International Journal of Engineering and Geosciences
https://dergipark.org.tr/en/pub/ijeg
e-ISSN 2548-0960

Machine learning assisted prediction of land surface temperature (LST) based on major air pollutants over the Annamayya District of India

Jagadish Kumar Mogaraju *[1]

[1] *International Union for Conservation of Nature Commission on Ecosystem Management, Agroecosystems, India, jagadishmogaraju@gmail.com*

Cite this study: Mogaraju, J. K. (2024). Machine learning assisted prediction of land surface temperature (LST) based on major air pollutants over the Annamayya District of India. International Journal of Engineering and Geosciences, 9 (2), 233-246

https://doi.org/10.26833/ijeg.1394111

Keywords
Machine learning
Geographic information systems
Sentinel-5-P
MODIS
Land surface temperature

Research Article

Received: 21.11.2023
Revised: 15.05.2024
Accepted: 04.06.2024
Published:24.07.2024

Abstract

Remote sensing (RS), Geographic information systems (GIS), and Machine learning can be integrated to predict land surface temperatures (LST) based on the data related to carbon monoxide (CO), Formaldehyde (HCHO), Nitrogen dioxide (NO_2), Sulphur dioxide (SO_2), absorbing aerosol index (AAI), and Aerosol optical depth (AOD). In this study, LST was predicted using machine learning classifiers, i.e., Extra trees classifier (ET), Logistic regressors (LR), and Random Forests (RF). The accuracy of the LR classifier (0.89 or 89%) is higher than ET (82%) and RF (82%) classifiers. Evaluation metrics for each classifier are presented in the form of accuracy, Area under the curve (AUC), Recall, Precision, F1 score, Kappa, and MCC (Matthew's correlation coefficient). Based on the relative performance of the ML classifiers, it was concluded that the LR classifier performed better. Geographic information systems and RS tools were used to extract the data across spatial and temporal scales (2019 to 2022). In order to evaluate the model graphically, ROC (Receiver operating characteristic) curve, Confusion matrix, Validation curve, Classification report, Feature importance plot, and t- SNE (t-distributed stochastic neighbour embedding) plot were used. On validation of each ML classifier, it was observed that the RF classifier returned model complexity due to limited data availability and other factors yet to be studied post data availability. Sentinel-5-P and MODIS data are used in this study.

1. Introduction

The land surface temperature (LST) can be considered an essential indicator in assessing the environmental conditions that might exist at a place at any point in time [1]. Satellite-derived products have helped us monitor the LST dynamics on a large scale for three decades, and more data is being added by the day, helping many researchers yield insights [2]. The urban centers are in focus since most of the population ingress is apparent there, causing congestion in the living space inside the regulatory frames of reference [3]. LST can be classified under critical geophysical parameters as it is dynamic across all the land-water interfaces we know [4]. It is a continuous task to monitor LST across spatial and temporal scales, which can assist governments in framing specific rules in extreme heat situations [5]. Aerosols suspended in the lower atmosphere can affect LST fluctuations and may induce a negative feedback mechanism on surface temperature that can influence the inhabitants [6]. The landscape dynamics initiated due to population ingress and unprecedented urban sprawl can alter local microclimate in terms of LST and Aerosol optical depth (AOD) [7]. The seasonal temperatures were lowered across the Indian subcontinent due to increased AOD levels [8]. Carbon monoxide (CO) emission can be anthropogenic, and its release into the lower atmosphere before an earthquake in 2001 was observed, reflecting that natural CO emission is possible [9]. The combined effect of forest fires and formaldehyde (HCHO) on surface temperature was observed using satellite-derived observations [10]. The fluctuation in HCHO concentrations in the troposphere was associated with the biogenic volatile organic compounds (BVOC) that were released by plants in response to abiotic stress (LST) [11]. The isoprene emissions released by the abiotically stressed plants were tracked using formaldehyde variations, and these emissions can affect radiative forcing on a large scale [12]. The surface temperatures are sensitive to the Nitrogen dioxide (NO_2) levels, and lightning-induced reactions can also favor NO_2 generation [13, 14]. The sulfur dioxide (SO_2) concentrations affect the radiative forcing, surface temperature, and climate change on a global scale [15,

16]. Plants cannot tolerate surface temperatures greater than 35°C with few exceptions and will release compounds that we can interpret as abiotic stress signals [17–23].

Machine learning (ML) algorithms were used in land cover and land use (LULC) dynamics in relation to surface temperature fluctuations assisted by remote sensing instruments. The Artificial Neural Networks (ANN) and Cellular Automata (CA) supported by Support Vector Machine (SVM) [24] were used to understand LST dynamics over urban landscapes [18, 25]. Soil Moisture and Ocean Salinity (SMOS) derived soil moisture, Moderate Resolution Image spectrometer (MODIS) for LST, along with generalized linear models (GLM) of Artificial Intelligence (AI) framework was used in hydrological applications [26]. The urban form metrics, i.e., building density (BD) and floor area ratio (FAR), were associated with LST fluctuations with the aid of machine learning [27]. The LST products of MODIS were successfully downscaled using ML frameworks, enabling high spatial and temporal resolutions for use in diverse applications [28]. Deep learning and neural networks (DL-NN) were used to optimize LST retrieval (MODIS) methods, and other geophysical parameters were also obtained easily [29]. Convolutional neural networks (CNN) of DL were used to get the LST values from Advanced microwave scanning radiometer 2 (AMSR2), and long-term LST disparities were easily studied [30]. XGB (Extra Gradient Boosting) regression and ANN-CA were used to extract urban heat islands using LST and LULC data [31]. Land use/Land cover variations and population can change LST [32]. Changes in indices such as normalized difference vegetation index (NDVI), normalized difference bareness index (NDBaI), normalized difference built-up index (NDBI), and modified normalized difference water index (MNDWI) can influence LST [33]. The effect of seasons in estimating LST was estimated using Landsat 8 data, and the Thermal sharpening (TsHARP) algorithm was also used [34]. The tendency of interrelationships between MNDWI, NDBaI, NDBI, and NDVI was apparent in the winter season than summer [35]. Fallow lands and settlements can enhance LST rate but water bodies, agricultural land and forests can reduce LST rate [36].

This work is done to possibly fill the research gap in predicting LST using major air pollutants, i.e., CO, HCHO, NO_2, SO_2, and other parameters like Aerosol optical depth (AOD) and Aerosol absorption index (AAI) using machine learning algorithms.

2. Method

2.1. Data

The datasets needed for this study are obtained from Earth Explorer (USGS) and Giovanni (NASA) websites [37, 38]. The required data is collected in the .csv and GeoTIFF formats so that the entire procedure involving processing, interpretation, and analysis can be done using in-house equipment. The satellite data is obtained from Sentinel-5- P (TROPOMI) [39] and MODIS platforms [40]. AAI, CO, HCHO, NO_2, and SO_2-related data is obtained from Sentinel-5- P, whereas Land surface temperature data is obtained from MODIS. AOD (500nm) related datasets are retrieved from MCD19A2.006: Terra & Aqua MAIAC. The spatial distribution maps of pollutants CO, HCHO, NO_2, SO_2, AAI, AOD, and LST were prepared to know the anomalies on a spatial scale. The LST values greater than 35°C are labeled as 'Hot,' and those less than 35°C are labeled as 'Normal.' These two classes were considered essential classes, and the datasets were passed onto the machine learning frameworks. Three ML algorithms, i.e., Extra Trees classifier (ET) [41, 42], Logistic regression (LR) [43, 44], and Random Forest (RF) [43], were used in this study as they performed better than others.

2.2. Study area

The study area selected for this study is Annamayya district, located in the southern part of Andhra Pradesh state, India [37], as shown in Figure 1. After reorganizing the Cuddapah and Chittoor districts of Andhra Pradesh, it is a newly formed district. The study area is about 7951 square kilometers and extends from 78° 18' 55" and 79° 20'26" E longitude and 13° 19' 55" and 14° 42' 32" of N latitude. Rajampet, Rayachoti, and Madanapalle are the main urban centers of the district, and most of the business activities are witnessed here [45]. The effects of the selected pollutants on the land surface temperature were studied for the years 2019, 2020,2021, and 2022. The time-averaged maps (annual) are prepared to know the spatial and temporal spread of the effects produced due to selected pollutants, and their effect on time-averaged LST is studied. The detailed methodology is given in Figure 2.

2.3. Interpretation of metrics and outputs

The evaluation metrics used in knowing the performance of the models are Accuracy, AUC (Area under the curve), Recall, Precision, F1 score (F-measure), Kappa, and MCC (Matthews correlation coefficient) values [46, 47]. After scrutinizing the evaluation metrics, accuracy values were given more weightage than other metrics. The results for each ML model are given through ROC (Receiver operating characteristic) curves, Confusion matrix, Validation curve, Classification report, Feature importance plot, and t-SNE (t-distributed stochastic neighbor embedding) manifold plots [48, 49]. The ROC curve illustrates model performance in a binary classification system with varied discrimination threshold or limit. It is generated using TPR or true positive rate and FPR or false positive rate. The primary advantage of an ROC curve is that there is no need for threshold optimization for each label or class. A confusion matrix can be employed to evaluate the classifier product quality. The diagonal values show us the points for which the label predicted = actual label. The off-diagonal values show the classifier's prediction error. If the diagonal values are higher than the off-diagonal values, then we can assume that there are more correct predictions. The F_1 score can be considered a harmonic mean of recall and precision (1-best, 0- worst). The kappa statistic vibrates between -1 (chance agreement) and +1 (complete agreement).

