International Journal of Environmental Trends (IJENT)
http://dergipark.gov.tr/ijent

International Journal of Environmental Trends (IJENT)

2024; 8 (2),76-88
ISSN: 2602-4160
Turkish JournalPark ACADEMIC

Research Article

Soil Temperature Prediction for Konya, Türkiye: Machine Learning Approaches

Kazım KUMAŞ[1], Ali AKYÜZ*[2]

[1] Bucak Emin Gulmez Vocational School of Technical Sciences, Burdur Mehmet Akif Ersoy University, Burdur, 15300, Türkiye, kkumas@mehmetakif.edu.tr *(ORCID Number 0000-0002-2348-4664:)*

[2] Bucak Emin Gulmez Vocational School of Technical Sciences, Burdur Mehmet Akif Ersoy University, Burdur, 15300, Türkiye, kkumas@mehmetakif.edu.tr, *(ORCID Number: 0000-0001-9265-7293)*

Abstract

Soil temperature is a critical parameter for agriculture meteorology applications. Although highly accurate, direct measurement may not be practical over large areas. The measurement process can also be costly and time-consuming. On the other hand, variables such as surface and soil properties that affect soil temperature can make it difficult to predict with physical models. Machine learning methods can overcome various limitations and predict targeted variables using complex non-linear relationships in the data distribution. For this purpose, it is used in many fields. Machine learning approaches are sensitive to input data and require many training data. This paper studied 5, 10, 20, and 50 cm soil temperature values of Konya province between 1960 and 2021 using machine learning algorithms (k-nearest neighbors, adaptive boosting, gradient boosting, light gradient boosting machine (LGBM)). The models were trained using data from 1960 to 2017, and the years 2019, 2020, and 2021 were predicted. In line with the successful results achieved, these models were used to predict the years 2022, 2023, 2024, and 2025.

Received
3 August 2024

Accepted
18 December 2024

Keywords
Konya
Machine Learning
Prediction
Soil Temperature

Konya, Türkiye için Toprak Sıcaklığı Tahmini: Makine Öğrenmesi Yaklaşımları

Özet

Toprak sıcaklığı tarımsal meteoroloji uygulamaları için kritik bir parametredir. Yüksek doğruluk oranına sahip olmasına rağmen, doğrudan ölçüm geniş alanlar için pratik olmayabilir. Ölçüm süreci maliyetli ve zaman alıcı olabilir. Öte yandan, toprak sıcaklığını etkileyen yüzey ve toprak özellikleri gibi değişkenler fiziksel modellerle tahmin etmeyi zorlaştırabilir. Makine öğrenmesi yöntemleri çeşitli sınırlamaların üstesinden gelebilir ve veri dağılımındaki karmaşık doğrusal olmayan ilişkileri kullanarak hedeflenen değişkenleri tahmin edebilir. Bu amaçla birçok alanda kullanılmaktadır. Makine öğrenmesi yaklaşımları giriş verilerine duyarlıdır ve çok sayıda eğitim verisi gerektirir. Bu makalede, makine öğrenmesi algoritmaları (k-en yakın komşular, adaptif yükseltme, gradyan yükseltme, hafif gradyan yükseltme makinesi (LGBM)) kullanılarak 1960-2021 yılları arasında Konya ilinin 5, 10, 20 ve 50 cm toprak sıcaklığı değerleri incelenmiştir. Modeller 1960-2017 yılları arasındaki veriler kullanılarak eğitilmiş ve 2019, 2020 ve 2021 yılları tahmin edilmiştir. Elde edilen başarılı sonuçlar doğrultusunda bu modeller kullanılarak 2022, 2023, 2024 ve 2025 yılları tahmin edilmiştir.

Anahtar Kelimeler
Konya
Makine Öğrenmesi
Tahmin
Toprak Sıcaklığı

Corresponding Author Email: aakyuz@mehmetakif.edu.tr.

INTRODUCTION

Soil temperature directly affects the development processes for plants in the agricultural field. Soil temperature determines the speed and rate of germination of seeds. It also has a direct effect on the development of plant roots. Microorganism activities in the soil are directly dependent on soil temperature [1]. Soil temperature affects water uptake, transpiration, and nutrient uptake by plants. Plants can take up nutrients more efficiently depending on soil temperature. The freezing and thawing processes of soil temperature affect soil structure and plant roots [2]. Soil temperature is a critical factor that directly affects productivity and plant health in agricultural production. Therefore, monitoring and managing soil temperature in agricultural activities is important to ensure optimal growth conditions for plants [3]. Since climate changes and agricultural practices can affect soil temperature, farmers need to plan to take this factor into account. High soil temperatures can affect soil structural integrity in several ways. Increasing temperatures can cause the expansion and contraction of soil particles, leading to cracks on the surface.

Furthermore, this can accelerate the evaporation of water, reducing the moisture content of the soil and causing a loss of organic matter, which leads to soil structural degradation and subsidence. High temperatures degrade soil structure, increasing the risk of dry layers and erosion, and can also lead to soil compacting. Measures such as irrigation management, vegetation protection, and regular soil moisture monitoring should be taken to reduce these negative impacts [4-6]. Soil temperature is also important for heat pumps, pipes, or ground heat exchangers [7,8].

Konya has an important place in Türkiye's agricultural sector. Konya has the largest surface area in Türkiye and is known for its large and fertile agricultural lands. These large lands allow for the cultivation of various agricultural products. This makes Konya the agricultural production center of Türkiye. Konya is a pioneer in producing cereal crops, especially wheat and barley. A significant portion of Türkiye's wheat production is supplied from Konya. Wheat is a critical crop for Türkiye's food security and economy. Konya also plays an important role in sugar beet and potato production. Sugar beet is a basic raw material for sugar production and is grown in Konya with high yields. Potatoes are both consumed in the domestic market and exported. Soil subsidence and giant sinkholes have occurred in Konya in recent years. These events may be caused by various factors resulting from global climate change. Among these reasons, soil temperature can also be listed [9-13].

Predicting soil temperature offers critical advantages ranging from agricultural production to environmental management [14]. In terms of agricultural planning and management, accurate forecasts increase crop productivity by optimizing planting and harvest timing. Soil temperature determines the water requirements of crops. Forecasts support the efficient use of water resources by optimizing irrigation schedules [15].

