

Comparison between convolutional neural networks and random forest for local climate zone classification in mega urban areas using Landsat images

Cheolhee Yoo^a, Daehyeon Han^a, Jungho Im^{a,*}, Benjamin Bechtel^b

^a School of Urban and Environmental Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, South Korea

^b Department of Geography, Ruhr-University Bochum, Bochum 44801, Germany



ARTICLE INFO

Keywords:

Local climate zone
Convolutional neural networks
Random forest
Urban climate
Landsat

ABSTRACT

The Local Climate Zone (LCZ) scheme is a classification system providing a standardization framework to present the characteristics of urban forms and functions, especially for urban heat island (UHI) research. Landsat-based 100 m resolution LCZ maps have been classified by the World Urban Database and Portal Tool (WUDAPT) method using a random forest (RF) machine learning classifier. Some studies have proposed modified RF and convolutional neural network (CNN) approaches. This study aims to compare CNN with an RF classifier for LCZ mapping in great detail. We designed five schemes (three RF-based schemes (S1–S3) and two CNN-based ones (S4–S5)), which consist of various combinations of input features from bitemporal Landsat 8 data over four global mega cities: Rome, Hong Kong, Madrid, and Chicago. Among the five schemes, the CNN-based one with the incorporation of a larger neighborhood information showed the best classification performance. When compared to the WUDAPT workflow, the overall accuracies for entire land cover classes (OA) and for urban LCZ types (i.e., LCZ1–10; OA_{urb}) increased by about 6–8% and 10–13%, respectively, for the four cities. The transferability of LCZ models for the four cities were evaluated, showing that CNN consistently resulted in higher accuracy (increased by about 7–18% and 18–29% for OA and OA_{urb}, respectively) than RF. This study revealed that the CNN classifier classified particularly well for the specific LCZ classes in which buildings were mixed with trees or buildings or plants were sparsely distributed. The research findings can provide a basis for guidance of future LCZ classification using deep learning.

1. Introduction

Although the ratio of urban areas to global land surface is just 3%, about 54% of the world's population live in urban centers; by 2050, that number will increase to nearly 65% (Cohen, 2015). Urbanization results in the increased absorption of solar radiation due to the expanded impervious area, the reduced sky view factor due to the greater number of (high-rise) buildings, and the release of artificial heat in the urban canyon especially in mega cities (Barnes et al., 2001; Giridharan et al., 2004; Han-qiu and Ben-qing, 2004; Rizwan et al., 2008). The urban heat island phenomenon (UHI), that is urban areas are warmer than the surrounding areas, is important these days as it interacts with other urban climate problems, such as heat waves and air pollution (Founda and Santamouris, 2017; Salata et al., 2017; Yadav et al., 2017; Fallmann et al., 2016). Different types of UHIs need to be differentiated, most importantly the surface temperature UHI (SUHI) and the air temperature UHI in the canopy layer, which is from the ground to the height of buildings.

Traditionally UHI studies analyze the temperature difference between urban and rural areas. These can be differentiated by satellite-based land cover data based on specific class types (i.e., typical land cover classification), is one of the possible solutions. Typical global land cover data used in existing UHI studies include the 500 m resolution MODIS land cover product (MCD12Q1) (Friedl et al., 2010), the 300 m resolution GlobCover 2009 dataset produced by ESA (Bontemps et al., 2011), and the Global Land Cover product (GLC or GlobalLand30) produced by Chen et al. (2015) with Landsat data for 30 m resolution (Mathew et al., 2018; Lauwaet et al., 2015; Liu et al., 2018b). However, these products have only one urban land cover class: "urban and built-up class" in MODIS, "artificial surfaces and associated areas" in GlobCover 2009, and "artificial surfaces" in GlobalLand30. Stewart and Oke (2012), however, explained that the thermal properties of urban areas vary with the height and density of the buildings in them. Thus, there is a limit to analyzing the detailed UHI characteristics of a city using global land cover products that have a single urban class.

In fact, many countries have produced their own detailed land cover

* Corresponding author.

E-mail address: ersgis@unist.ac.kr (J. Im).

data with at least several urban type classes. The national land cover product of the United States (NLCD2011), for example, has a total of 20 land cover classes and four of them are urban types based on the degree of development (i.e., high intensity, medium intensity, low intensity, and open space). The European CORINE (Co-ORdinated INformation on the Environment) land cover has 11 urban-related classes in its level 3 product. The Urban Atlas product also provides high-resolution land use maps of urban areas in European countries. Because urban classes vary by product, the use of the urban classes for studying global heat phenomena is relatively limited. Since the classification criteria of these products, such as NLCD2011, CORINE and Urban Atlas, focus only on the density of the impervious areas with consideration of land use information, the factors strongly linked to UHI—including the sky view factor and building height to street width ratio—were barely considered when the products were generated.

To overcome such an issue, researchers in the UHI field have designed a classification system that well fits this purpose. Local Climate Zone (LCZ) is a classification system designed by [Stewart and Oke \(2012\)](#) especially for UHI research. The LCZ consists of 10 urban LCZ types and 7 natural LCZ types. It has a culturally neutral framework which is generic and easy to understand for global urban climate studies ([Fig. 1](#)). [Bechtel et al. \(2015\)](#) devised a World Urban Database and Portal Tool (WUDAPT) method to construct a 100 m resolution pixel-based LCZ map using Landsat 8 images. Landsat 8 is a polar orbiting satellite sensor system that can capture global areas with a resolution of 30 m (for visible, NIR, and SWIR bands) to 100 m (for thermal bands) every 16 days. The WUDAPT method resamples the Landsat image of each city into 100 m resolution (i.e., using the zonal mean) to get the spectral information of local-scale urban structures. Local experts with deep knowledge of individual cities build LCZ reference polygons using high resolution Google Earth images. These polygons are then converted into 100 m resolution pixels and used for training and testing LCZ classification models with Landsat images. WUDAPT uses random forest (RF), a rule-based machine learning approach, for classification. The LCZ maps of many cities all over the globe (about 90 cities as of August 2018) have been built in this way and shared through the WUDAPT portal (<http://www.wudapt.org>) ([Bechtel et al., 2019](#)).

The LCZ maps produced by the WUDAPT method have been used to find several key parameters that affect UHI ([Giridharan and Emmanuel,](#)

[2018; Kaloustian and Bechtel, 2016](#)). Land surface temperature and air temperature have been analyzed for LCZ classes ([Beck et al., 2018; Wang et al., 2018; Cai et al., 2018](#)). Furthermore, the effect of respiratory particulate matter on land surface temperature has been discussed using various LCZ classes ([Ziaul and Pal, 2018](#)). The WUDAPT-based LCZ maps, however, are still limited in terms of classification accuracy. The average Overall Accuracy (OA) of the 90 LCZs uploaded on the WUDAPT portal is 74.5%, leaving much room for improvement. In particular, the average OA of the urban LCZ types (OA_{urb}) of the 90 LCZs is just 59.3%, which means that the urban LCZ types are not as accurate as the other general natural LCZ types such as forest and water. The low classification accuracy of urban features (i.e., urban LCZ types) is a major limitation for urban climate-related research.

Therefore, the WUDAPT community has encouraged scientists to explore various classification approaches to further improve LCZ classification ([Yokoya et al., 2018](#)). For example, [Danylo et al. \(2016\)](#) added various spectral metrics (i.e., zonal maximum and minimum) to the input variables of the RF classifier. Their OA improved by 2% when compared with the traditional WUDAPT method for LCZ classification in Kiev, Ukraine. [Verdonck et al. \(2017\)](#) extracted the spectral information (i.e., mean, minimum, maximum, median, and 25th and 75th quantile values) of neighboring pixels through a moving window approach. These six new features were used as input variables in the RF machine learning model. The OA of the LCZ classification of Antwerp, Brussels, and Ghent in Belgium were improved by 7.9%, 13.0%, and 5.4%, respectively, when compared to the original WUDAPT method. These studies improved LCZ classification by using additional input variables in a way that got more spectral features on a contextual domain into the RF classifier.

In recent years, deep learning models which exploit many layers of non-linear information have been widely used for image classification, object segmentation, and text determination ([Schmidhuber, 2015; LeCun et al., 2015; Wang et al., 2012](#)). Among various deep learning models, Convolutional Neural Networks (CNN) has been shown to exhibit high performance in image classification tasks ([Krizhevsky et al., 2012; Vedaldi and Lenc, 2015; Kim et al., 2018b](#)). CNN, a feedforward network with feature learning, extracts inherent spatial features at each layer. Theoretically, CNN has the ability of self-study and in-depth learning for feature extraction, weight sharing and dimension reduction

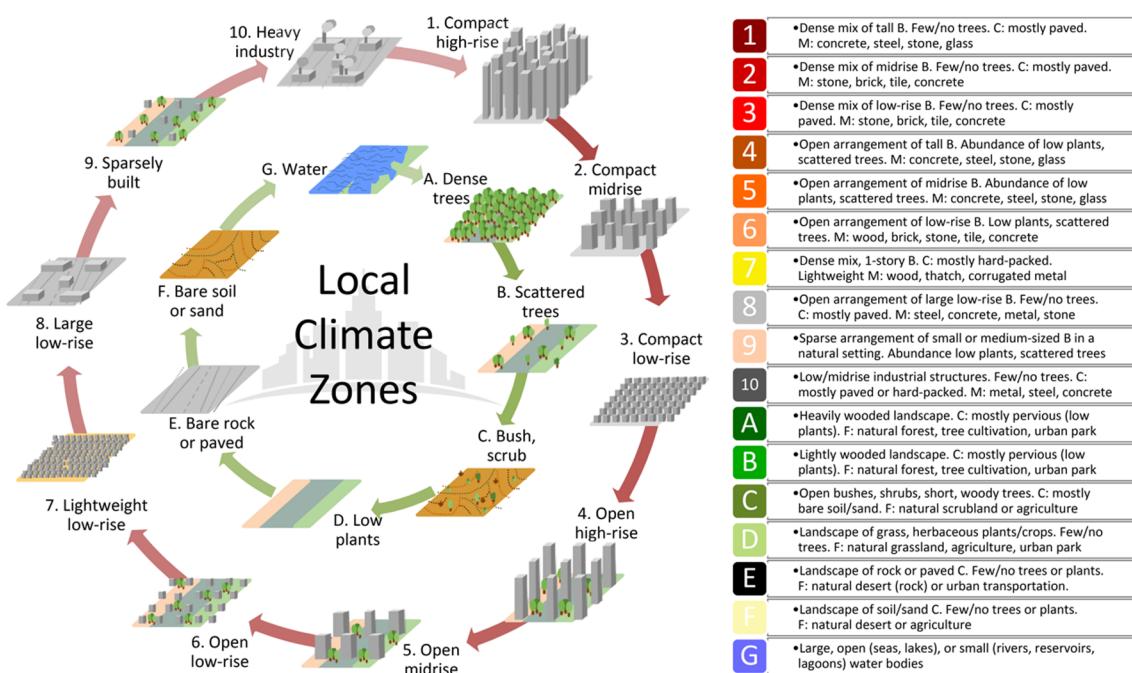


Fig. 1. The local climate zone (LCZ) types identified in urban climate research (from [Bechtel et al., 2017](#) after [Stewart and Oke, 2012](#)), © CC-BY 4.0.

by combining a backpropagation mechanism and a gradient descent optimization method. Back propagation gives an opportunity for backward feedback to enhance the reliability, and the gradient descent method is used in the self-training process.

