OPEN ACCESS
PAPER

Mongolian Journal of Geography and Geoecology - MJGG

Received: 09 May 2025; Accepted: 12 June 2025; Published online: 1 September 2025
https://doi.org/10.5564/mjgg.v62i46.4125

Estimating the spatial distribution of urban heat islands in six central districts of Ulaanbaatar city using machine learning algorithm

Byambadolgor Batdorj[1,*], Amarsaikhan Damdinsuren[1], Sainbayar Dalantai[1], Munkh-Erdene Altangerel[2]

[1]*Institute of Geography and Geoecology, Mongolian Academy of Sciences, Ulaanbaatar 15170, Mongolia*
[2]*Institute of Mathematics and Digital Technology, Mongolian Academy of Sciences, Ulaanbaatar, Mongolia*

***CORRESPONDING AUTHOR**
Byambadolgor Batdorj
byambadolgorb@mas.ac.mn
ORCID
0000-0002-9017-0729

CITATION
Byambadolgor B, Amarsaikhan D, Sainbayar D, Munkh-Erdene A (2025) Estimating The Spatial Distribution of Urban Heat Islands in Six Central Districts of Ulaanbaatar City Using Machine Learning Algorithm. *Mongolian Journal of Geograhy and Geoecology,* 62(46), 1–7.
https://doi.org/10.5564/mjgg.v62i46.4125

ABSTRACT

Urban Heat Islands (UHIs) represent a growing environmental challenge in rapidly urbanizing cities, particularly in extreme climate regions such as Ulaanbaatar, Mongolia. This study aims to model and analyze the spatial distribution of UHIs across six central districts of Ulaanbaatar using Landsat 5 and Landsat 8 satellite imagery. Land Surface Temperature (LST) was derived from Landsat 8, and its relationship with six spectral indices the Normalized Difference Vegetation Index (NDVI), Normalized Difference Built-up Index (NDBI), Modified Normalized Difference Water Index (MNDWI), Bare Soil index (BI), Normalized Differences Water Index (NDWI), and Normalized Building Soil Index (NDBSI) was assessed. Two ensemble machine learning algorithms, Random Forest (RF) and Extreme Gradient Boosting (XGB), were employed to model UHI patterns. RF performed better in 1994 (a coefficients of determination (R^2) = 0.72, a root-mean-square error (RMSE) = 1.89°C), while XGB showed superior performance in 2024 (R^2 = 0.75, RMSE = 1.97°C). BI and NDBSI were the most influential contributors to UHI intensity, whereas vegetated areas (NDVI, MNDVI) had consistent cooling effects. The spatial modeling revealed a clear intensification of UHI effects over time, particularly in built-up and bare land zones. This study is significant as it applies ML techniques to long-term satellite data in a cold-climate, data-limited urban setting. By providing a rare longitudinal perspective, it contributes to understanding UHI dynamics in Mongolia and demonstrates the potential of open-access remote sensing data combined with ML for urban climate assessments. The findings offer valuable insight for urban planners, emphasizing the critical role of green infrastructure in mitigating thermal stress and informing climate-resilient development strategies in rapidly growing cities.

KEYWORDS

Urban heat island, Machine learning, Land surface temperature

1. INTRODUCTION

Climate change is a critical global driver that significantly impacts the environment, ecosystems, and human societies, and it has been increasingly contributing to the intensification of the Urban heat island (UHI) phenomenon [1]. UHIs refer to the phenomenon where urban areas experience significantly higher temperatures than their surrounding rural regions due to human activities and alterations in land surface properties [2]. The increasing urbanization, characterized by the replacement of natural landscapes with impervious surfaces such as concrete and asphalt, exacerbates UHIs by altering thermal properties, reducing vegetation cover, and increasing heat retention. Understanding and mitigating UHIs is crucial for urban planning, public health, and climate change adaptation [3].

