ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES

MSc THESIS

Eman Ahmed Erfan Othman ABDALLA

DEEP LEARNING-BASED DAILY SOLAR IRRADIATION PREDICTION FOR ADANA REGION

DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING

ADANA-2023

ABSTRACT

MSc THESIS

DEEP LEARNING-BASED DAILY SOLAR IRRADIATION PREDICTION FOR ADANA REGION

Eman Ahmed Erfan Othman ABDALLA

ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES
DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING

Supervisor : Prof. Dr. İlyas EKER
Co-Supervisor : Asst. Prof. Dr. Kasım ZOR
Year: 2023, Page: 116
Jury : Prof. Dr. İlyas EKER
: Asst. Prof. Dr. Oğuzhan TİMUR
: Asst.Prof. Dr. Necdet Sinan ÖZBEK

Nature suffers because of the negative effects and insufficiency of fossil energy resources such as oil, coal, and natural gas, which is why renewable energy generation has received significant attention. To replace harmful fossil fuels with clean, safe, and sustainable energy, the photovoltaic (PV) market has grown rapidly in the last decades. A high-accuracy daily solar global horizontal irradiation (GHI) prediction model is a critical tool for the operation, maintenance, and optimization of PV systems so that developing an intelligent and robust GHI prediction model has become one of the most captivating topics among researchers in recent years. In this thesis, meteorological data of Adana region were obtained from MERRA-2 via introducing calendar data, then the dataset was wrangled to treat missing and erroneous data by using Python programming language in order to form a cleansing dataset that is normalized and split into 80% and 20% for training and testing dataset respectively. Moreover, four different feature selection scenarios were created depending on Pearson's correlation and the threshold limits were specified as 0, 0.10, 0.25, and 0.5. After day-head GHI prediction models were carried out using different deep learning-based models with a statistical and a traditional artificial neural network algorithm, namely multiple linear regression (MLR), multilayer perceptron (MLP), recurrent neural network (RNN), long short-term memory (LSTM), gated recurrent unit (GRU), bidirectional LSTM, and bidirectional GRU. The prediction results were evaluated by utilizing a variety of performance metrics, including R-squared (R^2), normalized mean absolute error (nMAE), mean absolute percentage error (MAPE), and normalized root mean squared error (nRMSE). The Bi-LSTM model has been observed to outperform the other models in terms of R^2 (85.495%) and MAPE (19.901%).

Key Words: Daily Global Horizontal Irradiation Prediction, Deep Learning, Recurrent Neural Networks, Long Short-Term Memory, Gated Recurrent Unit, Adana.

ÖZ

YÜKSEK LİSANS TEZİ

ADANA BÖLGESİ İÇİN DERİN ÖĞRENME TABANLI GÜNLÜK GÜNEŞ IŞINIMI TAHMİNİ

Eman Ahmed Erfan Othman ABDALLA

ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ ANA BİLİM DALI

Danışman : Prof. Dr. İlyas EKER
II. Danışman : Dr. Öğr. Üyesi Kasım ZOR
Yıl: 2023, Sayfa: 116
Jüri : Prof. Dr. İlyas EKER
: Dr. Öğr. Üyesi Oğuzhan TİMUR
: Dr. Öğr. Üyesi Necdet Sinan ÖZBEK

Petrol, kömür ve doğal gaz gibi fosil enerji kaynaklarının olumsuz etkileri ve yetersizliği doğaya zarar vermekte ve bu nedenle yenilenebilir enerji üretimi kaydadeğer şekilde önem kazanmıştır. Zararlı fosil yakıtları, temiz, güvenli ve sürdürülebilir enerjiyle değiştirmek için, fotovoltaik (FV) sektörü son yıllarda çok hızlı bir şekilde büyümüştür. Yüksek doğruluğa sahip günlük güneş küresel yatay ışınımı (KYI) tahmin modeli, FV sistemlerin çalıştırılması, bakımı ve optimizasyonu için kritik bir araçtır, öyle ki akıllı ve gürbüz KYI tahmin modeli son zamanlarda araştırmacılar arasında en ilgi çekici konulardan biri haline gelmiştir. Bu tezde, Adana bölgesinin meteorolojik verileri MERRA-2'den takvim değişkenleri eklenerek alınmış, daha sonra ham veri seti Python programlama dili aracılığıyla eksik ve hatalı veriler tedavi edilerek temiz veri seti oluşturulmuş ve bu veri seti normalize edildikten sonra bölünerek sırasıyla %80 ve %20 oranlarında eğitim ve test veri setlerine ayrılmıştır. Ayrıca, Pearson korelasyonuna göre 0, 0,10, 0,25 ve 0,50 eşikler belirlenerek, dört farklı öz nitelik seçimi senaryosu oluşturulmuştur. Ardından, gün öncesi KYI tahmin modellerinin gerçekleştirilmesinde değişik derin öğrenme tabanlı modeller ile bir istatistiki ve bir geleneksel yapay sinir ağı algoritması olarak adlandırılan çoklu doğrusal regresyon (ÇDR), çok katmanlı algılayıcı (ÇKA), yinelemeli sinir ağı (YSA), uzun kısa süreli bellek (UKSB), kapılı tekrarlayan hücre (KTH), çift yönlü UKSB ve çift yönlü KTH kullanılmıştır. Tahmin sonuçlarının değerlendirilmesinde belirleme katsayısı (BK), normalize edilmiş ortalama mutlak hata (nOMH), ortalama mutlak yüzdesel hata (OMYH) ve normalize edilmiş kök ortalama karesel hatayı (nKOKH) da içeren çeşitli performans ölçümlerinden yararlanılmıştır. Çift yönlü UKSB modelinin, BK (%85,495) ve OMYH (%19.901) performans ölçümlerine dayanarak diğer modellere göre daha yüksek performans sergilediği gözlemlenmiştir.

Anahtar kelimeler: Günlük Küresel Yatay Işınımı Tahmini, Derin Öğrenme, Yinelemeli Sinir Ağları, Uzun Kısa Süreli Bellek, Kapılı Tekrarlayan Hücre, Adana.

EXTENDED SUMMARY

Solar irradiation forecasting is a vital tool in grid-connected power systems for optimally dispatching solar and other energy sources to balance generation and demand. Improving solar forecasts will allow the electric grid to be more adaptable to changing conditions, resulting in fewer disruptions and lower overall operating costs.

Consumption of fossil fuels accounts for a significant portion of global and domestic energy requirements. The use of fossil fuels such as oil, coal, and natural gas is widely acknowledged to emit a large number of greenhouse gases into the atmosphere, resulting in highly negative environmental effects. Solar energy, which has begun to be used to meet the world's increasing energy needs, can be used to produce cleaner and carbon-free energy. The liberalization of the electricity market and rising demand for sustainable energy have shifted political and investment interests toward using renewable energy systems to meet electricity needs (Suttakul et al. 2022).

Power system advancement, high renewable energy penetration, and active consumer participation all lead to more complex power system operations. Physical and numerical modelling are heavily used in traditional power system analysis and control methods. These methods are unable to address recent hybrid smart-grid issues. Meteorological stations and electricity market both generate massive amounts of data at the same time. As a result, Deep Learning (DL) is an appropriate technology for smart-grid analysis's future development and success (A. Muhammad et al. 2019). There have been numerous recent studies on GHI forecasting using DL.

In this thesis, a research and application of day-ahead GHI prediction using deep learning-based techniques for Adana region. All data wrangling, employing and evaluating day-ahead GHI model computations were done on the Kaggle platform (Kaggle, 2022) using the Python programming language (Van Rossum, G., & Drake Jr, F. L., 1995). Tensorflow's Keras library (Chollet et al.2015) was used as the

backend for developing the GHI forecasting model. The Keras tuner tool was also used to tune DL hyperparameters. The Scikit-learn (Pedregosa et al.2011) library's functions were also used during the data processing phase to normalize and split the dataset. Matplotlib (Hunter, 2007)), Seaborn (Waskom, et al. 2017), and Pandas (McKinney et al.2010) were used extensively to plot the various graphs, as well as Pandas (Harris et al. 2020) to wrangle the dataset.

The meteorological raw data used in this thesis are daily resolution data from January 1, 2010, to September 30, 2022, from the GMAO (Global Modelling and Assimilation Office) of NASA (US National Aeronautics and Space Administration).

Furthermore, calendar parameters such as the year, the month of the year (MoY), the week of the year (WoY), and the day of the month (DoM) are extracted and added as input features, as well as enriching the dataset with some GHI lag variables to capture the seasonal trends in GHI.

To use a common scale without distorting differences in value ranges or losing information, the dataset was normalized between [0,1] by the Min-Max transformation function after removing outliers and erroneous data. The data set was divided into two parts for evaluation: training (80%) and testing (20%).

Furthermore, feature selection criteria based on Pearson's correlation were investigated in order to find the effects of GHI features. To do so, four different feature selection scenarios were created depending on Pearson’s correlation and the threshold limits were specified as 0, 0.10, 0.25, and 0.5. After day-head GHI prediction models were carried out using different deep learning-based models with a statistical and a traditional artificial neural network algorithm, namely multiple linear regression (MLR), multilayer perceptron (MLP), recurrent neural network (RNN), long short-term memory (LSTM), gated recurrent unit (GRU), bidirectional LSTM, and bidirectional GRU.

As a consequence, the coefficient of determination (R2), mean absolute percentage error (MAPE), normalized mean absolute error (nMAE), and normalized

root mean squared error (nRMSE) were used to evaluate the results, and future perspectives are discussed after the conclusions. It was observed that the Bi-LSTM model in the first scenario outperforms all other statistical and AI models by (R2= 85.495%, MAPE=19.901%).

GENİŞTİRİLMİŞ ÖZET

Güneş ışınımı tahmini, elektrik şebekelerine bağlı güç sistemlerinde, üretimi ve talebi dengelemek için güneş ve diğer enerji kaynaklarını en iyi şekilde kullanmanın önemli bir aracıdır. Güneş tahminlerini iyileştirmek, elektrik şebekesinin değişen koşullara daha uyumlu olmasını sağlayarak daha az kesinti ve genel işletme maliyetlerini azaltır.

Fosil yakıtların tüketimi, küresel ve iç enerji gereksinimlerinin önemli bir bölümünü oluşturur. Petrol, kömür ve doğal gaz gibi fosil yakıtların kullanılması, atmosfere çok sayıda sera gazı yaydığı ve bunun sonucunda da son derece olumsuz çevresel etkilere neden olduğu yaygın bir şekilde kabul edilmektedir. Dünyanın artan enerji ihtiyaçlarını karşılamak için kullanılan güneş enerjisi, daha temiz, karbonsuz enerji üretmek için kullanılmaktadır. Elektrik enerjisi pazarının serbestleştirilmesi ve sürdürülebilir enerji talebinin artması, siyasi ve yatırım çıkarlarını elektrik ihtiyaçlarını karşılamak için yenilenebilir enerji sistemleri kullanmaya kaydırdı.
Güç sisteminin ilerlemesi, yenilenebilir enerjinin yüksek nüfuzu ve aktif tüketici katılımı, daha karmaşık güç sistemi operasyonlarına yol açar. Fiziksel ve sayısal modelleme, geleneksel güç sistemi analizi ve kontrol yöntemlerinde yoğun bir şekilde kullanılır. Bu yöntemler, güncel melez akıllı şebeke sorunlarını giderememektedir. Meteoroloji istasyonları ve elektrik piyasası aynı anda büyük miktarda veri üretir. Sonuç olarak, derin öğrenme algoritması, akıllı şebekelerin analizinin gelecekteki gelişimi ve başarısı için uygun bir teknolojidir. Derin öğrenme kullanılarak KYI tahmini hakkında yakın zamanda çok sayıda araştırma gerçekleştirilmiştir.

Bu tezde, Adana bölgesinde derin öğrenme temelli algoritmalar kullanılarak, gün öncesi KYI tahmininin araştırılması ve uygulanması çalışması gerçekleştirilmiştir. Tüm veri manipülasyonu ve gün öncesi KYI model hesaplamaları Kaggle platformu üzerinde Python programlama dili kullanılarak gerçekleştirilmiştir. Tensorflow'un Keras kütüphanesi KYI tahmin modelinin geliştirilmesinde arka uç olarak kullanılmıştır. Keras ayarlama aracı derin öğrenme parametrelerinin belirlenmesinde görev almıştır. Sci-kit learn kütüphanesi fonksiyonları da, veri işleme fazında verilerin normalizasyonunu sağlamak ve veri setini ayırmak için tercih edilmiştir. Matplotlib, Seaborn ve Pandas modüllerinden çeşitli grafiklerin çiziminde faydalanılmış ve ayrıca Pandas kütüphanesi ek olarak ham veri setinin temiz veri setine dönüştürülmesi aşamasında da kullanılmıştır.

Bu tez içinde kullanılan meteorolojik ham veriler, NASA'nın 1 Ocak 2010 - 30 Eylül 2022 tarihleri arasındaki günlük çözüm verileridir. Ayrıca, yıl, yılın ayı, yılın haftası ve ayın günü, vb. gibi takvim değişkenleri giriş parametleri olarak eklenmiştir. KYI'daki mevsimsel trendleri yakalamak için veri seti geçmiş KYI verileri kullanılarak zenginleştirilmiştir.

Değer aralıklarındaki farklılıkları deforme etmeden veya bilgi kaybetmeden ortak bir ölçek kullanmak için veri seti, aykırı değerler ve hatalı veriler kaldırıldıktan sonra Min-Maks dönüştürme işlevi tarafından [0,1] arasında normalize edilmiştir. Veri seti değerlendirme için eğitim (%80) ve test (%20) setleri adında iki bölüme ayrılmıştır.

Ayrıca, Pearson korelasyonuna göre 0, 0,10, 0,25 ve 0,50 eşikler belirlenerek, dört farklı öz nitelik seçimi senaryosu oluşturulmuştur. Ardından, gün öncesi KYI tahmin modellerinin gerçekleştirilmesinde değişik derin öğrenme tabanlı modeller ile bir istatistiki ve bir geleneksel yapay sinir ağı algoritması olarak adlandırılan çoklu doğrusal regresyon (ÇDR), çok katmanlı algılayıcı (ÇKA), yinelemeli sinir ağı (YSA), uzun kısa süreli bellek (UKSB), kapılı tekrarlayan hücre (KTH), çift yönlü UKSB ve çift yönlü KTH kullanılmıştır.

Sonuç olarak, tespit katsayısı (R2), ortalama mutlak yüzde hatası (MAPE), normalleştirilmiş ortalama mutlak hata (nMAE) ve normalleştirilmiş ortalama karesi alınmış hata (nRMSE) sonuçları değerlendirmek için kullanılmıştır ve sonuçlar sonrasında gelecekteki perspektifler ele alınmıştır. İlk senaryodaki Bi-LSTM modelinin (R2 = %85.495, MAPE = %19.901) tarafından diğer tüm istatistiksel ve yapay zeka modellerinden daha iyi performans gösterdiği gözlemlenmiştir.

To my beloved family

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude to my supervisor, Prof. İlyas EKER, and my co-supervisor, Asst. Prof. Dr. Kasım ZOR, for their guidance and patience, as well as their insightful suggestions and remarks.

In addition, I would like to thank the jury members, Asst. Prof. Dr. Oğuzhan TİMUR and Asst. Prof. Dr. Necdet Sinan ÖZBEK, for providing insightful comments and suggestions and for evaluating my thesis.

I would also like to thank the Presidency for Turks Abroad and Related Communities (YTB) for their financial support, along with the Çukurova University Institute of Natural and Applied Sciences for their ongoing assistance.

Special thanks to m AAST college faculty members Dr. Mohamed NASSEF and Prof. Hamdy ASHOUR, and Dr. Moataz AMER for their unwavering support.

Furthermore, I would like to thank my best friend, with whom I have shared many iconic scenes and got encouragement.

Last but not least, I am eternally grateful to my family and my husband Mahmoud for their continuous support and unconditional love.

TABLE OF CONTENTS **PAGE**

LIST OF TABLES **PAGE**

LIST OF FIGURES **PAGE**

LIST OF SYMBOLS

B : Bias
CC : Cloud Cover
Ct : Long-Term State
D : Day of year
DoM : Day of month
DoW : Day of week
DT : Dry bulb Temperature
E : Evaporation
Ft : Forget Gate
H : Nodes in Hidden Layer
H : Hour
Ho : Extra-terrestrial Radiation
Ht : Hidden State
Ht : Short-Term State
ht-1 : Previous Hidden State from the Previous time-step
KT : Insolation Clearness
M : Month
MoY : Month of year
N : Number of observations
Ot : Output Gate
P : Pressure
RH : Relative Humidity
SD : Solar Duration
T : Air Temperature
Tavg : Average Air Temperature
Ti : Time
Tmax : Maximum Air Temperature

Tmin : Minimum Air Temperature

Tp : Panel Temperature

U : Input Weight Matrix

V : Output Weight

V : Visibility

WD : Wind Direction

WS : Wind Speed

X : Input Vector

Xi : Input Instance

xmax : Maximum Value of x

xmin : Minimum Value of x

Xnorm : Normalised Column Vector Converted From x

Xt : Input to the Hidden Layer

Y : Actual Output

Yi : Measured Output

ymax : Minimum Value of yi

ymin : Maximum Value of yi

A : Altitude

β0 : y-intercept

ϵ : Surface Emissivity

Λ : Longitude

Σ : Sigmoid Function

Φ : Latitude

Ω : Weight

Predicted Output

Mean of yi

LIST OF ABBREVIATIONS

Adj-SSE : Adjusted Sum of Squares Due to Error
AI : Artificial Intelligence
ALO : Ant Lion Optimizer
ALSTM : Adaptive Long Short-Term Memory
ANFIS : Adaptive Neuro Fuzzy Inference System
ANN : Artificial Neural Network
APB : Advanced Processor Build
ARIMA : Auto Regressive Integrated Moving Average
ARIMAX : Auto Regressive Integrated Moving Average with exogenous variables
ARMA : Auto Regressive Moving Average
Bi-GRU : Bidirectional Gated Recurrent Unit
Bi-LSTM : Bidirectional Long Short-Term Memory
BNN : Binarized Neural Network
BPNN : Back Propagation in Neural Network
CNN : Convolutional Neural Network
CNQR : Copula-based Nonlinear Quantile Regression
DBN : Deep Belief Network
DHI : Diffuse Horizontal Irradiance
DI : Decision Intelligence
DL : Deep Learning
DNI : Direct Normal Irradiance
DNN : Deep Neural Network
DP : Differential Privacy
ECMWF : European Centre for Medium- Range Weather Forecasts
EE : Electrical Energy
EF : Efficiency Factor
EI : Efficiency Index

EMAE : Envelope-weight Mean Average Error
ENN : Edited Nearest Neighbor
EP : Electrical Power
Evs : Enterprise Value
EWMA : European Wound Management Association
FA : Finite Automata
FCM : Fuzzy C-Means
FF : Firefly Algorithm
FFA : Feature Fusion Attention Network
FFNN : Feed-Forward Neural Network
FL : Federated Learning
FLC : Fast Learning Classifier
GA : Genetic Algorithm
GBDT : Gradient-Boosted Decision Trees
GFS : Global Forecast System
GHI : Global Horizontal Irradiation
GP : Gaussian Process Regression
GRNN : Generalized Regression Neural Network
GRU : Gated Recurrent Neural Network
GSR : Global Solar Radiation
GWO : Grey Wolf Optimization Algorithm
GWR : Geographically Weighted Regression
HarLin : Harmonic–Linear model
IDW : Inverse Distance Weighted technique.
IEA : International Energy Agency
IPSO : Improved Particle Swarm Optimization
KELM : kernels Extreme Learning Machines model
KNN : K-Nearest Neighbour algorithm
LRNN : Local Response Normalization Network

LSSVM : Least-Squares Support-Vector Machine
LSTM : Long Short-Term Memory
LSTNET : Long- and Short-term Time-series Network
MAE : Mean Absolute Error
MAPE : Mean Absolute Percentage Error
MARS : Multivariate Adaptive Regression Splines
MBE : Model-Based Enterprise
MENR : Ministry of Energy and Natural Resources
MGGP : Multi-Gene Genetic Programming
MISO : Multiple-Input-Single-Output
MLP : Multilayer Perceptron
MLR : Multiple Linear Regression
MPE : Mean Percentage Error
MSc : Master of Science
MSE : Mean Squared Error
N : Nebulosity
NARX : Nonlinear Autoregressive Exogenous model
NCEP : National Centres for Environmental Prediction
nMAE : Normalised Mean Absolute Error
nRMSE : Normalised Root Mean Squared Error
OLS : Ordinary Least Squares technique
PhD : Doctor of Philosophy
PR : Precipitation
PSO : Particle Swarm Optimization
PV : Photovoltaics
R : The correlation between the observed and predicted values
R2 : Coefficient of determination
RBF : Radial Basis Function model
RBPNN : Radial Basis Probabilistic Neural Network

RE	: Reinforcement learning
RES	: Global Renewable Energy
RF	: Random Forest algorithm
rMBE	: relative Mean Bias Error
RNN	: Recurrent Neural Network
RR	: Rainfall
rRMSE	: relative Root Mean Squared Error
RSE	: Relative Squared Error
SARIMA	: Seasonal Autoregressive Integrated Moving Average
SE	: Sum of Errors
SMDLPF	: Simplified Multidimensional Linear Prediction Filter model
SP	: Sun Position
SSE	: Sum of Squares Error
SVR	: Support Vector Regression
SW	: Sliding Window
t-stat	: Hypothesis Test Statistic
U95	: Uncertainty at 95%
WANN	: Wavelet Artificial Neural Network
WEO	: The World Economic Outlook
WNN	: Wavelet Neural Network
WOA	: The Whale Optimization Algorithm
WRBF	: Wavelet Radial Basis Function
WRF	: Weather Research and Forecast
XGBOOST	: Extreme Gradient Boost

1. INTRODUCTION

In this chapter the background, motivation, objectives, contributions, and thesis outline are emphasized respectively.

1.1 Background

Over the last two decades, solar energy resources have emerged as an important source of free electricity worldwide. According to recent studies, photovoltaic (PV) energy has grown at a compound annual growth rate of 40% over the last fifteen years Ghimire et al. (2018). According to the World Energy Outlook (WEO, 2022), as the world population grows over the next few decades, solar PV installed capacity will surpass wind before 2025, hydropower around 2030, and coal before 2040. (Conti et al. 2018). Similarly, the International Energy Agency (IEA) has developed an optimistic scenario in which renewable energy reaches 39% by 2050 (Birol 2017). The role of renewable energy in achieving long-term economic growth has recently piqued the interest of the energy sector. Türkiye is ranked fifth in Europe and 12th in the world regarding renewable energy installed capacity. Renewables accounted for 52% of Türkiye's installed power at the start of 2021(MENR, 2022), thanks to significant renewable energy potential. In 2021, Türkiye's photovoltaic power capacity reached 7.8 GW, up by 17 percent from the previous year (Statista, 2022).

Furthermore, solar energy extraction costs are relatively low due to emerging improved technologies, and large industrial-scale solar power plants provide low-cost electricity when compared to fossil fuel and nuclear systems. In the next 5 years, Solar is expected to account for approximately 48 percent of capacity additions in Türkiye between 2021 and 2026 (IEA, 2022). This would support Türkiye's Renewable Energy Target, which calls for large-scale renewable electricity generation up by 4800 GWh by 2026 and renewable energy sources (RES) to account for the rise of more than 60% from 2020 levels (IEA, 2022). This demonstrates that the global market share of solar energy is expected to increase further, and as a result, new and cost-effective technologies that assess long-term energy sustainability to promote clean energy production in all parts of the World (e.g. remote regions) are highly desirable.

Because of the decreasing trends in feed-in tariffs for solar PV power in many countries (including Türkiye), there has been an increased interest and need for versatile energy management schemes (EMS) for end-users to increase electricity generation and power transmission capacity from various regions, both locally and globally. to meet rising consumer energy demands, both remote and metropolitan Solar and conventional energy transmission and use can be monitored, controlled, and optimized using EMS. However, a prediction error on the power output from a PV system can have a negative impact on the system's economic profit. Given this, an accurate predictive tool for solar radiation can help to reduce the uncertainty of future power generation. Such tools can be used to investigate and assess the long-term viability of solar-powered energy installations in all regions, regardless of their location.

