

A GENERIC MACHINE LEARNING-BASED FRAMEWORK FOR PREDICTIVE MODELING OF LAND SURFACE TEMPERATURE

N. S. Rengma^{1*}, M. Yadav¹

¹Geographic Information System (GIS) Cell, Motilal Nehru National Institute of Technology Allahabad, Prayagraj-211004, Uttar Pradesh, India- (*nyenshu@gmail, ssm Yadav@mnnit.ac.in)

Commission IV, WG IV/3

KEY WORDS: Land surface temperature, Random forest, Spectral indices

ABSTRACT:

In the realm of data analytics, machine learning (ML) is one of the most successful techniques for making predictions. The ability of ML algorithms has also been studied in various aspects of land surface temperature (LST) besides its retrieval. The few investigations on LST retrieval using ML algorithms suggested that it may potentially obtain the LST values incorporating relevant variables of land surface parameters; however, the variables and ML models used differ, and so do their accuracies. The accuracy of the model is affected by the variable's contribution, its quality and quantity, and the fulfilment of each technique's assumptions. Hence this study provides a wide range of LST indicators to be employed for LST retrieval using a widely used ML algorithm, random forest. The ML algorithm framework for LST prediction is illustrated with significant spectral indices and terrain parameters across the highly industrialised district of Jharkhand, India. With the exception of one (aspect) variable, the analysis shows that all 20 variables that were included as independent factors were significant and equally contributed to the model. The model built with all the variables including the aspect of the terrain obtained an RMSE of 1.13 degree Celsius and R^2 of 0.48. However, after the removal of aspect, the model obtained an R^2 of 0.89 and RMSE of 0.74° C. The performance of the model on consecutive removal of lesser significant variables are evaluated and the study made clear how crucial it is to consider several environmental or land-use factors that could be pertinent to LST.

1. INTRODUCTION

Urbanisation and industrialisation drive rapid transformation of the land use and cover pattern (Kedia et al., 2021; Maheng et al., 2021). Although such development improves living standards in certain ways and causes social changes, it also impacts the environment and ultimately contributes to climate change. The human-induced changes in the land use/land cover (LULC) type transforms the pervious to impervious surfaces, which reduces the outgoing longwave terrestrial emissions while absorbing shortwave solar radiation, increasing the temperature of the atmosphere and the land surface. The local climate and rainfall patterns are altered by such temperature fluctuations (Eleftheriou et al., 2018; Ranagalage et al., 2018; Salwan et al., 2021). Several studies have found that LULC type significantly impacts the relative rise in land surface temperature (LST), particularly in urban areas. The world's population is expected to be exposed to human-induced climate change in metropolitan areas with a projected threefold increase in urban land use accompanied by heat stress (Seto et al., 2012; IPCC, 2014; Argüeso et al., 2015). LST massively contributes to the land surface processes, which impacts the Earth's environmental and climatic conditions (Anderson et al., 2008; Kustas and Anderson, 2009; Karnieli et al., 2010). Due to the significant heterogeneity of land use patterns and cover, understanding and keeping track of the dynamics of LST is crucial.

LST is typically obtained via remote sensors that collect data from one or more channels in the electromagnetic spectrum's thermal infrared window (Dar et al., 2019; Ermida et al., 2020). Numerous LST retrieval algorithms have been established for the Landsat satellite series (Sattari and Hashim, 2014; Duan et

al., 2020). Although earlier studies demonstrated a relationship between various spectral indices (especially LULC indices) and LST, they tend to be non-linear. Several techniques have been put forth to retrieve LST and quantify its relationship with associated variables. One such example is the geographically weighted regression (GWR) which uses linear models; however, they are insufficient to simulate the non-linear relations between LST and its indicators. Furthermore, multicollinearity has a significant negative impact on GWR and can result in predictions that are not accurate when the associated surface parameters have a high degree of correlation (Li et al., 2019; Zhao et al., 2018; Jia et al., 2021). Hence, researchers have implemented machine learning (ML) algorithms to address the shortcomings of traditional LST retrieval.

In the realm of data analytics, ML is one of the most successful techniques for making predictions using models and algorithms (Angra and Ahuja, 2017; Dhall et al., 2020). Although there is a paucity of research employing ML algorithm to retrieve LST, the technique has been used in other aspects of LST studies, such as spatial downscaling, simulation, addressing meteorological conditions, and similar tasks (Li et al., 2019; Buo et al., 2021; Maithani et al., 2022; Xu et al., 2021). Wang et al., (2022) employed land cover categories and Landsat bands to estimate the LST across the Tibetan plateau using random forest (RF). The LST by the RF trained model was the most accurate with the lowest root mean square error (RMSE) (1.89 Kelvin), according to Wang et al., who also obtained and examined the LST from the single channel technique, the linear regression model (2.77 Kelvin), and the moderate resolution imaging spectroradiometer product (MOD11A1) (3.62 Kelvin). In order to acquire the LST over Dehradun using an artificial neural network, Maithani et al., (2022) employed built-up

densities with a mean absolute error of 1.5° C and 0.9° C, while Rana and Suryanarayana (2022) employed four ML techniques, K nearest neighbour, neural network, regression tree, support vector machine incorporating three indices resulting with an RMSE of 0.54° C, 0.59° C, 0.89° C, and 0.61° C respectively. Mohammad et al., (2022) predicted the LST over a city in Ahmedabad with an RMSE of 0.03° C using XGB regressor.

