COMPUTATIONAL RESEARCH PROGRESS IN APPLIED SCIENCE & ENGINEERING (CRPASE)

CRPASE: TRANSACTIONS OF CIVIL AND ENVIRONMENTAL ENGINEERING

Journal homepage: http://www.crpase.com

CRPASE: Transactions of Civil and Environmental Engineering 8 (1) Article ID: 2747, 1–12, March 2022 ISSN 2423-4591

Research Article

Evaluation of Machine Learning Methods Application in Temperature Prediction

Babak Azari [1*], Khairul Hassan[1], Joel Pierce[1], Saman Ebrahimi[2]

[1] *Department of Civil Engineering, University of Memphis, Tennessee, United States, 3720 Alumni Ave, Memphis, TN 38152, USA*

[2] *School of Civil Engineering, College of Engineering, University of Tehran, Tehran, Iran*

Keywords

Machin learning,
Temperature prediction,
Ensemble Models,
Single Models.

Abstract

Machine Learning (ML) techniques for time series prediction are becoming increasingly accurate and helpful, particularly in considering climate change. As more methods are developed, it follows that differentiating between them is becoming increasingly more important as well. This work took a local temperature time series as a dependent variable and a collection of relevant climatology time series as independent variables and applied leading Machine Learning methods to them. The six methods tested included four simple models: Linear Regression (L.R.), k-Nearest Neighbor (kNN), Support Vector Machine (SVM), and Artificial Neural Network (ANN), In addition of two ensemble model methods: Random Forest (R.F.) and Adaptive Boosting (AdB). Results compared all the method's training and predictive performances to evaluate the method's overall effectiveness in forecasting the average daily temperature value. Actual data was used to train each of the mentioned ML methods, and then they were used to predict the future temperature in the study area. The analysis revealed that out of the six methods tested, the Artificial Neural Network outperformed the others in both training and prediction of temperature values in the Memphis, TN climate.

1. Introduction

Tracking and predicting temperature fluctuations plays a vital role in studying future climate patterns. These days, climate change and global warming are hot issues worldwide since their negative impact alters human lives [1]. Climate change is projected to increase the risk of water-related disasters such as urban floods and severe droughts [2–4] or has an intense effect on rivers' water quality in many regions around the globe [5]. Having a strong understating of the future, particularly around temperature change, is paramount in helping decision-makers evaluate and reduce the effects of climate change and increase the reliability and sustainability of infrastructure [6–8].

In assessing the quality of human life, the temperature is known to be essential. Therefore, a reliable and accurate framework to accurately predict the air temperature in the future is of high importance [9]. Such predictions with 100% accuracy may be

*Corresponding Author: Babak Azari
E-mail address: bazari@memphis.edu, ORCID: https://orcid.org/0000-0003-2476-5174

Received: 10 February 2022; Revised: 8 March 2022; Accepted: 12 March 2022
https://doi.org/ 10.52547/crpase.8.1.2747

Academic Editor: **Vahid Najafi Moghadam Gilani**

Please cite this article as: B. Azari, K. Hassan, J. Pierce, S. Ebrahimi, Evaluation of Machine Learning Methods Application in Temperature Prediction, Computational Research Progress in Applied Science & Engineering, CRPASE: Transactions of Civil and Environmental Engineering 8 (2022) 1–12, Article ID: 2747.

impossible, but the forecasting errors can be minimized, or the forecasting speed can be intensified [10, 11].

The current demand for this kind of information has promoted the development of new means of prediction and data processing and resulted in new specialized tools and methods being presented [12–14]. During the last twenty years, these new tools, called Machine Learning (ML) methods, have shown their capabilities and accuracy in different science and engineering fields, such as forecasting, predicting, and pattern recognition problems [15–18]. Many studies used ML methods and algorithms to predict natural phenomena such as temperature, dewpoint, precipitation, and soil temperature.

One of the long-standing and well-known ML methods used in many environmental areas is Linear Regression, applicable to single or multiple variable problems [19–21]. Another nonparametric ML method, k-Nearest Neighbor (kNN), has been used chronically in ML literature to solve regression problems [22–24]. Support Vector Machine (SVM) regression is another method applied to find relations between inputs and outputs [24, 25]. Finally, the Artificial Neural Network (ANN) is one of the most well-known and popular methods, capable of capturing nonlinear patterns in functional relationship features and targets. It has been applied to many ML problems in different areas [26–28].

Researchers used the methods mentioned above to predict natural phenomena during recent years. [29] used Coactive Neuro-Fuzzy Inference System (CANFIS) to forecast the soil temperature in an arid and semi-arid area. Jain et al. Jain et al. [30] developed an ANN model to predict wintertime air temperatures. They used the data from six hours prior to the prediction time, including relative humidity, wind speed, and solar radiation. Smith, McClendon, and Hoogenboom Smith, McClendon, and Hoogenboom [31] improved the former studies' predictions by extending the data lag to twenty-four hours, and they concluded that this duration improved the prediction accuracy. [32] used two ML methods, Support Vector Machine (SVM) regressions and Multi-layer Perceptron (M.L.P.), to predict the average mean monthly temperature in eight observational stations in Australia and two in New Zealand, with implications of detecting possible climate change signals in these regions. They concluded that SVM produced better predictions. [33] proposed a novel approach for air temperature prediction by combining the ANFIS and optimization algorithms. The proposed method showed an increase in the accuracy of the predictions. In another study, [34] used kNN and a pattern approximate matching machine-learning algorithm to predict the short-term air temperature.

Literature reviews showed that despite numerous studies focused on using specific ML methods on predicting air temperature, few studies are comparing and evaluating the application of multiple ML methods in the same study [35]. Since, temperature has a direct effect on precipitation, which will create drought event, so different ML approaches would help in forecasting drought event [36]. This work could be used as reference for other hydrological and hydro-metrological parameters like rainfall, streamflow measurement etc. for this study region. Therefore, this study aims at analyzing the trends and patterns of the daily temperature data of the Memphis International Airport Weather Station with multiple ML methods. First of all, the historical data obtained and analyzed. Then six different ML methods were trained using the actual data. So, the predictive capability of the ML methods was then evaluated and compared. The models ' inputs were chosen based on previous studies and available data, the relative humidity, wind speed, dewpoint degree, 1-hour precipitation, and barometric sea pressure time series. Various performance indicators for the six ML predictions were used to identify the best ML method in the study.

2. Methodology

2.1. Study Area

The Memphis International Airport is in the southern portion of Memphis, TN, as seen in Figure 1. Established in 1927, the airport now serves commercial and passenger flights in addition to being the "Super Hub" for FedEx. As the busiest cargo airport in the U.S. and with operations around the clock, the airport requires continuous data from the weather station, making it suitable for this project [37].

Memphis itself sits east of the Mississippi River in the Mississippi Embayment, and the surrounding region has low elevations and small geological reliefs [38]. The climate is humid, with average yearly precipitation near 143cm [39]. Temperatures are similar to most other eastern regions of the United States near the same latitude with a 30 year mean temperature norm of 16 - 18° C [40]. With a gentle climate and few natural borders to hinder its growth, Memphis and its suburbs extend urbanization for kilometers eastward away from the Mississippi River. The sprawling city supports a population of 650 thousand people according to the 2020 census, putting it within the top 30 largest cities by population in the U.S. [41].

From the Iowa Environmental Mesonet climate data repository, discrete data for the airport weather station (see 1.1.1.1Table 1) was mined for climate data from January 1st, 1980 to September 22nd, 2021 [42]. The available features were pared down, eliminating features that depicted the same base variable, that were recorded as categorical, or contained significant data gaps in the station's history.