1.2. Motivation

Global horizontal irradiance is the amount of solar energy (integrated over time) attenuated by all constituents of the atmosphere and falling on a horizontal surface of the Earth, and the primary source of energy for life on our planet. Data on solar irradiation provides information on how much energy strikes the earth. As a result, records of observed solar irradiation are required in a variety of fields, including energy planning, climate monitoring, and agricultural management. Global horizontal irradiance data in a specific geographical area is critical for proper investment identification, system design, projection, planning, and system continuity.

GHI forecasting is generally classified by its horizon (Hong et al. 2020) as shown in figure 1.1, short-term GHI forecasting is critical in grid operation because it allows operators to dispatch/schedule energy production with traditional fossil fuels (Alanazi and Khodaei, 2016).

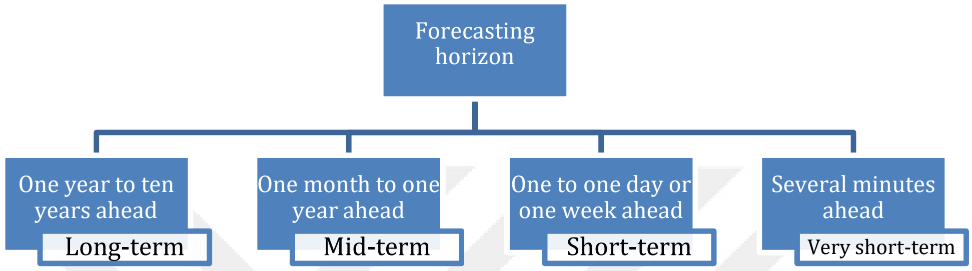

Figure 1.1. Illustration of different time horizon forecasting of GHI

In terms of investment, Increased generation from renewable resources, primarily solar and wind, exposes the power grid to more vulnerabilities, owing to variable generation, emphasizing the importance of accurate forecasting methods. As solar energy becomes more prevalent in the generation mix, grid operators will require greater visibility into how much solar power the system is producing. Then they can dispatch solar and other energy sources optimally to balance generation and demand. Improving solar forecasts will allow the electric grid to be more flexible and adapt to changing conditions, reducing disruptions and overall operating costs.

Furthermore, due to the cost and operation maintenance requirements, as well as the requirement to have installed the pyrometers with enough time in advance to have a record, it is not always possible to obtain global solar radiation data in this manner for academic research (Yildirim, 2016). For these reasons, solar energy planners are looking for alternative methods of obtaining short-term accurate solar radiation records.

Consequently, if we have a large enough amount of data and the data has a consistent pattern over a long period of time, we can make highly accurate predictions. However, making highly accurate predictions becomes difficult if the data format does not follow a consistent pattern and changes on a regular basis. The main issue with solar prediction is that power output is directly related to natural environmental factors during the relevant time period. Solar energy, for example, follows a predictable pattern throughout the year as solar power generation rises in the summer and falls in the winter. Simultaneously, if weather forecast data is also provided, solar energy generation can be predicted more accurately. Deep learning models, by definition, use the data structure learned from old data to make predictions accurately.

1.3. Objectives

The main objective of this thesis is to design accurate day ahead GHI prediction models for Adana using six different deep learning models (MLP, RNN, LSTM, BiLSTM, GRU, and BiGRU) and one statistical model (MLR) the results will be tested by performance metrics such as coefficient of determination (R^2), normalized mean absolute error (nMAE), mean absolute percentage error (MAPE), and normalized root mean square error (nRMSE).

The following are the objectives of this thesis:

- To summarize the literature and studies related to global horizontal irradiation prediction related to the thesis.
- To analyze and visualize MERRA-2 metrological data to provide a cleansing dataset to be used in the implemented prediction models.
- To design different feature selection scenarios depending on correlation between the input features and the target GHI.

- To analyze the effect of feature selection technique on the computation time and evaluation metrics.
- To implement a comprehensive study between seven different statistical and deep learning GHI day-ahead prediction models in Adana.
- To conclude the models' performance by testing the output by four evaluation metrics (R^2, nMAE, MAPE, nRMSE).
- To propose accurate day-ahead global horizontal irradiation models in Adana to be used as a database for daily generation power information.
- To provide day-ahead prediction to help investors in the investment sector in system design, projection, planning, and system reliability.

1.4. Contributions

The following are the thesis's main contributions:

– A comprehensive study of related literature until 2022 including journal publications and MSc theses concluding input features, output target, evaluated prediction models, and evaluation criteria.

– Six different deep learning algorithms are implemented including (MLP, RNN, LSTM, Bi-LSTM, GRU, and Bi-GRU) to design a day-ahead GHI model for Adana, Türkiye with output score up to (R^2= 85.495%, MAPE=19.901 %).

– A thorough examination of four different feature selection scenarios based on Pearson's correlation between the input features and the target GHI, demonstrating the significant impact of feature selection on both computational time and error reduction.

1.5. Thesis outline

The general methodology of the thesis is demonstrated in figure 1.1, the thesis consists of six Chapters as follows:

Chapter 1: This chapter provides an overview of the research's background, motivation, contributions, and the study's objectives as well as the thesis outline.

Chapter 2: A comprehensive summary of related literature, including journals, conference publications, and theses is presented.

Chapter 3: This Chapter describes the research topic, datasets, and general methodology used in this thesis, as well as the steps for dataset cleansing and visualization using python programming language.

Chapter 4: This Chapter presents the statistical method (MLR) and deep learning methods (MLP, RNN, LSTM, BiLSTM, GRU, and BiGRU) and their implementation and tuning steps, in addition to discussing the feature selection scenarios depending on their correlation with the target GHI.

Chapter 5: The results of the feature selection scenarios analysis, and the day-ahead GHI prediction models based on statistical and deep learning algorithms are discussed and concluded.

Chapter 6: This Chapter summarizes the results of this study, including conclusions, and recommendations.

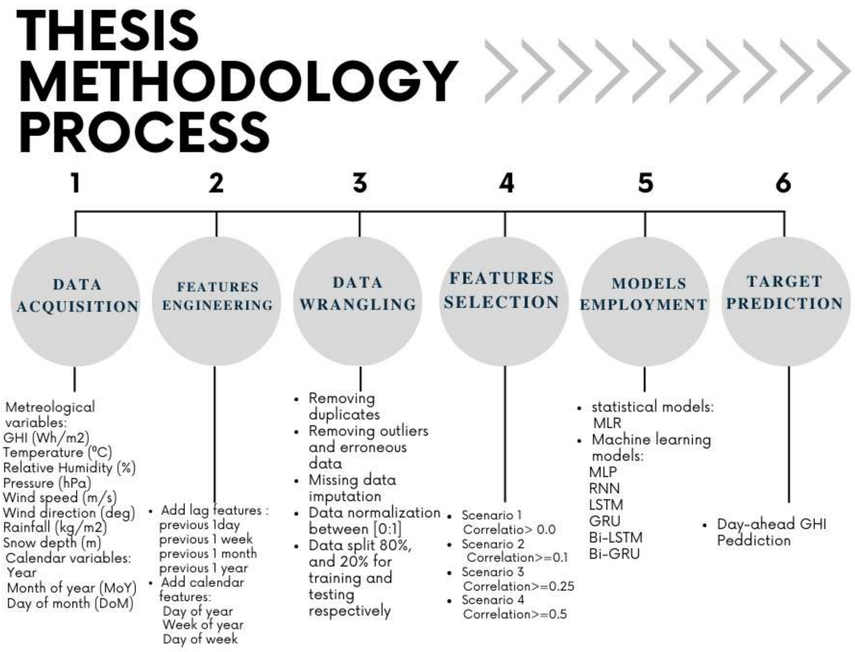

Figure 1.2. Demonstration of general thesis methodology

2. LITERATURE REVIEW

Solar prediction is essential for electric grid design; Very short-term prediction (from seconds to one hour) is beneficial for real-time electricity dispatch, optimal reserves, and smoothing solar and wind power production. Short-term prediction (from 1 hour to 24 hours) is useful for improving grid stability. Medium-term prediction (one week to one month) keeps power system planning and maintenance on track by predicting available electric power soon. Long-term prediction (from one month to one year) assists transmission and distribution authorities with electricity generation planning, as well as energy bidding and security operations (Akhter et al., 2019), and due to the lack of solar radiation measuring stations and issues of missing data, many studies are performed to estimate prediction model of solar power production. In this chapter past journals, conferences, and thesis publications as shown in figure 2.1 are reviewed to present a comprehensive point of view about the evaluated models related to the subject of the thesis.

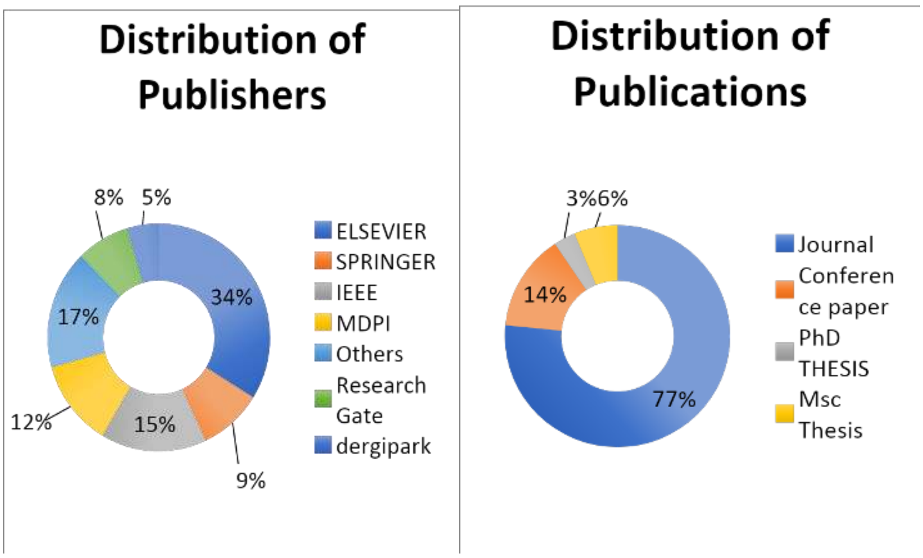

Figure 2.1. Distributions concerning publisher name and publication type

2.1 Related journal and conference articles in the literature

Barutcu B. et al. (2015) assessed an accurate short-term GHI prediction model by Advanced Research Weather Research and Forecasting (WRF-ARW) meteorological algorithm, and MISO-ANN as a post-processing tool. Güçlü et al. (2015) suggested a simple and effective approach for the estimation of solar irradiation values given the sunshine durations by considering the combined harmonic analyses and classical linear regression model (HarLin). Bakirci, K. (2015) achieved the best performance of diffuse solar radiation estimation by the third-order polynomial model based on sunshine duration and clearness index. Şenkal, O. (2015) estimated monthly daily average solar radiation based on the Resilient Propagation (RP) AI method using predicted precipitable water data by the same method based on meteorological and geographical data. Citakoglu, H. (2015) compared three artificial techniques (MLR, ANFIS, and ANN) and empirical models for solar radiation prediction in Türkiye using the calendar month number and pertinent meteorological data. Ozturk, M. (2015) developed 4 empirical models for estimating daily global solar radiation on a horizontal surface in Isparta, Türkiye, and compared them with other 46 existing models. Yaniktepe, and Genc (2015) presented a new third-order polynomial model to forecast global solar radiation on the horizontal surface which is more accurate than the models in the literature review. Akarslan et al. (2015) used the MDLPF approach that considered annual hourly solar radiations, extra-terrestrials, and their derivatives as multi-dimensional images and pixel values corresponding to the next hour's solar radiation. Teke et al. (2015) reviewed around 90 papers for global solar radiation estimation which were classified as linear, nonlinear, artificial intelligence, and fuzzy logic modelling techniques.

Çelik et al. (2016) presented an optimized Levenberg-Marquardt model to accurately estimate monthly global solar radiation in the Eastern Mediterranean Region of Türkiye based on different meteorological data combinations. Arslanoglu, N. (2016) confirmed the ability of the regression models to predict solar energy values precisely. Gairaa et al. (2016) combined linear autoregressive moving average

(ARMA) and nonlinear artificial neural network (ANN) models d to estimate the daily global solar radiation in Algeria. Arslanoglu, N. (2016) developed a more accurate and high-performance new regression model for monthly mean diffuse solar radiation on a horizontal surface. Kurtgoz, and Deniz (2016) developed four Different ANN models to estimate monthly mean daily GSR showing that extra-terrestrial solar radiation, sunshine hour, sine of declination angle and temperature parameters are the most relevant input parameters on the estimation of GSR, while wind speed and soil temperature input parameters are less influential and excluded.

Bakirci. (2017) derived three empirical models to estimate daily global solar radiation on a horizontal plane based on daily global solar radiation and the bright sunshine hours' values. Qudsia Memon, Nurettin Cetinkaya (2017) evaluated three different built-in learning algorithms MATLAB-ANN toolbox to predict the solar power production by using the real time data of a solar power plant. Emre Akarslan, Fatih Onur Hocaoglu. (2017) developed an adaptive and robust algorithm based on similarity assumption to estimate hourly solar irradiance from historical data.

Akarslan et al. (2018) proposed five different new models for hourly solar irradiance forecasting which are not site dependent as the Angstrom-Prescott models are. Bakirci et al. (2018) calculated monthly average daily diffuse solar radiation using six new feasible models based on the values the solar radiation obtained from NASA. Can Ekici & Ismail Teke (2018) present a model for Global solar radiation estimation from measurements of visibility and air temperature extremes parameters that simulates vapor and cloudiness. Bilal GUMUS and Hibetullah KILIC, (2018) evaluated a time series prediction model of monthly global solar radiation and sunshine duration based on exponentially weighted moving averages in south-eastern Türkiye. Gungor et al. (2018) compared 3 empirical sunshine-based models correlating the monthly mean daily global solar radiation on a horizontal surface with monthly mean sunshine records. Mohammed, H.N. and Mengüç, M.P. (2018) presented two empirical models to predict the monthly average daily of horizontal extra-terrestrial radiation that increased the efficiency 2% higher than those of the other approaches. Ayvazoğluyüksel, and Filik (2018) applied a comprehensive review of eleven different models to estimate hourly global solar radiation realizing that there is not a unique model that gives the most precise results in all locations as it depends on climatic conditions.

Karaveli, and Akinoglu (2018) carried out a comparison between the recent methodologies developed by the authors and some models taken from the related literature. Gumus, B., et al. (2018) proposed a time series based EWMA approach which is efficient in long and short-term GSR and SSD prediction Yıldırım et al. (2018) proposed two new models that give better results than 14 different models to predict daily global solar radiation based on different meteorological parameters in Eastern Mediterranean Region. Yıldırım et al.2018) created a MATLAB graphical user interface (GUI) program for estimating daily Global solar radiation in the Eastern Mediterranean region of Türkiye using ANN and Angström-Prescott models. (Jahani et al.2018) reviewed empirical and ANN methods for estimation of daily global solar radiation in Iran based on sunshine duration and the difference between the maximum and minimum temperature recorded by the Islamic Republic of Iran Meteorological Organization (IRIMO). (Khosravi et al.2018) Compared artificial intelligence methods in the estimation of daily global solar radiation in 10 cities in Iran and found that the GMDH model outperforms the other developed models. (Kaba et al.2018) Estimated daily global solar radiation using a deep learning model by applying 16 different combinations of inputs to find the optimum combination using the data belonging to 34 stations, which are almost uniformly distributed in Türkiye.

Marzouq et al. (2018) proposed a New daily global solar irradiation estimation model based on the automatic selection of input parameters using evolutionary artificial neural networks with different genetic operators including selection, crossover and mutation considering the coding of ANN as an individual in the pool of evolution. (Khosravi et al.2018) demonstrated that for the N1, SVR and MLFFNN models have the maximum performance to predict the hourly solar radiation in Abu Musa Island, Morocco. (Akbaba et al.2018) presented a Multilayer Perceptron (MLP) model to estimate daily global radiation in Türkiye that obtained lower errors than the universal models depending on five years of geographical and meteorological data obtained from Turkish State Meteorological Service (TSMS). (Aji et al.2018) verified Inverse Distance Weighting (IDW) model for Highly Accurate Daily Solar Radiation Forecasting using SW-SVR for Hybrid Power Plant at Yuan Chang Township Administration and Liujia Campus of Tainan in Indonesia. (Wang et al.2018) developed Comparative Study on KNN and SVM Based Weather Classification Models for Day Ahead Short-Term Solar PV Power Forecasting which proved SVM is prioritized for weather status pattern classification of DAST solar PV power forecasting because of its advantage in dealing with small sample scales for those newly built PV plants lacking in historical data. (Boussaada et al.2018) carried out Several simulations with different evaluation criteria are performed and evaluated using MSE and DMPE errors such as Nonlinear Autoregressive Exogenous (NARX) Neural Network Model for the Prediction of the Daily Direct Solar Radiation.

Lotfinejad et al. (2018) assessed a Comparative of Predicting Daily Solar Radiation Using Bat Neural Network (BNN), Generalized Regression Neural Network (GRNN), and Neuro-Fuzzy (NF) System. Lima et al.2018) developed Predictions using Multilayer Perceptron (MLP) Backpropagation Artificial Neural Network (ANN) with the advance of 1 hour for solar resource forecasting in equatorial areas, (Mohammad et al.2018) proposed Three structures of a multilayer artificial neural network for predicting the solar radiation of Baghdad City-Iraq. (Durrani et al.2018) proposed an irradiance forecast model based on multiple neural networks which have a MAPE of 18% to 21 % in summer and spring and an overall MAPE of 26% in photovoltaic yield prediction. Qing et al. (2018) enhanced hourly day-ahead solar irradiance prediction using weather forecasts by LSTM, for a case using 10 years of historical data to predict 1 year of irradiance data, the prediction RMSE using the proposed LSTM algorithm decreases by 42.9% against BPNN.

Yu et al. (2018) proposed ENN model is more predictive than the BPNN or RBFNN approaches in complicated weather cases where solar irradiance is random and highly volatile. (Sharma et al.2018) exhibit that from a few hours to two days ahead solar irradiance prediction is precisely estimated by machine learning-based models irrespective of seasonal variation in weather conditions. (Bouzgou et al.2018) provided slightly better performance for a considerably lower computing time by applying Wrapper Mutual Information for Fast Short-Term Global Solar Irradiance Forecasting. (Abdulhamid et al.2018) applied a fuzzy logic model to estimate daily global solar radiation over twelve Turkish cities based only on temperature and pressure data obtained from the Turkish State Meteorological Service (TSMS) from 2005 to 2010.

Incecik et al. (2019) investigated the forecasting skills of solar radiation by the WRF model with an optimum combination of parameterization schemes in terms of both partly cloudy and clear-sky conditions at five sites in the south-eastern Anatolia (SEA) region of Türkiye. Demirdelen et al. (2019) Generated Short-Term Power Forecasting Based on Artificial Neural Networks and Firefly Algorithm (FA) after processing input data by conventional ANN. Basaran et al. (2019) investigated boosting and bagging ensembles of 3 base artificial models to estimate hourly solar irradiance in five Turkish cities which improved the prediction accuracy. Ünlü, R. (2019) compared three artificial models (SVR, MLP, and LSTM) evaluating time series electricity prediction forecasting models. Kaygusuz, K. (2019) obtained solar power capacity for electricity generation in Türkiye based on the Geographically Weighted Regression (GWR) analysis method. Feng et al. (2019) proposed a hybrid MEA-ANN model which provided better estimations in temperate continental regions of China, compared with the existing machine learning and empirical models based only on temperature data. Zang H et al. (2019) obtained better estimation precision with empirical models of functional deep belief networks for estimating daily global solar radiation based on geographical and meteorological data of China.

Ghimire et al. (2019) concluded that an ANN approach integrated with ECMWF fields, incorporating physical interactions of Ira with atmospheric data, is an efficacious alternative to forecast solar energy and assist with energy modelling for solar-rich sites that have diverse climatic conditions to further support clean energy utilization. (Fan et al.2019) reviewed Comprehensively Empirical and machine learning models for predicting daily global solar radiation according to sunshine duration considering both the prediction accuracy and computational costs. (Ghimire et al.2019) also evaluated predictions of half-hourly, daily and monthly solar radiations whose input elements are defined by antecedent lagged GSR data. Essentially, for enhanced accuracy, the proposed model employs CNN to extract GSR data features and LSTM to encapsulate the features to generate a low latency-based time series GSR prediction.

Babu et al. (2019) established Wavelet Neural Network Model for Hourly Solar Radiation Forecasting from Daily Solar Radiation data of India. Yu et al. (2019) presented a comparison of deep learning methods (LSTM, ARIMA, SVR, BPNN, CNN, RNN) which approved LSTM Short-Term Solar Irradiance Forecasting Under Complicated Weather Conditions. (Sharadga et al.2019) enhanced that Artificial neural networks learn the complexity of time series data better than SARIM models. Thus, NNs are better than the suggested statistical models for PV power time series prediction. (Alsharif et al.2019) performed Time Series ARIMA Model for Prediction of Daily and Monthly Average Global Solar Radiation by Korea Meteorological Administration (KMA) GSR data in Seoul, South Korea. Husein et al. (2019) presented A Deep Learning Approach for Day-Ahead Solar Irradiance Forecasting for Microgrids Using Long Short-Term Memory Recurrent Neural Network and FFNN model in four different countries (Germany, USA, Switzerland, and South Korea). H. Zhou et al. (2019) evaluated the LSTM model which effectively forecasts PV power in all four seasons, and the forecasting curves of the LSTM model are very close to the actual curves in all different time intervals. (Rocha et al.2019) established the Estimation of daily, weekly, and monthly global solar radiation using ANNs and a long data set: a case study of Fortaleza, in the Brazilian Northeast region using 44-year-long data.

Çoban V. and Onar S.Ç. (2020) compared 11 machine learning techniques, evaluated by nRMSE and MAE to predict one-hour solar radiation in Istanbul, Türkiye. (Çiçek et al.2019) predicted global solar irradiance based on modelling and simulation analysis of an actual GRID-CONNECTED photovoltaic plant in Türkiye using clear-sky local meteorological station data located in Konya, Türkiye. Goksu et al. (2020) noticed that low radiation levels and the effect of temperature and humidity are noticeable on energy outputs in Artificial Neural Network based Solar Energy Potential prediction according to HELIOSAT processed data in Izmir, Türkiye. Gürel et al. (2020) preferred four different methods (feed-forward neural network, three Angstrom-type models, Holt-Winters, and RSM), and results showed that the ANN model gave the most accurate results although the dataset must be wide enough to train the network well. Özdemir et al. (2020) compared ANN and Bi-LSTM methods to predict cumulative installed solar power value showing the unexpected result that shows that the ANN method yields a better result than the Bi-LSTM.

Ceyhan Yildiz and Hakan Acikgoz. (2020) forecasted 2–4 hours ahead on-grid photovoltaic power generated by Kilis station, Türkiye based on Kernel Extreme Learning Machine (KELM) which is the improved version of the ELM. Kaygusuz, K. (2020) observed that the lowest RMSE and highest R^2 values have been looked for in Geographically Weighted Regression (GWR) when spatial interpolation of radiation data, aspect, latitude, relative humidity, and cloudiness was used as a secondary variable.

Colak et al. (2020) integrated grey wolf, ant lion, and whale optimization algorithms to the multilayer perceptron using meteorological data, also the effect of the activation functions was analyzed in detail. Ağbulut et al. (2020) proposed a new power coefficient equation for PV system performance which increases the accuracy of output power with the machine learning algorithms. Liu et al. (2020) compared the support vector machine with copula-based nonlinear quantile regression for estimating the daily diffuse solar radiation in four different regions in China which performed slightly better than CNQR models. (Marzouq et al.2020) evaluated a comparative study about short term solar irradiance forecasting via a novel evolutionary multi-model framework and performance assessment for sites with no solar irradiance data with three models (smart persistence, regression trees, and random forest), by proposed HAEANN models which showing better performances.

Mishra et al. (2020) intended a Deep Learning-Wavelet Transform integrated approach for Short-term solar PV power prediction which increased prediction precision and could meet the necessities of real-time developments. Tao et al. (2020) compared the proposed optimal extreme learning machine ELM model to MLR and ARIMA for Global solar radiation estimation and climatic variability analysis. Ojo et al. (2020) applied a nonlinear autoregressive neural network to the estimation of global solar radiation over Nigeria, these showed that the NARX model has higher values and lower RMSE values over all the climatic regions and entire Nigeria than the MLR mode. Aslam M et al. (2020) performed Comparative Study Deep Learning Models (FFNN, RNN, LSTM, GRU) for Long-Term Solar Radiation Forecasting Considering Microgrid Installation. (Brahma et al.2020) developed a comparison of the DNN models based on multiple horizons was also conducted. The results showed that forecasting tasks of shorter horizons show better accuracy while longer horizons require more complex models. (Wang et al.2020) forecasted A Day-ahead PV power based on the LSTM-RNN model and time correlation modification under a partial daily pattern prediction framework.

