# A WIND POWER FORECASTING APPROACH BASED ON NEURAL NETWORKS AND DATA DECOMPOSITION TECHNIQUES

**JOSEPH NG'ANG'A MATHENGE**

**MASTER OF SCIENCE**
**(Electrical Engineering)**

**JOMO KENYATTA UNIVERSITY**
**OF**
**AGRICULTURE AND TECHNOLOGY**

**2023**

# A Wind Power Forecasting Approach Based on Neural Networks and Data Decomposition Techniques

**Joseph Ng’ang’a Mathenge**

**A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Electrical Engineering of the Jomo Kenyatta University of Agriculture and Technology**

**2023**

# DECLARATION

This thesis is my original work and has not been presented for a degree in any other university.

Signature…..……………………….…..…..……………Date……………………….….………..

**Joseph Ng'ang'a Mathenge**

This thesis has been submitted for examination with our approval as university supervisors

Signature…..……………………….…..…..……………Date……………………….….………..

**Prof. David K. Murage, PhD**

**JKUAT, Kenya**

Signature…..……………………….…..…..……………Date……………………….….………..

**Prof. John N. Nderu, PhD**

**JKUAT, Kenya**

# DEDICATION

To my wife, daughter (Alba), mother, and brother.

# ACKNOWLEDGEMENT

I wish to thank the Almighty God for granting me the grace to complete this work. I would also like to sincerely thank my supervisors, Prof. J.N. Nderu and Prof. D.K Murage, for their patience and kindness as they guided and encouraged me throughout this research.

# TABLE OF CONTENTS

**DECLARATION........................................................................................ii**

**DEDICATION..........................................................................................iii**

**ACKNOWLEDGEMENT.................................................................................iv**

**TABLE OF CONTENTS..................................................................................v**

**LIST OF TABLES ....................................................................................ix**

**LIST OF FIGURES ....................................................................................x**

**LIST OF APPENDICES .............................................................................xiii**

**ABBREVIATIONS AND ACRONYMS.................................................................xiv**

**ABSTRACT.............................................................................................xvi**

**CHAPTER ONE ..........................................................................................1**

**INTRODUCTION.........................................................................................1**

1.1 Background ..............................................................................................1

1.2 Problem Statement ....................................................................................2

1.3 Justification ..............................................................................................3

1.4 Objectives.................................................................................................4

1.4.1 Main Objective .......................................................................................4

1.4.2 Specific Objectives ..................................................................................4

1.5 Project Scope.............................................................................................4

1.6 Thesis Outline........................................................................................5

**CHAPTER TWO..............................................................................................6**

**LITERATURE REVIEW ...................................................................................6**

2.1 Harnessing the Power of Wind....................................................................6

2.2 The Wind Power Equation and Wind Turbines' Power Characteristics ..................9

2.3 Wind Energy Conversion Systems (WECS) .........................................................12

2.4 Wind Power Forecasting .................................................................................15

2.5 Challenges of Wind Energy Integration in a Power System. ................................17

2.5.1 Wind Power Forecasting Approaches..............................................................18

2.5.2 Evaluation Criteria of the Accuracy of Wind Power Forecasts.......................20

2.6 The Artificial Neural Network (ANN) ...................................................................23

2.6.1 Basic Structure of the ANN ................................................................................23

2.6.2 ANN Activation Functions. .................................................................................26

2.6.3 Types of ANNs ......................................................................................................31

2.7 Data Decomposition Techniques.............................................................................40

2.7.1 Wavelet Transform ..............................................................................................43

2.7.2 Empirical Mode Decomposition (EMD) ..............................................................44

2.8 Previous Neural Network-Based Research on Wind Prediction. ...........................45

2.9 Summary and Research  Gaps .................................................................................48

**CHAPTER THREE** ........................................................................................ **49**

**METHODOLOGY** ........................................................................................ **49**

3.1 Outline of the Research Stages ................................................................ 49

3.2 Dataset Description and Handling ........................................................... 51

3.2.1 Data Description. .................................................................................. 51

3.2.2 Data Preparation .................................................................................. 53

3.2.3 Look Back period .................................................................................. 55

3.3 Data Decomposition. .................................................................................. 56

3.4 Modelling of the Wind Power Curve from Historical Data .................................. 58

3.5 Models and Model Parameters ................................................................... 59

**CHAPTER FOUR** ........................................................................................ **62**

**RESULTS, ANALYSIS AND DISCUSSION** ......................................................... **62**

4.1 Forecasting using the Direct and Indirect Approach ........................................ 62

4.1.1 Test on Effectiveness of Data Decomposition and Illustration of the Deficiency of LSTMs ........................................................................................ 62

4.2 Wind Power Forecasting using BiLSTM ........................................................ 67

4.2.1 BiLSTM Alone Prediction Results ............................................................ 67

4.2.2 BiLSTM_EMD ....................................................................................... 71

4.2.3 BiLSTM with Avrami Curve ..................................................................... 74

4.3 .Summary of Results .................................................................................. 80

**CHAPTER FIVE.........................................................................................82**

**CONCLUSION AND RECOMMENDATIONS ..................................................82**

5.1 Conclusion.......................................................................................................82

5.2 Recommendations ...........................................................................................83

**REFERENCES.........................................................................................................84**

**APPENDICES .........................................................................................................90**

# LIST OF TABLES

**Table 2.1:** Comparison Between Onshore and Offshore Wind Turbines. ........................ 8

**Table 2.2:** Types of WECS. ............................................................................ 14

**Table 2.3:** Wind Power Prediction Horizons and their Applications. ............................ 16

**Table 3.1:** Particulars of the Wind Dataset ...................................................... 51

**Table 3.2:** Conventional Approach of Interpreting Correlation Coefficients. ................ 52

**Table 3.3:** Pearson’s Correlation Results for AL_WIND_07_12 Dataset. ..................... 52

**Table 3.4:** AL_WIND_07_12 Correlation ...................................................... 56

**Table 3.5:** Partition of the Dataset into Train and Test Data.......................................... 58

**Table 4.1:** RMSE Values Comparison for the Models. ................................................. 66

**Table 4.2:** Comparison of Prediction Results of BiLSTM Alone .................................. 70

**Table 4.3:** Comparison of Prediction Results of BiLSTM_EMD .................................. 73

**Table 4.4:** Prediction Results of BiLSTM_EMD with Avrami Curve ............................ 79

**Table 4.5:** Comparison of Prediction Results for all BiLSTM Models .......................... 80

## LIST OF FIGURES

**Figure 2.1:** New Global Wind Power Installations in GW per Year. ...............................7

**Figure 2.2:** Wind Power Extraction Illustration According to the Betz Limit..................9

**Figure 2.3:** Wind Power Curve for a Vestas V52 - 850kW Wind Turbine.....................11

**Figure 2.4:** Parts of a Wind Turbine..................................................................13

**Figure 2.5:** Architecture of a Neuron. ...............................................................24

**Figure 2.6:** Flowchart of the ANN algorithm.......................................................25

**Figure 2.7:** Sigmoid Function and its Derivative ..................................................26

**Figure 2.8:** Hyperbolic Tangent Function and its Derivative .....................................27

**Figure 2.9:** ReLU Function and its Derivative......................................................28

**Figure 2.10:** PReLU Function and its Derivative...................................................29

**Figure 2.11:** ELU Function and its Derivative......................................................30

**Figure 2.12:** Swish and Mish Functions Comparison ..............................................31

**Figure 2.13:** Illustration of a Single-Layer FFNN. .................................................32

**Figure 2.14:** A Multilayer Perceptron ...............................................................33

**Figure 2.15:** Illustration of a CNN ...................................................................34

**Figure 2.16:** Illustration of an RNN (rolled) ........................................................34

**Figure 2.17:** A Fully Connected RNN (unrolled) ..................................................35

**Figure 2.18:** Chain of Simple Interconnected Neural Networks to Form an RNN.........36

**Figure 2.19:** Chain Structure of the LSTM....................................................................37

**Figure 2.20:** An LSTM Cell.............................................................................................37

**Figure 2.21:** Structure of a BiLSTM Network.................................................................39

**Figure 2.22:** An Illustration of Data Decomposition. ......................................................42

**Figure 3.1:** Illustration of the Various Stages of the Developed BiLSTM+EMD+Avrami Power Curve Model. ..........................................49

**Figure 3.2:** General Wind Power Forecasting Procedure ...............................................50

**Figure.3.3:** A Section of the AL_WIND_07_12 Wind Power Dataset before Normalization. ..........................................................................................54

**Figure 3.4:** A Section of Normalized AL_WIND_07_12 Wind Power Dataset.............55

**Figure 3.5:** Stages of Pre-Processing. .............................................................................53

**Figure 3.6:** LSTM/BiLSTM_EMD Forecast Procedure .................................................57

**Figure 4.1:** Graph of Wind Power vs Time for LSTM alone (168 look ahead instances)..........................................................................................63

**Figure 4.2:** Graph of Wind Power vs Time for LSTM+EMD (168 look ahead instances)..........................................................................................64

**Figure 4.3:** Graph of Wind Power vs Time for LSTM+EMD with Power Curve Layer (168 look ahead instances). ..............................................................65

**Figure 4.4:** Graph of Wind Power vs Time for Models Comparison. ............................66

**Figure 4.5:** 24 - Hours Ahead Wind Power Forecast......................................................68

**Figure 4.6:** 48 Hours Ahead Wind Power Forecast........................................................68

**Figure 4.7:** 72 Hours Ahead Wind Power Forecast ........................................................69

**Figure 4.8:** 168 Hours Ahead Wind Power Forecast ......................................................70

**Figure 4.9:** 24 Hours Ahead Wind Power Forecast ........................................................71

**Figure 4.10:** 48 Hours Ahead Wind Power Forecast ......................................................72

**Figure 4.11:** 72 Hours Ahead Wind Power Forecast ......................................................72

**Figure 4.12:** 168 Hours Ahead Wind Power Forecast ....................................................73

**Figure 4.13:** Scatter Plot of the Wind Power vs Wind Speed Data Before Curve Fitting (AL_WIND_07_12 Dataset)...........................................................74

**Figure 4.14**: Scatter Plot of the Wind Power vs Wind Speed Data After Curve Fitting with Avrami (AL_WIND_07_12 Dataset) .................................................75

**Figure 4. 15:** 24 - Hours Ahead Wind Power Forecast ...................................................76

**Figure 4.16:** 48 Hours Ahead Wind Power Forecast ......................................................77

**Figure 4.17:** 72 Hours Ahead Wind Power Forecast ......................................................78

**Figure 4.18:** 168 Hours Ahead Wind Power Forecast ....................................................79

# LIST OF APPENDICES

**Appendix I:** Wind Power Dataset Information ............................................................. 90

**Appendix II:** Developed MatLab Codes ......................................................................... 92

**Appendix III:** Conference and Journal Papers Published ........................................... 103

## ABBREVIATIONS AND ACRONYMS

| | |
|---|---|
| **ANFIS** | Adaptive Neuro-Fuzzy Inference System |
| **AWPO** | Actual Wind Power Output |
| **ANN** | Artificial Neural Network |
| **BP** | Back Propagation |
| **BPNN** | Back Propagation Neural Network |
| **DE** | Differential Evolution |
| **DFIG** | Doubly Fed Induction Generator |
| **DMT** | Data Mining Techniques |
| **EMD** | Empirical Mode Decomposition |
| **ERCOT** | Electric Reliability Council of Texas |
| **GA** | Genetic Algorithm |
| **GW** | Gigawatt |
| **HAWT** | Horizontal-Axis Wind Turbines |
| **ISO** | Independent System Operator |
| **KW** | Kilowatt |
| **MAE** | Mean Absolute Error |
| **MSE** | Mean Square Error |
| **POPF** | Probabilistic Optimal Power Flow |

| | |
|---|---|
| **PMSG** | Permanent Magnet Synchronous Generator |
| **PSO** | Particle Swarm Optimization |
| **RES** | Renewable Energy Sources |
| **SCIG** | Squirrel Cage Induction Generator |
| **SI** | Swarm Intelligence |
| **SSR** | Sum of Squared Residuals |
| **VAWT** | Vertical Axis Wind Turbine |
| **VOLL** | Value of lost load |
| **WT** | Wind Turbine |
| **WSF** | Wind Speed Forecast |

# ABSTRACT

In recent times, the world has inclined towards using renewable energy sources since they have relatively lower greenhouse gas emissions, occur freely in nature, and unlike fossil fuels, cannot be depleted. Wind power is one such renewable energy source that has attracted a lot of research and interest in the power industry. With the growing quantities of wind power generation incorporated into power systems, grid reliability is at risk since wind power is highly intermittent. Wind power forecasts help incorporate wind in a grid’s power mix more efficiently and reduce the quantity of power reserves allocated to cater to the intermittency of wind. This makes adopting more wind power resources into the grid more economical. This work developed an approach to wind power forecasting using Bidirectional Long Short-Term Memory (BiLSTM) Neural Networks hybridized with data decomposition techniques and a wind power curve layer. First, a wind power curve was modelled from the historical wind speed and wind power datasets using the Avrami equation and the best line of fit determined. Next, the wind time series data was decomposed into several Intrinsic Mode Functions (IMFs) and a Residual Function using Empirical Mode Decomposition (EMD). Finally, the BiLSTM model enhanced with the Avrami Power curve was used to forecast future wind power values. The developed model was tested on an online-based dataset and compared with the traditional LSTM, BiLSTM and hybrid (Bi)LSTM - data decomposition models. Using the developed BiLSTM + EMD enhanced with an Avrami Power Curve layer, wind power prediction RMSE improved by at least 50% for the 24-hour forecast compared to hybrid BiLSTM-data decomposition models.

# CHAPTER ONE

# INTRODUCTION

## 1.1 Background

Conventionally, power grids have been dominated by a few centrally located high-output generators that produce power to meet the entire load in the grid. Hydro, geothermal, coal, and nuclear power generators have been the primary electrical power sources in the grid. In recent times, however, there has been a gradual paradigm shift from the exclusive use of these classical electrical power sources to adopting more renewables in the power mix to meet the growing demand. This is in line with ensuring that the power industries reduce their environmental carbon footprint (Jarke & Perino, 2017). In addition, many grids have continued leaning towards more reliable sources of energy that do not stand to get depleted soon. Fossil fuel deposits are running low, and with increasing populations, they stand to get exhausted at some point in the future (Kalair Anam, Abas Naeem, Shoaib Saleem Muhammad, Raza Kalair Ali, 2020). Initiatives such as the Paris agreement and the Net Zero initiative have gone a long way in promoting the adoption of more renewables in a grid’s energy mix and reducing global warming (Oxford Net Zero, n.d.).

Solar, wind, and biofuels have attracted the most interest in this quest to go green in the 20th century (Lu et al., 2020). Solar and wind occur freely in nature and can be harnessed anywhere in the world. This has made them an attractive area of research for scholars globally who aim to maximize the benefits the world can reap from these free resources. Governments have incentivized energy institutions, including giving tax waivers and funding to encourage the exploration and exploitation of solar and wind energy. There have been technological advancements in the hardware technology used in solar panels, battery storage systems, and wind turbines. These advancements aim to improve the efficiency of renewable systems to ensure maximum power conversion from solar irradiation and wind into electric power. The popularity of Solar and Wind power varies by region and they are both considered to be important sources of renewable energy. In terms of their harnessing, solar energy is more predictable since it is naturally experienced during the day, allowing for effective planning and utilization. In contrast, the wind power is highly stochastic with wind patterns varying from day to day and even from hour to hour. In this regard, wind power is conventionally considered to be a non-dispatchable resource by power system operators (Hafner, Manfred Luciani, 2022).

Wind power’s intermittent and unpredictable nature has spurred significant interest in developing accurate prediction algorithms and models to ensure the optimal integration of wind as a dispatchable resource in a grid’s energy mix. Wind power uncertainty is a risk to grid stability in cases where wind power accounts for a significant amount of the grid’s total installed capacity. Voltage stability at the point of injection and transient stability become issues of concern regarding wind power integration into the grid. This challenge has called for the development of fast and accurate optimization tools capable of accurately predicting wind power, allowing for the best planning of grid operations and optimal usage of this free power as and when available.

## 1.2 Problem Statement

Penetration of renewable energy sources in power grids has been on the increase in recent years. Renewable Energy Sources (RES) are considered clean and with minimal to no pollution effects on the environment. Some RES, like wind, are intermittent, making their connection to the grid and dispatch problematic for grids with significant wind power. As more wind power is integrated into a power system, the operation of the grid is bound to be adversely affected by the unpredictable nature of this renewable resource. Despite the fast advances in the technologies behind harnessing wind power, the solutions to counter the intermittency of wind have not grown on a similar scale, and this remains a viable area of research. With global warming altering weather patterns, it becomes difficult to rely on traditional models as the only way to determine wind patterns and project the expected power. A more reliable forecast method that combines historical and real-time data becomes necessary to ensure optimal wind power dispatch in the grid.

The operating schedule of a wind farm connected to a grid is greatly affected by the accuracy of the method used to forecast the expected power (Y. Zheng et al., 2015). The wind farm can be underutilized, leading to the projected income from wind power not being achieved. On the other hand, expected wind power can be over-projected, leading to a deficit during dispatch in a power system. In such a case, the operating reserves from other generating sources must be over-engaged, diluting the economic advantages of wind power in the grid. Before a utility-scale wind farm is developed, feasibility studies are done to determine the wind speeds, direction, consistency, and intermittency levels (E & Diamond, 2011). However, despite the feasibility studies, this does not change the fact that wind is an intermittent natural resource, so wind power constantly fluctuates. Considering that wind farms have a viability period of between 20 to 30 years, it would be of great importance that when the wind turbines are in operation, utilization of the power output derived from them is optimal. Hence, developing accurate wind power prediction tools becomes necessary in present-day power grids as installed wind power capacity continues to increase.

## 1.3 Justification

Wind energy is a renewable energy resource that occurs intermittently in nature. Despite being a free resource, its intermittency makes it challenging to integrate into the grid. There arise challenges of power quality, stability, system operation, and control (Z. Chen & Member, 2005). Large-scale wind farms in the range of hundreds of megawatts have continued to be installed globally, with wind energy penetration increasing to over 30% of the total demand in some power systems like Denmark (Zhang, Shijie Wei, Jing Chen, Xi Zhao, 2020). Between 2010 and 2021, it was reported that the capacity of global wind energy installed grew by more than four times (Lee & Zhao, 2021). With such penetration levels, the security of a power system becomes a key concern and should be evaluated keenly to ensure that this free resource does not end up causing instability in previously stable systems.

The goal of sustainable development can only be attained by properly managing the available resources. Wind power is a free and available resource, and it should be well utilized as and when available. High-precision wind speed and wind power forecasting tools have become the most crucial aspect of any power system that deploys wind power in the range of Megawatts. Short-term forecasts extending from 1 hour to 72 hours are essential to plan the operations of a power system in terms of unit commitment and wind power dispatch. This helps ensure operations continuity and boosts the power system's reliability. In addition, accurate wind power prediction helps optimize the grid's operational costs by optimizing the grid’s power mix at any given time and optimizing any balancing power required to take care of the uncertainty of wind power. The developed hybrid algorithm in this thesis helps improve the accuracy of wind power forecasts, thus making integrating large-scale wind power more manageable and economical.

## 1.4 Objectives

### 1.4.1 Main Objective

To develop a wind power forecasting tool using Bidirectional Long Short-Term Memory (BiLSTM) Neural Networks hybridized with data decomposition techniques and a wind power curve layer for improved forecasts.

### 1.4.2 Specific Objectives

a) To develop a wind power forecasting model and implement it in a wind power forecasting tool using BiLSTM.
b) To enhance the developed BILSTM model with data decomposition techniques and a wind power curve layer.
c) To test and validate the developed model using the actual wind power data from a verifiable wind power dataset from the National Renewable Energy Laboratory.

## 1.5 Project Scope

The scope of this study entailed the development of an algorithm that gave improved wind power forecasts. The developed algorithm was then tested on a verifiable dataset. The simulation environment for this research was MATLAB.

### 1.6 Thesis Outline

Chapter 1 outlines the problem statement, justification, objectives and the scope of research.

Chapter 2 gives a literature review of the essential topics of this research. It starts by explaining the mechanics of wind power harnessing and the types of wind power conversion technologies. The various approaches to wind power forecasting based on neural network are then introduced after which the essence of data decomposition in time series prediction is presented.

Chapter 3 is the methodology section that gives the outline of the steps adopted in implementing the wind power forecasting tool using neural networks. The process of enhancing the base models with data decomposition and with a wind power curve layer is then presented.

Chapter 4 presents and discusses the results obtained in this thesis. The results are presented based on the methodology and the objectives of the study.

Chapter 5 outlines the conclusions of the thesis and gives recommendations for future research.

# CHAPTER TWO

# LITERATURE REVIEW

## 2.1 Harnessing the Power of Wind

Wind is the movement of air due to the sun’s uneven heating of the earth’s surface. In essence, there can never be uniform heating of the earth’s surface due to the varying altitudes of topological features; hence, the wind always exists in nature. Wind is random and uncontrollable, and it is a renewable energy resource that occurs freely in nature virtually in all places around the globe. It is relatively simple to harness and has a significantly lower carbon footprint on the environment. There has been tremendous growth in incorporating renewables such as wind energy in present-day power systems driven by public policies that champion climate change and reduction of the carbon footprint of various sectors on the environment.

In some countries, such as the United States, which have well-developed power systems, fossil fuels, mainly coal and natural gas, are among the largest sources of electricity (U.S Energy Information Administration, 2021). Deposits of these fossil fuels are limited in nature and hence stand to get depleted with the growing power demand. In addition, power stations running on fossil fuels are major emitters of $CO_2$ as a byproduct (EPA, 2019). $CO_2$ is a greenhouse gas that has been the most significant contributor to global warming in the last century. Recently, the world has sought to reduce its $CO_2$ emissions in the environment to reduce global warming. According to the Paris agreement, the world is looking to reduce global warming to below $2^oC$ by 2050 (Oxford Net Zero, n.d.). Renewables offer this much-needed path towards achieving net-zero emissions in the atmosphere since they are clean energy sources with minimal adverse environmental impacts.

According to the global wind report, the global wind industry experienced its most remarkable growth in 2020, where the total new wind power installations surpassed the 90GW mark that year. This was a 56.7% increase in new installations compared to the year 2019 (Lee & Zhao, 2021). The cumulative wind power installations as of 2020 now stood at 743GW, with China reporting cumulative grid-connected wind power of over 272GW, making it the global leader in wind power installations (Zhang, Shijie Wei, Jing Chen, Xi Zhao, 2020). Of the 95.3GW of new wind power installations in 2020, 88.4GW (92.8%) of this power came from onshore installations, which shows the popularity of onshore wind power plants worldwide. ***Figure 2.1*** shows the New global year-over-year wind power installations over the last seven years (Lee & Zhao, 2021)(Global Wind Energy Council, 2022):

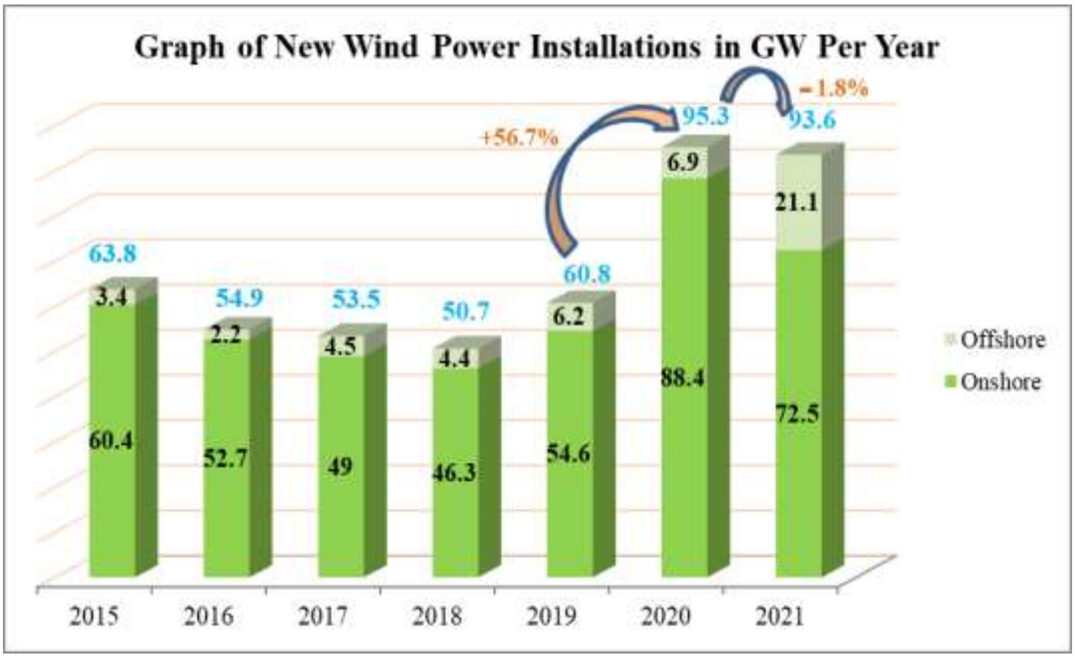


**Figure 2.1: New Global Wind Power Installations in GW per Year.**

One of the biggest challenges in using wind turbines is the fast change in the power output of these turbines (Ahmed & Al-ismail, 2020). The wind turbine power output depends on the wind speed – an irregular and random variable, making wind power an intermittent variable. Due to this variability of wind power, for the efficient running of wind power units connected to the grid, accurate forecasts of up to 48 hours are crucial (Lotfi et al., 2014). The knowledge and information regarding the future expected power generation from wind turbines is vital for the smooth operation of a power system.

Wind power uncertainty and the increased bulk wind power integration in power systems are bound to affect system stability. The uncertainty of wind power makes it a non-dispatchable resource [7]. As a result, there has been a lot of research focusing on the development of accurate prediction tools for wind power and improved methodologies of accurately determining and allocating reserves to cater to the intermittency of wind power, especially where large-scale wind power integration is involved. Advancing technologies in wind turbine manufacturing have greatly enhanced the process of harnessing wind energy hence popularizing wind power in present-day power grids. Wind power forecasting continues to be an attractive research area, especially with the increased integration of renewables in modern-day grids. Forecasts help power system operators make decisions regarding the grid's operation to guarantee reliability and ensure that generation follows the load demand.

Wind power is harnessed on two fronts:

i. Onshore
ii. Offshore

Onshore wind turbines are set up on land, while on the other hand, offshore wind turbines are installed in water bodies, e.g., seas and oceans.**Table 2.1** shows a comparison between onshore and offshore wind turbines:

**Table 2.1: Comparison Between Onshore and Offshore Wind Turbines (Blaabjerg & Ma Ke, 2017).**

| | **Onshore** | **Offshore** |
|---|---|---|
| 1 | Designed for moderate wind speeds | Designed for higher wind speeds |
| 2 | Comparatively lower installation costs | Higher installation costs compared to onshore turbines |
| 3 | Minimal erosion to turbine blades | More erosion of turbine blades due to moisture |
| 4 | Easy to access and maintain | Difficult to maintain and access |

In 2015, the average size of onshore wind turbines installed in the developed wind markets such as in Europe was 2.7 MW with the average for offshore turbines standing at 4.2 MW (Blaabjerg & Ma Ke, 2017). By 2017, the major wind turbine manufacturing companies are producing products ranging from 4 - 6 MW, indicating a growth in the wind power sector (Blaabjerg & Ma Ke, 2017). Offshore winds are more uniform and stronger than onshore winds and, thus, generate more power explaining why offshore turbines have higher output ratings compared to onshore turbines. However, onshore wind turbines still remain the most common wind turbine technology installed globally (Global Wind Energy Council, 2022).

## 2.2 The Wind Power Equation and Wind Turbines’ Power Characteristics

Harnessing wind power entails taping the kinetic energy in the wind and converting it into electric power. Extracting all the kinetic energy in a wind stream is practically impossible. Mathematically, if all the kinetic energy from a wind stream were extracted, the stream’s speed after the turbine would be zero (Aliprantis & Lafayette, 2014). This would imply a disruption in the flow of new wind streams across the wind turbine. German physicist Albert Betz researched on the maximum possible power that can be extracted from a mass of wind and called it the Betz limit. According to the Betz limit, a wind turbine cannot convert more than **59.3%** of the kinetic energy in the wind into mechanical energy to turn the generator rotor (Ragheb & Ragheb Magdi, 2011). This is illustrated in ***Figure 2.2***.

