UrbanReleaf: Enhancing Sustainable Urban Transformation with a Data-Driven Solution for Smart Depaving

Rafaela Morandi Santos[1], Lucas Sarmento[1], Mirela T. Cazzolato [2], Harlei Miguel de Arruda Leite[1]

[1]Instituto Tecnológico de Aeronáutica (ITA)
São José dos Campos – SP – Brazil

[2]Instituto de Ciências Matemáticas e de Computação (ICMC)
Universidade de São Paulo (USP) – São Carlos, SP – Brazil

{rafaela.santos.8932,lucas.sarmento.8956}@ga.ita.br
mirela@icmc.usp.br, harlei@ita.br

Abstract. *How does the proliferation of impervious surfaces affect the quality of life? Impermeabilization due to paving contributes to urban issues, such as increased temperatures, which form the so-called heat islands. It can also cause more intense and frequent floods, diminished biodiversity, and deterioration of air quality. In this context, depaving is a practical solution for increasing urban resilience. Given a large urban region covered by impervious surfaces, how can we select the best points for intervention? We propose the UrbanReleaf method to map excessively paved urban areas and simulate the environmental impact of implementing green infrastructure. UrbanReleaf identifies critical zones for nature-based interventions such as depaving and forecasts environmental outcomes, including surface temperature, soil moisture, and vegetation indices. The method achieves this by leveraging geospatial satellite imagery and machine learning regressors to analyze vegetation indices, land surface temperatures, and moisture content. Experimental results show that UrbanReleaf can support urban planners and policymakers with data-driven clues that can help mitigate problems caused by impermeabilized areas.*

1. Introduction

Urban expansion has led to an excessive proliferation of impervious surfaces, including asphalt and concrete. This problem is especially relevant in major cities worldwide, where impermeabilization contributes to environmental imbalance. This high degree of paving contributes to multiple urban issues, such as increased surface temperatures, more frequent and intense flooding, deterioration of air quality, and diminished biodiversity [Lee 2025]. The lack of vegetation and green infrastructure also affects public health. Elevated temperatures worsen respiratory and cardiovascular conditions, while limited access to natural spaces is linked to higher stress levels and reduced mental well-being [Zhang et al. 2012]. Green areas, on the other hand, have shown many benefits. They cool the urban microclimate, improve air quality by filtering pollutants, retain rainwater to prevent runoff and provide habitats for local fauna [Wen et al. 2025, Vedrí et al. 2025, Bowler et al. 2010]. In response, several cities around the world have invested in reintroducing green infrastructure through nature-based solutions.

Existing initiatives to address the problem of impermeable surfaces include converting underused paved areas into community parks, gardens, and rainwater-absorbing landscapes [Chan et al. 2018]. This practice, known as *depaving*, has gained traction as a practical and impactful strategy for increasing urban resilience [Nguyen and Park 2025, Huber et al. 2023, Meerow 2017]. However, despite its effectiveness, there remains a lack of tools to identify areas that are most suitable for intervention systematically. This work addresses this gap.

We propose UrbanReleaf, an automated, data-driven solution for mapping paved urban areas and evaluating the environmental impact of converting them into green spaces. By integrating remote sensing, geospatial analysis, and machine learning techniques, UrbanReleaf enables data-driven decision-making to support sustainable urban planning. Our method collects geo-referenced environmental indices, preprocesses the data, and predicts the Land Surface Temperature (LST) of a given region. This information is a powerful tool for guided intervention, enabling authorities to identify the most effective locations for depaving. The UrbanReleaf codebase is fully open-source at a GitHub repository[1], enabling replication of the methodology and adaptation to different cities and contexts.

Paper outline. Section 2 presents the relevant background and related work. Section 3 introduces the proposed method UrbanReleaf. Section 4 presents the experimental validation of our proposal. Finally, Section 5 gives the conclusion.

2. Background and Related Work

In this section, we present related work focused on enhancing depaving and the main concepts related to Regression models employed in our proposal.

