MULTIVARIATE FORECASTING OF GLOBAL HORIZONTAL IRRADIATION USING DEEP LEARNING ALGORITHMS

A THESIS SUBMITTED TO
THE BOARD OF GRADUATE PROGRAMS
OF
MIDDLE EAST TECHNICAL UNIVERSITY, NORTHERN CYPRUS CAMPUS

BY

NURAY VAKİTBİLİR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN SUSTAINABLE ENVIRONMENT AND ENERGY SYSTEMS PROGRAM

FEBRUARY 2021

Approval of the Board of Graduate Programs

Prof. Dr. Oğuz Solyalı
Chairperson

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science

Assoc. Prof. Dr. Ceren İnce Derogar
Program Coordinator

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Asst. Prof. Dr. Cem Direkoğlu
Supervisor

Examining Committee Members

Assoc. Prof. Dr. Murat Fahrioğlu
METU NCC, Electrical and Electronics Engineering

Asst. Prof. Dr. Cem Direkoğlu
METU NCC, Electrical and Electronics Engineering

Asst. Prof. Dr. Kamil Yurtkan
Cyprus International University, Computer Engineering

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Nuray Vakitbilir

Signature:

ABSTRACT

MULTIVARIATE FORECASTING OF GLOBAL HORIZONTAL IRRADIATION USING DEEP LEARNING ALGORITHMS

Vakitbilir, Nuray
Master of Science, Sustainable Environment and Energy Systems Program
Supervisor: Asst. Prof. Dr. Cem Direkoğlu

February 2021, 104 pages

Increasing photovoltaic (PV) panel instalments jeopardise the electrical grid frequency, especially in island countries, such as Cyprus. For a continuous growth in the PV instalments in Northern Cyprus as well as minimal usage of conventional energy sources in power generation, it is of utter importance for a grid manager to possess information on the energy production of PV panels, hence knowledge on received radiation, i.e. Global Horizontal Irradiation (GHI). Therefore, the prediction of GHI plays an essential role in the growth of renewable energy in Northern Cyprus. This study focuses on forecasting long-term and short-term GHI for Kalkanlı, Northern Cyprus. For long-term forecasting, a dataset is obtained from NASA while the short-term GHI prediction is carried out with a dataset recorded at METU NCC. Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) algorithms are employed for the long-term GHI forecasting. Support Vector Regression (SVR) is employed in addition to CNN and LSTM algorithms in the short-term GHI estimation. For both datasets, hybrid and stand-alone models are constructed, and their performances evaluated extensively. Additionally, seasonal

forecasting is carried out for the short-term GHI estimation with a hybrid model of CNN, LSTM and SVR.

Keywords: Global Horizontal Irradiation, Deep Learning, Time-Series Forecasting, Seasonal Forecasting, Hybrid Forecasting Algorithms

ÖZ

DERİN ÖĞRENME ALGORİTMALARI KULLANARAK KÜRESEL YATAY IŞINLAMANIN ÇOK DEĞİŞKENLİ TAHMİNİ

Vakitbilir, Nuray
Yüksek Lisans, Sürdürülebilir Çevre ve Enerji Sistemleri
Tez Yöneticisi: Yrd. Doç. Dr. Cem Direkoğlu

Şubat 2021, 104 sayfa

Artan fotovoltaik (PV) panel kurulumları, özellikle Kıbrıs gibi ada ülkelerinde elektrik şebekesi frekansını tehlikeye atıyor. Kuzey Kıbrıs'ta PV kurulumlarında sürekli bir büyüme ve aynı zamanda güç üretiminde geleneksel enerji kaynaklarının minimum kullanımı için, bir şebeke yöneticisinin PV panellerinin enerji üretimi hakkında bilgi sahibi olması, dolayısıyla alınan radyasyon, yani Küresel Yatay Işınlama (GHI) hakkında bilgi sahibi olması son derece önemlidir. Bu nedenle, GHI tahmini Kuzey Kıbrıs'ta yenilenebilir enerjinin büyümesinde önemli bir rol oynamaktadır. Bu çalışma, Kuzey Kıbrıs Kalkanlı için uzun vadeli ve kısa vadeli GHI tahminine odaklanmaktadır. Uzun vadeli tahminler için NASA'dan bir veri seti elde edilirken, kısa vadeli GHI tahmini ODTÜ KKK'da kaydedilen bir veri seti ile gerçekleştirilmiştir. Uzun vadeli GHI tahmini için Evrişimli Sinir Ağı (CNN) ve Uzun Kısa Süreli Bellek (LSTM) algoritmaları kullanılmıştır. Kısa vadeli GHI tahmininde CNN ve LSTM algoritmalarına ek olarak Destek Vektör Regresyonu (SVR) kullanılmıştır. Her iki veri kümesi için de hibrit ve bağımsız modeller oluşturulmuş ve performansları kapsamlı bir şekilde değerlendirmiştir. Ek olarak,

CNN, LSTM ve SVR'nin hibrit modeli ile kısa vadeli GHI tahmini için mevsimsel tahmin gerçekleştirilmiştir.

Anahtar Kelimeler: Küresel Yatay Işınlama, Derin Öğrenme, Zaman Serisi Tahmin, Mevsimsel Tahmin, Hibrit Tahmin Algoritmaları

To My Family

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my supervisor Asst. Prof. Dr. Cem Direkoğlu, for his valuable guidance, advice and constructive criticism that allowed me to grow as a researcher. His patience, motivation and great knowledge helped me during my research and writing of this thesis. I also would like to extend my gratitude towards my committee members, Assoc. Prof. Dr. Murat Fahrioğlu and Asst. Prof. Dr. Kamil Yurtkan.

Additionally, I would like to express my gratefulness to Asst. Prof. Dr. Bengü Bozkaya-Schrotter for allowing me to experience an academic career, for her constant support and guidance. Furthermore, I would like to thank Assoc. Prof. Dr. Onur Taylan for providing us with the necessary data. I would also like to thank Assoc. Prof. Dr. Kürşat Aker for helping me to establish the required skills to carry on my research.

I would like to thank my dear parents, Songül and Mehmet, for enabling me to achieve every single success of mine by creating opportunities that they never had. I appreciate your sacrifices every single day. I would also like to thank my brother, Vakkas, and sister-in-law, Ilper, for their continuous support, and my dear nephew, my little angel Mehmet Azis, for always being my source of cheer.

Above all, I would like to express my appreciation to my dearest friend and partner in life, Adnan Hilal, for his immense support, unending patience and love. Thank you for making this process to be a bearable and a joyful time.

TABLE OF CONTENTS

ABSTRACT v

ÖZ vii

ACKNOWLEDGEMENTS x

TABLE OF CONTENTS xi

LIST OF TABLES xiii

LIST OF FIGURES xv

LIST OF ABBREVIATIONS xviii

CHAPTERS

1 INTRODUCTION 1

1.1 Motivation 1

1.2 Objectives 3

1.3 Overview of the Thesis 4

1.4 Publication Related to this Study 4

2 BASIC CONCEPTS AND REVIEW OF RELATED WORK 5

2.1 Basic Concepts 5

2.1.1 Radiation 5

2.1.2 Forecasting Algorithms 7

2.1.3 Model Input Variables 11

2.1.4 Model Evaluation Metrics 12

2.2 Review of Related Work 15

2.2.1 GHI Estimation Around the World 15

2.2.2 GHI Forecasting for Mediterranean Region 19

2.2.3 Prediction of Radiation for Cyprus 21

2.3 Gaps in the Literature 22
3 MATERIALS AND METHODOLOGY 25
3.1 Study Area 25
3.2 Data 26
3.2.1 NASA Dataset 26
3.2.2 METU NCC Dataset 32
3.3 Forecasting Algorithms 37
3.3.1 Activation Functions 37
3.3.2 Convolutional Neural Networks (CNN) 38
3.3.3 Long Short-Term Memory (LSTM) 40
3.3.4 Support Vector Regression (SVR) 42
3.4 Experimental Setup 44
3.4.1 Data Preprocessing 44
3.4.2 Construction of Learning Algorithms 48
4 RESULTS AND DISCUSSION 61
4.1 GHI Prediction Analysis of NASA Dataset 61
4.2 GHI Forecasting Analysis of METU NCC Dataset 64
5 CONCLUSION AND RECOMMENDATIONS 75
REFERENCES 79
APPENDICES
A. Forecasting Results for NASA Dataset 87
B. Forecasting Results for METU NCC Dataset 95

LIST OF TABLES

TABLES

Table 2.1. Model performance classified according to nRMSE value 13
Table 2.2. Classification of prediction accuracy with corresponding MAPE value .. 14
Table 2.3. Summary of the research related to the GHI forecasting for various part of the world..... 18
Table 2.4. Summary of the studies concerning the GHI forecasting for several countries in the Mediterranean region..... 20
Table 2.5. Summary of radiation forecasting studies conducted for Cyprus 22
Table 3.1. A sample set of NASA data 27
Table 3.2. Units and data ranges of input variables for NASA dataset 28
Table 3.3. A sample set from the METU NCC dataset..... 33
Table 3.4. Input variables' units and data ranges in METU NCC dataset 33
Table 3.5. Input tensor dimensions of each dataset 48
Table 3.6. Training hyperparameters in each layer for the constructed forecasting algorithms..... 51
Table 3.7. Training parameters and input data information for the stand-alone algorithms of annual forecasting..... 53
Table 3.8. Training parameters and input data information for the hybrid algorithms, C-LSTM and CN-M, of annual forecasting 55
Table 3.9. Training parameters and input data information for the hybrid algorithm, CM-SVR, of annual forecasting..... 56
Table 3.10. Layers in hybrid CM-SVR algorithm for seasonal forecasting 58
Table 3.11. Training hyperparameters in each layer for the constructed forecasting algorithms..... 59
Table 4.1. Summary of GHI prediction model performances with corresponding dataset type, best results in each evaluation metric is shown in bold 63
Table 4.2. Summary of GHI prediction model performances in different time-leads for annual forecasting, best results are shown in bold..... 67

Table 4.3. Summary of GHI prediction model performances in different time-leads for seasonal forecasting..... 70
Table 4.4. Averaged evaluation metric results of all seasons in different forecasting horizons..... 73

LIST OF FIGURES

FIGURES

Figure 1.1. Short-term radiation forecasting scale for typical target applications [8] 3
Figure 2.1. Types of solar radiation incidents on a tilted surface [31] 6
Figure 2.2. Spatial distribution of averaged GHI around the world [33]..... 7
Figure 2.3. A linear ML algorithm example; linear regression fitting a line over a set of data points where y represents actual values, while X represents predicted values 8
Figure 2.4. Structure of a simple ANN showing the flow direction of information and errors though feed-forward and back-propagation, respectively 10
Figure 2.5. Bias-variance trade-off graph [6] 15
Figure 3.1. The spatial distribution of GHI over Cyprus island [33]..... 26
Figure 3.2. Temporal distribution of GHI throughout the NASA dataset 28
Figure 3.3. Change in GHI values over the year 2018..... 29
Figure 3.4. Distribution of daily GHI data throughout the whole NASA dataset .. 30
Figure 3.5. Temporal distribution of temperature over NASA dataset..... 31
Figure 3.6. Change in temperature values over the year 2018..... 32
Figure 3.7. Temporal distribution of GHI throughout the METU NCC dataset..... 34
Figure 3.8. Change in GHI values over a week in (a) February, (b) March, (c) June, and (d) October 35
Figure 3.9. Distribution of GHI data in 10-minute interval throughout the whole METU NCC dataset..... 36
Figure 3.10. Temporal distribution of temperature over METU NCC dataset..... 37
Figure 3.11. The architecture of a general 1D-CNN..... 40
Figure 3.12. Sample structure of LSTM unit..... 42
Figure 3.13. Sample structure of SVR model prediction [82] 44
Figure 3.14. Flow chart for the preprocessing procedure for both datasets..... **Error! Bookmark not defined.**

Figure 3.15. GHI distribution after the removal of night hours in METU NCC dataset 46
Figure 3.16. Flow chart showing the hybrid CN-M algorithm..... 50
Figure 3.17. Flow chart showing the hybrid CM-SVR algorithm..... 56
Figure 4.1. Flow chart of the forecasting procedure for the NASA dataset..... 62
Figure 4.2. APE frequency histograms generated by the results of testing set, a) CNN – exogenous, b) CNN – endogenous, c) LSTM – exogenous, d) LSTM – endogenous, e) CN-M..... 64
Figure 4.3. Flow chart of the forecasting procedure for the METU NCC dataset .. 66
Figure 4.4. APE frequency histograms generated by the results of the testing set for the stand-alone models over 30-minutes forecasting horizon, a) CNN – endogeneous, b) LSTM – exogeneous, c) SVR – exogeneous 68
Figure 4.5. APE frequency histograms generated by the results of the testing set for the hybrid models, a) C-LSTM, b) CN-M, c) CM-SVR 69
Figure 4.6. CM-SVR model prediction fitted over the actual data for the summer season for a clear-sky condition 71
Figure 4.7. CM-SVR model prediction fitted over the actual data for the winter season for a scattered clouds sky condition..... 71
Figure 4.8. APE frequency histograms generated by the results of the testing set for the seasonal forecasting with CM-SVR models, a) summer, b) fall, c) winter, d) spring 72
Figure A.1. Prediction performance of exogenous LSTM over a year 87
Figure A.2. Prediction performance of endogenous LSTM over a year 88
Figure A.3. Prediction performance of exogenous CNN over a year 88
Figure A.4. Prediction performance of endogenous CNN over a year 89
Figure A.5. Prediction performance of hybrid CN-M over a year 89
Figure A.6. Predicted and actual GHI values on a scattered plot for exogenous LSTM over the testing set 90
Figure A.7. Predicted and actual GHI values on a scattered plot for endogenous LSTM over the testing set 91

Figure A.8. Predicted and actual GHI values on a scattered plot for exogenous CNN over the testing set 92
Figure A.9. Predicted and actual GHI values on a scattered plot for endogenous CNN over the testing set 93
Figure A.10. Predicted and actual GHI values on a scattered plot for CN-M over the testing set 94
Figure B.1. Predicted and actual GHI values on a scattered plot for CNN over the testing set 95
Figure B.2. Predicted and actual GHI values on a scattered plot for LSTM over the testing set 96
Figure B.3. Predicted and actual GHI values on a scattered plot for SVR over the testing set 97
Figure B.4. Predicted and actual GHI values on a scattered plot for C-LSTM over the testing set..... 98
Figure B.5. Predicted and actual GHI values on a scattered plot for CN-M over the testing set 99
Figure B.6. Predicted and actual GHI values on a scattered plot for CM-SVR over the testing set..... 100
Figure B.7. Predicted and actual GHI values on a scattered plot for the summer season over the testing set..... 101
Figure B.8. Predicted and actual GHI values on a scattered plot for the fall season over the testing set 102
Figure B.9. Predicted and actual GHI values on a scattered plot for the winter season over the testing set 103
Figure B.10. Predicted and actual GHI values on a scattered plot for the spring season over the testing set 104

LIST OF ABBREVIATIONS

ABBREVIATIONS

2D-CNN	Two Dimensional Convolutional Neural Network
ANN	Artificial Neural Network
APE	Absolute Prediction Error
ARMA	Autoregressive Component Moving Average
CNN	Convolutional Neural Network
DHI	Diffuse Horizontal Irradiance
DL	Deep Learning
DNI	Direct Normal Irradiance
DRWNN	Diagonal Recurrent Wavelet Neural Network
FFNN	Feed-Forward Neural Network
GHI	Global Horizontal Irradiance
GRI	Groud Reflected Irradiance
GRU	Gated Recurrent Unit
GSR	Global Solar Radiation
IEA	International Energy Agency
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
METU NCC	Middle East Technical University Northern Cyprus Campus
ML	Machine Learning
MLFNN	Multilayer Feed-Forward
MLP	Multilayer Perceptron
MSE	Mean Squared Error
n-RMSE	Normalized Root Mean Square Error
NWP	Numerical Weather Prediction
PV	Photovoltaic
R^2	Coefficient of Determination

RBF	Radial Basis Function
RES	Renewable Energy Sources
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SLP	Single-Layer Perceptron
SVM	Support Vector Machine
SVR	Support Vector Regression
TRNC	Turkish Republic Of Northern Cyprus
WNN	Wavelet Neural Network
ReLU	Rectified Linear Unit

CHAPTER 1

INTRODUCTION

1.1 Motivation

Energy is an import aspect of economic growth [1]. Growth in the economy results in improved life quality which, along with increasing population, contributes to the rise in energy demand. Until 2019, global energy demand was increasing and expected to reach 30% by 2040 [2]. However, in the light of recent events, i.e. global pandemic and resulting worldwide lockdown, International Energy Agency (IEA) [3] has presented in the latest analysis that the global energy demand has decreased by 3.8% compared to 1st quarter of 2019, and will likely to decrease further by 6% shall the lockdowns continue in the coming months, and economic recoveries take place slowly. In the same analysis, it is reported that the demand for conventional energy sources, namely coal (by 8%), oil (by 5%), and natural gas (by 2%), as well as nuclear power, has decreased. In contrast, demand for Renewable Energy Sources (RES) increased by 1.5% so far.

As a result of global warming caused by greenhouse gas emissions from conventional energy sources, many countries have been focusing on RES to meet the increasing energy demand [4]. RES play a crucial role in combating global warming by reducing the energy produced from conventional sources [5]. Additionally, RES is suggested to be taking part in increasing life quality as well as contributing to the development of the economy [5]. More countries are expected to integrate renewable energy sources, specifically solar energy, into their energy supply in the coming years [4], [6]–[8].

Among the renewable energy sources, solar energy is the main focus of interest as there is tremendous growth in solar photovoltaic (PV) system installation in many countries [4], [7], [9], including Northern Cyprus. PV panels convert sunlight, i.e. solar radiation, into electricity [10]. However, the PV output is very intermittent and unstable, as it is a weather-dependent energy source [8], [11]. PV cannot produce electricity in the absence of sunlight. [8], [11]. However, the production quantity varies due to variables such as the variation of the ambient temperature, humidity, cloud movements, etc. [8], [10]. As its energy output is unstable, it makes the difficult task of balancing demand and supply of electricity in the isolated electrical grids, e.g. like in islands, even more challenging [6], [9], [10], [12]–[14].

Additionally, Voyant et al. [15] stated that a grid manager should know about a PV production at least one hour ahead due to delay in starting a power generation system. With the knowledge of PV power output, the grid manager would know the amount of power to be added to the grid at a particular time [16]. Thus, the production from conventional energy sources can be decreased according to the power output of the PVs. On another note, the PV power output knowledge is very valuable in the smart grids in terms of power scheduling, unit commitment and grid regulation [17].

Extensive integration of the solar energy to the existing or future electricity grids enhances the need for radiation forecasting as it helps mitigate the intermittency by giving information about the future energy production [6]–[9], [12], [18]–[22]. Forecasting solar radiation is a crucial and ongoing task on different time-horizons for various power system applications, as stated by many researchers [6], [8], [10], [13], [14], [22]–[25]. Figure 1.1 illustrates the application points of short-term GHI information concerning the prediction horizon.

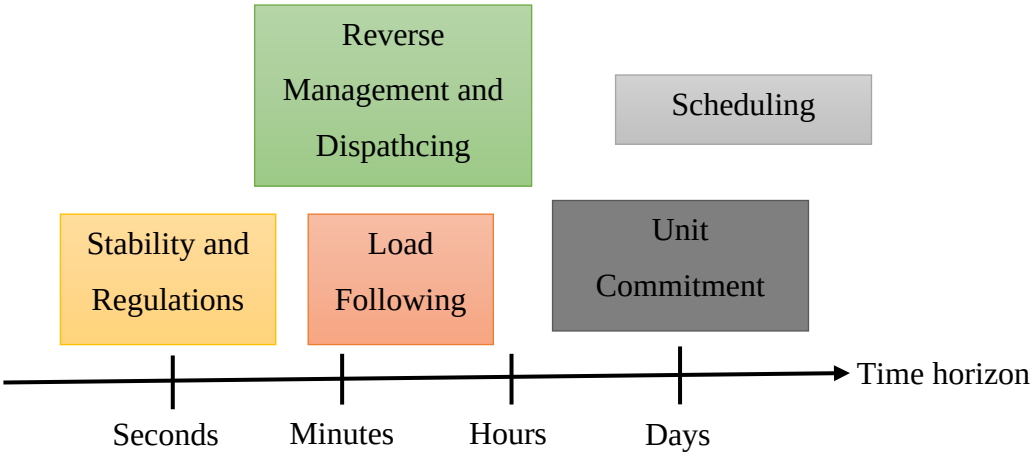

Figure 1.1. Short-term radiation forecasting scale for typical target applications [8]

In addition to the areas mentioned, knowledge in solar radiation data could be utilised in infrastructure and maintenance planning [18], [22], [26], [27], solar energy-related policymaking [20], and energy storage options which depend on the knowledge of solar energy production [6], [13]. Additionally, apart from utilisation in energy-related areas, radiation information is needed in various tools to access climate impacts on agriculture [28].

1.2 Objectives

This study aims to address the gap in the literature in terms of having an adequate GHI estimation model for Kalkanlı and surrounding regions for the safe and sustainable integration of solar energy to the electrical grid in Northern Cyprus. The effects of seasonality are also investigated. To achieve this objective, firstly, the datasets, which are NASA and Middle East Technical University Northern Cyprus Campus (METU NCC), are analysed and preprocessed so that the forecasting algorithms can easily interpret the feature of the data. Next, prediction algorithms are constructed using Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) and Support Vector Regression (SVR). Five algorithms are designed for the NASA dataset; four stand-alone algorithms: LSTM and CNN with two different datasets, and one hybrid algorithm of CNN and LSTM.

For the METU NCC dataset, two forecasting methods, i.e. annual and seasonal forecasting, are adopted. In annual forecasting, all samples in the dataset regardless of seasons are used to estimate GHI. Stand-alone algorithms of CNN, LSTM and SVR, and hybrid algorithms of CNN and LSTM, i.e. C-LSTM and CN-M, and SVR, i.e. CM-SVR, are constructed in annual forecasting. In seasonal forecasting, however, the dataset is separated into four sub-datasets depending on the months of the seasons, i.e. summer, fall, winter and spring. For each season, a hybrid algorithm of CNN, LSTM and SVR, i.e. CM-SVR, is constructed and trained with samples of the corresponding season separately. After training and finalising all models, the performances of each model are evaluated by employing commonly used evaluation metrics.

1.3 Overview of the Thesis

The thesis is made of five chapters. The background information and summaries of the related work on solar radiation are provided in Chapter 2. In Chapter 3, the study area and dataset descriptions are presented as well as detailed explanations for the prediction networks modelled for the study. Chapter 4 provides the results of the modelled forecasting networks, comparison, and discussion. Finally, conclusions and recommendations are presented in Chapter 5.

1.4 Publication Related to this Study

Vakitbilir N., Hilal A., Direkoğlu C. (2021) Prediction of Daily Solar Irradiation Using CNN and LSTM Networks. 14th International Conference on Theory and Application of Fuzzy Systems and Soft Computing – ICAFS-2020. ICAFS 2020. Advances in Intelligent Systems and Computing, vol 1306. Springer, Cham. https://doi.org/10.1007/978-3-030-64058-3_28 [29]

CHAPTER 2

BASIC CONCEPTS AND REVIEW OF RELATED WORK

In this section, initially, basic concepts related to radiation and radiation forecasting are briefly explained. Radiation prediction related research are then provided in the following section.

2.1 Basic Concepts

2.1.1 Radiation

Solar radiation that reaches the PV panel is classified into four categories. Direct Normal Irradiance (DNI) is the radiation type that reaches the surface without disruption, while Diffuse Horizontal Irradiance (DHI) is scattered by the atmosphere, e.g. by the clouds. The other type, Ground Reflected Irradiance (GRI), is reflected from the ground as the name suggests. The last type is the Global Horizontal Irradiance (GHI), which is the amount of radiation reaching a horizontal surface. Figure 2.1 illustrates the solar radiation incidents on a tilted surface.

DNI is a useful component for concentrating solar technologies. On the other hand, for PV panels, GHI is the relevant radiation type to be predicted. If GHI cannot be measured directly, it could be computed by (2.1) [30].

$$\mathrm{GHI} = \mathrm{DHI} + \mathrm{DNI} \cdot \cos(\theta) \tag{2.1}$$

where θ refers to the angle between the beam radiation and the vertical line, i.e. solar zenith angle.

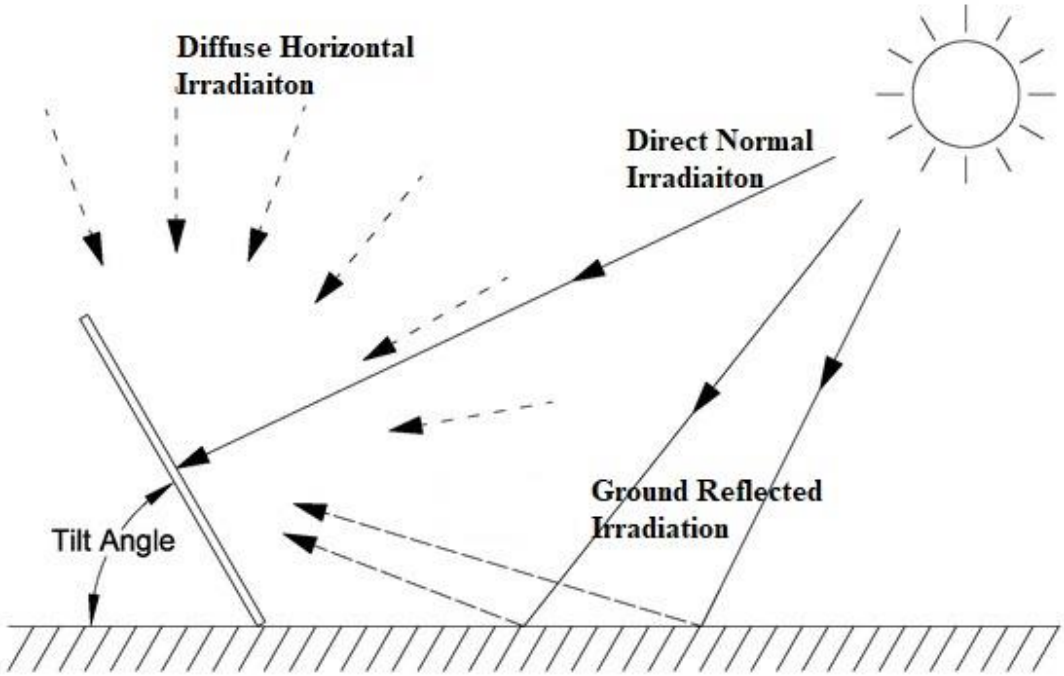

Figure 2.1. Types of solar radiation incidents on a tilted surface [31]

The amount of solar radiation reaching the ground changes drastically depending on the place on the Earth since the incoming radiation angle affects the climate of a location at different latitudes. In Figure 2.2, the spatial distribution of long-term averaged GHI values around the world are illustrated. As mentioned previously, PV production is strongly affected by local atmospheric conditions of a region [10], [19]–[21] since GHI is affected by weather variations, which also results in non-linear characteristics of GHI [32]. Therefore, solar radiation forecasting model should be developed for a specific region using the corresponding region's climatic variables [21].

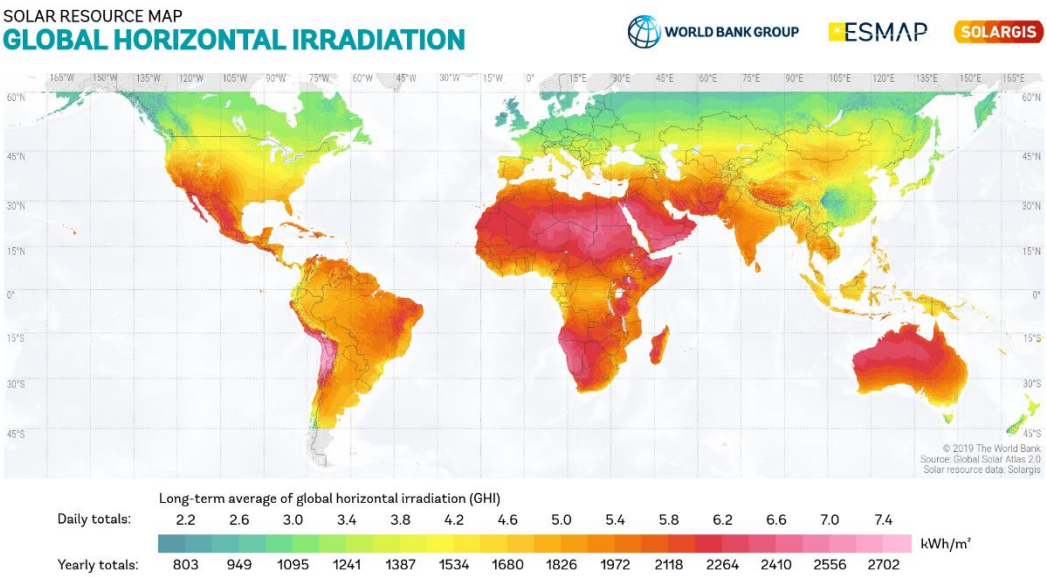

Figure 2.2. Spatial distribution of averaged GHI around the world [33]

2.1.2 Forecasting Algorithms

For the GHI forecasting, there are several approaches used in the literature. These approaches include the physical models, empirical models, statistical models, Numerical Weather Prediction (NWP) models, image-based models, and machine learning (ML) algorithms. Physical models, such as relative sunshine-based broadband model, the physical-based model for the tropical environment and the efficient physical-based model to name a few, correlate sky and atmospheric conditions to radiation through mathematical formulations [34].

Similar to physical models, empirical models also employ mathematical formulations, in which various meteorological parameters are used. Empirical methods are categorised mainly as sunshine-based, cloud-based, temperature-based, and other meteorological parameter-based models [35]. On the other hand, statistical models use statistics based on historical data, i.e. time-series data, to predict future values. Autoregressive model, persistence, and k-nearest neighbour interpolator are the common statistical models [36].

NWP models employ mathematical models utilising explanatory variables, e.g. cloud motion and direction, as input data [6], [32]. Image-based models use satellite images for prediction. They are proven to be very effective models for radiation forecasting; however, image availability, along with real-time image processing, among others, cause this model to be unpractical [32].

ML algorithms which are classified as artificial intelligence models solve problems that the explicit algorithms cannot represent [6]. Linear algorithms simply try to fit a line over a set of data points, as shown in Figure 2.3, by computing one weight for each input variable. Support Vector Regression (SVR), Linear Regression, Decision Tree, XGBoost, etc., are commonly used linear learning algorithms.

