = \frac{\partial E_k}{\partial h_k} \frac{\partial h_k}{\partial c_k} \dots \frac{\partial c_2}{\partial c_1} \frac{\partial c_1}{\partial W} = \frac{\partial E_k}{\partial h_k} \frac{\partial h_k}{\partial c_k} \left(\prod_{t=2}^k \frac{\partial c_t}{\partial c_{t-1}} \right) \frac{\partial c_1}{\partial W} \quad (3.14)$$

The W vector represents the weights of RNN and it is used to calculate c_t as follows:

$$c_t = \sigma(W_{rec}c_{t-1} + W_{in}x_t) \quad (3.15)$$

By calculating the derivative of c_t and replacing it in (3.14), the backpropagation gradient can be obtained as follows:

$$\frac{\partial E_k}{\partial W} = \frac{\partial E_k}{\partial h_k} \frac{\partial h_k}{\partial c_k} \left(\prod_{t=2}^k \tanh'(W_{rec}c_{t-1} + W_{in}x_t) W_{rec} \right) \frac{\partial c_1}{\partial W} \quad (3.16)$$

Larger k can vanish the second term value in (3.16) since the tanh activation function derivative value is smaller than 1. Moreover, the multiplication of derivatives term can explode when the W_{rec} is large enough to prevail the tanh small derivative effect. This process is called exploding gradients. The main problem of RNNs is vanishing gradients challenge which can

stop the learning process of network. The mathematical representation of vanishing gradients is shown as follows:

$$\left[\prod_{t=2}^k \sigma'(W_{rec}c_{t-1} + W_{in}x_t)W_{rec} \right] \rightarrow 0 \equiv \left[\frac{\partial E}{\partial W} = \sum_{t=1}^T \frac{\partial E_t}{\partial W} \rightarrow 0 \right] \quad (3.17)$$

$$\equiv W \leftarrow W - \alpha \frac{\partial E}{\partial W} \approx W$$

Hence, the networks weights update will stop and no reasonable learning will occur over time.

3.5 Long-Short Term Memory (LSTM)

As it was mentioned earlier, RNNs are widely used for analyzing time-series datasets. RNNs are enhanced version of multilayer neural networks which can consider previous states of system in predicting current or future states. This will act as a memory-based model that will be a suitable choice for sequential datasets. The simple structure of LSTM is shown in Fig. 3.12.

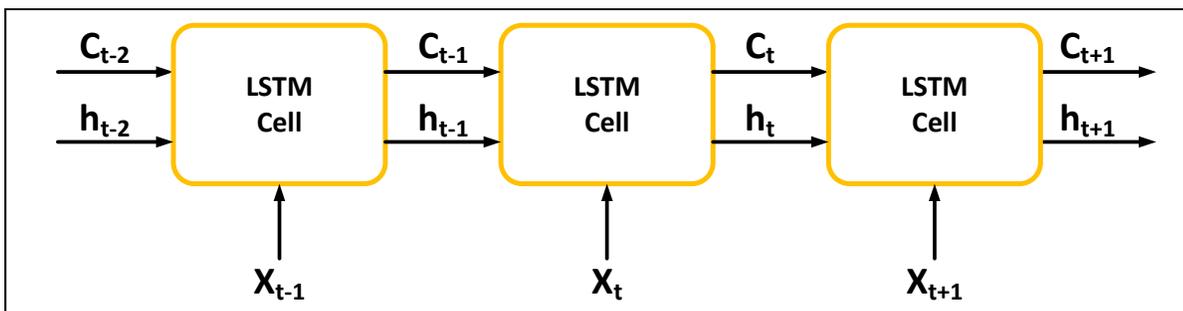


Figure 3.12 Simple LSTM structure

After developing RNNs, LSTMs are introduced specifically to handle datasets that have long-term dependencies. The LSTM structure consists of input, forget, and output gates and memory cell which is illustrated in Fig. 3.13.

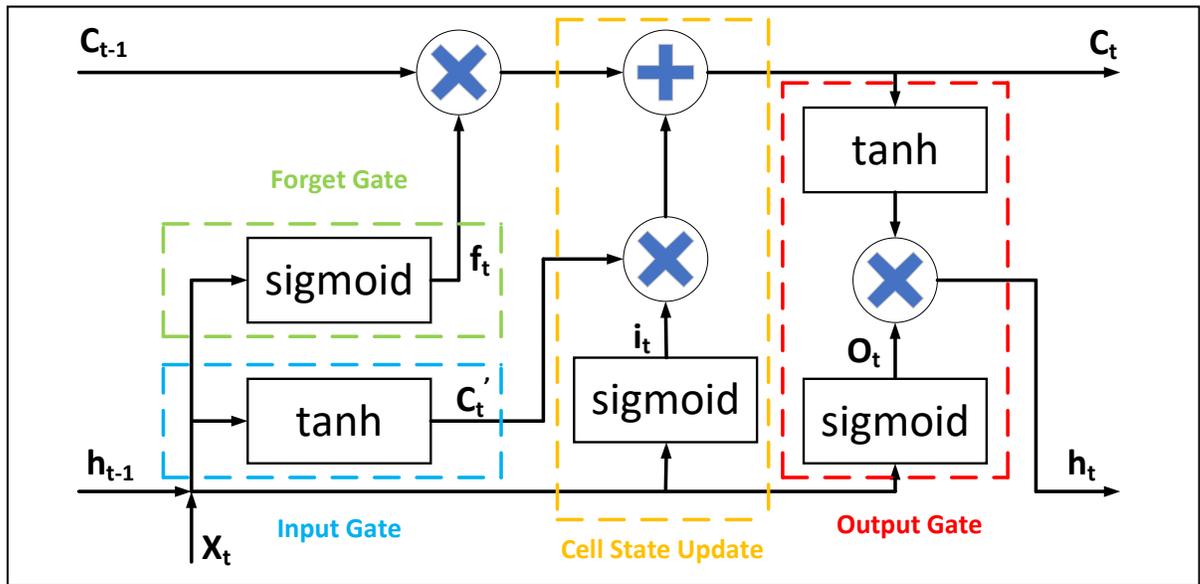


Figure 3.13 LSTM structure

The memory cell in each time step is responsible for deciding the eligible information for storing. The new information and previous output from last unit is fed to LSTM unit to pass from tanh layer and a sigmoid function layer to indicate which information should or should not be stored. Note that like other neural network the inputs should be scaled because sigmoid functions are operating between 0 and 1.

The forget layer consists of a sigmoid function and the information which is stored from previous LSTM units. The sigmoid function illustrates which information should be remembered for the next units. At the end, the output layer uses a tanh function to determine the information needed to pass to the output and next unit of LSTM. The mathematical equations of LSTM are elaborated in the following:

$$f_t = \sigma(W_F x_t + U_F h_{t-1} + b_F) \quad (3.18)$$

$$i_t = \sigma(W_I x_t + U_I h_{t-1} + b_I) \quad (3.19)$$

$$o_t = \sigma(W_O x_t + U_O h_{t-1} + b_O) \quad (3.20)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_C x_t + U_C h_{t-1} + b_C) \quad (3.21)$$

$$h_t = o_t \odot \tanh(c_t) \quad (3.22)$$

$$o_t = f(W_O h_t + b_O) \quad (3.23)$$

Where b_t is bias and x_t and h_t are input and hidden vectors, respectively. In addition, W and U are input to hidden and hidden to hidden weights matrices, respectively.

As it was mentioned earlier, RNNs suffer from vanishing gradients while LSTMs show excellent performance in handling this challenge. To calculate the gradients of error for LSTMs, we should obtain the derivative of cell state gradient which is a key term in (3.14). The derivative of LSTM cell state is calculated as follows (Hochreiter, 1998):

$$\begin{aligned} \frac{\partial c_t}{\partial c_{t-1}} &= \frac{\partial}{\partial c_{t-1}} [c_{t-1} \otimes f_t \oplus \tilde{c}_t \otimes i_t] = \frac{\partial}{\partial c_{t-1}} [c_{t-1} \otimes f_t] + \frac{\partial}{\partial c_{t-1}} [\tilde{c}_t \otimes i_t] \\ &= \frac{\partial}{\partial c_{t-1}} c_{t-1} + \frac{\partial f_t}{\partial c_{t-1}} c_{t-1} + \frac{\partial \tilde{c}_t}{\partial c_{t-1}} i_t + \frac{\partial i_t}{\partial c_{t-1}} \tilde{c}_t \end{aligned} \quad (3.24)$$

Where each term can be calculates as follows:

$$\begin{aligned} \frac{\partial f_t}{\partial c_{t-1}} c_{t-1} &= \frac{\partial}{\partial c_{t-1}} [\sigma(W_f [h_{t-1}, x_t])] c_{t-1} \\ &= \sigma'(W_f [h_{t-1}, x_t]) W_f \frac{\partial h_t}{\partial c_{t-1}} c_{t-1} \\ &= \sigma'(W_f [h_{t-1}, x_t]) W_f o_{t-1} \otimes \tanh'(c_{t-1}) c_{t-1} \end{aligned} \quad (3.25)$$

$$\begin{aligned}
\frac{\partial i_t}{\partial c_{t-1}} \tilde{c}_t &= \frac{\partial}{\partial c_{t-1}} [\sigma(W_i[h_{t-1}, x_t])] \tilde{c}_t \\
&= \sigma'(W_i[h_{t-1}, x_t]) W_i \frac{\partial h_t}{\partial c_{t-1}} \tilde{c}_t \\
&= \sigma'(W_i[h_{t-1}, x_t]) W_i o_{t-1} \otimes \tanh'(c_{t-1}) \tilde{c}_t
\end{aligned} \tag{3.26}$$

$$\begin{aligned}
\frac{\partial \tilde{c}_t}{\partial c_{t-1}} i_t &= \frac{\partial}{\partial c_{t-1}} [\sigma(W_c[h_{t-1}, x_t])] i_t \\
&= \sigma'(W_c[h_{t-1}, x_t]) W_c \frac{\partial h_t}{\partial c_{t-1}} i_t \\
&= \sigma'(W_c[h_{t-1}, x_t]) W_c o_{t-1} \otimes \tanh'(c_{t-1}) i_t
\end{aligned} \tag{3.27}$$

The derivative of cell state contains four sub-derivative terms which is computed as follows:

$$\begin{aligned}
\frac{\partial c_t}{\partial c_{t-1}} &= \sigma'(W_f[h_{t-1}, x_t]) W_f o_{t-1} \otimes \tanh'(c_{t-1}) c_{t-1} \\
&\quad + f_t \\
&\quad + \sigma'(W_i[h_{t-1}, x_t]) W_i o_{t-1} \otimes \tanh'(c_{t-1}) \tilde{c}_t \\
&\quad + \sigma'(W_c[h_{t-1}, x_t]) W_c o_{t-1} \otimes \tanh'(c_{t-1}) i_t
\end{aligned} \tag{3.28}$$

It is worth mentioning that forget gate activation function is included in the gradient which indicates the information that should be forgotten. This property helps LSTMs to prevent vanishing gradients by adjusting the forget gate parameters at any time step. Another equally noteworthy tool for preventing the vanishing gradients in LSTM is the additive structure of derivative of cell state in (3.28) which make LSTMs more robust against falling into vanishing gradients trap. This additive structure provides LSTMs with capability of four parameter tuning option.

