ACSIJ Advances in Computer Science: an International Journal, Vol. 4, Issue 6, No.18 , November 2015
ISSN : 2322-5157
www.ACSIJ.org

Select the most relevant input parameters using WEKA for models forecast Solar radiation based on Artificial Neural Networks

Somaieh Ayalvary[1], Zohreh Jahani[2] and Morteza Babazadeh[3]

[1] **Department of Computer Engineering, ISLAMIC AZAD UNIVERSITY*Babol Branch***
Langerud*, 013/ *4253*, *Iran
Somaieh.Ayalvary@gmail.com

[2] **Department of Computer Engineering, ISLAMIC AZAD UNIVERSITY*Babol Branch***
Gorgan*, 017/1235, *Iran
zohreh.jahanii@gmail.com

[3] **Department of Computer Engineering, ISLAMIC AZAD UNIVERSITY*Babol Branch***
Babol*, 011/3241, *Iran
Morteza_babazade@yahoo.com

Abstract
Forecasting solar radiation is important for many applications in research related to renewable energy. Solar radiation is forecasted by solar radiation forecast models including the traditional models and artificial neural network (ANN) based model. There are geographical and meteorological variables that affect the solar radiation, thus identifying the appropriate variables to forecast solar radiation correctly is an important issue in the research area. Accordingly Waikato Environment for Knowledge Analysis (WEKA) Software was used in 11 points in Guilan based on different weather conditions to find the most effective input parameters to forecast solar radiation in different ANN models. Input parameters include latitude, longitude, maximum wind speed, average temperatures in each month, the average maximum air temperature, average minimum air temperature, sunshine, monthly rainfall, maximum rainfall in a day for different cities of Gilan. In order to check the reliability of the forecasts by known parameters, three ANN models have developed (ANN-1, ANN-2 and ANN-3). The maximum MAPE for ANN-1, ANN-2 and ANN-3 equals 22.15%, 20.29% and 22.14%, respectively indicating 1.86% improvement in the accuracy in the prediction of ANN-2.
Keywords: *Neural Network, Data Mining, WEKA*

1. Introduction

Solar energy is a clean source of energy with high potential to meet the needs related to energy. The assessment of solar potential energy requires information on solar radiation in different places. In order to find the most relevant parameters, the variables must be selected by combining various input parameters so that the best forecast is made which is time consuming. Therefore, in this study WEKA software version 7.3.10 is used to forecast solar radiation at 11 points in Guilan Province with different climatic conditions which is discussed below. In order to check the accuracy of prediction, ANN models have been developed (ANN-1, ANN-2 and ANN-3). Model ANN- 1 was created using input parameters while model ANN-2 was provided by the most relevant parameters presented by WEKA software and model ANN-3 was presented by removing the relevant parameters that can be used to forecast solar radiation at some pints in Guilan. The paper is set out as follows; Evaluation studies in section 2 and databases and methods are presented in Section 3. Results in Section 4 and conclude have come in Section 5.

2. Research about Identify input parameters to forecast Solar radiation based on ANN

ANN models use different meteorological and geographical variables of an area as the input data to forecast solar radiation. Azeez [1] used back propagation neural network to estimate average monthly solar radiation in Gusau, Nigeria. The duration of radiation, the average prevailing temperature and relative humidity were considered as output. Statistical analysis (R = 99.96, MPE = 0.8512, RMSE = 0.0028) showed the best agreement between measured and estimated values of solar global radiation.

Linares-Rodriguez And colleagues [2] used From MLP model estimate solar radiation in Spain by using irradiation obtained by satellites. The input layer has 12 inputs (11 channels Meteosat and radiation of the sun on a clear day). RMSE is equal to 6.74%. The model works well in cloud and sunny weather conditions. According to studies, it was found that the accuracy of forecasting ANN model change by using geographic and meteorological variables as input parameters. To select the relevant input parameters, the researcher must use various combinations of input parameters to assess the accuracy of forecast models ANN. That It requires a lot of computational analysis. Therefore choosing the most relevant input parameters for the ANN models is and important research failure addressed in this study.