The final features included hourly temperatures in degrees Celsius, wind speed in miles per hour, relative humidity, dew point in degrees Celsius, 1-hour precipitation in mm, and sea pressure in millibar (1.1.1.1Table 2). Scrapper codes are implemented to take the raw data in multiple different time steps and average them into a daily time step. Gaps in the data at the hourly or 15-minute level were eliminated, while single-day gaps were bridged with an average between the day before and the day after the gap. The daily temperature was designated as the dependent variable, and the other variables were defined as independent variables.

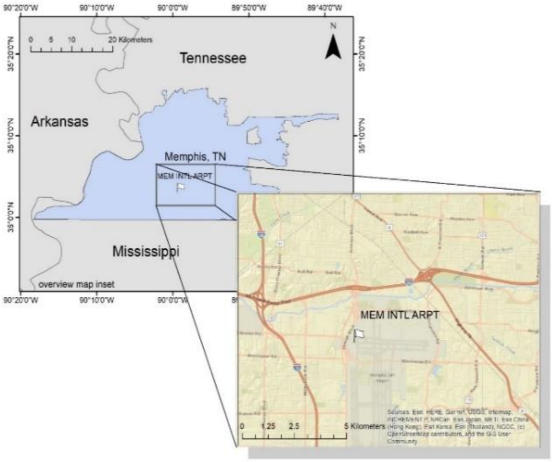

Figure 1. Inset map outlining the location of the Memphis International Airport in Memphis, TN.

Table 1. Memphis International Airport Weather Station Metadata.

ID	Latitude (Degree)	Longitude (Degree)	Elevation from sea level
MEM INTL ARPT	-89.985	35.0611	87

Table 2. Features and parameters used in this study

Data	Role	Temporal scale and Unit
Average Temperature	Target	Daily, °C
Dewpoint	Feature	Daily, °C
Relative humidity	Feature	Daily, %
Wind Speed	Feature	Daily, MPH
Barometric sea pressure	Feature	Daily, milibar
1-hour Precipitation	Feature	Daily, inch

2.2. Machine Learning Methods

2.2.1. Single Model ML Methods

Single models contain only one method in their procedure, unlike ensemble models, which contain multiple methods combined. In this study, four different single models were used and compared. This section discusses the single models used in this study.

Linear regression (L.R.)

Linear regression (L.R.) utilizes a linear curve based on the best fit criterion to estimate the trend of a dataset [43]. L.R. can be used with single or multiple variable datasets [35]. With a single variable, the method minimizes a single objective function based on that variable. A multiple variable regression also uses a single objective function, but the method combines the individual objective functions of each variable using a weighting scheme first [44]. The elasticity in the weighting schemed is determined using two parameters, L1 and L2. Once set up, the method minimizes the error between the line and the best fit criterion, producing the best fit line for prediction.

k-Nearest Neighbor (kNN)

The k-Nearest Neighbor (kNN) regression operates on the concept of spatial dependency [45]. The method averages the value of the nearest k neighbors around an event to predict its value, rewarding closer events and penalizing far events. The optimal number of neighbors (k) is determined in training the method, as it governs the predictive values and computational efficiency. Further optimization through various weighting schemes, transformation schemes, and rejecting schemes is common in literature, allowing this simple method to make more complex predictions and classifications [46].

Support Vector Machine (SVM)

The Support Vector Machine (SVM) regression method depends on feature classification and develops a hyperplane between classes. The vector lengths and variance between the features and the plane are minimized. Most common types of kernels can be used with SVM, including Euclidean, Gaussian, Exponential, and Dirichlet kernels [47]. The objective function for SVM regression includes a coefficient derived from the cost analysis that helps determine the flatness of the hyperplane formed [48]. This provides user input in modifying the SVM method to fit unique datasets.

Artificial Neural Network (ANN)

The ANN as a widely used machine learning technique, captures the nonlinear pattern between the input and target series by using the error backpropagation learning algorithm, also called the least-mean-square (L.M.S.) algorithm [35, 49, 50]. A typical feed-forward network consists of a hidden layer with nodes, input, and output layers [35]. The number of hidden layers and neurons in each hidden layer should be specified using the trial-and-error approach. The widely used feed-forward network multilayer perceptron (M.L.P.) was used for this study due to its popularity [51]. For training the ANN model, the Levenberg–Marquardt (L.M.) method is suitable for finding an optimum solution [49]. The Levenberg–Marquardt algorithm is often characterized as more stable, efficient, faster, and less likely trapped in local minima than other optimization algorithms [52].

2.2.2. Ensemble-Model ML methods

Nowadays, the application of ensemble ML models has increased dramatically in fields such as hydrology and hydrometeorology due to the high efficiency in prediction [53, 54]. Ensemble models combine several ML training algorithms to obtain higher training and testing accuracy [35, 54, 55]. These compound ensemble models are grouped into bagging, boosting, and stacking categories based on the types of models used [35]. Bagging and boosting methods use multiple similar models or homogeneous models, whereas stacking uses heterogeneous models [55]. This study will incorporate one bagging ensemble model (Random Forest) and one boosting ensemble model (Adaptive Boosting) to compare their predictive capability in reference to the temperature time series.

Random Forest

Random Forests (R.F.s) are bagging ensemble models that use several decision trees as base-learners to produce more accurate results [56]. Individual trees are created by the bootstrap sampling process from training data and use the set of random parameters as their roots and nodes [53, 54]. Multiple decision trees provide stability over a single tree by reducing over-fitting and averaging the results [53]. The three main parameters for the R.F.s are the number of trees in the forest at each binary node, the number of randomly selected predictors, and the minimal number of observations at the nodes of the trees [57, 58].

Adaptive Boosting (AdB)

Adaptive Boosting is a boosting ML algorithm, where strong learning algorithms effectively boost the weak learning algorithms. AdB needs to define the number of basic learners (n) as a parameter [35]. AdB creates basic learners with low accuracy during the training process and improves based on the previous ones [53]. By this process, the AdB dynamically updates the training weight by following the performance of the base learning algorithms [59].

2.3. Performance Measure

To compare the predictive capability of different ML algorithms, the following evaluation metrics are used, which have been used in many studies to evaluate the results in various fields: (1) Pearson's correlation coefficient (r), (2) coefficient of determination (R^2), (3) The index of agreement (d), (4) Nash-Sutcliffe efficiency (NSE), (5) Mean absolute error (M.A.E.), (6) Root mean square error (RMSE), (7) Percent bias (PBAIS), (8) and the RMSE-observations standard deviation ratio (R.S.R.) [32, 35, 60, 61].

Pearson's correlation coefficient (r) and coefficient of determination (R^2) both measure the degree of collinearity between observed and simulated values [35]. The correlation coefficient (r), which ranges from -1 to 1, measures the linear relationship between the parameters. Here, r = 0 means no linear relationship exists, whereas r = +1 or -1 mean positive and negative correlation respectively [62]. R^2 interprets the proportion of the variance in the observed values that is explained by the model. Typically, R^2 varies from 0 to 1 where the higher values suggest less variance in the observed dataset [61].