Numerous studies have used CNN for land cover classification from satellite images (Paoletti et al., 2018; Xu et al., 2018; Marcos et al., 2018), including recent applications for LCZ classification. Sukhanov et al. (2017) designed a multi-level ensemble model combining RF, Gradient Boosting Machines, and a simple CNN with small input data size (i.e., 3×3) to create LCZ maps, which was trained for five cities (i.e., Berlin, Rome, Paris, Sao Paulo and Hong Kong) and then tested over four different cities (i.e., Amsterdam, Chicago, Madrid and Xi'an). Qiu et al. (2018) used a residual convolutional neural network (ResNet) to conduct a systematic analysis of feature importance from multi-source datasets for LCZ classification across 9 cities located in Europe. Since RF is the most successfully used LCZ classifier so far, it is important to know the advantages and disadvantages of using CNN over RF for LCZ classification. However, there has been minimum exploration investigating LCZ classification performance between the CNN and RF classifiers.

This study aims to compare CNN with the RF classifier for LCZ classification. We designed five schemes, which consist of various combinations of input data over four global mega cities: Rome, Hong Kong, Madrid and Chicago. The objectives of this research were to: (1) examine five schemes in order to identify the effect of CNN when compared to other methods that employ RF classifiers, which were proposed in previous studies; (2) investigate a specific set of LCZ classes that produce high classification accuracies; (3) compare the LCZ map generated from two different types of classifiers with reference data; and (4) discuss the research direction of improving local climate zone classification methods for future use.

2. Study area and data

2.1. Study area

Rome, Hong Kong, Madrid, and Chicago were selected as our study areas (Fig. 2). These four cities represent various climatic (Table 1) and geographic characteristics. In addition, their urban structure differs, which enables us to verify the robustness of the proposed approaches.

Rome, the capital city of Italy, is in the midwestern region of the Italian peninsula, and the center of the city is about 24 km inland from the Mediterranean Sea. Rome has about 2.9 million residents living within an area of 1,285 km², making it Italy's largest and most populous city. The city has a monocentric urban structure with increasing densities toward the city center.

Hong Kong is located on the southern coast of China. The city covers about 1,104 km² of land, with 7.4 million residents. Hong Kong is known for its unique urban form and high-density land use. Most areas of the city are hilly, and just under a quarter of the study domain is habitable (i.e., built-up area).

Madrid is the capital city of Spain, a densely populated metropolis located in a relatively flat area lying in the center of the southern Meseta of the Iberian Peninsula. Madrid is the largest city in Spain, with 3.2 million residents living in 604 km². We selected the study region covering the Madrid metropolitan area, comprising monocentric Madrid and its surrounding municipalities called autonomous communities.

Chicago is the third largest city in the US, situated beside the huge Lake Michigan in Illinois. The city of Chicago has about 2.7 million residents in an area of 606 km². Chicago tends to have a regularly shaped street pattern and city blocks based on their grid plan (Ellickson, 2012). We selected the study region that includes the Chicago metropolitan area, comprising the city and its suburbs. The high-density urban center is located in the city of Chicago, while low-density suburban areas surrounding it.

2.2. Satellite input data

Two Landsat 8 images of different seasons for each city were downloaded from the US Geological Survey Earth Explorer site (<https://earthexplorer.usgs.gov>). The acquisition dates with clear sky conditions for the Landsat data are presented in Table 2. We chose two scenes per city close to summer and winter to consider seasonal effects, such as the phenology of vegetation, and to increase classification accuracy, as found by Bechtel et al. (2015). All Landsat images were first clipped covering each city and then atmospheric-corrected into scaled reflectance data using ENVI Fast Line-of-sight Atmospheric Analysis of Hypercubes (FLAASH). Nine of the 11 bands (bands 1–7, 10, and 11) in each Landsat 8 scene were used as input data. Bands 1–7 were the 30 m resolution Operational Land Imager (OLI) spectral bands, and bands 10 and 11 were 30 m resolution thermal bands interpolated from 100 m resolution data collected from Thermal Infrared Sensor (TIRS).

2.3. Reference data

LCZ reference data for the four cities are available from the 2017 IEEE GRSS data fusion contest organized by the Image Analysis and Data Fusion Technical Committee, in collaboration with WUDAPT and GeoWiki (Fig. 2). These data were extracted from the WUDAPT database and further revised to be as accurate as possible (Tuia et al., 2017; Yokoya et al., 2018). Due to unique urban structures and compositions, the number of LCZ classes differs from city to city. Rome has 10 LCZ classes (6 urban LCZ types and 4 natural LCZ types); Hong Kong has 13 LCZ classes (8 urban LCZ types and 5 natural LCZ types); Madrid has 14 LCZ classes (7 urban LCZ types and 7 natural LCZ types); and Chicago has 15 LCZ classes (9 urban LCZ types and 6 natural LCZ types). In addition, the number of polygons digitized for each LCZ class differs between both classes and cities. The polygons of each LCZ class were randomly divided into two parts: the first for training the models and the other for testing them. We tried to equally divide the polygons into the two sets, considering both the number of polygons and the number of 100 m resolution LCZ pixels within each polygon. It is well known that if the training and validation sample pixels share the same polygons, the classification accuracy can be inflated (Zhen et al., 2013). Some LCZ classes in each city, however, form a small number of polygons (fewer than 3), because the classes were not widely distributed within the city. Dividing these small numbers of polygons into two sets would make the models poorly trained. Therefore, we labeled these classes "red-star class". For the red-star classes, two sets were created by dividing the number of pixels of each polygon into two groups through a random sampling approach. The number of polygons and pixels of the two sets for each LCZ class for the four cities are shown in Table 3.

The Global Man-made Impervious Surface (GMIS) data were used to analyze the LCZ maps generated for each city. GMIS provides the 30 m resolution global fractional impervious cover for the year of 2010, which were derived from Landsat data (de Colstoun et al., 2017). To identify the medium-to-high density developed areas, we extracted the GMIS pixels which have an impervious fraction over 70% within the study domain for each city.

3. Methods

3.1. Random forest (RF) classifier

RF has been widely used in the remote sensing field for both classification (Sim et al., 2018; Park et al., 2018; Li et al., 2013) and regression (Lee et al., 2018; Yoo et al., 2018; Richardson et al., 2017). RF is an algorithm based on classification and regression trees (CART), which uses a recursive binary split method to reach final nodes in a tree structure (Breiman, 2001). RF produces numerous independent trees with randomly selected subsets through bootstrapping from training

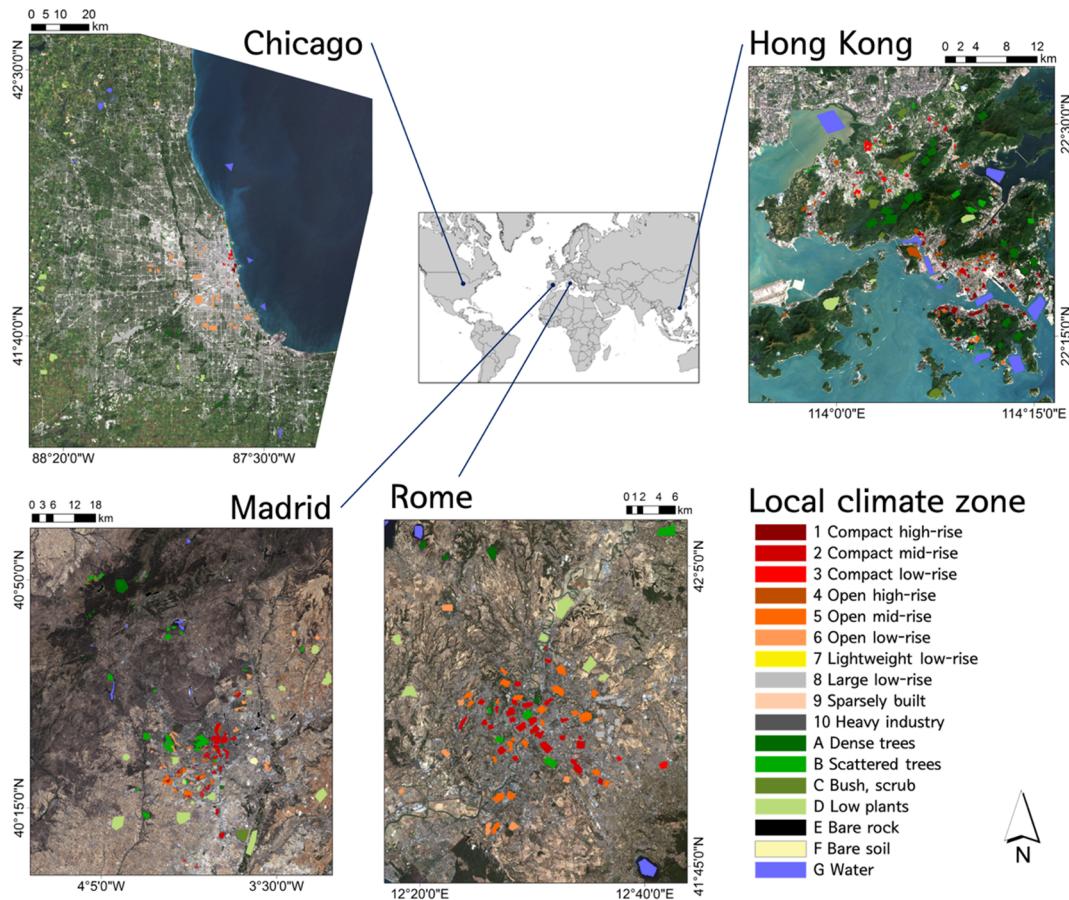


Fig. 2. Study area and Local Climate Zone (LCZ) reference data with legends.

samples and from input variables at every node of a tree. To achieve a final decision, RF adopts an ensemble approach from numerous trees through majority voting for classification.

In this study, the RF was implemented using a random forest package provided in R software (<https://www.r-project.org/>). All parameters except for the number of trees were set as the default values provided by the package (i.e., the number of training samples for each tree was 66.7% of the entire training samples, the number of randomly sampled variables as candidates at each split was the square root of the number of input variables, and the minimum size of the terminal node was 1). The number of trees (i.e., ntree) was selected at the modeling process.

3.2. Convolutional neural networks (CNN) classifier

CNN is a kind of artificial neural network and basically consists of convolutional layers, pooling layers, and fully connected layers. When compared to typical neural networks, the distinguishing feature of CNN is its use of convolutional layers. With the 3-dimensional input data (width, height, and channel), the output of a convolutional layer is transmitted to the next layer keeping the same 3-dimensional shape. The input and output data of the convolutional layer are called feature

Table 2
Selected winter and summer Landsat 8 scenes for each city.

	Scene 1	Scene 2
Rome	January 11, 2017	August 23, 2017
Hong Kong	February 12, 2018	October 23, 2017
Madrid	January 12, 2015	August 13, 2017
Chicago	February 03, 2017	September 12, 2016

maps. The convolution is performed with several filters (or kernels) over the input feature maps. Each moving filter sweeps the input feature maps conducting a dot-product with corresponding elements of the input feature maps, and then the total sum is obtained. The depth of the output feature maps is no longer the number of channels but the number of filters. For example, when 32 filters are used in the first convolutional layer, the output feature map has a depth of 32 regardless of the number of channels in the input feature map.