Although earlier studies demonstrated the potential of ML algorithms to retrieve the LST incorporating related variables like land surface parameters, the variables and the ML models utilised vary, and so do their accuracies. The variable's contribution and its quality and quantity together with the fulfilment of each technique's presumptions, all have an impact on the model's accuracy (Zhong et al., 2019). For example, it is generally known that ML models suffer when there are few observations available, and this issue could worsen as more predictors are added. Herein, a robust ML algorithm, RF is employed to extract LST in the study area. RF has been proven to be efficient for generating spatial predictions, first described in Breiman (2001), and several studies (Hengl et al., 2018; Meyer et al., 2019; Sekuli et al., 2020; Pouyan et al., 2022; Wang et al., 2022) have verified that it is a promising technique. Since LST retrieval is vital for a sustainable check and planning for every rapidly growing region, our prime goals are:

- to retrieve LST over the mineral-rich East Singhbhum district of Jharkhand, India,
- to quantify the relationship between the spatially varying LST and surface parameters,
- and to determine the effectiveness of the variables in ML model to predict the LST.

2. MATERIALS AND METHODS

2.1 Study site description

One of the highest mining potential districts, East Singhbhum, also called Purbi Singhbhum (Figure 1) is considered for this study. It is situated at the extreme southeast of Jharkhand, having a longitudinal and latitudinal extent of 86°04' – 86°54' E and 22°12' – 23°01' N, respectively. In terms of industrial development, and mining and quarrying, the district leads the state Jharkhand. Approximately 53% of the district's total area is made up of residual mountains and hills formed of granite, gneiss, and schist. It is situated on the Chhotanagpur plateau, surrounded by lush forest from west to east and the Dalma Range on the northern edge. The Subernarekha river runs across the district from west to the south-east. The annual rainfall ranges from 1200 to 1400 mm, and the climate is temperate. The minimum recorded winter temperature is 8 degrees Celsius, while summer temperatures reach 40 to 45 degrees Celsius. Minerals are in abundance in the district. The primary minerals include iron ore, copper, uranium, and gold kyanite (<https://jamshedpur.nic.in/about-district/>). The region is a heavily mineralised zone that has undergone substantial mineral extraction, which could result in collateral environmental damage (Singh et al., 2018). It is essential to monitor the growth and associated climate variables because Jharkhand's urbanisation has increased across all districts, with East Singhbhum experiencing a considerable growth rate (Kumar and Reshmi, 2018).

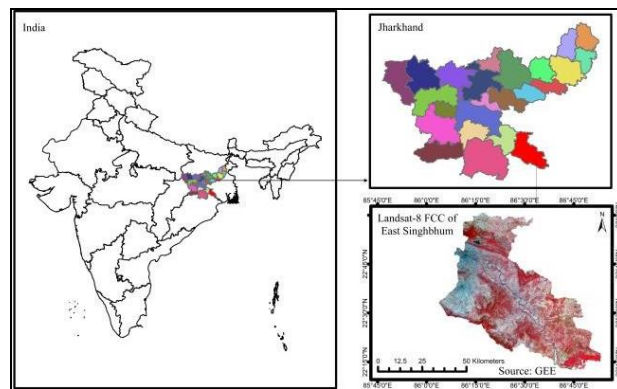


Figure 1. Location map of the study site

2.2 Dataset

For the LST derivation, this study employs the Landsat-8 operational land imager (OLI) and thermal infrared sensor (TIRS). In addition to Landsat data, the LST production code also uses two other datasets found in GEE: surface emissivity from the Advanced Spaceborne Thermal Emission and Reflection Radiometer Global Emissivity Database (ASTER GED) created by the National Aeronautics and Space Administration's (NASA) Jet Propulsion Laboratory, and atmospheric data from the reanalysis of the National Center for Environmental Prediction (NCEP) and National Center for Atmospheric Research (NCAR).

For the generation of the various LULC indices, we utilised the atmospherically corrected surface reflectance (SR) data from Landsat 8's collection 2, Tier 1 (highest available data quality) datasets that have been processed using a sophisticated data processing method and algorithm.

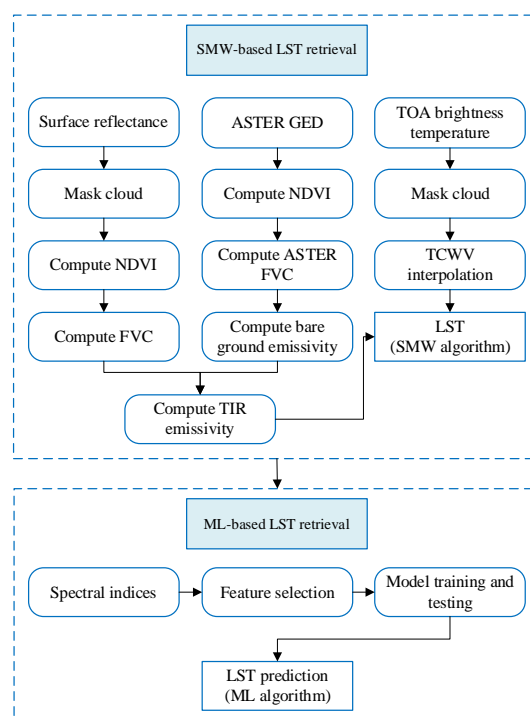


Figure 2. Overview of the methodology

2.3 Methodology

2.3.1 LST retrieval using SMW algorithm: Since the statistical mono-window (SMW) algorithm developed by the Climate Monitoring Satellite Application Facility (CM-SAF) method is a reliable technique to provide LST with high accuracy, it was applied to derive LST in this study using Google Earth Engine (Gorelick et al., 2017) (Figure 2). SMW algorithm is based on the empirical link between the top of atmosphere (TOA) brightness temperatures (BT) in a single thermal infrared channel and LST via a linear regression (Sun et al., 2004; Jiménez-Muñoz and Sobrino, 2003). Landsat's TOA-BT and SR collection have a cloud mask applied to them. The nearest two total column water vapour (TCWV) NCEP analysis times for each TOA-BT image are chosen and linearly interpolated to the Landsat observation time. The normalised difference vegetation index (NDVI) is computed using the SR data, which is then used to estimate the fractional vegetation cover (FVC).