Ustun et al. (2020) performed empirical models for estimating the daily and monthly global solar radiation for the Mediterranean and Central Anatolia regions of Türkiye (Urgup, Karaman, Nigde, Isparta, Iskenderun, and Adana). Jallal et al. (2020) proposed a Deep Learning Algorithm for hourly Solar Radiation Time Series Forecasting based on local station recorded data in Morocco. Jallal et al. (2020) obtained results demonstrate the accuracy and the stability of the proposed data-driven model to perform prediction in case of GSR measurements intermittence or sensor

Yildiz et al. (2020) forecasted 2–4 hours on-grid photovoltaic power generated by Kilis station, Türkiye based on three machine learning models (LM, ELM, KELM) ALSAFADI and BAŞARAN FİLİK. (2020) proved that ML methods outperform the CPRG model in the estimation of hourly global solar radiation for a local station located in ESKİŞEHİR TECHNICAL UNIVERSITY, Türkiye.
(Ozbek et al. 2021) proposed a deep learning model (LSTM) and two different data-driven methods (ANFIS–GP, ANFIS-FCM) to forecast one-hour-ahead electrical energy production from the solar-PV power plant using historical solar power time series data. (Ağbulut et al.2021) evaluated different machine learning algorithms by RapidMiner Studio to predict the daily global solar radiation of four cities in Türkiye (Kırklareli, Tokat, Nevsehir, and Karaman) depending on 2 years of meteorological data extracted from the Turkish State Meteorological Service (TSMS). (Boubaker et al.2021) proved that the LSTM model is the most investigated Deep Neural Networks type compared to GRU, BiLSTM, BiGRU, CNN, CNN-LSTM, and CNN-BiLSTM models for Predicting Solar Radiation at Hail Region. (AL-Rousan et al.2021) A Comparative Assessment of Time Series Forecasting Using NARX and SARIMA to Predict Hourly, Daily, and Monthly Global Solar Radiation Based on Short-Term Dataset.

Bounoua et al. (2021) Estimated daily global solar radiation using empirical and machine-learning methods, among these models, Random Forest (RF) proved to be the best performing in all stations giving a high degree of confidence in Morocco. (Mellit et al.2021) investigated DLNNs models which provided very good accuracy, particularly in the case of 1 min time horizon with one-step ahead (correlation coefficient is close to 1), while for the case of multi-step ahead (up to 8 steps ahead) the results are found to be acceptable (correlation coefficient ranges between 96.9% and 98%). In the following tables: a comprehensive study of daily, monthly, and hourly time horizon journal and conference publication models were stated in tables

2.1, 2.2, and 2.3 respectively, then the literature of PhD and MSc theses are expressed in table 2.4, also the results of the previous employed models in Adana are compared in table 2.5.

2.2. Related MSc and PhD theses in the literature

HB. YILDIRIM (2016) in her PhD thesis developed an ANN model to estimate monthly average daily global solar radiation using different combinations showing that monthly mean sunshine duration, month and altitude is the most optimal combination, Angström-Prescott model based daily global solar radiation models is developed for the first time giving satisfactory results in case of single city model despite the lower performance of group cities models. IT Yumru (2016) proposed parallel locally-connected LSTM (PLC-LSTM) architecture which a joint between the bias of ConvLSTM and the variance of FC-LSTM architectures to estimate day-ahead solar power forecasting, and verified that even simple fully connected ANN can obtain quite close to more complex architectures when making good input combination.

Bitirgen, K. (2018) incorporated atmospheric and topographical factors to obtain accurate solar potential prediction by isotropic and isotropic models, the Badescu model on a tilted surface gave the lowest error which is used in the second part of the thesis. Kasht, H. (2018) classified the days according to daily clearness index value using Support Vector Machines (SVM) algorithm then it is used to improve the estimation of hourly horizontal global solar irradiation. Sorkun, M. C. (2018) investigated LSTM and GRU models with optimal hyper-parameter tuning to forecast 1 hour-ahead global solar radiation by univariate and multivariate models verifying that Temperature and Nebulosity significantly affect the performance of the model.

Akbaba, E. C. (2019) compared multilayer perceptron (MLP) with classical estimation models showing the advantages of simplicity and high accuracy of MLP in solar irradiance prediction, beside that the effect of hardware on the prediction

efficiency, in the terms of the features he insured that setting H/H0 as target instead of H leads to lower errors. Kaya, Ö. (2019) optimized artificial intelligence model by genetic algorithm to estimate hourly global radiation for five different locations in Türkiye, the results if ANN-GA gives better results in case of prediction in one location while feature ranking method is better when prediction of different places is applied. Vatansever, C. (2019) integrated Sliding windows methodology 15-minute ahead solar irradiance forecasting after SVM feature selection method is applied to enhance the accuracy resulting 0.06 MAPE value. (Ghimire.2019) proved that CLSTM model outperform the standalone CNN, LSTM, DNN, Recurrent Neural Network (RNN), Gated Recurrent Unit Neural Network (GRU), Multilayer Perceptron (MLP) and Decision Tree (DT) for global solar radiation prediction. (Widerøe.2019) indicated that the elastic net algorithm is capable of forecasting power production in a one-hour time frame even in early morning and late evening hours where the algorithm struggles the most.
SARKEZ, A. E. (2020) designed ANN approach based on feed-forward, back propagation (BP), and tangent sigmoid (tansig) activation function at input layer and (purelin) at output layer to enhance solar energy systems Tripoli, Libya. (Talakoobi.2020) concluded that forecasting becomes more challenging when prediction horizon increases, also by clustering dataset into sunny and cloudy days and using a separate prediction model for each set, prediction accuracy was improved for all the models.

Kavakçı, G. (2021) proposed new method to improve time series solar forecasting by extracting mean and trend from historical data and set them as input features to the machine learning algorithm, then trend decomposition was applied to obtain more stable data and higher performance results, moreover trend data also estimated based on linear estimation algorithm. QASEM, M. (2021) created a deep neural network using IBM SPSS software for daily solar radiation forecasting based on weather features as input that outperforms the linear regression by accuracy of 98% instead of 96.71%. (Petros.2021) discussed the horizon approach and the weather conditions of each forecasted hour were considered that the horizon forecasting can provide better and faster results, than the simple multi-hour forecasting. (Paiva.2021) proposed MGGP model which presented a relevant reduction in error and proved to be a reliable and accurate approach for the analyzed localities indicating that both MGGP and ANN are suitable to diverse locations.

3. MATERIAL

In this chapter the data acquisition, cleansing, and visualization using investigated software environment are discussed. Data processing operation flows through sequential steps as shown in figure 3.1.

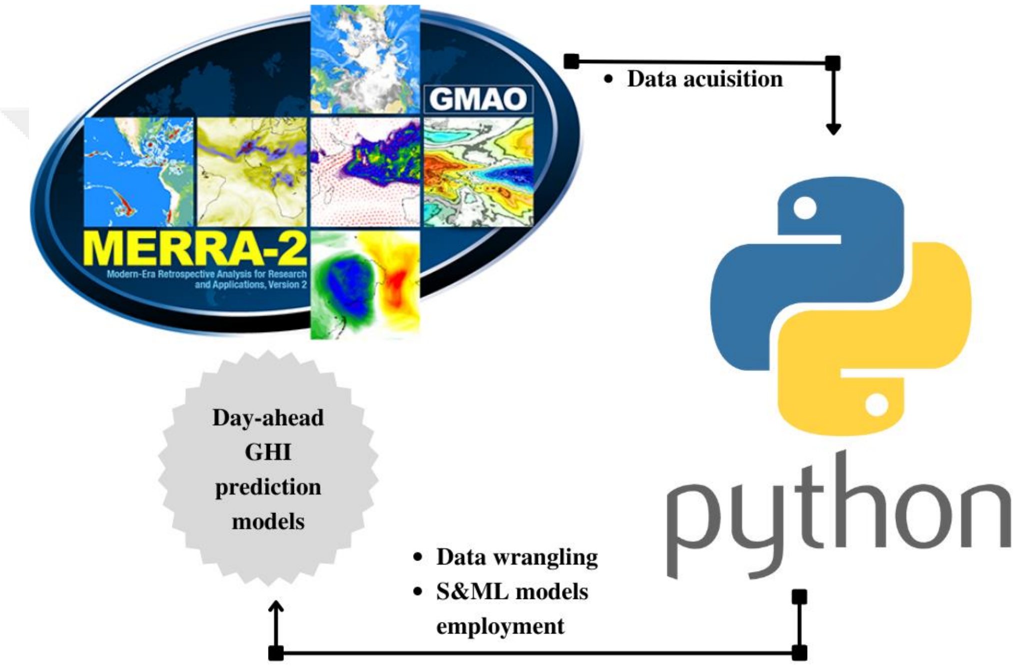

Figure 3.1. Illustration of data processing steps using python

3.1. Data description

The dataset used in this thesis is obtained from MERRA-2 (Modern-Era Retrospective analysis for Research and Applications, Version 2) for Adana city as shown in figure 3.2 with geographical coordinates of

- Latitude: 37.0◦N.
- Longitude: 35.20◦E.

- Altitude: 23 meters.

Figure 3.2. Geographical location of Adana, Türkiye.

Adana is a Turkish city located on the north-eastern coast of the Mediterranean Sea with sun duration hours up tov14 hr, 42 min per day (weather spark, 2022) as shown in figure 3.3

.

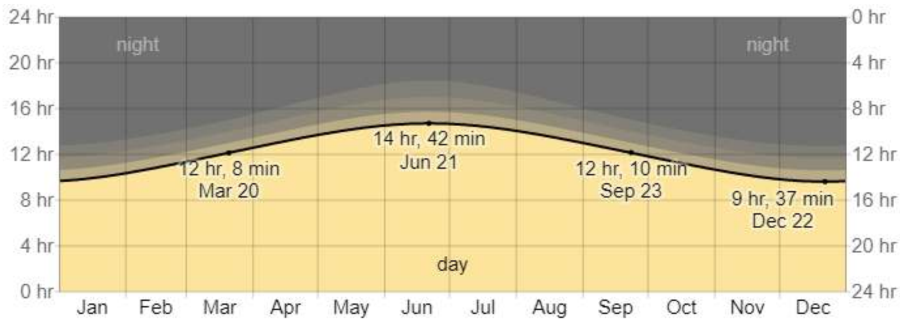

Figure 3.3. Hours of Daylight and Twilight in Adana

3.2. Data acquisition

MERRA-2 is the first long-term global reanalysis to incorporate space-based observations of aerosols and describe their interactions with other physical processes in the climate system that provides data starting in 1980 to the present (GMAO, 2015). The dataset covered daily meteorological data for period between January 2010 and September 2022. The meteorological data obtained from MERRA-2 consists of temperature (c) and relative humidity (%) at 2 meters above ground, pressure (kPa) at ground level, wind speed (m/s) and direction (◦) at 10 meters above ground, rainfall (kg/m^2), snowfall (kg/m^2), snow depth (m), and short-wave irradiation (Wh/m^2) for global horizontal irradiance (GHI). Moreover, Calendar parameters, such as the year, month of year (MoY), week of year (WoY), and day of month (DoM) are extracted and added as input features in addition to enriching the dataset by some lag variables of GHI to capture the double seasonal cycles in GHI in order the dataset as shown in table 3.1.

The meteorological data are studied to understand parameter behavior and correlations between them. The analysis results are shown in Table 3.2, also the graphs of the raw metrological dataset are illustrated in figures 3.4 to 3.12.

Table 3.1 Illustration of input features

Meterological variable	GHI (Wh/m^2)
	Temperature (ºC)
	Relative Humidity (%)
	Pressure (hPa)
	Wind speed (m/s)
	Wind direction (degree)
	Rainfall (kg/m^2)
	Snowfall (kg/m^2)
	Snow depth (m)
Calendar variables	Year
	Month of year (MoY)
	Week of year (WoY)
	Day of month (DoM)
	Day of week (DoW)
Lag variables	Previous 1 day GHI (Wh/m^2)
	Previous 1 week GHI (Wh/m^2)
	Previousv1 month GHI (Wh/m^2)

Table 3.2. Description of Metrological Variables

Metrological variable	Count	Mean	min	Max
Temperature (^0C)	4656.0	20.271490	-0.5270	37.043000
Relative Humidity (%)	4656.0	55.372762	16.5100	93.550000
Pressure (hPa)	4656.0	993.757678	972.2900	1012.120000
Wind speed (m/s)	4656.0	2.150292	0.0300	9.610000
Wind direction (degree)	4656.0	144.669942	0.0100	360.000000
Rainfall (kg/m2)	4656.0	1.457819	0.0000	56.069148
Snowfall (kg/m2)	4656.0	0.008659	0.0000	6.488402
Snow depth (m)	4656.0	0.000195	0.0000	0.024530
GHI (Wh/m^2)	4656.0	5302.725636	282.2607	9103.111000