Different types of LSTM are implemented in this project for predicting generator windings temperature. The developed LSTM models in this project are listed as follows:

1. **Stacked LSTM:** In this LSTM model, one hidden-layer is stacked over other hidden-layers while LSTM strategy will be the backbone of model.
2. **Bidirectional LSTM:** In this model, the backward and forward sequences are concatenated as the input data and fed to LSTM for predicting the target variable.
3. **Multivariate (multiple input) LSTM:** A LSTM model may have several sequential input variables which are used to predict the target variable.

Finally, these models will be evaluated based on several criteria such as Mean Square Error (MSE) and R-squared which shows the expected quadratic loss between predicted and real data.

CHAPTER 4

RESULTS & DISCUSSIONS

As it was mentioned earlier, LSTM technique is used in this project to predict the generator windings temperature; however, different types of LSTM can be implemented for this purpose. It is worth mentioning that two case scenarios are considered in this project to evaluate the models performance on healthy operation region (turbine No. 7) and abnormal operation state (turbine No. 67) for a wind farm in the North of Quebec Province, Canada.

4.1 Stacked LSTM

In this prediction problem, the input feature of Stacked LSTM is the same as target value which is generator windings temperature consists of 100K data (10-min sampling rate) provided by Power Factors. The actual temperature of generator windings for turbine No. 7 and No. 67 are shown in Fig. 4.1. and Fig. 4.2.

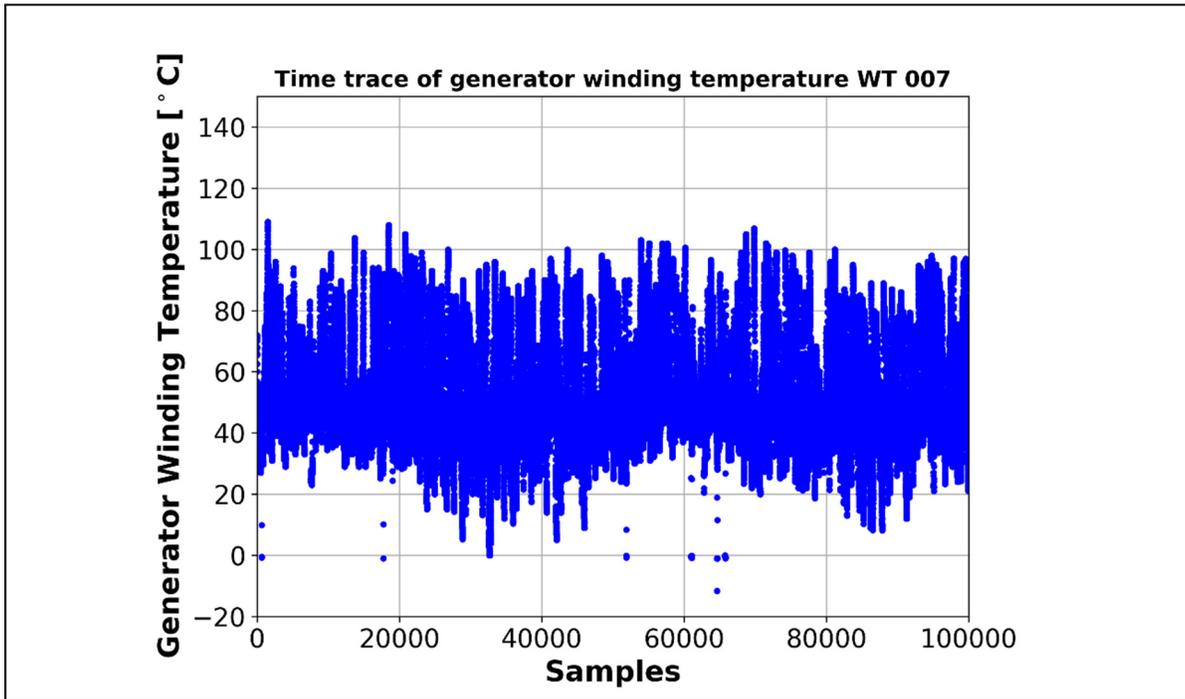


Figure 4.1 Generator windings actual temperature: WT 7

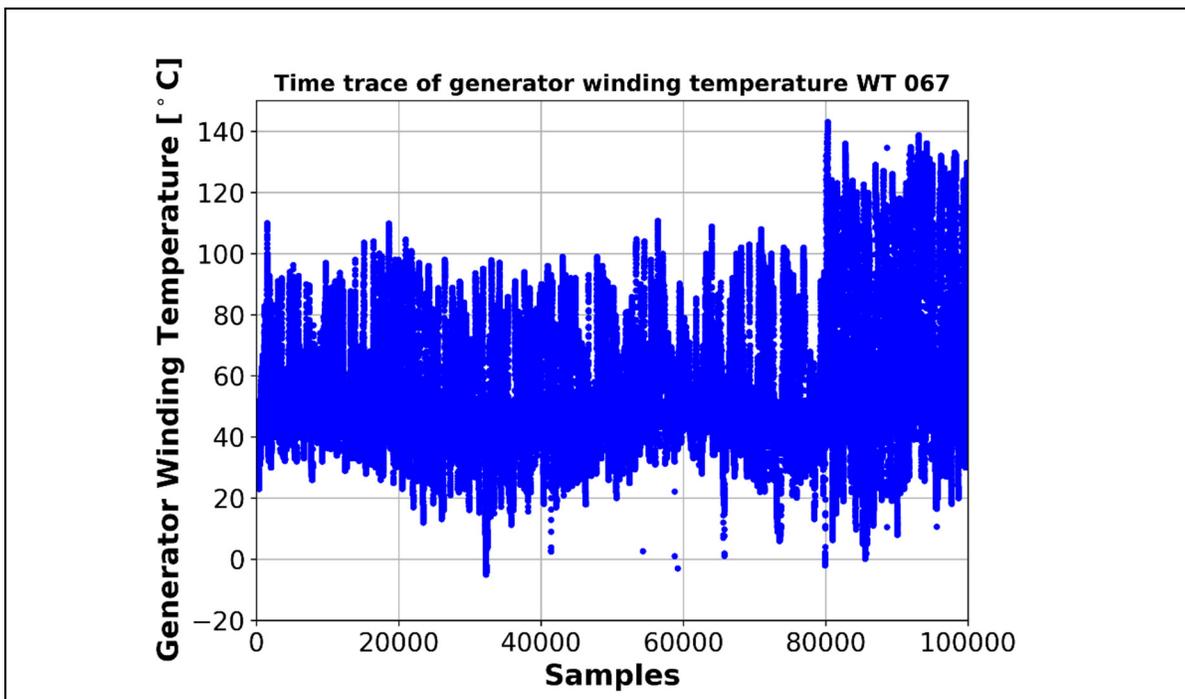


Figure 4.2 Generator windings actual temperature: WT 67

It can be inferred from Fig. 4.1. and Fig. 4.2., the temperature variation is affected by the seasonal cyclic variation of ambient temperature. Since the datasets nearly show 2 years of generator operation, two peaks and two dips demonstrate wind farm operation in summer and winter seasons, respectively. The generator wires in this wind farm are insulated with class F insulation material to prevent inner short circuit fault. National Electrical Manufacturers Association (NEMA) and IEC standards suggest that the maximum acceptable increase for insulation class F is 155°C (National Electrical Manufacturers Association, 2019), (International Electrotechnical Commission, 2004). When the generator operates above or near the prescribed threshold, the insulation system will age thermally from two to six times faster than normal temperature condition (insulation degradation doubles by every 6°C increase) (International Organization for Standardization, 2005), (McNutt, 1992).

As it was mentioned earlier, the chosen dataset is consisting of more than 100000 samples. This dataset is split into two sections: 1) Training dataset which the weights of model is trained in a way to minimize the defined loss function and 2) Testing dataset which is used to validate the developed model performance. To demonstrate the model proper performance, it is necessary to obtain acceptable values for model metrics. For a better representation of the proposed model in predicting the temperature of generator windings, wind turbine No. 67 from the same wind farm is considered as a case scenario. It is obvious from Fig. 4.2. that windings temperature follows a seasonal cyclic pattern the same as the previous case; however, there is a considerable increase in temperature variations after nearly 80000th sample for some reason and for some period it exceeds 140°C. For this purpose, the first 75000 samples which are 75% of total data is considered as the training dataset to train the model. This ratio is chosen in a way to guarantee that the mentioned incident is not included in the training dataset. The input training dataset is shown in Fig 4.3.

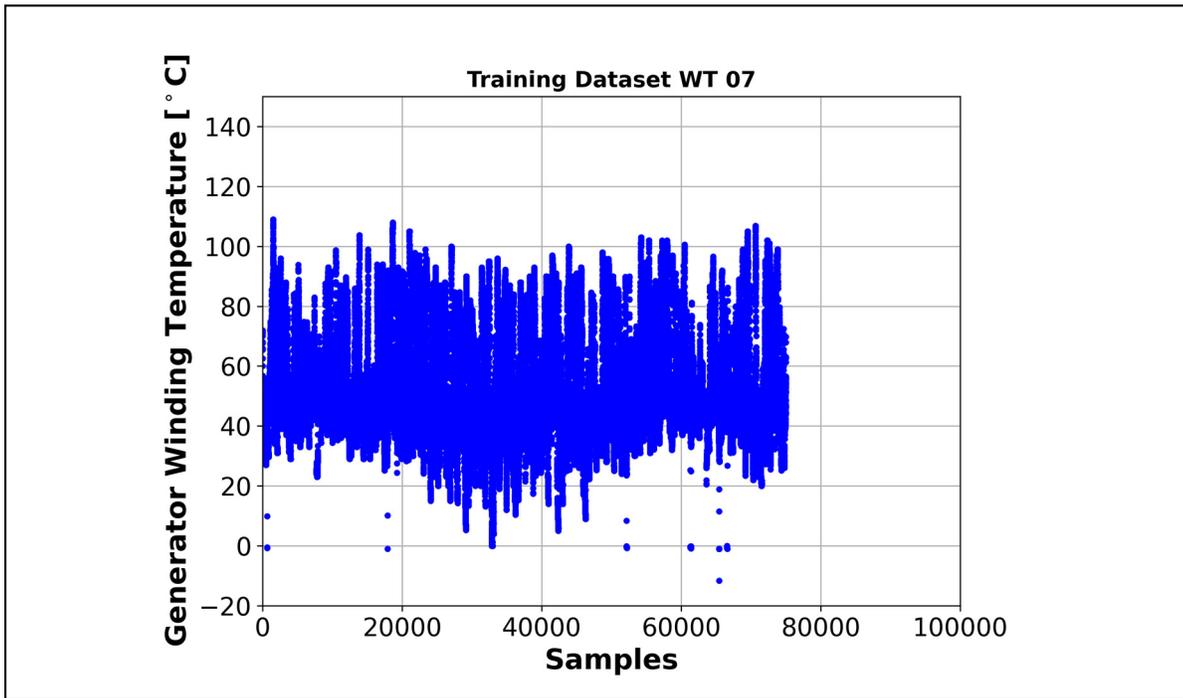


Figure 4.1 Training dataset: Generator Winding Temperature WT 7

Moreover, the testing dataset which is the last 25000 samples of generator windings temperature is shown in Fig. 4.4.