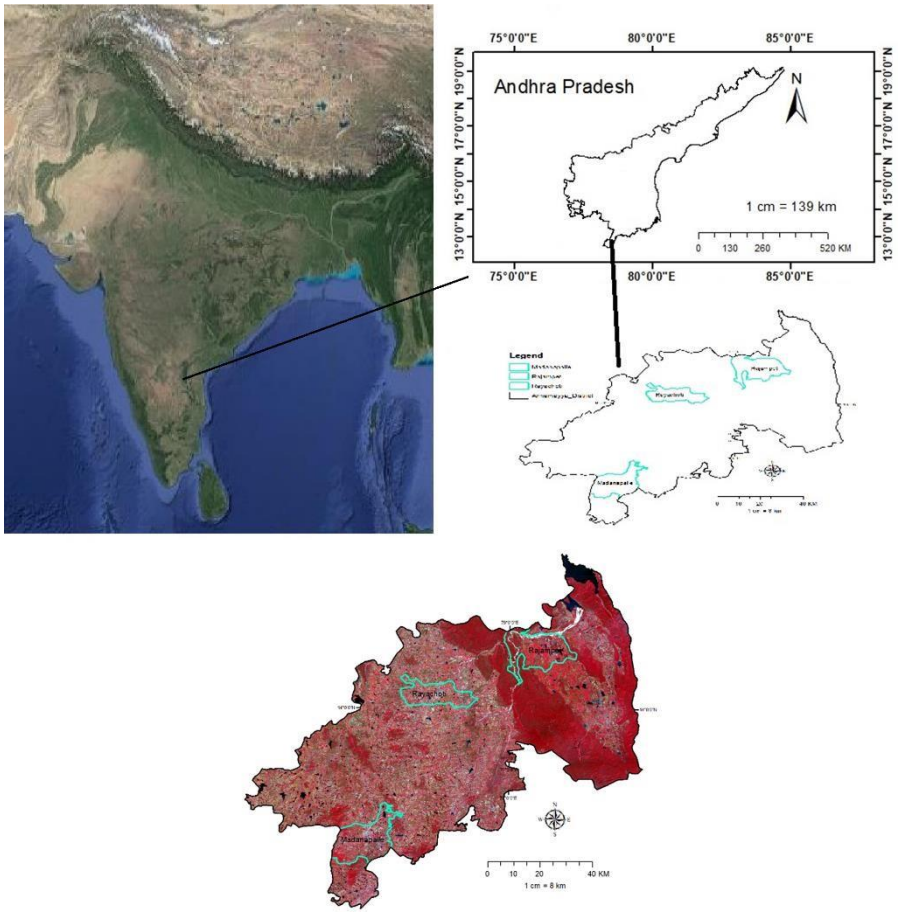


Figure 1. Study area with False colour composite.

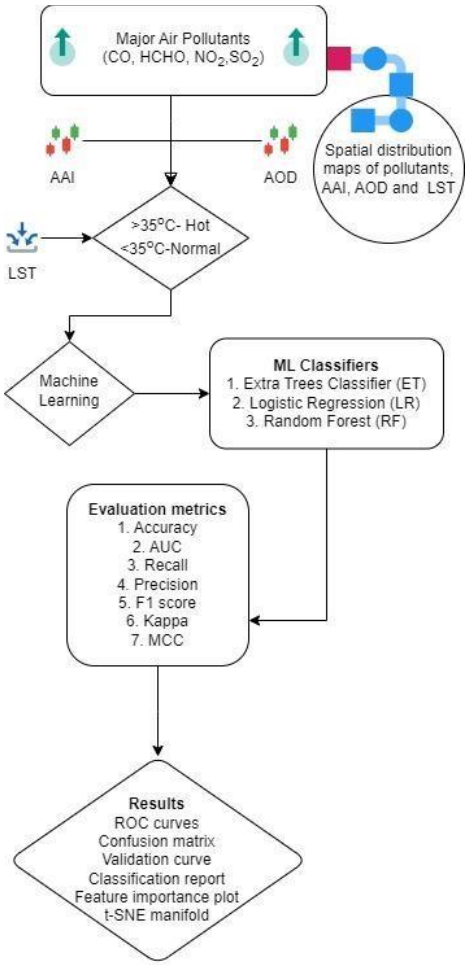


Figure 2. Methodology.

MCC score reflects the quality of the binary classification in this case. The primary advantage of the MCC metric is that it is a balanced measure and can be employed for classes of varied sizes. The MCC (phi coefficient) values fluctuate between -1 (inverse prediction), 0 (random prediction), and +1 (Perfect prediction). The validation and training score can explain the model's performance, i.e., underfitting, overfitting, and fit (well fitted). If the validation and training scores are low, we can assume that the estimator/model will be underfitting and, if high, overfitting or a well-fitted estimator. The classification report shows all the metrics, such as precision, recall, F1, and support (occurrence of each class in test data). The depth or relative rank of a specific feature can be used to evaluate the importance of that feature in relation to the predictability of the desired variable (target), and this can be shown in the form of a feature importance plot. t-SNE can be used to visualize the data (high-dimension) by transforming the similarities between the data points to cumulative probabilities, and it also lowers the Kullback-Leibler divergence.

2.4 Methodology

The data pertaining to the major air pollutants were subjected to preprocessing. Random points were selected for all the variables. In preprocessing, we removed outliers, eliminated perfect collinearity, and applied normalization and transformation. The independent variables are CO, HCHO, NO_2, SO_2, AAI, AOD, and LST alone is a dependent variable. The dataset was

subjected to a machine learning framework. ML classifiers were used to predict LST (at two levels), and three were selected among several classifiers. The evaluation metrics like MCC, Kappa, F_1 score, Precision, Recall, AUC, and accuracy were used to gauge prediction capability. The outputs were for selected ML classifiers. GIS maps were prepared for each variable to express their spatial distribution.

3. Results

3.1. GIS

The spatial distribution of the Absorbing aerosol index (AAI), i.e., the prevalence of aerosols, is given in Figure 3 (a-d). The Aerosol index can be considered as a qualitative index of the presence of aerosols generated from dust and biomass burning (plumes from volcanic ash are ignored as the study area is devoid of volcanic activity) [50]. The wavelength variations of Rayleigh scattering (UV) are minimal due to limited ozone absorption [51]. Top Of atmosphere (TOA) reflectance and Rayleigh scattering reflectance result in residual values where the positive values represent aerosols (UV-absorbing) [52]. The NE corners of the study area showed less aerosol index (-1.33844 in 2019) (-1.6613 in 2020) (-1.40316 in 2021) (-0.705651 in 2022) and high in E and central portions (-0.986181 in 2019) (-1.37002 in 2020) (-1.06346 in 2021) (-0.294711 in 2022). The spatial distribution of Aerosol optical depth (AOD) is given in Figure 4 (a-d). The AOD is low (593.324 in 2019) (508.243 in 2020) (795.5 in 2021) (791.5 in 2022) in the NE, S, and W and high in the N and central portions (68 in 2019) (250.053 in 2020) (183.5 in 2021) (68 in 2022). The spatial distribution of carbon monoxide is given in Figure 5 (a-d). Except for S and some portions in the NE (0.0328049 in 2019) (0.0333505 in 2020) (0.0317095 in 2021) (0.0310688) of the study area, most of the region exhibited high levels (0.0378867 in 2019) (0.0384048 in 2020) (0.0368594 in 2021) (0.0362371) of CO (mol/m^2). The spatial distribution of formaldehyde (HCHO) is given in Figure 6 (a-d). The HCHO (mol/m^2) concentration is higher in the NE portions (0.00018135 in 2019) (0.000181125 in 2020) (0.000175397 in 2021) (0.000180444 in 2022) and low in the S and E portions (0.00011805 in 2019) (0.000107572 in 2020) (0.00010903 in 2021) (0.00010737) of the study area. The spatial distribution of nitrogen dioxide NO_2 (mol/m^2) is given in Figure 7 (a-d). The N and NE corners exhibited high NO_2 concentrations (2.68345e-005 in 2019) (2.31525 e^{-005} in 2020) (2.62399 e^{-005} in 2021) (2.62361 e^{-005} in 2022) than other regions (2.08934 e^{-005} in 2019) (1.43907 e-005 in 2020) (1.68165 e^{-005} in 2021) (1.71763 e^{-005} in 2022) of the study area. The spatial distribution of sulphur dioxide (SO_2) (mol/m^2) is given in Figure 8 (a-d). The SO_2 concentration was high in the NE (0.000151513 in 2019) (0.00134435 in 2020) (0.000171003 in 2021) (0.000177544 in 2022), extending towards S and low in the central and E (-3.01868^{e-005} in 2019) (-6.71495 e^{-005} in 2020) (-4.07475^{e-005} in 2021) (-1.24852e^{-005} in 2022) portions of the study area. The spatial distribution of land surface temperature (LST) (°C) is given in Figure 9 (a-d). The E and central portions (43.88 °C in 2019) (43.28 °C in 2020) (41.70 °C in 2021) (41.59 °C in 2022) exhibited high temperatures, and NE portions (25.87 °C in 2019) (25.60 °C in 2020) (25.02 °C in 2021) (25.05 °C in 2022) exhibited low LST in the study area.

