REPUBLIC OF TURKEY

ADANA ALPARSLAN TÜRKEŞ SCIENCE AND TECHNOLOGY UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

DEPARTMENT OF INDUSTRIAL ENGINEERING

ANALYSIS OF METEOROLOGICAL DATA BY ARTIFICIAL INTELLIGENCE TECHNIQUES AND ESTIMATION OF GLOBAL SOLAR RADIATION

ÖZGE KAYA

MASTER OF SCIENCE

ADANA 2019

REPUBLIC OF TURKEY
ADANA ALPARSLAN TÜRKEŞ SCIENCE AND TECHNOLOGY UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
DEPARTMENT OF INDUSTRIAL ENGINEERING

ANALYSIS OF METEOROLOGICAL DATA BY ARTIFICIAL INTELLIGENCE TECHNIQUES AND ESTIMATION OF GLOBAL SOLAR RADIATION

ÖZGE KAYA
MASTER OF SCIENCE

SUPERVISOR
ASSIST. PROF.DR. FATİH KILIÇ
CO-SUPERVISOR
ASSOC. PROF. DR. İBRAHİM HALİL YILMAZ

ADANA 2019

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

[Signature]

Özge KAYA

ABSTRACT

ANALYSIS OF METEOROLOGICAL DATA BY ARTIFICIAL INTELLIGENCE TECHNIQUES AND ESTIMATION OF GLOBAL SOLAR RADIATION

Özge KAYA
Department of Industrial Engineering
Supervisor: Assist. Prof. Dr. Fatih KILIÇ
Co-Supervisor: Assoc. Prof. Dr. İbrahim Halil YILMAZ
September 2019, 49 pages

In addition to being the main source of life in the earth, the Sun planet has an endless source of renewable energy. Turkey is situated at the sunny belt of the earth i.e. 36−42° North latitudes and 26−45° East longitudes. The global irradiation measurement used in the calculation of solar thermal or electrical energy systems is very significant for the facility location in our country. The global radiation value is equal to the sum of the beam and the diffuse rays of the radiation from the sun. Its measurement is not as easy as the measurement and the estimation of other meteorological data (such as air temperature, humidity) and long-terms are required. In the literature, various corelations have been derived by use of meteorological radiation data and the predictions for distinct geographic locations were conducted. However, it is seen that the predictions made with these corelations involve high deviations. In this thesis, meteorological data of various regions in Turkey were taken from Meteonorm software and global solar radiation forecasts were made with the proposed Artificial Neural Networks (ANN) and Support Vector Machine (SVM) by optimized the Genetic Algorithm. Generated data from the software were divided into training and testing data sets, and the proposed ANN and SVM models were tested with the data. Better results were obtained from the proposed ANN model as compared with the SVM and the existing models in the literature. Mean average percentage error was obtained to be 3.44% for testing data by the proposed ANN model which utilizes the most relevant input variables determined by the genetic algorithm.

Keywords: ANN, SVM, Global Irradiation (GI), Genetic Algorithm

ÖZET

YAPAY ZEKÂ TEKNIKLERI ILE METEOROLOJIK VERILERIN ANALIZI VE GLOBAL GÜNEŞ IŞINIM TAHMİNİ

Özge KAYA
Endüstri Mühendisliği Anabilim Dalı
Danışman: Dr. Öğr. Üyesi Fatih KILIÇ
İkinci Danışman: Doç. Dr. İbrahim Halil YILMAZ
Eylül 2019, 49 sayfa

Dünyanın ana yaşam kaynağı olmasının yanı sıra Güneş gezegeni, bitmek tükenmek bilmeyen yenilenebilir enerji kaynağına sahiptir. Türkiye dünyanın güneş kuşağının içerisinde olan 36−42° kuzey enlemleri ve 26−45° doğu boylamlarında bulunmaktadır. Güneş enerjisi veya elektrik enerjisi sistemlerinin hesaplanmasında kullanılan küresel ışınım ölçümü ülkemizdeki enerji tesis yeri için çok önemli bir veridir. Küresel ışınım değeri, güneşten gelen ışınımın direkt (beam) ve gökyüzünde dağılan (diffuse) ışınların toplamına eşittir. Bu değerin ölçümü, diğer meteorolojik verilerin (hava sıcaklığı, nem gibi) ölçümü ve tahmini kadar kolay değildir ve uzun yıllar ölçümlerinin yapılması gerekmektedir. Literatürde, meteoroloji ışınım verileri kullanılarak çeşitli ilişkiler türetilmiş ve bu metotlar ile farklı coğrafik lokasyonlar için ölçüm verisi tahmini yapılmıştır. Fakat, bu ilişkilerle yapılan tahminlerde yüksek sapmalar içerdiği görülebilmektedir. Bu tez çalışmasında, Türkiye'deki çeşitli bölgelerin meteorolojik verileri Meteonorm yazılımından alınmış ve Genetik Algoritma ile optimize edilerek önerilen Yapay Sinir Ağları (YSA) ve Destek Vektör Makinesi (DVM) ile küresel güneş ışınımı tahminleri yapılmıştır. Meteonorm yazılımdan alınan veriler test ve eğitim verisi olarak ayrılarak önerilen ANN ve DVM modellerin performansı değerlendirilmiştir. Önerilen ANN modeli ile DVM ve literatürde mevcut modellerden daha iyi sonuçlar elde edilmiştir. Test verisi için Genetik algoritma ile girişleri optimize edilen ANN modeli ile elde edilen Ortalama Mutlak Yüzde Hata değeri %3.44 olmuştur.

Anahtar Kelimeler: YSA, DVM, Küresel Güneş Işınımı (KGI), Genetik Algoritma

“Dedication to my family…”

ACKNOWLEDGEMENTS

I would like to express my sincere gratefulness to my supervisor, Assist.Prof. Dr. Fatih KILIÇ for all her understanding, guidance, support and suggestions during the study.

I am very grateful to Assoc. Prof. Dr. Tolunay GÖÇKEN and. Prof. Dr. Ali KOKANGÜL for accepting to read and review this thesis and for their valuable suggestions and remarks.

I would like to thank Assoc. Prof. Dr. İbrahim Halil Yılmaz for their worthwhile assistance and supports.

Lastly, I am very grateful to my family and dearest friends for their enormous support and encouragement.

This thesis was financially supported by Adana Alparslan Türkeş Science and Technology University's Scientific Research Projects Coordination Unit with project number of 18103034.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

ANN	: Artificial Neural Networks
SVM	: Support Vector Machines
GA	: Genetic Algorithm
GI	: Global Irradiation
MLP	: Multi-Layer Perceptron
WEKA	: Waikato Environment for Knowledge Analysis
ρ	: Correlation coefficient
MAE	: Mean Absolute Percentage Error
RMSE	: Root Mean Square Error
RAE	: Relative Absolute Error
RRSE	: Root Relative Square Error

1. INTRODUCTION

1.1. Overview

Turkey's power consumption rose by 2.2% in 2018 compared to the previous year, 303.3 billion kWh, while electricity generation increased by 2.2% compared to the previous year and amounted to 303.9 billion kWh (Energy, 2019). 37.3% of Turkey's electricity production was from coal, 29.8% from natural gas and 19.8% from hydraulic energy in 2018 when 6.6% from wind, 2.6 % from the sun and 2.5 % were obtained from geothermal energy (Energy, 2019). Turkey generally imports the large part of its energy. Therefore, renewable energy is very important when taking into consideration the growth in primary energy demand and dependence on foreign sources.

Our earth has a number of renewable energy forms such as solar, wind, geothermal heat etc. Its major resources are wind and solar energies. The Sun is the only heat source of our Earth. Its giant power, which is about 1.75×10^5 TW (Kalogirou, 2009). Solar energy is one of renewable energy resources which can be managed to generate useful energy (Yılmaz, 2018b; Yılmaz et.al, 2018; Yılmaz and Mwesigye, 2018; Yılmaz and Söylemez, 2014). A part of the radiation given off by the Sun is lost in the atmosphere. The rest of it is known as global irradiation (GI) which boosts solar energy applications (Yılmaz, 2018a, 2018c, 2018d; Yılmaz and Söylemez, 2012). Solar energy applications are widely used in agriculture, industry and to meet the energy demand of cities. One of the solar energy applications is photovoltaic (PV) panels which are used to produce daily energy need. PV panels are preferred in buildings, agricultural fields and the areas where are not supplied with electricity. Since they can support the demanded power only enlarging the panel area. Firstly, PV panels are expensive and depreciation time of them must be calculated. GI is one of the parameters affecting the performance of PV panels the magnitude of GI depends upon the geometric relationships between the Sun and the Earth and the meteorological occasions happening in the atmosphere. Its measurement is performed by a pyranometer which is widely used equipment in meteorological stations to monitor the solar potential of a specific location. Setup, operating and maintenance costs of monitoring the solar potential are expensive thus there is not enough solar radiation data in many meteorological stations. Temperature and pressure, precipitation, humidity, wind direction and speed are typically measured in the meteorological stations of

Turkey. Empirical models and machine learning techniques can be used to estimate GI where it is not measured. Empirical models can be preferred for estimating GI if meteorological and geographical parameters are available. Furthermore, sunshine duration among these parameters has a strong relationship with GI.

Few studies have been made for either estimating the solar energy potential of Turkey or monthly average solar radiation for different locations in Turkey. Monthly average GI predictions were made for the locations of Sivas (Gurlek and Sahin, 2018), Isparta (Günoğlu et.al, 2011), seven cities from Mediterranean region of Anatolia (Koca et.al, 2011), 31 stations (Ozgoren et.al, 2012), 17 stations (Sozen et.al, 2004; Sözen et.al, 2004), 12 cities (Sözen et.al, 2005), 19 stations (Şenkal, 2010), Adana and İzmir (Şenkal et.al, 2010), 5 stations (Şenkal, 2015) and 68 stations (Yildiz et.al, 2013).

As clearly inferred from the literature review, the GI prediction has not been studied so far for the case of Turkey using the Artifial Neurol Network (ANN) and Support Vector Machine (SVM) by an optimized Genetic Algorithm (GA). The main contribution of this study is the prediction of GI applying the machine learning techniques by using the generated data from Meteonorm software.

1.2. Motivation of The Study

An accurate knowledge of the solar radiation data at a particular geographical location is of vital importance for the development of solar energy devices and for estimates of their performances. In this thesis, we aim to evaluate the solar potential of certain cities of Turkey using the ANN and SVM by the optimized GA based on GI prediction model. The reason I focus on ANN and SVM, they are effective tools for modeling nonlinear systems and require less input. Since neural networks require no prior assumptions concerning the data relationships, they become a useful tool to tackle the solar radiation modeling. It is also clear that the ANN method is suitable in regions where very large distances exist between meteorological stations in this thesis.

1.3. Thesis Outline

This thesis is organized as follows.

In Chapter 2, a literature review of GI prediction is performed. The rational and the history of these approaches are briefly described. Furthermore, current works on prediction of GI are also reviewed and focused on how to affect results to emphasize important of feature selection.

In Chapter 3, ANN, SVM combined with GA approach is explained in detail.

In Chapter 4, obtained results are evaluated using ANN and SVM and other models.

Finally, conclusions are given in Chapter 5

2. LITERATURE REVIEW

2.1. Previous Research

In the literature, the significant studies are shown in Table 2.1 for the estimation of GI using machine learning methods (Voyant et al., 2017). In these studies, different methods were used in the selection of parameters by using different data sets and the evaluation criteria related to the methods used are presented in this table. Turkey Demirtas et al. in private (Demirtas, 2012) in his study, air temperature, humidity and barometric pressure parameters were developed by k-nearest neighbor (k-NN) classifier as meteorological input variables. In this study, it was estimated that solar radiation estimation was affected by the number of closest neighbors, size of input parameters and distance metrics.

Using ANN except for global solar radiation prediction, Sözen et al. (2004) were the first researchers estimated that solar energy for Turkey (Sözen et al., 2004). Using the data for 17 provinces between 2000 and 2002, they made tests with different learning methods by modeling solar radiation value with ANN. The results are not made in the meteorological station in Turkey showed that the ANN model to assess the prospects of solar potential donors. Şenkal (2016) proposed an ANN model for estimating average solar radiation using atmospheric data for 5 provinces between 2004-2009 (Adana, Ankara, İstanbul, İzmir Samsun) (O Şenkal, 2016). Levenberg-Marquardt learning algorithm and logistic sigmoid transfer function were used in their study. The results of these 5 settlements yielded acceptable results between the measured and estimated values for all months.

Table 2.1. Significant studies on the solution for GI predictions (adapted from Voyant et al., 2017).