Pests and diseases become more active at certain soil temperatures. Accurate predictions ensure these threats are detected early, and preventive measures are taken. Early warning systems and accurate forecasts prevent economic losses by minimizing crop losses. Climate change can cause significant changes in soil temperatures. Temperature forecasts support long-term agricultural planning and the development of sustainable agricultural practices [16, 17].

Machine learning algorithms in agriculture, agronomy, and soil science can potentially transform these fields. By analyzing large data sets, these algorithms can more accurately predict soil health, plant growth, and the impacts of climate change. For example, modeling critical parameters such as soil temperature and moisture can improve agricultural productivity and optimize resource use. In addition, machine learning techniques offer important contributions to the early detection of plant diseases, predicting pests, and optimizing agricultural production. These technologies' integration supports sustainable agricultural practices and improves food security, providing more effective solutions to future agricultural challenges [18, 19].

Literature Review

Although there are studies on modeling and predicting subsoil temperature using artificial intelligence techniques and algorithms in the literature, no study on Konya has been found. Some of the articles we found in the literature on this subject are given chronologically.

Bilgili et al. proposed an artificial neural network model (ANN) to predict monthly average soil temperatures in the Aegean Region of Turkey. Soil temperatures and topographic information collected from various meteorological stations between 2000 and 2006 were used in the model training. When the prediction results of the model created with MATLAB program were compared with the actual values, it was found that the error values were within acceptable limits [20].
Ozturk et al. conducted the development of feed-forward artificial neural network (ANN) models to predict soil temperatures at 5, 10, 20, 50 cm depths with data collected from 66 monitoring stations across Turkey. The models

were trained and tested using standard geographical and meteorological data, resulting in a high correlation (98.91% - 95.37%) between ANN predictions and measured soil temperatures. The results show that ANN modeling is a reliable method for predicting monthly average soil temperature in areas of Turkey without monitoring stations [21].

Bilgili developed an ANN to predict the average soil temperature of the current month using the previous month's average meteorological data. A 3-layer feed-forward ANN structure was constructed using soil temperatures and meteorological data measured between 2000 and 2007 from Adana meteorological station and trained by a back propagation algorithm. The results show that the ANN model is reliable for predicting monthly average soil temperature [22].

Citakoglu developed ANN, ANFIS, and MLR models using 20 years of soil temperature data from 261 stations in Turkey. ANFIS best-predicted soil temperature with monthly minimum and maximum air temperatures, number of calendar months, soil depth, and monthly precipitation. In the model evaluation, the RMSE, MAE, and R^2 values of ANFIS were 1.99, 1.09, and 0.98, respectively, while the results of ANN and MLR models showed lower performance with higher error values and lower R^2 [23].

Behmanesh and Mehdizadeh used gene expression programming (GEP), artificial neural networks (ANN), and multiple linear regression (MLR) to predict soil temperature at six different depths (5, 10, 20, 30, 50 cm) at Sanandaj station in western Iran. The accuracy of the models was evaluated using various combinations of meteorological parameters and data sets from 1997 to 2008. The results showed that ANN outperformed the other methods and predicted soil temperature best [24].

Kara and Cemek used ANN to predict monthly average soil temperatures at various soil depths in the Central Black Sea region. The three-layer ANN structure created with meteorological data obtained between 1971 and 2015 was trained with the Levenberg-Marquardt algorithm and the coefficient of determination of the results at different depths was found to be 0.85-0.99, standard error 0.24-3.74 and mean absolute error 0.01-2.33. The results showed that ANN models can successfully predict monthly soil temperatures in the Central Black Sea [25].

Feng et al. evaluated four machine learning models (ELM, GRNN, BPNN, and RF) for half-hourly soil temperature forecasts at four different depths (2 cm, 5 cm, 10 cm, 20 cm) on the Loess Plateau of China. The models were trained using meteorological data such as air temperature, wind speed, relative humidity, solar radiation, and lack of vapor pressure, and all models provided high accuracy; the ELM model performed both faster and slightly better than the other models. Statistically significant agreement was found between measured and predicted values, and the ELM model stood out as the most effective model recommended for soil temperature prediction [26].

Pekel applied decision tree regression to predict soil moisture using parameters such as air temperature, time, relative humidity, and soil temperature. The use of decision tree regression gave effective results with a high coefficient of determination (R^2), low mean square error (MSE), and mean absolute error (MAE). The results show that the decision tree provided the highest fit values at five depth levels and successfully predicted soil moisture [27].

Alizamir et al. compared four machine learning techniques (ELM, ANN, CART, and GMDH) to predict monthly soil temperatures at four depths. ELM outperformed the other methods in general, and the model performance decreased with the increase in soil depth. It was concluded that soil temperatures at 5, 10, and 50 cm depths can be predicted using only air temperature data, while at 100 cm depth, soil temperatures can also be predicted using solar radiation and wind speed information [28].

Fathololoum et al. used machine-learning methods to model soil temperature (ST) in pot experiments in Ardebil, Iran. ANFIS was the best predictor of soil temperature at 5 cm depth (R^2=0.96, MAPE=10.5) under the influence of air temperature. The results show that ANFIS is an effective modeling tool for soil depths and locations where data gaps exist [29].

Benos et al. reviewed journal articles published between 2018 and 2020 according to PRISMA guidelines to examine how machine learning can be used in agriculture. The research showed that machine learning algorithms, especially Agricultural Neural Networks, are fruitful, with plants and animals such as corn, wheat, cattle, and sheep being the most studied. It was also noted that sensors connected to satellites and crewless aerial vehicles are used for data analysis, and the study is envisioned to be a useful guide to increasing the potential advantages of machine learning in agriculture [30].

Guleryuz used Bayesian Tuned Gaussian Process Regression (BT-GPR), Bayesian-tuned Support Vector Regression (BT-SVR), and Long Short-Term Memory (LSTM) models to predict soil temperature at Giresun and Bayburt stations in Turkey. The BT-GPR model provided the highest accuracy (RMSE=0.0439, R^2=0.9535, MAE=0.0344) in both semiarid and humid climates. The results show that the BT-GPR model provides superior performance in soil temperature prediction and is a valuable reference for future studies [31].

