



REPUBLIC OF TÜRKİYE

ALTINBAŞ UNIVERSITY

Institute of Graduate Studies

Electrical and Computer Engineering

**WEATHER PREDICTION BASED ON HYBRID  
DEEP LEARNING REGRESSION**

**Samer Shuaa Abdulnabi AL-MAYYAHİ**

Master's Thesis

Supervisor

Asst. Prof. Dr. Mesut ÇEVİK

Istanbul, 2023

# **WEATHER PREDICTION BASED ON HYBRID DEEP LEARNING REGRESSION**

**Samer Shuaa Abdulnabi AL-MAYYAH**

Electrical and Computer Engineering

Master's Thesis

ALTINBAŞ UNIVERSITY

2023

The thesis titled WEATHER PREDICTION BASED ON HYBRID DEEP LEARNING REGRESSION prepared by SAMER SHUAA ABDULNABI AL-MAYYAHİ and submitted on 27/03/2023 **accepted unanimously** for the degree of Master of Science in Electrical and Computer Engineering.

---

Asst. Prof. Dr. Mesut ÇEVİK

Supervisor

Thesis Defense Committee Members:

Asst. Prof. Dr. Mesut ÇEVİK

Department of Electrical and  
Electronics Engineering

Altınbaş University

---

Asst. Prof. Dr. Abdullahi Abdu  
IBRAHİM

Department of Computer  
Engineering

Altınbaş University

---

Asst. Prof. Dr. İzzet Paruğ DURU

Department of Medical  
Imaging Techniques

Istanbul Gedik University

---

I hereby declare that this thesis meets all format and submission requirements of a Master's thesis.

Submission date of the thesis to Institute of Graduate Studies: \_\_\_/\_\_\_/\_\_\_

I hereby affirm that all materials and knowledge presented in this graduation project were acquired legally and ethically. I further attest that all nonoriginal information and conclusions have been appropriately cited in the text, that all references listed in the Reference List have been appropriately cited in the text, and that I will abide by the aforementioned standards of behaviour and conduct when publishing my work.

Samer Shuaa Abdulnabi AL-MAYYAH

Signature

## **ACKNOWLEDGEMENTS**

Thank you, Allah Almighty, for every think in my life. Some of them were blessings, and some were Lessons, and I am grateful for everything you sent my way this thesis is dedicated to my parent for raising me to believe that anything was possible and for her constant love. To my advisor Dr. Mesut who have been my source of inspiration, strength, advice and encouragement to finish this study, and for the scientific committees who evaluate my study.



## ABSTRACT

### WEATHER PREDICTION BASED ON HYBRID DEEP LEARNING REGRESSION

AL-MAYYAH, Samer Shuaa Abdulnabi

M.Sc., Electrical and Computer Engineering, Altınbaş University,

Supervisor: Asst. Prof. Dr. Mesut ÇEVİK

Date: March / 2023

Pages: 68

All living organisms share the existence of nervous systems that enable them to deal and interact with the surrounding environment and help them control the vital processes necessary to sustain life. In higher organisms such as man, the human nervous system is considered the most complex of nervous systems at all, most of which are concentrated in the human brain, which is characterized by a nature of work that has led to the superiority of man over all other creatures in the abilities of understanding and recognizing shapes and symbols, learning, speaking, remembering, cognition, and precise control of the locomotor system, etc. To that of many qualities and capabilities that no other being other than humans can access, and the artificial system for forecasting weather is not new, but it is being used more and differently air. The experimental results showed that the proposed hybrid deep learning model has great performance and offers high prediction results compared to the machine learning algorithm, where the suggested deep model's MAE is always less than 15 and the worst result of RMSE is 22.09, on the other hand, the best results gained from the linear regression algorithm It equals 47.352 for MAE and 65.606 for RMSE. The proposed deep model is approved for its power to give high predictive results when the system is implemented during an epoch equal to 400, the back-to-back period equal to 10, where the obtained MAE result is 5.262, the RMSE is 18.243, and the MAPE value is 0.0000376, which the prediction result is considered almost complete. The result of this result agreed with the power of the deep learning model in performing data regression, giving impressive prediction results.

**Keywords:** Weather, Hybrid Deep, Regression, ML, DL.

# TABLE OF CONTENTS

	<u>Pages</u>
<b>ABSTRACT .....</b>	<b>vi</b>
<b>LIST OF TABLES.....</b>	<b>ix</b>
<b>LIST OF FIGURES.....</b>	<b>x</b>
<b>ABBREVIATIONS.....</b>	<b>xi</b>
<b>LIST OF SYMBOLS.....</b>	<b>xii</b>
<b>1. INTRODUCTION .....</b>	<b>1</b>
1.1 OVERVIEW .....	1
1.2 FORECASTING.....	1
1.3 PROBLEM STATEMENT.....	3
1.4 AIM OF THESIS .....	3
1.5 RELATED WORK.....	4
1.6 THESIS LAYOUT .....	9
<b>2. LITERATURE REVIEW .....</b>	<b>10</b>
2.1 INTRODUCTION .....	10
2.2 WEATHER FORECASTING OVERVIEW .....	10
2.2.1 Forecasting Weather Challenges.....	11
2.2.2 The Use of Software in Weather Prediction .....	14
2.3 DATA NORMALIZATION .....	17
2.4 AI AND MACHINE LEARNING.....	20
2.4.1 Machine Learning Linear Regression Algorithm .....	21
2.5 DEEP LEARNING DL OVERVIEW .....	24
2.5.1 Long Short-Term Memory LSTM.....	26
2.6 PERFORMANCE MEASUREMENTS .....	28
<b>3. MATERIALS AND METHODS .....</b>	<b>31</b>
3.1 INTRODUCTION .....	31

3.2 PROPOSED MODEL DESIGN AND IMPLEMENTATION.....	31
3.3 DATASET DESCRIPTION .....	33
3.4 DATASET NORMALIZATION PHASE .....	35
3.5 LINEAR REGRESSION MACHINE LEARNING ALGORITHM .....	35
3.6 REGRESSION BY PROPOSED HYBRID DEEP LEARNING MODEL.....	37
3.6.1 Lstm Layer .....	37
3.6.2 Dense Layers.....	39
<b>4. RESULTS AND DISCUSSION .....</b>	<b>41</b>
4.1 INTRODUCTION .....	41
4.2 PROPOSED SYSTEM IMPLEMENTATION ENVIRONMENT DESCRIPTION	41
4.3 IMPLEMENTATION RESULTS USING THE LINEAR REGRESSION MACHINE LEARNING ALGORITHM .....	41
4.4 DEEP HYBRID REGRESSION MODEL IMPLEMENTATION OUTCOMES ...	42
4.4.1 Training Phase Results.....	43
4.4.2 Consequences of the Trial Run .....	46
<b>5. CONCLUSION AND FUTURE WORKS.....</b>	<b>53</b>
5.1 CONCLUSIONS .....	53
5.2 FUTURE WORK.....	53
<b>REFERENCES .....</b>	<b>55</b>

## LIST OF TABLES

	<u>Pages</u>
Table 3.1: The Datasets Quantities Description. ....	34
Table 4.1: Linear Regression Algorithm Prediction Results. ....	42
Table 4.2: Proposed Deep Hybrid Model Layers. ....	42
Table 4.3: Training Results When Epoch =100, And Look Back =10. ....	43
Table 4.4: Training Results Within 100 Epochs. ....	44
Table 4.5: Training Results Within 200 Epoch. ....	44
Table 4.6: Training Results Within 300 Epoch. ....	44
Table 4.7: Training Results Within 400 Epoch. ....	45
Table 4.8: Training Results Within 500 Epoch. ....	45
Table 4.9: Testing Results Within 100 Epochs. ....	47
Table 4.10: Testing Results Within 200 Epochs. ....	47
Table 4.11: Testing Results Within 300 Epoch. ....	48
Table 4.12: Testing Results Within 400 Epoch. ....	48
Table 4.13: Testing Results Within 500 Epoch. ....	49

## LIST OF FIGURES

	<u>Pages</u>
Figure 2.1: Weather Factors. ....	10
Figure 2.2: The Visual Representation of the Linear Regression [67]. ....	22
Figure 2.3: The Simple Neural Network Structure. ....	23
Figure 2.4: Deep Neural Network Architecture. ....	25
Figure 2.5: Long Short-Term Memory Structure. ....	27
Figure 3.1: The Block Diagram of Weather Forecasting Within Machine Linear Regression Algorithm. ....	32
Figure 3.2: The Block Diagram of Weather Forecasting Using Proposed Deep Hybrid Model. ....	32
Figure 3.3: The Proposed Deep Hybrid Models' Layer's Structure. ....	40
Figure 4.1: MAE for all Epoch Value Using Look Back Equal to 40 in the Training Phase. ....	46
Figure 4.2: RMSE For all Epoch Value Using Look Back Equal to 40 in the Training Phase. ....	46
Figure 4.3: MAE for All Epoch Value Using Look Back Equal to 10 in the Testing Phase. ....	49
Figure 4.4: RMSE for all Epoch Value Using Look Back Equal to 10 in the Testing Phase. ....	50
Figure 4.5: MAPE for All Epoch Value Using Look Back Equal to 10 in the Testing Phase. ....	50
Figure 4.6: The Mean Absolute Error (MAE) for the Optimal Outcome of the Suggested Deep Model And LR Method. ....	51
Figure 4.7: Using The Suggested Deep Model and LR Algorithm, What Is The RMSE for The Best Possible Outcome. ....	51
Figure 4.8: As Assessed by the Maximal Average Performance Evaluation, The Deep Model and LR Algorithm Have the Best Overall Performance. ....	52

## ABBREVIATIONS

AI	:	Artificial Intelligence
ANN	:	Artificial Neural Network
DBN	:	Deep Belief Network
DL	:	Deep Learning
LSTM	:	Long Term Memory Network
MAE	:	Mean Absolute Error
MLP	:	Multilayer Perceptual
ML	:	Machine Learning
NWP	:	Numerical Weather Forecast
NZEB	:	Net Zero Energy Buildings
RMSE	:	Root Mean Squared Error
TCN	:	Temporal Convolutional Network
UAV	:	Unmanned Aerial Vehicle
MAPE	:	Mean Absolute Percentage Error

## LIST OF SYMBOLS

- + : Addition
- × : Multiplication



# 1. INTRODUCTION

## 1.1 OVERVIEW

In recent years, there has been an increasing focus on the capacity to forecast the behavior of a system. This skill has several applications, including the predicting of financial events, the preparation of product exchanges and staff presentations, and more. manage commodities and predict, in medicine to track the progression of disease, and in meteorology to forecast precipitation [1]. It may be challenging to foresee the future, particularly if you are only willing to explore model-based tactics. In the domains of business, economics, and management, the capacity to properly predict future occurrences is crucial for making many choices. The potential ramifications of recent estimates have captivated the attention of a substantial number of academics. If people were given prior notice of natural disasters, it might be possible to save a substantial number of lives each year. It should not come as a surprise that agricultural is one of the many applications where precise forecasts are quite important; in fact, agriculture is one of these applications. To increase the productivity of cereal grain cultivation, transport, and storage. It is crucial to have the ability to adjust to various types of storms. In the event that they were identified in time, it would have the potential to save a large number of lives. The effects of severe weather and other external factors may have both direct and indirect effects on the transportation industry and other types of management [2]. There has been an upsurge in rivalry between deep learning, machine learning, and AI technologies in recent years. This is due to the known difficulty and unpredictability of weather forecasting. Deep learning may be seen as a mathematical strategy for learning data representations, while conventional numerical approaches are, well, conventional. Comparatively, conventional numerical procedures are just that: traditional. Significant emphasis is placed on going through logically specified stages of schooling [3] Because deep learning algorithms have a track record of producing reliable and accurate findings, we will utilize them to produce predictions throughout.

## 1.2 FORECASTING

- a. As a result of partaking in this activity, one gains the capacity to predict events, which aids in giving a preferred adaptation, as well as the ability to eliminate or alter undesirable outcomes in favor of more desirable outcomes. Forecasting is a statistical activity used in

the business world to inform production, transportation, and personnel scheduling decisions, as well as to guide the development of long-term strategy. Since the Babylonians around 650 B.C. attempted to anticipate short-term changes in the weather based on the presence of clouds in the sky and other visual phenomena such as auras [4], it is clear that humans have always considered the act of trying to foretell beneficial. Accurate forecasting helps societies grow by increasing the rate at which their goals and desires are realized. A solid forecast may be used again to make the same sort of accurate forecasts, therefore it is an investment with a long-term payoff [5]. It is now common practice for forecasting to include an evaluation of the quality or amount of data employed in a given computation. This review may be included into the actual forecast. The fact that it is ultimately controlled by humans and has the capacity to introduce some degree of unpredictability into the future despite its widespread usage makes it susceptible to abuse. Frequently, predictions and fortune-telling are intertwined. The distinction is in the accuracy with which one can forecast what will occur in the future. Since the advent of more powerful technology, most notably computers, the use of prediction has expanded substantially. The entertainment industry use weather forecasting technologies and strategies the most. Popular are both the technical prognosis and the economic forecasts. The forecasting of these regions is crucial since even a little degree of precision may save a substantial amount of resources. Preventing crop loss and boosting agricultural productivity are two examples [6] of how this might be achieved. The administration of power grids, communication networks, and supply chains, to mention a few, rely heavily on predictions. As a consequence, the most accurate predictions do not aim to duplicate past events, but instead focus on the underlying patterns and relationships that may be seen in the data that came before. There are a lot of similarities between the various preprocessing tools and the personalized tactics used in time-series forecasting. There is only one effective way of prediction that can be used in the implementation and functioning of any given forecasting algorithm. Forecasters of time-series tend to construct and propose solutions that can handle a wide variety of issues and have a low error rate. They do this action in an attempt to increase their accuracy [7].

- b. Short term forecasts: These are used for project planning, production and delivery. Forecasting demand and supply requirements can also be part of planning processes, such as transportation and production planning, as well as inventory and replenishment.

- c. Medium-term forecasting: it is required to specify the resources needed in the future, to purchase equipment, hire workers or buy machinery and equipment.
- d. Long-term forecasting: used in strategic planning. As these decisions will take into account business and market opportunities, environmental issues and real estate.

A good forecast assumes good values and relationships between historical data, but does not repeat past events that will never happen again, a good forecast must have the following requirements:

- a. The prediction must be accurate.
- b. Weather is important in any forecast.
- c. Trustworthy and reliable policy.
- d. Ease of use is important in any forecasting process
- e. Forecasts should be presented in understandable units.
- f. Any policy must be written.

### **1.3 PROBLEM STATEMENT**

The absence of spatial or temporal autocorrelation in the data makes it challenging to construct reliable climate forecasts. Since, in addition to the distance between places, other factors in the sky, such as topography, prevailing winds, and water jets, can also affect the weather. To predict the weather traditionally, sophisticated numerical models are used to solve various equations believed to govern the air-ocean system. This is an expensive task, and advances in processing power are linked to advances in forecasting. Traditional climate forecasting involves the sophistication of physical modeling techniques for unsolved elements, in addition to computation. No fewer than four inputs are used in AI-based weather forecasting systems at present.

### **1.4 AIM OF THESIS**

Design and build weather forecasting models that are based on deep learning concepts and can overcome all the challenges of weather forecasting and use all weather methods to provide accurate, reliable and fast forecasts to avoid the dangers of weather. Climate has an impact on improving people's lives.

## 1.5 RELATED WORK

As a direct result of the enormous climatic changes that have happened over the past several decades, scientists have acquired a far stronger interest in climate forecasting and prediction. Several persons have built depth-based models and offered them as a mechanism to implement the following policies in order to achieve this purpose.

The work [8] done by Grover and coworkers has recently been published and is now freely available online. Using hybrid models, we were able to precisely forecast the weather in 2015. These models integrated learned prediction models with deep neural networks in order to produce joint statistics of climate change. Even though it is difficult to predict the impacts of climate change, these models helped us generate more accurate estimates. In this study, we illustrate how spatial inversion and long-range spatial dependence may be used to enhance the basic model. In addition, it learns and integrates, which are essential components for enhancing model parameters. The IGRA database consists of balloon observations gathered from sixty distinct locations around the United States. These balloons collect meteorological information like wind speed and direction, temperature, projected altitude, and dew point. The model incorporates several meteorological variables, including wind speed, air pressure, temperature, and dew point. Each of the proposed models consists of three primary elements: a set of climate change predictors, a method for adjusting the input of individual predictors so that their combined output is smooth and consistent with the constraints imposed by physical laws, and a set of deeply held beliefs that drive the search for a solution. The following is the sequence of these three components: A collection of climate change predictors, a technique for changing the input of individual predictors such that their aggregate output is smooth, and an ensemble of climate change forecasting models. As a consequence of climate change, the expected accuracy rates are 2.11 percent for six hours, 1.03 percent for twelve hours, and 1.01 percent for twenty-four hours. To create reliable daily precipitation forecasts, Hernandez et al. [9] introduced in 2016 a deep learning-based technique. Numerous forecasting jobs, in particular, may be completed with the aid of an automated regression, which reduces the number of features, finds improper connections between those characteristics, and provides a number of other advantages. One of the two networks that contribute to the system's overall structure is the auto-encoder network. The deep learning network shown here is merely one of the techniques used to evaluate and quantify data based on characteristics and time. To be able to think ahead and strategically

plan, the brain's cognitive network requires the cooperation of a variety of distinct nodes. After putting the recommended approach to the test, it was determined that its MSE was 40.11 and its RMSE was 6.