2.1. Solutions for Enhanced Depaving

The environmental consequences of urban impermeabilization, such as heat islands and flooding, are well-documented [Gill et al. 2007]. However, the integration of scalable technological solutions to guide green infrastructure planning remains limited. Existing depaving initiatives, like Depave Portland [Depave 2025] and Sponge Cities in China [Chan et al. 2018], have demonstrated localized success, but their implementation often relies on manual, community-based decision-making or top-down planning without algorithmic prioritization.

Recent advances in remote sensing and open-access satellite data platforms, such as Sentinel Hub and Landsat, offer an unprecedented opportunity to monitor urban environmental indicators over time and at scale [Lima et al. 2025] [Vasconcelos et al. 2023]. These platforms provide datasets with valuable information, including vegetation indices such as Normalized Difference Vegetation Index (NDVI), Land Surface Temperature (LST), and moisture content (SWIR). Such information can support data-driven decision-making. However, the potential of Artificial Intelligence (AI) to extract actionable information from the provided is still underexplored in the context of urban ecological restoration [Pimenow et al. 2025]. There is a lack of AI-powered tools capable of integrating geospatial data and producing predictive models that estimate the environ-

[1]The source code is available in `github.com/LucsarR/UrbanReleaf`

mental outcomes of interventions, such as reductions in temperature or improvements in vegetation cover.

UrbanReleaf addresses the literature gap by developing a data-driven solution that can identify critical urban areas with high impermeability and thermal stress. The method simulates the expected environmental impact of converting detected critical urban zones into green infrastructure. As a result, UrbanReleaf supports evidence-based planning and offers municipalities and environmental organizations a scalable, replicable solution for sustainable urban management.

2.2. Regression Models

Machine Learning (ML) approaches for Regression can predict data based on historical observations. The problem can be formally defined as learning a mapping function f that predicts a continuous output variable y based on one or more input variables X. The goal is to approximate the function so that when new input data X' is provided, the predicted output y' is as close as possible to the actual value [Bishop 2006].

Among the most commonly employed methods are Linear Regression and the Stochastic Gradient Descent (SGD) Regressor. Linear Regression is a fundamental algorithm that models the relationship between input variables and output by fitting a linear equation to the observed data. It calculates the best-fitting line that minimizes the sum of the squared differences between the predicted values and the actual values.

The SGD Regressor is an efficient and versatile optimization approach that can be used for linear Regression and other models. Instead of using the entire dataset to calculate the error gradient at each step (as traditional methods do), it updates the model's parameters using only a single, randomly selected training example at a time. This makes it much faster and particularly useful for very large datasets. In our context, these regression models are crucial for predicting future outcomes, enabling proactive decision-making, and optimizing operational strategies.

To evaluate the prediction quality, we consider standard metrics such as the Mean Squared Error (MSE) and the coefficient of determination (R^2). MSE quantifies the average magnitude of the prediction error and is given by $MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$. Low values for MSE indicate a better fit. Additionally, the R^2 measures the proportion of the variance in the data that is explained by the model, calculated as $R^2 = 1 - [\sum(y_i - \hat{y}_i)^2 / \sum(y_i - \bar{y})^2]$. R^2 values close to 1 signify that the model has high explanatory power over the observed data [Han et al. 2011].

3. The Proposed Method: UrbanReleaf

UrbanReleaf is a modular pipeline designed to process remote sensing data and identify critical urban areas for green infrastructure interventions. The system automates the acquisition of satellite imagery, extracts key environmental indicators, such as vegetation density and surface temperature, and employs machine learning models to predict thermal behavior under varying vegetation scenarios. These predictions inform urban depaving strategies by highlighting locations where green space integration may yield the most significant environmental benefits.

Figure 1 illustrates the architecture of UrbanReleaf, consisting of five main stages. The diagram illustrates the sequential flow of data, starting with satellite image acquisition, followed by environmental index extraction (*e.g.*, NDVI, LST), preprocessing and data cleaning, predictive modeling using machine learning algorithms, and finally, the visualization and interpretation of results. This modular structure facilitates extensibility, making it easier to integrate additional indicators or scale to different regions.