References	**Location**	**Evaluation criteria**	**Models**
Burrows (1997)	Canada	-	Regression tree and linear regression
Sözen et al. (2004)	Turkey	-	ANN
Mihalakakou et al. (2000)	Greece	-	ANN = AR
Podestá et al. (2004)	Argentina	-	Generalized regression is useful
Tso & Yau (2007)	China	-	Regression tree, ANN and linear regression
Paoli et al. (2010)	France	nRMSE = %21	ANN, AR, k-NN, Bayesian and Markov
Şenkal (2016)	Turkey	-	ANN

Marquez and Coimbra (2011)	USA	nRMSE= %17.7	ANN and persistence
Moreno et al. (2011)	Spain	-	ANN = generalized regression
Taieb et al. (2012)	Benchmark	MAPE = %18.95	MIMO-ACFLIN strategy (lazy learning) is the winner
Demirtas et al. (2012)	Turkey	nRMSE = %18	k-NN and ANN
Ferrari et al. (2012)	Italia	-	SVM, ANN, k-NN and persistence
Mori and Takahashi (2012)	Japan	-	Regression tree interesting to select variables
Olaiya & Adeyemo (2012)	Nigeria	nRMSE = %24	ANN and regression tree
Fernández et al. (2014)	Spain	-	SVM and persistence
Huang et al. (2014)	Australia	MAPE = %38	Random forest and linear regression
Křomer et al. (2014)	Canada	-	SVM and NWP
Long et al. (2014)	Macao	MAPE = %11.8	ANN, SVM, k-NN and linear regression
Salcedo-Sanz et al. (2014)	Spain	-	Extreme machine learning is useful
Zamo vd. (2014)	Benchmark	nRMSE = %10	Random Forest, SVM, generalized Regression, boosting, bagging and persistence
Almeida et al. (2015)	Spain	nMAE = %3.7-9.4	Quantile regression forests coupled with NWP give a good accuracy for PV prediction
Wolff et al. (2016)	Germany	nRMSE = %6.2	SVR and k-NN
Lauret et al. (2015)	French islands	nRMSE = %19.6	ANN = Gaussian = SVM and persistence
Yadav et al. (2014)	India	mMAPE = %6.89	J48, ANN
Lazzaroni et al. (2014)	Italia	-	SVR, ANN, AR, k-NN and persistence
McGovern et al., (2015)	Benchmark	nRMSE = %13	Regression tree and NWP
Pedro and Coimbra (2015)	USA	Skill over = %23.4	k-NN and persistence
Mori (2001)	Japan	nMAE = 1.75%	{regression tree-ANN} and ANN
Cao and Lin (2008)	China	R^2 = 0.72	{ANN-wavelet} and ANN
Reikard (2009)	USA	nRMSE = 26%	{ARMA} and ANN
Chaouachi et. al. (2010)	Japan	MAPE = 4%	{ANN} and ANN
Pagola et al. (2015)	Spain	-	{SVM-k-NN} and climatology
Chakraborty et. al. (2015)	USA	-	Bayesian and {SVM-ANN}
Hossain, et. al. (2012)	Australia	-	{ANN-least median square}, least median square, ANN and SVM
Bouzerdoum et al. (2013)	Italia	nRMSE = 9.4%	{SARIMA-SVM}, SARIMA and SVM
Chu et. al. (2013)	USA	Skill over persistence = 20%	{GA-ANN} and ANN
Prokop et. al. (2013)	Czech rep	-	{ANN-SVM} and SVM and ANN
Aggarwal & Saini (2013)	USA	-	{ANN-linear regression} and ANN and linear regression
Aggarwal & Saini (2013)	USA	-	{LSR-ANN}, regularized LSR and ordinary LSR

Alobaidi et al. (2014)	UAE	rRMSE = 9.1%	{ANNs} and ANN
Bilionis et al. (2014)	USA	-	{PCA-gaussian process} and NWP
Wu et al. (2014)	Singapore	-	{GA-kmean-ANN}, ANN and ARMA
Wu et al. (2014)	Malaysia	nRMSE = 5%	{GA-SVM-ANN-ARIMA}, SVM, ANN and ARIMA
Yang et al. (2014)	Taiwan	nMAE = 3%	{ANN-SVM}, SVM and ANN
Chu et al. (2015)	USA	Skill over persistence = 6%	{GA-ANN} and persistence
De Felice et al. (2015)	Italia	MAPE = 6%	SVM and linear model
Dong et al. (2015)	USA	nRMSE = 22%	{ANN-SVM} and ARMA
Samanta et al. (2015)	USA	-	{SVR}, SVR, {SVR-PCA}, ARIMA and linear regression

F.S.Rosen Tymvios et al. (2005) proposed models to solve the problem with angstrom linear approach and ANN techniques (F.S.Rosen Tymvios et al, 2005). In their studies, 3 Angstrom models and 7 ANN models were proposed. As a result of the study, ANN approach was found to be a good alternative to traditional approaches.

Yingni Jiang (2008) proposed an ANN model using meteorological data between 1995 and 2004 from 9 different stations in China (Jiang, 2008). 8 stations were used in the learning phase and 1 station in the test phase. The results obtained from the proposed ANN model were obtained better than the empirical regression model results.

Linares-Rodriguez et al. (2011) were made using the satellite imagery as an alternative method for the prediction of solar irradiation of the MLP model without surface measurements by re-analysis (Linares-Rodriguez et al., 2011). The proposed model was tested using the 9-year data of Spain's 83 weather stations. When modeling ANN, the location of the campus, day of the year, sky clearance and 4 meteorological data were used to estimate the spherical irradiation value. The root-mean square error value was 16.4% and the correlation coefficient was 94%.

Yadav et al. (2014) proposed a new model which selects most relevant input parameter using WEKA for artificial neural network based solar radiation prediction models. They use J48 algorithm to decrease input number for ANN. Their study shows important of the inputs for ANN.

Wu et al. (2014) proposed a genetic approach combing multi-model framework for solar radiation time series prediction in their study. Their model supposes several different patterns in the stochastic component of the solar radiation series. To discover the underlying pattern, GA is used to split the solar series dynamically, and the subsequences are further grouped into different clusters (Wu et al., 2014). Their model are called as GAMMF and they present performance of GAMMF is compared with other justified models such as ARMA, TDNN and hybrid model in their study. GAMMF has great advantage in accuracy and consistency according to other models.

Chu et al. presented the smart forecasting models for Direct Normal Irradiance (DNI) that combine sky image processing with ANN optimization schemes (Chu et al., 2013). They used the different methods to develop these season-specific models consist of sky image processing, deterministic and ANN forecasting models. GA is modeled to propose the optimal ANN-based forecasting models for each season. Their models archives that the combination of all these effects resulted in robust forecasting skill above the 20% level over persistence for 5 and 10 min forecast horizons (Chu et al., 2013).

In 2015, a method named as SVM–FFA is developed by hybridizing the SVMs with Firefly Algorithm (FFA) to predict the monthly mean horizontal global solar radiation by Olatomiwa et. al. (Olatomiwa et. al., 2015). Three meteorological data of sunshine duration, maximum temperature and minimum temperature are used as inputs. They present the predictions accuracy of the proposed SVM–FFA model is validated compared to those of ANN and Genetic Programming (GP) models. Their model archives that MAPE is %6.17 in training phase when MAPE is 11.51 in testing phase.

Marquez and Coimbra developed forecasting models for hourly solar irradiation using Artificial Neural Networks for lead times of up to 6 days (Marquez and Coimbra, 2011). Important component of their study is the development of a set of criteria for selecting parameter inputs. To select input variables, they use a version of the Gamma test combined with a genetic algorithm. Their developed forecasting models improve RMSEs for GHI by 10–15% over the reference model.

In previous studies, good results were obtained with ANN model. Yadav et al.’s study (Yadav et al, 2014) showed that performance values were increased by using the J48 algorithm and motivated us to achieve more effective results if better parameter selection was made with meta-heuristics. In this thesis, GA is used for optimization of feature selection and ANN’s parameters.

3. MATERIALS AND METHODS

3.1. Global Irradiation Data and Others Meteorical Parameters

GI data is very important for solar energy applications. GI directly affects the production of energy from solar energy source. But GI measurement is quite expensive process. Hence, others meteorological data can be used to predict GI since there are hidden relations between GI and other meteorological data. In Chapter 2, some important studies made for the cases of Turkey and other places show these relations.

Meteonorm generates accurate and representative typical years for any place on earth (Meteonorm, 2019). Meteonorm presents more than 30 different weather parameters for a selected place on earth. Figure 3.1 shows the screen of the Meteonorm software. Meteonorm can support various file format as exporting monthly, daily, and even hourly data.

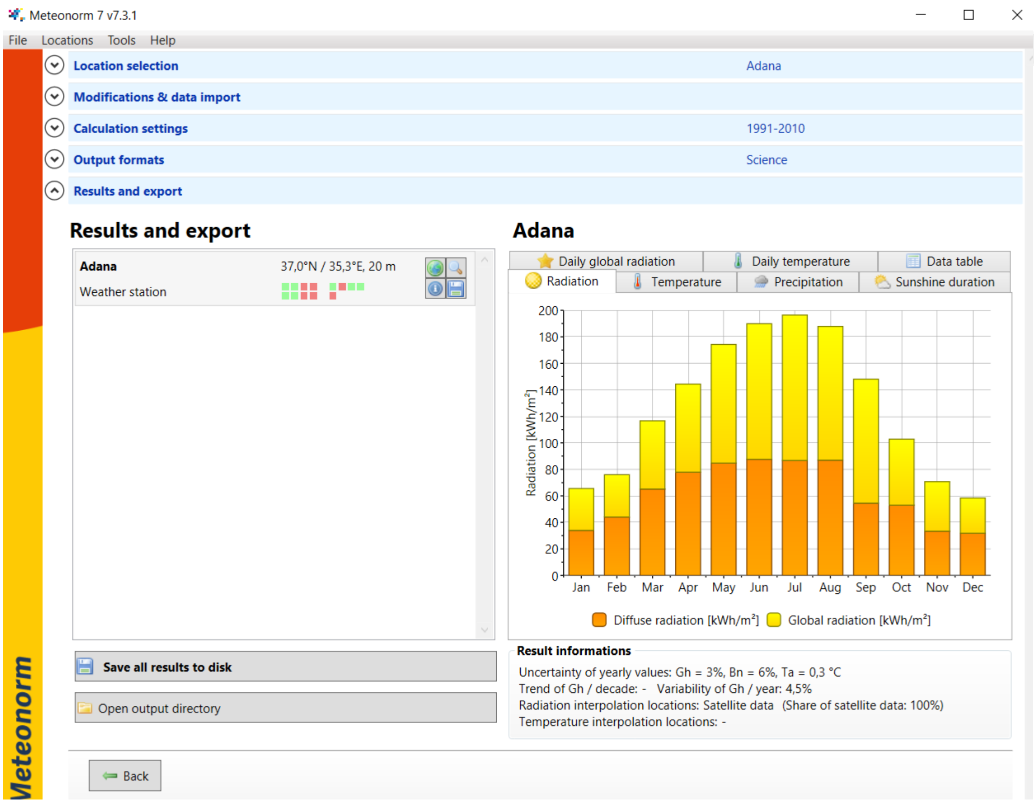


Figure 3.1. Interface of Meteonorm software.

In this study, the obtained data of Meteonorm software is used to predict GI. 29 selected cities' data located in Turkey are used for forecasting models as shown in Table 3.1. Yadav et al. (2014) used the temperature data (T (C), Tmax (C), Tmin (C)), the sun shine hour and monthly average daily solar radiation data (kWh/m^2/day) in their study. In addition to this, parameters in Table 3.3 are used in this study. The meteorological database of cities used in the study is two periods' average values from 1961 to 2009 in these cities. Figure 3.2 shows the map of the selected cities in Turkey. Furthermore, different locations away from each other are selected to test ANN which is educated with special small area for large area. Table 3.2 shows this location and Figure 3.3 shows the map of the locations.

Table 3.1. Geographical coordinates of 29 Turkey cities.

No	Location	Latitude	Longitude	Altitude
1	Adana/Incirlik AFB	37	35.417	66
2	Akhisar (TUR-AFB)	38.91	27.85	93
3	Ankara/Güvercinlik	39.93	32.75	819
4	Aydın	37.85	27.85	56
5	Bandırma (CIV/AFB)	40.31	27.96	49
6	Bodrum	37.03	27.43	26
7	Corum	40.55	34.967	776
8	Dalaman	36.7	28.78	2
9	Denizli	37.78	29.08	425
10	Erzurum	39.917	41.267	1758
11	Finike	36.3	30.15	2
12	Giresun	40.917	38.4	38
13	Gölcük/Dunmlpınar	40.71	29.81	18
14	Hakkari	37.567	43.767	1720
15	Hopa	41.4	41.43	33
16	İnebolu	41.98	33.78	64
17	İskenderun	36.58	36.16	3
18	Istanbul/Ataturk Ab	40.967	28.817	37
19	İzmir/Çiğli(CV/AFB)	38.5	27.01	5
20	İzmir \Menderes	38.26	27.15	120
21	Kahramanmaraş	37.6	36.933	549
22	Kırşehir	39.15	34.16	995
23	Merzifon (TUR-AFB)	40.85	35.58	535
24	Milas Bodrum Airp	37.23	27.68	8
25	Rize	41.03	41.5	4
26	Sanlıurfa	37.133	38.767	547
27	Sivas	39.75	37.017	1285
28	Tekirdağ	40.98	27.55	4
29	Yozgat	39.81	34.8	1298

Figure 3.2. Map of the selected weather stations in Turkey.