3. Methods

3.1. Solar radiation data source

11 selected locations in different climate zones in Gilan which were used to try and test the ANN models, came in Figure 3. Data from these stations Meteorological Organization in Gilan Province in the study for an average of four years, from 1387 to 1390, are presented in tables 11 and 12.

3.2. Select the input variables using WEKA

Choosing the input variables is the first step is to develop ANN models. The Input test data including temperature, maximum temperature, minimum temperature, altitude, hours of radiation, latitude and longitude for the solar radiation models are obtained by Table 1. In the process of selection of variables, the most relevant input variables should be evaluated to forecast solar radiation. To select the related input variables the feature evaluator and search method are selected as the result of which all variables are observed. The rank of each input variable as determined by WEKA to forecast solar radiation is presented in figures 3, 6 and 9.
The latitude variables have lowest rank. So, At Select the relevant input variables To calculate the accuracy of forecasting solar radiation, Latitude removed from the input vector X.And The accuracy of prediction was calculated by using ANN based on relevant input variable. After solving the problem of selection of variables, Three ANN (ie ANN-1, ANN-2 and ANN-3) were developed to calculate predictive accuracy. ANN-1 model uses variables of average monthly air temperature (T), the average minimum temperature (Tmin), the average maximum temperature (Tmax), the maximum wind speed (meters per second)) wind (m / s , sunshine (hours) (SH), the maximum daily rainfall (mm), rain (mm) day monthly rainfall (mm), rain (mm) month and latitude. ANN-2 model uses From The most relevant variables obtained from WEKA(Average air temperature in each month (T), the average minimum temperature (Tmin), the average maximum temperature (Tmax), the maximum wind speed (meters per second)) wind (m / s, sunshine (hours) (SH), the maximum daily rainfall (mm), rain (mm) day monthly rainfall (mm) rain (mm) month) and ANN-3 model uses variables, the average minimum temperature (Tmin), the average maximum temperature (Tmax), the maximum wind speed (meters per second)) wind (m / s, sunshine (hours) (SH), the maximum daily rainfall (mm), rain (mm) day and monthly rainfall (mm) rain (mm) month.[3,4]

3.3. Predictive models of solar radiation with selective input

ANN Models (ANN-1, ANN-2 and ANN-3) have been created by network fitness tools that are used to forecast.
The number of neurons in the hidden layer is evaluated by equation (1) [5,6] where H_n and S_n are the number of hidden layer neurons and sample data used in the ANN model, I_n and O_n also indicated input and output parameters.

$$Hn = (In + On) / 2 + \sqrt{Sn} \qquad (1)$$

Sensitivity tests to validate the number of hidden layer neurons is performed by calculating the change in prediction error (MAPE) at the time of change in the number of neurons in the hidden layer as ± 5 of hidden layer neurons calculated by the equation (1). The analysis of neurons' sensitivity is done for ANN models; MAPE was obtained by Equation 2 and ANN structure by minimum MAPE is used to forecast solar radiation.

$$\mathbf{MAPE} = \left(\left(1/n \sum_{i=1}^{n} |H/H0 - \hat{H} / \hat{H}0| \right) \times 100 \right) / \hat{H} / \hat{H}0 \qquad (2)$$

4. Discussion

In this system, we mean absolute percentage error (MAPE) considered as a parameter. To help ranking software "Weka" we find that At Model ANN1, latitude has least importance.

Table 1: meteorological data and geographic coordinates for the 11 cities in the Gilan ANN1

City	Wind	T	Tmax	Tmin	SH
Astara	9.7	15.96	19.42	12.43	144.57
Anzali	16.67	17.08	19.41	14.7	149.84
Jirandeh	24.16	12.72	17.05	8.34	221.83
Rudesar	9.16	16.97	20.51	13.4	136.23
Rasht	12.16	16.74	21.3	12.12	130.85
Kiashahr	12.08	16.87	20.44	13.27	127.85
Talesh	11.58	16.59	19.84	13.09	116.34
Masuleh	10.66	12.55	16.19	8.86	111.22
Manjal	17.83	18.41	23.53	13.24	233.61
Lahijan	7.83	16.91	21.38	12.39	141.2
Dealaman	13.08	12.07	16.9	7.21	153.48

Table 2: meteorological data and geographic coordinates for the 11 cities in the Gilan ANN1

