

Quantifying the effect of surrounding spatial heterogeneity on land surface temperature based on local climate zones using mutual information



Mrunali Vaidya ^{a,*}, Ravindra Keskar, Associate Professor ^a, Rajashree Kotharkar, Professor ^b

^a Department of Computer Science and Engineering, Visvesvaraya National Institute of Technology (VNIT), South Ambazari Road, Nagpur, Maharashtra, 440010, India

^b Department of Architecture and Planning, Visvesvaraya National Institute of Technology (VNIT), South Ambazari Road, Nagpur, Maharashtra, 440010, India

ARTICLE INFO

Keywords:

Local climate zone (LCZ)
Land surface temperature (LST)
Mutual information
Surrounding spatial heterogeneity
Random forest
Xgboost
Deep learning

ABSTRACT

Local climate zone (LCZ), a landscape classification scheme that segments urban region into 17 distinct zones, has become the standard for analyzing urban thermal environments since 2012. The characteristic features of each LCZ decide the standard regime for their Land Surface Temperature (LST). In heterogeneous cities, the variation in LST is observed for similar but spatially dispersed LCZs and has not yet been analysed by the researchers. The surrounding spatial heterogeneity could be partially responsible for such variation. Hence, we have presented a framework to analyze the LST variation of similar but spatially dispersed LCZs by considering the surrounding LCZ pattern¹ details in the study area of Nagpur (India). The framework uses machine learning techniques like random forest (RF), Xgboost, deep learning to confirm that the LST variation is due to surrounding LCZ pattern. Later it identifies the factors responsible for LST variation in each LCZ type using our proposed mutual information based Adjacent LCZ preference Estimator (ALPE). Deep learning model captures 90–92 % of the neighbourhood heterogeneity responsible for LST variation as compared to RF (0.83–0.79) and Xgboost (0.86–0.81). The reduction in bias (by 1.05 °C -1.39 °C) is observed while estimating LST incorporating surrounding LCZ pattern. This confirms that the external heterogeneity affects the LST of the corresponding LCZ. Further, the result analysis of ALPE suggests that characteristics of open LCZ types (when they are present in surrounding) highly influence the LST of compact and open LCZs.

1. Introduction

The notion of a Local climate zone (LCZ) has become a standard tool for investigating the nature of urban thermal environment since 2012 (Stewart & Oke, 2012). It is the unified classification scheme for urban climate, form, and functions (Xiang et al., 2023) based on ten built (LCZ 1–10) and seven natural (LCZ A–G) regions. Each LCZ type possesses similar characteristics of surface cover, material, and anthropogenic activities (Stewart & Oke, 2012; Kotharkar & Bagde, 2018a) at a local scale that influence the screen height temperature, thereby making it a prominent method to analyze the thermal contrast in urbanized areas. To date, multiple researchers have used the LCZ typology to study the impact of surface structure and cover properties (building height, tree height, sky view factor, vegetation density, pervious and impervious surface fraction, anthropogenic heat output, albedo, etc.) on the urban thermal environment via air temperature (Skarbit et al., 2017; Yang et al., 2018; Middel et al., 2014; Eimermacher, 2018; Fenner et al., 2017;

Yan et al., 2014; Shi et al., 2018) and land surface temperature (LST) (Cai et al., 2017; Geletić et al., 2016; Yang et al., 2020a; Yang et al., 2021; Zhou et al., 2022). LST-driven investigation for the LCZ is popular due to its complete spatial coverage (Geletić et al., 2016) over air temperature estimated from limited, fixed, and mobile measurement strategies (Ferreira & Durate, 2019). Though the LCZ is potentially applied at the mesoscale and micro-scale to predict outdoor temperature changes (Geletić et al., 2018), and to evaluate the effect of numerous land cover features on UHI (Thomas et al., 2014; Kotharkar & Bagde, 2018a; Anjos et al., 2020), the primary goal of LCZ focuses on analyzing urban thermal environment via investigating the relationship between spatially distributed LCZs and LST (Yang et al., 2020b).

The existing literature on the LCZ-LST relationship considers the factors that represent the spatial characteristics of individual LCZ in a single city at a study scale (fixed) lacking region scale (Yang et al., 2020b). However, due to the complex structure of urban surfaces in densely populated areas with diversified land cover categories (Xie

* Corresponding author.

E-mail addresses: dt18cse083@students.vnit.ac.in (M. Vaidya), rbkeskar@cse.vnit.ac.in (R. Keskar), rskotharkar@arc.vnit.ac.in (R. Kotharkar).

¹ LCZ pattern represents the group of LCZs surrounding any considered LCZ.

et al., 2020; Portela et al., 2020), the methodology of finding LST influencing factors can not be accomplished at single fixed scale. Generally, the primary focus of current LCZ-LST research is on LST variation (day-time, night-time, seasonal) among LCZs (Thomas et al., 2014; Unger et al., 2014; Zhao et al., 2019; Shi et al., 2018; Yang et al., 2020b; Krayenhoff & Voogt, 2016; Ochola et al., 2020; Yang et al., 2019) and very few studies give attention towards finding influencing factors of LST variation for spatially distributed individual LCZ (Yang et al., 2021). The standard method of identifying LCZ-LST relationship is a three step process utilising satellite data source like Landsat, Aster etc. for LST estimation. First, select the spatial grid scale (e.g. 150 m, 250 m, and 500 m) for the study area. Second, extract the explanatory features (representing land cover/surface characteristics) for LST within selected spatial scale. Third, build the model (random forest, Xgboost, deep learning, linear regression) to find out the LCZ-LST relationship. Typical features for LST variations are well demonstrated when the LCZs are correctly delineated. A GIS-based LCZ classification in Brno and Prague explains 89.3 % and 91.7 % LST variation when both LANDSAT and ASTER thermal images were used and proved that the LCZs can be better differentiated when LST is computed from Landsat over Aster (Geletič et al., 2016).

The general observation of LCZ-LST relationship states that, LST of LCZs varies with built and natural forms (Cai et al., 2017; Geletič et al., 2016; Zhao, 2018). In addition researchers have confirmed that- (1) There exist substantial differences in the LST of different LCZs (Geletič et al., 2019), the seasonal variation of the LST in LCZs did not follow a fixed order (Cai et al., 2018; Zhao et al., 2021) as well as there exists differences in day-time and night-time LST analysis (Ferreira & Durate, 2019). (2) LST variations were observed due to complex and diverse morphology (Cai et al., 2017; Cai et al., 2018). (3) LST of each LCZ is closely related to parameters such as geographical location, city size, building height and spacing, water permeability, and vegetation density (Thomas et al., 2014; Unger et al., 2014; Zhao et al., 2019). (4) LST of built LCZ is higher than that of natural LCZ (Shi et al., 2018; Yang et al., 2020b) and among natural LCZs, bare soil and land LST is higher than that of water bodies and forests (Krayenhoff & Voogt, 2016; Ochola

et al., 2020). (5) The effect of LST can be reduced by the wind as per the LCZ's building type and surface (Yang et al., 2019). The observations regarding the LCZ-LST relationship investigated by researchers till now are adequate only for homogeneous cities (cities that follow similar type of form and functions within a local scale) and not sufficient in depth and detail for heterogeneous urban forms (cities with mixed building types and physical functions within a local scale) (Yang et al., 2021). Hence, the current LCZ-LST research needs certain advancements to be further helpful for urban planners and policymakers to take decisions for reducing the adverse impact of urbanization.

One kind of possible enhancement is to analyze the LST variation within similar but spatially dispersed LCZs within the city by incorporating the information of the surrounding LCZ composition. Generally, the temperature behaviour of all similar LCZ patterns within the city is supposed to be identical as per the LCZ typology criteria, but this standard expectation is observed to be violated by complex and dense cities. Consider Fig. 1 showing the LST (day-time) variations among similar LCZ types located at different places within the city (Nagpur, India) in summer and winter 2019 surrounded by different combinations of other LCZs. Also the substantial LCZ-wise LST difference is observed from the LST statistics given in Table 2. The potential cause for such temperature variations may be the morphological features of individual LCZ and the effect of the characteristics of surrounding LCZs. There exists numerous works literature on relationship between LST-individual LCZ morphological parameters (Cai et al., 2017; Anjos et al., 2020; Geletič et al., 2019). For example, the effect of 2D/3D urban morphology on LST of LCZ is studied to identify warmest and coldest LCZ during summer (Zhou et al., 2022). The relationship between seasonal LST variation and spectral indices of various LCZs is analysed and observed varied LST increase in different cities and seasons (Xiang et al., 2023). But to the best of our knowledge, no one has focused on considering the surrounding LCZ details for analyzing the LST variation in LCZs for complex heterogeneous cities. The surrounding LCZ configuration may affect the LST of any corresponding LCZ according to their area and length of the shared boundary (see Fig. 4). The objective of this paper is twofold.

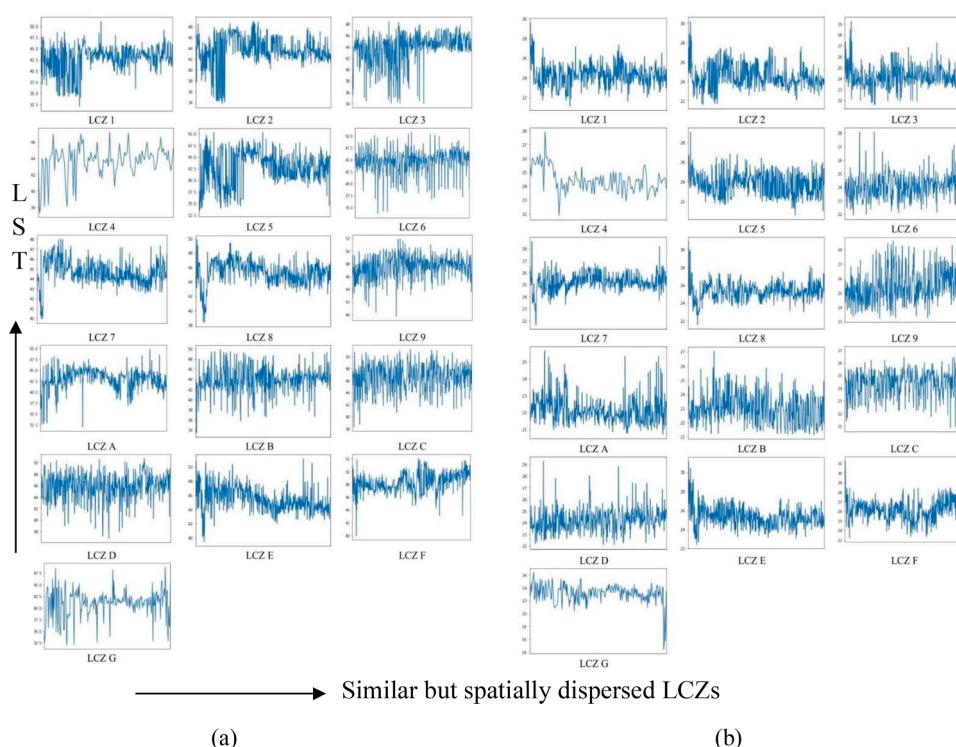


Fig. 1. Land surface temperature (day-time) variation of individual LCZs within the study area during 2019 year (a) summer variation (b) winter variation.

1. To evaluate the effect of surrounding LCZs on the LST variations in similar but spatially distributed LCZs in complex heterogeneous cities.
2. To identify the type of adjacent LCZs affecting the LST of individual LCZs in terms of their area and length of shared boundaries.

To summarize, existing research on LCZ-LST considers morphological parameters, spectral indices, and landscape metrics at a fixed spatial scale ignoring region scale. Although these parameters are explanatory for analyzing the LST variation according to spatial heterogeneity within LCZ, researchers did not focus on the surrounding LCZ composition and its effect on the LST of the considered LCZ. Hence, we have proposed a machine learning based methodology which validates that LST (day-time variation for summer and winter at regional scale) of individual LCZ is affected by its corresponding pattern of adjacent LCZ. The proposed methodology is tested for the heterogeneously built city Nagpur in India. It is divided into two parts.

1. The first part is the prediction in which we estimated the LST separately using two feature sets (<{LCZ} and {LCZ, adjacent LCZ pattern details}). The primary intention of the LST prediction part is to prove that, to some extent, the surrounding LCZ pattern is responsible for LST variation. The main focus of prediction phase is the limited feature set. In this paper, we have proposed the regression based model for LST prediction using LCZ and additional (but limited) information like adjacent LCZ pattern.
2. Second part deals with the analysis regarding LST of similar but spatially dispersed LCZ incorporating information about adjacent LCZs. The adjacent LCZ information is represented by two vectors representing ratio of area of each adjacent LCZ to the considered LCZ and the ratio of shared boundary with the considered LCZ. In analysis phase, LST extracted from Landsat 8 dataset is used to investigate the relationship between LCZ-LST-adjacent LCZs using mutual information based Adjacent LCZ preference estimator (ALPE). It ranks the adjacent LCZ in terms of their effect on the considered LCZ by considering their area and length (ratios) of shared boundary.

The organisation of this paper is as follows: [Sections 2 and 3](#) present

study area, input data and introduces methodology, followed by results and discussion in [Sections 4 and 5](#), respectively. Finally, the work is summarised and concluded in [Section 6](#).