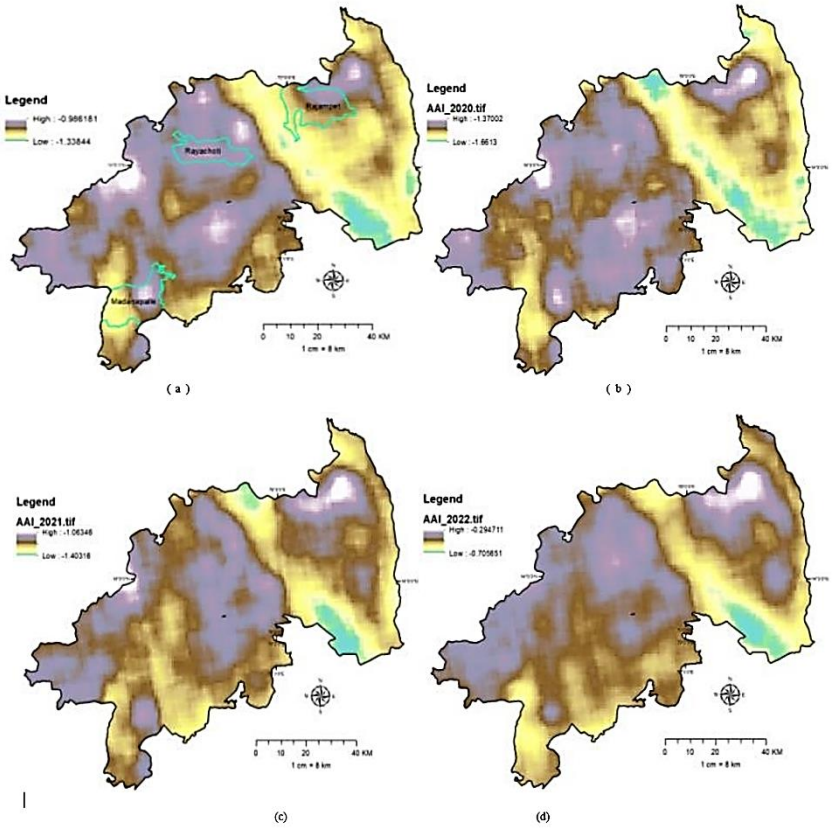


Figure 3. Absorbing Aerosol Index (AAI) (a) 2019, (b) 2020, (c) 2021), (d) 2022.

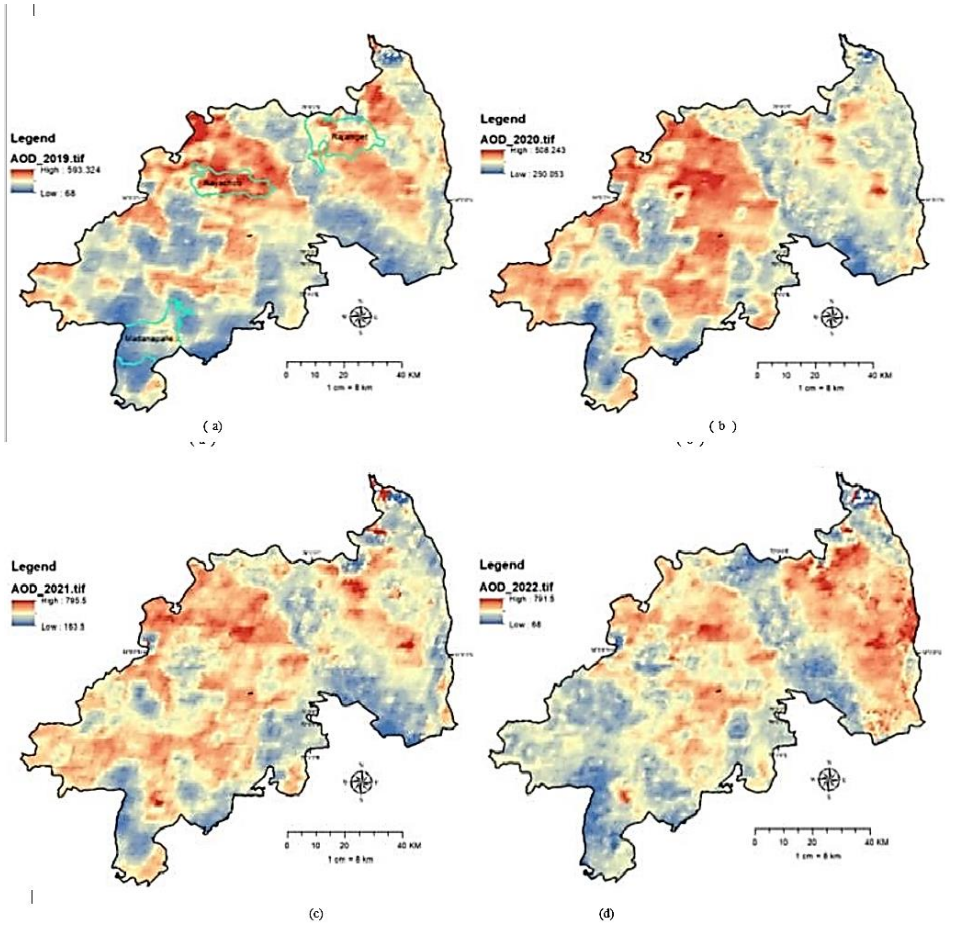


Figure 4. Aerosol optical depth (AOD) (a) 2019, (b) 2020, (c) 2021), (d) 2022.

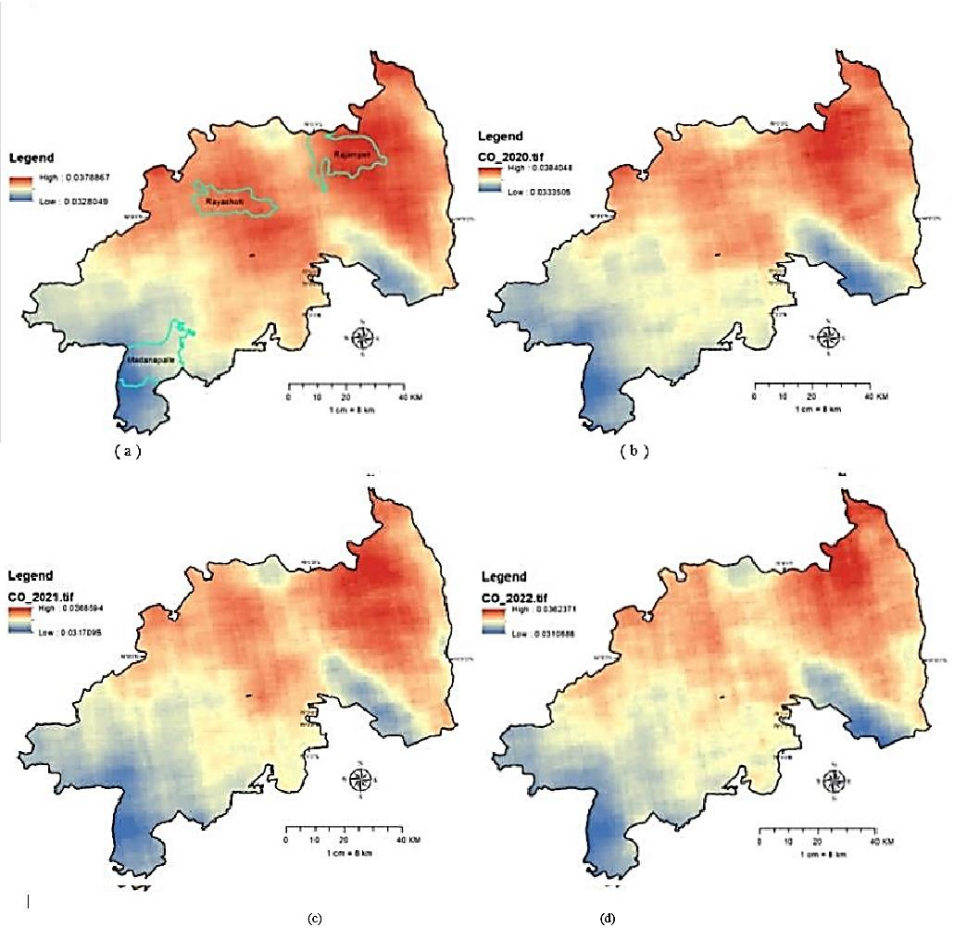


Figure 5. Carbon monoxide (CO) (mol/m^2) (a) 2019, (b) 2020, (c) 2021), (d) 2022.