City	Rain month	Rain day	Lat	Long	H/H0	Ĥ/Ĥ0
Astara	94.27	41.44	55	38	0.5	0.6026
Anzali	123.97	38.47	49	37	0.41	0.6026
Jirandeh	21.83	7.79	50	37	0.56	0.6043
Rudesar	109.17	42.66	50	37	0.5	0.6021
Rasht	102.06	32.23	49	37	0.48	0.6019
Kiashahr	83.22	21.2	56	32	0.51	0.6018
Talesh	102.79	39.9	48	37	0.42	0.6012
Masuleh	75.75	20.29	49	37	0.43	0.6009
Manjal	13.33	6.46	49	36	0.51	0.6045
Lahijan	111.13	35.65	50	37	0.34	0.6023
Dealaman	33.04	14.14	49	37	0.5	0.6028

If the amount of solar radiation on the earth's surface (H / H0) is considered in the WEKA class software, the ranking is obtained as follows:
Where the longitude, latitude and average temperature are the input values that have the least impact on the results and have a lower rank.
Table 3 presents the input variable ranks by WEKA algorithm to predict solar radiation in the model ANN1 that contains all input values (Latitude, longitude, maximum wind speed, with average temperatures in each month, the average maximum air temperature, average minimum air temperature, sunshine hours, rainfall monthly, maximum rainfall in a single day as our workload to the system.)

Table 3: The number of input variables by WEKA algorithm to predict solar radiation in the ANN1

Rank	*attributes*
0.083	Wind
0.0495	Rain(mm)month
0.0324	SH
0.0188	Rain(mm)day
-0.0178	Tmin
-0.0217	Tmax
-0.0328	T
-0.0587	Lat
-0.0611	Long

In ANN2 model the longitude and latitude parameters must be removed, the maximum wind speed, average air temperature in each month, the mean maximum air temperature, the average minimum air temperature, hours of sunshine, monthly rainfall and maximum rainfall in one day are entered to the system as the workload.
For this purpose we perform data mining for 11 cities in Guilan province and remove two cities in which the input values for longitude and latitude are more important (in Manjil and Lahijan). In this model the temperature parameter has the least importance.

Table 4: meteorological data and geographic coordinates for the city of Gilan 9 ANN2 model

City	Wind	T	Tmax	Tmin
Astara	9.7	15.96	19.42	12.43
Anzali	16.67	17.08	19.41	14.7
Jirandeh	24.16	12.72	17.05	8.34
Rudesar	9.16	16.97	20.51	13.4
Rasht	12.16	16.74	21.3	12.12
Kiashahr	12.08	16.87	20.44	13.27
Talesh	11.58	16.59	19.84	13.09
Masuleh	10.66	12.55	16.19	8.86
Dealaman	13.08	12.07	16.9	7.21

Table 5: meteorological data and geographic coordinates for the city of Gilan 9 ANN2 model

City	SH	Rain month	Rain day	H/H0	Ĥ/Ĥ0
Astara	144.57	94.27	41.44	0.5	0.6026
Anzali	149.84	123.97	38.47	0.41	0.6026
Jirandeh	221.83	21.83	7.79	0.56	0.6043
Rudesar	136.23	109.17	42.66	0.5	0.6021
Rasht	130.85	102.06	32.23	0.48	0.6019
Kiashahr	127.85	83.22	21.2	0.51	0.6018
Talesh	116.34	102.79	39.9	0.42	0.6012
Masuleh	111.22	75.75	20.29	0.43	0.6009
Dealaman	153.48	33.04	14.14	0.5	0.6028

Table 6: The number of input variables by WEKA algorithm to predict solar radiation in the ANN2

Rank	*attributes*
0.083	Wind
0.0495	Rain(mm)month
0.0324	SH
0.0188	Rain(mm)day
-0.0178	Tmin
-0.0217	Tmax
-0.0328	T

Temperature parameter should be eliminated in model ANN3,The maximum wind speed, the mean maximum air temperature, hours of sunshine, monthly rainfall and maximum rainfall in one day are entered to the system as the workload.
For this purpose one of the 11 cities in Guilan province in which temperature parameter has the highest importance (Jirandeh) is eliminated.

