Air temperature prediction using different machine learning models

Rana Muhammad Adnan[1], Zhongmin Liang[2], Alban Kuriqi[3], Ozgur Kisi[4], Anurag Malik[5], Binquan Li[6], Fatemehsadat Mortazavizadeh[7]
[1,2,6,7]College of Hydrology, Hohai University Nanjing 210098, China
[3]CERIS, University of Lisboa Lisbon 1049-001, Portugal
[4]Ilia State University Tbilisi, Georgia
[5]Punjab Agricultural University Bathinda-151001, Punjab, INDIA

Article Info

Article history:

Received Oct 18, 2020
Revised Jan 6, 2021
Accepted Jan 20, 2021

Keywords:

Astore river
CART
Gilgit river
GMDH-NN
LSSVM

ABSTRACT

Air temperature is an essential climatic component particularly in water resources management and other agro-hydrological/meteorological activities planning This paper examines the prediction capability of three machine learning models, least square support vector machine (LSSVM), group method and data handling neural network (GMDHNN) and classification and regression trees (CART) in air temperature forecasting using monthly temperature data of Astore and gilgit climatic stations of Pakistan. The prediction capability of three machine learning models is evaluated using different time lags input combinations with help of root mean square error (RMSE), the mean absolute error (MAE) and coefficient of determination (R2) statistical indicators. The obtained results indicated that the LSSVM model is more accurate in temperature forecasting than GMDHNN and CART models. LSSVM significantly decreases the mean RMSE of the GMHNN and CART models by 1.47-3.12% and 20.01-25.12% for the Chakdara and Kalam Stations, respectively.

Corresponding Author:

Binquan li
College of Hydrology
Hohai University
Nanjing 210098, China
Email: libinquan@hhu.edu.cn

1. INTRODUCTION

Air temperature represents a crucial meteorological variable that affects several meteorological and hydrologic processes at different spatial and temporal scale [1]. In case of water resources management or land evaluation, temperature is usually applied as an input variable to derive other parameters such as degree of soil degradation, vegetation growth and evapotranspiration resulting from moister content in soil and plants [2]. Accurate prediction of the temperature have an essential contribution for the planning of the most suitable agriculture site for crop planting, especially in those regions characterised by high intra- and inter-annual weather conditions variability [3].

Nevertheless, at many countries, in particular, undeveloped or developing countries distribution of the ground-based meteorological station is not uniformly distributed or insufficient to make a robust large-scale spatial characterization of the weather and climate conditions [4]. In recent years, increasing awareness towards global warming has drawn attention not only of the scientist but also decision-makers and other related stakeholders. According to different climate models, during this century is excepted a continues increase of the earth surface temperature at the global scale, which in turn may result in a wide range of consequences on ecosystem and humans as well [5]. In these circumstances, in addition to technology advancement, there are essential needs of developing and applying novel models that allow for accurate estimation and which help to address the variability of the air temperature more accurately. Concerning the modelling part, in recent years application of novel artificial intelligence-based algorithms has been successfully applied in a wide range of scientific domains, including water resources management, agro-hydrology and agro-meteorology among others [1-3]. In general, in the field of water resources, the majority of studies have compared the performance of the different algorithm on the prediction of different meteorological and hydrological variables. Kisi and Shiri [1] applied the adaptive neuro-fuzzy inference system (ANFIS) and artificial neural networks (ANN) algorithm to predict the long-term monthly temperature in an arid and semiarid region, and they found that the ANN model performed better than ANFIS. Sanikhani, *et al.* [2] evaluated the performance of four intelligent models, respectively: Multivariate adaptive regression splines (MARS), generalized regression neural network (GRNN), random forest (RF), and extreme learning machines (ELM) at a humid region to predict air temperature based only on geographic data. The findings of the study above showed that the GRNN model could perform better. Whereas, in the study conducted in [3], the authors assessed two intelligent algorithms, respectively; feed-forward back propagation (FFBP) and GRNN to predict several parameters of air temperature, which then were compared to multiple linear regression (MLR) model. The finding of that study showed that all models provide satisfactory results in terms of several performance criteria. Adnan, *et al.* [6] predicted reference evapotranspiration (ET0) in the China by evaluating different data driven models based solely on-air temperature data. Although there is plethora of studies that have already explored the potential use of several intelligent models at different sub-filed of water resources management, the present study aims to bring into light the potential of some other novel intelligent models which to the best of our knowledge are scarcely applied in the field of climatology, particularly in the air temperature prediction. In this study, we have evaluated the robustness of the least squares support vector machine (LSSVM), data handling-type neural network (GMDH-NN) and classification and regression tree (CART) for predicting long-term monthly air temperature. Yu, *et al.* [7] applied LSSVM to predict solar greenhouse temperatures, and they found that LSSVM could predict more accurately both maximum and minimum temperature. Similar conclusions were obtained from the other researcher as well [8-10]. While GMDH-NN was reported to predict accurately flow discharge [11], offshore wind speed [12], and hydraulic conductivity [13]. Finally, CART model has also been reported to provide satisfactory results in terms of the performance in a different field such as soil carbon prediction [14, 15], flood risk mapping [16], and earth surface temperature estimation among other [17]. Thus, the find of this study supported from similar conclusion reported in the previous studies will enrich further the diversity of intelligent models that could be applied in the field of water resources management.The successful aplications of LSSVM,GMDH-NN and CART models in literature compelled us to select in this study. To the best of our knowledge, there is no study reported in literature that compare the prediction accuracy of three slected models in air temperature modeling. This give imputes to research.