2. Study area and data

2.1. Study area

Nagpur, the third largest city in Maharashtra (India), is selected as the region of interest for the underlying task ([Fig. 2 \(a\)](#)). It is located at latitude 21.15 and longitude 79.09. The total population of Nagpur is recorded as per 2011 census is 2405,665 and the metropolitan zone of the city is estimated to be about 88 square miles. Its climate is distinguished as tropical wet and dry with extremely hot summers during March to June, and winters during November to February. Nagpur is selected because of its heterogeneous urban forms. It is described as one of the highly complex and heterogeneous cities in India and various mixed LCZs were observed in it such as LCZ 6₅, 6_B, 3₂ etc. ([Kotharkar & Bagde, 2018b](#)).

2.2. LCZ map

We have used Local Climate Zone (LCZ) typology for identifying land cover categories in the study area. The LCZ map for 2019 year has been incorporated for the day-time LST analysis in summer and winter season. We delineated the LCZ map by following remote sensing based classification approach and LCZ boundaries are identified using the concept of homogeneity ([Vaidya et al., 2024](#)). The LCZs observed during 2019 for the study area are shown in [Fig. 2 \(b\)](#). The percentage area contribution of each LCZ observed in 2019 is given in [Fig. 5 \(a\)](#). The LCZ map is prepared by following two steps. Initially, using spectral, textural and contextual features embedded in Landsat 8 scenes incorporating supervised learning, a pixel based classified Landsat 8 raster is created ([Vaidya et al., 2024](#)). Later, DEALB (Directional Edge Algorithm for Locating Border) is proposed to convert the pixel based outcome into homogeneous LCZ regions which are represented by Polygon geometry. DEALB is an entropy-based post-classification directional edge algorithm developed by us which creates homogeneous LCZ regions to

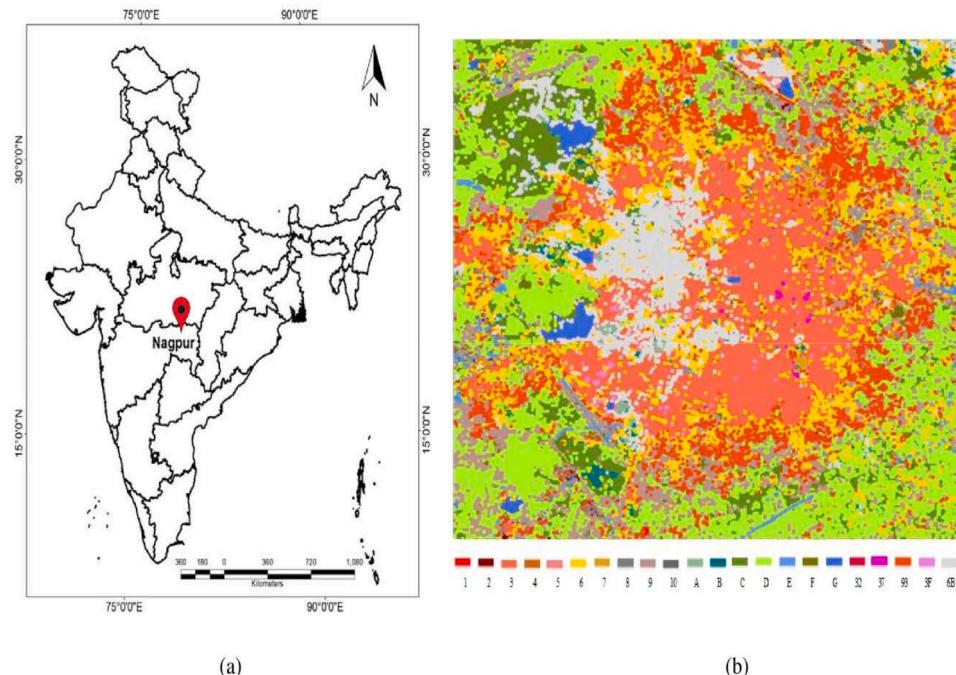


Fig. 2. Study area (a) Location of study area in administrative map of India (b) LCZ map of study area.

identify their boundaries. The foundation of DEALB lies in the formation and clustering of superpixels (a group of connected pixels exhibiting similar properties). An image (pixel-based LCZ raster) is initially partitioned into *superpixels* using directional edges at different angles (0, 90, 45, and 135°) within a spatial unit (grid) of 90 m. Next, similar but spatially cohesive superpixels are merged to form large homogeneous, coarser regions represented by polygons using clustering algorithm. The clustering algorithm within DEALB involves breadth-first search technique and deque data structure to maintain spatial cohesion among superpixels. The LCZ output represented by DEALB is effective for studying the LCZ-adjacent LCZ relationship.

2.3. LST raster

Generally, temperature pattern is not homogeneous within all areas in the city. LCZ, a systematic tool to classify any urban region for UHI study, is a benchmark to assess thermal behaviour of the cities by temperature observations at screen height (2–4 meter above ground) (Yang et al., 2018). Land surface temperature (LST) extracted from satellite imagery is an effective way to understand thermal behaviour of any city using LCZ (Middel et al., 2014; Cai et al., 2017; Geletić et al., 2016) as it is directly impacted by the transformation of natural landscapes to built-up forms (Chen et al., 2017; Jenerette et al., 2016), and influenced by land use/cover composition characteristics (Mills et al., 2021). LST is highly linked with morphological features such as albedo (Mills et al., 2021), building density and floor area ratio (FAR) (Guo et al., 2016; Giridharan et al., 2005), land use configuration and landscape structure (Zhao et al., 2018) which are nothing but the features for LCZ. Hence, understanding the relationship between LST and features (structural and thermal) explained by LCZ helps to design certain temperature mitigation plans. Also, the LST characteristics of LCZ classes are helpful to minimize the impact of urbanization by using urban vegetation and land use planning (Cai et al., 2018).

LST raster used in this paper is a GeoTIFF image generated from thermal infra-red sensor (TIRS, Band 10) and normalized difference vegetation index (NDVI, Band 4 and 5) values of Landsat 8 scenes. We have generated eight LST rasters (four per season, summer and winter) for day-time analysis by using the process explained in Section 3.1.

We have used Landsat 8 imagery for LST computation. Eight Landsat 8 scenes for two seasons (summer and winter) were used for day-time LST retrieval in 2019. The Landsat 8 scenes with cloud cover less than 5 % were downloaded from US Geological Survey Earth Explorer. The dates of acquisition of all Landsat 8 scenes are given in Table 1. The coverage area of all scenes was beyond the expected study area; all the scenes were cropped as per the region of interest. Among 11 bands of Landsat 8 image, band 1–9 are Operational Land Imager (OLI) spectral bands with 30 m resolution whereas band 10–11 are thermal bands with 100 m resolution which were resampled with pixels size of 30 m to match the optical bands (1–9).

3. Methodology

The proposed methodology is divided into two phases, LST prediction and LST analysis of each LCZ by considering adjacent LCZ

Table 1
Landsat 8 scenes and their utilization for the study area.

Season	Landsat 8 scene	Acquisition date
Summer	LC08_L1TP_144045_20190329_20200829	29th Mar 2019
	LC08_L1TP_148044_20190410_20200828	10th Apr 2019
	LC08_L1TP_144045_20190430_20200829	30th Apr 2019
	LC08_L1TP_144045_20190516_20200828	16th May 2019
Winter	LC08_L1TP_144045_20191124_20200825	24th Nov 2019
	LC08_L1TP_144045_20191210_20200824	10th Dec 2019
	LC08_L1TP_144045_20200111_20200823	11th Jan 2020
	LC08_L1TP_144045_20200127_20200823	27th Jan 2020

configuration. The first part is the prediction of LST given current LCZ and its neighbours. The prediction part is proposed first time in the context of LCZ to validate the LST variation. The analysis phase consists of estimating the affecting surrounding LCZ pattern on LST of each LCZ. The diagrammatic representation of all phases of methodology is depicted in Fig. 3 and the necessary modules are explained below.

3.1. LST computation

Land Surface Temperature (LST) (representing the degree of land surface warming (Hulley et al., 2019)) is the important consideration in urban climate studies (Ferreira & Durate, 2019). By using the surface reflectance values for band 4, 5 and 10 in Landsat 8 imagery, we have computed day-time LST as per following equations:

$$R = mf * \text{Band}10 + af \quad (1)$$

$$BT = \frac{K2}{\ln(K1 / (\text{radiance} + 1))} - 273.15 \quad (2)$$

$$PV = [(NDVI - NDVImin) / (NDVImax + NDVImin)]^2 \quad (3)$$

$$E = PV * 0.004 + 0.986 \quad (4)$$

$$LST = (BT + 1) + W * \left(\frac{BT}{14380} \right) * \ln(E) \quad (5)$$

Here, R is Band 10 radiance whereas, *mf* and *af* are the multiplicative and additive factors for band 10 fetched from the metadata file downloaded along with Landsat 8 scenes. For the study area they are 3.3420E-04 and 0.10000 respectively. *Band10* is the digital number. Eq. (2) computes the brightness temperature by using K1 and K2 constants for Band 10. In Eq. (2) *radiance* is the Band 10 radiance value. For the study area, K1 and K2 are 774.8853 and 1321.0789 respectively. Next, fractional vegetation (PV) and land surface emissivity (E) are calculated by using eq. (3) and (4) from normalized difference vegetation index (NDVI). Finally, we get LST using all the values calculated above from eq. (5). In eq. (5) W is the wavelength of emitted radiance which is 10.8 for band 10 of Landsat 8 image. The final result shows that the LST for the study area is in the range between 14.36 °C to 53.08 °C. The hottest and coolest LCZ observed for the study area in summer were LCZ A and LCZ 1 whereas LCZ G and LCZ 2 were the hottest and coolest LCZ in winter 2019. LCZ wise LST statistics is given in Table 2.

3.2. Building LCZ-LST repository

To verify the relationship of LCZ-LST-adjacent LCZs, it is crucial to map LST with the corresponding LCZs. In our study, we mainly focus on limited parameters bypassing the morphological parameters, spectral indices, landscape metrics and decision regarding the spatial scale. Proposed feature set includes the parameters: 1) LST, 2) The name of LCZ, 3) Area of LCZ, 4) Perimeter of LCZ, and 5) Adjacent LCZ pattern (ratio of area of adjacent LCZ to the area of corresponding LCZ, ratio of shared boundary length to the perimeter of corresponding LCZ). These parameters denote the measure of adjacency hence we hypothesized that they may partially affect the LST of considered LCZ. This limited feature set is capable in capturing the relationship of LST with landscape's internal and external spatial heterogeneity at regional scale because of following reasons. The reasons listed below are based on the observations of earlier work of authors (Vaidya et al., 2024).

1. The spectral, textural and contextual features representing internal variation within LCZ embedded in Landsat 8 dataset were already considered while creating LCZ classification (Vaidya et al., 2024).
2. The LCZ map represents homogeneous regions thereby making it efficient to work on polygon/regional level (Vaidya et al., 2024).

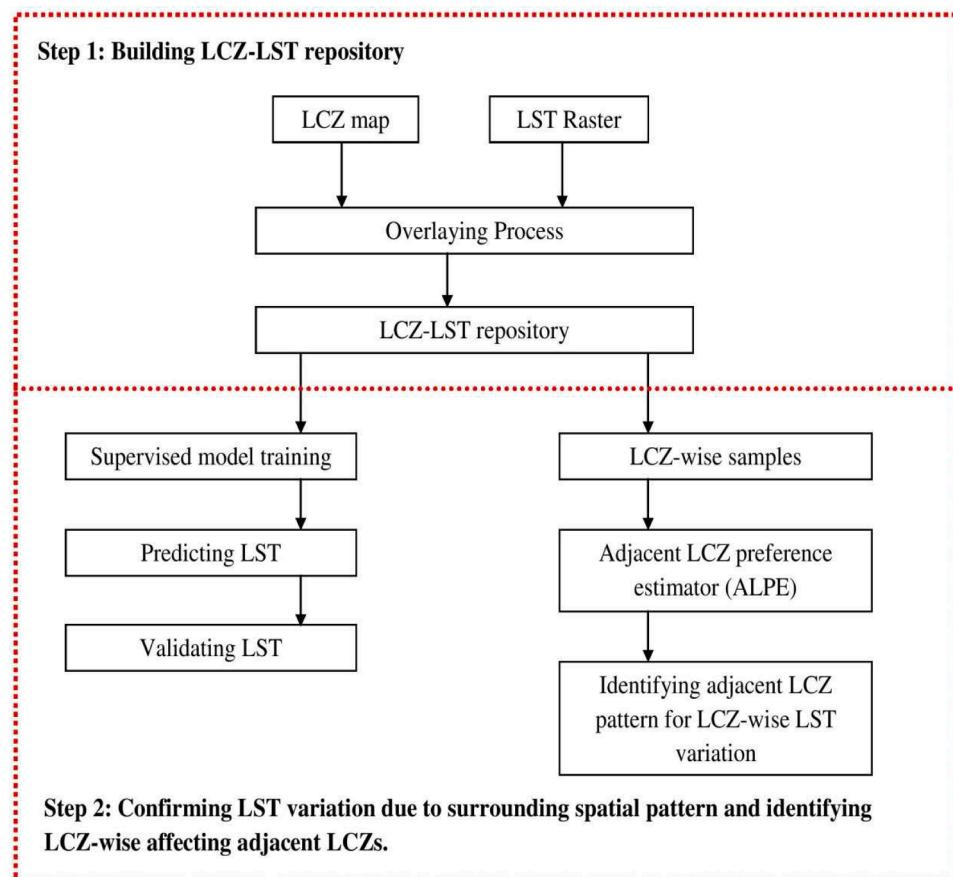


Fig. 3. Flowchart of the methodology.