Table 7: meteorological data and geographic coordinates for the 8 cities of Gilan in the ANN3

City	Wind	T	Tmax	Tmin	SH
Astara	9.7	15.96	19.42	12.43	144.57
Anzali	16.67	17.08	19.41	14.7	149.84
Rudesar	9.16	16.97	20.51	13.4	136.23
Rasht	12.16	16.74	21.3	12.12	130.85
Kiashahr	12.08	16.87	20.44	13.27	127.85
Talesh	11.58	16.59	19.84	13.09	116.34
Masuleh	10.66	12.55	16.19	8.86	111.22
Dealaman	13.08	12.07	16.9	7.21	153.48

Table 8: meteorological data and geographic coordinates for the 8 cities of Gilan in the ANN3

City	Rain month	Rain day	Lat	Long	H/H0	Ĥ/Ĥ0
Astara	94.27	41.44	55	38	0.5	0.6026
Anzali	123.97	38.47	49	37	0.41	0.6026
Rudesar	109.17	42.66	50	37	0.5	0.6021
Rasht	102.06	32.23	49	37	0.48	0.6019
Kiashahr	83.22	21.2	56	32	0.51	0.6018
Talesh	102.79	39.9	48	37	0.42	0.6012
Masuleh	75.75	20.29	49	37	0.43	0.6009
Dealaman	33.04	14.14	49	37	0.5	0.6028

Table 9: The number of input variables by WEKA algorithm to predict solar radiation in the ANN3

Rank	***attributes***
0.083	Wind
0.0495	Rain(mm)month
0.0324	SH
0.0188	Rain(mm)day
-0.0178	Tmin
-0.0217	Tmax

With the help of equation (2) to obtain the mean absolute percent error for three ANN1, ANN2 and ANN3 The mean absolute percent error for the ANN1:

$$MAPE = (1 / n\ \mathcal{E}\ ni = 1\ |\ (H / H0 - \hat{H} / \hat{H}0) / \hat{H} / \hat{H}0\ |) \times 100 = 22.15 \quad (3)$$

The mean absolute percent error for the ANN2:

$$MAPE = (1 / n\ \mathcal{E}\ ni = 1\ |\ (H / H0 - \hat{H} / \hat{H}0) / \hat{H} / \hat{H}0\ |) \times 100 = 20.29 \quad (4)$$

The mean absolute percent error for the ANN3:

$$MAPE = (1 / n\ \mathcal{E}\ ni = 1\ |\ (H / H0 - \hat{H} / \hat{H}0) / \hat{H} / \hat{H}0\ |) \times 100 = 22.14 \quad (5)$$

Model ANN2 is able to better forecast solar radiation for having the lowest error.
Based on the result of the calculation we find that the ANN2 model outperforms model ANN3 and the performance of model ANN3 is higher than ANN1.
The error results indicate that model ANN2has the lowest error and it is considered as the best model.
We want to obtain the maximum and minimum MAPE error obtained by equation (2) using algorithms GA and SA described below.
For this purpose the values of H / H0 and Ĥ / Ĥ0 are considered in the range between 0 and 1.
For each of these functions Simulated Annealing (SA) algorithms and genetic algorithm are used and compared.
This algorithm is made of two nested loops. First the temperature is very high (in our simulation randi (100,000)), so particle displacement is very high.

With decreasing temperature (Increasing the number of reps and getting closer to the answer), particle displacement is reduced and local search will occur.
One major difference between SA and GA is that SA is a single factor algorithm while GA is a population or multi-factor algorithm, thus GA gets a better answer and the possibility of local minimum is low. For example, in our simulations, SA algorithm is repeated more than 15,000 times in each run to reach the optimal solution, while the genetic algorithm is run 10 times and assuming a total of 40 children in each run, it is performed 400 times and has reached the optimal solution[7].