$$FVC = \left(\frac{NDVI - NDVI_b}{NDVI_v - NDVI_b} \right) \quad (1)$$

where $NDVI_b$ and $NDVI_v$ represent, respectively, the NDVI values of soil or bare land and densely vegetated pixels. The threshold values are set to $NDVI_b = 0.2$ and $NDVI_v = 0.9$ in accordance with Jiménez-Muñoz et al., (2009) since the highest NDVI values in our study area are 0.88, and those below 0.2 are non-vegetated surfaces. We used mean ASTER GEDv3 NDVI and NDVI calculated from Landsat to account for vegetation.

The equivalent Landsat emissivity is then calculated using the ASTER emissivity values.

$$\varepsilon_b = FVC\varepsilon_{b_veg} + (1 - FVC)\varepsilon_{b_bare} \quad (2)$$

where, for a specific spectral band b , ε_{b_veg} and ε_{b_bare} represent the emissivity of dense vegetation and bare soil, respectively. Because the emissivity of vegetated surfaces typically exhibits only minor fluctuations in the TIR area, vegetation is taken as 0.99 as recommended (Peres and DaCamara, 2005). The surface emissivity is derived from the ASTER GEDv3. Therefore, the emissivity is modified to correspond to Landsat's thermal bands using the coefficients provided by Malakar et al., 2018. The TOA BT of the Landsat TIR band is then subjected to the SMW algorithm. The TWVC from NCEP and the designated classes are then used to overlay the algorithm's coefficients over the Landsat image.

$$LST = A_i \frac{T_b}{\varepsilon} + B_i \frac{1}{\varepsilon} + C_i \quad (3)$$

where T_b is the TIR channel's TOA BT, and ε is the channel's surface emissivity. The Coefficients A_i , B_i , and C_i were computed from linear regressions of radiative transfer simulations following Martin et al., (2016) using a dataset of air temperature, water vapour, and ozone profiles generated by Borbas et al., 2011.

2.3.2 Machine learning-based LST retrieval: Random forest is a non-parametric, non-linear ML technique that creates decision trees on several samples and predicts continuous or categorical data by averaging the values or taking the majority vote, respectively (Breiman, 2001). The creation of RF models entails modelling the target variable, sometimes referred to as the dependent variable, using a dataset of predictor or independent variables. Each tree is built using several bootstrapped datasets, and predictors are randomly selected from the entire array to act as candidates for each split. The approach de-correlates each tree because they are not reliant on

the same factors, reducing the variance in the dataset (Zhong et al., 2021). Here, we created an RF regression model with LST as the target variable to investigate the spatial variation of the LST over the study area using spectral indices as predictor variables.

The SMW algorithm-retrieved LST values were utilised in conjunction with different spectral indices to train the ML algorithm. A total of 20 variables, comprising parameters of the terrain (aspect, elevation, slope), vegetation indices, namely enhanced vegetation index (EVI), normalised difference vegetation index (NDVI), and transformed vegetation index (TVI), soil or bare land indices viz bare soil index (BSI), normalised difference bareness index (NDBaI), normalised difference bare soil index (NDBSI), normalised multi-band drought index (NMDI), optimised soil adjusted vegetation index (OSAVI), soil adjusted vegetation index (SAVI), and urban characteristic indicators such as built up index (BUI), enhanced built-up and bareness index (EBBI), new built-up index (NBI), normalised built-up area index (NBAI), urban index (UI), and normalised difference built-up index (NDBI), and lastly the water indices including the modified normalised difference water index (MNDWI) and normalised difference water index (NDWI) were generated (Table 1) that were shown to have impact on the LST.

The values of all the parameters were then extracted at the corresponding coordinates of the 429 total points that were generated at equal intervals over the entire study region, ensuring local heterogeneity. The necessary packages for the random forest algorithms were installed and loaded using the RStudio integrated development environment of R to train the model (R Core Team, 2020). To ensure that the variables generated are relevant, they were passed to boruta algorithm, which uses a wrapper approach built around a random forest (Kursa and Rudnicki, 2010). It was cross-validated with the random forest variable importance ranking of the ranger package. The RF model was implemented using a computationally faster ranger package, which contains several parameters that can be fine-tuned. The most important parameters were considered when developing the model (Probst and Boulesteix, 2017; Probst et al., 2019):

Mtry- number of variables to be drawn at each split in a tree.

Node size- Minimum number of observations in a terminal node.

Ntree- number of trees in the forest.

Sample size: Number of observations drawn for each tree.

To determine whether the trained model's predictive ability can be extrapolated to an unknown dataset, we randomly split the sample points (429 points) into 70 and 30 per cent and used the later portion to validate. The predictor variables' imageries were then applied to predict the LST for the entire study region using the trained and validated model that has reached its maximum accuracy and stability based on root mean square error (RMSE) and r-square (R^2). In order to assess the effectiveness of the variables generated, the model was run repeatedly with all the variables generated with fine-tuned hyperparameters by consecutively removing the less significant ones.