Table 2

LST statistic for individual LCZ during 2019.

LCZ	LST					
	Summer			Winter		
	Max	Min	Mean	Max	Min	Mean
Built	1-Compact high-rise	53.08	32.04	43.22	29.65	21.14
	2-Compact mid-rise	49.68	33.81	43.90	32.20	20.86
	3-Compact low-rise	48.54	32.84	44.06	29.25	15.36
	4-Open high-rise	47.20	37.30	43.69	27.91	21.88
	5-Open mid-rise	50.35	32.54	43.11	29.05	20.60
	6-Open low-rise	51.82	33.74	44.31	28.07	18.98
	7-Lightweight low-rise	50.40	39.94	44.93	29.55	21.64
	8-Large low-rise	51.64	36.91	45.51	30.90	21.58
	9-Sparsely built	51.90	32.42	46.73	28.64	16.33
	10-Heavy industry	—	—	—	—	—
Natural	A-Dense trees	49.83	31.96	42.56	26.97	20.65
	B-Scattered trees	50.77	33.48	43.77	27.26	20.60
	C-Bush, scrub	51.95	36.76	46.38	27.74	20.89
	D-Low plants	52.43	36.10	46.01	30.95	21.28
	E-Bare rock or paved	51.43	38.89	45.74	31.17	21.11
	F-Bare soil or sand	52.05	33.48	48.28	31.29	15.32
	G-Water	48.80	32.04	41.06	26.42	14.36
Heterogeneous	3 _F	47.88	34.25	42.63	27.71	21.50
	6 _B	49.81	32.49	42.17	28.03	20.67
	3 ₂	47.92	28.87	41.55	25.55	21.79
	3 ₇	47.24	40.38	43.70	26.71	21.88
	9 ₃	50.31	38.10	45.32	27.97	21.79

In our study, the LST-LCZ repository is created by using LCZ map and LST raster for the region of interest. The LCZ map is overlaid on the LST raster to collect the LST within LCZ. For each LCZ (represented by polygon) all the adjacent LCZs are determined. Two vectors are used to

represent adjacent LCZs parameters namely, area and shared boundary with the considered LCZ. Table 3 shows the detailed list of parameters considered to accomplish the mentioned objective. Any LCZ may be surrounded by distinct LCZs or there may exist similar LCZs but at

Table 3
Feature set description.

Feature	Description	Representation
LST	Land surface temperature	Single valued
LCZ	Landscape pattern	Categorical
Area	Area of landscape pattern	Single valued
Perimeter	Perimeter of landscape pattern	Single valued
Adjacent_LCZ_Area (Ratio)	Adjacent landscape area	Vector
Adjacent_LCZ_shared_boundary (Ratio)	Adjacent landscape shared boundary	Vector

different locations as shown in Fig. 4. It shows that the LCZ A is surrounded by single instances of C and D and two instances of LCZ B. Consider the 5 distinct LCZs e.g. A, B, C, D and E. The area and boundary vectors (ratio) for LCZ A are represented as- Adjacent_LCZ_Area= [0, A_B, A_C, A_D, 0], Adjacent_LCZ_shared_boundary= [0, L_B, L_C, L_D, 0]. For the similar LCZ patterns weighted sum for the area and shared boundary are calculated. By following the above mentioned LCZ-LST repository creation process, we got 59,228 polygons representing different LCZs out of which 47,382 polygons were used for training and 11,846 were kept for testing the analysis models. The heterogeneous LCZs (3₂, 3₇, 6_B, 3_F) were omitted in analysing the effect of adjacent LCZs. The LCZ wise training and testing polygons distribution is given in Table 4. The algorithm used for generating LCZ-LST repository is as follows.

```

build_LCZ_LST_dataset(LCZ_map_shapefile, LST_raster_geotifffile)

• for each polygon/region in LCZ_map
  1. Extract the LST pixels within polygon
  //The resolution of each pixel within polygon is 30 m.

  2. Compute weighted average to get LST for LCZ polygon
  3. Identify adjacent LCZs for considering LCZ polygon
  4. For each adjacent LCZ
    a. Compute its area, shared boundary length (ratio) with current
       LCZ polygon
    b. Update the area and shared boundary vector for LCZ polygon
  5. Save the parameters calculated in step 2,3 and 4 for polygon

```

3.3. Model building

We used the regression model for the first part of our methodology i.

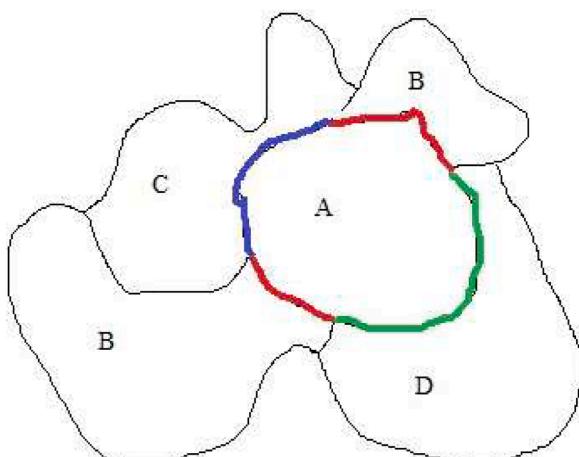


Fig. 4. Adjacent LCZ and their shared boundary representation.

Table 4
Train test distribution of LCZ polygons.

LCZ	Polygons used for model building		
	Total	Train	Test
1	924	739	185
2	933	746	187
3	3832	3066	766
4	111	89	22
5	1541	1233	308
6	10,836	8669	2167
7	1398	1118	280
8	1415	1132	283
9	8416	6733	1683
A	1281	1025	256
B	3957	3166	791
C	5644	4515	1129
D	10,554	8443	2111
E	1688	1350	338
F	684	547	137
G	256	205	51
3 ₂	149	119	30
3 ₇	436	349	87
3 _F	1974	1579	395
6 _B	3199	2559	640

e. prediction of LST. A statistical model describing relationship between dependent and independent features is termed as a regression model. The proposed methodology incorporated two types of supervised regression methods based on tree and deep learning techniques. Tree based learning is popular as it is capable to capture non-linear relationship while deep learning methodologies are capable to create new features from limited set to perform advanced analysis. Under tree based learning, we applied *random forest (RF)* and *Xgboost* regression techniques whereas feed-forward neural network was tested under deep learning mechanism. All the regression models proposed in this paper were trained and tested on the data distribution given in Table 4. The details of all regression models are represented in Table 5.

Random forest regression (RFR) is utilized for the proposed methodology, as it can quantitatively analyze the effect of each factor on LST (Wang et al., 2022). RFR is a non-linear machine learning algorithm and recently it has been applied in the fields like revealing the arsenic sources in private well (Giri et al., 2023), predicting meteorological drought and precipitation index (Elbeltagi et al., 2023) etc. The basis of RFR is the decision tree and it is able to remove multicollinearity among independent variables during the regression process. It tries to establish several parallel regression trees which later on combine to generate final predicted value. To maximize the accuracy of the proposed RFR model, we tuned two parameters, namely number of trees (*n_estimators*) and number of features to consider for best split (*max_features*). The *n_estimators* was tested from 50 to 500 with the interval of 50 with *max_features* set to a value equal to square root of total number of features. The optimal value of *n_estimators* for the proposed dataset is observed as 200.

The next tree based learning model proposed in this paper is XGBoost. It is a framework for joint decision making using multiple correlated regression trees. A regression tree input sample is correlated with the training and prediction of the previous regression tree, improving model performance and reducing the final variance of the model. The recent problems utilising Xgboost regression in climate

Table 5
Model description.

Model	Parameter details
Random forest	<i>n_estimators</i> =200, <i>max_features</i> =square root of total features used
Xgboost	<i>learning_rate</i> =0.2, <i>n_estimators</i> =100
Deep learning	(input, 2 hidden layer, output layer), optimizer=adam, <i>loss_function</i> = mean_squared_error

modelling include predicting global warming potential of thermal insulation materials (Tajuddeen et al., 2023), relationship between seasonal LST and spectral indices using LCZ (Xiang et al., 2023) etc. The XGBoost regressor implemented in this study by tuning the hyper-parameter namely *learning rate* and *n_estimators*. The parameter *n_estimators* set to default value of 100. The learning rate was tested from 0.1 to 0.3 and optimal results were obtained at 0.2.

Deep learning regression techniques are popular over machine learning methodologies like support vector machine, k nearest neighbour, linear regression etc. A deep learning model is able to learn more complex and advanced features from limited dataset. Over the last decade, deep learning is the state-of-the art in various remote sensing applications (Yuan et al., 2020). Under deep learning framework, we created a feed-forward neural network with input layer, two hidden layers and an output layer. We trained the model for 1800 epochs with *adam* optimizer and used mean squared error as the *loss function*.

3.4. Adjacent LCZ preference estimator (ALPE)

As described earlier, the goal of this study is to identify the effect of internal and external spatial heterogeneity on LST of each considered LCZ. For this, it is necessary to rank the features representing spatial heterogeneity. We bypassed the internal morphological factors of LCZ as LCZ itself represents a landscape pattern having certain range of values for different morphological parameters (sky view factor, building height, surface albedo etc.). For representing a surrounding/external heterogeneity of LCZ, adjacent LCZ types are extracted along with their sizes, length of shared borders. We have used two vectors to represent these features (see Table 3).

In this work we have proposed the module named Adjacent LCZ Preference Estimator (ALPE), to find out the substantial effect of neighbourhood LCZ pattern on the LST of an individual LCZ. It decides the preference in which adjacent LCZs affects LST of the individual LCZ. It is a mutual information-based model trained and tested on individual LCZ samples. It considers the LST (day-time for two seasons summer and winter) of each LCZ as a dependent parameter while independent parameters include area, length of shared boundary of each surrounding LCZ represented by ratio with the corresponding LCZ type. The proposed mutual information based ordering system measures the dependency between parameters using entropy estimation between K-nearest neighbours. The value of 'K' used for the proposed model is 3. The mutual information value equal to 0 between two features indicates they are independent and higher value indicates higher dependency. The ALPE module was tested for primary LCZ types (LCZ 1–9 and LCZ A–G) whereas heterogeneous LCZ samples were omitted. Though LCZ D (10,554) and 6 (10,836) samples are sufficient, other LCZ samples are limited. Hence, a repeated 5-fold cross validation (CV) mechanism is applied while ranking the features. During the repeated CV process, the available dataset is divided into 5 subsets. We took 4 subsets as training data, remaining one subset as testing data and then calculated the mutual information score for all the considered parameters. This process is repeated 5 times and the final ranking score for each feature is calculated from the average score of cross validation process.

4. Result

4.1. Model performance

To predict the LST by identifying its variation due to neighbourhood LCZ types, we have proposed three regression models based on tree and deep learning with LST as a dependent parameter and LCZ, LCZ's self area, LCZ's self perimeter, and surrounding LCZ pattern (area vector, shared boundary vector) as independent parameters. The R-squared (R^2), mean absolute percentage error (MAPE) and mean absolute error (MAE) are the metrics used to evaluate the performance of each model (see Table 6). The squared correlation between observed and predicted

Table 6
Model statistics.

Model	Summer			Winter		
	MAE	MAPE	R^2	MAE	MAPE	R^2
RF	1.04	2.7321	0.83	1.43	2.7456	0.79
Xgboost	0.96	2.5106	0.86	0.66	2.4628	0.81
Deep Learning	0.94	2.4355	0.92	0.59	2.2813	0.90

values is represented by R^2 . The mean absolute percentage error (MAPE) also known as mean absolute percentage deviation is generally used to measure the temperature forecast accuracy. MAPE less than 5 % is considered as forecast being acceptably accurate. Mean absolute error (MAE) represents the value of the error that can be expected for the forecast on average. The value of MAE closer to zero is considered as the accurate model. The LCZ-LST data with set of features mentioned in Section 3.2 is divided into training and test set with 80:20 ratios to build the mentioned regression models.

As per the error metric statistics given in Table 6, accuracy increases and bias decreases (i.e. error between actual and predicted value) when we switch from the tree to deep learning models. All R^2 values are more significant than 0.7 indicate that more than 70 % LST variations can be explained by the influencing independent factors under consideration. Furthermore, R^2 (0.92 for summer dataset and 0.90 for winter dataset) of deep learning model captures the 90–92 % of LST variation which is higher than RF (0.83, 0.79) and Xgboost (0.86, 0.81). This indicates that deep learning model is better at capturing relationship represented by adjacent LCZ pattern vector (Area vector, Shared boundary length vector) and hence, it is further utilised to verify LST variation using different feature sets representing LCZ surrounding information. The deep learning model is able to learn complex relationship for vector used to represent surrounding LCZ pattern details than tree based learning but it requires large amount of data.

4.2. Validation of LST variation due to surrounding LCZ pattern

Among the three models proposed for the LST prediction to validate its variation due to surrounding LCZ configuration details, the *deep learning* model results in lower prediction error (see Table 6) as compared to *random forest* and *Xgboost*. Hence, deep learning model is used for the study area for identifying whether the surrounding LCZ pattern is affecting the LST variation or not. The main objective is to check whether the LST can be predicted correctly with a single parameter i.e. name of LCZ or additional information is required. To test this, deep learning model (specified in Section 4.1) was tested with four kinds of feature sets described in Table 3 with LST as dependent parameter. For the performance testing, we used *bias* as a metric. Bias is the mean error calculated from actual and predicted values for the test samples. Table 7 describes the bias statistics for five types of feature set on the summer and winter LST in the study area. From the Table 7; it is observed that, with the LCZ name only, the LST can be estimated with accuracy 94.58 % (summer) and 89.28 % (winter) which are more than

Table 7
Feature wise accuracy and bias statistics for deep learning model.