The agreement index (d) is the standardized agreement measurement between the predictions and observations given by equation 1 and varies between 0 to 1 [63]. Here, d = 1 means a perfect agreement between the observed and predicted values, and 0 means no agreement [55, 61].

$$d = 1 - \frac{\sum_{i=1}^{n}\left(Y_i^{obs} - Y_i^{sim}\right)^2}{\sum_{i=1}^{n}\left(\left|Y_i^{sim} - Y^{mean}\right| + \left|Y_i^{obs} - Y^{mean}\right|\right)^2} \qquad (1)$$

Nash-Sutcliffe efficiency (NSE) is a normalized statistic determining the relative magnitude between the noise and information [64]. NSE variers between 0 to 1 for the acceptable performance and can be described by Eq.(20) [55, 62].

$$NSE = 1 - \left[\frac{\sum_{i=1}^{n}\left(Y_i^{obs} - Y_i^{sim}\right)^2}{\sum_{i=1}^{n}\left(Y_i^{obs} - Y^{mean}\right)^2}\right] \qquad (2)$$

Mean absolute error (M.A.E.), and root mean square error (RMSE) can be described using the Eqs. (3-4) below. Both M.A.E. and RMSE indicate perfect fit when close to 0. However, RMSE is a better tool than M.A.E. since it increases significantly when there are significant differences between the observed and simulated values [61, 65].

$$MAE = \frac{\sum_{i=1}^{n}\left|Y_i^{obs} - Y_i^{sim}\right|}{n} \qquad (3)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i^{obs} - Y_i^{sim}\right)^2}{n}} \qquad (4)$$

Percent bias (PBAIS) calculates the average tendency of the simulated data to be larger or smaller than their observed counterparts as in Eq. (5) [66]. Positive and negative values indicate underestimation and overestimation bias, respectively [61, 62].

$$PBAIS = \left[\frac{\sum_{i=1}^{n}\left(Y_i^{obs} - Y_i^{sim}\right) * (100)}{\sum_{i=1}^{n}\left(Y_i^{obs}\right)}\right] \qquad (5)$$

RMSE-observations standard deviation ratio (R.S.R.) is the ratio of the RMSE and the standard deviation of the measured data as shown in Eq. (6). Better model performance means lower R.S.R. or, in other words, lower RMSE [62].

$$RSR = \frac{RMSE}{STDEV_{obs}} = \frac{\left[\sqrt{\sum_{i=1}^{n}\left(Y_i^{obs} - Y_i^{sim}\right)^2}\right]}{\left[\sqrt{\sum_{i=1}^{n}\left(Y_i^{obs} - Y^{mean}\right)^2}\right]} \qquad (6)$$

For equations (1 through 6), n is the total number of observations, Y_i^{obs} is the *i*th observation of the temperature time series data, Y_i^{sim} is the ith simulated value of the temperature time series, Y^{mean} is the mean of the temperature time series. Table 3**Error! Reference source not found.** summarizes the criteria for the before mentioned statistical methods used in finding a better predictive model.

2.4. Machine Learning Tool for Implementation

In this study, data sorting and filtering were done using Excel VBA. First, statistical trends were conducted using a Microsoft Excel add-on called XLSTAT. Then ML algorithms were carried out using an open-source data mining software called Orange, developed by the University of Ljubljana [67, 68]. This software is based on the python programming language and is well used in different fields of hydrology by various researchers [35, 69, 70].

3. Results and Discussion

3.1. Data Correlation

Before applying learning algorithms to a data set, specific characterizations must be checked in the data. First, it is paramount to check the collinearity between the features themselves and between the features and the target.

Table 3. General performance ratings of the statistical indices (Adopted from [62])

Performance Rating	RANGE					
	RSR	NSE	PBIAS (%)	RMSE	MAE	INDEX OF AGREEMENT, d
Very Good	0 ~ 0.5	0.75 ~ 1	≤ ±10	0.1 ~ 0.25	0	1
Good	0.5 ~ 0.6	0.65 ~ 0.75	±10 ~ ±15	0.25 ~ 0.5	Not Defined	Not Defined
Satisfactory	0.6 ~ 0.7	0.5 ~ 0.65	±15 ~ ±25	0.5 ~ 0.75	Not Defined	Not Defined
Unsatisfactory	> 0.7	< 0.5	≥ ±25	0.75 ~ 1.0	1	0

Collinearity among the features and target data were evaluated and shown in Table 4. This table shows that the collinearity among the features was relatively low as required for learning methods. The sea pressure and dew point were higher than the rest but remained within an acceptable range.

3.2. Mann-Kendall Test

Second, the Mann–Kendall (M–K) tests for the time series data were developed and are shown inTable 5. The p-values of the Mann–Kendall (M–K) tests indicated trends in temperature and in each of the input features through time since the threshold value (p-value) was less than 0.05. The positive value of Kendall's τ indicated a positive upward trend; therefore, the temperature, dewpoint, and 1 hour precipitation time series, which were already known to have trends, were proven to have positive upward trends. On the other hand, relative humidity, wind speed, and sea pressure had downward trends. Sen's Slope, which refers to the slope of the trend, showed that the temperature time series had a positive trend of 0.005 0C per day, which is a clear indication of climate change in this region. The 1-hour precipitation trend, which had also been identified as a positive trend, was minimal as the Sen's Slope was only 0.00003 mm per day.

3.3. Training and Prediction Results

ML methods were used to predict the temperature pattern in the study area using the provided feature data sets. The combined data set contained 15241 records of daily temperature, dew point, relative humidity, wind speed, sea pressure, and depth of 1-hour precipitation from 1/1/1980 to 9/22/2021. The training set included 12785 records for the date range from 1/1/1980 to 12/31/2014. Training data were then randomly split into two parts, 80% for training, 20% for testing. Data from 1/1/2015 to 9/22/2021 were preserved for the validation/prediction step used in deciding on the best method **Error! Reference source not found.**). It was observed by [71] that considering the prior twenty-four hours of atmospheric data to predict the temperature at a certain time resulted in the most accurate forecasting. Therefore, in this study, we used the previous day's data, including dew point, relative humidity, sea pressure, and 1-hour precipitation, to predict the temperature for the day of interest.

Table 4. Features collinearity and correlation matrix

Features	Dewpoint	Relative Humidity	Wind Speed	Sea Pressure	1-hour Precipitation
Dewpoint	1.0000	-	-	-	-
Relative Humidity	0.3727	1.0000	-	-	-
Wind Speed	-0.1215	0.0827	1.0000	-	-
Sea Pressure	-0.5778	-0.3265	-0.2113	1.0000	-
1 hour Precipitation	0.0938	0.2761	0.0989	-0.1464	1.0000

Table 5. Mann–Kendall (M–K) Statistics and their corresponding p-value at 5% significance level for the Memphis International Airport Weather Station.

M-K test	Sen's Slope	Kendall's τ	p-value (two-tailed test)	alpha, α	Test interpretation
Temperature	0.00586	0.019	0.000	0.05	Trend in Series
Dewpoint	0.00356	0.012	0.027	0.05	Trend in Series
Relative Humidity	-0.00746	-0.016	0.004	0.05	Trend in Series
Wind Speed	-0.00694	-0.066	<0.0001	0.05	Trend in Series
Sea Pressure	-0.00434	-0.024	<0.0001	0.05	Trend in Series
1 hour Precipitation	0.00003	0.059	<0.0001	0.05	Trend in Series

Orange software (Version 3.29.3), a python-based and open-source application, trained the ML models. Figure 3 shows the schematic of the ML model training and prediction steps. In the first stage, the data sampler split the input data into training and test data sets. The training data set was introduced to the ML models in the second stage, and the training process started. For the third stage, each of the models was tested using the test data set. The training stage was done many times using trial and error to obtain the best results. When the best results for the training stage were achieved, the validation/prediction data set was introduced to the trained models, and the results were compared to the observed data set.

.

