Remote Sensing Applications: Society and Environment 41 (2026) 101889

Contents lists available at ScienceDirect

Remote Sensing Applications: Society and Environment

journal homepage: www.elsevier.com/locate/rsase

Urban local climate zone classification through deep learning using spatio-temporal thermal imagery

Michaja van Capel [a], Azarakhsh Rafiee [a,*], Roderik Lindenbergh [b]

[a] Architectural Engineering + Technology Department, Delft University of Technology, Delft, the Netherlands
[b] Department Geoscience & Remote Sensing, Delft University of Technology, Delft, the Netherlands

ARTICLE INFO

Keywords:
Urban local climate zone
Thermal satellite imagery
Deep learning
Spatio-temporal analysis
Classification
Thermal signature

ABSTRACT

Rapid urbanization challenges urban micro-climates, strains resources and affects public health. Understanding micro-climate dynamics is key to effective mitigation and sustainable development. Local Climate Zone (LCZ) classification supports climate-resilient planning but is complicated by the diversity and complexity of diverse urban landscapes and the coexistence of varying land uses and materials within small areas. While LCZ classification typically uses multispectral imagery, LiDAR, and land-use data, these sources often miss temporal thermal dynamic patterns. Thermal satellite imagery improves LCZ classification by distinguishing zones with similar structures but differing thermal behavior. This research proposes using deep learning-based multitemporal semantic segmentation to classify urban LCZs based solely on temporal thermal patterns from ECOSTRESS satellite imagery. The methodology is applied in a in a case study around the near coastal cities of Rotterdam and The Hague in The Netherlands and demonstrates how spatial and temporal factors (both diurnal and seasonal) influence the performance of the semantic segmentation model on different LCZ classes. The study shows that a U-Net architecture applied on spatio-temporal thermal imagery effectively classifies urban LCZs, achieving a test accuracy of 0.75. Temporal factors significantly impact model performance, with higher accuracies observed for daytime (0.8) and Spring/Summer imagery (0.78), as these conditions provide clearer thermal separability for distinguishing LCZs. The model achieved its highest test accuracy (0.83) when trained and tested on thermal images with the highest LST values. This suggests that focusing on high-value LST images with sufficient variability enhances classification performance compared to a generalized approach using the full dataset.

1. Introduction

In recent decades, urbanization has been rapidly transforming the global landscape, with a significant proportion of the world's population now living in urban areas (Zhang et al., 2022). Alongside, climate change is expected to increase both average global temperature as well as the amount of extreme weather events. Urban growth has led to numerous challenges, including negative effects on urban micro-climate and the well-being of urban inhabitants. Understanding and characterizing urban climate is crucial for mitigating these challenges as well as for creating and maintaining liveable urban environments (Ren et al., 2022). To this end, the concept of Local Climate Zones (LCZs) has emerged as a valuable framework for urban climate classification and analysis. In 2012, the

* Corresponding author.
E-mail addresses: michajavancapel@gmail.com (M. van Capel), a.rafiee@tudelft.nl (A. Rafiee), R.C.Lindenbergh@tudelft.nl (R. Lindenbergh).

https://doi.org/10.1016/j.rsase.2026.101889
Received 27 November 2025; Received in revised form 30 December 2025; Accepted 19 January 2026
Available online 22 January 2026

LCZ classification system was introduced by Stewart and Oke (2012). Before, climate classifications were typically formulated to describe climate zones at larger scales, making them ineffective when applied to smaller, micro-scale areas. Sites in cities with very different physical and climatological features were usually described only as "urban" or "rural". The LCZ system aims to overcome this (Aslam and Rana, 2022). Within this system, there are 17 distinct zones, each characterized by its unique combination of surface structure, cover, and human activity. By considering these factors, the LCZ system provides a more accurate and detailed representation of the climate within specific areas. LCZ 1–10 are different built-up classes, and LCZ A-G different land cover types. The different LCZs and their definition can be found in (Stewart and Oke, 2012).

Urban Local Climate Zones (LCZs) are crucial for understanding, analysing, and mitigating the complex interactions between urbanization and climate. LCZs enable researchers, planners, and policymakers to assess the impact of urban development on local and regional climate. LCZs are particularly important for studying the Urban Heat Island (UHI) effect, air quality dynamics, and energy consumption patterns (Shi et al., 2019; Du et al., 2023; Anjos et al., 2020), offering valuable insights for creating more sustainable and climate-resilient cities. Moreover, the LCZ framework supports global comparative studies, allowing for consistent methodologies in assessing urbanization impacts across diverse geographic and climatic contexts. Its applications extend to urban planning, climate modelling, and risk management, where LCZs help identify areas vulnerable to heat stress or flooding and guide intervention scenario analyses such as urban greening, use of reflective materials, or optimized building layouts. Furthermore, LCZ has been used for efficient population estimation, due to the robust relation between LCZ and population density (Ma et al., 2024). As cities continue to expand, the importance of LCZs in fostering sustainable urban environments becomes increasingly important. LCZ classification not only enhances our understanding of urban heat island effects but also supports urban planners and policymakers in designing interventions that promote thermal comfort and energy efficiency (Perera and Emmanuel, 2018; Lau et al., 2019; Ren et al., 2022). Furthermore, the LCZ framework has shown its potential in urban climate modelling (Bechtel et al., 2019), enabling researchers to integrate localized environmental data into broader climate simulations, thereby improving the accuracy of climate impact assessments.

1.1. LCZ and land surface temperature

Land Surface Temperature (LST) represents the radiant skin temperature of the Earth's land surface, as determined by the absorption, reflection and emission of solar radiation (Khan et al., 2021). Each LCZ type influences LST through its distinct physical characteristics. The composition and arrangement of land cover types directly influence the distribution of energy absorbed and emitted at the land surface, thereby establishing a correlation with LST variation. Understanding these correlations has important potential for managing urban heat islands, improving urban microclimates, addressing climate change impacts, and assessing environmental and ecological aspects of urban areas. The research by Cilek and Cilek (2021) shows differences in the mean LST values and differences per LCZ in Adana City, Turkey. For example: LCZs with high proportions of impervious surfaces, such as urban centers or compact high-rise areas, tend to have higher LSTs due to the absorption and retention of solar radiation by buildings and pavement. LCZs with more vegetation, such as parks or forests, generally have lower LSTs due to the cooling effect of vegetation through evapotranspiration and shading. Some LCZs do not show considerable differences in LST values when using one single thermal image (Lotfian et al., 2019). Another study by (Zhao et al., 2021) shows differences in seasonal LST variabilities per LCZ. A general trend is that the diurnal LST variation increases with the urbanization index (Chen et al., 2017).

Despite the observed differences between LCZs and thermal patterns, these studies consistently conclude that further information and more detailed investigations are necessary before the relationship between LCZs and LST can be fully understood and effectively used. This gap arises due to several factors. First, some LCZs do not show considerable differences in LST values, highlighting the complexity of thermal behavior within and between different urban areas. Second, studies are often limited to one city, which restricts the generalizability of findings across different urban contexts. Lastly, many studies rely on one or a few thermal images, which may not capture the full variability of temperature patterns over time.

1.2. LCZ classification

Classification of LCZs presents several challenges, primarily due to the complexity and variability of urban environments. One significant challenge is the accurate mapping and characterization of zones in highly heterogeneous cities where diverse land uses, building morphologies, and surface materials often coexist within small distance. Additionally, subjective interpretation of LCZ criteria, such as vegetation density or building height, can introduce inconsistencies. Temporal changes, such as seasonal variations in vegetation cover or ongoing urban development, also pose challenges in maintaining classification updates.

For LCZ classification, mainly multispectral satellite imagery, sometimes in combination with LiDAR data and land-use/land-cover maps, is used. For instance, Demuzere et al. (2019) mapped Europe into LCZs using techniques developed as part of the World Urban Database and Access Portal Tools (WUDAPT) project. The Random Forest classifier, as implemented in Google Earth Engine was used for training, enabling a LCZ classification based on Landsat satellite imagery. While the application of this data, especially if well integrated with Machine Learning techniques, has presented good results (Jing et al., 2019; Zhao et al., 2019; He et al., 2023), they cannot fully capture the temporal thermal dynamics and seasonal changes critical for fully characterizing LCZs. In addition to pixel-based classification, object-based classification methods have also been developed for LCZ classification. Yan et al. (2022) have systematically compared object-based and pixel-based LCZ mapping methods using a consistent set of multi-source input data to evaluate which paradigm yields higher classification accuracy. The study found that the considered object-based method generally outperformed pixel-based approaches in overall accuracy (with ~2–5 % higher OA) and notably showed advantages in classifying land

cover types involving building and also highlighted the importance of features such as building height and urban morphological parameters in the object-based framework. Ma et al. (2023) have improved LCZ mapping by enhancing the object-based image analysis (OBIA) framework with region-level features and urban morphological parameters such as sky view factor, building surface fraction, and permeable surface fraction. This approach was tested across three cities and demonstrated substantial improvements, particularly for built LCZ types, compared to conventional OBIA methods.

Through incorporating both multi-temporal thermal imagery, and spatial relations, LCZ classification methods can potentially achieve higher accuracy and objectivity in defining LCZs. While thermal imagery can directly capture fine-scale temperature variations within urban areas (Zhao et al., 2021), LCZ classifications—derived from multi-spectral satellite imagery—reflect urban morphology and may not accurately represent updated or localized microclimatic conditions. Furthermore, in highly heterogeneous and dynamic urban environments, overlapping LCZ types or mixed-use zones can challenge classification accuracy, even with detailed datasets. Multi-temporal thermal satellite imagery can significantly enhance the classification accuracy of urban LCZs through an improved differentiation of zones with similar morphological features but distinct thermal properties. For instance, Compact Low-Rise (LCZ 3) and Light Industry (LCZ 8) zones have morphological similarities in building height, building density and vegetation cover. However, their thermal behaviour is different due to different material properties, different anthropogenic heat emissions, roof characteristics and surface functionalities. Furthermore, thermal imagery is effective at capturing the effects of anthropogenic heat emissions, such as industrial activities or densely populated zones with high energy usage. This information is crucial for identifying zones with elevated nighttime temperatures and understanding urban heat island (UHI) dynamics.