Table 3.2. Geographical coordinates of 5 cities in Earth.

No	Location	Latitude	Longitude	Altitude
1	Aberdeen	57.167	-2.083	35
2	Ahmedabad	23.067	72.633	55
3	Bern-Liebefeld	46.928	7.42	565
4	Fresno CA	36.767	-119.717	102
5	Sydney	-33.867	151.2	42

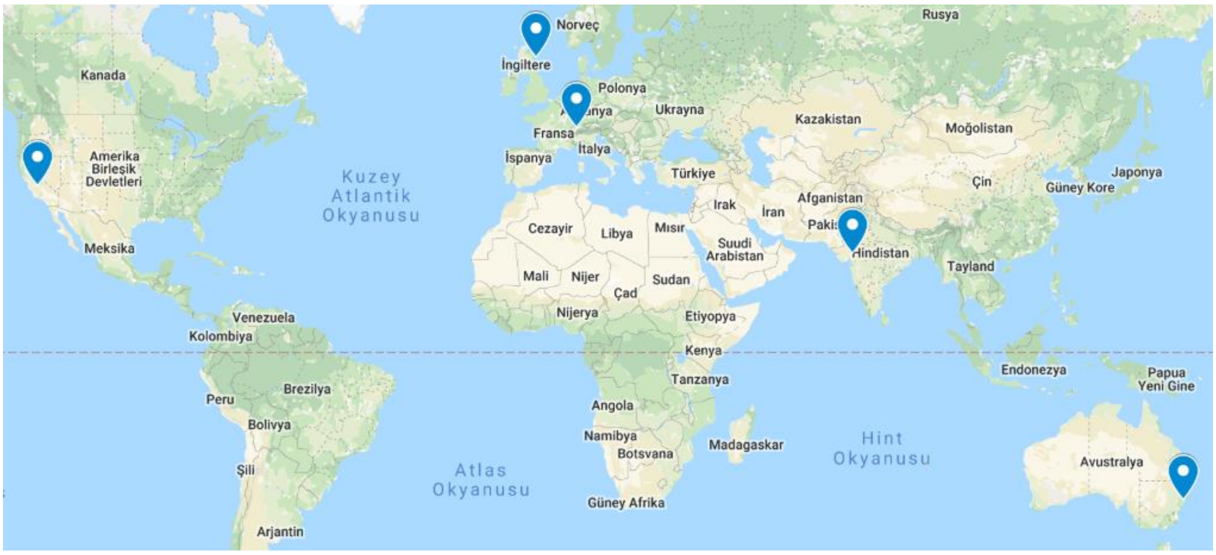


Figure 3.3. Map of the selected weather stations in the Earth.

Table 3.3. Parameters of input.

No	Name	Descriptions
1	Months (1-12: {0,1})	-
2	Ta	Air temperature [°C]
3	Ta min	10 y minimum [°C]
4	Ta dmin	Mean daily minimum Ta [°C]
5	Ta dmx	Mean daily maximum Ta [°C]
6	Ta max	10 y maximum (approx.) [°C]
7	RH	Relative humidity
8	SDm	Sunshine duration [h/day]
9	SDd	Sunshine duration per Day
10	SDastr	Sunshine duration, astronomic
11	RR	Precipitation
12	RD	Days with precipitation
13	FF	Wind speed [m/s]
14	DD	Wind direction
15	Latitude	the angular distance of a place north or south of the earth's equator
16	Longitude	the angular distance of a place east or west of the Greenwich meridian
17	Altitude	the height of an object or point in relation to sea level or ground level.

3.2. Nature-Inspired Algorithms

The Nature-inspired algorithms have been used widely in recent years to solve NP hard problems (Chiong, R., 2009). This study uses one of the most populars of the nature-inspired algorithms which are GA and ANN. GA is used for optimization problems whereas ANN is used for classification and prediction problems. In this study GA and ANN are used as hybrid and proposed a new model for the prediction of GI. In this section, this model is explained and a brief information about ANN and GA model have been presented.

The flow chart of GA with ANN is shown in Figure 3.4 to select the best input parameters and features.

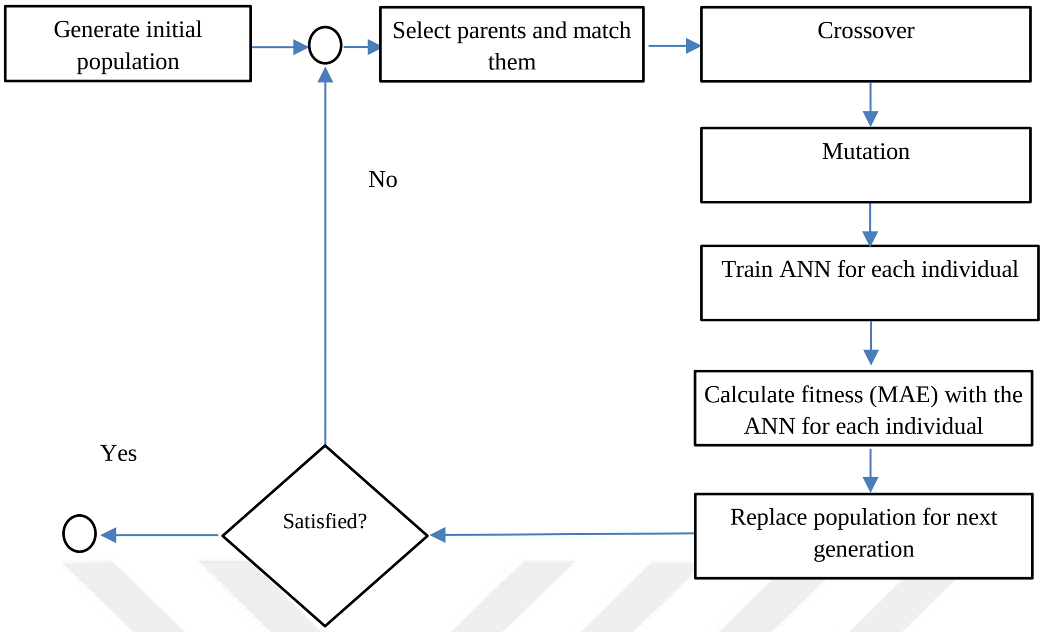


Figure 3.4. GA with ANN

3.3. Genetic Algorithm with ANN

GA modeling was inspired by natural evolution for optimization problems (Paszkowicz, 2009). GA is used to solve complex problem and search selection of the best solutions. GA is power optimization algorithm because it can be modeled as single-objective and multi-objective optimization and adapted easily for different problems. GA is based on population thus offers a set of solution for the problem. GA has four main operators which are generation of the initial solutions, crossover, mutation, selection of the individual, respectively. Algorithm1 shows the main steps to predict GSR. This algorithm is the application of GA. The algorithm needs parameters which are mutation probability, crossover probability, population size and generation size.

Algorithm1: Genetic Algorithm
(params: ***mutateProbability***, ***crossoverProbability***, ***numGen***, ***popSize***)

```
population = generate initial solution as popSize
while i ≠ numGen do
        increment i
        Offspring={}
        for each ind in population do
              parent1=ind
              parent2= select individual in population randomly
              newIndivual =null
              if rnd < crossoverProbability then
                            newIndivual ← Crossover(parent1, parent2)
              end if
              if rnd < mutateProbability then
                            if newIndivual is null then newIndivual = parent1
                            newIndivual ← Mutate(newIndivual)
              end if
              if newIndivual is not null then
                            append newIndivual to Offspring
              end if
        end for
        EvaluateFitnessANN(Offspring)
        Population ← SelectforNext(Population+Offspring, popSize)
end while
```

3.3.1. Representation

Each individual in a population consists of binary values and learning, momentum and number of hidden layer nodes. Each binary value shows whether input is selected. If the input is selected as ANN inputs, the value is “1” else “0”. Learning and momentum parameters are real numbers and compose the upper and lower limits. Number of hidden layer nodes is integer and makes up the upper and lower limits. To show each solution in a single list, by using this notation, the solution can be expressed as:

Structure of the solution for ANN optimization by GA is:

ANN inputs																												**L**	**M**	**#ofN**
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**
1	1	0	0	1	1	1	0	0	0	1	1	1	0	0	0	1	1	1	1	0	0	0	0	1	1	1	0	0.2	0.1	15

First 28 gene of the solution is dedicated to ANN input parameters, 29 gene for learning, 30 gene for momentum and 31 gene for node number in hidden layer are used.

3.3.2. Generation of the initial solutions

Generally, initial population is generated randomly by meta-heuristic algorithms. However, each individual of the population must be feasible for both the generation and process of them. For this reason, the structure of individual given in 3.3.1 must be saved. In this thesis, simple initial algorithm is used. Allocated each gene for ANN inputs is set “0” or “1” randomly. Other genes are assigned with values between given upper and lower limit by parameters.

3.3.3. Selection of parents

Selection parent operators are used to produce new offsprings. Selected parents are mated in crossover operators. In this thesis, each individual of the population are first parents in each iteration when second parent are selected randomly. Thus, each individual is given the opportunity to transfer its gens. This mechanism makes it difficult to catch local optimum and save diversity of the population.

3.3.4. Crossover

Crossover methods are used to produce new child from parents. Genes of the parents are mated then offsprings are generated. In the literature, there are different crossover methods. But some crossover methods are for specific problems. Single point, multi point, uniform, edge recombination are popular crossover techniques.

In this study, an individual has two parts. The first part is inputs whether weather conditions are used for ANN as for the second part is configuration information of ANN which are node number in hidden layer, learning and momentum. Thus, single point crossover is used for the first part. After that, each gene in the second part of the genes is selected randomly from each parent.

Figure 3.5 shows the example of crossover operation. Parent 1 and Parent 2 are mated and a new individual is generated. New individual are taken genes of parents which are blue.

Parent 1:

ANN inputs																												L	M	#ofN
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**
1	1	0	0	1	1	1	0	0	0	1	1	1	0	0	0	1	1	1	1	0	0	0	0	1	1	1	0	0.2	0.1	15

Parent 2:

ANN inputs																												L	M	#ofN
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**
0	0	0	1	1	0	1	0	1	0	1	0	1	0	1	1	0	0	1	0	0	1	0	1	1	0	1	0	0.4	0.5	10

New Indivual:

ANN inputs																												L	M	#ofN
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**
1	1	0	0	1	1	1	0	0	0	1	1	1	0	1	1	0	0	1	0	0	1	0	1	1	0	1	0	0.2	0.5	15

Figure 3.5. Crossover example

3.3.5. Mutation

Mutation operators are used to escape from local optimums for the problem. Mutation operators mutate individual slightly and within at a specified probability. Generally a gene or more than one is mutated when feasible new individuals are generated. Mutation pseudo code is given in Algorithm 2. Algorithm 2 is performed key to mutation probability. The algorithm provides next neighbor of the individual. In Figure 3.6, an example is given, Blue Columns are changed in new individual.

Algorithm2: Mutation operator(Individual p)

```
Select gen in first 28 gen of p
Filip-Flop the selected gen
if rnd(0,1) == 1 then
    newLearnVal=select new value between upper and lower limits
    currentLearnVal ← (currentLearnVal + newLearnVal) / 2
end if
if rnd(0,1) == 1 then
    newMomentumVal=select new value between upper and lower limits
    currentMomentumVal ← (currentMomentumVal + newMomentumVal) / 2
end if
if rnd(0,1) == 1 then
    newNodeNummerVal=select new value between upper and lower limits
    currentNodeNummerVal ← (currentNodeNummerVal + newNodeNummerVal) / 2
end if
```

Individual:

ANN inputs																												L	M	#ofN
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**
0	0	0	1	1	0	1	0	1	0	1	0	1	0	1	1	0	0	1	0	0	1	0	1	1	0	1	0	0.4	0.5	10

New Individual:

ANN inputs																												L	M	#ofN
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**
0	0	0	1	1	0	1	0	1	0	0	0	1	0	1	1	0	0	1	0	0	1	0	1	1	0	1	0	0.8	0.5	15

Figure 3.6. Mutation example

3.3.6. Replacement strategy

Selection method is used to generate the next population from the current population and offspring. The most popular of selection methods are Roulette Wheel Selection, selection of best individual and Rank Selection. In this study, selection of the best individual is used. This operation selects individuals which have the best fitness value for next generations.