Convolution reduces the size of the output feature maps. To prevent this, padding is widely used. Padding refers to filling the input feature maps with a specific value before doing the convolution. Padding is mainly used to adjust the spatial size of the output feature maps. The value to be filled can be determined according to the model, but zero-

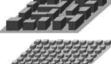
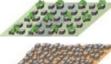
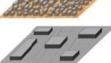
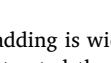
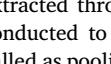
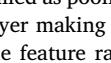
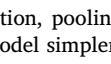
Table 1

The climatic characteristics of the cities. The classes in parentheses correspond to the Köppen-Geiger climate classification (Peel et al., 2007).

City	Description of climate
Rome	Mediterranean climate with dry summers and cool, humid winters (Csa)
Hong Kong	Humid subtropical climate with a hot and humid summer (Cfa)
Madrid	Inland Mediterranean climate, transitioning to a semi-arid climate in the eastern part of the city (Csa)
Chicago	Hot humid continental climate with distinct seasons such as warm to hot and humid summers and cold, snowy winters (Dfa)

Table 3

Training and test datasets of each LCZ type by city. The values in the training and test columns are the number of polygons. The number of the corresponding 100 m resolution pixels is shown in parentheses. * is allocated to the red-star classes, which have only a few reference polygons of the LCZ classes. The LCZ figures in the left column are from Stewart and Oke (2012).

LCZ	Rome		Hong Kong		Madrid		Chicago	
	Training	Test	Training	Test	Training	Test	Training	Test
	1	–	13 (318)	13 (313)	–	–	2* (228)	–
	2	13 (775)	12 (776)	6 (112)	5 (67)	12 (1567)	5 (5647)	2* (126)
	3	2* (104)	–	7 (195)	7 (131)	1* (92)	3 (128)	3 (123)
	4	–	9 (383)	10 (290)	3* (305)	–	2* (140)	–
	5	11 (749)	12 (746)	4 (76)	4 (50)	5 (715)	3 (620)	2* (104)
	6	3 (239)	4 (241)	7 (64)	6 (56)	6 (932)	6 (894)	10 (2059)
	7	–	–	–	–	–	–	–
	8	7 (235)	4 (194)	4 (86)	5 (51)	10 (1433)	12 (1380)	11 (2231)
	9	–	–	–	–	1* (82)	4 (422)	3 (429)
	10	2* (51)	–	5 (109)	4 (110)	–	2 (238)	2 (227)
	A	2 (146)	3 (138)	7 (832)	7 (784)	1 (1115)	3 (244)	6 (515)
	B	2 (293)	3 (262)	7 (207)	6 (200)	8 (1906)	8 (1888)	5 (188)
	C	–	5 (379)	4 (312)	2 (982)	2 (250)	–	–
	D	4 (512)	3 (472)	6 (332)	6 (236)	9 (3621)	6 (3517)	4 (1150)
	E	–	–	–	3 (324)	2 (312)	4 (115)	3 (86)
	F	–	–	–	–	1* (304)	3 (31)	2 (28)
	G	3* (485)	–	5 (1282)	4 (1054)	2 (391)	2 (385)	5 (967)
							4 (984)	

padding is widely used in various applications. Once feature maps are extracted through the convolutional layers, generally sub-sampling is conducted to reduce the size of data. This downsampling process is called as pooling. Pooling locally summarizes the output of the previous layer making translation invariance, which focuses on the presence of the feature rather than the location (Goodfellow et al., 2016). In addition, pooling helps to avoid the overfitting problem by making the model simpler. The number of weights to be optimized is significantly reduced at the pooling layers, creating a lower computational cost. Max pooling is commonly used based on the concept that the maximum values of a feature map can represent local features (Zhou and Chellappa, 1988). Finally, fully connected layers are used as the classifier using final output feature maps. By using the features from previous layers, fully connected layers determine the final class with the highest probability using a softmax function. It is a commonly used classifier in multi-class classification problems in neural networks (Goodfellow et al., 2016; Yu et al., 2017; Kim et al., 2018a). Fully connected layers consist of a set of weights to be optimized for a node. By using the features from previous layers, fully connected layers determine the final class with the highest probability using a softmax function. It is a commonly used classifier in multi-class classification problems in neural networks (Goodfellow et al., 2016; Yu et al., 2017; Kim et al., 2018a). Fully connected layers consist of a set of weights to be optimized for a node.

An activation function converts the sum of input data into an output result. To get the benefit of multiple layers on a neural network, it is essential to use a nonlinear activation function. The rectified linear unit

(ReLU) is the most popular activation function in deep learning for its excellent performance with a relatively simple structure (Glorot et al., 2011; LeCun et al., 2015).

All of the weights, such as filters in convolutional layers and nodes in the fully connected layers, are randomly initialized. By reducing the error between the estimated result and reference data, weights are gradually optimized. This iterative process is called backpropagation, which calculates the derivative of the error function to find the minimum error (Rumerhar, 1986; Goodfellow et al., 2016). All of the weights are updated by the optimization method using the calculated gradient.

In this study CNN was implemented using the Keras open-source library. There are many ways to construct the CNN architecture. Therefore, it is important to find an optimal model that works well with data considering their characteristics. Unfortunately, there is no way to directly find an optimal model in deep learning. A multitude of tests is typically conducted to find the optimal CNN parameters considering performance and efficiency. In this study, 32, 64, 128 and 256 filters at convolutional layers were tested to determine an optimal structure. We finally constructed a CNN model, which consisted of four convolutional layers with 32 3 × 3-sized filters. The ReLU activation function was adopted at each layer. Max pooling with a 2 × 2 window and a stride of 2 was performed after the second and fourth convolutional layers. A fully connected layer with 256 nodes was applied after the convolutional and max-pooling layers. A soft-max function was used to classify the LCZ type. The adaptive moment estimation (ADAM) optimizer was used to minimize the error function, which is typically used in neural

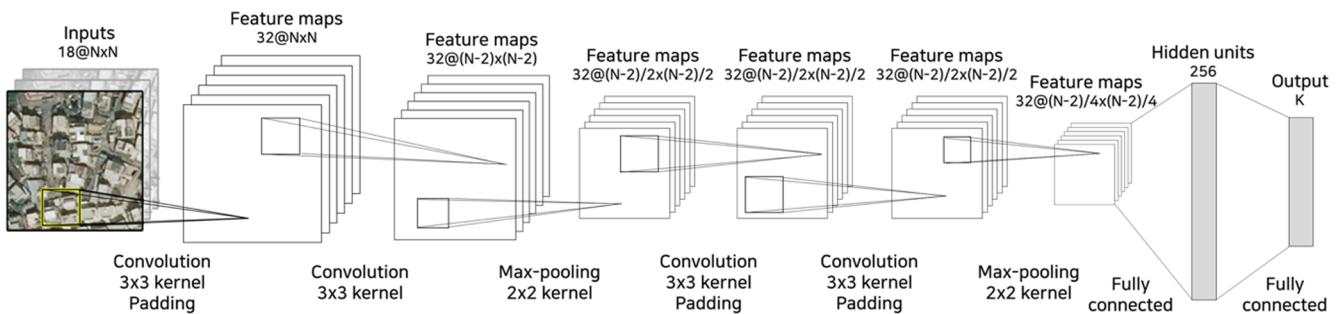


Fig. 3. The structure of CNN we designed in this study. N indicates the size of input image (i.e., a 10×10 size image has N of 10). The k in the last output means the number of LCZ types to be classified for each city.

networks especially for the classification task (Kingma and Ba, 2014). A graphics processor unit (GPU) of Nvidia GTX 1080Ti with 11 GB memory was used to speed up model training with 256 batch size and 1000 epochs. The final CNN structure used in this study is shown in Fig. 3.

3.3. Classification scheme design

To produce a 100 m resolution pixel-based LCZ map from Landsat images, this study used two classifiers—RF and CNN. We designed five classification schemes (three RF-based schemes (S1–S3) and two CNN-based ones (S4–S5) with different input features and classifiers (Fig. 4)).

3.3.1. Benchmark RF-based schemes (S1–S3)

RF is the classifier adopted by the existing LCZ classification community, including the WUDAPT method. We designed three schemes (S1–S3) with RF, based on the benchmark of existing studies. S1 corresponds to the existing WUDAPT method. The 30 m resolution Landsat images were bilinearly resampled to 10 m resolution, then resampled to 100 m resolution by a zonal mean function based on the LCZ grid area. S2 benchmarked the method proposed by Danylo et al. (2016), which achieved an increase in the classification accuracy by adding more spectral information to the WUDAPT model as input variables. The

10 m bilinear resampled Landsat images were resampled to 100 m, not only by zonal mean but also by maximum and minimum within the LCZ grid area. The three features were constructed for each Landsat band in S2. S3 benchmarked the method suggested by Verdonck et al. (2017). To consider the contextual characteristics of a feature, the mean, minimum, maximum, median, 25th and 75th quantile values of the nine pixels in a 3×3 window (i.e., one center pixel and its surrounding eight pixels) were calculated from 100-meter zonal-mean Landsat images. In each scheme, we used the features constructed from 18 bands (i.e., 9 bands for one scene) of two Landsat images in (or very close to) the winter and summer seasons (Table 2) as input variables. In summary, the number of input variables of each scheme was: 18, 54, and 108 for S1, S2, and S3, respectively (Table 4). We extracted the pixel values of the input variables at the location corresponding to LCZ reference pixels in each scheme.