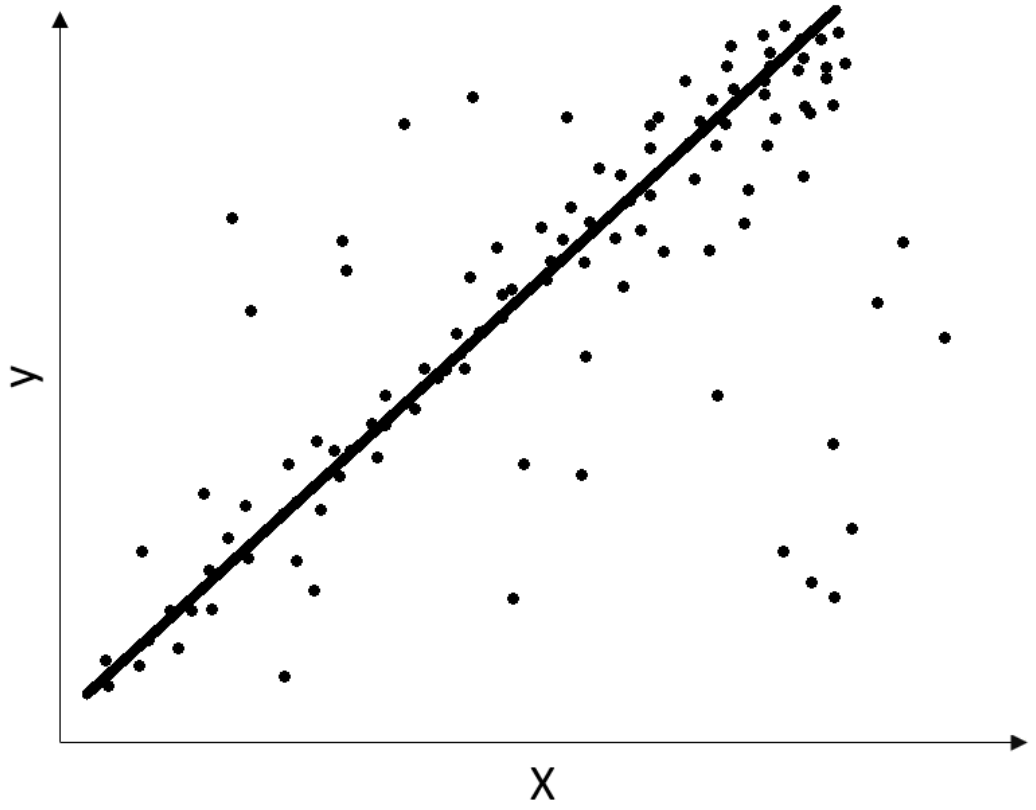

Figure 2.3. A linear ML algorithm example; linear regression fitting a line over a set of data points where y represents actual values, while X represents predicted values

Artificial Neural Network (ANN) algorithms, on the other hand, are non-linear self-adaptive ML techniques for processing the information just the way a brain does

[37]. They are composed of interconnected elements, commonly known as neurons, nodes or units [38]. The neurons placed in two or more layers interact through weighted connections, as shown in Figure 2.4. ANN learns from experience, i.e. past data, and develop a relationship between a set of input and output parameters even when the underlying relations are complex and non-linear [6], [21], [37]. Since the forecasting is the prediction of the future by understanding the past, ANN algorithms become a great applicant for the task [37]. When it comes to GHI estimation, many researchers [26], [39], [40] proved that ML algorithms outperform NWP models, as a result of non-linear and complex characteristics of GHI.

ANN's could be formed in two different ways. One way is a Feed-Forward Neural Network (FFNN), with two components: input and output vector. The information obtained from an input vector is used to calculate an output vector where the weights are decided explicitly [41], [42]. A simple FFNN is illustrated in Figure 2.4. Convolutional Neural Network (CNN), Single-Layer Perceptron (SLP), Multilayer Perceptron (MLP), Extreme Learning Machine, and Radial Basis Function (RBF) are some of the FFNN algorithms. The other way to form an ANN is Recurrent Neural Network (RNN), which contains loops where output is fed back to its input [43]. Long-Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) are the most effective RNN models.

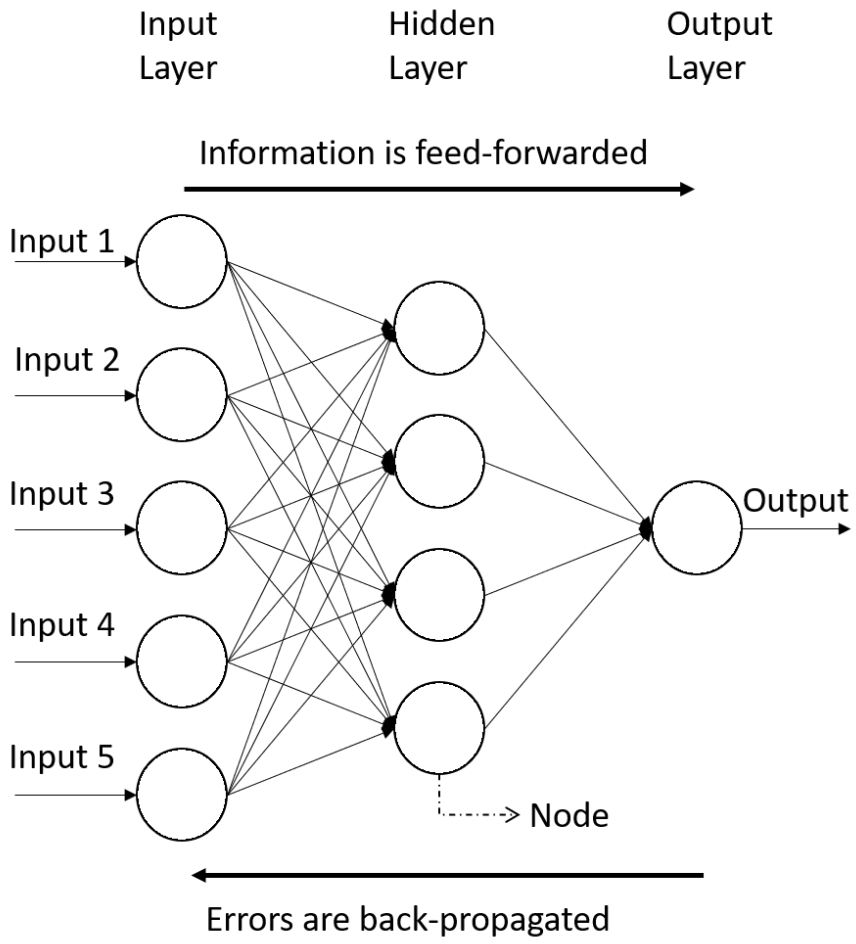

Figure 2.4. Structure of a simple ANN showing the flow direction of information and errors though feed-forward and back-propagation, respectively

ANN algorithms are made of 3 layers called the input layer, hidden layer and output layer, which can be seen in Figure 2.4. If the hidden layer consists of more than one layer, the ANN algorithm is called Deep Learning (DL) algorithm. The multiple layers in the algorithm improve the overall performance of the neural network enabling it to learn sophisticated relations and correlations that are way ahead of the traditional ML algorithms [22], [44]. In DL algorithms, weights are adjusted through gradient vector, which represents, for each weight, the increase or decrease in error if the weights were increased with a small amount [45].CNN, RNN, LSTM, Diagonal Recurrent Wavelet Neural Network (DRWNN), Wavelet Neural Network (WNN) are some of the DL algorithms.

In ML, the learning process can proceed through two different approaches: supervised and unsupervised learning. In supervised learning, the output value is introduced to the algorithm along with input data; hence the algorithm tries to minimise the error accordingly. In unsupervised learning, on the other hand, the ML algorithm is fed only by a group of patterns, and the algorithm itself tries to settle down to a steady-state after several iterations [42].

In supervised learning, depending on the output, the ML algorithm could be either classification or regression. In classification, the outcome is a categorical, i.e. discrete, value while in regression, the output is numerical, i.e. continuous, value. This study uses regression in forecasting algorithms.

2.1.3 Model Input Variables

The GHI forecasting analysis can be carried out in two different methods, i.e. annual forecasting and seasonal forecasting. In annual forecasting, all data points in a dataset are used to train and test the forecasting algorithm. On the other hand, in the seasonal forecasting, designed algorithms are trained and tested with separate sub-datasets that are created depending on the months of the season. Recent studies mostly conducted in the Mediterranean region have been utilising seasonal forecasting method.

Additionally, two different datasets can be used in GHI prediction. The first dataset is called time-series or historical data, which is a sequence of observations ordered through equally spaced time intervals. Time-series data is commonly used in the literature for short-term GHI prediction [26]. The second dataset includes meteorological variables, i.e. features. These variables include but not limited to sunshine duration, ambient temperature, relative humidity, wind speed, wind direction, pressure, date, time and so on [21]. Seldom, geographical variables, i.e. longitude, altitude and elevation, are used along with meteorological features in GHI forecasting.

Forecasting models can be based on time-series dataset or meteorological and geographical dataset. They can also be based on a hybrid dataset which considers both radiation and meteorological dataset as input features [6]. Aggarwal and Saini [46] refer to time-series forecasting, which uses past data of GHI values as input data, as endogenous forecasting and the hybrid forecasting as exogenous forecasting, which is also called multivariate forecasting. In this study, datasets are represented as endogenous or exogenous.

Although the endogenous dataset is commonly used in the literature, Ferrari et al. [36] suggest utilising meteorological data to improve the learning model as an outcome of their research.

2.1.4 Model Evaluation Metrics

Various most common score matrices in regression model evaluation are adapted in this study to evaluate the performance of the prediction models. These evaluation models are Mean Absolute Error (MAE) in Wm^{-2}, Mean Absolute Percentage Error (MAPE) in %, Root Mean Square Error (RMSE) in Wm^{-2}, normalised RMSE (n-RMSE) and Coefficient of Determination (R^2), the mathematical formulation of whom are illustrated in Equation (2.2) to (2.6), respectively [4].

$$\mathrm{MAE} = \frac{1}{\mathrm{N}}\sum_{\mathrm{i=1}}^{\mathrm{N}}\left|\mathrm{GHI_{r,i}} - \mathrm{GHI_{p,i}}\right| \tag{2.2}$$

$$\mathrm{MAPE} = \frac{1}{\mathrm{N}}\sum_{\mathrm{i=1}}^{\mathrm{N}}\left|\frac{\mathrm{GHI_{r,i}} - \mathrm{GHI_{p,i}}}{\mathrm{GHI_{r,i}}}\right| \tag{2.3}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{N}}\sum_{\mathrm{i=1}}^{\mathrm{N}}(\mathrm{GHI_{r,i}} - \mathrm{GHI_{p,i}})^2} \tag{2.4}$$

$$\mathrm{nRMSE} = \frac{1}{\mathrm{GHI_{r,i}}}\sqrt{\frac{1}{\mathrm{N}}\sum_{\mathrm{i=1}}^{\mathrm{N}}(\mathrm{GHI_{r,i}} - \mathrm{GHI_{p,i}})^2} \tag{2.5}$$

$$R^2 = \frac{\sum_{i=1}^{N}(GHI_{p,i} - GHI_{m,i})^2}{\sum_{i=1}^{N}(GHI_{r,i} - GHI_{m,i})^2} \quad (2.6)$$

where GHI_r, GHI_p and GHI_m are the i^{th} measured (real), predicted and mean GHI values, respectively, while N is the number of data points.

In forecasting problems where only small errors are accepted, RMSE is employed instead of its predecessor [8]. nRMSE is frequently calculated instead of RMSE for a meaningful comparison. Mohammadi et al. [47] defined ranges for nRMSE in order to measure a model's performance. The ranges are tabulated in Table 2.1.

Table 2.1. Model performance classified according to nRMSE value

nRMSE	*Model Precision*
< 0.10	Excellent
0.10 – 0.20	Good
0.20 – 0.30	Fair
> 0.30	Poor

RMSE and MAE being close to each other as value means that the forecast model has only small deviations from the real data [48]. Additionally, R^2 measures how well the predictions fit the data. In other words, it illustrates the difference between the predicted values and the variance of the errors [49]. The value of R^2 varies between zero to one where zero means that the regression forecasting poorly fit the data while one means perfect fit. Finally, Yadav & Chandel [21] has classified MAPE results of radiation forecasting in terms of forecasting accuracy, which is tabulated in Table 2.2. Although it is commonly used, MAPE has many disadvantages argued by several researchers [50]–[52]. Resulting in biased and underestimated results, and its inability to deal with zero predictions are the most prominent disadvantages of MAPE. Therefore, in this study, we do not consider MAPE in our model evaluation methods.

Table 2.2. Classification of prediction accuracy with corresponding MAPE value

% MAPE	*Prediction Accuracy*
≤ 10%	High
10% – 20%	Good
21% – 49%	Reasonable
≥ 50%	Inaccurate

In addition to error metrics, the absolute difference between prediction and forecast value, i.e. Absolute Prediction Error (APE) is evaluated through histograms. APE is calculated through Equation (2.7).

$$\mathrm{APE} = \left|\mathrm{GHI}_{r,i} - \mathrm{GHI}_{p,i}\right| \tag{2.7}$$

The input data is usually divided into three sets, that are called training, validation and testing sets. The training set is used to fit the model, and the validation set is used for hyperparameter tuning while the testing set is used to evaluate the final model [53]. Both training and testing sets are evaluated with an error metric. The resulting difference between the two error values gives an idea of the model's overall performance.

Lower training error compared to testing errors suggests overtrained, i.e. overfitted, model. Overtraining is the result of high variance in which the model learns outliers, i.e. noise. A overfitted model fails to generalise its output to fit unseen data [6]. On the other hand, when the model cannot recognise the underlying patterns in the dataset, the model said to be underfitted. Underfitting causes the model to lose its ability to determine the relationship between the actual and predicted output, which is also known as high bias.

A visual representation of the decrease in error value over the training and testing sets as the complexity of the model increases, which is also known as bias-variance trade-off, is illustrated in Figure 2.5.

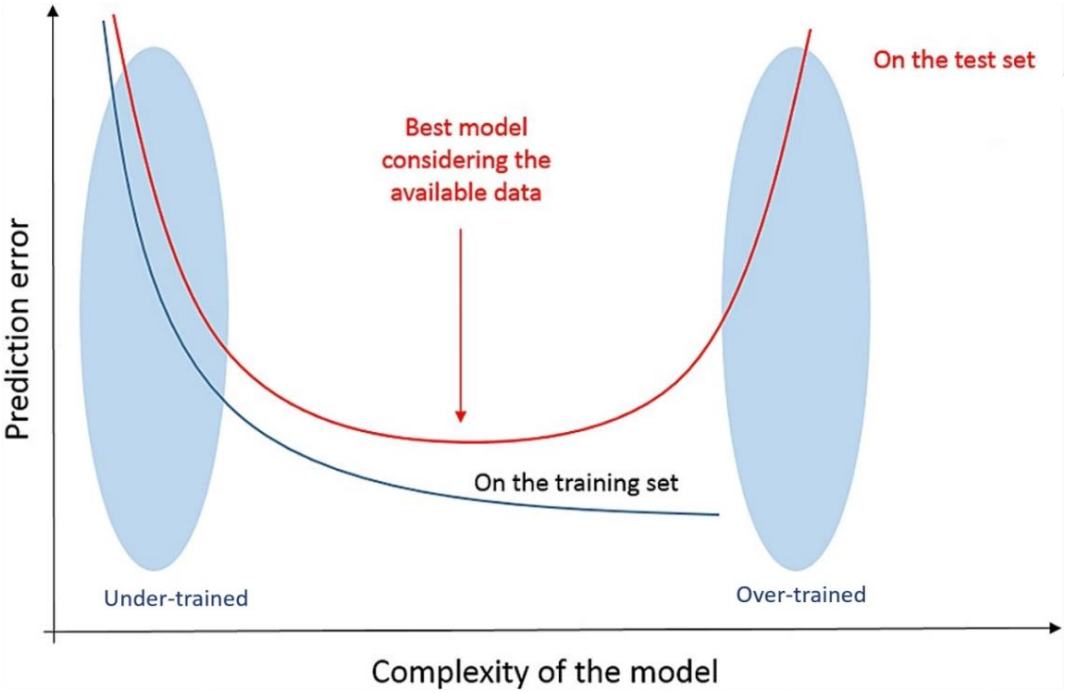

Figure 2.5. Bias-variance trade-off graph [6]

2.2 Review of Related Work

There is a vast extent of radiation forecasting studies in the literature, as mentioned earlier. For instance, some focuses on forecasting GHI, while others focus on DNI prediction. Furthermore, the methods used for radiation forecasting also differ significantly, e.g. numerical correlations, image classification, DL algorithms, etc. Literature review of this study focuses exclusively on ML employing studies. Selected studies are grouped depending on the region studied in the following sections.

2.2.1 GHI Estimation Around the World

In this section, the research concerning the radiation forecasting for the various parts of the world is presented in the following paragraphs.

Ghimire et al. [4] developed a combined model of CNN and LSTM, i.e. CLSTM, for prediction of half-hourly GHI for Alice Springs, Australia. They compared the novel model to stand-alone models that are CNN, LSTM, RNN and DNN. It is found that using hourly endogenous data of GHI give MAPE of 4.84% for CLSTM, which outperforms the rest of the models.

Alzahrani et al. [8] applied an LSTM model to forecast short-term GHI for a solar farm in Canada. The model uses GHI values of four climatic conditions: a few clouds, scattered clouds, overcast, and clear sky. The developed model was then compared to SVR and FFNN models. Results show that LSTM performs better than the other two models with nRMSE of 0.086.

Qing and Niu [13] built an LSTM model to predict hourly day-ahead GHI for Santiago, Cape Verde. They compared the LSTM model with persistence, LR, and Back-Propagated Neural Network (BPNN) models. The models take hourly temperature, dew point, humidity, visibility, wind speed, and meteorological type as input parameters, whereas GHI as the output parameter. It is found that the LSTM model performs better than the rest of the models with RMSE of 76.24 W m^{-2}.

Fan et al. [18] used SVR and XGBoost models to estimate daily GHI as a function of temperature and precipitation data for three different stations in China. Later, they compared the performance of these two models with four commonly used empirical models. It is found that SVR performs better than the rest of the models with an average value of 3.14 MJ m^{-2} day^{-1}.

Cao and Lin [54] elaborated DRWNN model to forecast hourly and daily GHI for Shanghai and Macau, China using GHI time-series data. They compared their result with sunshine based empirical models, i.e. Collares-Pereiraand Rabl, Ångström – Prescotta, and BPNN algorithm. RMSE is found to be 0.96 MJ m^{-2} day^{-1} and 0.048 MJ m^{-2} h^{-1} for daily and hourly GHI prediction, respectively. The results indicate that the DRWNN model performs better than the reference models.

Khosravi et al. [55] carried out a comparative study of three-layer Feed-Forward Neural Network (MLFNN), Radial Basis Function Neural Network (RBFNN), SVR, Fuzzy Inference System (FIS) and neuro-FIS (ANFIS) models to estimate GHI using seven years data for Abu Musa Island, Iran. They have applied two different datasets to each model. The first group uses the data of pressure, temperature, wind speed, relative humidity and local time, and the second one uses endogenous data of GHI. Results show that the models with meteorological input achieve relatively lower RMSE, and the lowest RMSE is obtained for SVR with 1.05 Wh m^{-2}.

Zang et al. [56] proposed CNN-LSTM hybrid model for forecasting short-term GHI for 34 stations in Texas, USA. In this proposed model, the two dimensional-CNN (2D-CNN) algorithm is fed with meteorological data while the LSTM algorithm receives GHI time series data as input. Outputs of these two algorithms are then combined in a fully connected layer before the final output. The hybrid model's performance is compared to hybrid 2D-CNN-ANN, hybrid ANN-LSTM, and stand-alone ANN, 2D-CNN, LSTM, and SVR. The resulting MAE and RMSE show that in almost all stations. The 2D-CNN-LSTM hybrid model achieves the lowest error values between 37.20 to 52 W/m^2 and 69.26 to 86.33 W/m^2, respectively. The authors also studied the effects of seasonal prediction. In seasonal forecasting, hybrid models perform better than the stand-alone models where 2D-CNN-LSTM (2CL) hybrid model achieving the lowest errors in most of the seasons and stations. The studies presented in this section are listed in Table 2.3 for easier observation and comparison.

Table 2.3. Summary of the research related to the GHI forecasting for various part of the world

Reference	*Dataset*	*Models*	*Best Result*	*Country*
Ghimire et al. [4]	Endogenous data	CLSTM CNN LSTM RNN DNN	$MAPE_{CLSTM}$ = 4.84%	Alice Springs, Australia
Alzahrani et al. [8]	Endogenous data	LSTM SVR FFNN	$nRMSE_{LSTM}$ =0.086	Canada
Qing and Niu [13]	Meteorological data	LSTM LR BPNN	$RMSE_{LSTM}$ = 76.24 W m^{-2}	Santiago, Cape Verde, USA
Fan et al. [18]	Meteorological data	SVR XGBoost	$RMSE_{SVR}$ = 3.14 MJ m^{-2} day^{-1}	China
Cao and Lin [54]	Endogenous data	DRWNN	$RMSE_{hourly}$ = 0.048 MJ m^{-2} h^{-1} $RMSE_{daily}$ = 0.96 MJ m^{-2} day^{-1}	Shanghai and Macau, China
Khosravi et al. [55]	Exogenous data	MLFNN, RBFNN, SVR, FIS, ANFIS	$RMSE_{SVR}$ = 1.05 Wh m^{-2}	Abu Musa Island, Iran
Zang et al. [56]	Meteorological data, Endogenous data	2CL, ANN-LSTM, ANN, 2D-CNN, LSTM, SVR	MAE_{2CL} = 37.20 – 52 W m^{-2} $RMSE_{2CL}$ = 69.26 – 86.33 W m^{-2}	Texas, USA

2.2.2 GHI Forecasting for Mediterranean Region

Cyprus is located in the Mediterranean region. Therefore, analysing the GHI prediction research for this area is essential to have a better understanding of the effects of the climate and incoming radiation angle on radiation forecasting. In this section, selected studies for the Mediterranean region are demonstrated in the following paragraphs.

Guariso et al. [14] estimated hourly-GHI using FFNN and LSTM. The prediction models utilise time-series data of GHI for Milan, Italy. They compared their models to persistence and clear-sky models. The MAE and RMSE of FFNN and LSTM models are found to be in similar ranges with average values of 41.46 and 82.83 W m^{-2}, respectively, which are relatively lower than achieved by the persistence and clear-sky models.

Voyant et al. [15] compared three different models on hourly GHI prediction for five locations in south-east France. The models are based on ANN and Autoregressive Component Moving Average (ARMA) algorithms. They also introduced seasonality into their forecasting models. For annual prediction in each location, the hybrid model of ANN and ARMA give better results than the stand-alone ANN and ARMA, with nRMSE ranging between 0.13 to 0.17. In seasonal estimation, the hybrid model performs the best in most of the locations and seasons. The lowest nRMSEs are achieved in summer season in all stations, while the highest nRMSEs are observed in the winter season.

Sozen et al. [25] applied ANN to estimate monthly GHI for Turkey using three years of longitude, latitude, altitude, month, mean sunshine duration, and mean temperature values of 17 stations all across Turkey. They compared their results to the that of in the literature. The results show that ANN-based models perform better than the classical regression models. MAPE varies from 2.92% to 6.74%.

Belaid and Mellit [57] built several SVR models for daily and mean-monthly forecasting of GHI for Ghardaïa, Algeria. The input data is an exogenous dataset

consisting of max, min, mean temperature values, and daily GHI. The SVR models are built using different combinations of input data. It is found that using temperature, GHI and sunshine duration as input variables give nRMSE as 0.13, and MAPE as 10.40%.

Mazorra Aguiar et al. [58] modelled annual and seasonal hourly-GHI forecasting for two stations in Gran Canaria Island, Spain. They compared the performance of an ANN model to persistence model and NWP model. The input data to the algorithms consist of GHI, humidity, temperature, and satellite images. Both in annual and seasonal forecasting, ANN model outperforms rest of the models. The resulting RMSE ranges from 82.90 to 105.30 W m^{-2}. In seasonal prediction, the lowest error is observed in the summer while the highest errors are observed in the winter season.

A comparative summary of the studies mentioned above conducted in the Mediterranean region is listed in Table 2.4.

Table 2.4. Summary of the studies concerning the GHI forecasting for several countries in the Mediterranean region

Reference	*Dataset*	*Models*	*Best Result*	*Country*
Guariso et al. [14]	Endogenous data	FFNN LSTM Persistence Clear-sky	$RMSE_{avg(FFNN\ \&\ LSTM)}$ = 82.83 W m^{-2} $MAE_{avg(FFNN\ \&\ LSTM)}$ = 41.46 W m^{-2}	Milan, Italy
Voyant et al. [15]	Exogenous data	ANN ARMA Hybrid	$nRMSE_{hybrid}$ = 0.13 – 0.17	South-east France
Sozen et al. [25]	Endogenous dataset	ANN	MAPE = 2.92% – 6.74%	Turkey
Belaid and Mellit [57]	Exogenous data	SVR	nRMSE = 0.13 MAPE = 10.40%	Ghardaïa, Algeria

Mazorra Aguiar et al. [58]	Exogenous data and satellite images	ANN Persistence NWP	$RMSE_{ANN}$ = 82.90 – 105.30 W m^{-2}	Gran Canaria Island, Spain

2.2.3 Prediction of Radiation for Cyprus

As mentioned previously, due to radiation's dependency on the local climate, radiation forecasting algorithms should be modelled using the region's climatic variables. In this section, the radiation prediction studies for Cyprus are presented.

Tymvios et al. [59] compared the performance of ANN and Ångström sunshine-based empirical model in forecasting GHI for Nicosia, the Republic of Cyprus (ROC) using hourly GHI and sunshine duration. The results show that ANN performs better with R^2 of 0.92 and normalised RMSE of 0.063.

Jacovides et al. [60] investigated several numerical correlations on forecasting hourly DNI for Athalassa, ROC. They used numerical models to measure GHI and DHI from radiometric data. The lowest nRMSE is obtained as 0.34.

Tapakis and Charalambides [61] predicted GHI for Limassol, ROC using cloud motion detection. The classification accuracy is found as 95%.

Kasht [62] carried out a sky condition classification and GHI prediction study using one-year data of hourly GHI time-series and temperature data for Kalkanlı, Northern Cyprus (TRNC). She firstly applied K-means cluster to the daily clearness index calculated from GHI values. The resulting cluster information of three sky conditions is fed to support vector machine (SVM) to further characterise the sky conditions. For GHI estimation, simple regression is applied on hourly GHI and temperature data of only clear sky days. The RMSE of GHI forecasting is achieved as 0.14 for June.

The studies mentioned previously for Cyprus are summarised in Table 2.5.

Table 2.5. Summary of radiation forecasting studies conducted for Cyprus

Reference	*Dataset*	*Models*	*Best Result*	*Country*
Tymvios et al. [59]	Exogenous data	ANN Ångström model	$R^2_{ANN} = 0.92$ $nRMSE_{ANN} = 0.063$	Nicosia, ROC
Jacovides et al. [60]	Radiometric data	Numerical correlations	nRMSE = 0.34	Athalassa, ROC
Tapakis and Charalambides [61]	Cloud images	Cloud motion detection	Accuracy = 95%	Limassol, ROC
Kasht [62]	Exogenous data	Simple regression	RMSE = 0.14	Kalkanlı, TRNC

Reviews of various forecasting algorithms for different radiation types are given in the references [6], [21], [63]–[66].

2.3 Gaps in the Literature

As a conclusion of the literature review, it is identified that an adequate GHI forecasting model is required for Northern Cyprus. The justification for this conclusion could be explained in the following paragraphs.

First of all, as mentioned previously, the received radiation is affected by the local climate of a region. Hence, a forecasting algorithm should be modelled for an area with data that belongs to that region.

Secondly, for the Mediterranean region, the seasons have distinguishable characteristics. For summer and most of the spring, clear sky condition is observed, which results in a smoother transition in the received GHI. Whereas, in the winter and most parts of fall, overcast sky condition is observed, in which there are sharp fluctuations in received GHI. Thus, seasonality plays an essential role in GHI

estimation in the Mediterranean region. Several authors [15], [56], [58], [67] from Mediterranean countries applied seasonality to their GHI prediction research in recent years, while Cyprus lacks such a study. Hence, seasonality prediction should be applied to GHI forecasting for Cyprus.

Additionally, in Cyprus island, the spatial distribution of GHI differs depending on the landform, which is illustrated in Figure 3.1. In other words, the received average GHI value for the mountains, i.e. Kyrenia Mountains and Trodos Mountains, is lower than than the plains of Mourphou. Hence, GHI estimation on a regional scale depending on a predominant landform is a necessity.

Finally, Sperati et al. [68] concluded in their benchmarking study within the European Actions "Weather Intelligence for Renewable Energies" framework that more research is needed on short-term energy forecasting using different models, locations, and data for a complete overview of all possible scenarios around the world representing all possible meteorological conditions.

Furthermore, CNN and LSTM algorithms recently started attracting attention for GHI forecasting. Researchers have been testing performances of these algorithms, as single and hybrid. CNN and LSTM algorithms have significant potential in terms of GHI prediction. Therefore, CNN and LSTM algorithms are employed in this study. Besides, single algorithms, hybrid of the two algorithms are constructed as well as a hybrid model combined with SVR, which is also a very well established algorithm for GHI prediction.

Therefore, in this study, it is aimed to obtain short-term and long-term GHI forecasting algorithm for Kalkanlı in order to help PV integration to the electrical grid for continuous sustainable renewable energy growth in Northern Cyprus.

CHAPTER 3

MATERIALS AND METHODOLOGY

This thesis aims to carry out a GHI forecasting study for Kalkanlı, Northern Cyprus, using machine learning algorithms. Information on the study area, data used, and the algorithms employed are presented in detail in this chapter in the following sections.

3.1 Study Area

This study is a case study for Middle East Technical University Northern Cyprus Campus (METU NCC) Kalkanlı, Northern Cyprus, placed in the northern part of Cyprus island. Cyprus is an Eastern Mediterranean island located at 35°N and 33°E in the Mediterranean Sea. Mediterranean climate dominates over the island, resulting in a semi-arid climate with average temperatures of 30℃ and 13℃ in summer and winter, respectively. The summers are dry and mostly sunny on the island, while the winters are rainy and cloudy.

Cyprus has excellent potential for receiving solar radiation. Figure 3.1 illustrates the spatial distribution of GHI potential over Cyprus. The yearly average GHI potential is 5.4 kWh m^{-2} [69].

In Northern Cyprus, the majority of the electricity demand has been supplied by conventional energy sources. In recent years, the developments and affordability of PV panels have resulted in many households to install PV panels over their rooftops. There are also two PV-farms in Northern Cyprus that are placed in METU NCC, Kalkanlı and Serhatköy. Kalkanlı shown with a red mark has a great photovoltaic (PV) power potential.

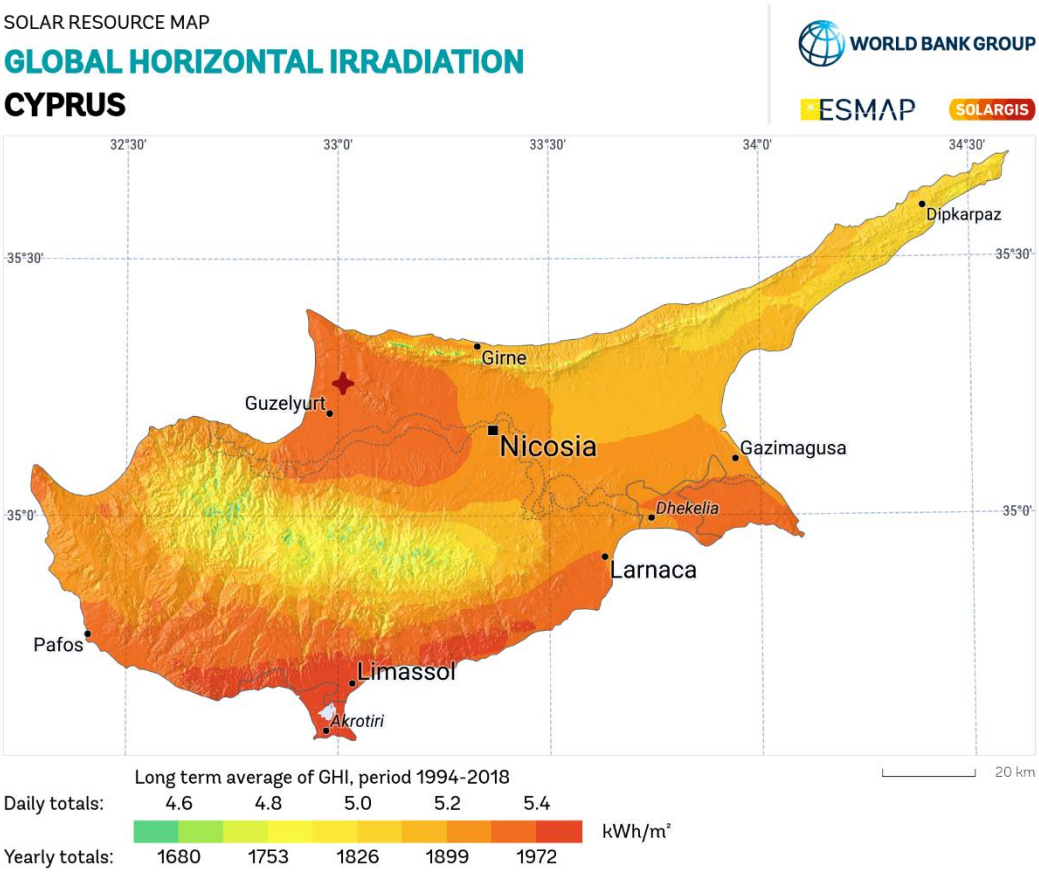

Figure 3.1. The spatial distribution of GHI over Cyprus island [33]

3.2 Data

This thesis is carried in two separate parts. Forecasting algorithms are built with data taken from the NASA web archive in the first part and with data received from METU NCC in the second part. The following sections explain each dataset in further detail.