Deep Belief Networks (DBN) and Block Boltzmann Machines have been combined in the hybrid model proposed in 2017 by Narejo et al. [10]. (RBM). Among the environmental characteristics that this model can forecast are temperature, pressure, and humidity, to mention a few. These restrictions included a ban on accessing certain locations. Initially, the RBM was trained alone, with no assistance from humans. After the RBMs have been merged, their combined resources may be used to build the DBN. To instruct DBN to provide accurate weather forecasts, a very particular mechanism is used. Three layers comprise a DBN: the input layer, the hidden layer, and the output layer. These layers are hidden from the user. RBMs ranging in size from 500 to 3000 are at the center of an irreversible revolutionary development that has resulted in the formation of a number of hidden layers. With RMSE values of  $9.06e-4$ ,  $1.5e-3$ , and  $0.58e-4$  for the expected temperature, humidity, and pressure, respectively, the findings indicated that the forecasting model of the plan was very accurate.

In 2018, Rodriguez et al. [11] proposed a technique that takes use of a deep neural network to learn high-level representations based on low-level predictions. This technique was evaluated after its creation. They demonstrated the effectiveness of their strategy by providing precise weather predictions. Input to a low-frequency forecasting method is often supplied by a multivariate model with a delta function close to the value of the most recent strike. Cross-validation may be used to discover the best number of levels for a variable. We recommend adding an extra forty blankets on the bed for maximum comfort. Three unique time intervals are used to train the model: (1) training, including forecasting and monitoring, between January 1, 1981 and December 31, 2003; (2) validations, between January 1, 2004 and December 31, 2004; and (3) trials, between January 1, 2005 and December 31, 2005. Beginning on January 1, 1981, and concluding on December 31, 2003, the training period runs from 1981 to 2003. In order to do method comparisons in the confined space, the root of the RMSE cost function was used, and the resulting value was merely 24 meters. The Temporal Long-Short-Term Memory (LSTM) algorithm maintains a record of the network's data's temporal correlation. Wilson et al. [12] introduced a unique deep learning technique for weather forecasting in 2018 using two-weight convolutional graph LSTM WGC-LSTM.

This method was implemented in 2018. Considered crucial to the functionality of the model. Since training the learning methodology by applying the graph categorization based on the borders between the sites is inadequate, the presented method takes into consideration the graph's adjacency matrix as a data-learnable example. This is performed since teaching the method by categorizing graphs is inadequate. Due to the inadequacy of training the learning strategy through graph classification based on site boundaries, this is the case. It may be difficult to learn anything new by trial and error, and it may also be difficult to alter one's behavior after becoming aware of several hurdles. Both of these are difficult to do. By conducting an experiment using two distinct sets of real-world meteorological data, it was discovered that WGCLSTM performs better than the other reference methodologies. This was shown to be true in the majority of circumstances. Mehrkanon et al. [13] have created projections for 2019 in particular. We developed a unique method for weather forecasting by merging data from many weather stations with a convolutional neural network (CNN) model. For instance, the recently developed deep learning method employs 1D-, 2D-, and 3D-CNN optimization models. In order to give a more accurate portrayal of the environment, each of the climate models incorporates past and future readings from user-created weather stations into a single database. It has been shown that the consistency of the features that are learned from a variety of input levels and the incorporation of a feature-learning component have a major effect on the prediction performance. At a total of six different sites in the Netherlands and Belgium, two separate research projects will use the data available on the Weather Underground website. The just-described 3D-CNN model is often more accurate than other approaches for forecasting temperatures 1–10 days into the future. As such, Bose et al. According to his recommendation, a weather forecasting system based on deep learning should be implemented in 2019. He has just recently finished a large amount of research and data collection using an actual meteorological data set to show the potential of learning models for deep learning. It was an effort to show that learning models are successful in certain circumstances. This research seeks to enhance the representation of data in predictions and examine the impact of a lack of data on predicting. Three unique layers make up a neural network, which may be categorized as follows: The structure of the input layers consists of simply a single layer. Whenever we are forced to make judgments at the input level, the input matrix acts as a helpful mental tool. Due to the fact that the window size for this experiment is set to 12, The mathematical layers of a neural network are completely

invisible to a human researcher's naked sight. Each increasingly buried layer's nodes are connected to those of the layers immediately above and below it. In an attempt to improve the accuracy of weather forecasts, a sliding window approach using matrices was developed. In order to generate an accurate forecast for the month in question, weather data from the past 12 months, commencing with the month in question, are fed into a deep learning model. After that time, it will be allowed to transfer data for a whole year into the designated repository. This layer consists of one hundred nodes, and the Linear Unit ReLU activation function is what ties them all together. As seen by the gathered meteorological data, expanding the training set has the potential to boost the performance of the trained model. The accuracy of the trained model is not meaningfully constrained by the baseline data. Roy and his colleagues want to examine the use of deep learning models, which have the potential to surpass the capabilities of classic two-dimensional models, in order to enhance weather forecasting by the year 2020. Everything said will be carried out in full. During the course of our research, we used many various types of deep neural networks. These included the Multilayer Perceptual Neural Network (MLP), the Long-Term Memory Network (LSTM), and a CNN plus LSTM combo. Since January 1, 2009, we have received daily weather updates from John F. Kennedy International Airport in New York, and this will continue until January 1, 2019. The John F. Kennedy International Airport is the world's busiest airport. Accurate weather forecasts are vital for the safe operation of any airplanes entering North America. Overall, the complexity of the model adds to an improvement in the accuracy of the forecasts. The CNN+LSTM model produces the most accurate results when compared to other models. MAPE found that the MLP model had an accuracy of 89.15 percent, the LSTM model had an accuracy of 95.03 percent.

Czibula and his colleagues predict that [16]'s full functioning will be realized by 2020. NowDeepN is a model based on supervised learning that employs several deep neural networks that act in tandem. The authors recommend using this approach to produce real-time value forecasts with radar products. NowDeepN's patterns might possibly be used by meteorologists to foretell the future path of catastrophic catastrophes. The radar's results have been enhanced by a data purification method that eliminates any mathematical or meteorological flaws from the data. After the radar data was cleaned up, the RMSE for NowDeepN reduced to 5.34, a reduction of less than 4%. Using an Adaptive Multi-Attention and Residual Convolutional Neural Network, Yan et al. [17] suggested the construction of a

rainfall model by 2020. This model would be used to forecast future precipitation. In addition to radar signals situated at three separate heights, the grid also includes numerical information such as cloud movement speed. The objective of doing research on occurrences of severe weather is to better prepare communities for the impacts of adverse rainfall patterns. The proposed model was created using a minimal number of MAR-CNN nodes and is based on a wide range of various deep learning methodologies. Single-channel techniques incorporate a convolutional model, an attention module, a detector, and a GBDT. Now that we've reached this position, we may consider ourselves to have advanced up a rung. We use RMSA to demonstrate the predictive ability of our model, and the score we arrive at is 7.90. In 2021, Grünquist et al. [18] introduced the first hybrid model to the globe. This model employs deep neural network-based post-processing techniques in combination with a modified version of the conventional weather algorithm. This allows the model to account for nonlinear interactions, which are often ignored by conventional numerical and structural approaches. Have an impact on how knowledge is conveyed around the globe. The capability of the data processing pipeline to make previously inaccessible amounts of data readily available has resulted in a considerable decrease in training time. U-Net applications use both locally interconnected fabrics and residual neural networks with Inception modules. Using a mixed model to make CRPS predictions increases accuracy by around 14%. Hughey et al. [19] propose developing a data-driven, small-scale weather prediction model using LSTM and Temporal Convolutional Networks in 2021. (TCN). Utilizing a multi-level depth model is recommended when trying to provide precise weather predictions. This kind of model takes into account the frequency with which severe weather occurs over a certain period of time. On the depicted deep learning network, there are several inputs in addition to a number of outputs. In addition to LSTM and TCN research, we conduct TCN and LSTM studies with many inputs but a single output. TCN can map input sequences of any length to output sequences of the same length, similar to how RNN exploits increasing complexity to provide universal access, general acceptance, and the potential for effectiveness. TCN may also accept input sequences of arbitrary length. a great deal of past expertise in the area Extensive testing demonstrates that the proposed weighted model is more accurate at predicting the weather 12 hours in advance than the model that is typically applied and is rather complex.

## **1.6 THESIS LAYOUT**

In addition to the first chapter, which is enough to give a short introduction about the concerns of the thesis, and what is the main problem, and the purpose behind the proposed work, there are four other chapters as follows:

The second chapter: enough to provide the theoretical framework and the main ideas of this thesis.

The third chapter: concerns to provide the main aspects of the forecasting model, including algorithms and criteria.

Chapter four: shows the results of experiments conducted using the recommended approach.

Chapter Five: Present and present a set of conclusions and recommendations that can be applied to the project.

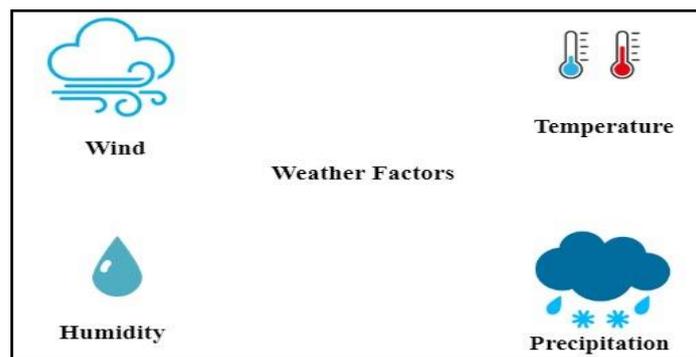
## 2. LITERATURE REVIEW

### 2.1 INTRODUCTION

Numerous experts are becoming more interested in climate forecasting due to its global relevance and potential influence on human existence. In this chapter, an introduction to what climate is and what causes it, as well as the challenges and applications of climate forecasting, is introduced. A description of the design process, machine description and deep learning concepts are presented in this chapter of the review. Also, a brief description of the types of performance metrics used to measure the success of climate prediction models.

### 2.2 WEATHER FORECASTING OVERVIEW

When we discuss the weather, we are talking to the condition of the atmosphere at a certain location and time. This condition may fluctuate hourly, daily, and seasonally. When we discuss the weather, we are talking to the exact state of the atmosphere. It is an experience that never ends, requires a great deal of hard work, and consists of many moving parts, complexity, and turmoil [20]. Time can be measured using a number of factors including Temperature, rain, humidity, and wind speed and direction, which can be expressed numerically or more in the form of names, that is, in general, have characteristic values that are names, but sometimes, because and they describe the details of time [21].



**Figure 2.1:** Weather Factors.

- a. Temperature: defines the amount of cold and temperature of a place or environment measured and expressed in degrees Fahrenheit F of Celsius © [22].

- b. Humidity: a refers to the quantity of moisture present in the air. The sensation of chilliness is created by both high humidity and high relative humidity; however, during the summer, the humidity is often low and causes just a little amount of chill.
- c. Rain: shows any condensation product of atmospheric water, which falls on the surface of the earth, which is a lot of rain, snow and cold that fell in one place in a specific time, it important for some. External events and some company activities depend on it, and it is used for forecasting, up to one year [24].
- d. "Wind" refers to the directional and/or velocimetric motion of air at or near the surface of the Earth as a consequence of pressure differentials in the horizontal and vertical directions. The term "wind speed" refers to the rate at which weather-related air is transported from one location to another. If you see where the wind is blowing from, you may determine its direction [25]. In the discipline of weather forecasting, science and technology are used to predict the weather at a specific location and at a certain time in the future [26]. Despite the fact that climate forecasting dates back to the 19th century [27], it wasn't until the 1970s and 1980s that numerical weather prediction (NWP) led to significant advancements. In order to do this, one must speculate about the future evolution of the climate. Numerous sources, including ground observers, ship observers, airplane observers, radio sounders, Doppler radar, and satellites, may provide information on the present condition of the atmosphere. The data is subsequently sent to the meteorological office, where it is collated, evaluated, and displayed in a number of ways based on specified criteria. Today's advanced computers transfer thousands of atmospheric observations to atmospheric maps [29]. Meteorology is the foundation of scientific forecasting, which revolves around predicting the weather conditions for a particular area. Predicting climate as humans do is an example of judgment in the face of uncertainty [30]. weather forecasting challenges.

### **2.2.1 Forecasting Weather Challenges**

- a. Handling Large Data Sets: The ongoing development of climate statements and models necessitates the use of existing data, and the number and diversity of environmental information is expanding considerably, placing enormous demands on the infrastructure that can exchange, handle, and store this data. informative quantity It is essential to offer a substantial amount of storage space for the output files that are generated by

monitoring. Predictions of the weather based on numerical analysis It is probable that period operations on NWP's data will need a substantial amount of processing power. In order to input the data and obtain a solution to the nonlinear Navier-Stokes equation, an abundance of data [31] is required. Having access to a big quantity of data over an extended period of time presents exciting organizational issues.

- b. **Access To Historical Data:** The overwhelming majority of ports lack full historical documentation. Due to the lack of a substantial amount of data at the time the forecast is to be provided, Nerdette observed that the limited historical records are inadequate to provide the necessary models for predicting extreme weather and severe weather. There are just a few of event-related systems that can function well with limited data access. The surface information has been damaged the most severely. As a consequence, a large number of ports have been either completely deactivated or partially implemented. To get a comprehensive knowledge of a system, one must start from a higher vantage point and go downward. It is well known, for example, that the equator and pole are produced by the geographic region storm line, which comes from the Intertropical Convergence Zone and, as a consequence, the geographic region jet at 700–600 hpa and the wing of the Japanese tropical jet at 100–150 hpa. This is how the equator and the pole are formed. Carlson and Fortune published in 1980 (1969) It is probable that constructing a map of rain bands in West Africa would need the modelling of these well-known possibilities in order to pinpoint their precise positions. In the current state of affairs, a lack of expertise in the area of high speed will provide a substantial impediment. Several of these aspects, however, have the potential to be improved via the use of distant data collecting through proxy satellites. Because satellite recordings are only kept for extremely brief periods of time, they will not be able to capture the nuances of specific storms in great detail. As a result of this information being processed [32].
- c. **Technical Barriers:** High accuracy fashions are a key requirement of climate forecasting for correct and complete forecasting. Therefore, thinking about the long-time period situations of the superior fashions of the weather gadget is important. In addition, many bodily and chemical strategies have been included into the version in order that we will take a look at the subtle traits of weather behaviour. A climate provider middle have to meet all the necessities for a vital meeting. The underlying technical infrastructure have to be included to aid the simulation and evaluation process. They have to are expecting

every day overall performance inside a slender strolling window; The middle calls for the maximum updated gadget information as soon as and plenty of important repayment to make sure they may be 100% to be had each day. The gadget is designed for maximum parts dealing with the climate station. They want sturdy software program that may get better from hardware disasters due to the fact the executions are long; And they'll manipulate and examine the big quantity of statistics that has been created thru many years of simulation [33]. However, those predictions are starting to go through in the place and hyper-place, in particular whilst the fast instability makes adjustments take place quicker than the weather fashions can are expecting. Therefore, the destiny of forecasting calls for abilities that we couldn't consider whilst we first began out forecasting.