3.3.2. CNN-based schemes (S4–S5)

We proposed two different schemes based on CNN. The 30 m resolution Landsat images were bilinearly resampled to 10 m, allowing 100 (10×10) pixels to be placed in a single 100 m LCZ grid. Each 10 m resolution image was normalized using the min-max approach, to reduce training time (Ba et al., 2016). In the case of S4, the 10×10 size features of 10 m resolution Landsat images in each LCZ reference pixel

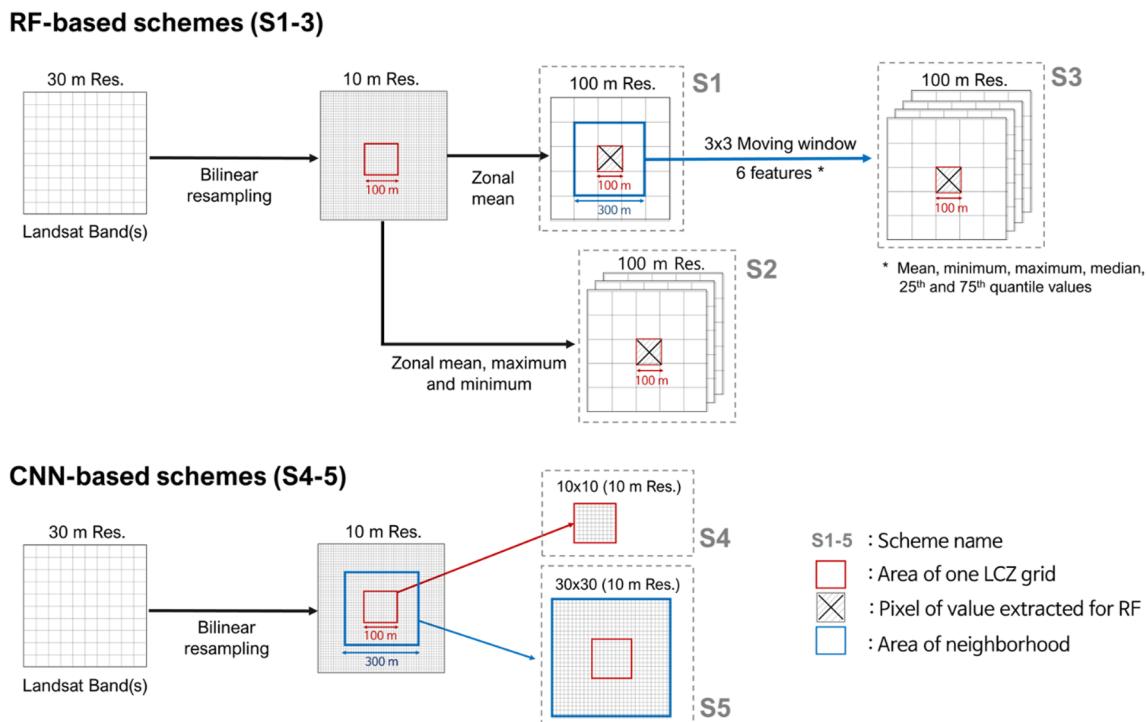


Fig. 4. The schematic process flow showing how to prepare the input features for each scheme.

Table 4

Summary of each scheme with input feature types.

Scheme	Classifier	# of input features	Feature types (spatial resolution)
S1	RF	18	Zonal mean (100 m) Zonal mean, maximum and minimum (100 m) Mean, minimum, maximum, median, 25th and 75th quantile values of 3 × 3 moving window (100 m) 10 × 10 sized (10 m) 30 × 30 sized (10 m)
S2	RF	54	
S3	RF	108	
S4	CNN	18	
S5	CNN	18	

Table 5

Accuracy assessment results for five schemes of four cities with the average statistic values from 10 times runs for each scheme. The numbers in parentheses are standard deviations of OA with 10 times runs.

Scheme	Rome			Hong Kong		
	OA (σ) %	OA _{urb} %	OA _{nat} %	OA (σ) %	OA _{urb} %	OA _{nat} %
S1 (RF)	72.05 (0.13)	68.17	79.13	71.58 (0.43)	52.96	79.27
S2 (RF)	72.45 (0.19)	68.17	80.27	71.42 (0.18)	56.73	77.49
S3 (RF)	75.36 (0.25)	73.76	78.27	75.37 (0.06)	64.34	79.93
S4 (CNN)	73.32 (0.64)	72.22	75.33	74.84 (0.52)	54.62	83.20
S5 (CNN)	80.34 (1.04)	81.99	77.34	79.80 (0.63)	65.15	85.85
Scheme	Madrid			Chicago		
	OA (σ) %	OA _{urb} %	OA _{nat} %	OA (σ) %	OA _{urb} %	OA _{nat} %
S1 (RF)	82.75 (0.18)	76.65	87.07	84.22 (0.06)	80.96	90.02
S2 (RF)	84.41 (0.07)	79.76	87.71	87.28 (0.09)	85.22	90.96
S3 (RF)	85.78 (0.14)	81.58	88.76	89.66 (0.10)	88.71	91.36
S4 (CNN)	85.33 (0.42)	80.67	88.64	86.18 (0.21)	83.44	91.07
S5 (CNN)	89.72 (0.41)	88.18	90.82	90.85 (0.38)	90.46	91.54

area were extracted and fed into CNN. The final scheme (S5) takes into consideration the surrounding area of a focus pixel (i.e., the same area of the moving window in S3). We extracted the 30 × 30 size features of 10 m resolution Landsat images and fed them into the CNN classifier. In summary, the S4 has a 10 × 10-sized 10 m resolution feature for each band, while the S5 has a 30 × 30-sized 10 m resolution feature for each band as input variables (Table 4). After the 10 m resolution images were fed into the CNN model, the fully connected layers could make a final decision of one LCZ class for each image in order to produce a 100 m resolution LCZ map. The number of trainable parameters of the S4 and S5 for the four cities are summarized in Table S1.

3.4. Modelling and accuracy assessment

A randomly selected 90% of the training samples (i.e., training set) were used to train the models and the remaining 10% were used to identify the optimum parameter values for the models. Through this process, we selected the optimal number of RF trees (i.e., ntree) within 100–1000 based on overall accuracy (OA) for each RF-based scheme (S1–3). In the case of the CNN-based schemes (S4–5), the model resulting in the best accuracy based on the 10% samples in 1000 epochs was selected. We ran the models ten times for each scheme to examine the robustness of the methods, and assessed accuracy using the separate test datasets (Table 3). For an assessment of accuracy, we used not only OA but also OA_{urb}, which is the accuracy among the urban LCZ types (LCZs 1–10) and OA_{nat}, which is the accuracy among the natural LCZ types (LCZs A–G). In addition, we obtained the F1-score (Eq. (1)) from user's accuracy (UA) and producer's accuracy (PA) of each LCZ class to further examine the classification accuracy by class. As the F1-score is the harmonic mean of UA and PA, the score is not only an indicator of the classification capability but also able to explain how similar the two values (i.e., UA and PA) are (Sokolova and Lapalme, 2009).

$$F1 = \frac{(2 \times UA \times PA)}{(UA + PA)} \quad (1)$$

Finally, we selected one model among the 10 simulated models to map LCZ for each city, based on the highest value of the sum of OA and OA_{urb} for each scheme. We also conducted McNemar's test to evaluate the significance of the differences in the classification results by scheme.

3.5. Transferability experiments

We further compared the transferability between CNN and RF classifiers by applying the LCZ models developed for three cities to the remaining city based on the best performing RF and CNN models from the experiment of individual cities. In other words, reference data of one city was used to evaluate the transferability of the LCZ model that was developed using reference data of the other three cities as training samples shown in Table 3. The procedure for designing the models for transferability test is the same as that documented in Section 3.4. Considering the different LCZ types by city, only the LCZ labels belonging to the test city were selected when training the LCZ models.

4. Results and discussion

4.1. Overall performance of the schemes

Table 5 shows the accuracy assessment results of the five schemes for four cities. When compared to S1 (i.e., the WUDAPT method), S2 showed an increase of OA of 2% for Madrid and of 3% for Chicago, which agrees with the findings from Danylo et al. (2016). Moreover, it should be noted that the OA_{urb} of S2 significantly increased when compared to that of S1 for Hong Kong, Madrid, and Chicago. Interestingly, the OA_{nat} did not significantly increase for all cities in S2. This suggests that putting various spectral information (i.e., maximum and minimum) as input variables might contribute to the increase in the accuracy for urban LCZ types (i.e., LCZ1–10), which have more heterogeneous spectral characteristics.

For Hong Kong, while the OA_{urb} of S2 was higher than that of S1, the OA_{nat} of S2 is lower than that of S1. For natural LCZ types (i.e., LCZA–Z) in Hong Kong, adding manually extracted features as input data rather decreased the accuracy. Moreover, there was no significant accuracy difference between S1 and S2 for Rome. This implies that including more contextual information as input variables in the RF does not always guarantee improving classification accuracy.

Unlike S1 and S2, where we manually selected input features, the CNN-based S4 can automatically learn multi-level features from the original input images. It is not surprising, therefore, that S4 shows higher OA value than S1 for all four cities. In addition, S4 showed higher OA than S2, except for Chicago; one possible reason is that the added contextual information in S2 was meaningful enough to improve accuracy in Chicago where the city blocks have regular arrangements. Athiwaratkun and Kang (2015) showed that using well-learned features as input variables in RF yielded higher accuracy than CNN.

The influence of considering neighborhood pixels as input features is seen in both RF- and CNN-based schemes. S3 produced the highest OA value among the three RF-based schemes (S1–3), which is consistent with Verdonck et al. (2017). The CNN-based S5, with 30 × 30-sized input features, showed the highest accuracy among the five schemes, by increasing OA in all cities by 5–8% when compared to the

WUDAPT method (S1). These findings agree well with [Zhang and Tang \(2019\)](#), who showed that accuracy improved when the surrounding areas of the center target pixel were fed into CNN. In particular, when comparing the accuracy difference between S2 and S3 with that between S4 and S5, the influence of contextual information appeared more effective in CNN than in RF. This implies that the input features of S5 that consider the surrounding areas of the target LCZ pixels (i.e., 30×30 -sized images) contributed to learning more meaningful features in the convolutional layers of CNN than those of S4 (i.e., 10×10 -sized images). Since the wider areas were considered more in S5 than in S4, the CNN could learn more significant patterns, probably due to the broader information integrated by combining local patterns, especially for urban features ([Min et al., 2017](#)). Moreover, the OA_{urb} of S5 increased by about 10–13% compared to that of S1 for four cities. The increasing rate of OA_{urb} between S5 and S1 is much higher than that of OA_{nat}, implying that CNN-based S5 can be considered as the most effective LCZ mapping model for the mega urban areas.

The imbalance problem of accuracy by class occurred when the number of samples differed greatly among classes, resulting in poor performance over the minority classes ([Huang et al., 2016](#); [Jeatrakul et al., 2010](#); [Zhou and Liu, 2006](#)). In particular, RF is known to be less sensitive to unbalanced sample size than neural network-based CNN ([Liu et al., 2018a](#); [Liu et al., 2013](#)). In Rome, S4 showed a higher OA_{urb} but a lower OA_{nat} than S2. For Hong Kong, on the other hand, S4 showed the opposite pattern. One reason may be that the ratio of samples among the LCZ classes varies by city. The sample sizes in [Table 3](#) show that Rome has fewer samples of natural classes, and more samples of urban classes than natural classes. In Hong Kong, however, the number of samples in the urban LCZ types was very small, while the number of samples in the natural LCZ types was much larger than that of urban classes. When training CNN, LCZ classes with a relatively large number of samples could be more correctly classified than weakly represented LCZ classes. Such an imbalance problem of training sample size by class seemed to be mitigated in S5 when compared to S4. Consequently, the consideration of neighborhood pixels in CNN led to the good classification of the LCZ classes even with a small sample size (i.e., natural classes for Rome, urban classes for Hong Kong).

The standard deviations of the results in [Table 5](#) show that the CNN-based schemes (S4–S5) yielded a higher variation of accuracy than those of the RF-based schemes (S1–S3). This implies that RF produces more consistent results than CNN, because RF is an ensemble-based model ([Lebedev et al., 2014](#); [Kursa, 2014](#); [Khoshgoftaar et al., 2007](#)). Despite the relatively high standard deviation in the S5 results, most of the S5 classifications produced higher accuracy than those of the RF-based schemes.