2. RESEARCH METHODS

2.1. Least square support vector machine (LSSVM)

The concept of least square support vector machine (LSSVM) was first introduced by Suykens and Vandewalle [18] to solve the classification and regression issues based on linear separation theory [19]. LSSVM is the modified version of the support vector machine (SVM), but both work on different principles in solving the hyperplane [20]. In the SVM model, quadratic programming is used to optimize the parameters of the hyperplane, while the LSSVM model uses linear programming to solve the problems [21]. Figure 1 illustrates the hierarchical network of the LSSVM model with inputs (Xi) and output (Y) series. The regression function is defined in (1) with algebraic function F(X) and permissible error (ε) (1)-(2) [22]:

$$SY = F(X) + \varepsilon \tag{1}$$

$$F(X) = \omega^T \delta(X) + \beta \tag{2}$$

where, Y = dependent variable, X = independent variable, ω^T = weighted factor, δ = kernel function, and β = characteristic constant of regression function or bias term.

A variety of kernel functions (i.e. linear, polynomial, radial basis, and spline) have been available to solve the regression or mapping problems [23-25]. The current research utilized radial basis function (RBF) as kernel function for mapping the data into a high dimensional feature space and expressed by (3) as:

$$K(X, X_i) = exp\left(-\frac{\|X - X_i\|^2}{\mu}\right) \quad (3)$$

where, K = gram matrix obtained according to the samples, and μ = RBF kernel function parameter.