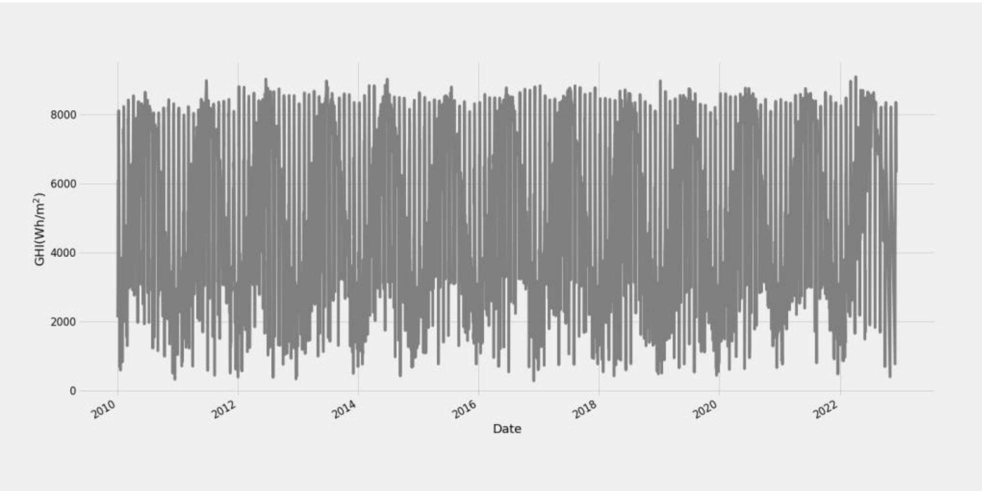

Figure 3.4. Graph of GHI (Wh/m^2) data

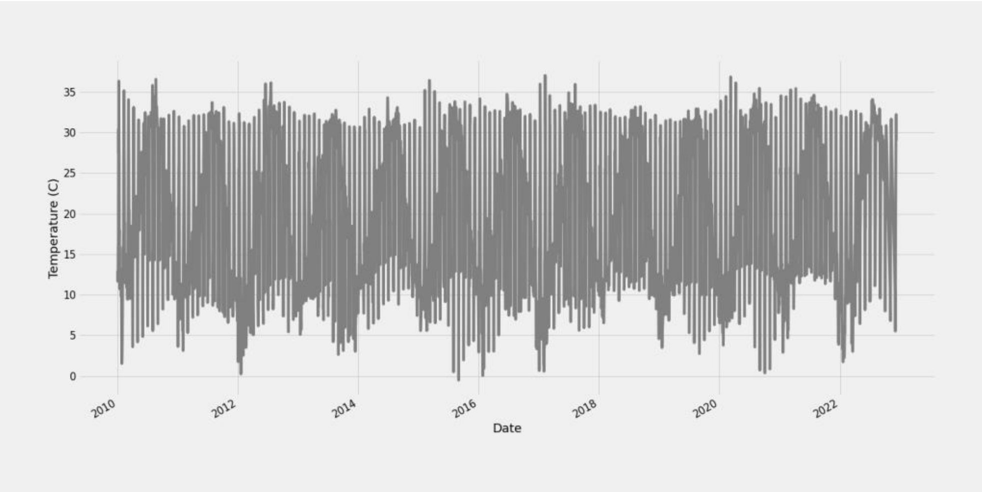

Figure 3.5. Graph of Temperature (ºC) data

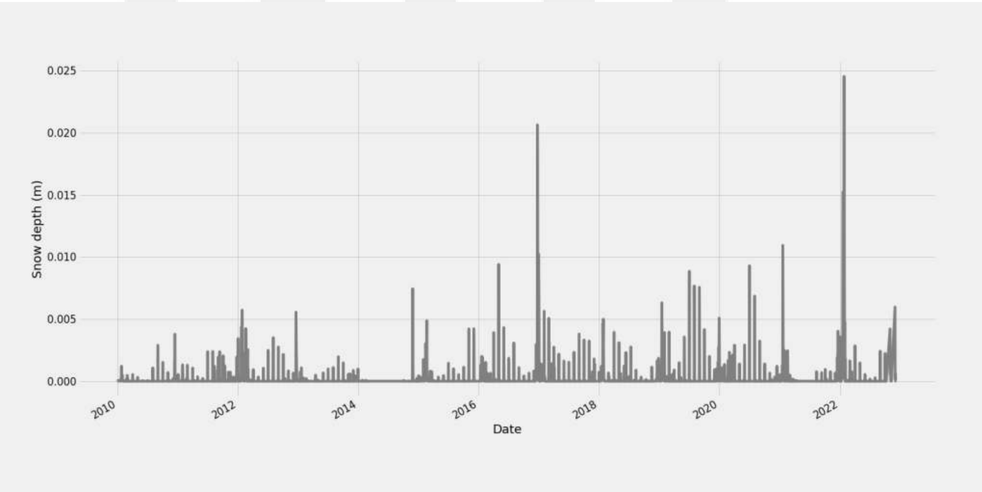

Figure 3.6. Graph of Snow depth (m) data

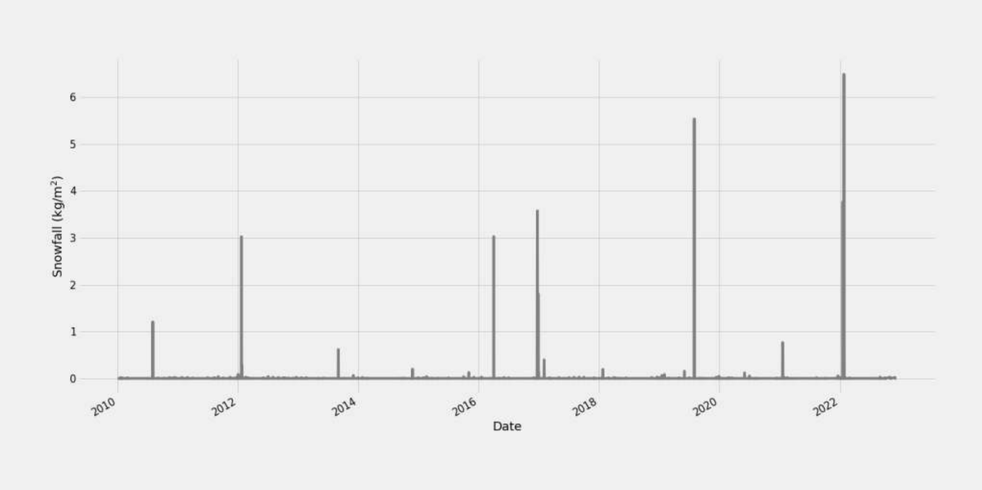

Figure 3.7. Graph of Snowfall (Kg/m2) data

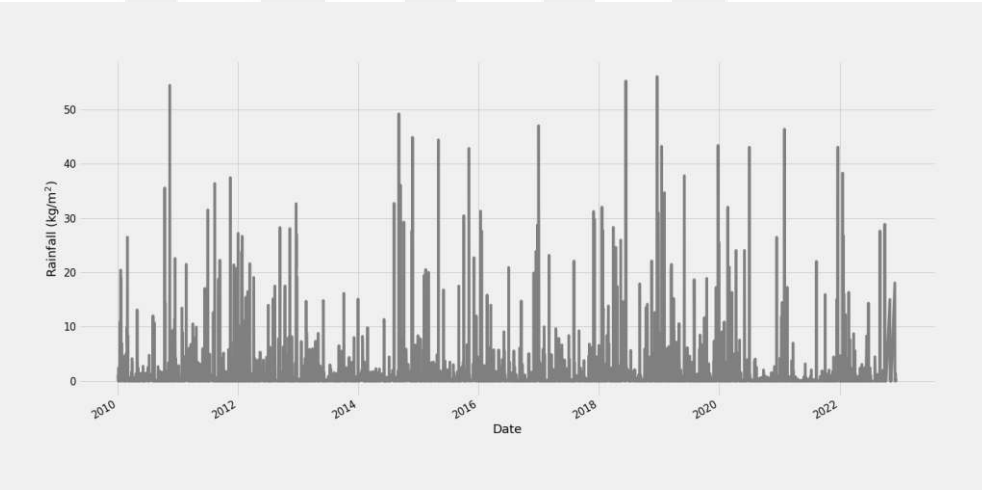

Figure 3.8. Graph of Rainfall (Kg/m2) data

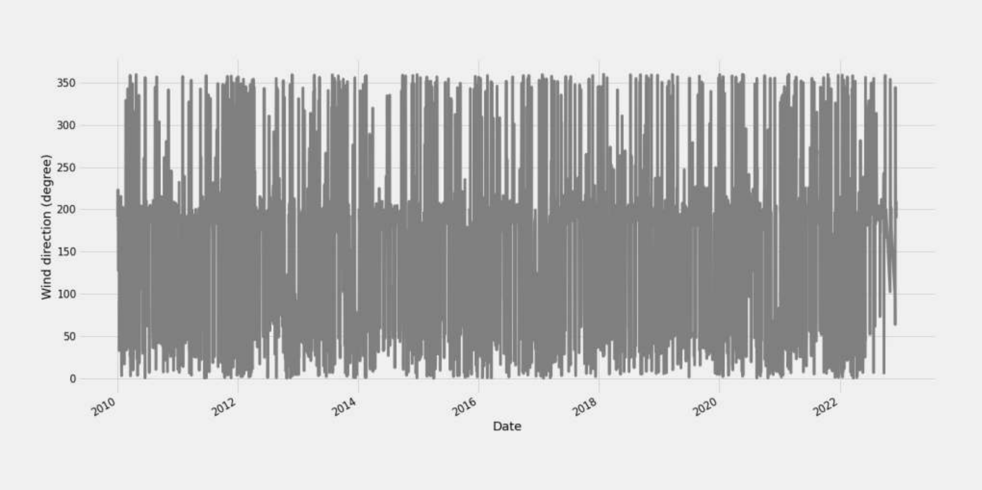

Figure 3.9. Graph of Wind direction (degree) data

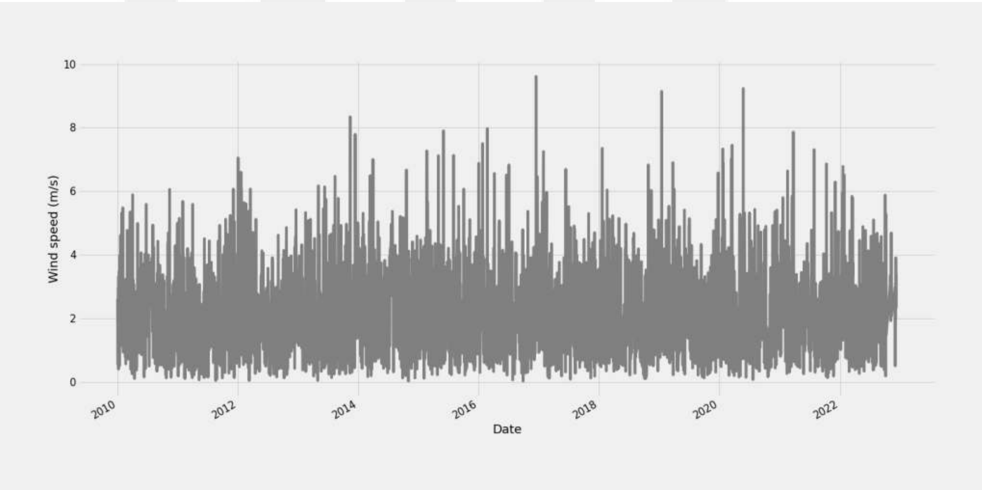

Figure 3.10. Graph of Wind speed (m/s) data

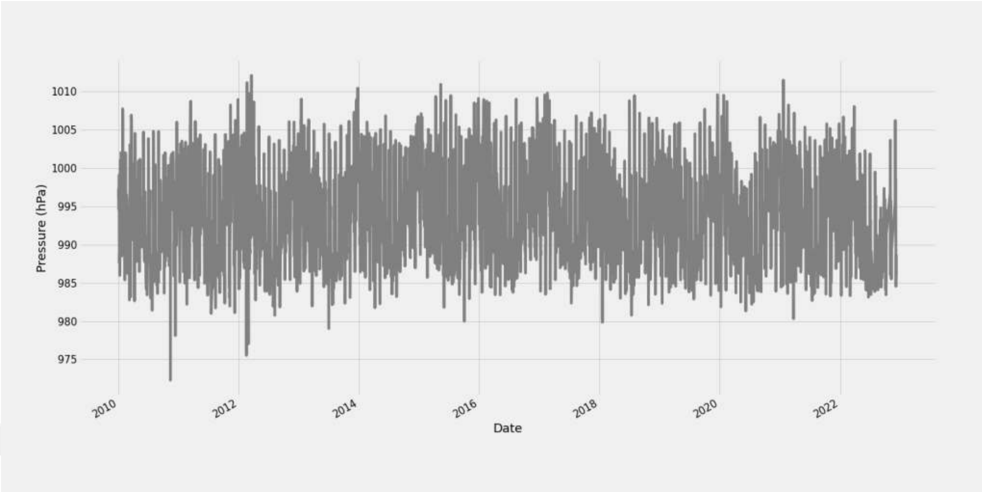

Figure 3.11. Graph of Pressure (hPa) data

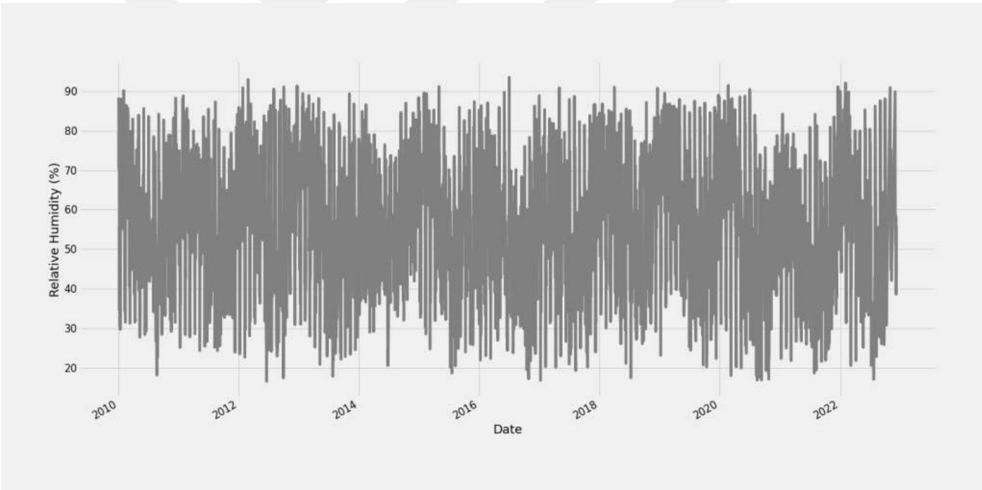

Figure 3.12. Graph of Relative humidity (%) data

3.3. Data wrangling

Eliminating inaccurate data from a set is a procedure known as data cleansing. Data that is wrong, duplicated, incomplete, corrupt, or incorrectly structured are just a few examples of the different kinds of data that can be regarded as invalid. Data cleansing is seen to be crucial, to guarantee data integrity and produce accurate and dependable results. The procedures to properly prepare data are shown below.

- Removing duplicates: duplicate entries can ruin the split between train, validation, and test sets. This can result in biased performance estimates, resulting in unsatisfactory models through output.
- Removing outliers: A dataset may contain extreme values that are outside of the expected range and differ from the other data. These are known as outliers. Understanding and even removing these outlier values can often improve machine learning modelling and model skills in general.

According to Tukey's rule; values more than 1.5 times the interquartile range from the quartiles — either below Q1 − 1.5IQR, or above Q3 + 1.5IQR — are considered

outlier values as shown in figure 3.13.

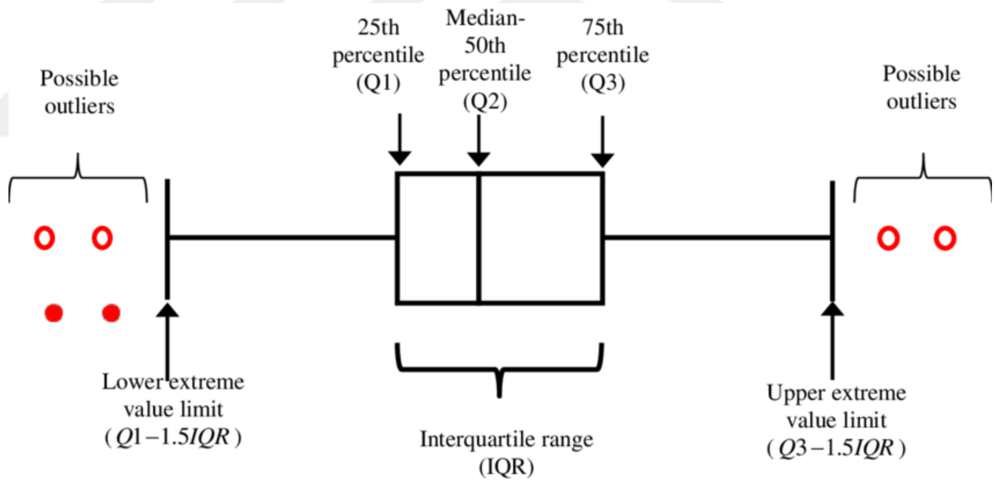

Figure 3.13. Demonstration of outlier detection rule

- Handling missing values: Missing data causes a variety of issues. First, the lack of data reduces statistical power, which is the likelihood that the test will reject the null hypothesis when it is false. Second, lost data can cause bias in parameter estimation. Third, it has the potential to reduce sample representativeness. In the cleansing phase, missing data are handled by a linear interpolation algorithm.

- Data normalization: Normalization is the process of converting the columns in a dataset to the same scale. It is essential when the ranges of characteristics differ. When you use un-normalized inputs to activation functions, you can get stuck in a very flat region of the domain when the ranges of characteristics are different and may not learn at all. Dataset is normalized to the value between 0 and 1 by applying the following formula, Where y_{max}= 1, and y_{min}= 0.

$$x_{norm} = (y_{max} - y_{min}) * (\frac{x - x_{min}}{x_{max} - x_{min}}) + y_{min} \quad (3.1)$$

- Data split: data splitting function is applied to create training and testing sets with 80% and 20% of the total data set, respectively.
- Data de-normalization: normalized data must be de-normalized before evaluation to bring each variable its real value to calculate the error at the same unit.

3.4. Software environment utilization in the thesis

Python is a general-purpose, high-level, interpreted programming language. Its design philosophy prioritizes code readability by employing significant indentation. Python's advantages for machine learning and AI-based projects include its simplicity and consistency, access to great libraries and frameworks for AI and deep learning, flexibility, platform independence, and a large community. These factors contribute to the language's overall popularity.

All computations in this thesis were carried out on a personal laptop running Windows 7 Ultimate, with a 2.1GHz (Intel(R)CoreTM) processor and a memory size of 2 GB, using the Kaggle platform (kaggle, 2010) and Python scripts. Kaggle is a data scientist and machine learning enthusiast community platform. Users can collaborate with other users, find and publish datasets, use GPU integrated notebooks, and compete with other data scientists to solve data science challenges on Kaggle. Kaggle Kernels are essentially browser-based Jupyter notebooks. These kernels are completely free to use. Are available for a variety of programming languages such as Python, R, and others.

4. METHODS

In this chapter the statistical and deep learning models structure, implementation and tuning are discussed, moreover feature selection strategy is explained.

4.1. Statistical models

4.1.1. Multiple linear regression (MLR)

Multiple linear regression (MLR), also known as multiple regression, is a statistical technique that predicts the outcome of a response variable using several explanatory variables proposed by (Pearson and Filon, 1898). Multiple linear regression attempts to model the linear relationship between explanatory (independent) and response (dependent) variables. Multiple regression is essentially an extension of ordinary least-squares (OLS) regression in that it involves more than one explanatory variable. The formula of MLR is expressed as follows

$$y_i = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \ldots\ldots + \beta_n * x_n + \epsilon \quad (4.1)$$

where i=n observations are made y_i= the dependent variable. x_i denotes explanatory variables. β_0= y-intercept (constant term).

4.2. Machine learning mode

In the following section the machine learning models including MLP, RNN, LSTM, GRU, Bi-LSTM, and Bi-GRU are explained in addition to their hyper-parameters tuning process as shown in figure 4.1.

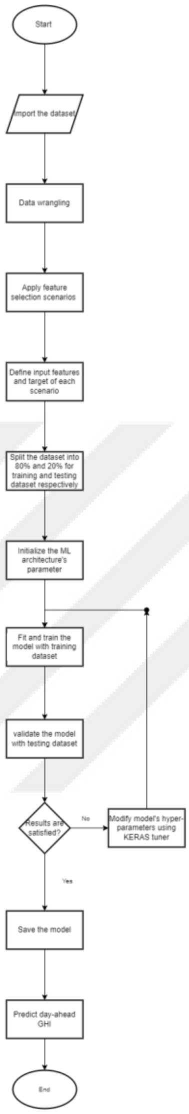

Figure 4.1.Flow chart of machine learning models designing steps

4.2.1. Multilayer perceptron (MLP)

Deep Learning (DL) may be a machine learning method using multi-layer deep ANN architectures. ANN consists of three layers, the input layer, the hidden layer, and also the output layer. DL is the ANN model with multiple hidden layers. the primary study involving the DL structure was applied by (Ivakhnenko and Lapa, 1965). whether or not the DL concept was supported in the 1960s, the increase passed off in recent years. this is often because there's no processor power to coach the deep architects, and there's no sufficient amount of knowledge. Nowadays the rise of processor power and the reaching of enormous dimensions of knowledge generated by digitization in many areas has profiled the required infrastructure for the DL. These developments have enabled DL to become widely employed in the sector, like computer vision, text processing, translation, and statistic prediction.

Multilayer perceptron is occasionally colloquially stated as "vanilla" neural networks, mainly once they have a single hidden layer. MLP is a feedforward artificial neural network that generates a set of outputs from a set of inputs proposed by (Rosenblatt, 1958). An MLP is defined by several layers of input nodes that are linked together as a directed graph between the input and output layers. Backpropagation is used by MLP to train the network.

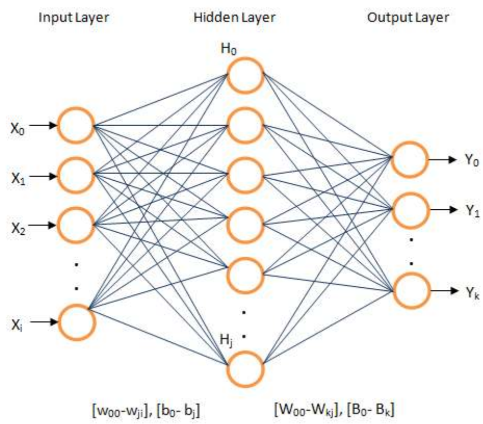

Figure 4.2. Structure of multilayer perceptron unit

An MLP as shown in figure 4.2 includes at least 3 layers of nodes: an input layer, a hidden layer, and an output layer. Except for the input nodes, every node is a neuron that makes use of a nonlinear activation function, the mechanism of MLP is described as the following equations:

$$Y_K = \sigma(\sum_0^k \sum_0^j (W_{KJ} * H_J) + B_K)) \qquad (4.2)$$

$$H_K = \sigma(\sum_0^i \sum_0^j (W_{JI} * X_I) + b_I)) \qquad (4.3)$$

Where X_0-X_i are inputs, H_0-H_j nodes in hidden layer, w_{ji} is weight between j^{th} hidden node and i^{th} input, b_j is bias for H_j, and σ is the activation function that could be sigmoid, tanh, or ReLU. The essential reason for choosing MLP in this thesis as it can apply complex non-linear problems and provide quick predictions after training.

4.2.2. Recurrent Neural Networks (RNN)

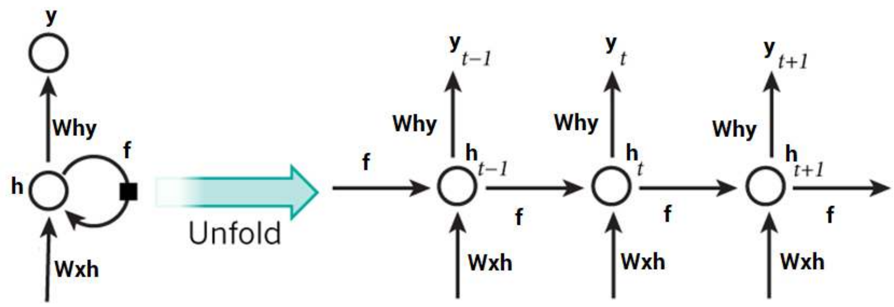

Figure 4.3. Structure of RNN unit

RNN is a feedback extension of a conventional MLP that can use the last time step output as input at each node proposed by Rumelhart et al. (1986) as shown in figure 4.3 RNN is useful for sequential data because it handles variable-length sequences by utilising a recurrent hidden state whose activation at each time-step is dependent on the activation at the previous time-step.

An RNN is made up of one hidden layer that keeps hidden states and one fully-connected layer that gets output from the hidden states. Figure 8 depicts the hidden unit of an RNN unit. The following is the definition of the hidden state ht:

$$h_0=0 \tag{4.4}$$

$$h_t = \tanh(U.x_t + W.h_{t-1}) \tag{4.5}$$

where x_t is the input to the hidden layer, h_{t-1} is the previous hidden state from the previous time-step, and U is the input weight matrix.

After obtaining the hidden state, RNN output o_t is calculated directly from the current hidden state in a fully-connected layer as follows:

$$o_t = \sigma(V.h_t) \tag{4.6}$$

where the sigmoid activation function and V is the output weight.

Unfortunately, it was discovered that it is difficult to train RNNs to capture long-term dependencies because gradients tend to either vanish or explode most of the time. To address these issues, the LSTM unit was introduced, and more recently, gated recurrent units (GRUs) followed LSTM. These two RNN extensions were shown to solve gradient problems by capturing long-term dependencies.

4.2.3. Long Short-Term Memory (LSTM)

LSTM is a recurrent neural network that improves on the efficiency of traditional sequence learning mechanisms that presented by (Hochreiter, and Schmidhuber, 1997). The problem of vanishing and exploding gradients persists in RNNs, prompting the development of LSTM to address these issues.

As illustrated in Figure 4.4, LSTM adds three new computation components to the RNN: the input gate, the forget gate, and the output gate. The forward pass equations are as follows:

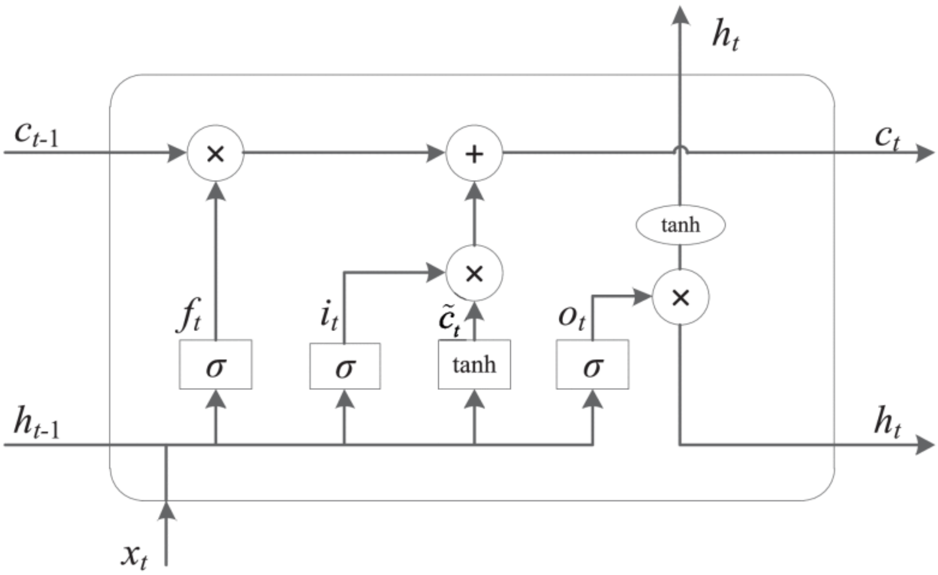

Figure 4.4 Structure of LSTM unit

Input gate: $i_t = \sigma(\omega_i \times [h_{t-1}, x_t] + b_i)$ (4.7)

Forget gate: $f_t = \sigma\left(\omega_f \times [h_{t-1}, x_t] + b_f\right)$ (4.8)

Output gate: $o_t = \sigma(\omega_o \times [h_{t-1}, x_t] + b_o)$ (4.9)

Intermediate Cell State: $\tilde{c}_t = \tanh(\omega_c \times [h_{t-1}, x_t] + b_c)$ (4.10)

long-term memory: $c_t = f_t \times c_{t-1} + i_t \times \tilde{c}_t$ (4.11)

Short-term memory: $h_t = o_t \times \tanh(c_t)$ (4.12)

Where: w_f, w_i, w_c and w_o are weight matrices. b_f, b_i, b_c and b_o are bias vectors. $\tilde{c}_t$ is a new candidate state, and σ () is the Sigmoid activation function that generates values between 0 and 1 to decide which information is important and which can be discarded.

It processes the current input and the previous state before the input gate. It determines which parts of A_t will be added in the long-term state C_t. F_t is the forget gate in charge of deciding which parts of C_{t-1} should be erased and discarding the unnecessary parts. The output gate O_t locates the parts of C_t to be read and displays them as output. As a result, the cells share a short-term state H_t, and the memories are dropped and added by the respective gates in a long-term state C_t.

4.2.4. Bidirectional LSTM (Bi-LSTM)

Bi-LSTM is a modified version of LSTM proposed by (Schuster, and Paliwal, 1997). It has two separate hidden layers; first, it computes the forward hidden sequence, then the backward hidden sequence, and finally, it combines both to compute the output as shown in figure 4.5.

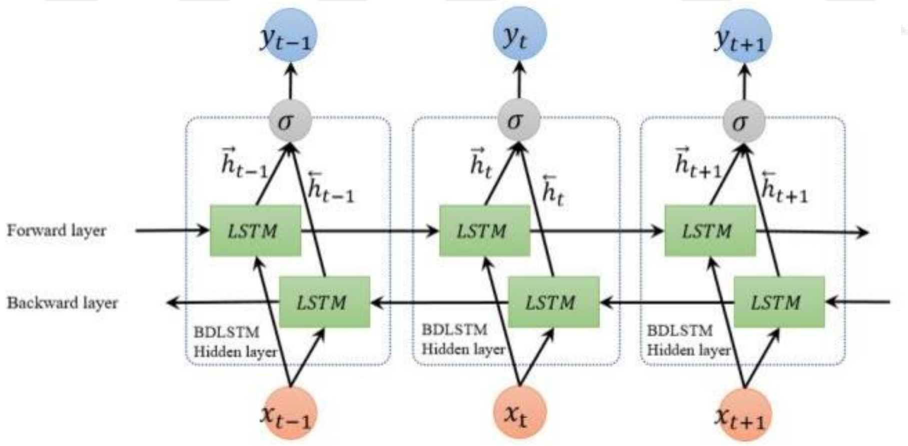

Figure 4.5 Structure of Bi-LSTM unit

The inputs are processed by LSTM in strict temporal order. This indicates that the current input is contextualized by previous inputs but not by future inputs. To address this shortcoming, the bidirectional LSTM model was developed. It duplicates the LSTM processing chain so that the inputs are processed both forward and backward in time. This enables the network to consider the future context.

4.2.5. Gated Recurrent unit (GRU)

The gated recurrent unit can be thought of as a simplified version of LSTM proposed by (Cho et al. 2014) in which cell states are not explicitly used. The main simplification is the merging of both state vectors into a single vector. Figure 4.6 describes a GRU cell, with the equations used during the forward pass shown below.

Update gate: $$z_t = \sigma(\omega_{xz}.[h_{t-1} + x_t]) \tag{4.13}$$

Reset gate: $$r_t = \sigma(\omega_{xr}.[h_{t-1} + x_t]) \tag{4.14}$$

Cell state: $$c_t = f_t, c_{t-1} + i_t, \tilde{c}_t \tag{4.15}$$

New state: $$h_t = (1 - z_t) \otimes \tanh(\omega_{xc}.h_{t-1}(z_t \otimes g_t)) \tag{4.16}$$

Output gate: $$o_t = \sigma(\omega_o \times [h_{t-1}, x_t]) \tag{4.17}$$

Where: w_{xz}, w_{hz}, w_{xr}, w_{hr}, w_{hc}, and w_{xc} are weight matrices. and σ () is the Sigmoid activation function that generates values between 0 and 1 to decide which information is important and which can be discarded.

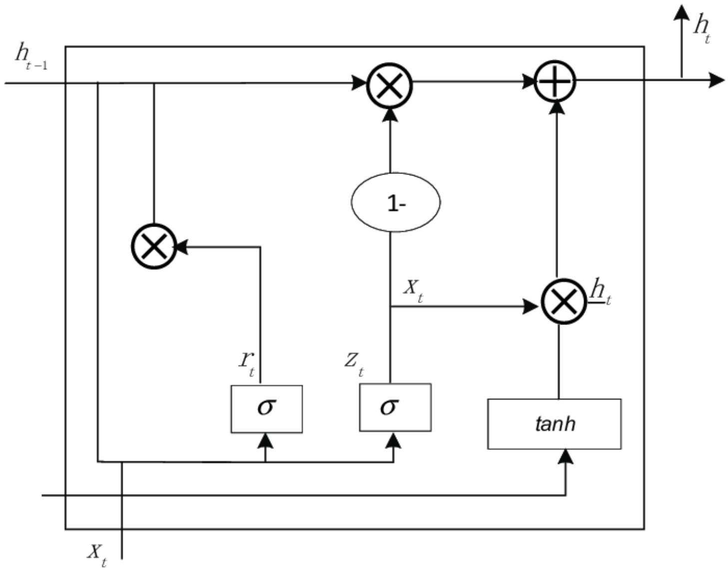

Figure 4.6 Structure of GRU unit

A single gate controller controls both the input gate and the forget gate. When the output of a gate controller is 1, the input gate is opened while the forget gate is closed, and vice versa when the output is 0. This means that when a memory is required to be stored, the location to be stored is erased first. It is, in fact, a commonly used variant of LSTM. There is no output gate, and the entire state vector is output every time step. However, a new gate controller controls the portion of the previous state that will be displayed on the main layer.

4.2.6. Bidirectional GRU (Bi-GRU)

Bi-GRU is a modified version of GRU with the same structure as Bi-LSTM (Schuster, and Paliwal, 1997), but with fewer parameters as shown in figure 3.10. It is made up of two distinct hidden layers. The procedure is very similar to Bi-LSTM.