Bilgili et al., using daily soil temperature data from the Sivas meteorological observation station, six different machine learning techniques such as ANFIS-FCM, ANFIS-GP, ANFIS-SC, FNN, ENN, and LSTM were used to predict the soil temperature of the day before. The performance of the models was evaluated using four statistical metrics: mean absolute error, root mean square error (RMSE), Nash-Sutcliffe efficiency coefficient, and correlation coefficient. The results showed that ANFIS-FCM, ANFIS-GP, ANFIS-SC, ENN, FNN, and LSTM models performed satisfactorily at all depths and provided highly accurate predictions with high RMSE and R values [32].

Elmi et al. compared soil temperatures measured at three different depths (5, 50, 100 cm) in Kuwait between 2007 and 2016 using a regional climate model (RegCM4) and regression models. While the RegCM4 model predicted soil temperatures well near the surface but inadequate in deeper soil layers, the linear scaling (LS) method improved these predictions. These findings suggest that the RegCM4 model can reliably predict soil temperature in arid ecosystems [33].

In this study, 5, 10, 20, 50 cm subsoil temperature values of Konya province between 1960-2021 were modeled with machine learning algorithms (k-nearest neighbors (KNN), Adaptive Boosting (Adabust), Gradient Boosting (GB), Light Gradient Boosting Machine (LGBM)). Accordingly, the models were trained using data between 1960 and 2017; 2019, 2020, and 2021 were predicted. In line with the successful results obtained, these models were used to predict 2022,2023,2024,2025.

MATERIAL AND METHOD

Monthly average soil temperatures at different depths were obtained from the Turkish State Meteorological Service for Konya province [34]. The subsoil temperatures obtained are monthly for 5 cm, 10 cm, 20 cm, and 50 cm. The data covers the years 1960-2021. The monthly average soil temperature of 5 cm contains data for 744 months and missing data for some years. For 131 months between 1960 and 2017, the records in the relevant data set, for which there was no data, were deleted and cleaned. After the data cleaning, the remaining 565 data between 1960-2017 were used for training machine learning algorithms, while 36 data between 2019-2021 were tried to be predicted. 2018 was not predicted because there were no measurement values at the station in that year. For the Monthly Average 10 cm Soil Temperature, the records in the relevant dataset, which had no data for 143 months between 1960-2017, were deleted and cleaned. 556 data were used for training, while 36 data between 2019 and 2021 were used for forecasting. For the monthly average 20 cm soil temperature, the records in the relevant dataset, which did not have data for 135 months between 1960-2017, were deleted and cleaned. 561 data were used for training, while 36 data between 2019 and 2021 were used for prediction. For the monthly average of 50 cm soil temperature and monthly average of 100 cm soil temperature, the records in the relevant data set, for which data were not available for 132 months between 1960 and 2017, were deleted and cleaned. Afterwards, 564 data were used for training, while 36 data between 2019 and 2021 were used for prediction. In order to make future forecasts (2022-2025), the problem was considered a time series. For this reason, year and month information are used as attributes for model training, while the output values are subsoil temperatures.

This study has used regression-based machine learning algorithms to predict soil temperatures consisting of real number values. As machine learning algorithms, K-Nearest Neighbors (KNN), Adaptive Boosting (ADABOOST), Gradient Boosting (GB), and Light Gradient Boosting Machine (LGBM) were preferred. These machine learning algorithms offer several advantages in regression problems. While KNN is preferred due to its simplicity and interpretability, it makes predictions using the data directly. ADABOOST improves model performance and provides high accuracy by successively strengthening weak learners. GB combines successive models using gradient descent to minimize errors and make robust predictions. LGBM is a faster and more efficient version of GB, performing highly on large data sets and optimizing memory usage. These machine learning algorithms were chosen because they offer clear advantages in regression problems with speed and low error rates. A description of these methods were given below.

The KNN algorithm is a simple and effective machine-learning method that is used in classification and regression tasks. Methodologically, KNN makes predictions based on the labels or values of the K nearest neighbours of a given sample. The algorithm calculates the distances between instances in the dataset, usually using Euclidean distance, and collects the data of the K nearest neighbours for a new instance. This approach can directly model the effect of local data structures without requiring a priori knowledge of the data distribution and can provide high performance. However, for large data sets, the computational cost can be high, and the choice of K value can significantly affect the model's accuracy [35, 36].

ADABOOST is an efficient algorithm among ensemble learning methods that builds a strong classifier from a combination of weak learners. The basic principle of AdaBoost is to iteratively train each weak learner (usually short decision trees). The training data starts with equal weights in the first step, and the first weak learner is trained. During training, the differences between the predictions and the actual labels are calculated, and the errors are weighted. The weights of the instances that make incorrect predictions are increased, while those that make correct predictions are decreased. These weight updates help the next weaker learner to correct errors. This process is repeated until training is complete, and each weak learner is weighted according to its ability to correct errors. With this method, AdaBoost creates a high-performance classifier from the combination of weak models and improves model accuracy by increasing generalization ability [35, 37].

The GB algorithm is a powerful ensemble learning method widely used in machine learning. GB allows each new model to correct the errors of previous models using a sequential training process. In the first step, a simple prediction is made, usually with the mean of the target variable (for regression) or the most common class (for classification). The difference between these predictions and the actual values is calculated to obtain the error terms (residual). The error terms are attempted to be estimated by a new weak learner (usually a short decision tree), and each new model is trained to correct the existing error. The predictions of the new model are combined with the predictions of the previous models, and the process continues until a certain number of iterations is reached or the model's performance becomes satisfactory. The GB algorithm provides high prediction accuracy and, thanks to its flexibility, can be effectively applied to various data types and problems [38, 39].

LGBM is a gradient-boosting framework for fast and efficient training on large data sets. LGBM applies the gradient boosting algorithm on decision trees and optimizes the tree structures to improve the model's prediction performance. In particular, using a histogram-based approach partitions features into specific intervals and builds decision trees according to them, speeding up the training process and reducing memory usage. It also performs data reads in blocks, enabling fast and efficient data processing even with large data sets. LGBM can efficiently process large data sets with techniques that optimize the model's memory usage, offer high accuracy and generalization capability, and have a wide range of applications with various types of boosting and model parameters [40,41].