While in previous LCZ-LST studies only one snapshot, or diurnal, seasonal and annual patterns were analyzed (Liu et al., 2018), this research aims to explore the suitability of multitemporal deep learning based semantic segmentation to harness the potential of multi-temporal thermal imagery for LCZ classification. In this approach, stacks of thermal images will serve as input data, enabling the model to effectively capture temporal relations while maintaining spatial coherence. The semantic segmentation component facilitates the contextual understanding of the spatial thermal patterns, incorporating the spatial urban configuration. The multi-temporal incorporates the temporal (day, night, seasonal) thermal behaviour. This spatio-temporal integration enables a multi-aspect thermal behaviour analysis which can significantly enhance urban LCZ classification. The impact of spatial (semantic segmentation) and temporal components are investigated in this study through different experiments.

2. Data

2.1. Study area and time span selection

The study area and time span of the data area are selected having enough spatial extent to accommodate a substantial number of patches for the Deep Learning model. However, it should not be excessively large to ensure frequent coverage within the satellite sensor's field of view, facilitating comparable LST measurements across the entire area. Moreover, the study area should exhibit diversity in terms of LCZs to enable the differentiation of various LCZs. The selected study area is shown in Fig. 1. It ranges from Amsterdam to Rotterdam and therefore yields several large urban areas with varying distance from the coastline.

2.2. Data collection

This study makes use of ECOSTRESS thermal infrared imagery (Land Surface Temperature, level 2 product), which is open-source and has frequent coverage (multiple times per week for the Netherlands). ECOSTRESS, mounted on the International Space Station (ISS), provides thermal imagery with a spatial resolution of 70 m and a high temporal resolution, capturing data at different times of the day due to the orbit of ISS, which is particularly beneficial for monitoring diurnal temperature variations (Hook et al., 2019). An

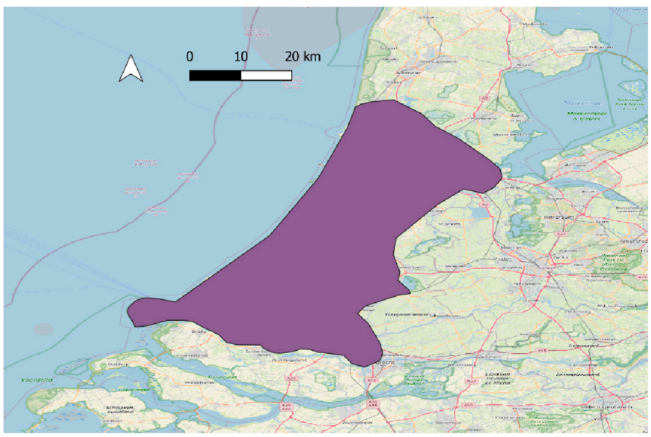


Fig. 1. Study area incorporating the cities of Amsterdam, Rotterdam and The Hague in the Western part of The Netherlands.

example of ECOSTRESS thermal imagery (Land Surface Temperature), applied in this study is depicted in Fig. 2.

The imagery is accessed using the online open source tool: Application for Extracting and Exploring Analysis Ready Samples (AppEARS), that was developed by NASA (EARTHDATA, 2015). The data is downloaded by the selected study area and time span and manually processed. The following files are excluded from the final dataset:

- Thermal images that do not cover the entire study area. For every pixel the LST value at every measuring time is desired, to train the model with the same amount of input parameters per pixel.
- Thermal images with clouds. The thermal sensor cannot see through clouds. The study area is a coastal area where clouds are common.
- Thermal images with missing data. The data of ECOSTRESS is transferred in data bundles. Occasionally, a single data bundle is corrupted as it is transferred from the instrument to the ground data system. This can result in small patches of missing data within a file.
- Images with other artefacts. For example, occasionally the ISS must adjust the position of some of its solar panel arrays, these may pass into the ECOSTRESS field of view. This results in obstructions in the resulting images.

3. Methodology

Fig. 3 presents the methodological framework of this study.

The components of the methodological framework are described as follows.

3.1. Data pre-processing

3.1.1. Multi-temporal image stack

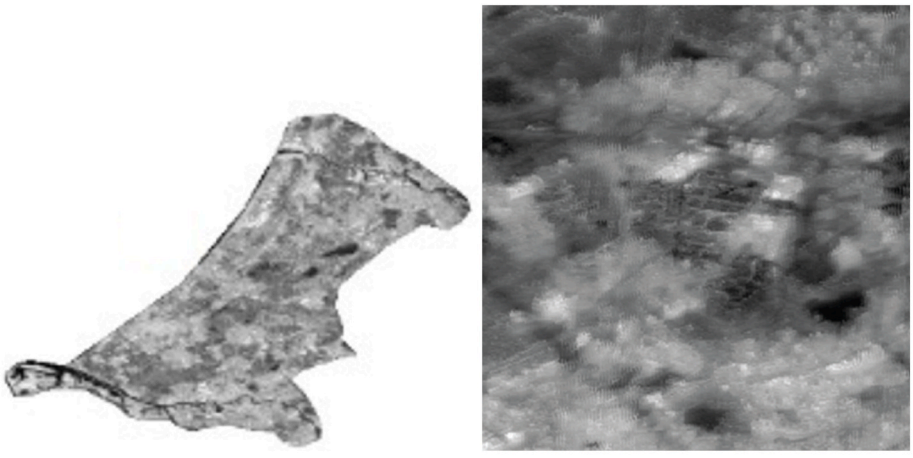
To incorporate multi-temporal thermal datasets into the deep learning model, they were aligned and stacked into one common raster with different time step values given as different band values. Every pixel has a vector of LST values from each of the input thermal images. Different stacks are created for different experiments.

3.1.2. Data normalization

Subsequent to raster stacking, normalization is applied to each band independently to enhance the speed of model convergence. The normalization was performed using linear scaling to re-map input values to a range between 0 and 1. The scaling function used is:

$$val_{out} = (val_{in} - c)\left(\frac{b-a}{d-c}\right) \quad \text{(Equation 1)}$$

Here val represents the original pixel index and val_{out} the normalized pixel value, while a and b are the lower and upper values of the desired range, which are set to 0 and 1. The lower and upper values of the original range are denoted by c and d respectively, which are set to the minimum and maximum of each band (Kadunc, 2022).



a b

Fig. 2. Example of ECOSTRESS thermal imagery of a) the study area and b) zoomed-in.

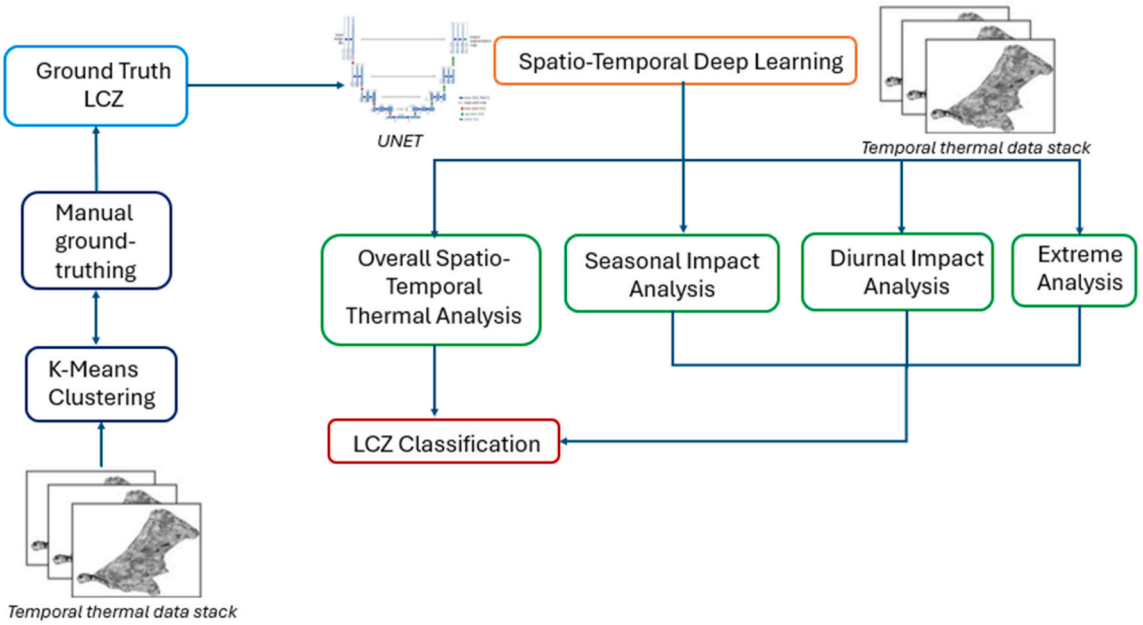


Fig. 3. Methodological framework, including overall spatio-temporal thermal analysis for LCZ classification, as well as experiments for the seasonal, diurnal and extreme data impact analyses on LCZ classification performance.

3.2. Training

In this study, labelling training data is challenging, due to lack of existing ground truth data. Since there is not a unified or standard LCZ map, labelling of the training data is implemented through an unsupervised clustering approach followed by manual ground truthing. That is, ground truth LCZs are generated based on thermal behaviour, using unsupervised clustering, by grouping similar thermal time series together and discovering underlying patterns (Aghabozorgi et al., 2015). This is a common approach, first discussed in (Lasserre et al., 2006), and since then applied for various classification problems with a lack of ground truth data, including remote sensing problems (Wen et al., 2025). In a first step, weak labels are derived using an unsupervised classification method (i.e. clustering), followed by a quality control step, which could consist of manual inspection of clustered results (adapted here), or automatically using relatively pure cluster samples, that is, samples close to the mean or majority of the data points in a cluster. In a second step, the labels obtained in the first step are fed to a supervised classification method. The quality control step ensures that labels used in the second step are indeed representative for typical thermal behaviour, and do not belong to outliers. For unsupervised clustering, The K-means clustering algorithm is applied to the pre-processed LST time series data. The choice of the number of clusters is initially based on existing LCZ classifications, followed by experimental adjustments. Following the application of K-means, ISODATA is used to further refine the clustering results, which adapts the number of clusters based on both the variability between clusters and the inter-cluster variability, thereby enhancing the flexibility and robustness of the clustering process. To ensure the reliability of the ground truth data, the clusters, asgenerated by K-means and refined by ISODATA, are manually inspected and suitable training data is selected accordingly.