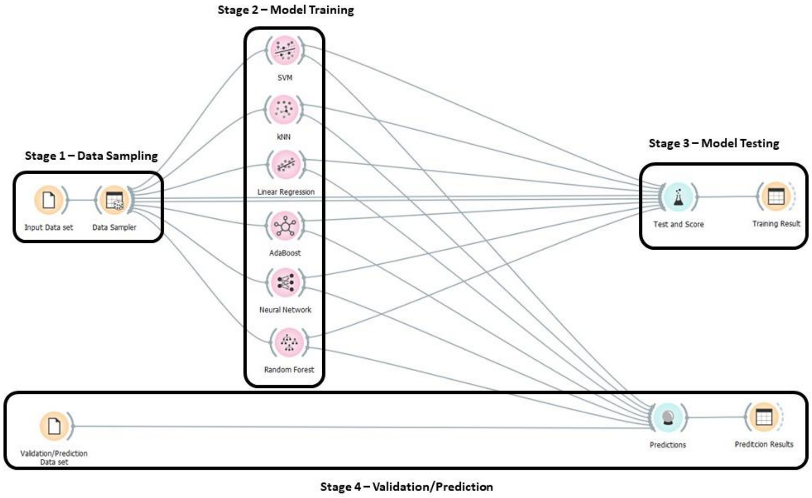

Figure 2. Schematic of the workflow and the Orange Software workspace

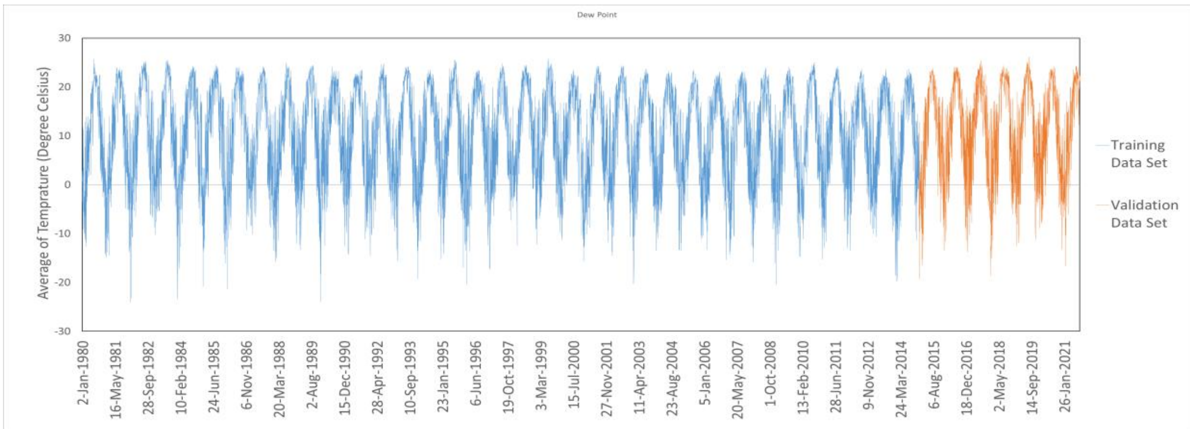

Figure 3. Temperature time series including the Training and Validation data sets

Training step accuracy was measured by seven unique indexes, including the coefficient of determination (R^2), the index of agreement (d), the Nash-Sutcliffe efficiency (NSE), the mean absolute error (M.A.E.), the root mean square error (RMSE), the percent bias (PBAIS), and the RMSE-observations standard deviation ratio (R.S.R.). In this section, the results for every method were investigated more in detail.

The first ML method was kNN, which was used to predict the temperature data. The optimal number of neighbors (k) after running the algorithm many times was six. The selected distance metric for the kNN method was Euclidian and uniform weighting was also used. Table 7.Figure 4 (a) shows the correlation between the observed value and test results of the kNN method. The plot shows a perfect fit, but the accuracy for temperatures less than zero was slightly lower than temperatures greater than zero.

The SVM method features were set and calibrated by trial and error. This produced cost and ε values of 5.2 and 0.1, respectively. The kernel function, Radial Basis Function (RBF), was chosen for this study, and the gamma value was left to be calculated automatically by the software. Optimization tolerance and the iteration limit were 0.001 and 600, respectively. Table 7.Figure 4 (b) shows the correlation between the observed values and the SVM test results.

The neural network used in this study is the multiple layer perceptron (M.L.P.) with backpropagation. Setting up the parameters of the ANN was done using multiple model runs. The number of hidden layers and the neurons were changed through these runs to obtain the best results for the test data. The best test results were observed for the model run with ten hidden layers and 20 neurons in each layer. Table 7.Figure 4 (c) depicts the correlation between the observed values and the ANN results. The plot shows a perfect fit for the training step. As can be seen from the picture, ANN had problems with temperatures over 30º C.

Table 7.Figure 4 (d) shows the correlation results for the Linear Regression, which indicates a perfect fit. Parameters for the Linear Regression were also determined by trial and error with a final value for the alpha of 0.012 and L1 and L2 values for elastic mixing of 0.5.

The first ensemble model used in this study was the Random Forest model. The number of Trees was set to 20 based on many trial and error runs. Table 7.Figure 4 (e) shows the correlation between the Random Forest results and the observed temperature. As established from Table 7.Figure 4 (e), this method was trained very well, and the results were highly accurate. However, the accuracy is slightly lower for temperatures less than -5 degrees C.

The second ensemble model tested was the Adaptive Boosting (AdB) model. The best results were obtained when the number of estimators equaled 80, and the learning rate was 1. The classification algorithm was set as SAMME (Stagewise Additive Modeling using a Multi-class Exponential loss function), and a square regression loss function was used [72]. SAMME classification algorithm is an extension on AdB that improves the ability of this method in multi-class classification and is used here to acquire the best results [72]. Table 7.Figure 4 (f) shows the AdB results versus the observation data. Results are similar to that of Random Forest as the training step was very accurate and successful despite slightly lower accuracy for temperatures less than -5 degrees C.

As Table 7.Figure 4 shows, all the methods were well trained and acceptable. However, the SVM results were significantly inferior to the other methods. The ANN method with an R^2 = 0.9040 and the SVM method with an R^2 = 0.5560 were respectively the best and the worst trained models based on the R^2 value. As Table 7.Figure 4 shows, all the methods were well trained and acceptable. However, the SVM results were significantly inferior to the other methods. The ANN method with an R^2 = 0.9040 and the SVM method with an R^2 = 0.5560 were respectively the best and the worst trained models based on the R^2 value.

Table 6. Performance ratings for the ML models for the test dataset. (Highlighted cells show the best method in each performance indicator column.)

METHOD	RSR	NSE	MAE	INDEX OF AGREEMENT, d	PBIAS (%)	RMSE	R2
KNN	0.0236	0.9764	1.0180	0.9940	-0.7324	1.3631	0.8893
SVM	0.3767	0.6233	5.0778	0.9012	28.1201	5.4490	0.5560
ANN	0.0173	0.9827	0.8852	0.9957	-0.4695	1.1694	0.9040
Linear Regression	0.0211	0.9789	0.9861	0.9947	-0.1969	1.2896	0.9010
Random Forest	0.6329	0.3671	5.3492	0.8253	-0.5337	7.0636	0.8930
AdB	0.0136	0.9864	0.7650	0.9966	-1.0957	1.0371	0.8860

However, R^2 is not enough to compare the training results. Therefore, six more performance indexes were used to compare the methods. **Error! Reference source not found.**Table 6 shows the RSR., NSE, M.A.E., Index of agreement, PBIAS, and RSME values for all the methods. According to **Error! Reference source not found.**, all the methods showed an acceptable performance except SVM. Among the methods, ANN showed the best performance based on three indicators: NSE = 0.9827, Index of Agreement = 0.9957, and an R^2 = 0.9040. AdB method was the best method based on a different set of three indicators: RSR = 0.0136, MAE = 0.7650, and RMSE = 1.0371. The linear regression method worked best based on the PBIAS indicator alone, with an absolute value of 0.1969. The SVM method resulted in the weakest results by means of all seven of the performance indicators. From the training results, it was concluded that the ANN and AdB models provided the best-trained networks.