3.4. Artificial Neural Networks

The human brain is that has an exceptionally complicated structure which is able to think, recognize, keep in mind and analyze. Especially human brain has strong learning process. Therefore, human can response in case of unobserved conditions. The biological neuron cells are dependable for information exchanges to the brain. A commonplace organic nerve cell called a neuron is fundamentally comprised from a cell body, a signal carrying axon, and a cluster of dendrites as seen from Figure 3.7. Neurons carry on as chip in computer phrasing. Each neuron gets the flag of numerous other neurons all through the dendrites. In the event that this flag is solid sufficient, the neuron is enacted and produces a yield flag which is transmitted through the axons.

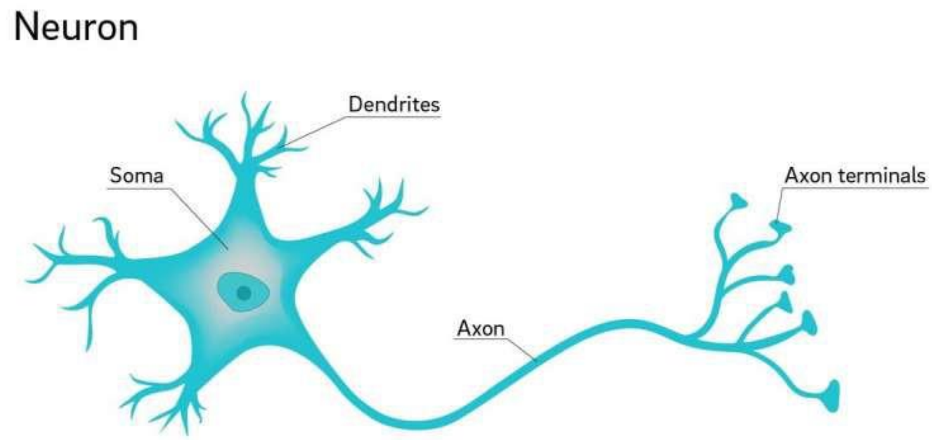


Figure 3.7. Biological Neuron (Towardsdatascience, 2019).

ANNs are mathematically modeled inspring from biological nervous systems. One of the most comprehensive definitions of ANN in the literature is that ANNs are distributed, adaptive, generally nonlinear learning machines composed of many different processing elements (or neurons) (Principe et al., 2000).

First studies on neural network concept date back to the pioneering work of McCulloh and Pitts in 1943. In their research, McCulloh and Pitts created a neuron model that was the mathematical imitation of the functioning of the biological neuron. Moreover, it has been proved that artificial neuron could, in principle, compute any arithmetic or logical function (Hagan et al., 1996). A simple artificial neural network is modeled in figure 3.8. This neural network consists of inputs, weights, total function, junction function, activation function and output (S. Samarasinghe, 2007)

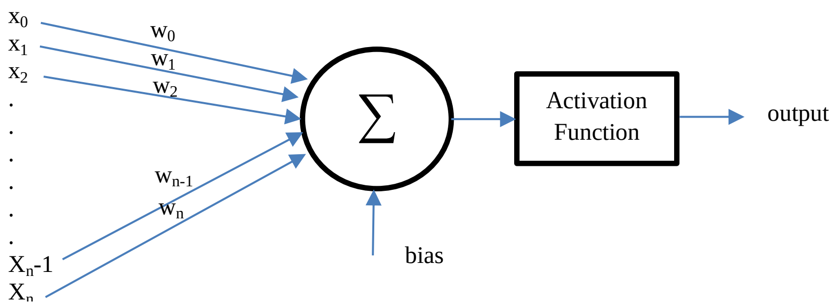


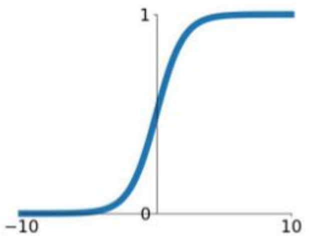
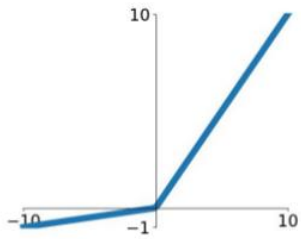
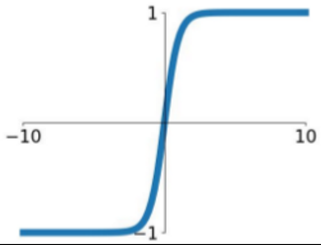
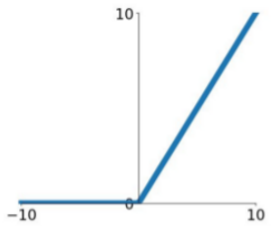
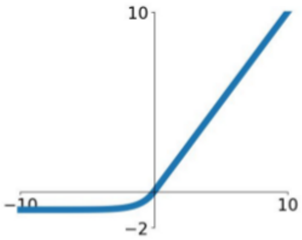
Figure 3.8. Structure of an artificial neuron.

In Figure 3.8, an artificial neuron has two important missions which are of receiving from input variables and producing an output value. Each input affects the output as its synaptic weight and determined as net function. The net function determines how to receive the signals from the input layer by synaptic weights (or simply called weights). Signals are transferred to neuron as

$$net = \sum_{1}^{n} (w_i \; x \; x_i) + b \qquad (1)$$

where n is number of the input variables, x_i is the i. value of the input vector, w_i is the synaptic weights of the th i pair of the input vector, and b is the bias or threshold term of the neuron. The neuron uses activation function to produce an output. The activation function are selected related with the problem. Table 3.4 shows list of the activation functions.

Table 3.4. The list of the activation functions (Farhadi, 2017)

Name	Output	Name	Output
Sigmoid		Leaky RelU Max(0.1x, x)	
tanh		unthresholded linear units	
RelU max(0,x)		ELU $\begin{cases} x, \\ x \propto (e^x - 1), \end{cases}$	

Neural Network Topologies and Structures: ANN can be modelled as different structures. Firstly, Feed-forward Neural Networks (FFNNs) and Recurrent Neural Networks (RNNs) can be divided.

FFNNs are artificial neural networks where the connections between units do not form a cycle (Ben-Bright et al, 2017). Single or multiple layers structures can be used in FFNNs. FFNNs use learning algorithms to compute weight of the connections and to construct relation between input and outputs.

Recurrent Neural Networks (RNNs) has feedback loop to use output of them (Ben-Bright et al, 2017). Furthermore, RNNs have internal memory to use input and output in different times and do not need to use learning algorithms (Hu and Hwang, 2002). But RNNs are more complex than FFNNs according to their structures.

Secondly, the major models of the ANN are Single Layer Perceptron, Multilayer Perceptron, Radial Basis Function Network, The Kohonen Network. In this study, Multi-Layer Perceptron are used to predict GI.

3.4.1. Feed-Forward Neural Network Architectures

ANNs have layered structure which is responsible for data transfer. A typical neural network is comprised of an input layer, one or more hidden layers and an output layer. Input layer contains the input variable vector that will be introduced to the network. The hidden layer which is built of the neurons is the processor layer of the network and responsible for transfer of data. Each neuron in the hidden layer has the net and transfer functions. An output layer produces the network output by processing the data which are transferred from the hidden layer. The first neural network model developed by McCulloch and Pitts had composed of only one neuron which was connected to the input layer. At approximately the same time, Rosenblatt (1958) also proposed the first generation neural networks known as the perceptron. In the perceptron model, a single neuron with a weighted linear net function and transfer function is utilized (Hu and Hwang, 2002). The difference between the two models is that the former model suggested the mathematical explanations of a neuron whereas the latter proposed a learning algorithm which provides convergence when the training samples are linearly separable.

3.4.2. Multi-Layer Perceptron Models

Single layer models are insufficient to map non-linear relation between the input and outputs, network models with multi-layer neurons are often employed (Zhang et al, 2019). A Multilayer Perceptron (MLP) neural network model consists of layered network of neurons as shown in Figure 3.9. Each neuron in an MLP has nonlinear transfer function that is often continuously differentiable (Hu and Hwang, 2002).

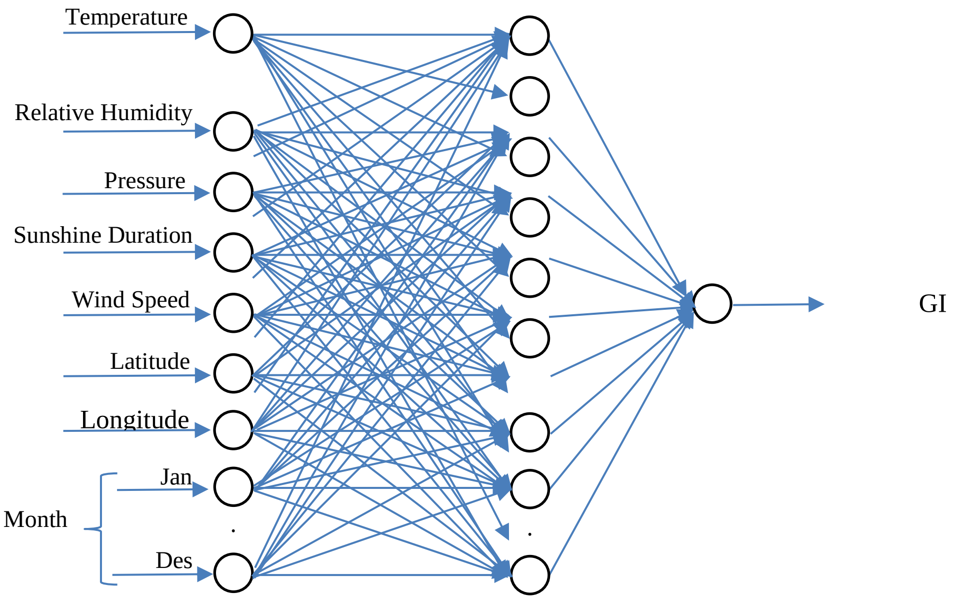


Figure 3.9. A MLP example

3.4.3. Learning Strategies and Back-Propagation Learning

In the neural network literature, there are two fundamental learning strategies which depend on the neural network architecture.

a) Supervised Learning: Supervised learning or sometimes called active learning implies the availability of an external agent that means having knowledge of a set of input- output examples (Haykin, 1994). A learning algorithm such as backpropagation algorithm is utilized to perform supervised learning in most feedforward MLPs.

b) Unsupervised Learning: In unsupervised learning strategy, there is no external teacher which defines the environment to perform learning process. In other words, there are no specific examples of the function to be learned by the network. Self-organizing networks are trained in the manner of unsupervised learning by employing a special learning algorithm such as Hebbian Learning algorithm (Kroese and Smagt, 1996).

Back-propagation is important of neural network training and widely used. It is the act of adjusting the loads of a neural net dependent on the mistake rate (for example misfortune) acquired in the previous epoch

In this thesis, backpropagation algorithm is used to train ANN.

3.5. Support Vector Machines

A SVM performs classification by constructing an N-dimensional hyperplane that optimally separates clusters of vector into two non-overlapping groups (Vapnik, 1995; Vapnik, 1997, Cristianini and Taylor, 2000). SVM is proposed by Vapnik (Vapnik, 1995) and SVM were used to solve the classification problems with the introduction of the ε-insensitive loss function, after it is extended to use in regression problems (Vapnik, 1997; Yao et al, 2017). Figure 3.10 shows example of hyperplanes for a SVM trained with samples from two classes (Bhavsar & Panchal, 2012). Kernel functions (sigmoid kernel, radial basis function (RBF) kernel, polynomial kernel and Gaussian kernel) and hyper-parameters (ε, c) are used such that a bound on the Vapnik Chervonenkis (VC) dimension is minimized (Suykents and Vandewalle, 1998). SVM models are closely related to artificial neural networks and SVM model which uses a sigmoid kernel function is a two-layer, feed-forward neural network (Swiercz anet l, 2008; Haykin, 1998; Wallace and Anagnostopoulos, 2010) SVM has same principles for regression and classification problems. Usually SVM are modeled as supervised learning model with using input data. Similar to ANN, data is split into a training and a test data. For training data (x_i, y_i) , …, (x_n, y_n) , where x_i are input values which are temperature, humidity, wind and pressure. y_i is GI data and output value for SVM.