3.3. Data splitting

The dataset was divided into training, test, and validation sets with a 70/15/15 split. This higher percentage for test and validation data, compared to the more common 10 % in literature, was chosen due to the expected large number of classes. Ensuring that all classes are present in the test dataset makes it representative of the study area. A high-quality test dataset is crucial for evaluating the effectiveness of the classification model, as it provides a reliable basis for comparing predictions (Bai et al., 2021).

3.4. U-net architecture

The U-Net architecture employed in this research is adapted from Bhatia (2021) based on the original U-Net architecture by Ronneberger et al. (2015). The model is composed of five encoder blocks and five decoder blocks using ReLU as its activation function (Bishop and Bishop, 2023). For training, the model makes use of the Adam optimizer. The U-Net trains the detection of the classes present in the ground truth data in the training data set for a fixed number of epochs. The validation batches provide unbiased insights into training progress. First, the model outputs a class membership probability per class per pixel of the test data set. The pixel will be assigned to the class with the largest probability. The final classification outputted by the model is a multi-class LCZ prediction. The

tuning of the hyperparameters batch size, loss function and learning rate has been performed by experimentation and model performance evaluation as in Durrani et al. (2023).

3.5. Evaluation

In order to evaluate the quality of the segmentation outputs, thermal images from the test datasets are provided to the UNet and the predicted masks are compared to the ground truth labels of these images. For the evaluation of this multi-class classification, we have employed F1-score per class. The F1-score is a metric that combines both recall and precision as a single value. It is computed as:

$$F1 = \frac{2 \times TP}{2 \times TP + FP + FN} \quad \text{(Equation 2)}$$

using True Positives (TP), False Positives (FP) and False Negatives (FN).

Additionally, the macro F1-score metric is also computed. The F1-score metric was developed for single-label information retrieval, but there are variants of the F1-score for the multi-class models. Macro F1 calculates the F1 score for each class independently and then averages these scores, giving equal weight to each class regardless of its size (Zhang et al., 2015).

$$MacroF1 = \frac{1}{N} \sum_{i=1}^{N} F1_i \quad \text{(Equation 3)}$$

3.6. Experimental settings for temporal and extreme LST value impact on model performance

Evaluating the impact of seasonal and diurnal variations in temporal thermal satellite data is important for accurately assessing their influence on LCZ classification performance. Seasonal patterns capturing long-term shifts in surface temperature are key indicators of broader climatic regimes, while diurnal behavior reflects short-term temperature fluctuations. By isolating these two temporal components, it becomes possible to identify which contributes more significantly to classification accuracy. Furthermore, evaluating classification performance using thermal imagery of peak temperature periods supports exploring whether, and in which temporal resolution, extreme heat conditions enhance the differentiation of LCZ classes. These experiments aid in optimizing the temporal resolution and sampling strategies for satellite data used in climate-related applications.

In this research, we have performed several experiments to explore the impact of seasonal, diurnal and extreme high LST values on the model performance on each LCZ class. The first experiment employs the full temporal dataset in the semantic segmentation model. The outcome of this experiment is used as the reference for other experiments. The *seasonal* experiment comprises separate training and testing of the deep learning semantic segmentation model for Spring/Summer versus Autumn/Winter thermal satellite imagery. The *diurnal* experiment includes separate training and testing for daytime versus nighttime thermal satellite data. The e*xtreme* experiment comprises different (sub) experiments on separate training and testing for varying numbers of the employed thermal imagery during peak temperature periods.

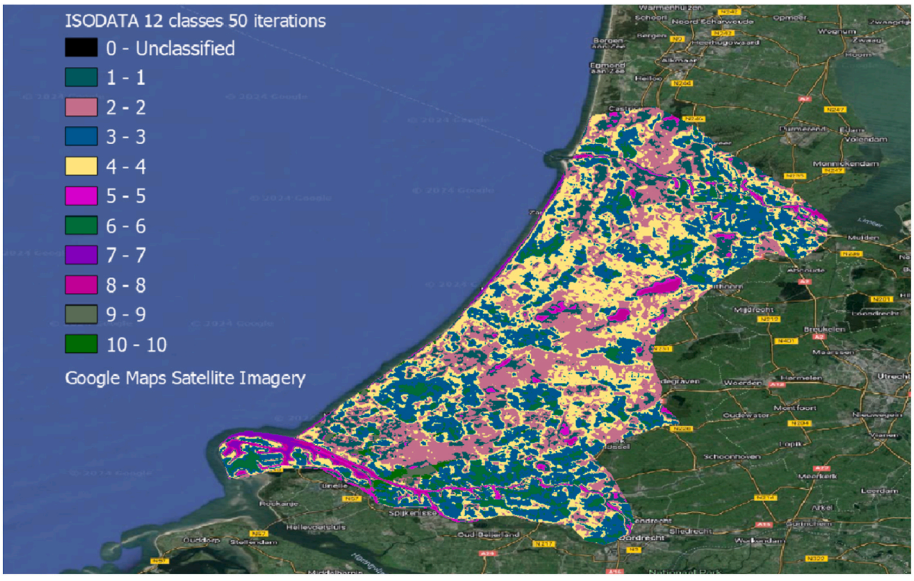


Fig. 4. ISODATA clustering results obtained from clustering pixel wise time series built from 46 Land Surface Temperature Ecostress images from 2021, 2022 and 2023.

4. Results and discussions

4.1. Training dataset

The LST data of 3 years (2021, 2022 and 2023 respectively) was used for clustering. From this time span, 46 useful images were selected, excluding the files described in Section 2. The ISODATA included 12 classes, 50 iterations, maximum standard deviation 0.0001 and minimum class size 10 pixels. This ISODATA result was used to create the training data set (Fig. 4). Although the number of input classes was set to 12, the algorithm produced 9 distinguishable classes, as class 9 and 10 both represent greenhouses. The areas covered by the classes were analyzed with aerial imagery from Google Maps, and labelled according to the observed dominant land cover. The class descriptions and statistics are depicted in Table 1. Note that class distribution is not balanced as some classes contain significantly more pixels than others.

Our class definitions are adapted to the dominant land-cover and urban morphology characteristics present in the Rotterdam–The Hague region, which leads to a slight difference with LCZ taxonomy (LCZ 1–10, A–G), proposed by Stewart and Oke (2012). In terms of correspondence, our forest/meadow classes (Classes 1 and 2) align most closely with LCZ A (dense trees) and LCZ B/D (scattered trees and low plants), differentiated primarily by vegetation density rather than structural height. The residential class (Class 3) broadly corresponds to open low-rise and mid-rise built types (LCZ 5 and LCZ 6), while the residential with green space class (Class 4) is most similar to LCZ 6 (open low-rise) with a higher pervious surface fraction and vegetation cover. The city centre/industrial class (Class 6) aggregates characteristics of compact mid-rise/low-rise built types (LCZ 2–3) and large low-rise or heavy industry (LCZ 8–10), reflecting their similar thermal behavior at ECOSTRESS resolution. Water-related classes (Classes 5, 7, and 8) correspond directly to LCZ G (water) but are subdivided based on depth and thermal dynamics, which are particularly relevant for coastal and estuarine environments.

Due to the relationship between LCZ and surface structure, we have performed statistical analyses of morphological indices (building height and building density) for each class. Table 2 presents the mean building density and height for each class and Fig. 5 presents the building density distribution for each class.

Fig. 6 present the building height distribution per class.

4.2. Hyperparameter configuration

A set of hyperparameters—including learning rate, patch size, and loss function—was evaluated to identify the optimal configuration for the U-Net model. Different values for the learning rate and patch size were tested, along with various loss functions. Each adjustment was made individually, with all other parameters held constant.

4.2.1. Learning rate

The model's performance in terms of loss and accuracy are evaluated for learning rates values of 0.1, 0.01, 0.001, and 0.0001 to tune the learning process of the Adam optimizer. Accuracy, often referred to as overall accuracy (OA), represents the percentage of pixels that are correctly classified out of the total number of pixels. It is calculated by dividing the number of true positives (TP) by the total number of pixels. The loss function used in all runs is *Sparse Categorical Cross entropy* and the patch size 64. The results are presented in Fig. 7.

A learning rate of 0.001 results in the best model performance. This learning rate was selected because it was the only one for which the loss function converged to a stable value within 100 epochs. Additionally, the test accuracy achieved with a learning rate of 0.001 was the highest among all tested values indicating that the model not only learned effectively but also generalized well to new data. The other learning rates either caused significant fluctuations in the loss function or resulted in slow progress and sub-optimal accuracy.

4.2.2. Patch size

The model's performance in terms of loss and accuracy for patch sizes of 64, 128 and 256 are presented in Fig. 8. The loss function

Table 1
Class descriptions and statistics.

Class Id	Class Description	Pixel Nr.	Ratio in %
0	Unclassified	–	–
1	Dense forest/meadows, often next to water	52846	9.1
2	Less dense forest/meadows	146680	25.2
3	Residential area	154387	26.5
4	Residential area with a lot of green space	155032	26.7
5	Shallow water	13010	2.2
6	City centre/industrial area	45612	7.8
7	Deepest water/sea water	845	0.1
8	Deep water	5804	1.0
9	greenhouses	3339	0.6
10	greenhouses	4077	0.7

Table 2
Mean building density for each class.