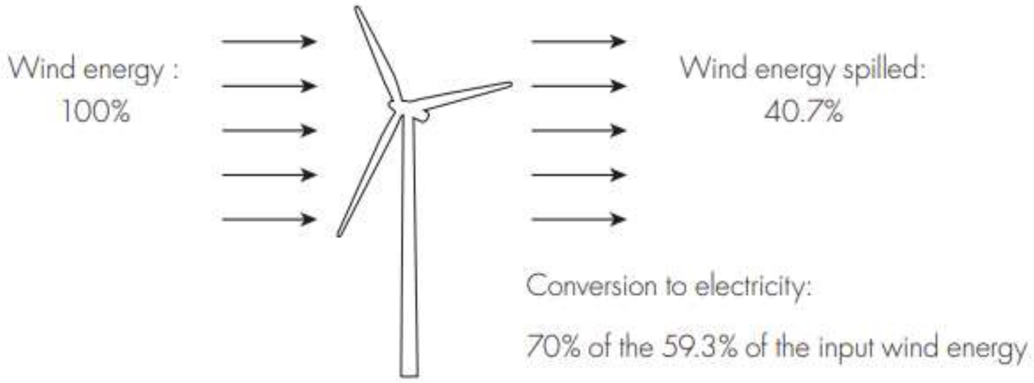

Wind energy :
100%
Wind energy spilled:
40.7%
Conversion to electricity:
70% of the 59.3% of the input wind energy

**Figure 2.2: Wind Power Extraction Illustration According to the Betz Limit (Ragheb & Ragheb Magdi, 2011).**

Out of the 59.3% of the wind’s mechanical energy extracted by the wind turbine, it converts part of it into electrical power based on its efficiency. For example, with a 70% efficiency, the power coefficient ($C_p$) of the wind turbine becomes $70\ x\ 59.3\% = \mathbf{41.51}\%$**.**

The wind power equation is derived from the equation for kinetic energy (K.E) given by

$$.K.E = \frac{1}{2}mv^2 \qquad (2.1)$$

Where: $m$ is the mass of an object in kg

$v$ is the velocity of the object in m/s

The mass of flowing air particles used to generate the kinetic energy is a function of time since air is continuously flowing and is given as:

$$\frac{dm}{dt} = \rho(t)Av \qquad (2.2)$$

Where: $m$ is the mass of air in kg

$t$ is the time in $s$

$v$ is the velocity of air

$\rho(t)$ is the density of air in kg$/m^3$ at a given time $(t)$.

$A$ is the area swept by the wind turbine blades in $m^2$.

From equation (2.1), the average electrical power $(P_{av})$ in Watts developed by wind is given as:

$$P_{av} = C_p \frac{1}{2}\frac{dm}{dt}v^2 \qquad (2.3)$$

Substituting equation (2.2) in (2.3), we have:

$$P_{av} = C_p * \frac{1}{2}\rho Av.v^2 = \frac{1}{2}C_p\rho Av^3 \qquad (2.4)$$

Where $C_P$ is the Power Coefficient

Equation 2.4 above shows the **cubic relationship** between wind speed and wind power, making the wind speed, $v$, the most significant variable in determining the power output of a wind turbine (Rafaat & Hussein, 2018). A graph of $P_{av}$ against $v$ is known as the wind power generation curve. There are three critical zones in a wind power generation curve:

i. Cut in speed
ii. Constant Cp region
iii. Constant Power output region.

These zones are illustrated in **Figure 2.3** next:

**Power Curve V52 - 850 kW**

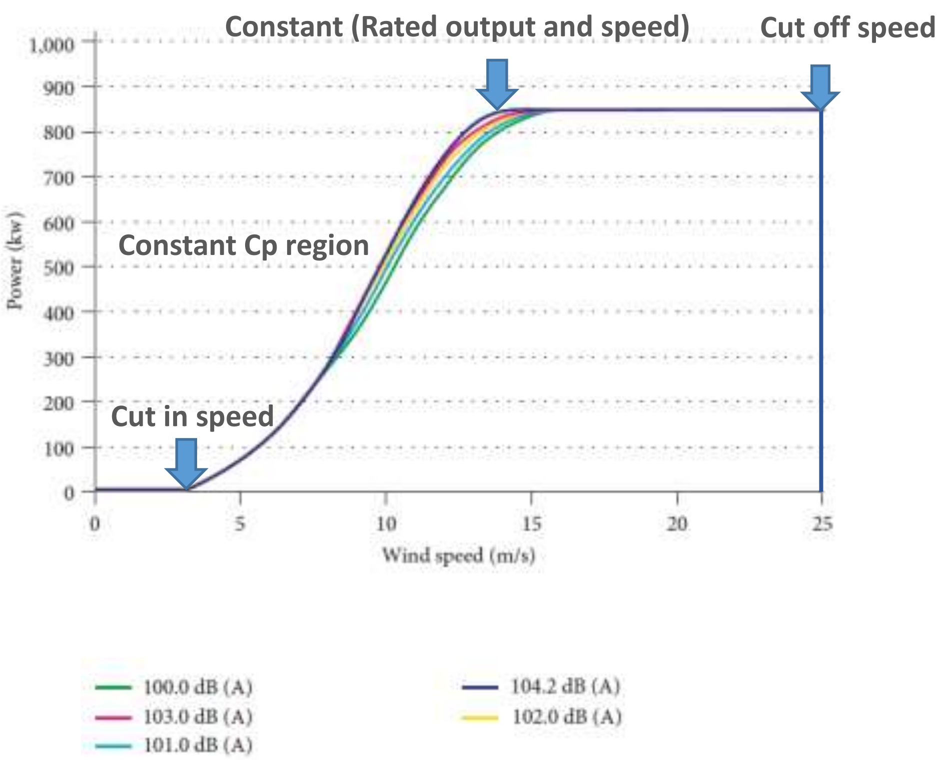


**Figure 2.3: Wind Power Curve for a Vestas V52 - 850kW Wind Turbine (Vestas Wind Systems, n.d.).**

The cut-in speed is the speed at which the wind turbine has sufficient mechanical energy to start generating electric power. Below the cut-in speed, the generator output is zero since the energy in the wind is too small to overcome mechanical losses and start generating. In the Constant Cp region, the wind power has a quasi-cubic increase for every step increase in the wind speed. When the wind speed reaches the rated speed, the wind generator's power output becomes constant, and this zone is referred to as the constant power region. The output remains constant until the cut-off speed, where the turbine disengages from the generator and generation stops. Beyond the cut-off speed, the wind speed is too fast and stands to damage the generator; hence, generation is stopped.

## 2.3 Wind Energy Conversion Systems (WECS)

Wind Turbines have increased in size over the years, having started from a few KW to the MW systems of today. New combinations and configurations of generators and convertor technologies help achieve high and stable output power. One of the most notable modifications has been the introduction of pitch-able blades. These allow the wind generator to vary its output by controlling the angle of attack of the turbine blades on the wind (Beainy et al., 2016). The future of harnessing wind power is going offshore since winds are stronger and more consistent at sea than on land. However, for now, most of the installations that have been done are onshore.

Wind turbines are usually clustered together to form wind farms, after which the power obtained from the wind farms is aggregated and injected into the grid. According to the history of wind turbines, there have been two types of WTs – fixed speed and variable speed. Fixed-speed WTs are an earlier technology that has gradually been replaced by variable-speed WTs. Variable speed WTs can achieve better aerodynamic efficiency for varying wind speeds. Usually, large-scale WTs have a horizontal axis structure and are referred to as Horizontal Axis Wind Turbines (HAWT). Other configurations are the Vertical Axis Wind Turbines (VAWT). HAWTs are more common globally due to their high power conversion efficiencies (Johari et al., 2018). The blades of a WT can reach up to 140 m in diameter for rotational speeds of between 5 and 25 rpm. ***Figure 2.4*** below shows the parts of a wind turbine:

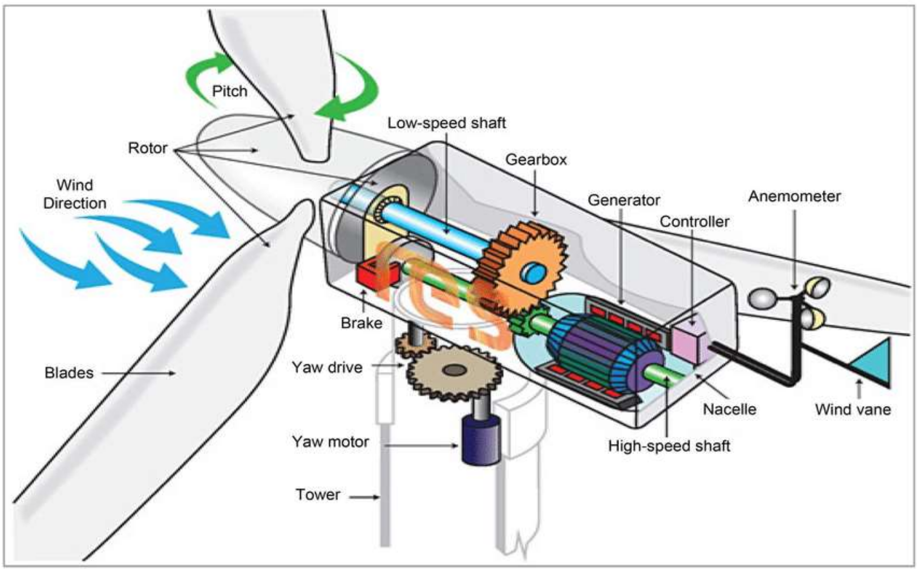


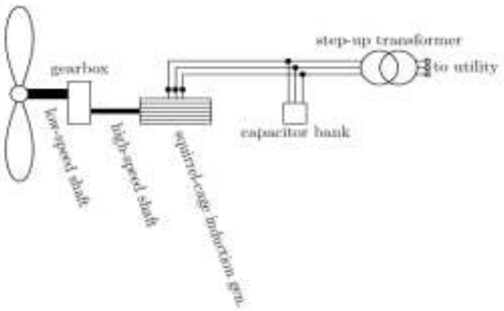
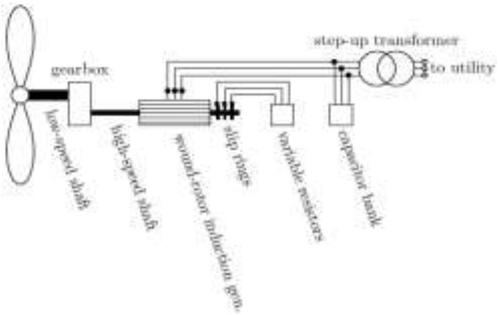
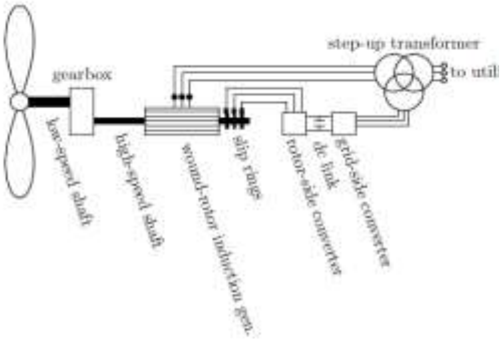
**Figure 2.4: Parts of a Wind Turbine (Rafaat & Hussein, 2018)**

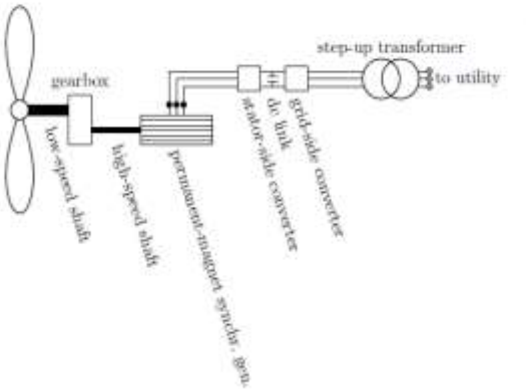
There are four types of WECS used in geared operations (Aliprantis & Lafayette, 2014):

i. Squirrel Cage Induction Generator (SCIG) or Type I wind turbines.
ii. Wound Rotor Induction Generator (WRIG) or Type II wind turbines.
iii. Doubly Fed Induction Generators (DFIG) or Type III wind turbines.
iv. Permanent Magnet Synchronous Generators (PMSG) or Type IV wind turbines.

**Table 2.2** summarises these four types of WECS and their characteristics (Aliprantis & Lafayette, 2014) (Beainy et al., 2016).

**Table 2.2: Types of WECS.**

| S/No | Type | Characteristics | Illustration |
|---|---|---|---|
| i. | Squirrel Cage Induction Generator (SCIG) or Type I wind turbine | • Simple design<br>• Squirrel cage rotor<br>• Fixed speed operation<br>• Limited controllability, hence inefficient at varying wind speeds. |   step-up transformer<br>to utility<br>gearbox<br>capacitor bank<br>low-speed shaft<br>high-speed shaft<br>squirrel-cage induction gen. |
| ii. | Wound Rotor Induction Generator (WRIG) or Type II wind turbine | • Wound rotor with external resistor bank with variable resistors.<br>• Some control over output by varying the rotor resistance. |   step-up transformer<br>to utility<br>gearbox<br>low-speed shaft<br>high-speed shaft<br>wound-rotor induction gen.<br>slip rings<br>variable resistors<br>capacitor bank |
| iii. | Doubly Fed Induction Generators (DFIG) or Type III wind turbines | • Similar topology to the WRIG, but has its rotor is connected to the grid via power electronic devices.<br>• Higher level of control, hence higher efficiency. |   step-up transformer<br>to util<br>gearbox<br>low-speed shaft<br>high-speed shaft<br>wound-rotor induction gen.<br>slip rings<br>rotor-side converter<br>dc link<br>grid-side converter |

| S/No | Type | Characteristics | Illustration |
|---|---|---|---|
| iv. | Permanent Magnet Synchronous Generators (PMSG) or Type IV Wind Turbines | • Use permanent magnets to produce the magnetic field.<br>• They have higher controllability and, hence, are more efficient. |   step-up transformer<br>gearbox<br>to utility<br>low-speed shaft<br>high-speed shaft<br>permanent-magnet synchr. gen.<br>stator-side converter<br>dc link<br>grid-side converter |

## 2.4 Wind Power Forecasting

Wind power is a stochastic variable and requires accurate forecasting techniques to determine a future value close to its actual value over a given horizon (Shao et al., 2016). Wind power forecasting is a time series analysis problem since wind data fits the description of a time series. A time series is a set of ordered data that spreads over a given period and has equally spaced observation points (Ivanovic & Kurbalija, 2016). Time series analysis entails using previously observed values to predict future values. It is the process of analyzing time series data to extract information and forecast future instances of the series. In time series analysis, a one model fits all approach cannot be assumed since each data set has unique intrinsic characteristics that may not be common to other data sets (Shrestha & Bhatta, 2017). Therefore, a chosen model must be customized and tailored to fit the specific problem to which it shall be applied.

When analyzing time series data, three crucial pieces of information can be drawn from the data (Ivanovic & Kurbalija, 2016):

i. Autocorrelation - refers to the degree of similarity between a given time series and a lagged version of itself.

ii. Seasonality - involves the time series data having patterns that repeat at regular intervals in the timeframe of interest. For example, when the data peaks and dips are observed at similar instances, it is concluded that the time series exhibits seasonality tendencies.

iii. Stationarity - Measures how little the time series mean and variance change with time. If these data values do not change much throughout the time series (i.e., they are seen to be constant over time), then the data is said to have high stationarity.

Time series analysis explains what happened in the past and why and then endeavours to predict what will happen in the future. Such decisions are crucial in power systems since they assist operators in making the best operational decisions to optimize system efficiency and reliability.

Wind power can be predicted in several forecast horizons, as summarized in ***Table 2.3*** (Zhao et al., 2011):

**Table 2.3: Wind Power Prediction Horizons and their Applications.**

| Time Horizon | Forecast Range | Applications in Power Systems |
|---|---|---|
| Very Short Term | Few seconds - 30 mins | Turbine control<br>Market clearing |
| Short Term | 30 mins - 48(or 72) hours | Dispatch planning |
| Medium Term | 48(or 72) hours - 1 week | Unit commitment<br>Maintenance planning |
| Long Term | 1 week - 1 year (or more) | Feasibility studies for wind farm<br>Scheduling for optimal costs |

In as much as these time scales are important, the short-term forecast is the most important for the operation of the wind farms and the grid, and this research focuses on testing the developed forecast algorithms on the short-term prediction horizon. Should wind power be suddenly lost a few minutes to dispatch, thermal generators have to be engaged to kick in and provide the deficit power. Operating thermal units is expensive, and thus, this leads to the increase in per unit cost of electricity. With accurate short-term forecasts, a power system operator can plan for cheaper alternatives in the absence or dip in wind power (Zhu & Genton, 2012). The advantages of accurate short-term predictions include the following:

i. Efficient operation of the wind turbine units/farms ensuring optimal power is obtained from them.

ii. Savings resulting from reduced thermal capacity that is committed as spinning reserve.

## 2.5 Challenges of Wind Energy Integration in a Power System.

There are several challenges posed by integrating wind power into a power system. These challenges are the genesis of the research revolving around wind power forecasting and prediction. The challenges include (Zhu & Genton, 2012):

i. **High variability** – Wind can be blowing at one moment and, in the next, suddenly disappear. This characteristic is unlike conventional sources of power generation, where the power system operators have some control over the power being generated at any particular time.

ii. **Limited predictability** – Weather forecasting is becoming more complicated with the day due to global warming disrupting the normal weather patterns that have existed through the years. Forecasting wind power has now been reduced to a game of chance rather than a science.

iii. **Limited dispatchability** – Due to the intermittency of wind power, power operators consider it a non-dispatchable resource. Such a fluctuating power resource is bound to cause substantial stability issues in a power system when there is a sudden influx of wind power or when a considerable amount of power is suddenly lost when the wind disappears.

iv. **Limited storage capability** – There are very few ways wind power can be stored. They are either too limited such that they do not make use of the total wind power produced or are too expensive to implement. Unlike coal and natural gas, wind cannot be stored physically and used to generate power later.

The main objective of power system operation is to balance the demand and supply of electric power with a generation mix that gives minimum costs. The constraints are the line capacities, among other contingencies. On different time scales, the Independent System Operator (ISO) is tasked with balancing supply and demand and keeping the generation costs at a minimum. These time scales can be a day ahead, hourly, or even minutes ahead. This indicates that if a wind power prediction system is to be developed, it must be able to forecast within minutes to allow the system operators to incorporate the wind power into the grid seamlessly. Power plants provide power based on the projected load demands at any given time. On top of that, they also offer other ancillary services that include frequency regulation and provision of reserve capacity (spinning reserve) to guarantee system security and reliability. Online plants must adjust their output by increasing or decreasing their production to keep the frequency at a nominal level.

### 2.5.1 Wind Power Forecasting Approaches

Wind power forecasting approaches can be categorized into two based on the forecasting data used (Bokde et al., 2018):

a) **The indirect method** - Entails forecasting future **wind speed** values, after which appropriate transformations are applied to get the corresponding values of **wind power**.

b) **The direct method** - Here, wind power is forecast **directly** without predicting wind speed.

According to the approach used in wind power prediction, the various methods can be classified into:

i. **Persistence Methods** - Assumes that wind power at a given observation point in the future, $t + 1$, is the same as the measured power at the current time, $t$. The accuracy of this method deteriorates fast with an increase in the prediction scope/timescale (Hanifi et al., 2020). If the current wind power at time $t$ is $P(t)$, then wind power at a future time $t + \Delta t$ is assumed to be $P(t + \Delta t) = P(t)$. For this assumption to hold, the forecasting time frame needs to be very small; hence, its accuracy is only viable for ultra-short-term time frames.

ii. **Physical Methods** - Utilize complex mathematical models to perform their prediction. Physical methods deal with numerical weather prediction data from meteorological services that study the behaviour and physics of the lower boundary of the atmosphere to predict future weather patterns. This method considers the topology of a wind farm's location. The main disadvantage of physical models is that they need very accurate online and offline data (Zhao et al., 2011). This method is also computationally tasking and requires considerable computing resources (Hanifi et al., 2020). Physical methods present better performance for medium-term and long-term forecasts (Y. Wang et al., 2021). They, however, have poor performance in shorter-term forecasts.

iii. **Statistical Methods** - These are based on establishing the linear and non-linear relationships between weather parameters such as wind speed, direction, and temperature with the generated wind power. Statistical methods require historical data to train the models to determine these relationships. Statistical models are ideal for short-term wind prediction, and prediction accuracy decreases as the forecast time scale increases (Hanifi et al., 2020). They are simple to model and require short computational durations. The models are occasionally tuned by comparing the predicted and measured power to ensure that the forecasts continuously improve as time progresses. Statistical models are divided into time series models and artificial neural network models.

**a) Time series models**

These are mathematical models proposed by George Box and Gwilym Jenkins and use historical data to develop a mathematical model that forecasts future time series instances (Hanifi et al., 2020). The Box - Jenkins models are based on the ARIMA models to find the best fit for a time series based on historical values (Lawan et al., 2014). The general form of the model is described as follows:

$$X_t = \sum_{i=1}^{p} \varphi_i X_{t-i} + \alpha_t - \sum_{j=1}^{q} \theta_j \alpha_{t-j} \tag{2.5}$$

Where:

$X_t$ - is the wind power forecast at the time $t$

$\varphi_i$ - is the autoregressive parameter

$\theta_j$ - moving average parameter

$\alpha_t$ - white noise

$p$ - order of the autoregressive model

$q$ - order of the moving average model

If $p = 0$, the model becomes a Moving Average (MA) model, and if $q = 0$, the model becomes an Auto-Regressive (AR) model.

**b) Artificial Neural Network Models**

Artificial Neural Networks (ANNs) are some of the most popular methods used in wind power forecasting (Hanifi et al., 2020). The strengths of ANNs come from their ability to establish non-linear relationships between input features and the predicted variable(s) without any need for mathematical formulations (Lawan et al., 2014).

**iv. Hybrid Methods**

It involves combining physical and statistical models or two or more physical or statistical models. The aim is to have one model improve on the weakness of the other model they are hybridized with. In doing this, the overall prediction accuracy of the hybrid is improved.

### 2.5.2 Evaluation Criteria of the Accuracy of Wind Power Forecasts.

Since wind power is a highly fluctuating variable, wind power forecasting cannot be an exact science. The forecast wind power at a time $t$ needs to be compared to the actual power from the dataset and the error computed. The less the error, the more accurate the prediction model.  The most common methods used in evaluating the accuracy of a wind forecast are summarized below (Zhao et al., 2011):

a) Mean Error (ME)

$$ME = \frac{1}{N}\sum_{t=1}^{N}(X_{pred,t} - X_{obs,t}) \tag{2.6}$$

Where:

$X_{pred,t}$ - Is the forecast value for time t

$X_{obs,t}$ - Is the observed (actual) value at time t

$N-$ is the number of prediction points

b) Mean Absolute Error (MAE)

$$MAE = \frac{1}{N}\sum_{t=1}^{N}\left|(X_{pred,t} - X_{obs,t})\right| \tag{2.7}$$

Where:

$X_{pred,t}$ - Is the forecast value for time t

$X_{obs,t}$ - Is the observed (actual) value at time t

$N-$ is the number of prediction points

c) Mean Square Error (MSE)

$$MSE = \frac{1}{N}\sum_{t=1}^{N}(X_{pred,t} - X_{obs,t})^2 \tag{2.8}$$

Where:

$X_{pred,t}$ - Is the forecast value for time t

$X_{obs,t}$ - Is the observed (actual) value at time t

$N-$ is the number of prediction points

d) Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(X_{pred,t} - X_{obs,t})^2} \tag{2.9}$$

Where:

$X_{pred,t}$ - Is the forecast value for time t

$X_{obs,t}$ - Is the observed (actual) value at time t

$N-$ is the number of prediction points

e) Normalized Mean Absolute Error (NMAE)

$$NMAE = \frac{1}{P_{inst}}\left\{\frac{1}{N}\sum_{t=1}^{N}\left|(X_{pred,t} - X_{obs,t})\right|\right\} \tag{2.10}$$

Where:

$P_{inst}$ - Installed capacity of the wind farm.

f) Normalized Root Mean Square Error (NRMSE)

$$NRMSE = \frac{1}{P_{inst}}\sqrt{\frac{1}{N}\sum_{t=1}^{N}(X_{pred,t} - X_{obs,t})^2} \tag{2.11}$$

Where:

$P_{inst}$ - Installed capacity of the wind farm.

g) Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{100}{N}\sum_{t=1}^{N}\left|\frac{(X_{pred,t} - X_{obs,t})}{(X_{obs,t}}\right| \tag{2.12}$$

Where:

$X_{pred,t}$ - Is the forecast value for time t

$X_{obs,t}$ - Is the observed (actual) value at time t

$N-$ is the number of prediction points

## 2.6 The Artificial Neural Network (ANN)

### 2.6.1 Basic Structure of the ANN

The ANN is a non-linear mapping architecture mimicking a human's central nervous system operation. The basic structure of an ANN is the neuron. It mimics the biological neuron, and its illustration is shown in ***Figure 2.5:*** *Architecture of a Neuron.* The first artificial neural network model was created by McCulloch and Pitts back in 1943 (Singh, 2016). The ideas behind that first model are still in use today. An ANN can also be defined as a massively paralleled processor consisting of simple processing units that can learn by experience and use that knowledge to make future decisions (Sharkawy, 2020). The ANN is a robust prediction tool for situations where the relationship between data is unknown and seeks to be established. ANNs learn from any correlated patterns observed between input data sets and target values in the training dataset. Once trained, an ANN can then predict subsequent future outcome(s) from the determined pattern/connections it extablishes from the training dataset.

ANNs are well suited to deal with data considered vague, noisy and data that, at times, changes erratically. Wind power exhibits such characteristics due to the fluctuating nature of wind speed. Wind power has a quasi-cubic relationship with wind speed.  Thus, ANNs are ideal for a wind speed/ power data prediction tool since such data is complex and often non-linear (Singh, 2016). Since a neural network consists of highly interconnected nodes, it can learn and generalize training patterns from the training data, just like the human brain. This vital learning capability of ANNs is the main advantage that makes it the best tool for forecasting applications.