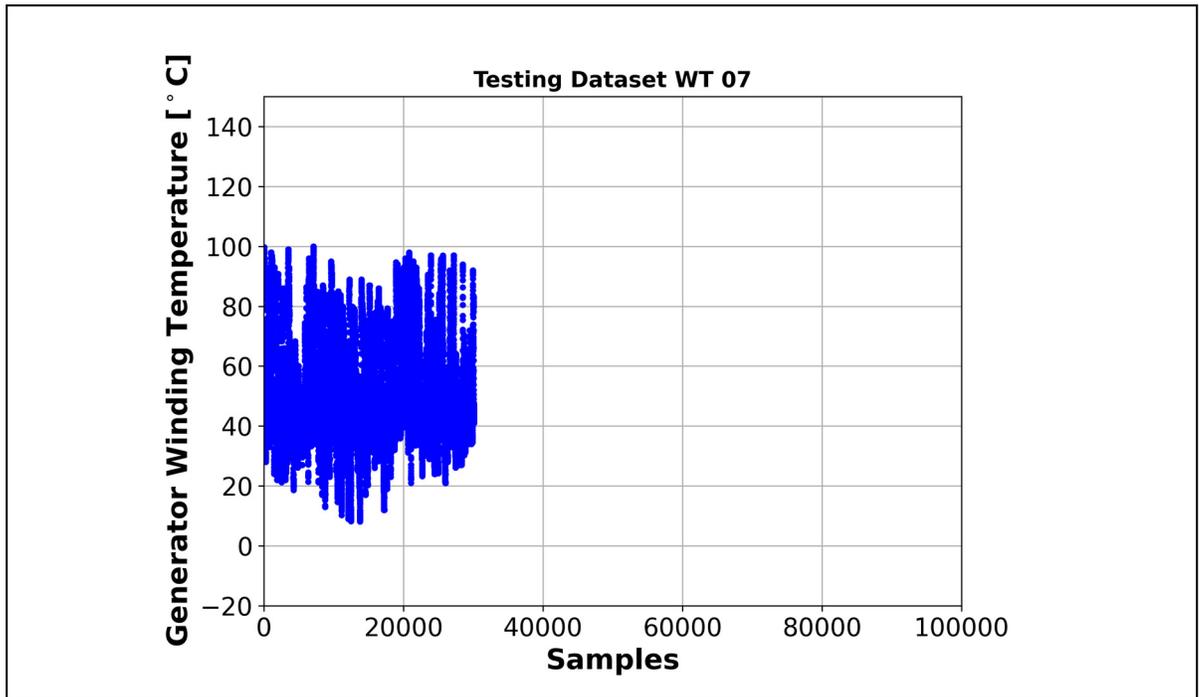


Figure 4.4. Testing dataset: Generator Winding Temperature WT 7

The predictions of Stacked LSTM model for wind turbine No. 7 and No. 67 are depicted in Fig. 4.5. and Fig. 4.6.

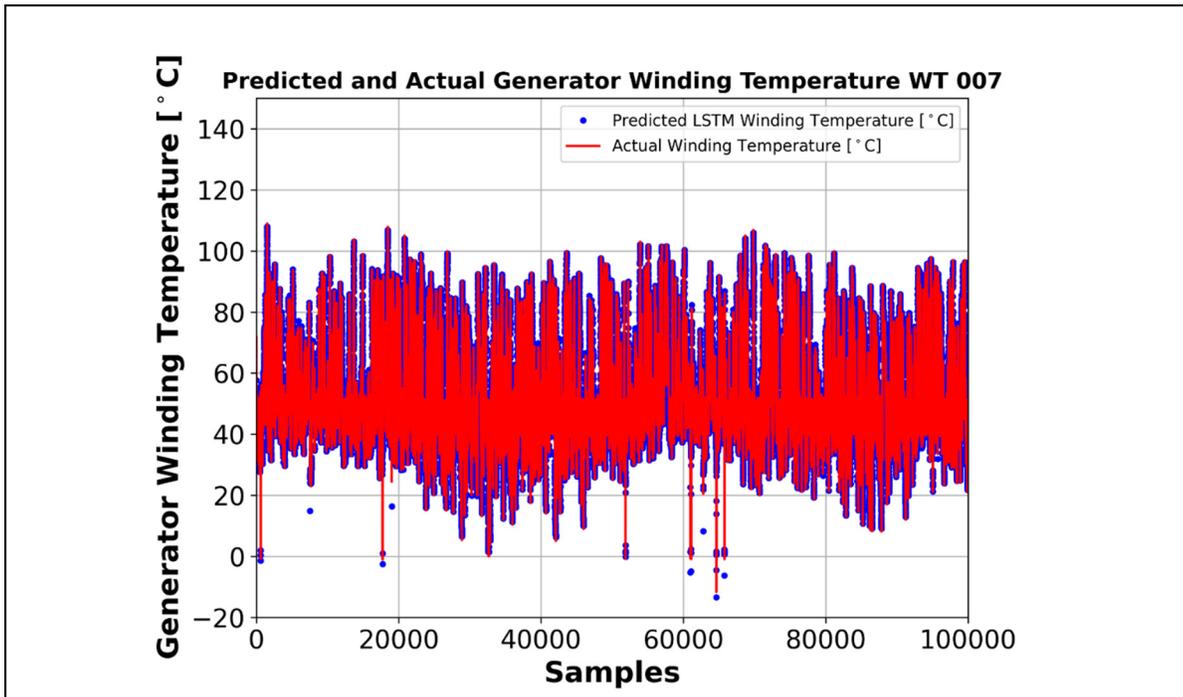


Figure 4.5 Predicted generator windings temperature by Stacked LSTM: WT 7

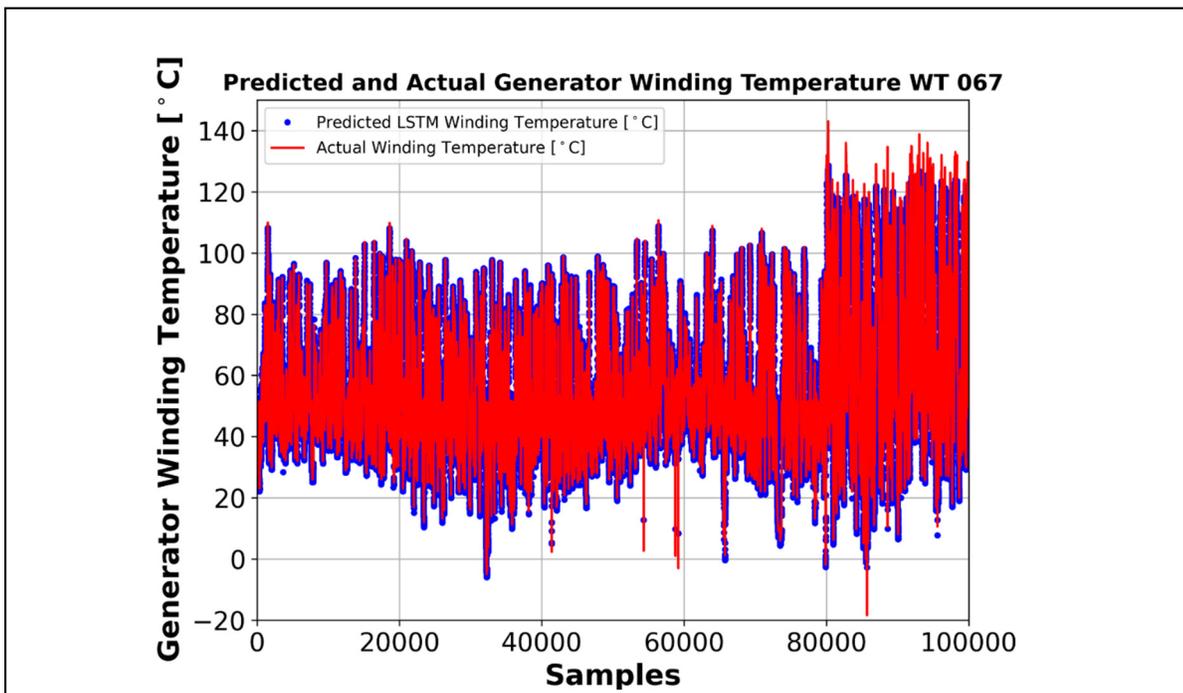


Figure 4.6. Predicted generator windings temperature by Stacked LSTM: WT 67

These figures show that Stacked LSTM model succeeded in predicting generator winding temperature with considerable precision for wind turbine No. 7. Moreover, the predictions for wind turbine No. 67 demonstrate that before the 80000th sample model predicted the temperature precisely; however, deviations between predicted and real temperature increases after 80000th sample due to unknown reason which increased the actual windings temperature.

4.2 Bidirectional LSTM

Another type of LSTM is bidirectional LSTM which consists of two hidden layers in the opposite direction for prediction of same value. In the other word, this LSTM exploits backward (past) and forward (future) states of system for predicting the current value. It should be mentioned that since Bidirectional LSTM uses forward states of system, it shows its best performance once datasets are complete; hence, it is not a proper choice for real-time prediction or monitoring problems while it shows an excellent performance for Speech Recognition and Handwritten Recognition. The structure of Bidirectional LSTM is shown in Fig. 4.7.

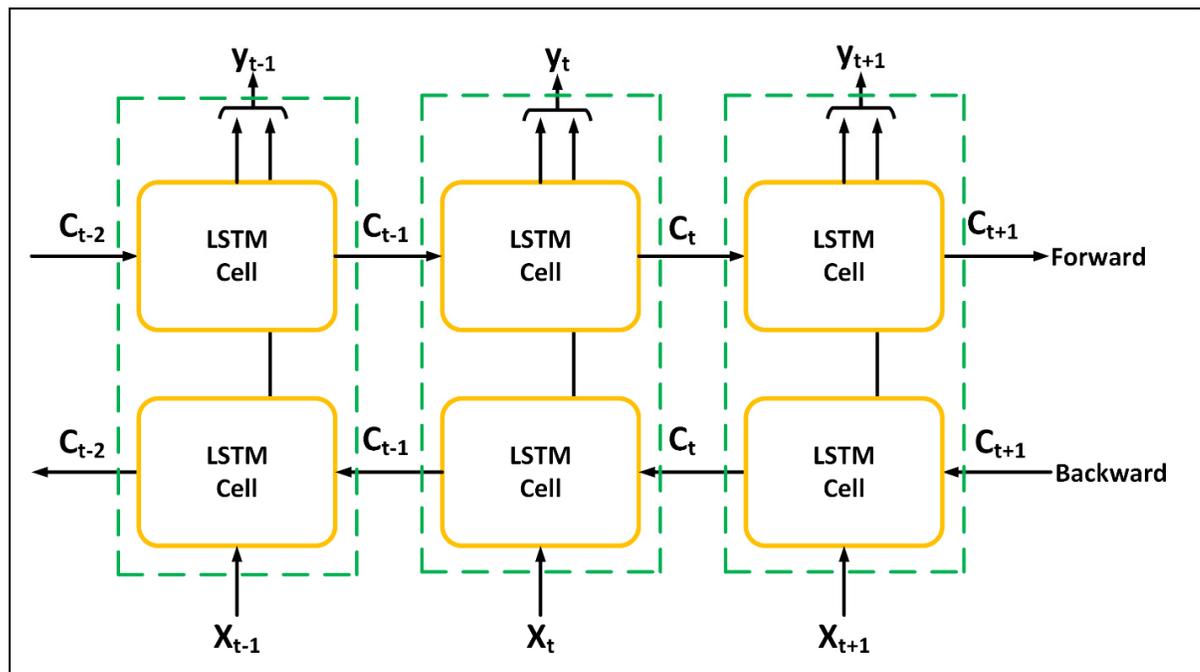


Figure 4.7. Bidirectional LSTM structure

Same case scenarios are chosen for analyzing Bidirectional LSTM. The results of Bidirectional LSTM model for wind turbine No. 7 and No. 67 are shown in Fig. 4.8. and Fig. 4.9., respectively.