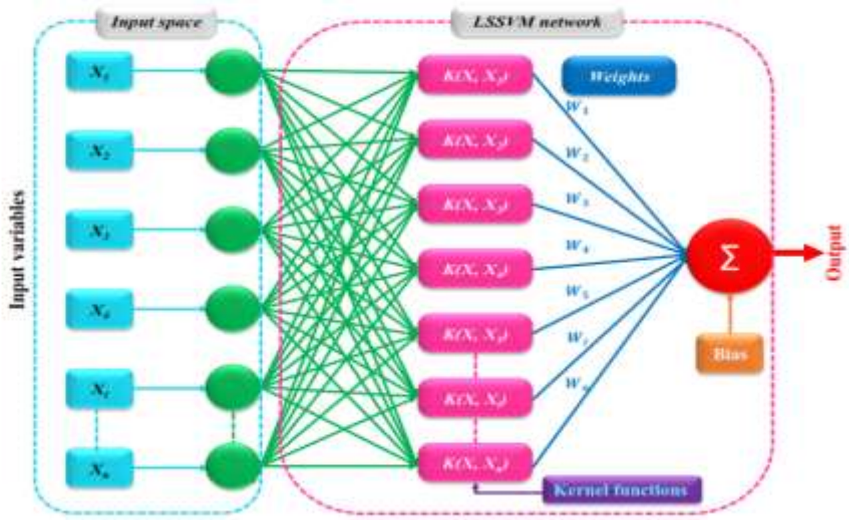

Figure 1. The architecture of LSSVM model

2.2. Group method and data handling neural network (GMDHNN)

The group method and data handling (GMDH) approach was invented by Ivakhnenko [26] for mathematical modelling of multivariate complex systems [27]. The GMDH model is reliant on efficacy with multi-input and single-output data sets based on reference polynomial functions [28]. The connection between the multi-input-single-output expressed by the Volterra function series which is discrete analogous of the polynomial of Kolmogorov-Gabor [26] and expressed as:

$$y = a + \sum_{i=1}^{m} b_i x_i + \sum_{i=1}^{m}\sum_{j=1}^{m} c_{ij} x_i x_j + \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m} d_{ijk} x_i x_j x_k + \cdots$$

where, y is the output vector; a, b and c are the coefficients of the polynomial; x_i, x_j, x_k are multi inputs (i, j, k = 1, 2, 3, …, m). In this research, the GMDH integrated with neural network (NN) for modelling monthly temperature at Astore and Gilgit meteorological stations. Figure 2 shows the flowchart of the implemented GMDHNN model.

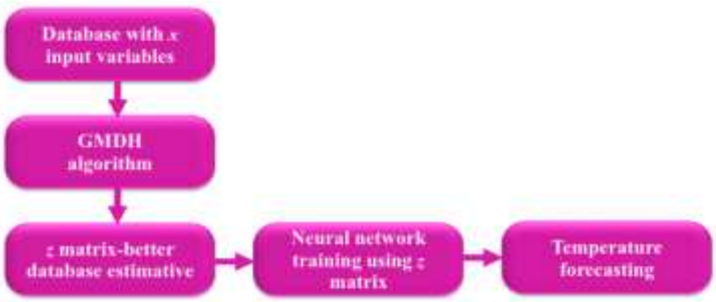

Figure 2. The architecture of GMDHNN model

2.3. Classification and regression trees (CART)

Breiman *et al.* [29] exposed the classification and regression trees (CART) model for splitting a sample into the gradually reduced subclasses (or generates the binary decision trees) [30, 31]. Furthermore, the CART model resolves the classification problems either from absolute (categorical) or continuous dependent variables [32]. The CART model constructs a classification tree based on the absolute dependent variable and a regression tree-based on the continuous dependent variable [32]. Figure 3 demonstrates the working principle of the CART model. The most commonly used evaluation function for splitting in the CART model is the impurity GINI index and written as [31]:

$$GINI(k) = 1 - \sum_i p_i^2 \quad (4)$$

where, p_i = probability of class i in node k.

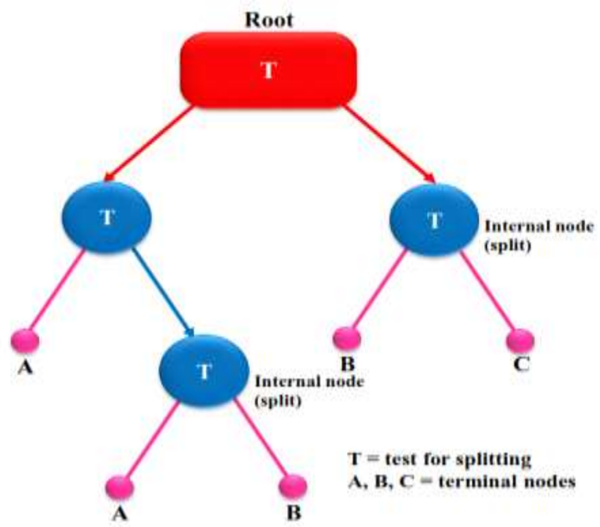

Figure 3. The architecture of CART model

3. CASE STUDY AND PERFORMANCE INDICATORS

Monthly temperature data got from two stations Astore and Gilgit, Pakistan, were used in the study as shown in Figure 4. Statistical characteristics of the temperature data (1975-2008) are summed up in Table 1. From the table, it is seen that the range of training data set does not cover those of the testing and validation for the Astore and Gilgit stations, respectively. Due to this, the applied models may get difficulties in catching peak temperature values in the testing/validation stages. The other important issue which can be derived from the table that both stations highly skewed temperature data (skewness > 7/8).