The neuron is the basic information processing unit in a neural network, and its architecture is shown in ***Figure 2.5*** below (Sharkawy, 2020):

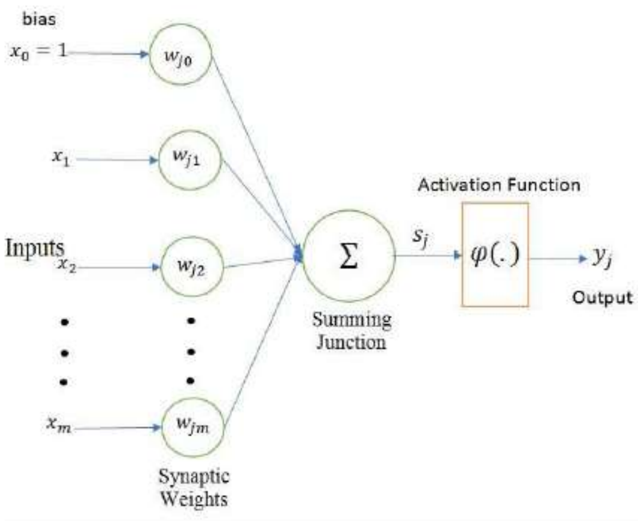


**Figure 2.5: Architecture of a Neuron.**

Where:

- $x_1$ to $x_m$ represent the inputs to the neuron
- $x_0$ represents the bias - The bias helps in establishing meaningful activation of neurons. It stipulates the threshold above which a neuron is activated; otherwise, the neuron remains inactive.
- $w_{j0}$ to $w_{jm}$ represent respective synaptic weights connecting each input to neuron $j$
- $y_j$ is the output signal of the neuron

The output of the summation junction $s_j$ is given as:

$$s_j = \sum_{m=0}^{m} w_{jm} x_m \qquad (2.13)$$

The weighted sum is then passed through an activation function, $\varphi$, to squash it between a given small range of values. The output from the activation function is the output of the neuron, $y_j$.

$$y_j = \varphi(s_j) \qquad (2.14)$$

The flow chart of the ANN algorithm is shown in ***Figure 2.6***:

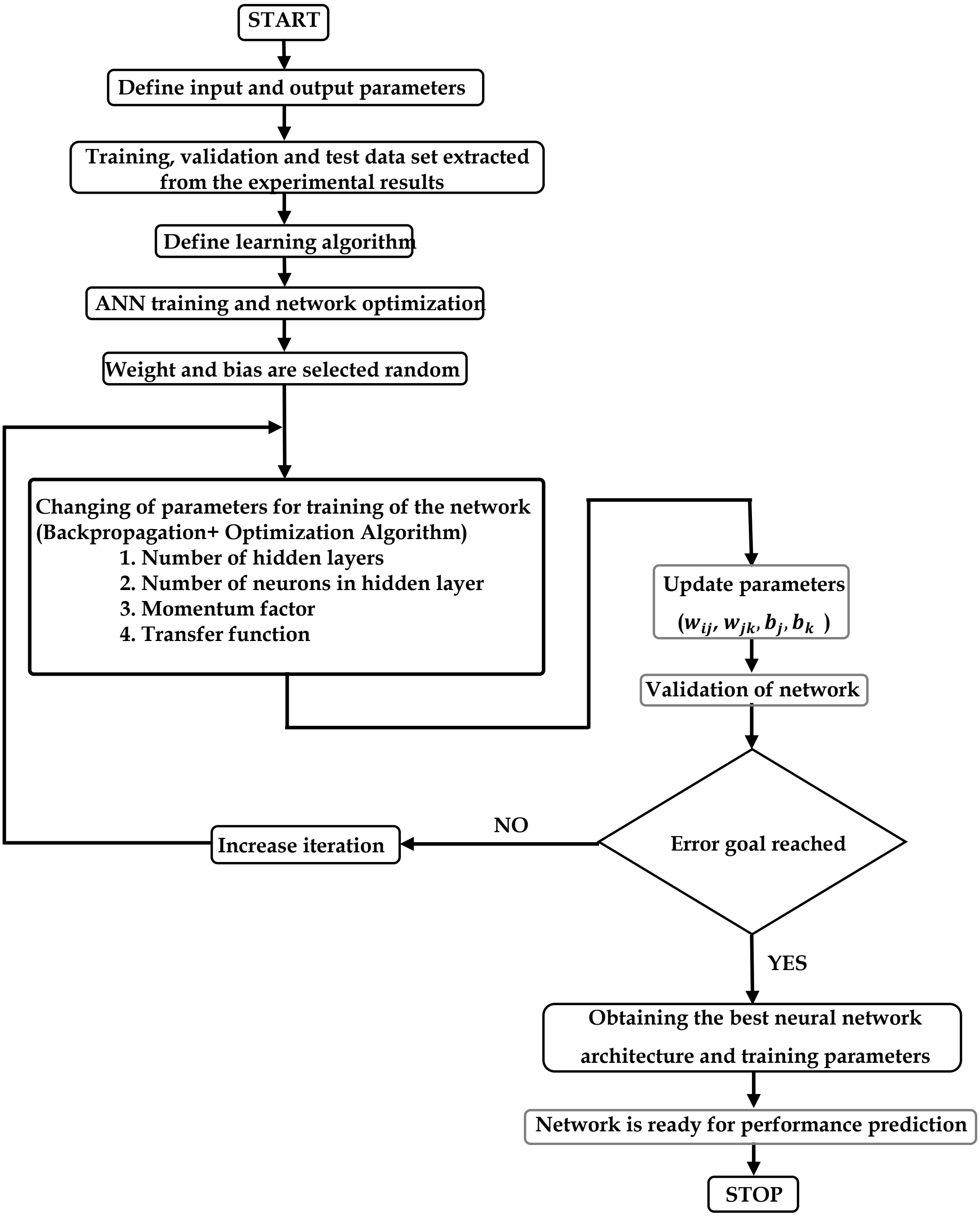

START
Define input and output parameters
Training, validation and test data set extracted from the experimental results
Define learning algorithm
ANN training and network optimization
Weight and bias are selected random
Changing of parameters for training of the network (Backpropagation+ Optimization Algorithm)
1. Number of hidden layers
2. Number of neurons in hidden layer
3. Momentum factor
4. Transfer function
Update parameters
$(w_{ij}, w_{jk}, b_j, b_k)$
Validation of network
Error goal reached
NO
Increase iteration
YES
Obtaining the best neural network architecture and training parameters
Network is ready for performance prediction
STOP

**Figure 2.6: Flowchart of the ANN algorithm.**

### 2.6.2 ANN Activation Functions.

Eight activation functions can be used with neural networks (Rasamoelina, 2020):

i. The sigmoid/logistic activation function

It handles inputs ranging from $-\infty$ to $+\infty$ and translates them to the range $[0; 1]$.

The sigmoid function is defined as:

$$\sigma_x = \frac{1}{1+e^{-x}} \qquad (2.15)$$

A plot of the sigmoid function and its derivative is shown in ***Figure 2.7*** below:

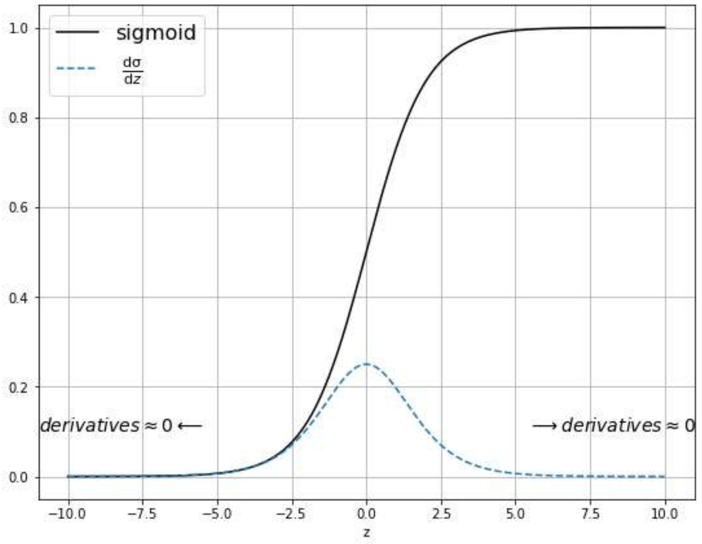


**Figure 2.7: Sigmoid Function and its Derivative**

The sigmoid function squashes a large input space into a small range between 0 and 1. As the input to the sigmoid function increases, the derivative tends to zero. Consequently, the sigmoid function suffers from the vanishing gradient problem in deep networks since the derivative tends to zero, making training ineffective.

ii. Hyperbolic tangent function

It handles inputs ranging from $-\infty$ to $+\infty$ and translates them to the range $[-1, +1]$.

The hyperbolic tangent function is defined as:

$$\tanh(x) = \frac{2}{1+e^{-2x}} - 1 \tag{2.16}$$

A plot of the hyperbolic tangent function and its derivative is shown in ***Figure 2.8*** :

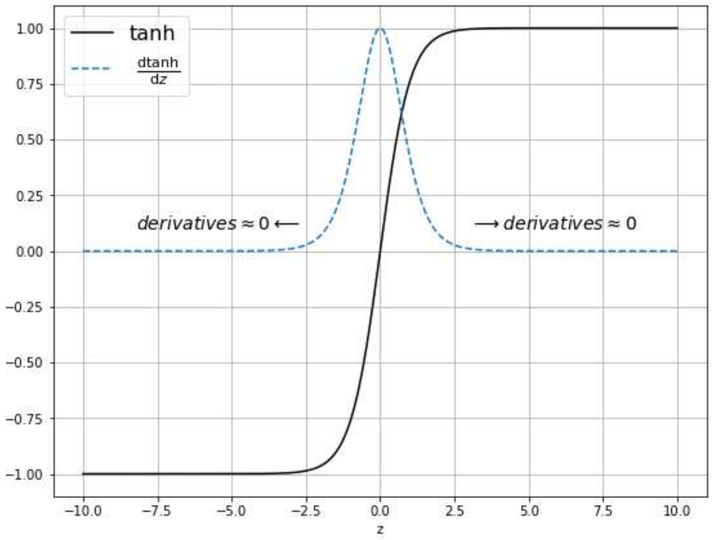


**Figure 2.8: Hyperbolic Tangent Function and its Derivative**

iii. The rectified linear unit (ReLU)

It is the most common activation function, especially in deep neural networks, due to its superior performance over the sigmoid and hyperbolic tangent functions. It handles inputs ranging from $-\infty$ to $+\infty$ and translates them to the range $[0, +\infty]$. It is defined as:

$$ReLU(x) = \max(0, x) \tag{2.17}$$

ReLU is zero for all negative values and linear for positive ones. Since RNNs can have very large outputs, the ReLu activation function is not recommended since it might lead to exploding gradients compared to activation functions with bounded output values. The disadvantage of the ReLU is that all negative input values are transformed into zeros. This implies that once a neuron gets negative, then it is unlikely to recover. This phenomenon is called the dying ReLU and is a unique case of the vanishing gradient problem.

A plot of the ReLU function and its derivative is shown in ***Figure 2.9*** below:

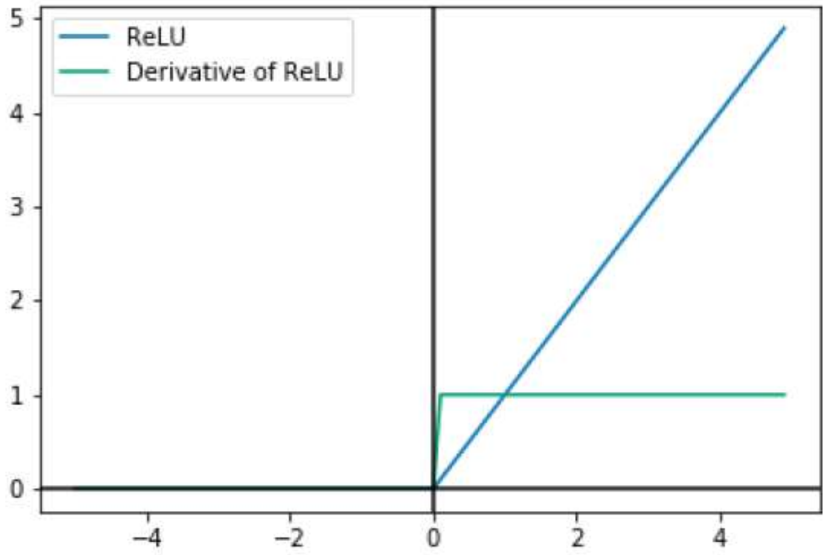


**Figure 2.9: ReLU Function and its Derivative**

iv. Parametric Leaky ReLU

Seeks to resolve the dying ReLU problem by giving the negative input values a negative slope instead of squashing them to zero. It is defined as:

$$PReLU(x) = \max(0, x) + \alpha \cdot \min(0, x) \qquad (2.18)$$

A plot of the PReLU function and its derivative is shown in ***Figure 2.10***:

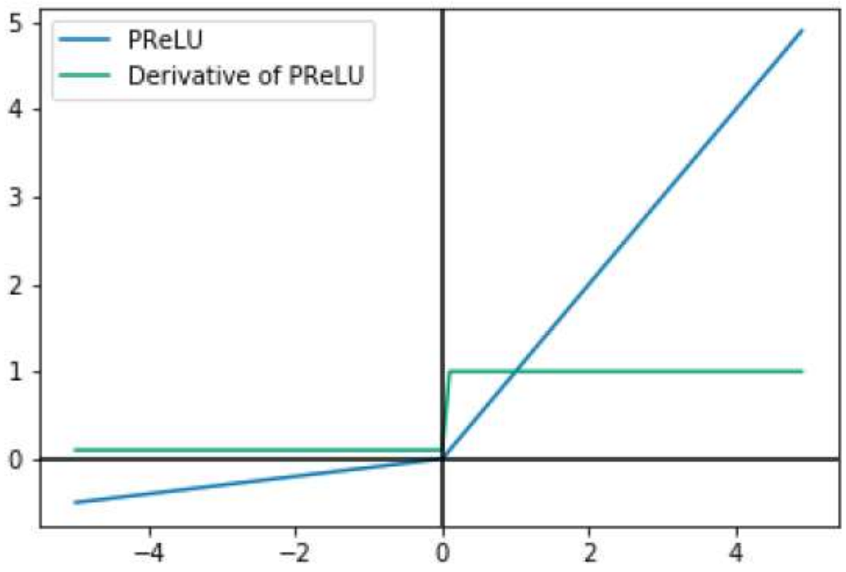


**Figure 2.10: PReLU Function and its Derivative**

v. Exponential Linear Unit (ELU)

It is an improvement over the traditional ReLu. For input values greater than zero, it follows the same rules as ReLU. However, the ELU increases exponentially for negative values to resolve the dying ReLU problem of the traditional ReLU.

It is defined as:

$$ELU(x) = \max(0, x) + \min(0, \alpha(e^{x} - 1)) \qquad (2.19)$$

A plot of the ELU function and its derivative is shown in ***Figure 2.11*** next:

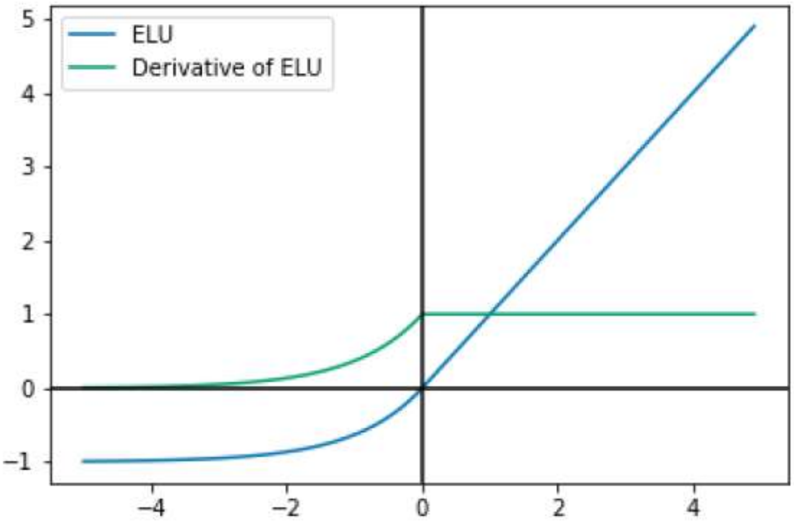


**Figure 2.11: ELU Function and its Derivative**

vi. Scaled Exponential Linear Unit (SELU)

It was proposed for use with self-normalized neural networks. It is defined as:

$$SELU(x) = \gamma \cdot (\max(0, x) + \min(0, \alpha(e^x - 1))) \tag{2.20}$$

Where: $\alpha = 1{:}\,67326324235437728481704299167 17$ and $\gamma = 1{:}\,0507009873554804934193349852946$

Not many studies have been done on the use of SELU.

vii. Swish Function

It doesn’t have an upper bound but instead has a lower bound. It is defined as:

$$Swish(x) = \frac{x}{1+e^x} \tag{2.21}$$

viii. Mish function

It is defined as:

$$Mish(x) = xtanh(\log{(1 + e^x)}) \tag{2.22}$$

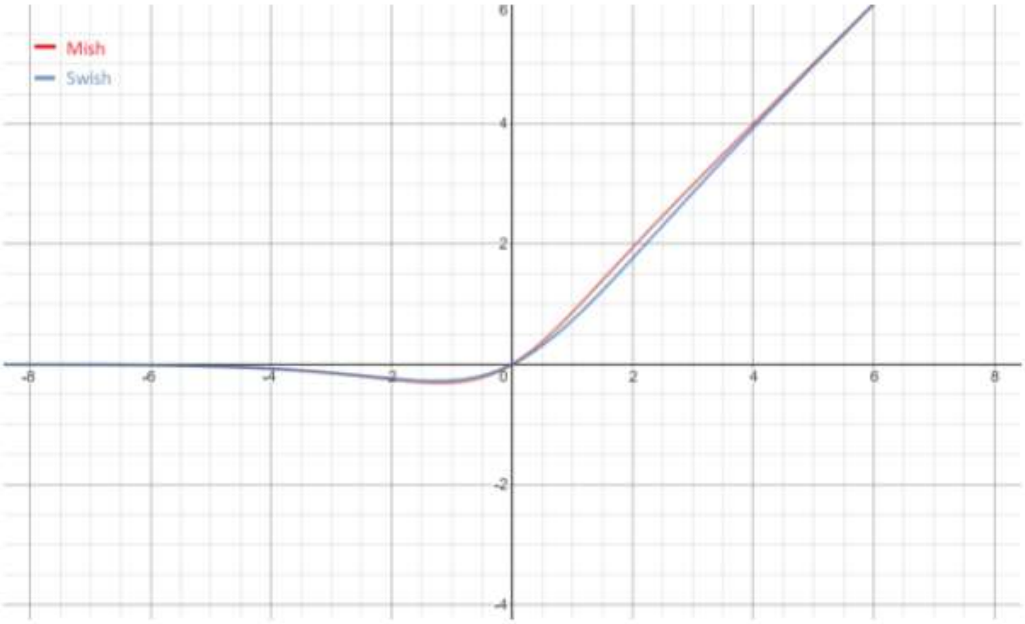


**Figure 2.12: Swish and Mish Functions Comparison**

### 2.6.3 Types of ANNs

Artificial Neural Networks can be broadly categorized into:

i. Feedforward Neural Networks (FFNNs)
ii. Radial Basis Neural Networks (RBFNNs)
iii. Convolutional Neural Networks (CNNs)
iv. Recurrent Neural Networks (RNNs)

The RNNs are optimized to handle time series data and form the basis of the research for this thesis. Next is a discussion of each of the four categories of neural networks.

#### 2.6.3.1 Feedforward Neural Network (FFNN)

It is one of the simplest neural networks. In the FFNN, the data flows in one direction from the input node(s) to the output node(s). Neurons are interconnected by weights that form some weighted associations between the inputs and outputs. The network learns by comparing the processed output vs the actual/target output. If this value is less than the desired threshold, the network then adjusts the values of the interconnecting weights according to a set learning rate based on the error values. ***Figure 2.13*** below shows an illustration of a single-layer FFNN.

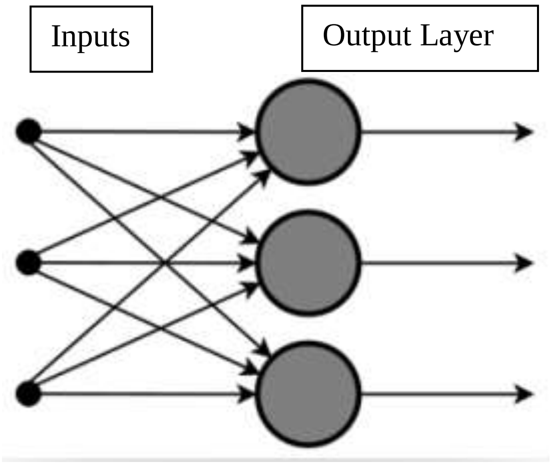


**Figure 2.13: Illustration of a Single-Layer FFNN.**

In the FFNN, the product of the inputs and the corresponding weights connecting them to a given node are calculated and then added together. This sum is then fed to the output, as shown in ***Figure 2.13***. If the FFNN has more than one layer, it is referred to as a Multilayer Perceptron (MLP).

The MLP has three or more layers and is used to classify data that cannot be linearly separated. Every single node in a given layer is connected to each node in the next layer; hence, the MLP is considered a fully connected ANN. An MLP uses a non-linear activation function, e.g., hyperbolic tangent or the logistic function. FFNNs mainly find applications in general regression and classification problems. ***Figure 2.14*** below shows the illustration of an MLP:

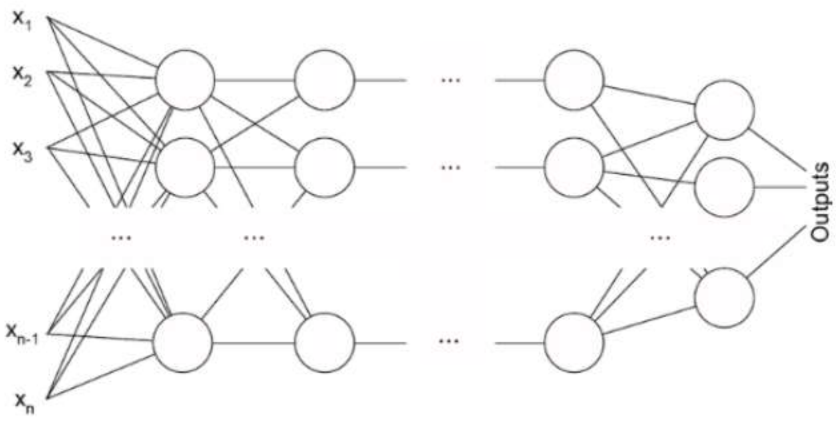


**Figure 2.14: A Multilayer Perceptron**

### 2.6.3.2 Radial Basis Function Neural Network (RBFNN)

RBFNNs are a class of FFNNs. The design of his network is such that it tries to establish the best curve of fit in a high-dimensional space. In getting the surface that provides the best line of fit for the training data, RBFNNs can use that to forecast future values of a given quantity. RBFNNs are applied in power restoration systems to ensure power restoration is done in the shortest time possible.

### 2.6.3.3 Convolutional Neural Network (CNN)

Convolutional networks are very effective in image and video recognition. They are mainly used for object detection and image classification problems. A CNN is a variant of the MLP in that it has several layers. The convolutional layers can either be wholly connected or pooled. ***Figure 2.15*** shows a CNN.

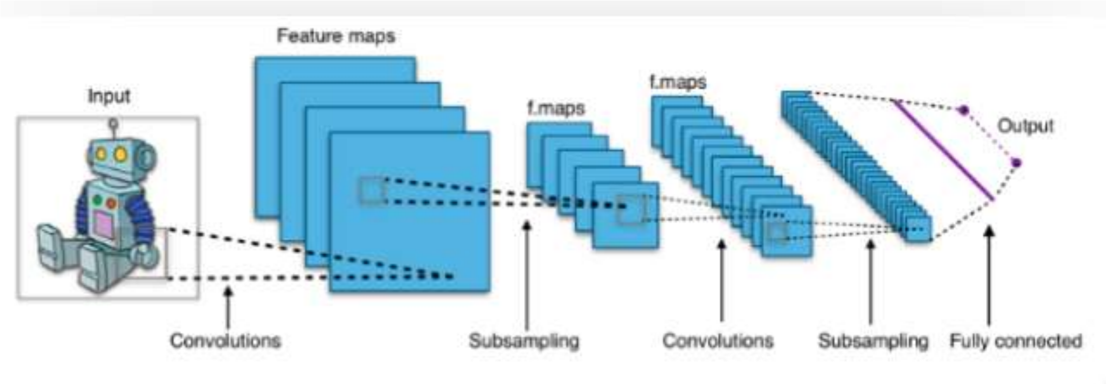


**Figure 2.15: Illustration of a CNN**

### 2.6.3.4 The Recurrent Neural Network (RNN)

The RNN is a type of neural network where the output of a given layer is saved and fed back to the network as an input. RNNs were created to solve the FFNN's problem of not handling sequential data well. The FFNN only considers the current inputs and cannot memorize previous inputs. Also, the input nodes of the FFNN are independent; hence, the neural network cannot learn any temporal relationship present in the input data (Y. Liu et al., 2019). ***Figure 2.16*** below shows a simple illustration of an RNN.

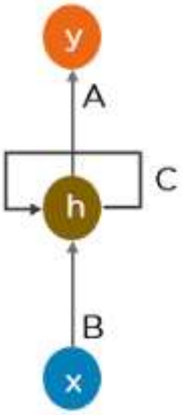


**Figure 2.16: Illustration of an RNN (rolled)**

Expanding/unrolling the illustration in ***Figure 2.16*** to show how the output of the hidden layer is fed back as an input in the next time step, we get the neuron in ***Figure 2.17***:

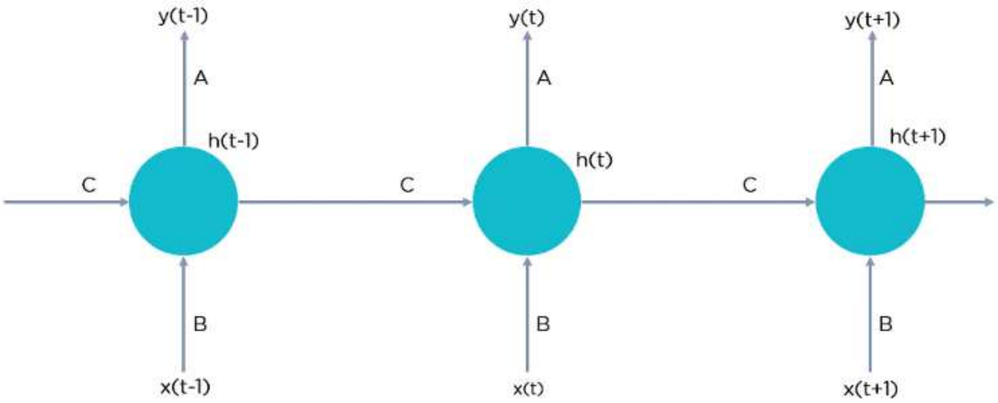


**Figure 2.17: A Fully Connected RNN (unrolled)**

Where:

$h(t)$ is the new state of the network

$x(t)$ is the input

$y(t)$ the output

At the time $(t)$, the state of the network is going to be determined by the input $x(t)$ as well as the previous/old state of the network $h(t-1)$, i.e.,

$$h(t) = f\ (x(t), h(t-1)) \qquad (2.23)$$

Therefore, an RNN forms a chain of interconnected modules of the simple neural network, as shown in ***Figure 2.18***:

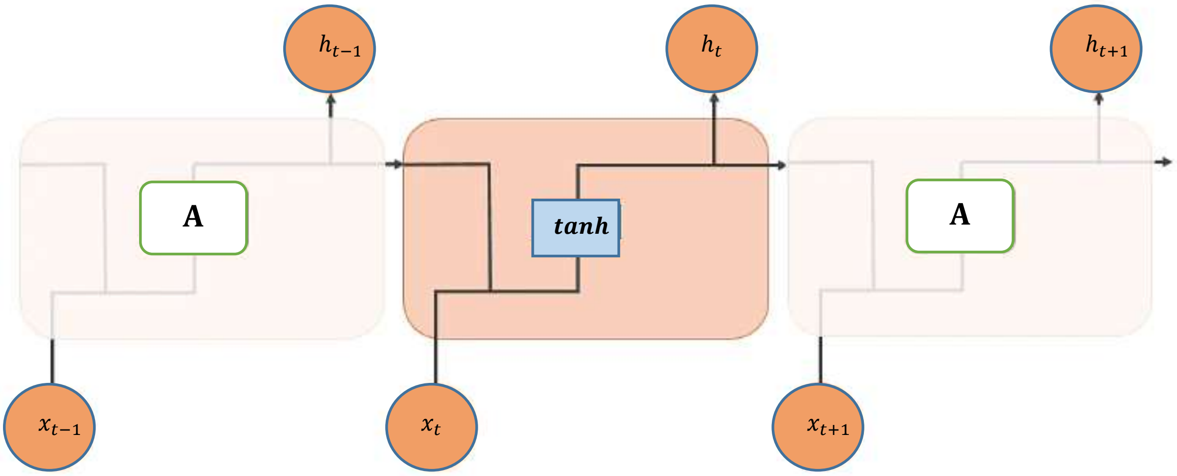


**Figure 2.18: Chain of Simple Interconnected Neural Networks to Form an RNN**

The old state of the network is fetched back to improve the current output of the network. This makes RNNs well-adapted to deal with time series data, image captioning, and natural language processing. Such data usually correlates between previously observed data and the current data.