In recent years, the integration of multi-source remote sensing (RS) data with machine learning (ML) techniques has significantly advanced the study of natural phenomena, providing powerful and flexible tools for mapping, assessment, and modeling of complex environmental processes. Numerous studies have demonstrated the utility of ML algorithms in environmental applications. For example, Talukdar et al [4] applied various ML classifiers for land cover and land use (LULC) classification, achieving an accuracy of 94.3% with RF and 88.5% with Support Vector Machine (SVM). Sainbayar [5] utilized multiple ML algorithms to model atmospheric CO_2 concentrations, with the Light Gradient Boosting Machine (LGBM) yielding an the coefficient of determination (R^2) value of 0.73. Yoo et al. [6] employed ML techniques to enhance the spatial resolution of Moderate Resolution Imaging Spectroradiometer (MODIS) nighttime land surface temperature (LST) data by integrating Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) thermal imagery, wherein the Local Linear Forest (LLF) method achieved an R^2 range of 0.85–0.92 and a Root mean square error (RMSE) between 1.5°C and 2.0°C. Similarly, Arulananth et al. [7] leveraged hyperspectral imaging and artificial intelligence to detect and interpret urban thermal patterns, reporting classification accuracies exceeding 90%, which outperformed traditional methods. The aim of this study is to map and analyze the spatial distribution of UHIs in Ulaanbaatar using Landsat imagery and advanced ML algorithms.

2. DATASETS AND RESEARCH METHODS

2.1. Datasets

For this study, the city of Ulaanbaatar—the coldest capital city in the world—was selected as the study area due to its unique climatic conditions and rapid urbanization. Situated in the Tuul River valley at an elevation of approximately 1,350 meters above sea level (47.8864°N, 106.9057°E), Ulaanbaatar experiences a harsh semi-arid continental climate. The city has an average annual temperature of −0.4°C, with extreme winter lows reaching −40°C and summer highs occasionally exceeding 35°C. As of 2024, Ulaanbaatar is home to approximately 1.6 million residents, representing nearly 48% of Mongolia's total population, thereby making it the nation's most densely populated urban center [8]. To assess the spatial distribution and intensity of the UHI effect in a representative and comprehensive manner, six districts of Ulaanbaatar—Bayanzurkh, Chingeltei, Sukhbaatar, Songinokhairkhan, Khan-Uul, and Bayangol—were selected as the study area. These districts were purposefully chosen because they encompass both the highly urbanized central zones and the more vegetated and less densely built-up peripheral areas of the city. This spatial variation in land use, building density, and green cover provides an ideal basis for analyzing how different surface types contribute to urban heat dynamics.

To assess the spatial distribution and intensity of the UHI phenomenon over time, multi-temporal satellite imagery from the Landsat program was utilized. Specifically, Landsat 5 Thematic Mapper (TM) data acquired on September 5, 1994, and Landsat 8 Operational Land Imager (OLI) data acquired on September 7, 2024, were used. For Landsat OLI, LST is derived using the single-channel method with thermal Band 10, while Band 11, which is also a thermal band, is excluded from the LST calculation. The native spatial resolutions of Landsat TM (120 m) and Landsat OLI (100 m) were resampled in Collection 1 to 30 m, which is the spatial resolution that we used. In this study, data from the visible, near-infrared (NIR), shortwave infrared (SWIR), and thermal spectral bands were utilized. Several vegetation indices, such as the Normalized Difference Vegetation Index (NDVI), Bare Soil Index (BI), Normalized Difference Water Index (NDWI), Normalized Difference Bare Soil Index (NDBSI), Modified Normalized Difference Vegetation Index (MNDVI) and Normalized Difference Built-up Index (NDBI), were used to

model the spatial distribution of the UHI using machine learning algorithms [9].

2.2. Research methods

The standard method for retrieving LST from raw Landsat datasets requires the conversion of the digital number (DN) values of the thermal bands (band 6 in Landsat TM and Bands 10 and 11 in Landsat OLI) first into absolute radiance values. These radiance values are then used to derive at satellite brightness temperature, calculated under an assumption of unity emissivity and using pre-launch calibration constants [10]. This process is followed by a correction for spectral emissivity according to the nature of the landscape. In this study, we used the preprocessed Band 10 of Landsat-OLI, containing top of atmosphere brightness temperature values expressed in Kelvin to produce an LST map for each of the study areas. After converting the brightness temperature values into degrees Celsius (°C), the emissivity-corrected LST was calculated [11].

RF is an ensemble method that constructs multiple decision trees using random subsets of the data and performs analysis by aggregating their predictions, typically by averaging in regression tasks or voting in classification tasks. This approach offers high accuracy and is particularly suitable for handling large-scale datasets [12].