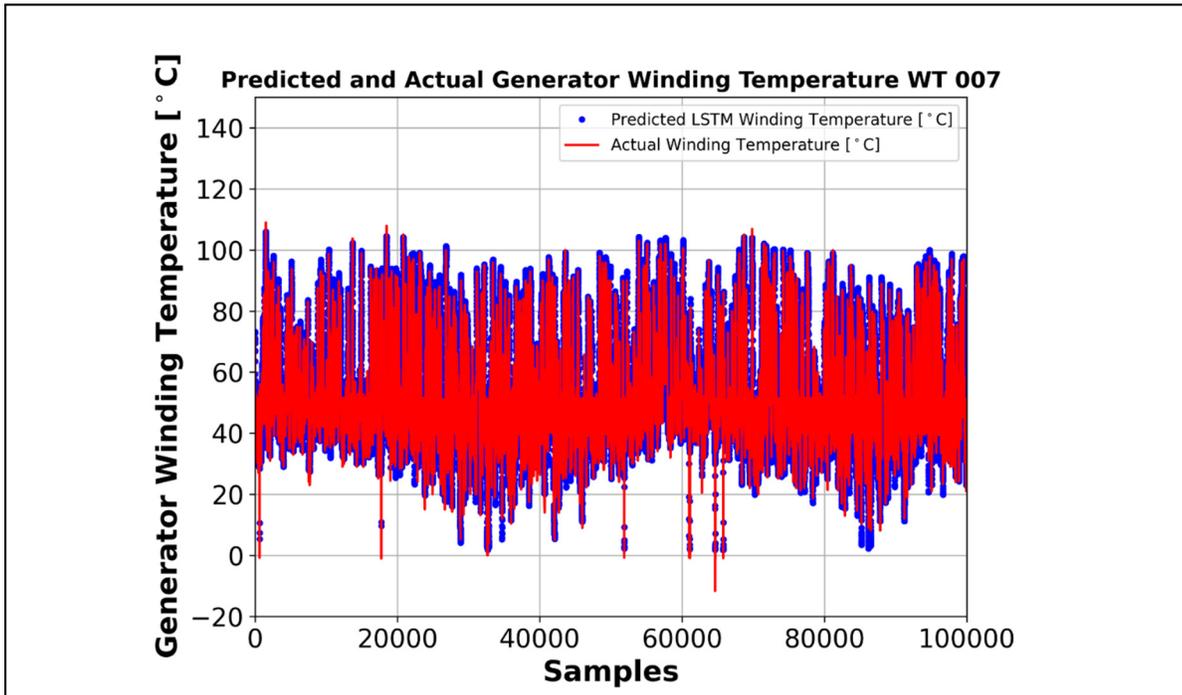


Figure 4.8. Predicted generator windings temperature by Bidirectional LSTM: WT 7

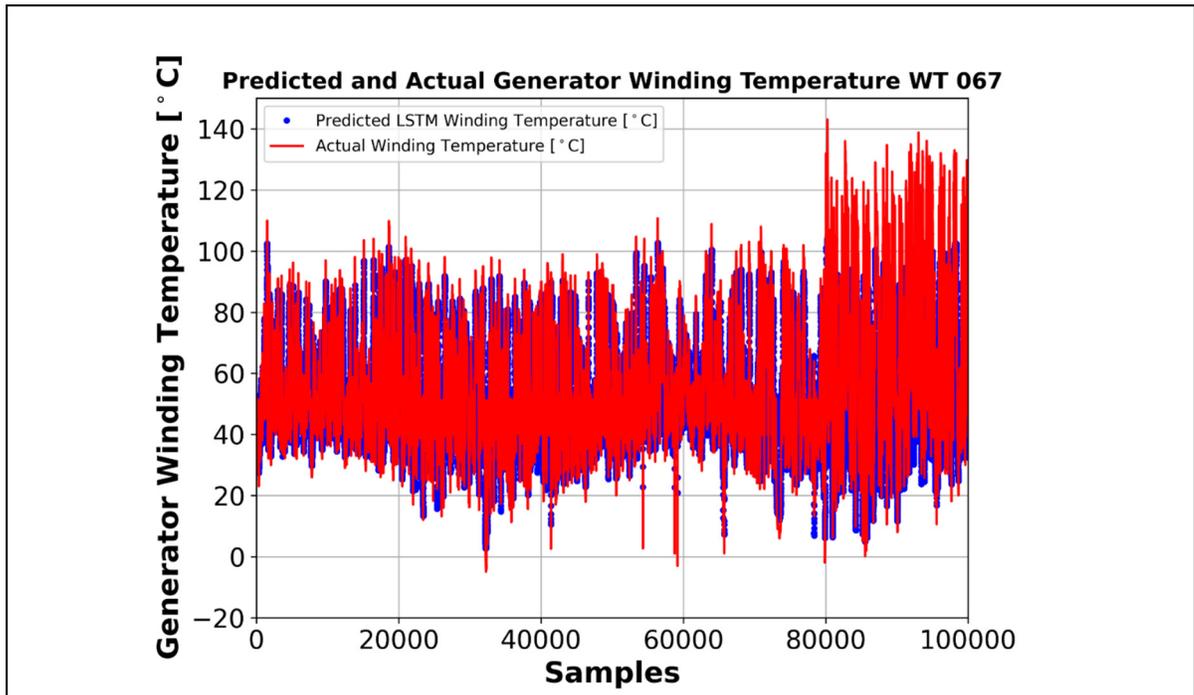


Figure 4.9. Predicted generator windings temperature by Bidirectional LSTM: WT 67

It can be understood from Fig. 4.8. that Bidirectional LSTM shows an acceptable performance in predicting the windings temperature when system is operating in healthy operation state; however, the model precision considerably decreases when thermal anomaly occurs after the 80000th sample in Fig 4.9.

4.3 Multivariate LSTM

As it was mentioned earlier, the first step for preprocessing the datasets are exploiting the correlation analysis. In this project, real datasets are used from a wind farm in the North of Quebec Province, Canada, which are collected in collaboration with Power Factors. These SCADA-based datasets are collected every 10 minutes for nearly 2 years and include 35 parameters. The prior engineering knowledge and correlation analysis are exploited to mitigate the input parameters of model for enhancing speed and performance. Among the 35 provided parameters, 7 parameters that are considered to be the most relevant ones for this modeling are chosen based on the engineering knowledge. The candidate parameters are listed in Table 4.1.

Table 4.1 The candidate parameters for inputs and output

Input Features	Target Variable
Nacelle Temperature	Temperature of Generator Windings
Generator Stator Current	
Produced Active Power	
Reactive Power	
Rotational Speed of Generator	
Temperature of Generator Cooling System	
Generator Stator Voltage	

The correlation analysis is exploited to increase computational speed by decreasing the number of input features. One simple way to identify whether features are linearly related to each other is by looking at the two-dimensional figures of them. The scatter figures of candidate features are shown in Fig. 4.10.

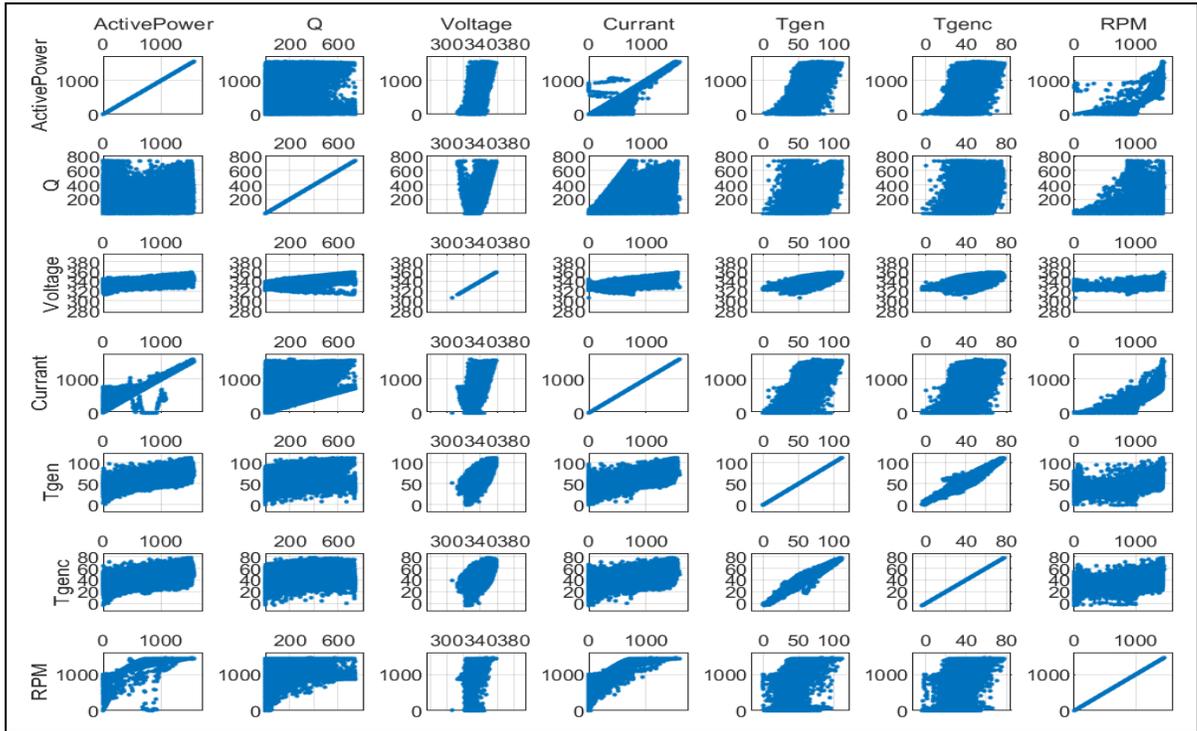


Figure 4.10 Scatter plots of candidate parameters

Where Q, Tgen, Tgenc, and RPM are reactive power, generator windings temperature, generator cooling system temperature, and generator rotational speed, respectively. Highly scattered plots show that features have low correlation and linear plots indicate that features are highly correlated. These visual representations of correlated features are supported by the mathematical calculation based on (3.1.) which its results are shown in Table 4.2.

Table 4.2 Correlation values of candidate parameters

	Active Power	Reactive Power	Voltage	Current	Tgen	Tgenc	RPM
Active Power	1	0.5637	0.8193	0.9881	0.8338	0.6644	0.8951
Reactive Power	0.5637	1	0.8161	0.6540	0.6117	0.4762	0.4109
Voltage	0.8193	0.8161	1	0.8742	0.8055	0.6482	0.6963
Current	0.9881	0.6540	0.8742	1	0.8544	0.6864	0.789
Tgen	0.8338	0.6117	0.8055	0.8544	1	0.9426	0.6598
Tgenc	0.6644	0.4762	0.6864	0.6864	0.9426	1	0.5575
RPM	0.8951	0.4109	0.789	0.789	0.6598	0.5575	1

Active Power, Voltage, Current, and Tgenc are highly correlated with Tgen, which is the target variable in this study. Moreover, the correlation analysis shows that voltage, current, and

rotational speed have a high correlation with produced active power. Hence, the candidate features for the regression model narrow down to the temperature of generator cooling system and produced active power. The generator cooling system temperature and produced active power for wind turbine No. 7 and No. 67 are shown in Fig. 4.11., Fig. 4.12., Fig. 4.13., and Fig. 4.14., respectively.