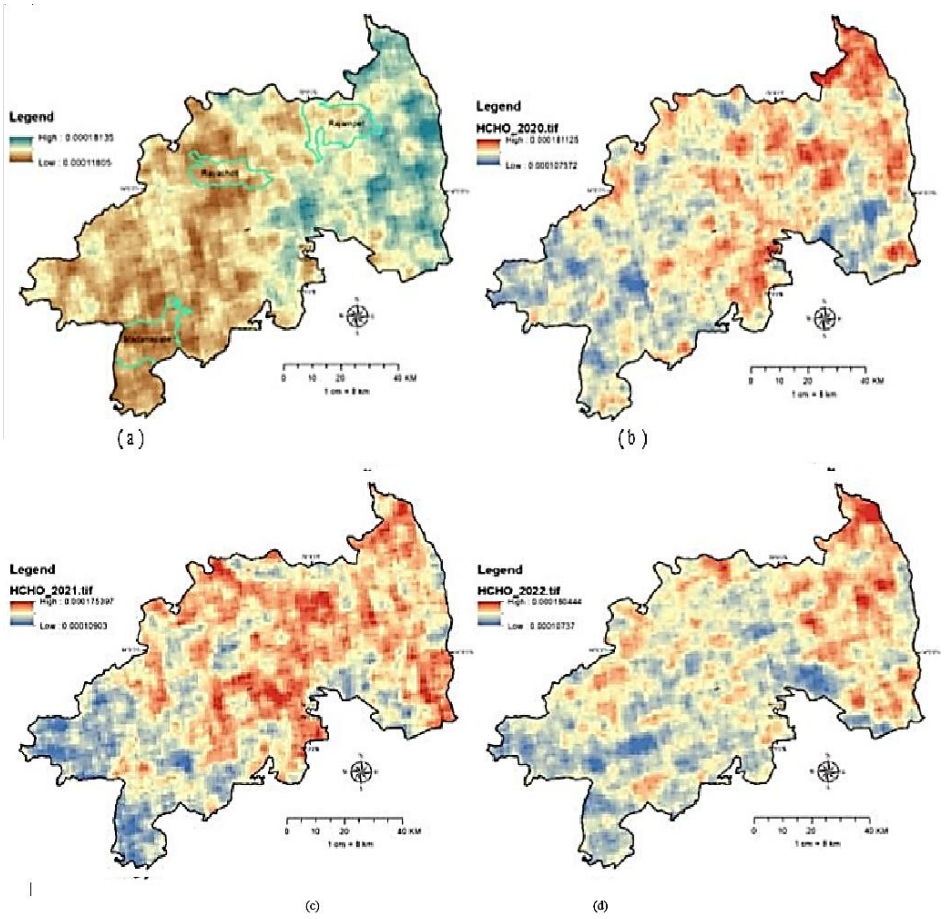


Figure 6. Formaldehyde (HCHO) (mol/m^2) (a) 2019, (b) 2020, (c) 2021), (d) 2022.

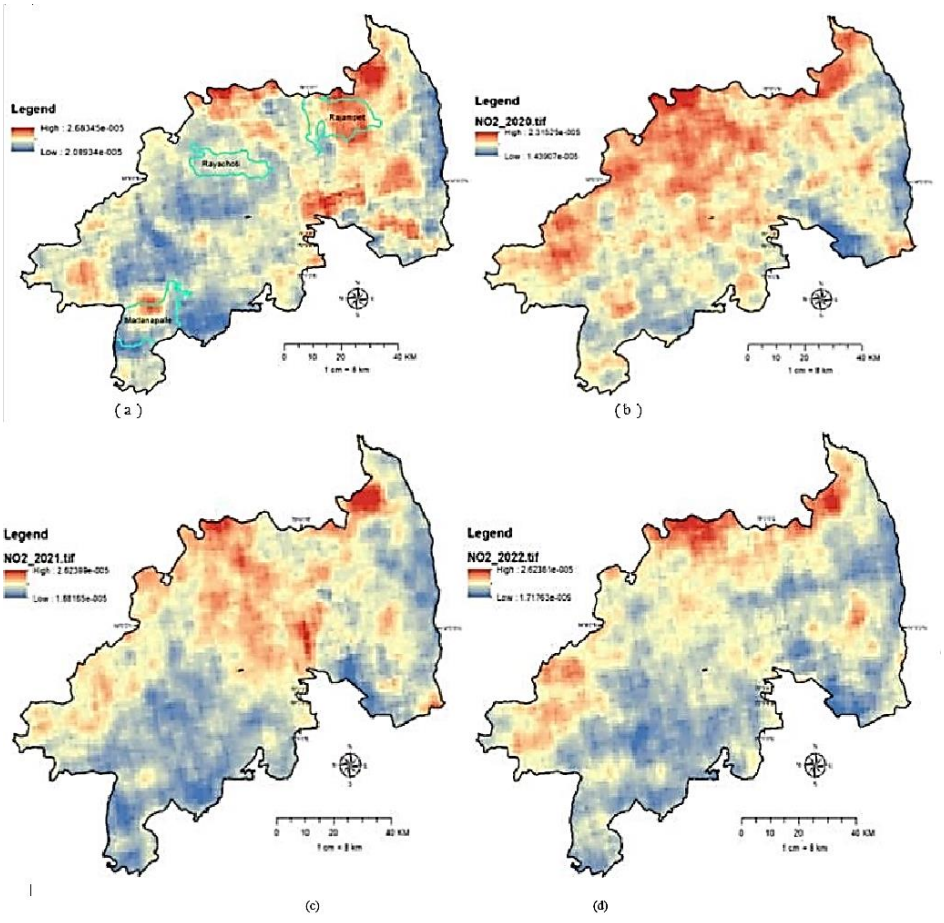


Figure 7. Nitrogen dioxide (NO_2) (mol/m^2) (a) 2019, (b) 2020, (c) 2021), (d) 2022.

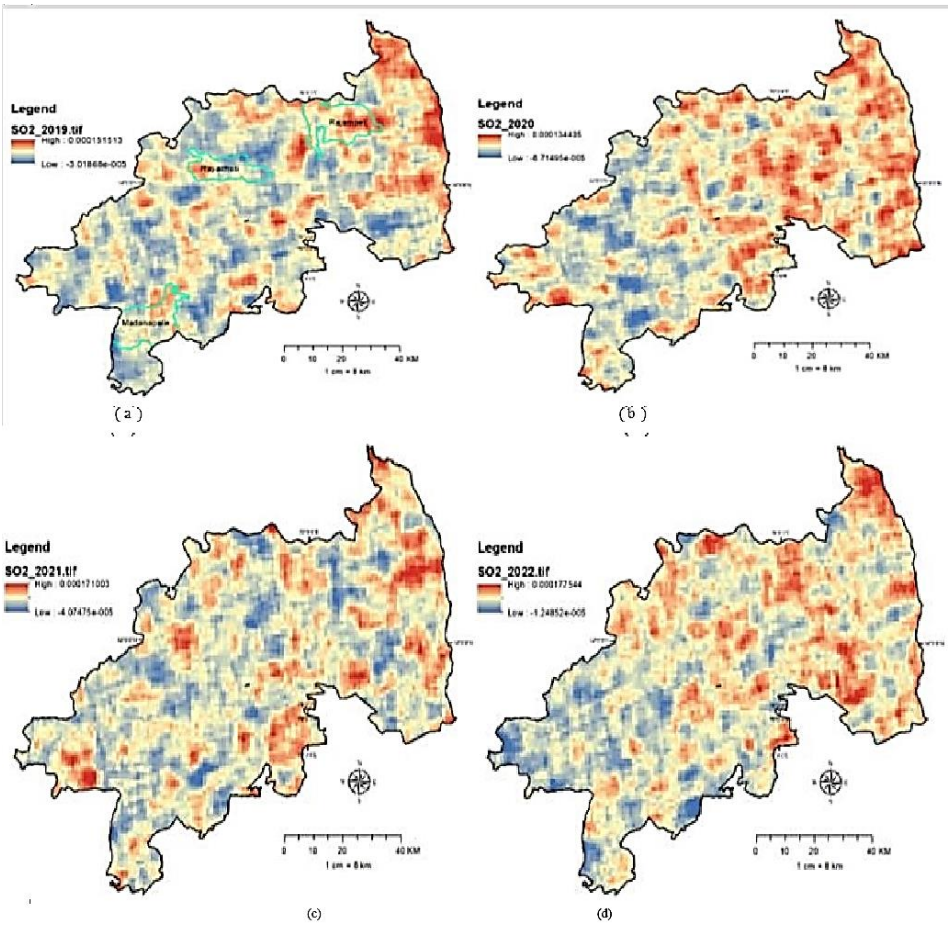


Figure 8. Sulfur dioxide (SO_2) (mol/m^2) (a) 2019, (b) 2020, (c) 2021), (d) 2022.