Standard RNNs have a few limitations:

i. They suffer from the vanishing gradient problem - RNNs work on time-dependent and sequential data problems. Gradients carry the information transmitted over time across the RNN to improve future updates. If the gradient becomes too small, the updates to the network become insignificant, limiting the network's ability to learn through long data sequences.

ii. They suffer from the exploding gradient problem - When training an RNN, if the slope of the data keeps growing exponentially and does not decay, this results in an exploding gradient. When large error gradients accumulate, huge updates are made to the networks for every step, resulting in longer training times and inferior performance. To resolve the limitations of the RNNs, the Long Short-Term Memory (LSTM) neural network was developed. The LSTM has a long-term memory to learn any long-term dependencies in a dataset and store that information for future reference. For the LSTM, instead of having a chain of single neural networks, it has four layers that interact and communicate, as shown in ***Figure 2.19*** below:

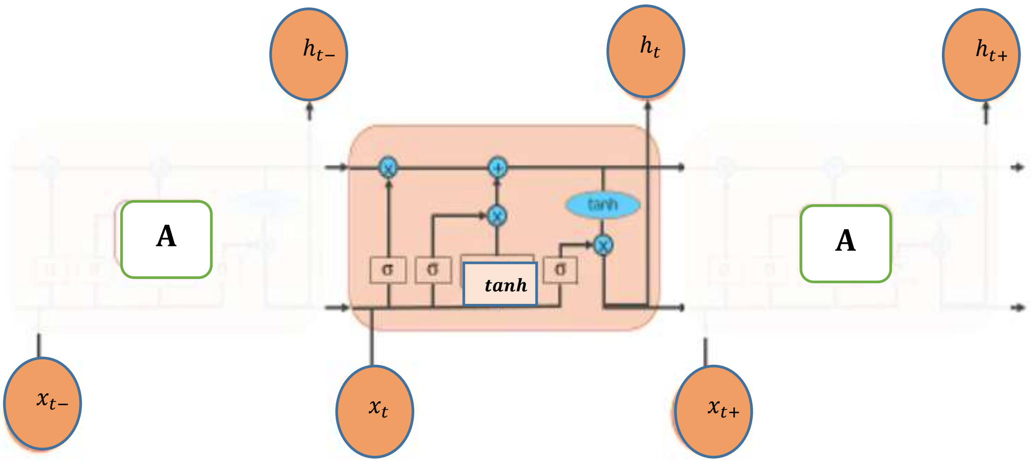


**Figure 2.19: Chain Structure of the LSTM**

The structure of a single LSTM cell is shown in ***Figure 2.20***, and the working of how it processed information is explained next.

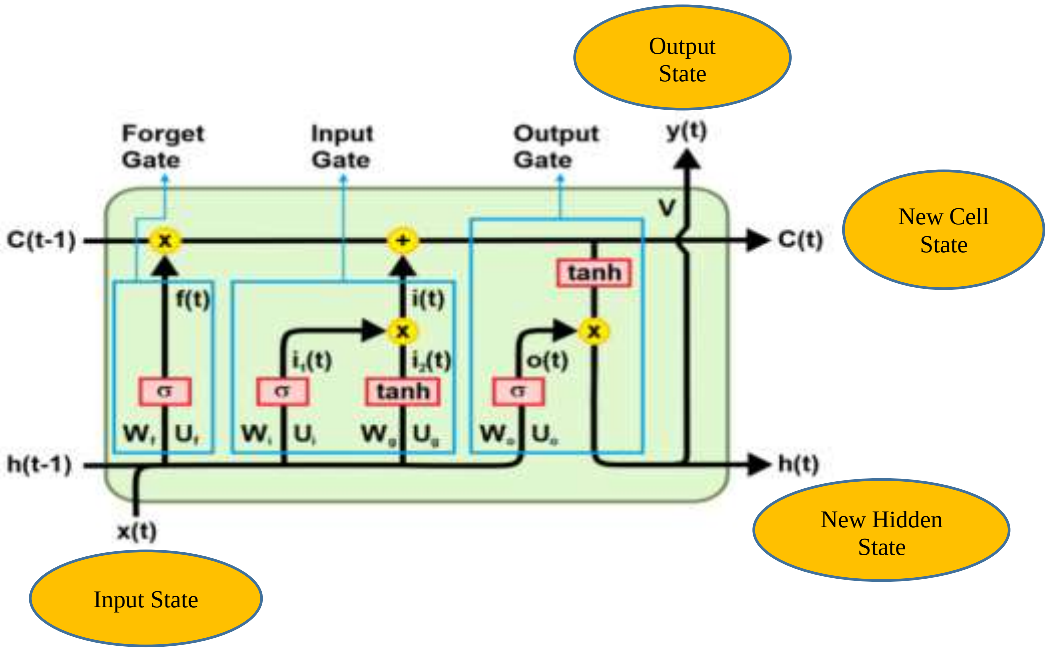


**Figure 2.20: An LSTM Cell**

The information flow of the LSTM cell, as shown in ***Figure 2.20***, is from left to right. The working of the LSTM is as explained below:

**Stage 1: The Forget Gate**

Here, the LSTM decides what to remember or forget from the information received from the previous timestep. The output of the forget gate is formulated as below:

$$f(t) = \sigma(U_f x(t) + W_f h(t-1)) \tag{2.24}$$

The previous hidden state and the current input are passed through a sigmoid function to determine the relevance of the old information to the current input. The sigmoid function gives an output of a 0 or a 1. If the forget gate output $f(t) = 0$ or close to a zero, that information is irrelevant and forgotten. If the forget gate output $f(t) = 1$ or close to one, then that information is remembered.

**Stage 2: The Input Gate**

The input gate updates the cell state with the current input. Initially, the previous hidden state and current input are given as inputs to a sigmoid function. The closer the output is to a 1, the more relevant the information is to the network. In this stage, to further improve the tuning of the network, the previous hidden state and current input are passed through a tanh to squeeze the values between +1 and -1. This gives a weighting to the variables based on their relevance or level of importance. The two outputs are then multiplied element by element. The sigmoid output then determines what information to keep from the tanh output.

$$i_1(t) = \sigma(U_i x(t) + W_i h(t-1)) \tag{2.25}$$

$$i_2(t) = \tanh(U_g x(t) + W_g h(t-1)) \tag{2.26}$$

$$i(t) = i_1(t) + i_2(t) \tag{2.27}$$

**Stage 3: Calculation of the new cell state**

After we get the output of our input gate, we calculate the new cell state as below:

$$c(t) = f(t)c(t-1) + i(t) \qquad (2.28)$$

**Stage 4: The Output Stage**

First, the current cell state part that makes it to the output is determined. The cell state first passes through a tanh function to squash the values between $+1$ and $-1$. The previous hidden state and the current input are summed and passed through a sigmoid function. After that, the two outputs are multiplied element by element, and this gives the matrix of the information to be contained in the next hidden state $h(t)$.

$$o(t) = \sigma(W_o h(t-1) + U_o x(t)) \qquad (2.29)$$

$$h(t) = \tanh\big(c(t)\big) * o(t) \qquad (2.30)$$

The Bidirectional LSTM is an improvement of the LSTM. It combines the LSTM network and the bidirectional RNNs, allowing it to learn long-term dependencies in a dataset (Dolatabadi et al., 2020). The BiLSTM allows better training of the network by traversing the training data twice, i.e. from left to right (forward) and from right to left (backward), hence the term “Bi”.BiLSTM models have better accuracy than LSTM but reach equilibrium more slowly (Namini et al., n.d.). The structure of a BiLSTM network for three consecutive steps is as shown below (Yanga, Mo Wang, 2022):

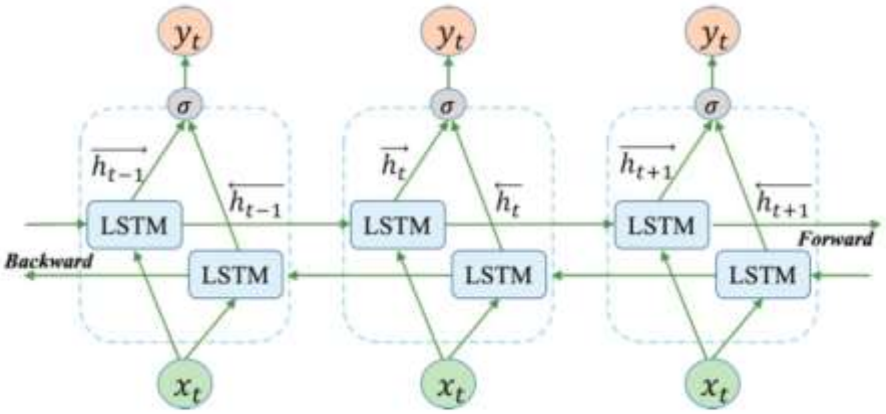


**Figure 2.21: Structure of a BiLSTM Network**

In ***Figure 2.21*** above, the BiLSTM Network is presented, illustrating its bidirectional data processing capabilities. This is achieved by having two LSTM layers - one for each direction. During the forward pass through the data, the forward LSTM processes the input sequence from $x_1, x_2, \dots \dots \dots \dots x_t$ , and processes it sequentially from $t = 1$ to $t = T,$ where T is the sequence length. The backward LSTM processes the data from time $t = T$ backwards to time $t = 1$ (Y. Wang et al., 2021). After that, the forward and backward LSTM outputs are concatenated at each time stamp. This results in a combined hidden state for each time stamp $[h_t, h_t']$ from the forward LSTM and the backward LSTM respectively. The information in the concatenated hidden layers is then used in subsequent time steps for predictions in the time series data. The capability of the BiLSTM networks to capture contextual information and connections from past and future data points makes the BiLSTM a robust tool for time series prediction. The BiLSTM has been successfully implemented in time series forecasting problems and have been shown to have superior performance over the LSTM models (Dolatabadi et al., 2020). By sweeping through the data in the forward and backward directions, the BiLSTM network can identify some additional connections in time series data that cannot be identified by the traditional LSTM and is recommended as a better tool for time series forecasting (Namini et al., 2019).

### 2.7 Data Decomposition Techniques.

Time series decomposition is a statistical task that involves breaking down data into several high and low-frequency components to extract information on seasonality or trends from the data (Chourasia, 2020). Time series data is usually very noisy and complex, and decomposition helps break it down into simpler components. Decomposition gives a better understanding of the data by highlighting any repetitive seasons in the data or increasing or decreasing trends over time. Data decomposition reduces the non-stationarity and non-linearity of a time-series dataset, making it easier to predict than the original data series (Y. Liu et al., 2019).

Data decomposition models can be classified into *additive* and *multiplicative* models. In the additive model, the variance of the data does not change much over time, and the trend line is linear (Chourasia, 2020). To get the original time series, one only needs to add the lower frequency components the data is decomposed into. i.e.,

$$y(t) = s(t) + T(t) + R(t) \qquad (2.31)$$

Where: $y(t)$ - Original time series

$s(t)$- Seasonal component

$T(t)$ - Trend component

$R(t)$ - Residual component

In the multiplicative model, the seasonal and trend components also increase as the data increases over time. Hence, to reconstitute the original time series, one multiplies all the components the trend was broken down into as shown below (Chourasia, 2020):

$$y(t) = s(t) \text{ x } T(t) \text{ x } R(t) \qquad (2.32)$$

The working of data decomposition is illustrated in ***Figure 2.22***:

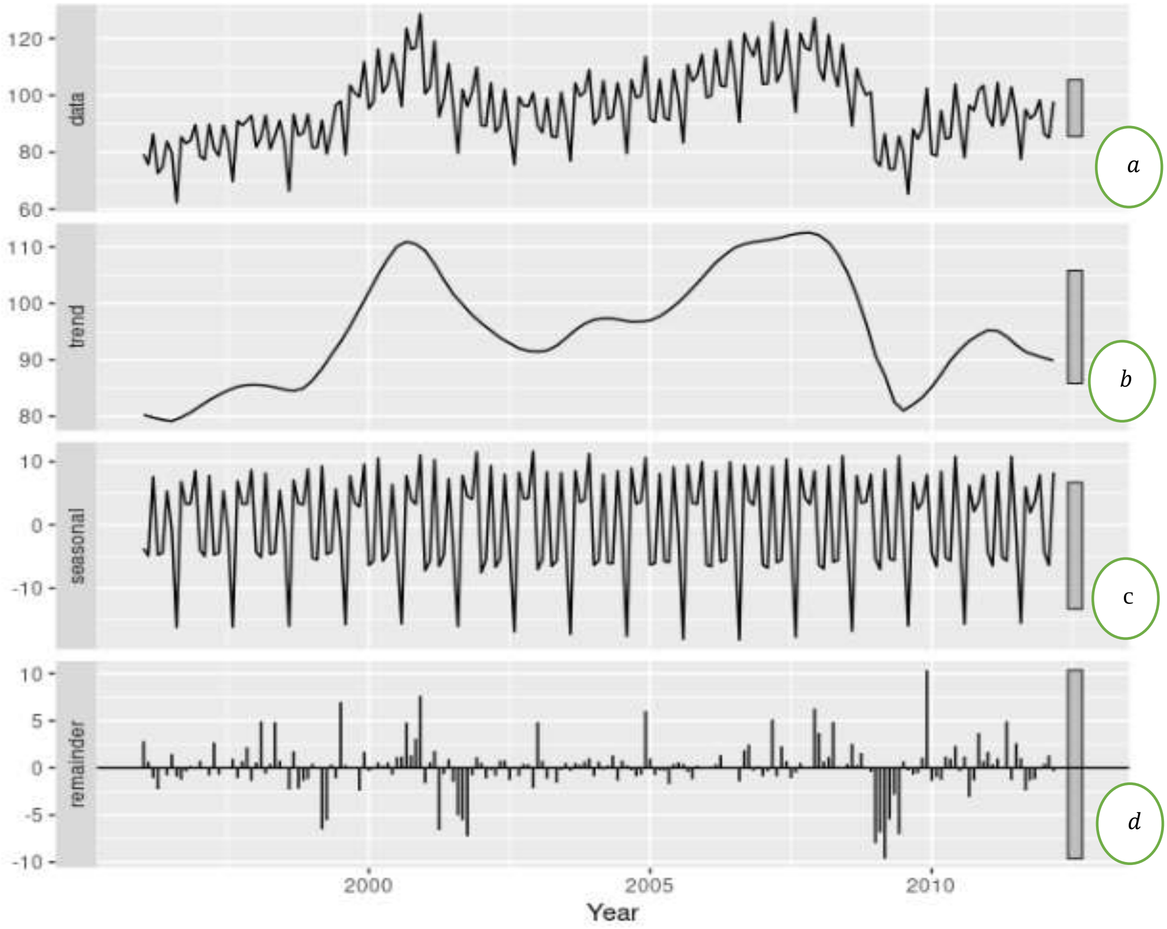


**Figure 2.22: An Illustration of Data Decomposition (McHugh, 2020).**

***Figure 2.22*** above illustrates how data decomposition can bring out clear patterns of seasonality from a dataset. The original time series data is shown in the graph $(a)$. The trend component is extracted as illustrated in $(b)$ and shows any long-term directional property of the time series data. It indicates whether the data is increasing, decreasing or remaining constant relative to time. The extracted seasonal component is illustrated in $(c)$ and highlights the repetitive patterns or events occurring in the data at fixed intervals such as daily, weekly or yearly cycles. These repetitive patterns indicate the cycles of seasons, especially in time series data that rely on weather parameters. The residual or remainder is shown in $(d)$ and is commonly referred to as the noise. It represents the unexplained variability in the data once the seasonal and trend components have been accounted for.

There has been a lot of research on the significance of data decomposition of wind time series data on the accuracy of forecasts. Most researchers hybridize their algorithm of choice with a given data decomposition technique, and the results that have been posted show the improvements that are achieved with the inclusion of data decomposition. The original time series/data is broken down into simpler constituent components, which help improve the prediction of the data since the selected algorithm can now be optimized to forecast each of the components, after which the components are reconstituted to get the forecast series. The data decomposition techniques that have been previously used include the wavelet transform, empirical mode decomposition, variational mode decomposition, and their variants.

### 2.7.1 Wavelet Transform

It consists of two variants. The Continuous Wavelet Transform (CWT) and the Discrete Wavelet Transform (DWT). For a given input signal $W(t)$, the two variants of wavelet transform can be expressed as shown next:

$$CWT(u,v) = \frac{1}{\sqrt{u}}\int_{-\infty}^{\infty} W(t)\psi^{*}(\frac{t-v}{u})dt \qquad (2.33)$$

Where: $u$ - scale factor

$v$ - translation parameter

$\psi(t)$ - represents the mother wavelet

$$DWT(x,y) = 2^{-\frac{x}{2}}\sum_{j=0}^{L-1} W(t)\psi^{*}(\frac{t-y.2^{x}}{2^{x}})dt \qquad (2.34)$$

Where: $L$ - length of the input signal $W(t)$

$x$ - scale factor

$y$ - translation parameter

$\psi(t)$ - represents the mother wavelet

DWT is more popular than CWT due to its computational efficiency and data compression capability. Wavelet transform breaks down the data into a single low-frequency component and multiple high-frequency components. The accuracy of WT depends on the selection of the mother wavelet and the number of levels of decomposition chosen. The available mother wavelet options include Daubechies, Haar, Morlet, and Mexican Hat (Jaseena & Kovoor, 2021).

### 2.7.2 Empirical Mode Decomposition (EMD)

Hilbert Huang developed EMD in 1998. It entails the decomposition of a time series into finite nearly monotonic intrinsic mode functions (IMFs) and one residual component, as shown below:

$$W(t) = \sum_{i=1}^{n} IMF_i(t) + r_n(t) \tag{2.35}$$

Where: $W(t)$ - input signal

$IMF_i(t)$ - intrinsic mode functions

$r_n(t)$ - residual component

$n$ - number of IMFs

Mode mixing refers to the scenario where an IMF obtained using EMD has components of different/various frequencies. EMD offers advantages such as adaptability to non-linear and non-stationary data, making it suitable for analyzing real-world signals like financial, biomedical, and environmental data (Zhang et al., 2018). EMD does not rely on predetermined basis functions and automatically extracts the inherent oscillatory modes within the data. It is widely used in signal processing, fault diagnosis, and trend analysis, providing a powerful tool for uncovering hidden patterns and extracting valuable information from complex and dynamic datasets.

## 2.8 Previous Neural Network-Based Research on Wind Prediction.

In (Singh, 2016), the FFNN was used to predict wind power and gave MSE and MAE values of 13.5 and 9.35 for a short-term forecast showcasing the promising capabilities of using FFNNs in wind power forecasting. This research did not, however, explore further ways that the FFNNs would be enhanced for better forecasts and did not compare its performance against other prediction approaches. In (Kassa et al., 2016), ANFIS was used to predict wind power and compared against the BPNN and GA-BPNN hybrid models. It was observed that ANFIS performed better than the BPNN but only showed slightly better performance compared to the GA-BPNN hybrid model. The BPNN model had 11 hidden layer neurons and was trained using the Levenberg-Marquardt algorithm with tangent sigmoid and pure linear activation functions at the input and output. This paper highlighted the importance of hybridization since the GA_BPNN model had better performance than BP alone, and its performance was almost equal to the ANFIS. In (Q. Chen et al., 2018), ANFIS, ANN, and ARIMA were used to predict wind power. With ten hidden layers and trained using Bayesian Regularization, the ANN was the best prediction algorithm for 1-hour ahead forecasts with RMSE and MAE values of 18.1 and 12.3. ANFIS came in at a close second with RMSE and MAE values of 18.4 and 12.2. This research showed the robust capabilities of models incorporating neural networks compared to statistical methods.

In (Shao et al., 2016), a neural network was used together with wavelet decomposition to forecast wind speed. The neural network had three layers with seven input neurons, 12 hidden layer neurons, and one output neuron for spring data. The network had nine-input neurons and 14 hidden layer neurons for the rest of the three seasons. The proposed network was enhanced using wavelet decomposition and adaptive boosting and outperformed fuzzy logic and classical FFNNs. The hybrid network with wavelet decomposition outperformed the base models, showcasing the advantages of improving forecast accuracy by incorporating data decomposition.

In (J. Wang, 2014), Wavelet Transform (WT) was combined with a two-hidden layer neural network to forecast wind speed. The data was decomposed into four components using wavelet transform. A neural network with two hidden layers was used to predict each of the obtained sub-series, after which the obtained forecasts were combined to get the final prediction. The proposed WT-TNN approach outperformed the NN, and the lowest RMSE and MAE values obtained for the four test datasets were 0.3317 and 0.2312. The TNN's corresponding RMSE and MAE values were 0.8019 and 0.4767, respectively. The superiority of the WT-TNN again showed the significance of including a data decomposition technique in a forecast.

In (Berrezzek et al., 2019), the Discrete Wavelet Transform (DWT) was combined with ANNs for wind speed forecasting. A 5-level data decomposition was performed using DWT to give five detail coefficients and an approximation coefficient for each data value. Each detail and approximation coefficient had its neural network consisting of five inputs, two hidden layers, and one output. Five previous daily average wind speeds were used to forecast the wind speed for the sixth day, i.e., a sliding window of 5 days was used. After the prediction, all the outputs were reconstructed to give the final forecast. The dataset consisted of 4375 daily average wind speeds spanning twelve years from 2007 to 2018. The training data was from 2007 to 2017 (4015 data points), while the testing data used was for 2018 (360 points). The least RMSE and MAPE values posted for the test sets were 0.1265 and 0.0371. Once more, the superior performance of incorporating data decomposition in prediction using neural networks was illustrated.

In (Z. Zheng et al., 2012), empirical mode decomposition was used with RBFNN to forecast wind power with three input parameters - wind power, speed, and direction. EMD decomposed the data into eight intrinsic mode functions (IMFs) and one residual function. 900 samples were used to test the model, and 74 were used for testing. An RBFNN network was used for each of the obtained data sub-series, after which the values were combined to give the final prediction. The EMD_RBFNN model was compared to the RBFNN. EMD_RBFNN posted MAE and MSE values of 22.075 and 2.485, respectively, while RBFNN alone posted MAE and MSE values of 34.905 and 4.293, respectively. The results again highlighted the advantage of data decomposition and showed the strength of EMD as a data decomposition tool.

Wind speed was forecast using EMD, ANNs, and ARIMA in (Bantupalli, 2017), and the results were compared against ANN and EMD_ANN. EMD_ANN outperformed ANN alone, posting RMSE and MAE of 2.058 and 0.713 against RMSE and MAE of 2.949 and 1.456, respectively. The EMD_ANN_ARIMA was the best amongst the three, with RMSE and MAE of 0.890 and 0.527, respectively. In (H. Liu et al., 2019), a hybrid method combining EEMD and neural networks (Group Method of Data Handling networks) was used to forecast wind speed. The researcher compared the proposed method with GMDH neural networks alone, GMDH neural networks with wavelet packet decomposition, SVM, and ELM. The proposed method posted minimum RMSE and MAPE values of 0.0113 and 0.6337. These results were better than the algorithms used without any data decomposition techniques. Based on the results, EEMD performed better than WT in the same hybrid scenarios. In (Bantupalli, 2017) and (H. Liu et al., 2019), the strength of using a data decomposition technique based on Empirical Mode Decomposition was highlighted.

In (Quan & Shang, 2021), a method based on variational mode decomposition (VMD) and the Bare Bones Fireworks Algorithm (BBFA) was proposed to forecast wind speed. Again, compared to the same approaches but with no data decomposition involved, the hybrid approaches using data decomposition came out superior. VMD_BBFA had the least RMSE and MAE values of 0.20 and 0.17. In (Shahid et al., 2021), LSTM and GA were combined to predict wind power for seven wind farms and showed an improvement of between 6% - 30% in wind power forecast accuracy. The LSTM network had three layers, a batch size of 30 and 10 epochs. The window size used is three timesteps. In (Wu, Qianyu Guan1, Fei Lv, Chen Huang, n.d.), a hybrid of CNN-LSTM was used in wind power forecasting. The CNN was used to extract the temporal and spatial correlations between the parameters affecting wind power and the wind turbines' location on a wind farm. The CNN_LSTM improved the forecast accuracy by 14% compared to the LSTM model, highlighting the advantages of hybridized models in wind power forecasting.

## 2.9 Summary and Research Gaps

Much research has been done on the wind power prediction problem, as discussed in section 2.8. However, it is noted that there are opportunities for further research in improving wind power forecasts. Wind power forecasting is an area of research with significant untapped potential, as it has not yet reached its maximum capabilities. Unlike load forecasting, which is now a mature science and has been used for a long time, wind forecasting is still a relatively new phenomenon and new models continue to be developed. This is due to the fact that wind power **prediction errors** are still significant and can be improved further with newer technologies. Load forecasting has been refined and tuned to errors of 1-3% of the actual values, while wind power forecasting errors are much higher.

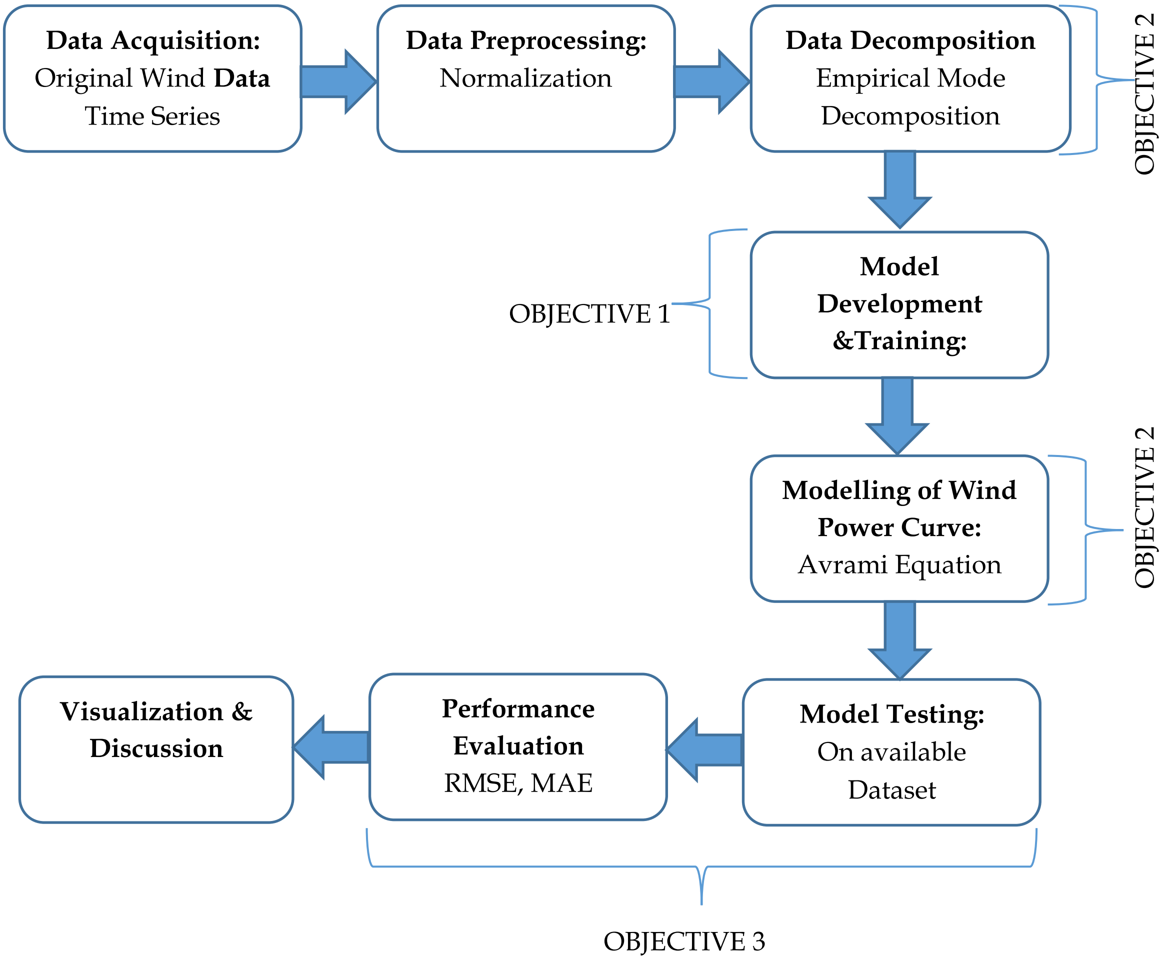
With precise wind power forecasts, curtailment of wind power and over-allocation of spinning reserves can be reduced significantly, allowing for more accurate power dispatch. To guarantee the optimal operation of a power system, more precise wind power prediction tools and approaches must be developed to ensure optimal utilization of this free resource when available. The future of wind power prediction lies in using online/real-time data, especially for short-term wind forecasting, which is the gap this research aims to fill. LSTM-based neural networks are noted to have superior capabilities in handling time series data due to their ability to capture short and long-term dependencies without suffering from the vanishing and exploding gradient limitations of RNNs. The BiLSTM is a further enhancement of the LSTM since it allows better training of a network by traversing through the training data twice. EMD, as a data decomposition method, has shown great adaptability in handling non-linear, complex, rapidly changing time series data such as wind power data. It has good time-frequency localization capabilities and is very effective in capturing and highlighting non-linear patters in a dataset. In this research, a novel approach using BiLSTM, data decomposition and a wind power curve modelled from the Avrami equation is used to develop a much more accurate wind power prediction approach.