#	Feature set	Summer		Winter	
		Bias (°C)	Accuracy (%)	Bias (°C)	Accuracy (%)
F1	{LCZ}	2.35	94.58	1.64	89.28
F2	{LCZ, Self area, Self perimeter}	1.53	95.97	0.74	90.56
F3	{LCZ, Self area, surrounding LCZ patterns (area)}	1.11	96.18	0.66	92.79
F4	{LCZ, Self area, surrounding LCZ patterns (area + shared boundary length)}	0.96	97.2	0.59	96.94

satisfactory but the bias reduces (by 1.39 °C in summer and 1.05 °C in winter) when the learning model is fed with additional information of adjacent LCZs. This proves that, the characteristics of neighbouring LCZs have certain influence on the LST of considered LCZ in the study area.

4.3. LCZ map statistics

LCZ map configuration is the primary requirement for LCZ-LST. We generated LCZ map (for the year 2019) for the study area by using remote sensing based classification technique (Vaidya et al., 2024) mentioned in Section 2.2. The Generated LCZ map is represented by Fig. 2(b). In LCZ map, a large proportion of LCZ 3 was discovered in the center surrounded by small portions of LCZ 6, 9₃. Towards the periphery, existence of natural class LCZ D and F mixed with urban LCZs like 6, 3 and 9₃ was observed. The land cover area contribution (%) of each LCZ and their surrounding LCZ pattern observed during 2019 for the study area is given in Fig. 5 (a) and Fig. 5 (b). Five different colors are used in Fig. 5(a) and (b) to represent top 5 surrounding LCZs according to their area and length of shared boundary with considered LCZ. The color follows ranking specified as order shown on the top of the Fig. 5(a) and (b). It shows the percentage contribution of shared boundary length with other LCZs for each LCZ. The top 5 largest surrounding LCZs (area-wise and shared boundary wise) are represented by five different colours in Fig. 5 (a) and (b). Fig. 5(a) shows that, the built LCZs (1–9) in the study area are surrounded by large proportion of compact low-rise LCZ (LCZ 3) and open low-rise LCZ (LCZ 6) whereas in natural LCZ type (A-G) the surrounding pattern constitutes LCZ C (Bush, scrub), D (Low plants), G (water) and 6_B (Open low-rise + scattered trees). From the Fig. 5(b), it can be stated that, the built LCZs (1–9) shares the largest boundary proportion with LCZ 3 (compact low-rise), 6(open low-rise), D (low plants), 9₃ (sparsely built + compact low-rise) and 6_B (Open low-rise + scattered trees) whereas natural LCZs (A-G) shares the largest boundary length with natural LCZs like LCZ C (Bush, scrub) and D (low plants).

4.4. Impact of adjacent LCZs on LST of individual LCZs

Generally, the built-type LCZ recorded higher LST than natural-type LCZ (Zhao, 2018), whereas cities with complex urban morphology experienced an inconsistent LST pattern (Zhao et al., 2018). Researchers have identified these types of observations by investigating the relationship between LST and different LCZs. For the given study area, it is observed that similar LCZ types located at different places in the city

experience substantial LST variation when they are surrounded by a different combination of LCZ pattern (see Fig. 1). In Section 4.2 (LST prediction), we proved that the bias in estimating LST is reduced by adding neighbourhood LCZ pattern information. Hence, it is necessary to study the effect of surrounding spatial heterogeneity for each LCZ. The effect of surrounding LCZ pattern is analysed in this study by using two parameters i.e. the area of LCZ and its length of shared boundary with the considered LCZ (ratio). The substantial effect of surrounding LCZ pattern on LST of each LCZ is estimated by identifying the importance value of a particular feature. The ALPE module estimates the rank of the feature by averaging the mutual importance value collected from 5-fold cross validation process. The feature set for ALPE consist of LST as dependent variable and LCZ-wise independent features i.e. Area of LCZ, Adjacent LCZ pattern (vector representing area and length of shared boundary). We calculated the affecting adjacent features for the basic LCZs (LCZ 1–10 and LCZ A-G) of LCZ map for the 2019 of the study area in summer as well as winter. The ALPE result (LCZ-wise mutual information score for each feature considered) for summer-time LST variation for the study area on two independent feature sets (area and shared boundary length), is given in Table 8 (A) and (B). Table 9 (A) and (B) shows ALPE result for winter-time LST variation.

Table 8 (a) and 9 (a) shows the area-wise surrounding LCZs responsible for summer and winter time LST variation in each LCZ. From the Table 8 (a), it is observed that, in built LCZs (1–9), the topmost surrounding LCZs (area-wise) responsible for summer-time LST variation of each LCZ include LCZ 5 (open mid-rise), 6 (open low-rise), 9 (sparsely built), 6_B (open low-rise + sparsely built), 9₃ (sparsely built + compact low-rise) and D (low plants) whereas in natural LCZ types, the LCZ 3 (compact low-rise), 6 (open low-rise), 6_B (open low-rise + scattered trees), 9 (sparsely built), A (dense trees) and B (scattered trees) are responsible for LST variation. From Table 9 (A), it is observed that, topmost surrounding LCZ pattern (area-wise) responsible for winter-time LST variation of spatially dispersed built LCZs (1–9) consist of LCZ 6 (Open low-rise), 6_B (open low-rise + scattered trees), 8 (Large low-rise), 9₃ (sparsely built + compact low-rise) and E (Bare rock or paved). Surrounding LCZs (area-wise) responsible for Winter-time LST variation in natural LCZs includes LCZ 6 (open low-rise), 6_B (open low-rise + scattered trees), 8 (large low-rise), 9 (sparsely built), B (scattered trees) and D (low plants). To summarise, in summer as well as winter, LCZs (area-wise) which are open in nature are responsible for LST variation in LCZs.

Table 8 (b) and 9 (b) shows the length of shared boundary-wise

LCZ	Area (%)	Surrounding Area(%)																				
		Built LCZ									Natural LCZ											
		1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	G	3 _F	6 _B	3 ₂	3 ₇	9 ₃
1	0.4	0	2.36	18.4	0.08	4.98	8.19	1.37	2.69	1.56	0.2	0.42	1.91	9.16	7.62	0.01	33.6	0.36	2.48	0.02	0.15	4.43
2	0.36	7.09	0	23.9	0.29	2.38	7.92	2.38	5.99	2.15	0	0.64	1.25	10.7	9.55	0.91	12.4	1.01	2.6	0.03	0.28	3.56
3	7.71	2.15	1.07	0	0.1	1.91	35.5	3.18	2.24	2.39	0.67	1.57	1.88	12.9	8.2	0.04	4.25	1	0.67	0.03	0.66	13.6
4	0.04	1.54	2.49	45.3	0	0.43	25.6	0.06	8.6	0.55	0	3.41	2.18	0.68	2.7	0	0.52	0.34	0.71	0.03	0	4.88
5	0.87	1.62	0.51	15.4	0.02	0	5.99	0.29	0.54	1.26	9.47	1.33	14.6	8.06	1.17	0.01	25.6	0.27	11	0.06	0.01	2.93
6	9.85	0.42	0.21	21.4	0.06	0.47	0	0.64	0.67	4.87	1.75	3.74	11.7	20.2	1.86	0.07	14	0.85	4.78	0.01	0.1	12.3
7	0.66	1.35	0.76	35.4	0.01	0.25	14.5	0	3.53	3.26	0.02	0.09	2.22	9.93	15.2	0.36	0.02	0	0.13	0.01	2.11	10.8
8	0.65	2.8	1.49	21.5	0.11	0.46	13.9	3.23	0	7.12	0	0.44	3.41	19	10.4	1.79	0.03	0.07	0.11	0	0.19	14
9	6.05	0.19	0.12	3.2	0.02	0.16	11.7	0.2	0.54	0	0	0.27	23.4	37.3	3.36	3.42	0.17	0.02	0.08	0	0.01	15.9
10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
A	1.89	0.02	0	0.48	0	0.29	3.5	0.01	0	0	0	10.6	12.8	4.45	0.08	0	52.5	0.05	15.2	0.01	0	0.04
B	2.79	0.05	0.01	2.53	0.02	0.19	7.69	0.02	0.06	0.57	14.4	0	17.3	11.1	0.39	0	34.9	0.11	10.3	0	0	0.37
C	15.36	0.05	0.01	1.32	0.02	0.36	9.32	0.04	0.12	11.8	3.05	6.84	0	46.9	1.24	0.11	10.8	0.01	2.32	0	0	5.67
D	37.21	0.23	0.21	6.77	0.05	0.35	13.9	0.32	0.73	15.3	1.35	4.22	31.9	0	4.45	0.81	7.68	0.2	1.75	0	0.02	9.77
E	1.53	1.33	0.89	21	0.02	0.24	10.2	3.62	2.37	10.4	0	0.75	7.51	29.7	0	1.57	2.13	0.02	0.56	0	0.17	7.46
F	0.55	0.01	0.99	2.61	0	0.05	0.22	0.64	0.3	41.1	0	0.01	1.06	37.2	9.28	0	0.03	0	0.01	0	0	4.54
G	1.41	4.46	1.3	12.4	0.02	4.72	8.22	0.12	0.07	1.03	22.3	3.28	6.45	8.81	1.91	0.01	0	0.68	23.5	0.03	0.27	0.49
3 _F , 6 _B , 3 ₂ , 3 ₇ , 9 ₃	12.68																					

Fig. 5a. Area contribution (%) of each LCZ and its surrounding LCZ pattern during 2019 showing Top 5 (specified by different colors) largest surrounding LCZ (area-wise) for individual LCZ.

1:Green 2:Yellow 3:Blue 4:Purple 5:Red

LCZ	Adjacent LCZs shared border proportion (%)																				
	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	G	3 ₁	3 ₂	9 ₃	3 _F	6 _B
1	0	10.39	19.33	0.56	8.42	11.42	3.94	6.65	1	0.15	0.71	0.35	9.98	8.8	0.03	7.03	0.15	0.35	6.36	1.12	3.27
2	11.83	0	21.12	1.17	4.01	8.39	4.88	9.32	1.74	0	0.23	0.1	8.42	8.79	2.21	1.91	0.1	0.23	10.86	2.94	1.74
3	2.71	2.61	0	0.39	3.18	38.23	6.46	4.47	1.55	0.1	0.92	0.26	7.5	6.03	0.09	0.56	0.15	1.52	13.63	3.4	6.23
4	4.96	9.14	24.54	0	2.09	21.41	0.26	9.14	0.78	0	0.26	1.04	12.53	3.39	0	0.26	0.52	0	8.36	1.04	0.26
5	5.66	2.37	15.23	0.16	0	12.27	0.63	1.7	1.94	1.05	2.55	10.09	11.93	1.38	0.06	3.01	0.73	0.06	5.32	1.27	22.59
6	0.74	0.47	17.4	0.15	1.17	0	1.58	1.66	7.93	0.55	5.7	5.16	20.63	1.68	0.03	0.24	0.05	0.27	25.02	3.56	0.03
7	2.57	2.8	29.98	0.02	0.61	16.09	0	9.08	3.26	0.02	0.04	0.56	5.71	13.06	0.36	0.02	0.08	5.69	9.94	0	0.13
8	3.7	4.51	17.53	0.57	1.39	14.26	7.67	0	6.94	0	0.29	0.82	14.25	8.38	2.26	0.03	0.02	0.27	16.8	0.23	0.08
9	0.09	0.14	1.04	0.01	0.27	11.66	0.47	1.19	0	0	0.42	15.18	42.35	1.81	0.28	0	0	0.04	22.38	0.05	0.04
A	0.07	0	0.33	0	0.76	4.22	0.01	0	0	0	34.55	9.27	6.75	0	0	2.1	0.01	0	0.01	0.22	41.69
B	0.14	0.04	1.34	0.01	0.77	18.15	0.01	0.11	0.91	14.42	0	15.63	21.39	0.34	0.01	0.61	0.02	0.01	0.65	0.5	24.95
C	0.04	0.01	0.21	0.01	1.7	9.16	0.1	0.17	18.31	2.15	8.7	0	52.28	0.26	0.05	0.06	0	0	4.75	0.01	2.02
D	0.53	0.39	2.84	0.08	0.94	17.15	0.47	1.38	23.94	0.73	5.58	24.49	0	4.47	1.83	0.29	0	0.02	12.71	0.64	1.51
E	3.43	3	16.66	0.15	0.8	10.19	7.78	5.91	7.47	0	0.65	0.9	32.62	0	2.04	0.38	0.02	0.48	6.92	0.06	0.55
F	0.04	2.42	0.81	0	0.11	0.51	0.7	5.14	37.44	0	0.04	0.59	43.02	6.57	0	0.02	0	0	2.57	0	0.04
G	0.04	2.4	0.8	0.02	0.11	0.51	0.69	5.09	37.09	0.48	0.04	0.58	42.61	6.51	0.02	0	0.04	0.36	2.55	0.04	0.04

Fig. 5b. Length of shared boundary contribution (%) for each LCZ with its surrounding LCZ pattern during 2019 showing Top 5 (specified by different colors) largest surrounding LCZ (shared boundary length-wise) for individual LCZ.

Table 8a

LCZ-wise ALPE result for summer-time LST by considering the area of adjacent LCZs.