3.2.1 NASA Dataset

One of the datasets used in the first part of this study is obtained from NASA as daily values, available at [70]. NASA supplies data for various meteorological variables and radiation types. However, only relative humidity (RH), pressure (P), average,

minimum and maximum temperature (T, Tmin, and Tmax, respectively), wind speed (WS), wind direction (WD), radiation (GHI), and corresponding month (M) and date (D) information for Kalkanlı at the latitude of 35.26 and longitude of 33.02, Northern Cyprus between 1983 and 2019 are extracted for this study. The radiation data is originally obtained as kWh m^{-2} day^{-1} but converted to W m^{-2} for consistency with the literature. A sample set of features representing the data used in the first part of this study is given in Table 3.1.

Table 3.1. A sample set of NASA data

M	*D*	*RH*	*P*	*T*	T_{max}	T_{min}	*WD*	*WS*	*GHI*
7	1	50.7	99.38	27.38	31.78	23.3	270.1	5.23	704.17
7	2	63.13	99.27	24.99	28.44	22.08	270.3	6.96	620.83
7	3	65.57	99.14	24.92	28.25	21.94	266.38	7.35	606.67
7	4	64.92	99.13	24.32	27.07	21.63	263.84	7.18	641.67
7	5	61.97	99.22	25.94	29.18	22.82	270.54	6.11	679.17

The average GHI for the study area is 431 W m^{-2} over the whole dataset. In Table 3.2, the minimum and maximum data points of the input variables are listed along with their units. The variables' units are given only for convenience and do not affect the learning algorithms.

Table 3.2. Units and data ranges of input variables for NASA dataset

Input variables	*Units*	*Data range*
GHI	W m^{-2}	7.50 – 766.67
T	℃	4.61 – 33.29
T_{min}	℃	2.70 – 29.78
T_{max}	℃	6.35 – 38.72
Wind speed	m/s	0.56 – 13.48
Wind direction	degrees	0.00 – 359.95
RH	%	34.21 – 89.80
P	kPa	97.32 – 101.59

Figure 3.2 demonstrates the temporal distribution of GHI data points over the whole NASA dataset in three-dimension. The highest data points of GHI are observed during the summer season, while the lowest points are obtained in the winter season.

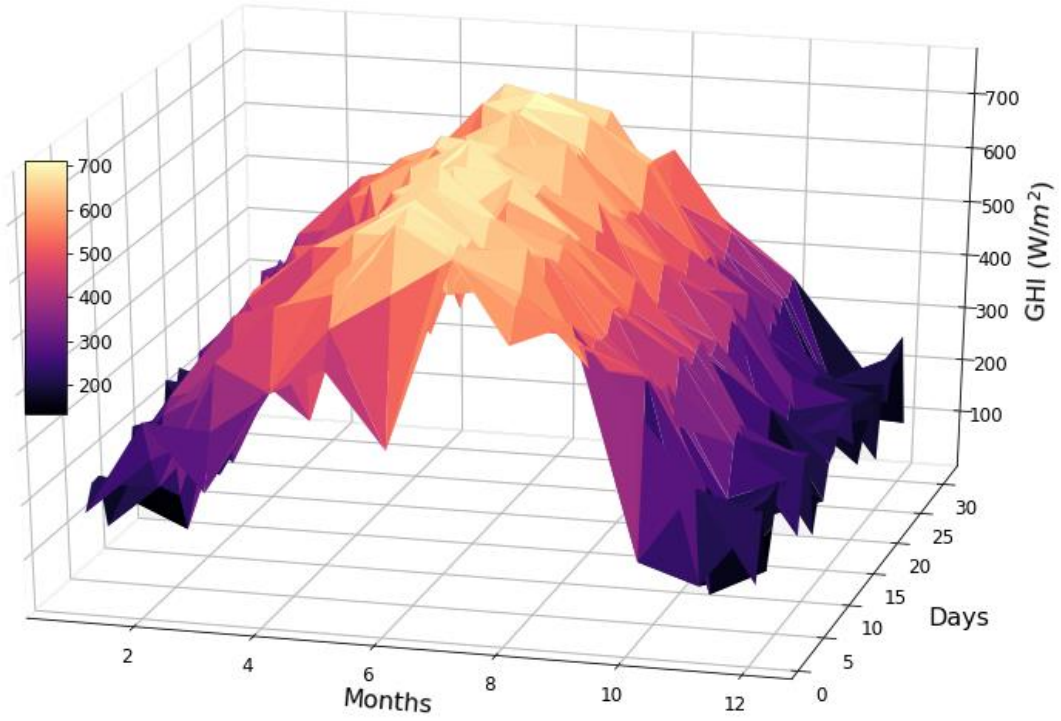

Figure 3.2. Temporal distribution of GHI throughout the NASA dataset

A two-dimensional representation of the change in GHI over the year 2018 is shown in Figure 3.3, where the fluctuations in GHI can be observed more clearly. In the winter season, observed GHI is relatively lower than that observed in the Summer season. On the other hand, the fluctuations in GHI is higher in the fall, winter, and spring season while in the summer season, the GHI values are more steady.

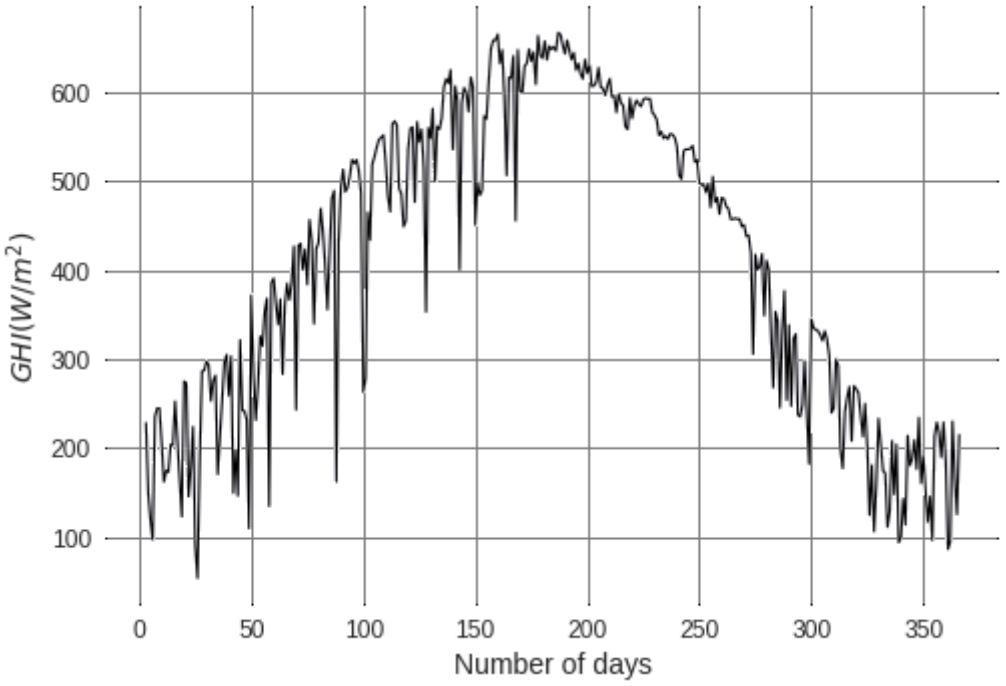

Figure 3.3. Change in GHI values over the year 2018

A box plot illustrating the distribution of GHI values for each month for the whole dataset is displayed in Figure 3.4. From the box plot, it is observed that most of the months except the months of summer have a wide range of data points, resulting from the sharp fluctuations in the GHI values.

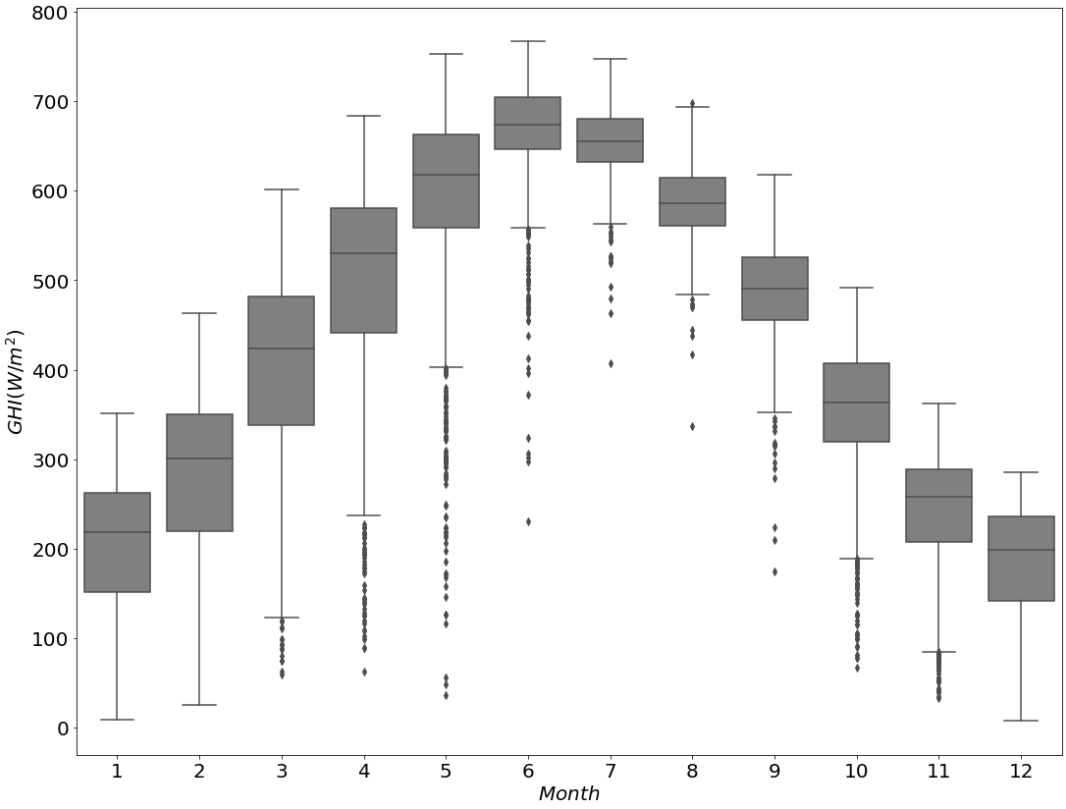

Figure 3.4. Distribution of daily GHI data throughout the whole NASA dataset

In Figure 3.5, the temporal distribution of temperature data points over the whole NASA dataset is illustrated. The highest data points of temperature, similar to GHI values, are observed during the summer season, while the lowest points are recorded in the winter season.

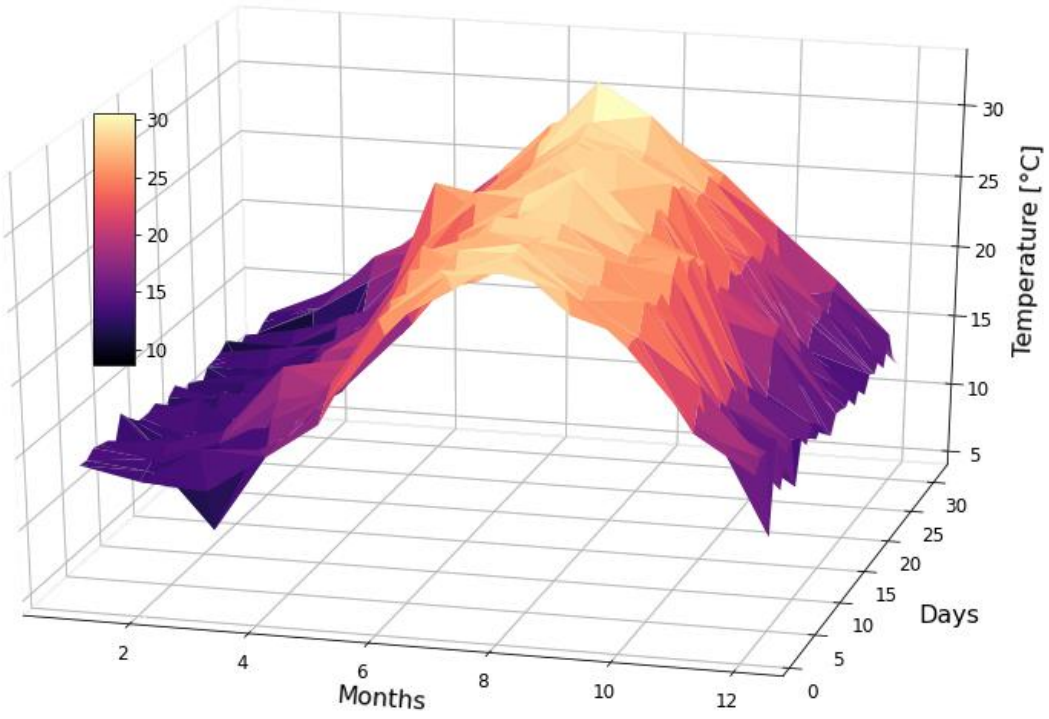

Figure 3.5. Temporal distribution of temperature over NASA dataset

Figure 3.6 presents a two-dimensional representation of temperature change over the year 2018. When Figure 3.3 and Figure 3.6 are compared, it can be concluded that there is a connection between GHI and temperature since both curves show a similar trend.

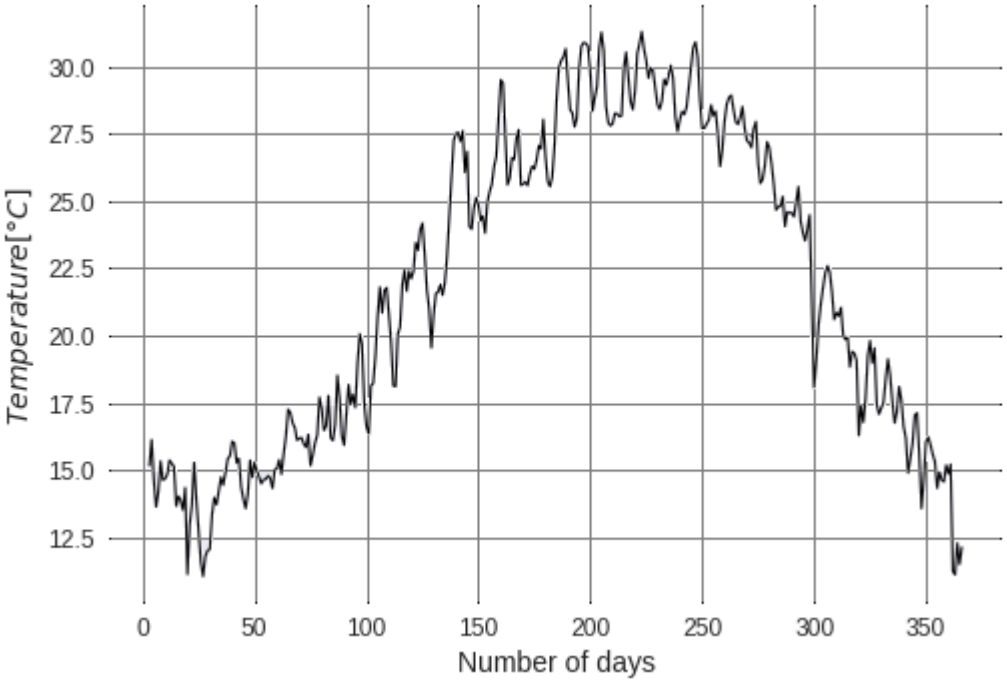

Figure 3.6. Change in temperature values over the year 2018

3.2.2 METU NCC Dataset

The second dataset used for the second part of this study is recorded in the PV farm in METU NCC, Kalkanlı. There were two separate data files initially. The first data file contained meteorological variables such as temperature, wind speed on various height, humidity, etc., recorded at the wind tower between 2013 and 2017. In contrast, the second data file included GHI and DNI data recorded on the PV farm between 2010 and 2017. All data are recorded at 10 minutes intervals. Both data files are matched and combined in a single dataset. The resulting dataset consists of radiation (GHI), relative humidity (RH), pressure (P), temperature (T), wind speed at various elevation (WS60, WS50, WS40, and WS30), wind direction (WD), with the corresponding month (M), day (D), hour (H), and minute (M) information between 2013 and 2017.

Table 3.3 shows a sample set of the variables available in the METU NCC dataset. First and last three rows of data are shown for a complete representation. It should be noted that the time is in the 24-hour format.

Table 3.3. A sample set from the METU NCC dataset

M	*D*	*H*	*M*	*WS60*	*WS50*	*WS40*	*WS30*	*WD*	*T*	*RH*	*P*	*GHI*
2	19	6	50	1.81	2.03	2.34	2.23	69	10.53	79.83	1001	23.00
2	19	7	0	1.65	1.66	1.52	1.44	94	11.63	76.42	1001	46.00
2	19	7	10	0.96	0.77	0.71	0.68	103	12.48	75.32	1002	91.00
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
6	15	19	20	4.48	4.34	4.13	3.81	291	24.20	69.09	998	82.00
6	15	19	30	3.25	3.13	2.95	2.78	280	23.96	69.95	997	41.00
6	15	19	40	3.05	2.84	2.63	2.33	263	23.83	70.68	997	20.00

The average GHI is 409 W m^{-2} over the whole METU NCC dataset, while it is 510 W m^{-2} for the summer season and 275 W m^{-2} for the winter season. Minimum and maximum available data points for each variable are listed in Table 3.4

Table 3.4. Input variables' units and data ranges in METU NCC dataset

Input variables	*Units*	*Data range*
GHI	W m^{-2}	2.00 – 1247.00
T	℃	-0.34 – 41.78
Wind speed	m/s	0.24 – 25.15
Wind direction	degrees	0.00 – 359.00
RH	%	6.59 – 95.95
P	mbar	975 – 1195

Figure 3.7 demonstrates the temporal distribution of GHI data points over the whole METU NCC dataset in three-dimension. The summer season during noon, highest GHI values are obtained. Figure 3.7 illustrates that the change in GHI has higher

fluctuations during winter and spring seasons, while a smoother curve is obtained during summer to fall seasons.

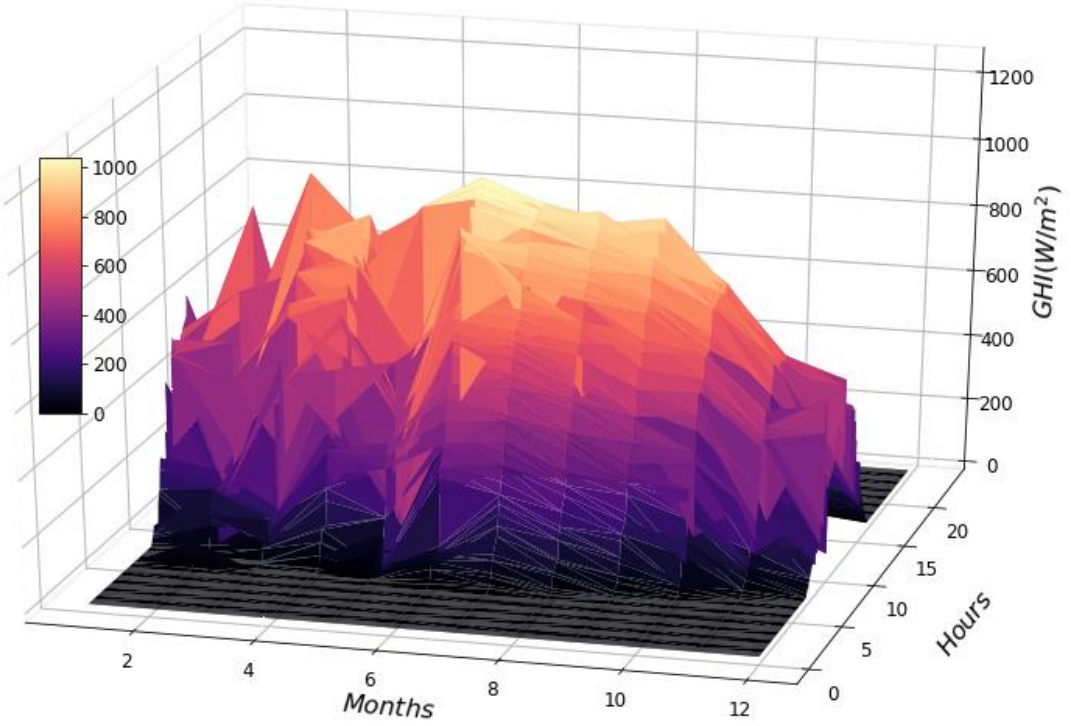

Figure 3.7. Temporal distribution of GHI throughout the METU NCC dataset

Two-dimensional representations of the change in GHI values over a week in February, March, July and October months to represent all four seasons are illustrated in Figure 3.8 - (a), (b), (c) and (d), respectively. The x-axes show the continuous number of data points in a 10-minutes time interval, whereas the y-axes show the change in GHI amounts. The fluctuations that are observed in Figure 3.7 in the winter and spring seasons, i.e. November to May, are clearly illustrated on a daily basis in Figure 3.8 - (a) and (b). These fluctuations in the winter season are due to overcast sky condition. The smooth curves observed in Figure 3.7 in the summer and fall seasons,i.e. June to October, on the other hand, are shown in detail in Figure 3.8 - (c) and (d) which are a result of clear sky.

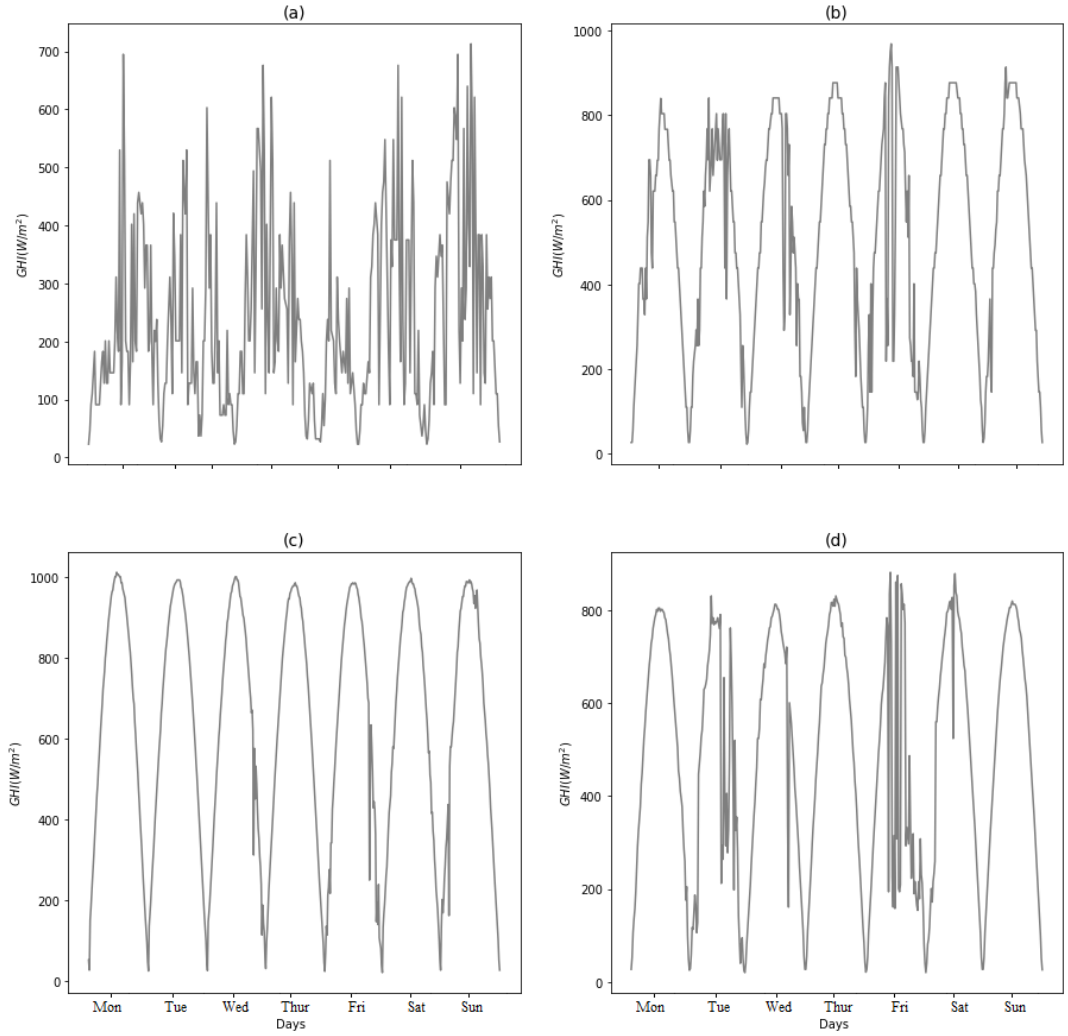

Figure 3.8. Change in GHI values over a week in (a) February, (b) March, (c) June, and (d) October

A box plot illustrating the distribution of GHI values over a day for the whole dataset is displayed in Figure 3.9. It is observed from the plot that the outliers are occurring prominently in the morning and the afternoon. The outliers are mostly caused by the difference between summer and winter seasons.

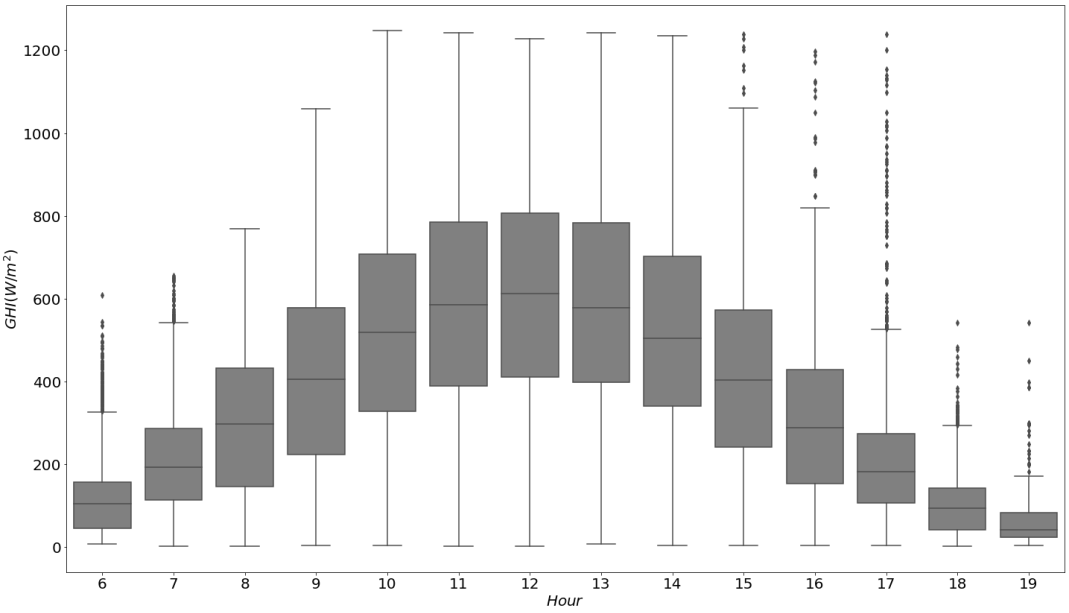

Figure 3.9. Distribution of GHI data in 10-minute interval throughout the whole METU NCC dataset

In Figure 3.10, the temporal distribution of temperature data points over the whole METU NCC dataset is illustrated. The highest data points of temperature, similar to GHI values, are observed during the summer season, while the lowest points are recorded in the winter season.

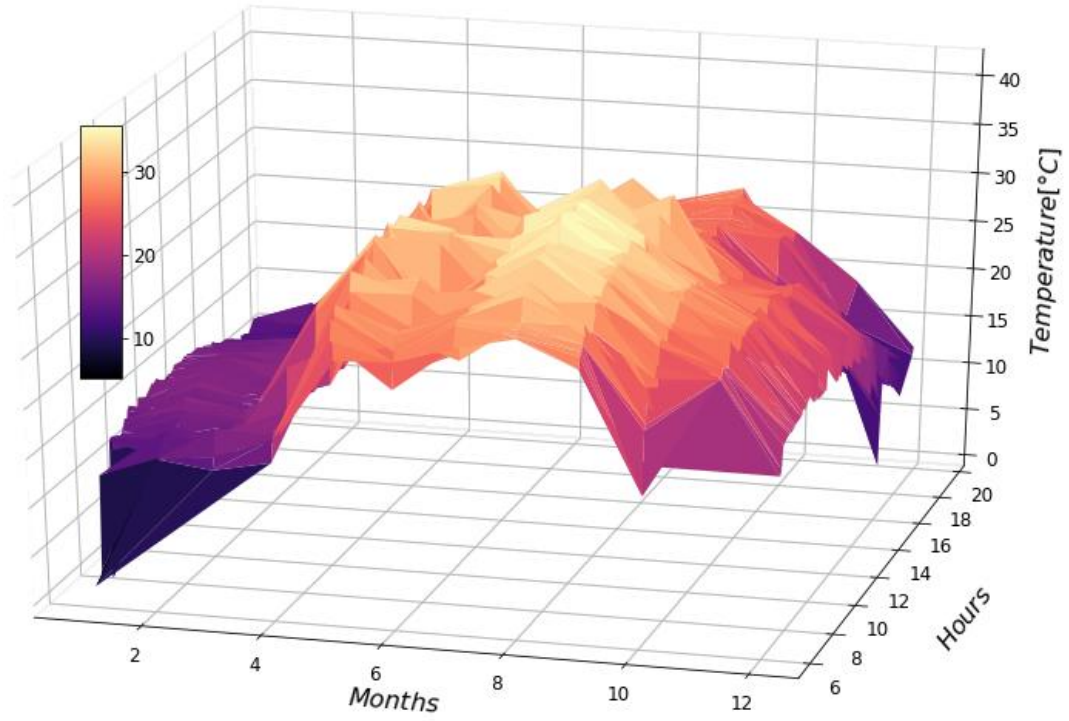

Figure 3.10. Temporal distribution of temperature over METU NCC dataset

3.3 Forecasting Algorithms

CNN, LSTM and SVR algorithms are employed in this study for GHI forecasting in Kalkanlı. The algorithms and the main activation functions used in the algorithms are elaborated in the following chapters.

3.3.1 Activation Functions

Activation functions are mathematical equations, usually non-linear, that derive a node's output in neural networks [71]. There are various activation functions such as sigmoid function, threshold function, piecewise linear function etc. [71]. In this study, Rectified Linear Unit (ReLU), and sigmoid function (σ), which are defined in Equations (3.1) and (3.2), respectively, are employed in the forecasting algorithms.

$$\mathrm{ReLu}(x) = \max(0, x) \tag{3.1}$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{3.2}$$

3.3.2 Convolutional Neural Networks (CNN)

CNN is one of the most popular ANN algorithms commonly used in the deep learning field [72]. One of the main advantages of using CNN is its powerful ability to achieve non-linear feature extraction [73]. Various architectures of CNN are available in the literature, resulting from different combinations and numbers of layers. These layers are mainly made of convolutional layer, pooling layer and fully connected layer, i.e. dense layer [74].

3.3.2.1 Convolutional Layer

The convolutional layer's main aim is to generate feature maps from the input data using filters, i.e. neurons composed of kernels [4]. Each kernel is used to generate one feature map. An activation function is applied to introduce non-linearity to the convolutional layer. ReLu, Sigmoid and tanh are the commonly used algorithms as activation functions in CNN. Each kernel represents a different weight matrix. The weight values and a bias term are updated during the training phase. The mathematical formula of the convolutional layer is shown in Equation (3.3) [22], [43],

$$y_{i,j,k}^{l} = F((w_k^l)^T x_{ij}^l + b_k^l) \tag{3.3}$$

where the weight and bias of *k*th convolutional kernel in the *l*th layer are represented as w_k^l and b_k^l, respectively. x_{ij}^l is the input patch in the *l*th layer, concentrated at the location *(i,j). F()* represents the activation function. All regions of the input are shared with the weight w_k^l which reduces the training time and the complexity of the

network. In Figure 3.11-a, the convolutional layer is presented in which all the mentioned steps to produce future maps are shown.