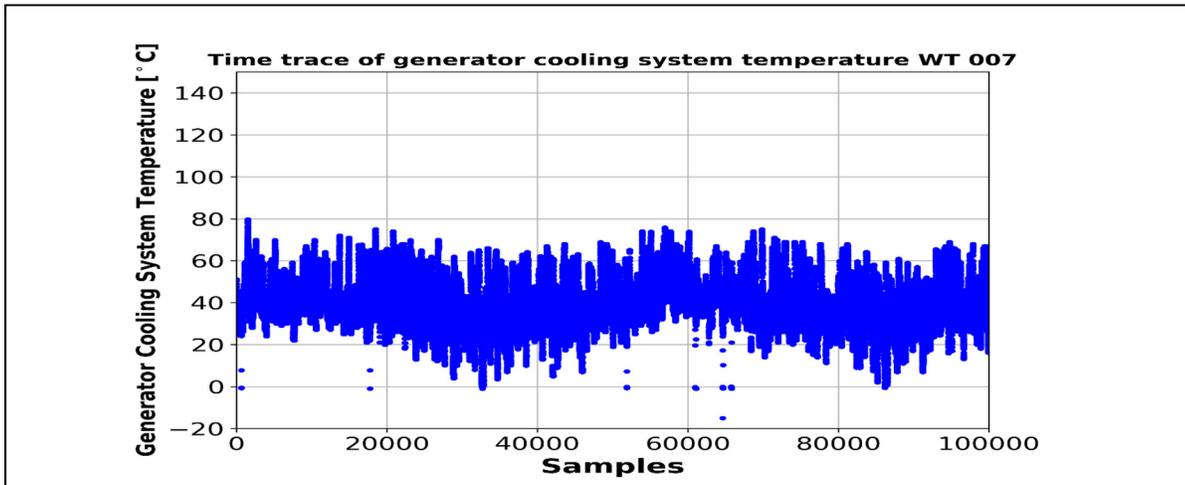


Figure 4.11. Generator cooling system temperature: WT 7

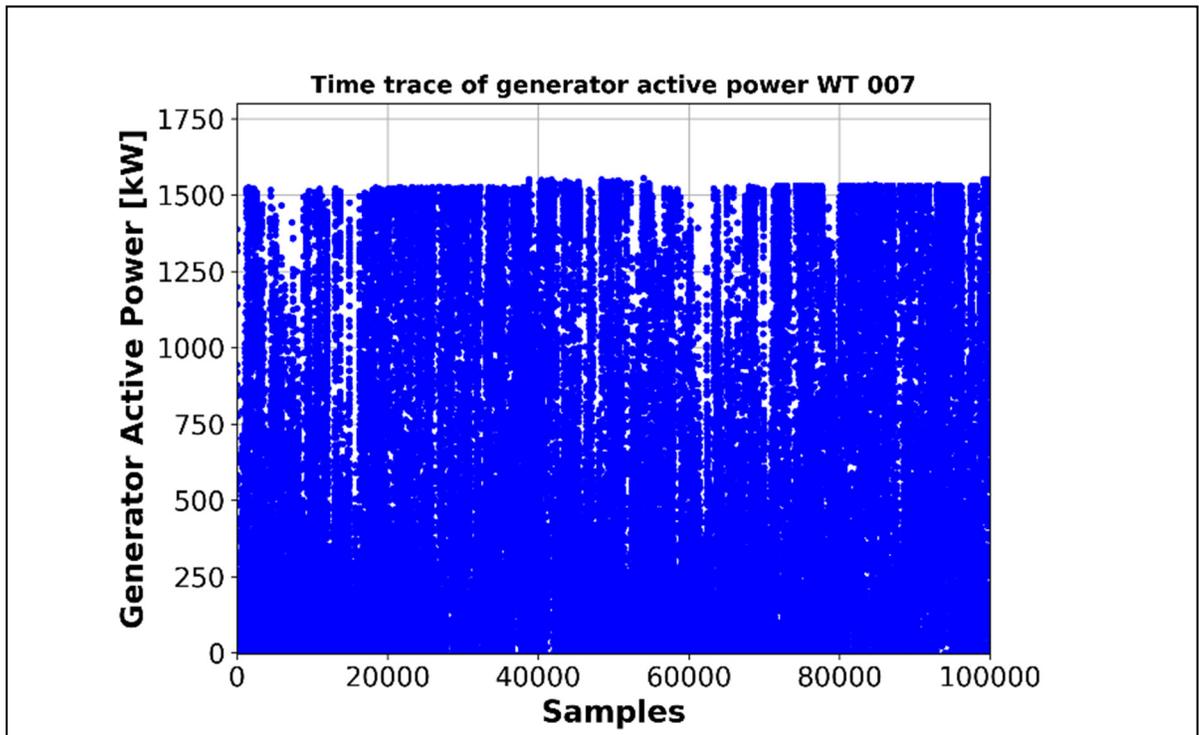


Figure 4.2 Produced active power: WT 7

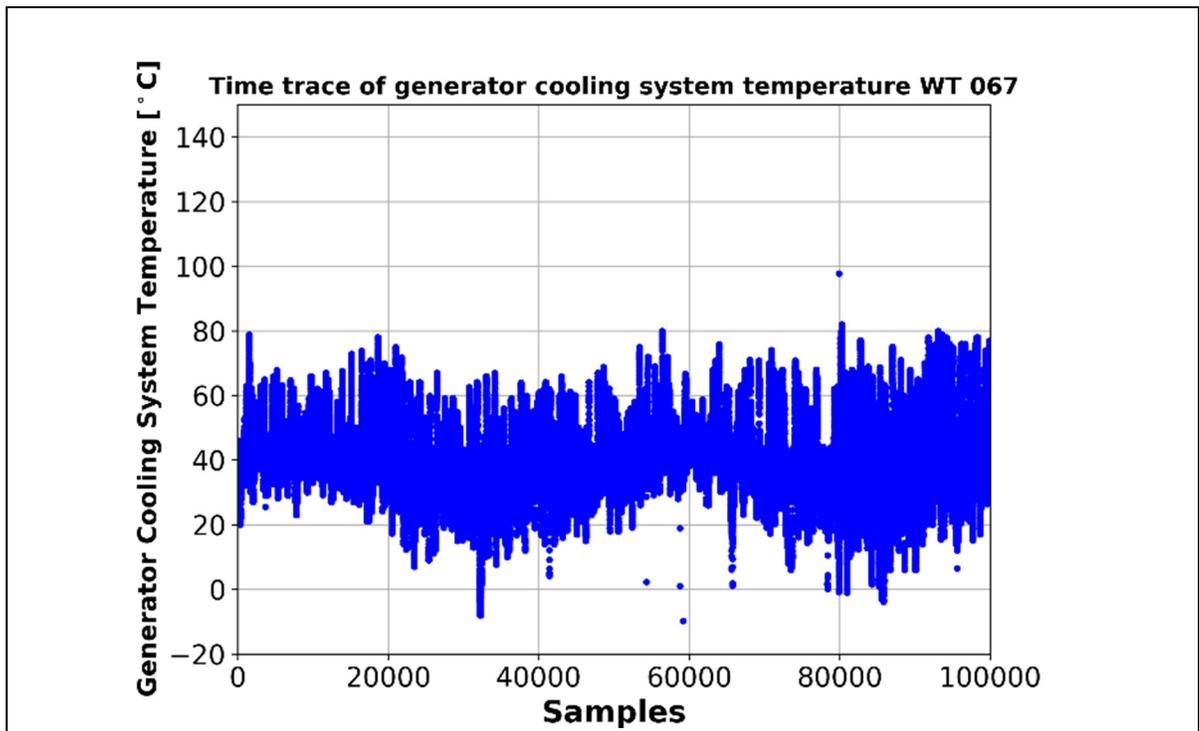


Figure 4.3 Generator cooling system temperature: WT 67

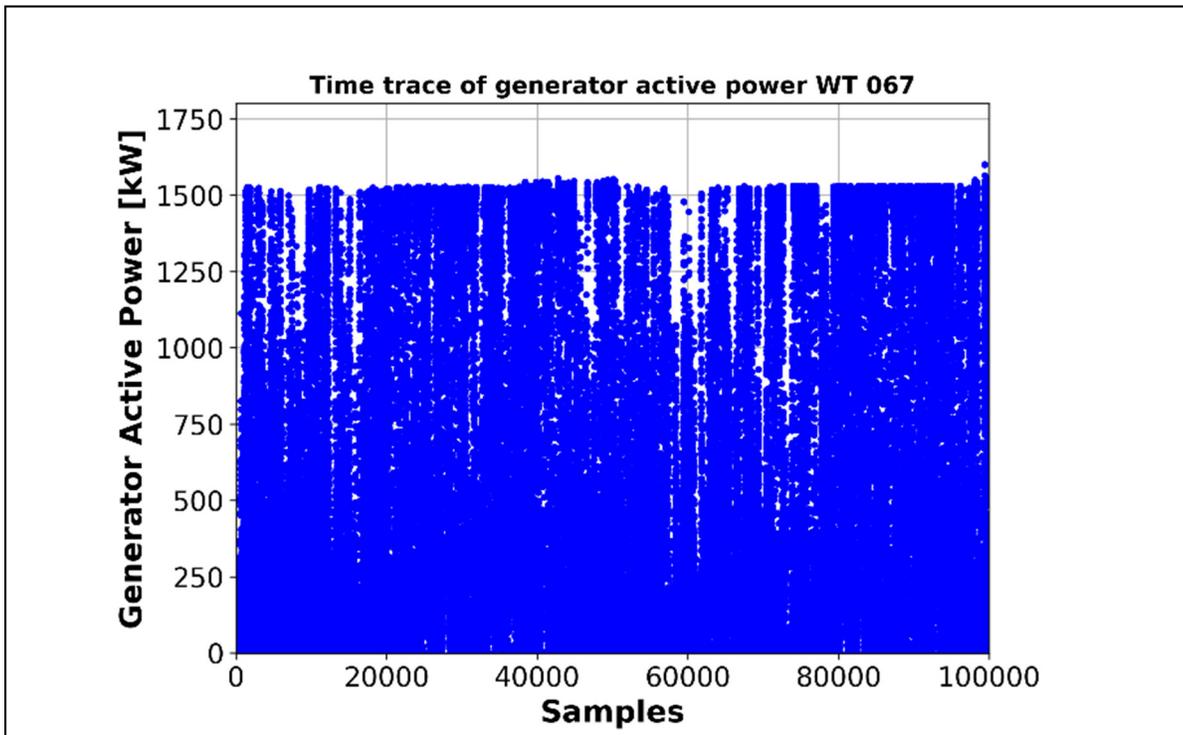


Figure 4.14 Produced active power: WT 67

These figures demonstrate that there is not any visual difference between candidate parameters variations in case scenarios 1 and 2. The results of developed Multivariate LSTM for Case scenarios 1 and 2 are shown in Fig. 4.15. and Fig. 4.16.