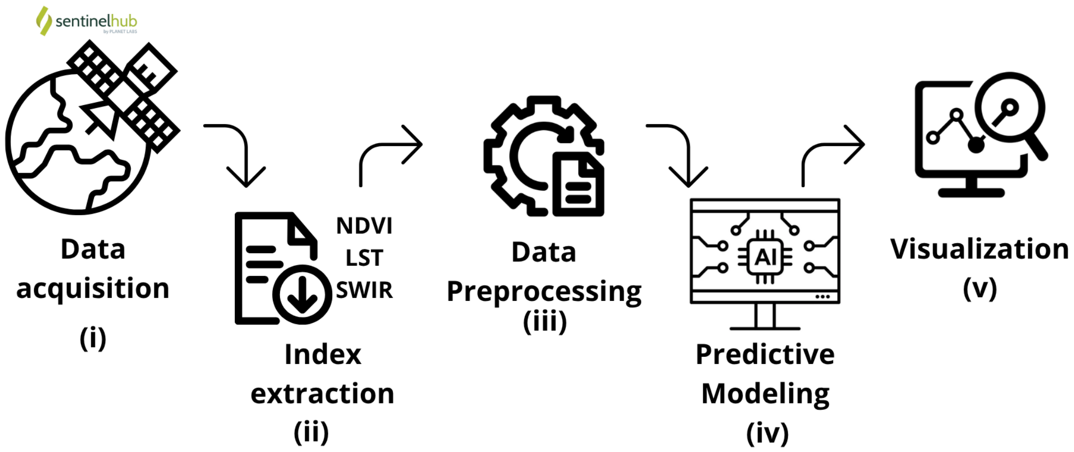

Figure 1. Overview of the UrbanReleaf system architecture.

First, (i) environmental data are acquired through the Sentinel Hub API [Sentinel Hub 2025], granting access to multi-temporal satellite imagery. The system focuses primarily on two sensors: Sentinel-2, used to obtain NDVI and SWIR data, and Landsat 8/9, used to extract Land Surface Temperature (LST). Bounding boxes define the geographic region of interest, and images are filtered based on seasonal variability and cloud coverage. Next, (ii) custom *evalscripts* are applied to generate environmental indices, including NDVI (Normalized Difference Vegetation Index), LST, and SWIR (Short-Wave Infrared). The resulting *.tiff* files are (iii) flattened into one-dimensional arrays, spatially aligned, and preprocessed. This step also includes the removal of missing values and outliers, followed by normalization using *StandardScaler* to enhance model convergence and interpretability.

The core predictive task is performed in stage (iv), and is framed as a Regression problem aimed at estimating surface temperature variations based on NDVI and other geospatial features. UrbanReleaf employs two baseline models: Linear Regression for transparency and interpretability and SGDRegressor to handle large datasets through incremental learning efficiently. The model's performance is evaluated using the Mean Squared Error (MSE) and the coefficient of determination (R^2). To support interpretation and decision-making, (v) the Regression outputs are visualized using scatter plots, loss curves, and heatmaps that compare predicted and actual temperatures. These visualizations not only enable technical validation but also facilitate communication with non-expert users, such as city planners and policymakers.

UrbanReleaf has a modular architecture that allows for the incorporation of additional features, such as carbon sequestration potential, air quality metrics, or socioeconomic indicators, thereby expanding the system's utility in sustainable urban planning.

4. Experimental Analysis

This section presents the experimental analysis carried to validate UrbanReleaf with data collected from a real-world scenario.