3. RESULTS AND DISCUSSION

To demonstrate the spatial distribution of surface temperature in response to the surface parameters of the study area, LST map of the year 2021 was generated using the SMW algorithm.

Table 1. Spectral indices of the surface parameters

Spectral indices (References)	Equations*
BSI (Abdullah et al., 2022)	$\frac{(\rho_{SWIR1} + \rho_R) - (\rho_{NIR} + \rho_B)}{(\rho_{SWIR1} + \rho_R) + (\rho_{NIR} + \rho_B)}$
BUI (Javed et al. 2021)	$2 \times \frac{(\rho_R \times \rho_{SWIR2}) - (\rho_{SWIR1} \times \rho_{SWIR1})}{(\rho_R + \rho_{SWIR1}) \times (\rho_{SWIR1} + \rho_{SWIR2})}$
EBBI (Ramaiah et al., 2020)	$\frac{(\rho_{SWIR1} - \rho_{NIR})}{10 \times \sqrt{\rho_{SWIR1} + \rho_{TIR}}}$
EVI (AlAdemomi et al., 2020)	$\frac{2.5(\rho_{NIR} - \rho_R)}{\rho_{NIR} + 6\rho_R - 7.5\rho_B + 1}$
MNDWI (Rahman et al., 2017, Khan et al., 2021)	$\frac{\rho_G - \rho_{SWIR1}}{\rho_G + \rho_{SWIR1}}$
NBAI (Javed et al., 2021)	$\frac{\rho_{SWIR2} - \rho_{SWIR1}}{\rho_G}$ $\rho_{SWIR2} + \frac{\rho_{SWIR1}}{\rho_G}$
NBI (Javed et al., 2021)	$\frac{\rho_R \times \rho_{SWIR1}}{\rho_{NIR}}$
NDBaI (Essa et al., 2012; Guha and Govil, 2021)	$\frac{\rho_{SWIR2} - \rho_{SWIR1}}{\rho_G}$ $\rho_{SWIR2} - \frac{\rho_{SWIR1}}{\rho_G}$
NDBI (Rahman et al., 2017, Kafy et al., 2022)	$\frac{\rho_{SWIR1} - \rho_{NIR}}{\rho_{SWIR1} + \rho_{NIR}}$
NDBSI (Rahman et al., 2017; Kafy et al., 2022)	$\frac{(\rho_R + \rho_{SWIR1}) - (\rho_{NIR} + \rho_B)}{(\rho_R + \rho_{SWIR1}) + (\rho_{NIR} + \rho_B)}$
NDVI (Jia et al., 2021, Rana and Suryanarayana, 2022)	$\frac{\rho_{NIR} - \rho_R}{\rho_{NIR} + \rho_R}$
NDWI (Jia et al., 2021, Rana and Suryanarayana, 2022)	$\frac{\rho_G - \rho_{NIR}}{\rho_G + \rho_{NIR}}$
NMDI (Guha and Govil, 2021)	$\frac{[\rho_{NIR} - (\rho_{SWIR1} + \rho_{SWIR2})]}{[\rho_{NIR} + (\rho_{SWIR1} + \rho_{SWIR2})]}$
OSAVI (Zou et al., 2020)	$1.6 \left(\frac{\rho_{NIR} - \rho_R}{\rho_{NIR} + \rho_R + 0.16} \right)$
SAVI (Rahman et al., 2017; Ramaiah et al., 2020)	$\frac{1.5(\rho_{NIR} - \rho_R)}{\rho_{NIR} + \rho_R + 0.5}$
TVI (Gong et al., 2006)	$0.5 [120(\rho_{NIR} - \rho_G) - 200(\rho_R - \rho_G)]$
UI (Khan et al., 2021)	$\frac{\rho_{SWIR2} - \rho_{NIR}}{\rho_{SWIR2} + \rho_{NIR}}$

* $\rho_B, \rho_G, \rho_R, \rho_{NIR}, \rho_{SWIR1},$ and ρ_{SWIR2} are respectively, the blue, green, red, near infrared, short wavelength infrared 1, and short wavelength 2 bands of Landsat 8.

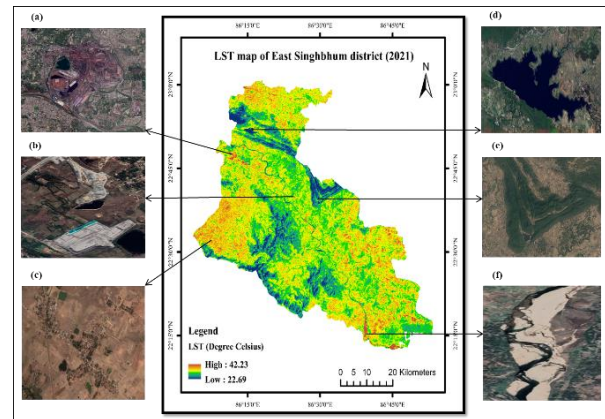


Figure 3. LST variation in response to LULC; (a) Industrial and settlement area (b), quarrying/mining area (c) agricultural and settlement (d), waterbody (e) Vegetation, (f) Bare ground

The mean of the LST for the year 2021 retrieved by the SMW algorithm ranged between 22.69 to 42.23 degree Celsius. The distinct thermal signatures exhibited in the LST maps are due to the different attributes of the various land use/cover type. Throughout the period, temperatures were high in places with various factories and industries, mining and quarrying locations, urban settlements, and agricultural lands. On the other hand, waterbody and vegetation have the lowest temperature throughout the area. The graph shown in Figure 4 depicts a positive correlation between urban spectral indices' (BSI, EBBI, MNDWI, NBAI, NBI, NDBaI, NDBI, NDBSI, NMDI, and UI) and the LST. In the case of vegetation and waterbody, the opposite is supposed to be true. Here the vegetation indices are negatively correlate with LST because wherever there is vegetation cover, the surface temperature is low (Mishra et al., 2021). Similarly, the NDWI-LST relation is low as water temperature varies with its surroundings, implying a negative correlation between them.