XGB is a powerful algorithm based on the ensemble gradient boosting technique, designed to iteratively grow decision trees in a way that minimizes prediction error. In the XGB algorithm, each decision tree is grown sequentially, where the output of each previous tree is used to model the residual errors, enabling the subsequent tree to focus on correcting those errors, thus iteratively minimizing overall prediction error. The final solution is derived by combining the predictions from all decision trees in the model [13].

To evaluate the suitability and predictive performance of the selected vegetation indices in modeling UHI distribution, several statistical performance metrics were computed: R^2 and RMSE [14]. For model training and evaluation, 80% of the dataset was used for training, while the remaining 20% was reserved for testing the generalization capability of each model [15].

3. RESULT AND DISCUSSION

In this study, the appropriateness of the selected indices for modeling the spatial distribution of the UHI in Ulaanbaatar was evaluated using data from the years 1994 and 2024. For the modeling, a total of 818 LST samples were selected as the dependent variable, while six vegetation indices were used as independent variables. RF and XGB models were employed for analysis. As shown in Table 1, for the year 1994, the RF model demonstrated superior explanatory power compared to the XGB model, with a higher coefficient of determination (R^2 = 0.722). RF model exhibited lower prediction error, as indicated by a lower RMSE value. In contrast, for the year 2024, the XGB model outperformed the RF model, achieving an R^2 of 0.682 and an RMSE of 1.972°C.

Table 1. Performance of models predicting the UHI

1994	Train		Test	
	RF	XGB	RF	XGB
R^2	0.772	0.612	0.722	0.532
RMSE	1.832	2.322	1.892	2.722
2024	**Train**		**Test**	
	RF	XGB	RF	XGB
R^2	0.752	0.832	0.612	0.682
RMSE	1.652	1.332	1.982	1.972

The relationships between LST and various influencing factors were analyzed using Spearman's rank correlation coefficient, with a threshold value set at r = ±0.2. As illustrated in Figure1, the BI, NDVSI, and NDBI indices exhibited strong positive correlations with each other (r = 0.71–0.78, p < 0.01). These indices also showed similarly strong positive correlations with LST.

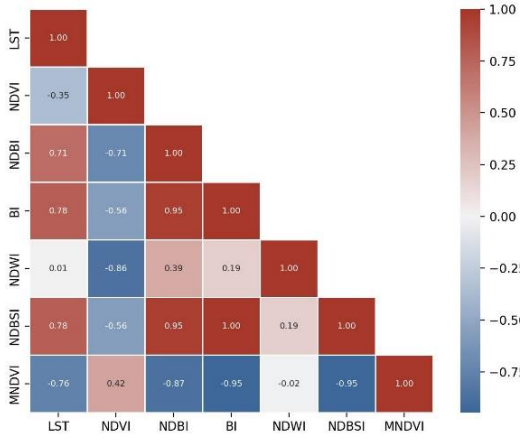

Figure 1. Correlation matrix of vegetation indices for the year 1994

Conversely, the MNDVI demonstrated a strong negative correlation, while the NDVI showed a weak negative association. It is important to note that Spearman's correlation coefficient indicates the

strength and direction of monotonic relationships but does not provide causal explanations for UHI processes. In Figure 2, the variable importance ranking within the spatial distribution model of the UHI indicates that the NDBSI and BI indices had the highest relative importance. In opposition to this, the MNDVI, NDBI, NDVI, and NDWI indices demonstrated lower significance within the model.

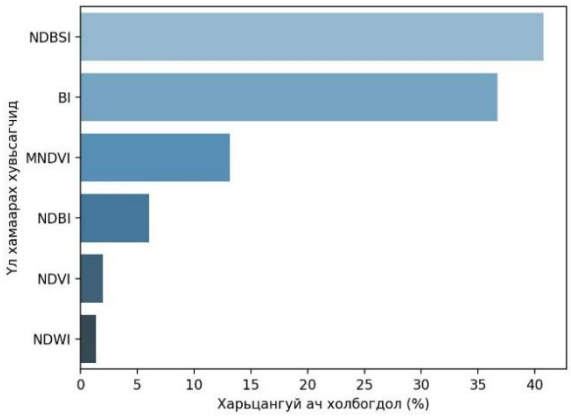

Figure 2. Variable importance ranking of vegetation indices in the RF prediction model

As shown in Figure 3, LST exhibits strong positive correlations with the BI, NDVSI, and NDBI indices ($r = 0.71–0.78$, $p < 0.01$), suggesting that these indices are critical for identifying impervious surfaces.