# CHAPTER THREE

# METHODOLOGY

## 3.1 Outline of the Research Stages

The developed hybrid approach utilizes the BiLSTM neural network and empirical mode decomposition to forecast wind power. The developed method follows the following stages, as illustrated in ***Figure 3.1***, and the detailed flowchart is shown in ***Figure 3.2***.



Data Acquisition: Original Wind Data Time Series
Data Preprocessing: Normalization
Data Decomposition Empirical Mode Decomposition
OBJECTIVE 2
Model Development &Training:
OBJECTIVE 1
Modelling of Wind Power Curve: Avrami Equation
OBJECTIVE 2
Model Testing: On available Dataset
Performance Evaluation RMSE, MAE
Visualization & Discussion
OBJECTIVE 3

**Figure 3.1: Illustration of the Various Stages of the developed BiLSTM+EMD+Avrami Power Curve Model.**

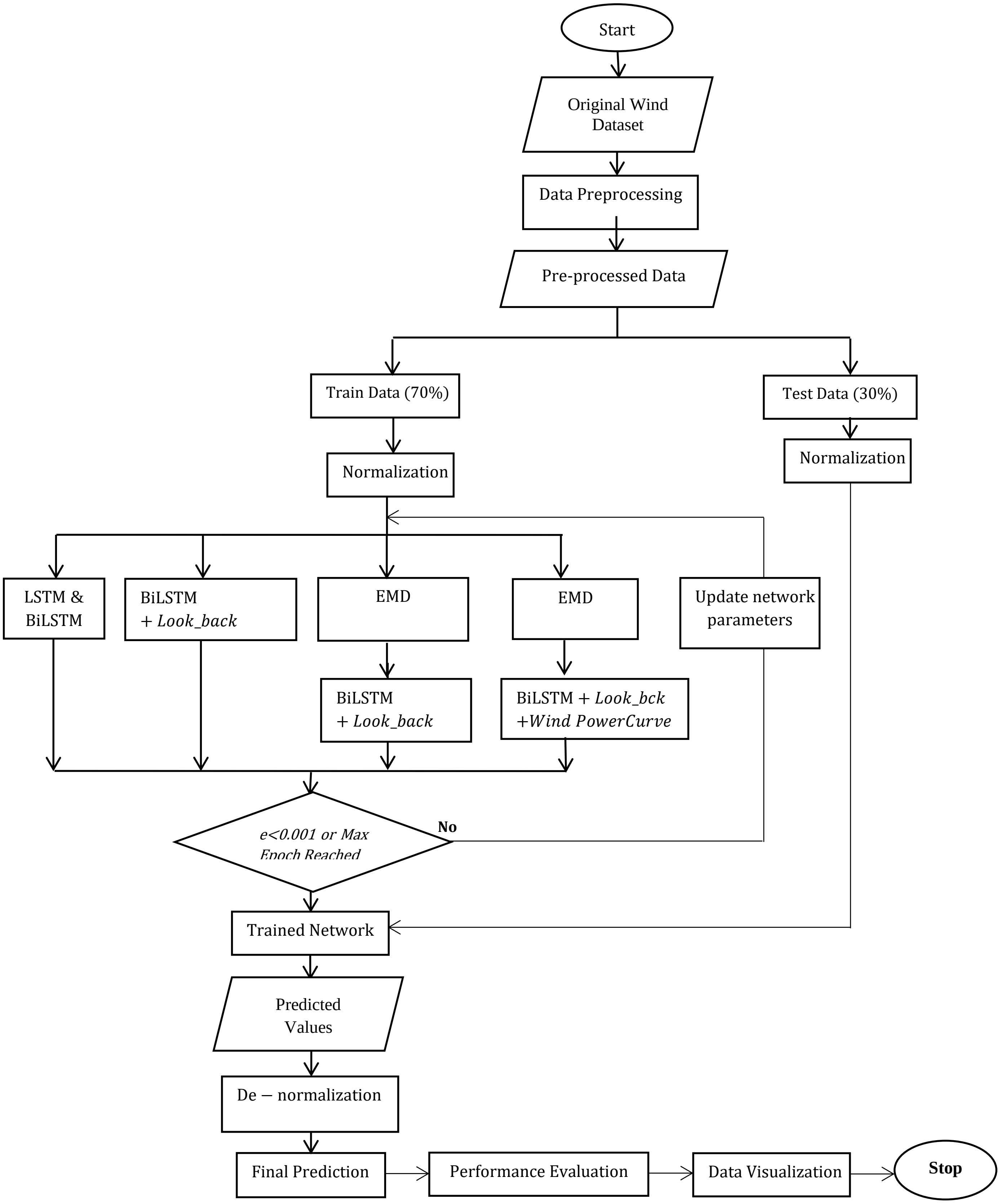


**Figure 3.2: General Wind Power Forecasting Procedure**

## 3.2 Dataset Description and Handling

### 3.2.1 Data Description.

The selected dataset is an online-based dataset used by various researchers looking into the wind power forecasting problem. The particulars of the dataset are summarized in ***Table 3.1*** below:

**Table 3.1: Particulars of the Wind Dataset**

| **Data Name** | **Source** | **Data Resolution** | **Span** |
|---|---|---|---|
| AL_WIND_07_12 | National Renewable Energy Laboratory (**NREL**) | **Hourly** observations | 01/01/2007 00:00:00 to 31/12/2012 23:00:00 (**52560** observations equivalent to 6 years of data) |

*** Dataset available on**: https://github.com/ShashwatArghode/Wind-Energy-Prediction-using-LSTM/blob/master/AL_WIND_07_12.xlsx
or on
https://developer.nrel.gov/docs/wind/wind-toolkit/wtk-download/

The AL_WIND_07_12 wind dataset consists of the following variables: a timestamp, air temperature ($^{o}$C), pressure (atm), wind speed (m/s), wind direction (deg) and wind power (kW). Correlation is a measure of the strength of association between two variables. The correlation between other variables and wind power is determined. According to (Schober et al., 2018), correlation can be categorized from negligible correlation to very strong correlation, as presented in ***Table 3.2***:

**Table 3.2: Conventional Approach of Interpreting Correlation Coefficients.**

| Absolute Value of Correlation Coefficient | Interpretation |
|---|---|
| 0.00 - 0.10 | Negligible Correlation |
| 0.10 - 0.39 | Weak Correlation |
| 0.40 - 0.69 | Moderate Correlation |
| 0.70 - 0.89 | Strong Correlation |
| 0.90 - 1.00 | Very Strong Correlation |

The correlation of the various weather parameters/variables against wind power in the AL_WIND_07_12 dataset is determined using Pearson's correlation, and the results are summarized in ***Table 3.3*** below:

**Table 3. 3: Pearson's Correlation Results for AL_WIND_07_12 Dataset.**

| Pearson's Correlation Test Results (AL_WIND_07_12) | | |
|---|---|---|
| Wind **Speed** vs Wind Power | 0.9435 | Very Strong Positive Correlation |
| Wind **Direction** vs Wind Power | -0.0090 | Negligible Negative Correlation |
| Wind **Temperature** vs Wind Power | -0.2608 | Weak Negative Correlation |
| Wind **Pressure** vs Wind Power | -0.0764 | Negligible Negative Correlation |

As observed in ***Table 3.3*** above, wind speed is the variable that has the strongest correlation to wind power, and this can be explained by the fact that wind speed has a quasi-cubic relationship with wind power, as shown in *Equation 3.1* below:

$$P_{av} = \frac{1}{2} C_p \rho A v^3 \qquad (3.1)$$ (****Retrieved from Equation 2.4)***

Where: $P_{av}$ is Average Power and $v$ is the wind speed.

The dataset was split into 70% training and 30% testing data for each test. Empirical analysis demonstrates that optimal outcomes are achieved when 20-30% of the observed data points are reserved for testing, and the remaining 70-80% are used for training. In time series data analysis, the choice of 70% training and 30% testing data split acknowledges the temporal dependencies within the data. The smaller testing set allows models to learn from a more extended history, capturing time-evolving patterns and ensuring better forecasting accuracy over time (Gholamy, Afshin Kreinovich, Vladik Kosheleva, 2018).

### 3.2.2 Data Preparation

#### 3.2.2.1 Data Preprocessing

Data preprocessing in time series data forecasting is one of the most important steps before any forecasting is done. It involves estimating missing values, identifying and filling existing outliers, and data feature scaling. ***Figure 3.3*** below shows these stages of data pre-processing.

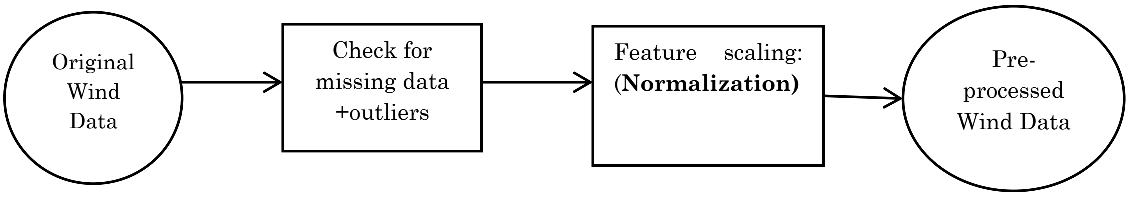

Original Wind Data
Check for missing data +outliers
Feature scaling: (**Normalization)**
Pre-processed Wind Data

**Figure 3.3: Stages of Pre-Processing.**

The following procedure is followed in the data preprocessing:

i. Checking for missing data - This is done by using the *rmmissing* function in MATLAB. The function has the option of removing rows or columns with missing data. One could specify the *Min Num Missing* threshold below which rows or columns with missing data are not deleted. The dataset used did not have any missing data.

ii. Handing Outliers - The *isoutlier* MATLAB function is used to identify the outliers in the wind dataset, after which any outliers are *clipped*. The AL_WIND_07_12 dataset did not have any outliers.

iii. Feature scaling - This is done to prevent the domination of one variable on the output. There are two types of standard feature scaling options, Standardization and normalization, and their formulas are represented in Equations 3.2 and 3.3 below:

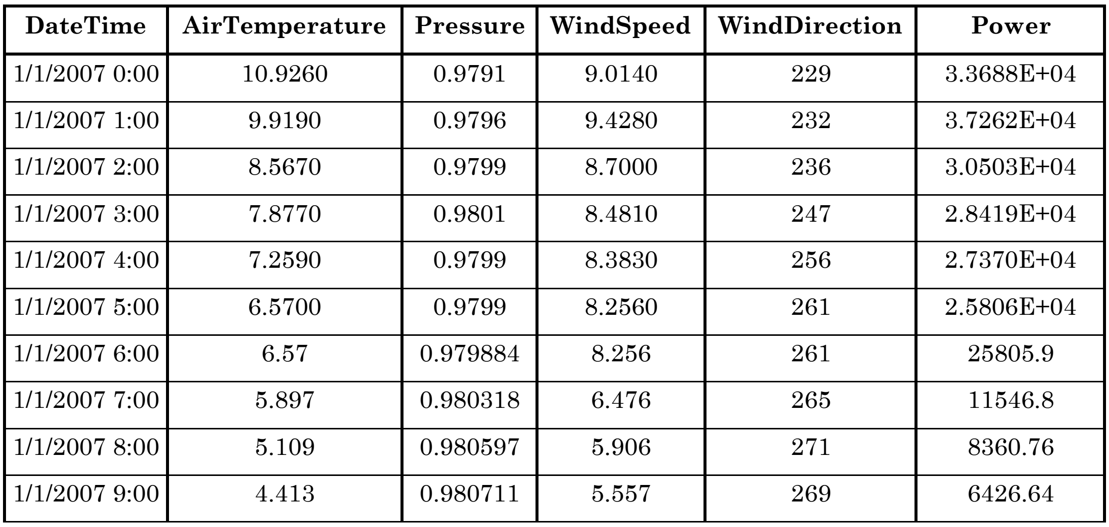
$$x_{transformed} = \frac{x - mean(x)}{std.dev(x)} \qquad (3.2)$$

$$x_{transformed} = \frac{x - min(x)}{\max(x) - \min(x)} \qquad (3.3)$$

Where:

$x$ is the data before scaling and $x_{transformed}$ is the data after scaling.

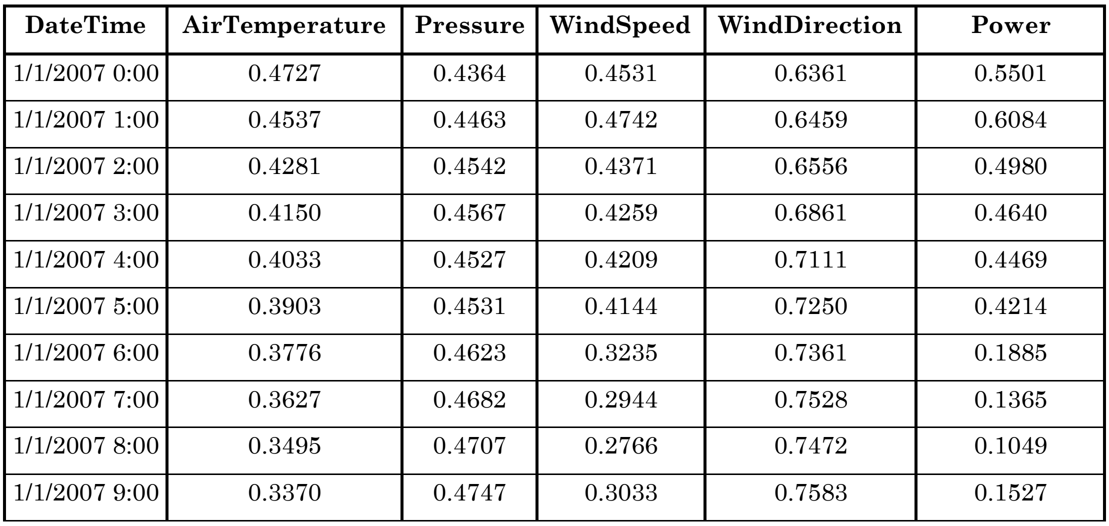
Standardization scales the data to the range of $-1 < x < 1$, while normalization squashes the data to the range $0 < x < 1$. For this work, **normalization** was used as the feature scaling method. This ensured that all the data values were scaled to 0 and 1. Normalization ensures that no one factor or parameter is weighted higher than others based on their numerical values. For example, in the absence of feature scaling, the wind direction would be weighted higher than wind speed since wind direction values (angles) typically have larger values compared to wind speed values. ***Figure 3.4*** below is a section of the un-normalized data in the AL_WIND_07_12 dataset for illustration.



| DateTime | AirTemperature | Pressure | WindSpeed | WindDirection | Power |
|---|---|---|---|---|---|
| 1/1/2007 0:00 | 10.9260 | 0.9791 | 9.0140 | 229 | 3.3688E+04 |
| 1/1/2007 1:00 | 9.9190 | 0.9796 | 9.4280 | 232 | 3.7262E+04 |
| 1/1/2007 2:00 | 8.5670 | 0.9799 | 8.7000 | 236 | 3.0503E+04 |
| 1/1/2007 3:00 | 7.8770 | 0.9801 | 8.4810 | 247 | 2.8419E+04 |
| 1/1/2007 4:00 | 7.2590 | 0.9799 | 8.3830 | 256 | 2.7370E+04 |
| 1/1/2007 5:00 | 6.5700 | 0.9799 | 8.2560 | 261 | 2.5806E+04 |
| 1/1/2007 6:00 | 6.57 | 0.979884 | 8.256 | 261 | 25805.9 |
| 1/1/2007 7:00 | 5.897 | 0.980318 | 6.476 | 265 | 11546.8 |
| 1/1/2007 8:00 | 5.109 | 0.980597 | 5.906 | 271 | 8360.76 |
| 1/1/2007 9:00 | 4.413 | 0.980711 | 5.557 | 269 | 6426.64 |

**Figure 3.4: A Section of the AL_WIND_07_12 Wind Power Dataset before Normalization.**

Note the large magnitude of the wind **direction** values compared to wind **speed** values. If both of these parameters are being used as input variables, they need to be normalized to ensure that one parameter does not overshadow the impact of another on the variable being predicted. ***Figure 3.5*** below presents the same data with the variables now normalized. Note how the distribution of all variables (other than the timestamp) is now a value between 0 and 1.



| DateTime | AirTemperature | Pressure | WindSpeed | WindDirection | Power |
|---|---|---|---|---|---|
| 1/1/2007 0:00 | 0.4727 | 0.4364 | 0.4531 | 0.6361 | 0.5501 |
| 1/1/2007 1:00 | 0.4537 | 0.4463 | 0.4742 | 0.6459 | 0.6084 |
| 1/1/2007 2:00 | 0.4281 | 0.4542 | 0.4371 | 0.6556 | 0.4980 |
| 1/1/2007 3:00 | 0.4150 | 0.4567 | 0.4259 | 0.6861 | 0.4640 |
| 1/1/2007 4:00 | 0.4033 | 0.4527 | 0.4209 | 0.7111 | 0.4469 |
| 1/1/2007 5:00 | 0.3903 | 0.4531 | 0.4144 | 0.7250 | 0.4214 |
| 1/1/2007 6:00 | 0.3776 | 0.4623 | 0.3235 | 0.7361 | 0.1885 |
| 1/1/2007 7:00 | 0.3627 | 0.4682 | 0.2944 | 0.7528 | 0.1365 |
| 1/1/2007 8:00 | 0.3495 | 0.4707 | 0.2766 | 0.7472 | 0.1049 |
| 1/1/2007 9:00 | 0.3370 | 0.4747 | 0.3033 | 0.7583 | 0.1527 |

**Figure 3.5: A Section of Normalized AL_WIND_07_12 Wind Power Dataset.**

### 3.2.3 Look Back period

Look back refers to the number of observations in the past that are used to make predictions for future data points. The choice of a look-back period is critical since it significantly impacts the prediction accuracy of a developed model. A longer look-back captures more historical information but risks introducing potentially irrelevant data. A very short look-back period results in a simpler model but one that might not be able to identify seasonal information in a dataset. Wind power is a non-stationary time series, which means that at a given future time $t$, previous observations of the time series variable affect future value observation at time $t$ (Y. Liu et al., 2019). For the scenario where historical wind power observations shall be used to predict future wind power, the look-back period is a crucial parameter to determine. In (Mathenge et al., 2021), this was examined for a look-back period of 1 and produced better forecast results than the forecast without a look-back. Autocorrelation exists in time series dataset observations at time t and those that occurred earlier, i.e., at time $t-1$, $t-2$, $t-3$, $t-n$ where $n$ is the look back period where the obervations at time $t-n$ still bear a strong correlation with the obervations at time $t$. To determine the most ideal look back period for the dataset, pearson' correlation is used and the autocorrelation results are obtained as below:

**Table 3.4: AL_WIND_07_12 Correlation**

| **Dataset: AL_WIND_07_12** | | |
|---|---|---|
| Look_back steps | Autocorrelation | Interpretation |
| 1 | 0.9250 | Very strong correlation |
| 2 | 0.8368 | Strong Correlation |
| 3 | 0.7594 | Strong Correlation |
| 4 | 0.6878 | Moderate Correlation |
| 5 | 0.6197 | Moderate Correlation |
| 6 | 0.5562 | Moderate Correlation |
| 7 | 0.4976 | Moderate Correlation |
| 8 | 0.4455 | Moderate Correlation |
| 9 | 0.4006 | Moderate Correlation |
| 10 | 0.3616 | Weak Correlation |

**Note: The AL_WIND_07_12 dataset data observations are spaced every 1 hour.*

### 3.3 Data Decomposition.

Empirical Mode Decomposition (EMD) was used to break down the input data into more basic, less noisy signals. The original data signal is broken down into more physically meaningful components of different frequencies, making it easier to identify trends and patterns. In this research, data decomposition is applied for the scenario where wind speed is used to forecast wind power. The wind speed and power are decomposed into different frequency components; then, corresponding mode components are used as input and output. Since time series decomposition is a statistical task that involves breaking down data into several high and low-frequency components, decomposition of the data makes it easier to identify trends on seasonality or other trends from the data.  Time series data is usually very noisy and complex, and decomposition helps break it down into simpler components for the neural network-based algorithms to handle.

A flowchart of the prediction process that entails decomposition by EMD is illustrated in ***Figure 3.6***:

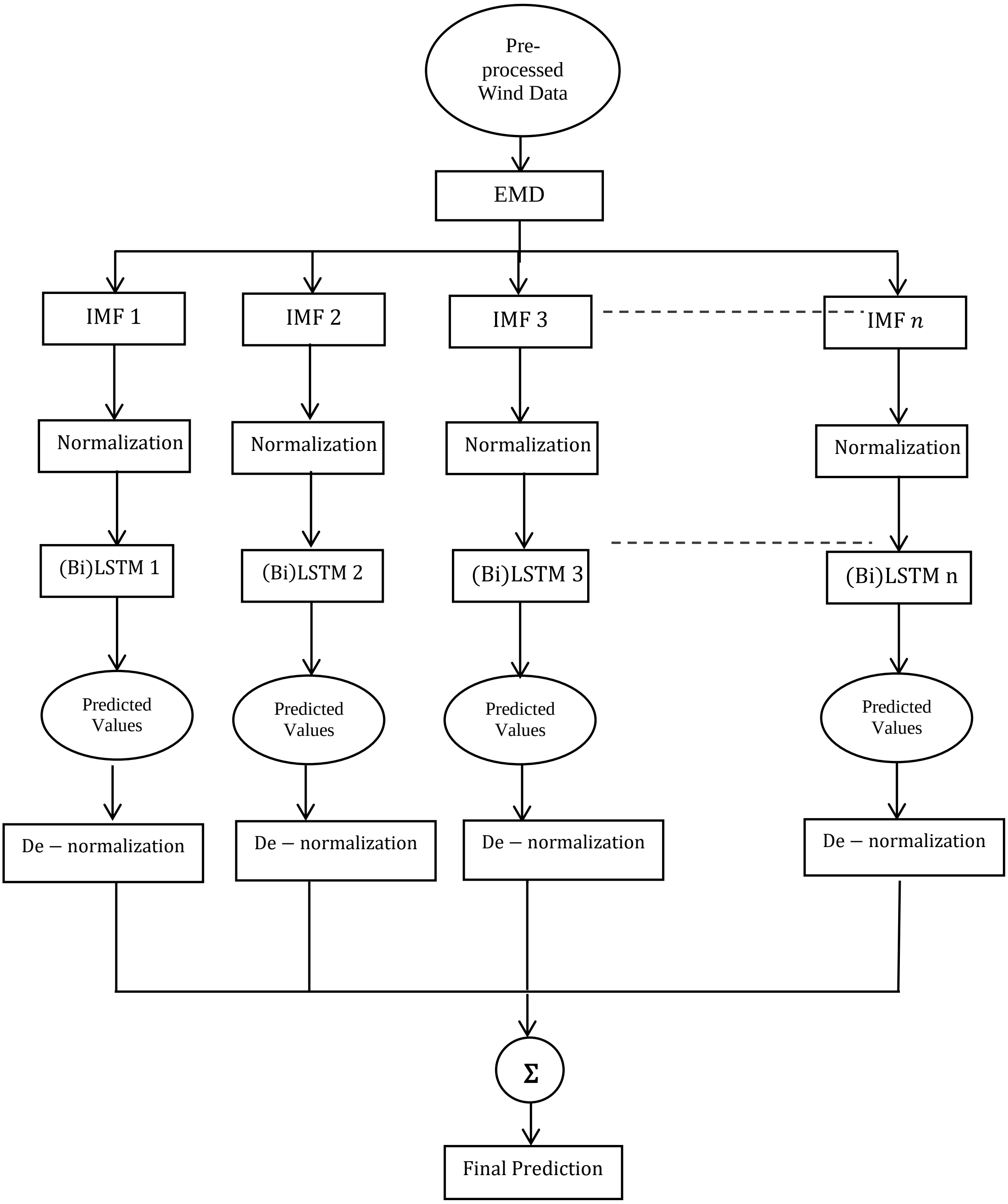


**Figure 3.6: LSTM/BiLSTM_EMD Forecast Procedure**

### 3.4 Modelling of the Wind Power Curve from Historical Data

A wind power curve is a graphical illustration of the relationship between wind speed and wind power. The actual wind power curve for a wind turbine or a wind farm is usually not an exact match to the wind power curve from the manufacturer. The manufacturer wind power curve is derived from turbines operated in ideal conditions and with wind turbine blade angles positioned to extract as much energy as possible from the wind. For this research, the wind power curve was reconstructed from the historical training data (i.e., the data section used for training). First, the wind speed vs wind power scatter plot was drawn in Excel. The wind turbine power curve presents itself in a graph mimicking the Avrami equation. The Avrami equation defines the kinetics behind crystallization and is also applied to define other applications involving changes such as chemical reactions (Finke & Ozkar, 2017). The best line of fit through the data was determined, and the Avrami equation coefficients were incorporated into the MATLAB prediction algorithm.

For each test, the dataset was split into 70% training and 30% testing data. Based on this split, the AL_WIND_07_12 wind dataset uses approximately four years of data for training and data from the final two years is used for testing. The partition of the data into train and test sets is illustrated in ***Table 3.5*** below:

**Table 3.5: Partition of the Dataset into Train and Test Data**

| | | |
|---|---|---|
| **AL_WIND_07_12** | Train Data | $0.7 * 52560 = \mathbf{36792}$ |
| **Dataset** | Test Data | $0.3 * 52560 = \mathbf{15768}$ |

The Avrami equation is defined as:

$$y = 1 - e^{-kt^n} \qquad (3.1)$$

Where: y is the fraction of completed transformation at time t

k is a rate constant
t is time
n is the growth dimensionality.

The equation, as defined in Equation 3.1  above, is constrained between 0 and 1. Since our wind data exceeds these boundaries, we modify the Avrami equation to encompass the scope of wind power. The modified avrami equation is shown below:

$$y = A\ (1 - e^{-kt^{n}}) \qquad (3.2)$$

Where $A, k, t$ are fitting constants and $t\ =\ wind\ speed$

The fitting constants are determined by minimizing the sum square residual where:

$$residual\ =\ actual\ power\ -\ avrami\ power\ estimate$$

(3.3)

The fitting of the Avrami equation is done in Excel using the ***Solver*** tool**,** and the obtained Avrami equation is coded into the forecasting algorithm in MATLAB. For the AL_WIND_07_12 wind dataset, the Avrami equation for minimized Sum of Squared Residuals (SSR) was obtained as follows:

$$\boldsymbol{y = 59711.44\ \ (1 - e^{-k(8.06\ x10^{-05})\cdot\ t^{\,4.203855}})} \qquad (3.4)$$

The wind power curve, as determined by the Avrami equation, is used to provide a wind power estimate for wind power at time $t$. This new input provides a variable that improves the BiLSTM prediction accuracy for wind power at time $t$ and with a look-back period of $n$ time steps.

### 3.5 Models and Model Parameters

Based on repeated experiments, the hyperparameters for the LSTM and BiLSTM were set as follows:

Number of neurons in the hidden layer = 150

Iteration epochs = 10

Look back period/Sequence length = 6

The **Adam** optimizer was used since it helps accelerate the learning speed of LSTM while reducing the optimization resources required. It combines the benefits of two gradient descent algorithms: AdaGrad and RMSProp (Y. Liu et al., 2019).

Adam helps in adjusting individual parameters (weights) during training. It dynamically adapts the learning rates based on the previous gradients and the moving averages of the parameters as the network passes through the training data. This adaptability makes Adam makes it well-suited for optimizing complex and high-dimensional models, leading to faster convergence and improved training performance.

Adam follows the following steps during the training of the LSTMs and BiLSTMs (Y. Liu et al., 2019):

i. Initialization: Adam initializes variables to zero. These variables are meant to keep track of the moving averages of past gradients.
ii. Computing Gradients: At each training iteration, gradients are computed for the model's weights based on the present batch of training data.
iii. Updating of Moving Averages: Adam updates the moving averages of past gradients and squared gradients using exponential decay. Doing this helps the optimizer remember information about past gradients while at the same time giving more weight to recent gradients.
iv. Bias Correction: Adam performs bias correction to adjust the estimates to avoid biased estimates in the initial training steps when moving averages are close to zero.
v. Updating of Weights: The optimizer calculates the weight update using the moving averages and the current gradients. It uses the moving averages to adapt the learning rates for each parameter, ensuring that parameters with different scales receive appropriate updates.