Class Id	Class Description	Mean building density	Mean building height
1	Dense forest/meadows, often next to water	9.47	7.66
2	Less dense forest/meadows	8.17	6.84
3	Residential area	26.14	9.66
4	Residential area with a lot of green space	12.09	7.07
5	Shallow water	1.79	5.67
6	City centre/industrial area	35.41	11.51
7	Deepest water/sea water	0.01	–
8	Deep water	0.19	2.11
9	Greenhouses	17.22	7.31
10	Greenhouses	19.52	7.66

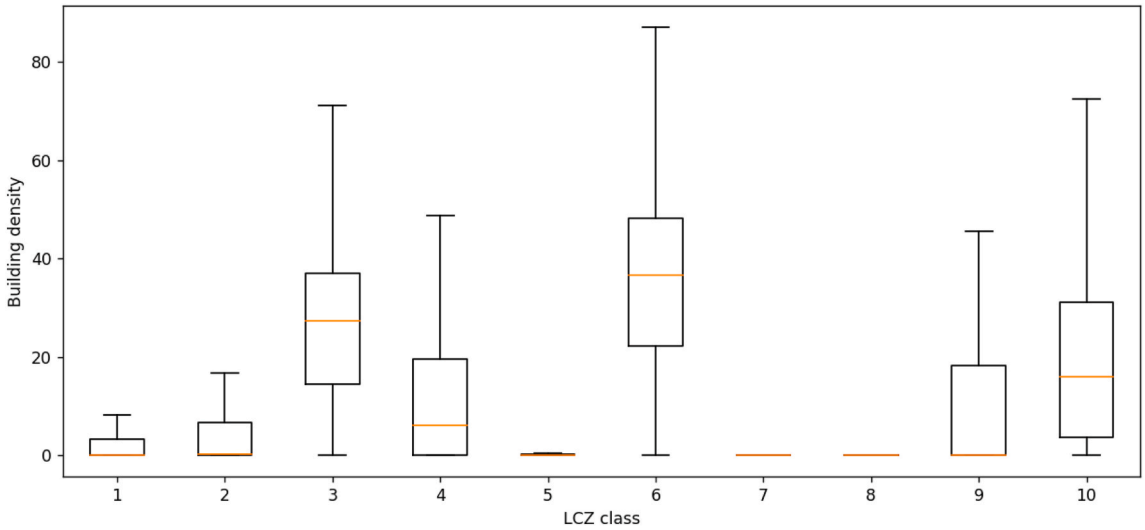


Fig. 5. Building density distribution for each class.

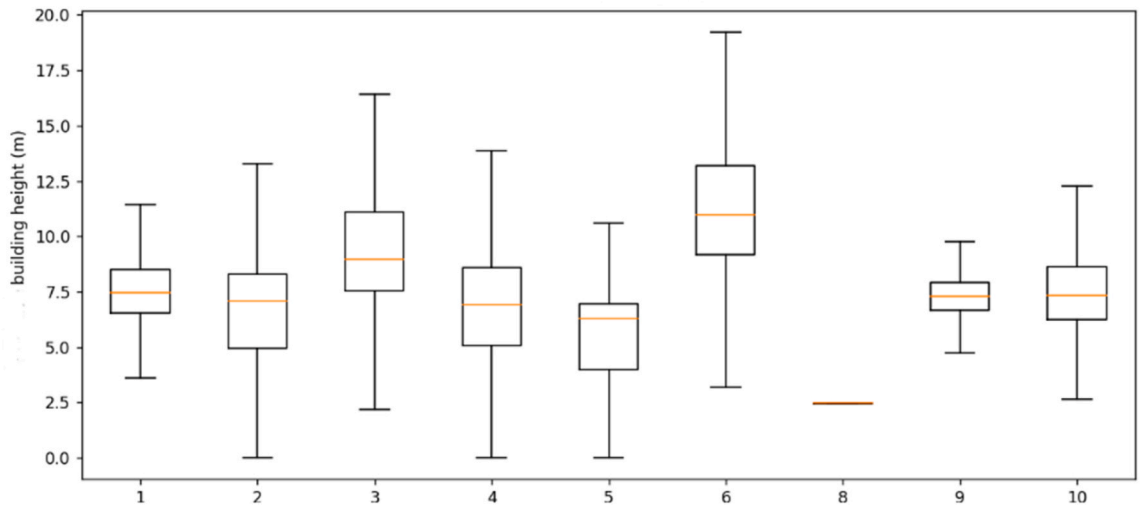


Fig. 6. Building height distribution for each class.

used in the runs is again the *Sparse Categorical Cross entropy* and the learning rate the selected value of 0.001.

The results show that a patch size of 64 is the best choice for the model. Although a patch size of 256 yielded the highest overall accuracy, it learned much slower, with the loss function converging at a slower rate. Additionally, using such a large patch size resulted

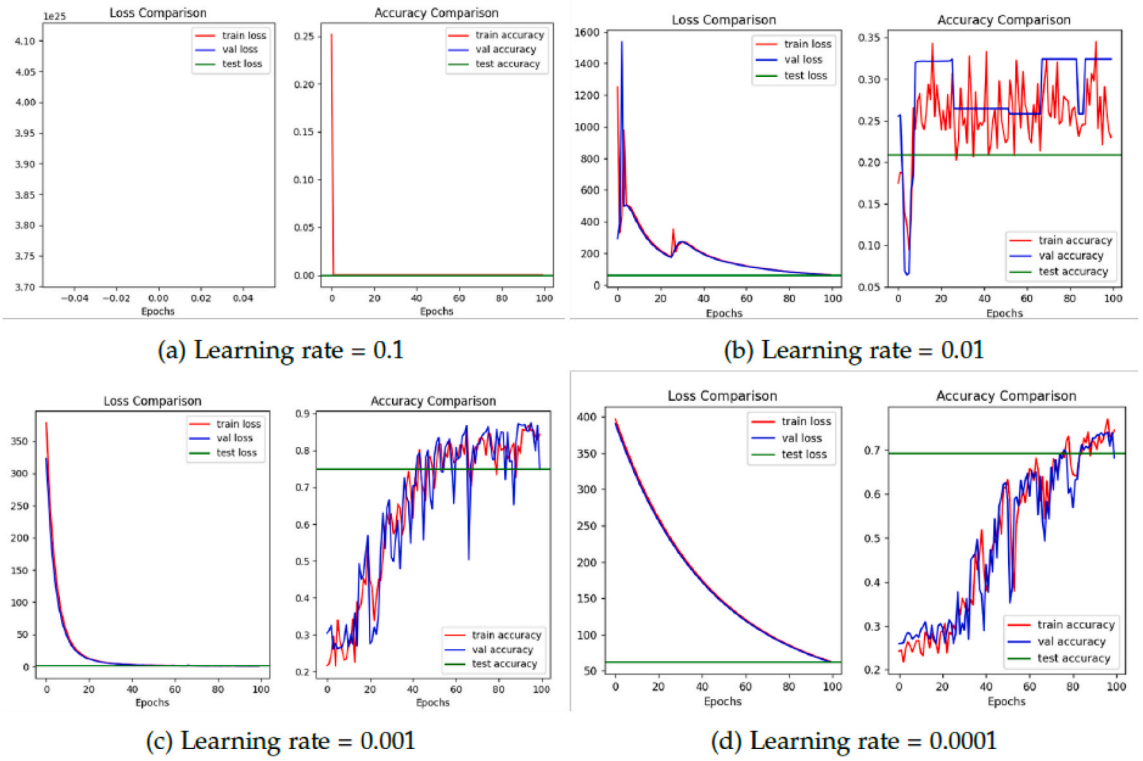


Fig. 7. Model's performance according to loss and accuracy with different learning rates.

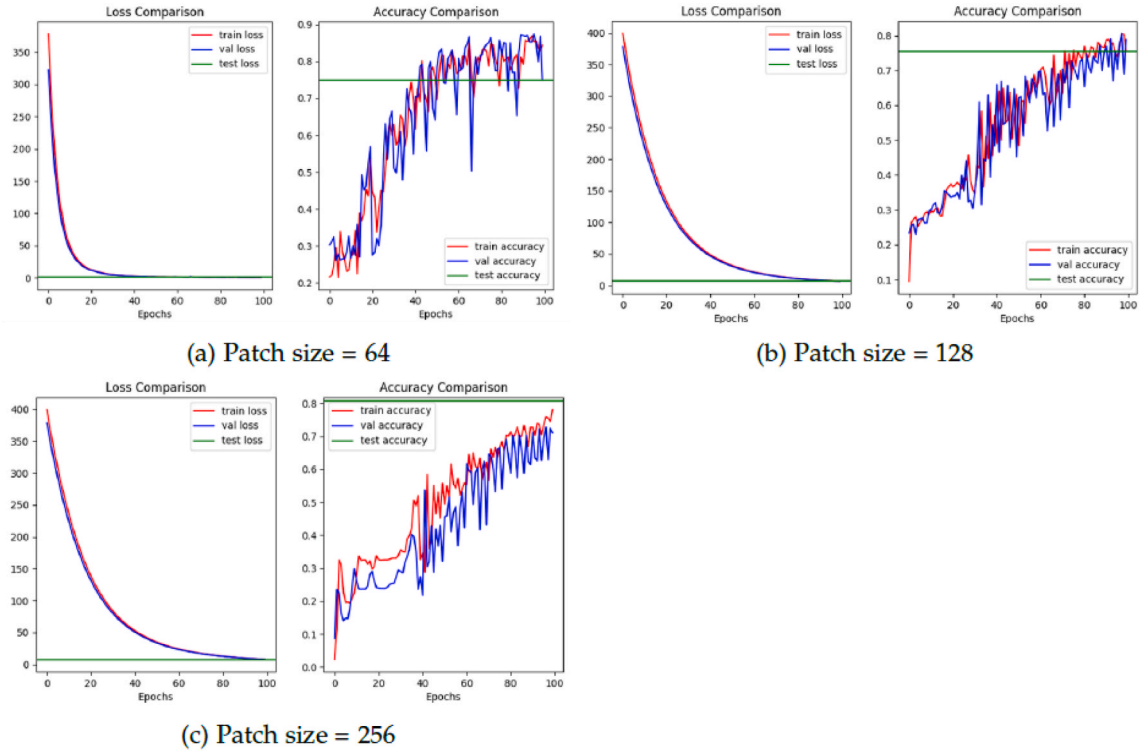


Fig. 8. Model's performance according to loss and accuracy for different patch sizes.

in fewer patches in the training data, as each 256x256 patch covers a significant area with a spatial resolution of 70-m pixels. This limitation in the number of patches negatively impacted the training process. Therefore, a patch size of 64 was selected, balancing learning speed and the amount of training data available, leading to a more efficient and robust model performance.

4.2.3. Loss function

The model's performance in terms of loss and accuracy are also compared with regard to two multi-class loss functions: Kullback-Leibler (*KL) Divergence* and Sparse *Categorical Cross entropy*. The patch size was set to 64 and the learning rate to 0.001. The results are presented in Fig. 9.

The model's performance when using the KL Divergence loss function is not satisfactory. The gradients provided by *KL Divergence* during backpropagation might be less informative or more unpredictable compared to those from *Sparse Categorical Cross entropy*. This can make the optimization process less efficient, causing the model to struggle with learning the correct patterns from the data. Given this, *Sparse Categorical Cross entropy* was selected for its more reliable and meaningful results. This loss function is well-suited for classification tasks, ensuring effective model performance and interpretability.