Simulation equation 2 with SA

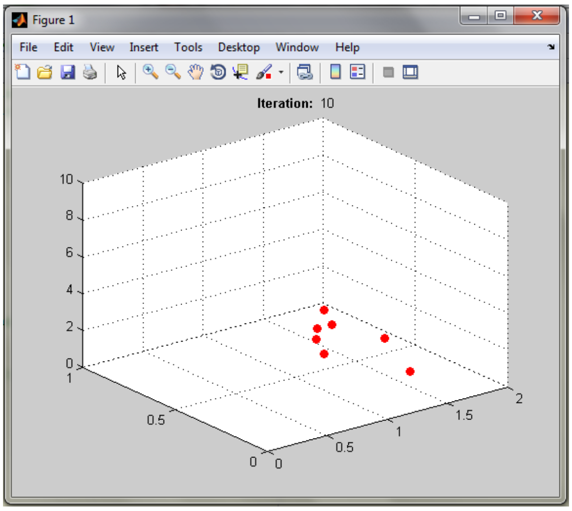

Figure 1: Output Simulation shows the minimum MAPE error.

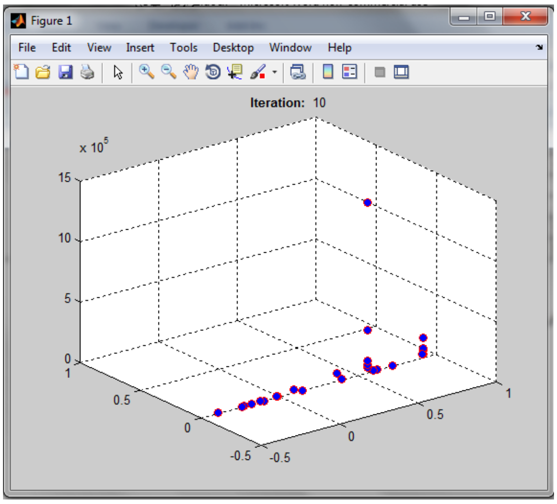

Figure 2: Output Simulation shows the maximum MAPE error

Table 10: Abbreviations used

Lat	Latitude
Wind	Longitude
T	The average air temperature per month
Tmax	The average maximum air temperature
Tmin	The average minimum air temperature
SH	Sunshine hours(hrs)
Rain month	Monthly Rainfall(mm)
Rain day	The maximum daily rainfall(mm)
MAPE	The mean absolute percentage error
H/H0	Monthly average daily solar radiation forecast data for the month i
Ĥ/Ĥ0	Monthly average daily solar radiation data measured for month i

Figure 3: The above image is a map of Gilan That shows the selected cities to examine and test the model ANN.

Table 11: meteorological data and geographic coordinates for the 11 cities in Gilan

City	Wind	T	Tmax	Tmin	SH
Astara	9.7	15.96	19.42	12.43	144.57
Anzali	16.67	17.08	19.41	14.7	149.84
Jirandeh	24.16	12.72	17.05	8.34	221.83
Rudesar	9.16	16.97	20.51	13.4	136.23
Rasht	12.16	16.74	21.3	12.12	130.85
Kiashahr	12.08	16.87	20.44	13.27	127.85
Talesh	11.58	16.59	19.84	13.09	116.34
Masuleh	10.66	12.55	16.19	8.86	111.22
Manjal	17.83	18.41	23.53	13.24	233.61
Lahijan	7.83	16.91	21.38	12.39	141.2
Dealaman	13.08	12.07	16.9	7.21	153.48

Table 12: meteorological data and geographic coordinates for the 11 cities in Gilan

City	Rain month	Rain day	Lat	Long	H/H0
Astara	94.27	41.44	55	38	0.5
Anzali	123.97	38.47	49	37	0.41
Jirandeh	21.83	7.79	50	37	0.56
Rudesar	109.17	42.66	50	37	0.5
Rasht	102.06	32.23	49	37	0.48
Kiashahr	83.22	21.2	56	32	0.51
Talesh	102.79	39.9	48	37	0.42
Masuleh	75.75	20.29	49	37	0.43
Manjal	13.33	6.46	49	36	0.51
Lahijan	111.13	35.65	50	37	0.34
Dealaman	33.04	14.14	49	37	0.5

The first model based on the parameters of the sundial, estimates the amount of radiation on a horizontal surface is Angstrom empirical equation (2) and Prescott (8).

$$\hat{H} / \hat{H}0 = a + b\,(\tilde{n} / \tilde{N}) \qquad (6)$$

In the above equation Ĥ represents the total daily radiation per month, Ĥ0 represents measured radiation outside the atmosphere, ñ is the average monthly hours of sunshine daily, Ñ is the average monthly peak sunshine hours (during the day).[8]