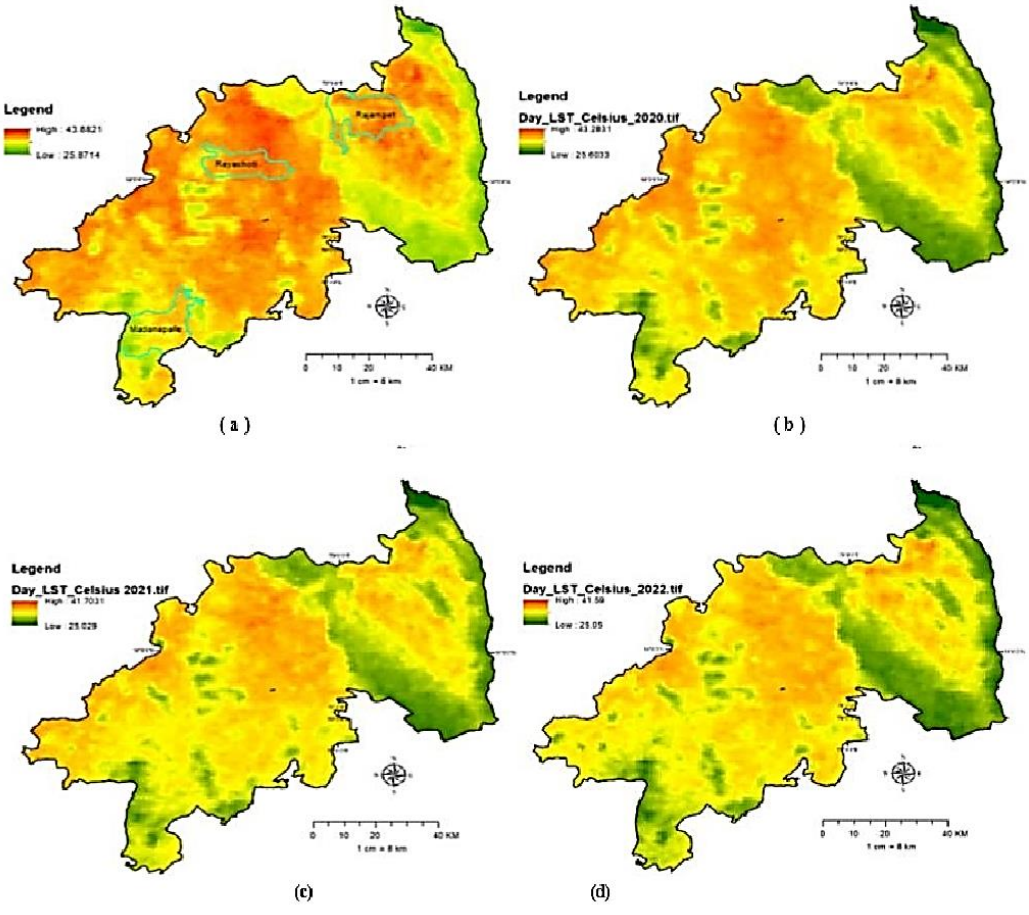


Figure 9. Land surface temperature (LST) (°C) (a) 2019, (b) 2020, (c) 2021), (d) 2022.

3.2. Machine learning

The ROC curve of the Extra Trees classifier (Figure 10a) shows AUC of 0.89 for both the 'Hot' and 'Normal classes.

The micro-average ROC curve reflects AUC of 0.91, and the macro-average curve shows 0.90. The AUC values show that the classifier predicted the classes well (0.89). The confusion matrix is given in Figure 10b. The training curve reached a high level at maximum depths of 6 to 10, whereas the cross-validation curve reached a high level at maximum depths of 4, 7, and 8. This shows that neither of the curves reached a high score quickly (relatively), and hence, through the validation curve given in Figure 10c, it is apparent that model performance is appropriate. The classification report containing some of the essential metrics is given in the Figure 10d. The feature importance plot given in Figure 10e shows that AAI and AOD have more influence on LST prediction. The t-SNE manifold plot shows minimal clustering and some negligible overlaps and is given in Figure 10f.

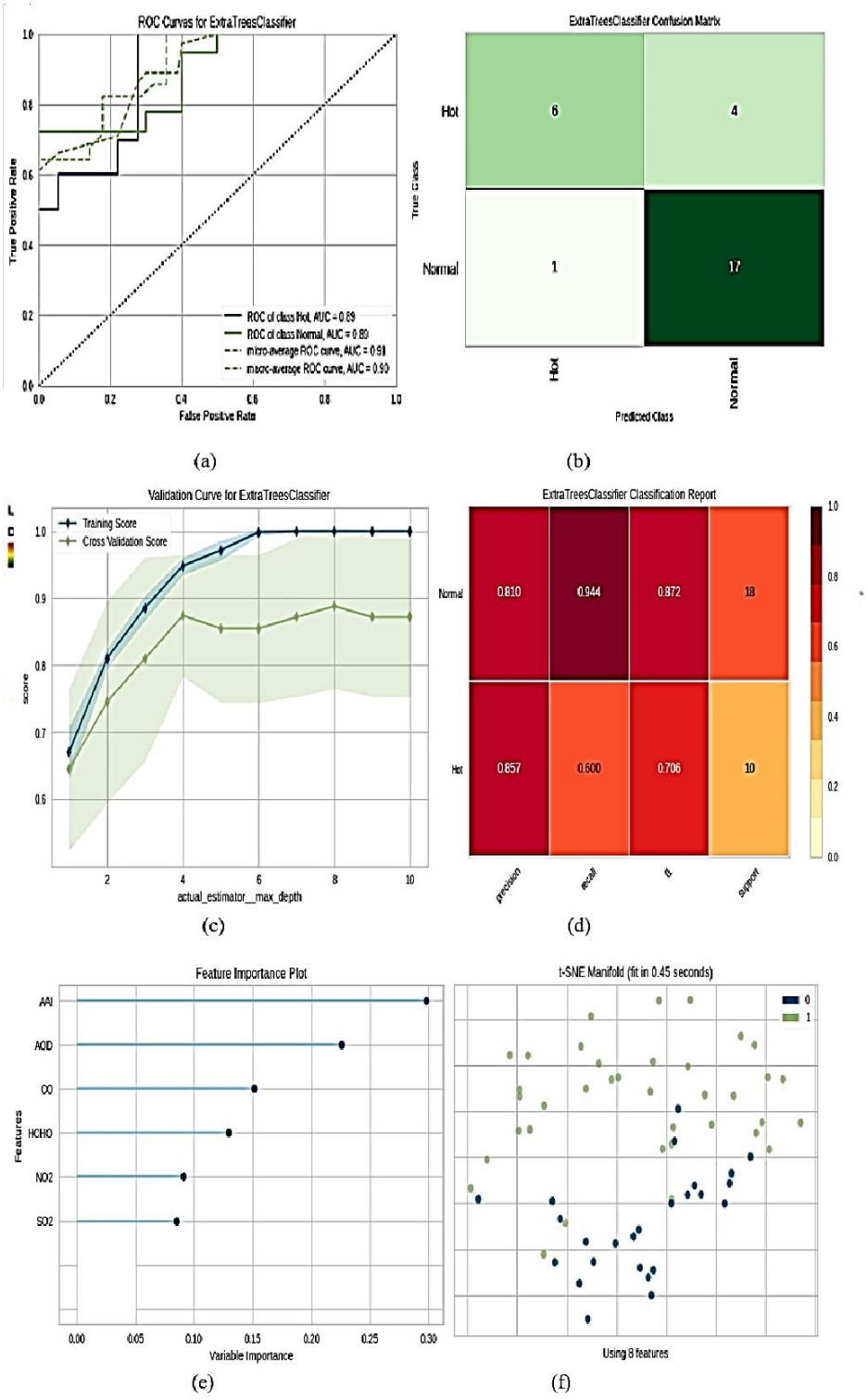


Figure 10. Extra Trees Classifier (ET) (a) ROC curve, (b) Confusion matrix, (c) Validation curve, (d) Classification report, (e) Feature importance plot, (f) t-SNE plot.

The ROC curve of the logistic regression classifier (Figure 11a) shows AUC of 0.96 for both the 'Hot' and 'Normal' classes. The micro-average ROC curve reflects AUC of 0.96, and the macro-average curve shows 0.97. The AUC values obtained show that the classifier performed better (0.96) in predicting the classes. The confusion matrix is given in Figure 11b. The training curve and cross-validation curve travelled steadily all along. This shows that neither of the curves reached a high score quickly (relatively) nor reached a low score immediately (Figure 11c). It is apparent that model performance is appropriate. The classification report containing some of the essential metrics is given in Figure 11d. The feature importance plot in Figure 11e shows that AAI has more influence on LST prediction. The t-SNE manifold plot shows a minimal clustering and is given in Figure 11f.