The next step was to investigate each trained model's performance in predicting temperature for future dates. Daily data from 1/1/2015 to 9/22/2021 were used to validate and choose the best method for average daily temperature prediction. The independent variables of relative humidity, dewpoint, wind speed, sea pressure, and 1-hour precipitation for the mentioned period were introduced to the trained models, and the resulting temperature predictions were compared to the observed temperatures.

Error! Reference source not found. shows the correlation between the predicted values and the observed values for all six models. As can be seen from the pictures, the highest R2 was achieved using ANN and Linear Regression with R2 = 0.8960 and R2 = 0.8930, respectively. On the other hand, SVM resulted in the least accurate prediction with an R2 = 0.4980. It also could be seen in **Error! Reference source not found.** (b) that the SVM predictions had multiple values that did not lie close to the trendline, indicating several outputs with unacceptable accuracy. For further comparison, the other six performance indexes were calculated and reported in Table 7.

In **Error! Reference source not found.**, the ANN and Linear Regression methods outperformed the other methods, even the ensemble model methods. The R.S.R. index reported the ANN and Linear Regression predictions to be the best among the methods. The NSE index also indicated the superiority of the ANN and Linear Regression methods. The M.A.E., Index of Agreement, RMSE, and R^2 showed the same story. The one indicator that differed from this trend was the PBIAS, which indicated that the Random Forest had a slightly lower absolute value than the ANN or Linear Regression methods. Overall, the training step for ANN showed the best performance in the prediction step, while the training step for the SVM method had the weakest predictions

Although the performance of most of the methods was outstanding and accurate, ANN showed the most accurate results to predict the temperature. **Error! Reference source not found.** shows the ranking of the methods based on the prediction step's performance. Overall, ANN obtained the best prediction accuracy among all the methods. **Error! Reference source not found.** compares the ANN results with observed values.

Table 7. Performance ratings for the ML models for validation/prediction data set. (Highlighted cell shows the best method in each performance indicator.)

METHOD	RSR	NSE	MAE	INDEX OF AGREEMENT, d	PBIAS (%)	RMSE	R2
KNN	0.1141	0.8859	2.2618	0.9690	-0.2966	3.0621	0.8860
SVM	0.5020	0.4980	5.5916	0.8534	26.9938	6.4242	0.4980
ANN	0.1045	0.8955	2.1217	0.9719	-0.5290	2.9307	0.8960
Linear Reg.	0.1071	0.8929	2.2066	0.9715	0.2811	2.9679	0.8930
Random Forest	0.1077	0.8923	2.1805	0.9715	0.2345	2.9758	0.8920
Adaptive Boosting	0.1147	0.8853	2.2208	0.9701	-0.5605	3.0709	0.8850

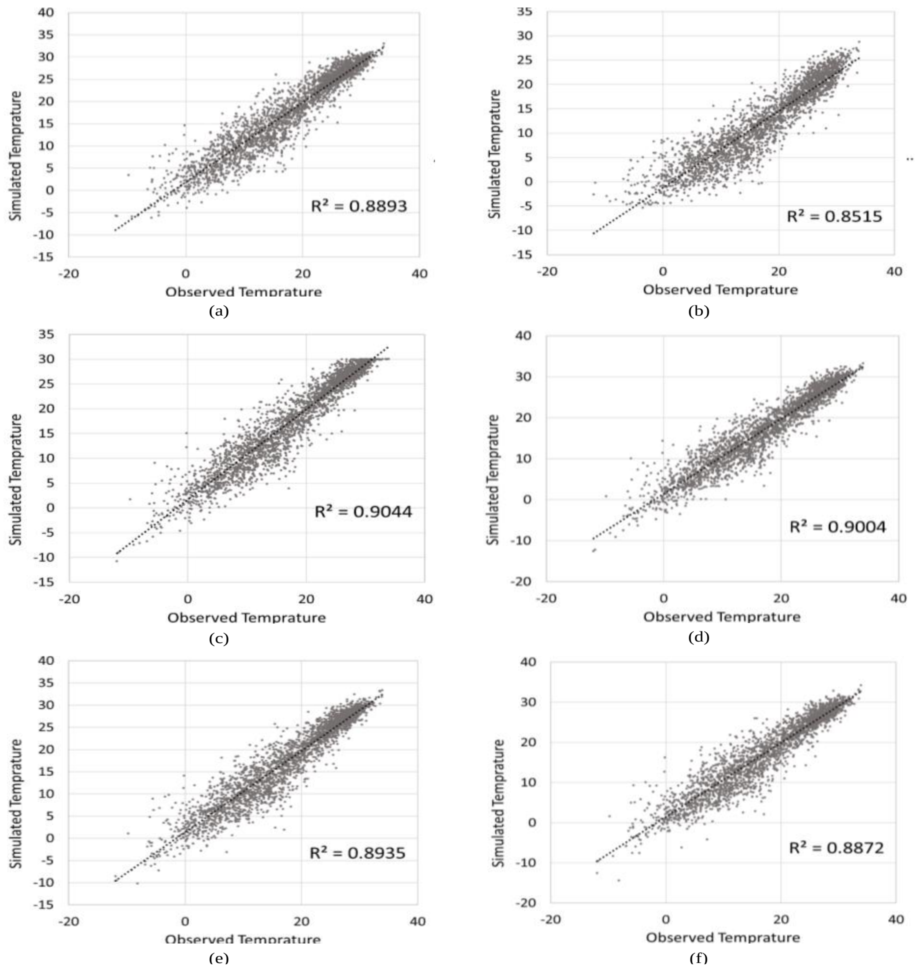

Figure 4. Correlation between observed and test outputs for (a) kNN and (b) SVM, (c) ANN, (d) Linear regression, (e) Random Forest, and (f) AdB