- d. Availability Of Policy Types: Creating physical identity remains a major challenge for artists and limits the potential for prediction. These long-term weather phenomena are among the most advanced and demanding mathematical and scientific problems. The inclusion of fresh physics concepts into numerical prediction models has greatly enhanced the complexity of these models. Using multiple models, one for each of the figure's starting points, researchers study a vast array of climate-related issues. The task of the prophet is to explain and analyze the various things: a little ability is needed to provide information about the prototype, but it is possible to find the correct model, especially for this time, where the element is chosen [34]. So rapid growth is a continuing challenge for IT.
- e. Complexity: The mathematical models and algorithms used for weather forecasting have become more sophisticated throughout time. Consequently, these models and algorithms may now include enormous datasets into their projections. The terms "data volume," "data variety," "data speed," "data variability," and "data accuracy" highlight the difficulties associated with combining many data sources to provide real-time predictions. As an instance, observations improve projections in at least three ways. These include (but are not limited to): (1) educating artificial intelligence algorithms; (2) identifying present conditions to offer data for current forecasts; and (3) performing and publishing numerical weather predictions. After the forecast has been created, further data will be necessary for verification and validation [35], and it may be difficult to apply

many data reassessments appropriately for these reasons without compromising other factors.

- f. Price Transfer: Climate impacts individuals, corporations and the financial system each day. Inaccurate climate forecasts can cause regulations or even entire cessation of air and street travel. These networks are very touchy to terrible climate. Without the constraints set by Earth science, even a normal weather system has the ability to wreak havoc on people's, communities', and even nations' day-to-day operations at a great financial cost. As a result, the increasing complexity of it does not seem to be a highly valued commodity. The loss of reliable, cheap and predictable area has been a first-rate impediment to the implementation of NASA's Earth technology program. [36].

### **2.2.2 The Use of Software in Weather Prediction**

Numerous businesses have benefited from accurate weather predictions, including trade, tourism, sports, and the energy industry, to mention a few. In this part, we might discuss the effects that climate change has had on various locations.

- a. Company: Agriculture/Food Industry): Based on population boom, there should be boom in meals manufacturing to fulfill meals needs. With the assist of huge records for climate forecasting, this could be completed very well. Climate forecasting can assist farmers put together for droughts, floods and soil erosion [37]. Rainfall forecasting is essential on the subject of meals security, these forecasts assist in dealing with the transport of meals to regions wherein shortages are expected, and it allows farmers to make selections approximately their plants and prices of meals merchandise also can be estimated. Large supermarkets use weather predictions to plan their inventory management. They are aware that a winter with lower temperatures will result in increased revenue from the sale of soup components, while a summer with higher temperatures will result in an increase in the number of customers who are eager to buy those components. The planning of stock management at large supermarkets sometimes involves consulting weather predictions. A tasty delicacy that is refreshingly chilled and a barbecue [38].
- b. Tourism Industry: "Travel for fun and freedom" is one kind of tourism, but there are many other reasons why people travel, including to learn something new, to pay their loved ones a visit, for business, or even for their health. Because tourism is one of the

nation's businesses that is both rapidly expanding and substantial, his impact on the country's economy is significant. The World Travel and Tourism Council estimates that tourism contributed 9.8 percent of the global GDP during the 2015–2016 fiscal year [39]. The weather may have both positive and bad effects on tourists; thus, it is essential to monitor the forecast. Therefore, accurate climate forecasting is becoming more crucial for the safety and comfort of passengers, in addition to defending major tourist areas such as glaciers, coral reefs, and animals, the park also protects wildlife. Accurate weather forecasts may enable businesses and visitors alike to better manage their activities and itineraries [40]. This advantage is available to both parties.

- c. Construction Industry: Construction can have an effect on wind, temperature and humidity, likewise because the weather. Therefore, so as to defend workers' jobs and assets from the risks of the weather, the simplest offered data is needed. Weather forecasts can facilitate establish risks before they happen and you'll be able to set up around them, saving your cash and redundant period of time or injury or worse to workers. foretelling can help in building an honest house. MPC Model prognostic Control could be a key issue for implementing internet Zero Energy Buildings NZEB in cities. The wind gathers up and pushes the top higher than the plain because, when contemplating the use of cranes in construction, the underline the need of taking gas pressure into consideration. Therefore, it is impossible to evaluate if a surface may be used without first preparing it. Prediction is vital for functioning in an efficient and risk-free way. Flood forecasting ought to be taken seriously as they assist build property and resilient buildings for the future. Weather forecasts are important to organize for reconstruction if severe conditions are expected; the main points of serious rain should be taken into consideration in the construction of dams as a result of it will harm vehicles, instrumentation and cause several deaths [41].
- d. Sports industry: Since weather conditions such as lightning and rain may wreak havoc on sporting events, many organizations hire their own meteorological crew. Soon, a forecaster's broadcast will be required at a select number of significant events to shelter attendees from lightning strikes and limit the amount of time spent inside during recreational activities. A rain forecast at some point of the event might also additionally suggest whilst the courtroom docket may be covered. Small modifications in the wind

also can have a huge effect at the effects of sailing; Many Olympic sailing groups appoint their personal weathermen [42].

- e. Transportation: Global climate forecasts may be used to tell transport businesses and operators approximately the threat of storms, letting them determine whilst to apply ships and which routes to take [43]. In intense climate conditions, there may be a put off with inside the flight agenda, however we recognition on retaining the identical agenda as quickly as viable to lessen the flight time and stoppages and for that we want a few tools. Commonly, commercial pilots are shielded from dangerous situations such as poor weather, vision constraints, and periodic air strain. This is done in order to safeguard both life and property. Because drivers might suffer devastating consequences in the event of heavy rains and ice, roadways must be equipped with adequate detours depending on the projected weather conditions. Ariel's unmanned aerial vehicle UAV gives excessive manoeuvrability however faces harsh climate conditions. For a hit UAV missions, take into account climate forecasting at some point of threat assessment and task planning [44].
- f. Disaster Management: Even though natural disasters are unavoidable; it is feasible to save a substantial number of lives and reduce property damage and economic losses by sensible preparation and prevention. In order to develop precise predictions about the effectiveness of herbal medicines, extensive study is necessary. The National Oceanic and Atmospheric Administration utilizes cutting-edge computer technology and methods to enhance its capacity to predict storms such as Hurricane/Superstorm Sandy [45]. This makes it easier to plan for future natural catastrophe consequences. The process of forecasting the path of a storm requires collecting a significant amount of data, using high-precision models that account for the physics on a regional scale, and making precise measurements of the initial circumstances. Despite the fact that the forecast's accuracy and delay rely on the kind of risk. In addition to storms, natural disasters such as landslides and floods may inflict catastrophic injury or even death to humans. Tornadoes have killed many human beings within the past [46].
- g. Power: Due to the increasing relevance of the use of renewable energy, the length of time spent amassing the greatest possible quantity of energy from renewable power reasserts is presently the single most important factor. Wind and solar power supply the bulk of the world's energy; but, due to their great reliance on the weather, accurate

weather forecasting is essential for meeting energy demands from these sources. In particular, many "clever grid" techniques and contingency control are primarily based totally on short-time period forecasts of renewable power generation, which additionally require correct forecasts of wind speed, wind coverage, temperature and different comparable variables. Therefore short-time period climate forecasting permits the green integration of renewables into destiny clever power systems [47]. Electric groups could make heaps of greenbacks primarily based totally on one wintry weather climate forecast, as it lets in them to shop for gas at a decrease charge earlier than call for will increase in bloodless climate. In addition, climate elements frequently have an effect on load forecasting, wherein using energy by residential purchasers varies with the climate, in order that the weight profile will extrude as an end result impact of time [48].

### **2.3 DATA NORMALIZATION**

Normalization describes the process used to organize the characteristics of the data specified in the model to increase the correlation of the characteristic types, and also helps to simplify the data. Organized data and high-quality standards reduce the chance of incorrect data, such as keeping information in one place. Combined normalization performed in two steps; re-scaling to make classification easier, and to make the similarity of values in the form easier despite the value being compared at different scales. A modification of this kind is made to the dataset whenever a request to search for interaction is made. To normalize a dataset, one must edit the values of the columns that include integers and normal numbers without altering the difference between the two data types. Using a normalizing characteristic that is related to scaling may improve one's understanding of time. Performance and is considered an important step in the design of multidimensional datasets. There are a large number of data normalizations [50]:

- a. **Min-Max Scaler:** Min-Max Scalar normalization is a prominent data processing method that is often used. The strategy is implemented by first applying a linear transformation to the raw data, and then applying a scaler function on the difference between the old values of some feature or output and the newly computed values of that feature or output. This is one of the most used data processing methods. Typically, estimates are evaluated on a scale ranging from best to worst using a value between 0 and 1 (or -1 and -1) as the scale's value. When the min-max method is used, you are

moved into a new category, but none of the data-related characteristics that comprise your profile are altered [51]. When the feature has a constant value in the determined data, it will be dismissed, because it does not provide any information, and Min-max normalization has the advantage of maintaining a good relationship with the data. computed as follow [52]:

$$X = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (2.1)$$

Where  $X_{min}$  denote the minimum value in X Feature, and  $X_{max}$  is the maximum value in X Feature.

- b. Standard Scaler: It is considered to be a featuring scaling technique that normalize all the features by eliminating its mean, and perform variance scaling into one count. The computation of a normalized value based only on the mean and variance has a number of benefits, including linearity, rapidity, reversibility, and scalability. Among the other advantages is the capability to scale to a substantial degree. This approach has a number of drawbacks, the most significant of which is that it is too sensitive to outliers and works best when applied to equally distributed data. a given component, its mean and standard deviation must first be provided. a Before you can scale a particular component using [53], you must first supply the mean and standard deviation of that component.

$$X = \frac{X - \mu}{\sigma} \quad (2.2)$$

For the mean value, we have and for the standard deviation, we have

- c. Maximum Absolute MA Scaler: It is a technique for normalizing feature values that divides each instance by the feature's absolute maximum value. Since it is linear, reversible, and extensible, the MA scaler is able to process dispersed data. Nevertheless, MA Scaler has some sensitivity to outliers, and every feature has a non-zero mean, indicating that it would be necessary to include a bias factor in the models. X MA is the absolute maximum number of  $X_i$  observations that may be included in the X sample, every normalized observation  $X_i$  can be determined as follow [54]:

$$X = \frac{X_i}{X_{MA}} \quad (2.3)$$

- d. **Quantile Transform:** To scale each feature contained inside an implementation of an inverse cumulative distribution function, it is necessary to first get quantile information. This information is necessary for scaling each feature. Even though it has the potential to be a robust approach against marginal outliers, it is non-linear and the information has low reversibility. Nevertheless, QT may have performance difficulties that must be considered. It has trouble dealing with sparse data, produces a non-sparse output, and has the ability to distort the linear liaisons between various types of samples that belong to the same feature. If the following requirements are met, it is feasible to compute the Quantile Transform, abbreviated QT: (a) the samples must be ordered so that it is possible to estimate their cumulative distribution function (CDF); and (b) the QT for the samples in the input ranges must be equal to its CDF. (The QT can be calculated using the CDF's boundary conditions even if the samples originate from beyond the input domain.
- e. **Z-Score:** After computing the Mean and Standard Deviation of each feature using a set of training data, which helps us understand how much each feature deviates from the Mean, we may normalize the data vector by modifying the values of each feature. This enables us to assess how much each characteristic deviates from the Mean. By comparing two distinct sets of data side-by-side, one may determine whether or not they are similar. The normalized data are marked by  $X'$ , the input variable is indicated by  $x_i$ , the arithmetic mean is denoted by  $\mu$ , and the standard deviation is similarly denoted by  $\sigma$ . Using this method to generate features will result in features with a mean of zero and a standard deviation of one [56]. The chosen dataset's feature vectors must be normalized prior to further processing. As a design metric, we are now calculating and storing the mean and standard deviation of each training data feature. This data will be used in the future, and can be obtained as follow:

$$X' = \frac{x_i - \mu_i}{\sigma_i} \quad (2.4)$$

- f. **Median Method:** The median strategy entails identifying the value at which each input is most often seen and then applying that value to each sample. This method is good for learning the features of the distribution or computing the ratio of two samples in a hybrid term [57], and it is also very resistant to a wide range of modifications. The approach is quite robust. The number of outliers and the median have no relationship. There may be

additional benefits associated with a well-managed distribution system, and can be obtained as follow:

$$X = \frac{X_i}{\text{mediana}_i} \quad (2.5)$$

## 2.4 AI AND MACHINE LEARNING

In 1990, Warren McCulloch and Walter Pitts introduced a sub-situational method to time series forecasting using what they termed an artificial neural network. This technique was created by the couple (ANN). There is a need for a solution to complex classification and regression problems [58] since AI has been successfully used in a vast array of businesses. In numerous subfields of artificial intelligence, including automated recognition, classification, prediction, regression, and autonomous diagnosis, as well as in other disciplines, machine learning has been used to generate discoveries that have garnered significant attention. It is feasible to use machine learning (ML) to automatically extract pertinent facts from a given data set. In this instance, machine learning provides accurate predictions based on historical data and human intuition. This is especially advantageous owing to the fact that individuals are prone to make mistakes during data analysis or when trying to establish connections between a large number of different types of qualities. [59]. Each dataset characteristic used by machine learning algorithms [60] falls into one of three categories: continuous, categorical, and binary. There are three major classifications of machine learning algorithms: supervised, unsupervised, and semi-supervised. The method in which the algorithms interact with a certain data collection comprised of instances and attributes differentiates these groups.

- a. ML Supervised: It is a kind of Machine Learning technique that enables a machine to learn a desired function from a set of instances. Examples are provided to educate the function to the system. The bulk of training data consists of paired input items, which are often vectors, and their desired outputs. The outcomes of the function may have a continuous value named as regression, or can predict a determined class of label of the entered and used object named as classification. The supervised learner operates by predicting the count of the function for any valid entered object after obtaining a set of training samples, it is also utilized examples that are labelled X and Y to give a

prediction about their relationship to attain it in both classification, and regression [61]. Supervised learning suppose that all training instances are labelled in advance.

- b. **ML Unsupervised:** Human-supplied labels are discarded in this kind of machine learning. In contrast to supervised approaches, unsupervised learning does not need an instructor to lead students through the processes required to accomplish a task like as classification or regression using a small, well selected data set. Instead, unsupervised learning use a well picked data set. Obtaining the Xs is all that is necessary in order to get the finest possible performance feedback function. In an unsupervised learning environment, it is more difficult to get information on the distribution of unlabelled samples; yet, such information may be useful for compression, clustering, feature extraction, etc. In addition to its many other benefits, one of its primary advantages is that it does not need an expert or a review of the data to classify occurrences. As a consequence, it is more objective and can work effectively even without any prior information. [62].
- c. **Semi-Supervised Machine Learning:** It attains popularity in the last few years, and operate by assuming that both labeled and unlabelled training instances are existed, and considered to be the extension of both supervised, and unsupervised learning approaches, due to the way that is work. When there is a substantial volume of untagged data, it is employed. It is vital to gather labeled data to increase one's capacity to distinguish between data points belonging to distinct categories. In contrast, the primary goal of collecting unlabeled data is to expose the region's underlying geometrical structure. This is possible by analyzing the data without labels. Despite the substantial cost investment and time commitment required for data categorization, semi-supervised machine learning algorithms are advantageous in a variety of settings [63]. It is not only possible but also often useful to utilize unlabelled data to impact the learning process.

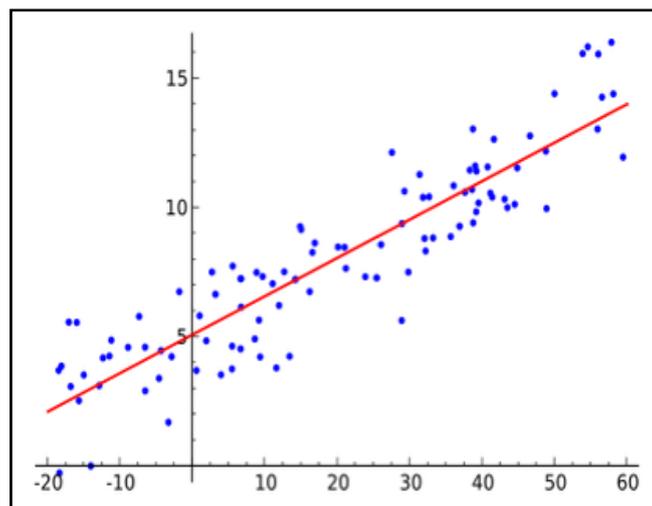
#### **2.4.1 Machine Learning Linear Regression Algorithm**

Generally, it's one of the plainest, and most popular algorithms of machine learning. It was first suggested by Francis Galton [64], and considered as a mathematical technique attained in order to do a predictive analysis, which permits projections of the mathematical variables whether its continuous, or real kinds. The Linear regression is utilized as a common approach in the mathematical research fields, that allows to measure the prognosticated effects, and

mapping them to many entered variables, in which it is an approach for data evaluation, that model these determination relationships that are linear between the intended variables whether they are dependent, or independent. This method would represent the relationships that exist between variables that are independent, and dependent by analysing, and learning procedures to the obtained training results. The goal of regression analysis defined in predicting a continuous intended variable, in which in another area named as classification that is predict the label from a set that is restricted. Linear regression is an example of supervised learning, which is a collection of approaches that need a dataset containing labels to "train" a model, which is then used to predict labels in a separate dataset, known as "testing" [65]. Among the methods for supervised learning is linear regression. By consuming the variable  $y$  that is dependent, and the variable  $x_i$  which is independent, in which  $x_i, i=1,2, 3$ , that will affect on  $y$ . the trivial linear regression is defined as follow [66].