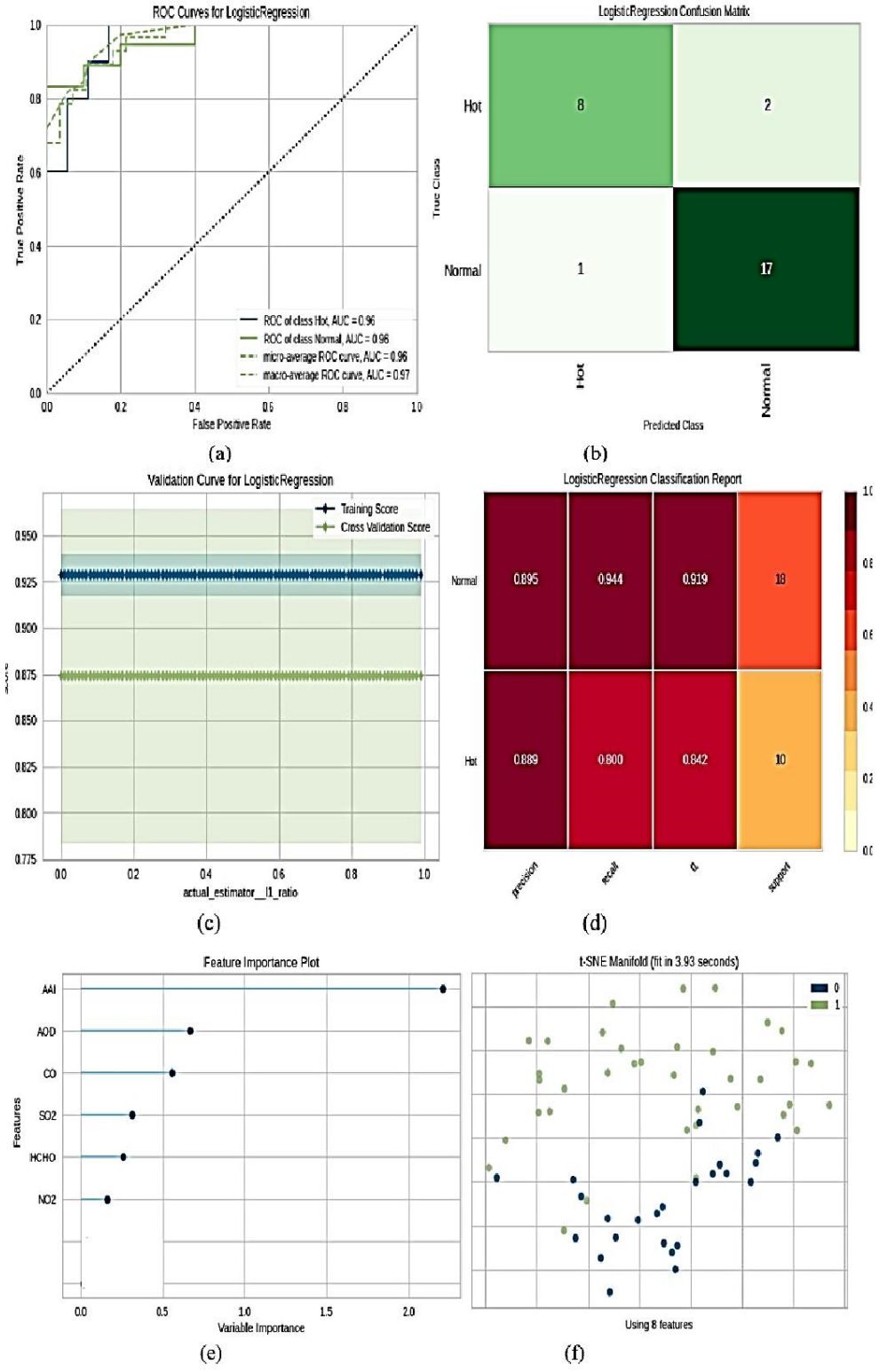


Figure 11. Logistic Regression (LR) (a) ROC curve, (b) Confusion matrix, (c) Validation curve, (d) Classification report, (e) Feature importance plot, (f) t-SNE plot.

The ROC curve of the Random Forest classifier (Figure 12a) shows AUC of 0.94 for both the 'Hot' and 'Normal' classes. The micro-average ROC curve reflects AUC of 0.93, and the macro-average curve shows 0.94. The AUC values obtained show that the classifier performed appropriately (0.94) in predicting the classes. The confusion matrix is given in Figure 12b. The training curve reached a high level at a maximum depth of 5, whereas the cross-validation curve reached a high level at a maximum depth of 2. This shows that the cross-validation curve reached a high score quickly (relatively) (Figure 12c). It is apparent that model complexity is present. This can be due to limited data available for analysis or other factors. The classification report containing some of the essential metrics is given in Figure 12d. The feature importance plot in Figure 12e shows that AAI and AOD have more influence on LST prediction. The t-SNE manifold plot shows minimal clustering and negligible overlaps and is given in Figure 12f. The comparison of evaluation metrics for each model is given in Figure 13. The logistic regression classifier performed well when accuracy was considered. Even with other evaluation metrics taken into consideration, the logistic regression classifier performed well, i.e., AUC (0.9611), Recall (0.9444), Precision (0.08947), F1 score (0.9189), kappa (0.7614), and MCC score (0.7638).

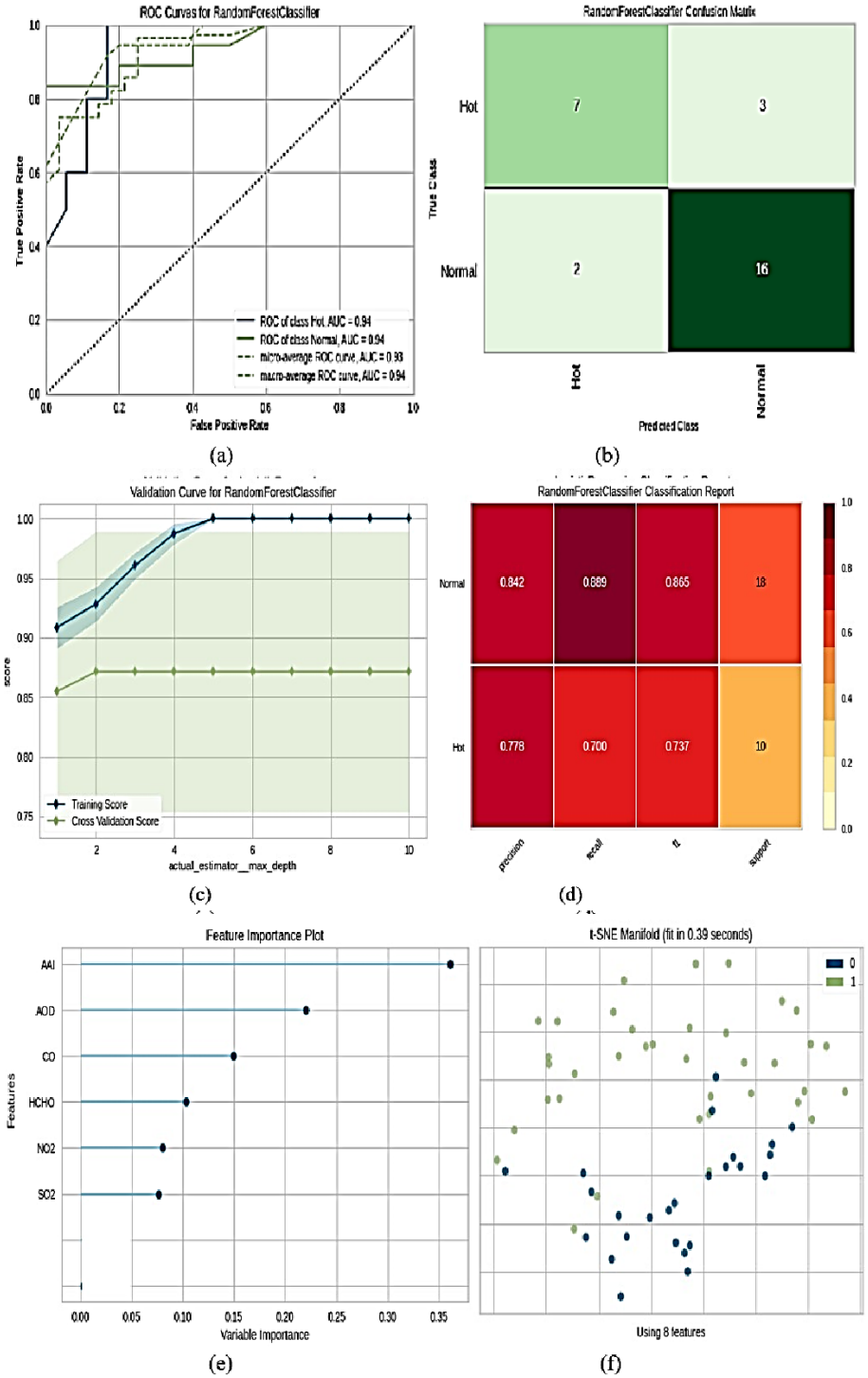
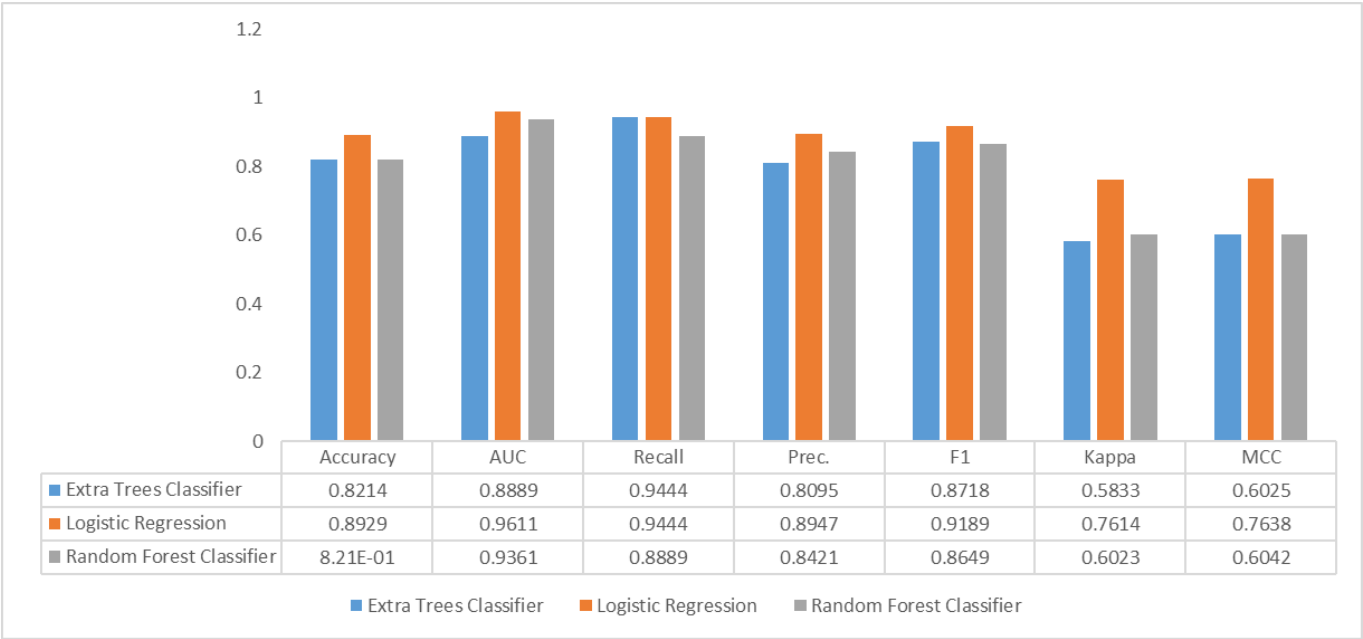


Figure 12. Random Forest Classifier (RF) (a) ROC curve, (b) Confusion matrix, (c) Validation curve, (d) Classification report, (e) Feature importance plot, (f) t-SNE plot.