4.1. Setup

The UrbanReleaf pipeline was implemented in Python, using libraries such as *scikit-learn*, *rasterio*, and *matplotlib*. Environmental data were collected through the Sentinel Hub API, using *evalscripts* to extract vegetation index (NDVI) from Sentinel-2 imagery and land surface temperature (LST) from Landsat-8 imagery. The two evaluated regression models were the linear regression baseline and the Stochastic Gradient Descent (SGD) Regressor, both trained over 100 iterations. The prediction quality was evaluated using the Mean Squared Error (MSE) and the coefficient of determination (R^2) metrics.

4.2. Data Collection and Preprocessing

As a case study, the region of interest was defined as a bounding box over Curitiba, Brazil, and the selected imagery corresponds to the period of December 2023. The resulting *GeoTIFF* files were spatially aligned, flattened into one-dimensional arrays, and cleaned to remove invalid values. A feature matrix was then constructed with NDVI as the primary input and LST as the target. Features were normalized using the *StandardScaler* method, and the dataset was split into training and testing sets in an 80/20 ratio.

4.3. Results

In this subsection, we show a case study to validate UrbanReleaf. Figure 2 shows a side-by-side comparison between (a) the actual Land Surface Temperature (LST) map of Curitiba in December 2023, and (b) the predicted LST generated by the UrbanReleaf's model trained on NDVI data. The predicted map preserves the spatial temperature gradient and highlights the model's ability to generalize heat distribution patterns from vegetation coverage.

We observe that both the ground truth and predicted maps reveal that regions with lower NDVI values are systematically hotter (depicted as reddish areas). This behavior typically corresponds to highly urbanized or paved zones

The patterns confirm the phenomenon of urban heat islands and demonstrate that NDVI is a meaningful predictor of thermal behavior. On the other hand, greener areas (with higher NDVI values) consistently exhibit lower surface temperatures (depicted as bluish areas), reinforcing the environmental relevance of vegetation in regulating urban microclimates.

Similar studies have also reported consistent relationships between urban land cover patterns and surface temperature. For example, Stamou and Manika [Stamou and Manika 2013] analyzed remote sensing data for the cities of Thessaloniki and Volos, Greece, and identified a positive correlation between densely built-up areas and higher surface temperatures, as well as a negative correlation between green areas and surface temperature. Although the methods and geographic context differ, the results obtained in this study follow a similar trend, reinforcing that building density and vegetation presence are key factors in shaping the urban microclimate.

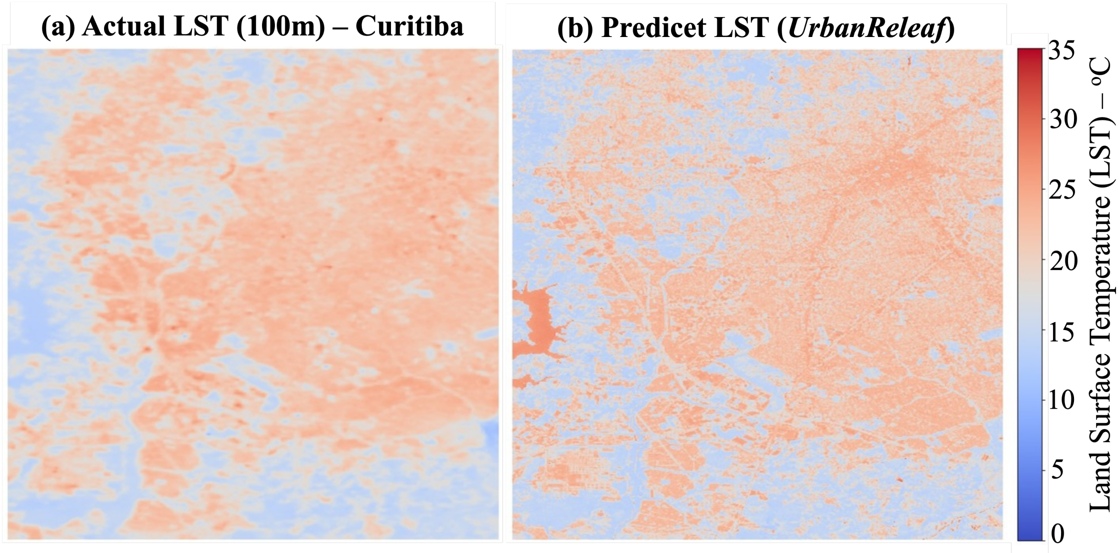

Figure 2. Comparison between Land Surface Temperature values. In (a), the map shows actual (measured) LST values (Ground Truth). The values were collected from Curitiba, Brazil, in December 2023. In (b), the map shows LST values predicted by UrbanReleaf. By comparing both maps we observe great similarity between actual and predicted LST.