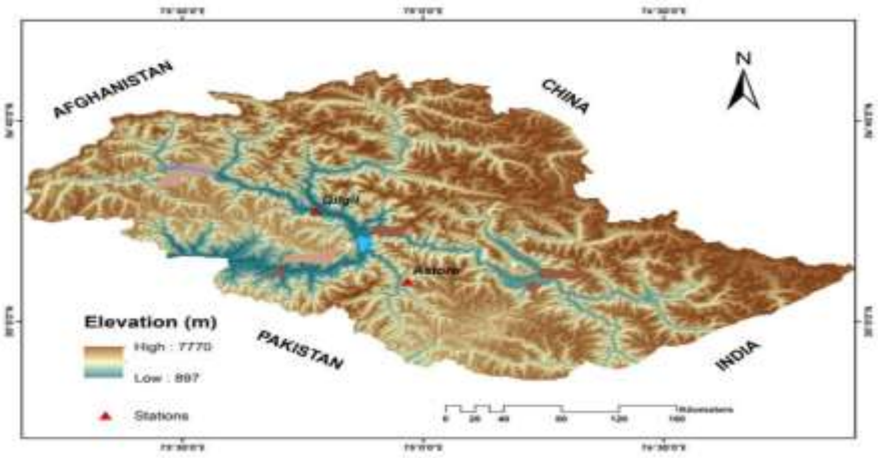

Figure 4. Study Area

In Table 1, Avg, Mx, Mn, Sk and St denote the mean, maximum, minimum, skewness coefficient and standard deviation of data, respectively. Distinct input lags were used as inputs to the models and results were evaluated using root mean square error (RMSE), mean absolute error (MAE) and determination coefficient (R^2). These indexes have been used extensively in literature [33-36]. The RMSE and MAE statistics are expressed (6) till (7):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(T_{i,o}-T_{inm})^2}{N}} \tag{5}$$

$$MAE = \frac{\sum_{i=1}^{N}|T_{i,o}-T_{i,m}|}{N} \tag{6}$$

$$R^2 = \left[\frac{\sum_{i=1}^{N}(K_{iO}-\bar{K}_O)(K_{iM}-\bar{K}_{iM})}{\sqrt{\sum_{i=1}^{N}(K_{iO}-\bar{K}_O)^2\sum_{i=1}^{N}(K_{iM}-\bar{K}_{iM})^2}}\right]^2 \tag{7}$$

where N is data quantity, $T_{i,o}$ is observed temperature, $T_{i,m}$ is modeled temperature.

Table 1. Monthly temperature statistics of both Stations

Station	Data set	Avg	Mx	Mn	Sk	St
Astore Station						
	Train	9.7	-4.4	24.3	-0.03	8.26
	Valid	9.5	-6.5	23.2	-0.08	7.96
	Test.	10.3	-5.2	23.4	-0.16	7.81
Gilgit Station						
	Train	15.7	1.9	30.7	-0.06	8.50
	Valid	15.6	1.1	29.0	-0.05	8.07
	Test.	16.1	2.4	27.9	-0.16	7.82

4. RESULTS AND DISCUSSION

In the presented work, three machine learning (ML) methods, LSSV, GMDHNN and CART, were implemented for temperature prediction. Table 2 provides the validation and test statistics of the three ML methods for the Astore station. In Table 2, lg1, lg2, lg3, lg4 and lg5 denote the 1st, 2nd, 3rd, 4th and 5th time lags of data, respectively. It is observed from the table that models' accuracies decrease by increasing input lags. It should be noted that the inputs beyond the 5th lag did not considerably improve the ability of LSSVM and GMDHNN in prediction of temperature and therefore, they were not included in the research. As evident from the results, LSSVM with 5 input lags has the lowest RMSE (1.753), MAE (1.492) and the highest R^2 (0.956) than those of the GMDHNN and CART values. The scatterplot comparison of the predictions provided by the three machine learning methods is made in Figure 5 for the Astore station. It is seen that the LSSVM has less scattered estimates with higher R^2 compared to other methods.