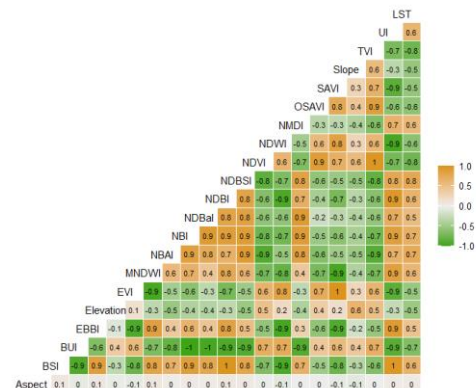


Figure 4. Correlation matrix of the LST and the LULC indices

The variables that are highly correlated, whether positively or negatively, are shown to be important for the prediction of the LST by Boruta algorithm (Figure 5). With the exception of the aspect, which is due to the

planar study region (70% of the study area has nearly level or very gentle slopes), all variables were determined to be significant for the LST prediction.

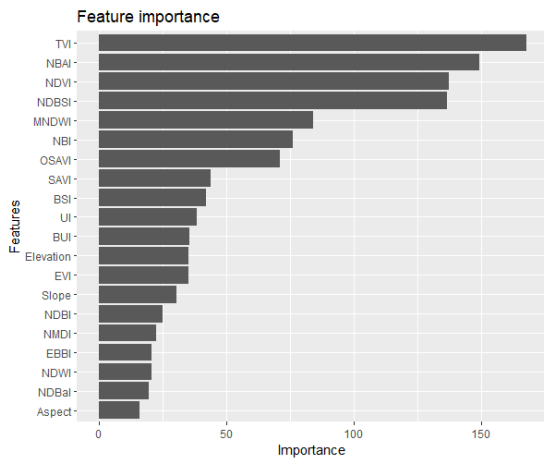


Figure 5. Feature importance ranking

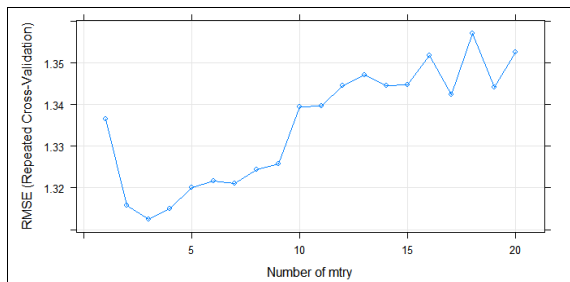


Figure 6. Grid search for the number of variables at each split (mtry)

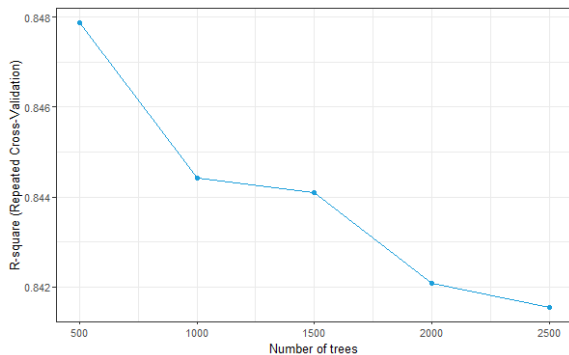


Figure 7. Grid search for number of trees (ntree)

To predict the LST, the RF model was built using the variables selected by the feature selection algorithm. The Mtry, minimum node size, sample size, and ntree values were obtained via the hyperparameter grid search for optimal parameters as 3 (Figure 6), 10, 0.7 (70%), and 500 (Figure 7), respectively. Less correlated trees are produced by using smaller sample sizes, larger node size values, and smaller mtry values (Probst et al., 2019). These trees are more distinct from one another; thus, they produce more distinctive predictions. The model predicted the LST on the testing dataset with an R square of 0.89 and RMSE of 0.74 degree Celsius. The SMW algorithm has an LST ranging from 22.69 to 42.23 degree

Celsius, while the RF predicted LST ranges from 24.75 to 41.09 degree Celsius. The range difference in the temperatures are due to the deficient coverage of the temperature above 41 degrees Celsius (0.01 square kilometre) and below 24 degrees Celsius (0.06 square kilometre), which prevented the model from being trained with smaller samples. It was not possible to infer that expanding the dataset would increase the model's accuracy as proposed by Wang et al., (2022). Because when we increased the training points in our case, the majority of the region was predicted with hardly any spatial variance of the LST.