#	Built LCZ									Natural LCZ						
	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	G
A_1	0.015	0.076	0.014	0.020	0.033	0.001	0.006	0.047	0.003	0.001	0.001	0.000	0.009	0.038	0.000	0.061
A_2	0.041	0.000	0.011	0.007	0.022	0.002	0.019	0.004	0.001	0.001	0.001	0.001	0.006	0.021	0.067	0.037
A_3	0.081	0.114	0.036	0.112	0.111	0.041	0.072	0.092	0.031	0.006	0.032	0.006	0.050	0.154	0.012	0.018
A_4	0.000	0.001	0.002	0.067	0.001	0.000	0.000	0.009	0.000	0.000	0.000	0.000	0.001	0.004	0.000	0.000
A_5	0.114	0.071	0.073	0.118	0.094	0.006	0.000	0.021	0.003	0.028	0.012	0.009	0.005	0.011	0.000	0.098
A_6	0.051	0.020	0.089	0.000	0.141	0.037	0.025	0.076	0.199	0.023	0.046	0.080	0.170	0.106	0.022	0.000
A_7	0.043	0.102	0.042	0.000	0.024	0.022	0.024	0.015	0.005	0.000	0.001	0.002	0.010	0.053	0.020	0.000
A_8	0.098	0.096	0.039	0.039	0.025	0.020	0.048	0.000	0.009	0.000	0.007	0.002	0.016	0.037	0.051	0.000
A_9	0.005	0.069	0.044	0.000	0.049	0.059	0.064	0.105	0.018	0.000	0.030	0.036	0.057	0.103	0.072	0.000
A_A	0.000	0.000	0.009	0.000	0.068	0.008	0.000	0.000	0.000	0.027	0.105	0.066	0.011	0.000	0.000	0.122
A_B	0.005	0.005	0.029	0.011	0.011	0.008	0.000	0.003	0.007	0.081	0.019	0.083	0.068	0.008	0.000	0.000
A_C	0.002	0.004	0.010	0.040	0.204	0.054	0.014	0.003	0.062	0.068	0.081	0.020	0.053	0.035	0.005	0.000
A_D	0.038	0.119	0.036	0.042	0.170	0.021	0.091	0.084	0.026	0.015	0.076	0.045	0.018	0.061	0.040	0.045
A_E	0.076	0.086	0.068	0.000	0.038	0.018	0.021	0.044	0.011	0.000	0.008	0.008	0.015	0.030	0.054	0.058
A_F	0.001	0.061	0.002	0.000	0.001	0.001	0.015	0.061	0.055	0.000	0.000	0.000	0.026	0.052	0.033	0.000
A_G	0.052	0.044	0.019	0.000	0.082	0.009	0.000	0.000	0.000	0.049	0.017	0.003	0.006	0.006	0.001	0.043
A_3F	0.036	0.056	0.065	0.016	0.027	0.073	0.000	0.024	0.005	0.000	0.014	0.001	0.045	0.003	0.000	0.038
A_6B	0.061	0.063	0.122	0.019	0.289	0.093	0.000	0.000	0.002	0.118	0.147	0.064	0.079	0.037	0.001	0.087
A_32	0.001	0.005	0.004	0.000	0.024	0.002	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000
A_37	0.000	0.002	0.021	0.000	0.005	0.001	0.079	0.011	0.002	0.000	0.000	0.000	0.002	0.010	0.000	0.005
A_93	0.091	0.131	0.114	0.000	0.116	0.104	0.049	0.024	0.109	0.000	0.012	0.033	0.065	0.071	0.070	0.000

surrounding LCZs responsible summer and winter-time LST variation in each LCZ. The topmost LCZs (length of shared boundary-wise) responsible for summer-time LST variation in each LCZ are LCZ 2 (compact mid-rise), 3_F (compact low-rise + bare soil or sand), 3₇ (compact low-rise + lightweight low-rise), 6 (open low-rise), 6_B (open low-rise + scattered trees), 9 (sparsely built), A (dense trees), D (low plants) and E (bare rock or paved). From Table 9 (B), it is observed that, the length of shared boundary of LCZ 3_F (compact mid-rise), 6_B (open low-rise + scattered trees), 8 (Large low-rise), 9 (sparsely built), 9₃ (sparsely built + compact low-rise), A (dense trees), C (Bush, scrub), D (low plants), E (bare rock) and F (bare soil or sand) are responsible for winter-time LST variation in each LCZ.

Further, the mutual information score generated for two feature sets (surrounding LCZ pattern area vector, surrounding LCZ pattern shared boundary vector) on summer and winter LCZ-wise LST variation dataset are sorted and ranked. Top 10 features based on surrounding LCZs area responsible for LST variation in each LCZ are shown in Table 10 ((a) and (b)). The percentage area contribution of surrounding LCZs for each LCZ in study area is shown in Fig. 5 (a). The cumulative observations from Fig. 5 (a) and Table 10 ((a) and (b)) for individual LCZ are observed and

listed below.

- The LCZ 1 (compact high-rise) is surrounded by the combination of all other LCZs with largest proportion of LCZ G (33.6 %) and LCZ 3 (18.4 %). Though LCZ G and 3 exist in majority proportion surrounding LCZ 1, their impact on LST of LCZ 1 is lowest. With less area coverage area, LCZ 5 (4.98 %), LCZ 8 (2.69 %), LCZ 9₃ (4.43 %), LCZ 6_B (2.48 %), and LCZ E (7.62 %) are contributing more in LST variation.
- The adjacent LCZ pattern for LCZ 2 (compact mid-rise) consist of largest portion of LCZ 3 (23.9 %) but it ranked 3rd and 4th in LST variation analysis for summer and winter. LCZ D and E are the third and fourth largest surrounding LCZs for LCZ 2 in terms of their coverage area as well as effect in LST variation. LCZ 9₃ and 8 becomes the topmost contributing LCZs in analysing LST variation of LCZ 2 even with less area coverage of 8.56 % and 5.99 % respectively.
- The summer-time as well as winter-time LST variation of LCZ 3 (compact low-rise) is mostly explained by LCZ 6_B with very less area

Table 8b

LCZ-wise ALPE result for summer-time LST by considering the shared boundary of adjacent LCZs.

#	Built LCZ									Natural LCZ						
	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	G
L_1	0.000	0.074	0.005	0.016	0.019	0.000	0.011	0.014	0.000	0.002	0.000	0.001	0.006	0.033	0.000	0.019
L_2	0.012	0.000	0.000	0.007	0.018	0.001	0.008	0.015	0.002	0.000	0.002	0.000	0.005	0.012	0.057	0.017
L_3	0.013	0.016	0.001	0.016	0.031	0.020	0.032	0.039	0.018	0.001	0.013	0.001	0.026	0.031	0.029	0.024
L_4	0.004	0.005	0.003	0.007	0.000	0.002	0.000	0.008	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000
L_5	0.069	0.041	0.053	0.084	0.005	0.002	0.000	0.003	0.001	0.014	0.013	0.008	0.005	0.009	0.000	0.007
L_6	0.000	0.021	0.020	0.036	0.121	0.005	0.015	0.057	0.090	0.000	0.034	0.025	0.045	0.037	0.010	0.045
L_7	0.045	0.074	0.007	0.000	0.012	0.015	0.000	0.000	0.002	0.000	0.000	0.001	0.004	0.008	0.000	0.000
L_8	0.054	0.034	0.011	0.037	0.016	0.014	0.035	0.000	0.008	0.000	0.003	0.000	0.011	0.030	0.027	0.000
L_9	0.000	0.029	0.027	0.000	0.036	0.022	0.041	0.061	0.001	0.000	0.018	0.017	0.051	0.041	0.041	0.000
L_A	0.000	0.000	0.003	0.000	0.027	0.001	0.000	0.000	0.000	0.012	0.009	0.052	0.009	0.000	0.000	0.000
L_B	0.003	0.006	0.011	0.000	0.010	0.003	0.000	0.003	0.009	0.020	0.003	0.035	0.025	0.006	0.000	0.000
L_C	0.005	0.000	0.008	0.010	0.186	0.028	0.021	0.002	0.020	0.017	0.025	0.003	0.031	0.009	0.001	0.000
L_D	0.040	0.097	0.010	0.040	0.057	0.006	0.041	0.026	0.032	0.012	0.034	0.015	0.003	0.069	0.006	0.000
L_E	0.074	0.072	0.028	0.000	0.027	0.018	0.014	0.000	0.001	0.000	0.004	0.000	0.009	0.000	0.021	0.011
L_F	0.000	0.035	0.001	0.000	0.000	0.001	0.011	0.033	0.048	0.000	0.000	0.000	0.025	0.041	0.001	0.000
L_G	0.018	0.050	0.009	0.000	0.068	0.006	0.000	0.000	0.000	0.048	0.011	0.001	0.008	0.002	0.000	0.006
L_3F	0.027	0.041	0.060	0.000	0.022	0.065	0.000	0.016	0.003	0.000	0.010	0.000	0.033	0.001	0.000	0.081
L_6B	0.039	0.018	0.064	0.000	0.060	0.039	0.000	0.005	0.001	0.015	0.040	0.025	0.060	0.034	0.001	0.025
L_32	0.003	0.000	0.002	0.000	0.011	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
L_37	0.005	0.001	0.017	0.000	0.000	0.001	0.060	0.005	0.001	0.000	0.000	0.000	0.000	0.006	0.000	0.004
L_93	0.055	0.081	0.046	0.000	0.067	0.040	0.015	0.009	0.026	0.000	0.005	0.019	0.008	0.044	0.031	0.000

Table 9a

LCZ-wise ALPE result for winter-time LST by considering the area of adjacent LCZs.

#	Built LCZ									Natural LCZ						
	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	G
A_1	0.020	0.008	0.016	0.019	0.010	0.003	0.005	0.032	0.000	0.000	0.000	0.000	0.003	0.025	0.001	0.171
A_2	0.041	0.048	0.012	0.000	0.007	0.005	0.026	0.023	0.003	0.000	0.000	0.000	0.006	0.004	0.002	0.053
A_3	0.033	0.081	0.019	0.074	0.033	0.027	0.025	0.045	0.010	0.011	0.023	0.001	0.017	0.057	0.000	0.009
A_4	0.000	0.004	0.000	0.172	0.006	0.001	0.000	0.002	0.000	0.000	0.000	0.000	0.001	0.002	0.000	0.001
A_5	0.085	0.041	0.054	0.000	0.071	0.001	0.007	0.000	0.003	0.003	0.008	0.000	0.005	0.008	0.002	0.105
A_6	0.047	0.046	0.094	0.050	0.083	0.019	0.036	0.068	0.115	0.071	0.124	0.059	0.079	0.080	0.010	0.000
A_7	0.052	0.068	0.120	0.000	0.009	0.042	0.020	0.034	0.004	0.000	0.002	0.002	0.012	0.047	0.000	0.000
A_8	0.089	0.127	0.071	0.047	0.043	0.024	0.048	0.007	0.006	0.000	0.006	0.001	0.024	0.058	0.080	0.000
A_9	0.026	0.054	0.018	0.000	0.053	0.044	0.005	0.024	0.013	0.000	0.032	0.081	0.087	0.016	0.064	0.003
A_A	0.000	0.000	0.007	0.000	0.049	0.021	0.000	0.000	0.000	0.041	0.131	0.085	0.014	0.000	0.000	0.115
A_B	0.008	0.002	0.023	0.003	0.003	0.045	0.002	0.003	0.005	0.043	0.032	0.103	0.082	0.005	0.000	0.045
A_C	0.018	0.002	0.000	0.002	0.118	0.012	0.000	0.006	0.054	0.036	0.033	0.010	0.038	0.015	0.019	0.000
A_D	0.039	0.105	0.023	0.015	0.103	0.009	0.022	0.007	0.011	0.055	0.144	0.059	0.006	0.040	0.061	0.047
A_E	0.088	0.127	0.089	0.045	0.030	0.027	0.076	0.046	0.021	0.000	0.011	0.006	0.043	0.027	0.009	0.067
A_F	0.000	0.063	0.003	0.000	0.001	0.001	0.025	0.056	0.064	0.000	0.000	0.003	0.044	0.073	0.036	0.000
A_G	0.028	0.039	0.006	0.000	0.038	0.009	0.000	0.000	0.000	0.023	0.016	0.002	0.004	0.000	0.000	0.042
A_3F	0.030	0.044	0.064	0.038	0.027	0.040	0.001	0.012	0.003	0.010	0.018	0.001	0.021	0.003	0.000	0.001
A_6B	0.053	0.059	0.121	0.000	0.228	0.086	0.000	0.004	0.003	0.054	0.104	0.078	0.043	0.028	0.001	0.181
A_32	0.000	0.003	0.006	0.000	0.019	0.002	0.000	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.006
A_37	0.000	0.000	0.029	0.000	0.004	0.005	0.046	0.012	0.001	0.000	0.000	0.000	0.000	0.010	0.000	0.005
A_93	0.041	0.102	0.087	0.245	0.093	0.092	0.035	0.027	0.061	0.000	0.026	0.048	0.058	0.018	0.012	0.005

coverage (6.67 %) as compared to LCZ 6 which is the largest LCZ with area coverage (35.6 %).