3.3.2.2 Pooling Layer

The primary purpose of pooling layer is to decrease the resolution of the feature map. Usually, this layer is used between two convolutional layers. The mathematical representation of the pooling layer is shown in Equation (3.4) [22], [43],

$$P^{l}_{i,j,k} = Pool(y^{l}_{m,n,k}) \tag{3.4}$$

where, $(m,n) \in R_{i,j}$ which represent the region around the location (i,j). *Pool ()* describes the type of pooling operation used in the layer. Average pooling and max pooling are the pooling operations that are used most often. The pooling layer usually increases network accuracy while decreasing the training time by reducing the number of parameters in the network [43]. Both the pooling function block and the downsized matrices are presented in Figure 3.11-b.

3.3.2.3 Fully Connected Layer

This layer's main task is to perform high-level reasoning by transporting the learned feature in the network to one space [74], as shown in Figure 3.11-c. The fully connected layer, also called dense layer, connects each neuron from the previous layers to every neuron in the current layer to create meaningful global information. Typically one or more dense layers are presented in CNN models after convolution and pooling layers [43]. The last dense layer generates the network output.

There are various branches of CNN available. One dimensional CNN (1D-CNN) and two-dimensional CNN (2D-CNN) are commonly used CNN algorithms in the literature. The former is widely used to process numerical data such as meteorological variables and energy production, while the latter is frequently used for image and text processing. Both types of CNN models are composed of the same

main layers. However, the main difference occurs in the convolutional layer. The kernels slide in two dimensions on the input data in the 2D-CNN while in 1D-CNN, the sliding happens only in one dimension [73]. A sample for a 1D-CNN is illustrated in Figure 3.11 with all four main layers. In this study, various 1D-CNN algorithms are constructed for predictions using 1D convolutional and 1D pooling layers available in Keras library [75] and is simply called CNN for convenience. In 1D-CNN, convolution involves sliding the filter over the input data which performs shift-multiply-sum procedure. In our implementation, this is done with cross-correlation (used in typical CNNs). The output data length is made equal to length of the input data using padding operation in our 1D-Convolutional layer implementation.

If needed, the output length can be made equal to the input length using padding and this mode is called ”same” padding convolution in keras

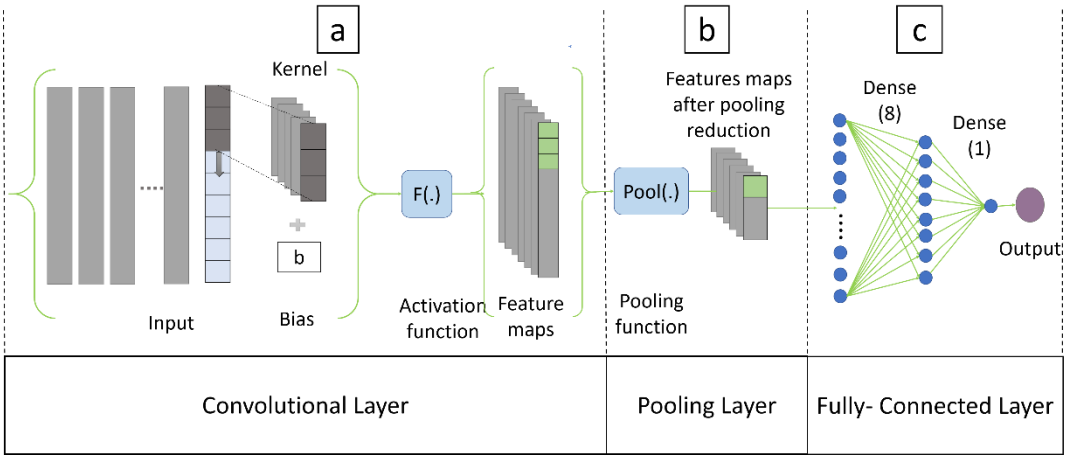

Figure 3.11. The architecture of a general 1D-CNN

3.3.3 Long Short-Term Memory (LSTM)

LSTM is one of the RNN architectures. RNN algorithms are capable of deriving relations between consecutive events. On the other hand, they become insufficient

when relating the long-range events because of gradient vanishing or gradient exploding [22]. Gradient vanishing refers to the fast-exponential decrease of the gradient norm to zero. In contrast, the exploding gradient refers to an opposite event, resulting in a network that cannot learn from long data sequences [76]. LSTM is introduced to overcome the gradient vanishing and exploding problems with its memory cell, first introduced by Hochreiter and Schmidhuber [77], and extra forget gate included by Gers et al. [78]. Memory blocks in LSTM that include input, output and forget gate allow updating and controlling information flow in separate blocks [4].

Figure 3.12 illustrates a sample structure for an LSTM block. Forget gate f_t, input gate i_t, intermediate state g_t and output gate o_t formulated in Equation (3.5) to Equation (3.8), respectively,

$$f_t = \sigma(W_{fx}X_t + W_{fh}h_{t-1} + b_f) \tag{3.5}$$

$$i_t = \sigma(W_{ix}X_t + W_{ih}h_{t-1} + b_i) \tag{3.6}$$

$$g_t = ReLu(W_{gx}X_t + W_{gh}h_{t-1} + b_g) \tag{3.7}$$

$$o_t = \sigma(W_{ox}X_t + W_{oh}h_{t-1} + b_o) \tag{3.8}$$

where σ refers to the non-linear activation function (sigmoid function), W_x and W_h are the weight matrices, and b is the bias of the relevant gates, X_t refers to input of the current time-step while h_{t-1} is the output of the previous time-step. Forget gate decides which information to keep from the previous memory cell (m_{t-1}), while the input gate determines the information to preserve in the current memory cell (m_t). m_t is then calculated as given in Equation (3.9),

$$m_t = g_t \odot i_t + m_{t-1} \odot f_t \tag{3.9}$$

where $\odot$ refers to Hadamard product. Then, the output gate decides of which memory cell to pass as output (h_t) as formulated in Equation (3.10),

$$h_t = ReLu(m_t) \odot o_t \tag{3.10}$$

The process given from Equation (3.5) to (3.10) continues taking place in the next time steps. Weights and biases are adjusted during the training by minimising the differences between the actual data and the predicted LSTM output. The predicted output of the LSTM ($\bar{y}_t$) is calculated by Equation (3.11),

$$\bar{y}_t = W_y h_t \tag{3.11}$$

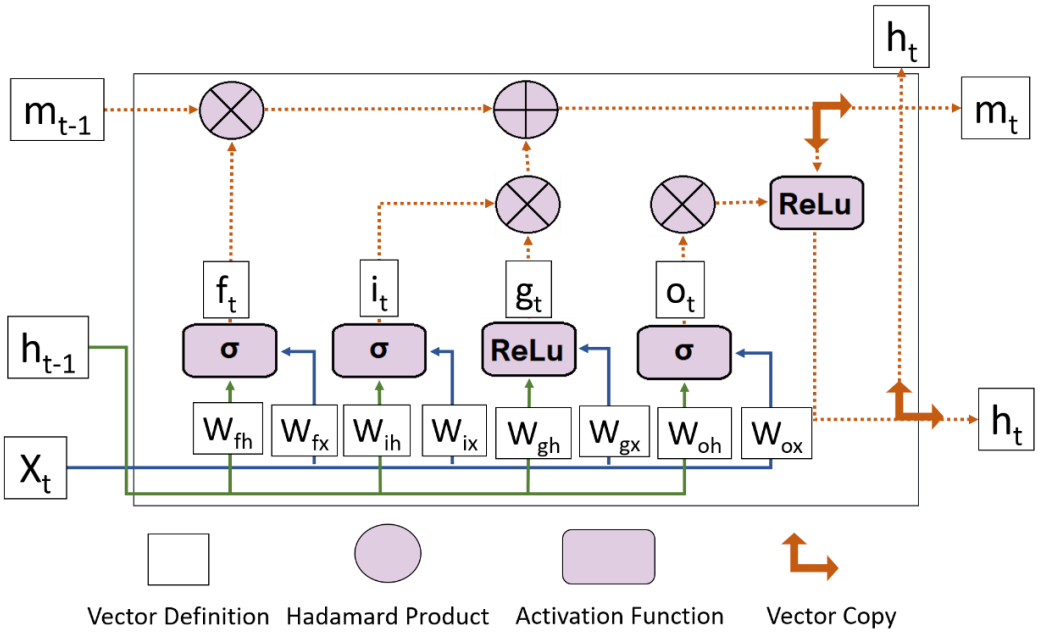

Figure 3.12. Sample structure of LSTM unit

3.3.4 Support Vector Regression (SVR)

SVR is an ML algorithm in which the input vectors are mapped into high dimensional feature space by non-linear mapping Φ where linear regression occurs to find a relationship between input and output vectors [79]. SVR algorithm assigns a linear hyperplane, called a decision boundary to estimate input and output data relation, which is then used to predict future values represented in Equation (3.12) [80],

$$f(x) = w \cdot \phi(x) + b \tag{3.12}$$

where w denotes learned weight vector, b is the threshold, $\Phi(x)$ is the mapping function, and $f(x)$ represents the predicted value.

There is a trade-off between minimising the training error and good generalisation behaviour. The algorithm aims to maximise the distance between the decision boundary and data points in order to control the trade-off and obtain a hyperplane with a good regression performance[81]. Equation (3.13) illustrates compound risk $R_{reg}(f)$ to balance the trade-off [81],

$$R_{reg}(f) = \frac{C}{N}\sum_{i=1}^{N} L_{\varepsilon}(f(x_i), y_i) + \frac{1}{2}\|w\|^2 \tag{3.13}$$

where C denotes to regularisation parameter, N is the sample size, $L_{\varepsilon}(f(x_i),y_i)$ refers to Vapnik's ε-insensitive loss function while $\|w\|^2$ is the complexity term related to the complexity of the model. $R_{reg}(f)$ results from model complexity and training errors and should be kept as low as possible. Vapnik's ε-insensitive loss function is defined by Equation (3.14) [79].

$$L_{\varepsilon}(f(x) - y) = \begin{cases} |f(x) - y| - \varepsilon & \text{for } |f(x) - y| \geq \varepsilon \\ 0 & \text{otherwise} \end{cases} \tag{3.14}$$

where ε is the maximum error specified by the user to achieve the model's desired error, $f(x)$ is the predicted and y is the actual value. When SVR is training, it solves Equations (3.15) and (3.16) [82].

$$\text{minimize } \frac{C}{N}\sum_{i=1}^{N}(\xi_i^* + \xi_i) + \frac{1}{2}\|w\|^2 \tag{3.15}$$

$$\text{subject to } \begin{cases} y_i - \langle w, x_i\rangle - b \leq e + \xi_i^* \\ \langle w, x_i\rangle + b - y_i \leq e + \xi_i \end{cases} \tag{3.16}$$

where ξ_i is the distance between the bounds, which are defined by ε, and the predicted values outside the bounds. A simple SVR model is illustrated in Figure 3.13.

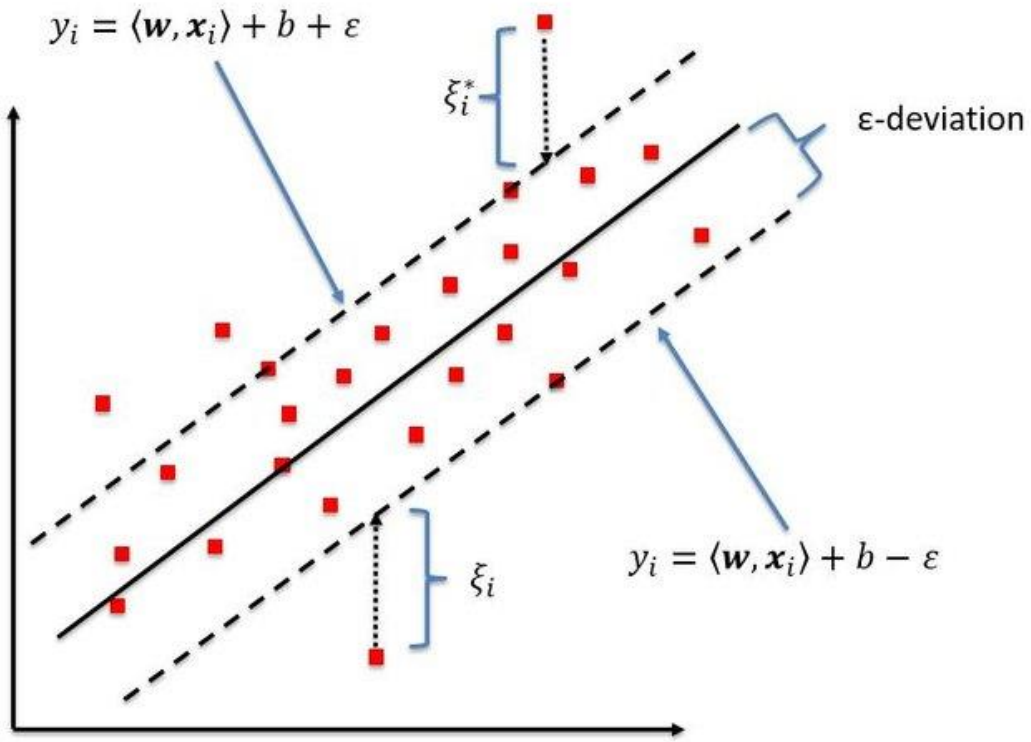

Figure 3.13. Sample structure of SVR model prediction [82]

3.4 Experimental Setup

The procedures followed for preprocessing the datasets and constructing the learning algorithms are explained in detail in the following sections. Forecasting models are constructed separately for each dataset.

3.4.1 Data Preprocessing

Preprocessing is transforming the data so that the algorithm can easily interpret the features of the data. Prior to the construction of the learning algorithms, each dataset is preprocessed in accordance with the steps shown in **Error! Reference source not found.**, which is established following Alzahrani et al. [8].

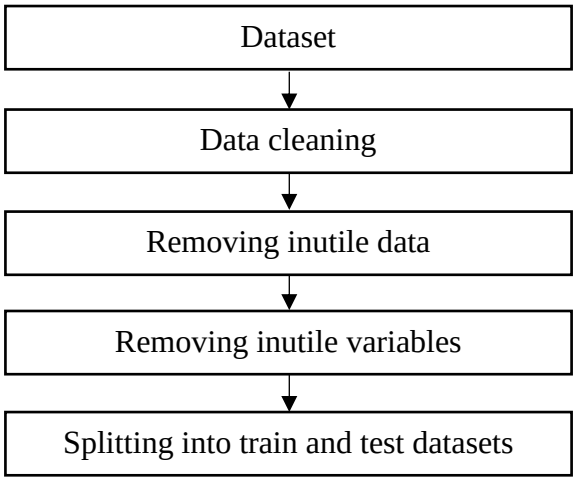

Figure 3.14. Flow chart for the preprocessing procedure for both datasets

In data cleaning step, missing data and incorrect data readings are eliminated from the dataset or replaced with the average of previous data points. For the NASA dataset, the missing data points exist as -999. The rows with a value of -999 are removed the dataset. Missing values are observed before 1st June 1983 and after 31st August 2019, so removal of these points did not affect the continuity of the dataset. In METU NCC dataset, there were several data points missing in GHI and temperature. The data points where several days of data are missing are filled using the average of previous years on the same date. The remaining points are filled by linear interpolation.

Following the data cleaning step, the non-useful data, i.e. night hours where GHI values are recorded as zero, are removed from the datasets. This step is particularly important as removing night hours leave only the meaningful data improving the prediction model's performance [8]. For the METU NCC dataset, data is trimmed between 6-7 am and 5-7 pm depending on the season. This step is skipped in NASA dataset since the data are daily values. Figure 3.15 illustrates the three-dimensional distribution of GHI over the whole METU NCC after the night hours are removed.

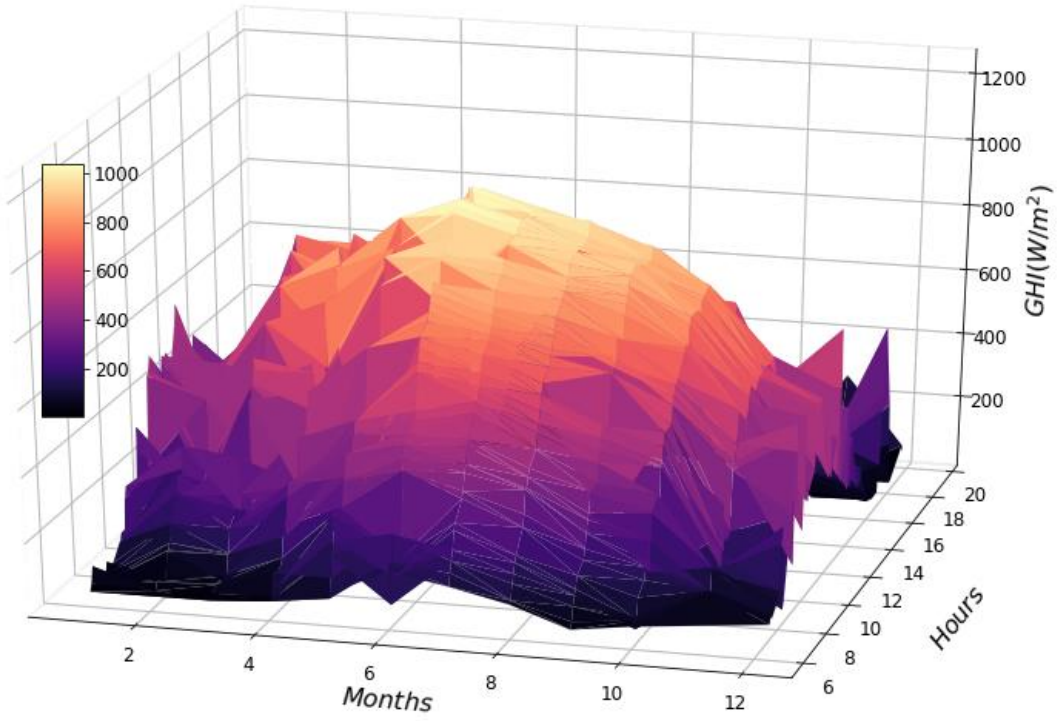

Figure 3.15. GHI distribution after the removal of night hours in METU NCC dataset

Next, the variables used as input for the algorithms are finalised in the variable removing step, which reduces the dimensionality of the dataset to increase the performance of the model. The removing inutile variables step is carried out manually taking similar studies from literature as a reference point. For this purpose, meteorological data in NASA dataset consist of mean, maximum and minimum temperature, wind speed, wind direction, pressure, relative humidity, and the corresponding month and day data along, as shown in Table 3.1. Similarly, the mentioned variables of NASA dataset are included in METU NCC dataset as well except for minimum and maximum temperature. The corresponding hour and minute information are added to METU NCC dataset, additionally. However, wind speed variables at 60, 50 and 40 meters are removed from the dataset since they have similar values as wind speed at 30 meters.

The variable removing step is followed by splitting the input dataset into training and testing datasets. The training dataset consists of 80% of the whole dataset, while the remaining 20% is left for the testing set, following the rule of thumb.

Additionally, 10% of the training set is used as a validation set in the training process to guide the parameter optimisation in each algorithm. Validation is used to prevent overfitting, and adjust hyperparameters such as learning rate, number of hidden layers and units in these layers, number of epochs and size of mini batches. The validation and testing sets are used instead of a cross-validation methods since these methods tend to have high computational cost on deep learning algorithms. This rule is applied to all datasets in all algorithms.

NASA dataset is used both exogenously and endogenously. There are meteorological variables and GHI data in the exogenous dataset, as mentioned in Table 3.1, while in the endogenous dataset, only GHI time-series data exist. For both of the dataset, separate forecasting algorithms are constructed.

METU NCC dataset is used as exogenous and endogenous in two different prediction methods. The first method is a widely used annual prediction method. The second method is seasonal prediction recently gaining attention from researchers to forecast GHI, specifically in the Mediterranean region. In seasonal forecasting, unlike annual forecasting, the years in the dataset is separated into seasons. As mentioned in the literature review, this method mostly results in better prediction results in the Mediterranean region. In annual forecasting, the performance of stand-alone models are compared to hybrid models' performance. On the other hand, only hybrid algorithms are developed for seasonal forecasting.

The dimension of the input tensor of all training and testing sets are in the form of (sample_size, time-step, features). Detailed information on the input dimensions for both NASA dataset and METU NCC exogenous dataset are given in Table 3.5. In the endogenous dataset, the number of features is one which refers to GHI time-series. The output of all algorithms is a scalar value.

Table 3.5. Input tensor dimensions of each dataset

Dataset	*Method*		*Sample_size*			*Time-step*	*Features*
			Training	*Validation*	*Testing*		
NASA	Annual		9507	1056	2641	7	10
METU NCC	Annual		75173	8353	20882	10,30,60	
	Seasonal	Summer	21098	2344	5861		
		Fall	16074	1786	4466		
		Winter	14676	1631	4077		
		Spring	23317	2591	6478		

3.4.2 Construction of Learning Algorithms

In this section, the construction of the prediction algorithms and the finalised model parameters are discussed. The algorithms are implemented in Python 3.7 [83], using freely available Keras [84], Tensor Flow [85] and Sklearn [86] libraries. For each dataset, different algorithms are built and presented in separate sections. The first section is about the prediction algorithms built for the NASA datasets, followed by the second section that covers the algorithms constructed for METU NCC datasets.

In this study, it is aimed to obtain a model that is able to predict short-term GHI using 10-minute interval data, i.e. METU NCC dataset. However, algorithm constructions are applied with the NASA dataset as well in order to have guiding models for the regions in Northern Cyprus where ground-level data collection is absent.

In the training process for deep learning algorithms, MSE is used for hyperparameter tuning, while the MAE is used as the objective function to be minimised in each dataset. Additionally, the learning rate for all the algorithms is kept at 0.01 while the activation function in CNN and LSTM layers are selected as ReLU. Also, the pooling layer in CNN is not added to the constructed algorithms since it decreases the algorithms' performance considerably.

3.4.2.1 Algorithms for NASA Dataset

In this study, we initially started working with NASA dataset for long-term GHI forecasting. For both exogenous and endogenous datasets generated from NASA data, separate CNN, LSTM and hybrid CNN and LSTM algorithms are developed. Algorithms are trained for GHI forecasting at 7-day lead times, i.e. time-step. In other words, with NASA dataset, we aimed to predict the coming eighth day, since GHI forecasting with a week ahead horizon could give valuable information on solar energy availability and sustainability during a robust solar-powered system design [4].

The parameters of each algorithm are optimised manually through several trials. Additionally, the layers and neurons are also configurated randomly until the optimum bias-variance is achieved, as was shown in Figure 2.5. Exogenous algorithms are trained with input dimensions (10563, 7, 10), and tested with dimensions (2641, 7, 10) where numbers represent the number of samples, the time-step, and the number of features, respectively. Whereas endogenous algorithms are trained with dimensions (10563, 7, 1), and tested with dimensions (2641, 7, 1). In all of the algorithms, the optimiser is selected as Adam after grid search.

CNN algorithm for the exogeneous dataset is composed of two convolutional layers followed by two fully connected layers. As mentioned previously, the loss is calculated by MAE while the metric is chosen as MSE in every neural network. The batch size and number of epochs are set as 550 and 100, respectively.

Another CNN algorithm is constructed for the endogenous dataset, which has three convolutional layers and three dense layers. As mentioned earlier, the pooling layer is not added to this algorithm as well due to performance issues. In this CNN algorithm, the batch size is optimised as 800 while the epochs are set to 100.

Following the construction of stand-alone CNN models, stand-alone LSTM algorithms are constructed for both datasets. The LSTM algorithm for the exogenous dataset is built with three LSTM layers and three dense layers. The recurrent

activation of the LSTM layer is kept as the sigmoid function. The batch size and number of epochs are set as 450 and 90, respectively.

The second LSTM algorithm is established for the endogenous dataset. The algorithm also has three LSTM layers, followed by three dense layers. The recurrent activation of the LSTM layer is again left as the sigmoid function. The batch size and number of epochs are set as 500 and 90, respectively.

Finally, a hybrid model of CNN and LSTM is designed where the outputs of CNN and LSTM are merged and fed into fully connected layers for a final GHI prediction. This hybrid algorithm is called CN-M for convenience. The CNN part of the hybrid algorithm is built to have two convolutional layers. LSTM, similarly, consists of three layers. The outputs of these two algorithms are then merged and fed into three fully connected layers. CNN is fed with exogenous data, and LSTM is fed with endogenous data. A flow chart showing the construction of the hybrid algorithm is illustrated in Figure 3.16. The batch size and number of epochs are set as 750 and 50, respectively.

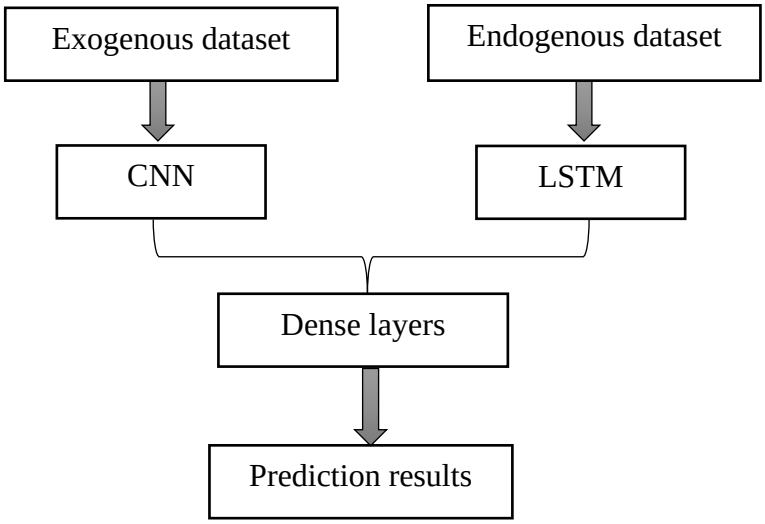

Figure 3.16. Flow chart showing the hybrid CN-M algorithm

Optimised hyperparameters of each layer in all NASA algorithms are presented in Table 3.6 along with the number of neuron and batch size.

Table 3.6. Training hyperparameters in each layer for the constructed forecasting algorithms

Layer (L)	*CNN - Exogeneous*	*CNN - Endogeneous*	*LSTM - Exogeneous*	*LSTM - Endogeneous*	*CN-M*
L1	Input	Input	Input	Input	Input
L2	Conv1d (f=64, k=4)	Conv1d (f=64, k=4)	LSTM (u=50)	LSTM (u=25)	Conv1d (f=68, k=3)
L3	Conv1d (f=128, k=8)	Conv1d (f=128, k=8)	LSTM (u=150)	LSTM (u=100)	Conv1d (f=128, k=3)
L4	Flatten	Conv1d (f=240, k=8)	Dense (40)	Dense (50)	Conv1d (f=240, k=3)
L5	Dense (100)	Flatten	Dense (20)	Dense (20)	Flatten
L6	Dense (50)	Dense (100)	Dense (1)	Dense (1)	LSTM (u=5)
L7	Dense (1)	Dense (50)	-	-	LSTM (u=10)
L8	-	Dense (1)	-	-	LSTM (u=30)
L9	-	-	-	-	Merge
L10	-	-	-	-	Dense (40)
L11	-	-	-	-	Dense (20)
L12	-	-	-	-	Dense (1)
Epochs	100	100	90	90	50
Batch size	550	800	450	500	750
Optimiser	Adam	Adam	Adam	Adam	Adam

where Conv1d refers to the convolutional layer of CNN, k is the kernel size, and u refers to the unit number in LSTM layer.

3.4.2.2 Algorithms for METU NCC Dataset

After forecasting with NASA dataset, we moved on to constructing algorithms for short-term GHI prediction with METU NCC dataset. There are two methods of GHI

prediction, as mentioned before, namely annual and seasonal forecasting. Exogenous and endogenous datasets formed from METU NCC data are used in both methods. In annual forecasting, all data points in the dataset are used for training and testing. However, in seasonal forecasting, data points are separated according to seasons, and the algorithms are built for each season.

The following sections cover algorithm construction of stand-alone CNN, LSTM and SVR algorithms as well as hybrid CNN-LSTM and CNN-LSTM-SVR algorithms for annual forecasting followed by hybrid algorithms for the seasonal forecasting. Each algorithm is trained on multiple time horizons, i.e. 10 minutes, 30 minutes and 60 minutes to estimate coming 10^{th}-minute value. In other words, with 10-minute lead time, the algorithm predicts the 20^{th} minute, while with the 30-minute horizon it predicts the coming 40^{th} minute. The parameters of each algorithm are optimised manually through many trials. Additionally, the layers and neurons are also configurated through grid search. Similar to NASA algorithms, the activation function of CNN and LSTM is selected as ReLU while the optimiser is chosen as Adam.

3.4.2.2.1 Annual Forecasting

In this forecasting method, all samples in the METU NCC dataset is divided into training, testing and validation sets. Construction of the stand-alone and hybrid algorithms are explained in detail in the following sections.

3.4.2.2.1.1 Stand-alone Algorithms

Stand-alone algorithms of CNN, LSTM and SVR are designed in order to compare with the performances of the hybrid algorithms. Due to high computational cost, manual grid search is applied to select the hyperparameters in each layer of CNN and LSTM algorithms, and SVR following heuristics.

Firstly, CNN algorithm is developed for the exogenous METU NCC dataset. It is designed to have three convolutional layers connected to three fully connected layers. Hyperparameters in each layer and parameters are listed in Table 3.7.

Next, LSTM algorithm is developed with the endogenous dataset. It contains two LSTM layers with two dense layers, which are shown in Table 3.7. The number of epochs and batch size are set as 100 and 300, respectively.

Finally, SVR algorithm is developed for the endogenous dataset. C and ε are selected as 10 and 0.05, and the kernel is set as RBF.

Table 3.7. Training parameters and input data information for the stand-alone algorithms of annual forecasting

Layer (L)	*CNN*	*LSTM*
L1	Input	Input
L2	Conv1d (f=50, k=4)	LSTM (u=7)
L3	Conv1d (f=100, k=6)	LSTM (u=10)
L4	Conv1d (f=150, k=8)	Dense (20)
L5	Flatten	Dense (1)
L6	Dense (50)	-
L7	Dense (10)	-
L8	Dense (1)	-
Epochs	100	100
Batch size	500	300
Optimiser	Adam	Adam

3.4.2.2.1.2 Hybrid Algorithms

Hybrid algorithms are the main objective of this study. In the following paragraphs, three different hybrid algorithm construction are introduced.

The first hybrid algorithm is a combination of CNN and LSTM, which is named C-LSTM for convenience. In this algorithm, CNN layers are used for feature extraction

step before LSTM layers. There are three convolutional layers connected to two LSTM layers by a time distributed layer. Two fully connected layers follow LSTM layers to give the prediction output. Convolutional layers are fed with the exogenous dataset. The batch size and epochs are chosen as 150 and 100, respectively.

The second hybrid algorithm is designed to feed CNN only exogenous data while feeding endogenous data to LSTM and feed their output to dense layers for final output. The algorithm is called CN-M for convenience, as mentioned before, and it has the same construction flow as the one shown in Figure 3.16. The CNN is composed of four convolutional layers, while LSTM has three layers. The outputs of both layers are merged and fed to two dense layers. In this algorithm, batch size and epochs are optimised as 150 and 300, respectively. Hyperparameters in each layer of both hybrid forecasting algorithms are summarised in Table 3.8.