Using the most accurate model among the 10 simulations for each of the five schemes, the significance of the accuracy differences between the classifications was assessed by McNemar's Chi-squared test ([Fig. 5](#)). In Rome, S1 yielded an outcome comparable to S2, and the performance of both S2 and S3 were similar to that of S4. In the case of Hong Kong, S1 and S2 showed similar results, as did S3 and S4. For Madrid and Chicago, all classifications were statistically different, except for

the S3/S4 pair for Madrid. S5, the CNN scheme, achieved significantly higher accuracy than the other schemes in all four cities ([Table 5](#)). S5 is considered to have a great utility in LCZ classification because it consistently shows statistical significance with the other schemes, resulting in the highest classification accuracy ([Table 5](#) and [Fig. 5](#)). Interestingly, the accuracy difference between S4 and S3 was not significant in Rome, Hong Kong, and Madrid. This implies that the RF model considering the neighborhood area (300×300 m; S3) produced a similar performance with the CNN model without considering such a large neighborhood of the LCZ grid (100×100 m; S4). In the case of Chicago, however, S3 and S4 resulted in a significant difference, showing higher accuracy of the RF model for S3 than the CNN for S4 ([Table 5](#)). In Chicago where the city has been developed based on regular grids, increasing features (i.e., spectral and neighboring information) could bring sufficiently high accuracy in RF. This is particularly true in light of the accuracy differences between the pair S2 and S1 and the pair S3 and S1, which are both the highest among four cities in [Table 5](#).

4.2. Classification accuracy per class

[Fig. 6](#) shows the F1-score for each of the four cities on the average of 10-time runs per scheme. [Figs. 7–10](#) show the confusion matrices of the most accurate models among the 10-time runs of S3, which is the best scheme among the RF-based schemes, and S5, which is the best of the CNN-based schemes, as shown in [Table 5](#).

It should be noted that CNN-based S5 showed the highest F1-score among the five schemes on LCZ5 and LCZ6 for all cities except the red-star classes. For LCZ5 and LCZ6, where the abundant trees are mixed with openly arranged low or mid-rise buildings, the RF-based S3 misclassified them as other classes, such as densely packed buildings (i.e., LCZ1–3) or natural LCZ types ([Figs. 7–10](#)). On the other hand, the CNN-based S5 classified the classes more accurately than S3. One possible reason is that CNN can learn the regions of mixed pixels with buildings in the images by incorporating the surrounding area information. [Awrangjeb et al. \(2012\)](#) reported that using the building edge information improved the detection performance of the trees, which were misclassified as buildings. For S3 in Rome, LCZ6 (open low-rise) was confused by various classes, especially LCZB (scattered trees) in [Fig. 7a](#). However, the accuracy of LCZ6 clearly increased in S5 ([Fig. 7b](#)). Among the urban LCZ types in Hong Kong, LCZ5 (open mid-rise) and LCZ6 (open mid-rise) showed higher F1-scores in S5 than those in the other schemes. In [Fig. 8a](#), LCZ5 was often confused with LCZ4 (open high-rise) and LCZA (dense trees) in S3, and LCZ6 was confused with LCZD (low plants) in S3. [Fig. 8b](#) shows that such misclassification of LCZ5 and LCZ6 happened much less in S5 than in S3. In Madrid, the confusion of LCZ5 with LCZ2 in S3 significantly improved in S5, and the confusion between LCZ6 and LCZ5 in S3 also clearly improved in S5 ([Fig. 9a–b](#)).

The LCZ9 (sparsely built) in the CNN-based schemes showed higher F1-scores than the RF-based schemes did. Especially in Chicago, LCZ9 tended to be misclassified as LCZD (low plants) in S3, but this error was observed to be reduced in S5 ([Fig. 10a–b](#)). In fact, LCZ9 shows a unique spatial structure where small-sized buildings are sparsely built among

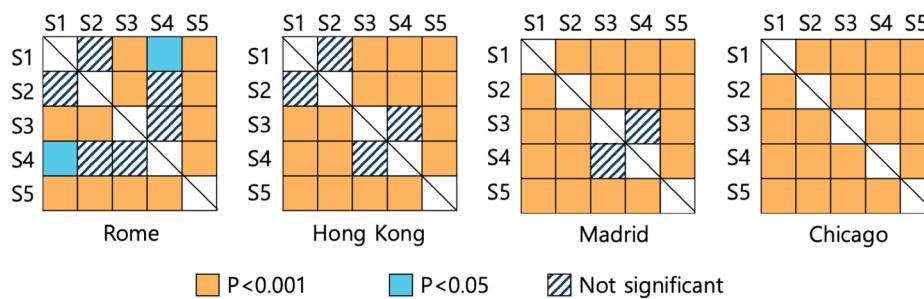


Fig. 5. Results of McNemar's Chi-squared test. The orange squares indicate the significant accuracy difference at the 99% confidence level, while the blue ones at the 95% confidence level. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

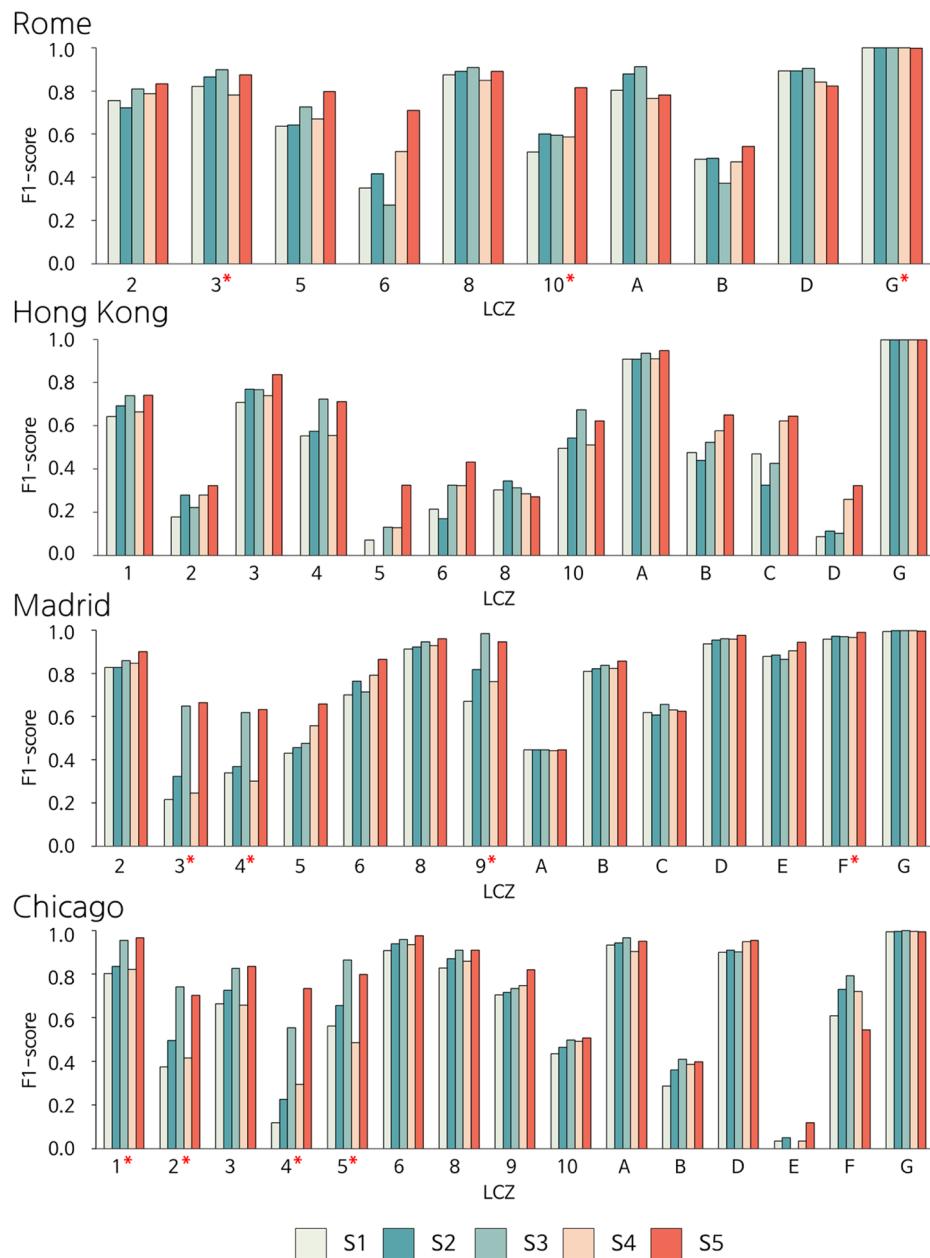


Fig. 6. Comparison of F1-score between the five schemes (S1–5) for the four cities. The F1-scores were averaged from 10 runs for each scheme. * indicates the red-star classes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

vegetation. Thus, it is not surprising that the CNN classifiers classified this class well, which is consistent with the results of Fu et al. (2018), who reported that CNN showed the higher classification accuracy for the mixed objects when compared to RF.

One possible reason that the CNN's F1-score was relatively lower than that of RF in LCZA in Rome and LCZ8 in Hong Kong could be a data imbalance problem due to the relatively small number of training samples of the classes (Table 3). In Hong Kong, however, LCZ5 and LCZ6, which have open arrangement of buildings mixed with tress, confirmed that the F1-scores of CNN were higher than those of RF, even if the number of samples was small. In this study, the data imbalance problem might exert a relatively weak influence on accuracy because LCZ classes with a small number of polygons were classified into the red-star class in each city.

We calculated the F1-score difference between S3 and S2 (RF schemes) and S5 and S4 (CNN schemes) to identify the neighboring

effects for all LCZs, except red-star classes. Interestingly, the class yielding the highest difference for each city is LCZ6 in Rome, LCZ5 in Hong Kong and Madrid, and LCZ3 in Chicago (Figure S1) for the difference between S5 and S4 (CNN schemes). This result implies that the consideration of neighborhoods could be more effective for urban types classes, especially when using CNN. Moreover, LCZ5 and LCZ6, which are open arrangement classes, showed significantly improved accuracy when using CNN with the incorporation of neighborhood information for all cities except Chicago, where the accuracy is still good enough, even before the consideration of neighboring areas.

The results of the red-star class should be carefully interpreted in Fig. 6. A positive bias may appear because the training and test sets were randomly stratified samples within one polygon (Verdonck et al., 2017). In particular, the red-star classes tended to have a higher F1-score than the other schemes in S3 and S5, which incorporate the neighboring areas into their classifications.