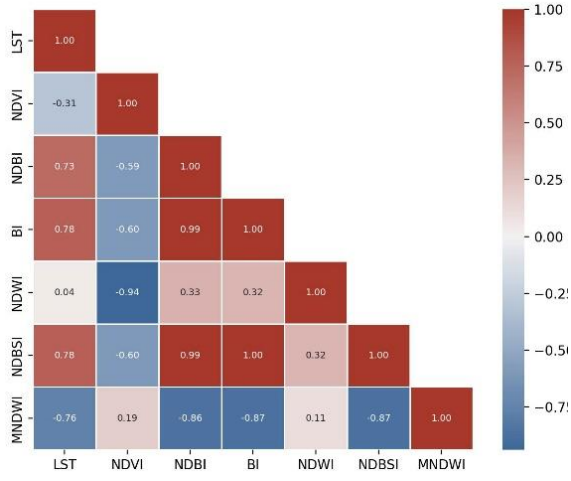

Figure 3. Correlation matrix of vegetation indices for the year 2024

Figure 4 shows that modeling the spatial distribution of the UHI in 2024, the XGB model outperformed the RF model. Within this modeling framework, the MNDWI and BI indices demonstrated high importance, while the NDVI, NDBI, and NDWI indices showed relatively low significance. Notably, the NDBSI index exhibited no measurable contribution to the model. Based on the modeled spatial distribution of UHIs in 1994 and 2024, both RF and XGB models indicate that LST is consistently higher over bare land and impervious surfaces, and lower in areas with significant green space area. In 1994, RF predicted LSTs ranging from 4.8°C to 16.4°C, while XGB produced a narrower range of 8.2°C to 14.4°C.

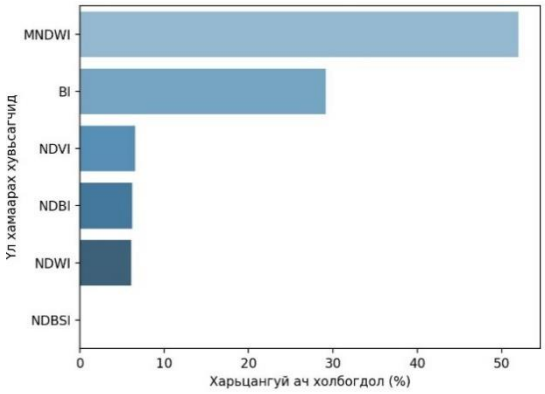

Figure 4. Variable importance ranking of vegetation indices in the XGB prediction model

For 2024, RF predictions ranged from 18.3°C to 29.5°C, whereas XGB estimated temperatures between 21.2°C and 25.7°C (Figure 5 and 6). These results underscore the growing intensity of UHI effects over time, particularly in the context of expanding impervious surfaces. Furthermore, the consistent observation of lower LSTs in vegetated areas highlights the critical role of green infrastructure in mitigating UHI intensity. This aligns with previous studies [16-17] that emphasize the cooling effect of urban vegetation and its contribution to climate resilience in rapidly urbanizing regions. UHI research has largely focused on developed countries where high-resolution data, extensive infrastructure, and computational resources are more readily available. For instance, Chakraborty et al. [18] and Debbage et al. [19] applied RF and Gradient Boosting Machines to predict urban thermal hotspots in the United States with high accuracy ($R^2 = 0.80–0.85$). Similarly, Voogt and Oke [20] employed traditional in-situ measurements and thermal cameras in Canada to quantify UHI, although their spatial coverage was limited. These studies demonstrate the effectiveness of ML models in UHI detection under optimal data conditions. In contrast, our study demonstrates the applicability of ensemble ML algorithms—RF and XGB—in a data-limited, cold-climate city using freely available Landsat satellite data. The models achieved