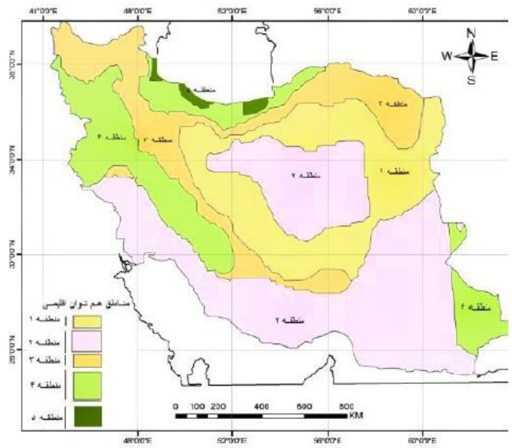

Figure 4: climatic classification

In the above map Iran is divided into five regional areas Gilan province is in the region of five. The parameters a and b are coefficients fixed equation That at Region Five are Equal 0.404 and 0.204 respectively.

$$H / H0 = (KT) \times (TD)^{0.5} \qquad (7)$$

$$TD = Tmax - Tmin \qquad (8)$$

In the above equation are H and H0 radiation reaching the Earth and Extraterrestrial radiation in calories per square centimeter per day (Cal \ cm2 day), respectively. TD is Daily temperature range (0C) and KT is constant coefficient equation that

For coastal and non-coastal are 0.19 and 0.16 respectively.