Table 2. The statistics of three machine learning models using different temperature input combinations-Astore Station

Model	Model	Validation period			Test period		
	inputs	RMSE	MAE	R^2	RMSE	MAE	R^2
LSSVM	Ig1	4.35	3.85	0.702	4.27	3.71	0.689
	Ig2	2.14	1.57	0.935	2.15	1.86	0.912
	Ig3	1.76	1.45	0.951	1.89	1.50	0.945
	Ig4	1.64	1.33	0.958	1.76	1.41	0.954
	lg5	1.56	1.25	0.962	1.75	1.38	0.956
GMDHNN	Ig1	4.24	3.71	0.717	4.20	3.70	0.712
	Ig2	1.95	1.57	0.940	2.34	1.67	0.925
	Ig3	1.82	1.49	0.948	1.87	1.54	0.943
	Ig4	1.85	1.50	0.946	1.86	1.48	0.945
	lg5	1.62	1.26	0.959	1.93	1.49	0.946
CART	Ig1	4.54	3.89	0.689	4.58	3.77	0.673
	Ig2	2.47	1.91	0.919	3.96	2.33	0.777
	Ig3	2.34	1.68	0.919	2.65	1.84	0.889
	Ig4	1.98	1.47	0.940	2.25	1.80	0.925
	lg5	1.98	1.50	0.940	2.36	1.84	0.919

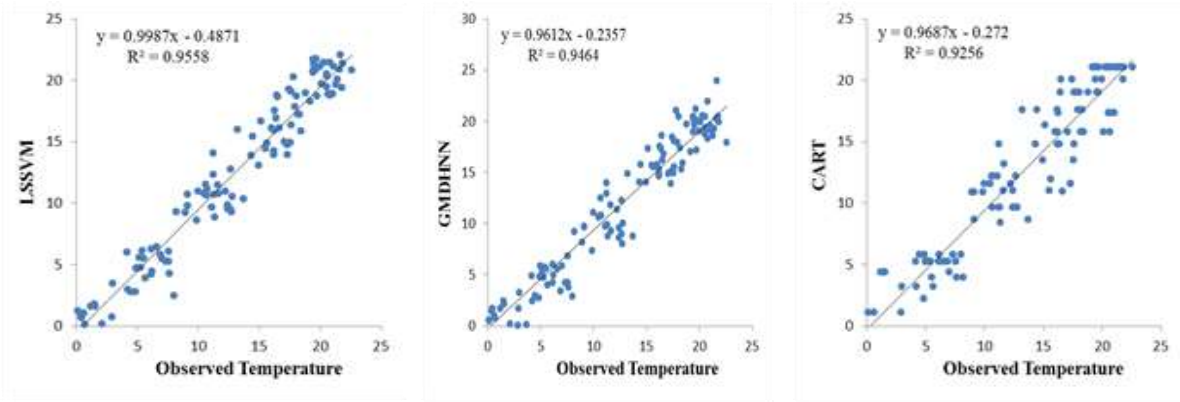

Figure 5. Scatterplots by LSSVM, GMDHNN and CART models for Astore Station

Table 3 presents the accuracy of three methods with different input lags for the Gilgit station. In this station, also LSSVM performs better than the GMDHNN and CART, however, the MAE (1.222) of GMDHNN is slightly lower than the LSSVM. CART provided the worst results in temperature prediction. Figure 6 compares the predictions provided by the three machine learning methods in the scatterplot form for the Gilgit station. It is clear that the LSSVM provided less scattered estimates than the other two methods. The reason of worse results provided by CART model is due to linear structure.