4.3. Experiments

4.3.1. Full dataset

The performance of the model when being trained and tested with the full dataset of 46 thermal images is presented in Fig. 10. The optimal hyperparameter combination that was selected in Section 4.2 is used for this run: learning rate = 0.001, patch size = 64 and loss function = *Sparse Categorical Cross entropy*, respectively. 70 % of the full dataset was used for training, 15 % for validation and 15 % for testing.

The loss functions converge smoothly, indicating effective learning and gradual optimization of the model parameters. The accuracy values and macro F1 score also converge towards a steady value, indicating that the model's performance stabilizes over time. The test accuracy value is 0.748 and the test macro F1 score is 0.59. The macro F1 score value is significantly smaller than the accuracy value, as it assigns each class the same weight. This indicates that the performance of some classes that do not appear often in the dataset is less favourable than the average performance of other classes. The test F1 score per class is presented in Table 3.

The majority of the classes provide a high test F1 score. However, classes 7, 9 and 10 have a test F1 score of 0. These classes are defined as “Deepest water/sea water” and “A few greenhouses, does not appear often” for both class 9 and 10. This can be explained by low representation of these classes in the training and/or test data, or its complex nature. Class 7, 9 and 10 all have a representation smaller than 1 % of the total number of pixels. This results in the model not distinguishing these classes from other classes at all. Other classes yield more desirable test F1 scores, averaging between 0.7 and 0.9.

4.3.1.1. Probability distribution. The model assigns a probability value to each class for every pixel, and the pixel classified into the class with the highest probability. This allows the model to express varying levels of certainty about a pixel's class. Performance varies by class, as shown in Fig. 11, which illustrates the distribution of maximum probability values per class in the test data. These distributions align with test F1 scores: classes with lower F1 scores, which are more frequently misclassified, also show lower model certainty. This may be due to the underrepresentation of classes 7, 9, and 10 or the complexity of distinguishing their behaviour from other classes. Notably, class 9 is absent in the test data.

4.3.2. Seasonal influence

The impact of using thermal imagery from different seasons, for training and testing, has been explored. The initial training dataset was split in two parts: Spring/Summer and Autumn/Winter. These two subsets were used as training and testing input for two separate runs, using the optimal hyperparameters settings explored in Section 4.2. The results show that the Spring/Summer dataset leads to better results, with higher accuracy and the F1 scores. The results of the two are presented in Fig. 12.

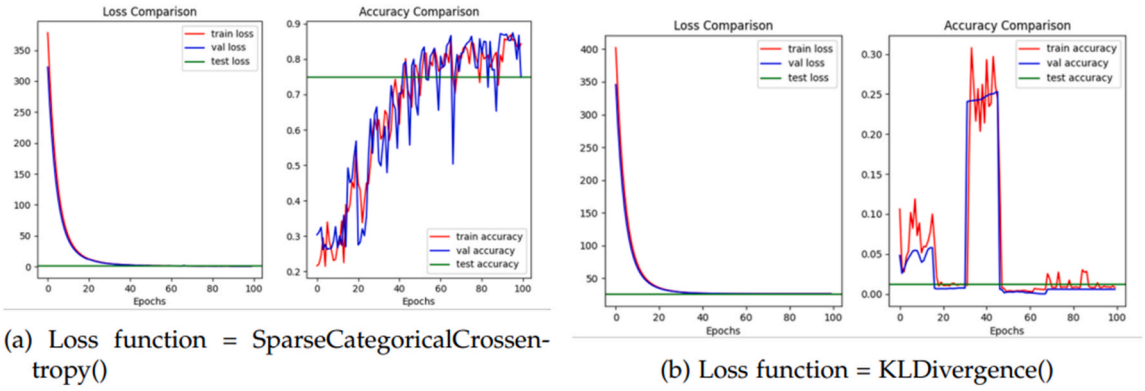


Fig. 9. Model's performance according to loss and accuracy with different loss functions.

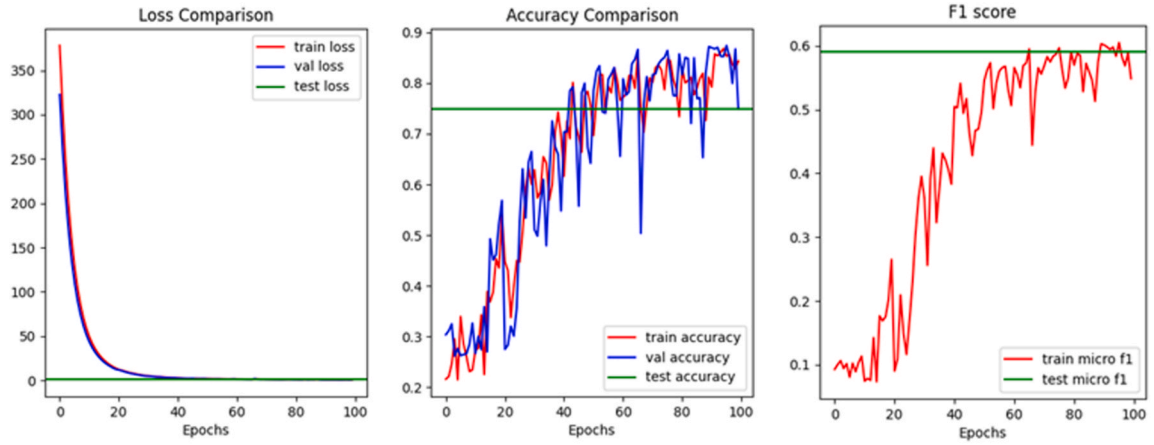


Fig. 10. Training and testing with full dataset results.

Table 3
Test F1 score per class for "Full Dataset" experiment.

Class Number	Test F1 score (per class)
1	0.698
2	0.856
3	0.888
4	0.862
5	0.765
6	0.877
7	0.000
8	0.918
9	0.000
10	0.000

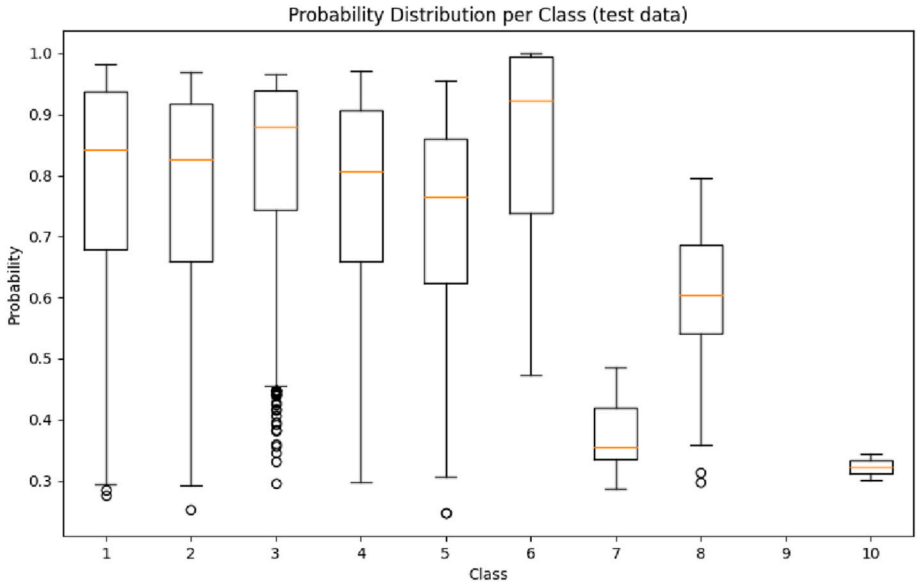


Fig. 11. Probability distribution per class for the test data.

In both runs, the loss functions converge smoothly, indicating effective learning and gradual optimization of the model parameters. The training and validation losses decrease and stabilize over time, suggesting that the model is not overfitting or underfitting significantly. When looking at the accuracy values, they also converge towards a steady value, indicating that the model's performance

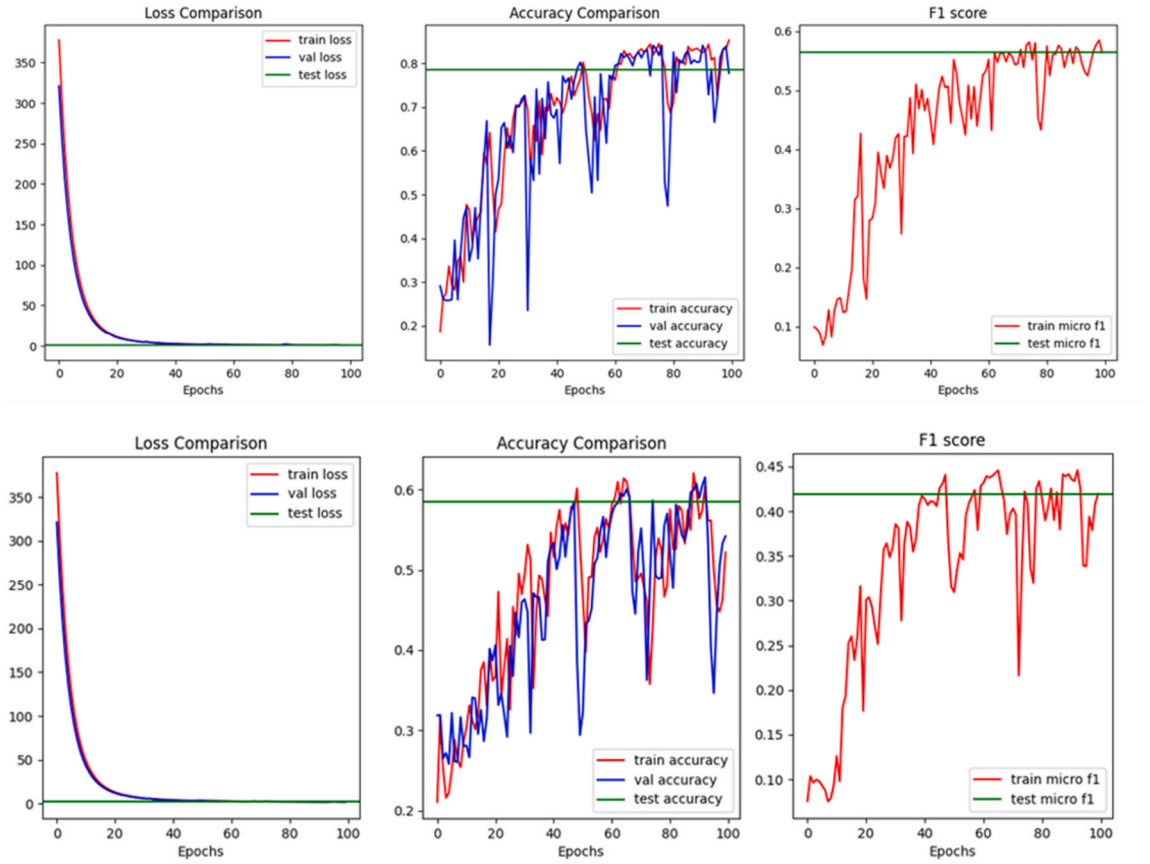


Fig. 12. Training and testing with data from Spring/Summer (above) and Autumn/Winter (below).