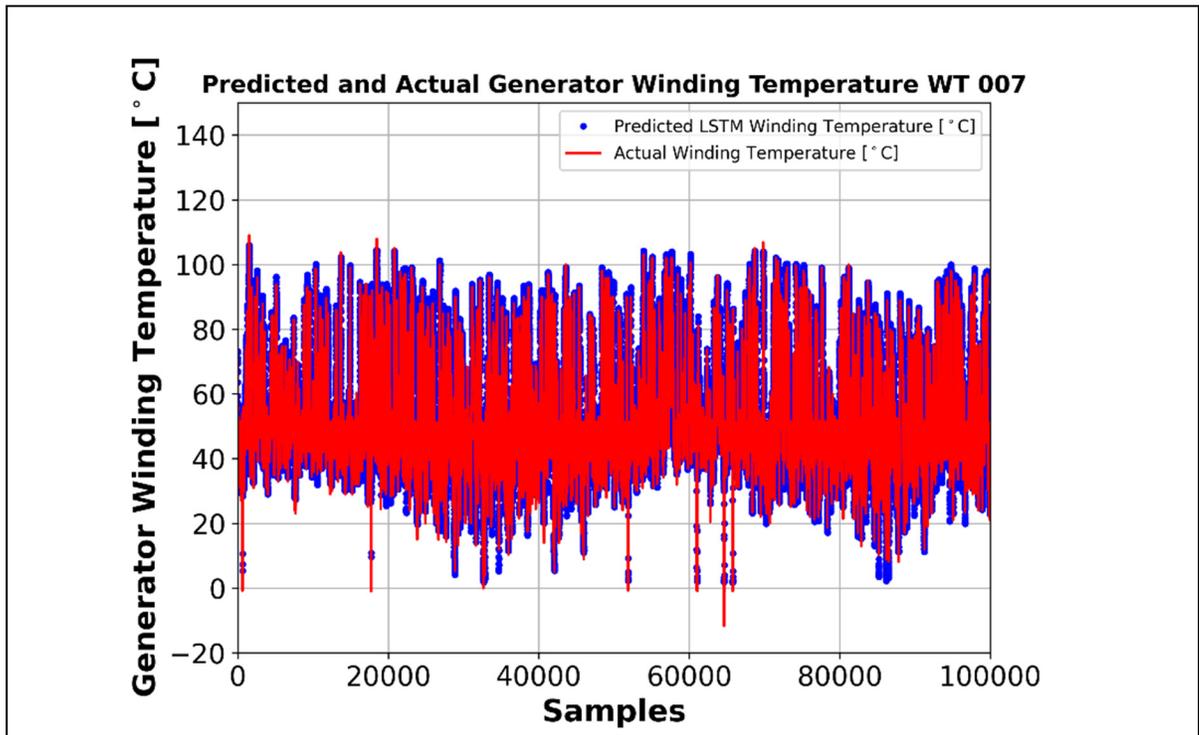


Figure 4.4 Predicted generator windings temperature by Multivariate LSTM: WT 7

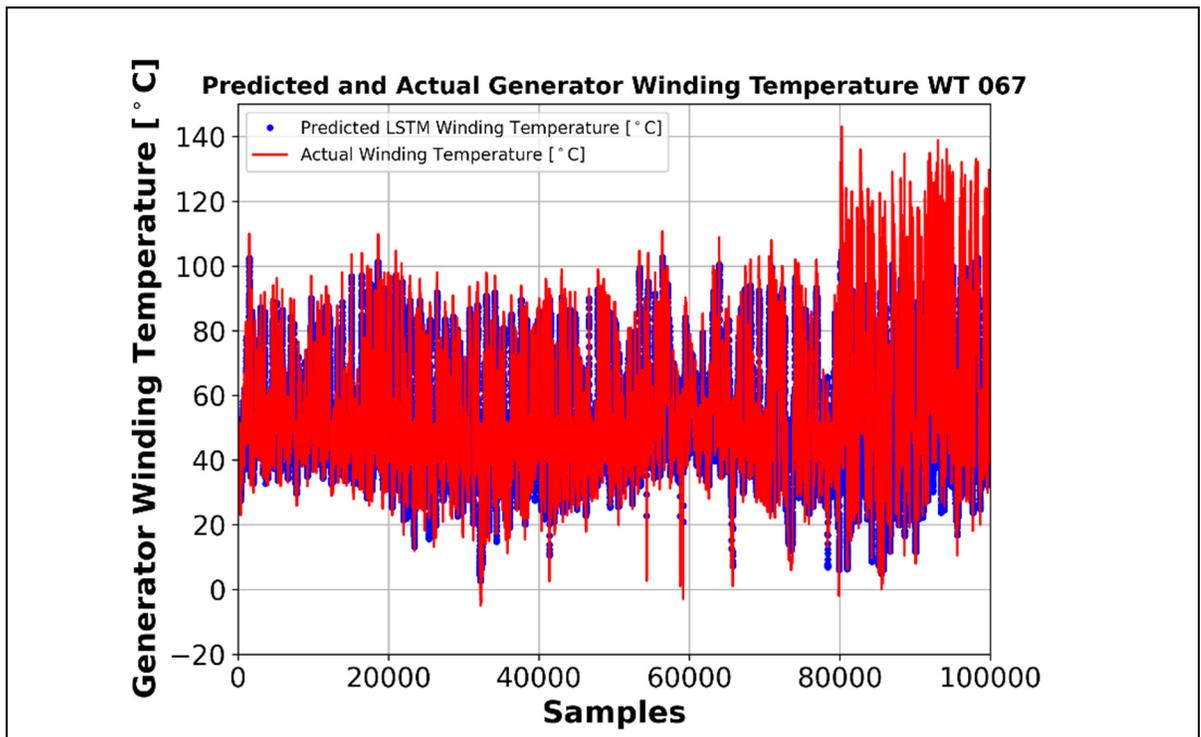


Figure 4.5 Predicted generator windings temperature by Multivariate LSTM: WT 67

One of the key assumptions of regression is that the model residuals should follow a normal distribution in order to guarantee proper performance of the model. Hence, the residuals histogram of model is extracted to ensure that they have a normal distribution. The residual histograms are shown in Fig. 4.17.

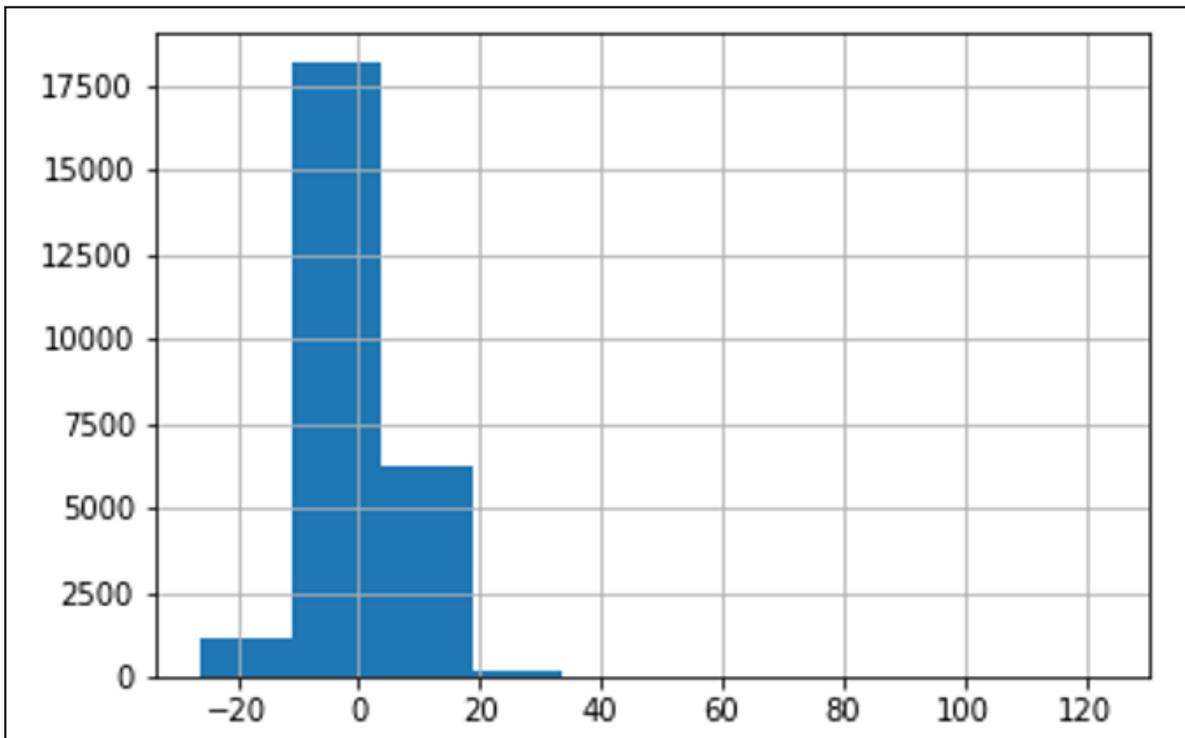


Figure 4.6 Histogram of model residuals

To have a better representation, the density plot of model residuals is drawn in Fig. 4.18.

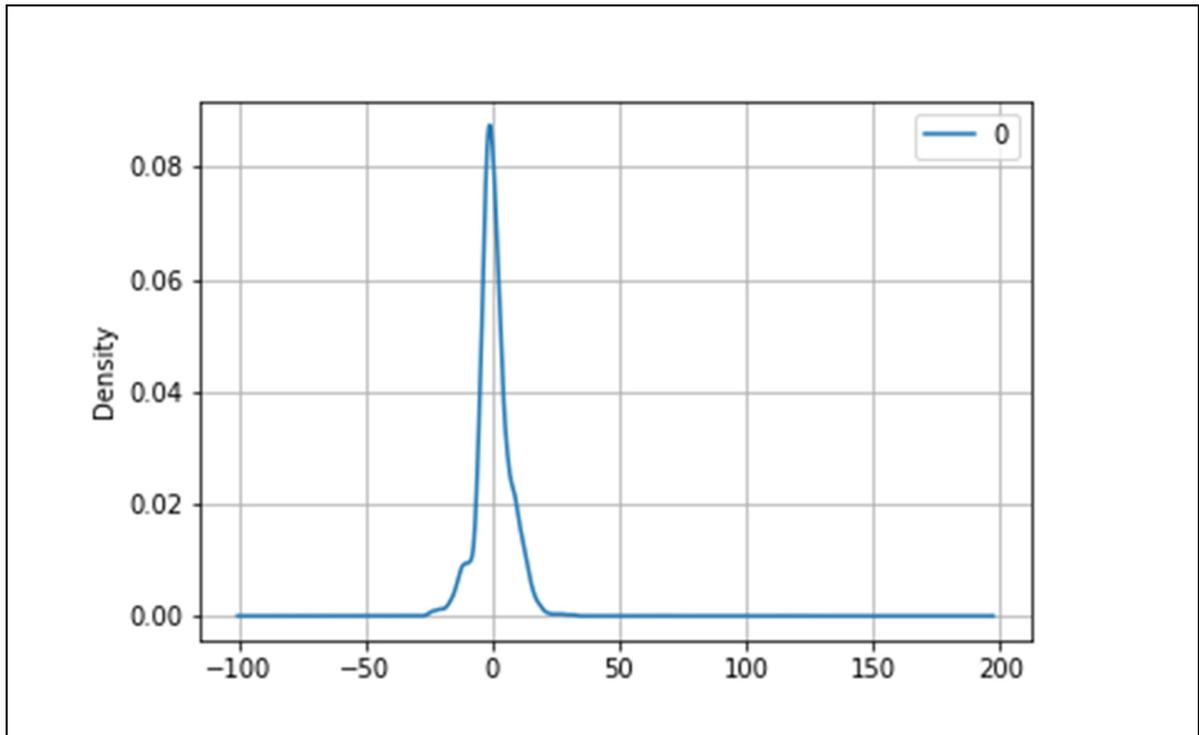


Figure 4.7. Density plot of model residuals

It can be seen that the residuals follow a normal distribution which indicates one of the key assumptions of regression is met in this model.