Table 3. The statistics of three machine learning models using different temperature input combinations-Gilgit Station

Model	Model inputs	Validation period			Test period		
		RMSE	MAE	R^2	RMSE	MAE	R^2
LSSVM	Ig1	4.30	3.79	0.718	4.19	3.65	0.715
	Ig2	2.10	1.73	0.933	1.95	1.66	0.930
	Ig3	1.56	1.25	0.964	1.59	1.30	0.961
	Ig4	1.41	1.10	0.971	1.55	1.21	0.967
	lg5	1.39	1.05	0.972	1.51	1.23	0.968
GMDH NN	Ig1	4.17	3.09	0.735	4.06	3.03	0.724
	Ig2	1.89	1.54	0.945	2.08	1.55	0.938
	Ig3	1.64	1.33	0.959	1.67	1.34	0.955
	Ig4	1.58	1.24	0.969	1.60	1.24	0.961
	lg5	1.42	1.08	0.962	1.54	1.22	0.963
CART	Ig1	4.93	3.90	0.705	4.92	3.82	0.630
	Ig2	2.48	1.52	0.911	2.54	1.66	0.903
	Ig3	1.59	1.25	0.963	1.81	1.40	0.953
	Ig4	1.76	1.25	0.954	1.89	1.51	0.950
	lg5	1.54	1.16	0.966	2.33	1.66	0.930

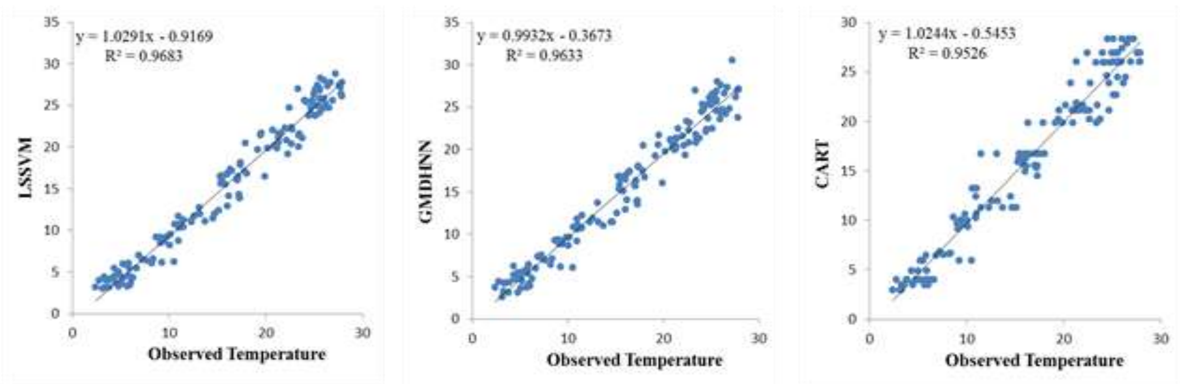

Figure 6. Scatterplots by LSSVM, GMDHNN and CART models for Astore Station

5. CONCLUSION

The prediction accuracy of three machine learning models, LSSVM, GMDHNN and CART for predicting air temperature of astore and gilgit stations of Pakistan was evaluated in this paper. The LSSVM, GMDHNN and CART models were examined by using five different time lags input combinations with help of RMSE, MAE and R^2 performance evaluation indexes. It is found that LSSVM model provided more accurate results than GMDHNN and CART models whereas CART provided worst results for both stations. Both models (GMDHNN and CART) reduced the mean RMSE by 1.47-3.12% and 20.01-25.12% for the Chakdara and Kalam Stations, respectively using the LSSVM model.

ACKNOWLEDGEMENTS

This work was supported by the National Key R&D Program of China (2016YFC0402706), and National Natural Science Foundation of China (41730750). The authors thank the staff of WAPDA for providing river flows data of both basins.