RESULT AND DISCUSSION

The data set for training and testing machine learning models is considered a time series. For this reason, the training set consists of data between 1960-2017, while the test set consists of data between 2019-2021. 2018 could not be used because there was no data available at the station in that year. In line with the success of the machine learning model, the model was also used to predict the years 2022-2025.

In order to evaluate the machine learning models, MAE, RMSE, R^2, popularly preferred in regression problems, were used as evaluation metrics. In the literature, various metrics are used to analyze the performance of models trained by machine learning. In this study, root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination (R2) have been used to identify the best model. The formulas for, RMSE, MAE, and R2 are presented in (Equations (1)-(3)) respectively.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2} \quad (1)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right| \quad (2)$$

In Equation (1-4), Y is the target value, $\hat{Y}$ is the predicted value, and n is the number of samples.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y}_i)^2} \quad (3)$$

In Eq. (3), Y represents the target value, $\hat{Y}$ the predicted value, $\bar{Y}$ the mean of the target value, and n the number of samples.

Analyzing the success results in the training set, KNN obtained an R^2 value of 0.9593 for 5 cm soil temperature, whereas the R^2 value for the GB method was 0.9604. For other methods, this value is above 0.98. For 10 cm soil temperature, R^2 values of machine learning models are 0.9526 and above. For 20 cm soil temperature, R^2 values are 0.9577 and above. Finally, for the 50 cm soil temperature, the KNN method obtained a value of 0.9615, while the other methods obtained results above this value. Considering the test results for 2019-2021, the lowest R^2 value for soil temperature at 5 cm depth was obtained as 0.6388 with KNN, while the highest R^2 value was obtained as 0.9330 with LGBM. Table 1 shows the MAE, RMSE, and R^2 values obtained for the training and testing KNN, ADABOOST, GB, and LGBM algorithms.

Figure 1. compares the prediction values obtained with the successful LGBM algorithm with the original temperature values. In Figure 1.a, the vertical axis shows the soil temperature in units of ^{0}C, and the horizontal axis shows the months between 2019 and 2021 (the same representation was used in all figures). It is seen that the original temperature measurement values and the prediction values had a good correlation. When the box plot in Figure 1.b is analyzed for LGBM, the lower quartile, or first quartile Q_1=2.25 ^{0}C, median (Q_2)=13.13 ^{0}C, third quartile, or upper quartile Q_3=23.01 ^{0}C. The minimum temperature value was 0.85 ^{0}C and the maximum temperature value was 30.09 ^{0}C.

For soil temperature at 10 cm depth, the R^2 values of machine learning models for the test set were 0.9219 and above, and the most successful result was obtained with LGBM as 0.9466. Figure 2. compares the prediction values obtained with the LGBM algorithm with the actual temperature values. This method's minimum temperature value was 2.38 ^{0}C, and the maximum temperature value was 28.82 ^{0}C. Q_1= 3.40, Q_2 = 12.84 and Q_3 = 23.18 ^{0}C, while the average temperature value was 14.70 ^{0}C (Figure 2.b).

For 20 cm soil temperature, R^2 values are 0.9060 and above and LGBM obtained 0.9382. Figure 3. a shows the comparison of LGBM with real data. It can be considered to be generally in line. In the box plot in Figure 3.b, the minimum and maximum temperature values are 2.54, 27.66 0C, while Q_1= 3.55, Q_2=12.43, Q_3=22.85 ^{0}C. For 50 cm soil temperature, the lowest R^2 value was 0.6288 with KNN, and the highest value was 0.8848 with ADABOOST (Figure 4.a). When the box plot in Figure 4.b is analyzed, Q_1= 5.94, Q_2= 13.85, Q_3= 22.96 ^{0}C, and the minimum and maximum temperature values were found as 3.72, 24.94 ^{0}C and the average temperature value was 14.46 ^{0}C.

Table 1. The MAE, RMSE and R^2 values for the training and tests for the models

Experiment		**TRAIN**			**TEST**		
	Algorithm	**RMSE**	**MAE**	**R^2**	**RMSE**	**MAE**	**R^2**
T=5 cm	KNN	1.956	1.523	0.9593	5.447	4.471	0.6388
	ADABOOST	1.928	1.526	0.9604	2.651	2.045	0.9145
	GB	1.363	1.094	0.9802	3.018	2.541	0.8891
	LGBM	1.287	1.020	0.9824	2.346	2.003	0.9330
T=10 cm	KNN	1.951	1.515	0.9526	2.441	1.976	0.9219
	ADABOOST	1.646	1.314	0.9663	2.300	1.624	0.9306
	GB	1.132	0.907	0.9841	2.372	1.945	0.9262
	LGBM	1.065	0.848	0.9859	2.018	1.602	0.9466
T=20 cm	KNN	1.768	1.383	0.9577	2.529	2.185	0.9060
	ADABOOST	1.668	1.332	0.9624	2.205	1.678	0.9285
	GB	1.058	0.851	0.9849	2.338	1.914	0.9197
	LGBM	0.978	0.778	0.9871	2.051	1.630	0.9382
T=50 cm	KNN	1.474	1.134	0.9615	4.043	3.603	0.6288
	ADABOOST	1.423	1.154	0.9641	2.252	1.908	0.8848
	GB	0.906	0.715	0.9854	2.802	2.386	0.8217
	LGBM	0.828	0.646	0.9878	2.319	2.039	0.8778