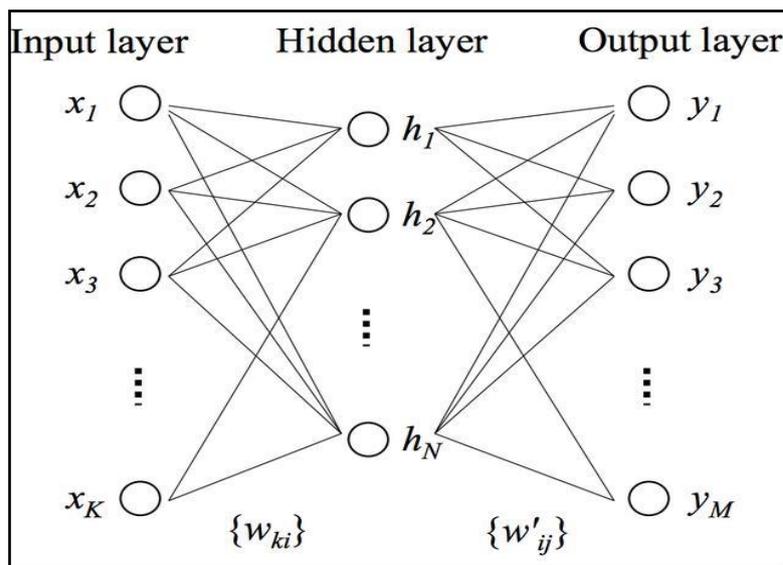
$$y = \alpha_0 + \alpha_1 x_1 + e \quad (2.6)$$

Where  $y$  represents the variable within a dependent type, and  $x$  is the independent one.  $\alpha_0$  is the constant parameter, which represents the cross line of the regression that lay on the vertical axis. the  $\alpha_1$  is the coefficient regression that is the slope degree of the line of. As an example of how random parameters might possibly influence a dependent variable, we will utilize the random error value, which will be denoted by the letter.



**Figure 2.2:** The Visual Representation of The Linear Regression [67].

a. Neural Network (NN): The neural network is of the grid-based and non-linear kind, which have the ability to perform curve fitting of type non-linear, by representing the paradigm belonging to information processing, which is basically inspired by the way of the humans' brain biological behavior, that have the ability to process a specific part of information in every time. It has the ability to perform nonlinear mathematical function on the entries [68], in which the entries may be the outcomes of other kinds of network elements, or may refer to the intended network entries. Basically, the NN is parallel in wide range, within processor that is highly distributed, which is composed of plain units of processing. It normal behavior have an innate inclination to keep the experiential knowledge and make it obtainable for utilization, in which its looks like the brain of the human beings in two manners, first its gain its knowledge within a learning procedure from surrounded environment, second the power of connections between Inter-neuron, noted as synaptic weights, that are obtained in order to save the obtained knowledge. mainly, the building unit of the NN is the neuron, which is named as a node of the network that in general conserves either a function used in calculation or values that are resulted, while the edges linking the mentioned nodes are named as weights [69]. Figure 2.3 is a representation of a basic artificial neural network. This structure consists of three separate layers, which have been designated as input, concealed, and output, respectively.

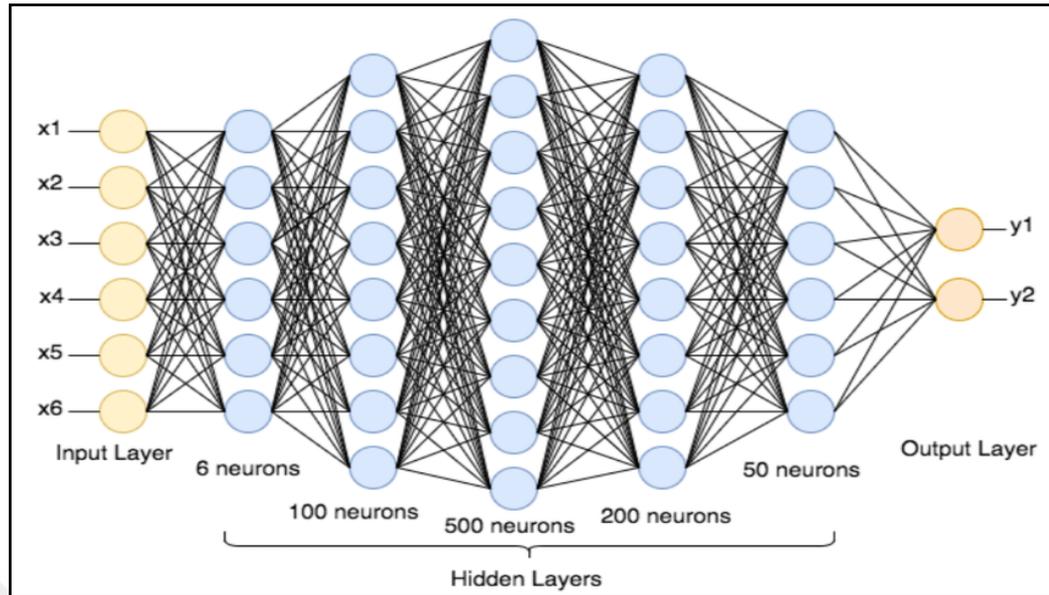


**Figure 2.3:** The Simple Neural Network Structure.

## 2.5 DEEP LEARNING DL OVERVIEW

Deep learning is a technology that enables computers to learn in a manner akin to how people learn, using complex mathematical formulae and massive amounts of data to reach definitive conclusions. The educational use of deep neural networks is gaining prominence. The neural network models that come under the heading of "deep learning" [70] consist of an input layer, a configurable number of hidden layers, and an output layer. These components comprise this tiered structure: Each layer of the human brain contains neurons with comparable physical and cognitive properties. The past three years have seen unprecedented success [71]. Figure 2.4 depicts the major contrast between a typical neural network and a deep learning neural network. One kind of network is known as a "stacked neural network," and it is often referred to as "deep learning" since it consists of several layers (DL). Taking the construction of a building into account, and the methods are meant to be used, the deep learning architecture can be classified into three kinds including [72]:

- a. Generative architectures, which are determined in order to differentiate the correlation that have superior-order of the properties belonging to the intended data or joint statistical apportionments of the visible data and their related classes. The utilization of the Bayes rule will convert this kind of architecture into a discriminative kind.
- b. Discriminative architectures, are obtained to give a discriminative potentiality for pattern classification, within characterizing the rearward distributions of classes that are exists and depends on the obvious data.
- c. Hybrid architectures, where the aim of it is differentiation along within assessment often in an important manner with the outputs of the generative architectures within better improvements or/and adjustments.



**Figure 2.4:** Deep Neural Network Architecture.

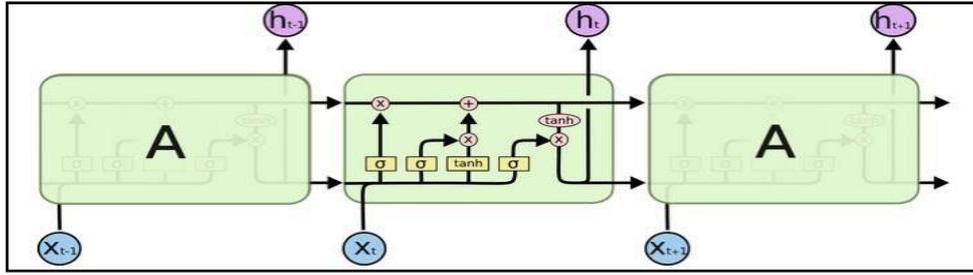
Multiple researchers provide a highly succeeded result in the domain of image classification, predictions of stock market, spontaneous recognition of speech, and face, processing of naturalistic language, due to it provide an illustration model that can learn from a set of features to make its own the ability to find many levels of properties in automatic way, as well as it doesn't require a lot of engineering procedures, so it can maximize the amount of data that can be obtained [73]. Whenever the model architecture has a high complexity, it will have resulted in better performance results, rather than superficial learning especially when trying to solve a complicated problem. The DL in general have multiple advantages that can be which are illustrated briefly as follow [74]:

- a. The DL can detect the properties of the non-linear data easily, due to vast amount of hidden layers that have in its structure, also have the ability to perform analyze procedure in a powerful manner on the internal correlation that exists between entered vector, and related outcome.
- b. The DL is capable to learn in an easy manner using the extracted features, in which the information belonging to the features are transformed between the multiple number of layers that have to show a new feature domain
- c. Due to the use of a significant quantity of data for training the necessary train model, deep learning is able to extract and assign a large number of important relevant characteristics from data. This is made feasible by the usage of vast amounts of data.

Convolutional Neural Network, Autoencoder, Restricted Boltzmann Machine (RBM), and Long Short-Term Memory (LSTM) are the most widely used deep learning neural network architectures [75]. Nonetheless, a significant number of different deep learning algorithms operating on a variety of underlying architectures have been developed and made accessible for use in a vast array of application domains.

### **2.5.1 Long Short-Term Memory LSTM**

Hochreiter and Schmidhuber published the first LSTM [76]. It is a robust Recurrent Neural system developed to solve difficulties with vanishing gradients that emerge during learning with long-term dependencies, especially when the minimum time periods are too lengthy. This was achieved by shortening the minimum time periods. This was not possible prior to the publishing of the very first LSTM. Memory "blocks" are basically extremely tiny networks that are interconnected in a repetitive pattern to form the larger network. A memory block is a collection of cells capable of storing data for a defined amount of time. These cells collaborate to construct the block. In addition, three gates govern the flow of information between neurons and the memory cells associated with each neuron. These circuit components are known by their individual names, such as the input gate, output gate, and forget gate. At each of the LSTM's gates, the data received from the input neuron is replicated. In addition, each of them is related to an activation function. The LST is not like the normal recurring unit, which overwrites its tone periodically. Instead, the LST composes the tenor. By choosing the gates, the LSTM unit may choose whether or not to retain the present memory. For example, if the input gate indicates a maximum activation, the input is stored in the memory cell. The memory will be purged if the output gate announces a minimal activation. If the output gate of the memory cell detects this high level of activity, it will transfer previously stored information to following neurons [77]. The memory cell will be rewritten, however, if the forget gate detects the appropriate amount of activity. Figure 2.5 demonstrates the development of an LSTM, which reveals that instead of a single layer of neural network, it has a structure with four intriguing linking layers [78].



**Figure 2.5:** Long Short-Term Memory Structure.

Each line in Figure 2.5 is an example of a vector that was planned. These vectors link the network output of one node to the network inputs of other nodes. In contrast to the yellow boxes, which represent forward operations such as vector addition, the pink circles represent the learned layers of the neural network. An example of a forward operation is the addition of vectors. The connected lines indicated that a string was being concatenated, whilst the forked lines revealed that the string's contents were being reproduced and sent to many destinations simultaneously. Additionally, the vertical line that goes from top to bottom and the individual cell states of the LSTMs have a significant role. This behavior is depending on the degree of control exerted over the gates. LSTMs may either delete or add information to the state of a cell. Before creating or purchasing an LSTM network, it is essential to define the types of data that may be erased from the cell without damaging its integrity [79]. Before any work on constructing or acquiring an LSTM network can begin, this step must be accomplished. The sigmoid function computes the method for recognizing and rejecting data after obtaining the previous LSTM unit's output at time  $t-1$ , represented by  $h_{t-1}$ , and the current input at time  $t$ , represented by  $X_t$ . At time  $t$ , these two pieces of information are received. In addition, the sigmoid function employs the forget gate  $f_t$ , where  $f_t$  is a vector with values between 0 and 1 associated with each value in the cell state  $C_{t-1}$ , to decide what proportion and in what sequence of the previous output should be eliminated. This is achieved by combining the sigmoid function with the forget gate function  $f_t$ .

$$f_t = \sigma W_f[h_{t-1}, X_t] + b_f \quad (2.7)$$

Here,  $\sigma$  represents the sigmoid function,  $W_f$  and  $b_f$  represent the weight matrices of the forget gate, and  $b_f$  signifies the forget gate's bias.

Before a decision can be made, the information from the new input  $X_t$  can be assimilated into the present state of the cell, and the alterations can be maintained, the subsequent actions

must be executed. This structure consists of both the sigmoid layer and the tanh layer. First, the sigmoid layer determines whether the new data should be utilized to update the model or disregarded (a value of 0 or 1), and then it utilizes the tanh function to give a weight between -1 and 1 to the values that have been propagated according on their relevance. The range of weights is from -1 to 1. Multiplying the two functions together and adding this new memory to the more established memory  $C(t-1)$  yields the value  $C_t$ , which represents the most recent cellular state.

$$i_t = \sigma W_i[h_{t-1}, X_t] + b_i \quad (2.8)$$

$$N_t = \tanh W_n[h_{t-1}, X_t] + b_n \quad (2.9)$$

$$C_t = C_{t-1}f_t + N_t i_t \quad (2.10)$$

$C(t-(t-1))$  and  $C_{t-1}$  are the states of the cell at those times;  $W$  and  $b$  are the cell's weight matrices and bias; and  $t$  is the current time. where  $C(t-(t-1))$  and  $C_{t-1}$  represent the cell's states at those periods.

The state of the filtered cell, represented by the symbol  $O_t$ , ultimately determines the values of the output  $h_t$ . The sigmoid layer is responsible for determining which parts of the cell's condition should be collected first. Based on the current state of the cell, the tanh layer generates a new number between -1 and 1 called  $C_t$ , which is then multiplied by the output of the sigmoid gate, which is denoted by  $O_t$  [80].

$$O_t = \sigma W_o[h_{t-1}, X_t] + b_o \quad (2.11)$$

$$h_t = O_t \tanh C_t \quad (2.12)$$

where,  $W_o$  and  $b_o$  are the weight matrices and bias, in a sequence, belonging to output gate

## 2.6 PERFORMANCE MEASUREMENTS

These measurements have been collected only once for the purpose of analyzing the suggested model. There are a range of measures used to assess the performance of a model. In addition to the mean absolute error, the root mean squared error, or RMSE, is calculated (MAE), as follow:

- a. Mean Absolute Error (MAE): Calculates the difference between two continuous variables by comparing and computing the outcomes of two comparisons. If two observations have

the same original sin and X and Y are viewed as two independent variables, the observations are a pair. Using a gauge in place of several other metrics is an example of Y being used in place of X. Other instances include the contrast between prophecy and observation, later and earlier periods, and substitute measurements. Imagine that a scatter plot, I, has n points and their associated coordinates. Let's also assume that I exists (xi, yi). Mean absolute error is a statistical metric that measures the distance between each data point and the Y=X line, commonly known as the line of One-to-One correspondence. You must be acquainted with this background material in order to comprehend the MAE in its entirety. In addition to that MAE is the mean distance that is horizontal between every one of the points and the Y=X line:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |e_i| \quad (2.13)$$

Supposing that is already have an n samples of model errors which are calculated like  $e_i$ ,  $i = 1, 2, \dots, n$  [81].

- b. Mean Square Error MSE: The difference between a group's mean values and its standard deviation, also called the Mean Squared Deviation. MSD, or mean squared error, is a statistic that may be used to assess the accuracy of an estimation method while dealing with an unknown variable. MSE will always have a positive value since Error is squared; nonetheless, values with a smaller absolute difference are desired. Mean Squared Error (MSE) is the second moment of error in the graph of a function. It provides us the capacity to simultaneously account for both the variation and bias of the estimator. Variance refers to the degree to which estimates vary from one data sample to the next (how far from the mean we are) [82]:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n X_i - Y_i^2 \quad (2.14)$$

Where  $X_i$  denotes to the original sample data, and  $Y_i$  referred to the processed one

- c. Root Mean Squared Error (RMSE): The square root of the standard deviation is applied to the data in order to get the mean square error. Calculating the root-mean-squared error (RMSE), also known as the sum of the squares of the differences between the numbers predicted by the hypothetical model and the actual values, is one way to assess the degree of congruence between the observed data and the projected model. Calculating the

average squared difference between the actual values and the values predicted by the model yields the root mean square error (RMSE). The root-mean-squared error is often used to evaluate the "fitness of fit" of the general regression model (RMSE). In the application of regression systems, except the relationship is perfect, the predicted number, are more or less variant from the real defined observations. These differentiations are referred to the prediction errors or superfluties. These superfluties are measured by computing the vertical distances between the real intended values, and the line of regression. Large distances mean large amount of errors [83]. As a result, the RMSE is the average distance that is vertical of the intended real data points, from the appropriate line., and can be computed as follow:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n X_i - Y_i^2} \quad (2.15)$$

- d. Mean Absolute Percentage Error (MAPE :To get the outcome, the total absolute inaccuracy throughout all time periods is divided by the values that were reported to be present during the time in question. The next step is to do the appropriate calculations to get the mean of these constant percentages. This strategy is crucial because the size of the prediction variable affects the degree of inaccuracy when comparing the predicted value to the original value [84]. The MAPE is calculated over the first N periods of a single series, and its principal use is projecting future values. The formula for the anticipated error is Time t = et = At - Ft.