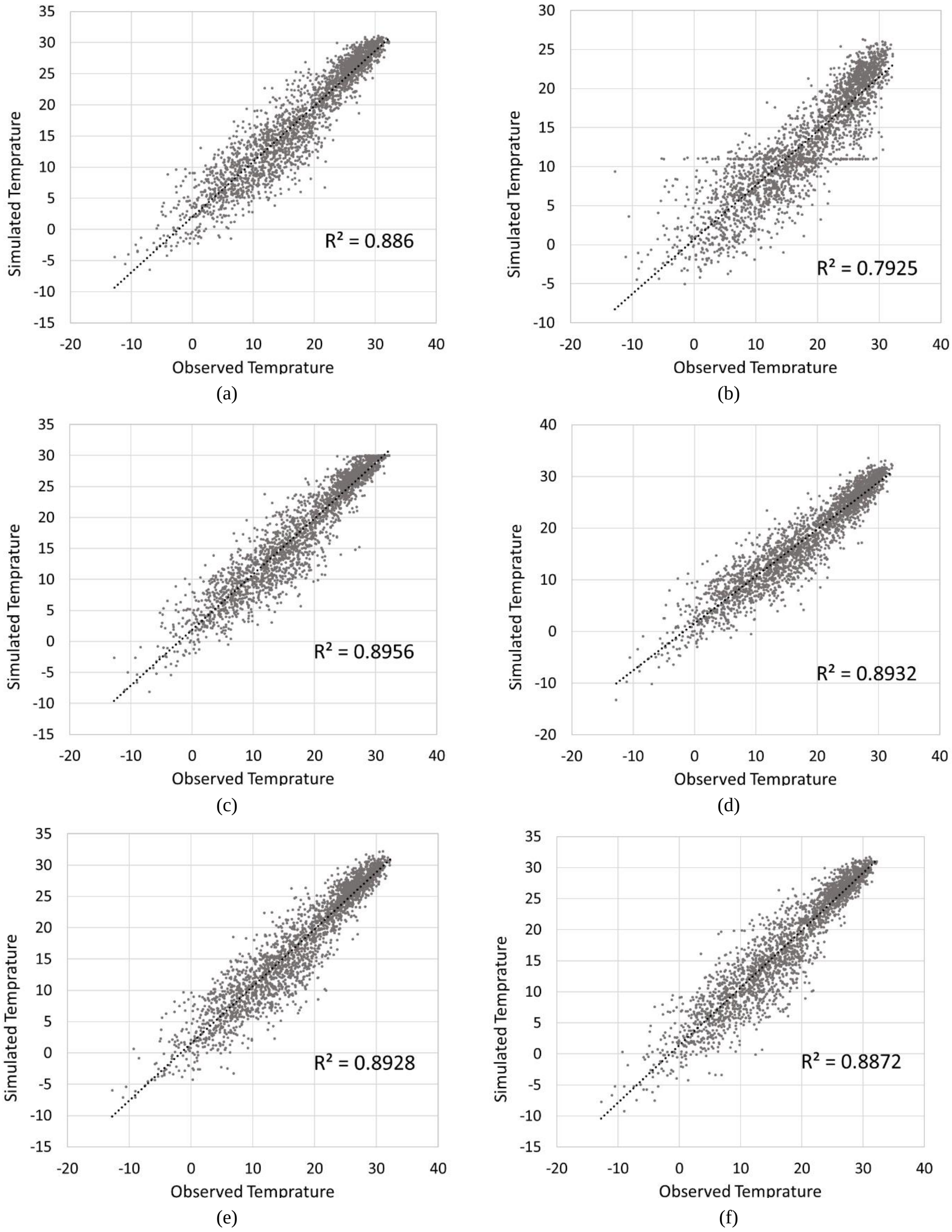

Figure 5. Correlation between observed and validation outputs for (a) kNN, (b) SVM, (c) ANN, (d) Linear regression, (e) Random Forest, and (f) AdB.

Table 8. ML methods ranking based on their performance

Rank	RSR	NSE	MAE	INDEX OF AGREEMENT, d	PBIAS (%)	RMSE	R2	Overall
1	ANN	ANN	ANN	ANN	Random Forest	ANN	ANN	ANN
2	Linear Regression	Linear Regression	Random Forest	Linear Regression	Linear Regression	Linear Regression	Linear Regression	Linear Regression
3	Random Forest	Random Forest	Linear Regression	Random Forest	KNN	Random Forest	Random Forest	Random Forest
4	KNN	KNN	AdB	AdB	ANN	KNN	KNN	KNN
5	AdB	AdB	KNN	KNN	AdB	AdB	AdB	AdB
6	SVM	SVM	SVM	SVM	SVM	SVM	SVM	SVM

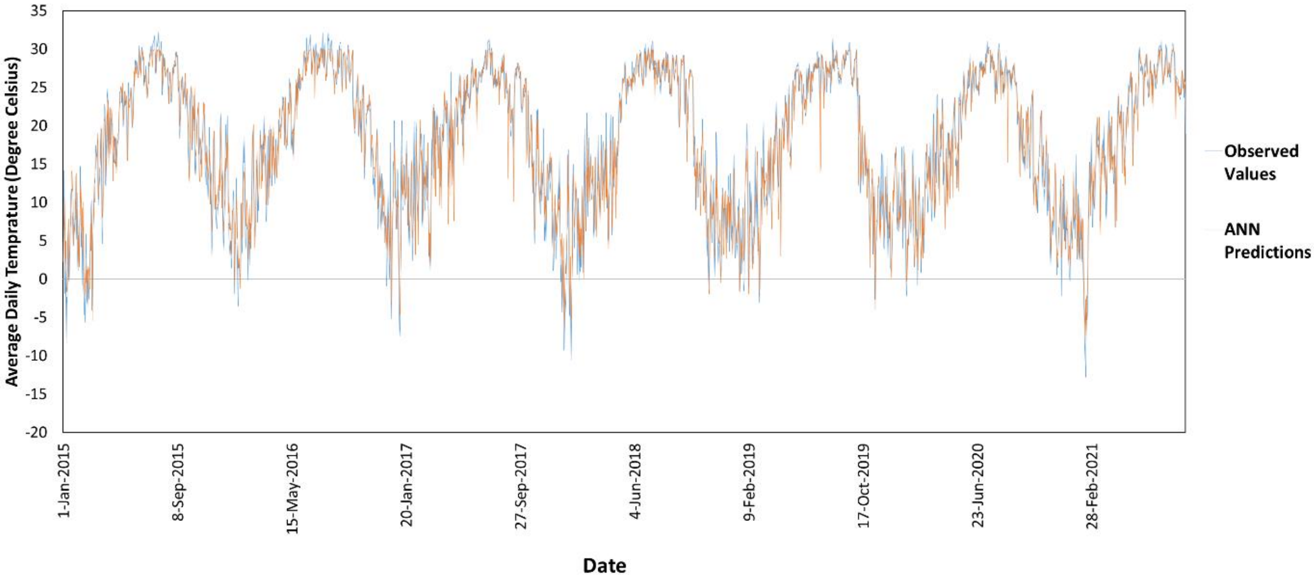

Figure 6. Comparing ANN results with the observed values

4. Conclusion

In this study, meteorological time series data were extracted for the city of Memphis from the International Airport Weather Station, including air temperature, dewpoint, relative humidity, wind speed, sea pressure, and 1-hour precipitation values. Data covered almost 21 years, from 1/1/1980 to 9/22/2021. First, collinearity and correlation between the features were evaluated, revealing no sign of correlation. Using the Mann-Kendall test, trends within the data were studied. Mann-Kendal's test showed that all the time series have trends, which validated climate change in the area. Relative humidity, wind speed, and sea pressure showed negative trends while temperature, dewpoint, and precipitation trends were positive.

The primary purpose of this study was to compare the capability of different ML methods in predicting air temperature for studying climate change.

Therefore, six ML methods, including LR. kNN, SVM, ANN, Random Forest, and AdB were used to predict the temperature. Data from 1/1/1980 to 12/31/2014 were selected as training data. Then, 20% of the training data set was randomly selected for the test, and the remaining 20% were used to train the ML methods. The test results showed that, except for SVM, the methods were well trained. SVM had the least R^2 value, which was equal to 0.5560. To help decide the best trained method, the six other performance indicators were included: R.S.R., NSE, M.A.E., Index of Agreement, PBIAS (%), and RMSE. Each of the performance indexes claimed different ML methods as the best trained. However, three out of six performance indicators showed ANN as the most suitable method.

The next step was to predict the temperatures for the validation/prediction period from 1/1/2015 to 9/22/2021. As expected, the R2 results revealed that SVM had the least accuracy in predicting the temperature. On the other hand, the other five methods performed well, with similar results to one another in terms of R2. Therefore, the other six performance indexes were calculated for the results of each ML method. The analysis showed that in the prediction step, ANN was superior to all other methods in five out of six performance indicators. Random Forest was the only method that obtained better results than ANN in terms of the PBIAS index. However, the difference between the PBIAS of ANN and Random Forest was very slight. By considering the training and validation/prediction results, the ANN method's superiority over the other ML methods in air temperature prediction was concluded in **Error! Reference source not found.**.