Table 3.8. Training parameters and input data information for the hybrid algorithms, C-LSTM and CN-M, of annual forecasting

Layer (L)	*C-LSTM*	*CN-M*	
L1	Input	Input	Input
L2	Conv1d (f=28, k=3)	Conv1d (f=12, k=9)	LSTM (u=10)
L3	Conv1d (f=50, k=5)	Conv1d (f=50, k=6)	LSTM (u=30)
L4	Conv1d (f=68, k=8)	Conv1d (f=100, k=3)	LSTM (u=150)
L5	Flatten	Conv1d (f=150, k=3)	Flatten()
L6	TimeDistributed	Flatten()	-
L7	LSTM (u=5)	Merge	
L8	LSTM (u=12)	Dense (50)	
L9	Dense (30)	Dense (1)	
L10	Dense (1)	-	
Epochs	100	100	
Batch size	800	500	
Optimiser	Adam	Adam	

The final hybrid algorithm is designed similar to CN-M algorithm except CNN and LSTM are constructed as algorithms, and their output is merged and fed to an SVR model that is named CM-SVR. A flow chart simply illustrating the construction of the CM-SVR algorithm is presented in Figure 3.17. CNN algorithm is made of three convolutional layers. The output of the convolutional layer is then flattened and fed to two fully connected layers. The batch size and number of epochs of CNN are set as 500 and 100, respectively. LSTM algorithm consists of two LSTM layers, followed by two fully connected layers. The number of epochs and batch size optimised as 100 and 300. These two algorithms are trained, and the resulting predictions are combined in an array to feed to SVR algorithm. In SVR algorithms, hyperparameters C and ε are set as 14 and 0.05, respectively while the kernel is chosen as RBF. Table 3.9 lists the details on hyperparameters of each layer in all algorithms.

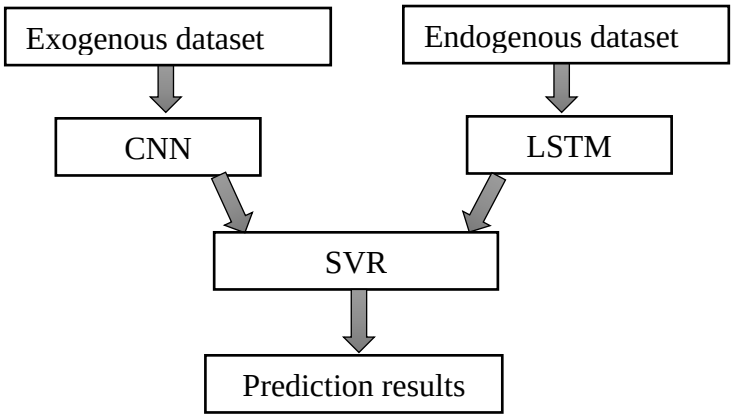

Figure 3.17. Flow chart showing the hybrid CM-SVR algorithm

Table 3.9. Training parameters and input data information for the hybrid algorithm, CM-SVR, of annual forecasting

Layer (L)	*CNN*	*LSTM*
L1	Input	Input
L2	Conv1d (f=50, k=4)	LSTM (u=7)
L3	Conv1d (f=100, k=6)	LSTM (u=10)
L4	Conv1d (f=150, k=8)	Dense (20)
L5	Flatten	Dense (1)
L6	Dense (50)	-
L7	Dense (10)	-
L8	Dense (1)	
Epochs	100	100
Batch size	500	300
Optimiser	Adam	Adam
Merge	Output 1	Output 2
SVR	C=14, kernel=RBF, ε=0.05	

3.4.2.2.2 Seasonal Forecasting

In seasonal forecasting method, the METU NCC dataset is divided into four sub-datasets. The division is based on months depending on the seasons, namely summer, fall, winter and spring. For seasonal forecasting, only CM-SVR algorithms are developed for each season, and each algorithm is built separately with the corresponding season. CNN algorithm is fed with the exogeneous dataset while the LSTM is fed with the endogenous dataset. Finally, SVR model is fed with an array prepared by combining outputs of CNN and LSTM.

The summer season dataset is made of the months June, July and August. CNN algorithm is made of three convolutional layers. The output of the convolutional layer is then flattened and fed to three dense layers. LSTM algorithm consists of two layers, followed by two fully connected layers. Finally, in SVR model, the hyperparameters C and ε are set as 2 and 0.05, respectively while the kernel is chosen as RBF.

For the fall season, only September, October and November months are included in the dataset. CNN, LSTM and SVR algorithms are designed similar to the algorithms for the summer season. CNN algorithm is made of four convolutional layers. The output of the convolutional layer is then flattened and fed to two fully connected layers. There are two layers in LSTM algorithm, which are followed by two fully connected layers. Finally, in SVR model, the hyperparameters C and ε are set as 1 and 0.05, respectively while the kernel is chosen as RBF.

In the winter season, December, January and February months are selected in the dataset. There are three convolutional and two dense layers in CNN while LSTM algorithm has three LSTM layers and two dense layers. C and ε in SVR are set as 10 and 0.03, respectively.

Finally, the spring dataset is composed of March, April and May. CNN algorithm is composed of three convolutional layers and three dense layers. LSTM algorithm as well has three LSTM layers and three dense layers. C and ε in SVR are set as 4 and

0.05, respectively. A detailed list of all hyperparameters in each layer in all algorithms is presented in Table 3.10.

Table 3.10. Layers in hybrid CM-SVR algorithm for seasonal forecasting

Layer (L)	*Summer*	*Fall*	*Winter*	*Spring*
L1	Input	Input	Input	Input
	Conv1d	Conv1d	Conv1d	Conv1d
L2	(f=50, k=4)	(f=20, k=4)	(f=50, k=4)	(f=50, k=4)
	Conv1d	Conv1d		Conv1d
L3	(f=100, k=6)	(f=200, k=6)	Dropout (0.2)	(f=100, k=6)
	Conv1d	Conv1d	Conv1d	Conv1d
L4	(f=150, k=8)	(f=75, k=6)	(f=70, k=6)	(f=150, k=6)
		Conv1d	Conv1d	
L5	Flatten	(f=150, k=8)	(f=100, k=8)	Flatten
L6	Dense (50)	Flatten	Flatten	Dense (50)
L7	Dense (10)	Dense (10)	Dense (15)	Dense (10)
L8	Dense (1)	Dense (1)	Dense (1)	Dense (1)
Epochs	100	100	100	100
Batch size	500	500	400	500
Optimiser	Adam	Adam	Adam	Adam
L9	Input	Input	Input	Input
L10	LSTM (u=7)	LSTM (u=10)	LSTM (u=3)	LSTM (u=20)
L11	LSTM (u=10)	LSTM (u=15)	LSTM (u=7)	LSTM (u=100)
L12	Dense (20)	Dense (20)	LSTM (u=11)	LSTM (u=15)
L13	Dense (1)	Dense (1)	Dense (20)	Dense (30)
L14	-	-	Dense (1)	Dense (20)
L15	-	-	-	Dense (1)
Epochs	100	100	150	100
Batch size	300	300	600	500
Optimiser	Adam	Adam	Adam	Adam

SVR	C=2.0, kernel=RBF, ε=0.05	C=1.0, kernel=RBF, ε=0.03	C=10.0, kernel=RBF, ε=0.03	C=4.0, kernel=RBF, ε=0.05

Table 3.11 lists the search space of the hyperparameters used in grid search for all algorithms constructed in this study.

Table 3.11. Training hyperparameters in each layer for the constructed forecasting algorithms

Algorithm Hyperparameters	*Search Space*
Conv1D - filters	[10, 12, 28, 50, 64, 70, 75, 100, 128, 150, 200, 240, 300]
Conv1D - kernel size	[3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20]
LSTM - units	[3, 4, 5, 7, 10, 15, 20, 25, 30, 40, 50, 100, 150, 200, 240]
Dense	[10, 15, 20, 25, 30, 40, 50, 100, 150]
Batch size	[30, 50, 100, 200, 300, 400, 450, 500, 550, 600, 750, 800]
Number of Epochs	[10, 20, 50, 70, 90, 100, 150]
Activation function	[ReLU, sigmoid, softmax, softplus]
C	[1, 2, 3, 4, 5, 10, 11, 12,13, 14, 15, 16, 20]
ε	[0.01, 0.02, 0.03, 0.04, 0.05, 0.1, 0.15,0.35, 0.2]
Kernel	[RBF, poly, linear, sigmoid]

CHAPTER 4

RESULTS AND DISCUSSION

This chapter presents GHI forecasting results and performance assessment for both NASA and METU NCC datasets with various learning algorithms at different forecasting horizons. The first section presents the GHI forecasting analysis for the NASA dataset, while the second section covers the forecasting results and discussion for the METU NCC dataset.

4.1 GHI Prediction Analysis of NASA Dataset

The NASA dataset for Kalkanlı is obtained from [70] and preprocessed prior to the training of the forecasting algorithms, as mentioned in section 3.4.1. Two different datasets, i.e. exogenous and endogenous datasets, are created from the NASA dataset. The exogenous dataset consists of meteorological variables and GHI data, which are listed in Table 3.1, whereas the endogenous dataset is made of only GHI data.

For GHI prediction with NASA dataset, CNN and LSTM algorithms are employed. Both CNN and LSTM algorithms are constructed for each dataset, adding up to four different forecasting models. In addition to these models, a hybrid algorithm of CNN and LSTM, i.e. CN-M, is constructed feeding exogeneous dataset to CNN and endogenous dataset to LSTM algorithms and combining their output in a fully connected layer for a final result. After the construction of the algorithms, i.e. CNN, LSTM and CN-M, in section 3.4.2.1 for the NASA dataset, the training is initialised. In training, the time-step is chosen as seven days lead in order to deliver useful information on radiation availability for a future solar-powered system design. Hyperparameter of the algorithms is optimised by manual grid search due to the high

computational cost of an automated grid search. After the algorithm training, the resulting models are evaluated using MAE, RMSE, n-RMSE and R^2 in the testing stage. A flow diagram summarising the process of GHI forecasting from dataset to final result for NASA dataset is illustrated in Figure 4.1.

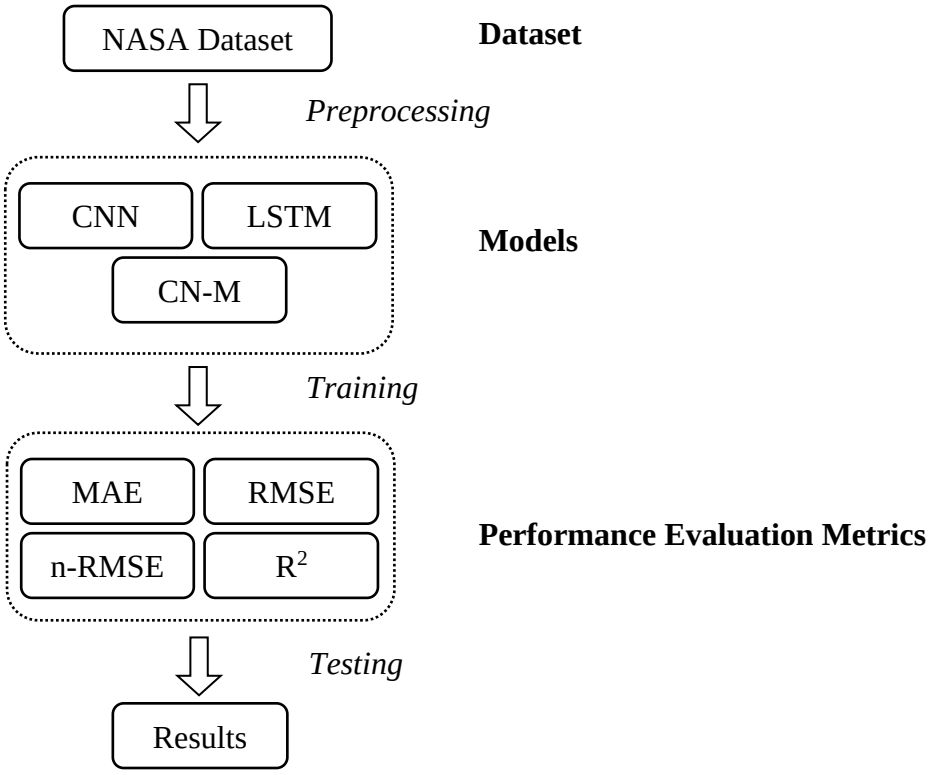

Figure 4.1. Flow chart of the forecasting procedure for the NASA dataset

The evaluation results of each forecasting model are tabulated in Table 4.1, where the best results in each evaluation metric are shown in bold. The reported evaluation results are the mean of five separate runs for each model where data is randomly partitioned into training, validation and test set. The hybrid model achieves slightly better results than the remaining four models. It also has an MAE of 19.9 and 21.7 W m^{-2} for training and validation sets, respectively. However, according to Table 2.1, the n-RMSE results of all models are in the good model precision range with slight differences among each other. Also, the prediction outcome of each model has small deviations from the real data since MAE and RMSE values are in close range. Additionally, R^2 results illustrate that all models fit the data very good.

Table 4.1. Summary of GHI prediction model performances with corresponding dataset type, best results in each evaluation metric is shown in bold

Models	*MAE* *(W m^{-2})*	*RMSE* *(W m^{-2})*	*n-RMSE*	R^2
CNN – exogeneous	21.5	34.0	0.16	0.86
CNN – endogeneous	19.7	31.2	0.15	0.87
LSTM – exogeneous	21.0	33.3	0.16	0.85
LSTM – endogeneous	21.2	31.3	0.15	0.87
CN-M	**19.3**	**30.4**	**0.14**	**0.88**

In order to assess the performance of the algorithms better and understand the error distribution, the APE of each model with the testing dataset is drawn on a histogram. The histograms of each model are illustrated in Figure 4.2, where the y-axis shows the APE frequency in percentages, while the x-axis represents the APE ranges in 10 W m^{-2} interval. The hybrid model and CNN model constructed with the endogenous dataset result in around 50% of APE as less than 10 W m^{-2}. In other words, these models will predict the GHI with a maximum deviation error of 10 W m^{-2} half of the prediction time. While both models result in relatively similar results, model preference could be made based on computation time which is 2 seconds in the hybrid model while it is 6 seconds in the CNN model.

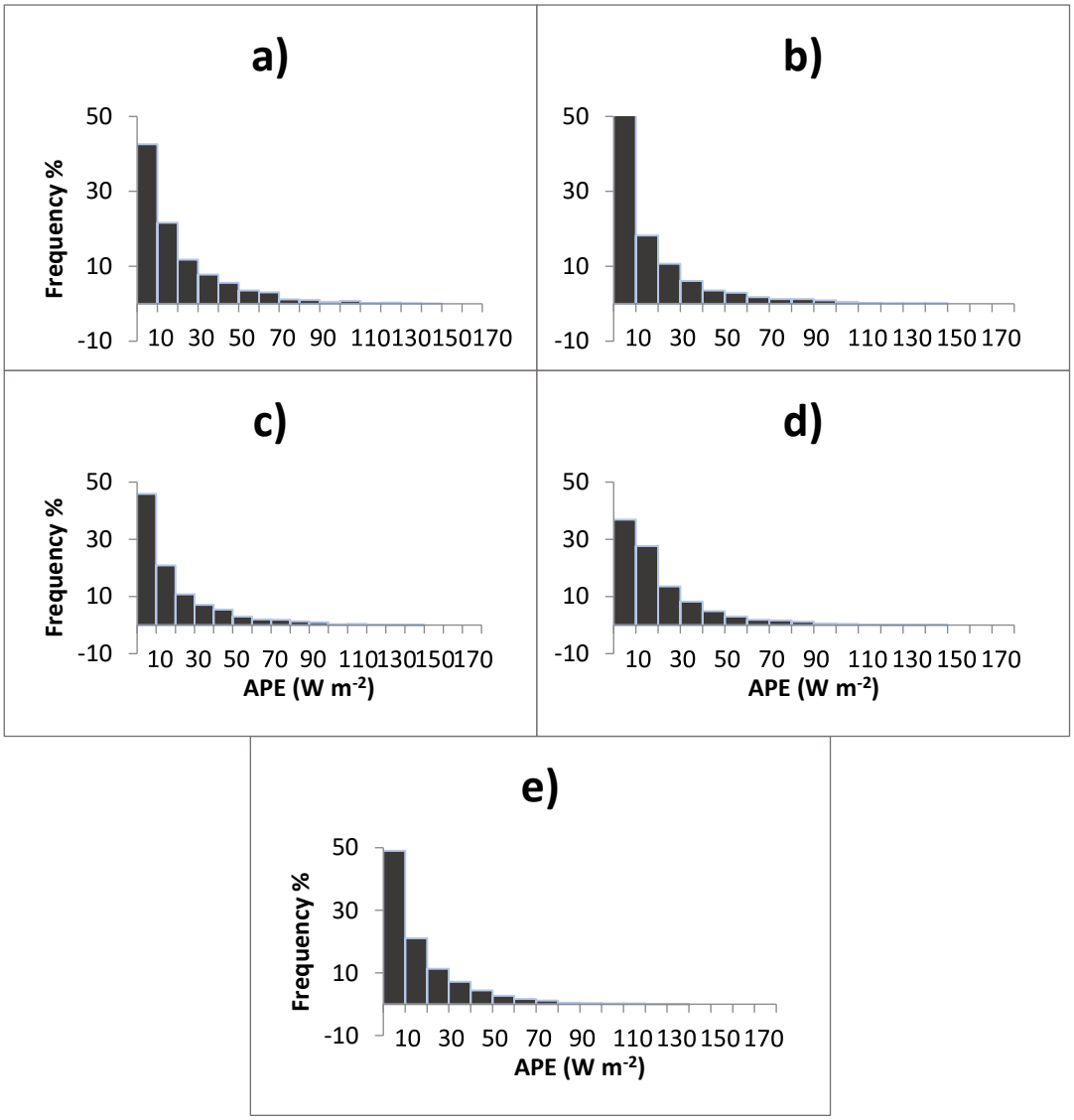

Figure 4.2. APE frequency histograms generated by the results of testing set, a) CNN – exogenous, b) CNN – endogenous, c) LSTM – exogenous, d) LSTM – endogenous, e) CN-M

4.2 GHI Forecasting Analysis of METU NCC Dataset

GHI and meteorological data with a 10-minute time interval for METU NCC, Kalkanlı, are used in this part of the study. The dataset is preprocessed, and two datasets are created from it, similar to the NASA dataset. The exogenous dataset

consists of meteorological variables and GHI data, listed in Table 3.3, whereas the endogenous dataset is made of only GHI data.

CNN, LSTM and SVR algorithms are employed for the GHI forecasting with METU NCC dataset. There are two different forecasting methods followed in this part, i.e. annual and seasonal forecasting, Hybrid algorithms of CNN, LSTM, and SVR, namely C-LSTM, CN-M and CM-SVR, are created and compared with the performance of stand-alone algorithms, i.e. CNN, LSTM and SVR, in the annual forecasting part. In the seasonal forecasting part, the performance of models for each season is evaluated through the hybrid algorithm CM-SVR. After the construction of the algorithms for the METU NCC dataset, the training is initialised. The constructed algorithms are trained on several time-horizons, i.e. time interval of 10 minutes, 30 minutes and 60 minutes. Hyperparameter of the algorithms is optimised by manual grid search due to high computational cost. After the algorithm training, the resulting models are evaluated using MAE, RMSE, n-RMSE and R^2 in the testing stage. A flow diagram summarising the process of GHI forecasting from dataset to final result for METU NCC dataset is illustrated in Figure 4.3.

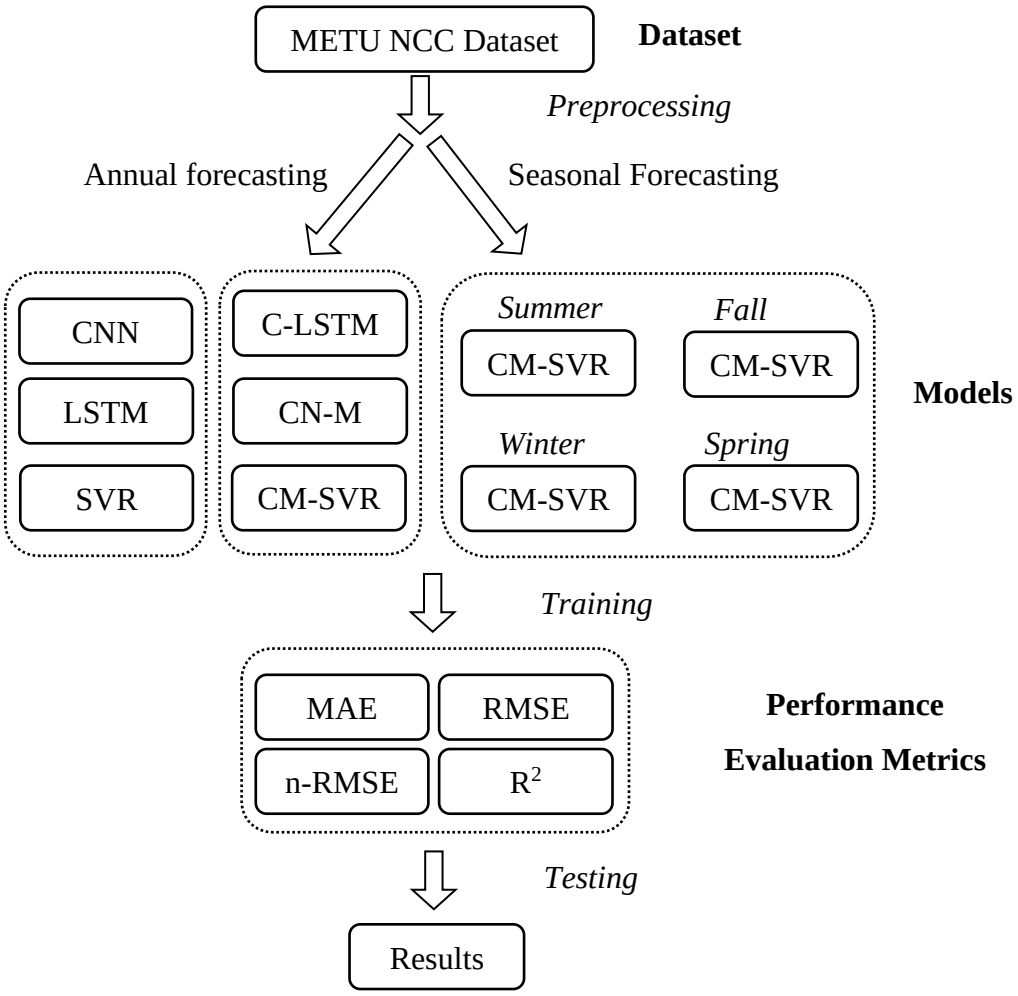

Figure 4.3. Flow chart of the forecasting procedure for the METU NCC dataset

Testing results of stand-alone and hybrid prediction algorithms for the METU NCC dataset are listed in Table 4.2 with corresponding evaluation metric score in three forecasting horizons, where the best results are shown in bold. Similar to NASA dataset, the reported evaluation results of METU NCC dataset are the mean of five separate runs for each model where data is randomly partitioned into training, validation and test set. In terms of evaluation metrics' results, the models perform similarly. Similar to the NASA dataset models, the models for the METU NCC dataset also have a good model precision according to Table 2.1, with n-RMSE changing between 0.14 to 0.17. CN-M achieves better results with respect to MAE among the hybrid algorithms. CN-M also has an MAE of 28.3 and 37.8 W m^{-2} for

training and validation sets, respectively. Additionally, R^2 results illustrate that all models fit the data quite well. On the other hand, all evaluation results suggest that CM-SVR performs poorly in 10-minutes forecasting lead.

Table 4.2. Summary of GHI prediction model performances in different time-leads for annual forecasting, best results are shown in bold

Evaluation Metrics	*MAE (W m^{-2})*			*RMSE (W m^{-2})*			*n-RMSE*			*R^2*		
Lead time (minutes)	*10*	*30*	*60*	*10*	*30*	*60*	*10*	*30*	*60*	*10*	*30*	*60*
CNN	30.9	30.1	30.0	53.9	53.2	**52.5**	0.15	0.15	**0.14**	0.95	0.95	0.95
LSTM	37.0	33.8	32.0	56.4	57.4	56.2	0.15	0.16	0.15	0.95	0.95	0.95
SVR	37.2	-	-	56.5	-	-	0.16	-	-	0.95	-	-
C-LSTM	31.0	30.3	38.3	53.9	53.4	62.9	0.15	**0.14**	0.17	**0.96**	**0.96**	0.94
CN-M	31.0	**29.6**	31.0	53.9	53.0	53.6	**0.14**	**0.14**	0.15	**0.96**	0.95	**0.96**
CM-SVR	49.3	30.1	30.1	89.4	53.1	53.9	0.24	**0.14**	**0.14**	0.88	**0.96**	**0.96**

Figure 4.4 and Figure 4.5 demonstrate the frequency of APE on a histogram for stand-alone and hybrid models, respectively. Although the models performed similarly in terms of evaluation metrics, stand-alone CNN and SVR result in poor prediction outputs. On the other hand, stand-alone LSTM has a high MAE result in the 30-minute horizon, indicating a high average error magnitude. However, LSTM histogram shows that more than 40% of the APE is 10 W m^{-2}.

Among the hybrid models, CN-M and CM-SVR result in approximately 45% of APE less than 10 W m^{-2}, performing better than the rest of the models when evaluation metrics' results are also considered. Similar to NASA dataset, model preference could be made based on computation time which is approximately 20 seconds in the CM-SVR while 18 seconds in the CN-M model.

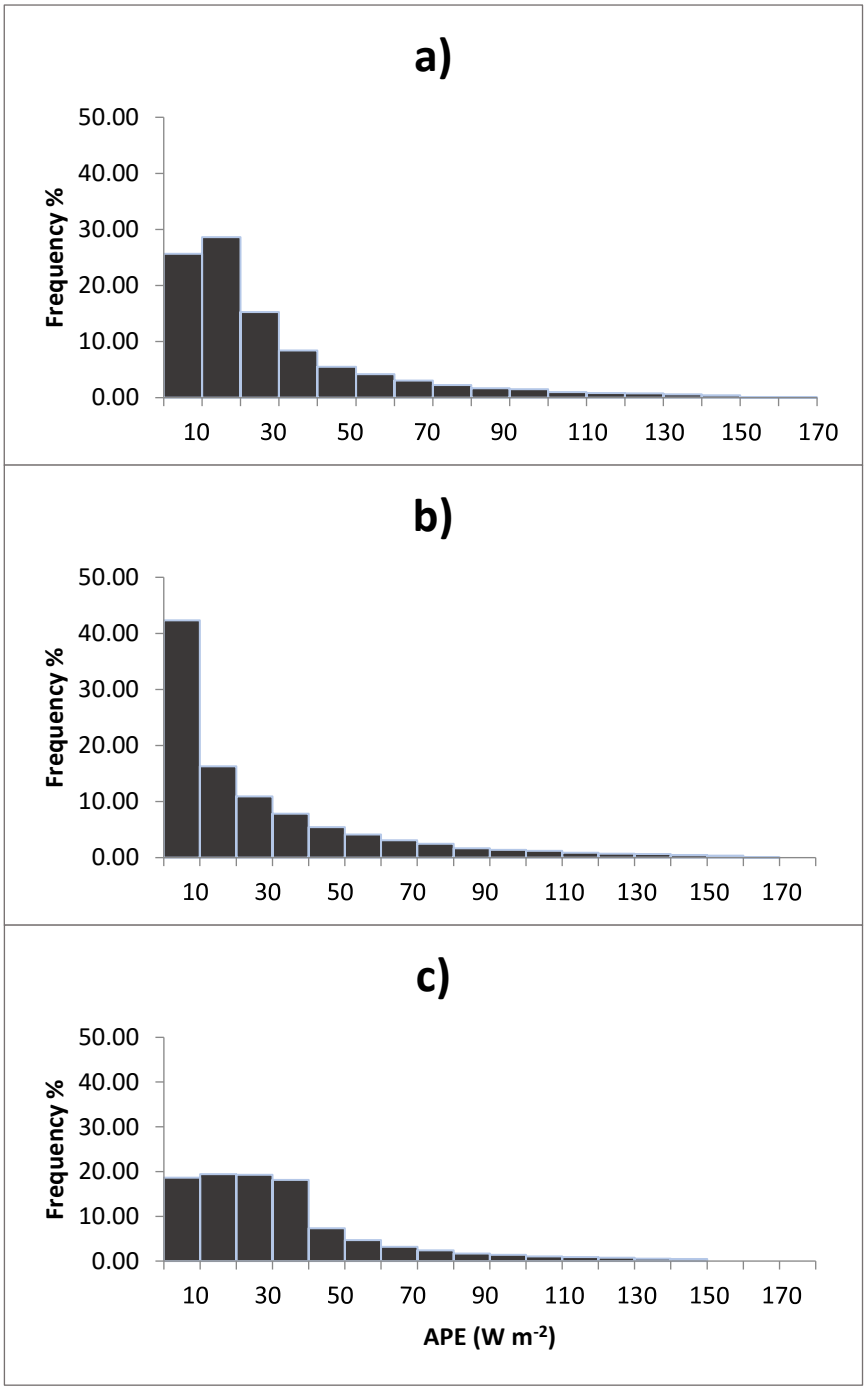

Figure 4.4. APE frequency histograms generated by the results of the testing set for the stand-alone models over 30-minutes forecasting horizon, a) CNN – endogeneous, b) LSTM – exogeneous, c) SVR – exogeneous

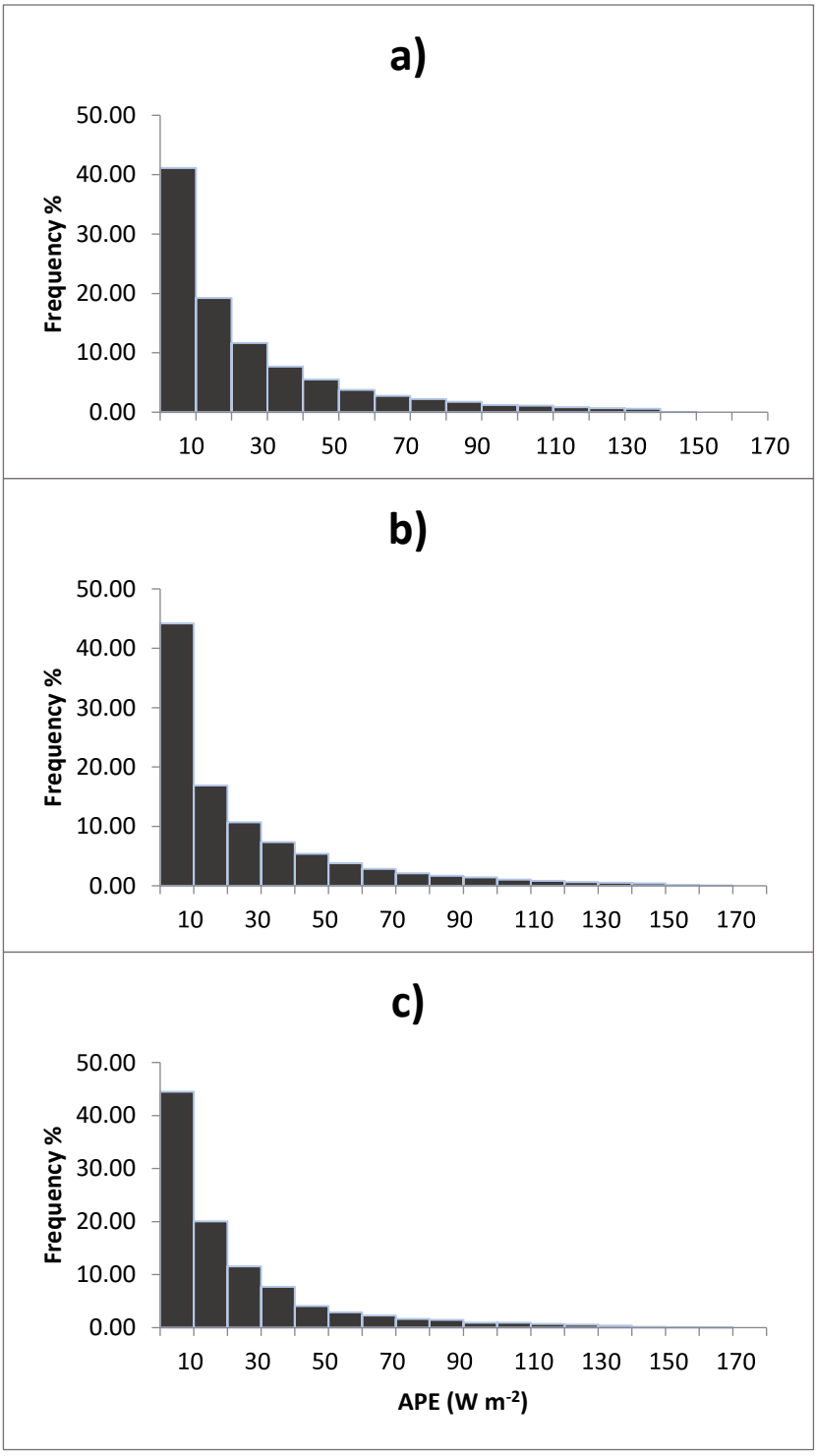

Figure 4.5. APE frequency histograms generated by the results of the testing set for the hybrid models, a) C-LSTM, b) CN-M, c) CM-SVR

Evaluation metric results of the testing set on CM-SVR for the seasonal prediction with METU NCC dataset are presented in Table 4.3 with corresponding evaluation metric score in three forecasting horizons. The CM-SVR model for the summer season produces the best results among all seasons as expected since clear sky conditions occur most prominently in the summer months, as illustrated in Figure 4.6. On that note, the n-RMSE of the summer model has excellent precision, according to Table 2.1. On the other hand, the model of the winter season results in higher error outputs as a result of high fluctuations in GHI, i.e. overcast sky condition, as shown in Figure 4.7. Hence, the model has a fair precision with an n-RMSE of 0.24. Additionally, the forecasting model for the spring season behaves similar to the model of the summer season, while the fall model performs closer to the winter model. Similar patterns are observed in R^2 results that the summer and spring models fit almost perfectly on the actual data.