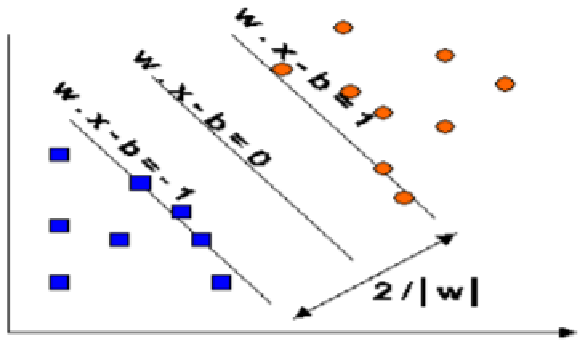


Figure 3.10. Maximum margin hyperplanes for a SVM trained with samples from two classes (Bhavsar & Panchal, 2012)

3.6. Feature Ranking

Feature Selection (FS) and Feature Ranking (FR) are used to select a subset of properties in a dataset or rank properties in a dataset that are related with output values. In this thesis, Correlation Feature Selection (CFS) and Information Gain (IG) are selected to rank the features. Selected a subset of available features using obtained meteorological by these methods are used for ANN and SVM and performance results of the inputs are given in Chapter 4.

3.6.1. Correlation Feature Selection

CFS is used to choose a subset of features which are highly co-related to class (Mukherjee & Sharma, 2012; Bhosale & Ade, 2014). In each subset attribute are selected taking account of the redundancy ranking between them and conjectural ability of each individual variable. A function which calculates best individual feature is (Bhosale & Ade, 2014):

$$Merit(S_k) = \frac{K * cr_{fc}}{\sqrt{K + K * (K-1) * cr_{ff}}} \tag{2}$$

where, Merit is the heuristic merit of feature subset S having K variable, cr_{fc} is the average feature class correlation and cr_{ff} is average feature-feature co-relation. The numerator shows the conjectural ability of features and the denominator indicates the ranking of redundancy between the inputs.

3.6.2. Information Gain

Information Gain (IG) is used to measure the relationship between inputs and outputs for classification or regression problems. IG calculates the entropy value for each input. The entropy value is calculated using Eq. (3) and gives an information about the most useful feature for the problems. The entropy values having higher values are more related with the output. If the target outputs can take on k different values, then the entropy of the feature relative to this k - wise classification is defined as (Novakovic, 2009):

$$Entropy(S) = \sum_{c=1}^{k} -[p_c * (log_2 * p_c)] \tag{3}$$

where, P_c is a proportion of S having a bearing upon class c.

$$IG(S, a) = \mathrm{E(S)} - \mathrm{S(S|a)} \tag{4}$$

where a is input value. Eq. (4) is used to measure IG that training example S obtains from observation that a random input V takes some value (Novakovic, 2009).

3.7. WEKA

Waikato Environment for Knowledge Analysis (Weka) is a suite of machine learning software written in Java, developed at the University of Waikato (Weka, 2019). WEKA has several machine learning algorithm and is popular for data mining. ARFF and CVS files are analyzed with using WEKA interface. Furthermore, WEKA API can be used in JAVA. Figure 3.11 shows the interface of Weka.

In this study, Weka API is used in JAVA platform. GA is coded with Java. WEKA’s ANN, SVM and Ranking algorithm API are used as Fitness function. Obtained results are saved to SQL Server for each iteration.

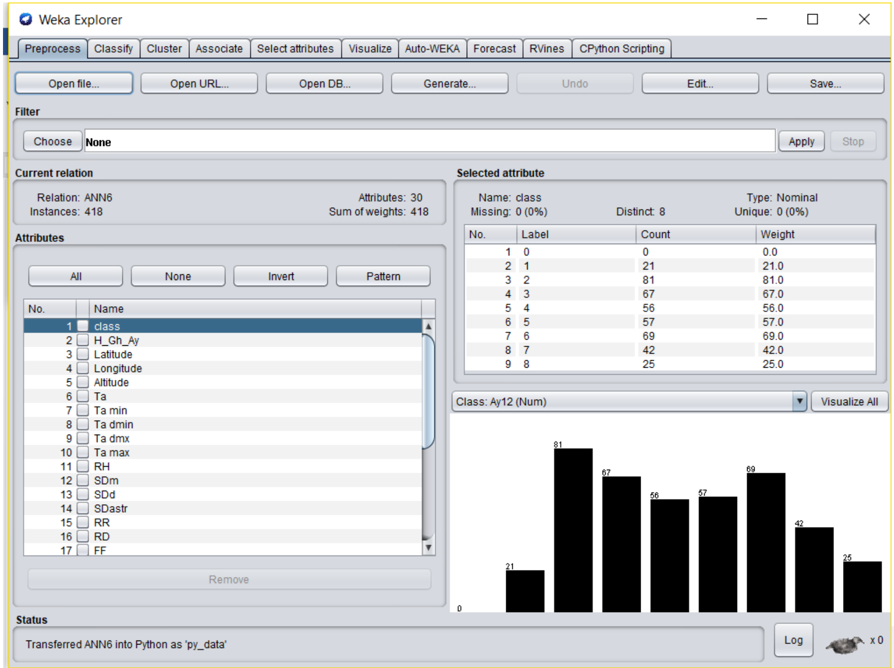


Figure 3.11. Weka interface.

3.7.1. WEKA API

WEKA application programming interface (API) provides programmers to use its sub modules. WEKA has several machine learning algorithms such as ANN, SVM, Correlation Feature Selection, Information Gain and regression analysis.

The most components of WEKA API are Instances, Filter, Classifier/Clusterer, Evaluating and Attribute selection (WEKA, 2019). In this section, used modules are explained how to use them in this thesis.

WEKA supports different well known file formats such as excel, data and json. But programmers generally use ARFF (Attribute-Relation File Format) file which is a text file that describes a list of instances sharing a set of attributes (ARFF, 2019). Instances, ArffLoader and ArffSaver are must be included to use ARFF file and load and save data operation in java code as follows:

```
import weka.core.*;
import weka.core.Instances;
import weka.core.converters.ArffLoader;
import weka.core.converters.ArffSaver;
```

Instances class is used to employe obtained data from observation from the problem domain. Firstly the data must be loaded into memory. An object (named loader in the following code) is created from ArffLoader class and filename (including its path) are allocated to the object using setFile method. All data are loaded with getDataSet method of ArffLoader.

```
ArffLoader loader   = new ArffLoader();
loader.setFile(new File(testFilename));
Instances ins = loader.getDataSet();
```

MultilayerPerceptron class provides artificial neural network model which uses backpropagation to learn a multi-layer perceptron and its nodes in this network are all sigmoid function (but when the class is numeric, in case of the output nodes become unthresholded linear units). Parameters are important for ANN models. setOptions method of MultilayerPerceptron is used to set parameters that supports learning rate (-L), momentum (-M), number of epochs (-N), percentage size of validation set (-V), seed (-S), threshold for number of consecutive errors (-E), The hidden layers to be created for the network (-H)

```
MultilayerPerceptron m = new MultilayerPerceptron();
String optStr= -L 0.12 -M 0.11 -N 5000 -V 0 -S 0 -E 20 -H 15";
m.setOptions(Utils.splitOptions(optStr));
m.buildClassifier(train);
```

SMOreg is used to perform SVM. SMOreg provides few parameters such as the complexity constant C (-C), whether to 0=normalize/1=standardize/2=neither (-N), classname and parameters (-I), classname and parameters (-K). its buildClassifier method is used to train model with train data.

```
SMOreg s = new SMOreg();
String optStr -C 1.0 -N 0 -I
"weka.classifiers.functions.supportVector.RegSMOImproved -T 0.001
-V -P 1.0E-12 -L 0.001 -W 1" -K
"weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C
250007";
s.setOptions(Utils.splitOptions(optStr));
s.buildClassifier(train);
```

Evaluation class is used for evaluating machine learning models. Firstly, an object is created from the class. The object receives two parameters which are model and data. When the model can be trained model with train data, the data is to be test data. Well-known evaluation values can be obtained from the object that are Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Relative Absolute Error (RAE) and Root Relative Squared Error (RRSE).

```
Evaluation eval = new Evaluation(train);
eval.evaluateModel(mlp, test);
double corCoef = eval.correlationCoefficient();
double MAE=eval.meanAbsoluteError();
double RMSE=eval.rootMeanSquaredError();
double RAE=eval.relativeAbsoluteError();
double RRSE = eval.rootRelativeSquaredError();
```

4. RESULTS AND DISCUSSIONS

In this thesis, a solution approach for prediction GI via GA is proposed. The performance of GA is investigated using Meteonorm Software data. The algorithms are coded in the Java language and are run on a computer with i7 processor technology, 2.6 GHz Turbo processor speed and 8 GB RAM capacity. The algorithms are tested with the parameters illustrated in Table 4.1. These are mutation probability, crossover probability, population size and generation size for GA, learning rate, momentum Factor and number of hidden layer for ANN.

Table 4.1. Parameters of GA and ANN algorithm.

GA Parameters		ANN Parameters	
Parameters	**Values**	**Parameters**	**Values**
Mutation probability	0.1	Learning rate	0.1−1
Crossover probability	0.7	Momentum Factor	0.1−1
Population Size	50	# of Hidden Layer	5−50
Generation Size	100	Max. Input Number	40

4.1. Performance comparison functions

The prediction performance is evaluated based on six metrics, namely, R^2, Correlation coefficient (ρ), Mean Absolute Percentage Error (MAE), root mean square error (RMSE), relative absolute error (RAE) and root relative square error (RRSE).

$$r = \frac{\sum(GSR_a - \overline{GSR_a})(GSR_p - \overline{GSR_p})}{\sqrt{\sum(GSR_a - \overline{GSR_a})^2 \sum(GSR_p - \overline{GSR_p})^2}} \tag{5}$$

$$\rho = \frac{\frac{1}{n}\sum_{i=1}^{n}(GSR_{i,a} - \overline{GSR_a})(GSR_{i,p} - \overline{GSR_p})}{\sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}(GSR_{i,a} - \overline{GSR_a})^2\right)\ \left(\frac{1}{n}\sum_{i=1}^{n}(GSR_{i,p} - \overline{GSR_p})^2\right)}} \tag{6}$$

$$MAPE = \frac{100}{n}\sum_{i}^{n}\frac{[GSR_{i,a} - GSR_{i,p}]}{GSR_{i,a}} \tag{7}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(GSR_{i,a} - GSR_{i,p})^2}{n}} \tag{8}$$

$$RAE = \frac{\sum_{i}^{n}\left[GSR_{i,a} - GSR_{i,p}\right]}{\sum_{i}^{n}\left[\overline{GSR_p} - GSR_{i,a}\right]} \tag{9}$$

$$RRSE = \sqrt{\frac{\sum_{i=1}^{n}(GSR_{i,a} - GSR_{i,p})^2}{\sum_{i=1}^{n}(\overline{GSR_p} - GSR_{i,a})^2}} \tag{10}$$

where, $GSR_{i,a}$ is GSR i^{th} actual input value, $GSR_{i,p}$ is GSR i^{th} predicted value, n is number of instance size, $\overline{GSR_p}$ is average of GSR's predicted value, $\overline{GSR_a}$ is average of GSR's actual value.

4.2. Computational Results and Analyses

Obtained feature from Meteonorm software are ranked by Info Gain Attribute Eval and Correlation Ranking Filter and given in Table 4.2. First 11 rows of the table shows the same features for InfoGainAttributeEval and Correlation Ranking Filter. And the features have more than 0.5 rank value for Info Gain Attribute Eval. For this reason, these features are selected for input parameters as the first model (it is named as Model 1) and the first proposed model are tested to compare with others.

Table 4.2. Parameters of input.

No	InfoGainAttributeEval		Correlation Ranking Filter	
	Name	Rank	Name	Rank
1	SDastr	1.80618	SDastr	0.3384
2	SDm	1.65331	SDd	0.3203
3	SDd	1.60248	SDm	0.3192
4	Ta max	1.00717	Ta max	0.292
5	Ta dmx	0.88317	Ta dmx	0.2894
6	Ta	0.88165	Ta	0.2812
7	RD	0.74723	Ta dmin	0.2604
8	Ta min	0.63957	Ta min	0.2558
9	Ta dmin	0.63894	RD	0.2398
10	RR	0.6257	RH	0.2215
11	RH	0.50224	RR	0.2116
12	month 12	0.22005	month 12	0.1763
13	Latitude	0.18458	month 3	0.1648
14	month 10	0.15227	month 10	0.1611
15	month 7	0.15072	month 2	0.1595
16	month 3	0.15003	month 11	0.1566
17	month 6	0.14023	month 1	0.1531
18	month 1	0.136	month 6	0.1487
19	month 2	0.13503	month 5	0.1365
20	month 11	0.12386	month 8	0.1286
21	month 8	0.11482	month 4	0.1284
22	month 5	0.11034	month 7	0.1262
23	month 4	0.11014	month 9	0.1258
24	month 9	0.10219	Latitude	0.0995
25	FF	0.06141	DD	0.0681
26	Altitude	0	Longitude	0.063
27	Longitude	0	FF	0.0462
28	DD	0	Altitude	0.0394

The second model is proposed by GA with ANN and it is named as Model 2. The detailed information about founded the best parameters by the proposed algorithms and parameters of Yadav’s ANN models (Yadav et al., 2014) are presented in Table 4.3. These parameters are used to compare performance of the models.