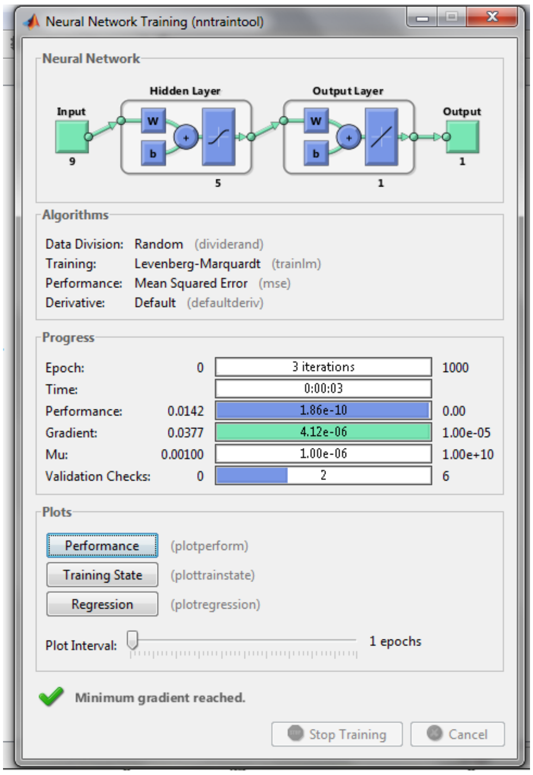

Figure 5: Evaluation of neural networks ANN-1

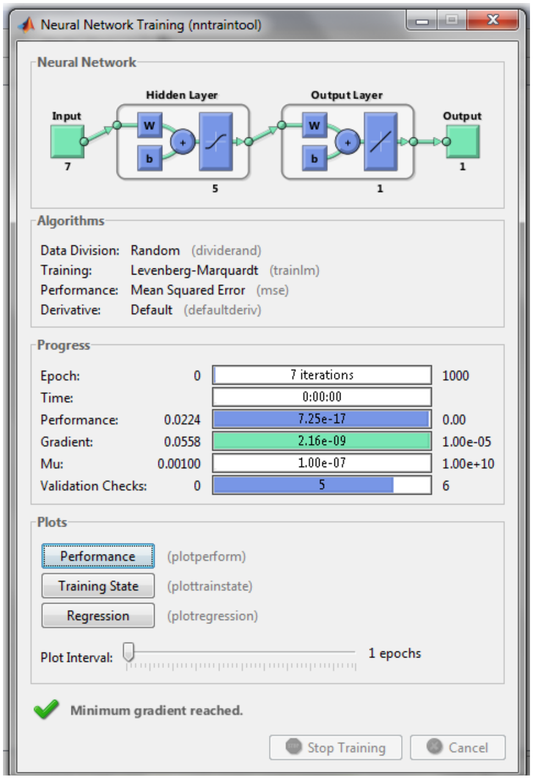

Figure 6: Evaluation of neural networks ANN-2

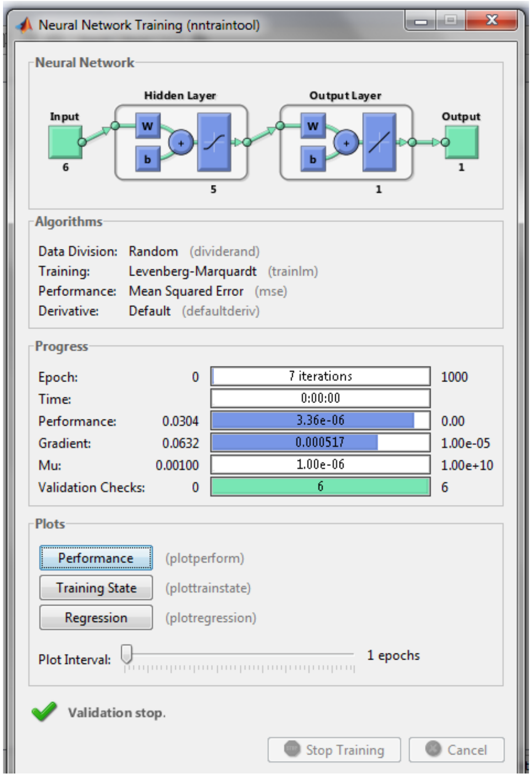

Figure 7: Evaluation of neural networks ANN-3

5. Conclusion

The present study shows the robust nature of WEKA in evaluating the most effective parameter to forecast solar radiation using ANN. It was found that the most relevant parameters to forecast solar radiation include temperature, maximum temperature, minimum temperature, altitude, and the hours of sunshine.Maximum MAPE for models ANN-1, ANN-2 and ANN-3 equals 22.15%, 20.29% and 22.14%, respectively indicating a high level of accuracy for ANN-2 that have used the most relevant input variables. ANN-2 model developed could be used to forecast solar radiation at any location in Guilan. Further studies can be made for more accurate estimation of the solar potential of the area. The future study should be focused on finding the most relevant parameters of the meteorological variables with improved forecast accuracy in different ANN models. We made a 5 layer neural network by data collected from 11 cities in Guilan. Input reduction and elimination of longitude and latitude from the input data caused the model ANN2 to be built more quickly than ANN1. As a result in this model we observe higher performance ability (Throughput).[9]

References

[1] Azeez MAA. Artificial neural network estimation of global solar radiation using meteorological parameters in Gusau, Nigeria. Arch Appl Sci Res 2011;3 (2):586–95.

[2] Linares-Rodriguez A, Ruiz-Arias JA, Pozo-Vazquez D, Tovar-Pescador J. An artificial neural network ensemble model for estimating global solar radiation from Meteosat satellite images. Energy 2013;61:636–45.

[3] Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA Data Mining Software: An Update; SIGKDD Explorations 2009; 11(1). [accessed 02.02.2013] ⟨http://dx.doi.org/http://www.cs.waikato.ac.nz/~eibe/ pubs/weka_update.pdf⟩.

[4] Witten IH, Frank E, Hall MA. Data mining: practical machine learning tools and techniques. A Book of Morgan Kaufmann Publishers. 3rd ed.2011.

[5] Chow SKH, Lee EWM, Li DHW. Short-term prediction of photovoltaic energy generation by intelligent approach. Energy Build 2012;55:660–7.

[6] Frederick M. Neuroshell 2 Manual, Ward Systems Group Inc., 1996.

[7] Bakirci K. Models of solar radiation with hours of bright sunshine: a review. Renew Sustain Energy Rev 2009;13:2580–8.

[8] Khatib T, Mohamed A, Sopian K. A review of solar energy modeling techni-ques. Renew Sustain Energy Rev 2012;16:2864–9.

[9] Rehman S, Mohandes M. Estimation of diffuse fraction of global solar radiation using artificial neural networks. Energy Sources, Part A 2009;31: 974–84.