Table 4.3. Summary of GHI prediction model performances in different time-leads for seasonal forecasting

Evaluation Metrics	*MAE (W m^{-2})*			*RMSE (W m^{-2})*			*n-RMSE*			*R^2*		
Lead time (minutes)	*10*	*30*	*60*	*10*	*30*	*60*	*10*	*30*	*60*	*10*	*30*	*60*
Summer	24.5	25.0	25.3	40.8	41.5	41.4	0.08	0.08	0.08	0.97	0.97	0.97
Fall	34.3	34.6	34.6	53.5	53.7	53.7	0.21	0.21	0.21	0.88	0.88	0.88
Winter	35.0	35.0	35.4	55.3	56.0	55.2	0.24	0.24	0.24	0.88	0.89	0.89
Spring	31.5	30.3	30.0	61.0	61.4	60.7	0.13	0.12	0.12	0.95	0.95	0.95

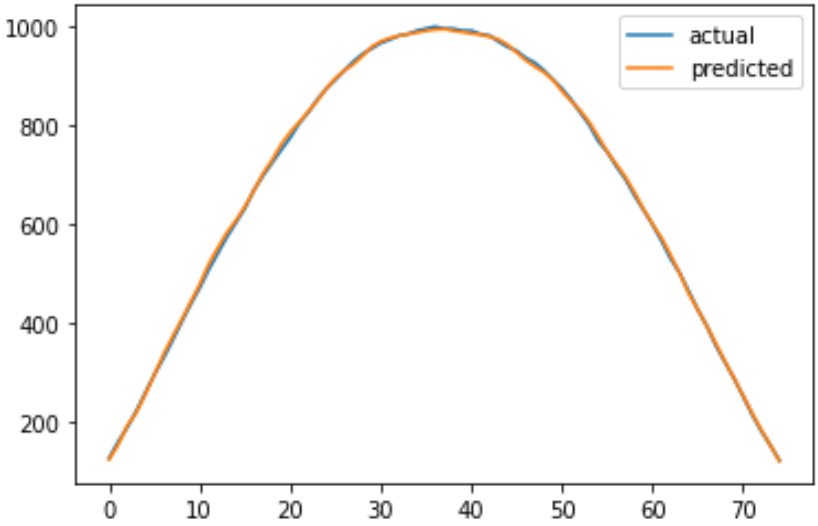

Figure 4.6. CM-SVR model prediction fitted over the actual data for the summer season for a clear-sky condition

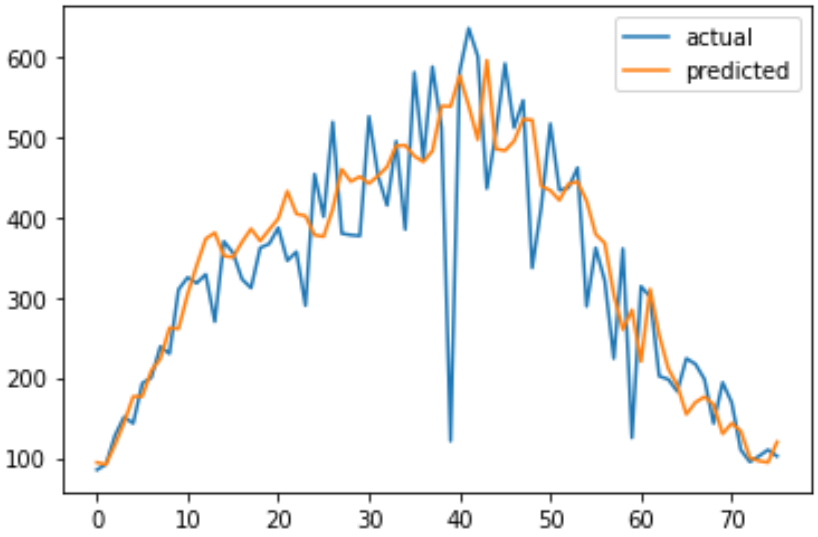

Figure 4.7. CM-SVR model prediction fitted over the actual data for the winter season for a scattered clouds sky condition

Figure 4.8 demonstrate the frequency of APE on histograms for each season separately. Evaluation metric results comply with APE results for the winter season. Also, the models for the summer and spring predict 40% of the results with a deviation of less than 10 W m^{-2}, performing better than the rest of the seasons.

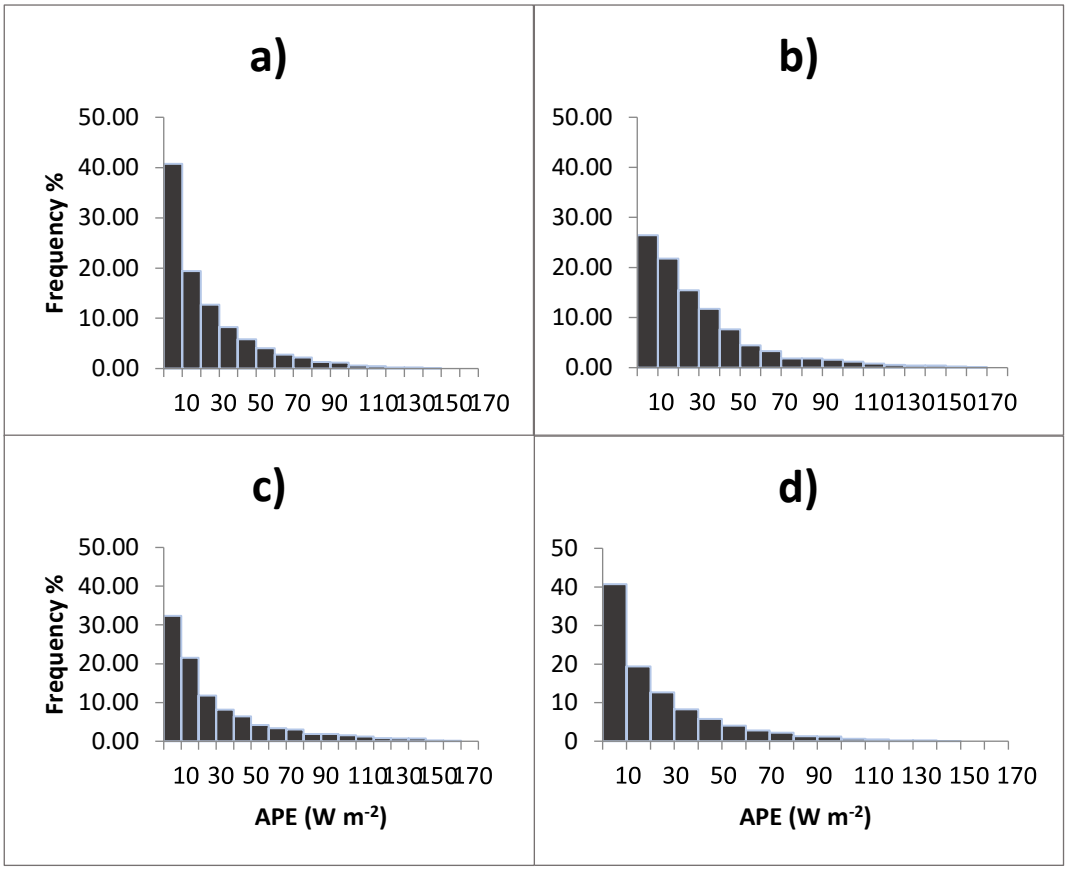

Figure 4.8. APE frequency histograms generated by the results of the testing set for the seasonal forecasting with CM-SVR models, a) summer, b) fall, c) winter, d) spring

The evaluation results of seasonal forecasting models are averaged and presented in Table 4.4 in order to compare the annual and seasonal forecasting models. On average, the seasonal forecasting models performs similar to the annual forecasting algorithm. In addition to the performances, all models in seasonal forecasting compute results in 1 second, making them the fastest forecasting models compared to stand-alone and hybrid models of annual forecasting. Hence, seasonal forecasting could be preferable to annual forecasting for the Meditteranean region, where the seasons have distinct patterns in GHI fluctuations.

Table 4.4. Averaged evaluation metric results of all seasons in different forecasting horizons

Evaluation Metrics	*MAE (W m^{-2})*			*RMSE (W m^{-2})*			*n-RMSE*			*R^2*		
Lead time (minutes)	*10*	*30*	*60*	*10*	*30*	*60*	*10*	*30*	*60*	*10*	*30*	*60*
Averaged	31.3	31.2	31.3	52.8	53.2	52.8	0.16	0.16	0.16	0.92	0.92	0.92

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

In this study, it is aimed to deliver models to predict short-term and long-term GHI for Kalkanlı region for the sustainable and continuous growth of PV panels in Northern Cyprus. Because knowledge of solar radiation, i.e. GHI, information is essential, especially for the case of Northern Cyprus where there are limited energy production sources and a high rate of PV installation over rooftops. This study also contributes to ongoing research on developing prediction algorithms to accurately estimate solar radiation and energy production by testing different hybrid forecasting algorithms.

Two different datasets from two different sources are used to construct forecasting algorithms. The first dataset is obtained from the website of NASA. It contains daily values of meteorological variables and GHI from 1983 to 2019. The dataset is initially preprocessed. Following the preprocessing step, two separate datasets are created from the NASA data, namely exogenous dataset, which contains meteorological variables and GHI, and endogenous dataset made of past data of GHI. The second dataset is obtained from METU NCC; hence it is called METU NCC dataset. It contains meteorological variables and GHI values over the 10-minute interval from 2013 to 2017. The METU NCC dataset is also preprocessed, similar to NASA dataset. Exogenous and endogenous datasets are also created from METU NCC dataset.

Five prediction models are developed for the NASA dataset over seven days forecasting horizon. Four stand-alone CNN and LSTM models for exogenous and endogenous datasets and one CNN-LSTM hybrid model, i.e. CN-M, are constructed. The resulting errors indicate that although the differences are relatively small, the hybrid model is preferable as it has a lower computation time. One the other hand,

the CNN model with the endogenous dataset could be desirable in regions where meteorological variables are absent.

For the METU NCC dataset, two different forecasting methods are followed, i.e. annual forecasting and seasonal forecasting. In annual forecasting, all data is used for training and testing. Overall, six different models are designed for annual forecasting; three stand-alone models, i.e. CNN, LSTM and SVR, and three hybrid models, i.e. C-LSTM, CN-M and CM-SVR. Among the models created with METU NCC dataset, CN-M performs relatively better than the remaining models with a lower computational cost.

In seasonal forecasting, four sub-datasets are created based on the seasons. For each season, a CM-SVR model is designed. Evaluation results suggest that the summer model achieves the lowest error, while the winter model results in the highest error. For the Mediterranean region, where seasons have distinct sky condition patterns, it could be preferable to have different separate models for each season. When compared to the performance of annual forecasting models, seasonal models perform similarly on average. However, low computation cost makes the seasonal models desirable.

The importance of this study is that it provides information on future GHI, which is the main parameter on PV power generation. The information on PV power output enables the power generation utility to maximise the use of PV panels and to decrease the use of conventional energy sources that contribute to global warming. In other words, a better prediction of GHI allows better planning of power generation from conventional sources. Hence, the energy production units with better efficiencies could be utilised. The PV power output knowledge is also an important factor in smart grids.

In this study, we successfully constructed effective stand-alone and hybrid models for GHI forecasting. In future studies, the dataset sensitivity of the forecasting models in different parts of the world could be investigated. The performances of other related machine learning algorithms combined with deep learning algorithms

may also be examined. Additionally, the effects of longer forecasting horizons as well as different input variables could be analysed. Finally, the forecasting of PV output may be studied.

REFERENCES

[1] M. Ben Amar, “Energy consumption and economic growth: the case of African countries,” Autumn, 2012.

[2] IEA, “World Energy Outlook 2019,” *IEA, Paris*, 2019. [Online]. Available: https://www.iea.org/reports/world-energy-outlook-2019. [Accessed: 03-Jul-2020].

[3] IEA, “Global Energy Review 2020,” *IEA, Paris*, 2020. [Online]. Available: https://www.iea.org/reports/global-energy-review-2020/global-energy-and-co2-emissions-in-2020#abstract. [Accessed: 03-Jul-2020].

[4] S. Ghimire, R. C. Deo, N. Raj, and J. Mi, “Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms,” *Appl. Energy*, vol. 253, p. 113541, Nov. 2019.

[5] M. Demirtas, M. Yesilbudak, S. Sagiroglu, and I. Colak, “Prediction of solar radiation using meteorological data,” *2012 Int. Conf. Renew. Energy Res. Appl. ICRERA 2012*, pp. 1–4, 2012.

[6] C. Voyant *et al.*, “Machine learning methods for solar radiation forecasting: A review,” *Renew. Energy*, vol. 105, pp. 569–582, 2017.

[7] S. Salcedo-Sanz, C. Casanova-Mateo, A. Pastor-Sánchez, and M. Sánchez-Girón, “Daily global solar radiation prediction based on a hybrid Coral Reefs Optimization - Extreme Learning Machine approach,” *Sol. Energy*, vol. 105, pp. 91–98, 2014.

[8] A. Alzahrani, P. Shamsi, C. Dagli, and M. Ferdowsi, “Solar Irradiance Forecasting Using Deep Neural Networks,” in *Procedia Computer Science*, 2017, vol. 114, pp. 304–313.

[9] B. Elliston and I. MacGill, “The potential role of forecasting for integrating solar generation into the Australian National Electricity Market,” *Sol. 2010, Aust. Sol. Energy Soc.*, no. December 2010, pp. 1–11, 2010.

[10] S. Ferrari *et al.*, “Illuminance prediction through Extreme Learning Machines,” *2012 IEEE Work. Environ. Energy, Struct. Monit. Syst. EESMS 2012 - Proc.*, pp. 97–103, 2012.

[11] L. Mazorra-Aguiar and F. Díaz, “Solar radiation forecasting with statistical models,” in *Green Energy and Technology*, no. 9783319768755, Springer Verlag, 2018, pp. 171–200.

[12] A. Fouilloy *et al.*, “Solar irradiation prediction with machine learning: Forecasting models selection method depending on weather variability,” *Energy*, vol. 165, pp. 620–629, Dec. 2018.

[13] X. Qing and Y. Niu, “Hourly day-ahead solar irradiance prediction using

weather forecasts by LSTM," *Energy*, vol. 148, pp. 461–468, 2018.

[14] G. Guariso, G. Nunnari, and M. Sangiorgio, "Multi-Step Solar Irradiance Forecasting and Domain Adaptation of Deep Neural Networks," *Energies*, vol. 13, no. 15, p. 3987, Aug. 2020.

[15] C. Voyant, M. Muselli, C. Paoli, and M. L. Nivet, "Hybrid methodology for hourly global radiation forecasting in Mediterranean area," *Renew. Energy*, vol. 53, pp. 1–11, May 2013.

[16] R. Nageem and R. Jayabarathi, "Predicting the Power Output of a Grid-Connected Solar Panel Using Multi-Input Support Vector Regression," in *Procedia Computer Science*, 2017, vol. 115, pp. 723–730.

[17] C. Wan, J. Zhao, Y. Song, Z. Xu, J. Lin, and Z. Hu, "Photovoltaic and solar power forecasting for smart grid energy management," *CSEE J. Power Energy Syst.*, vol. 1, no. 4, pp. 38–46, Jan. 2016.

[18] J. Fan *et al.*, "Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China," *Energy Convers. Manag.*, vol. 164, no. February, pp. 102–111, 2018.

[19] M. Lazzaroni, S. Ferrari, V. Piuri, A. Salman, L. Cristaldi, and M. Faifer, "Models for solar radiation prediction based on different measurement sites," *Meas. J. Int. Meas. Confed.*, vol. 63, pp. 346–363, 2015.

[20] R. Marquez and C. F. M. Coimbra, "Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the NWS database," *Sol. Energy*, vol. 85, no. 5, pp. 746–756, 2011.

[21] A. K. Yadav and S. S. Chandel, "Solar radiation prediction using Artificial Neural Network techniques: A review," *Renewable and Sustainable Energy Reviews*, vol. 33. Elsevier Ltd, pp. 772–781, 01-May-2014.

[22] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 841–851, Jan. 2019.

[23] G. Reikard, "Predicting solar radiation at high resolutions: A comparison of time series forecasts," *Solar Energy*, vol. 83, no. 3. pp. 342–349, 2009.

[24] H. T. C. Pedro and C. F. M. Coimbra, "Nearest-neighbor methodology for prediction of intra-hour global horizontal and direct normal irradiances," *Renew. Energy*, vol. 80, pp. 770–782, 2015.

[25] A. Sözen, E. Arcaklioglu, and M. Özalp, "Estimation of solar potential in Turkey by artificial neural networks using meteorological and geographical data," *Energy Convers. Manag.*, vol. 45, no. 18–19, pp. 3033–3052, Nov. 2004.

[26] P. Krömer, P. Musilek, E. Pelikan, P. Krc, P. Jurus, and K. Eben, “Support Vector Regression of multiple predictive models of downward short-wave radiation,” *Proc. Int. Jt. Conf. Neural Networks*, pp. 651–657, 2014.

[27] F. J. L. Lima, F. R. Martins, E. B. Pereira, E. Lorenz, and D. Heinemann, “Forecast for surface solar irradiance at the Brazilian Northeastern region using NWP model and artificial neural networks,” *Renew. Energy*, vol. 87, pp. 807–818, Mar. 2016.

[28] G. P. Podestá, L. Núñez, C. A. Villanueva, and M. A. Skansi, “Estimating daily solar radiation in the Argentine Pampas,” *Agric. For. Meteorol.*, vol. 123, no. 1–2, pp. 41–53, 2004.

[29] N. Vakitbilir, A. Hilal, and C. Direkoğlu, “Prediction of Daily Solar Irradiation Using CNN and LSTM Networks,” Springer, Cham, 2021, pp. 230–238.

[30] “PV Performance Modeling Collaborative | Global Horizontal Irradiance.” [Online]. Available: https://pvpmc.sandia.gov/modeling-steps/1-weather-design-inputs/irradiance-and-insolation-2/global-horizontal-irradiance/. [Accessed: 28-Nov-2020].

[31] K. Mallon, F. Assadian, and B. Fu, “Analysis of On-Board Photovoltaics for a Battery Electric Bus and Their Impact on Battery Lifespan,” *Energies*, vol. 10, no. 7, p. 943, Jul. 2017.

[32] F. V. Gutierrez-Corea, M. A. Manso-Callejo, M. P. Moreno-Regidor, and M. T. Manrique-Sancho, “Forecasting short-term solar irradiance based on artificial neural networks and data from neighboring meteorological stations,” *Sol. Energy*, vol. 134, pp. 119–131, Sep. 2016.

[33] “Global Solar Atlas.” [Online]. Available: https://globalsolaratlas.info/map. [Accessed: 01-Oct-2020].

[34] L. Feng, W. Qin, L. Wang, A. Lin, and M. Zhang, “Comparison of Artificial Intelligence and Physical Models for Forecasting Photosynthetically-Active Radiation,” *Remote Sens.*, vol. 10, no. 11, p. 1855, Nov. 2018.

[35] F. Besharat, A. A. Dehghan, and A. R. Faghih, “Empirical models for estimating global solar radiation: A review and case study,” *Renewable and Sustainable Energy Reviews*, vol. 21. Pergamon, pp. 798–821, 01-May-2013.

[36] S. Ferrari, M. Lazzaroni, V. Piuri, L. Cristaldi, and M. Faifer, “Statistical models approach for solar radiation prediction,” in *Conference Record - IEEE Instrumentation and Measurement Technology Conference*, 2013, pp. 1734–1739.

[37] G. Zhang, B. Eddy Patuwo, and M. Y. Hu, “Forecasting with artificial neural networks: The state of the art,” *Int. J. Forecast.*, vol. 14, no. 1, pp. 35–62, Mar. 1998.

[38] A. T. C. Goh, “Back-propagation neural networks for modeling complex systems,” *Artif. Intell. Eng.*, vol. 9, no. 3, pp. 143–151, Jan. 1995.

[39] A. McGovern, D. J. Gagne, J. Basara, T. M. Hamill, and D. Margolin, “Solar energy prediction : An international contest to initiate interdisciplinary research on compelling meteorological problems,” *Bull. Am. Meteorol. Soc.*, vol. 96, no. 8, pp. 1388–1393, 2015.

[40] I. Bilionis, E. M. Constantinescu, and M. Anitescu, “Data-driven model for solar irradiation based on satellite observations,” *Sol. Energy*, vol. 110, pp. 22–38, 2014.

[41] F. S. Wong, “Time series forecasting using backpropagation neural networks,” *Neurocomputing*, vol. 2, no. 4, pp. 147–159, Jul. 1991.

[42] D. Svozil, V. Kvasnička, and J. Pospíchal, “Introduction to multi-layer feed-forward neural networks,” in *Chemometrics and Intelligent Laboratory Systems*, 1997, vol. 39, no. 1, pp. 43–62.

[43] A. Burkov, “The Hundred-Page Machine Learning Book,” 2019.

[44] H. Shi, M. Xu, and R. Li, “Deep Learning for Household Load Forecasting-A Novel Pooling Deep RNN,” *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 5271–5280, Sep. 2018.

[45] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553. Nature Publishing Group, pp. 436–444, 27-May-2015.

[46] S. K. Aggarwal and L. M. Saini, “Solar energy prediction using linear and non-linear regularization models: A study on AMS (American Meteorological Society) 2013-14 Solar Energy Prediction Contest,” *Energy*, vol. 78. pp. 247–256, 2014.

[47] K. Mohammadi, S. Shamshirband, M. H. Anisi, K. Amjad Alam, and D. Petković, “Support vector regression based prediction of global solar radiation on a horizontal surface,” *Energy Convers. Manag.*, vol. 91, pp. 433–441, Feb. 2015.

[48] A. Gensler, J. Henze, B. Sick, and N. Raabe, “Deep Learning for solar power forecasting - An approach using AutoEncoder and LSTM Neural Networks,” in *2016 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2016 - Conference Proceedings*, 2017, pp. 2858–2865.

[49] R. H. Inman, H. T. C. Pedro, and C. F. M. Coimbra, “Solar forecasting methods for renewable energy integration,” *Progress in Energy and Combustion Science*, vol. 39, no. 6. pp. 535–576, Dec-2013.

[50] P. Goodwin and R. Lawton, “On the asymmetry of the symmetric MAPE,” *Int. J. Forecast.*, vol. 15, no. 4, pp. 405–408, Oct. 1999.

[51] S. Makridakis, “Accuracy measures: theoretical and practical concerns,” *Int.*

J. Forecast., vol. 9, no. 4, pp. 527–529, Dec. 1993.

[52] C. Tofallis, "A better measure of relative prediction accuracy for model selection and model estimation," *J. Oper. Res. Soc.*, vol. 66, no. 8, pp. 1352–1362, Aug. 2015.

[53] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. One Lake Street Upper Saddle River: Prentice Hall Press, 2009.

[54] J. Cao and X. Lin, "Study of hourly and daily solar irradiation forecast using diagonal recurrent wavelet neural networks," *Energy Convers. Manag.*, vol. 49, no. 6, pp. 1396–1406, 2008.

[55] A. Khosravi, R. N. N. Koury, L. Machado, and J. J. G. Pabon, "Prediction of hourly solar radiation in Abu Musa Island using machine learning algorithms," *J. Clean. Prod.*, vol. 176, pp. 63–75, Mar. 2018.

[56] H. Zang, L. Liu, L. Sun, L. Cheng, Z. Wei, and G. Sun, "Short-term global horizontal irradiance forecasting based on a hybrid CNN-LSTM model with spatiotemporal correlations," *Renew. Energy*, vol. 160, pp. 26–41, Nov. 2020.

[57] S. Belaid and A. Mellit, "Prediction of daily and mean monthly global solar radiation using support vector machine in an arid climate," *Energy Convers. Manag.*, vol. 118, pp. 105–118, Jun. 2016.

[58] L. Mazorra-Aguiar, B. Pereira, M. David, F. Díaz, and P. Lauret, "Use of satellite data to improve solar radiation forecasting with Bayesian Artificial Neural Networks," *Sol. Energy*, vol. 122, pp. 1309–1324, Dec. 2015.

[59] F. Tymvios, C. Jacovides, S. Michaelides, and C. Scouteli, "Comparative study of Ångström's and artificial neural networks' methodologies in estimating global solar radiation," *Sol. Energy*, vol. 78, no. 6, pp. 752–762, Jun. 2005.

[60] C. Jacovides, F. Tymvios, V. Assimakopoulos, and N. Kaltsounides, "Comparative study of various correlations in estimating hourly diffuse fraction of global solar radiation," *Renew. Energy*, vol. 31, no. 15, pp. 2492–2504, Dec. 2006.

[61] R. D. Tapakis and A. G. Charalambides, "Monitoring Cloud Motion in Cyprus for Solar Irradiance Prediction," *Conf. Pap. Energy*, vol. 2013, pp. 1–6, 2013.

[62] H. Kasht, "Sky conditions classification and estimation of solar radiation for clear sky days," Middle East Technical University Northern Cyprus Campus, 2018.

[63] A. Mellit, "Artificial Intelligence technique for modelling and forecasting of solar radiation data: a review," *Int. J. Artif. Intell. Soft Comput.*, vol. 1, no. 1, p. 52, 2008.

[64] D. S. Kumar, G. M. Yagli, M. Kashyap, and D. Srinivasan, "Solar irradiance

resource and forecasting: a comprehensive review," *IET Renew. Power Gener.*, vol. 14, no. 10, pp. 1641–1656, Jul. 2020.

[65] M. Guermoui, F. Melgani, K. Gairaa, and M. L. Mekhalfi, "A comprehensive review of hybrid models for solar radiation forecasting," *Journal of Cleaner Production*, vol. 258. Elsevier Ltd, p. 120357, 10-Jun-2020.

[66] M. Guermoui, F. Melgani, and C. Danilo, "Multi-step ahead forecasting of daily global and direct solar radiation: A review and case study of Ghardaia region," *J. Clean. Prod.*, vol. 201, pp. 716–734, Nov. 2018.

[67] L. Benali, G. Notton, A. Fouilloy, C. Voyant, and R. Dizene, "Solar radiation forecasting using artificial neural network and random forest methods: Application to normal beam, horizontal diffuse and global components," *Renew. Energy*, vol. 132, pp. 871–884, Mar. 2019.

[68] S. Sperati, S. Alessandrini, P. Pinson, and G. Kariniotakis, "The 'Weather Intelligence for Renewable Energies' Benchmarking Exercise on Short-Term Forecasting of Wind and Solar Power Generation," *Energies*, vol. 8, no. 9, pp. 9594–9619, Sep. 2015.

[69] M. Ilkan, E. Erdil, and F. Egelioglu, "Renewable energy resources as an alternative to modify the load curve in Northern Cyprus," *Energy*, vol. 30, no. 5, pp. 555–572, 2005.

[70] "POWER Data Access Viewer." [Online]. Available: https://power.larc.nasa.gov/data-access-viewer/?fbclid=IwAR1yPlfK_3RPZbL3RWwHIrizUeq8SugivFCDN7ASnIeuC8lfO-3TJSlrlRg. [Accessed: 09-May-2020].

[71] J. D. Rios, A. Y. Alanis, N. Arana-Daniel, and C. Lopez-Franco, "Artificial neural networks," in *Neural Networks Modeling and Control*, E. N. Sanchez, Ed. Elsevier, 2020, pp. 117–124.

[72] H. Zang *et al.*, "Hybrid method for short-term photovoltaic power forecasting based on deep convolutional neural network," *IET Gener. Transm. Distrib.*, vol. 12, no. 20, pp. 4557–4567, Nov. 2018.

[73] F. Wang *et al.*, "Generative adversarial networks and convolutional neural networks based weather classification model for day ahead short-term photovoltaic power forecasting," *Energy Convers. Manag.*, vol. 181, no. August 2018, pp. 443–462, 2019.

[74] J. Gu *et al.*, "Recent advances in convolutional neural networks," *Pattern Recognit.*, vol. 77, pp. 354–377, May 2018.

[75] "Conv1D layer." [Online]. Available: https://keras.io/api/layers/convolution_layers/convolution1d/. [Accessed: 14-Feb-2021].

[76] R. Pascanu, T. Mikolov, and Y. Bengio, “On the difficulty of training recurrent neural networks,” 2013.

[77] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[78] F. A. Gers, J. Schmidhuber, and F. Cummins, “Learning to forget: Continual prediction with LSTM,” in *IEEE Conference Publication*, 1999, vol. 2, no. 470, pp. 850–855.

[79] K. R. Müller, A. J. Smoła, G. Rätsch, B. Schölkopf, J. Kohlmorgen, and V. Vapnik, “Predicting time series with support vector machines,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 1997, vol. 1327, pp. 999–1004.

[80] M. Wauters and M. Vanhoucke, “Support Vector Machine Regression for project control forecasting,” *Autom. Constr.*, vol. 47, pp. 92–106, Nov. 2014.

[81] R. C. Deo, X. Wen, and F. Qi, “A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset,” *Appl. Energy*, vol. 168, pp. 568–593, Apr. 2016.

[82] T. Kleynhans, M. Montanaro, A. Gerace, and C. Kanan, “Predicting Top-of-Atmosphere Thermal Radiance Using MERRA-2 Atmospheric Data with Deep Learning,” *Remote Sens.*, vol. 9, no. 11, p. 1133, Nov. 2017.

[83] M. F. Sanner, “Python: A programming language for software integration and development,” *Journal of Molecular Graphics and Modelling*, vol. 17, no. 1. pp. 57–61, 1999.

[84] N. Ketkar and N. Ketkar, “Introduction to Keras,” in *Deep Learning with Python*, Apress, 2017, pp. 97–111.

[85] M. Abadi *et al.*, “TensorFlow: A system for large-scale machine learning,” *Proc. 12th USENIX Symp. Oper. Syst. Des. Implementation, OSDI 2016*, pp. 265–283, May 2016.

[86] F. Pedregosa *et al.*, “Scikit-learn: Machine learning in Python,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.