stabilizes over time. However, with the Autumn/Winter dataset, the accuracy values show higher variability across epochs, and more outliers can be observed. This suggests that the model may have more difficulty in learning consistently from this dataset, and that the Autumn/Winter dataset might be less representative for the classification task. This is also emphasized by the lower test accuracy value, specifically 0.586 for Autumn/Winter, compared to 0.785 for Spring/Summer. A similar proportion is observed when examining the test macro F1 score. The Autumn/Winter dataset results in a test macro F1 score of 0.42, compared to 0,56 for the Spring/Summer dataset. These values are lower than the accuracy values, as each class has the same weight in this calculation. The unique F1 scores per class also show different performance per training and testing input dataset (Table 4).

Almost all classes are classified better by the model that is trained and tested with the Spring/Summer dataset. This could be due to the larger temperature variations between objects due to warmer ambient temperatures in Spring/- Summer. This results in more distinct thermal signatures in the images, enabling the model to more effectively differentiate between different classes. This finding is also presented by Du et al. (2020), showing that LCZs are differentiated better in Summer than in other seasons regarding LST. Some classes that are not better classified in Spring/Summer compared to Autumn/Winter are misclassified entirely, achieving an F1 score of 0 or nearly 0. This can be explained by the low representation of these classes in the training dataset. Class 7, 9 and 10 (“Deepest water/sea water”, “A few greenhouses” and “A few greenhouses” respectively) have fewer pixels in the mask that is used for training. With few examples, the model doesn't have enough data to learn the patterns and characteristics of that class effectively. The test F1 scores that are 0 can also be explained by the fact that the testing is done with only 15 % of the initial dataset and sometimes this subset does not contain any pixels of these underrepresented classes.

Especially class 1 and 4 show a significant performance difference. Class 1 is described as “Dense forest/meadows, often next to

Table 4
Number of images and test accuracy for different selections.

Selection of Images	Number of Images	Test Accuracy
Spring/Summer	39	0.785
Spring/Summer random selection 1	7	0.759
Spring/Summer random selection 2	7	0.783
Autumn/Winter	7	0.586

water", and class 4 as "Residential area with a lot of green/meadows". These classes may be harder to differentiate when using only Autumn/Winter images, as their distinguishing feature of being relatively cooler compared to other classes is more apparent during warmer periods. When training the model with only lower LST values, this class might not be as distinguishable. Fig. 13 presents the actual mask of one patch with its classification result after training the model with Autumn/Winter images. Class 1 (blue) and class 4 (orange) are generally misclassified as class 2 ("Less dense forest/meadows" in yellow). This implies that the tree coverage has less impact on the thermal signature in Autumn/Winter and is therefore less distinguishable. The classes that show the smallest performance difference are class 5 ("Shallow water"), class 6 ("City centre/industrial area") and class 8 ("Deep water"). This implies that these classes are comparably distinguishable in Autumn/Winter than in Spring/Summer.

It must be noted that the seasonal distribution of datasets used in this experiment are imbalanced. Specifically, the Autumn/Winter dataset comprises 7 images, whereas the Spring/Summer dataset contains 39 images. This imbalance can significantly influence the performance of the model. To mitigate this effect and ensure a fair comparison, additional experiments are conducted where two random subsets of 7 images were selected from the Spring/Summer dataset. These subsets were then used to match the number of images in the Autumn/Winter dataset, allowing to isolate and evaluate the impact of dataset size on model performance. The model performance of these runs is expressed in test accuracy in Table 4.

The results show that training and testing the model with a random selection of 7 Spring/- Summer images does not change the larger test accuracy compared to training and testing the model with Autumn/Winter images (Table 5). Similar performance differences can also be observed here. This implies that the performance differences observed were due to using images from different seasons, rather than the number of images used.

4.3.3. Daytime vs. nighttime

To explore the impact of diurnal fluctuations of LST on LCZ classification performance, thermal imagery from different times (daytime vs. nighttime), as training and testing dataset, is leveraged. The initial training dataset was split into two subsets, one consisting of thermal images taken at nighttime and one consisting of thermal images taken at daytime. The subsets were used as training and testing input for two separate runs, using the optimal hyperparameters settings (Section 4.2). When comparing the results of the two runs, the daytime subset results in better model performance, with higher accuracy and F1 scores. The results are presented in Fig. 14.

The loss functions converge smoothly, indicating effective learning and gradual optimization of the model parameters. The training and validation losses decrease and stabilize over time, suggesting that the model is not overfitting or underfitting significantly. Accuracy values also converge towards a steady value, indicating that the model's performance stabilizes over time. The performance of the nighttime run converges to a significantly lower value compared to the daytime run, specifically 0.4764 versus 0.8001. The same proportion can be observed with the test macro F1 score. The nighttime dataset results in a test macro F1 score of 0.2109 and the daytime dataset in 0.5648. These values are lower than the accuracy values, because each class has the same weight in this calculation. The unique F1 scores per class, indicating different performance per training and testing input dataset are presented in Table 6.

Almost all classes are classified better by the model that is trained and tested with the daytime dataset. The model performs generally worse across all classes when using only night images. Half of the classes are not distinguished in the test images. The most significant performance differences between training/testing with images from daytime and nighttime can be observed for the classes 1 ("Dense forest/meadows, often next to water"), 4 ("Residential area with a lot of green/meadows"), 6 ("City centre/industrial area") and 8 ("Deep water"). These classes were classified satisfactory with images from daytime, and with images from nighttime the test F1 scores are zero or close to zero. It is remarkable that these classes contain the generally cold (1, 4, 8) but also a generally warmer class (6). To get an idea to which other classes these classes are misclassified, the same patch and its predicted masked images is shown in Fig. 15. The class "City centre/industrial area" (in red) is misclassified as the class 3: "Residential area" (in green) and a small part of class 5: "Shallow water" (in purple). Class 2: "Less dense forest/meadows" (in yellow) and class 1: "Dense forest/meadows, often next to water" (in blue) is misclassified as class 4: "Residential area with a lot of green/meadows" (in orange). A general trend that can be concluded, is that the classes that show more "extreme" behavior (warmer or cooler than other classes), are misclassified as classes with less fluctuations and more average values. This might be because at night the LST values of classes are more similar to each other.

It must be noted that the diurnal distribution of datasets used for this experiment are imbalanced. For this experiment, there are only 5 images taken at nighttime, whereas the daytime dataset contains 41 images. To mitigate this effect and ensure a fair comparison, additional experiments are conducted where two random subsets of 5 images were selected from the daytime dataset. These subsets were then used to match the number of images in the nighttime dataset, allowing to isolate and evaluate the impact of dataset size on model performance. The model performance of these runs is expressed in test accuracy (Table 7).

A slight performance difference between the different random selections of 5 images and the dataset of 41 images can be noted, but a significant performance difference when only using nighttime images. This implies that the observed performance differences were a result of using images from different times rather than the number of images used.

4.3.4. Extreme LST value analysis

In previous experiments, it was concluded that training and testing the model with images from days and on average warmer seasons lead to better results. To explore the impact of extreme higher LST values on LCZ classification performance, only thermal imagery containing peak LST values are used for this experiment. The thermal signatures, containing average LST values per thermal imagery, have shown four peaks. It was also concluded from the previous experiments that using more images does not always lead to a better performance of the model (Sections 4.3.2 and 4.3.3). Therefore the model is trained and tested with different numbers of images from the observed peaks. With only one image (the image with the maximum LST values), one image per peak (four images with the

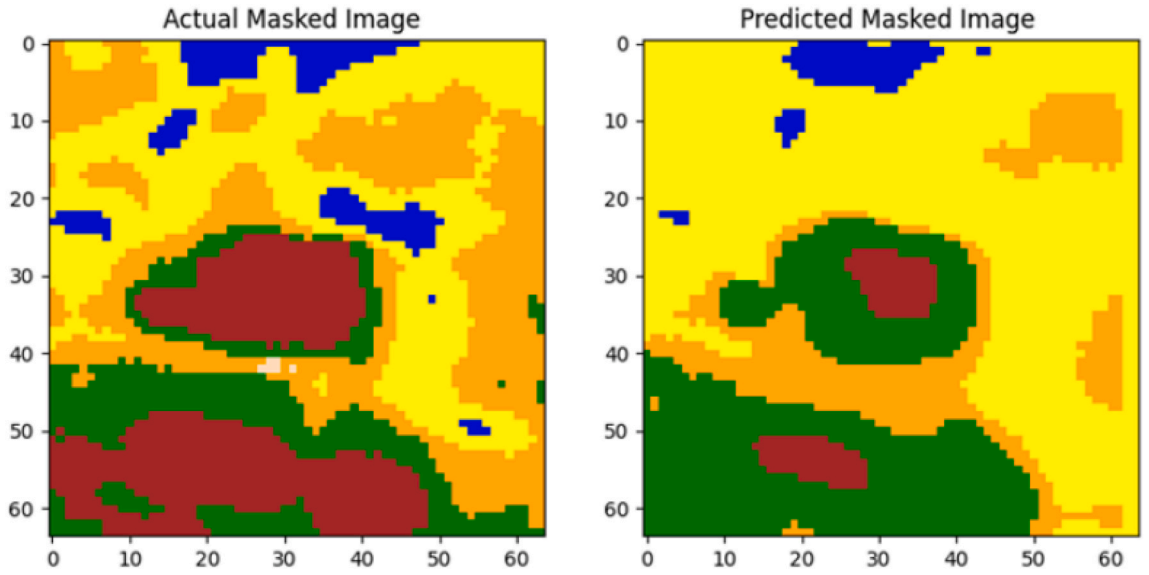


Fig. 13. Actual mask and classification result after training the model with Autumn/Winter images.