In conclusion, the data preprocessing steps to be followed are discussed in this chapter, with the LSTM and BiLSTM identified as choices of Recurrent Neural Networks to be used for the prediction. The forecast accuracy of these two networks is then enhanced using data decomposition using EMD and then further improved by adding mathematical modelling of the wind power curve based on the Avrami Equation. The following models were tested on the AL_WIND_07_12 wind power dataset, with their results presented and discussed in ***Chapter 4***.

i. LSTM_EMD

ii. BiLSTM

iii. BiLSTM_EMD

iv. BiLSTM_EMD_with Avrami Power Curve

# CHAPTER FOUR

# RESULTS, ANALYSIS AND DISCUSSION

## 4.1 Forecasting using the Direct and Indirect Approach

Two wind power forecasting approaches are discussed in the literature review: Indirect and direct. The **indirect process** entails forecasting future wind **speed** values, after which appropriate transformations are applied to get the corresponding values of wind power. On the other hand, in the **direct method**, wind power is forecast directly without predicting wind speed first.

This section presents wind power forecast results using Bidirectional Long Short Term Neural Networks (BiLSTMs) and Empirical Mode Decomposition (EMD) techniques. After that, the Avrami wind power curve modelled from the historical wind speed vs wind power datasets is used to improve the forecasts. To validate the effectiveness of the approaches, a comparison is made for the direct vs indirect wind power forecasting approaches for the same models and the results discussed. The indirect approach that used BiLTSM + EMD + Avrami Curve was discovered to have the best accuracy compared to other models for the same dataset and forecast horizons.

### 4.1.1 Test on Effectiveness of Data Decomposition and Illustration of the Deficiency of LSTMs

Simulations in this section were carried out on a section of the dataset. The forecasting algorithm used was the LSTM, and it was later hybridized using EMD and a wind power curve. The LSTM was used alone to predict 168 look-ahead instances (equivalent to 1 week for the AL_WIND_07_12 dataset with hourly observations). ***Figure 4.1*** shows the plot of the predicted vs actual wind power forecast using LSTM alone and with wind power as the input parameter.

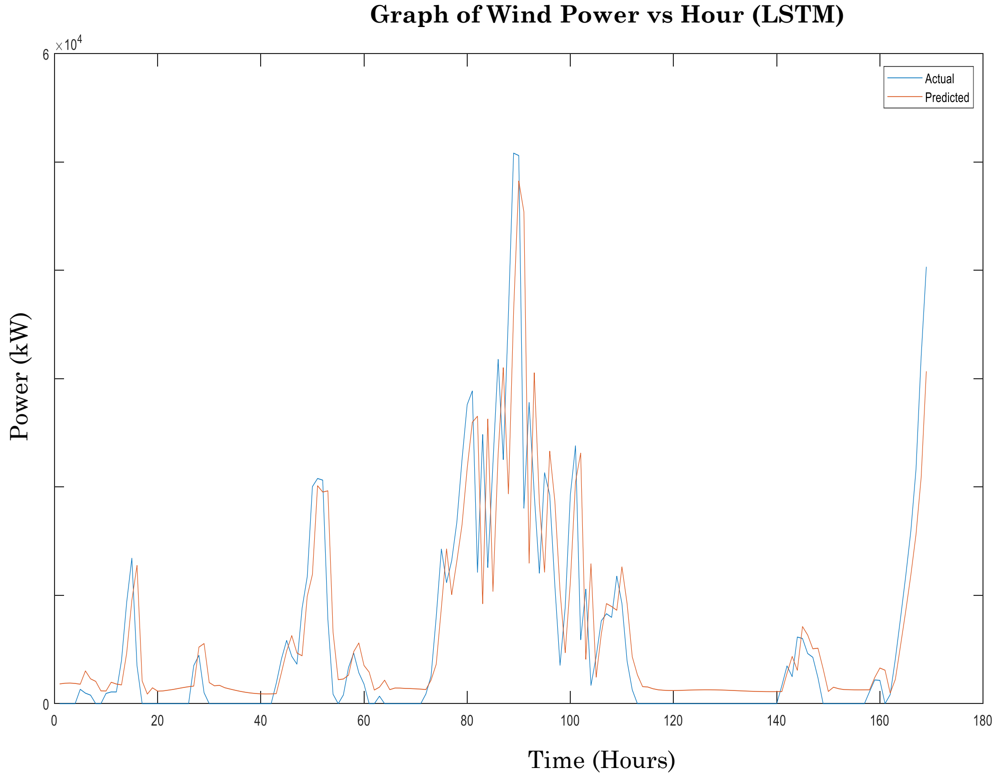


**Figure 4.1: Graph of Wind Power vs Time for LSTM Alone (168 Look Ahead Instances)**

*Observation:* The predicted Wind Power shows good accuracy relative to the actual values. However, a **time lag** is noted in the LSTM prediction compared to the actual wind power data. This is caused by the **high variability** of the wind power data and can be resolved by breaking down the wind power time-series data into simpler forms using data decomposition techniques. ***Figure 4.2*** shows the improvement in the forecast when data decomposition is used to break down the data before the prediction.

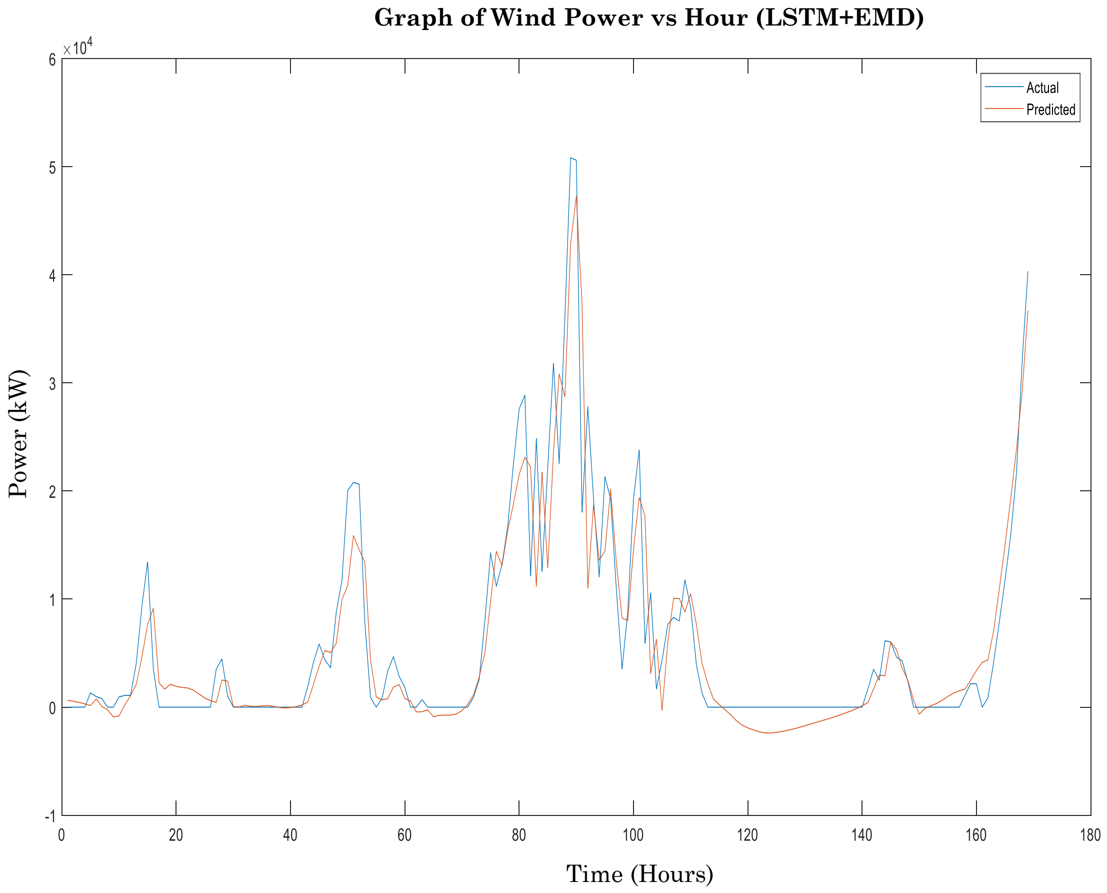


**Figure 4.2: Graph of Wind Power vs Time for LSTM+EMD (168 Look Ahead Instances)**

*Observation:* The time lag phenomenon visible in the LSTM prediction graphs in *Fig 4.1* is now suppressed by EMD, and the prediction is now better. A deviation in the prediction below the zero mark is noted in some instances, e.g. at hour 70 and hour 130, and this results from the fact that in the decomposed data, some of the components are negative. It is however observed that the prediction from the LSTM is suffering from overshoots and undershoots. This is attributed to the fact that the Wind Power dataset is highly dynamic and therefore the LSTM network is not able to adapt quickly to the data changes. To correct the undershoot, a simple curve correction definition was used to limit the prediction from swinging below the zero mark, and the final updated prediction is as shown in ***Fig 4.3***:

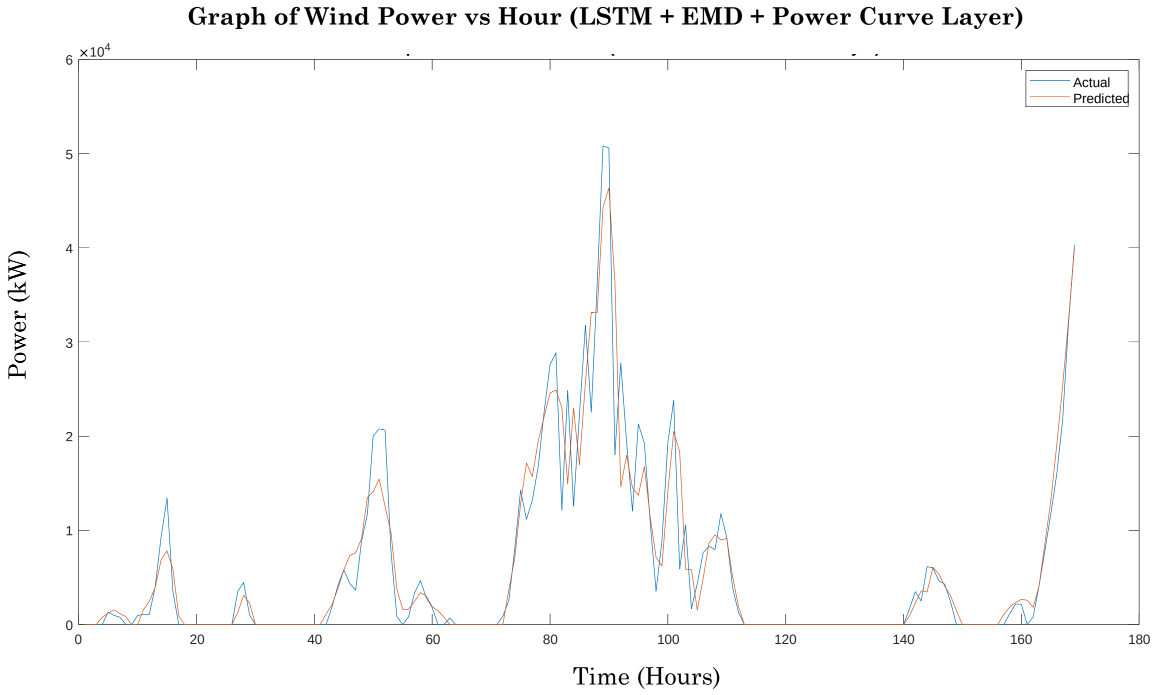


**Figure 4.3: Graph of Wind Power vs Time for LSTM+EMD with Power Curve Layer (168 look ahead instances).**

The three approaches (LSTM, LSTM + EMD and LSTM + EMD + Curve correction) were then compared, and a graph of a prediction of 100 future instances was plotted, as shown in ***Fig 4.4*** below.

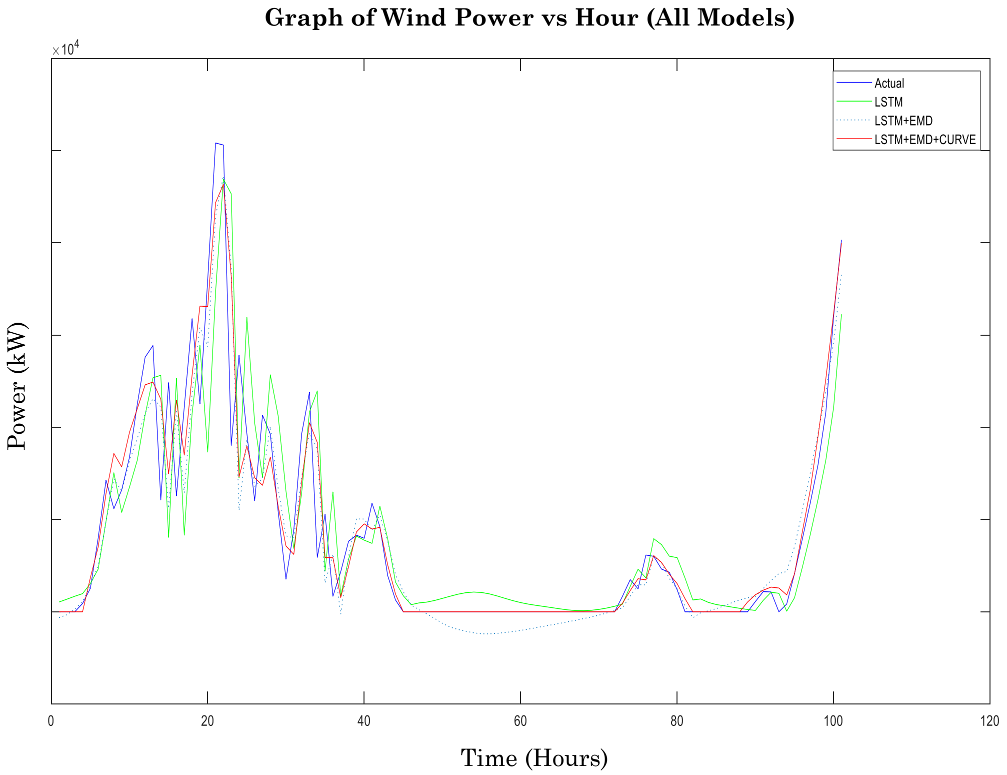


**Figure 4.4: Graph of Wind Power vs Time for Models Comparison.**

*100 observations sampled for plotting from the 876 predictions made.

Observation: The LSTM + EMD with Power Curve model gives the **best** prediction compared to the other two models since it follows the actual wind power better than the other two models. LSTM +EMD + Power curve gives a **22.858% improved** forecast accuracy compared to LSTM alone and a **13.895%** better forecast compared to LSTM +EMD. The impeovements in the forecast expecially at the peaks are noted when EMD is included in the forecast model. ***Table 4.1*** summarizes the RMSE values for the forecast.

**Table 4.1: RMSE Values Comparison for the Models.**

| S/No | Model | RMSE (kW) |
|---|---|---|
| 1. | LSTM | 6184 |
| 2. | LSTM+EMD | 5540 |
| 3. | LSTM+EMD+Curve Correction | 4770 |

The importance of this section was to show the limitation of LSTMs and how that limitation was overcome. In the next section, the BiLSTM was used as the forecasting network, EMD as the data decomposition technique and the Avrami approach as the method of obtaining the wind power curve from the historical training wind data. The BiLSTM model offers better prediction accuracy than LSTM since it traverses the data backwards and forward during training [34], giving it a better understanding of the relationship between various variables to determine more precise patterns and connections in the dataset. Therefore, this research now focuses on the BiLSTM networks from this point onwards.

## 4.2 Wind Power Forecasting using BiLSTM

### 4.2.1 BiLSTM Alone Prediction Results

#### 4.2.1.1 Prediction on AL_Wind_07_12 Dataset

Dataset: AL_Wind_07_12 Dataset

Predicted Variable: Wind Power (t)

Input Variables: Air Temperature, Pressure, Wind Speed & Wind Direction

Hidden layer neurons = 150 (Determined through trial and error)

Look_back/Sequence length = 6

Average simulation time: = 390 seconds

The results obtained from this forecast are as shown next:

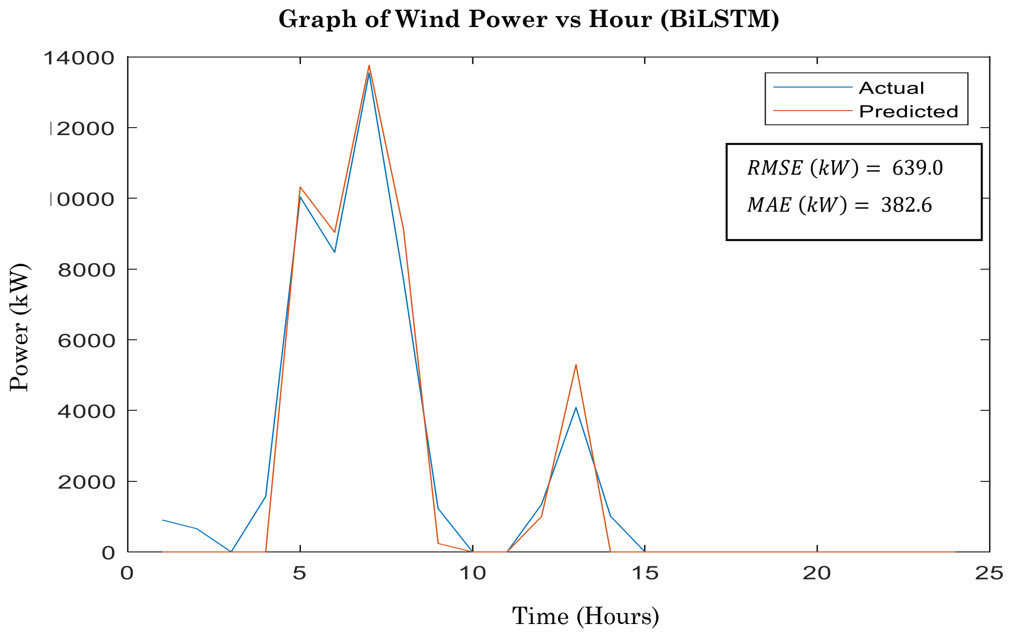


**Figure 4.5: 24 - Hours Ahead Wind Power Forecast**

*Observation*: For the 24 hour ahead forecast using BiLSTM, the RMSE and MAE values obtained when the actual and predicted data are compared is 639 kW and 383.6 kW respectively. The actual and predicted datasets have a correlation of 0.9898 indicating that the prediction model was able to strongly capture the underlying patterns in the dataset.

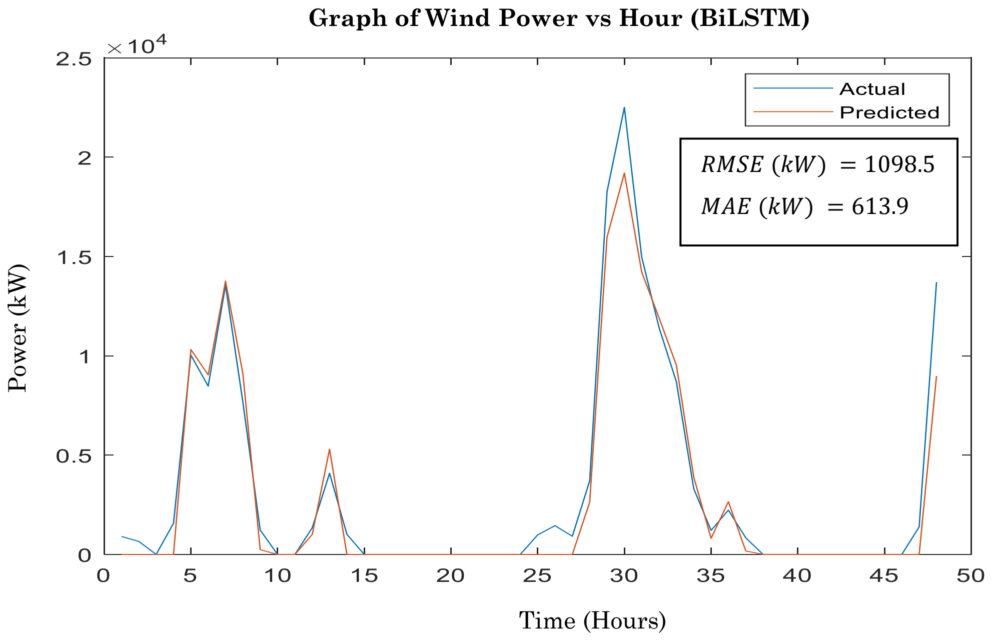


**Figure 4.6: 48 Hours Ahead Wind Power Forecast**

*Observation*: For the 48-hour ahead forecast using BiLSTM, the RMSE and MAE values obtained when the actual and predicted data are compared are 1098.5 kW and 613.9 kW, respectively. Compared to the 24-hour forecast, the prediction accuracy is observed to deteriorate.

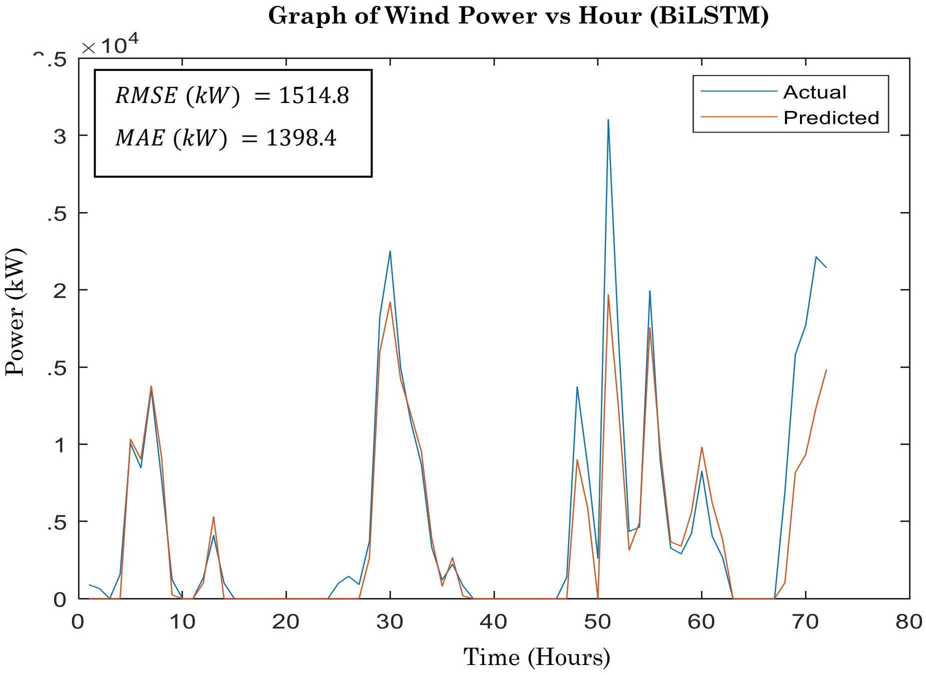


**Figure 4.7: 72 Hours Ahead Wind Power Forecast**

*Observation*: For the 72-hour ahead forecast using BiLSTM, the RMSE and MAE values obtained when the actual and predicted data are compared are 1514.8 kW and 1398.4 kW, respectively. Compared to the 48-hour forecast, the prediction accuracy is noted to deteriorate further.

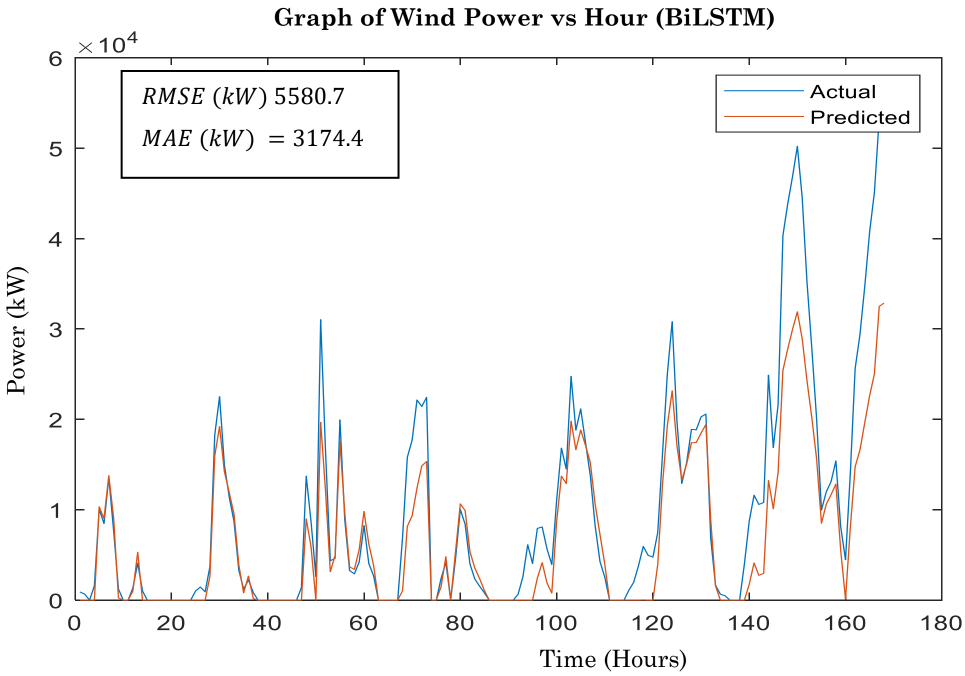


**Figure 4.8: 168 Hours Ahead Wind Power Forecast**

*Observation*: For the 168-hour ahead forecast using BiLSTM, the RMSE and MAE values obtained when the actual and predicted data are compared are 5580.7 kW and 3174.4 kW, respectively. The forecast accuracy is seen to deteriorate even further compared to the 24, 48 and 72-hour forecasts. ***Table 4.2*** below compares the prediction results obtained from BiLSTM.

**Table 4. 2: Comparison of Prediction Results of BiLSTM Alone**

| **Look Ahead (hrs)** | **RMSE (kW)** | **MAE (kW)** | $\boldsymbol{R^2}$ |
|---|---|---|---|
| 24 | 639.0 | 382.6 | 0.9898 |
| 48 | 1098.5 | 613.9 | 0.9827 |
| 72 | 1514.8 | 1398.4 | 0.9478 |
| 168 | 5580.7 | 3174.4 | 0.9568 |

*Observation:* The BiLSTM network has its best prediction accuracy during the 24-hour look ahead test. According to literature, as the forecast horizon increases, the forecast accuracy diminishes. This shows that the BiLSTM model is well-suited for short-term forecasts.

### 4.2.2 BiLSTM_EMD

Dataset: AL_Wind_07_12 Dataset

Predicted Variable: Wind Power (t)

Input Variables: Wind Power (t-1, t-2, ….)

Hidden layer neurons = 150 (Determined through trial and error)

Average simulation time: = 10 minutes

The results obtained from this forecast are as shown below:

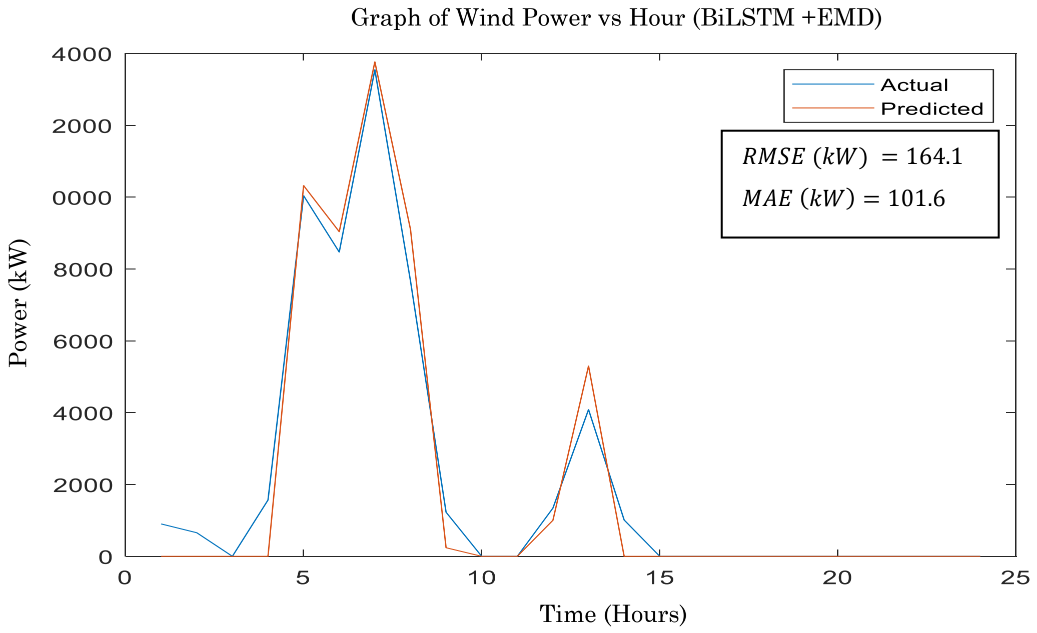


**Figure 4.9: 24 Hours Ahead Wind Power Forecast**

*Observation*: For the 24-hour ahead forecast using BiLSTM + EMD, the RMSE and MAE values obtained when the actual and predicted data are compared are 164.1 kW and 101.6 kW, respectively. This forecast is comparatively better than the forecast using BiLSTM alone.