APPENDICES

A. Forecasting Results for NASA Dataset

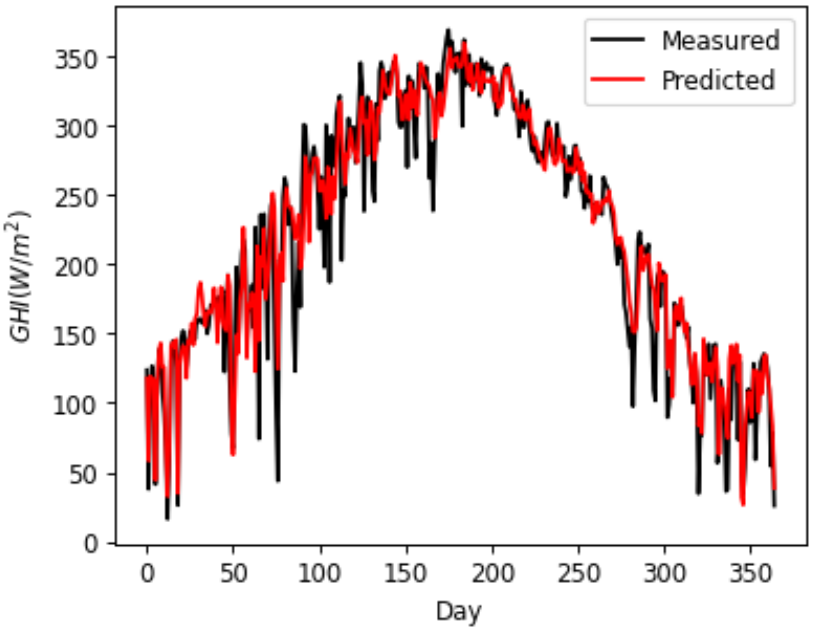

Figure A.1. Prediction performance of exogenous LSTM over a year

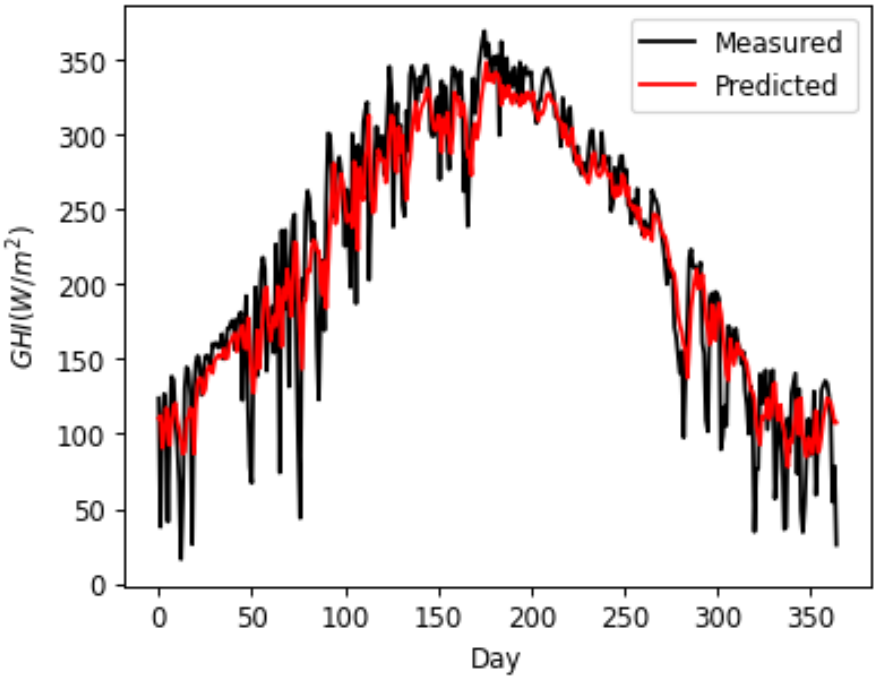

Figure A.2. Prediction performance of endogenous LSTM over a year

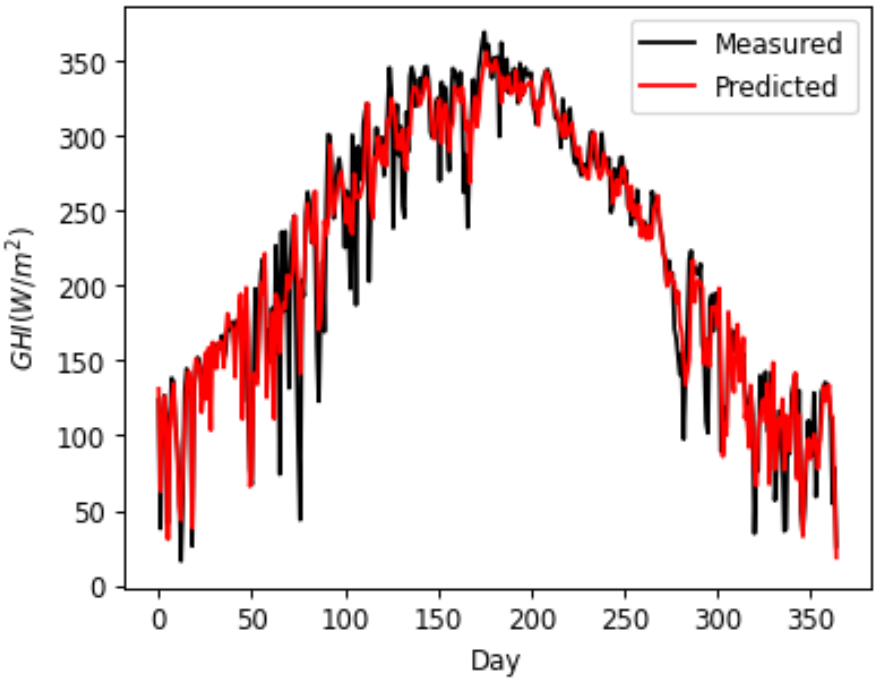

Figure A.3. Prediction performance of exogenous CNN over a year

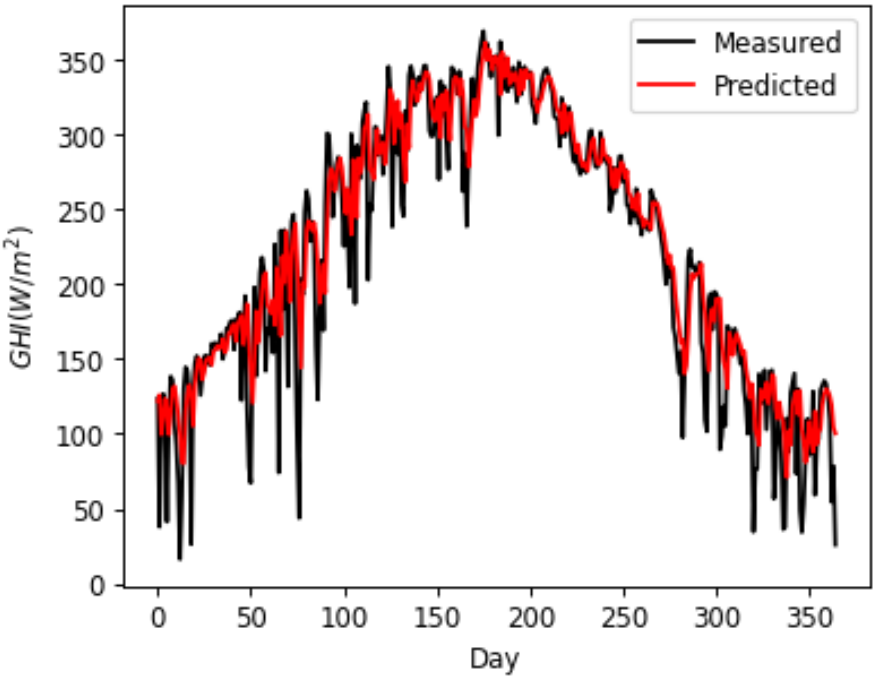

Figure A.4. Prediction performance of endogenous CNN over a year

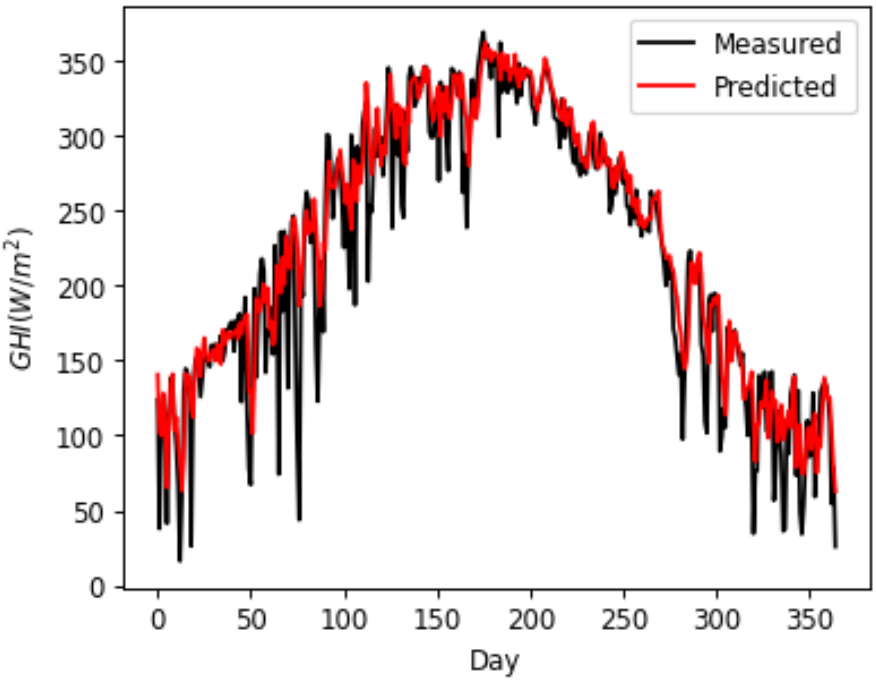

Figure A.5. Prediction performance of hybrid CN-M over a year

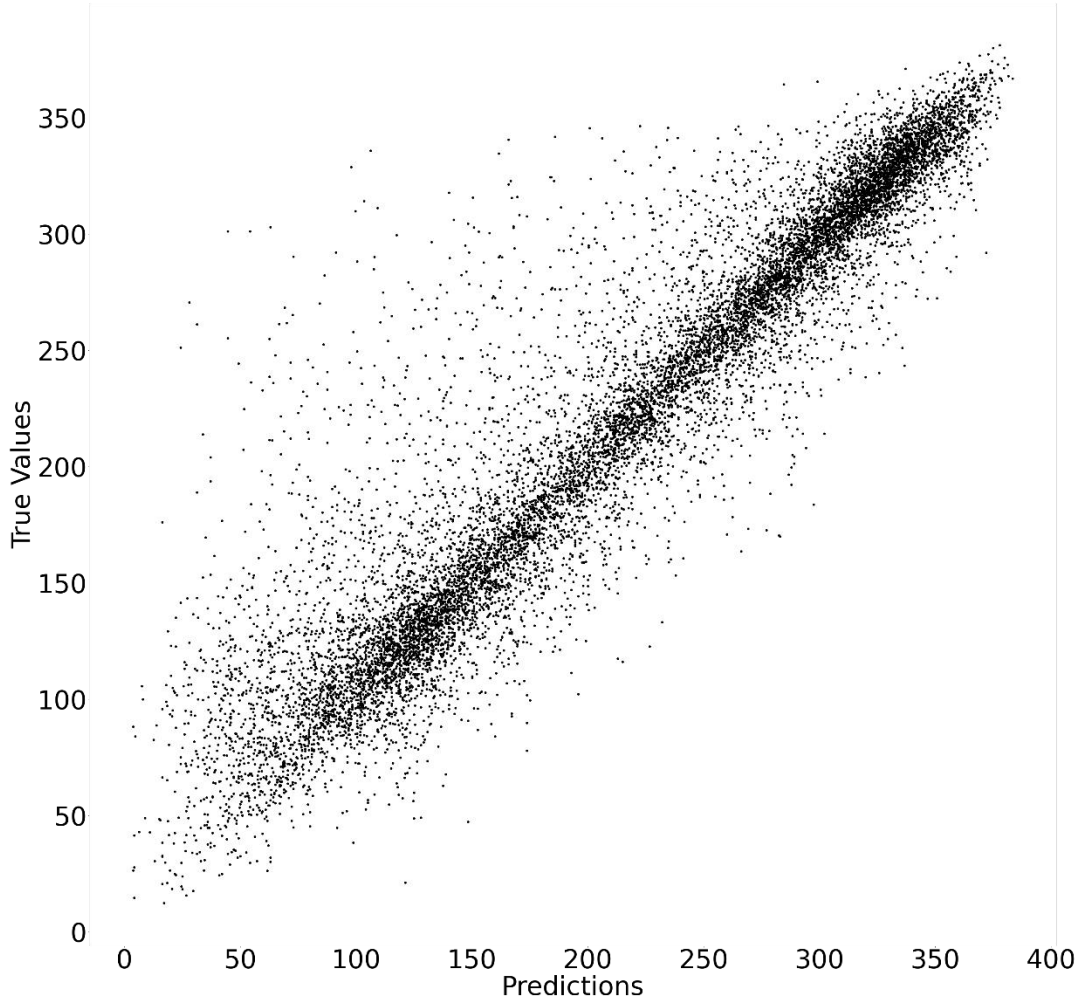

Figure A.6. Predicted and actual GHI values on a scattered plot for exogenous LSTM over the testing set

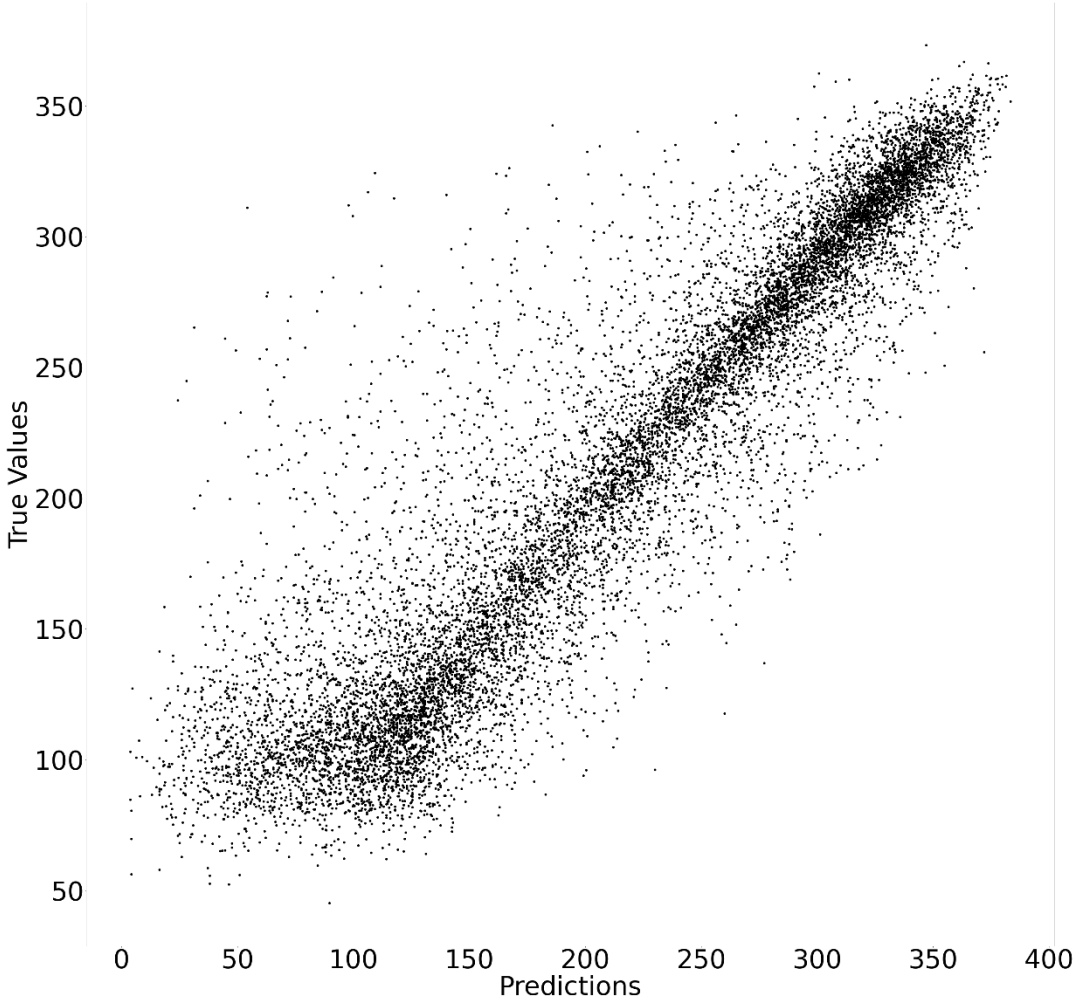

Figure A.7. Predicted and actual GHI values on a scattered plot for endogenous LSTM over the testing set

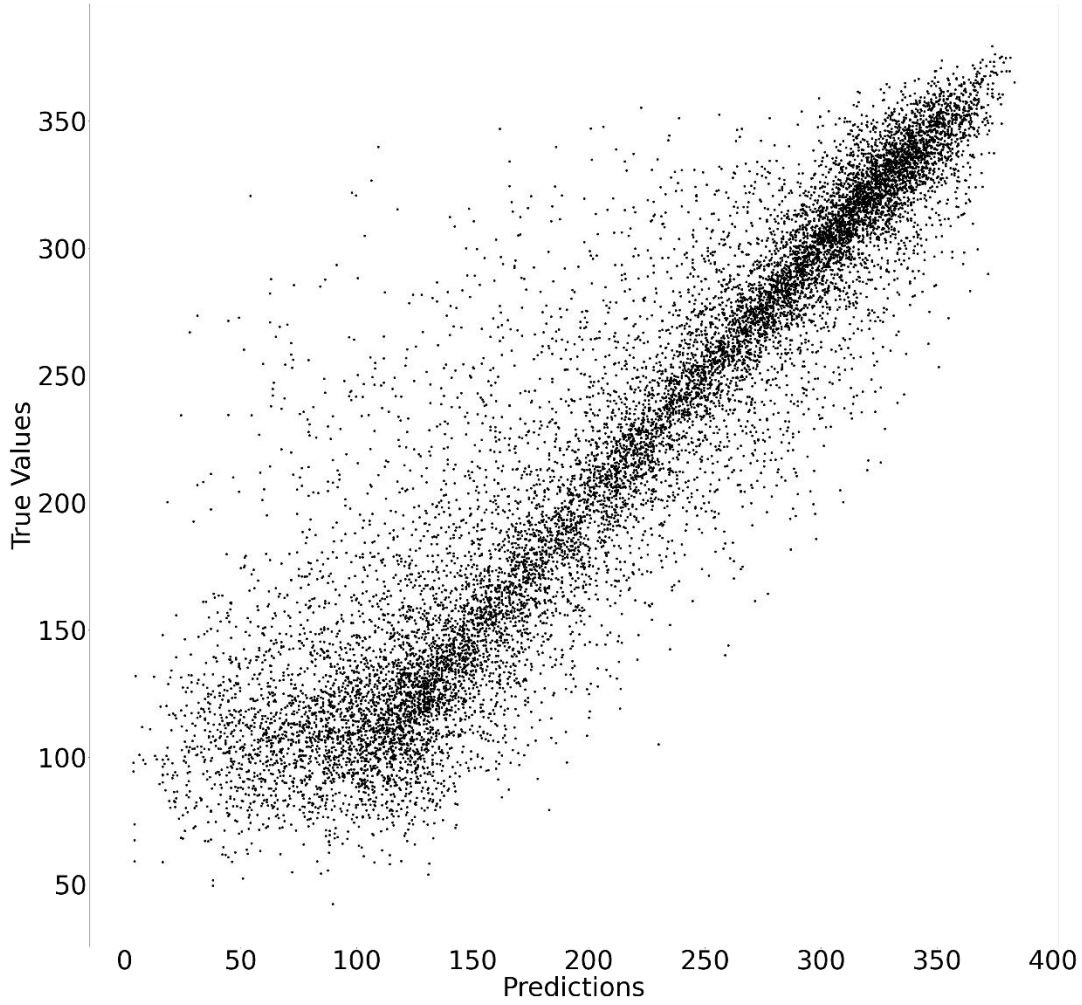

Figure A.58. Predicted and actual GHI values on a scattered plot for exogenous CNN over the testing set

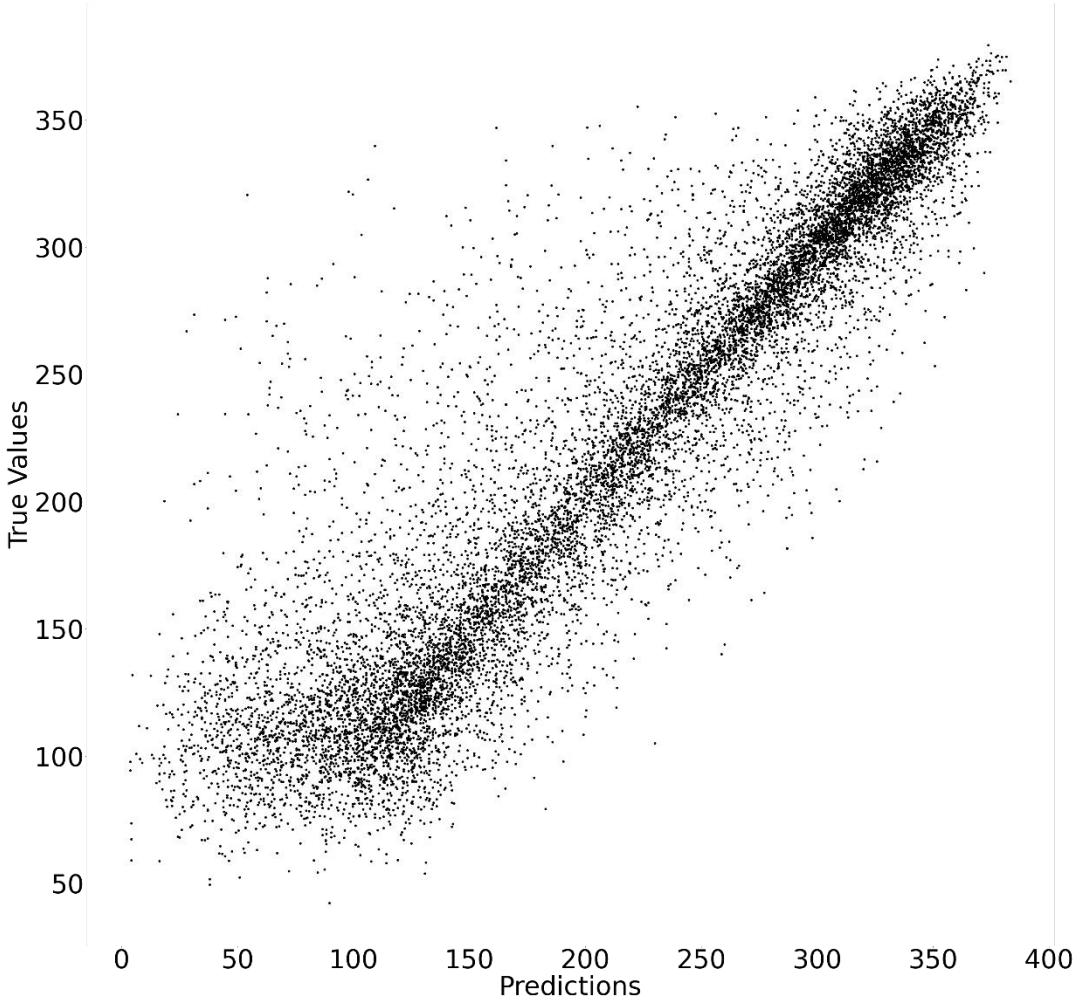

Figure A.9. Predicted and actual GHI values on a scattered plot for endogenous CNN over the testing set

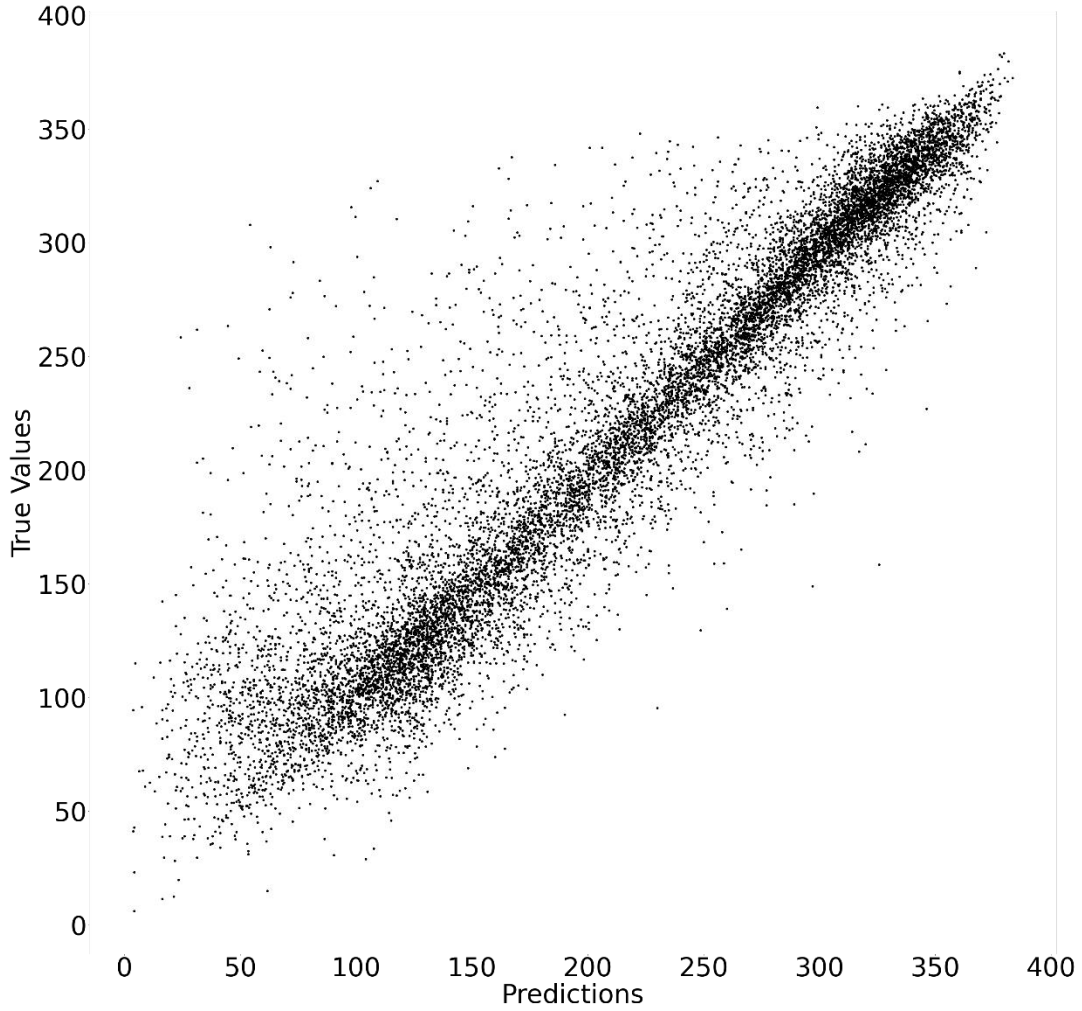

Figure A.10. Predicted and actual GHI values on a scattered plot for CN-M over the testing set

B. **Forecasting Results for METU NCC Dataset**

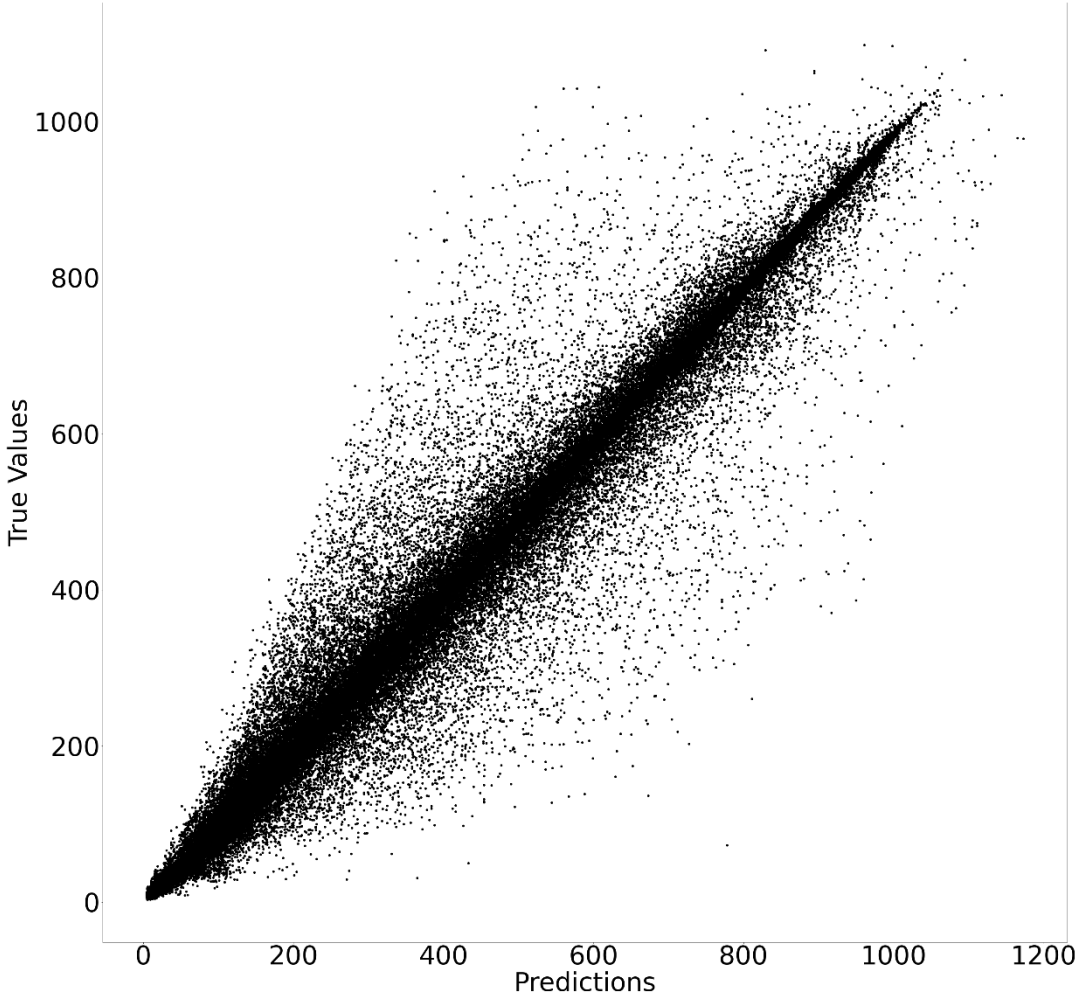

Figure B.1. Predicted and actual GHI values on a scattered plot for CNN over the testing set