Table 5
Test F1 score per class for Spring/Summer and Autumn/Winter.

Class #	Test F1 score		Test F1 score (7 images)	
	Spring/Summer	Autumn/Winter	Spring/Summer	Autumn/Winter
1	0.7522	0.3374	0.6230	0.3374
2	0.8345	0.6245	0.8253	0.6245
3	0.8159	0.6596	0.8307	0.6596
4	0.7180	0.4392	0.7581	0.4392
5	0.7648	0.6781	0.6948	0.6781
6	0.8038	0.7308	0.8052	0.7308
7	0.0556	0.0000	0.0000	0.0000
8	0.7790	0.7011	0.7714	0.7011
9	0.0000	0.0000	0.0000	0.0000
10	0.0000	0.0000	0.0000	0.0000

highest LST values for that peak) and all images of the peaks (14 images) specifically.

The test OA is comparable for the first two experiments and very high for the last experiment (Table 8). In the first experiment, using a single image with high LST values, the test accuracy achieved is 0.260. This suggests that training and testing on a single image, while providing some discriminatory power, lacks generalizability due to limited data diversity. The second experiment includes four images with high LST values, the accuracy improves slightly to 0.285. This increase is not significant but might indicate that incorporating more samples helps capture broader patterns. The dataset still falls short of robust performance. The third experiment, using 14 images with high LST values and emphasizing higher variability (not only the maximum values), significantly boosts accuracy to 0.83. This substantial improvement underscores the importance of dataset diversity in enhancing model generalization. By exposing the model to a wider range of LST variations across multiple images, it can learn more comprehensive features and achieve higher accuracy in classification tasks.

Remarkably, the accuracy of 0.83 in the third experiment surpasses that achieved when training and testing on the full dataset (Section 4.3.1), indicating that focusing on images with high LST values and ensuring variability can yield superior performance compared to a more generalized approach across all dataset samples.

The higher accuracy observed for daytime, warm-season, and high-LST imagery indicates that the model is able to distinguish LCZ classes more effectively under strong thermal conditions. Daytime and warm-season conditions amplify surface–atmosphere interactions that are directly linked to LCZ-defining properties such as building density, material thermal inertia, vegetation fraction, and impervious surface coverage. Under these conditions, LCZs with similar geometric structure but different surface compositions (e.g., compact mid-rise vs. open mid-rise, or built-up vs. vegetated zones) exhibit more distinct diurnal heating patterns, suggesting that improved separability is not only due to increased contrast but also to more informative temporal thermal signals tied to LCZ characteristics. This improvement reflects the model's ability to exploit informative temporal thermal patterns, which facilitates more reliable differentiation of LCZs by the model. The improved performance observed for daytime, warm-season, and high-LST imagery has important implications for the broader applicability of thermal-only LCZ classification. These conditions enhance LCZ separability

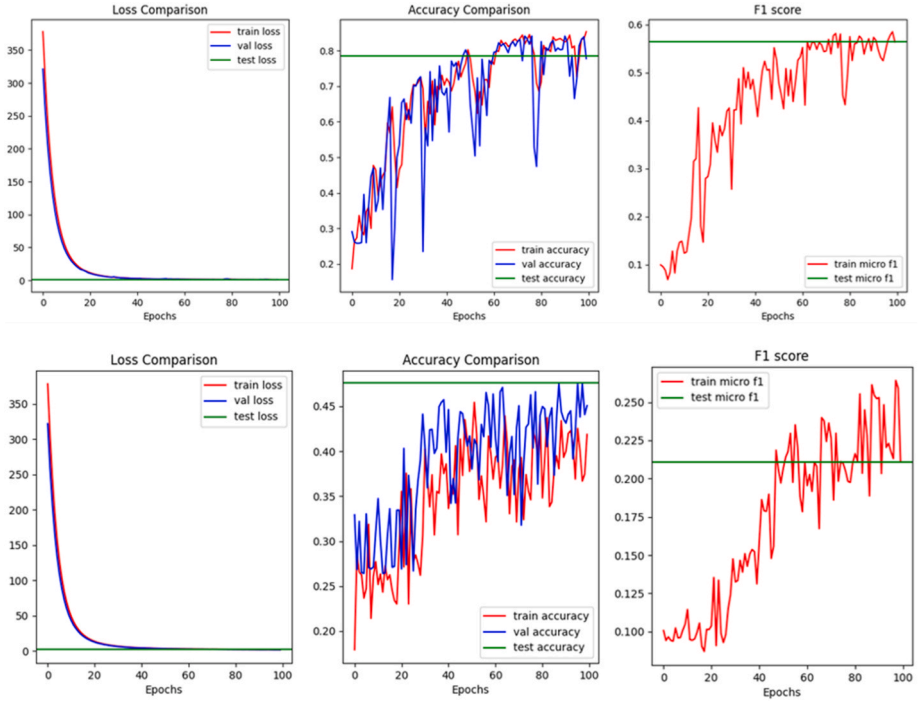


Fig. 14. Training and testing with data from daytime (above) and nighttime (below).

Table 6
Test F1 score per class for daytime and nighttime.

Class Number	Test F1 score (per class)	
	Daytime	Nighttime
1	0.7594	0.0513
2	0.8505	0.5874
3	0.8296	0.6596
4	0.7842	0.1409
5	0.7143	0.5035
6	0.7273	0.0000
7	0.0510	0.0000
8	0.7690	0.0000
9	0.0000	0.0000
10	0.0000	0.0000

by amplifying surface–atmosphere interactions linked to urban form and materials. Consequently, broader applicability may require context-aware training strategies, such as season-specific ones.

5. Conclusions

The results of this study indicate that spatio-temporal analysis of thermal imagery through a U-Net architecture is effective in classifying urban LCZs. The significant impact of temporal factors, such as diurnal and seasonal variations, underscores the importance of considering time-series data in LCZ classification. The observed higher performance with daytime and Spring/Summer imagery suggests that these conditions provide clearer, more distinguishable thermal signatures, facilitating better differentiation between LCZs. The findings align with previous research indicating the utility of remote sensing data in urban climate studies. Unlike traditional LCZ classification methods that rely on multi-spectral data, this approach focuses on thermal behavior, offering a new perspective. The improved classification accuracy during daytime and warmer seasons supports studies by Stewart and Oke (2012), Lotfian et al. (2019) and Zhao et al. (2021), which emphasize the role of thermal properties in defining urban LCZs. While the U-Net architecture employed in this study is a well-established deep learning model for semantic segmentation, the primary contribution of this research resides in systematically analyzing the role of multitemporal thermal information for LCZ classification. By using ECOSTRESS land surface temperature data exclusively, this study investigates how diurnal and seasonal thermal dynamics influence the separability of LCZ classes and model performance. This focus on temporal thermal behavior provides insights that complement morphology-based LCZ

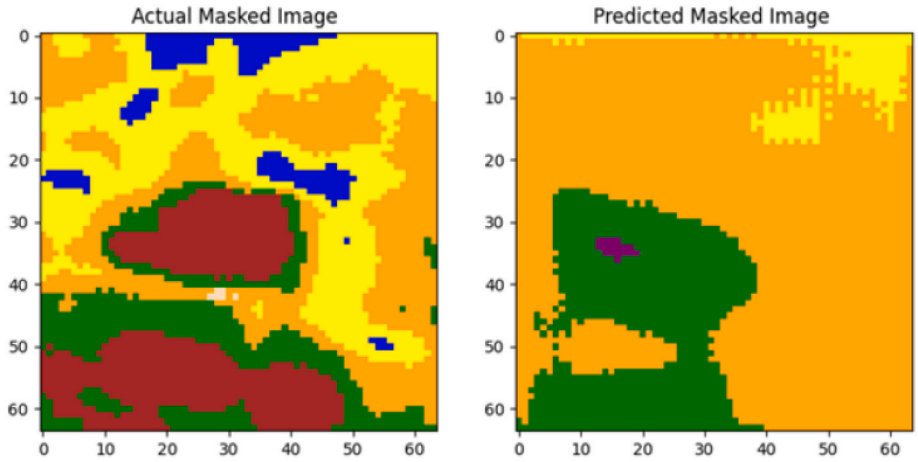


Fig. 15. Actual mask and classification result after training the model with nighttime images.

Table 7
Number of images and test accuracy for different selections.

Selection of Images	Number of Images	Test Accuracy
Daytime	41	0.8001
Daytime random selection 1	5	0.7998
Daytime random selection 2	5	0.7657
Nighttime	5	0.4764

Table 8
Test accuracy values for different image selections.

Selection of Images	Number of Images	Test Accuracy
Maximum	1	0.260
Maximum per peak	4	0.285
All peaks	14	0.834

approaches and highlight the added value of thermal observations for urban climate-oriented LCZ mapping. The result of the experiment on using extreme LST values for LCZ classification indicated that the optimal performance of the model is a trade-off between sufficient temporal variability and distinctive LST behavior on periods of peak thermal intensity. This provides an essential insight for optimizing the sampling strategies for satellite imagery, considering both classification accuracy as well as data and computational costs.

Although ECOSTRESS LST data have a spatial resolution of 70 m and the model produces pixel-wise predictions, the LCZ classification is not interpreted at an isolated per-pixel scale. The U-Net architecture incorporates spatial context through its convolutional receptive fields, allowing each prediction to reflect neighbourhood characteristics, extending beyond the individual pixels, approaching the typical LCZ scale (100–400 m). The finer resolution of ECOSTRESS helps capture urban thermal heterogeneity, especially in complex and mixed-use environments such as Rotterdam and The Hague. The resulting LCZ maps can also be aggregated to coarser spatial units when strict adherence to LCZ scale guidelines is required.

The main advantage of the proposed approach lies in its exclusive use of multitemporal land surface temperature (LST) data, which directly capture urban thermal behaviour rather than inferring it from physical form or land-use proxies. While methods based on high resolution optical imagery, LiDAR, or vector building data effectively represent urban morphology, they often lack information on diurnal and seasonal thermal dynamics that are central to urban climate processes. By leveraging ECOSTRESS thermal imagery, our approach distinguishes LCZs with similar structural characteristics but differing thermal responses, which is particularly relevant for micro-climate analysis and heat mitigation studies. Furthermore, using only ECOSTRESS thermal time-series data avoids dependency on data sources that are often unavailable or inconsistent across cities, enhancing transferability and applicability in data-scarce regions.