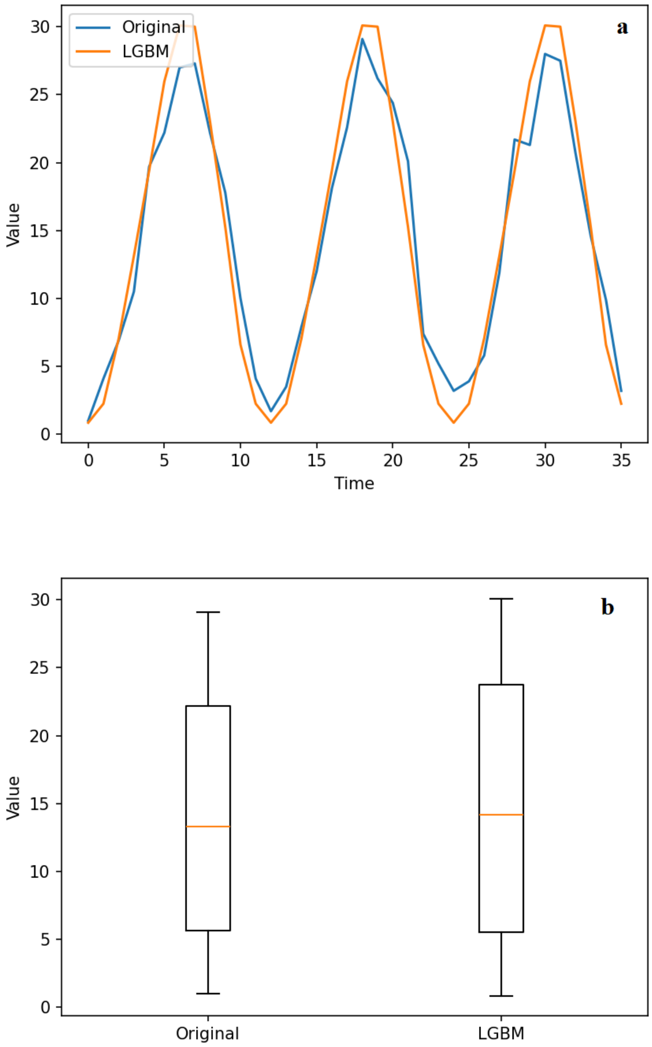
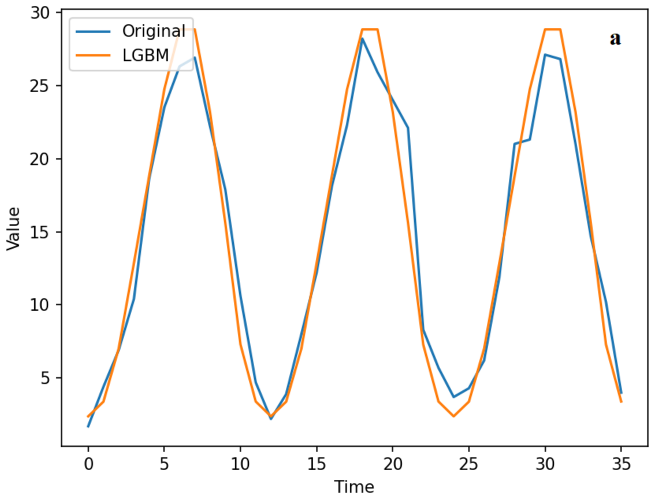


Figure 1. (a) Comparison of LGBM prediction results with original data for 5 ^{0}C soil temperature **(b)** Box plots of LGBM's predictions with original data.



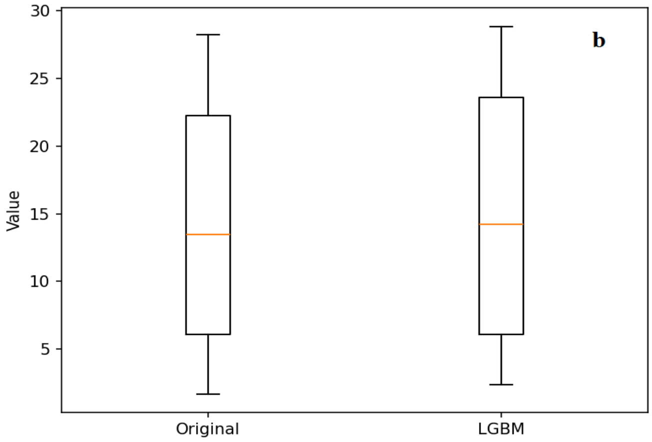
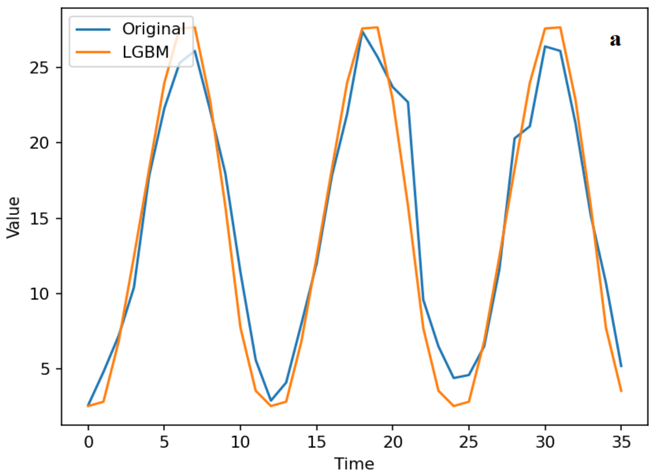


Figure 2. (a) Comparison of LGBM prediction results with original data for 10 ^{0}C soil temperature **(b)** Box plots of LGBM's predictions with original data.



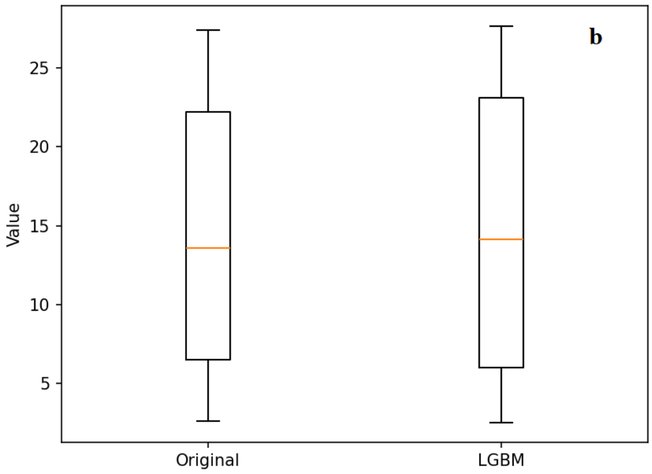
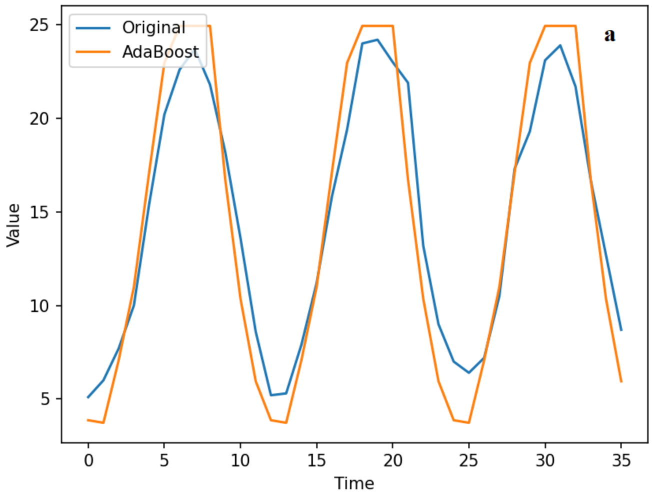


Figure 3. (a) Comparison of LGBM prediction results with original data for 20 ^{0}C soil temperature **(b)** Box plots of LGBM's predictions with original data.