Where:

$A_t$  = represent the real observation at time  $t$

$F_t$  = *represnt*the forecast intended for period  $t$ .

$$\text{The percentage error} = \frac{A_t - F_t}{A_t} \times 100 \quad (2.16)$$

As a result, the MAPE for period t is calculated as follow:

$$(APE_t) = \left| \frac{A_t - F_t}{A_t} \right| \times 100 \quad (2.17)$$

$$MAPE = \sum_{t=1}^N APE_t / N \quad (2.18)$$

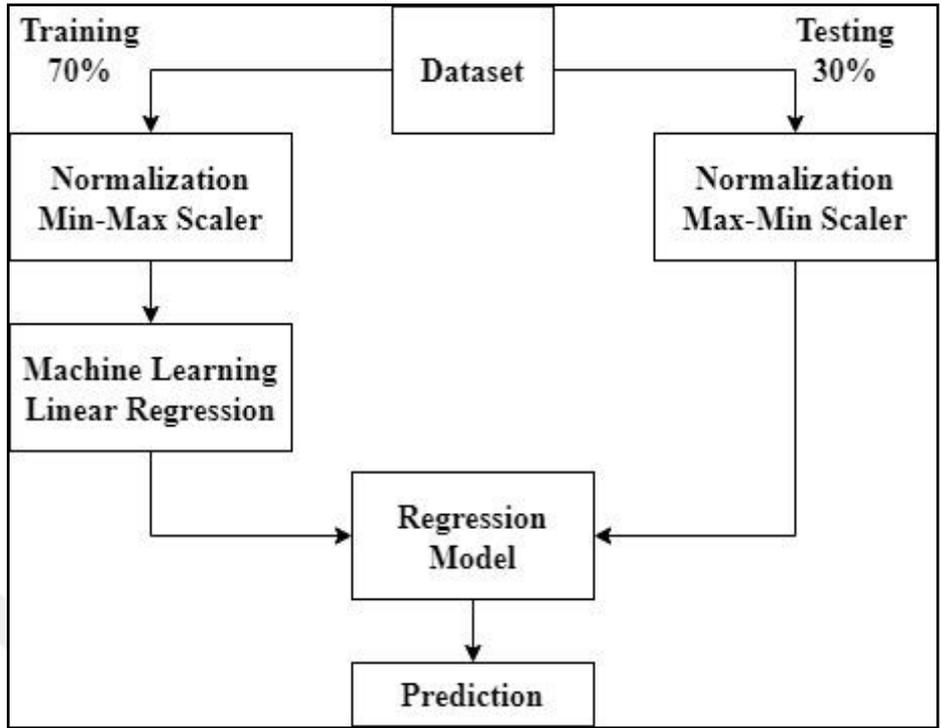
### **3. MATERIALS AND METHODS**

#### **3.1 INTRODUCTION**

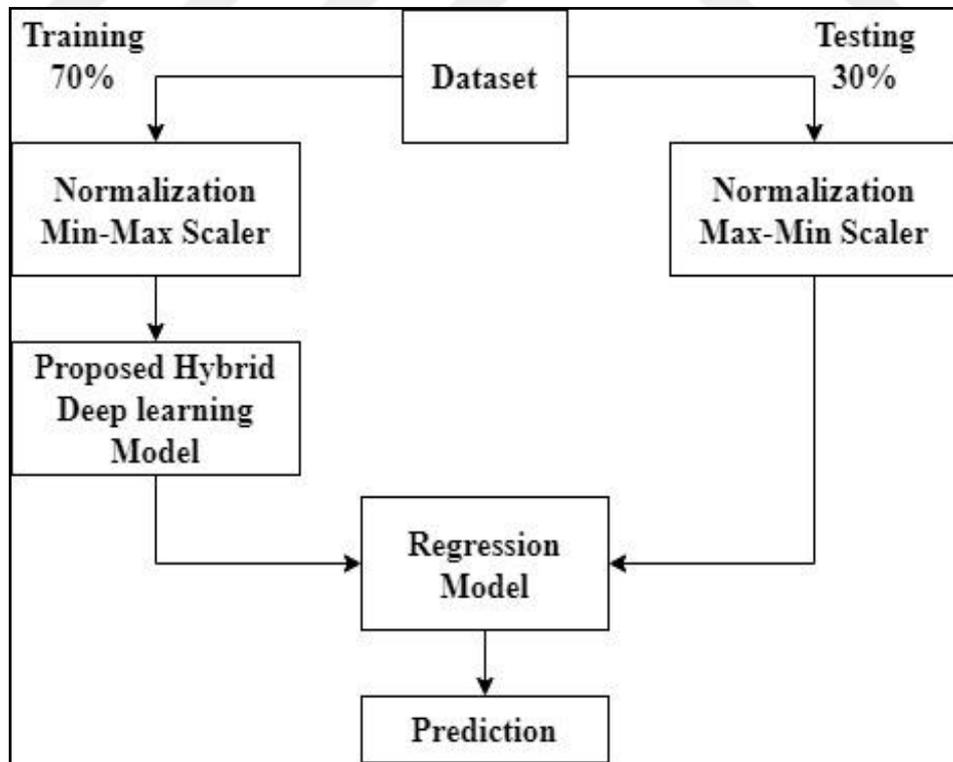
Nowadays, forecasting weather considered to a significant matter in multiple fields like economic, electricity load, and social expansion, and mainly a lot of things in the life is depending on the weather conditions. o weather forecasting gives a future tendency, and patterns of the weather conditions and changes may happen in the near and far future in order to take into account all the necessary operations to face these changes in the weather. in this chapter all the operations, this describes not just the suggested model for weather forecasting, but also its underlying methodology. This section provides not only an introduction to the proposed deep hybrid model, but also a detailed explanation of the dataset utilized in the process of producing weather predictions.

#### **3.2 PROPOSED MODEL DESIGN AND IMPLEMENTATION**

The proposed model design considered as an important operation, in which it must illustrated the in an applicable manner, because of its significant effect when implementing it in the real-world applications. The proposed system aims to perform a weather changes forecasting, within the use of the deep learning concept, in which a proposed model, that based on the deep hybrid component including LSTM along with Dense layers, is used in order to provide a data regression on the intended dataset, that contain a description about some weather condition changes. The system will go along through two phases after splitting the data into two groups including; 30% is Testing, and 70% is the Training. Data normalization is performed through the use of Max-Min normalization technique, in which all the dataset elements are first of all normalized before getting into the regression phase. The data regression phase accomplished using two techniques, at first the machine learning linear regression algorithm is used, and second a proposed deep hybrid model is implemented, As shown in Figure 3.1, and 3.2 sequentially.



**Figure 3.1:** The Block Diagram of Weather Forecasting Within Machine Linear Regression Algorithm.



**Figure 3.2:** The Block Diagram of Weather Forecasting Using Proposed Deep Hybrid Model.

In order to demonstrate the utility of the proposed deep learning model, the mix deep learning with more traditional machine learning on the same data set. In the next section labelled "Algorithm," the components of an accurate weather prediction algorithm are described in further detail.

**Algorithm 3.1:** The Proposed Weather Forecasting System.

**Input:** weather dataset

**Output:** Prediction result

**Begin**

- A. The data set should be divided in half, with 70% used for training and 30% for testing.
- B. Normalize the two groups values using the Min-Max scaler normalization technique
- C. Use the testing group as an input to the regression technique machine learning, deep hybrid model to build a regression model
- D. Use the testing group values as an entry to the regression model for testing the success of the proposed system
- E. Return a prediction result

**End**

### 3.3 DATASET DESCRIPTION

[85] provides an example of the JENA climate 2009–2016 dataset that would be used for the planned weather forecasting. This section will provide that example. The Meteorological Station in Jena, Germany, which is part of the Max Planck Institute for Biogeochemistry, has been collecting local climatic data for quite some time. Over the course of many years, 14 distinct variables were observed and recorded at 10-minute intervals for the Jena Climate dataset. Such components include the wind direction, air temperature, barometric pressure, and relative humidity. This set of data includes every occurrence that occurred between January 1, 2009 and December 31, 2016. Below the names of each column are detailed descriptions of the data structure and values associated with that column.

**Table 3.1:** The Datasets Quantities Description.

Index	Features	Format	Description
1	Date Time	02.02.2009 00:11:00	Date and time indication
2	p mbar	997.52	Internal pressure is often measured in pascals, a SI-derived unit of pressure. The millibar is the standard unit of measurement for atmospheric pressure in meteorological reporting.
3	T degC	-8.03	Temperature in Celsius
4	Tpot K	265.5	Temperature in Kelvin
5	Tdew degC	-8.7	Temperature in Celsius relative to humidity. Dew Point is a measure of the absolute amount of water in the air, the DP is the temperature at which the air cannot hold all the moisture in it and water condenses.
6	rh %	93.4	Relative Humidity is a measure of how saturated the air is with water vapor, the %RH determines the amount of water contained within collection objects.
7	VPmax mbar	3.34	Saturation vapor pressure
8	VPact mbar	3.12	Vapor pressure
9	VPdef mbar	0.23	Vapor pressure deficit
10	sh g/kg	1.95	Specific humidity
11	H2OC mmol/mol	3.13	Water vapor concentration
12	rho g/m ** 3	1307.76	Airtight
13	wv m/s	1.04	Wind speed
14	max. wv m/s	1.76	Maximum wind speed
15	wd deg	152.3	Wind direction in degrees

### 3.4 DATASET NORMALIZATION PHASE

Converting the form of the data, like normalization may ameliorate the data mining approaches in both terms of accuracy, and efficiency including; Neural Networks, clustering, classifiers, nearest neighbor, etc. the mentioned approached give a result in a better form, especially if the normalization operation implemented on the intended data to be analyzed. The training of the Neural Network may have more efficiency, assuming that both network entries and targets are subjected to a certain preprocessing technique. Normalization of raw data inputs has a major impact on the data and plays a crucial part in making the data acceptable for use in training. Due to the lack of consistency in the data, training neural networks without the normalization strategy, which is used to normalize the input, would result in a decrease in training speed. Algorithm 3.2 illustrates the Min-Max normalization method that was used to prepare the forecasting dataset for examination.

#### Algorithm 3.2: Min-Max Normalization

**Input:** weather forecasting dataset element vector **S**

**Output:** normalized elements vector **N**

**Begin**

Initialize  $X_{min}$  = minimum value of features in vector **S**

Initialize  $X_{max}$  = maximum value of features in vector **S**

**For** each feature **x** in vector **S** **do**

    Apply the normalization Min-Max scaler

$$X = \frac{X - X_{min}}{X_{max} - X_{min}}$$

    Add the normalized value to the vector **N**

**N.** add  $X$

**End For**

Return **N**

**End**

### 3.5 LINEAR REGRESSION MACHINE LEARNING ALGORITHM

The core principles, methods, and algorithmic foundations of machine learning have proven effective in a broad range of contexts. The technique of linear regression, which serves the

statistical purpose of creating a linear relationship between a number of predictions, is widely used and quite elementary in its application to machine learning. In this research project, a weather forecast is generated using a linear regression machine learning technique that takes use of the dataset's standardized values. The linear regression approach is shown with an example in Algorithm 3.3.

<b>Algorithm 3.4: Linear Regression</b>
<p><b>Begin</b></p> <p>1: Data Set Size n, Rows I Columns n Read:</p> <p style="padding-left: 20px;">Read <math>x_i</math> and <math>y_i</math></p> <p style="padding-left: 20px;">Next i</p> <p>End for</p> <p>2: Initialize:</p> <p style="padding-left: 20px;"><math>sum_x = 0</math></p> <p style="padding-left: 20px;"><math>sum_{x^2} = 0</math></p> <p style="padding-left: 20px;"><math>sum_y = 0</math></p> <p style="padding-left: 20px;"><math>sum_{xy} = 0</math></p> <p>3: Compute the intended Summations</p> <p style="padding-left: 20px;">For i=1 to n:</p> <p style="padding-left: 40px;"><math>sum_x = sum_x + x_i</math></p> <p style="padding-left: 40px;"><math>sum_{x^2} = sum_{x^2} + x_i^2</math></p> <p style="padding-left: 40px;"><math>sum_y = sum_y + y_i</math></p> <p style="padding-left: 40px;"><math>sum_{xy} = sum_{xy} + x_i * y_i</math></p> <p style="padding-left: 20px;"><b>End for</b></p> <p>4: Compute the interception of the <b>J</b> line, and the slop of the line <b>K</b></p> <p style="padding-left: 20px;"><math>K = (n * sum_{xy} - sum_x * sum_y) / (n * sum_{x^2} - sum_x * sum_x)</math></p> <p style="padding-left: 20px;"><math>J = (sum_y - K * sum_x) / n</math></p> <p>5: Compute the dependent variable count</p> <p style="padding-left: 20px;"><math>y = J + K * x</math></p> <p><b>End</b></p>

### **3.6 REGRESSION BY PROPOSED HYBRID DEEP LEARNING MODEL**

The Neural Networks methods are the most prevalent in the models of Artificial Intelligence AI, that utilized in order to give a solution for the problems of time series that are nonlinear. furthermore, if there are variety of hidden layers, the neural networks do not operate in an acceptable manner, due to the technique of the back propagation. It may consume too long time, and in many cases it gets local minimum with a poor count, and slow concourse, because of the random initialization. In order to solve the complexity problems facing when dealing with machine learning, deep learning architectures considered to be as the best choice for it.

#### **3.6.1 Long Short-Term Memory Layer**

The long short-term memory (LSTM) network, abbreviated LSTM for brevity, is one of the basic neural network types that may be employed in deep architecture for data regression. This investigation proposes a model that is said to be LSTM and dense network based. The objective of this model is to expedite the process of data regression and provide meteorologists with reliable predictions. At each and every stage of the development procedure, the backpropagation network model was used. The objective of this method is to achieve the smallest difference possible between the initial output vector of the found network and the desired output vector by adjusting the weights associated with such communications in an iterative way. This article provides a concise introduction to the deep LSTM, concentrating on the most significant aspects and parameters of this neural network. Activation Function: it's basically existing in the core of deep networks, and give the network the ability to learn tyrannically the complicated mappings operation. When this occurs, the neural network is incapable of generalizing its learning and can only learn linear correlations between the input and the desired output, as opposed to the more generalized connections it is capable of learning under normal circumstances. It is composed mostly of mathematical equations and is linked to every neuron in the network. It decides whether or not a neuron in the network needs to be active (or "fired") at any particular moment, based on the relevance of each neuron's input to the prediction of the desired model. It is not possible to have two distinct forms of activation function perform identically, even if they have the same structure. This makes the selection of the proper one a crucial process; hence, it may be notified as a hyper parameter whose principal control is dependent on the set duties

and data. The activation function is not just a practical way to enable NNs to do their thing; it also has the theoretical consequence of normalizing the output of each neuron to a range between -1 and 1 or 1 and 0. In other words, the activation function has the capacity to normalize each neuron's output. There are many distinct sorts of activation functions in the suggested system. Below are some examples of these functions:

- a. Linear: it is the simplest activation function, which it is simply a linear transform on the data, that is also noticed as an identity function, in which the activation is proportionate, and suits to the input. Basically, the restricted form solution for such networks, which may be expressed as  $f(x)=x$ , and the linear activation used for linear regression problems are both instances of linear activation. The input is represented by the letter  $f$  and the output by the letter  $x$ .
- b. Rectified Linear Unit ReLU: it's a kind of the functions of activation that is non-linear, which employed the neural network having a variety of –layers. and the major concern here is that the ReLU function does not give the ability to activated all the neurons at the same intended amount of time. This function can be represented as:

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases} \quad (3.1)$$

And its gradient is obtained as

$$f(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x < 0 \\ \text{undefiend} & \text{if } x = 0 \end{cases} \quad (3.2)$$

Comparing the training times of deep neural networks to those of more conventional neural networks demonstrates that the ReLU function may significantly expedite the process. since the derivative of ReLU equal to 1 for a positive obtained value. constant, In the training phase, deep neural networks do not need additional time to compute error form. The ReLU function does not urge the problem of vanishing gradient when the count of layers grows more, due to this function does not have a close upper and lower bound in their ranges. There are many purposes, which make the ReLU get a preferred, as follow:

- i. Gradient stability: the ReLU will never vanish, because it has always a value especially a one value when the entry is positive despite of the deepness of the network that have been obtained.