The model achieved an R^2 of 0.41 in predicting LST from NDVI features, indicating moderate predictive capability. The greening intervention simulation suggested possible temperature reductions in selected areas of Curitiba.

This study constitutes a proof of concept, developed and evaluated using data from a single city (Curitiba) within a limited time frame. The predictive performance of the model, with an R^2 of 0.41, is modest, reflecting the inherent complexity of modeling land surface temperature solely from spectral indices. The results should therefore be interpreted with caution, as variations are likely when applying the methodology to other geographic contexts or temporal periods. Future work may benefit from more robust modeling strategies, such as ensemble learning or convolutional neural networks, in combination with additional predictors (e.g., socioeconomic, hydrological, or urban morphology data) to enhance accuracy and generalizability.

In the Brazilian context, UrbanReleaf could be directly integrated into municipal planning tools and public policy frameworks. For example, its predictive maps could support *Planos Diretores Municipais* by identifying priority zones for depaving and green space creation, guiding zoning changes toward sustainable land use. In flood-prone cities, the tool could be linked to *Planos Municipais de Redução de Riscos* to prioritize nature-based drainage solutions. For urban arborization programs, such as *Plano Municipal de Arborização Urbana* or São Paulo's *Programa Ambientes Verdes e Saudáveis* (PAVS), UrbanReleaf could help select planting locations with the greatest cooling and hydrological benefits. Furthermore, integration into platforms like the *Sistema Nacional de Informações sobre Meio Ambiente* (SINIMA) or municipal GIS dashboards would enable continuous monitoring of environmental indicators, allowing city officials to evaluate the effectiveness of interventions and adjust strategies accordingly.

5. Conclusions

This work proposes UrbanReleaf, a data-driven approach to assist in identifying urban areas where green infrastructure interventions, such as depaving, can have the most significant environmental impact. By leveraging open-access satellite imagery, geospatial analysis, and machine learning techniques, we demonstrated the ability to predict Land Surface Temperature (LST) variations based on vegetation indices.

In future work, we aim to expand the model by incorporating additional environmental indices, such as SWIR for soil moisture and air quality proxies, and explore Deep Learning architectures like Convolutional Neural Networks (CNNs) for spatial pattern recognition. We also plan to apply the methodology to other Brazilian cities, validate predictions with ground-truth meteorological data, and integrate stakeholder feedback from public institutions and NGOs.

The UrbanReleaf platform and codebase are open-source, encouraging community contributions and adaptation of the framework to different urban contexts.

Acknowledgments

We thank ITA/ITAEx/Nu Pagamentos S.A. for the financial assistance. We also thank the support of the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) – Finance Code 001, the São Paulo Research Foundation (FAPESP) – grants #2024/15430-5 and #2024/13328-9, the Brazilian National Council for Scientific and Technological Development (CNPq), and the University of São Paulo (PRPI Ordinance No. 1032) – grant #207.

References

Bishop, C. M. (2006). Pattern recognition and machine learning. *Springer google schola*, 2:1122–1128. ISBN: 0387310738.

Bowler, D. E., Buyung-Ali, L., Knight, T. M., and Pullin, A. S. (2010). Urban greening to cool towns and cities: A systematic review of the empirical evidence. *Landscape and Urban Planning*, 97(3):147–155. DOI: 10.1016/j.landurbplan.2010.05.006.