Table 4.3. Found Best Parameters using GA and Yadav's used Parameters for ANN.

Parameters	Model 1 by by Ranking Methods	Model 2 by GA	Yadav's ANN1	Yadav's ANN2	Yadav's ANN3
Learning rate	0.15	0.15	-	-	-
Momentum Factor	0.11	0.11	-	-	-
# of Hidden Layer	32	32	9	10	10
Inputs	SDastr, SDm, SDd, Tamax, Tadmx, Ta, RD, Tamin, Tadmin, RR,RH	Latitude, Longitude, Altitude, Ta min, Ta dmx, SDd, SDastr, RD	T, Tmin, Tmax, Altitude, SH, Lat and Long	T, Tmin, Tmax, Altitude, SH	T, Tmin, Tmax and Altitude

Firstly, SVM and ANN are tested with identical inputs in this study. Table 4.4 presents the results of these comparisons for the Turkey case. Evaluation functions are computed for both training and testing. Input Type column shows the selected features for ANN and SVM in Table 4.5. All inputs, Model 1 and Model 2 are proposed in this thesis. Obtained results by the models are presented in Table 4.5. ANN Models achieves better results than SVM for all cases. The best MAPE is % 5.60 when all features are used for testing data. But better results are obtained with using Model 2 which has less features and the best MAPE value is %3.44 for testing data also its other evaluation values are the best according to other's results. And the row has the best values that are presented as bold.

Table 4.4. Performance statistics of the SVM and ANN models for proposed best inputs and all inputs.

Model	Inp.Type	#of Ins.		ρ	R^2	MAE	MAPE	RMSE	RAE	RRSE
ANN	All	418	Training	1	1	0.08	%2.28	0.10	%4.72	%5.25
	All	278	Testing	0.98	0.97	0.23	%6.01	0.36	%13.96	%19.00
SVM	All	418	Training	0.99	0.98	0.20	%5.28	0.29	%12.00	%14.91
	All	278	Testing	0.99	0.98	0.21	%5.60	0.29	%13.09	%15.73
ANN	Model 1	418	Training	0.99	0.98	0.21	%5.80	0.29	%12.39	%14.84
	Model 1	278	Testing	0.98	0.97	0.27	%7.71	0.37	%16.54	%19.60
SVM	Model 1	418	Training	0.98	0.96	0.29	%7.62	0.38	%16.94	%19.50
	Model 1	278	Testing	0.98	0.96	0.30	%7.97	0.39	%18.56	%20.85
ANN	Model 2	418	Training	1	0.99	0.11	%3.05	0.18	%6.68	%9.11
	Model 2	278	Testing	**0.99**	**0.99**	**0.13**	**%3.44**	**0.20**	**%8.01**	**%10.67**
SVM	Model 2	418	Training	0.98	0.97	0.27	%7.58	0.35	%15.82	%18.13
	Model 2	278	Testing	0.98	0.97	0.25	%6.72	0.33	%15.19	%17.40

Table 4.5 are shows the obtained results by the proposed ANN models and Yadav's models for Turkey. The obtained results by ANN Model 2 are better than Yadav's Models and ANN Model 1.

Table 4.5. Comparison of Proposed ANN models and Yadav's Models

Model		**ρ**	**R^2**	**MAE**	**MAPE**	**RMSE**	**RAE**	**RRSE**
ANN Model1	Training	0.99	0.98	0.21	%5.80	0.29	%12.39	%14.84
	Testing	0.98	0.97	0.27	%7.71	0.37	%16.54	%19.60
ANN Model2	Training	1	0.99	0.11	%3.05	0.18	%6.68	%9.11
	Testing	**0.99**	**0.99**	**0.13**	**%3.44**	**0.20**	**%8.01**	**%10.67**
ANN1	Training	0.98	0.95	0.33	%9.08	0.43	%19.61	%22.30
	Testing	0.96	0.91	0.43	%12.34	0.55	%26.26	%29.55
ANN2	Training	0.96	0.91	0.45	%13.62	0.59	%26.91	%30.53
	Testing	0.95	0.90	0.50	%14.44	0.62	%30.29	%33.17
ANN3	Training	0.89	0.80	0.74	%23.04s	0.89	%43.78	%46.42
	Testing	0.88	0.77	0.76	%23.09	0.92	%46.16	%48.99

Figure 4.1 – 4.6 show comparison of the predicted GI and actual GI with using testing data for Turkey. While y-axis represents the predicted GI, x-axis represents the measured or actual GI values in these figures. It is clearly seen that disparities between predictions and measurements are lower in ANN relative to SVM. Furthermore, predicted GI values are close to actual GI values.

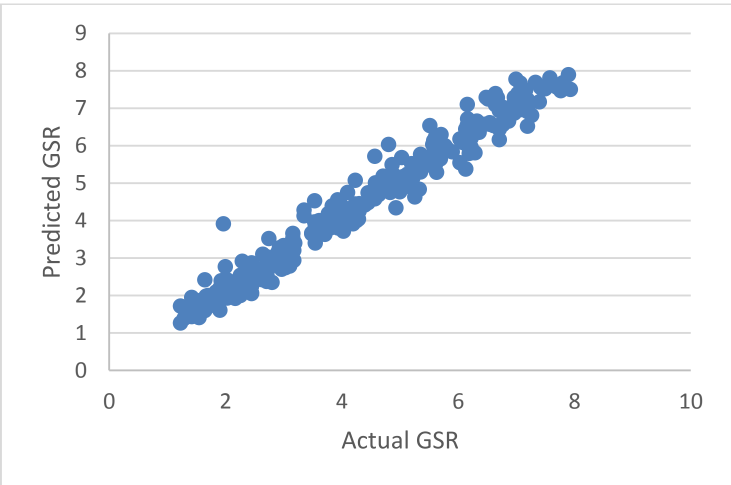


Figure 4.1. Comparison of measured and estimated GI for ANN_{model1} for Turkey.

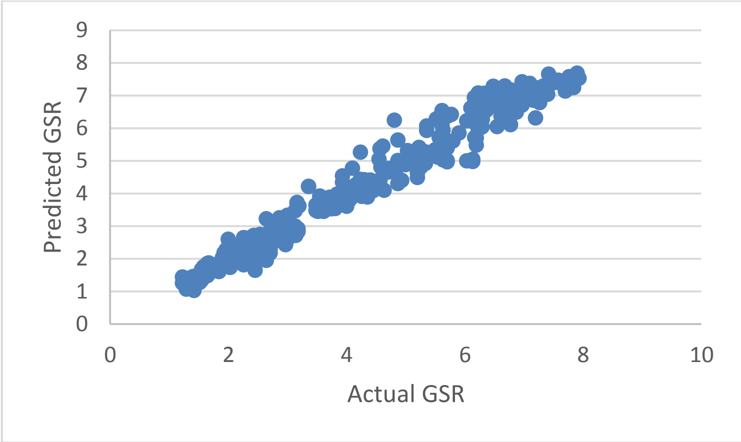


Figure 4.2. Comparison of measured and estimated GI for SVM_{model1} Turkey.

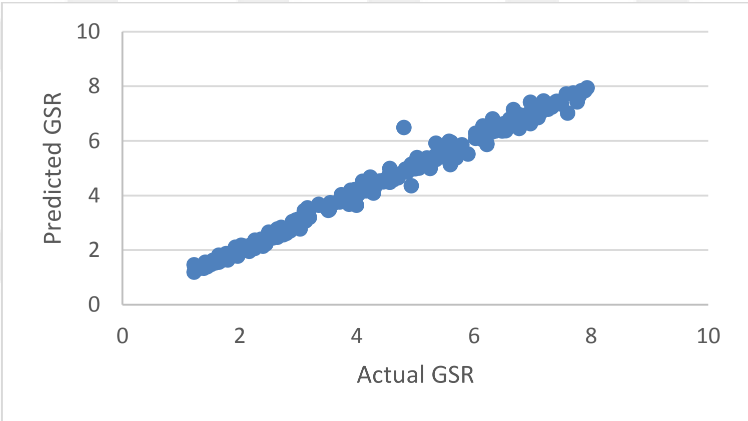


Figure 4.3. Comparison of measured and estimated GI for ANN_{model2} Turkey.

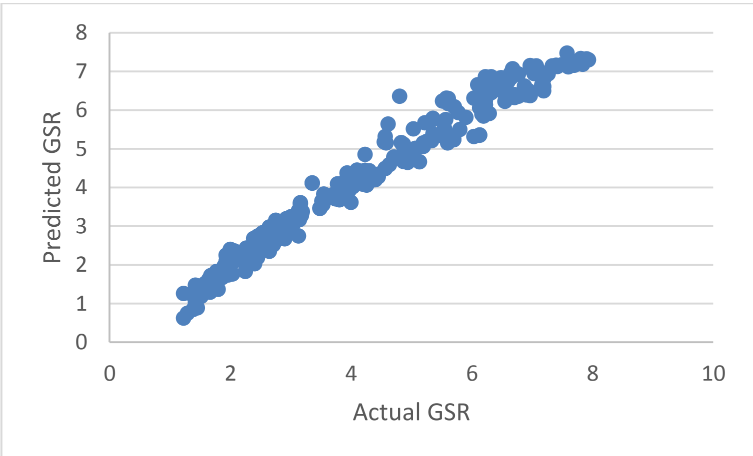


Figure 4.4. Comparison of measured and estimated GI for SVM_{model2} Turkey.

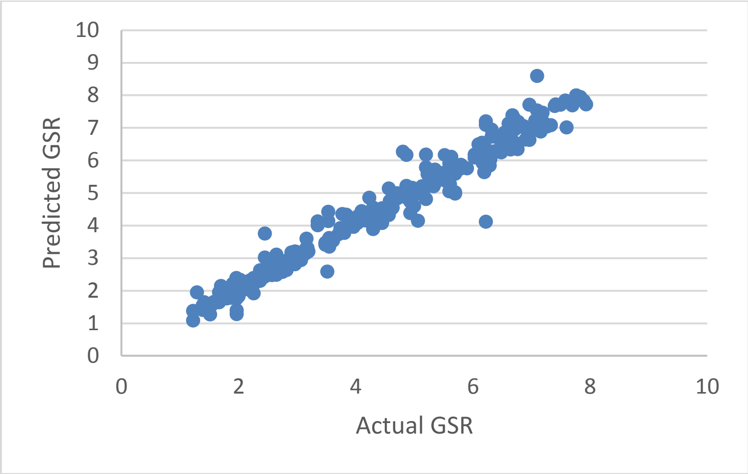


Figure 4.5. Comparison of measured and estimated GI for ANN_{all} Turkey.

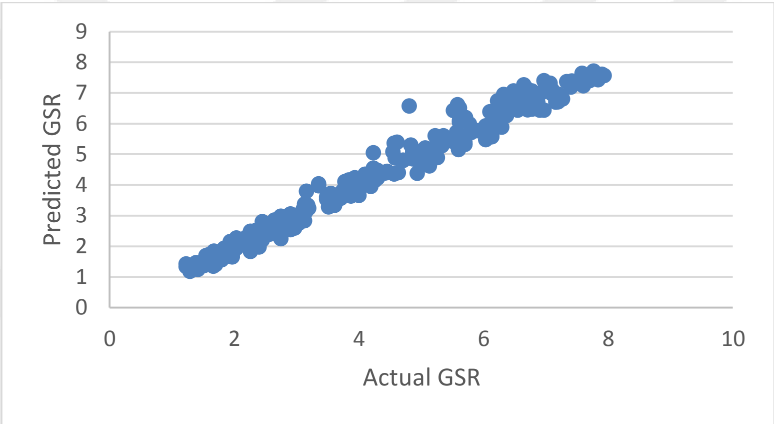


Figure 4.6. Comparison of measured and estimated GI for SVM_{all} Turkey.

In this thesis, trained ANN models by Turkey case are tested for different locations in the world. The obtained results are given in Table 4.6. ANN Model 1 and ANN Model 2 are used for this comparison. Furthermore, ANN Model 3 is proposed by Latitude, Longitude and Altitude in ANN Model 2 is removed. Because latitude and longitude are angular data and trained ANN by specific locations produces high error results. In this practice, ANN Model 1 has the best evaluation values than others and its results can be acceptable.

Table 4.6. Comparison of Trained ANN Models using Turkey Region are applied for different locations.