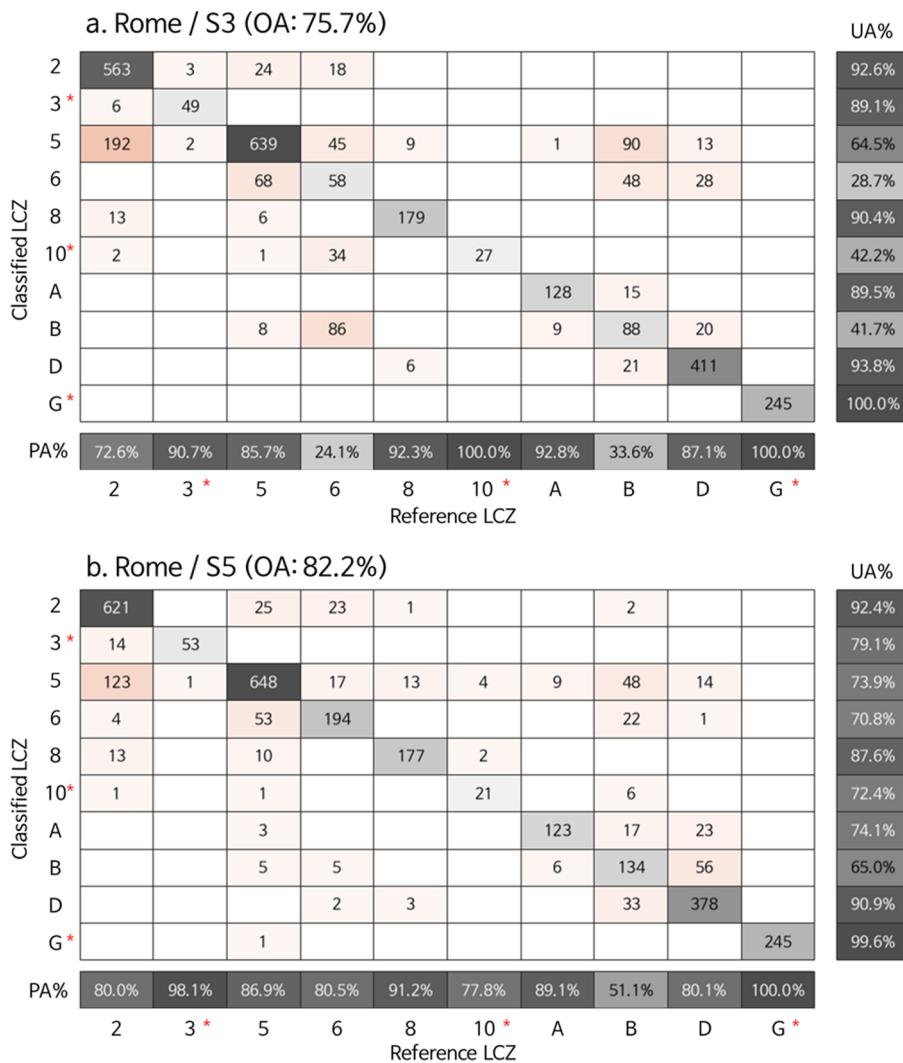


Fig. 7. Confusion matrices of the most accurate model among the 10-time runs of S3, the best scheme among the RF-based schemes, and S5, the best of the CNN-based schemes for Rome. * indicates the red-star classes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.3. Mapping LCZ for four cities

Figs. 11 and 12 show the 30 m resolution GMIS with LCZ references and the developed LCZ maps for the four cities from S3 and S5, which have the highest accuracy among the RF- and CNN-based schemes, respectively. We divided the results of the two classified maps into four cases and then calculated their ratios, as shown in Table 6.

The two maps of Rome have different classification results for the suburban areas consisting of open arrangements that appear at a distance from the monocentric city center. In Rome, the study domain denoted by the blue box (middle bottom) in S3 was classified as LCZ5 (open mid-rise), while that in S5 was classified as LCZD (low plants). In addition, the east bottom part of the study domain bound by yellow box shows an amount of open low-rise areas in S5, while S3 tended to show this area as scattered trees. On the other hand, S3 classified the northeastern part more as open mid-rise classes than in S5, which classified the area as low plants and scattered trees. When compared to the impervious cover and Google Earth images (not shown), S5 seemed to classify the built-type classes better than the S3 did, while S3 tended to be confused between vegetation and mixed buildings. These results correspond to the accuracy assessment where the F1-scores of LCZ5 and LCZ6 showed better performance in CNN than in RF in Rome. It should be noted that LCZs 5, 6, B and D in Rome appear as the dominant LCZ

classes in the classified maps (Table S2). Considering that LCZs 5 and 6 are open-arrangement urban LCZ types, CNN's ability to classify these types of LCZs better than RF seems to account for the map disparities for Rome. Two maps of Rome were more likely to be classified differently between urban and natural LCZ types: 21.01% (Table 6).

In Hong Kong, the non-residential (i.e., hilly) and habitable areas are clearly distinguished from each other, so the difference within each natural and urban LCZ type is somewhat larger than that between natural and urban LCZ types in the two maps (Table 6). Especially, the maps of S3 and S5 in Hong Kong exhibited differences within natural LCZ types, such as low plants, trees, and bushes, in some areas based on visual inspection. The classification accuracy of LCZD in Hong Kong was higher in the CNN-based schemes than the RF-based schemes, considering both the F1-scores and the confusion matrices. The region of the study bounded by a black box in S5 was classified more as LCZD (low plants) than LCZC (bush, scrub), as opposed to its more predominant classification of LCZC (bush, scrub) in S3. As in the land use map of Hong Kong (Chan et al., 2016), these areas are dominated by grassland, which implies that CNN could distinguish between plants and scrub better than RF. Unlike RF, less confusion occurred between LCZC and LCZD in CNN, which corresponds to the results of the confusion matrices for not only Hong Kong but also Madrid (Figs. 8 & 9).

In Madrid, various small clustered areas called autonomous

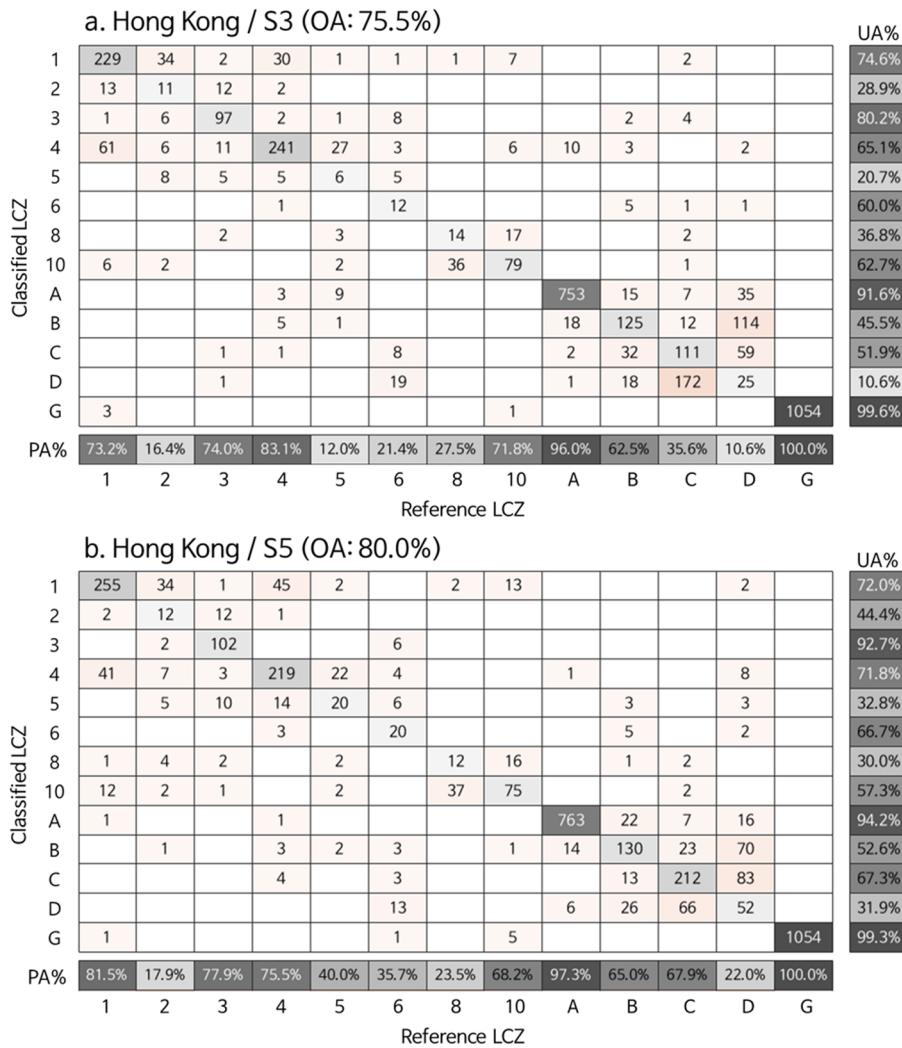


Fig. 8. Confusion matrices of the most accurate model among the 10-time runs of S3, the best scheme among the RF-based schemes, and S5, the best of the CNN-based schemes for Hong Kong. * indicates the red-star classes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

communities surround the city-core. This is a conurbation in which extended suburbs and villages comprise dense built-type classes, such as compact mid-rise; identified using Google Earth images (not shown). These clusters appear more distinctly in S5 than in S3 (bounded by a blue box), and were compared to the impervious cover, which clearly showed that the S5 classified the clusters of these dense buildings relatively well. In RF, these compact clusters were misclassified as open arrangement buildings, which corresponds well to the accuracy assessment result showing that LCZ2 was more often confused among other urban LCZ types in RF compared to CNN (Fig. 9). It is interesting that LCZ2 was not confused that often with natural LCZ types in the confusion matrix of S3, but the generated LCZ map of S3 showed some misclassification of LCZ2, often confused with LCZB. This could be because there are few reference samples to test in these cluster areas. In the case of Madrid, the ratio of natural LCZ types in the study region is remarkably high, resulting in a classification difference among natural LCZ types that is the highest, at 14.65% (Table 6). The different classification of urban and natural LCZ types in the two maps (~5.02%; Table 6), could originate from the municipalities surrounding Madrid, which were better classified in CNN than in RF due to the textural patterns over large areas.

The promising results for LCZ classification by CNN could be useful data for the various urban climate studies especially for the regions with abundant LCZ classes mixed with different objects (i.e., buildings

with trees and bare-soil with shrubs). Although to a lesser extent than Rome, the two maps of Chicago show a difference in the open arrangement of low-density suburban areas surrounding the high-density urban center, particularly in their different classifications for LCZ types: Urban and Natural (10.43% in Table 6). However, it should be noted that CNN could result in low user's accuracy. In Chicago, LCZ9 (sparsely built) was distributed widely in the middle top of the study domain (bounded by a black box) of the maps of S3 and S5. The CNN-based S5 has the advantage of catching the sparse buildings between some of the trees and plants by object detection, but LCZ9 seems instead to be over-classified on the map of S5 when compared to S3. This also corresponds well with the result of the confusion matrices in Chicago (Fig. 10), as S5 showed a higher producer's accuracy, but a lower user's accuracy than S3 for LCZ9. Although this paper used only the Landsat data corresponding to the WUDAPT method, if input variables, such as Sentinel-1 backscattered data, were used additionally to explain the characteristics of buildings (Koppel et al., 2017; Demuzere et al., 2019), the limitation of the CNN could be improved.

The CNN-based classification is known to take more time from the training stage to the mapping stage than the RF. Nonetheless, the CNN-based S5 had high classification accuracy and was of high value in classifying specific LCZ types where the objects were mixed, when compared to those of the RF-based S3.