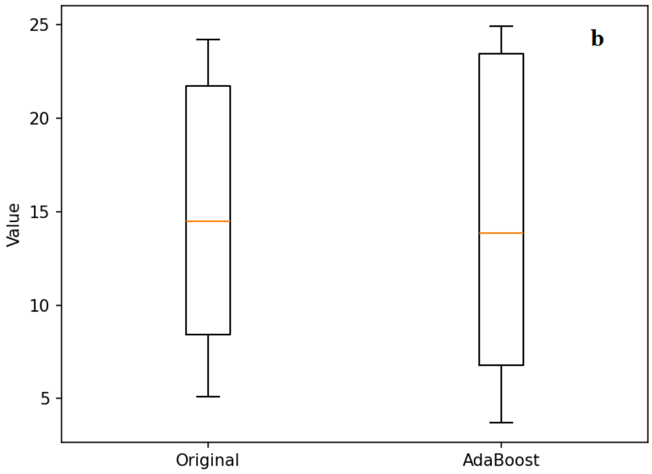


Figure 4. (a) Comparison of ADABOOST predictions with original data for 50 ^{0}C soil temperature **(b)** Box plots of ADABOOST's predictions with original data.

For the future predictions for the years 2022-2025, the LGBM and ADABOOST algorithms were selected, which gave successful results in the comparison. Figure 5 shows the prediction of the LGBM algorithm for a soil temperature of 5 cm. The vertical axis of Figure 5 shows the soil temperature (^{0}C), while the horizontal axis shows 48 months for the years 2022-2025 (the same notation was also used for Figures 6-8). The LGBM method predicts a minimum of 1.36 ^{0}C, a maximum of 28.61 ^{0}C, and an average of 14.48 0C for 5 cm soil temperature between these years. For 10 cm soil temperature, the minimum values were 2.59 ^{0}C, the maximum of 27.81 ^{0}C, and the average was 14.64 ^{0}C (Figure 6), while for 20 cm soil temperature, the minimum values were 3.40 ^{0}C, the maximum 26.90 ^{0}C and the average 14.76 0C (Figure 7). Finally, for 50 cm soil temperature, the minimum ADABOOST estimate was 4.92 ^{0}C, the maximum was 25.210C, and the average was 14.71 ^{0}C (Figure 8).

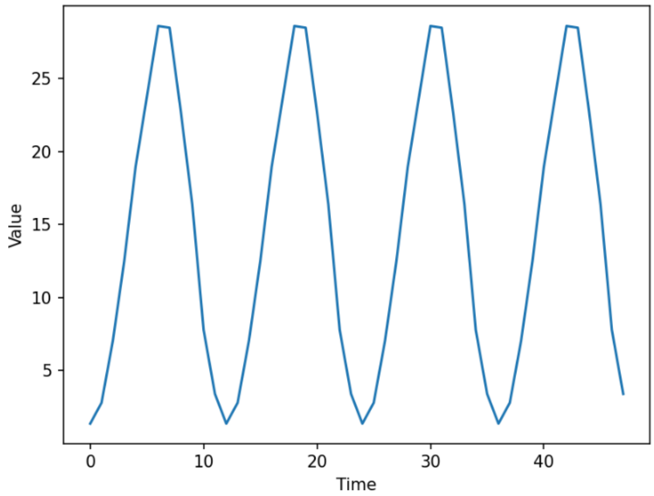


Figure 5. LGBM predictions for 5 cm soil temperature (2022-2025).

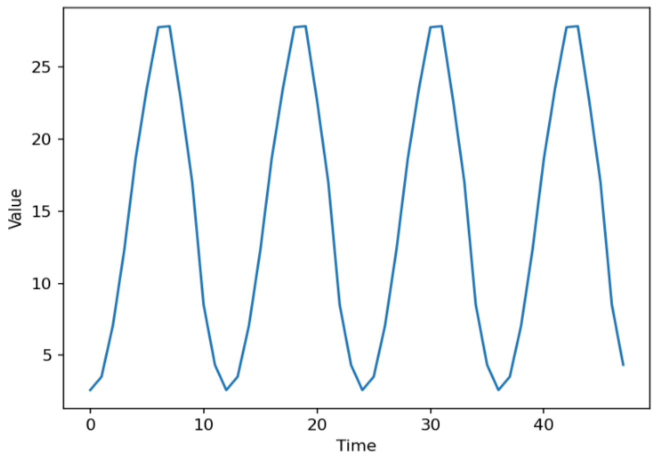


Figure 6. LGBM predictions for 10 cm soil temperature (2022-2025).

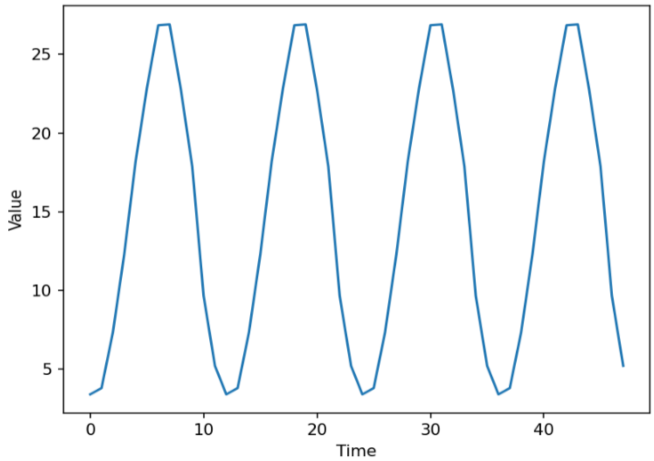


Figure 7. LGBM predictions for 20 cm soil temperature (2022-2025).