- For LCZ 4 (open low-rise), LCZ 3 with largest surrounding area of (45.3 %) becomes the second most influential parameter in explaining LST variation of summer and winter. Open mid-rise LCZ 5 with very less area coverage (0.43 %) becomes the topmost influential parameter in summer whereas in winter its effect nullifies.
- For LCZ 5 (open mid-rise), the set of top 3 surrounding/influencing LCZs is same in winter and summer i.e. LCZ 6_B (11 %), C (14.6 %) and D (8.06 %). In case of LCZ 6 (open low-rise), LCZ 3 contribution in surrounding LCZ is highest i.e. 21.4 % but the combination of open and compact i.e. 9₃ with coverage area of 12.3 % along with 6_B (4.78 %) becomes the most LST influencing feature in summer as well as in winter.
- LST of LCZ 7 (lightweight low-rise) is influenced by LCZs which are open, compact + open and compact in nature. Though, bare rock and paved (LCZ E) exist in very less proportion (15.2 %) in surrounding

to LCZ 7 but it ranked 1st in terms of its effect on LST of LCZ 7 in winter.

- Summer-time LST of LCZ 8 (large low-rise) is highly affected by sparsely built area (LCZ 9) even with less coverage whereas winter-time variation is explained by open low-rise (LCZ 6).
- LCZ 9 is surrounded by large proportion of natural LCZs as compared to built LCZs and in both summer as well as winter.
- LCZ 6 with highest coverage area percentage (11.7 %) among built LCZ affects the LST.
- LCZ A (dense trees) and LCZ B (scattered trees) are surrounded by large proportion of 6_B (15.2 and 10.3 %) among built and heterogeneous LCZs and is topmost LCZ responsible for their LST variation in summer.
- In winter, LCZ 6 (3.5, 7.69 %) and LCZ D (4.45, 11.1 %) influence the LST of LCZ A and B.
- Though LCZ G is covering largest area (52.5, 39.4 %) in surrounding LCZs for LCZ A and B it is not the influencing parameter for their LST variation.

Table 9b

LCZ-wise ALPE result for winter-time LST by considering the shared boundary of adjacent LCZs.

#	Built LCZ									Natural LCZ						
	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	G
L_1	0.000	0.008	0.010	0.000	0.007	0.002	0.002	0.009	0.000	0.008	0.002	0.000	0.003	0.013	0.000	0.012
L_2	0.006	0.000	0.002	0.000	0.003	0.002	0.032	0.026	0.003	0.000	0.000	0.000	0.005	0.000	0.000	0.005
L_3	0.000	0.029	0.000	0.046	0.020	0.008	0.023	0.012	0.000	0.002	0.012	0.000	0.003	0.016	0.008	0.018
L_4	0.006	0.008	0.000	0.000	0.000	0.000	0.000	0.007	0.000	0.000	0.000	0.000	0.001	0.003	0.000	0.000
L_5	0.061	0.013	0.045	0.007	0.005	0.002	0.011	0.000	0.000	0.005	0.004	0.000	0.004	0.000	0.000	0.000
L_6	0.000	0.023	0.018	0.000	0.092	0.001	0.011	0.046	0.043	0.055	0.041	0.015	0.020	0.005	0.010	0.015
L_7	0.040	0.057	0.046	0.000	0.008	0.031	0.000	0.008	0.001	0.000	0.000	0.000	0.006	0.021	0.000	0.000
L_8	0.044	0.082	0.019	0.008	0.020	0.018	0.053	0.000	0.004	0.000	0.003	0.002	0.021	0.024	0.053	0.000
L_9	0.016	0.025	0.013	0.010	0.026	0.015	0.003	0.000	0.000	0.000	0.023	0.031	0.052	0.012	0.021	0.000
L_A	0.000	0.000	0.002	0.000	0.014	0.013	0.000	0.000	0.000	0.012	0.033	0.071	0.008	0.000	0.000	0.000
L_B	0.000	0.002	0.010	0.000	0.004	0.039	0.001	0.001	0.002	0.000	0.005	0.050	0.029	0.003	0.000	0.021
L_C	0.008	0.000	0.000	0.004	0.097	0.006	0.000	0.000	0.020	0.010	0.008	0.001	0.008	0.005	0.018	0.000
L_D	0.041	0.079	0.001	0.025	0.038	0.001	0.000	0.022	0.015	0.028	0.057	0.006	0.001	0.023	0.000	0.000
L_E	0.067	0.091	0.038	0.037	0.026	0.020	0.025	0.044	0.010	0.000	0.007	0.002	0.024	0.001	0.037	0.073
L_F	0.000	0.037	0.003	0.000	0.000	0.013	0.047	0.052	0.000	0.000	0.001	0.039	0.053	0.000	0.000	0.000
L_G	0.020	0.028	0.007	0.000	0.028	0.003	0.000	0.000	0.000	0.025	0.005	0.000	0.001	0.000	0.000	0.000
L_3F	0.007	0.025	0.060	0.018	0.025	0.037	0.000	0.006	0.001	0.008	0.009	0.000	0.011	0.000	0.000	0.006
L_6B	0.033	0.022	0.057	0.000	0.075	0.041	0.000	0.002	0.002	0.010	0.045	0.039	0.033	0.029	0.001	0.007
L_32	0.001	0.000	0.004	0.000	0.008	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003
L_37	0.004	0.000	0.031	0.000	0.001	0.003	0.046	0.010	0.002	0.000	0.000	0.000	0.000	0.007	0.000	0.004
L_93	0.020	0.085	0.032	0.055	0.088	0.027	0.000	0.012	0.013	0.000	0.011	0.025	0.018	0.024	0.000	0.001

Table 10a

Top 10 surrounding LCZs (area-wise) for summer-time LST variation in individual LCZ.

1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	G
A_5	A_93	A_6B	A_5	A_6B	A_93	A_D	A_9	A_6	A_6B	A_6B	A_B	A_6	A_3	A_9	A_A
A_8	A_D	A_93	A_3	A_C	A_6B	A_37	A_3	A_93	A_B	A_A	A_6	A_6B	A_6	A_93	A_5
A_93	A_3	A_6	A_D	A_3F	A_3	A_D	A_C	A_C	A_C	A_A	A_B	A_9	A_2	A_6B	A_6
A_3	A_7	A_5	A_C	A_6	A_9	A_9	A_6	A_F	A_G	A_D	A_6B	A_93	A_93	A_E	A_1
A_E	A_8	A_8	A_93	A_C	A_93	A_F	A_3	A_5	A_5	A_6	A_D	A_9	A_D	A_8	A_E
A_6B	A_E	A_3F	A_1	A_3	A_3	A_8	A_1	A_D	A_6	A_3	A_9	A_C	A_7	A_D	A_D
A_G	A_1	A_9	A_6B	A_G	A_7	A_6	A_E	A_E	A_D	A_9	A_93	A_3	A_F	A_6	A_3F
A_6	A_5	A_7	A_3F	A_A	A_D	A_E	A_3F	A_8	A_3	A_B	A_5	A_3F	A_1	A_7	A_2
A_7	A_9	A_8	A_B	A_9	A_8	A_2	A_93	A_B	A_1	A_G	A_E	A_F	A_8	A_3	A_3
A_2	A_6B	A_D	A_2	A_E	A_E	A_F	A_5	A_3F	A_2	A_3F	A_3	A_8	A_6B	A_C	A_37

Table 10b

Top 10 surrounding LCZs (area-wise) for winter-time LST variation in individual LCZ.

1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	G
A_8	A_8	A_6B	A_93	A_6B	A_93	A_E	A_6	A_6	A_D	A_B	A_9	A_6	A_8	A_6B	A_6B
A_E	A_E	A_7	A_3	A_C	A_6B	A_8	A_F	A_F	A_D	A_A	A_A	A_B	A_F	A_9	A_1
A_5	A_D	A_6	A_6	A_D	A_B	A_37	A_E	A_93	A_6B	A_6	A_9	A_6	A_8	A_D	A_A
A_6B	A_93	A_E	A_8	A_93	A_9	A_6	A_3	A_C	A_B	A_6B	A_6B	A_93	A_3	A_C	A_5
A_7	A_3	A_93	A_E	A_6	A_7	A_93	A_7	A_E	A_C	A_C	A_6	A_F	A_7	A_93	A_E
A_6	A_7	A_8	A_3F	A_9	A_3F	A_2	A_1	A_D	A_G	A_9	A_D	A_E	A_D	A_6	A_2
A_93	A_F	A_3F	A_1	A_A	A_E	A_F	A_93	A_3	A_3	A_93	A_93	A_6B	A_6B	A_E	A_D
A_2	A_6B	A_5	A_D	A_8	A_3	A_3	A_9	A_8	A_3F	A_3	A_E	A_C	A_1	A_2	A_B
A_D	A_9	A_37	A_B	A_G	A_8	A_D	A_2	A_B	A_5	A_3F	A_F	A_8	A_93	A_5	A_3
A_3	A_6	A_B	A_C	A_3	A_A	A_5	A_37	A_7	A_7	A_G	A_7	A_3F	A_9	A_6B	A_32

- LCZ C is surrounded by large proportion of LCZ D (46.9 %) but it ranked 5th and 6th in summer as well as winter LST variation analysis.
- LCZ D is affected by LCZ 6 (13.9 %) in summer and LCZ 9 (15.3 %) in winter. It proves that, the anthropogenic activities in LCZ 6 are influencing the LST in summer.
- The surrounding LCZs are affecting the LST of LCZ E in the order of their surrounding area proportion in summer.
- The summertime LST of LCZ F is influenced by LCZ 9 whose contribution is higher in F's surrounding pattern (41.1 %) whereas winter-time LST is influenced by LCZ 8 with very less area contribution (0.3 %).
- In winter, LST of LCZ G is affected by surrounding LCZ 6_B (23.5 %) and in summer it is affected by LCZ A (22.3 %).

Top 10 features (surrounding LCZ pattern length of shared boundary vector) responsible for LST variation in each LCZ are shown in [Table 11 \(A\)](#) and [\(b\)](#)). The percentage contribution of surrounding LCZs in terms of their length of shared boundary for each LCZ in study area is shown in [Fig. 5\(b\)](#). The cumulative observations from [Fig. 5\(b\)](#) and [Table 11 \(A\)](#) and [\(b\)](#)) are as follows.

- For compact LCZs (LCZ 1, 2 and 3), in summer as well as in winter, the length of shared boundary with LCZ D, E, 6_B and 3_F is responsible for LST variation even with small shared boundary as compared to LCZ 3, 6 which are sharing large boundaries with LCZ 1.
- LCZ 4 is sharing largest boundaries with LCZ 3 (24.54 %), 6 (21.41 %) and D (12.53 %) but LCZ 5 shared boundary (2.09 %) is

Table 11a

Top 10 surrounding LCZs (shared boundary-wise) for summer-time LST variation in individual LCZ.

1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	G
L_E	L_D	L_6B	L_5	L_C	L_3F	L_37	L_9	L_6	L_G	L_6B	L_A	L_6B	L_D	L_2	L_3F
L_5	L_93	L_3F	L_D	L_6	L_93	L_9	L_6	L_F	L_B	L_D	L_B	L_9	L_93	L_9	L_6
L_93	L_7	L_5	L_8	L_G	L_6B	L_D	L_3	L_D	L_C	L_6	L_6B	L_6	L_9	L_93	L_6B
L_8	L_1	L_93	L_6	L_93	L_C	L_8	L_F	L_93	L_6B	L_C	L_6	L_3F	L_F	L_3	L_3
L_7	L_E	L_E	L_3	L_6B	L_9	L_3	L_D	L_C	L_5	L_9	L_93	L_C	L_6	L_8	L_1
L_D	L_G	L_9	L_1	L_D	L_3	L_C	L_3F	L_3	L_D	L_3	L_9	L_3	L_6B	L_E	L_2
L_6B	L_3F	L_6	L_C	L_9	L_E	L_6	L_2	L_B	L_1	L_5	L_D	L_B	L_1	L_6	L_E
L_3F	L_5	L_37	L_2	L_3	L_7	L_93	L_1	L_8	L_3	L_G	L_5	L_F	L_3	L_D	L_5
L_G	L_F	L_8	L_4	L_A	L_8	L_E	L_93	L_3F	L_2	L_3F	L_3	L_8	L_8	L_C	L_37
L_3	L_8	L_B	L_7	L_E	L_G	L_F	L_4	L_7	L_4	L_A	L_1	L_A	L_2	L_6B	L_4

Table 11b

Top 10 surrounding LCZs (shared boundary-wise) for winter-time LST variation in individual LCZ.

1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	G
L_E	L_E	L_3F	L_93	L_C	L_6B	L_8	L_F	L_F	L_6	L_D	L_A	L_9	L_F	L_8	L_E
L_5	L_93	L_6B	L_3	L_6	L_B	L_37	L_6	L_6	L_D	L_6B	L_B	L_F	L_6B	L_E	L_B
L_8	L_8	L_7	L_E	L_93	L_3F	L_2	L_E	L_C	L_G	L_6	L_6B	L_6B	L_93	L_9	L_3
L_D	L_D	L_5	L_D	L_6B	L_7	L_E	L_2	L_D	L_6B	L_A	L_9	L_B	L_8	L_C	L_6
L_7	L_7	L_E	L_3F	L_D	L_93	L_3	L_D	L_93	L_C	L_9	L_93	L_E	L_D	L_6	L_1
L_6B	L_F	L_93	L_9	L_G	L_E	L_F	L_3	L_E	L_1	L_3	L_6	L_8	L_7	L_3	L_6B
L_93	L_3	L_37	L_8	L_E	L_8	L_5	L_93	L_8	L_3F	L_93	L_D	L_6	L_3	L_6B	L_3F
L_G	L_G	L_8	L_5	L_9	L_9	L_6	L_37	L_2	L_5	L_3F	L_8	L_93	L_1	L_1	L_2
L_9	L_3F	L_6	L_C	L_3F	L_A	L_9	L_1	L_B	L_3	L_C	L_E	L_3F	L_9	L_2	L_37
L_C	L_9	L_9	L_1	L_8	L_3	L_1	L_7	L_6B	L_93	L_E	L_F	L_C	L_37	L_4	L_32

influencing the LST in summer and LCZ 9₃ shared boundary (8.36 %) is influencing LST in winter.