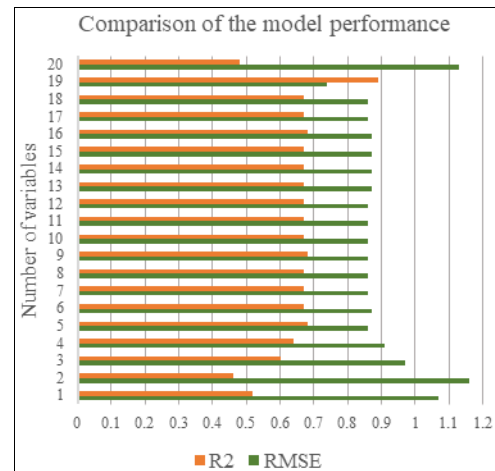


Figure 8. Comparison of the model performance based on the variables' importance

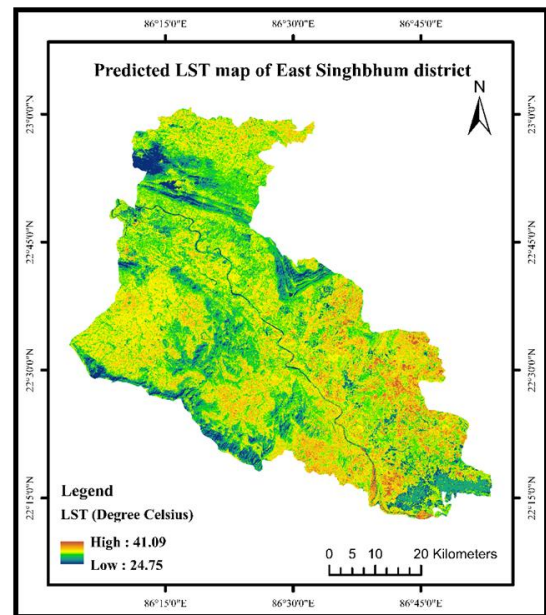


Figure 9. Predicted LST map of East Singhbhum district using RF algorithm

To evaluate the effectiveness of the variables, the models were run repeatedly with all the variables and gradually removed one lesser significant variable at a time. The model built with all the variables including aspect obtained an RMSE of 1.13 and R² of 0.48, which when

removed the model improved. Using all the other variables by removing the terrain aspect obtained an R^2 of 0.89 and RMSE of 0.74. The performance of the model on consecutive removal of lesser significant variables are displayed in figure 8.

4. CONCLUSION

Given the likelihood that global temperatures will continue to rise, it is crucial to detect and monitor LST frequently. The LST variation in response to surface parameters using the spectral indices was determined in the mineral-abundant and industrially developed district of Jharkhand, East Singhbhum. The study has also quantified the relationship between the spatially varying LST and surface parameters. The analysis showed substantial variations in the temperatures of each type of land use and cover. The ML algorithm predicted the LST incorporating terrain parameters and significant spectral indices of vegetation, urban, soil or bare land and water, thus demonstrating the potential of the algorithm to predict the LST with known errors. Keeping in mind that the variables used by the ML algorithm are not redundant, other surface parameters may be investigated. Furthermore, to prevent environmental extremes, the LST must be closely monitored due to the anticipated increase in population and rising urbanisation and industry. The study's findings may also be helpful in developing site-specific adaptation plans/strategies to mitigate environmental effects and enhance urban residents' quality of life. City planners and policymakers may take specific steps to make the city less sensitive to climate change by assessing the parameter (LST) that is influenced by compact land use patterns.

REFERENCES

- Abdullah, S., Barua, D., Abdullah, S., Abubakar, M., Rabby, Y. W., 2022. Investigating the impact of land use/land cover change on present and future land surface temperature (LST) of Chittagong, Bangladesh. *Earth Systems and Environment*, 6(1), 221-235.
- AlAdemomi, A. S., Okolie, C. J., Daramola, O. E., Agboola, R. O., Salami, T. J., 2020. Assessing the relationship of LST, NDVI and EVI with land cover changes in the Lagos Lagoon environment. *Quaestiones Geographicae*, 39(3), 87-109.
- Anderson, M. C., Norman, J. M., Kustas, W. P., Houborg, R., Starks, P. J., Agam, N., 2008. A thermal-based remote sensing technique for routine mapping of land-surface carbon, water and energy fluxes from field to regional scales. *Remote Sensing of Environment*, 112(12), 4227-4241.
- Angra, S., Ahuja, S., 2017. Machine learning and its applications: A review. In *2017 international conference on big data analytics and computational intelligence (ICBDAC)* (pp. 57-60). IEEE.
- Argüeso, D., Evans, J. P., Pitman, A. J., Di Luca, A., 2015. Effects of city expansion on heat stress under climate change conditions. *PLoS one*, 10(2), e0117066.
- Borbás, E. E., Seemann, S. W., Kern, A. N. I. K. O., Moy, L. E. S. L. I. E., Li, J., Gumley, L. I. A. M., & Menzel, W. P., 2011. MODIS atmospheric profile retrieval algorithm theoretical basis document. *Electronic resource*, http://modis-atmos.gsfc.nasa.gov/MOD07_L2/atbd.html, (10.08. 2015).
- Breiman, L., 2001. Random forests. *Machine learning*, 45(1), 5-32.
- Buo, I., Sagris, V., Burdun, I., Uuemaa, E., 2021. Estimating the expansion of urban areas and urban heat islands (UHI) in Ghana: a case study. *Natural Hazards*, 105(2), 1299-1321.
- Dar, I., Qadir, J., Shukla, A., 2019. Estimation of LST from multi-sensor thermal remote sensing data and evaluating the influence of sensor characteristics. *Annals of GIS*, 25(3), 263-281.
- Dhall, D., Kaur, R., Juneja, M., 2020. Machine learning: a review of the algorithms and its applications. *Proceedings of ICRIC 2019*, 47-63.
- Duan, S. B., Han, X. J., Huang, C., Li, Z. L., Wu, H., Qian, Y., Gao, M., Leng, P., 2020. Land surface temperature retrieval from passive microwave satellite observations: State-of-the-art and future directions. *Remote Sensing*, 12(16), 2573.
- Eleftheriou, D., Kiachidis, K., Kalmintzis, G., Kalea, A., Bantasis, C., Koumadoraki, P., Spathara, M. E., Tsolaki, A., Tzampazidou, M. I., Gemitzi, A., 2018. Determination of annual and seasonal daytime and nighttime trends of MODIS LST over Greece-climate change implications. *Science of the total environment*, 616, 937-947.
- Ermida, S. L., Soares, P., Mantas, V., Göttsche, F. M., Trigo, I. F., 2020. Google earth engine open-source code for land surface temperature estimation from the landsat series. *Remote Sensing*, 12(9), 1471.
- Essa, W., Verbeiren, B., van der Kwast, J., Van de Voorde, T., Batelaan, O., 2012. Evaluation of the DisTrad thermal sharpening methodology for urban areas. *International Journal of Applied Earth Observation and Geoinformation*, 19, 163-172.
- Gong, P., Chen, Z., Tang, H., Zhang, F., 2006. Land cover classification based on multi-temporal MODIS NDVI and LST in Northeastern China. In *2006 IEEE International Symposium on Geoscience and Remote Sensing* (pp. 1149-1152). IEEE.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*.
- Guha, S., Govil, H., 2021. Annual assessment on the relationship between land surface temperature and six remote sensing indices using Landsat data from 1988 to 2019. *Geocarto International*, 1-20.
- Hengl, T., ussbaum, M., Wright, M. N., Heuvelink, G. B., räler, B., 2018. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, 6, e5518.
- IPCC (2014) Climate change 2014: impacts, adaptation, and vulnerability. Part A: global and sectoral aspects. Contribution of working group II to the fifth assessment report of the intergovernmental panel on climate change [Field, C.B., V.R. Barros, D.J. Dokken, K.J. Mach, M.D. Mastrandrea, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L. White (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA

- Javed, A., Cheng, Q., Peng, H., Altan, O., Li, Y., Ara, I., Huq, E., Ali, Y., Saleem, N., 2021. Review of Spectral Indices for Urban Remote Sensing. *Photogrammetric Engineering & Remote Sensing*, 87(7), 513-524.
- Jia, H., Yang, D., Deng, W., Wei, Q., Jiang, W., 2021. Predicting land surface temperature with geographically weighed regression and deep learning. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 11(1), e1396.
- Jiménez-Muñoz, J. C., Sobrino, J. A., 2003. A generalised single-channel method for retrieving land surface temperature from remote sensing data. *Journal of geophysical research: atmospheres*, 108(D22).
- Jiménez-Muñoz, J. C., Sobrino, J. A., Plaza, A., Guanter, L., Moreno, J., Martínez, P., 2009. Comparison between fractional vegetation cover retrievals from vegetation indices and spectral mixture analysis: Case study of PROBA/CHRIS data over an agricultural area. *Sensors*, 9(02), 768-793.
- Kafy, A. A., Saha, M., Rahaman, Z. A., Rahman, M. T., Liu, D., Fattah, M. A., Rakib, A. A., AlDousari, A. E., Rahaman, S. K., Hasan, M. Z., Ahasan, M. A. K., 2022. Predicting the impacts of land use/land cover changes on seasonal urban thermal characteristics using machine learning algorithms. *Building and Environment*, 217, 109066.
- Karnieli, A., Agam, N., Pinker, R. T., Anderson, M., Imhoff, M. L., Gutman, G. G., Panov, N., Goldberg, A., 2010. Use of NDVI and land surface temperature for droughtassessment: Merits and limitations. *Journal of Climate*, 23,618–633.
- Kedia, S., Bhakare, S. P., Dwivedi, A. K., Islam, S., Kaginalkar, A., 2021. Estimates of change in surface meteorology and urban heat island over northwest India: Impact of urbanisation. *Urban Climate*, 36, 100782.
- Khan, M. S., Ullah, S., Chen, L., 2021. Comparison on Land-Use/Land-Cover Indices in Explaining Land Surface Temperature Variations in the City of Beijing, China. *Land*, 10(10), 1018.
- Kumar, G., Reshmi, R. S., 2018. Urbanization and development in selected Empowered Action Group (EAG) states in India. *Jharkhand Journal of Development and Management Studies*, 16(1), 7641-7658.
- Kursa, M. B., Rudnicki, W. R., 2010. Feature selection with the Boruta package. *Journal of statistical software*, 36, 1-13.
- Kustas, W., Anderson, M., 2009. Advances in thermal infrared remote sensing for land surface modeling. *Agriculturaland Forest Meteorology*, 149, 2071–2081.
- 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.
- Maheng, D., Pathirana, A., & Zevenbergen, C., 2021. A preliminary study on the impact of landscape pattern changes due to urbanisation: Case study of Jakarta, Indonesia. *Land*, 10(2), 218.
- Maithani, S., Nautiyal, G., Sharma, A., Sharma, S. K., 2022. Simulation of Land Surface Temperature Patterns Over Future Urban Areas—A Machine Learning Approach. *Journal of the Indian Society of Remote Sensing*, 1-18.
- Malakar, N. K., Hulley, G. C., Hook, S. J., Laraby, K., Cook, M., Schott, J. R., 2018. An operational land surface temperature product for Landsat thermal data: Methodology and validation. *IEEE Transactions on Geoscience and Remote Sensing*, 56(10), 5717-5735.
- Martins, J. P., Trigo, I. F., Bento, V. A., Da Camara, C., 2016. A Physically constrained calibration database for land surface temperature using infrared retrieval algorithms. *Remote Sensing*, 8(10), 808.
- Meyer, H., Reudenbach, C., Wöllauer, S., Nauss, T., 2019. Importance of spatial predictor variable selection in machine learning applications—Moving from data reproduction to spatial prediction. *Ecological Modelling*, 411, 108815.
- Mishra, R., Singh, E., Kumar, A., Kumar, S., 2021. Application of remote sensing for assessment of change in vegetation cover and the subsequent impact on climatic variables. *Environmental Science and Pollution Research*, 28(31), 41675-41687.
- 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 characterise the surface urban heat island phenomena over Ahmedabad city, India. *Urban Climate*, 42, 101116.
- Peres, L. F., DaCamara, C. C., 2005. Emissivity maps to retrieve land-surface temperature from MSG/SEVIRI. *IEEE Transactions on Geoscience and Remote Sensing*, 43(8), 1834-1844.
- Pouyan, S., Rahmanian, S., Amindin, A., Pourghasemi, H. R., 2022. Spatial and seasonal modeling of the land surface temperature using random forest. In *Computers in Earth and Environmental Sciences* (pp. 221-234). Elsevier.
- Probst, P., Wright, M. N., Boulesteix, A. L., 2019. Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: data mining and knowledge discovery*, 9(3), e1301.
- R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rahman, M. T., Aldosary, A. S., Morteja, M. G., 2017. Modeling future land cover changes and their effects on the land surface temperatures in the Saudi Arabian eastern coastal city of Dammam. *Land*, 6(2), 36.
- Ramaiah, M., Avtar, R., Rahman, M., 2020. Land cover influences on LST in two proposed smart cities of India: Comparative analysis using spectral indices. *Land*, 9(9), 292.
- Rana, V. K., Suryanarayana, T. M. V., 2022. Quantifying the effects of urban land forms on land surface temperature and modelling the spatial variation using machine learning. *Geocarto International*, 37(12), 3540-3564.
- Ranagalage, M., Estoque, R. C., Zhang, X., Murayama, Y., 2018. Spatial changes of urban heat island formation in the