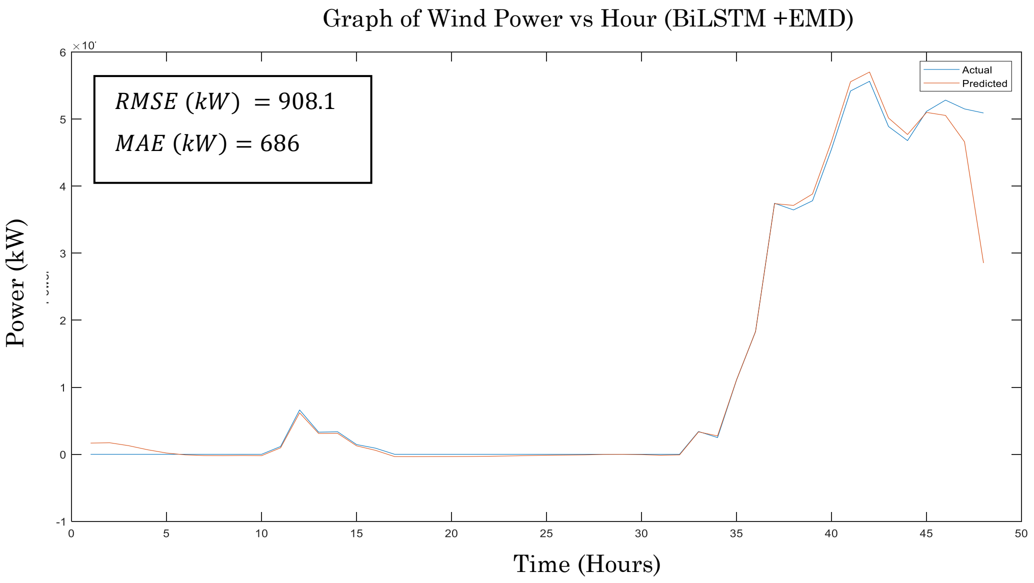


**Figure 4.10: 48 Hours Ahead Wind Power Forecast**

*Observation*: For the 48-hour ahead forecast using BiLSTM + EMD, the RMSE and MAE values obtained when the actual and predicted data are compared are 908.1 kW and 686 kW, respectively. The forecast is noted to deteriorate compared to the 24-hour forecast.

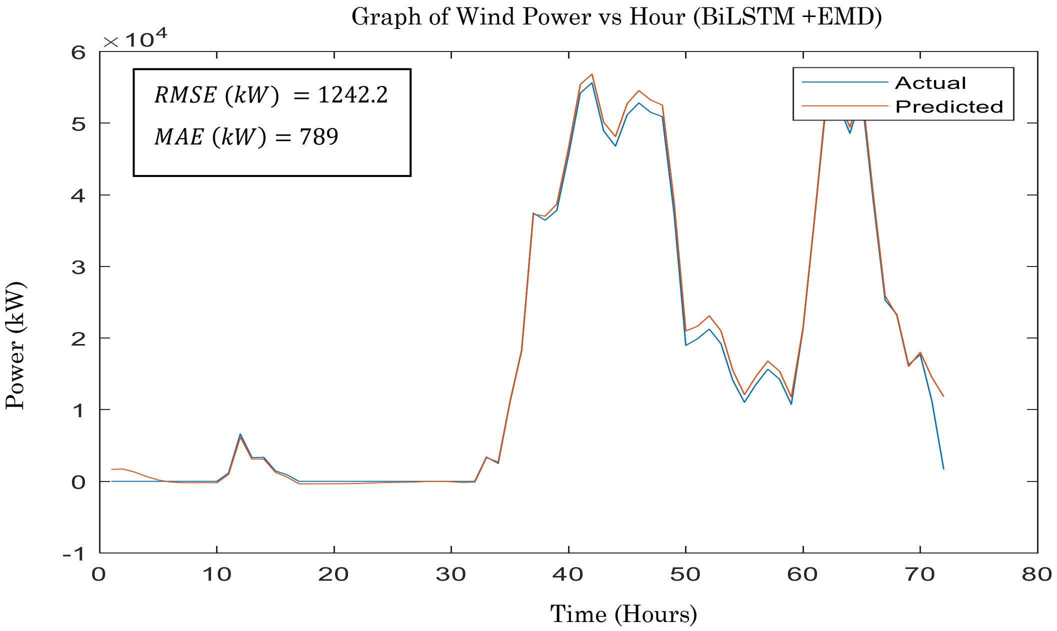


**Figure 4.11: 72 Hours Ahead Wind Power Forecast**

*Observation*: For the 72-hour ahead forecast using BiLSTM + EMD in ***Figure 4.11*** above, the RMSE and MAE values obtained when the actual and predicted data are compared are 1242.2 kW and 789 kW, respectively. The forecast is noted to deteriorate further in comparison to the 48-hour forecast.

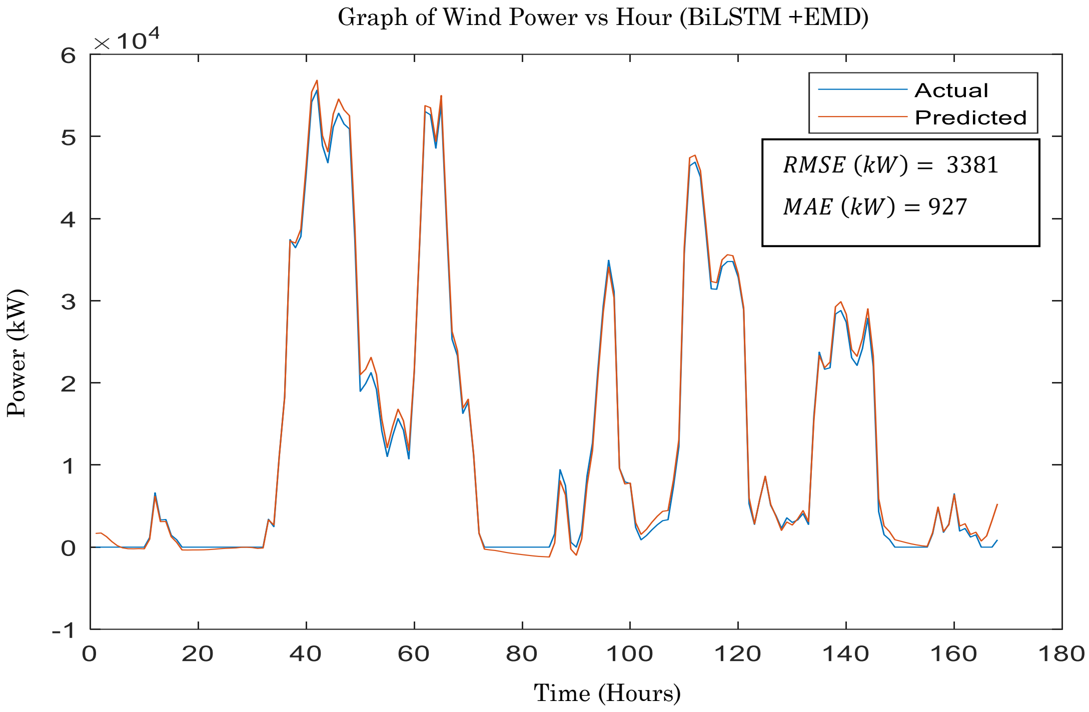


**Figure 4.12: 168 Hours Ahead Wind Power Forecast**

*Observation*: For the 168-hour ahead forecast using BiLSTM + EMD, the RMSE and MAE values are noted to deteriorate further and settle at 3381 kW and 927 kW, respectively. The results obtained from the BiLSTM_EMD model are as shown in ***Table 4.3*** below:

**Table 4.3: Comparison of Prediction Results of BiLSTM_EMD**

| Look Ahead (hrs) | RMSE (kW) | MAE (kW) | $R^2$ |
|---|---|---|---|
| 24 | 164.1 | 101.6 | 0.9989 |
| 48 | 908.1 | 686 | 0.9977 |
| 72 | 1242.2 | 789 | 0.9866 |
| 168 | 3381 | 927 | 0.9868 |

*Observation:* The BiLSTM_EMD model has its best prediction accuracy during the 24-hour look ahead test. The forecast accuracy deteriorates as the forecast period increases. This is an expected phenomenon from literature where the forecast accuracy diminishes as the forecast horizon increases. EMD is seen to improve the accuracy of the BiLSTM model working alone.

### 4.2.3 BiLSTM with Avrami Curve

Dataset: AL_Wind_07_12 Dataset

Predicted Variable: Wind Power (t)

Input Variable: Wind Power (t-1, t-2, ….)

Hidden layers = 150 (Determined through trial and error)

Look Back/ Sequence Length = 6

The first step in implementing the BiLSTM with the Avrami model was determining the parameters for the Avrami equation that give the best line of fit on the training data drawn from the wind power dataset. A scatter plot of the training data before curve fitting is shown in ***Figure 4.13*** below:

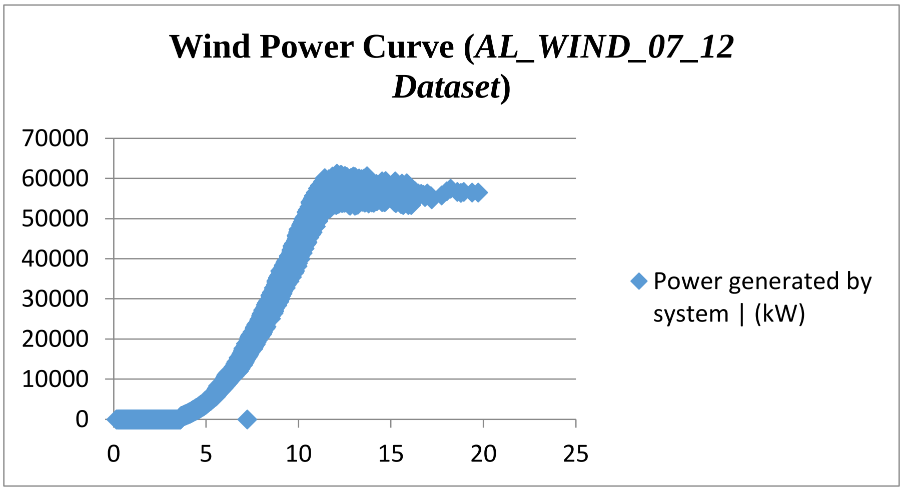


**Figure 4.13: Scatter Plot of the Wind Power vs Wind Speed Data before curve fitting (AL_WIND_07_12 Dataset)**

After the optimal Avrami equation parameters are determined using Excel's solver, a curve of best fit is plotted as shown in ***Figure 4.14***:

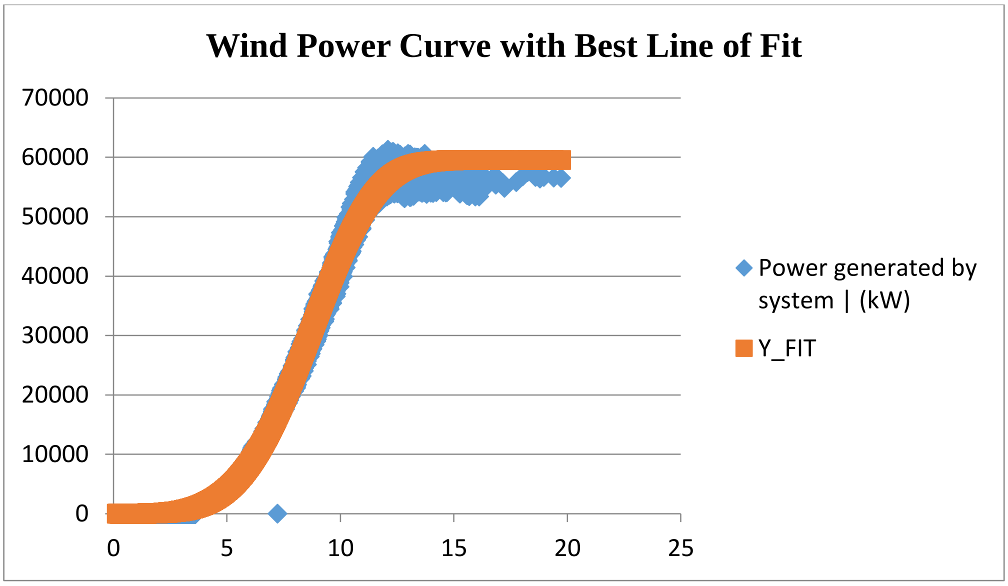


**Figure 4.14: Scatter Plot of the Wind Power vs Wind Speed Data after curve fitting with Avrami (*AL_WIND_07_12 Dataset)***

The obtained Avrami Equation parameters that best fit the historical wind power, as shown in Equation (4.1) are then incorporated into the forecast methodology:

$$\boldsymbol{y} = \boldsymbol{A}\left(1 - e^{-kt^{n}}\right) = \mathbf{59711.44}\ \left(\mathbf{1} - \boldsymbol{e}^{-\boldsymbol{k}\left(\mathbf{8.06}\ \boldsymbol{x}\mathbf{10}^{-\mathbf{05}}\right)\cdot\ \boldsymbol{t}^{\mathbf{4.203855}}}\right) \qquad (4.1)$$

The results obtained from the BiLSTM EMD + Avrami curve forecast are shown in ***Figure 4.15 - Figure 4.18*** for look-ahead forecasts of 24 hours to 168 hours:

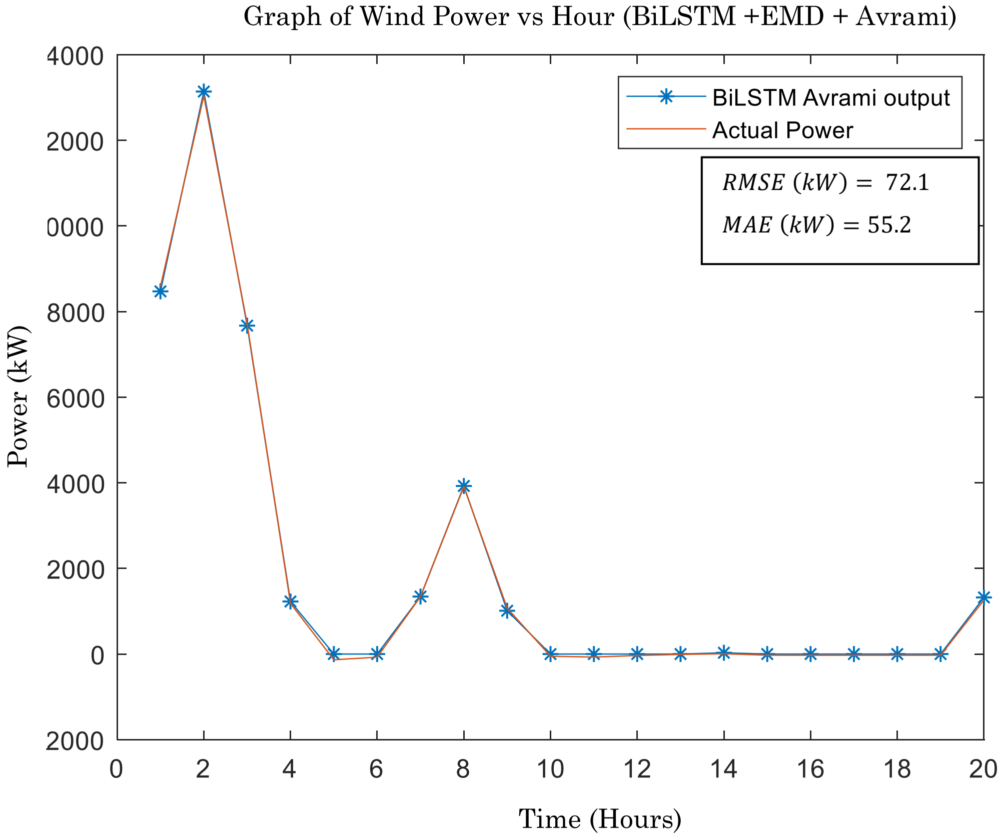


**Figure 4. 15: 24 - Hours Ahead Wind Power Forecast**

*Observation*: For the 24-hour ahead forecast using BiLSTM + EMD +Avrami Curve, the RMSE and MAE values obtained are 72.1 kW and 55.2 kW, respectively. This forecast is the best when compared to the BiLSTM and the BiLSTM + EMD models for the same forecast period of 24 hours.

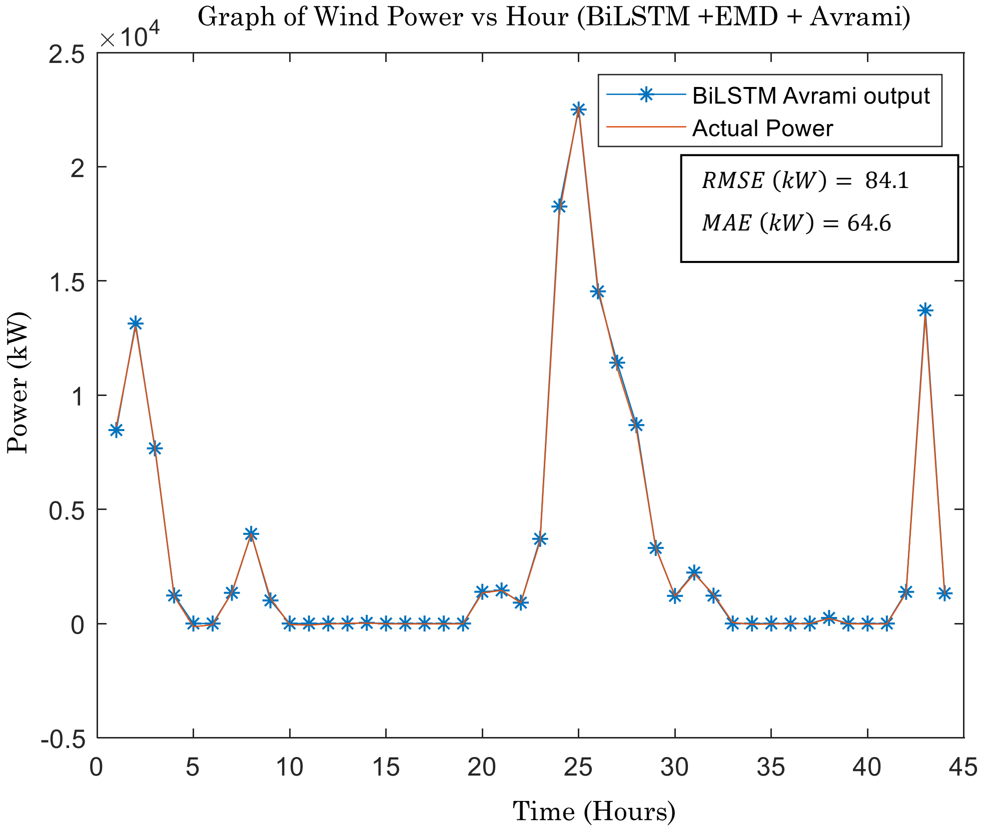


**Figure 4.16: 48 Hours Ahead Wind Power Forecast**

*Observation*: For the 48-hour ahead forecast, the RMSE and MAE values obtained are 84.1 kW and 64.6 kW, respectively. This is a slight deterioration compared to the 24-hour forecast, indicating that the BiLSTM + EMD with Avrami curve would have superior accuracy for the 48-hour forecast.

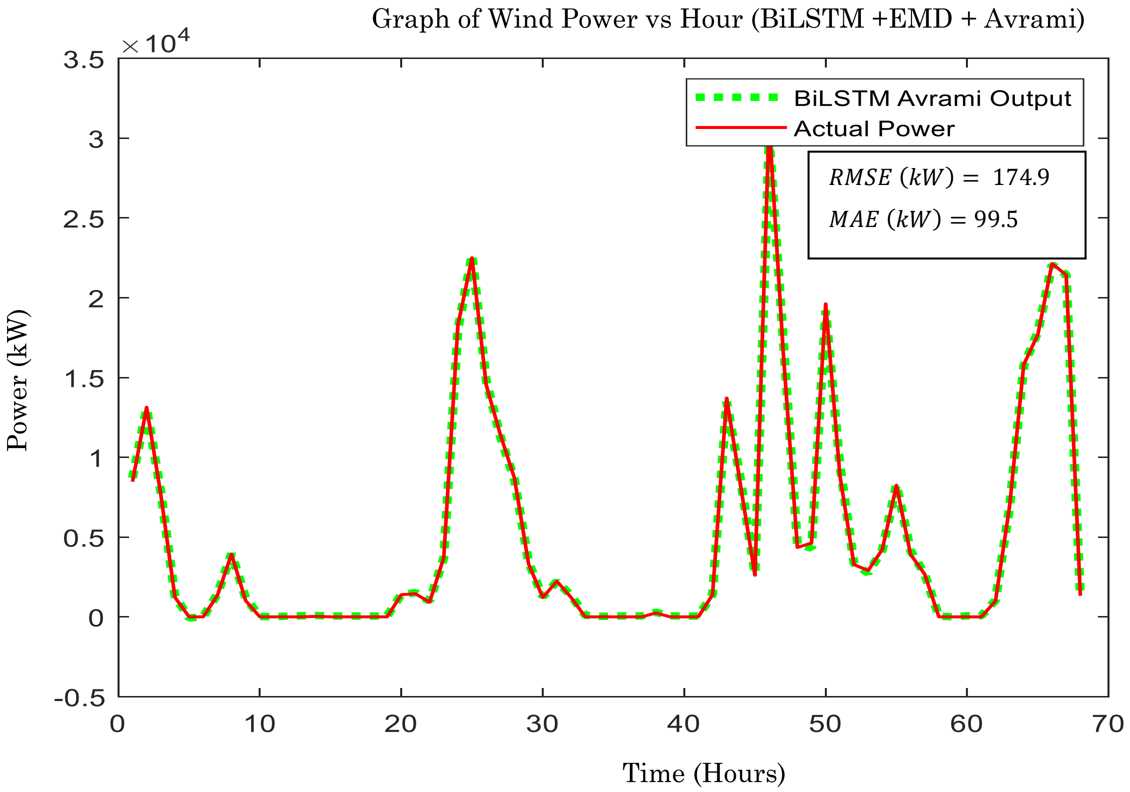


**Figure 4.17: 72 Hours Ahead Wind Power Forecast**

*Observation*: For the 72-hour ahead forecast, the RMSE and MAE values obtained are 174.9 kW and 99.5 kW, respectively. A deterioration in the forecast accuracy is noted when the results are compared to the 48-hour forecast.

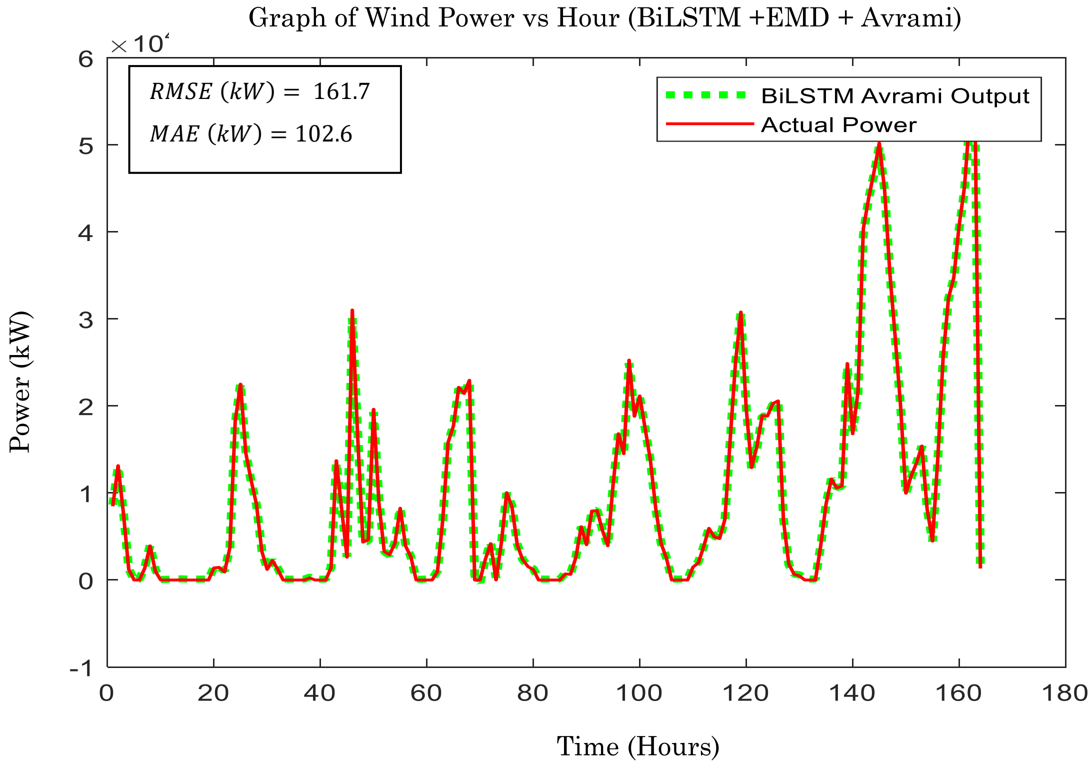


**Figure 4.18: 168 Hours Ahead Wind Power Forecast**

*Observation*: The 168-hour look ahead forecast has almost similar accuracy to the 72-hour forecast. Compared to the 24-hour and the 48-hour forecasts, deterioration in the forecast accuracy is noted.

**Table 4.4: Prediction Results of BiLSTM_EMD with Avrami Curve**

| Look Ahead (hrs) | RMSE (kW) | MAE (kW) | $R^2$ |
|---|---|---|---|
| 24 | 72.1 | 55.2 | 0.9999 |
| 48 | 84.1 | 64.6 | 0.9999 |
| 72 | 174.9 | 99.5 | 0.9997 |
| 168 | 161.7 | 102.6 | 0.9999 |

*Observation:* The BiLSTM_Avrami model has its best prediction accuracy during the 24-hour look ahead test. The forecast accuracy deteriorates as the forecast period increases. The inclusion of the concept of the Avrami Equation is seen to improve the accuracy of the BiLSTM model in predicting future wind power.

## 4.3.Summary of Results

The best RMSE and MAE values were observed when using the BiLSTM + EMD + Avrami model, as seen in ***Table 4.5*** below:

**Table 4.5: Comparison of Prediction Results for all BiLSTM Models**

| | BiLSTM | | BiLSTM + EMD | | BiLSTM + EMD + Avrami | |
|---|---|---|---|---|---|---|
| **Look Ahead (hrs)** | **RMSE (kW)** | **MAE (kW)** | **RMSE (kW)** | **MAE (kW)** | **RMSE (kW)** | **MAE (kW)** |
| 24 | 639.0 | 382.6 | 164.1 | 101.6 | 72.1 | 55.2 |
| 48 | 1098.5 | 613.9 | 908.1 | 686 | 84.1 | 64.6 |
| 72 | 1514.8 | 1398.4 | 1242.2 | 789 | 174.9 | 99.5 |
| 168 | 5580.7 | 3174.4 | 3381 | 927 | 161.7 | 102.6 |

The RMSE value of the BiLSTM + Avrami Power curve is 72.1kW, making it the lowest among the hybrid approaches for the 24-hour forecast. Compared to the BiLSTM + EMD hybrid approach, the percentage improvement in the RMSE of the forecasts resulting from inclusion of the Avrami curve is calculated as:

$$\frac{164.1\ kW - 72.1\ kW}{164.1\ kW}\ x\ 100 = \mathbf{56.1}\%$$

This shows the superiority of modelling the wind power curve using the Avrami equation and the potential to improve neural network-based forecasts using mathematical modelling. The improvements that data decomposition brings to wind power forecasting are seen when the predictions from LSTM vs LSTM+EMD and BiLSTM vs BiLSTM + EMD approaches, respectively. The latter two models with EMD have lower RMSE values than the base models. Neural networks perform better when the time series data is broken down/decomposed into its constituent frequency components. The data becomes easier to understand and to extract meaningful information from a time series. Therefore, the developed indirect approach to wind power forecasting based on BiLSTM+EMD+Avrami Power Curve is an effective approach to wind power prediction.