- LCZ 5 is sharing the largest boundary with 6B (22.59 %) but it is not responsible for summertime and wintertime LST variation.
- LCZ 6's LST is influenced when it shares the boundary with low-rise LCZs.
- LCZ 7 is sharing largest boundary with LCZ 3 (29.98 %) but shared boundary length with 3₇ (5.69 %) and 8 (9.08 %) are responsible for LST variation in LCZ 7.
- In winter, LCZ F influences the LST of both LCZs 8 and 9.
- LCZ A and B are sharing the largest border with LCZ 6_B (41.69 %, 24.95 %) but the LST of LCZ B is only influenced by LCZ 6_B.
- LCZ C's LST is affected by LCZ A's shared border length even with very less proportion of 2.15 %.

Further, the local climate zones observed during 2019 for the study area are grouped into five categories namely compact LCZs (LCZ 1–3), open LCZs (LCZ 4–6), low-rise LCZs (LCZ 7–8), other (LCZ 9) and natural LCZs (LCZ A–G). The ALPE module with the configuration mentioned in methodology section is applied on each group to find their LST influencing factors for summer and winter by using the combined features i.e. area of adjacent LCZ pattern and length of shared boundary of adjacent LCZ pattern. This type of grouped results is generally helpful for policy makers rather than individual LCZ analysis in order to make any urban planning decision. The graphical representation showing ranked (as per mutual information values) LST influencing factors for five groups is shown in Fig. 6. From the Fig. 6 (a), it is observed that, LST of compact type LCZs is influenced by area of surrounding LCZs rather than length of shared boundary with surrounding LCZs in summer and winter. In summer, area of LCZs which are open in nature like LCZ 9₃, 6_B, 5, 6 are contributing towards LST variation whereas in winter area of low-rise LCZs (LCZ 7,8) are the topmost LST influencing factors. Fig. 6(b) shows that, the summer and winter LST variation of open LCZ types is observed by similar type of surrounding pattern i.e. LCZ 6_B and 9₃. Their area and shared boundary length both are contributing in equal manner. The LST of low-rise LCZs is influenced by area of natural LCZs (Fig. 6(c)) in summer (LCZ D) and winter (LCZ E). In other LCZ type, area and length of shared boundary of open low-rise (LCZ 6) are the largest

parameters responsible for LST variation including area of compact low-rise (LCZ 3). The summertime LST variation in natural LCZs (Fig. 6 (d)) is due to surrounding area and shared boundary length with LCZ 6_B (Open low-rise + scattered trees) whereas their wintertime LST variation is due to area of surrounding LCZ 9 (sparsely built).

5. Discussion

In urban thermal environment research, LCZ acts as a common objective mapping scheme (Abdi et al., 2020; Bechtel et al., 2019). The factors influencing urban climate can be identified by assessing the relationship between urban structures and LST at LCZ scale (Yang et al., 2021). In terms of the research on temperature variation, an inconsistent LST variation is observed within similar but spatially dispersed LCZ for the study area (see Fig. 1). The proposed methodology verifies such inconsistent LST variation is due to the surrounding LCZs via LST prediction phase. Also it identifies the surrounding LCZs responsible of LST variation in each LCZ excluding the need of grid scale (i.e. it works on polygon geometry thus no need to define the scale). We identified the influencing factors for inconsistent LST variation using various machine learning and deep learning techniques by incorporating the adjacent LCZ pattern information in terms of two parameters i.e. area and length of shared boundary, in addition to the LCZ (LCZ name/type) and LCZ's self area. It has been observed that the addition of these two parameters results in decreased bias (0.96, 0.59) in LST prediction. This indicates that the surrounding LCZ pattern is also responsible for LST of any LCZ along with its own characteristics and area. It is observed for the study area that when any LCZ type is surrounded by the LCZ types which are open and compact with low-rise in nature, its LST is highly variable. This is because; in open and low-rise LCZs the wind velocity varies thereby causing variations in LST.

The substantial LST variation was observed within similar but spatially dispersed LCZs for the study area (see Fig. 1). The proposed ALPE (Adjacent LCZ estimator) was used to identify the surrounding LCZs responsible for LST variation in each LCZ for the seasons, summer and winter respectively. From the result analysis of ALPE, some points can be summarised. First - for any influencing LCZ, its shared boundary length with surrounding LCZ types and their area are the influential

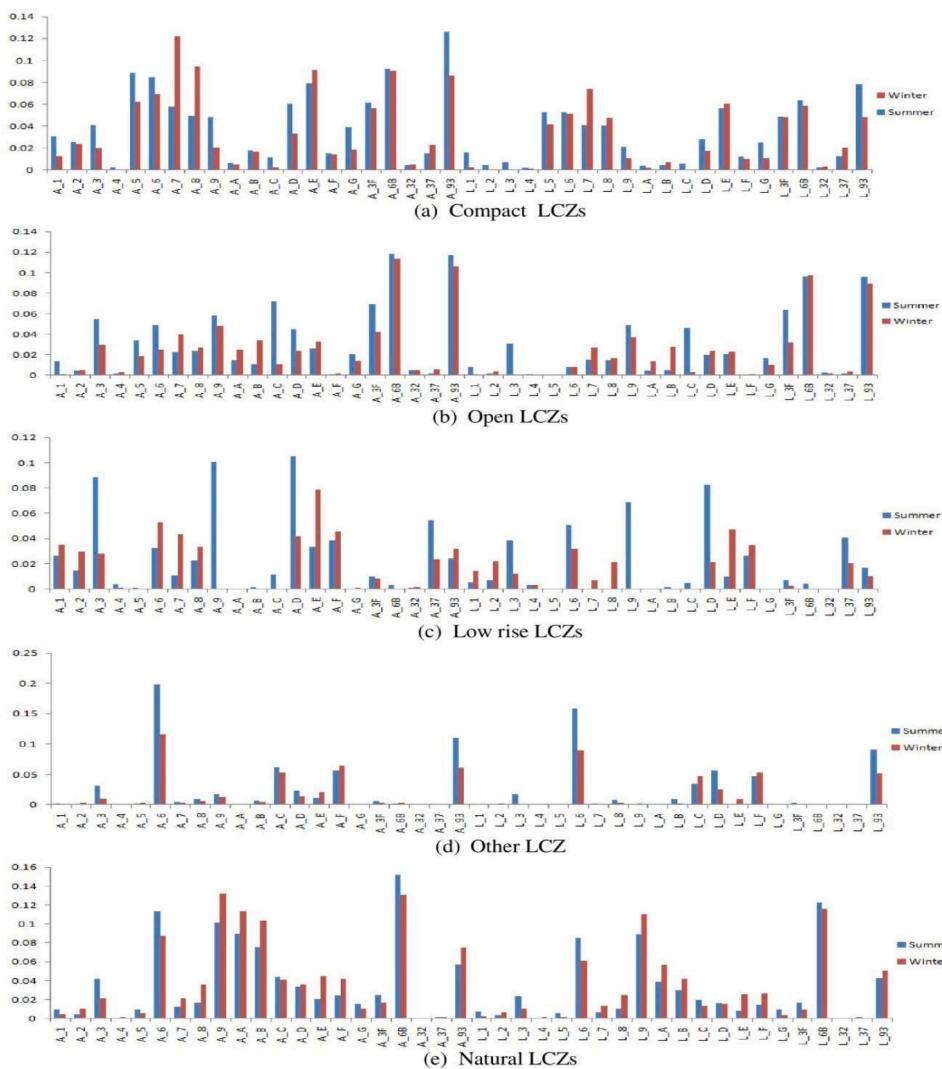


Fig. 6. LST influencing factors for five groups of LCZs.

factors for LST. Second- the length of shared boundary in built LCZ types (LCZs which are generally open in nature) is important over their area. Table 12 shows the category-wise ALPE result analysis for top 5 LST influencing parameters. From the Table 12, it can be concluded that areas of open types of LCZ highly influence the LST of compact type of LCZs. Hence, to reduce the adverse effect of LST on compact and open LCZ types, the planning strategies need to consider open and low-rise LCZs. In addition to the above specified observations, some general conclusions can be made from the proposed methodology and result analysis.

Table 12
Top 5 features affecting LST for different LCZ categories.

LCZ category	Summer		Winter	
	Area (Ratio)	Shared boundary length (Ratio)	Area (Ratio)	Shared boundary length (Ratio)
Compact (1-3)	9 ₃ , 6 _B , 5, 6, E	—	7, 8, 6 _B , 9 ₃ , E	—
Open (4-6)	6 _B , 9 ₃ , C	6 _B , 9 ₃	6 _B , 9 ₃ , 9	6 _B , 9 ₃
Low-rise (7,8)	D, 9, 3	D, 9	E, 6, F, 7	E
Other (9)	6, 9 ₃ , C	6, 9 ₃	6, F, 9 ₃	6, F
Natural (A-G)	6 _B , 6, 9, A	6 _B	9, 6 _B , A	6 _B , 9

1. LST variation of similar but spatially dispersed LCZs within study area is due to the effect of adjacent LCZ pattern.
2. LCZ name/type only is able to predict the LST with satisfactory accuracy (94.58 %, 89.28 %) but the bias indicating difference between actual and predicted LST is reduced (0.96, 0.59) by adding adjacent LCZ pattern information. Adding this information to statistical model enhances the LST prediction accuracy (97.2 %, 96.94 %) for the study area.
3. Surrounding LCZs responsible for LST variation of corresponding LCZ in summer and winter are different.
4. LST of individual LCZ is affected by surrounding LCZs which are open in nature. Open LCZs are the LCZ with open area like LCZ 5, 6, and F.

Each LCZ unit is an aggregation of spatial characteristics. Along with standard LCZs (Built LCZ 1–10, Natural LCZ A–G), there also exist mixed/heterogeneous classes (6_B, 9₃, 3₂, 3₇ and 3_F) in the study area. These mixed LCZ classes denote an area consist of two LCZs with different morphological characteristics. Such areas are formed due to variation in socio-functional attributes. The proposed study, analyzes the LST variation for standard LCZ types by considering their surrounding spatial heterogeneity represented by LCZ types (spatial aggregation of standard and mixed/heterogeneous LCZs) and quantified by the attributes {area, ratio of length of shared boundary}. When

parameters like the area (ratio) of nearby LCZs is considered, the majority of open-type LCZs (5, 6, 9, 6_B, 8, 9₃, D, E) are found to have an impact on the variance of LST in the LCZ under consideration in the summer and winter. When it comes to the ratio of shared boundary between LCZs, both compact and open type LCZs (2, 3F, 37, 6, 6_B, 9, A, D, E) are responsible for summer-time LST variation whereas, open type LCZs (6_B, 8, 9, 9₃, D, E, F) are found to influence winter-time LST. Further, we have analyzed the LST influencing neighborhood LCZs for mixed class 9₃ and observed that LCZ 6, 3 and 9 are mostly responsible for its LST variance. Also, we found the mixed LCZs in the LST influencing set, but their open type nature highly determines the variation for the study area. Though, we have considered all the surrounding LCZ types, the mixed LCZs may pursue an additional relationship with LST. The proportion of spatial heterogeneity in mixed LCZs surrounding the considered LCZs can be correlated with LST of the considered LCZ. Identifying such relationship can be the future perspective of LCZ-LST research. The proposed methodology is useful for analysing the effect of spatial aggregation on various domains such as urban development, urban planning, mapping disease risk etc.

To the best of our knowledge, this work is the first to propose an LST analysis model based on the concept of similar but spatially dispersed LCZs in heterogeneous urban environment. The results of this study will be helpful for urban planners and policymakers to better design the plans for any LCZ by considering its adjacent LCZ pattern. Also, the proposed methodology needs to be extended by combining different cities LST statistics to investigate the generalized observation. As said earlier, this is the first work of its kind hence the proposed work and its findings can be used as a benchmark and can be verified, validated by further work done in this area of research.

6. Conclusion

In this paper, the effect of adjacent LCZ patterns on LST of individual LCZ is verified in addition to LST prediction with limited input parameters for the Nagpur study area. We used adjacent LCZ preference estimator (ALPE) along with different machine learning and deep learning techniques to accomplish this task. The objective to propose the methodology mentioned in this paper is to validate the temperature variation of similar but spatially dispersed LCZ pattern is due to their adjacent surrounding LCZs. To test the effect of surrounding LCZs, we used limited feature set that includes LCZ, LCZ area, adjacent LCZ patterns area (vector), adjacent LCZ patterns shared boundary length (vector).

We first conducted the study to check and validate the temperature variation of similar but spatially dispersed LCZs in heterogeneously structured city like Nagpur with minimum features and to a significant extent, we proved that the neighbourhood LCZ pattern is responsible for the day-time variation in study area. As we know this research is limited to heterogeneous city Nagpur and in that it considers only the day-time temperature variations, but night-time temperature variation and different cities climatic condition may lead to different result. Therefore, comparable studies need to be conducted by considering places with varying climatic conditions to get standard conclusions on LCZ –adjacent LCZ pattern relationship. In future, the methodology shall be extended to combine the dataset for many heterogeneously built cities to test the mentioned hypothesis.