Model	ρ	R^2	MAE	MAPE	RMSE	RAE	RRSE
ANN Model 1	**0.94**	**0.89**	**0.50**	**%15.27**	**0.69**	**%30.21**	**%33.85**
ANN Model 2	0.34	0.113	6.45	%189.48	9.18	%388.67	%453.19
ANN Model 3	0.92	0.84	0.57	% 19.50	0.87	% 34.39	% 43.0

Figure 4.7 – 4.10 show the comparison of the predicted GI and actual GI with using testing data for 5 different locations and ANN Models. ANN Model 1 has better predicted values than the others.

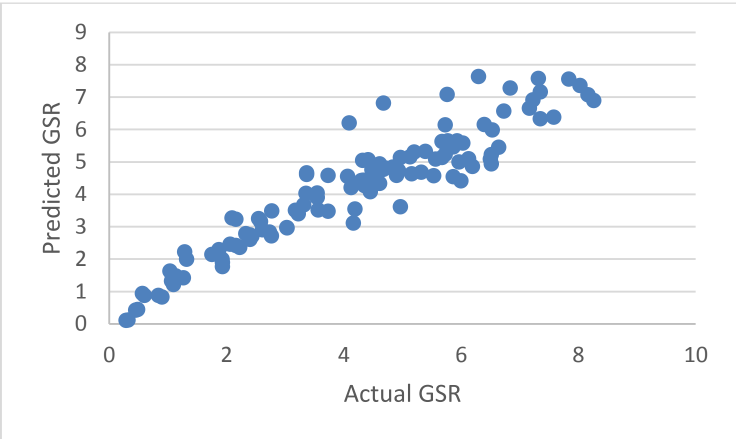


Figure 4.7. Comparison of measured and estimated GI for ANN_{Model1} using different locations.

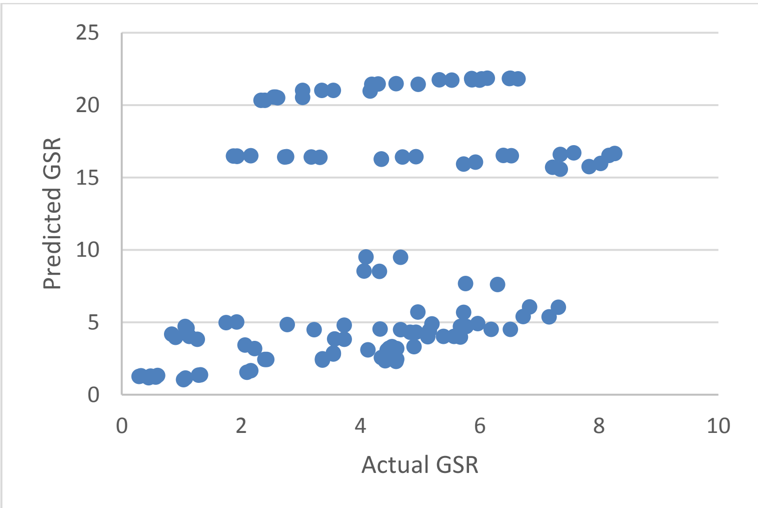


Figure 4.8. Comparison of measured and estimated GI for ANN_{Model2} using different locations.

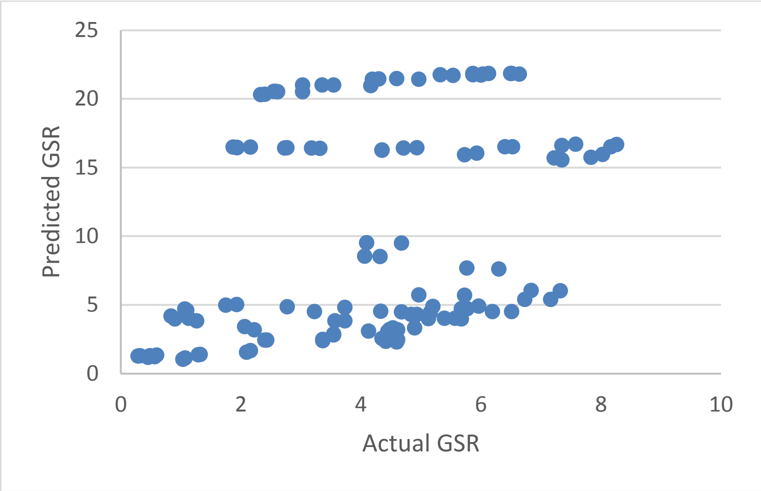


Figure 4.9. Comparison of measured and estimated GI for ANN_{model3} using different locations.

5. CONCLUSIONS

This study presents GA with ANN and feature ranking methods for GI estimation that use adapted crossover and mutation operation to solve this problem. The algorithms are tested on obtained meteorological data from Meteonorm software in this study using proposed ANN models and ANN models in Yadav’s study. GA is designed to perform single-objective optimization. The experiments performed on different locations to show the real potential of the proposed methods. Compared to the Yadav’s models, better MAPE, MAE, RAE, RRSE, and RMSE values are obtained. For example, the best MAPE is 3.44 using optimized ANN by GA. MAPE can be acceptable to use ANN Model 2. Another point is that the testing data usually are selected in the area which is close to training data. In this thesis, trained ANN by using locations in Turkey are tested using 5 location in the world. The MAPE value has %15.21 when ANN Model 1 proposed by using feature ranking methods so that reasonable results are obtained. It can be concluded that optimized ANN model by GA with ANN method can presents better results using training and testing data in specific location when feature ranking methods present better results using training and testing data in different location. The proposed model is used a place where GI is not measured. Electricity suppliers, farmers and etc. can use the proposed model to see the efficiency and electricity production amount. This study has still potential to be improved by employing different strategies such as using different artificial intelligence methods and classifying monthly meteorological parameters. The proposed models will be tested on daily data and using the real-time meteorological data measured instantaneously by the Turkish State Meteorological Service in future. Also, new models and methods can be developed to extend other meta-heuristic algorithms in the future.

REFERENCES

Aggarwal, S. K., & Saini, L. M. (2013). Solar energy prediction using linear and non-linear regularization models: a study on AMS (American Meteorological Society) 2013-14 solar energy prediction contest, Energy, 78, 247-256.

Almeida, M. P., Perpinan, O., & Narvarte, L. (2015). PV power forecast using a nonparametric PV model. Solar Energy, 115, 354-368.

Alobaidi, M.H., Marpu, P.R., Ouarda, T.B.M.J., & Ghedira, H. (2014). Mapping of the solar irradiance in the UAE using advanced artificial neural network ensemble, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7 3668-3680.

ARFF (2019), https://www.cs.waikato.ac.nz/~ml/weka/arff.html, (Access Date: 13 May 2019).

Aston Zhang and Zachary C. Lipton and Mu Li and Alexander J. Smola, Mu Li,& Alex J. Smola (2019). Dive into Deep Learning.

Baozhen Yao, Chao Chen, Qingda Cao, Lu Jin & Mingheng Zhang, Short-Term Traffic Speed Prediction for an Urban Corridor, Computer-Aided Civil and Infrastructure Engineering, 32, 154–169.

Ben-Bright, B., Zhan, Y., Ghansah, B., Amankwah, R., Wornyo, D. K., & Ansah, E. (2017). Taxonomy and a Theoretical Model for Feedforward Neural Networks, International Journal of Computer Applications, 163, 0975 – 8887.

Bhosale D., & Ade R. (2014). Roshani AdeFeature Selection based Classification using Naive Bayes, J48 and Support Vector Machine, International Journal of Computer Applications, Volume 99, 0975 – 8887.

Bilionis, I., Constantinescu, E.M. & Anitescu, M. (2014). Data-driven model for solar irradiation based on satellite observations, Sol. Energy 110, 22-38.

Bouzerdoum, M., Mellit, A. & Massi Pavan, A. (2013). A hybrid model (SARIMA-SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant, Sol. Energy 98 (Part C) 226-235.

Burrows, W. R. (1997). CART regression models for predicting UV radiation at the ground in the presence of cloud and other environmental factors. Journal of Applied Meteorology, 36(5), 531-544.

Cao, J., & Lin, X. (2008). Study of hourly and daily solar irradiation forecast using diagonal recurrent wavelet neural networks, Energy Convers. Manag. 49, 1396-1406.

Chakraborty, P., Marwah, M., Arlitt, M.F., & Ramakrishnan, N. (2012). Fine-grained photovoltaic output prediction using a bayesian ensemble, AAAI'12 Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence (pp. 274-280).

Chaouachi, A., Kamel, R.M., & Nagasaka, K. (2010). Neural network ensemble-based solar power generation short-term forecasting, JACIII. 14 , 69-75.

Chiong, R. (2009), Nature-Inspired Algorithms for Optimisation, Springer ISBN: 978-3-642-00266-3.

Chu, Y., Pedro, H., Li, M., & Coimbra, C. (2015). Real-time forecasting of solar irradiance ramps with smart image processing, Solar Energy 114 (2015) 91–104.

Chu, Y., Pedro, M. & Coimbra, C. (2013). Hybrid intra-hour DNI forecasts with sky image processing enhanced by stochastic learning, Solar Energy, 98, 592–603.

Cristianini N, & Shawe-Taylor J. An introduction to support vector machines (and other kernel-based learning methods) Cambridge: Cambridge University Press; 2000.

Demirtas, M., Yesilbudak, M., Sagiroglu, S., & Colak, I. (2012). Prediction of solar radiation using meteorological data. In Renewable Energy Research and Applications (ICRERA), 2012 International Conference on (pp. 1-4). IEEE.

Dong, Z., Yang, D., Reindl, T. & Walsh, W.M. (2015). A novel hybrid approach based on self-organizing maps, support vector regression and particle swarm optimization to forecast solar irradiance, Energy, Volume 82, 570-577.

Dr.Saurabh Mukherjee, Neelam Sharma, "Intrusion Detection Using Naive Bayes Classifier With Feature Reduction", Elsevier, pp. 119 – 128, 2012.

Energy (2019), https://www.enerji.gov.tr/tr-TR/Sayfalar/Elektrik, (Access Date: 13 May 2019).

F. Farhadi (2017), Learning Activation Functions In Deep Neural Networks, Université De Montréal.

Felice, M. D., Petitta, M. & Ruti, P.M. (2015). Short-term predictability of photovoltaic production over Italy, Renew. Energy, 80, 197-204.

Fernández, Á., Gala, Y., & Dorronsoro, J. R. (2014). Machine learning prediction of large area photovoltaic energy production. In International Workshop on Data Analytics for Renewable Energy Integration (pp. 38-53). Springer, Cham.

Ferrari, S., Lazzaroni, M., Piuri, V., Salman, A., Cristaldi, L., Rossi, M. & Poli, T. (2012). Illuminance prediction through extreme learning machines. In Environmental Energy and Structural Monitoring Systems (EESMS), 2012 IEEE Workshop on (pp. 97-103). IEEE.

Gaston, M., Pagola, I., Fernandez-Peruchena, C.M., Ramírez, L. & Mallor, F. (2010). A new Adaptive methodology of Global-to-Direct irradiance based on clustering and kernel machines techniques, in: 15th SolarPACES Conf., Berlin, Germany, 2010, (pp. 11693).

Günoğlu, K., Mavi, B. & Akkurt İ. (2011). Yapay Sinir Ağları (YSA) yöntemi ile global radyasyon tahmini, E-Journal of New World Sciences Academy, 6(2), 527-532.

Gurlek, C. & Sahin, M. (2018). Estimation of the global solar radiation with the artificial neural networks for the city of Sivas, European Mechanical Science, 2(2), 46-51.

Hagan, M. T., Demuth, H. B. & Beale, M. (1996). Neural Network Design, PWS Publishing Company, Boston, MA.

Haykin S. Neural networks. A comprehensive foundation. New York: Prentice Hall; 1999.

Himani Bhavsar, Mahesh H. Panchal (2012). A Review on Support Vector Machine for Data Classification, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 1, Issue 10, December 2012.

Hossain, R., Amanullah, M.T. & Ali, S. (2012). Hybrid prediction method of solar power using different computational intelligence algorithms, in: Univ. Power Eng. Conf. AUPEC 2012 22nd Australas., 2012, (pp. 1-6).

Hu, Y. & Hwang, J.N. (Eds.) (2002) Handbook of Neural Network Signal Processing. Boca Raton, FL: CRC Press.

Huang, J., Troccoli, A., & Coppin, P. (2014). An analytical comparison of four approaches to modelling the daily variability of solar irradiance using meteorological records. Renewable Energy, 72, 195-202.

Jiang, Y. (2008). Prediction of monthly mean daily diffuse solar radiation using artificial neural networks and comparison with other empirical models, Energy Policy, 36, 3833–3837.

Kalogirou, S. A. (2009). Solar Energy Engineering: Processes and Systems, 1st ed. Academic Press.