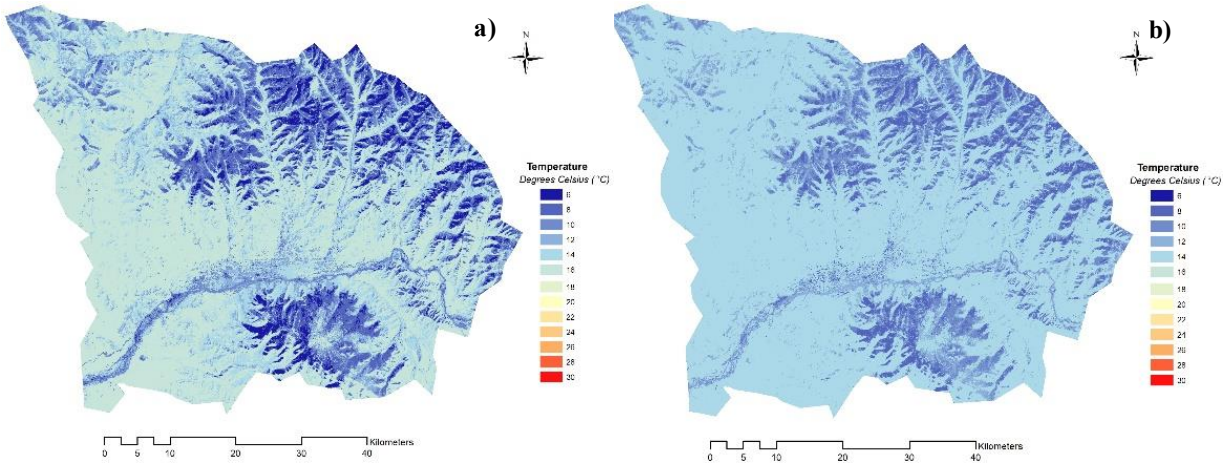

Figure 5. Predicted spatial distribution of the UHI in Ulaanbaatar for the year 1994 using (a) RF model and (b) XGB model

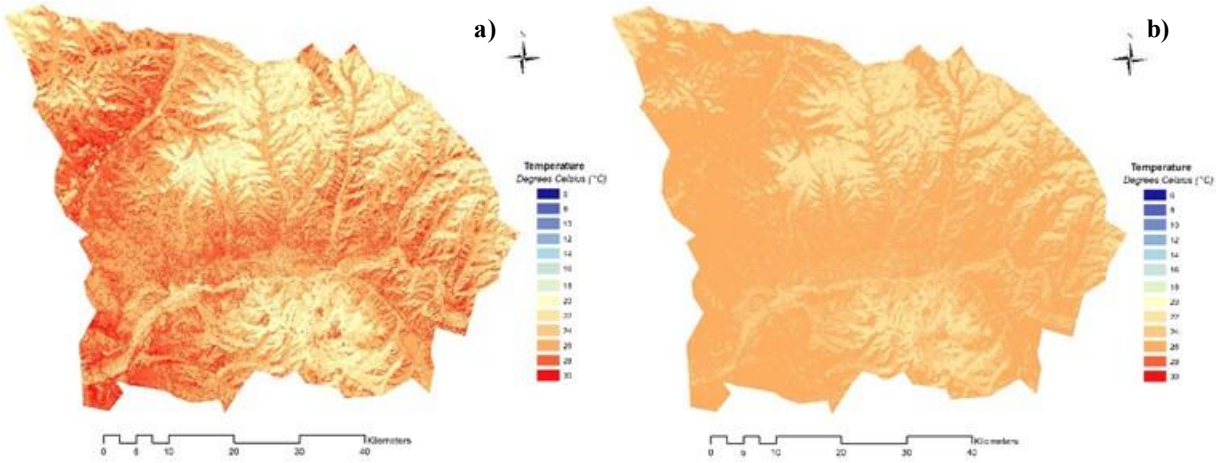

Figure 6. Predicted spatial distribution of the UHI in Ulaanbaatar for the year 2024 using (a) RF model and (b) XGB model

R^2 values of 0.72 (1994) and 0.75 (2024), with corresponding RMSE values of 1.89°C and 1.97°C, respectively. These results indicate a strong model fit and show that robust UHI modeling is achievable even under constraints typical of developed countries. The successful integration of open-source satellite imagery and ML techniques presents a replicable, cost-effective framework for UHI assessment in similar urban contexts. When compared with studies from developing countries, our results remain competitive. Hu et al. [21], using XGB in Wuhan, China, reported $R^2 = 0.81$ and RMSE = 1.4°C, slightly outperforming our model. Phan et al. [22] applied RF, SVM, and ANN in Ho Chi Minh City, Vietnam, achieving $R^2 = 0.78$ for RF. In Lahore, Pakistan, Tanoori et al. [23] implemented a hybrid RF–ANN model and obtained a high R^2 of 0.86, benefiting from the inclusion of socio-economic variables and high-resolution data. Although our model does not incorporate socio-economic or ancillary datasets, it still produces reasonably high accuracy solely from RS–derived indices, confirming the potential of RS-based ML modeling in resource-constrained settings.