REFERENCES

[1] Kisi, O.; Shiri, J., "Prediction of long-term monthly air temperature using geographical inputs," *International Journal of Climatology*, vol. 34, pp. 179-186, 2014. doi:10.1002/joc.3676.
[2] Sanikhani, H., *et al*,. "Survey of different data-intelligent modeling strategies for forecasting air temperature using geographic information as model predictors," *Computers and Electronics in Agriculture*, vol. 152, pp. 242-260, 2018, doi:10.1016/j.compag.2018.07.008
[3] Ustaoglu, B., Cigizoglu, H.K., Karaca, M., "Forecast of daily mean, maximum and minimum temperature time series by three artificial neural network methods," *Meteorological Applications*, vol. 15, pp. 431-445, 2008, doi:10.1002/met.83
[4] Kuriqi, A., "Assessment and quantification of meteorological data for implementation of weather radar in mountainous regions MAUSAM,", vol. 67, pp. 789-802, 2015. doi:551.501.8: 551.585.7.
[5] Mooney, H., *et al*,. "Biodiversity, climate change, and ecosystem services," *Current Opinion in Environmental Sustainability*, vol. 1, pp. 46-54, 2009. doi:10.1016/j.cosust.2009.07.006.
[6] Adnan, R. M., *et al*,. "Reference Evapotranspiration Modeling Using New Heuristic Methods," *Entropy*, vol. 22, no. 5, p. 547, 2020. doi: 10.3390/e22050547
[7] Yu, H.; Chen, Y.; Hassan, S.G.; Li, D., "Prediction of the temperature in a Chinese solar greenhouse based on LSSVM optimized by improved PSO," *Computers and Electronics in Agriculture*, vol. 122, pp. 94-102, 2016. doi:10.1016/j.compag.2016.01.019.
[8] Zhang, Y., Li, H., Wang, Z., Zhang, W., Li, J., "A preliminary study on time series forecast of fair-weather atmospheric electric field with WT-LSSVM method," *Journal of Electrostatics*, vol. 75, pp. 85-89, 2015. doi:10.1016/j.elstat.2015.03.005
[9] Baghban, A., *et al*,. "Estimation of air dew point temperature using computational intelligence schemes," *Applied Thermal Engineering*, vol. 93, pp. 1043-1052, 2016.
[10] Bhagwat, P.P.; Maity, R., "Hydroclimatic streamflow prediction using Least Square-Support Vector Regression," *ISH Journal of Hydraulic Engineering*, vol. 19, pp. 320-328, 2013. doi:10.1080/09715010.2013.819705
[11] Adnan, R. M., Liang, Z., Parmar, K. S., Soni, K., & Kisi, O., "Modeling monthly streamflow in mountainous basin by MARS, GMDH-NN and DENFIS using hydroclimatic data," *Neural Computing and Applications*, pp. 1-19, 2020.
[12] Panapakidis, Michailides, Angelides," A Data-Driven Short-Term Forecasting Model for Offshore Wind Speed Prediction Based on Computational Intelligence," *Electronics*, vol. 8, 2019. doi:10.3390/electronics8040420
[13] Adnan, R. M., Khosravinia, P., Karimi, B., & Kisi, O., "Prediction of hydraulics performance in drain envelopes using Kmeans based multivariate adaptive regression spline," *Applied Soft Computing*, p. 107008, 2020.
[14] Keskin, H., Grunwald, S., Harris, W.G., "Digital mapping of soil carbon fractions with machine learning," *Geoderma*, vol. 339, pp. 40-58, 2019. doi:10.1016/j.geoderma.2018.12.037
[15] Cao, B., Domke, G.M., Russell, M.B., Walters, B.F., "Spatial modeling of litter and soil carbon stocks on forest land in the conterminous United States," *Sci Total Environ*, vol. 654, pp. 94-106, 2019. doi:10.1016/j.scitotenv.2018.10.359.
[16] Darabi, H., *et al*,. "Urban flood risk mapping using the GARP and QUEST models: A comparative study of machine learning techniques," *Journal of Hydrology*, vol. 569, pp. 142-154, 2019. doi:10.1016/j.jhydrol.2018.12.002
[17] Ebrahimy, H., Azadbakht, M., "Downscaling MODIS land surface temperature over a heterogeneous area: An investigation of machine learning techniques, feature selection, and impacts of mixed pixels," *Computers & Geosciences*, vol. 124, pp. 93-102, 2019. doi:10.1016/j.cageo.2019.01.004
[18] J. A. K. Suykens, J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, 1999.
[19] Adnan, R. M., *et al*,. "Comparison of LSSVR, M5RT, NF-GP, and NF-SC Models for Predictions of Hourly Wind Speed and Wind Power Based on Cross-Validation," *Energies*, vol. 12, no. 2, p. 329, 2019.