Overall, the approach proves to be robust and adaptable, offering valuable insights for urban planning and environmental monitoring. The ability to distinguish LCZs based on their thermal behavior provides urban planners with valuable insights into the thermal characteristics of different urban areas. This information can inform strategies to mitigate urban heat islands and enhance urban resilience to climate change.

The key limitation of this study is related to the temporal resolution, which is restricted by the thermal sensor, which, despite daily

coverage, is hindered by operational challenges (e.g. irregular overpass times) and the inability of thermal sensors to penetrate cloud cover. These create temporal gaps in the dataset, potentially reducing the model's effectiveness in capturing and classifying temporal variations in LST. Integrating ECOSTRESS imagery with other thermal data sources can enhance observation frequency, mitigate temporal gaps, and provide a more continuous record of LST variations. However, managing different spatial resolutions and sensor characteristics remains a challenge and requires careful handling to ensure consistency and reliability.

Future research should prioritize the integration of weather conditions, as LST trends in thermal imagery are significantly influenced by factors such as cloud cover, precipitation, and wind speed. Incorporating meteorological data would improve the model's ability to classify land cover types under diverse weather conditions more accurately. The goal of this research was exploring the potential of spatio-temporal behavior analysis of thermal data for LCZ classification and therefore only thermal imagery was leveraged. While the results indicated that temporal thermal imagery leads to high accuracy in LCZ classification, combining thermal imagery with other geospatial data sources (e.g. multispectral imagery, topographic data, and land use/land cover maps) can complement thermal data and further refine LCZ classifications.

CRediT authorship contribution statement

Michaja van Capel: Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Azarakhsh Rafiee:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Roderik Lindenbergh:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data and source code are available at: https://github.com/mvancapel/LCZ-classification.

References

Aghabozorgi, S., Shirkhorshidi, A.S., Wah, T.Y., 2015. Time-series clustering–a decade review. Inf. Syst. 53, 16–38.

Anjos, M., Targino, A.C., Krecl, P., Oukawa, G.Y., Braga, R.F., 2020. Analysis of the urban heat island under different synoptic patterns using local climate zones. Build. Environ. 185, 107268.

Aslam, A., Rana, I.A., 2022. The use of local climate zones in the urban environment: a systematic review of data sources, methods, and themes. Urban Clim. 42, 101120. https://doi.org/10.1016/J.UCLIM.2022.101120.

Bai, Y., Chen, M., Zhou, P., Zhao, T., Lee, J., Kakade, S., Wang, H., Xiong, C., 2021. How important is the train-validation split in meta-learning?. In: Meila, M., Zhang, T. (Eds.), Proc. 38th Int. Conf. Mach. Learn., vol. 139. PMLR, pp. 543–553. URL. https://proceedings.mlr.press/v139/bai21a.html.

Bechtel, B., Alexander, P.J., Beck, C., Böhner, J., Brousse, O., Ching, J., et al., 2019. Generating WUDAPT Level 0 data–current status of production and evaluation. Urban Clim. 27, 24–45.

Bhatia, V., 2021. U-net implementation from scratch using tensorflow. URL. https://medium.com/geekculture/u-net-implementation-from-scratch-using-tensorflow-b4342266e406.

Bishop, C.M., Bishop, H., 2023. Deep Learning: Foundations and Concepts. Springer Nature.

Chen, Y.C., Chiu, H.W., Su, Y.F., Wu, Y.C., Cheng, K.S., 2017. Does urbanization increase diurnal land surface temperature variation? Evidence and implications. Landsc. Urban Plann. 157, 247–258.

Cilek, M.U., Cilek, A., 2021. Analyses of land surface temperature (LST) variability among local climate zones (LCZs) comparing Landsat-8 and ENVI-met model data. Sustain. Cities Soc. 69, 102877.

Demuzere, M., Bechtel, B., Middel, A., Mills, G., 2019. Mapping Europe into local climate zones. PLoS One 14, e0214474.

Du, P., Chen, J., Bai, X., Han, W., 2020. Understanding the seasonal variations of land surface temperature in Nanjing urban area based on local climate zone. Urban Clim. 33, 100657.

Du, R., Liu, C.H., Li, X.X., Lin, C.Y., 2023. Effect of local climate zone (LCZ) and building category (BC) classification on the simulation of urban climate and air-conditioning load in Hong Kong. Energy 271, 127004.

Durrani, A.R., Minallah, N., Aziz, N., Frnda, J., Khan, W., Nedoma, J., 2023. Effect of hyper-parameters on the performance of ConvLSTM based deep neural network in crop classification. PLoS One 18, e0275653.

EARTHDATA. https://appeears.earthdatacloud.nasa.gov/, 2015.

He, G., Dong, Z., Guan, J., Feng, P., Jin, S., Zhang, X., 2023. SAR and multi-spectral data fusion for local climate zone classification with multi-branch convolutional neural network. Remote Sens. 15 (2), 434.

Hook, S.J., Cawse-Nicholson, K., Barsi, J., Radocinski, R., Hulley, G.C., Johnson, W.R., et al., 2019. In-flight validation of the ECOSTRESS, Landsats 7 and 8 thermal infrared spectral channels using the Lake Tahoe CA/NV and Salton Sea CA automated validation sites. IEEE Trans. Geosci. Rem. Sens. 58 (2), 1294–1302.

Jing, H., Feng, Y., Zhang, W., Zhang, Y., Wang, S., Fu, K., Chen, K., 2019. Effective classification of local climate zones based on multi-source remote sensing data. In: IGARSS 2019–IEEE Int. Geosci. Remote Sens. Symp. IEEE, pp. 2666–2669.

Kadunc, N.O., 2022. How to normalize satellite images for deep learning. URL. https://medium.com/sentinel-hub/how-to-normalize-satellite-images-for-deep-learning-d5b668c885af.

Khan, A., Chatterjee, S., Weng, Y., 2021. 2 - characterizing thermal fields and evaluating UHI effects. In: Khan, A., Chatterjee, S., Weng, Y. (Eds.), Urban Heat Island Modeling for Tropical Climates. Elsevier, pp. 37–67.

Lasserre, J.A., Bishop, C.M., Minka, T.P., 2006. Principled hybrids of generative and discriminative models. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), vol. 1. IEEE, pp. 87–94.

Lau, K.K.L., Chung, S.C., Ren, C., 2019. Outdoor thermal comfort in different urban settings of sub-tropical high-density cities: an approach of adopting local climate zone (LCZ) classification. Build. Environ. 154, 227–238.

Liu, H., Zhan, Q., Yang, C., Wang, J., 2018. Characterizing the spatio-temporal pattern of land surface temperature through time series clustering: based on the latent pattern and morphology. Remote Sens. 10 (4).

Lotfian, M., Ingensand, J., Composto, S., 2019. The relationship between land surface temperature and local climate zone classification: a case study of the canton Geneva, Switzerland. URL. https://api.semanticscholar.org/CorpusID:199514954.
Ma, L., Yan, Z., He, W., Lv, L., He, G., Li, M., 2023. Towards better exploiting object-based image analysis paradigm for local climate zones mapping. ISPRS J. Photogrammetry Remote Sens. 199, 73–86.
Ma, L., Zhou, L., Blaschke, T., Yan, Z., He, W., Lu, H., et al., 2024. Projecting high resolution population distribution using local climate zones and multi-source big data. Remote Sens. Appl.: Soc. Environ. 33, 101077.
Perera, N.G., Emmanuel, R., 2018. A "Local Climate Zone" based approach to urban planning in Colombo, Sri Lanka. Urban Clim. 23, 188–203.
Ren, Z., Fu, Y., Dong, Y., Zhang, P., He, X., 2022a. Rapid urbanization and climate change significantly contribute to worsening urban human thermal comfort: a national 183-city, 26-year study in China. Urban Clim. 43, 101154.
Ren, J., Yang, J., Zhang, Y., Xiao, X., Xia, J.C., Li, X., Wang, S., 2022b. Exploring thermal comfort of urban buildings based on local climate zones. J. Clean. Prod. 340, 130744.
Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. CoRR. https://doi.org/10.48550/arXiv.1505.04597 abs/1505.04597.
Shi, Y., Ren, C., Lau, K.K.L., Ng, E., 2019. Investigating the influence of urban land use and landscape pattern on PM2.5 spatial variation using mobile monitoring and WUDAPT. Landsc. Urban Plann. 189, 15–26.
Stewart, I.D., Oke, T.R., 2012. Local climate zones for urban temperature studies. Bull. Am. Meteorol. Soc. 93, 1879–1900. https://doi.org/10.1175/BAMS-D-11-00019.1.
Wen, L., Ryan, S., Powell, M., Ling, J.E., 2025. From clusters to communities: enhancing wetland vegetation mapping using unsupervised and supervised synergy. Remote Sens. 17 (13), 2279.
Yan, Z., Ma, L., He, W., Zhou, L., Lu, H., Liu, G., Huang, G., 2022. Comparing object-based and pixel-based methods for local climate zones mapping with multi-source data. Remote Sens. 14 (15), 3744.
Zhang, D., Wang, J., Zhao, X., 2015. Estimating the uncertainty of average F1 scores. In: *Proc. 2015 Int. Conf. Theory Inf. Retr.* (ICTIR '15). ACM, pp. 317–320. https://doi.org/10.1145/2808194.2809488.
Zhang, Y., Li, Y., Chen, Y., Liu, S., Yang, Q., 2022. Spatiotemporal heterogeneity of urban land expansion and urban population growth under new urbanization: a case study of Chongqing. Int. J. Environ. Res. Publ. Health 19, 7792.
Zhao, N., Ma, A., Zhong, Y., Zhao, J., Cao, L., 2019. Self-training classification framework with spatial-contextual information for local climate zones. Remote Sens. 11 (23), 2828.
Zhao, Z., Sharifi, A., Dong, X., Shen, L., He, B.J., 2021. Spatial variability and temporal heterogeneity of surface urban heat island patterns and the suitability of local climate zones for land surface temperature characterization. Remote Sens. 13, 4338.