	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
Extra Trees Classifier	0.8214	0.8889	0.9444	0.8095	0.8718	0.5833	0.6025
Logistic Regression	0.8929	0.9611	0.9444	0.8947	0.9189	0.7614	0.7638
Random Forest Classifier	8.21E-01	0.9361	0.8889	0.8421	0.8649	0.6023	0.6042

Figure 13. Model comparison (ET, LR, RF).

4. Discussion

Machine learning tools can be used with ease, and this can be due to their reliability and cost-reducing functionalities [53, 54]. Artificial Intelligence includes machine learning tools that deal with structured and unstructured data and help researchers across various domains achieve appropriate results [55, 56]. This study used AI (ML) tools to predict land surface temperatures based on air quality parameters like CO, HCHO, NO_2, SO_2, AAI, and AOD. Prior to ML treatment, the data is derived from remote sensing and GIS platforms. It was observed that Machine learning tools could predict land surface temperatures effectively and within low-cost frameworks. The main limitations of this study are: 1. Limited availability of ground data, 2. Lack of air pollution sensors in the study area, 3. Excess reliance on satellite-based data, 4. Problems with cross-validation (satellite versus ground data). These problems might be solved by installing air pollution monitoring stations and training the personnel to use mobile monitoring stations.

5. Conclusion

This work aimed at the integration of geographic information systems (GIS) and Machine learning (ML) frameworks to predict land surface temperatures (LST) over the Annamayya district of India (Andhra Pradesh state). After systematic analysis and interpretation in this study, it is concluded that machine learning tools can easily predict land surface temperatures with reliable results. The prediction capabilities can be improved once more data across spatial and temporal scales is available.

Acknowledgement

The authors sincerely thank the Government of India, NASA, USGS, ISRO, Government of Andhra Pradesh, District administration of Annamayya district for helping in this research work through data supply.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Hulley, G. C., Ghent, D., Göttsche, F. M., Guillevic, P. C., Mildrexler, D. J., & Coll, C. (2019). Land surface temperature. Taking the Temperature of the Earth, 57-127. https://doi.org/10.1016/B978-0-12-814458-9.00003-4
2. Tran, D. X., Pla, F., Latorre-Carmona, P., Myint, S. W., Caetano, M., & Kieu, H. V. (2017). Characterizing the relationship between land use land cover change and land surface temperature. ISPRS Journal of Photogrammetry and Remote Sensing, 124, 119-132. https://doi.org/10.1016/j.isprsjprs.2017.01.001
3. Pal, S., & Ziaul, S. K. (2017). Detection of land use and land cover change and land surface temperature in English Bazar urban centre. The Egyptian Journal of Remote Sensing and Space Science, 20(1), 125-145. https://doi.org/10.1016/j.ejrs.2016.11.003
4. Li, Z. L., Wu, H., Duan, S. B., Zhao, W., Ren, H., Liu, X., ... & Zhou, C. (2023). Satellite remote sensing of global land surface temperature: Definition, methods, products, and applications. Reviews of Geophysics, 61(1). https://doi.org/10.1029/2022RG000777
5. Mumtaz, F., Tao, Y., de Leeuw, G., Zhao, L., Fan, C., Elnashar, A., ... & Wang, D. (2020). Modeling spatio-temporal land transformation and its associated impacts on land surface temperature (LST). Remote Sensing, 12(18), 2987. https://doi.org/10.3390/rs12182987
6. Walker, J. C., Hays, P. B., & Kasting, J. F. (1981). A negative feedback mechanism for the long-term stabilization of Earth's surface temperature. Journal of Geophysical Research: Oceans, 86(C10), 9776-9782. https://doi.org/10.1029/JC086iC10p09776
7. Xiang, Y., Ye, Y., Peng, C., Teng, M., & Zhou, Z. (2022). Seasonal variations for combined effects of landscape

metrics on land surface temperature (LST) and aerosol optical depth (AOD). Ecological Indicators, 138, 108810. https://doi.org/10.1016/j.ecolind.2022.108810

8. Roy, S. S. (2008). Impact of aerosol optical depth on seasonal temperatures in India: a spatio-temporal analysis. International Journal of Remote Sensing, 29(3), 727-740. https://doi.org/10.1080/01431160701352121
9. Singh, R. P., Kumar, J. S., Zlotnicki, J., & Kafatos, M. (2010). Satellite detection of carbon monoxide emission prior to the Gujarat earthquake of 26 January 2001. Applied Geochemistry, 25(4), 580-585. https://doi.org/10.1016/j.apgeochem.2010.01.014
10. Marbach, T., Beirle, S., Liu, C., Platt, U., & Wagner, T. (2008). Biomass burning emissions from satellite observations: synergistic use of formaldehyde (HCHO), fire counts, and surface temperature. In Remote Sensing of Fire: Science and Application, 7089, 131-140. https://doi.org/10.1117/12.793654
11. Morfopoulos, C., Müller, J. F., Stavrakou, T., Bauwens, M., De Smedt, I., Friedlingstein, P., ... & Regnier, P. (2022). Vegetation responses to climate extremes recorded by remotely sensed atmospheric formaldehyde. Global Change Biology, 28(5), 1809-1822. https://doi.org/10.1111/gcb.15880
12. Zheng, Y., Unger, N., Barkley, M. P., & Yue, X. (2015). Relationships between photosynthesis and formaldehyde as a probe of isoprene emission. Atmospheric Chemistry and Physics, 15(15), 8559-8576. https://doi.org/10.5194/acp-15-8559-2015
13. Ramanathan, V., Callis, L. B., & Boughner, R. E. (1976). Sensitivity of surface temperature and atmospheric temperature to perturbations in the stratospheric concentration of ozone and nitrogen dioxide. Journal of the Atmospheric Sciences, 33(6), 1092-1112. https://doi.org/10.1175/1520-0469(1976)033<1092:SOSTAA>2.0.CO;2
14. Schumann, U., & Huntrieser, H. (2007). The global lightning-induced nitrogen oxides source. Atmospheric Chemistry and Physics, 7(14), 3823-3907. https://doi.org/10.5194/acp-7-3823-2007
15. Conley, A. J., Westervelt, D. M., Lamarque, J. F., Fiore, A. M., Shindell, D., Correa, G., ... & Horowitz, L. W. (2018). Multimodel surface temperature responses to removal of US sulfur dioxide emissions. Journal of Geophysical Research: Atmospheres, 123(5), 2773-2796. https://doi.org/10.1002/2017JD027411
16. Ward, P. L. (2009). Sulfur dioxide initiates global climate change in four ways. Thin Solid Films, 517(11), 3188-3203. https://doi.org/10.1016/j.tsf.2009.01.005
17. Abidin, M. R., Nur, R., Mayzarah, E. M., & Umar, R. (2021). Estimating and monitoring the land surface temperature (LST) using Landsat OLI 8 TIRS. International Journal of Environment, Engineering and Education, 3(1), 17-24. https://doi.org/10.55151/ijeedu.v3i1.43
18. Kafy, A. A., Shuvo, R. M., Naim, M. N. H., Sikdar, M. S., Chowdhury, R. R., Islam, M. A., ... & Kona, M. A. (2021). Remote sensing approach to simulate the land use/land cover and seasonal land surface temperature change using machine learning algorithms in a fastest-growing megacity of Bangladesh. Remote Sensing Applications: Society and Environment, 21, 100463. https://doi.org/10.1016/j.rsase.2020.100463
19. Roberts, D. A., Dennison, P. E., Roth, K. L., Dudley, K., & Hulley, G. (2015). Relationships between dominant plant species, fractional cover and land surface temperature in a Mediterranean ecosystem. Remote Sensing of Environment, 167, 152-167. https://doi.org/10.1016/j.rse.2015.01.026
20. Sekertekin, A., Kutoglu, S. H., & Kaya, S. (2016). Evaluation of spatio-temporal variability in Land Surface Temperature: A case study of Zonguldak, Turkey. Environmental Monitoring and Assessment, 188, 1-15. https://doi.org/10.1007/s10661-015-5032-2
21. Wan, Z., Zhang, Y., Zhang, Q., & Li, Z. L. (2004). Quality assessment and validation of the MODIS global land surface temperature. International Journal of Remote Sensing, 25(1), 261-274. https://doi.org/10.1080/0143116031000116417
22. Zaitunah, A., Silitonga, A. F., & Syaufina, L. (2022). Urban greening effect on land surface temperature. Sensors, 22(11), 4168. https://doi.org/10.3390/s22114168
23. Ziaul, S., & Pal, S. (2018). Analyzing control of respiratory particulate matter on Land Surface Temperature in local climatic zones of English Bazar Municipality and Surroundings. Urban Climate, 24, 34-50. https://doi.org/10.1016/j.uclim.2018.01.006
24. Mahdavifard, M., Ahangar, S. K., Feizizadeh, B., Kamran, K. V., & Karimzadeh, S. (2023). Spatio-Temporal monitoring of Qeshm mangrove forests through machine learning classification of SAR and optical images on Google Earth Engine. International Journal of Engineering and Geosciences, 8(3), 239-250. https://doi.org/10.26833/ijeg.1118542
25. Ebrahimy, H., & Azadbakht, M. (2019). Downscaling MODIS land surface temperature over a heterogeneous area: An investigation of machine learning techniques, feature selection, and impacts of mixed pixels. Computers & Geosciences, 124, 93-102. https://doi.org/10.1016/j.cageo.2019.01.004
26. Srivastava, P. K., Han, D., Ramirez, M. R., & Islam, T. (2013). Machine learning techniques for downscaling SMOS satellite soil moisture using MODIS land surface temperature for hydrological application. Water Resources Management, 27, 3127-3144. https://doi.org/10.1007/s11269-013-0337-9
27. Sun, Y., Gao, C., Li, J., Wang, R., & Liu, J. (2019). Quantifying the effects of urban form on land surface temperature in subtropical high-density urban areas using machine learning. Remote Sensing, 11(8), 959. https://doi.org/10.3390/rs11080959
28. Li, W., Ni, L., Li, Z. L., Duan, S. B., & Wu, H. (2019). Evaluation of machine learning algorithms in spatial downscaling of MODIS land surface temperature. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12(7), 2299-2307. https://doi.org/10.1109/JSTARS.2019.2896923