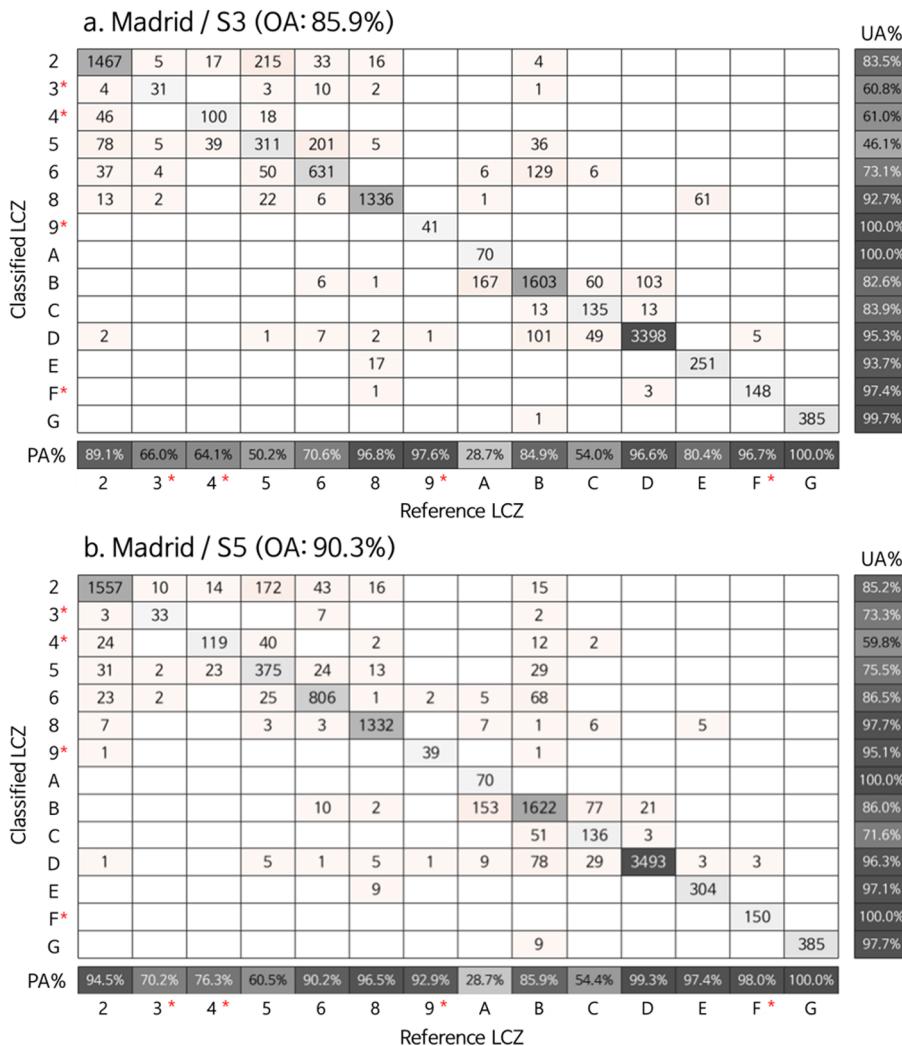


Fig. 9. Confusion matrices of the most accurate model among the 10-time runs of S3, the best scheme among the RF-based schemes, and S5, the best of the CNN-based schemes for Madrid. * indicates the red-star classes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.4. Evaluation of model transferability

Table 7 shows the transferability assessment results based on the two best performing schemes from the experiment of individual cities. Interestingly, the CNN-based scheme (S5) showed a distinctly higher performance than the RF-based scheme (S3) both for the OA and OA_{urb}. In particular, the significant improvement of OA_{urb} for all four cities was found corresponding to the findings in our single-city experiments, which implies the superiority of the object detection-based characteristics of CNN classifiers. In recent years, research on a transferability framework has been attempted, with LCZ reference samples of specific cities trained and applied to other cities (Demuzere et al., 2019; Qiu et al., 2019; Yokoya et al., 2018). For example, Demuzere et al. (2019) examined global transferability of LCZ models using RF classifiers with the Google Earth Engine. However, they found the transferability of the LCZ models was still challenging because the accuracies of their models were generally poor (average OA of the 15 cities close to 50%). The results of this present study identified the advantages of using CNN classifiers over RF in the transferability framework of LCZ classification, especially for urban-type LCZ classification. When compared to the single-city experiment results in Table 5, the accuracy of the transferability experiment was a bit lower, varying by city, possibly due to the limited coverage of reference data for training. It is crucial to construct thorough and sufficient reference data of LCZ classes for various urban

structural types over the globe to improve the transferability of LCZ models.

4.5. Novelty, limitations, and future directions

To our knowledge, this is the first study to compare and discuss LCZ classification results between RF and CNN classifiers, in detail. Although some previous studies tried to compare the LCZ classification results among different classifiers including basic machine learning algorithms (i.e., RF, Support Vector Machine (SVM) and Neural Networks (NN)), they didn't examine deep learning-based classifiers (Bechtel and Daneke, 2012; Bechtel et al., 2016). More recently, a few studies on LCZ classification using CNN classifiers have been conducted (Sukhanov et al., 2017; Qiu et al., 2018). However, they did not fully compare the classification performance with the existing models using RF classifiers. Furthermore, this paper compared the results using different sizes of input data (i.e., 10 × 10 and 30 × 30) fed into the CNN classifiers. The positive effect of an increasing input patch size in CNN has been proven in different studies (Hamwood et al., 2018), which is also shown in the LCZ mapping in this study. In particular, the specific LCZ classes (i.e. LCZ5 and LCZ6) that have a favorable impact using CNN were identified when an increasing size of input data was applied when compared to the impact of RF under the same conditions. This result can provide meaningful guidance for the continued research of

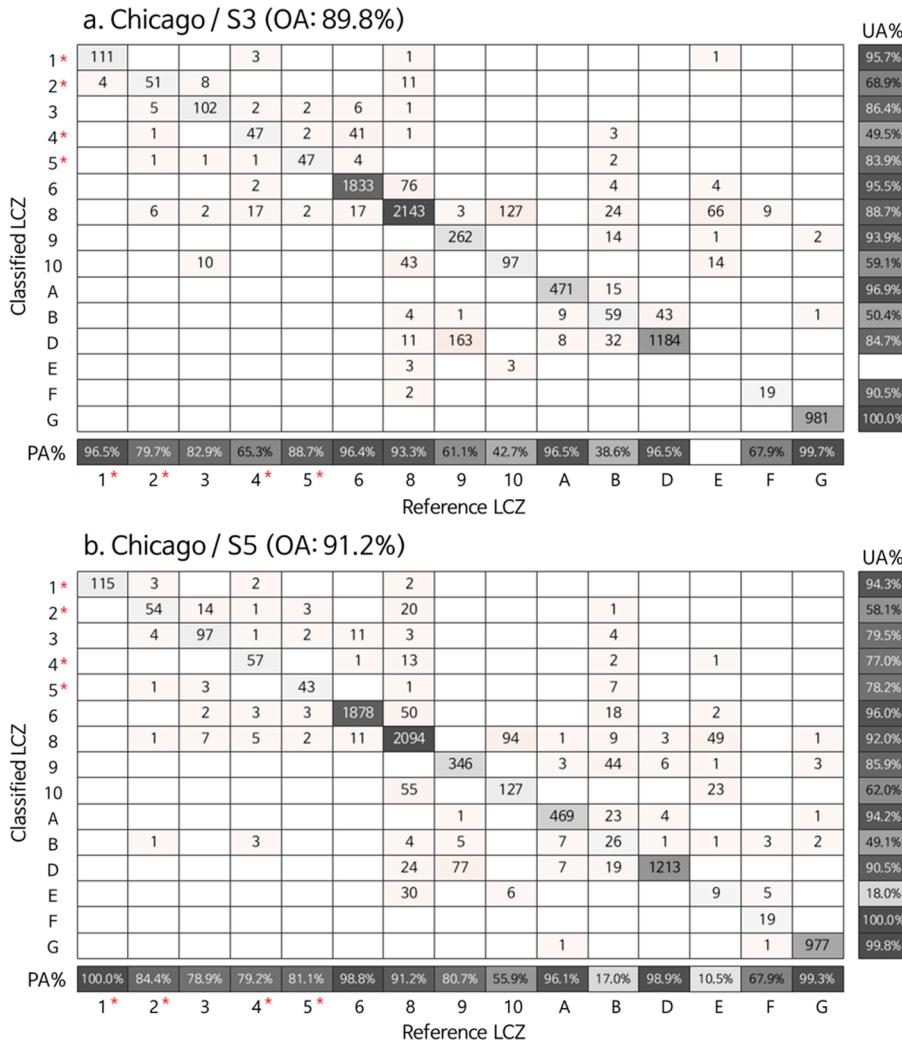


Fig. 10. Confusion matrices of the most accurate model among the 10-time runs of S3, the best scheme among the RF-based schemes, and S5, the best of the CNN-based schemes for Chicago. * indicates the red-star classes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

LCZ classification using CNN classifiers. In addition, many LCZ classification studies have focused on only OA for their accuracy assessment. In this study, OA_{urb}, the overall accuracy between the urban LCZ types, was also carefully examined when comparing the accuracy of the proposed schemes. The validity of the results of the LCZ classification using two classifiers was strengthened by applying it to four cities with different urban structures and geographical characteristics in various continents such as Europe, Asia, and America.

The major limitation of this study is the small sample size of the specific LCZ classes (i.e., red-star classes). In this study, reference LCZ data were provided by the IEEE data fusion contest to ensure the reliability of the data. In order to validate LCZ classification with a minimum bias, the reference polygons should be divided into training and test sets. For the red-star classes, however, we divided the datasets by stratified random sampling among the pixels in a polygon, because of the limited number of polygons. The red-star classes are likely to have a positive bias in their classification results, so care is needed in any interpretation. Further improvement in accuracy for the LCZ classes with a small number of samples (i.e., LCZE in Chicago) is expected through the utilization of data augmentation methods discussed by Yokoya et al. (2018). The CNN-based S5 showed higher accuracy in four cities when compared to other RF-based schemes, but we could not pinpoint which objects contributed to the detection of each LCZ class in CNN. The use of high spatial resolution satellite data (i.e., Sentinels) in

future LCZ classification will improve the object detection ability of CNN classifiers. In addition, using high-resolution images will enable a more detailed analysis, especially if heat maps of CNN classifiers are used. It is also possible to make Landsat images as higher-resolution images by using pan-sharpening techniques (Xing et al., 2018; Gilbertson et al., 2017; Rahaman et al., 2017). Recently, in the deep learning field, CNN and other machine learning classifiers are being combined to construct better models (Zhang et al., 2018; Soltan et al., 2014). These techniques can be applied to the field of LCZ classification as well.

When it comes to the CNN model, the fully connected network (FCN) is adopted in recent land cover classification, with the aspect of the semantic segmentation (Mohammadianesh et al., 2019; Wurm et al., 2019; Yue et al., 2019). FCN has the advantage of learning spatial relationships at different scales (Volpi and Tuia, 2016), which can be expected to yield improved performance in LCZ classification by taking into account the various size and shape of each LCZ class in future work.