[20] L. T. Qin, S. S. Liu, H. L. Liu, and Y. H. Zhang, "Support vector regression and least squares support vector regression for hormetic dose-response curves fitting," *Chemosphere*, 2010.

[21] A. Seyedzadeh, *et al*,. "Artificial intelligence approach to estimate discharge of drip tape irrigation based on temperature and pressure," *Agricultural Water Management*, 2019.

[22] S. Maroufpoor, E. Maroufpoor, O. Bozorg-Haddad, J. Shiri, and Z. Mundher Yaseen, "Soil moisture simulation using hybrid artificial intelligent model: Hybridization of adaptive neuro fuzzy inference system with grey wolf optimizer algorithm," *Journal of Hydrology*, 2019.

[23] K. R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf, "An introduction to kernel-based learning algorithms," *IEEE Transactions on Neural Networks*. 2001.

[24] H. Azimi, H. Bonakdari, and I. Ebtehaj, "Design of radial basis function-based support vector regression in predicting the discharge coefficient of a side weir in a trapezoidal channel," *Applied Water Science*, 2019.

[25] D. K. Wornyo and X. J. Shen, "Coupled least squares support vector ensemble machines," *Information (Switzerland)*, 2019.

[26] A. G. Ivakhnenko, "Polynomial Theory of Complex Systems," *IEEE Transactions on Systems, Man and Cybernetics*, 1971.

[27] H. Azimi, H. Bonakdari, I. Ebtehaj, B. Gharabaghi, F. Khoshbin, "Evolutionary design of generalized group method of data handling-type neural network for estimating the hydraulic jump roller length," *Acta Mechanica*, 2018.

[28] N. Nariman-Zadeh, A. Darvizeh, M. E. Felezi, and H. Gharababaei, "Polynomial modelling of explosive compaction process of metallic powders using GMDH-type neural networks and singular value decomposition," *Modelling and Simulation in Materials Science and Engineering*, 2002.

[29] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, "Classification and Regression Trees," New York: Chapman & Hall/CRC Press, 1984.

[30] M. Ture, I. Kurt, A. Turhan Kurum, and K. Ozdamar, "Comparing classification techniques for predicting essential hypertension," *Expert Systems with Applications*, 2005.

[31] C. C. Tsai, M. C. Lu, and C. C. Wei, "Decision tree-based classifier combined with neural-based predictor for water-stage forecasts in a river basin during typhoons: A case study in Taiwan," *Environmental Engineering Science*, 2012.

[32] M. T. Dastorani, J. Mahjoobi, A. Talebi, and F. Fakhar, "Application of machine learning approaches in rainfall-runoff modeling (case study: Zayandeh_rood basin in Iran)," *Civil Engineering Infrastructures Journal*, vol. 51, no. 2, pp. 293-310, 2018.

[33] Alizamir, M., Kisi, O., Muhammad Adnan, R., & Kuriqi, A., "Modelling reference evapotranspiration by combining neuro-fuzzy and evolutionary strategies," *Acta Geophysica*, vol. 68, pp. 1113-1126, 2020.

[34] Kisi, O., Shiri, J., Karimi, S., & Adnan, R. M., "Three different adaptive neuro fuzzy computing techniques for forecasting long-period daily streamflows," In *Big data in engineering applications*, pp. 303-321. Springer, Singapore, 2018.

[35] Adnan, R.M., Yuan, X., Kisi, O., Adnan, M, Mehmood, A., "Streamflow forecasting of poorly gauged mountainous watershed by least square support vector machine, fuzzy genetic algorithm and M5 model tree using climatic data from nearby station," *Water Resources Management*, vol. 32, no. 14, pp. 4469-4486, 2018.

[36] Yuan, X., Chen, C., Lei, X., Yuan, Y., & Adnan, R. M., "Monthly runoff forecasting based on LSTM–ALO model," *Stochastic environmental research and risk assessment*, vol. 32, no. 8, pp. 2199-2212, 2018.