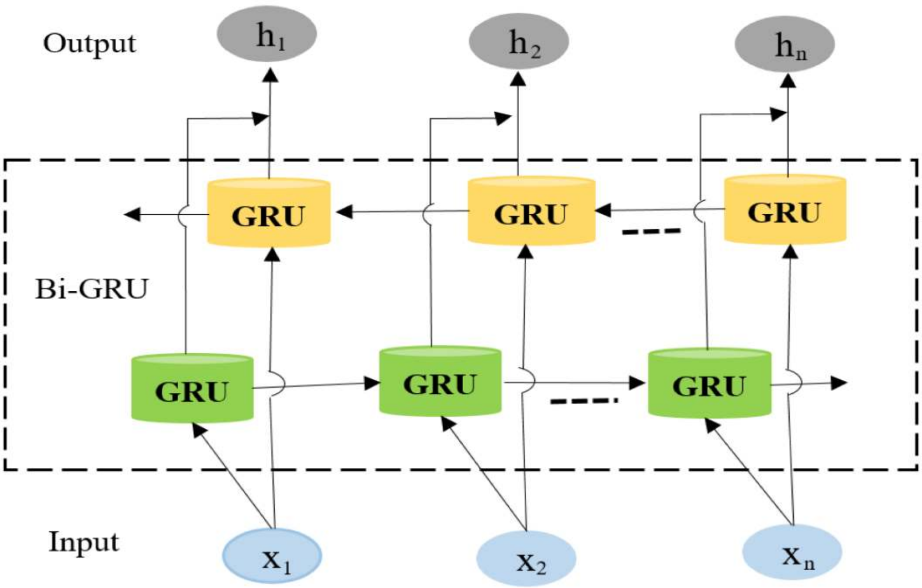

Figure 4.7 Structure of Bi-GRU unit

4.3. Feature selection criteria

Even if ML-based models support feature extraction, it may be beneficial to remove or process irrelevant features before training the model to:

- Reduce memory and time consumption because DL requires a large dataset.
- Discover the most and least important characteristics.

In order to design feature selection scenarios depending on their Pearson's correlation matrix stated in figure 4.8; four different scenarios based on the correlation threshold between the input features and GHI.

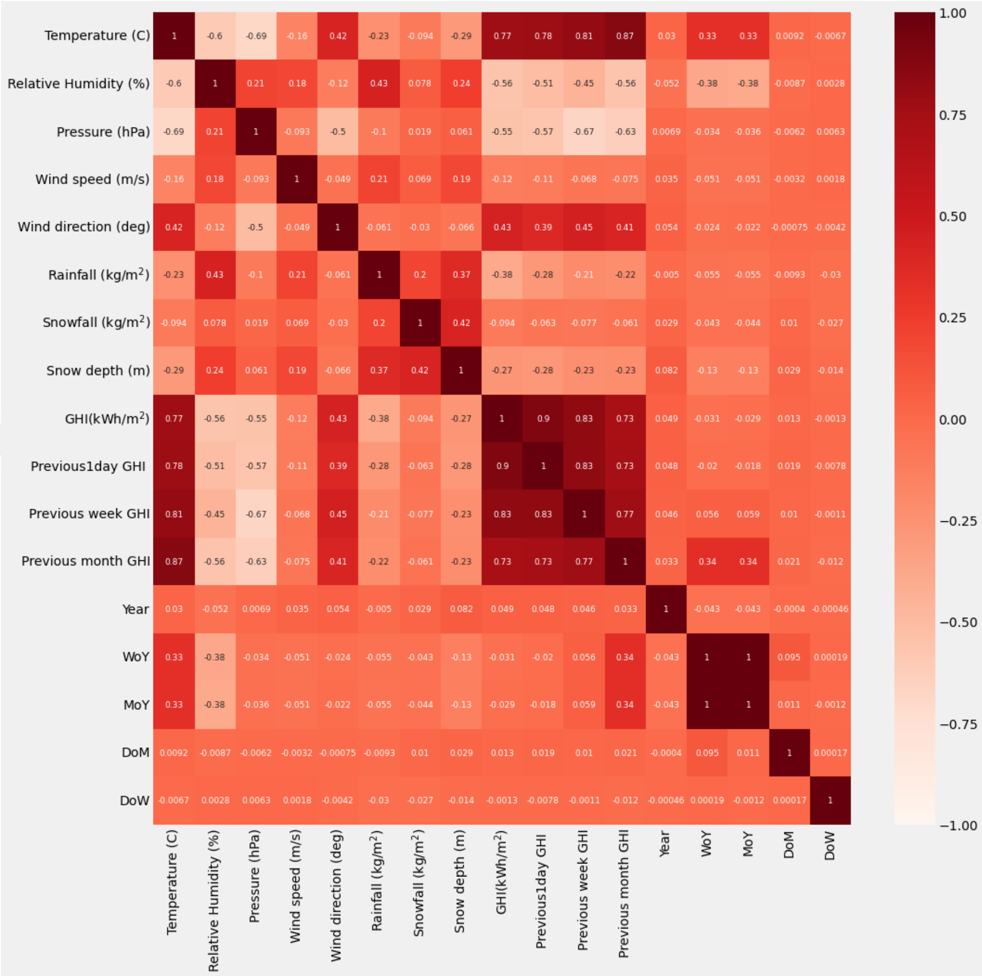

Figure 4.8. Pearson's correlation matrix of input features

The correlation criteria in this thesis are built on using all data without filtering in the first scenario, and then according to the correlation threshold setting the dataset is filtered in each scenario, table 4.1 extracts and summarizes the correlation values between each features and GHI in descending order, then the four scenarios are stated below.

Table 4.1 Feature's correlation with GHI sorted in descending order

Feature	Correlation with GHI	Feature	Correlation with GHI
GHI (Wh/m^2) **Target variable**	1.000000	Relative Humidity (%)	-0.561878
Previous 1 day GHI (Wh/m^2)	0.897130	Pressure (hPa)	-0.547156
Previous 1 week GHI (Wh/m^2)	0.830737	Rainfall (kg/m2)	-0.384175
Temperature (^{o}C)	0.768524	Snow depth (m)	-0.268701
Previous 1 month GHI (Wh/m^2)	0.733267	Wind speed (m/s)	-0.124803
Wind direction (degree)	0.433324	Snowfall (kg/m2)	-0.094202
Year	0.049019	Week of year (WoY)	-0.031458
Day of month (DoM)	0.013238	Month of Year (MoY)	-0.029120
		Day of week (DoW)	-0.001260

- Scenario 1: selects all the features without any filtering corresponding to correlation |threshold| > 0, including [4656,18] matrix dataset.
- Scenario 2: selects only features with correlation |threshold| >= 0.1, including [4656,12] matrix dataset and dropping the following features from scenario 1:
    - Snowfall (kg/m^2)
    - Year
    - WoY
    - MoY
    - DoM
    - DoW
- Scenario 3: selects only features with correlation |threshold| >= 0.25, including [4656,11] matrix dataset and dropping the following columns from scenario 2:

  - Wind speed (m/s)
- Scenario 4: selects only features with correlation |threshold|>= 0.5, including [4656,8] matrix dataset and dropping the following features from scenario 3:
  - Wind direction (degree)
  - Rainfall (kg/m^2)
  - Snow depth (m)

4.4. Deep learning hyper-parameter tuning

Hyperparameters are variables that govern the training process and the topology of a DL model, such as the number and width of hidden layers. Tuning hyperparameters is a critical component of controlling the behavior of a machine learning model. If the model's hyper-parameters are not properly tuned, its estimated model parameters produce suboptimal results since it doesn't minimize the loss function.

The Keras Tuner is a library that assists you in selecting the best set of hyperparameters for your TensorFlow program (Keras, 2015). Hyperparameter tuning, also known as hyper tuning, is the process of selecting the best set of hyperparameters for your machine learning (ML) application.

To optimize the deep learning model Keras tuner library was investigated to tune the hyper-parameters by using hp.int (), hp.float(), and hp.choice() functions for integer, float, and string parameters ranges respectively to highlight the optimal value of each parameter from the suggested range as stated in table 4.1 in order to design the DL model. İn table the range of each parameter

Table 4.2. Description of deep learning models' hyper-parameters range

Parameter	**Range**
Number of hidden layers	1:4
Number of neurons in the hidden layers	2:200
Weight initialization	Truncated normal, glorot normal, glorot uniform, he normal, he uniform.
Activation function	relu, sigmoid, softmax, tanh.
Optimizer	SGD, RMSprop, Adam, Adadelta, Adagrad, Adamax, Nadam, Ftrl.
Learning rate	0.0001: 0.01
Number of batch size	8:512
Number of epochs	1:500
Loss	Mean squared error, Mean absolute error, Mean absolute percentage error, Mean squared logarithmic error,
Dropout rate	0.0:0.5

Four different datasets were formed as a result of designing four different feature selection scenarios based on the correlation threshold, so the hyper parameter tuning process was used for each scenario. The tuned hyper-parameters for each model across the four scenarios are listed in the tables below including the total elapsed time during tuning each model.

Table 4.3. Tuned hyper-parameters for each ML model of scenario1

Parameter	MLP	RNN	LSTM	GRU	Bi-LSTM	Bi-GRU
Number of layers	1	1	1	1	2	2
Number of neurons in the hidden layer	84	100	100	100	100,2	30,100
Weight initialization	Glorot normal	Glorot normal	He normal	He normal	He normal	He normal
Activation function	relu	relu	relu	relu	Relu	relu
Optimizer	Adam	Adam	Adam	Adam	Adam	Adam
Learning rate	0.01	0.01	0.01	0.01	0.01	0.01
Number of batch size	8	8	8	8	8	8
Number of epochs With patience= 10	63	48	75	45	45	51
Loss	Mean squared error	Mean squared error	Mean squared error	Mean squared error	Mean squared error	Mean squared error
Dropout	0.0	0.0	0.5	0.0	0.0	0.0
Total tuning time	04m 11s	10m 40s	16m 35s	22m 03s	32m 47s	38m 27s

Table 4.4 Tuned hyper-parameters for each ML model of scenario 2

Parameter	MLP	RNN	LSTM	GRU	Bi-LSTM	Bi-GRU
Number of layers	1	1	1	3	3	3
Number of neurons in the hidden layer	200	2	2	82,90,100	28,2,2	100,2,20
Weight initialization	He normal	He normal	Glorot normal	Glorot normal	He normal	He normal
Activation function	relu	relu	Relu	Relu	tanh	relu
Optimizer	Adam	Adam	Adam	Adam	Adam	Adam
Learning rate	0.01	0.01	0.01	0.01	0.01	0.01
Number of batch size	8	8	8	8	8	8
Number of epochs With patience= 10	46	31	178	45	78	60
Loss	Mean squared error	Mean squared error	Mean squared error	Mean squared error	Mean squared error	Mean squared error
Dropout	0.0	0.0	0.5	0.0	0.0	0.0
Total tuning time	04m 13s	08m 47s	12m 41s	22m 03s	24m 38s	33m 19s

Table 4.5 Tuned hyper-parameters for each ML model of scenario 3

Parameter	**MLP**	**RNN**	**LSTM**	**GRU**	**Bi-LSTM**	**Bi-GRU**	
Number of layers	3	2	1	2	1	3	
Number of neurons in the hidden layer	200,100,100	26.200	178	200,56	200	100,2,58	
Weight initialization	He normal	Glorot normal	He normal	Glorot normal	Glorot normal	He normal	
Activation function	Relu	Relu	relu	Tanh	Relu	relu	
Optimizer	Adam	Adam	Adam	Adam	Adam	Adam	
Learning rate	0.01	0.001	0.01	0.01	0.01	0.01	
Number of batch size	8	8	8	8	8	8	
Number of epochs With patience=10	54	46	49	60	78	54	
Loss	Mean squared error	Mean squared error	Mean squared error	Mean squared error	Mean squared error	Mean squared error	Mean squared error
Dropout	0.0	0.0	0.5	0.2	0.0	0.5	
Total tuning time	03m 18s	06m 48s	14m 18s	16m 49s	16m 36s	24m 31s	

Table 4.6. Tuned hyper-parameters for each ML model of scenario 4

Parameter	**MLP**	**RNN**	**LSTM**	**GRU**	**Bi-LSTM**	**Bi-GRU**
Number of layers	2	2	1	2	1	2
Number of neurons in the hidden layer	200,100	26.200	172	80,82	200	100,100
Weight initialization	He normal	Glorot normal	Glorot normal	Glorot normal	He normal	Glorot normal
Activation function	Relu	Relu	Tanh	Relu	Relu	Tanh
Optimizer	Adam	Adam	Adam	Adam	Adam	Adam
Learning rate	0.01	0.001	0.01	0.01	0.01	0.01
Number of batch size	8	8	8	8	8	8
Number of epochs With patience=10	42	42	61	63	63	38
Loss	Mean squared error	Mean squared error	Mean squared error Mean squared error	Mean squared error	Mean squared error	Mean squared error
Dropout	0.0	0.0	0.2	0.0	0.0	0.0
Total tuning time	03m 42s	07m 13s	11m 12s	13m 49s	18m 44s	19m 32s

5. RESULTS AND DISCUSSION

In this chapter, the evaluation criteria, and hyper-parameter tuning are explained, then the results of both daily and day-ahead GHI prediction models are compared and discussed based on their performance evaluation.

5.1. Evaluation Criteria

Several evaluation metrics such as R^2, nMAE, MAPE, and nRMSE, are employed to compare the performances of both statistical and deep learning algorithms.

- R^2: represents the coefficient of determination which measures the goodness of fit in the model that determines the proportion of variance in the dependent variable that can be predicted by the independent variable.
- nMAE: absolute percentage error measures the accuracy of the prediction for continuous variables. and represents the amount of error we can expect from the prediction model on average by calculating the average magnitude of the absolute errors between the predicted and actual values.
- MAPE: mean absolute percentage error is the most widely common measure for checking forecasting model accuracy as it is interpretable and scale-independent performance metrics.
- nRMSE: normalized root mean squared error is used to evaluate the quality of the prediction model by measuring the standard deviation of the prediction error to represent the concentration of the data around the best-fitted line.

Unlike the nMAE, MAPE, and nRMSE, The higher the R^2, the greater accuracy of the prediction model. Formulas of R^2, MAE, MAPE, and nRMSE are as follows,

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{5.1}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{5.2}$$

$$MAPE\ (\%) = \frac{100}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{|y_i|} \tag{5.3}$$

$$nRMSE\ (\%) = \frac{100}{\bar{y}_i}\sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{5.4}$$

where y_i is the actual value, $\hat{y}$ is the predicted value, y is the mean of y_i, and n presents the number of samples.

5.2. Analyses of Experimental Results

In this section, comparison of the experimental results is evaluated by R^2, MAPE, nMAE, and nRMSE,

5.2.1. Experimental results of the different feature selection scenarios

In this section the results of the four scenarios of the feature selection criteria that explained in chapter 4 after employing day-ahead statistical and deep learning models are compared by evaluation metrics as shown in tables 5.1 to table 5.4.

Table 5.1 Results of scenario 1 dataset

	Model	R2 (%)	MAPE (%)	nMAE (%)	nRMSE (%)	CT (seconds)
SCENARIO 1	MLR	83.026	22.149	12.423	17.824	0.004
	MLP	85.280	20.965	11.182	16.599	22.256
	RNN	83.400	22.089	12.092	17.626	106.012
	LSTM	84.451	20.315	11.476	17.059	269.327
	GRU	84.944	20.269	11.721	16.787	205.43
	Bi-LSTM	85.495	19.901	11.684	16.477	345.545
	Bi-GRU	83.702	21.283	13.004	17.465	920.448
	Average	84.328	20.996	11.94	17.12	267.003

Table 5.2 Results of scenario 2 dataset.

	Model	R2 (%)	MAPE (%)	nMAE (%)	nRMSE (%)	CT (seconds)
Scenario 2	MLR	82.749	22.297	12.399	17.969	0.011
	MLP	84.269	21.245	11.516	17.159	15.016
	RNN	80.972	23.641	14.128	18.872	27.451
	LSTM	83.365	21.811	13.400	17.645	232.134
	GRU	84.249	21.342	11.630	17.170	315.513
	Bi-LSTM	84.355	20.066	11.498	17.112	422.790
	Bi-GRU	84.136	20.712	11.569	17.231	483.362
	Average	83.442	21.588	12.306	17.594	213.754

Table 5.3. Results of scenario 3 dataset.

	Model	R2 (%)	MAPE (%)	nMAE (%)	nRMSE (%)	CT (seconds)
SCENARIO 3	MLR	82.727	22.322	12.406	17.980	0.004
	MLP	82.820	22.249	12.079	17.932	20.923
	RNN	83.140	21.419	11.726	17.764	81.887
	LSTM	83.666	20.892	12.012	17.485	109.353
	GRU	82.776	22.453	11.936	17.955	256.5
	Bi-LSTM	84.122	20.884	11.593	17.239	289.003
	Bi-GRU	84.308	21.346	11.742	17.138	395.377
	Average	83.366	21.652	11.928	17.642	164.721

Table 5.4. Results of scenario 4 dataset.

	Model	R2 (%)	MAPE (%)	nMAE (%)	nRMSE (%)	CT (seconds)
SCENARIO 4	MLR	82.680	22.343	12.455	18.005	0.012
	MLP	83.174	21.702	11.983	17.746	13.907
	RNN	82.834	21.829	12.120	17.925	61.91
	LSTM	83.979	21.409	11.727	17.316	108.194
	GRU	83.380	22.637	12.193	17.637	167.517
	Bi-LSTM	82.311	22.416	12.266	18.196	182.832
	Bi-GRU	83.600	21.927	12.615	17.520	187.591
	Average	83.137	22.038	12.194	17.764	103.138

5.2.2. Summary of Results feature selection scenarios

Several feature selection scenarios were used in conjunction with S&ML techniques to determine the individual impact of each scenario on prediction performance using R^2, MAPE, nMAE, and nRMSE. Table 5.5 summarizes the results of the four scenarios and their rankings based on average MAPE values for each feature selection scenario for all S&ML techniques.

Table 5.5. summary of feature selection scenarios results

Scenario	R2 (%)	MAPE (%)	nMAE (%)	nRMSE (%)	CT (seconds)
Scenario 1	**84.328**	**20.996**	**11.94**	**17.12**	267.003
Scenario 2	83.442	21.588	12.306	17.594	213.754
Scenario 3	83.366	21.652	11.928	17.642	164.721
Scenario 4	83.137	22.038	12.194	17.764	**103.138**
Average	83.568	21.569	12.092	17.53	187.154

Regarding the table although scenario 1 took first place in terms of experimental results for the evaluation of their overall performance on all S&ML techniques, the other scenarios reduce computational time by reducing model complexity as a result of reducing the number of input features and selecting only high correlation features and ignoring low correlation features according to the threshold of each scenario, as shown in the table 5.6 below.

Table 5.6. Relation between results of each scenario regarding to scenario 1

Scenario	R2 (%)	MAPE (%)	nMAE (%)	nRMSE (%)	CT (seconds)
Scenario 1	100%	100%	100%	100%	100%
Scenario 2	-1.05%	+2.82%	+3.065%	+2.768%	-19.943%
Scenario 3	-1.14%	+3.124%	-0.101%	+3.049%	-38.307%
Scenario 4	-1.412%	+4.963%	+2.127%	+3.761%	-61.371%

As shown in the table 5.6 above, while decreasing the input features as the correlation threshold increased in each scenario affected and increased the error value, it had a significant effect on the computational time reduction. This significant impact can be seen between scenarios 1 and 4; increasing the threshold correlation in scenario 4 to select only features with correlation greater than 0.5 increased the MAPE value by 4.963% while decreasing the computational time by 61.371%.

5.2.3. Experimental Results of S&ML models

Seven different statistical and machine learning models are used to predict day-ahead GHI value based on the different feature selection scenarios mentioned above. The tables 5.7 to 5.13 below show the experimental results of each statistical and machine learning model, sorted by average MAPE value.

Table 5.7 Experimental results of Bi-LSTM models

Scenario	R2 (%)	MAPE (%)	nMAE (%)	nRMSE (%)	CT (seconds)
Scenario 1	85.495	19.901	11.684	16.477	345.545
Scenario 2	84.355	20.066	11.498	17.112	422.790
Scenario 3	84.122	20.884	11.593	17.239	289.003
Scenario 4	82.311	22.416	12.266	18.196	182.832
Average	84.071	20.817	11.76	17.256	310.043

Table 5.8. Experimental results of LSTM models

Scenario	R2 (%)	MAPE (%)	nMAE (%)	nRMSE (%)	CT (seconds)
Scenario 1	84.451	20.315	11.476	17.059	269.327
Scenario 2	83.365	21.811	13.400	17.645	232.134
Scenario 3	83.666	20.892	12.012	17.485	109.353
Scenario 4	83.979	21.409	11.727	17.316	108.194
Average	83.865	21.107	12.154	17.376	179.752

Table 5.9. Experimental results of Bi-GRU models

Scenario	R2 (%)	MAPE (%)	nMAE (%)	nRMSE (%)	CT (seconds)
Scenario 1	83.702	21.283	13.004	17.465	920.448
Scenario 2	84.136	20.712	11.569	17.231	483.362
Scenario 3	84.308	21.346	11.742	17.138	395.377
Scenario 4	83.600	21.927	12.615	17.520	187.591
Average	83.937	21.317	12.233	17.339	496.695

Table 5.10. Experimental results of MLP

Scenario	R2 (%)	MAPE (%)	nMAE (%)	nRMSE (%)	CT (seconds)
Scenario 1	85.280	20.965	11.182	16.599	22.256
Scenario 2	84.269	21.245	11.516	17.159	15.016
Scenario 3	82.820	22.249	12.079	17.932	20.923
Scenario 4	83.174	21.702	11.983	17.746	13.907
Average	83.886	21.54	11.69	17.359	18.026

Table 5.11. Experimental results of GRU models

Scenario	R2 (%)	MAPE (%)	nMAE (%)	nRMSE (%)	CT (seconds)
Scenario 1	84.944	20.269	11.721	16.787	205.43
Scenario 2	84.249	21.342	11.630	17.170	315.513
Scenario 3	82.776	22.453	11.936	17.955	256.5
Scenario 4	83.380	22.637	12.193	17.637	167.517
Average	83.837	21.675	11.87	17.387	236.24

Table 5.12. Experimental results of RNN models

Scenario	R2 (%)	MAPE (%)	nMAE (%)	nRMSE (%)	CT (seconds)
Scenario 1	83.400	22.089	12.092	17.626	106.012
Scenario 2	80.972	23.641	14.128	18.872	27.451
Scenario 3	83.140	21.419	11.726	17.764	81.887
Scenario 4	82.834	21.829	12.120	17.925	61.91
Average	82.587	22.245	12.517	18.047	69.315

Table 5.13. Experimental results of MLR models

Scenario	R2 (%)	MAPE (%)	nMAE (%)	nRMSE (%)	CT (seconds)
Scenario 1	83.026	22.149	12.423	17.824	0.004
Scenario 2	82.749	22.297	12.399	17.969	0.011
Scenario 3	82.727	22.322	12.406	17.980	0.004
Scenario 4	82.680	22.343	12.455	18.005	0.012
Average	82.796	22.278	12.421	17.945	0.008

Table 5.14. Summary of statistical and machine learning results

Model	R2 (%)	MAPE (%)	nMAE (%)	nRMSE (%)	CT (seconds)
Bi-LSTM	84.071	20.817	11.76	17.256	310.043
LSTM	83.865	21.107	12.154	17.376	179.752
Bi-GRU	83.937	21.317	12.233	17.339	496.695
MLP	83.886	21.54	11.69	17.359	18.026
GRU	83.837	21.675	11.87	17.387	236.24
RNN	82.587	22.245	12.517	18.047	69.315
MLR	82.796	22.278	12.421	17.945	0.008

According to the average value of MAPE, the Bi-LSTM model ranks first among all statistical and machine learning models (R2= 84.071%, and MAPE= 20.817%), followed by the LSTM, and then by the Bi-GRU, while MLR model ranks the first regards to the computational time, it ranks the last place corresponding to the other employed ML models.

6. CONCLUSIONS

As solar energy becomes more prevalent in electrical energy generation sources, a robust prediction model has become an essential tool for grid operators to dispatch solar and other energy sources optimally to balance generation and demand and increase system reliability. In this thesis, day-ahead GHI prediction models were developed in Adana, Türkiye, using S and ML techniques. To do so, daily horizon meteorological variables such as GHI, temperature, pressure, relative humidity, rainfall, snow depth, wind speed, and wind direction, and calendar variables such as day of month, day of week, week of year, day of month, month of year, and year are acquired from MERRA-2. Then the acquired data was processed to produce a cleansed data set. Moreover, four different feature selection scenarios based on the correlation between input features and target GHI were designed and applied to form four different datasets, following that, day-ahead GHI are predicted using MLR as a statistical technique as well as ML-based techniques such as MLP, RNN, LSTM, Bi-LSTM, GRU, and Bi-GRU. Finally, the obtained results are evaluated, with R^2, MAPE, MAE, and nRMSE.

As a result, the Bi-LSTM in first feature selection scenario (R^2= 85.495%, MAPE=19.901%) model outperformed the other statistical and machine learning models in terms of correlation and resulting error.