29. Wang, S., Ma, Y., Wang, Z., Wang, L., Chi, X., Ding, A., ... & Zhang, Y. (2021). Mobile monitoring of urban air quality at high spatial resolution by low-cost sensors: impacts of COVID-19 pandemic lockdown. Atmospheric Chemistry and Physics, 21(9), 7199-7215. https://doi.org/10.5194/acp-21-7199-2021
30. Tan, J., NourEldeen, N., Mao, K., Shi, J., Li, Z., Xu, T., & Yuan, Z. (2019). Deep learning convolutional neural network for the retrieval of land surface temperature from AMSR2 data in China. Sensors, 19(13), 2987. https://doi.org/10.3390/s19132987
31. Mohammad, P., Goswami, A., Chauhan, S., & Nayak, S. (2022). Machine learning algorithm based prediction of land use land cover and land surface temperature changes to characterize the surface urban heat island phenomena over Ahmedabad city, India. Urban Climate, 42, 101116. https://doi.org/10.1016/j.uclim.2022.101116
32. Pandey, A., Mondal, A., Guha, S., Singh, D., & Kundu, S. (2023). Analysis of the Variability in Land Surface Temperature Due to Land Use/Land Cover Change for a Sustainable Urban Planning. Journal of Landscape Ecology, 16(3), 20-35. https://doi.org/10.2478/jlecol-2023-0015
33. Pandey, A., Mondal, A., Guha, S., Upadhyay, P. K., & Singh, D. (2023). A long-term analysis of the dependency of land surface temperature on land surface indexes. Papers in Applied Geography, 9(3), 279-294. https://doi.org/10.1080/23754931.2023.2187314
34. Pandey, A., Mondal, A., Guha, S., Upadhyay, P. K., Rashmi, & Kundu, S. (2023). Analysis of spectral indices-based downscaled land surface temperature in a humid subtropical city. International Journal of Image and Data Fusion, 14(4), 336-358. https://doi.org/10.1080/19479832.2023.2252818
35. Pandey, A., Mondal, A., Guha, S., & Upadhyay, P. K. (2022). A seasonal investigation on land surface temperature and spectral indices in Imphal City, India. Journal of Landscape Ecology, 15(3), 1-18. https://doi.org/10.2478/jlecol-2022-0015
36. Pandey, A., Mondal, A., Guha, S., Upadhyay, P. K., & Singh, D. (2022). Land use status and its impact on land surface temperature in Imphal city, India. Geology, Ecology, and Landscapes, 1-15. https://doi.org/10.1080/24749508.2022.2131962
37. https://earthexplorer.usgs.gov/
38. https://giovanni.gsfc.nasa.gov/giovanni/
39. https://developers.google.com/earth-engine/datasets/catalog/sentinel-5p
40. https://developers.google.com/earth-engine/datasets/catalog/modis
41. Khan, P. W., Byun, Y. C., & Jeong, O. R. (2023). A stacking ensemble classifier-based machine learning model for classifying pollution sources on photovoltaic panels. Scientific Reports, 13(1), 10256. https://doi.org/10.1038/s41598-023-35476-y
42. Mogaraju, J. K. (2023). Application of machine learning algorithms in the investigation of groundwater quality parameters over YSR district, India. Turkish Journal of Engineering, 7(1), 64-72. https://doi.org/10.31127/tuje.1032314
43. Gong, X., Huang, Y., Duong, J., Leng, S., Zhan, F. B., Guo, Y., ... & Luo, L. (2023). Industrial air pollution and low birth weight in New Mexico, USA. Journal of Environmental Management, 348, 119236. https://doi.org/10.1016/j.jenvman.2023.119236
44. Mogaraju, J. K. (2024). Machine learning empowered prediction of geolocation using groundwater quality variables over YSR district of India. Turkish Journal of Engineering, 8(1), 31-45. https://doi.org/10.31127/tuje.1223779
45. Raju, R. S., Raju, G. S., & Rajasekhar, M. (2023). A study on hydro-geological characterization through Dar-Zarrouk parameters in hard rock terrain of Mandavi River Basin, Andhra Pradesh, India. Arabian Journal of Geosciences, 16(8), 453. https://doi.org/10.1007/s12517-023-11573-0
46. Naidu, G., Zuva, T., & Sibanda, E. M. (2023, April). A review of evaluation metrics in machine learning algorithms. Computer Science On-line Conference, 15-25. https://doi.org/10.1007/978-3-031-35314-7_2
47. Psaros, A. F., Meng, X., Zou, Z., Guo, L., & Karniadakis, G. E. (2023). Uncertainty quantification in scientific machine learning: Methods, metrics, and comparisons. Journal of Computational Physics, 477, 111902. https://doi.org/10.1016/j.jcp.2022.111902
48. Islam, S. M., & Talukder, K. H. (2023). Exploratory Analysis of Smartphone Sensor Data for Human Activity Recognition. IEEE Access, 99481 – 99498. https://doi.org/10.1109/ACCESS.2023.3314651
49. Hughes, B. K., Wallis, R., & Bishop, C. L. (2023). Yearning for machine learning: applications for the classification and characterisation of senescence. Cell and Tissue Research, 394(1), 1-16. https://doi.org/10.1007/s00441-023-03768-4
50. Annapurna, S. M., Anitha, M., & Kumar, L. S. (2024). Composition and source based aerosol classification using machine learning algorithms. Advances in Space Research, 73(1), 474-497. https://doi.org/10.1016/j.asr.2023.09.068
51. Choi, M., Zhang, J., Zhang, Y., Fan, J., Li, X., & Ying, Q. (2024). Impact of wildfires on regional ozone and PM2. 5: Considering the light absorption of Brown carbon. Atmospheric Environment, 316, 120196. https://doi.org/10.1016/j.atmosenv.2023.120196
52. Zhen, Z., Chen, S., Yin, T., & Gastellu-Etchegorry, J. P. (2023). Globally quantitative analysis of the impact of atmosphere and spectral response function on 2-band enhanced vegetation index (EVI2) over Sentinel-2 and Landsat-8. ISPRS Journal of Photogrammetry and Remote Sensing, 205, 206-226. https://doi.org/10.1016/j.isprsjprs.2023.09.024
53. vom Scheidt, F., & Staudt, P. (2024). A data-driven Recommendation Tool for Sustainable Utility Service Bundles. Applied Energy, 353, 122137. https://doi.org/10.1016/j.apenergy.2023.122137
54. Moghadam, S. M., Yeung, T., & Choisne, J. (2023). A comparison of machine learning models' accuracy in predicting lower-limb joints' kinematics, kinetics, and muscle forces from wearable sensors. Scientific Reports, 13(1), 5046. https://doi.org/10.1038/s41598-023-31906-z

55. Javed, A. R., Saadia, A., Mughal, H., Gadekallu, T. R., Rizwan, M., Maddikunta, P. K. R., ... & Hussain, A. (2023). Artificial intelligence for cognitive health assessment: state-of-the-art, open challenges and future directions. Cognitive Computation, 15(6), 1767-1812. https://doi.org/10.1007/s12559-023-10153-4

56. Singh, P. (2023). Systematic review of data-centric approaches in artificial intelligence and machine learning. Data Science and Management, 6(3), 144-157. https://doi.org/10.1016/j.dsm.2023.06.001