Chan, F., Griffiths, J., Higgitt, D., Xu, S., Zhu, F., Tang, Y.-T., Xu, Y., and Thorne, C. (2018). "sponge city" in china—a breakthrough of planning and flood risk management in the urban context. *Land Use Policy*, In Press. DOI: 10.1016/j.landusepol.2018.03.005.

Depave (2025). Depave: Urban re-greening and community revitalization. `https://www.depave.org/`. Accessed: 2025-07-06.

Gill, S., Handley, J., Ennos, R., and Pauleit, S. (2007). Adapting cities for climate change: The role of the green infrastructure. *Built Environment*, 33:115–133. DOI: 10.2148/benv.33.1.115.

Han, J., Kamber, M., and Pei, J. (2011). *Data Mining: Concepts and Techniques, 3rd edition*. Morgan Kaufmann. ISBN: 978-0123814791.

Huber, M., Kumar, V., Steele-Dunne, S. C., and Rommen, B. (2023). Sentinel-1 insar coherence as an indicator of monitor farming activities. In *IEEE International*

Geoscience and Remote Sensing Symposium, IGARSS 2023, Pasadena, CA, USA, July 16-21, 2023, pages 429–432. DOI: 10.1109/IGARSS52108.2023.10281522.

Lee, J. (2025). Estimating near-surface air temperature from satellite-derived land surface temperature using temporal deep learning: A comparative analysis. *IEEE Access*, 13:28935–28945. DOI: 10.1109/ACCESS.2025.3539581.

Lima, L. B., Franca Rocha, W. J., Souza, D. T., Lobao, J. S., de Santana, M. M., Cambui, E. C., and Vasconcelos, R. N. (2025). Urban quality: A remote-sensing-perspective review. *Urban Science*, 9(2):31. DOI: 10.3390/urbansci9020031.

Meerow, S. (2017). Spatial planning for multifunctional green infrastructure: Growing resilience in detroit. *Landscape and Urban Planning*, 159:62–75. DOI: 10.1016/j.landurbplan.2016.10.005.

Nguyen, K. and Park, C. J. (2025). On calibration of prompt learning using temperature scaling. *IEEE Access*, 13:31171–31182. DOI: 10.1109/ACCESS.2025.3538617.

Pimenow, S., Pimenowa, O., Prus, P., and Niklas, A. (2025). The impact of artificial intelligence on the sustainability of regional ecosystems: Current challenges and future prospects. *Sustainability*, 17(11):4795. DOI: 10.3390/su17114795.

Sentinel Hub (2025). Sentinel hub api documentation. `https://www.sentinel-hub.com/develop/api/`. Accessed: 2025-07-06.

Stamou, A. and Manika, S. (2013). Estimation of land surface temperature and urban patterns relationship for urban heat island studies.

Vasconcelos, F. F., Ramos, V. T., and Coutinho, F. J. (2023). Os desafios e soluções para a implementação de big data analytics em cidades inteligentes. In *Simpósio Brasileiro de Banco de Dados (SBBD)*, pages 50–56. SBC. DOI: 10.5753/sbbd_estendido.2023.233368.

Vedrí, J., Niclòs, R., Pérez-Planells, L., Valor, E., Luna, Y., and Estrela, M. J. (2025). Empirical methods to determine surface air temperature from satellite-retrieved data. *Int. J. Appl. Earth Obs. Geoinformation*, 136:104380. DOI: doi.org/10.1016/j.jag.2025.104380.

Wen, Z., Zhuo, L., Gao, M., and Han, D. (2025). How can we improve data integration to enhance urban air temperature estimations? *Int. J. Appl. Earth Obs. Geoinformation*, 140:104599. DOI: 10.1016/j.jag.2025.104599.

Zhang, B., Xie, G., Zhang, C., and Zhang, J. (2012). The economic benefits of rainwater-runoff reduction by urban green spaces: A case study in beijing, china. *Journal of environmental management*, 100:65–71. DOI: 10.1016/j.jenvman.2012.01.015.