4.4 Performance Comparison

The three kinds of developed LSTM model are trained roughly in the same condition to make the comparison fair. In the following table, the hyperparameters of developed LSTM model topology is presented.

Table 4.3 Tuned hyperparameters for developed LSTM models

Hyperparameters	Designated values
Learning Rate	0.001
Training dataset ratio	75%
Time steps	12
Number of hidden layers	3
Number of neurons in each hidden layer	30
Dropout rate	0.1
Optimizer	ADAM
Epochs	10
Batch size	256
Loss function	Mean Square Error

The developed LSTM topology followed the same structure suggested in Table 4.3. The results from Stacked, Bidirectional, and Multivariate LSTMs are compared based on MSE and R-squared criteria. Moreover, the obtained LSTM models results are compared with MLR results to show excellent performance of LSTM over Linear Regression for handling nonlinear time-series datasets. The tabulated error comparison of developed models is shown in Table 4.4.

Table 4.4 MSE and R-squared values for different models

	Case 1				Case 2			
	MLR	Stacked LSTM	Bidirectional LSTM	Multivariate LSTM	MLR	Stacked LSTM	Bidirectional LSTM	Multivariate LSTM
MSE	8.71	1.81	1.83	5.01	153.38	8.5	6.83	33.84
R²	0.997698	0.999378	0.999371	0.998423	0.985088	0.998712	0.998767	0.986900

The graphical representation of models errors is shown in Fig. 4.19.



Figure 4.8 The graphical representation of models performance

Where S_LSTM, B_LSTM, M_LSTM, and MLR are stand for Stacked LSTM, Bidirectional LSTM, Multivariate LSTM, and Multiple Linear Regression, respectively. It can be understood from Fig. 4.4., Fig.4.7., and Fig. 4.14. that the developed LSTM models show acceptable performance in predicting the winding temperature in abnormal condition; however, Stacked and Bidirectional LSTM have lower MSE in Case 1 since the target and feature parameters are the same. On the other hand, Multivariate LSTM uses other parameters as input features which obviously have lower correlation with target parameter. Hence, the MSE error for Multivariate LSTM is higher than the other two LSTMs. Moreover, Bidirectional LSTM has lower MSE error than Stacked LSTM in Case 2 since it is using forward states of system for predicting the current target variable. In terms of scalability of model, it is worth mentioning that since the model shows acceptable results for other case scenarios and other wind turbines, it can be used for large scale of wind farms with a lot of wind turbines.

CONCLUSION

As it was mentioned earlier, large portion of costs of a wind farm is related to condition monitoring and O&M. Among different existing strategies for mitigation of O&M costs, the methods which use SCADA-based datasets are more popular since these systems are already installed which can be interpreted as decrease in expenses. In this project, several classic approaches for condition monitoring along side the most common failures in wind farms are described and it deduced that the classic methods cannot play a cardinal role in mitigation of related costs. In the recent decade, ML algorithms emerged as one of the most potent and proficient tools to overcome many industrial challenges. Hence, many DL models are developed in the literature to make condition monitoring in wind farms easier and more accurate. In this project, the main established methods for condition monitoring in wind farms are investigated and RNN is set as the focal point of interest due to their considerable capabilities for handling time-series data.

In this project, several preprocessing techniques for missing value and outliers and scaling techniques are exploited to prepare the data for ML models. As the first model, a simple linear regression model is developed to predict the winding temperature of generators. Moreover, correlation analysis is used to reduce the dimensionality of inputs features for Multivariate modeling from more than 35 features to 2 variables. Furthermore, different types of LSTM such as Stacked LSTM, Bidirectional LSTM, and Multivariate LSTM are developed to predict temperature with higher precision. To demonstrate the effectiveness of LSTM models, these models are evaluated based on different criteria such as MSE and R-squared. For this purpose, two case scenarios are considered based on the performance of a wind farm in the North of Quebec Province, Canada in collaboration with Power Factors. The first case scenario is designed to show the performance of models in normal operation mode and the second one assessed models' performance in the generator's over-temperature operation mode. The results showed that in the first case scenario Stacked and Bidirectional LSTMs have the same and the best results while Bidirectional LSTM showed the best performance in the second scenario due to its ability.

RECOMMENDATIONS

Based on the obtained results from this project, the following possible areas for future works can be explored:

- Since the discussed models in this project are designed specifically for time-series datasets, they can be used for fault and underperformance detection in other equipment of wind farms.
- The proposed models in this project can be exploited for other wind turbine or wind farms and can be reevaluated in different case scenarios to demonstrate their proper performance.
- The developed models can be deployed in a real-time analysis to evaluate their performance in real-life situation.
- As a suggestion, different classification algorithms can be developed to enhance the output of existing LSTM model. The classification models can be used to determine the root causes of failures or anomalies in the output of LSTM models.
- Dimensionality reduction is an important topic which can be more scrutinized by exploiting different correlation techniques.
- Convolutional Neural Networks (CNNs) can be used for dimensionality reduction in the input of LSTM model and its performance can be compared with the correlation analysis.
- Another important area for future work is sensitivity analysis to the hyperparameters of proposed models. Grid Search techniques can be used to find the most optimized values for hyperparameters.
- Other types of RNNs can be developed for predicting the temperature of generator windings. The obtained results can be compared with them to show the models performance.

BIBLIOGRAPHY

- Artigao, E., Honrubia-escribano, A. & Gomez-lazaro, E. (2018). Current signature analysis to monitor DFIG wind turbine generators : A case study. *Renewable Energy*, 116, pp. 5–14. DOI: <https://doi.org/10.1016/j.renene.2017.06.016>.
- Artigao, E., Honrubia-Escribano, A., Martín-Martínez, S. & Gómez-Lázaro, E. (2020), The Use of Electrical Measurements of Wind Turbine Generators for Drive Train Condition Monitoring, In Maalawi, K. (Ed) *Design Optimization of Wind Energy Conversion Systems with Applications*, (pp. 69-90), United Kingdom : Butterworth-Heinemann. DOI: 10.5772/intechopen.90127.
- Aziz, U., Charbonnier, S., Bérenguer, C., Lebranchu, A. & Prevost, F. (2019). *SCADA data based realistic simulation framework to evaluate environmental impact on performance of wind turbine condition monitoring systems*. The 4th Conference on Control and Fault Tolerant Systems (SysTol), Casablanca, Morocco, pp. 360-365. DOI: 10.1109/SYSTOL.2019.8864769.
- Blancke, O., Merkhouf, A., Amyot, N., Hudon, C. & Haddad, K. (2016). *Strategic Fault Diagnosis Approach for Hydrogenerator Shaft Current Discharges*. 2016 XXII International Conference on Electrical Machines (ICEM), Lausanne, pp. 2346–2351. DOI: 10.1109/ICELMACH.2016.7732849.
- Breiman, L. & Friedman, J.H. (1997). Predicting Multivariate Responses in Multiple Linear Regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(1), pp. 3-54. DOI:10.1111/1467-9868.00054.
- Canadian Wind Energy Association (2020). Installed Capacity. Retrieved from: <https://canwea.ca/wind-energy/installed-capacity/#:~:text=Continuing%202018's%20growth%2C%20Canada%20finished,over%20%241%20billion%20of%20investment>.
- Canadian Wind Energy Association (2017). Canada's growing wind turbine fleet generates new business opportunities. Retrieved from: <https://canwea.ca/operations-and-maintenance/>.
- Deng, L., Hou, Z., Liu, H. & Sun, Z. (2019), *A Fractional Hilbert Transform Order optimization Algorithm Based DE for Bearing Health Monitoring*. Chinese Control Conference (CCC), Guangzhou, China, pp. 2183-2186. DOI: 10.23919/ChiCC.2019.8865403.

- Finnegan, W., Flanagan, T. & Goggins, J. (2020). *Development of a Novel Solution for Leading Edge Erosion on Offshore Wind Turbine Blades*. Proceedings of the 13th International Conference on Damage Assessment of Structures, Singapore, pp. 517-528. DOI: https://doi.org/10.1007/978-981-13-8331-1_38.
- García Vera, Y.E., Dufo-López, R. & Bernal-Agustín, J.L. (2019) Energy Management in Microgrids with Renewable Energy Sources: A Literature Review. *Applied Sciences*. 9(18), p. 3854. DOI: <https://doi.org/10.3390/app9183854>.
- Gong, X., Qiao, W. & Member, S. (2015). Current-Based Mechanical Fault Detection for Direct-Drive Wind Turbines via Synchronous Sampling and Impulse Detection. *IEEE Transactions on Industrial Electronics*, 62(3), pp. 1693–1702. DOI: 10.1109/TIE.2014.2363440.
- Gray, C. & Watson, S. (2010). Physics of failure approach to wind turbine condition based maintenance. *Wind Energy*, 13(5), pp. 395–405. DOI: <https://doi.org/10.1002/we.360>.
- Hahn, B., Durstewitz, M., & Rohrig, K. (2006). *Reliability of Wind Turbines Experiences of 15 years with 1500 WTs*. Retrieved from: https://www.researchgate.net/publication/46383070_Reliability_of_wind_turbines_Experiences_of_15_years_with_1500_WTs.
- Halpin, S., Harley, K., Jones, R. & Taylor, L. (2008). Slope-permissive under-voltage load shed relay for delayed voltage recovery mitigation. *IEEE Transactions on Power Systems*, 23(3), pp. 1211–1216. DOI: 10.1109/TPWRS.2008.926409.
- Helander, J., Member, S., Ericsson, A., Member, S., Gustafsson, M., Member, S., Martin, T., Member, S., Sjöberg, D., Larsson, C. & Member, S. (2017) Compressive Sensing Techniques for mm-Wave Nondestructive Testing of Composite Panels. *IEEE Transactions on Antennas and Propagation*, 65(10), pp. 5523–5531. DOI: 10.1109/TAP.2017.2738034.
- Hinton, G., Srivastava, N., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *ArXiv*. Retrieved from: <https://arxiv.org/pdf/1207.0580.pdf>.
- Hochreiter, S. (1998). The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*. 6(2), pp. 107-116. DOI: <https://doi.org/10.1142/S0218488598000094>.