CRediT authorship contribution statement

Mrunali Vaidya: Writing – original draft, Methodology. **Ravindra Keskar:** Writing – review & editing, Supervision. **Rajashree Kotharkar:** Validation.

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 will be made available on request.

References

- Abdi, B., Hami, A., & Zarehaghi, D. (2020). Impact of small-scale tree planting patterns on outdoor cooling and thermal comfort. *Sustainable Cities and Society*, 56, Article 102085. <https://doi.org/10.1016/j.scs.2020.102085>
- 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. *Building and Environment*, 185, Article 107268. <https://doi.org/10.1016/j.buildenv.2020.107268>
- Bechtel, B., Demuzere, M., Mills, G., Zhan, W., Sismanidis, P., Small, C., & Voogt, J. (2019). SUHI analysis using Local Climate Zones—A comparison of 50 cities. *Urban Climate*, 28, Article 100451. <https://doi.org/10.1016/j.uclim.2019.01.005>
- Cai, M., Ren, C., & Xu, Y. (2017). Investigating the relationship between Local Climate Zone and land surface temperature. In 2017 Joint Urban Remote Sensing Event (JURSE) 2017 Mar 6 (pp. 1–4). IEEE. <https://doi.org/10.1109/JURSE.2017.7924622>
- Cai, M., Ren, C., Xu, Y., Lau, K. K., & Wang, R. (2018). Investigating the relationship between local climate zone and land surface temperature using an improved WUDAPT methodology—A case study of Yangtze River Delta, China. *Urban Climate*, 24, 485–502. <https://doi.org/10.1016/j.uclim.2017.05.010>
- Chen, Y., Duan, S. B., Ren, H., Labed, J., & Li, Z. L. (2017). Algorithm development for land surface temperature retrieval: Application to Chinese Gaofen-5 data. *Remote Sensing*, 9(2), 161. <https://doi.org/10.3390/rs9020161>
- Eimermacher, J. (2018). Micro-scale variability of air temperature within a local climate zone in Berlin, Germany, during Summer. *Climate*, 6, 5. <https://doi.org/10.3390/cli6010005>
- Elbeltagi, A., Pande, C. B., Kumar, M., Tolche, A. D., Singh, S. K., Kumar, A., & Vishwakarma, D. K. (2023). Prediction of meteorological drought and standardized precipitation index based on the random forest (RF), random tree (RT), and Gaussian process regression (GPR) models. *Environmental Science and Pollution Research*, 30 (15), 43183–43202. <https://doi.org/10.1007/s11356-023-25221-3>
- Fenner, D., Meier, F., Bechtel, B., Otto, M., & Scherer, D. (2017). Intra and inter local climate zone variability of air temperature as observed by crowdsourced citizen weather stations in Berlin, Germany. In *Meteorologische Zeitschrift*, 26 pp. 525–547. <https://doi.org/10.1127/metz/2017/0861>
- Ferreira, L. S., & Duarte, D. H. (2019). Exploring the relationship between urban form, land surface temperature and vegetation indices in a subtropical megacity. *Urban Climate*, 27, 105–123. <https://doi.org/10.1016/j.uclim.2018.11.002>
- Geletić, J., Lehnert, M., & Dobrovolný, P. (2016). Land surface temperature differences within local climate zones, based on two central European cities. *Remote Sensing*, 8 (10), 788. <https://doi.org/10.3390/rs8100788>
- Geletić, J., Lehnert, M., Savić, S., & Milošević, D. (2018). Modelled spatiotemporal variability of outdoor thermal comfort in local climate zones of the city of Brno, Czech Republic. *Science of the Total Environment*, 624, 385–395. <https://doi.org/10.1016/j.scitotenv.2017.12.076>
- Geletić, J., Lehnert, M., Savić, S., & Milošević, D. (2019). Inter-/intra-zonal seasonal variability of the surface urban heat island based on local climate zones in three central European cities. *Building and Environment*, 156, 21–32. <https://doi.org/10.1016/j.buildenv.2019.04.011>
- Giri, S., Kang, Y., MacDonald, K., Tippett, M., Qiu, Z., Lathrop, R. G., & Obropta, C. C. (2023). Revealing the sources of arsenic in private well water using Random Forest Classification and Regression. *Science of The Total Environment*, 857, Article 159360. <https://doi.org/10.1016/j.scitotenv.2022.159360>
- Giridharan, R., Lau, S. S., & Ganesan, S. (2005). Nocturnal heat island effect in urban residential developments of Hong Kong. *Energy and Buildings*, 37(9), 964–971. <https://doi.org/10.1016/j.enbuild.2004.12.005>
- Guo, G., Zhou, X., Wu, Z., Xiao, R., & Chen, Y. (2016). Characterizing the impact of urban morphology heterogeneity on land surface temperature in Guangzhou, China. *Environmental Modelling & Software*, 84, 427–439. <https://doi.org/10.1016/j.envsoft.2016.06.021>
- Hulley, G. C., Ghent, D., Götsche, F. M., Guillevic, P. C., Mildrexler, D. J., & Coll, C. (2019). Land surface temperature. *Taking the temperature of the earth 2019 Jan 1* (pp. 57–127). Elsevier. <https://doi.org/10.1016/B978-0-12-814458-9.00003-4>
- Jenerette, G. D., Harlan, S. L., Buyantuev, A., Stefanov, W. L., Declet-Barreto, J., Ruddell, B. L., Myint, SW, Kaplan, S., & Li, X. (2016). Micro-scale urban surface temperatures are related to land-cover features and residential heat related health impacts in Phoenix, AZ USA. *Landscape Ecology*, 31, 745–760. <https://doi.org/10.1007/s10980-015-0284-3>
- Kotharkar, R., & Bagade, A. (2018a). Local Climate Zone classification for Indian cities: A case study of Nagpur. *Urban Climate*, 24, 369–392. <https://doi.org/10.1016/j.uclim.2017.03.003>
- Kotharkar, R., & Bagade, A. (2018b). Evaluating urban heat island in the critical local climate zones of an Indian city. *Landscape and Urban Planning*, 169, 92–104. <https://doi.org/10.1016/j.landurbplan.2017.08.009>

- Krayenhoff, E. S., & Voogt, J. A. (2016). Daytime thermal anisotropy of urban neighbourhoods: Morphological causation. *Remote Sensing*, 8(2), 108. <https://doi.org/10.3390/rs8020108>
- Middel, A., Häb, K., Brazel, A. J., Martin, C. A., & Guhathakurta, S. (2014). Impact of urban form and design on mid-afternoon microclimate in Phoenix Local Climate Zones. *Landscape and Urban Planning*, 122, 16–28. <https://doi.org/10.1016/j.landurbplan.2013.11.004>
- Mills, G., Futcher, J., & Stewart, I. D. (2021). The urban heat island: Its energetic basis and management. In M. Palme, & A. Salvati (Eds.), *Urban microclimate modelling for comfort and energy studies*. Cham: Springer. https://doi.org/10.1007/978-3-030-65421-4_3
- Ochola, E. M., Fakharizadehshirazi, E., Adimo, A. O., Mukundi, J. B., Wesonga, J. M., & Sodoudi, S. (2020). Inter-local climate zone differentiation of land surface temperatures for Management of Urban Heat in Nairobi City, Kenya. *Urban Climate*, 31, Article 100540. <https://doi.org/10.1016/j.uclim.2019.100540>
- Portela, C. I., Massi, K. G., Rodrigues, T., & Alcântara, E (2020). Impact of urban and industrial features on land surface temperature: Evidences from satellite thermal indices. *Sustainable Cities and Society*, 56, Article 102100. <https://doi.org/10.1016/j.scs.2020.102100>
- Shi, Y., Lau, K. K., Ren, C., & Ng, E. (2018). Evaluating the local climate zone classification in high-density heterogeneous urban environment using mobile measurement. *Urban Climate*, 25, 167–186. <https://doi.org/10.1016/j.uclim.2018.07.001>
- Skarbit, N., Stewart, I. D., Unger, J., & Gál, T. (2017). Employing an urban meteorological network to monitor air temperature conditions in the 'local climate zones' of Szeged, Hungary. *International Journal of Climatology*. <https://doi.org/10.1002/joc.502337:582-96>
- Stewart, I. D., & Oke, T. R. (2012). Local climate zones for urban temperature studies. *Bulletin of the American Meteorological Society*, 93(12), 1879–1900. <https://doi.org/10.1175/BAMS-D-11-00019.1>
- Tajuddeen, I., Sajadian, S. M., & Jafari, M. (2023). Regression models for predicting the global warming potential of thermal insulation materials. *Buildings*, 13(1), 171. <https://doi.org/10.3390/buildings13010171>
- Thomas, G., Sherin, A. P., Ansar, S., & Zachariah, E. J (2014). Analysis of urban heat island in Kochi, India, using a modified local climate zone classification. *Procedia Environmental Sciences*, 21, 3–13. <https://doi.org/10.1016/j.proenv.2014.09.002>
- Unger, J., Lelovics, E., & Gál, T. (2014). Local Climate Zone mapping using GIS methods in Szeged. *Hungarian Geographical Bulletin*, 63(1), 29–41. <https://doi.org/10.15201/hungeobull.63.1.3>
- Vaidya, M., Keskar, R., & Kotharkar, R. (2024). Classifying heterogeneous urban form into local climate zones using supervised learning and greedy clustering incorporating Landsat dataset. *Urban Climate*, 53, Article 101770. <https://doi.org/10.1016/j.uclim.2023.101770>
- Wang, Q., Wang, X., Zhou, Y., Liu, D., & Wang, H. (2022). The dominant factors and influence of urban characteristics on land surface temperature using random forest algorithm. *Sustainable Cities and Society*, 79, Article 103722. <https://doi.org/10.1016/j.scs.2022.103722>
- Xiang, Y., Tang, Y., Wang, Z., Peng, C., Huang, C., Dian, Y., Teng, M., & Zhou, Z. (2023). Seasonal Variations of the Relationship between Spectral Indexes and Land Surface Temperature Based on Local Climate Zones: A Study in Three Yangtze River Megacities. *Remote Sensing*, 15(4), 870. <https://doi.org/10.3390/rs15040870>
- Xie, P., Yang, J., Wang, H., Liu, Y., & Liu, Y. (2020). A New method of simulating urban ventilation corridors using circuit theory. *Sustainable Cities and Society*, 59, Article 102162. <https://doi.org/10.1016/j.scs.2020.102162>
- Yan, H., Fan, S., Guo, C., Wu, F., Zhang, N., & Dong, L. (2014). Assessing the effects of landscape design parameters on intra-urban air temperature variability: The case of Beijing, China. *Building and Environment*, 76, 44–53. <https://doi.org/10.1016/j.buildenv.2014.03.007>
- Yang, X., Yao, L., Jin, T., Peng, L. L., Jiang, Z., Hu, Z., & Ye, Y (2018). Assessing the thermal behavior of different local climate zones in the Nanjing metropolis, China. *Building and Environment*, 137, 171–184. <https://doi.org/10.1016/j.buildenv.2018.04.009>
- Yang, J., Zhan, Y., Xiao, X., Xia, J. C., Sun, W., & Li, X. (2020a). Investigating the diversity of land surface temperature characteristics in different scale cities based on local climate zones. *Urban Climate*, 34, Article 100700. <https://doi.org/10.1016/j.uclim.2020.100700>
- Yang, J., Ren, J., Sun, D., Xiao, X., Xia, J. C., Jin, C., & Li, X (2021). Understanding land surface temperature impact factors based on local climate zones. *Sustainable Cities and Society*, 69, Article 102818. <https://doi.org/10.1016/j.scs.2021.102818>
- Yang, J., Wang, Y., Xiu, C., Xiao, X., Xia, J., & Jin, C (2020b). Optimizing local climate zones to mitigate urban heat island effect in human settlements. *Journal of Cleaner Production*, 275, Article 123767. <https://doi.org/10.1016/j.jclepro.2020.123767>
- Yang, J., Jin, S., Xiao, X., Jin, C., Xia, J. C., Li, X., & Wang, S (2019). Local climate zone ventilation and urban land surface temperatures: Towards a performance-based and wind-sensitive planning proposal in megacities. In *Sustainable Cities and Society*, 47, Article 101487. <https://doi.org/10.1016/j.scs.2019.101487>
- Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., & Gao, J (2020). Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sensing of Environment*, 241, Article 111716. <https://doi.org/10.1016/j.rse.2020.111716>
- Zhao, N., Ma, A., Zhong, Y., Zhao, J., & Cao, L. (2019). Self-training classification framework with spatial-contextual information for local climate zones. *Remote Sensing*, 11(23), 2828. <https://doi.org/10.3390/rs11232828>
- Zhao, C. (2018). Linking the local climate zones and land surface temperature to investigate the surface urban heat island, a case study of San Antonio, Texas, US. In *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4 pp. 277–283. <https://doi.org/10.5194/isprs-annals-IV-3-277-2018>
- 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 Sensing*, 13 (21), 4338. <https://doi.org/10.3390/rs13214338>
- Zhao, C., Jensen, J., Weng, Q., & Weaver, R. (2018). A geographically weighted regression analysis of the underlying factors related to the surface urban heat island phenomenon. *Remote Sensing*, 10(9), 1428. <https://doi.org/10.3390/rs10091428>
- Zhou, L., Yuan, B., Hu, F., Wei, C., Dang, X., & Sun, D. (2022). Understanding the effects of 2D/3D urban morphology on land surface temperature based on local climate zones. *Building and Environment*, 208, Article 108578. <https://doi.org/10.1016/j.buildenv.2021.108578>