# CHAPTER FIVE

# CONCLUSION AND RECOMMENDATIONS

## 5.1 Conclusion

This work successfully developed an approach to wind power forecasting based on BiLSTM and EMD. Data preprocessing was done to ensure no missing data values, after which feature scaling was done on the data using normalization. The preprocessed data was then used in various models based on BiLSTM alone, BiLSTM and with EMD and BiLSTM_EMD with the Avrami Power Curve. The output of all these models was the predicted wind power. In the adopted forecasting model, past wind power forecasts $(t-1, t-2, ..)$ were used to predict wind power at time $t$. The Avrami power curve was used to approximate wind power at time $t$ given wind speed to ensure that the BiLSTM neural network could converge faster and offer more accurate results. The wind power curve was modelled from the historical train data in the dataset to ensure as much factual information was captured as compared to using the manufacturer's wind power curve. The modelled wind power curve considers the wind farm operation dynamics that could have affected the wind power output, such as the prevailing weather conditions at that time and other operational constraints.

The observed value error values using the BiLSTM + EMD + Avrami model was an RMSE value of 72.1kW for a 24-hour ahead forecast. This was a more accurate forecast than the BiLSTM+EMD model and BiLSTM alone. The use of EMD does, however, come at a small cost. When the data is decomposed into its various mode components, each input mode component data and its corresponding output data is forecast by its own BiLSTM network, increasing the computation times. However, a longer execution time is worth it since the forecast model's accuracy is improved.

With improved forecasts, integrating large-scale wind power in present-day grids becomes easier and more economical since improved accuracy of wind forecasts means less balancing power requirements.

## 5.2 Recommendations

This research only focused on one type of data decomposition technique - EMD. The thesis was primarily focused on developing an approach to wind power forecasting that uses the Avrami equation to model the wind power curve of a wind farm and help improve forecasts. The essence of data decomposition in time series prediction cannot be overlooked, and future research can focus on exploring alternative data decomposition methods to improve prediction along with the Avrami wind power curve. The forecasting approach adopted in this research can also be implemented on a different development platform, such as Python and comparing computational resources against MATLAB.

## REFERENCES

Ahmed, S. D., & Al-ismail, F. S. M. (2020). *Grid Integration Challenges of Wind Energy : A Review*. *8*(type 1), 10857–10878.

Aliprantis, D. C., & Lafayette, W. (2014). *Fundamentals of Wind Energy Conversion for Electrical Engineers*.

Bantupalli, M. K. (2017). *decomposition with ANN and ARIMA models*.

Beainy, A., Maatouk, C., Moubayed, N., & Kaddah, F. (2016). Comparison of different types of generator for wind energy conversion system topologies. *2016 3rd International Conference on Renewable Energies for Developing Countries (REDEC)*, 1–6. https://doi.org/10.1109/REDEC.2016.7577535

Berrezzek, F., Khelil, K., & Bouadjila, T. (2019). *Efficient Wind Speed Forecasting Using Discrete Wavelet Transform and Artificial Neural Networks*. *33*(6), 447–452.

Blaabjerg, F., & Ma Ke. (2017). *Wind Energy Systems*. *105*(11), 2116–2131. https://ieeexplore.ieee.org/abstract/document/7927779

Bokde, N., Feijoo, A., Villanueva, D., & Kulat, K. (2018). A Novel and Alternative Approach for Direct and Indirect Wind-Power Prediction Methods _ Enhanced Reader.pdf. *Energies*, *11*.

Chen, Q., Town, C., Town, C., & Town, C. (2018). Wind Power Forecasting Wind Power Forecasting. *IFAC-PapersOnLine*, *51*(28), 414–419. https://doi.org/10.1016/ j.ifacol.2018.11.738

Chen, Z., & Member, S. (2005). *Issues of Connecting Wind Farms into Power Systems*. 1–6.

Chourasia, A. (2020). *Decomposition in Time Series Data*. https://medium.com/ analytics-vidhya/decomposition-in-time-series-data-b20764946d63

Dolatabadi, A., Abdeltawab, H., & Abdel-Rady, Y. (2020). *Hybrid Deep Learning-Based Model for Wind Speed Forecasting Based on DWPT and Bidirectional LSTM Network*. 229219–229232. https://doi.org/10.1109/ACCESS. 2020.3047077

E, K., & Diamond. (2011). Global Warming's Impact on Wind Speeds: Long-Term Risks for Wind Farms May Impact Guarantees and Wind Derivatives Tied to Wind Energy Production. *40th Annual Conference on Environmental Law*.

EPA. (2019). *The Sources and Solutions: Fossil Fuels*. https://www.epa.gov/ nutrientpollution/sources-and-solutions-fossil-fuels

Finke, R. G., & Ozkar, S. (2017). Silver Nanoparticles Synthesized by Microwave Heating: A Kinetic and Mechanistic Re-Analysis and Re-Interpretation. *The Journal of Physical Chemistry*, *111*, 121. https://doi.org/10.1021/acs.jpcc. 7b06323

Gholamy, Afshin Kreinovich, Vladik Kosheleva, O. (2018). *Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation*.

Global Wind Energy Council. (2022). *GWEC | GLOBAL WIND REPORT 2022*.

Hafner, Manfred Luciani, G. (2022). The Integration of Non-Dispatchable Renewables. In *The Palgrave Handbook of International Energy Economics* (pp. 270–296).

Hanifi, S., Liu, X., Lin, Z., & Lotfian, S. (2020). A Critical Review of Wind Power Forecasting Methods—Past, Present and Future. *Energies*, *13*(15), 3764. https://doi.org/10.3390/en13153764

Ivanovic, M., & Kurbalija, V. (2016). Time series analysis and possible applications. *2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 473–479. https://doi.org/ 10.1109/MIPRO.2016.7522190

Jarke, J., & Perino, G. (2017). Do Renewable Energy Policies Reduce Carbon Emissions? On Caps and Inter-Industry Leakage. *Journal of Environmental Economics and Management*. https://doi.org/10.1016/j.jeem.2017.01.004

Jaseena, K. U., & Kovoor, B. C. (2021). Decomposition-based hybrid wind speed forecasting model using deep bidirectional LSTM networks. *Energy Conversion and Management*, *234*(November 2020), 113944. https://doi.org/10.1016/ j.enconman.2021.113944

Johari, M. K., Jalil, M. A. A., Faizal, M., & Shariff, M. (2018). *Comparison of horizontal axis wind turbine ( HAWT ) and vertical axis wind turbine ( VAWT ). October*. https://doi.org/10.14419/ijet.v7i4.13.21333

Kalair Anam, Abas Naeem, Shoaib Saleem Muhammad, Raza Kalair Ali, K. N. (2020). Role of Energy Storage Systems in Energy Transition from Fossil Fuels to Renewables. *Energy Storage*, *3*(1). https://doi.org/10.1002/est2.135

Kassa, Y., Zhang, J. H., Zheng, D. H., & Dan Wei. (2016). Short term wind power prediction using ANFIS. *2016 IEEE International Conference on Power and Renewable Energy (ICPRE)*, 388–393. https://doi.org/10.1109/ICPRE. 2016.7871238

Lawan, S. M., Abidin, W. A. W. Z., Chai, W. Y., Baharun, A., & Masri, T. (2014). *Different Models of Wind Speed Prediction; A Comprehensive Review*. *5*(1).

Lee, J., & Zhao, F. (2021). Global Wind Report 2021. In *Global Wind Energy Council*.

Liu, H., Duan, Z., Wu, H., Li, Y., & Dong, S. (2019). Wind speed forecasting models based on data decomposition , feature selection and group method of data handling network. *Measurement*, *148*, 106971. https://doi.org/10.1016/ j.measurement. 2019.106971

Liu, Y., Guan, L., Hou, C., Han, H., Liu, Z., & Sun, Y. (2019). Wind Power Short-Term Prediction Based on LSTM and Discrete Wavelet Transform. *MDPI*

*Journal of Applied Sciences*, *9*. https://doi.org/10.3390/app9061108

Lotfi, H., Lotfi, T., & Hesari, S. (2014). *Short and Mid-Term Wind Power Forecasting with ANN-PSO*. *September*.

Lu, Y., Khan, Z., Alvarado, M., Zhang, Y., & Huang, Z. (2020). A Critical Review of Sustainable Energy Policies for the Promotion of Renewable Energy Sources. *MDPI Journal of Sustainability, 12*(12).

Mathenge, J. N., Murage, D. K., & Nderu, J. N. (2021). *A Short-Term Wind Power Forecasting Approach*. *5*(1).

McHugh, J. (2020). *An Overview of Different Time Series Models for Forecasting*. https://medium.com/@teamastertoast/an-overview-of-different-time-series-models-for-forecasting-c06c071c7684

Namini, S. S., Tavakoli, N., & Namin, A. S. (2019). The Performance of LSTM and BiLSTM in Forecasting Time Series. *IEEE International Conference on Big Data*.

Namini, S. S., Tavakoli, N., & Namin, Ak. (n.d.). *The Performance of LSTM and BiLSTM in Forecasting Time Series*. https://doi.org/10.1109/BigData 47090.2019. 9005997

Oxford Net Zero. (n.d.). *WHAT IS NET ZERO?* https://netzeroclimate.org/what-is-net-zero/

Quan, J., & Shang, L. (2021). *An Ensemble Model of Wind Speed Forecasting Based on Variational Mode Decomposition and Bare-Bones*. *2021*.

Rafaat, S. M., & Hussein, R. (2018). *Power Maximization and Control of Variable-Speed Wind Turbine System Using Extremum Seeking*. 51–69. https://doi.org/10.4236/jpee.2018.61005

Ragheb, A. M., & Ragheb Magdi. (2011). Wind Turbines Theory - The Betz Equation and Optimal Rotor Tip Speed Ration. In *Fundamental and Advanced*

*Topics in Wind Power* (pp. 19–29).

Rasamoelina, D. (2020). *A Review of Activation Function for Artificial Neural Network*. 281–286.

Schober, P., Boer, C., & Schwarte, L. A. (2018). *Correlation Coefficients: Appropriate Use and Interpretation*. *XXX*(Xxx), 1–6. https://doi.org/10.1213/ANE. 0000000000002864

Shahid, F., Zameer, A., & Muneeb, M. (2021). A novel genetic LSTM model for wind power forecast. *Energy*, *223*, 120069. https://doi.org/10.1016/j.energy .2021.120069

Shao, H., Deng, X., & Cui, F. (2016). *Short-term wind speed forecasting using the wavelet decomposition and AdaBoost technique in wind farm of East China*. *10*, 2585–2592. https://doi.org/10.1049/iet-gtd.2015.0911

Sharkawy, A. (2020). *Principle of Neural Network and Its Main Types : Review Principle of Neural Network and Its Main Types : Review*. *August*, 7–19. https://doi.org/10.15377/2409-5761.2020.07.2

Shrestha, M. B., & Bhatta, G. R. (2017). Selecting appropriate methodological framework for time series data analysis. *The Journal of Finance and Data Science, 4*(2), 71–89. https://doi.org/https://doi.org/10.1016/j.jfds.2017.11.001

Singh, V. (2016). *Application of Artificial Neural Networks for Predicting Generated Wind Power*. *7*(3), 250–253.

U.S Energy Information Administration. (2021). *Electricity explained Electricity in the United States*. https://www.eia.gov/energyexplained/electricity/electricity-in-the-us.php

Vestas Wind Systems. (n.d.). *V52-850 kW The turbine that goes anywhere*.

Wang, J. (2014). *A Hybrid Wavelet Transform Based Short-Term Wind Speed*. *2014*, 12–17.

Wang, Y., Zou, R., Liu, F., Zhang, L., & Liu, Q. (2021). A review of wind speed and wind power forecasting with deep neural networks. *Applied Energy*, *304*(April). https://doi.org/10.1016/j.apenergy.2021.117766

Wu, Qianyu Guan1, Fei Lv, Chen Huang, Y. (n.d.). *IET Renewable Power Gen - 2021 - Wu - Ultra-short-term multi-step wind power forecasting based on CNN-LSTM.pdf*.

Yanga, Mo Wang, J. (2022). Adaptability of Financial Time Series Prediction Based on BiLSTM. *Procedia Computer Science*, 18–25.

Zhang, Shijie Wei, Jing Chen, Xi Zhao, Y. (2020). China in global wind power development: Role, status and impact. *Renewable and Sustainable Energy Reviews*, *127*. https://doi.org/https://doi.org/10.1016/j.rser.2020.109881

Zhang, Y., Zhang, C., Sun, J., & Guo, J. (2018). Improved wind speed prediction using empirical mode decomposition. *Advances in Electrical and Computer Engineering*, *18*(2), 3–10. https://doi.org/10.4316/AECE.2018.02001

Zhao, X., Wang, S., & Li, T. (2011). Review of Evaluation Criteria and Main Methods of Wind Power Forecasting. *Energy Procedia*, *12*, 761–769. https://doi.org/10.1016/j.egypro.2011.10.102

Zheng, Y., Hill, D. J., Meng, K., Luo, F. J., & Dong, Z. Y. (2015). Optimal Short-term Power Dispatch Scheduling for a Wind Farm with Battery Energy Storage System. *IFAC-PapersOnLine*, *48*(30), 518–523. https://doi.org/10.1016/j.ifacol.2015.12.432

Zheng, Z., Chen, Y., Zhou, X., & Huo, M. (2012). *Short-Term Wind Power Forecasting Using Empirical Mode Decomposition and RBFNN*.

Zhu, X., & Genton, M. G. (2012). *Short-Term Wind Speed Forecasting for Power System Operations*. 2–23. https://doi.org/10.1111/j.1751-5823.2011.00168.x

# APPENDICES

## Appendix I: Wind Power Dataset Information

| Data Name | Source | Data Resolution | Span |
|---|---|---|---|
| AL_WIND_07_12 | National Renewable Energy Laboratory (**NREL**) | **Hourly** observations | 01/01/2007 00:00:00 to 31/12/2012 23:00:00 (**52560** observations equivalent to 6 years of data) |

**Dataset available on**: https://github.com/ShashwatArghode/Wind-Energy-Prediction-using-LSTM/blob/master/AL_WIND_07_12.xlsx

**or on**

https://developer.nrel.gov/docs/wind/wind-toolkit/wtk-download/

| | | |
|---|---|---|
| **AL_WIND_07_12 Dataset** | Train Data | 01/01/2007 - 31/12/2010 |
| | Test Data Sample | 01/01/2011 - 31/12/2012 |

**The test data was drawn from the Test Data Sample depending on the look ehead period required.*

**Validation data is the Actual Wind Power Data from the Test Data Sample and this is compared to the forecast values.*

***First 20 Sampled Data from the 24,48 and 72 hour prediction Horizons (BiLSTM_Avrami Model) (Power in kW)**

| Y_Test Raw | Y_Pred_Raw (24 hour forecast) | Y_Pred_Raw (48 hour forecast) | Y_Pred_Raw (72 hour forecast) |
|---|---|---|---|
| 8471.07 | 8581.556668 | 8648.437668 | 8701.334475 |
| 13137.4003 | 13036.30908 | 13012.05692 | 13058.99244 |
| 7670.32 | 7709.916107 | 7815.397603 | 7842.965344 |
| 1227.85 | 1183.407032 | 1319.599016 | 1295.435147 |
| 0 | -133.6945785 | 5.730502639 | -73.74393028 |
| 0 | -72.68382779 | 96.76048335 | 6.480254714 |
| 1342.74 | 1340.724405 | 1444.953205 | 1411.443003 |
| 3925.305645 | 3911.072786 | 3977.590559 | 3971.059796 |
| 1011.39 | 1071.63913 | 1193.544613 | 1183.317366 |
| 0 | -50.83392647 | 66.07001573 | 24.92782338 |
| 0 | -70.57782557 | 71.36599356 | 3.676662795 |
| 0 | -33.77885638 | 107.8803941 | 48.38349291 |
| 0 | -8.76316408 | 112.3915692 | 70.47809249 |
| 30.13869999 | 6.79428416 | 131.1815443 | 85.88701059 |
| 0 | -27.25801826 | 100.0849004 | 52.80431793 |
| 0 | -28.487898 | 98.56577477 | 51.11746846 |
| 0 | -28.39138763 | 98.92871535 | 51.30204738 |
| 0 | -28.20178925 | 99.11112679 | 51.54275291 |
| 0 | -28.00374907 | 99.12789632 | 51.70120787 |
| 1322.365922 | 1261.394481 | 1444.994273 | 1404.362953 |

## Appendix II: Developed MatLab Codes

### BiLSTM Alone

```
clear all;

close all;

dataWind=readtable('AL_WIND_07_12.csv');

time=dataWind.DateTime;

dataWind.DateTime=[];

record=dataWind{:,:};

cos_angle = cosd(record(:,4));

record(:,4) = cos_angle;

%%Normalizing the data

for i=1:size(record,2)

    record(:,i)=(record(:,i)-min(record(:,i)))/(max(record(:,i))-
min(record(:,i)));

end

record=record(1:100,:);

% %Partition Data for Training and Testing

numObservations=size(record,1);

idxTrain = 1:floor(0.7*numObservations);

idxTest = floor(0.7*numObservations):numObservations;

%separate the dataset into training and testing

dataTrain = record(idxTrain,:);

dataTest = record(idxTest,:);

%separate the training dataset into inputs and outputs

XTrain = dataTrain(:,1:4); %select input variables

YTrain = dataTrain(:,5);   %select target variable (power)
```

```
for i=1:size(XTrain,1)
    XTrain2(i)={XTrain(i,:)}; %Convert the inputs into a cell array.
end
XTrain=XTrain2;
YTrain=YTrain';
%% Train
inputSize = 1;
numHiddenUnits = 150;
layers = [ ...
    sequenceInputLayer(inputSize)
    bilstmLayer(numHiddenUnits,'OutputMode','last')
    fullyConnectedLayer(1)
    regressionLayer];
maxEpochs = 10;
options = trainingOptions('adam', ...
    'MaxEpochs',maxEpochs, ...
    'InitialLearnRate',0.001, ...
    'GradientThreshold',1, ...
    'Shuffle','never', ...
    'Plots','training-progress',...
    'SequenceLength',6, ...
    'Verbose',0);
net = trainNetwork(XTrain,YTrain',layers,options);
trainY=double(predict(net,XTrain));
%Test network
%separate the testing dataset into inputs and outputs
look_ahead=24; %Repeat for 48 hours, 72 hours, 168 hours.
```

```
XTest = dataTest(1:look_ahead,1:4);

YTest = dataTest(1:look_ahead,5);

for i=1:size(XTest,1)

XTest2(i)={XTest(i,:)};

end

XTest=XTest2;

YTest=YTest';

YPred = predict(net,XTest);

YPred=YPred';

for i=1:size(YPred,2)

    if YPred(i)<0

        YPred(i)=0;

    end

end

% Getting actual test values

for i=1:size(YTest,2)

YTest_actual(i)=min(dataWind.Power)+(YTest(i)*(max(dataWind.Power)-
min(dataWind.Power)));

end

% Getting actual predicted values

for i=1:size(YTest,2)


YPred_actual(i)=min(dataWind.Power)+(YPred(i)*(max(dataWind.Power)-
min(dataWind.Power)));

end

%Plot the actual figures

figure

plot(YTest_actual)
```

```
hold on

plot(YPred_actual)

hold off

legend('Actual','Predicted')

xlabel('Hours')

ylabel('Power')

title('Graph of Wind Power vs Hour (BiLSTM)')

figure, ploterrhist(YPred_actual-YTest_actual)

figure, plot(YPred_actual,'-o')

hold on

plot(YTest_actual,'-^')

title('Train Results')

xlabel('Time')

ylabel('Power Output');

legend('biLSTM output','Actual Demand')

%Performance evaluation of the model

rmse_power=sqrt(mean((YPred_actual-YTest_actual).^2))

MAE = mean(abs(YTest_actual-YPred_actual))

MAPE = mean(abs((YTest_actual-YPred_actual)/YTest_actual))

corrcoef(YPred_actual,YTest_actual)
```

## BiLSTM EMD

```
clear all;

close all;

dataWind=readtable('AL_WIND_07_12.csv');   %Has   52560   observation
points

speedData=dataWind.Power; % power selected as the variable to be
interrogated.

sample_data=speedData(1:8760);%Pick a sample of data
```

```
% %Partition Data for Training and Testing
numObservations=size(sample_data,1);
idxTrain = 1:floor(0.7*numObservations);
idxTest = floor(0.7*numObservations):numObservations;
%Feature Scaling of the sampled dataset
min_sample=min(sample_data);
max_sample=max(sample_data);
sample_dataSTD=(sample_data-min_sample)./(max_sample-min_sample);
%
%emd on dataTrain
[imf_train,res_train]=emd(sample_dataSTD');
sample_dataEMD=[imf_train,res_train];
%separate the dataset into training and testing
dataTrain = sample_dataEMD(idxTrain,:);
dataTest = sample_dataEMD(idxTest,:);
XTrain=dataTrain(1:end-1,:)';
YTrain=dataTrain(2:end,:)';
numFeatures=size(XTrain,1);
numResponses=size(XTrain,1);
numHiddenUnits=150;
layers = [
    sequenceInputLayer(numFeatures)
    bilstmLayer(numHiddenUnits,'OutputMode','sequence')     %formerly
128 on 23.3.2022
    dropoutLayer(0.2)
    fullyConnectedLayer(numResponses)
    regressionLayer];
```

```
options = trainingOptions('adam', ...
    'MaxEpochs',100, ...
    'MiniBatchSize',30, ...
    'InitialLearnRate',0.01, ...
    'GradientThreshold',1,...
    'Shuffle','never', ...
    'Plots','training-progress', ...
    'Verbose',0);
net=trainNetwork(XTrain,YTrain,layers,options)
dataTest=dataTest';
look_ahead=24;
dataTest2=dataTest(:,1:look_ahead+1);
XTest=dataTest2(:,1:end-1);
YTest=dataTest2(:,2:end);
YPred = predict(net,XTest);
%Get the sums
for j=1:size(YPred,2)
    YPredSum(j)=sum(YPred(:,j));
    YTestSum(j)=sum(YTest(:,j));
end
YPredRaw=YPredSum;
YTestRaw=YTestSum;
YPredRaw=(YPredSum*(max_sample-min_sample))+min_sample;
YTestRaw=(YTestSum*(max_sample-min_sample))+min_sample;

YPredRaw2=YPredRaw(1:end);
```

```
YTestRaw2=YTestRaw(1:end);

%Plot the figure

figure

plot(YTestRaw2)

hold on

plot(YPredRaw2)

hold off

%Plot the figure

figure

plot(YTestRaw2)

hold on

plot(YPredRaw2)

hold off

legend('Actual','Predicted')

xlabel('Hours')

ylabel('Power')

title('Graph of Wind Power vs Hour (BiLSTM + EMD)')

%Performance evaluation of the model

rmse_power=sqrt(mean((YPredRaw2-YTestRaw2).^2))

MAE = mean(abs(YPredRaw2-YTestRaw2))

MAPE = mean(abs((YPredRaw2-YTestRaw2)/YTestRaw2))

corrcoef(YPredRaw2,YTestRaw2)
```

**BiLSTM Avrami**

```
clc

clear

close all
```

```
%%

record_actual=csvread('AL_WIND_07_12 - No Labels.csv');

record=csvread('AL_WIND_07_12 - No Labels.csv');

look_back = 6;

sizeData=size(record,1);

avrami_data=xlsread('AL_WIND_07_12_Avrami.xlsx');

% %Based on the look_back, we replace the values of power at time t
with the

% %estimates from AVRAMI

for i=look_back:look_back:size(record,1)

     record(i,5)=avrami_data(i,6);

end

%%Normalizing the data

for i=1:size(record,2)

    record(:,i)=(record(:,i)-min(record(:,i)))/(max(record(:,i))-
min(record(:,i)));

end

numObservations=size(record,1);

idxTrain = 1:floor(0.7*numObservations);

idxTest = floor(0.7*numObservations):numObservations;

in=record(idxTrain,5);

out=record(idxTrain,5);

i=1;

while ~isempty(in)

    pick=look_back;

    if pick<size(in,1)

        X_train{i}=(in(1:pick,:))';

        Y_train(i)=out(pick); %Estimate provided from Avrami Eqn
        in(1,:)=[];

        out(1,:)=[];

        i=i+1;

    else

        X_train{i}=in';

        Y_train(i)=out(end);

        break;

    end

end

%% Train

inputSize = 1;

numHiddenUnits = 150;

layers = [ ...

    sequenceInputLayer(inputSize)

    bilstmLayer(numHiddenUnits,'OutputMode','last')

    fullyConnectedLayer(1)

    regressionLayer];

maxEpochs = 10;

options = trainingOptions('adam', ...

    'MaxEpochs',maxEpochs, ...

    'InitialLearnRate',0.001, ...

    'GradientThreshold',1, ...

    'Shuffle','never', ...

    'Plots','training-progress',...

    'SequenceLength',6, ...

    'Verbose',0);

net = trainNetwork(X_train,Y_train',layers,options);
```

```
trainY=double(predict(net,X_train));

figure, ploterrhist(trainY-Y_train')

figure, plot(trainY,'-o')

hold on

plot(Y_train,'-^')

title('Train Results')

xlabel('Time')

ylabel('Power Output');

legend('biLSTM output','Actual Demand')

%% test network

in_test=record(min(idxTest):min(idxTest)+24,5);

out_test=record(min(idxTest):min(idxTest)+24,5);

%Preparing time series data for specified look_back

i=1;

while ~isempty(in_test)

    pick=look_back;

    if pick<size(in_test,1)

        X_test{i}=(in_test(1:pick,:))';

        Y_test(i)=out_test(pick);

        in_test(1,:)=[];

        out_test(1,:)=[];

        i=i+1;

    else

        X_test{i}=in';

        Y_test(i)=out(end);

        break;

    end
```

```
end

testY=double(predict(net,X_test));

figure, ploterrhist(testY-Y_test'), title('test')

figure, plot(testY,'-o')

hold on

plot(Y_test,'-^')

title('Test Results')

xlabel('Time')

ylabel('Output Power');

legend('biLSTM output','Actual Demand')

% Getting actual test values

for i=1:size(Y_test,2)


Y_test_raw(i)=min(record_actual(:,5))+(Y_test(i))*(max(record_actual(:,5))-min(record_actual(:,5)));

     i=i+1;

end

 %Getting actual predicted values

 for i=1:size(testY,1)

Y_predicted_raw(i)=min(record_actual(:,5))+(testY(i))*(max(record_actual(:,5))-min(record_actual(:,5)));

     i=i+1;

 end

%Performance evaluation of the model

rmse_power=sqrt(mean((Y_predicted_raw-Y_test_raw).^2))

MAE = mean(abs(Y_test_raw-Y_predicted_raw))

MAPE = mean(abs((Y_test_raw-Y_predicted_raw)/Y_test_raw))

corrcoef(Y_predicted_raw,Y_test_raw)
```

## Appendix III: Conference and Journal Papers Published

From this research, two conference papers and two journal papers were published.

Mathenge J.N, Murage D.K & Nderu J.N. (2021). A Short-Term Wind Power Forecasting Approach using ANFIS. *International Research Journal of Innovations in Engineering and Technology*, 5 (5), 35 - 42. https://doi.org/10.47001/IRJIET/2021.505007

Mathenge J.N, Murage D.K, Nderu J.N & Muriithi C.M. (2018). A Review on Artificial Neural Network Models for Short-Term Wind Power Prediction. Proceedings of the 2018 Sustainable Research and Innovation Conference, 54 - 58. https://sri.jkuat.ac.ke/jkuatsri/index.php/sri/article/view/37

Mathenge J.N, Murage D.K, Nderu J.N & Muriithi C.M. (2018). Grid Integration of Large Capacity Wind Power: A Review. Proceedings of the 2018 Sustainable Research and Innovation Conference, 75 - 79.https://sri.jkuat.ac.ke/jkuatsri/index.php/sri/article/view/41

Mathenge J.N, Murage D.K & Nderu J.N. (2023). A Short-Term Hybrid Wind Power Forecasting Approach using BiLSTM_EMD and the Avrami Curve. *International Research Journal of Innovations in Engineering and Technology*, 7 (10), 376 - 392. https://doi.org/10.47001/IRJIET/2023.710051