Koca, A., Oztop, H. F., Varol, Y. & Koca G. O. (2011). Estimation of solar radiation using artificial neural networks with different input parameters for Mediterranean region of Anatolia in Turkey, Expert Systems with Applications, 38(7), 8756-8762.

Kroese, B. & Smagt, V. D. (1996). An Introduction to Neural Networks (8th ed.). University of Amsterdam , Amsterdam, NL.

Křomer, P., Musílek, P., Pelikán, E., Krč, P., Juruš, P., & Eben, K. (2014, July). Support Vector Regression of multiple predictive models of downward short-wave radiation. In Neural Networks (IJCNN), 2014 International Joint Conference on (pp. 651-657). IEEE.

Lauret, P., Voyant, C., Soubdhan, T., David, M., & Poggi, P. (2015). A benchmarking of machine learning techniques for solar radiation forecasting in an insular context. Solar Energy, 112, 446-457.

Lazzaroni, M., Ferrari, S., Piuri, V., Salman, A., Cristaldi, L., & Faifer, M. (2015). Models for solar radiation prediction based on different measurement sites. Measurement, 63, 346-363.

Linares-Rodríguez, A., Ruiz-Arias, J. A., Pozo-Vázquez, D. & Tovar-Pescador, J. (2011). Generation of synthetic daily global solar radiation data based on ERA-Interim reanalysis and artificial neural networks, Energy, 36, 5356-5365.

Long, H., Zhang, Z., & Su, Y. (2014). Analysis of daily solar power prediction with data-driven approaches. Applied Energy, 126, 29-37.

Marquez, R., & Coimbra, C. F. (2011). Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the NWS database. Solar Energy, 85(5), 746-756.

McGovern, A., Gagne, D. J., Basara, J., Hamill, T. M., & Margolin, D. (2015). Solar energy prediction: An international contest to initiate interdisciplinary research on compelling meteorological problems. Bulletin of the American Meteorological Society, 96(8), 1388-1395.

Meteonorm (2019), https://meteonorm.com/en/, (Access Date: 13 May 2019).

Mihalakakou, G., Santamouris, M., & Asimakopoulos, D. N. (2000). The total solar radiation time series simulation in Athens, using neural networks. Theoretical and Applied Climatology, 66(3-4), 185-197.

Miroslaw Swiercz, Jan Kochanowicz, John Weigele, Robert Hurst, David S. Liebeskind, Zenon Mariak, Elias R. Melhem, and Jaroslaw Krejza, Learning Vector Quantization Neural Networks Improve Accuracy of Transcranial Color-coded Duplex Sonography in Detection of Middle Cerebral Artery Spasm—Preliminary Report, Neuroinformatics. 2008 Winter; 6(4): 279–290.

Moreno, A., Gilabert, M. A., & Martínez, B. (2011). Mapping daily global solar irradiation over Spain: a comparative study of selected approaches. Solar Energy, 85(9), 2072-2084.

Mori, H., & Takahashi, A. (2012). A data mining method for selecting input variables for forecasting model of global solar radiation. In Transmission and Distribution Conference and Exposition (T&D), 2012 IEEE PES (pp. 1-6). IEEE.

Mori, N.K.H. (2001). Optimal Regression Tree Based Rule Discovery for Short-term Load Forecasting, vol. 2, 2001, pp. 421-426.

Novakovic, J. (2009), Using Information Gain Attribute Evaluation to Classify Sonar Targets, 17th Telecommunications forum TELFOR 2009, Serbia, Belgrade, November (pp. 24-26).

Olaiya, F., & Adeyemo, A. B. (2012). Application of data mining techniques in weather prediction and climate change studies. International Journal of Information Engineering and Electronic Business, 4(1), 51.

Olatomiwa, L., Mekhilef, S., Shamshirband, S., Mohammadi, K., Petkovic, D., & Sudheer, C., (2015). A support vector machine–firefly algorithm-based model for global solar radiation prediction. Solar Energy 115 (2015) 632–644.

Ozgoren, M., Bilgili, M., & Sahin B. (2012). Estimation of global solar radiation using ANN over Turkey, Expert Systems with Applications, 39(5), 5043-5051.

Paoli, C., Voyant, C., Muselli, M., & Nivet, M. L. (2010). Forecasting of preprocessed daily solar radiation time series using neural networks. Solar Energy, 84(12), 2146-2160.

Paszkowicz, W. (2009). Genetic Algorithms, a Nature-Inspired Tool: Survey of Applications in Materials Science and Related Fields, Materials and Manufacturing Processes Volume 24, 2009 - Issue 2, 174-197.

Pedro, H. T., & Coimbra, C. F. (2015). Nearest-neighbor methodology for prediction of intra-hour global horizontal and direct normal irradiances. Renewable Energy, 80, 770-782.

Podestá, G. P., Núñez, L., Villanueva, C. A., & Skansi, M. A. (2004). Estimating daily solar radiation in the Argentine Pampas. Agricultural and Forest Meteorology, 123(1-2), 41-53.

Principe, J.C., Euliano, N.R. & Lefebvre, W.C. (2000). Neural and Adaptive Systems: Fundamentals Through Simulations. New York: Wiley.

Prokop, L., Misak, S., Snasel, V., Platos, J., & Kroemer, P. (2013). Supervised learning of photovoltaic power plant output prediction models, Neural Netw. World 23 (2013) 321-338.

Reikard, G. (2009). Predicting solar radiation at high resolutions: a comparison of time series forecasts, Sol. Energy, 83, 342-349.

Salcedo-Sanz, S., Casanova-Mateo, C., Pastor-Sánchez, A., & Sánchez-Girón, M. (2014). Daily global solar radiation prediction based on a hybrid Coral Reefs Optimization–Extreme Learning Machine approach. Solar Energy, 105, 91-98.

Samanta, M., Srikanth, B., & Yerrapragada, J.B. (2014). Short-Term Power Forecasting of Solar PV Systems Using Machine Learning Techniques.

Sandhya Samarasinghe (2007), Neural Networks for Applied Sciences and Engineering From Fundamentals to Complex Pattern Recognition, Taylor & Francis.

Suykens, J.A.K. & Vandewalle, J. (1999). Least Squares Support Vector Machine Classifiers, Neural Processing Letters 9: 293–300.

Şenkal, O. (2010). Modeling of solar radiation using remote sensing and artificial neural network in Turkey, Energy, 35(12), 4795-4801.

Şenkal, O. (2015). Solar radiation and precipitable water modeling for Turkey using artificial neural networks, Meteorology and Atmospheric Physics, 127(4), 481-488.

Şenkal, O. (2016). Solar Radiation Modeling for Turkey Using Atmospheric Parameters with Artificial Neural Networks. Çukurova Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi, 31(2), 179-185.

Şenkal, O., Şahin, M. & Peştemalci V. (2010). The estimation of solar radiation for different time periods, Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 32(13), 1176-1184.

Sözen, A., Arcaklioğlu, E., & Özalp, M. (2004). Estimation of solar potential in Turkey by artificial neural networks using meteorological and geographical data, Energy Conversion and Management, 45(18-19), 3033-3052.

Sözen, A., Arcaklioğlu, E., & Özalp, M. (2004). Estimation of solar potential in Turkey by artificial neural networks using meteorological and geographical data. Energy Conversion and Management, 45(18-19), 3033-3052.

Sözen, A., Arcaklıoğlu, E., Özalp, M., & Çağlar, N. (2005). Forecasting based on neural network approach of solar potential in Turkey, Renewable Energy, 30(7), 1075-1090.

Sozen, A., Ozalp, M., Arcaklioglu, E., & Galip Kanit E. (2004). A study for estimating solar resources in Turkey using artificial neural networks, Energy Sources, 26(14), 1369-1378.

Suykens J. A. K. & Vandewalle, J. (1999). Least Squares Support Vector Machine Classifiers, Neural Processing Letters, Volume 9, Issue 3, 293–30.

Taieb, S. B., Bontempi, G., Atiya, A. F., & Sorjamaa, A. (2012). A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. Expert systems with applications, 39(8), 7067-7083.

Towardsdatascience (2019), https://towardsdatascience.com/deep-learning-versus-biological-neurons-floating-point-numbers-spikes-and-neurotransmitters-6eebfa3390e9

Tso, G. K., & Yau, K. K. (2007). Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. Energy, 32(9), 1761-1768.

Tymvios, F., Jacovides, C.P., Michaelides, S., & Skouteli, C. (2004). A comparative study of Ǻngstrom's and artificial neural networks' methodologies in estimating global solar radiation, Solar Energy 78 (6), 752-762.

Vapnik, V. (1995). The Nature of Statistical Learning Theory (second ed.). Springer.

Vapnik, V., Golowich, S., & Smola, A. (1997). Support vector method for functionmapproximation, regression estimation, and signal processing. In M. Mozer, M. Jordan, and T. Petsche (Eds.), Advances in Neural Information Processing Systems 9, Cambridge, MA, pp. 281–287. MIT Press.

Voyant, C., Notton, G., Kalogirou, S., Nivet, M.L., Paoli, C., Motte, F. & Fouilloy, A. (2017). Machine learning methods for solar radiation forecasting: A review, Renewable Energy 105, 569-582.

Wallace, M., Anagnostopoulos, I.E., Mylonas, P., & Bieliková, M., (2010). Semantics in Adaptive and Personalized Services: Methods, Tools and Applications, Springer.

Weka (2019), https://www.cs.waikato.ac.nz/ml/weka/, (Access Date: 13 May 2019).

Wolff, B., Lorenz, E., & Kramer, O. (2016). Statistical learning for short-term photovoltaic power predictions. In Computational sustainability (pp. 31-45). Springer, Cham.

Wu, J., Chan, C. K., Zhang, Y., Xiong, B. Y., & Zhang, Q. H. (2014). Prediction of solar radiation with genetic approach combing multi-model framework, Renewable Energy, 66, 132-139.

Wu, J., Chan, C.K., Zhang, Y., Xiong, B.Y., & Zhang, Q.H. (2014). Prediction of solar radiation with genetic approach combing multi-model framework, Renew. Energy 66 132-139.

Yadav, A. K., Malik, H., & Chandel, S. S. (2014). Selection of most relevant input parameters using WEKA for artificial neural network based solar radiation prediction models. Renewable and Sustainable Energy Reviews, 31, 509-519.

Yang, H.T., Huang, C.M., Huang, Y.C., & Pai, Y.S. (2014). A weather-based hybrid method for 1-day ahead hourly forecasting of PV power output, IEEE Trans. Sustain. Energy 5, 917-926.

Yildiz, B., Şahin, M., Şenkal, O., Pestemalci, V., & Emrahoğlu, N. (2013). A Comparison of Two Solar Radiation Models Using Artificial Neural Networks and Remote Sensing in Turkey, Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 35(3), 209-217.

Yılmaz İ. H. (2018a). The effect of hot water consumption profile on the long-term system performance of a residential solar hot water system, SolarTR2018, Solar Conference and Exhibition. Istanbul, (pp. 135-142).

Yılmaz İ. H. (2018c). Optimization of an integral flat plate collector-storage system for domestic solar water heating in Adana, Anadolu University Journal of Science and Technology A- Applied Sciences and Engineering, 19(1), 165-176.

Yılmaz İ. H. (2018d). Residential use of solar water heating in Turkey: A novel thermo-economic optimization for energy savings, cost benefit and ecology, Journal of Cleaner Production, 204, 511-524.

Yılmaz, İ. H. & Mwesigye A. (2018). Modeling, simulation and performance analysis of parabolic trough solar collectors: A comprehensive review, Applied Energy, 225, 135-174.

Yılmaz, İ. H. & Söylemez M. S. (2014). Thermo-mathematical modeling of parabolic trough collector, Energy Conversion and Management, 88, 768-784.

Yılmaz, İ. H. & Söylemez, M. S. (2012). Design and computer simulation on multi-effect evaporation seawater desalination system using hybrid renewable energy sources in Turkey, Desalination, 291, 23-40.

Yılmaz, İ. H. (2018b). Optimization of a molten salt concentrating solar power tower heliostat field: Case study for Adana, SolarTR2018, Solar Conference and Exhibition. Istanbul, (pp. 155-159).

Zamo, M., Mestre, O., Arbogast, P., & Pannekoucke, O. (2014). A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production, part I: Deterministic forecast of hourly production. Solar Energy, 105, 792-803.

CURRICULUM VITAE

Özge KAYA was born in Adana, Turkey, in 1993. She received a B.S. degree from the Department of Industrial Engineering from Erciyes University, Turkey, in 2016.

She worked as an manufacturing engineer at Kurmel Group Aktav Akustik Malz. San. Ve Tic A.Ş. from 2017 to 2018. Since 2018, she has been working as manufacturing engineer at Aşut Fiberglass.

Her research interests are in forecasting, computer software and optimization.