From this study, it was observed that the ANN ML method was the best method for predicting temperature values when using dewpoint, relative humidity, wind speed, sea pressure, and 1-hour precipitation values as dependent variables. Future studies in climatology in regions around Memphis, TN should consider using ANN.

Conflict of interest statement

The authors declare no conflict of interest.

References

[1] S. Thober, R. Kumar, N. Wanders, A. Marx, M. Pan, O. Rakovec, L. Samaniego, J. Sheffield, E. F. Wood, M. Zink, Multi-model ensemble projections of European river floods and high flows at 1.5, 2, and 3 degrees global warming, Environ. Res. Lett. 13 (2018) 014003.

[2] K. Mohammed, A. K. M. S. Islam, G. M. T. Islam, L. Alfieri, M. J. U. Khan, S. K. Bala, M. K. Das, Future Floods in Bangladesh under 1.5°C, 2°C, and 4°C Global Warming Scenarios, J. Hydrol. Eng. 23 (2018) 04018050.

[3] B. Azari, M. Tabesh, Urban storm water drainage system optimization using a sustainability index and LID/BMPs,

Sustain. Cities Soc. 76 (2022) 103500.

[4] K. Azizi, C. I. Meier, Improving the Characterization of Urban Flash Floods through Application of Local Knowledge, in AGU Fall Meeting Abstracts, (2020), 2020, H162-0011.

[5] S. Ebrahimi, M. Khorram, Variability effect of hydrological regime on river quality pattern and its uncertainties: case study of Zarjoob River in Iran, J. Hydroinformatics 23 (2021) 1146–1164.

[6] M. Tabesh, B. Azari, H. Rezaei, Hydraulic performance assessment of stormwater networks, in Proceedings of the International Conference on Civil, Architecture and Disaster Management (ICCADM-16), Hong Kong, (2016), 17–18.

[7] B. Azari, M. Tabesh, Storm Water Drainage Networks Hydraulic Performance Assessment, International Conference On Sustainable Development And Urban Construction, Iran. (2015)

[8] K. Azizi, C. I. Meier, Urban Pluvial Flood Risk Assessment: Challenges and Opportunities for Improvement Using a Community-Based Approach, in World Environmental and Water Resources Congress 2021, (2021), 350–361. doi: 10.1061/9780784483466.033.

[9] D. Absalon, B. Ślesak, Air temperature increase and quality of life in an anthropogenically transformed environment: A case study, Polish J. Environ. Stud. 21 (2012) 235–239.

[10] Q. Song, B. S. Chissom, Forecasting enrollments with fuzzy time series — Part I, Fuzzy Sets Syst. 54 (1993) 1–9.

[11] S.-M. Chen, J.-R. Hwang, Temperature prediction using fuzzy time series, IEEE Trans. Syst. Man, Cybern. Part B 30 (2000) 263–275.

[12] N. Zhong, S. S. Yau, J. Ma, S. Shimojo, M. Just, B. Hu, G. Wang, K. Oiwa, Y. Anzai, Brain Informatics-Based Big Data and the Wisdom Web of Things, IEEE Intell. Syst. 30 (2015) 2–7.

[13] Z. Su, Q. Xu, Q. Qi, Big data in mobile social networks: a QoE-oriented framework, IEEE Netw. 30 (2016) 52–57.

[14] J. Wen, J. Yang, B. Jiang, H. Song, H. Wang, Big Data Driven Marine Environment Information Forecasting: A Time Series Prediction Network, IEEE Trans. Fuzzy Syst. 29 (2021) 4–18.

[15] M. Alizamir, O. Kisi, A. N. Ahmed, C. Mert, C. M. Fai, S. Kim, N. W. Kim, A. El-Shafie, Advanced machine learning model for better prediction accuracy of soil temperature at different depths, PLoS One 15 (2020) 1–25.

[16] B. Azari, M. Tabesh, Optimal design of stormwater collection networks considering hydraulic performance and BMPs, Int. J. Environ. Res. 12 (2018) 585–596.

[17] A. Rahman, A. Khan, A. A. Raza, Parkinson's disease detection based on signal processing algorithms and machine learning, CRPASE Trans. Electr. Electron. Comput. Eng. 6 (2020) 141–145.

[18] S. R. Moosavia, D. A. Woodb, S. A. Samadanic, Modeling Performance of Foam-CO2 Reservoir Flooding with Hybrid Machine-learning Models Combining a Radial Basis Function and Evolutionary Algorithms, methods 4 (2020) 5.

[19] K. A. Marill, Advanced Statistics: Linear Regression, Part II: Multiple Linear Regression, Acad. Emerg. Med. 11 (2004) 94–102.

[20] M. M. Ahmadi, H. Shahriari, Y. Samimi, A novel robust control chart for monitoring multiple linear profiles in phase II, Commun. Stat. - Simul. Comput. (2020) 1–12.

[21] S. Mehri, M. M. Ahmadi, H. Shahriari, A. Aghaie, Robust process capability indices for multiple linear profiles, Qual. Reliab. Eng. Int. 37 (2021) 3568–3579.

[22] J. Liang, Q. Yang, T. Sun, J. D. Martin, H. Sun, L. Li, MIKE 11 model-based water quality model as a tool for the evaluation of water quality management plans, J. Water Supply Res. Technol. 64 (2015) 708–718.

[23] I. Saini, D. Singh, A. Khosla, QRS detection using K-Nearest Neighbor algorithm (KNN) and evaluation on standard ECG databases, J. Adv. Res. 4 (2013) 331–344.

[24] X. Wang, L. Ma, X. Wang, Apply semi-supervised support vector regression for remote sensing water quality retrieving, in 2010 IEEE International Geoscience and Remote Sensing Symposium, (2010), 2757–2760. doi: 10.1109/IGARSS.2010.5653832.

[25] S. Liu, H. Tai, Q. Ding, D. Li, L. Xu, Y. Wei, A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction, Math. Comput. Model. 58 (2013) 458–465.

[26] A. Najah, A. El-Shafie, O. A. Karim, A. H. El-Shafie, Application of artificial neural networks for water quality prediction, Neural Comput. Appl. 22 (2013) 187–201.

[27] K. P. Singh, A. Basant, A. Malik, G. Jain, Artificial neural network modeling of the river water quality—A case study, Ecol. Modell. 220 (2009) 888–895.

[28] B. Fotovvati, M. Balasubramanian, E. Asadi, Modeling and Optimization Approaches of Laser-Based Powder-Bed Fusion Process for Ti-6Al-4V Alloy, Coatings 10 (2020)

[29] P. Hosseinzadeh Talaee, Daily soil temperature modeling using neuro-fuzzy approach, Theor. Appl. Climatol. 118 (2014) 481–489.

[30] A. Jain, R. W. McClendon, G. Hoogenboom, R. Ramyaa, Prediction of frost for fruit protection using artificial neural networks, Am. Soc. Agric. Eng. St. Joseph, MI, ASAE Pap. (2003) 3–3075.

[31] B. A. Smith, R. W. McClendon, G. Hoogenboom, Improving air temperature prediction with artificial neural networks, Int. J. Comput. Intell. 3 (2006) 179–186.

[32] S. Salcedo-Sanz, R. C. Deo, L. Carro-Calvo, B. Saavedra-Moreno, Monthly prediction of air temperature in Australia and New Zealand with machine learning algorithms, Theor. Appl. Climatol. 125 (2016) 13–25.