4. CONCLUSION

This study investigated the spatial distribution and temporal evolution of the UHI phenomenon in Ulaanbaatar, Mongolia, using Landsat satellite imagery from 1994 and 2024, in conjunction with ML algorithms RF and XGB. Despite the data and resource constraints common in developing and cold-climate regions, both models demonstrated reliable predictive performance, with RF achieving $R^2 = 0.72$ in 1994 and XGB attaining $R^2 = 0.75$ in 2024.

The findings indicate a significant intensification of UHI effects over the past three decades, particularly over impervious surfaces such as bare soil and built-up areas, while vegetated zones consistently exhibited lower LSTs. Among the spectral indices analyzed, the

BI and NDBSI had the highest positive correlation with LST, highlighting the critical role of surface materials in influencing urban heat dynamics. Conversely, vegetation indices such as NDVI and MNDVI demonstrated cooling effects, emphasizing the importance of green infrastructure.

Compared to similar studies in both developed and developing countries, this research achieved comparable model accuracy using only open-access data, and contributed a rare long-term perspective on UHI dynamics in a high-altitude, cold urban environment. The successful application of RF and XGB models in this context demonstrates the potential of cost-effective and scalable ML-based approaches for urban climate assessments in data-limited regions. These findings offer valuable implications for urban planners and policy makers. The consistent identification of thermal hotspots and the mitigating role of vegetation suggest that the integration of green spaces into urban development strategies should be a top priority. Future research could benefit from incorporating socio-economic variables, higher-resolution imagery, or real-time climate data to further enhance model accuracy and policy relevance.

REFERENCES

[1] P. Hoffmann, O. Krueger, and K. H. Schlünzen, "A statistical model for the urban heat island and its application to a climate change scenario," *Int. J. Climatol*, vol. 32, no. 8, pp. 1238–1248, May 2011. Available: doi: 10.1002/joc.2348.

[2] X. Zhou and H. Chen, "Impact of urbanization-related land use land cover changes and urban morphology changes on the urban heat island phenomenon," *Sci. Total Environ*, vol. 635, pp. 1467–1476, Sep. 2018. Available: doi: 10.1016/j.scitotenv.2018.04.091.

[3] B. Y. Tam, W. A. Gough, and T. Mohsin, "The impact of urbanization and the urban heat island effect on day to day temperature variation," *Urban Clim.*, vol. 12, pp. 1–10, Jun. 2015. Available: doi: 10.1016/j.uclim.2014.12.004.

[4] S. Talukdar et al., "Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A Review," *Remote Sens*, vol. 12, no. 7, p. 1135, Apr. 2020. Available: doi: 10.3390/rs12071135.

[5] D. Sainbayar, "Modeling the spatial distribution of atmospheric carbon dioxide concentration" in Ph.D. dissertation, School of Science, National University of Mongolia, Ulaanbaatar, Mongolia, 2024, pp. 115.

[6] C. Yoo, J. Im, D. Cho, Y. Lee, D. Bae, and P. Sismanidis, "Downscaling MODIS nighttime land surface temperatures in urban areas using ASTER thermal data through local linear forest," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 110, pp. 102827, Jun. 2022,. Available: doi: 10.1016/j.jag.2022.102827.

[7] T. S. Arulananth et al., "Analysis of Reason to Global Warming Based on Heat Pattern Using Hyperspectral Imaging: Artificial Intelligence Application," *Remote Sens Earth Syst Sci*, vol. 7, pp 379–388, Sep. 2024, Available: doi: 10.1007/s41976-024-00130-2.

[8] National Statistics Office of Mongolia. (2024). Statistical yearbook 2024 pp. 40. [Online]. Available: https://www.nso.mn/mn/statistic/file-library/view/82183302.