The thesis's future prospects are as follows:

- The proposed models could be applied to the solar energy investment market to improve scheduling and system reliability.
- According to the hyper-parameters tuning process, Adam and mean squared error ranks the best parameters for all models through all scenarios in both of optimizers and loss respectively.
- Due to the lack of solar measurement installation and the availability of other meteorological variables such as temperature and humidity, the

proposed models can be used as database generators for research and planning.

- Additional feature selection scenario can be designed with different correlation thresholds in order to reach the most optimum model regarding to both performance and time.
- By including some geographical parameters such as latitude, longitude, and altitude, the GRU model can be applied to a larger region that has similar weather conditions to Adana, such as the east Mediterranean region.
- The employed models and data processing were written in the Python programming language on the Kaggle environment; the performed study analysis can be utilized in the same environment in other languages such as R.
- Another hybrid deep learning version can be used to build the GHI prediction model, and a different hyper parameter tuner, such as GridSearchCv, can be used instead of the Keras tuner to improve model’s performance.

REFERENCES

A. Muhammad, J. M. Lee, S. W. Hong, S. J. Lee and E. H. Lee, "Deep Learning Application in Power System with a Case Study on Solar Irradiation Forecasting," 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), 2019, pp. 275-279, doi: 10.1109/ICAIIC.2019.8668969.

Abadi, Mart'in, Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., … others. (2016). Tensorflow: A system for large-scale machine learning. In 12th $USENIX$ Symposium on Operating Systems Design and Implementation ($OSDI$ 16) (pp. 265–283).

Abdulhamid, I.A.M., Sahiner, A. and Ozturk, M., 2018. Temperature and pressure-based pressure estimation of daily global solar radiation over Türkiye using fuzzy logic modelling. International Journal of Recent Scientific Research. http://dx.doi.org/10.24327/ijrsr.2019.1001.3083

Adana Climate, Weather By Month, Average Temperature (Turkey) - Weather Spark. (n.d.). https://weatherspark.com/y/99280/Average-Weather-in-Adana-Turkey-Year-Round.

Adnan Sözen; Erol Arcaklioğlu; Mehmet Özalp (2004). Estimation of solar potential in Türkiye by artificial neural networks using meteorological and geographical data. 45(18-19), 3033–3052. doi:10.1016/j.enconman.2003.12.020

Adnan Sözen; Erol Arcaklioğlu; Mehmet Özalp; E.Galip Kanit (2004). Use of artificial neural networks for mapping of solar potential in Türkiye. 77(3), 273–286. doi:10.1016/s0306-2619(03)00137-5

Ağbulut, Ü., Gürel, A.E. and Biçen, Y., 2021. Prediction of daily global solar radiation using different machine learning algorithms: Evaluation and comparison. Renewable and Sustainable Energy Reviews, 135, p.110114.https://doi.org/10.1016/j.rser.2020.110114

Ahmet Duran Şahin (2007). A new formulation for solar irradiation and sunshine duration estimation. , 31(2), 109–118. doi:10.1002/er.1229

Ahmet Koca; Hakan F. Oztop; Yasin Varol; Gonca Ozmen Koca (2011). Estimation of solar radiation using artificial neural networks with different input parameters for the Mediterranean region of Anatolia in Türkiye. , 38(7), 8756–8762. doi:10.1016/j.eswa.2011.01.085

Aji, P., Wakamori, K. and Moreno, H., 2018, November. Highly accurate daily solar radiation forecasting using SW-SVR for hybrid power plants in Indonesia. In 2018 4th international conference on nano electronics research and education (ICNERE) (pp. 1-5). IEEE. 10.1109/ICNERE.2018.8642593

Akarslan, E. and Hocaoglu, F.O., 2017. A novel method based on similarity for hourly solar irradiance forecasting. Renewable Energy, 112, pp.337-346. https://doi.org/10.1016/j.renene.2017.05.058

Akarslan, E., Hocaoglu, F.O. and Edizkan, R., 2018. Novel short term solar irradiance forecasting models. Renewable Energy, 123, pp.58-66. https://doi.org/10.1016/j.renene.2018.02.048

Akbaba, E.C., 2019. Using artificial neural network (ANN) techniques for solar irradiation predictions (Master's thesis, Middle East Technical University).

Akbaba, E.C., Yüce, E. and Akinoglu, B.G., 2018, December. Deep learning algorithm applied to daily solar irradiation estimations. In 2018 6th International Renewable and Sustainable Energy Conference (IRSEC) (pp. 1-4). IEEE. 10.1109/IRSEC.2018.8702963

Akhter, M.N. et al. (2019), "Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques", IET Renewable Power Generation, Vol. 13/7, Institution of Engineering and Technology, Stevenage, United Kingdom, pp. 1 009–1 023, https://ieeexplore.ieee.org/document/8707118

AL-Rousan, N. and Al-Najjar, H., 2021. A Comparative Assessment of Time Series Forecasting Using NARX and SARIMA to Predict Hourly, Daily, and

Monthly Global Solar Radiation Based on Short-Term Dataset. Arabian Journal for Science and Engineering, 46(9), pp.8827-8848. https://doi.org/10.1007/s13369-021-05669-6

Alsafadi, M. and Filik, Ü.B., 2020. HOURLY GLOBAL SOLAR RADIATION ESTIMATION BASED ON MACHINE LEARNING METHODS IN ESKISEHIR. Eskişehir Technical University Journal of Science and Technology A-Applied Sciences and Engineering, 21(2), pp.294-313. https://doi.org/10.3390/sym11020240

Alsafadi, M. and Filik, Ü.B., 2020. HOURLY GLOBAL SOLAR RADIATION ESTIMATION BASED ON MACHINE LEARNING METHODS IN ESKISEHIR. Eskişehir Technical University Journal of Science and Technology A-Applied Sciences and Engineering, 21(2), pp.294-313.https://doi.org/10.18038/estubtda.650497

Antonopoulos, V.Z., Papamichail, D.M., Aschonitis, V.G. and Antonopoulos, A.V., 2019. Solar radiation estimation methods using ANN and empirical models. Computers and Electronics in Agriculture, 160, pp.160-167. https://doi.org/10.1016/j.compag.2019.03.022

Aslam M, Lee J-M, Kim H-S, Lee S-J, Hong S. Deep Learning Models for Long-Term Solar Radiation Forecasting Considering Microgrid Installation: A Comparative Study.Energies. 2020; 13(1):147. https://doi.org/10.3390/en13010147.

Aslam, M., Lee, J.M., Kim, H.S., Lee, S.J. and Hong, S., 2019. Deep learning models for long-term solar radiation forecasting considering microgrid installation: A comparative study. Energies, 13(1), p.147.https://doi.org/10.3390/en13010147

B.G. Akinoğlu; A. Ecevit (1990). A further comparison and discussion of sunshine-based models to estimate global solar radiation. , 15(10), 865–872. doi:10.1016/0360-5442(90)90068-d

Babu, M.K. and Ray, P., 2019, March. A wavelet neural network model for hourly solar radiation forecasting from daily solar radiation. In 2019 IEEE 5th International Conference for Convergence in Technology (I2CT) (pp. 1-5). IEEE. 10.1109/I2CT45611.2019.9033864

Bakirci, K. and Kirtiloglu, Y., 2018. Prediction of diffuse solar radiation using satellite data. International Journal of Green Energy, 15(2), pp.76-79. https://doi.org/10.1080/15435075.2018.1423976

Bakirci, K., 2015. Models for the estimation of diffuse solar radiation for typical cities in Türkiye. Energy, 82, pp.827-838. https://doi.org/10.1016/j.energy.2015.01.093

Bakirci, K., 2017. Prediction of global solar radiation and comparison with satellite data. Journal of Atmospheric and Solar-Terrestrial Physics, 152, pp.41-49. https://doi.org/10.1016/j.jastp.2016.12.002

Bakirci, Kadir (2015). Models for the estimation of diffuse solar radiation for typical cities in Türkiye. Energy, 82(), 827–838. doi:10.1016/j.energy.2015.01.093

Balitskiy, S., Bilan, Y., Strielkowski, W. and Štreimikienė, D., 2016. Energy efficiency and natural gas consumption in the context of economic development in the European Union. Renewable and Sustainable Energy Reviews, 55, pp.156-168. https://doi.org/10.1016/j.rser.2015.10.053

Barutcu, B., Tanriover, S.T., Sakarya, S., Incecik, S., Sayinta, F.M., Caliskan, E., Kahraman, A., Aksoy, B., Kahya, C. and Topcu, S., 2015. Improving WRF GHI Forecasts with Model Output Statistics. In Progress in Clean Energy, Volume 1 (pp. 291-299). Springer, Cham. https://doi.org/10.1007/978-3-319-16709-1_20

Basaran, K., Özçift, A. and Kılınç, D., 2019. A new approach for prediction of solar radiation using ensemble learning algorithms. Arabian Journal for Science and Engineering, 44(8), pp.7159-7171.

Baykal, E., 2019. Creating organisational commitment through spiritual leadership: Mediating effect of meaning at work. Business and Management Studies: An International Journal. http://doi.org/10.15295/bmij.v7i2.1113

Bitirgen, K., 2018. Solar radiation estimation for modelling of PV arrays and calculation of solar energy potential based on ArcGIS (Doctoral dissertation, Anadolu University (Türkiye)).

Boubaker, S., Benghanem, M., Mellit, A., Lefza, A., Kahouli, O. and Kolsi, L., 2021. Deep neural networks for predicting solar radiation at Hail Region, Saudi Arabia. IEEE Access, 9, pp.36719-36729. 10.1109/ACCESS.2021.3062205

Boubaker, S., Benghanem, M., Mellit, A., Lefza, A., Kahouli, O., & Kolsi, L. (2021). Deep Neural Networks for Predicting Solar Radiation at Hail Region, Saudi Arabia. IEEE Access, 9, 36719–36729. doi:10.1109/access.2021.3062205.

Boussaada, Z., Curea, O., Remaci, A., Camblong, H. and Mrabet Bellaaj, N., 2018. A nonlinear autoregressive exogenous (NARX) neural network model for the prediction of daily direct solar radiation. Energies, 11(3), p.620. https://doi.org/10.3390/en11030620

Bouzgou, H. and Gueymard, C.A., 2019. Fast short-term global solar irradiance forecasting with wrapper mutual information. Renewable energy, 133, pp.1055-1065. https://doi.org/10.1016/j.renene.2018.10.096

Brahma, B. and Wadhvani, R., 2020. Solar irradiance forecasting based on deep learning methodologies and multi-site data. Symmetry, 12(11), p.1830.https://doi.org/10.3390/sym12111830

Brahma, Banalaxmi, and Rajesh Wadhvani. 2020. "Solar Irradiance Forecasting Based on Deep Learning Methodologies and Multi-Site Data" Symmetry 12, no. 11: 1830. https://doi.org/10.3390/sym12111830.

Bulut, H.; Büyükalaca, O.; Yılmaz, A. (2009). Generation of Typical Solar Radiation Year for the Mediterranean Region of Türkiye. International Journal of Green Energy, 6(2), 173–183. doi:10.1080/15435070902784970

Çelik, Ö., Teke, A. and Yıldırım, H.B., 2016. The optimized artificial neural network model with Levenberg–Marquardt algorithm for global solar radiation estimation in Eastern Mediterranean Region of Türkiye. Journal of cleaner production, 116, pp.1-12. https://doi.org/10.1016/j.rser.2017.08.033

Çelik, Özgür; Teke, Ahmet; Yildirim, H. Başak, 2015. The optimized artificial neural network model with Levenberg-Marquardt algorithm for global solar radiation estimation in Eastern Mediterranean Region of Türkiye. Journal of Cleaner Production, (), S0959652615019071–. doi:10.1016/j.jclepro.2015.12.082

Cheng, H., Chen, C., Wu, S., Mirza, Z.A. and Liu, Z., 2017. Energy evaluation of cropping, poultry rearing, and fish raising systems in the drawdown zone of Three Gorges Reservoir of China. Journal of Cleaner Production, 144, pp.559-571. https://doi.org/10.1016/j.jclepro.2016.12.053

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

Chollet, F. et al., 2015. Keras. GitHub. url= https://github.com/fchollet/keras

Chollet, F., & others. (2015). Keras. GitHub. Retrieved from https://github.com/fchollet/keras.

ÇİÇEK, O., MILLAD, M.A.M. and ERKEN, F., 2019. Energy prediction based on modelling and simulation analysis of an actual grid-connected photovoltaic power plant in Türkiye. European Journal of Technique (EJT), 9(2), pp.159-174. https://doi.org/10.36222/ejt.593250

Citakoglu, H., 2015. Comparison of artificial intelligence techniques via empirical equations for prediction of solar radiation. Computers and Electronics in Agriculture, 118, pp.28-37. https://doi.org/10.1016/j.compag.2015.08.020

Çoban, V. and Onar, S.Ç., 2019, July. Solar radiation prediction based on machine learning for Istanbul in Türkiye. In the International Conference on

Intelligent and Fuzzy Systems (pp. 197-204). Springer, Cham. https://doi.org/10.1007/978-3-030-23756-1_25

Çoban, V. and Onar, S.Ç., 2019, July. Solar radiation prediction based on machine learning for Istanbul in Türkiye. In the International Conference on Intelligent and Fuzzy Systems (pp. 197-204). Springer, Cham. https://doi.org/10.1007/978-3-030-23756-1_25

Colak, M., Yesilbudak, M. and Bayindir, R., 2019, November. Forecasting of daily total horizontal solar radiation using grey wolf optimizer and multilayer perceptron algorithms. In 2019 8th International Conference on Renewable Energy Research and Applications (ICRERA) (pp. 939-942). IEEE.

Colak, M., Yesilbudak, M. and Bayindir, R., 2020. Daily photovoltaic power prediction enhanced by hybrid GWO-MLP, ALO-MLP and WOA-MLP models using meteorological information. Energies, 13(4), p.901. . https://doi.org/10.3390/en13040901

Demirdelen, T., Ozge Aksu, I., Esenboga, B., Aygul, K., Ekinci, F. and Bilgili, M., 2019. A new method for generating short-term power forecasting based on artificial neural networks and optimization methods for solar photovoltaic power plants. In Solar photovoltaic power plants (pp. 165-189). Springer, Singapore. https://doi.org/10.1007/978-981-13-6151-7_8

Durrani, S.P., Balluff, S., Wurzer, L. and Krauter, S., 2018. Photovoltaic yield prediction using an irradiance forecast model based on multiple neural networks. Journal of Modern Power Systems and Clean Energy, 6(2), pp.255-267. 10.1007/s40565-018-0393-5

Ekici, C. and Teke, I., 2019. Global solar radiation estimation from measurements of visibility and air temperature extremes. Energy Sources, Part A: Recovery, Utilisation, and Environmental Effects, 41(11), pp.1344-1359. https://doi.org/10.1080/15567036.2018.1548513

Fan, J., Wu, L., Zhang, F., Cai, H., Zeng, W., Wang, X. and Zou, H., 2019. Empirical and machine learning models for predicting daily global solar radiation from sunshine duration: A review and case study in China. Renewable and Sustainable Energy Reviews, 100, pp.186-212. https://doi.org/10.1016/j.rser.2018.10.018

Farhath, Z.A., Arputhamary, B. and Arockiam, L., 2016. A survey on ARIMA forecasting using time series models. Int. J. Comput. Sci. Mobile Comput, 5, pp.104-109.

Feng, Y., Gong, D., Zhang, Q., Jiang, S., Zhao, L. and Cui, N., 2019. Evaluation of temperature-based machine learning and empirical models for predicting daily global solar radiation. Energy conversion and management, 198, p.111780. https://doi.org/10.1016/j.enconman.2019.111780

Gairaa, K., Khellaf, A., Messlem, Y. and Chellali, F., 2016. Estimation of the daily global solar radiation based on Box–Jenkins and ANN models: A combined approach. Renewable and Sustainable Energy Reviews, 57, pp.238-249. https://doi.org/10.1016/j.rser.2015.12.111

Gelaro, R.et al. (2017). The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), Journal of Climate, 30(14), 5419-5454.

Ghazvinian, Hamidreza; Mousavi, Sayed-Farhad; Karami, Hojat; Farzin, Saeed; Ehteram, Mohammad; Hossain, Md Shabbir; Fai, Chow Ming; Hashim, Huzaifa Bin; Singh, Vijay P.; Ros, Faizah Che; Ahmed, Ali Najah; Afan, Haitham Abdulmohsin; Lai, Sai Hin; El-Shafie, Ahmed; Li, Yang (2019). Integrated support vector regression and an improved particle swarm optimization-based model for solar radiation prediction. PLOS ONE, 14(5), e0217634–. doi:10.1371/journal.pone.0217634

Ghimire, S. (2019). Predictive modelling of global solar radiation with artificial intelligence approaches using MODIS satellites and atmospheric reanalysis data for Australia (Doctoral dissertation, University of Southern Queensland).https://doi.org/10.1016/j.jclepro.2019.01.158

Ghimire, S., Deo, R. C., Raj, N., & Mi, J. (2019). Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms. Applied Energy, 253, 113541. doi:10.1016/j.apenergy.2019.11354.

Ghimire, S., Deo, R.C., Downs, N.J. and Raj, N., 2019. Global solar radiation prediction by ANN integrated with the European Centre for medium range weather forecast fields in solar rich cities of Queensland Australia. Journal of cleaner production, 216, pp.288-310. https://doi.org/10.1016/j.jclepro.2019.01.158

Ghimire, S., Deo, R.C., Raj, N. and Mi, J., 2019. Deep solar radiation forecasting with convolutional neural networks and long short-term memory network algorithms. Applied Energy, 253, p.113541. https://doi.org/10.1016/j.apenergy.2019.113541

Global Modelling and Assimilation Office (GMAO) (2015), inst3_3d_asm_Cp: MERRA-2 3D IAU State, Meteorology Instantaneous 3-hourly (p-coord, 0.625x0.5L42), version 5.12.4, Greenbelt, MD, USA: Goddard Space Flight Centre Distributed Active Archive Centre (GSFC DAAC), Accessed Enter User Data Access Date at doi: 10.5067/VJAFPLI1CSIV.

Global Modelling and Assimilation Office (GMAO) (2015), inst3_3d_asm_Cp: MERRA-2 3D IAU State, Meteorology Instantaneous 3-hourly (p-coord, 0.625x0.5L42), version 5.12.4, Greenbelt, MD, USA: Goddard Space Flight Centre Distributed Active Archive Centre (GSFC DAAC), Accessed 01.05.2022 at doi: 10.5067/VJAFPLI1CSIV.

Goksu, B., Bayraktar, M. and Pamik, M., 2020. Prediction of solar energy potential with artificial neural networks. In Environmentally-Benign Energy Solutions (pp. 247-258). Springer, Cham. . https://doi.org/10.1007/978-3-030-20637-6_13

Goksu, B., Bayraktar, M. and Pamik, M., 2020. Prediction of solar energy potential with artificial neural networks. In Environmentally-Benign Energy Solutions (pp. 247-258). . https://doi.org/10.1007/978-3-030-20637-6_13

Goncu, O. , Koroglu, T. & Ozdıl, N. F. (2021). Estimation of hourly global solar radiation using an artificial neural network in Adana province, Türkiye . Journal of Thermal Engineering , Volume: 7 Supplement: 14 - December 2021 , 2017-2030 . DOI: 10.18186/thermal.1051313

Güçlü, Y.S., Dabanlı, İ., Şişman, E. and Şen, Z., 2015. HARmonic–LINear (HarLin) model for solar irradiation estimation. Renewable Energy, 81, pp.209-218. https://doi.org/10.1016/j.renene.2015.03.035

Güçlü, Yavuz Selim; Dabanlı, İsmail; Şişman, Eyüp; Şen, Zekai (2015). HARmonic–LINear (HarLin) model for solar irradiation estimation. Renewable Energy, 81(), 209–218. doi:10.1016/j.renene.2015.03.035

Güçlü, Yavuz Selim; Yeleğen, Mehmet Öner; Dabanlı, İsmail; Şişman, Eyüp (2014). Solar irradiation estimations and comparisons by ANFIS, Angström–Prescott and dependency models. Solar Energy, 109(), 118–124. doi:10.1016/j.solener.2014.08.027

Gumus, B. and Kilic, H., 2018. Time dependent prediction of monthly global solar radiation and sunshine duration using exponentially weighted moving average in southeastern Türkiye. Thermal Science, 22(2), pp.943-951. https://doi.org/10.2298/TSCI160107228K

Gungor, A., Gokcek, M., Yalcin, F., Kocer, A., Yaka, I.F. and Sardogan, G.T., 2015. Determining the best model for estimating the monthly mean daily global solar radiation on a horizontal surface–A case study in Nigde, Türkiye. World Journal of Engineering. 10.1260/1708-5284.12.3.307

H. Ögelman; A. Ecevit; E. Tasdemiroğlu (1984). A new method for estimating solar radiation from bright sunshine data. , 33(6), 619–625. doi:10.1016/0038-092x(84)90018-5

H. Zhou, Y. Zhang, L. Yang, Q. Liu, K. Yan and Y. Du, "Short-Term Photovoltaic Power Forecasting Based on Long Short Term Memory Neural Network and Attention Mechanism," in IEEE Access, vol. 7, pp. 78063-78074, 2019, doi: 10.1109/ACCESS.2019.2923006.

Hai, T., Sharafati, A., Mohammed, A., Salih, S.Q., Deo, R.C., Al-Ansari, N. and Yaseen, Z.M., 2020. Global solar radiation estimation and climatic variability analysis using extreme learning machine based predictive models. IEEE Access, 8, pp.12026-12042 doi:10.1109/ACCESS.2020.2965303

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., … Oliphant, T. E. (2020). Array programming with NumPy. Nature, 585, 357–362. https://doi.org/10.1038/s41586-020-2649-2.

Hochreiter, S. and Schmidhuber, J. (1997).“long short-term memory”. *Neural Computation*, *9*(8). 1735e1780, https://doi.org/10.1162/neco.1997.9.8.1735.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in Science & Engineering, 9(3), 90–95.

Hüsamettin Bulut; Orhan Büyükalaca (2007). Simple model for the generation of daily global solar-radiation data in Türkiye. , 84(5), 477–491. doi:10.1016/j.apenergy.2006.10.003

Husein, M. and Chung, I.Y., 2019. Day-ahead solar irradiance forecasting for microgrids using a long short-term memory recurrent neural network: A deep learning approach. Energies, 12(10), p.1856. https://doi.org/10.3390/en12101856

Husein, Munir, and Il-Yop Chung. 2019. "Day-Ahead Solar Irradiance Forecasting for Microgrids Using a Long Short-Term Memory Recurrent Neural Network: A Deep Learning Approach" Energies 12, no. 10: 1856. https://doi.org/10.3390/en12101856.

Incecik, S., Sakarya, S., Tilev, S., Kahraman, A., Aksoy, B., Caliskan, E., ... & Odman, M. T. (2019). Evaluation of WRF parameterizations for global horizontal irradiation forecasts: A study for Türkiye. Atmósfera, 32(2), 143-158. https://doi.org/10.20937/atm.2019.32.02.05

Inci Türk Toğrul; Hasan Toğrul (2002). Global solar radiation over Türkiye: comparison of predicted and measured data. , 25(1), 55–67. doi:10.1016/s0960-1481(00)00197-x

Inci Turk Toğrul; Hasan Toğrul; Duygu Evin (2000). Estimation of global solar radiation under clear sky radiation in Türkiye. , 21(2), 271–287. doi:10.1016/s0960-1481(99)00128-7

J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," preprint arXiv (2015).

Jahani, B. and Mohammadi, B., 2019. A comparison between the application of empirical and ANN methods for estimation of daily global solar radiation in Iran. Theoretical and Applied Climatology, 137(1), pp.1257-1269. https://doi.org/10.1007/s00704-018-2666-3

Jallal, M.A., El Yassini, A., Chabaa, S., Zeroual, A. and Ibnyaich, S., 2020. A Deep Learning Algorithm for Solar Radiation Time Series Forecasting: A Case Study of El Kelaa Des Sraghna City. Rev. d'Intelligence Artif., 34(5), pp.563-569.10.1109/IRSEC53969.2021.9741202

Jallal, M.A., El Yassini, A., Chabaa, S., Zeroual, A. and Ibnyaich, S., 2020. AI Data Driven Approach-Based Endogenous Inputs for Global Solar Radiation Forecasting. Ingénierie des Systèmes d Inf., 25(1), pp.27-34.https://doi.org/10.18280/isi.250104

K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," preprint arXiv (2014).