- Hu, Y. & Chen, L. (2018). A nonlinear hybrid wind speed forecasting model using LSTM network, hysteretic ELM and Differential Evolution algorithm. *Energy Conversion and Management*, 173, pp. 123-142.
DOI: <https://doi.org/10.1016/j.enconman.2018.07.070>.
- Inductive Automation (2018). What is SCADA?.
Retrieved from: <https://inductiveautomation.com/resources/article/what-is-scada>.
- International Electrotechnical Commission (IEC), (2004). *Rotating electrical machines – Part1: Rating and performance (IEC 60034-1)*. Standard IEC 60034-1. Retrieved from: <https://archive.org/details/gov.in.is.iec.60034.1.2004/page/n3>.
- International Organization for Standardization. (2005). *Reciprocating internal combustion engine driven alternating current generating sets -- Part 3: Alternating current generators for generating sets (ISO 8528-3)*. Standard ISO 8528-3. Retrieved from: <https://www.iso.org/standard/39045.html>.
- Jantara, V. & Papaalias, M. (2020), Chapter 5 - Wind turbine gearboxes: Failures, surface treatments and condition monitoring, In Papaalias, M., Márquez, F. & Karyotakis, A. (Eds) *Non-Destructive Testing and Condition Monitoring Techniques for Renewable Energy Industrial Assets*, (pp. 69-90), United Kingdom : Butterworth-Heinemann.
- Jiang Y. & Srivastava, A. (2020). Data-Driven Event Diagnosis in Transmission Systems With Incomplete and Conflicting Alarms Given Sensor Malfunctions. *IEEE Transactions on Power Delivery*, 35(1), pp. 214-225. DOI: 10.1109/TPWRD.2019.2947671.
- Kingma, D. & Ba, J. (2015). *Adam: A Method for Stochastic Optimization*. *International Conference on Learning Representations*. The 3rd International Conference for Learning Representations, San Diego. Retrieved from: <https://arxiv.org/abs/1412.6980>.
- Kusiak, A. & Verma, A. (2012). A Data-Mining Approach to Monitoring Wind Turbines. *IEEE Transactions on Sustainable Energy*, 3(1), pp. 150–157.
DOI: 10.1109/TSTE.2011.2163177.
- Lei, J., Liu, C. & Jiang, D, (2019). Fault diagnosis of wind turbine based on Long Short-term memory networks. *Renewable Energy*, 133, pp. 422-432.
DOI: <https://doi.org/10.1016/j.renene.2018.10.031>.
- Lu, D., Member, S., Qiao, W., Gong, X. & Member, S. (2013). *Current-Based Fault Detection for Wind Turbine Systems via Hilbert-Huang Transform*. 2013 IEEE Power & Energy Society General Meeting, Vancouver, pp. 1–5. DOI: 10.1109/PESMG.2013.6672999.

- Luo, J., Lin, K., Li, J., Xue, Y. & Zhang, X. (2019). *Cost analysis and comparison between modular multilevel converter (MMC) and modular multilevel matrix converter (M3C) for offshore wind power transmission*. The 15th IET International Conference on AC and DC Power Transmission (ACDC 2019), Coventry, UK, pp. 1-6. DOI: 10.1049/cp.2019.0063.
- McCulloch, W. & Pitts, W. (1943). A logical calculus of ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5, pp. 115-133. Retrieved from: <https://link.springer.com/article/10.1007/BF02478259>.
- McNutt, W. (1992). Insulation thermal life considerations for transformer loading guides. *IEEE Transactions on Power Delivery*, 7(1), pp. 392-401. DOI: 10.1109/61.108933.
- Moeini, R., Entezami, M., Ratkovac, M., Tricoli, P., Hemida, H., Hoeffler, R. & Baniotopoulos, C. (2019). Perspectives on condition monitoring techniques of wind turbines. *Wind Engineering*, 43(5), pp. 539–555. DOI: <https://doi.org/10.1177/0309524X18807028>.
- Mustafa, A., Barabadi, A. & Markeset, T. (2019). *Risk assessment of wind farm development in ice proven area*. Proceedings of International Conference on Port and Ocean Engineering under Arctic Conditions, Norway. Retrieved from : <http://www.poac.com/Papers/2019/pdf/POAC19-113.pdf>.
- National Electrical Manufacturers Association (NEMA), (2019). *NEMA MG 1: Motors and Generators (MG 1-2009)*, Standard MG 1-2009. Retrieved from: <https://law.resource.org/pub/us/cfr/ibr/005/nema.mg-1.2009.pdf>.
- Orleans, N. (2013). Using SCADA data for wind turbine condition monitoring – a review. *IET Renewable Power Generation*, 11(4), pp. 382 - 394. DOI: 10.1049/iet-rpg.2016.0248.
- Østergaard, P., Duic, N., Noorollahi, Y., Mikulcic, H. & Kalogirou, S. (2020). Sustainable development using renewable energy technology, *Renewable Energy*, 146, pp. 2430-2437. DOI: <https://doi.org/10.1016/j.renene.2019.08.094>.
- Patrizi, G., Ciani, L., Guidi, G. & Bartolini, A. (2019). *Condition Monitoring of Wind Farm based on Wireless Mesh Network*. The 16th IMEKO TC10 Conference, Berlin, Germany, pp. 39-44. Retrieved from : <https://www.imeko.org/publications/tc10-2019/IMEKO-TC10-2019-004.pdf>.

- Papatzimos, A., Thies, P., Lonchamp, J., Joly, A. & Dawood, T. (2019). *Data-Informed Lifetime Reliability Prediction for Offshore Wind Farms*. IEEE International Conference on Prognostics and Health Management (ICPHM), San Francisco, CA, USA, pp. 1-8. DOI: 10.1109/ICPHM.2019.8819378.
- Peyghami, S., Blaabjerg, F. & Palensky, P. (2020), Incorporating Power Electronic Converters Reliability into Modern Power System Reliability Analysis. *IEEE Journal of Emerging and Selected Topics in Power Electronics*. DOI: 10.1109/JESTPE.2020.2967216.
- Qiao, W. & Lu, D. (2015). A Survey on Wind Turbine Condition Monitoring and Fault Diagnosis — Part II: Signals and Signal Processing Methods. *IEEE Transactions on Industrial Electronics*, 62(10), pp. 6546–6557. DOI: 10.1109/TIE.2015.2422394.
- Salem, A., Abu-siada, A. & Islam, S. (2017). Improved condition monitoring technique for wind turbine gearbox and shaft stress detection. *IET Science, Measurement & Technology*, 11(4), pp. 431-437. DOI: 10.1049/iet-smt.2016.0338.
- Santos, M. & González, M. (2019). Factors that influence the performance of wind farms, *Renewable Energy*, 135, pp. 643-651.
DOI: <https://doi.org/10.1016/j.renene.2018.12.033>.
- Sarma, N., Djurovi, S., Zappal, D., Crabtree, C., Mohammad, A. & Tavner, P. (2019). Electrical & mechanical diagnostic indicators of wind turbine induction generator rotor faults. *Renewable Energy*, 131, pp. 14–24.
DOI: <https://doi.org/10.1016/j.renene.2018.06.098>.
- Setiawan, A., Sugeng, Koesoema, K., Bakhri, S. & Aditya, J. (2018). *The SCADA system using PLC and HMI to improve the effectiveness and efficiency of production processes*. The 1st Siliwangi International Conference on Innovation in Research 2018 (SICIR), Bandung, Indonesia. 550, 012008. DOI: <https://doi.org/10.1088/1757-899X/550/1/012008>.
- Shen, J., Francis, R., Miller, L., Carriveau, R., Ting, D. S., Rodgers, M. & Davis, J. (2019). Geographic information systems visualization of wind farm operational data to inform maintenance and planning discussions. *Wind Engineering*.
DOI : <https://doi.org/10.1177/0309524X19862757>.
- Shi, X., Lei, X., Huang, Q., Huang, S., Ren, K. & Hu, Y. (2018). Hourly Day-Ahead Wind Power Prediction Using the Hybrid Model of Variational Model Decomposition and Long Short-Term Memory. *Energies*, 11(11), pp. 3227.
DOI: <https://doi.org/10.3390/en11113227>.

- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15, pp. 1929–1958.
- Thompson, B. (2005). Canonical Correlation Analysis, In Everitt, S. & Howell, C. (Eds) *Encyclopedia of Statistics in Behavioral Science*. DOI: 10.1002/0470013192.bsa068.
- Wang, J., Liang, Y., Zheng, Y., Gao, R. & Zhang, F. (2020), An integrated fault diagnosis and prognosis approach for predictive maintenance of wind turbine bearing with limited samples. *Renewable Energy*, 145, pp. 642-650.
DOI: <https://doi.org/10.1016/j.renene.2019.06.103>.
- Watson, S., Xiang, B., Yang, W., Tavner, P., Member, S. & Crabtree, C. (2010). Condition Monitoring of the Power Output of Wind Turbine Generators Using Wavelets. *IEEE Transactions on Energy Conversion*, 25(3), pp. 715–721.
DOI: 10.1109/TEC.2010.2040083.
- Weller, N. (2009). Acceleration Enveloping-Higher Sensitivity, Earlier Detection. Retrieved from: <https://www.pumpsandsystems.com/acceleration-enveloping-higher-sensitivity-earlier-detection>.
- Yang, S. & Shen, L. (2020). Loss-Based Control Charts for Monitoring Non-Normal Process Data. *IEEE Access*, 8, pp. 91163-91169. DOI: 10.1109/ACCESS.2020.2989400.
- Yang, W., Court, R. & Jiang, J. (2013). Wind turbine condition monitoring by the approach of SCADA data analysis. *Renewable Energy*, 53, pp. 365–376.
DOI: <https://doi.org/10.1016/j.renene.2012.11.030>.
- Ye, X. & Zhou L. (2013). *Using SCADA Data Fusion by Swarm Intelligence for Wind Turbine Condition Monitoring*. Paper presented at 2013 Fourth Global Congress on Intelligent Systems, Hong Kong, (pp. 210–215), DOI: 10.1109/GCIS.2013.40.
- Yucui, W. & Yonggang, L. (2016). Diagnosis of Short Circuit Faults Within Turbogenerator Excitation Winding Based on the Expected Electromotive Force Method. *IEEE Transactions on Energy Conversion*, 31(2), pp. 706–713.
DOI: 10.1109/TEC.2016.2521422.
- Zaremba, W., Sutskever, I. & Vinyals, O. (2014). Recurrent Neural Network Regularization. *ArXiv, abs/1409.2329*.

- Zeng, J., Member, S., Lu, D., Member, S. & Zhao, Y. (2013) *Wind Turbine Fault Detection and Isolation Using Support Vector Machine and a Residual-Based Method*. 2013 American Control Conference, Washington, (pp. 3661–3666). DOI: 10.1109/ACC.2013.6580398.
- Zhang, D., Yuan, J., Zhu, J., Ji, Q., Zhang, X. & Liu, H. (2020). Fault Diagnosis Strategy for Wind Turbine Generator Based on the Gaussian Process Metamodel. *Mathematical Problems in Engineering*. 2020. DOI : <https://doi.org/10.1155/2020/4295093>.
- Zhang, Q., Li, F., Long, F. & Ling, Q. (2018). Vehicle Emission Forecasting Based on Wavelet Transform and Long Short-Term Memory Network. *IEEE Access*, 6, pp. 56984-56994. DOI: 10.1109/ACCESS.2018.2874068.
- Zimroz, R., Urbanek, J., Barszcz, T., Bartelmus, W., Martin, N., Zimroz, R., Urbanek, J., Barszcz, T., Bartelmus, W. & Millioz, F. (2012). Measurement of instantaneous shaft speed by advanced vibration signal processing - application to wind turbine gearbox. *Metrology and Measurement Systems*, 18(4), pp. 701-712. DOI: 10.2478/v10178-011-0066-4.