- ii. low Computational complexity: The ReLU is simple in implementing and demand to provide the output only the operation of max, unlike other kinds of functions which need to perform some complicated operations such as division. Another purpose of making the ReLU is a favorable one that the same pretext is valid for the gradient computation.
- iii. Expansion: The ReLU result in layer variance as just like if the input close or turn into zero value, so as a result the connection becomes not pertinent to the intended model, so this property provides the ability to analyses the significant features or variables. In addition to the benefits that are provided by implementing the ReLU, but also suffers from a set of defects, in which it can cause the neurons to be dead when the outcome of a specific activation turns into zero, then its gradient may die evermore. Another disadvantage of the ReLU, that it is non-zero central in its nature, due to that the activations are have a tendentious to be a positive value.
- c. Leaky ReLU is used to sidestep the issue of ReLU dying out. The most major flaw of ReLU is that it always reaches saturation, despite the fact that it is not being actively employed. This is a need for its functioning. As a result, the training's tempo is dropped to zero. To address this problem, the activation has been modified such that it no longer contains a hard zero, but rather a leak. Due to this breach, the ReLU's effective range has been expanded and mathematically defined as follow:

$$f(x) = \begin{cases} 0.01x & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases} \quad (3.3)$$

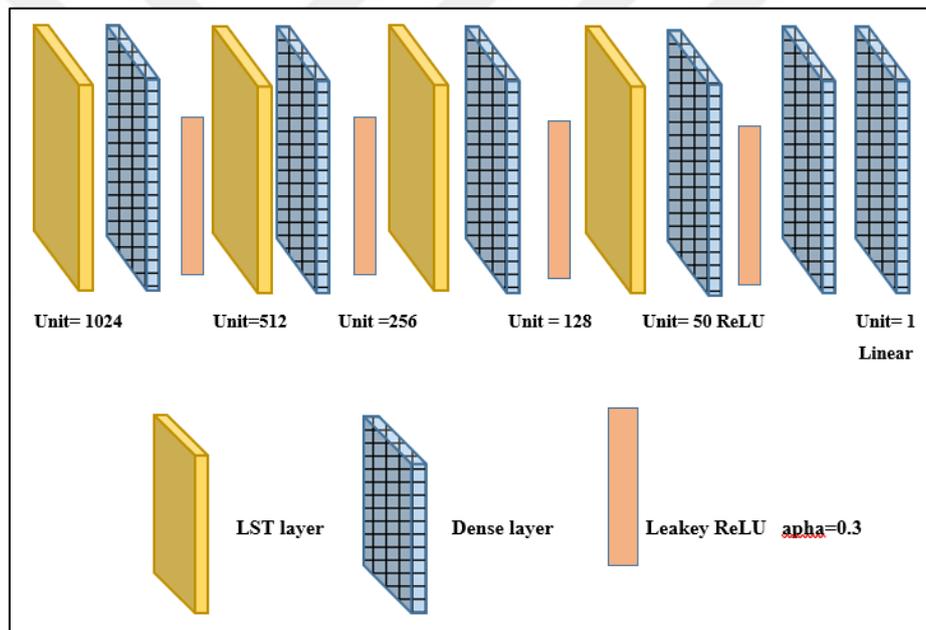
### 3.6.2 Dense Layers

It is also known as the fully connected layer or the block of dense in NN, that composed of n count of dense layers. These dense layers are communicated within the dense circuits, like in every layer in the dense obtain and get a feature map from all prior layers and receive it's feature maps to all posterior layers. The dimensionality of the outcome obtained from the former layer is changed in this kind of layer as a result the model will have the ability in defining the relationships exists between the values describing the data in a simple manner when the intended model is operating. A model's dense-layer neurons may do matrix-vector multiplication because they get input from every neuron in the layer underneath them. This is the technique through which the row vector of an item corresponds to the column vector given by layers above it. In accordance with the fundamental rule of matrix-vector

multiplication, which states that the row vector must include as many types of columns as the column vector, the dense layer will produce an N-dimensional vector as its output. The output must be an N-dimensional vector under this criterion. Within the thick layer function of Keras is a forwarding operation implementation. bias, which represents the vector acquired from the layer, may only be utilized if it is set to True, as well as activation, which represents the element-wise values sent via the activation parameter. Both of these values are only used if bias equals True.

Figure 3.3 depicts the hybrid deep model that has been approved for use in weather forecasting in its entirety. This model contains fourteen layers in total.

- i. Four LSTM layers
- ii. Six Dense layers
- iii. Four Leaky-ReLU layers



**Figure 3.3:** The Proposed Deep Hybrid Models' Layer's Structure.

## **4. RESULTS AND DISCUSSION**

### **4.1 INTRODUCTION**

In this section of the inquiry, we will examine the proposed model for weather forecasting and present a preview of the predicted outcomes after its implementation. In addition, details on the computer and programming language that would be used to implement the proposed system are provided. The root mean square error, the mean absolute error, and the mean absolute percentage error represent the results of weather forecasts generated by machine learning and the recommended deep model, respectively.

### **4.2 PROPOSED SYSTEM IMPLEMENTATION ENVIRONMENT DESCRIPTION**

It is important to define the operating and application environment of the system. This is because, as shown in this thesis, the environment in which the system is executed has a considerable influence on both the system's performance and the quality of the outputs obtained. Python, especially version 3.6.5, is utilized as the primary programming language to build the system's capabilities. All planned system tasks will be performed on a Lenovo laptop equipped with a seventh-generation CORE i7 CPU, 16 gigabytes of random access memory (RAM), an NVIDIA GTX 6G graphics card, an SSD, and a 64-bit version of Windows 10.

### **4.3 IMPLEMENTATION RESULTS USING THE LINEAR REGRESSION MACHINE LEARNING ALGORITHM**

Linear regression algorithms considered to be the simplest and popular algorithm of machine learning for performing data regression procedure. The linear regression utilized in this thesis in order to preview the gathered results from it and compare them to those resulted from the proposed deep model to preview its success in providing the best prediction results for weather forecasting. Table 4.1, preview the obtained results for data regression from the machine learning algorithm.

**Table 4.1:** Linear Regression Algorithm Prediction Results.

MAE	RMSE	MAPE
47.352	65.606	0.0002623

#### 4.4 DEEP HYBRID REGRESSION MODEL IMPLEMENTATION OUTCOMES

When applied to a dataset consisting of LSTM and Dense, the described hybrid deep learning model may be used to increase the accuracy of weather forecasts. The Root Mean Square Error, the Mean Absolute Percentage Error, and the Mean Absolute Error are well-established metrics used to measure the model's predictive accuracy. During the lifetime of a deep learning model, both the training and testing phases generate unique sets of predictions. It is necessary to have a precise estimate of the amount of time needed for both instruction and evaluation. Table 4.2 demonstrates the recommended design for the model's various tiers. This is a quality shared by the vast majority of deep models.

**Table 4.2:** Proposed Deep Hybrid Model Layers.

No.	Layer Type	Parameter	Output shape
1	LSTM	4321280	None, 1, 1024
2	Dense	1049600	None, 1, 1024
3	Leaky_ReLU	-	None, 1, 1024
4	LSTM	3147776	None, 1, 512
5	Dense	262656	None, 1, 512
6	Leaky_ReLU	-	None, 1, 512
7	LSTM	787456	None, 1, 256
8	Dense	65792	None, 1, 256
9	Leaky_ReLU	-	None, 1, 256
10	LSTM	197120	None, 128
11	Intensive	16512	None, 128
12	Leaky_ReLU	-	None, 129
13	Intensive	6450	None, 51
14	Intensive	51	None, 2

**Table 4.23:** Proposed Deep Hybrid Model Layers. “Table Continued”

Total Number of Parameters count	9,854,693
Trainable Parameters count	9,854,693
No- Trainable Parameters count	0

The utilized total number of data are separated as mentioned earlier in chapter 3 into training and testing and a detailed description about the total amount of data in both sets and the date and time of each are described forward

#### 4.4.1 Training Phase Results

The first phase of the deep model is the training phase on the 70 percent subset from the general dataset. The system has been trained each time with a different epoch value, and look back, in order to provide the best training result, within the best time required. The utilized epochs range starts within 100 epochs, as a minimum value, and every time increased within 100 reaching to the final epoch value that equal to 500. Table 4.3, illustrated the results when implementing the training phase of the proposed system using epoch equal to 100, and look back value equal to 10.

**Table 4.3:** Training Results When Epoch =100, and Look Back =10.

Mean Absolute Error (MAE)	Root Mean Squared Error (RMSE)
5.09079233444326	16.405096367151398
Time = 88589.95 seconds	

As noticed from the previous table the error rate acquired from the two measurements considered to be un acceptable value.

In order to give better performance, the system is trained by increasing both of epochs, and look back value. Table 4.4, and 4.5 illustrate the results comes from training the system within different look back value in 100, and 200 epoch values sequentially.

**Table 4.4:** Training Results within 100 Epochs.

Look back	MAE	RMSE	Time
10	5.090	16.405	88589.95
20	4.3206	10.924	6508.74
30	12.290	6.0188	5711.86
40	8.7317	12.901	4869.84
50	12.086	15.299	7125.30

**Table 4.5:** Training Results Within 200 Epoch.

Look back	MAE	RMSE	Time	Look back	MAE
10	5.020	16.414	9055.91	10	5.020
20	5.081	10.716	88809.90	20	5.081
30	9.221	13.268	7010.93	30	9.221

From the previous results, indicates that increasing the epoch value has a significant influence without substantially altering the findings. hence, the epoch is raised to get more desired results, and more accurate ones, as illustrated in table 4.6, 4.7, and 4.8 for different epoch values equal to 300, 400, and 500.

**Table 4.6:** Training Results Within 300 Epoch.

Look back	MAE	RMSE	Time
10	4.615	16.349	5882.70
20	5.861	11.437	11855.75
30	6.673	11.553	7530.17
40	5.500	10.767	18404.54
50	5.959	10.935	6702.38

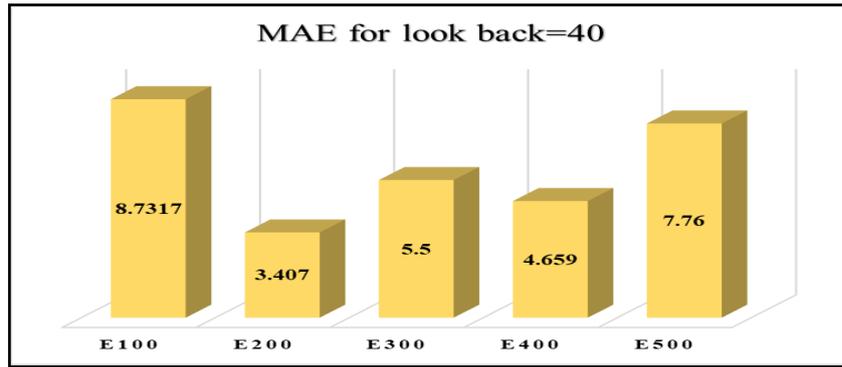
**Table 4.7:** Training Results Within 400 Epoch.

Look back	MAE	RMSE	Time
10	4.494	16.341	4792.02
20	8.252	13.253	16743.95
30	5.769	11.498	6967.99
40	4.659	10.496	7479.41
50	8.307	11.850	9996.48

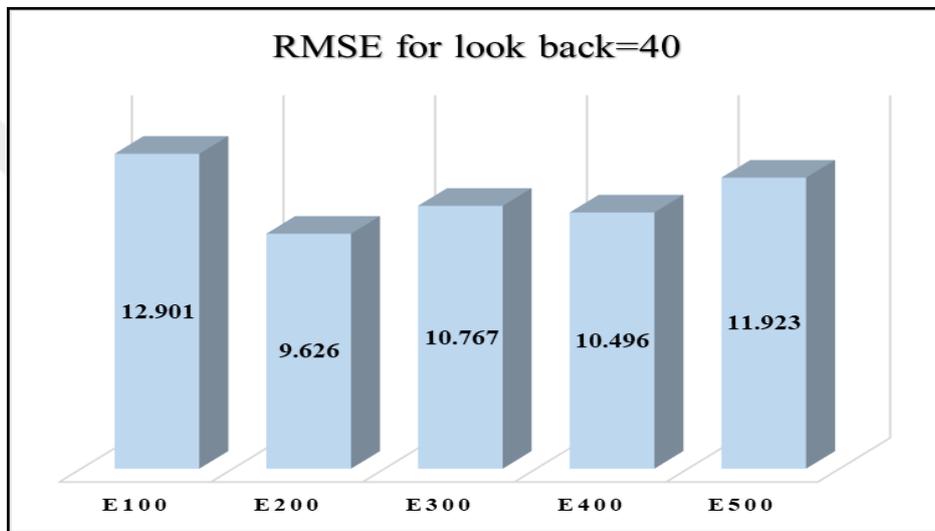
**Table 4.4:** Training results within 500 epoch.

Look back	MAE	RMSE	Time
10	7.382	16.909	5886.55
20	5.414	10.895	35192.07
30	3.795	11.151	4733.94
40	7.760	11.923	6393.03
50	4.358	10.753	6193.44

As shown in the previous results, whenever the epoch increased the training phase behaviour is enhanced, and the time required for training the system become less. As well as the lock back have an effect on the utilized results in which it has also an effect on the accuracy of the training. In general, the best attained result in the training phase is appeared when the epoch equal to 200, and the lock back is 40 as shown in figure 4.1, and 4.2, that show the MAE, and RMSE when the look back equal to 40 for all five-epoch count in the training phase.



**Figure 4.1:** MAE for All Epoch Value Using Look Back Equal to 40 in the Training Phase.



**Figure 4.2:** RMSE for all epoch value using look back equal to 40 in the training phase.

#### 4.4.2 Consequences of the Trial Run

The testing phase of the proposed system is crucial since it demonstrates the accuracy and precision of the system's weather predictions. During the testing phase, the results were also analysed in terms of root mean square error, mean absolute percentage error, and mean absolute error, as shown in the tables below for epochs of 100 and 500 and look back values between 10 and 50. This examination was conducted in order to compare the findings' precision.

**Table 4.5:** Testing Results Within 100 Epochs.

Look back	MAE	RMSE	MAPE
10	5.957	18.462	0.0000399
20	6.990	18.390	0.0000483
30	8.610	20.106	0.0000747
40	11.363	20.101	0.0000518
50	14.502	22.090	0.0000657

It noticed from the previous table of prediction results, the best results obtained when the epoch equal to 100 for look back is 10, and whenever the look back increased in the epoch 100 the results become worse and the testing prediction results reaches its final results in the look back equal to 50.

**Table 4.6:** Testing Results Within 200 Epochs.

Look back	MAE	RMSE	MAPE
10	5.842	18.292	0.0000443
20	7.830	18.136	0.0000444
30	11.671	19.754	0.0000791
40	6.355	17.374	0.0000580
50	9.111	19.386	0.0000712

The previous illustrated table, for epoch 200 for testing results, shows that the best result acquired when the look back equal to 10, as well as the worst case shown when the look back equal to 30 but it's still better than the worst case in epoch 100 and look back 50.

**Table 4.7:** Testing Results Within 300 Epoch.

Look back	MAE	RMSE	MAPE
10	5.454	18.245	0.0000400
20	8.029	18.322	0.0000461
30	9.536	19.578	0.00007912
40	8.239	19.079	0.0000525
50	8.721	19.466	0.000060

This table shows an enhancement in the acquired results for all look back values, and the best result obtained when the look back equal to 10, while the worst one is when the look back is 30.