Kaba, K., Sarıgül, M., Avcı, M. and Kandırmaz, H.M., 2018. Estimation of daily global solar radiation using deep learning models. Energy, 162, pp.126-135. https://doi.org/10.1016/j.energy.2018.07.202

Kaba, K., Sarıgül, M., Avcı, M., & Kandırmaz, H. M. (2018). Estimation of daily global solar radiation using deep learning model.Energy,162,126–135. doi:10.1016/j.energy.2018.07.202.

Kaba, Kazım; Sarıgül, Mehmet; Avcı, Mutlu; Kandırmaz, H. Mustafa (2018). Estimation Of Daily Global Solar Radiation Using Deep Learning Model. Energy, (), S0360544218315007–. doi:10.1016/j.energy.2018.07.202

Kadir Bakirci (2009). Correlations for estimation of daily global solar radiation with hours of bright sunshine in Türkiye. , 34(4), 485–501. doi:10.1016/j.energy.2009.02.005

Kaggle: Your Machine Learning and Data Science Community. (n.d.). https://www.kaggle.com/.

Karaveli, A.B. and Akinoglu, B.G., 2018. Comparisons and critical assessment of global and diffuse solar irradiation estimation methodologies. International Journal of Green Energy, 15(5), pp.325-332. Gumus, B. and Kilic, H., 2018. Time dependent prediction of monthly global solar radiation and sunshine duration using exponentially weighted moving average in southeastern Türkiye. Thermal Science, 22(2), pp.943-951. https://doi.org/10.1080/15435075.2018.1452743

Kasht, H., 2018. Sky conditions classification and estimation of solar radiation for clear sky days (Master's thesis, Middle East Technical University).

Kavakçı, G., 2021. Time series prediction of solar power generation using trend decomposition (Master's thesis, Middle East Technical University).

Kaya, Ö., Analysis of meteorological data by artificial intelligence techniques and estimation of global solar radiation(Master's thesis, Fen Bilimleri Enstitüsü).

Kaygusuz, K., 2019. Prediction of the solar photovoltaic potential for electricity generation in Türkiye. Journal of Engineering Research and Applied

Science, 8(2), pp.1271-1278. http://journaleras.com/index.php/jeras/article/view/182

Khoshvaght-Aliabadi, M. and Arani-Lahtari, Z., 2016. Proposing new configurations for twisted square channel (TSC): Nanofluid as working fluid. Applied Thermal Engineering, 108, pp.709-719. https://doi.org/10.1016/j.applthermaleng.2016.07.173

Khosravi, A., Koury, R.N.N., Machado, L. and Pabon, J.J.G., 2018. Prediction of hourly solar radiation in Abu Musa Island using machine learning algorithms. Journal of Cleaner Production, 176, pp.63-75. https://doi.org/10.1016/j.jclepro.2017.12.065

Khosravi, A., Nunes, R.O., Assad, M.E.H. and Machado, L., 2018. Comparison of artificial intelligence methods in estimation of daily global solar radiation. Journal of cleaner production, 194, pp.342-358. Kaba, K., Sarıgül, M., Avcı, M. and Kandırmaz, H.M., 2018. Estimation of daily global solar radiation using deep learning models. Energy, 162, pp.126-135. https://doi.org/10.1016/j.jclepro.2018.05.147

Kisi, Ozgur (2014). Modelling solar radiation of the Mediterranean region in Türkiye by using a fuzzy genetic approach. Energy, 64(), 429–436. doi:10.1016/j.energy.2013.10.009

Kisi, Ozgur; Alizamir, Meysam; Trajkovic, Slavisa; Shiri, Jalal; Kim, Sungwon (2020). Solar Radiation Estimation in Mediterranean Climate by Weather Variables Using a Novel Bayesian Model Averaging and Machine Learning Methods. Neural Processing Letters, (), –. doi:10.1007/s11063-020-10350-4

Kisi, Ozgur; Heddam, Salim; Yaseen, Zaher Mundher (2019). The implementation of univariable scheme-based air temperature for solar radiation prediction: New development of dynamic evolving neural-fuzzy inference system model. Applied Energy, 241(), 184–195. doi:10.1016/j.apenergy.2019.03.089

Kurtgoz, Y. and Deniz, E., 2016. Global solar radiation estimation using artificial neural networks by the addition of nearby meteorological stations' solar radiation data and exergy of solar radiation: a case study. International Journal of Exergy, 21(3), pp.315-330.

Kurtgoz, Yusuf (2018). Exergetic, Energetic and Environmental Dimensions || Comparison of ANN, Regression Analysis, and ANFIS Models in Estimation of Global Solar Radiation for Different Climatological Locations. , (), 133–148. doi:10.1016/B978-0-12-813734-5.00008-1

Lima, M.A.F.B., Carvalho, P.C., Braga, A.P.D.S., Ramírez, L.M.F. and Leite, J.R., 2018. MLP back propagation artificial neural network for solar resource forecasting in equatorial areas. Renewable Energy and Power Quality Journal (RE&PQJ), 1, pp.175-80. https://doi.org/10.24084/repqj16.253

Liu, Y., Zhou, Y., Chen, Y., Wang, D., Wang, Y. and Zhu, Y., 2020. Comparison of support vector machine and copula-based nonlinear quantile regression for estimating the daily diffuse solar radiation: A case study in China. Renewable Energy, 146, pp.1101-1112.https://doi.org/10.1016/j.renene.2019.07.053

Lotfinejad, M.M., Hafezi, R., Khanali, M., Hosseini, S.S., Mehrpooya, M. and Shamshirband, S., 2018. A comparative assessment of predicting daily solar radiation using bat neural network (BNN), generalised regression neural network (GRNN), and neuro-fuzzy (NF) system: A case study. Energies, 11(5), p.1188. https://doi.org/10.3390/en11051188

M. Alanazi and A. Khodaei, "Day-ahead solar forecasting using time series stationarization and feed-forward neural network", *Proc. 48th North Am. Power Symp. (NAPS)*, pp. 1-6, 2016.

Marzouq, M., Bounoua, Z., El Fadili, H., Mechaqrane, A., Zenkouar, K. and Lakhliai, Z., 2019. New daily global solar irradiation estimation model based on automatic selection of input parameters using evolutionary artificial neural networks. Journal of Cleaner Production, 209, pp.1105-1118. https://doi.org/10.1016/j.jclepro.2018.10.254

Marzouq, M., El Fadili, H., Zenkouar, K., Lakhliai, Z. and Amouzg, M., 2020. Short term solar irradiance forecasting via a novel evolutionary multi-model framework and performance assessment for sites with no solar irradiance data. Renewable Energy, 157, pp.214-231. https://doi.org/10.1016/j.renene.2020.04.133

McKinney, W., & others. (2010). Data structures for statistical computing in python. In Proceedings of the 9th Python in Science Conference (Vol. 445, pp. 51–56).

Mehmet Bilgili; Muammer Ozgoren (2011). Daily total global solar radiation modelling from several meteorological data. , 112(3-4), 125–138. doi:10.1007/s00703-011-0137-9

Mellit, A., Pavan, A. M., & Lughi, V. (2021). Deep learning neural networks for short-term photovoltaic power forecasting. Renewable Energy, 172, 276–288. doi:10.1016/j.renene.2021.02.166.

Mellit, A., Pavan, A.M. and Lughi, V., 2021. Deep learning neural networks for short-term photovoltaic power forecasting. Renewable Energy, 172, pp.276-288https://doi.org/10.1016/j.renene.2021.02.166

MENR (Republic of Türkiye Ministry of Energy and Natural Resources). solar info bank, https://enerji.gov.tr/bilgi-merkezi-enerji-gunes-en; 2022 [accessed 3 May 2022].

Mishra, M., Byomakesha Dash, P., Nayak, J., Naik, B., & Kumar Swain, S. (2020). Deep Learning and Wavelet Transform integrated approach for Short-term solar PV power prediction. Measurement, 108250. doi:10.1016/j.measurement.2020.108250.

Mishra, M., Dash, P.B., Nayak, J., Naik, B. and Swain, S.K., 2020. Deep learning and wavelet transform integrated approach for short-term solar PV power prediction. Measurement, 166, p.108250.https://doi.org/10.1016/j.measurement.2020.108250

Mohammad, A.T., Mahmood, I. and Mahjoob, S., 2018. Three structures of a multilayer artificial neural network for predicting the solar radiation of Baghdad City-Iraq. International Journal of Applied Engineering Research, 13(6), pp.3759-3766.

Mohammed, H.N. and Mengüç, M.P., 2018. Solar radiation exergy and quality performance for Iraq and Türkiye. International Journal of Exergy, 25(4), pp.364-385.

Muammer Ozgoren; Mehmet Bilgili; Besir Sahin (2012). Estimation of global solar radiation using ANN over Türkiye. , 39(5), 5043–5051. doi:10.1016/j.eswa.2011.11.036

Ojo, O.S. and Adeyemi, B., 2020. Application of nonlinear autoregressive neural network to estimation of global solar radiation over Nigeria. Journal of Advances in Science and Engineering, 3(2), pp.68-79. https://doi.org/10.37121/jase.v3i2.98

Ozan Şenkal (2010). Modelling of solar radiation using remote sensing and artificial neural network in Türkiye. , 35(12), 4795–4801. doi:10.1016/j.energy.2010.09.009

Ozan Şenkal; Tuncay Kuleli (2009). Estimation of solar radiation over Türkiye using artificial neural network and satellite data. , 86(7-8), 1222–1228. doi:10.1016/j.apenergy.2008.06.003

Ozbek, A., Yildirim, A. and Bilgili, M., 2021. Deep learning approach for one-hour ahead forecasting of energy production in a solar-PV plant. Energy Sources, Part A: Recovery, Utilisation, and Environmental Effects, pp.1-16.https://doi.org/10.1080/15567036.2021.1924316

Ozturk, M., 2015. An evaluation of global solar radiation empirical formulations in Isparta, Türkiye. Energy Sources, Part A: Recovery, Utilisation, and

Environmental Effects, 37(22), pp.2474-2486. https://doi.org/10.1080/15567036.2012.745037

Paiva, G. M. D. ,2021. Machine learning methods applied to intraday solar forecasting for power system operation. Federal University of Goias (UFG).

Pearson, K., & Filon, L. N. G. (1898). VII. Mathematical contributions to the theory of evolution.IV. On the probable errors of frequency constants and on the influence of random selection on variation and correlation. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, (191), 229-311.

Pedregosa, F et al. ,2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, vol.12, pp. 2825—2830.

Pedregosa, F., Varoquaux, Ga"el, Gramfort, A., Michel, V., Thirion, B., Grisel, O., … others. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12(Oct), 2825–2830.

Petros, T. and Vlahavas,2021. I., RENEWABLE ENERGY FORECASTING. MSc thesis, Aristotle University Of Thessaloniki

Qing, X. and Niu, Y., 2018. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. Energy, 148, pp.461-468. https://doi.org/10.1016/j.energy.2018.01.177

R. Tuğrul Oğulata; S. Noyan Oğulata (2002). Solar radiation on Adana, Türkiye. , 71(4), 351–358. doi:10.1016/s0306-2619(01)00050-2

Ramazan, Ü.N.L.Ü., 2019. A comparative study of machine learning and deep learning for time series forecasting: a case study of choosing the best prediction model for Türkiye electricity production. Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 23(2), pp.635-646. https://doi.org/10.1007/s13369-019-03841-7

Ramazan, Ü.N.L.Ü., 2019. A comparative study of machine learning and deep learning for time series forecasting: a case study of choosing the best prediction model for Türkiye electricity production. Süleyman Demirel

Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 23(2), pp.635-646. https://doi.org/10.19113/sdufenbed.494396

Rocha, P.A., Fernandes, J.L., Modolo, A.B., Lima, R.J., da Silva, M.E. and Bezerra, C.A., 2019. Estimation of daily, weekly and monthly global solar radiation using ANNs and a long data set: a case study of Fortaleza, in Brazilian Northeast region. International Journal of Energy and Environmental Engineering, 10(3), pp.319-334. https://doi.org/10.1007/s40095-019-0313-0.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, *65*(6), 386.

Rumelhart, D.E., Hinton, G.E., & Williams, R.J. (1986). Learning representations by back-propagating errors. *Nature, 323*, 533-536.

Sahan, Muhittin; Yakut, Emre; Sarpün, I.H.; Tel, E.; Aydin, A.; Kaplan, A. (2016). Estimation of monthly global solar radiation in the eastern Mediterranean region in Türkiye by using artificial neural networks. EPJ Web of Conferences, 128(), 06001–. doi:10.1051/epjconf/201612806001

Şahin, Mehmet; Kaya, Yılmaz; Uyar, Murat; Yıldırım, Selçuk (2014). Application of extreme learning machine for estimating solar radiation from satellite data. International Journal of Energy Research, 38(2), 205–212. doi:10.1002/er.3030

SARKEZ, A.E., 2020. SOLAR RADIATION PREDICTION IN SOLAR ENERGY SYSTEMS USING ANN MODELS: A CASE STUDY (Doctoral dissertation) Institute of Graduate Programs Department of Science.

Schuster, M., & Paliwal, K. K. (1997). Networks bidirectional reccurent neural. *IEEE Trans Signal Proces*, *45*, 2673-2681.

Şen, Zekâi (2016). Probabilistic innovative solar irradiation estimation. International Journal of Energy Research, (), –. doi:10.1002/er.3610

SEN, ZEKAI; OZTOPAL, AHMET; SAHIN, AHMET (2004). Solar Irradiation Estimation from Sunshine Duration by Geno-Fuzzy Partial Approach. Energy Sources, 26(4), 377–386. doi:10.1080/00908310490266814

Şenkal, O. (2016). Solar radiation modeling for Türkiye using atmospheric parameters with artificial neural networks. Çukurova Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi, 31(2), 179-185.

Şenkal, O., 2015. Solar radiation and precipitable water modelling for Türkiye using artificial neural networks. Meteorology and Atmospheric Physics, 127(4), pp.481-488. https://doi.org/10.1007/s00703-015-0372-6

Şenkal, O.; Şahin, M.; Peştemalci, V. (2010). The Estimation of Solar Radiation for Different Time Periods. Energy Sources, Part A: Recovery, Utilisation, and Environmental Effects, 32(13), 1176–1184. doi:10.1080/15567030902967850

Sepp Hochreiter and Juergen Schmidhuber (1997) “Long Short-Term Memory” Neural Computation 9(8):1735-1780, 1997.

Sepp Hochreiter, Yoshua Bengio, Paolo Frasconi, and Juergen Schmidhuber (2001) “Gradient flow in recurrent nets: the difficulty of learning long-term dependencies”. In S. C. Kremer and J. F. Kolen, eds., A Field Guide to Dynamical Recurrent Neural Networks. IEEE press.

Sharadga, H., Hajimirza, S. and Balog, R.S., 2020. Time series forecasting of solar power generation for large-scale photovoltaic plants. Renewable Energy, 150, pp.797-807. https://doi.org/10.1016/j.renene.2019.12.131

Sharma, A. and Kakkar, A., 2018. Forecasting daily global solar irradiance generation using machine learning. Renewable and Sustainable Energy Reviews, 82, pp.2254-2269. https://doi.org/10.1016/j.rser.2017.08.066

Sorkun, M.C., Paoli, C. and Incel, Ö.D., 2017, November. Time series forecasting on solar irradiation using deep learning. In 2017 10th international conference on electrical and electronics engineering (ELECO) (pp. 151-155). IEEE 10.1109/ICRERA47325.2019.8997040

SOZEN, ADNAN; OZALP, MEHMET; ARCAKLIOĞLU, EROL; GALIP KANIT, E. (2004). A Study for Estimating Solar Resources in Türkiye Using Artificial Neural Networks. Energy Sources, 26(14), 1369–1378. doi:10.1080/00908310490441935

Suttakul, P., Fongsamootr, T., Wongsapai, W., Mona, Y., & Poolsawat, K. (2022). Energy consumptions and CO2 emissions of different powertrains under real-world driving with various route characteristics. *Energy Reports*, *8*, 554-561.

Teke, A., Yıldırım, H.B. and Çelik, Ö., 2015. Evaluation and performance comparison of different models for the estimation of solar radiation. Renewable and sustainable energy reviews, 50, pp.1097-1107. https://doi.org/10.1016/j.rser.2015.05.049

Teke, Ahmet; Yıldırım, H. Başak (2014). Estimating the monthly global solar radiation for Eastern Mediterranean Region. Energy Conversion and Management, 87(), 628–635. doi:10.1016/j.enconman.2014.07.052

Ustun, I., Karakus, C. and Yagli, H., 2020. Empirical models for estimating the daily and monthly global solar radiation for the Mediterranean and Central Anatolia region of Türkiye. International Journal of Global Warming, 20(3), pp.249-

Vakitbilir, N., Hilal, A. and Direkoğlu, C., 2020, August. Prediction of daily solar irradiation using CNN and LSTM networks. In the International Conference on Theory and Applications of Fuzzy Systems and Soft Computing (pp. 230-238). Springer, Cham.https://doi.org/10.1016/j.apenergy.2019.113541

Van Rossum, G., & Drake Jr, F. L. (1995). Python reference manual. Centrum voor Wiskunde en Informatica Amsterdam.

Vatansever, C., 2019. Short-term solar irradiance forecasting with deep neural networks. Vickers, N.J., 2017., pp.R713-R715. https://doi.org/10.1016/j.cub.2017.05.064

Wang, F., Xuan, Z., Zhen, Z., Li, K., Wang, T. and Shi, M., 2020. A day-ahead PV power forecasting method based on LSTM-RNN model and time correlation modification under partial daily pattern prediction framework. Energy Conversion and Management, 212, p.112766.https://doi.org/10.1016/j.enconman.2020.112766

Wang, F., Xuan, Z., Zhen, Z., Li, K., Wang, T., & Shi, M. (2020). A day-ahead PV power forecasting method based on LSTM-RNN model and time correlation modification under partial daily pattern prediction framework. Energy Conversion and Management, 212, 112766. doi:10.1016/j.enconman.2020.11276.

Wang, F., Zhen, Z., Wang, B. and Mi, Z., 2017. Comparative study on KNN and SVM based weather classification models for day ahead short term solar PV power forecasting. Applied Sciences, 8(1), p.28. https://doi.org/10.3390/app8010028

Waskom, M et al. (2017). mwaskom/seaborn: v0.8.1 (September 2017). Zenodo. https://doi.org/10.5281/zenodo.883859.

Widerøe, B.M., 2019. Forecasting Electricity Production from Photovoltaic Solar Panels using Elastic Net Regularisation (Master's thesis, The University of Bergen).

Y. -Y. Hong, J. J. F. Martinez and A. C. Fajardo, "Day-Ahead Solar Irradiation Forecasting Utilizing Gramian Angular Field and Convolutional Long Short-Term Memory," in IEEE Access, vol. 8, pp. 18741-18753, 2020, doi: 10.1109/ACCESS.2020.2967900.

Yaniktepe, B. and Genc, Y.A., 2015. Establishing a new model for predicting the global solar radiation on a horizontal surface. International Journal of Hydrogen Energy, 40(44), pp.15278-15283. https://doi.org/10.1016/j.ijhydene.2015.02.064

Yıldırım, H. Başak; Teke, Ahmet; Antonanzas-Torres, Fernando (2017). Evaluation of classical parametric models for estimating solar radiation in the Eastern

Mediterranean region of Türkiye. Renewable and Sustainable Energy Reviews, (), S1364032117311784–. doi:10.1016/j.rser.2017.08.033

Yıldırım, H.B., Çelik, Ö., Teke, A. and Barutçu, B., 2018. Estimating daily Global solar radiation with graphical user interface in Eastern Mediterranean region of Türkiye. Renewable and Sustainable Energy Reviews, 82, pp.1528-1537. https://doi.org/10.1016/j.rser.2017.06.030

Yıldırım, H.B., Çelik, Ö., Teke, A. and Barutçu, B., 2018. Estimating daily Global solar radiation with graphical user interface in Eastern Mediterranean region of Türkiye. Renewable and Sustainable Energy Reviews, 82, pp.1528-1537. https://doi.org/10.1016/j.rser.2017.06.030

Yıldız, A. (2019). PREDICTING THE ENERGY PRODUCTION OF A ROOFTOP PV PLANT BY USING DIFFERENTIAL EVOLUTION ALGORITHM . Turkish Journal of Engineering , 3 (3) , 106-109 . DOI: 10.31127/tuje.466953

Yildiz, B. Y.; Şahin, M.; Şenkal, O.; Pestemalci, V.; Emrahoğlu, N. (2013). A Comparison of Two Solar Radiation Models Using Artificial Neural Networks and Remote Sensing in Türkiye. Energy Sources, Part A: Recovery, Utilisation, and Environmental Effects, 35(3), 209–217. doi:10.1080/15567036.2011.650276

Yildiz, C. and Acikgoz, H., 2021. A kernel extreme learning machine-based neural network to forecast very short-term power output of an on-grid photovoltaic power plant. Energy Sources, Part A: Recovery, Utilisation, and Environmental Effects, 43(4), pp.395-412. http://journaleras.com/index.php/jeras/article/view/182

Yildiz, C. and Acikgoz, H., 2021. A kernel extreme learning machine-based neural network to forecast very short-term power output of an on-grid photovoltaic power plant. Energy Sources, Part A: Recovery, Utilisation, and Environmental Effects, 43(4), pp.395-412. https://doi.org/10.1080/15567036.2020.1801899

Yu, Y., Cao, J. and Zhu, J., 2019. An LSTM short-term solar irradiance forecasting under complicated weather conditions. IEEE Access, 7, pp.145651-145666. 10.1109/ACCESS.2019.2946057

Yu, Y., Cao, J., & Zhu, J. (2019). An LSTM Short-Term Solar Irradiance Forecasting Under Complicated Weather Conditions. IEEE Access, 7, 145651–145666. doi:10.1109/access.2019.2946057.

Yu, Y., Cao, J., Wan, X., Zeng, F., Xin, J. and Ji, Q., 2018. Comparison of short-term solar irradiance forecasting methods when weather conditions are complicated. Journal of Renewable and Sustainable Energy, 10(5), p.053501. https://doi.org/10.1063/1.5041905

Zang, H., Cheng, L., Ding, T., Cheung, K.W., Wang, M., Wei, Z. and Sun, G., 2020. Application of functional deep belief network for estimating daily global solar radiation: A case study in China. Energy, 191, p.116502. https://doi.org/10.1016/j.energy.2019.116502

Zekai Şen (1998). Fuzzy algorithm for estimation of solar irradiation from sunshine duration. , 63(1), 39–49. doi:10.1016/s0038-092x(98)00043-7

Zhang, Y., Wei, J., Zhu, Y. and George-Ufot, G., 2020. Untangling the relationship between Corporate Environmental Performance and Corporate Financial Performance: The double-edged moderating effects of environmental uncertainty. Journal of Cleaner Production, 263, p.121584. https://doi.org/10.10 16/j.jclepro.2020.121584

Zhou, H., Zhang, Y., Yang, L., Liu, Q., Yan, K. and Du, Y., 2019. Short-term photovoltaic power forecasting based on long short term memory neural network and attention mechanism. Ieee Access, 7, pp.78063-78074. 10.1109/ACCESS.2019.2923006

Zimmer-Gembeck, M.J. and Helfand, M., 2008. Ten years of longitudinal research on US adolescent sexual behavior: Developmental correlates of sexual intercourse, and the importance of age, gender and ethnic background.

Developmental review, 28(2), pp.153-224. https://doi.org/10.1016/j.dr.2007.06.001

BIOGRAPHY

Eman Ahmed Erfan was born in Alexandria. She received a BSc degree from the Department of Electrical and Control Engineering from AAST, Egypt, in 2017.

She worked as a teaching assistant in Arab Academy for Science, Technology, and Maritime Transport from 2017 to 2018.

Her research interests include sustainable energy, artificial intelligence, solar irradiation prediction, and industrial automation.