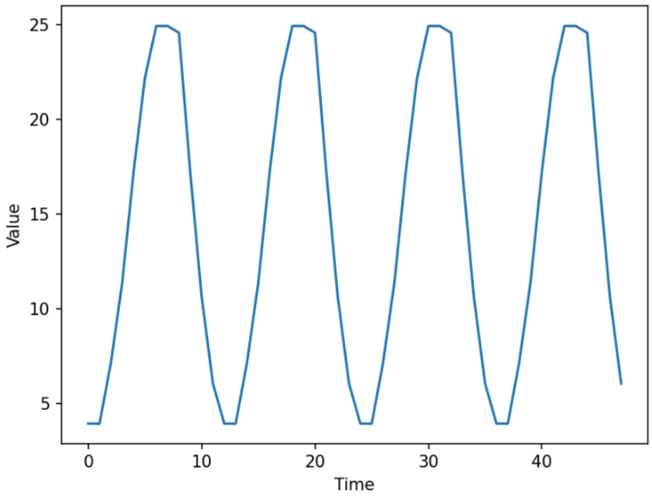


Figure 8. ADABOOST predictions for 50 cm soil temperature (2022-2025).

CONCLUSION

Soil temperature is a parameter that significantly affects soil's physical, chemical and biological properties. Meteorological variables, topographic conditions, and soil properties influence this temperature. The complexity of these interactions makes it difficult to estimate soil temperature. Although direct measurement of soil temperature over large areas can be done with thermistors, thermocouples, and similar instruments, it is costly and time-consuming and may not be practical. Therefore, soil temperature is often estimated using physical and statistical models that balance resolution, accuracy, and computational efficiency. Physically based models, while widely used in soil temperature prediction, are known for their high data requirements and complexity.

On the other hand, empirical methods require less data and offer a simple approach; however, they are based on site-specific statistical relationships. Especially for deep soil temperature predictions, combining available climate data with time information is important to achieve the best results. Therefore, improving both machine learning and statistical methods to take periodicity into account can improve the accuracy of soil temperature forecasts.

In this study, especially in the test set, the prediction values of 0.933 R^2 and above obtained for 5, 10, and 20 cm soil temperature show that machine learning models can successfully perform future predictions. For 50 cm soil temperature, ADABOOST was the most successful method and obtained an R^2 value of 0.8848. Although this value seems low compared to other soil temperatures, it is acceptable. Since the results obtained and the training of the machine learning models were carried out for Konya province, the model should be retrained with the relevant data for other provinces or Türkiye in general, and the success rate should be analyzed.

ACKNOWLEDGMENT

The authors thank the Turkish State Meteorological Service for the data used in the study.

REFERENCES

[1] Heinze, J., Gensch, S., Weber, E., & Joshi, J. 2017. Soil temperature modifies effects of soil biota on plant growth. Journal of Plant Ecology, 10(5), 808-821.
[2] Sharma, P. K., & Kumar, S. 2023. Soil Temperature and Plant Growth. In Soil Physical Environment and Plant Growth: Evaluation and Management (pp. 175-204). Cham: Springer International Publishing.
[3] Costa, J. M., Egipto, R., Aguiar, F. C., Marques, P., Nogales, A., & Madeira, M. 2023. The role of soil temperature in mediterranean vineyards in a climate change context. Frontiers in Plant Science, 14, 1145137.
[4] Ribeiro Filho, J. C., Andrade, E. M. D., Guerreiro, M. S., Palácio, H. A. D. Q., & Brasil, J. B. 2023. Soil–Water–Atmosphere Effects on Soil Crack Characteristics under Field Conditions in a Semiarid Climate. Hydrology, 10(4), 83.
[5] Xu, S., Nowamooz, H., Lai, J., & Liu, H. 2023. Mechanism, influencing factors and research methods for soil desiccation cracking: a review. European Journal of Environmental and Civil Engineering, 27(10), 3091-3115.