**Table 4.8:** Testing Results Within 400 Epoch.

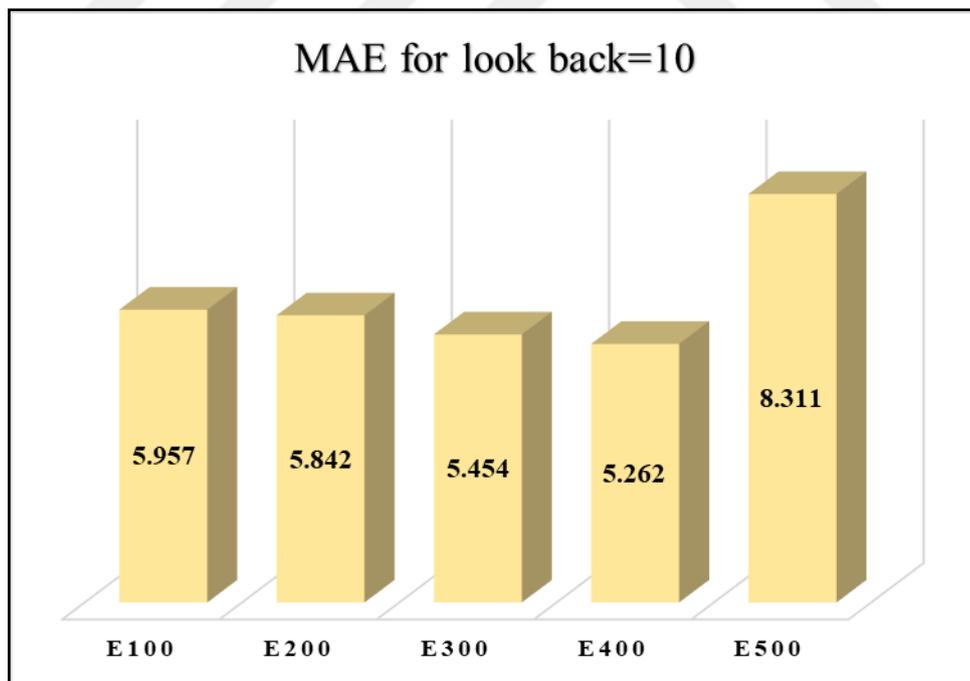
Look back	MAE	RMSE	MAPE
10	5.262	18.243	0.0000376
20	9.880	20.249	0.0000387
30	8.803	19.268	0.0000557
40	7.851	21.696	0.00006271
50	11.258	20.269	e-057.529

As illustrated above results from table 4.12 when the epoch equal to 400 the best result is when the look back equal to 10, and this is the best result gained so far, while the worst result is shown for this epoch when the look back equal to 50.

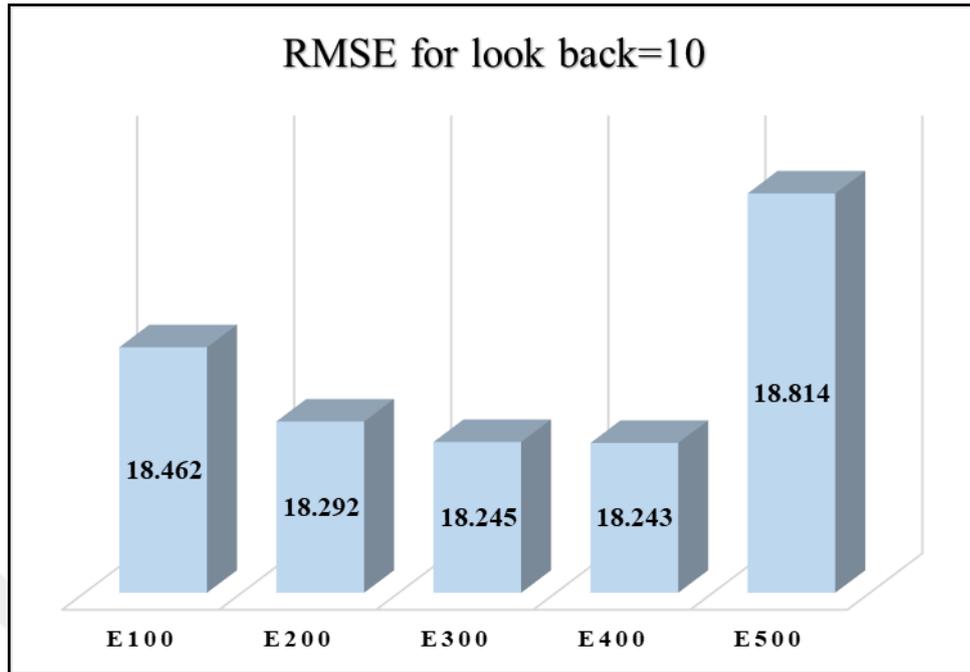
**Table 4.9:** Testing Results Within 500 Epoch.

Look back	MAE	RMSE	MAPE
10	8.311	18.814	0.0000376
20	8.204	18.957	0.0000522
30	6.328	19.737	0.0000587
40	10.156	21.891	0.0000498
50	7.332	19.728	e-059.54053

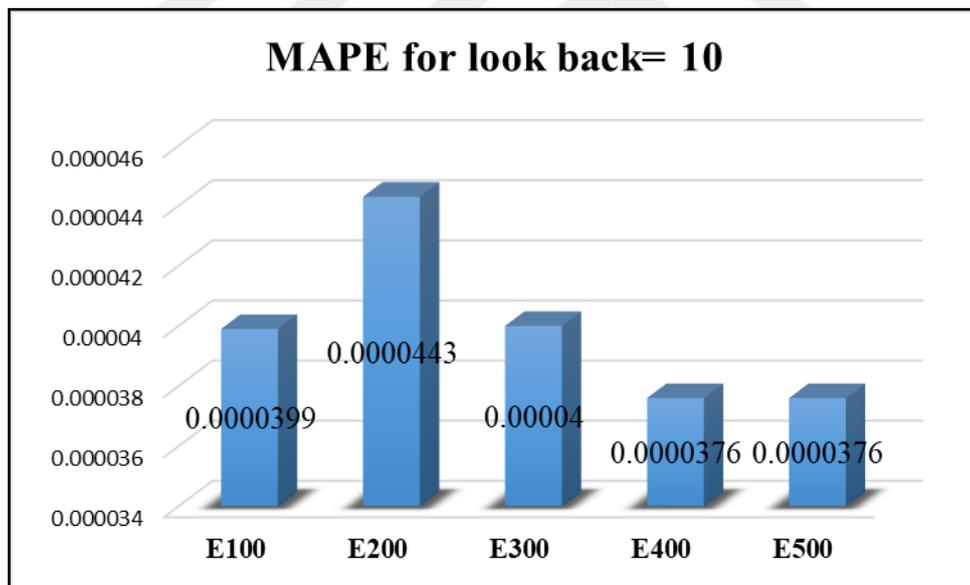
When the epoch equal to 500 some look back have been decreased compared to epoch 400, and others are increased. So far the best obtained result from implementing the proposed prediction system for all epoch values from 100 to 500, using the variant look back count from 10 to 50 appeared when using the epoch equal to 400 and look back equal to 10. Figure 4.3, 4.4, and 4.5, illustrated the MAE, RMSE, and MAPE for all epochs when utilizing the look back equal to 10.



**Figure 4.3:** MAE For All Epoch Value Using Look Back Equal to 10 in The Testing Phase.

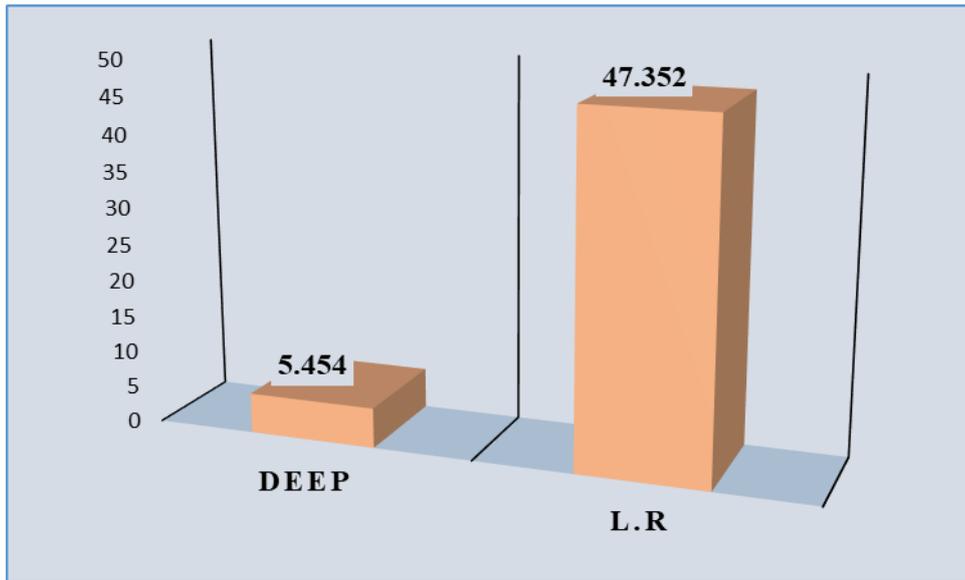


**Figure 4.4:** RMSE for All Epoch Value Using Look Back Equal to 10 in the Testing Phase.

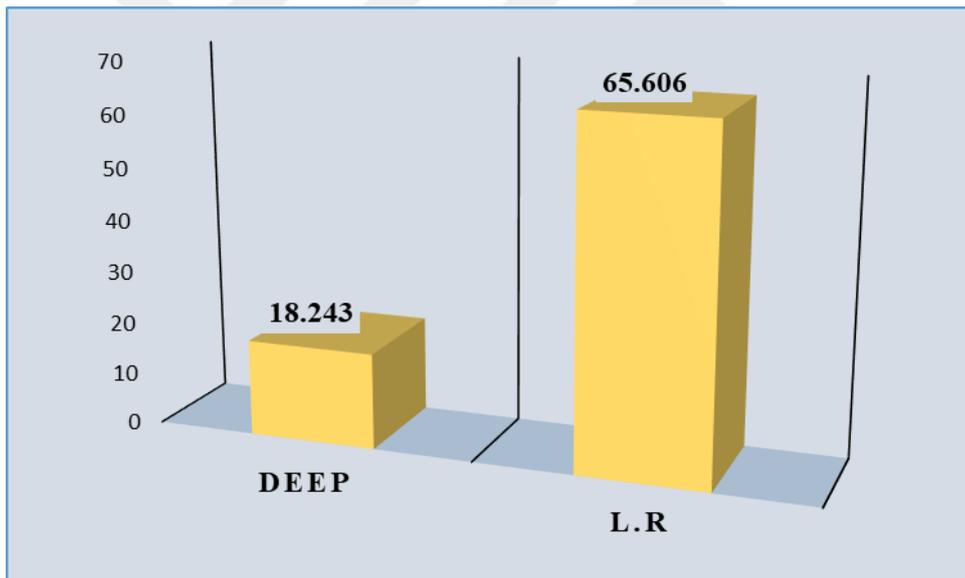


**Figure 4.5:** MAPE for all Epoch Value Using Look Back Equal to 10 in the Testing Phase.

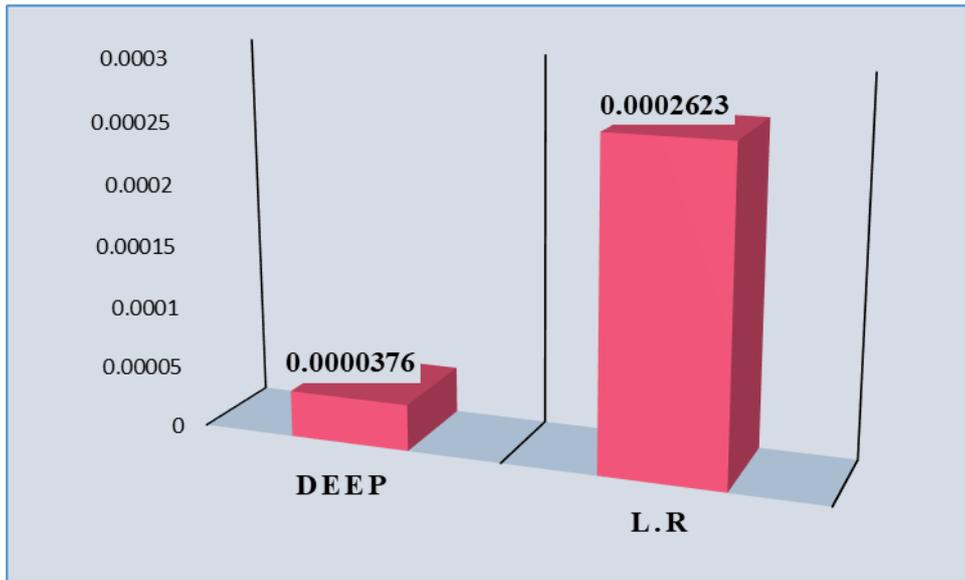
Comparing the best acquired result from the deep model within the linear regression algorithm, there is a big difference, and much better regression and prediction performance when utilizing the proposed deep model rather than the machine learning linear algorithm. Figure 4.6, and 4.7, and 4.8 illustrated the MAE, RMSE, and MAPE for the highest results comes from the proposed deep model compared within the Linear Regression algorithm



**Figure 4.6:** The Mean Absolute Error (MAE) for the Optimal Outcome of the Suggested Deep Model And LR Method.



**Figure 4.7:** Using the Suggested Deep Model and LR Algorithm, what is the RMSE for the Best Possible Outcome.



**Figure 4.8:** As Assessed by the Maximal Average Performance Evaluation, the Deep Model and LR Algorithm Have the Best Overall Performance.

## **5. CONCLUSION AND FUTURE WORKS**

### **5.1 CONCLUSIONS**

It is generally acknowledged that precise weather forecasting is required for a wide range of human activities. The writers of this book present a dependable and accurate model for weather forecasting as a method for extracting the most relevant information from the available data. The hybrid deep learning model that combines a long short-term memory network with a dense network and a leaky-reLU function serves as the basis and key component of the proposed system. The dataset was separated into a training set and a test set as the first phase of the procedure. While the testing team was responsible for giving the results of the needed predictions to validate the system's performance, the training team was responsible for developing the classifier. In addition, a basic machine learning technique known as linear regression is used to the data utilized for weather forecasting. The accuracy of a prediction may be measured using three distinct metrics: root-mean-squared error (RMSE), mean absolute error (MAE), and mean predictive error (MPE) (MAPE). Experiments reveal that the proposed hybrid deep learning model outperforms prior machine learning techniques and generates reliable predictions. The proposed deep model has an MAE that cannot exceed 15 and an RMSE that cannot exceed 22,09. To put this into perspective, the most precise results that can be produced by linear regression are an MAE of 47,352 and an RMSE of 65,606. The numbers may be found in the table shown below. Implementing the system with an epoch of 400 and a look-back period of 10 demonstrates the effectiveness of the proposed deep model for producing accurate predictions. The MAE was 5.262, the RMSE was 18.243, and the MAPE was 0.000376. This prediction comes really close to becoming accurate. This research generated a very accurate forecast, proving that a deep learning model is effective for completing a data regression. In addition, the model was capable of performing the data regression.

### **5.2 FUTURE WORK**

Multiple enhancements and additions, as well as use abilities can be performed using the proposed deep hybrid regression model, some of feature works that could be implemented are illustrated as follow:

- a. Implementing the proposed deep model on another type of weather dataset.
- b. Implementing the proposed system for performing a data regression for multiple kind of data such as electricity forecasting, stock market forecasting, electricity price forecasting, etc.
- c. Using the provided deep model, create a mobile application that delivers weather predictions. Create a smartphone application capable of delivering weather forecasts using the given deep model.
- d. Either the current layers of the stated model should be replaced with an alternative deep learning network for weather forecasting, such as a Conventional Neural Network, or more layers should be added to the model.