APPENDICES

Table Related daily time-horizon prediction models in the literature 1

Reference	Used model(s)	Input variable(s)	Target values	Evaluation metrics
Barutc et al. 2015	WRF-ARW, MISO - ANN	GHI, φ, λ, α	3 days ahead GHI	MBE, rMBE, RMSE, rRMSE
Güçlü et al .2015	HarLin, ANFIS	SD	Daily GHI	MAE
Yaniktepe and Genc, 2015	Third order polynomial model	Historical GHI	global solar radiation on horizontal surface	MABE, MAPE, RMSE
Teke et al. 2015	ANN, FL	geographical and meteorological data	global solar radiation	MAPE, RMSE, MBE, MPE, R2
Gairaa et al. 2016	ARMA-ANN	historical daily global solar radiation	daily global solar radiation	R2, nMBE, MPE, MBE, RMSE, nRMSE, t-stat
Bakirci, 2017	Empirical Model	SD	Daily GHI	MBE, RMSE, MPE, t-stat
Cetinkaya et al. 2017	LM	RH, Tmin, Tmax, GSR, WS, WSmax	solar power production	MSE
Ekici and Teke, 2018	Buckingham Theorem	V, DP, α, Tmax, Tmin	GSR	MBE, RMSE, MPE, and NSE

Table 1 Continue

Reference	Used model(s)	Input variable(s)	Target values	Evaluation metrics
Gumus et al. 2018	EWMA	GSR, SD	GSR	R2, and MAPE
Yıldırım et al. 2018	two proposed models	SD, T, RH, PR	Daily global solar radiation	MBE, MAE, RMSE, CV-MAE, CV-RMSE, and R2
Yıldırım et al. 2018	ANN, Angström-Prescott	T, SD, φ, λ, α, RH	GSR	SSE, R2, RMSE
Kaba et al. 2018	CNN	GSR, CC, SD, Tmin, Tmax, φ, λ, α	GSR	R, MAE, RMSE
Jahani et al. 2018	MLP, MLP-GA	SD, Tmax, Tmin, KT	GSR	R2, RMSE, MBE
Abdulhamid et al. 2018	FL	T, RH	GHI	MAE, MSE, RMSE, MAPE R2
Khosravi et al. 2018	GMDH, MLFFNN, ANFIS, ANFIS-PSO, ANFIS-GA, ANFIS-ACO	M, D, T, Tmax, Tmin, P, RH, WS, KT, φ, λ	GSR	R2, MSE, RMSE

Table 1 Continue

Reference	Used model(s)	Input variable(s)	Target values	Evaluation metrics
Kaba et al. 2018	CNN	GSR, CC, SD, Tmin, Tmax, φ, λ, α	GSR	R, MAE, RMSE
Marzouq et al.2018	MLP, K-NN	Y, TS, M, D, RHmax, RHmin, Tmax, Tmin, R, WS, WD, SD	GHI	R2, RMSE, MABE
Akbaba et al. 2018	MLP	SD, φ, T, Ho	GSR	RMSE, MAE, R2, EVS, FIE_H, FIE-O, FIE-OH, LINCC
Aji et al. 2018	IDW	GSR	GSR	MAPE
Aji et al. 2018	SW-SVR	T, RH, Ws, P	GSR	MAPE, R2
Wang et al. 2018	SVM, KNN	D, δ, φ, hour angle	GHI	R2, RMSE, SSE, Adj-SSE
Boussaada et al. 2018	NARX		GHI	RMSE, DMPE
Lotfinejad et al. 2018	BNN, GRNN, ANFIS	SD, T, Ws, RH, R	GSR	R2, MPE
Lima et al. 2018	MLP	T, GSR	HGSR	MAPE
Mohammad et al. 2018	MLP	Tmax, SD, RH, Ws	GSR	R2, MSE

Table 2-1 Continue

Reference	Used model(s)	Input variable(s)	Target values	Evaluation metrics
Durrani et al. 2018	ANN	T, P, RH, CC, SD, Ws, WD, zenith angle, azimuth angle Ho, D	GHI	MAE, MAPE, RMSE, R
Qing et al. 2018	LSTM, MLR, BPNN	D, H, M, T, DP, RH, Ws, weather type, V	GHI	RMSE
Yu et al. 2018	ENN, BPNN, RBPNN	T, M, WEATHER TYP	GHI	MAPE
Sharma et al. 2018	FoBa, leap Forward, Spikeslab, Cubist model	GHI	GHI	R2, RMSE, R, ACCURAC Y
Abdulhamid et al. 2018	FL	T, RH	GHI	MAE, MSE, RMSE, MAPE, R2

Table 1 continue

Reference	Used model(s)	Input variable(s)	Target values	Evaluation metrics
Demirdelen et al. 2019	ANN, ANN-PSO, ANN-FA	T, Tp	GHI	R2, MAPE
Ünlü, 2019	SVR, MLP, and LSTM	Historical EE	EE	R2, MAE, MSE, and RMSE
Kaygusuz, 2019	GWR	φ, RH, CC	Annual GSR	R2, and RMSE
Feng et al. 2019	ANN, MEA-ANN, RF, WNN, HS, BC, Fan, Jahani	PR, RH, Tmin, Tmax, φ, λ, α Ws	GHI	R2, RMSE, rRMSE, MBE, Ns, MSE, MAE
Antonopoulos et al. 2019	MLP, MLR	Tmax, Tmin, T, RH, Ws	GSR	R2, RMSE, EF
Zang et al.2019	SVR, GPR, ANFIS	φ, λ, α, Tmax, Tmin, Tavg, SD, Ws, RH	GSR	MAE, RMSE, MAPE, R
Ghimire et al. 2019	ANN, SVR, GPML, GP, ARIMA, TSFS, TMHS, TMCH, TMBC, TMGO, TMLI	Tmax, Tmin, R, E, P, Tmax, RHmax, RHmin, Tmin, RH	GSR	R, RMSE, MAE
Fan et al. 2019	MLP, RBF, GRNN, ELM, SVM, LSSVM, M5T, RF, GBDT, XGBoost, ANFIS, MARS	SD, Ho, φ, λ, α	GSR	MBE, RMSE, nRMSE, R2, t-statistic, U95

Table 2-1 continue

Reference	Used model(s)	Input variable(s)	Target values	Evaluation metrics
Ghimire et al. 2019	LSTM, CLSTM, CNN, DNN, MLP, DT, GRU, RNN	Tmax, Tmin, Tavg, SD	GSR	MAE, RMSE, MAPE, KGE, APB
Yu et al. 2019	LSTM, ARIMA, SVR, BPNN, CNN, RNN	T, RH, ZENITH, DP, CT, WS, WD, PR	GHI	MAE, MAPE, RMSE, R2
Sharadga et al. 2019	ARMA, ARIMA, SARIMA, LSTM, BI-LSTM, LRNN-LM, MLP-LM, LRNN- BR, FCM, FF-LM, MLP- BR, FF-BR	POWER	POWER	R, RMSE, Ti
Alsharif et al. 2019	SARIMA, ARIMA	Historical GSR	GSR	R2, RMSE
Husein et al. 2019	FFNN, LSTM	DT, DP, RH	GHI	RMSE, MAE, S

Table 1 continue

Reference	Used model (s)	Input variable(s)	Target values	Evaluation metrics
H. Zhou et al. 2019	ALSTM, PM, ARIMAX, MLP, LSTM	GHI	GHI	MAPE, RMSE, and MAE
Rocha et al. 2019	MLP	D, Tmax, Tmin, PR, CC, Ho, RH, E, WD	GSR	MAPE, Nrmse
Goksu et al. 2020	ANN	T, RH, GSR	POWER	R, MSE
Özdemir et al. 2020	ANN, Bi-LSTM	Historical EP	EP	RE
Kaygusuz, 2020	GWR	GHI, SD, φ, RH and CC	EP	R2 and RMSE
Colak et al. 2020	GWO-MLP, ALO-MLP, WOA-MLP	T, RH, GHI, DHI	Daily PE	R2, MAE and MAPE
Ustun et al. 2020	Empirical models	Ho, TS, CC, SD, T, RH	GHI	MBE, MPE, RMSE, R2
Yildiz et al. 2020	LM, ELM, KELM	T, RH, GSR, Ws, P, CC, R	POWER	R, MAE, RMSE
Goksu et al. 2020	ANN	T, RH, GSR	EP	R, MSE
Ağbulut et al. 2020	SVM, ANN, DL, K-NN	DHI, DNI, TI, and Tp	EP	R2, RMSE, MBE, t-stat, MABE

Table 1 continue

Reference	Used model (s)	Input variable(s)	Target values	Evaluation metrics
Vakitbilir et al. 2020	CNN, LSTM	GHI, Tmax, Tmin, Ws, WD, RH, P	GHI	MAE, MAPE, RMSE, R2
Mishra et al. 2020	WT-LSTM-dropout	T, V, CC, DP, RH, P, WS, SE	GSR	RMSE, MAPE, MAE, R2
Liu et al. 2020	CNQR, SVM-FFA	KT, SD, T, RH, φ, λ, α	Hd	R2, MAPE, MBE, rRMSE
Tao et al. 2020	MLR, ARIMA, EML	Tmax, Tmin, Tavg, RH, Ws, SD	GSR	R, RMSE, MAE, Willmott's Index
Ojo et al. 2020	MLR, NARX	Tmin, Tmax, WS, RH	GSR	R, R2, RMSE, MBE, t-test
Aslam M et al. 2020	FFNN, RNN, LSTM, GRU	Historical GHI	GHI	RMSE
Brahma et al. 2020	LSTM, GRU, CNN, BI-LSTM, ATTENTION-LSTM	GHI	GHI	MSE, RMSE
Wang et al.2020	LSTM, BPNN, SVM, Persistent	DNI, T	DNI	MAE, RMSE, R
Ustun et al. 2020		Ho, TS, CC, SD, T, RH	GHI	MBE, MPE, RMSE, R2

Table 1 continue

Reference	**Used model (s)**	**Input variable(s)**	**Target values**	**Evaluation metrics**
Yildiz et al. 2020	LM, ELM, KELM	T, RH, GSR, Ws, P, CC, R	POWER	R, MAE, RMSE
Ağbulut et al. 2021	SVM, ANN, DL, K-NN	Tmin, Tmax, CC, Ho, SD	GSR	R2, RMSE, rRMSE, MBE, MABE, t-stat, MAPE
Ağbulut et al. 2021	SVM, ANN, K-NN	Tmin, Tmax, CC, Ho, SD	GSR	R2, RMSE, rRMSE, MBE, MABE, t-stat, MAPE
Boubaker et al. 2021	LSTM, GRU, BiLSTM, BiGRU, CNN, CNN-LSTM, CNN-BiLSTM	Historical GHI	GHI	R, RMSE, MAPE
AL-Rousan et al.2021	NARX, SARIMA, ANFIS	GHI	GHI	R2, RMSE, MSE, MBE, MAE
Bounoua et al.2021	MLP, Boosting, Bagging, RF, Empirical models	D, T, RH, WS, WD, P, H0, Tmax, Tmin, Tdiff, Tmin, Tmax	GSR	R, nRMSE, nMAE

Table 2 Related monthly time-horizonprediction in the literature.

Reference	Used model(s)	Input variable(s)	Target values	Evaluation metrics
Bakirci, 2015	Empirical models	SD, GSR, GHI	Monthly DHI	MBE, RMSE, MAPE
Şenkal, 2015	ANN	α, M, T, RH, P, PR	Monthly GHI	RMSE
Citakoglu, 2015	ANFIS, ANN, MLR	M, T, RH	Monthly GHI	MAE, MARE, RMSE, R2
Ozturk, 2015	Empirical models	S	Monthly GHI	R, MBE, MPE, RMSE
Çelik et al. 2016	LM	SD, T, φ, λ, α, M	monthly GSR	MAPE, R2
Arslanoglu, 2016	Empirical models	meteorologica l data	Monthly GHI	R2, MABE, MPE, MAPE, MBE, RMSE, and t-stat
Bakirci et al. 2018	six new models	K, GHI	Monthly DHI	R2, MBE, RMSE, MAPE, and t-stat
GUMUS and KILIC, 2018	EWMA	SD	MONTHL Y GHI	R2, R, MAPE, and MAE
Gungor et al. 2018	Empirical models	SD	Monthly GHI	R2, MBE, RMSE, and t-stat
Mohammed and Mengüç, 2018	Empirical models	S, T, WS	Monthly GHI	RMSE, MBE, MABE, MPE, and MAPE

Table 2 Continue

Reference	Used model(s)	Input variable(s)	Target values	Evaluation metrics
Arslanoglu, N. (2016)	Empirical models	GHI, SD	Monthly DHI	MABE, MAPE, RMSE, MBE, MPE, and t-stat
Karaveli, and Akinoglu (2018)	MCDL, MCDQ, MCGD	historical global and diffuse solar irradiation	Monthly GHI, DHI	MBE, and RMSE
Gürel et al. 2020	ANN, Angstrom-models, HW, and RSM	P, RH, WS, T, and SD	Monthly GHI	R2, MBE, RMSE
Mellit et al. 2021	LSTM, BI-LSTM, GRU, BI-GRU, CNN, NAR-ENN, stacked-LSTM	Monthly GHI	Monthly EP	R, RMSE, MAE, MAE, and TI
Kurtgoz and Deniz, 2016	ANN	GHI, SD, δ, WS, T, Ts	Monthly GSR	RMSE, MAPE, and R

Table 3 Related hourly time-horizon prediction in literature.

Reference	Used model(s)	Input variable(s)	Target values	Evaluation metrics
Akarslan et al. 2015	MDLPF	Hourly GSR, and GHI	Hourly GHI	RMSE
Hocaoglu et al.2017	Adaptive model	Historical GSR	Hourly GSR	R2, MBE, RMSE, MAPE, and t-stat
Akarslan et al. 2018	Five new models	SD, GSR, GHI, K	Hourly GSR	R2, RMSE, nRMSE, MBE
Ayvazoğluyüksel, and Filik, 2018	eleven different models	Hourly and Daily GSR	hourly GSR	nMAE, and WMAE
Khosravi et al. 2018	MLFFNN, RBFNN, SVR, FIS, ANFIS	TI, T, P, WS, RH	Hourly GSR	R, MSE, RMSE
Bouzgou et al. 2018	ELM, MLP	Historical GHI	Hourly GHI	R2, RMSE, nRMSE, MAPE
Incecik et al. 2019	WRF	NCEP-GFS data	Hourly GHI	MBE, RMSE, and rRMSE

Table 3 Continue

Reference	Used model(s)	Input variable(s)	Target values	Evaluation metrics
Basaran et al. 2019	SVR, ANN, and DT	5 years meteorological data	Hourly GSR	R2, and RMSE
Çoban and Onar, 2020	RF, MLP, GP, SVR, RT, PRUNED-RT, BOOST-RT, BAGGED-RT, ARIMA	HGHI	Hourly GHI	nRMSE, MAE
Yildiz and Acikgoz, 2020	KELM	GSR, H, T	Hourly EP	R, MAE, and RMSE
Marzouq et al. 2020	RT, SP, RF, HAEANN 6	φ, λ, α,	Hourly GSR	nRMSE, nMAE
Jallal et al. 2020	ENN	T, GSR	Hourly GSR	R, MSE
Jallal et al. 2020	MLP	HHGSR	Hourly GSR	MSE
Ozbek et al. 2021	LSTM, ANFIS, FCM, ANFIS–GP	Historical GSR	Hour-ahead EE	RMSE, MAE, and R
ALSAFADI et al.2020)	ANN, RT, SVR	solar time, solar hour angle, D, λ, φ, RH, T, P	Hourly GSR	R, RMSE, MABE, MBE
Çiçek et al. 2019	Statistical model	T, WS	Hourly GHI	AE

Table 4 Related MSc and PhD inliterature.

Reference	Used model(s)	Input variable(s)	Target value(s)	Evaluation metrics
YILDIRIM,2016	ANN, and Angstrom-Prescot	SD, RH, PR and T	Daily GHI	MBE, MAE, RMSE, and R2
Yumru, 2016	PLC-LSTM	CHI and SP	Day-ahead GSR	WMAPE, Pbias, DWMAPI med, DWMAPEsd, DPBiasmed, and DPBiassd
Bitirgen, 2018	ANN	WS, P, WS, P, T	Hourly GSI	MBE, MAPE, RMSE, t-stat
Kasht, 2018	SVM	T, K	Hourly GHI	MBE and RMSE
Sorkun, 2018	LSTM and GRU	T, P, H, N, WS, WD, KT, R, and GSR	Hourly GSR	MSE and NRMSE
Akbaba, 2019	MLP	SD, Dl, T, φ	Daily GHI	RMSE, MAE, R2, EVS, FIE (0.5), FIE (1), and FIE (1.5)
Vatansever, 2019	SW, K-NN, SVM, and RF	T, RH, WS, WSmax, WD, P, PR	15-minute ahead GSR	MAE, MSE, MAPE, and R2

Table 4 continue

Reference	Used model (s)	Input variable(s)	Target values	Evaluation metrics
Kaya, 2019	ANN-GA, SVM	SDastr, SDm, SDd, Tamax, Tadmx, Ta, RD, Tamin, Tadmin, RR, RH, φ, λ, α	Daily GHI	MAPE, MAE, RAE, RRSE, and RMSE
Ghimire, 2019	ANN, ELM, SVR, DBN, DNN, CNN, LSTM	Historical GSR	Monthly GSR	RMSE, MAE, R, WI, EI, APB
Widerøe, 2019	ENR	T, RH, CC, GSR, TS	EP	MSE
SARKEZ, 2020	LM	RH, P, T	Monthly GHI	R, and MSE
Talakoobi, 2020	LSTM, LSTNET, TCN	GHI, RH, T, R	Monthly EE	R, RSE
Kavakçı, 2021	ANN, RF, KNN, SVR, and LGBM	M, Historical GSR, GHI, DNI, DHI	Hourly GSR	MAE and RMSE
QADEM, 2021	ANN	SD, WSmax, WD, H, Tmax	Daily GHI	RMSE, MABE, MBE, and t-statistic

Table 4 Continue

Reference	Used model(s)	Input variable(s)	Target values	Evaluation metrics
Petros, 2021	RF, GBR, XBOOST, LSVR, RBF-SVR, LSTM, BiLSTM, CNN	T, H	Daily GSR	MAE, MSE, RMSE, R2, MAPE, MMR, MASE
Paiva, 2021	MGGP	GHI, T, P, RH, WS	Hourly GHI	RMSE, MAE, OMAE, EMAE

The results of the previous employed solar estimation models in Adana are compared in table 5.

Table 5 A comparison between previous employed models for Adana.

Reference	Used model(s)	Input variable(s)	Target value(s)	Evaluation metrics
Ögelman et al. 1984	Empirical models	SD, RH, WS, T, Ts	Monthly GSR	Graphically
Akinoğlu et al. 1990	Empirical models	SD	Monthly and daily GSR	MBE=0.087, RMSE=1.16, MABE=0.885
Şen, 1998	FL	SD	Monthly GHI	Graphically
Toğrul et al. 2000	Empirical model	SD	Daily GHI	MBE= -0.572, MABE= 1.373
Oğulata, 2002	Empirical model	GHI, DHI, DNI	Hourly and monthly GHI, DHI, DNI	Graphically
Türk Toğrul and Toğrul, 2002	Empirical model	Dl	Monthly GHI	MBE= -0.034 RMSE= 0.213
Sözen et al. 2004	ANN	φ, λ, α, M, SD, T	Monthly GSR	R2= 0.9987, RMS=2.002, MAX_ERROR=6.096, MAPE=2.42, COV=0.037
Sözen et al. 2004	ANN	φ, λ, α, M, SD, T	Monthly GSR	R2=99.8937%, MAPE<6.7

Table 5 Continue

Reference	**Used model(s)**	**Input variable(s)**	**Target value(s)**	**Evaluation metrics**
Sözen et al. 2004	ANN	φ, λ, α, M, SD, T	Monthly GSR	RMS= 2.0024, R2=0.9988, COV= 0.009, MAPE=2.41
Şen, 2004	Geno-fuzzy	SD	Daily GHI	RMSE= 0.050
Bulut and Büyükalaca, 2007	trigonometric function	D	Daily GHI	MAE= 2.57, MRE=20, RMSE=3.36, R=0.87
Şahin, 2007	Empirical model	SD	Monthly GHI	R2= 0.94, MBE= 0.361, RMSE= 12.627, RE= 9.556
Şenkal et al. 2009	ANN	φ, λ, α, M, DHI, DNI	Monthly GHI	R=98.15, RMS= 4.21
Bakirci, 2009	Empirical equations	SD, Daily-GHI	Monthly GSR	R=0.9818, RMSE=1.0218, MRE=0.0687, MABE=0.8931, MBE=0.8911
Bulut et al. 2009	Finkelstein-Schafer	Daily-GHI	Daily GHI	Graphically

Şenkal et al. 2010	ANN	φ, λ, α, M, T	Monthly GHI	R2=99.75, RMS= 0.0144, COV= 0.0075, MAPE=0.0260
Şenkal, 2010	GRNN	φ, λ, α, T, ε3, ε4	Monthly GHI	MSE=0.0075, RMSE= 0.2598
Koca et al. 2011	ANN	φ, λ, α, M, CC	Monthly GHI	R2=0.9945, COV=0.0529, RMSE= 7.7940
Ozgoren et al. 2012	MNLR, ANN	Tmin, Tmax, T, TS, RH, WS, R, P, VP, CC, SD	Monthly GSR	MAPE= 4.55, R2= 0.9868
yildiz et al. 2018	ANN	φ, λ, α, M, T	Monthly GHI	R2= 82.37%, MBE=2.67, RMSE= 3.81

Table 5 Continue

Reference	**Used model(s)**	**Input variable(s)**	**Target value(s)**	**Evaluation metrics**
Şahin et al. 2014	ELM, ANN	T, RH, SD, CC	Daily GSR	RMSE= 0.060, MBE= 0.004, R2=0.980
Kisi, 2014	FG, ANN, ANFIS	φ, λ, α, M	Monthly GHI	RMSE= 4.29, MAE=4.04, R2=0.987
Teke and Başak, 2014	Linear, quadratic and cubic equations	SD	Monthly GSR	MAPE= 5%
Güçlü et al. 2014	Angstrom–Prescott, FL-ANFIS	SD	Daily GHI	R2= 0.9681, MAE=0.4602
Bakirci and Kadir, 2015	Empirical models	SD, GSR, GHI	Monthly DHI	MBE=0.00015, RMSE=0.0002, MABE=0.0002
Çelik et al. 2015	ANN-LM	T, SD, M, α	Monthly GSR	R2= 99.8%, MAPE= 2.8087%
Güçlü et al. 2015	Har-Lin, ANFIS	SD	Daily GHI	R2= 0.9681, MAE=0.4602
Sahan et al.2016	ANN	φ, λ, α, M, T, RH	Monthly GHI	R2=0.971, RMSE=1.007, MSE=1.013, MABE=0.814, MAPE=0.056
Şenkal, 2016	ANN	T, P, RH, PR, M	Monthly GHI	MBE=-0.6929, RMSE=

				2.4002, R2= 0.9599
Şen, 2016	CDFs	SD	Daily GHI	MAPE= 2.21%
Yıldırım et al. 2017	classical parametric models	T, Tmax, Tmin, PR, SD, RH	Daily GHI	R2= 0.965, RMSE= 1.183, MAE= 0.947, MBE= 0.142
Yıldız, 2019	DE	Daily-EE	Monthly EE	APE= 7.4 %
Kaba et al. 2018	ANN	AF, GHI, SD, CC, Tmax, Tmin	Daily GSR	R2 =0.960, MAE=1.04, RMSE= 1.04
kurtgoz, 2018	ANN, MLRA, and ANFIS	SD, RH, WS, T, Ts	Monthly GSR	RMSE=0.83, MAPE=4.01, and R=0.9902

Table 5 Continue

Reference	Used model(s)	Input variable(s)	Target value(s)	Evaluation metrics
Kisi et al. 2019	DENFIS, MARS, M5T and LSSVR	Tmax, Tmin	Monthly GHI	RMSE=1.45, MAE= 1.12, NSE=0.938
Ghazvinian et al. 2019	M5T, GP, SVR-PSO, SVR-IPSO, SVR-GA, SVR-FFA, MARS	Tmax, Tmin, SD, WS, RH	Daily GHI	MAE=4.01, RMSE= 5.02, MBE= 0.002, NSE= 0.991, TIME= 26, R2= 0.9894
Ustun et al. 2020	ANFIS, Angström-Prescott, dependency	T, RH, CC, PR	Daily and monthly GSR	MBE=3.021, RMSE= 1.6836, R2= 0.705
Kisi et al. 2020	ANN, ELM, RBF, WANN, WELM, WRBF	Tmax, Tmin, SD, WS, RH	Monthly GSR	RMSE=10.08, NSE= 0.947, R2=0.987
Goncu et al. 2021	ANN	WS, WD, RH, P, T	Hourly GSR	R2= 0.87313, MSE= 58.262 MW/m2