Colombo District, Sri Lanka: Implications for sustainability planning. *Sustainability*, 10(5), 1367.

Salwan, A. A., Ahmed, A. A., & Salim, M. A., 2021. Using ArcGIS Software and Remote Sensing Technology to Predict Land Surface Temperature (LST) for Monitoring Ecological and Climate Change in Hor Al-Dalmaj, Southern Iraq. In *IOP Conference Series: Earth and Environmental Science* (Vol. 790, No. 1, p. 012076). IOP Publishing.

Sattari, F., Hashim, M., 2014. A brief review of land surface temperature retrieval methods from thermal satellite sensors. *Middle-East Journal of Scientific Research*, 22(5), 757-768.

Sekulić, A., Kilibarda, M., Heuvelink, G., Nikolić, M., & Bajat, B., 2020. Random forest spatial interpolation. *Remote Sensing*, 12(10), 1687.

Seto, K. C., Güneralp, B., Hutyrá, L. R., 2012. Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. *Proceedings of the National Academy of Sciences*, 109(40), 16083-16088.

Singh, U. K., Ramanathan, A. L., Subramanian, V., 2018. Groundwater chemistry and human health risk assessment in the mining region of East Singhbhum, Jharkhand, India. *Chemosphere*, 204, 501-513.

Sun, D., Pinker, R. T., Basara, J. B., 2004. Land surface temperature estimation from the next generation of Geostationary Operational Environmental Satellites: GOES M–Q. *Journal of Applied Meteorology and Climatology*, 43(2), 363-372.

Wang, X., Zhong, L., Ma, Y., 2022. Estimation of 30 m land surface temperatures over the entire Tibetan Plateau based on Landsat-7 ETM+ data and machine learning methods. *International Journal of Digital Earth*, 15(1), 1038-1055.

Xu, S., Zhao, Q., Yin, K., He, G., Zhang, Z., Wang, G., Wen, M., Zhang, N., 2021. Spatial downscaling of land surface temperature based on a multi-factor geographically weighted machine learning model. *Remote Sensing*, 13(6), 1186.

Zhao, C., Jensen, J., Weng, Q., Weaver, R., 2018. A geographically weighted regression analysis of the underlying factors related to the surface urban heat island phenomenon. *Remote Sensing* 10(9), 1428.

Zhong, S., Zhang, K., Bagheri, M., Burken, J. G., Gu, A., Li, B., Ma, X., Marrone, B. L., Ren, Z. J., Schrier, J., Shi, W., Tan, H., Wang, T., Wang, X., Wong, B. M., Xiao, X., Yu, X., Zhu, J. J., Zhang, H., 2021. Machine learning: new ideas and tools in environmental science and engineering. *Environmental Science & Technology*, 55(19), 12741-12754.

Zou, F., Li, H., Hu, Q., 2020. Responses of vegetation greening and land surface temperature variations to global warming on the Qinghai-Tibetan Plateau, 2001–2016. *Ecological Indicators*, 119, 106867.