## REFERENCES

- [1] Y. Wang, S. Zhang, D. Wang, and Y. Liu, "Experimental study on the influence of temperature and humidity on the thermal conductivity of building insulation materials," *Energy Built Environ.*, Mar. 2022, doi: 10.1016/j.enbenv.2022.02.008.
- [2] Raju, VN Ganapathi, K. Prasanna Lakshmi, Vinod Mahesh Jain, Archana Kalidindi, and V. Padma. "Study the Influence of Normalization/Transformation Process on the Accuracy of Supervised Classification." In 2020 Third International Conference on Smart Systems and Inventive Technology ICSSIT, pp. 729-735. IEEE, 2020.
- [3] A. M. Abderrazak and F. Hachouf, "Texture Synthesis Using Improved Transfer Learning," in 2022 4th International Conference on Pattern Analysis and Intelligent Systems (PAIS), Oct. 2022, pp. 1–5. doi: 10.1109/PAIS56586.2022.9946881.
- [4] K. Kar, N. Thakur, and P. Sanghvi, "Prediction of Rainfall Using Fuzzy Dataset," *Int. J. Comput. Sci. Mob. Comput.*, vol. 8, no. 4, pp. 182–186, 2019.
- [5] Ang, Jun Chin, Andri Mirzal, Habibollah Haron, and Haza Nuzly Abdull Hamed. "Supervised, unsupervised, and semi-supervised feature selection: a review on gene selection." *IEEE/ACM transactions on computational biology and bioinformatics* 13, no. 5 2015: 971-989.
- [6] T. Anjali, K. Chandini, K. Anoop, and V. L. Lajish, "Temperature Prediction using Machine Learning Approaches," in 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), Jul. 2019, pp. 1264–1268. doi: 10.1109/ICICICT46008.2019.8993316.
- [7] L. Bengtsson and K. I. Hodges, "Can an ensemble climate simulation be used to separate climate change signals from internal unforced variability?" *Clim. Dyn.*, vol. 52, no. 5–6, pp. 3553–3573, Mar. 2019, doi: 10.1007/s00382-018-4343-8.
- [8] Bhattacharya, Sweta, et al., "Deep Neural Networks Based Approach for Battery Life Prediction." *CMC-COMPUTERS MATERIALS & CONTINUA* 69, no. 2 2021: 2599-2 615.
- [9] Booz, Jarrett, Wei Yu, Guobin Xu, David Griffith, and Nada Golmie. "A deep learning-based weather forecast system for data volume and recency analysis." In 2019 International Conference on Computing, Networking and Communications ICNC , pp. 697-701. IEEE, 2019.

- [10] G. Dzemyda and M. Sabaliauskas, "On the Computational Efficiency of Geometric Multidimensional Scaling," in 2021 2nd European Symposium on Software Engineering, Nov. 2021, pp. 136–141. doi: 10.1145/3501774.3501794.
- [11] D. Zimmer, "Machine Learning for the Construction of Force Fields," pp. 1–10, 2021.
- [12] M. Cerjan, A. Petričić, and M. Delimar, "HIRA Model for Short-Term Electricity Price Forecasting," *Energies*, vol. 12, no. 3, p. 568, Feb. 2019, doi: 10.3390/en12030568.
- [13] S. Han, Y. Qiao, J. Yan, Y. Liu, L. Li, and Z. Wang, "Mid-to-long term wind and photovoltaic power generation prediction based on copula function and long short term memory network," *Appl. Energy*, vol. 239, pp. 181–191, Apr. 2019, doi: 10.1016/j.apenergy.2019.01.193.
- [14] Czibula, Gabriela, Andrei Mihai, and Eugen Mihuleț. "Nowdeepn: An ensemble of deep learning models for weather nowcasting based on radar products' values prediction." *Applied Sciences* 11, no. 1 2020: 125.
- [15] Dalton, Benjamin. "Data Mining: A Preprocessing Engine." *Solid State Technology* 62, no. 4 2019: 09-16.
- [16] Davis, Robert E., Glenn R. McGregor, and Kyle B. Enfield. "Humidity: A review and primer on atmospheric moisture and human health." *Environmental research* 144 2016: 106-116.
- [17] K. S. Rani, M. Kumari, V. B. Singh, and M. Sharma, "Deep Learning with Big Data: An Emerging Trend," in 2019 19th International Conference on Computational Science and Its Applications (ICCSA), Jul. 2019, pp. 93–101. doi: 10.1109/ICCSA.2019.00005.
- [18] Ferreira, Pedro, Duc C. Le, and Nur Zincir-Heywood. "Exploring feature normalization and temporal information for machine learning based insider threat detection." In 2019 15th International Conference on Network and Service Management CNSM , pp. 1-7. IEEE, 2019.
- [19] X. Liu, "Research on the Relevant Matching Method of Internet MediaInformation-Stock Assets Based on the Combination of Literal and Semantic," *J. Educ. Humanit. Soc. Sci.*, vol. 4, pp. 228–233, Nov. 2022, doi: 10.54097/ehss.v4i.2771.
- [20] Gökhan, A. K. S. U., Cem Oktay Güzeller, and Mehmet Taha Eser. "The effect of the normalization method used in different sample sizes on the success of artificial neural network model." *International Journal of Assessment Tools in Education* 6, no. 2 2019: 170-192.

- [21] Grönquist, Peter, Chengyuan Yao, Tal Ben-Nun, Nikoli Dryden, Peter Dueben, Shigang Li, and Torsten Hoefler. "Deep learning for post-processing ensemble weather forecasts." *Philosophical Transactions of the Royal Society A* 379, no. 2194 2021: 20200092.
- [22] Grover, Aditya, Ashish Kapoor, and Eric Horvitz. "A deep hybrid model for weather forecasting." In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 379-386. 2015.
- [23] Y. Sun, W. Jiang, and C. Zeng, "Do Government Subsidies Affect Analyst Forecast Accuracy?" *SSRN Electron. J.*, 2022, doi: 10.2139/ssrn.4045559.
- [24] Haupt, Sue Ellen, and Branko Kosovic. "Big data and machine learning for applied weather forecasts: Forecasting solar power for utility operations." In *2015 IEEE Symposium Series on Computational Intelligence*, pp. 496-501. IEEE, 2015.
- [25] Alowaid Alotaibi, Mohd Khairul Hafiz Muda, Faizal Mustapha, Izhal Abdul Halin, and Noorfaizal Yidris, "Perpetual Motion Wind Turbine Generator for Novelty Energy Harvesting System; Conceptual Design Approach," *J. Adv. Res. Fluid Mech. Therm. Sci.*, vol. 94, no. 2, pp. 166–173, May 2022, doi: 10.37934/arfmts.94.2.166173.
- [26] Hernández, Emilcy, Victor Sanchez-Anguix, Vicente Julian, Javier Palanca, and Néstor Duque. "Rainfall prediction: A deep learning approach." In *International Conference on Hybrid Artificial Intelligence Systems*, pp. 151-162. Springer, Cham, 2016.
- [27] Hewage, Pradeep, Marcello Trovati, Ella Pereira, and Ardhendu Behera. "Deep learning-based effective fine-grained weather forecasting model." *Pattern Analysis and Applications* 24, no. 1 2021: 343-366.
- [28] S. Bin Islam, M. M. Hasan, and M. M. Khan, "Prediction of Stock Market Using Recurrent Neural Network," in *2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, Oct. 2021, pp. 0479–0483. doi: 10.1109/IEMCON53756.2021.9623206.
- [29] [https://datacadamia.com/data\\_mining/simple\\_regression](https://datacadamia.com/data_mining/simple_regression)
- [30] <https://www.kaggle.com/datasets/stytch16/jena-climate-2009-2016>
- [31] Huang, Gao, Shiji Song, Jatinder ND Gupta, and Cheng Wu. "Semi-supervised and unsupervised extreme learning machines." *IEEE transactions on cybernetics* 44, no. 12 2014: 2405-2417.

- [32] Hyndman, Rob J., and George Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2018.
- [33] M. All, "Sikkim University," vol. 2015, no. 1, pp. 2015–2016, 2015.
- [34] Irie, Kazuki, Zoltán Tüske, Tamer Alkhouli, Ralf Schlüter, and Hermann Ney. "LSTM, GRU, highway and a bit of attention: an empirical overview for language modeling in speech recognition." In *Interspeech*, pp. 3519-3523. 2016.
- [35] Jain, Himanshi, and Raksha Jain. "Big data in weather forecasting: Applications and challenges." In *2017 International conference on big data analytics and computational intelligence ICBDAC*, pp. 138-142. IEEE, 2017.
- [36] S. Makigaki and T. Ishida, "Sequence alignment using machine learning for accurate template-based protein structure prediction," *Bioinformatics*, vol. 36, no. 1, pp. 104–111, Jan. 2020, doi: 10.1093/bioinformatics/btz483.
- [37] Janicki, Marcin. "Methods of weather variables introduction into short-term electric load forecasting models-a review." *Przegląd Elektrotechniczny* 93 2017: 70-73.
- [38] Kappal, Sunil. "Data normalization using median median absolute deviation MMAD based Z-score for robust predictions vs. min–max normalization." *London Journal of Research in Science: Natural and Formal* 2019.
- [39] Khair, Ummul, Hasanul Fahmi, Sarudin Al Hakim, and Robbi Rahim. "Forecasting error calculation with mean absolute deviation and mean absolute percentage error." In *Journal of Physics: Conference Series*, vol. 930, no. 1, p. 012002. IOP Publishing, 2017.
- [40] Kocharekar, Abhijit, Bharat V. Nemade, Chetan G. Patil, Durgesh D. Sapkale, and Sagar G. Salunke. "Weather prediction for tourism application using arima." *Weather* 6, no. 11 2019.
- [41] Kuster, Corentin, Yacine Rezgui, and Monjur Mourshed. "Electrical load forecasting models: A critical systematic review." *Sustainable cities and society* 35 2017: 257-270.
- [42] Le, Xuan-Hien, Hung Viet Ho, Giha Lee, and Sungho Jung. "Application of long short-term memory LSTM neural network for flood forecasting." *Water* 11, no. 7 2019: 1387.

- [43] R. Li, H. Zhang, N. Yu, R. Li, and Q. Zhuang, "A fast approximation for 1-D inversion of transient electromagnetic data by using a back propagation neural network and improved particle swarm optimization," *Nonlinear Process. Geophys.*, vol. 26, no. 4, pp. 445–456, Nov. 2019, doi: 10.5194/npg-26-445-2019.
- [44] Libbrecht, Maxwell W., and William Stafford Noble. "Machine learning applications in genetics and genomics." *Nature Reviews Genetics* 16, no. 6 2015: 321-332.
- [45] J. T. Halloran and D. M. Roche, "GPU-Accelerated SVM Learning for Extremely Fast Large-Scale Proteomics Analysis," *Mlcb.Github.Io*, 2019.
- [46] Liu, Weibo, Zidong Wang, Xiaohui Liu, Nianyin Zeng, Yurong Liu, and Fuad E. Alsaadi., "A survey of deep neural network architectures and their applications." *Neurocomputing* 234 2017: 11-26.
- [47] M. Fayad, M.-Y. Hachani, A. Mostefaoui, S. Chouali, and R. Yahiaoui, "Elderly Fall Detection: A Lightweight Kinect Based Deep Learning Approach," in *Proceedings of the 20th ACM International Symposium on Mobility Management and Wireless Access on 20th ACM International Symposium on Mobility Management and Wireless Access*, Oct. 2022, pp. 89–95. doi: 10.1145/3551660.3560911.
- [48] Z. Zhou, "Automatic Machine Learning-Based Data Analysis For Video Game Industry," in *2022 2nd International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI)*, Sep. 2022, pp. 732–737. doi: 10.1109/CEI57409.2022.9950194.
- [49] Mehrkanoon, Siamak., "Deep shared representation learning for weather elements forecasting." *Knowledge-Based Systems* 179 2019: 120-128.
- [50] Mei, Chao, Jiahong Liu, Hao Wang, Zhiyong Yang, Xiangyi Ding, and Weiwei Shao. "Integrated assessments of green infrastructure for flood mitigation to support robust decision-making for sponge city construction in an urbanized watershed." *Science of the Total Environment* 639 2018: 1394-1407.
- [51] Narejo, Sanam, and Eros Pasero. "Meteonowcasting using deep learning architecture." *IJACSA International Journal of Advanced Computer Science and Applications* 8, no. 8 2017.

- [52] C. E. Bunn, K. S. Repasky, R. A. Stillwell, M. Hayman, and S. M. Spuler, "Diode-Laser-Based Micro-Pulse Differential Absorption Lidar (DIAL) for Thermodynamic Profiling of the Lower Troposphere," *Opt. Sensors Sens. Congr. (ES, FTS, HISE, Sensors)* (2019), Pap. ETu2A.4, vol. Part F172-ES 2019, p. ETu2A.4, Jun. 2019, doi: 10.1364/ES.2019.ETU2A.4.
- [53] T. Kim, N. Ho, D. Kim, and S.-Y. Yun, "Benchmark Dataset for Precipitation Forecasting by Post-Processing the Numerical Weather Prediction," Jun. 2022, doi: 10.48550/arxiv.2206.15241.
- [54] L. J. Speight, M. D. Cranston, C. J. White, and L. Kelly, "Operational and emerging capabilities for surface water flood forecasting," *Wiley Interdiscip. Rev. Water*, vol. 8, no. 3, p. e1517, May 2021, doi: 10.1002/WAT2.1517.
- [55] Phyoo, Pyae Pyae. "Deep Learning for Short-Term Electricity Load Forecasting." Phd Diss., Thammasat University, 2017.
- [56] Rodrigues, Eduardo Rocha, Igor Oliveira, Renato Cunha, and Marco Netto. "DeepDownscale: a deep learning strategy for high-resolution weather forecast." In 2018 IEEE 14th International Conference on e-Science e-Science , pp. 415-422. IEEE, 2018.
- [57] Rong, Shen, and Zhang Bao-wen. "The research of regression model in machine learning field." In *MATEC Web of Conferences*, vol. 176, p. 01033. EDP Sciences, 2018.
- [58] Roy, Debneil Saha. "Forecasting the air temperature at a weather station using deep neural networks." *Procedia Computer Science* 178 2020: 38-46
- [59] Ryu, Seunghyoung, Jaekoo Noh, and Hongseok Kim. "Deep neural network-based demand side short term load forecasting." *Energies* 10, no. 1 2017: 3.
- [60] Maciej Serda et al., "Synteza i aktywność biologiczna nowych analogów tiosemikarbazonowych chelatorów żelaza," *Uniw. śląski*, vol. 7, no. 1, pp. 343–354, 2013, doi: 10.2/JQUERY.MIN.JS.
- [61] Schmidhuber, Jürgen. "Deep learning in neural networks: An overview." *Neural networks* 61 2015: 85-117.
- [62] G. P. McMillan and J. B. Cannon, "Bayesian Applications in Auditory Research," *J. Speech, Lang. Hear. Res.*, vol. 62, no. 3, pp. 577–586, Mar. 2019, doi: 10.1044/2018\_JSLHR-H-ASTM-18-0228.
- [63] Shrestha, Ajay, and Ausif Mahmood. "Review of deep learning algorithms and architectures." *IEEE Access* 7 2019: 53040-53065.

- [64] E. Tochimoto, "Environmental Controls on Tornadoes and Tornado Outbreaks," <https://doi.org/10.1080/07055900.2022.2079472>, vol. 60, no. 3–4, pp. 399–421, 2022, doi: 10.1080/07055900.2022.2079472.
- [65] Staudemeyer, Ralf C., and Eric Rothstein Morris. "Understanding LSTM--a tutorial into Long Short-Term Memory Recurrent Neural Networks." arXiv preprint arXiv:1909.09586 2019.
- [66] Stewart, Margaret C., and B. Gail Wilson. "The dynamic role of social media during Hurricane# Sandy: An introduction of the STREMI model to weather the storm of the crisis lifecycle." *Computers in Human Behavior* 54 2016: 639-646.
- [67] K. Kraehnert et al., "Insurance against Extreme Weather Events: An Overview," *Rev. Econ.*, vol. 72, no. 2, pp. 71–95, Aug. 2021, doi: 10.1515/ROE-2021-0024/PDF.
- [68] Taylor, Sean J., and Benjamin Letham. "Forecasting at scale." *The American Statistician* 72, no. 1 2018: 37-45.
- [69] M. Tajiki, G. Schoups, H. J. Hendricks Franssen, A. Najafinejad, and A. Bahremand, "Recursive Bayesian Estimation of Conceptual Rainfall-Runoff Model Errors in Real-Time Prediction of Streamflow," *Water Resour. Res.*, vol. 56, no. 2, Feb. 2020, doi: 10.1029/2019WR025237.
- [70] G. Ganjir, S. Pattnaik, and D. Trivedi, "Characteristics of dynamical and thermodynamical variables during heavy rainfall events over the Indian region," *Dyn. Atmos. Ocean.*, vol. 99, p. 101315, Sep. 2022, doi: 10.1016/J.DYNATMOCE.2022.101315.
- [71] Wilson, Tyler, Pang-Ning Tan, and Lifeng Luo. "A low rank weighted graph convolutional approach to weather prediction." In 2018 IEEE International Conference on Data Mining ICDM, pp. 627-636. IEEE, 2018.
- [72] Z. Xiao, B. Liu, H. Liu, and D. Zhang, "Progress in climate prediction and weather forecast operations in China," *Adv. Atmos. Sci.*, vol. 29, no. 5, pp. 943–957, Sep. 2016, doi: 10.1007/S00376-012-1194-9/METRICS.
- [73] Yan, Qing, Fuxin Ji, Kaichao Miao, Qi Wu, Yi Xia, and Teng Li. "Convolutional residual-attention: a deep learning approach for precipitation nowcasting." *Advances in Meteorology* 2020.