[6] Jabbarzadeh, M., Sadeghi, H., Tourchi, S., & Darzi, A. G. 2024. Thermo-hydraulic analysis of desiccation cracked soil strata considering ground temperature and moisture dynamics under the influence of soil-atmosphere interactions. Geomechanics for Energy and the Environment, 38, 100558.
[7] Ozsoy, A., & Yildirim, R. 2018. The performance of ground source heat pipes at low constant source temperatures. International journal of green energy, 15(11), 641-650.
[8] Ozsoy, A., & Yildirim, R. 2016. Prevention of icing with ground source heat pipe: A theoretical analysis for Turkey's climatic conditions. Cold Regions Science and Technology, 125, 65-71.
[9] Sungur, C., & Altun, A. A. 2010. Konya Bölgesindeki Don Olaylarına Karşı Mistleme Sisteminin Yapay Sinir Ağları İle Modellenmesi. Selcuk Journal of Agriculture & Food Sciences/Selcuk Tarim ve Gida Bilimleri Dergisi, 24(4).
[10] Osmanlı, N., & Karakayacı, Ö. 2023. Ekonomik Coğrafya Odağında Kırsalı Yeniden Düşünmek: Konya Örneği. Konya Sanat, (6), 195-215.
[11] Orhan, O., Haghshenas Haghighi, M., Demir, V., Gökkaya, E., Gutiérrez, F., & Al-Halbouni, D. (2023). Spatial and Temporal Patterns of Land Subsidence and Sinkhole Occurrence in the Konya Endorheic Basin, Turkey. Geosciences, 14(1), 5.
[12] Sarı, F., & Yalcin, M. 2023. Maximum entropy model-based spatial sinkhole occurrence prediction in Karapınar, Turkey. Kuwait Journal of Science, 50(1B).
[13] Bayrak, S., & Şeflek, A. Y. 2023. General Characteristics of Konya Agricultural Machinery Manufacturing Industry. Selcuk Journal of Agriculture and Food Sciences, 37(3), 531-545.
[14] Fan, Y., Wang, X., Funk, T., Rashid, I., Herman, B., Bompoti, N., ... & Li, B. 2022. A critical review for real-time continuous soil monitoring: Advantages, challenges, and perspectives. Environmental Science & Technology, 56(19), 13546-13564.
[15] Basso, B., & Liu, L. 2019. Seasonal crop yield forecast: Methods, applications, and accuracies. advances in agronomy, 154, 201-255.
[16] Sharma, P. K., & Kumar, S. 2023. Soil Temperature and Plant Growth. In Soil Physical Environment and Plant Growth: Evaluation and Management (pp. 175-204). Cham: Springer International Publishing.
[17] Costa, C. A., Guiné, R. P., Costa, D. V., Correia, H. E., & Nave, A. 2023. Pest control in organic farming. In Advances in Resting-state Functional MRI (pp. 111-179). Woodhead Publishing.
[18] Araújo, S. O., Peres, R. S., Ramalho, J. C., Lidon, F., & Barata, J. 2023. Machine learning applications in agriculture: current trends, challenges, and future perspectives. Agronomy, 13(12), 2976.
[19] Folorunso, O., Ojo, O., Busari, M., Adebayo, M., Joshua, A., Folorunso, D., ... & Olabanjo, O. 2023. Exploring machine learning models for soil nutrient properties prediction: A systematic review. Big Data and Cognitive Computing, 7(2), 113.
[20] Bilgili, M., Şimşek, E., & Şahin, B. 2010. Ege Bölgesindeki Toprak Sıcaklıklarının Yapay Sinir Ağları Yöntemi İle Belirlenmesi. Isı Bilimi ve Tekniği Dergisi, 30(1), 121-132.
[21] Ozturk, M., Salman, O., & Koc, M. 2011. Artificial neural network model for estimating the soil temperature. Canadian journal of soil science, 91(4), 551-562.
[22] Bilgili, M. 2011. "The use of artificial neural networks for forecasting the monthly mean soil temperatures in Adana, Turkey," Turkish Journal of Agriculture and Forestry: Vol. 35: No. 1, Article 9.
[23] Citakoglu, H. 2017. Comparison of artificial intelligence techniques for prediction of soil temperatures in Turkey. Theoretical and Applied Climatology, 130, 545-556.
[24] Behmanesh, J., & Mehdizadeh, S. 2017. Estimation of soil temperature using gene expression programming and artificial neural networks in a semiarid region. Environmental Earth Sciences, 76, 1-15.
[25] Kara, F., & Cemek, B. 2019. Orta Karadeniz Bölgesi illerine ait toprak sıcaklıklarının yapay sinir ağları yöntemi ile tahmin edilmesi. Derim, 36(2), 192-198.
[26] Feng, Y., Cui, N., Hao, W., Gao, L., & Gong, D. 2019. Estimation of soil temperature from meteorological data using different machine learning models. Geoderma, 338, 67-77.
[27] Pekel, E. 2020. Estimation of soil moisture using decision tree regression. Theoretical and Applied Climatology, 139(3), 1111-1119.
[28] Alizamir, M., Kisi, O., Ahmed, A. N., Mert, C., Fai, C. M., Kim, S., ... & El-Shafie, A. 2020. Advanced machine learning model for better prediction accuracy of soil temperature at different depths. PLoS One, 15(4), e0231055.
[29] Fathololoumi, S., Vaezi, A. R., Alavipanah, S. K., Montzka, C., Ghorbani, A., & Biswas, A. 2020. Soil temperature modeling using machine learning techniques. Desert, 25(2), 185-199.
[30] Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. 2021. Machine learning in agriculture: A comprehensive updated review. Sensors, 21(11), 3758.
[31] Guleryuz, D. (2022). Estimation of soil temperatures with machine learning algorithms—Giresun and Bayburt stations in Turkey. Theoretical and Applied Climatology, 147(1), 109-125.
[32] Bilgili, M., Ünal, Ş., Şekertekin, A., & Gürlek, C. 2023. Machine learning approaches for one-day ahead soil temperature forecasting. Journal of Agricultural Sciences, 29(1), 221-238.

[33] Elmi, A., Anwar, S. A., & Al-Dashti, H. 2024. Measurements and Regional Climate Modeling of Soil Temperature with and Without Bias Correction Method Under Arid Environment: Can Soil Temperature Outperform Air Temperature as a Climate Change Indicator? Environmental Modeling & Assessment, 29(2), 279-289.
[34] Turkish State Meteorological Service (TSMS), Ankara, Turkey
[35] Pence, I., Kumas, K., Siseci Cesmeli, M., & Akyüz, A. 2024. Detailed analysis of Türkiye's agricultural biomass-based energy potential with machine learning algorithms based on environmental and climatic conditions. Clean Technologies and Environmental Policy, 1-19.
[36] Acito, F. 2023. k Nearest Neighbors. In Predictive Analytics with KNIME: Analytics for Citizen Data Scientists (pp. 209-227). Cham: Springer Nature Switzerland.
[37] Pence, I., Kumaş, K., Cesmeli, M. S., & Akyüz, A. 2024. Future prediction of biogas potential and CH_4 emission with boosting algorithms: the case of cattle, small ruminant, and poultry manure from Turkey. Environmental Science and Pollution Research, 31(16), 24461-24479.
[38] Pence, I., Yıldırım, R., Cesmeli, M. S., Güngör, A., Akyüz, A. 2023. Evaluation of machine learning approaches for estimating thermodynamic properties of new generation refrigerant R513A. Sustainable Energy Technologies and Assessments, 55, 102973.
[39] Guillen, M. D., Aparicio, J., & Esteve, M. 2023. Gradient tree boosting and the estimation of production frontiers. Expert Systems with Applications, 214, 119134.
[40] Akyüz, A. 2024. Effect of Temperature on Ultrasonic Degradation of Sodium Poly (Styrene Sulfonate): Analysis of Online Viscometric Data with Theoretical Models and Machine Learning Approaches. Journal of Macromolecular Science, Part B, 1-25.
[41] Renaud, J., Karam, R., Salomon, M., & Couturier, R. 2023. Deep learning and gradient boosting for urban environmental noise monitoring in smart cities. Expert Systems with Applications, 218, 119568.