[9] D. Amarsaikhan, A. Enkhmanlai, G. Tsogzol, A. Munkh-Erdene, E. Jargaldalai, D. Enkhjargal & B. Byambadolgor, "Urban land use change study in Ulaanbaatar city using RS and GIS," Journal of Institute of Mathematics and Digital Technology, 5(1), 40-49, 2023. Available: doi: 10.5564/jimdt.v5i1.3317

[10] R. C. Estoque and Y. Murayama, "Monitoring surface urban heat island formation in a tropical mountain city using Landsat data (1987–2015)," *ISPRS J. Photogramm. Remote Sens*., vol. 133, pp. 18–29, 2017. Available: doi: 10.1016/j.isprsjprs.2017.09.008.

[11] B. Byambadolgor, D. Amarsaikhan, "Mapping the spatial distribution of urban heat islands in Darhan and Erdenet cities," Mongolia, Full paper published in Proceeding of the ACRS, Ulaanbaatar, Mongolia, 2022.

[12] M. Y. Yasin, J. Abdullah, N. M. Noor, M. M. Yusoff, and N. M. Noor, "Landsat observation of urban growth and land use change using NDVI and NDBI analysis," IOP Conference Series: Earth and Environmental Science, vol. 1067, no. 1, p. 012037, Oct. 2022. Available: doi: 10.1088/1755-1315/1067/1/012037.

[13] K.S. Arunab and A. Mathew, "Exploring spatial machine learning techniques for improving land surface temperature prediction," *Kuwait Journal of Science*, vol. 51, no. 3, pp. 100242–

100242, May 2024. Available: doi: 10.1016/j.kjs.2024.100242.

[14] E. Nyamjargal, D. Amarsaikhan, A. Munkh-Erdene, V. Battsengel, and Ch. Bolorchuluun, "Object-based classification of mixed forest types in Mongolia," *Geocarto International*, vol. 35, no. 14, pp. 1615–1626, Jun 2019. Available: doi: 10.1080/10106049.2019.1583775.

[15] M. K. Uçar, M. Nour, H. Sindi, and K. Polat, "The Effect of Training and Testing Process on Machine Learning in Biomedical Datasets," *Mathematical Problems in Engineering*, vol. 2020, pp. 1–17, May 2020, Available: doi:10.1155/2020/2836236.

[16] D. Zhou, S. Zhao, S. Liu, L. Zhang, and C. Zhu, "Surface urban heat island in China's 32 major cities: Spatial patterns and drivers," Remote Sensing of Environment, vol. 152, pp. 51–61, Sep. 2014, doi: 10.1016/j.rse.2014.05.017.

[17] X. Li, Y. Zhou, G. R. Asrar, M. Imhoff, and X. Li, "The surface urban heat island response to urban expansion: A panel analysis for the conterminous United States," *Sci. Total Environ*, vol. 605–606, pp. 426–435, Dec. 2017, Available: doi: 10.1016/j.scitotenv.2017.06.229.

[18] S. Razzaghi Asl, "Rooftops for Whom? Some Environmental Justice Issues in Urban Green Roof Policies of Three North American Cities," *Environmental Policy and Law*, pp. 1–12, Dec. 2022. Available: doi: 10.3233/epl-220015.

[19] N. Debbage, C. Shepherd, M. Hall, and S. C. Cordero, "Integrating socio-environmental data in machine learning for urban heat island mapping," Computers, Environment and Urban Systems, vol. 83, p. 101521, 2020.

[20] J. A. Voogt and T. R. Oke, "Thermal remote sensing of urban climates," *Remote Sensing of Environment,* vol. 86, no. 3, pp. 370–384, Aug. 2003. Available: doi: 10.1016/s0034-4257(03)00079-8.

[21] Y. Hu, J. Zhang, and Y. Lin, "A machine learning approach for urban land surface temperature modeling: A case study in Wuhan, China," *Remote Sens*, vol. 12, no. 17, p. 2784, 2020.

[22] C. C. Pham, N. T. Nguyen, Q. T. Tran, and N. N. Nguyen, "Prediction of surface urban heat island in Ho Chi Minh City, Viet Nam using remote sensing and logistic regression model," *IOP Conference Series Earth and Environmental Science*, vol. 1349, no. 1, pp. 012031–012031, May 2024. Available: doi: 10.1088/1755-1315/1349/1/012031.

[23] Ghazaleh Tanoori, A. Soltani, and Atoosa Modiri, "Machine Learning for Urban Heat Island (UHI) Analysis: Predicting Land Surface Temperature (LST) in Urban Environments," *Urban clim*, vol. 55, pp. 101962–101962, May 2024. Available: doi: 10.1016/j.uclim.2024.101962.