[33] A. Azad, H. Kashi, S. Farzin, V. P. Singh, O. Kisi, H. Karami, H. Sanikhani, Novel approaches for air temperature prediction: A comparison of four hybrid evolutionary fuzzy models, Meteorol. Appl. 27 (2020) e1817.

[34] Y. Wang, Y. Bai, L. Yang, H. Li, Short time air temperature prediction using pattern approximate matching, Energy Build. 244 (2021) 111036.

[35] M. Moeini, A. Shojaeizadeh, M. Geza, Supervised Machine Learning for Estimation of Total Suspended Solids in Urban Watersheds, Water 13 (2021) 147.

[36] P. Kumar, S. F. Shah, M. A. Uqaili, L. Kumar, R. F. Zafar, Forecasting of Drought: A Case Study of Water-Stressed Region of Pakistan, Atmosphere (Basel). 12 (2021) 1248.

[37] Memphis - Shelby County Airport Authority, Airport History, Memphis International Airport, (2021).

[38] D. D. Graham, W. S. Parks, Potential for leakage among principal aquifers in the Memphis area, Tennessee, Memphis, TN, (1986). [Online]. Available: http://pubs.er.usgs.gov/publication/wri854295

[39] D. Larsen, J. Bursi, B. Waldron, S. Schoefernacker, J. Eason, Recharge pathways and rates for a sand aquifer beneath a loess-mantled landscape in western Tennessee, U.S.A., J. Hydrol. Reg. Stud. 28 (2020) 100667.

[40] PRISM Climate Group, 30-Year Normals, PRISM Climate Group, (2015).

[41] U.S. Census Bureau, City and Town Population Totals: 2010 - 2019, U.S. Census Bureau, 2021.

[42] Department of Agronomy, Tennessee ASOS: MEM Memphis International Airport Weather Data 2015, Iowa State University: Environmental Mesonet, (2021). https://mesonet.agron.iastate.edu/request/download.phtml?network=TN_ASOS

[43] S. C. Chapra, Applied numerical methods with MATLAB for engineers and scientists. New York, N.Y.: McGraw-Hill Education, (2018).

[44] Shojaeizadeh, Geza, McCray, Hogue, Site-Scale Integrated Decision Support Tool (i-DSTss) for Stormwater Management, Water 11 (2019) 2022.

[45] T. J. Oyana, Spatial Analysis with R: Statistics, Visualization, and Computational Methods. CRC press, (2020).

[46] L. E. Peterson, K-nearest neighbor, Scholarpedia 4 (2009) 1883.

[47] B. Schölkopf, A. J. Smola, B. Francis, Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, (2002).

[48] A. J. Smola, B. Schölkopf, A tutorial on support vector regression, Stat. Comput. 14 (2004) 199–222.

[49] I. N. Daliakopoulos, P. Coulibaly, I. K. Tsanis, Groundwater level forecasting using artificial neural networks, J. Hydrol. 309 (2005) 229–240.

[50] N. T. Duy, Apply the Artificial Neural Network to Diagnose Potential Fault of Power Transformer Based on Dissolved Gas-in-oil Analysis Data, Comput. Res. Prog. Appl. Sci. Eng. Trans. Electr. Electron. Comput. Eng. 6 (2020) 127–131.

[51] C. W. Dawson, R. L. Wilby, Hydrological modelling using artificial neural networks, Prog. Phys. Geogr. 25 (2001) 80–108.

[52] A. M. Melesse, S. Ahmad, M. E. McClain, X. Wang, Y. H. Lim, Suspended sediment load prediction of river systems: An artificial neural network approach, Agric. Water Manag. 98 (2011) 855–866.

[53] M. Zounemat-Kermani, O. Batelaan, M. Fadaee, R. Hinkelmann, Ensemble machine learning paradigms in hydrology: A review, J. Hydrol. 598 (2021) 126266.

[54] S. Ardabili, A. Mosavi, A. R. Várkonyi-Kóczy, Advances in machine learning modeling reviewing hybrid and ensemble methods, in International Conference on Global Research and Education, (2019), 215–227.

[55] K. A. Nguyen, W. Chen, B.-S. Lin, U. Seeboonruang, Comparison of Ensemble Machine Learning Methods for Soil Erosion Pin Measurements, ISPRS Int. J. Geo-Information 10 (2021) 42.

[56] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.

[57] L. T. Pham, L. Luo, A. Finley, Evaluation of random forests for short-term daily streamflow forecasting in rainfall-and snowmelt-driven watersheds, Hydrol. Earth Syst. Sci. 25 (2021) 2997–3015.

[58] B. Li, G. Yang, R. Wan, X. Dai, Y. Zhang, Comparison of random forests and other statistical methods for the prediction of lake water level: a case study of the Poyang Lake in China, Hydrol. Res. 47 (2016) 69–83.

[59] T. Yang, X. Liu, L. Wang, P. Bai, J. Li, Simulating hydropower discharge using multiple decision tree methods and a dynamical model merging technique, J. Water Resour. Plan. Manag. 146 (2020) 4019072.

[60] B. Azari, Stormwater Collection System Optimal Design by Considering BMPs and Uncertainties, University of Tehran, (2016).

[61] A. H. Nury, K. Hasan, M. J. Bin Alam, Comparative study of wavelet-ARIMA and wavelet-ANN models for temperature time series data in northeastern Bangladesh, J. King Saud Univ. - Sci. 29 (2017) 47–61.

[62] D. N. Moriasi, J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, T. L. Veith, Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, Trans. ASABE 50 (2007) 885–900.

[63] C. J. Willmott, On the validation of models, Phys. Geogr. 2 (1981) 184–194.

[64] J. E. Nash, J. V Sutcliffe, River flow forecasting through conceptual models part I—A discussion of principles, J. Hydrol. 10 (1970) 282–290.

[65] P. Amarasinghe, N. S. Abeygunawardana, T. N. Jayasekara, E. Edirisinghe, S. K. Abeygunawardane, Ensemble models for solar power forecasting—a weather classification approach, AIMS Energy 8 (2020) 252–271.

[66] H. V. Gupta, S. Sorooshian, P. O. Yapo, Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration, J. Hydrol. Eng. 4 (1999) 135–143.

[67] J. Demšar, T. Curk, A. Erjavec, Č. Gorup, T. Hočevar, M. Milutinovič, M. Možina, M. Polajnar, M. Toplak, A. Starič, Orange: data mining toolbox in Python, J. Mach. Learn. Res. 14 (2013) 2349–2353.

[68] A. Naik, L. Samant, Correlation review of classification algorithm using data mining tool: WEKA, Rapidminer, Tanagra, Orange and Knime, Procedia Comput. Sci. 85 (2016) 662–668.

[69] J. Hill, J. Drake, B. Sleep, L. Margolis, Influences of four extensive green roof design variables on stormwater hydrology, J. Hydrol. Eng. 22 (2017) 4017019.

[70] E. Aytaç, Unsupervised learning approach in defining the similarity of catchments: Hydrological response unit based k-means clustering, a demonstration on Western Black Sea Region of Turkey, Int. Soil Water Conserv. Res. 8 (2020) 321–331.

[71] B. A. Smith, R. W. Mcclendon, G. Hoogenboom, Improving Air Temperature Prediction with Artificial Neural Networks, Int. J. Comput. Inf. Eng. 1 (2007) 3159.

[72] T. Hastie, S. Rosset, J. Zhu, H. Zou, Multi-class AdaBoost, Stat. Interface 2 (2009) 349–360.