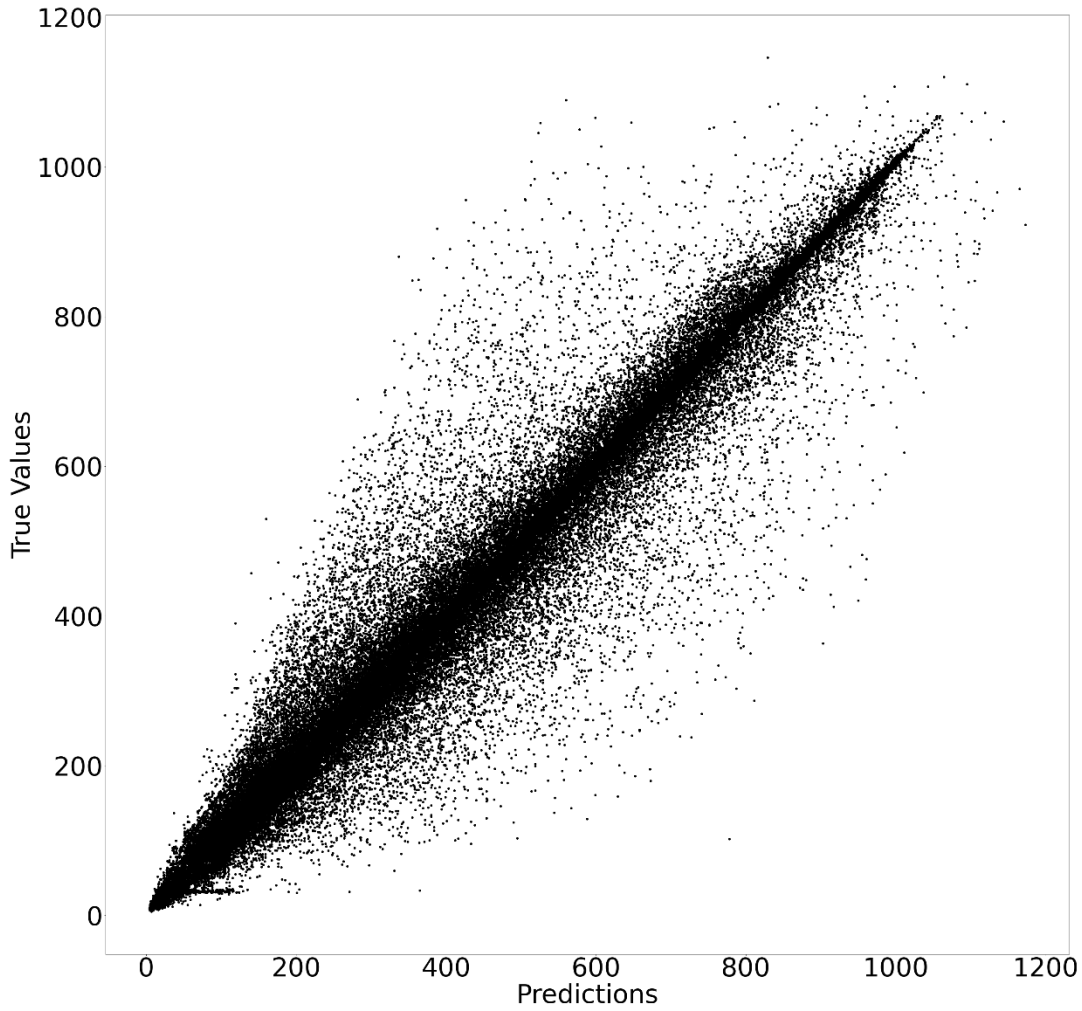

Figure B.2. Predicted and actual GHI values on a scattered plot for LSTM over the testing set

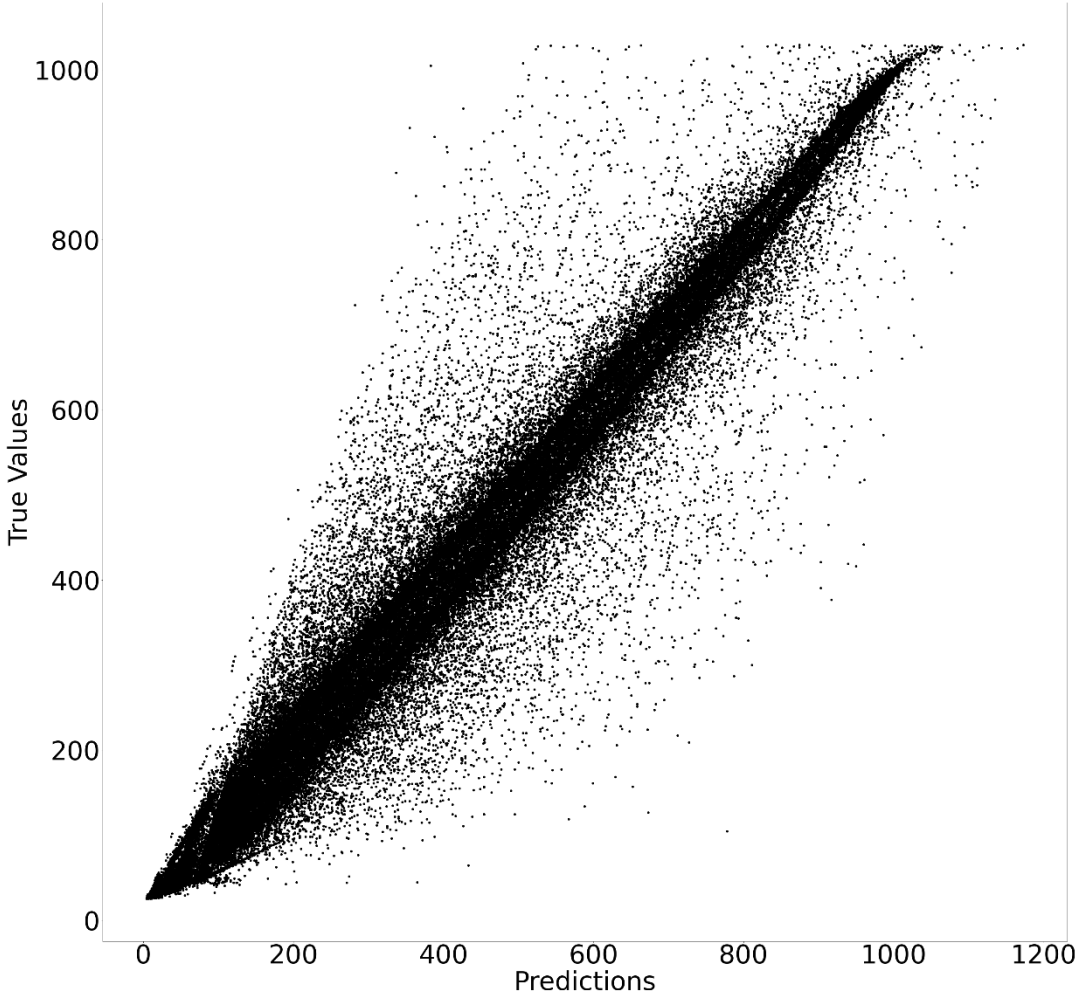

Figure B.3. Predicted and actual GHI values on a scattered plot for SVR over the testing set

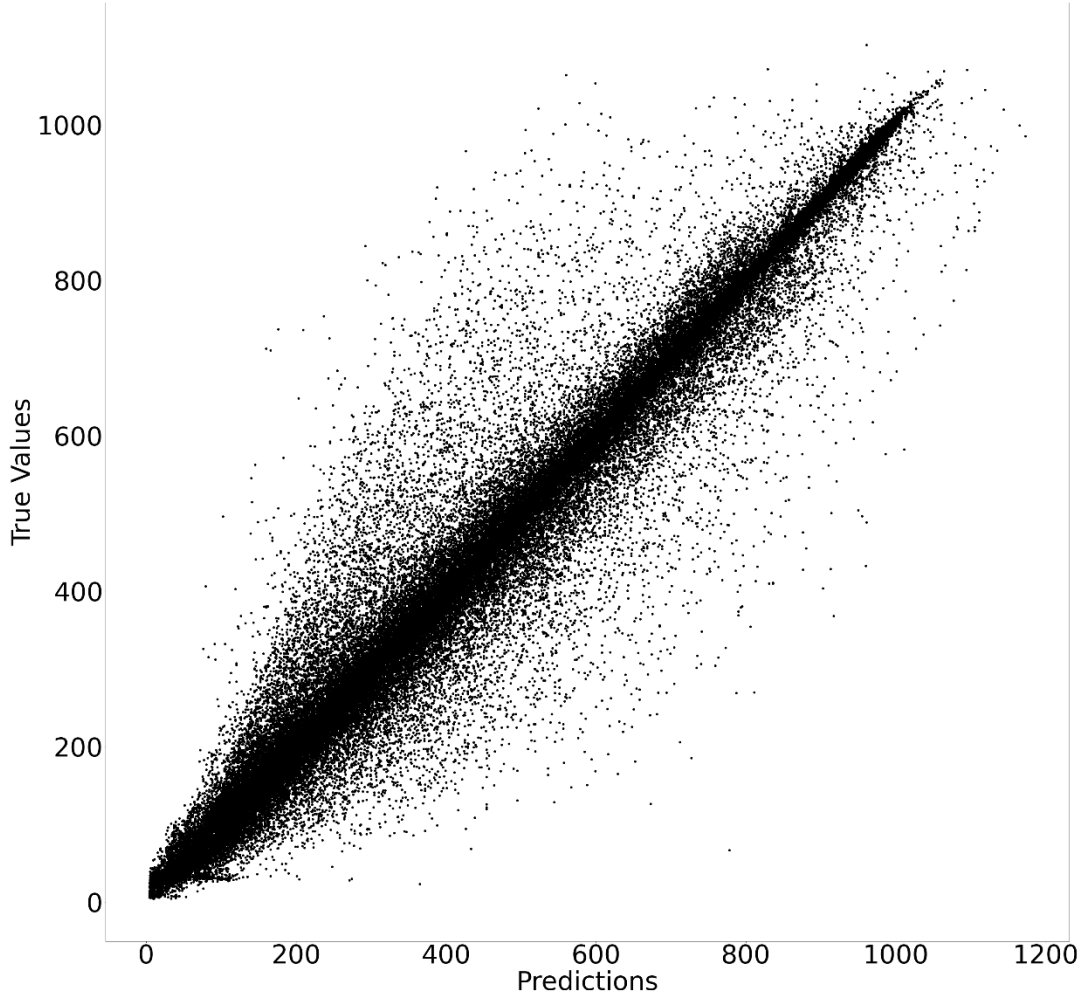

Figure B.4. Predicted and actual GHI values on a scattered plot for C-LSTM over the testing set

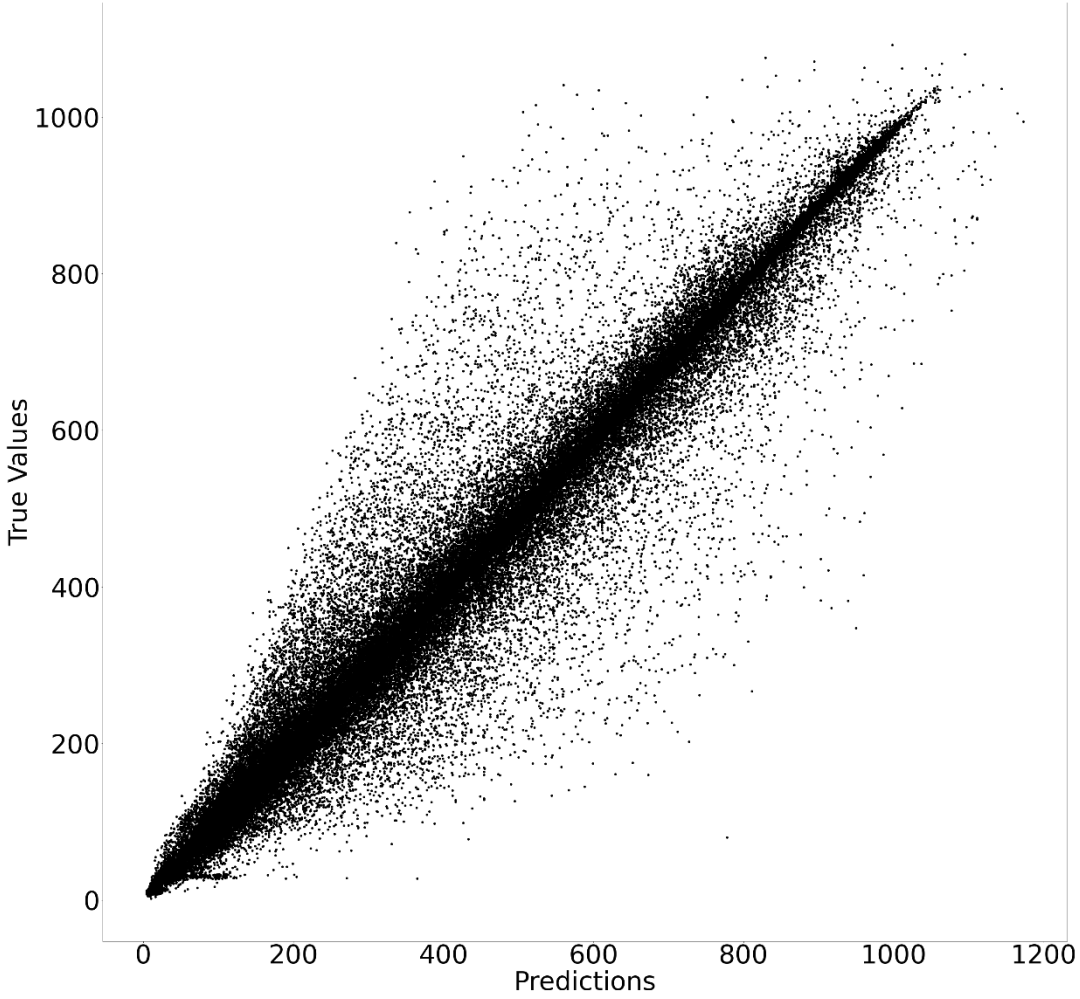

Figure B.5. Predicted and actual GHI values on a scattered plot for CN-M over the testing set

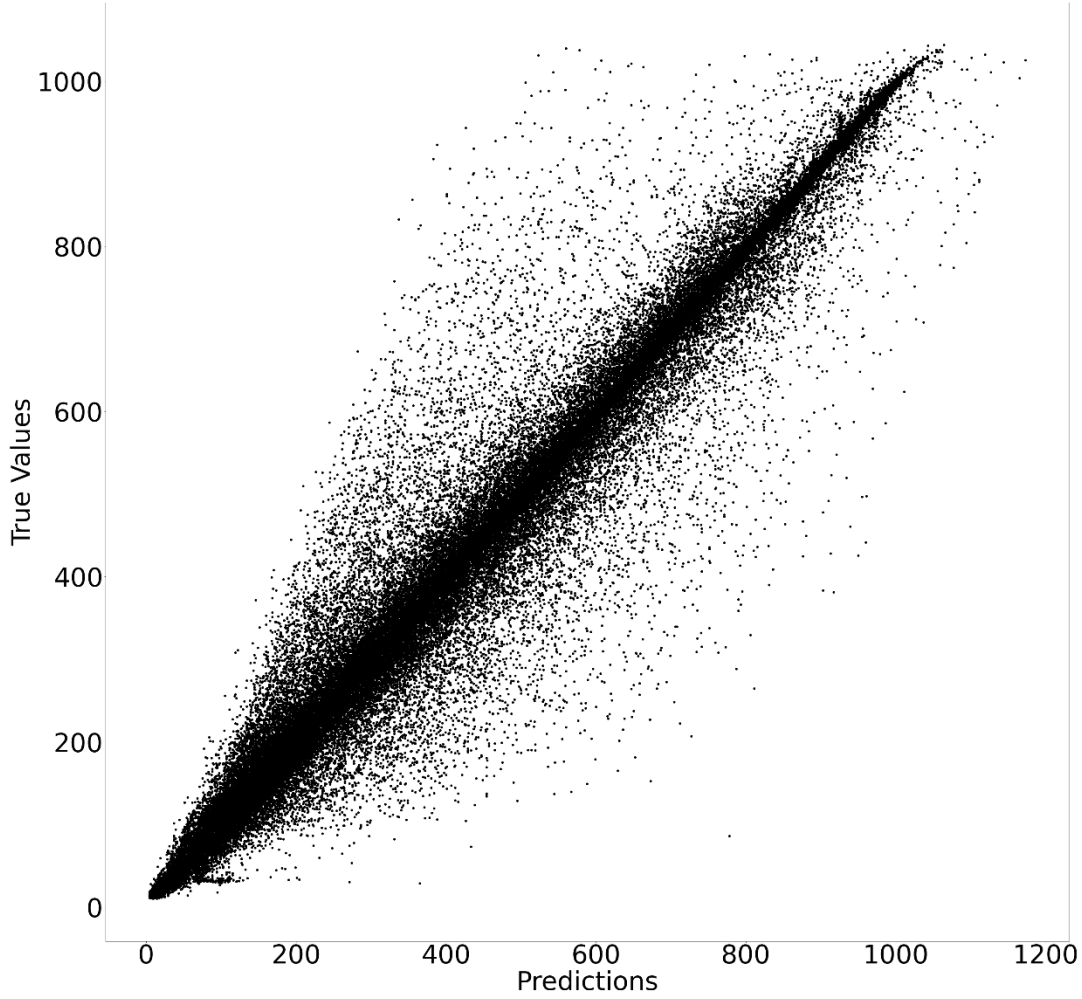

Figure B.6. Predicted and actual GHI values on a scattered plot for CM-SVR over the testing set

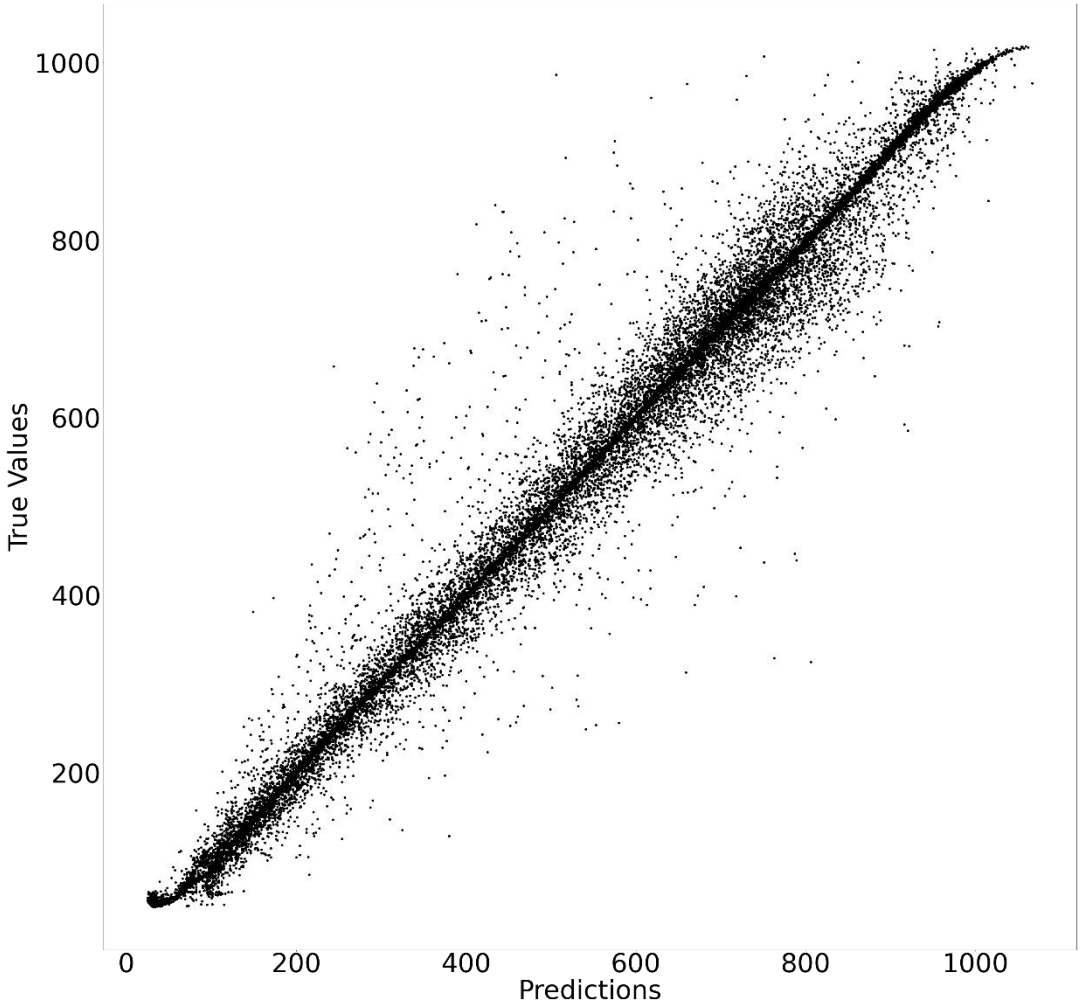

Figure B.7. Predicted and actual GHI values on a scattered plot for the summer season over the testing set

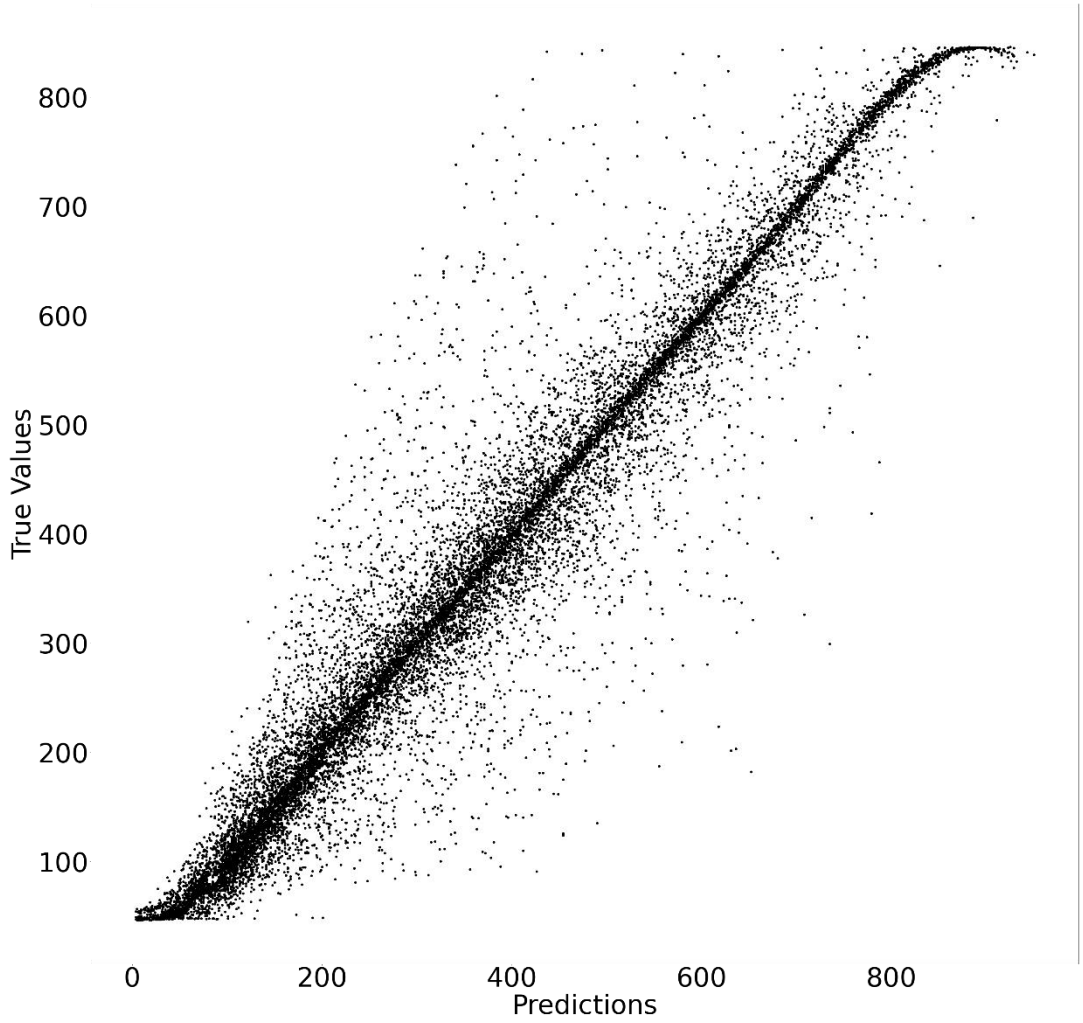

Figure B.8. Predicted and actual GHI values on a scattered plot for the fall season over the testing set

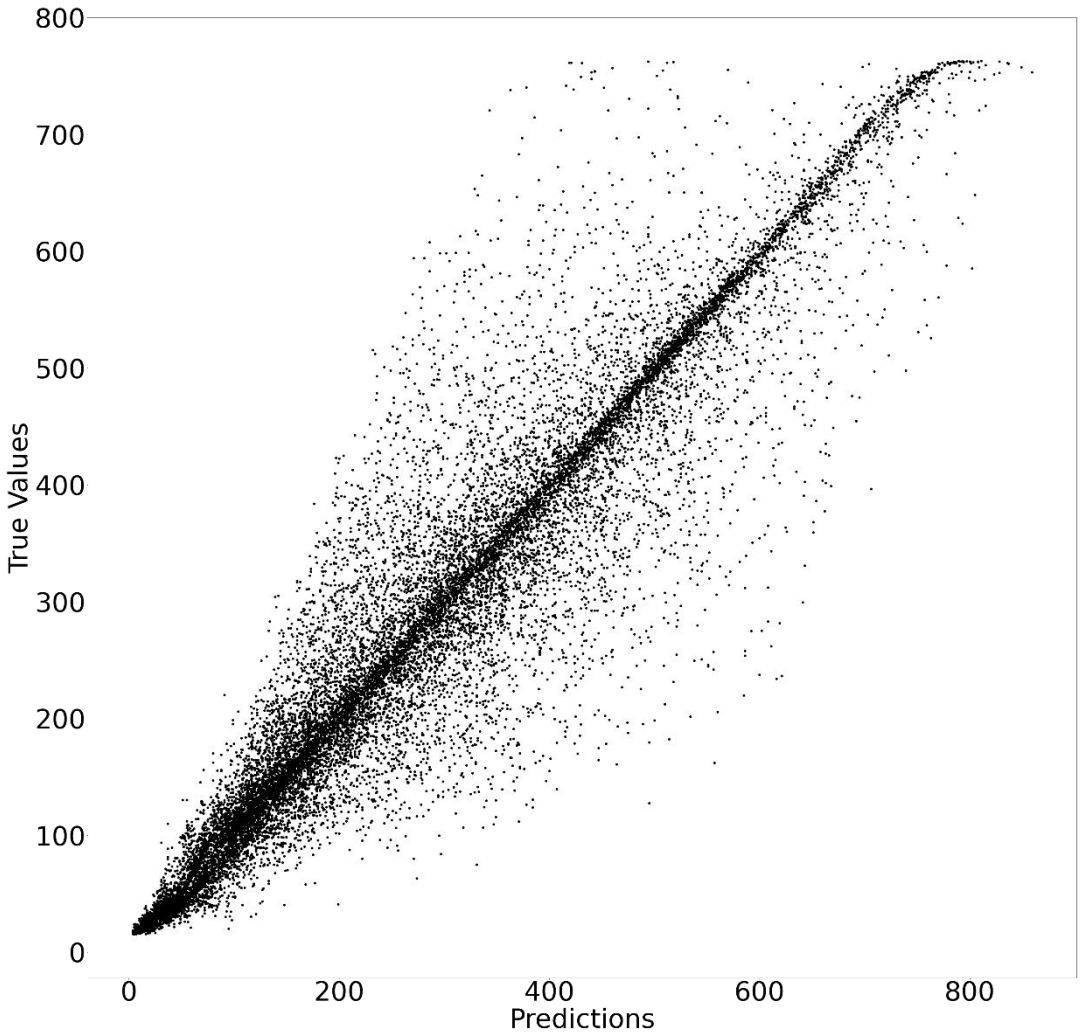

Figure B.9. Predicted and actual GHI values on a scattered plot for the winter season over the testing set

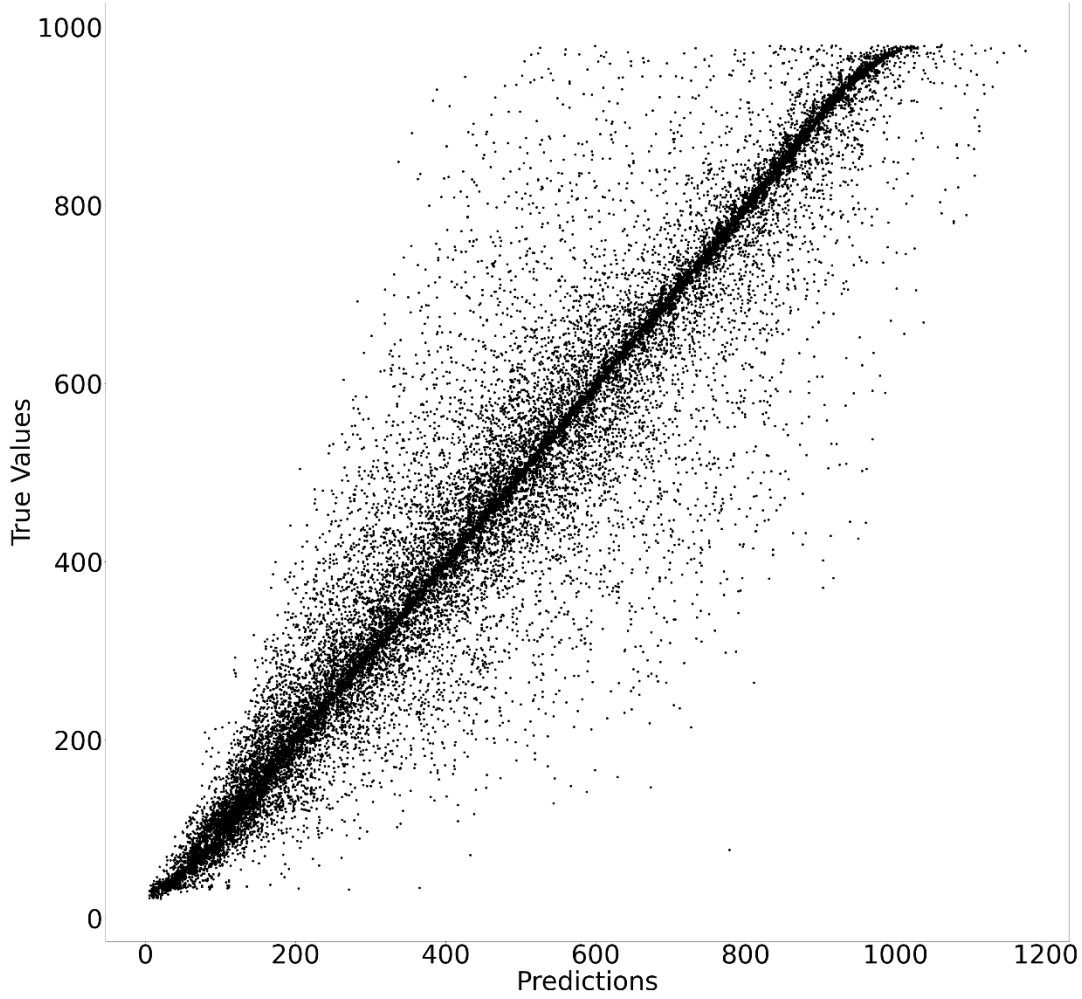

Figure B.10. Predicted and actual GHI values on a scattered plot for the spring season over the testing set

ORTA DOĞU TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜ

TEZ İZİN FORMU / THESIS PERMISSION FORM

PROGRAM / PROGRAM

Sürdürülebilir Çevre ve Enerji Sistemleri / Sustainable Environment and Energy Systems ☑

Siyaset Bilimi ve Uluslararası İlişkiler / Political Science and International Relations ☐

İngilizce Öğretmenliği / English Language Teaching ☐

Elektrik Elektronik Mühendisliği / Electrical and Electronics Engineering ☐

Bilgisayar Mühendisliği / Computer Engineering ☐

Makina Mühendisliği / Mechanical Engineering ☐

YAZARIN / AUTHOR
Soyadı / Surname : Vakitbilir
Adı / Name : Nuray
Programı / Program : Sustainable Environment and Energy Systems

TEZİN ADI / TITLE OF THE THESIS (**İngilizce** / English) :
Multivariate Forecasting of Global Horizontal Irradiation Using Deep Learning Algorithms

TEZİN TÜRÜ / DEGREE: **Yüksek Lisans** / Master ☑ **Doktora** / PhD ☐

1. **Tezin tamamı dünya çapında erişime açılacaktır. /** Release the entire work immediately for access worldwide. ☐

2. Tez iki yıl süreyle erişime kapalı olacaktır. / Secure the entire work for patent and/or proprietary purposes for a period of **two years.** * ☑

3. Tez altı ay süreyle erişime kapalı olacaktır. / Secure the entire work for period of **six months**.* ☐

Yazarın imzası / Author Signature ..... **Tarih** / Date 11/03/2021

Tez Danışmanı / Thesis Advisor Full Name: Asst. Prof. Dr. Cem Direkoğlu

Tez Danışmanı İmzası / Thesis Advisor Signature: .....

Eş Danışmanı / Co-Advisor Full Name: .....

Eş Danışmanı İmzası / Co-Advisor Signature: .....

Program Koordinatörü / Program Coordinator Full Name: Assoc. Prof. Dr. Ceren İnce Derogar

Program Koordinatörü İmzası / Program Coordinator Signature: .....