5. Conclusion

In this study, we compared the two classifiers, RF and CNN, for LCZ classification in four mega cities—Rome, Hong Kong, Madrid, and Chicago—using bitemporal Landsat images. A total of five schemes

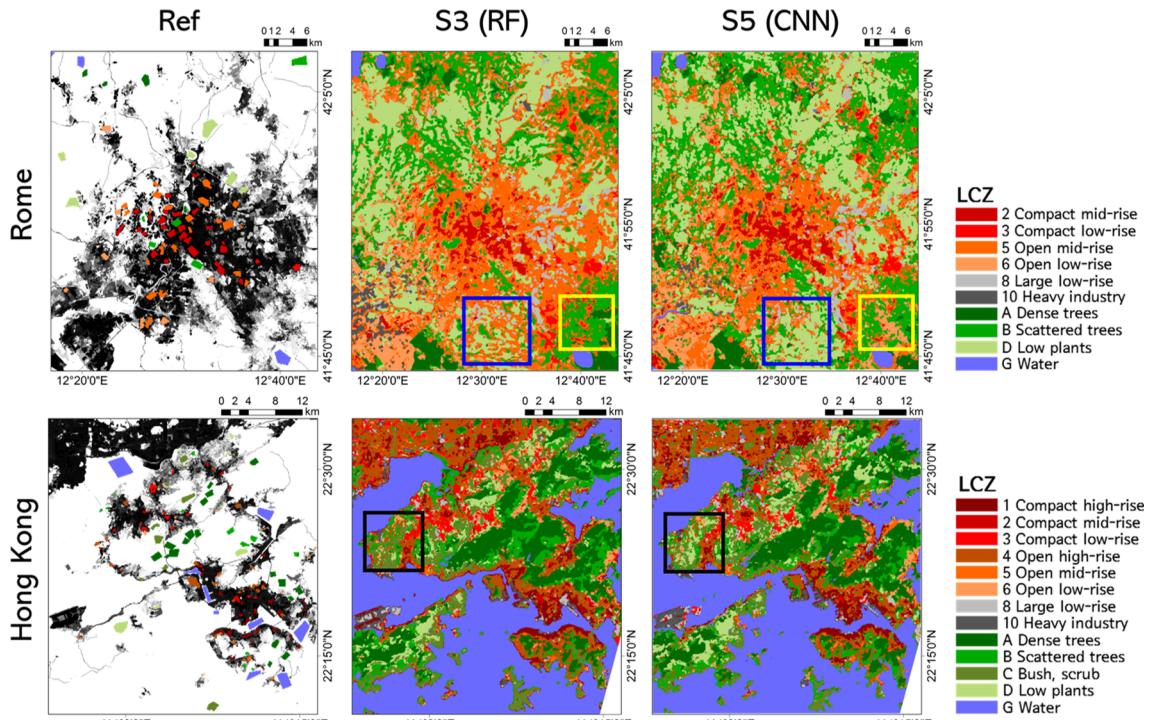


Fig. 11. LCZ maps of the best classification model among the 10-time runs of S3, the best scheme among the RF-based schemes, and S5, the best of the CNN-based schemes for Rome and Hong Kong. Impervious covers from GMIS and LCZ reference datasets are also presented.

were constructed and compared. Three RF-based schemes (S1–3) were benchmarked based on previous LCZ classification research studies. Two CNN-based schemes (S4–5) were benchmarked using different input feature sizes. Among the five schemes, S5 showed the best classification performance. When compared to the existing WUDAPT

workflow (i.e., S1), the OA and OA_{urb} of S5 increased by about 6–8% and 10–13%, respectively, for the four cities. This study has revealed that the CNN classifiers were particularly good at classifying the specific LCZ classes in which buildings were mixed with trees or buildings and trees were sparsely distributed. We also found that the

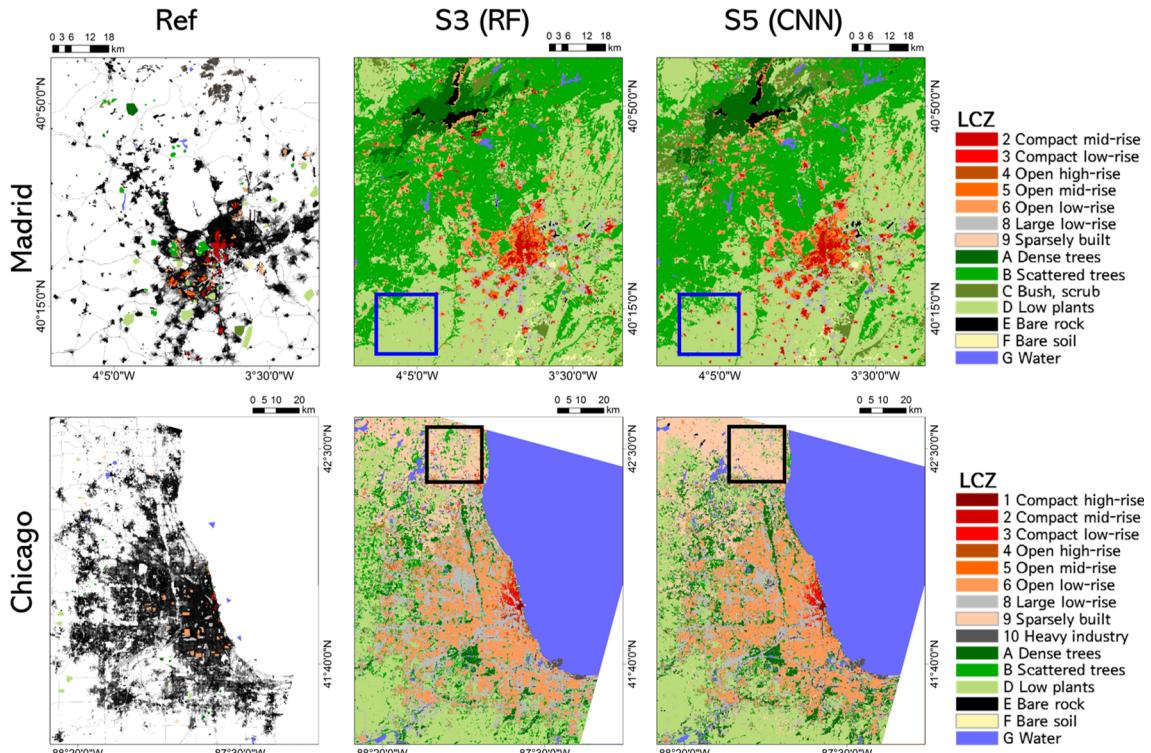


Fig. 12. LCZ maps of the best classification model among the 10-time runs of S3, the best scheme among the RF-based schemes, and S5, the best of the CNN-based schemes for Madrid and Chicago. Impervious covers from GMIS and LCZ reference datasets are also presented.

Table 6

The percentages of the LCZ differences between two classified maps (S3 and S5) for four cities shown in Figs. 11 and 12.

	Rome	Hong Kong	Madrid	Chicago
Classification within the same LCZ	60.21%	73.47%	77.87%	80.31%
Different classification within Urban LCZ types	11.29%	8.09%	2.46%	5.70%
Different Classification within Natural LCZ types	7.49%	11.10%	14.65%	3.56%
Different Classification for LCZ types: Urban and Natural	21.01%	7.33%	5.02%	10.43%

Table 7

Transferability assessment results by test city based on S3 and S5, the best performing RF and CNN schemes from the single city experiments, respectively. The overall accuracies were extracted from the best model among 10-time runs.

Scheme	Rome			Hong Kong		
	OA %	OA _{urb} %	OA _{nat} %	OA %	OA _{urb} %	OA _{nat} %
S3 (RF)	45.20	43.33	48.61	52.03	5.34	71.31
S5 (CNN)	62.69	67.42	54.07	58.68	28.18	71.27
Scheme	Madrid			Chicago		
	OA %	OA _{urb} %	OA _{nat} %	OA %	OA _{urb} %	OA _{nat} %
S3 (RF)	60.64	47.68	69.83	27.24	6.89	63.45
S5 (CNN)	78.03	77.08	78.71	41.52	25.13	70.70

classification performance of CNN significantly improved when the input features were created with consideration of the larger neighborhood areas. The results from the transferability experiment of the LCZ models supported the superiority of the CNN approach over RF in terms of both OA and OA_{urb} for all four cities. In the future, the CNN-based approach will become more advantageous when incorporating higher-resolution satellite images (i.e., Sentinels) and additional spatio-temporal features.

Acknowledgements

This research was supported by the Space Technology Development Program and the Basic Science Research Program through the National Foundation of Korea (NRF) funded by the Ministry of Science, ICT, & Future Planning and the Ministry of Education of Korea, respectively (Grant: NRF-2017M1A3A3A02015981; NRF-2017R1D1A1B03028129), and the Korea Meteorological Administration Research and Development Program under Grant KMIPA 2017-7010. CY was also supported by Global PhD Fellowship Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education (NRF-2018H1A2A1062207). We also would like to thank WUDAPT (the World Urban Database and Access Portal Tools project, www.wudapt.org), the IEEE GRSS Image Analysis and Data Fusion Technical Committee, and all the contributors for LCZ ground-truth samples, in particular Chao Ren, Dragan Milosevic, Guillaume Dumas, and Maria De Fatima Andrade.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.isprsjprs.2019.09.009>.

References

- Athiwaratkun, B., Kang, K., 2015. Feature representation in convolutional neural networks. arXiv preprint arXiv:1507.02313.
- Awrangjeb, M., Zhang, C., Fraser, C.S., 2012. Building detection in complex scenes thorough effective separation of buildings from trees. Photogramm. Eng. Remote Sens. 78, 729–745.
- Ba, J.L., Kiros, J.R., Hinton, G.E., 2016. Layer normalization. arXiv preprint arXiv:1607.06450.
- Barnes, K.B., Morgan, J., Roberge, M., 2001. Impervious surfaces and the quality of natural and built environments. Department of Geography and Environmental Planning, Towson University, Baltimore.
- Bechtel, B., Alexander, P.J., Beck, C., Böhner, J., Brousse, O., Ching, J., Demuzere, M., Fonte, C., Gál, T., Hidalgo, J., 2019. Generating WUDAPT Level 0 data—Current status of production and evaluation. Urban Clim. 27, 24–45.
- Bechtel, B., Alexander, P.J., Böhner, J., Ching, J., Conrad, O., Feddema, J., Mills, G., See, L., Stewart, I., 2015. Mapping local climate zones for a worldwide database of the form and function of cities. ISPRS Int. J. Geo-Inf. 4, 199–219.
- Bechtel, B., Daneke, C., 2012. Classification of local climate zones based on multiple earth observation data. IEEE J-Stars 5, 1191.
- Bechtel, B., Demuzere, M., Sismanidis, P., Fenner, D., Brousse, O., Beck, C., Van Coillie, F., Conrad, O., Keramitsoglou, I., Middel, A., 2017. Quality of crowdsourced data on urban morphology—The human influence experiment (HUMINEX). Urban Sci. 1, 15.
- Bechtel, B., See, L., Mills, G., Foley, M., 2016. Classification of local climate zones using SAR and multispectral data in an arid environment. IEEE J-Stars 9, 3097–3105.
- Beck, C., Straub, A., Breitner, S., Cyrys, J., Philipp, A., Rathmann, J., Schneider, A., Wolf, K., Jacobbeit, J., 2018. Air temperature characteristics of local climate zones in the Augsburg urban area (Bavaria, southern Germany) under varying synoptic conditions. Urban Clim. 25, 152–166.
- Bontemps, S., Defourny, P., Bogaert, E.V., Arino, O., Kalogirou, V., Perez, J.R., 2011. GLOBECOVER 2009—Products description and validation report.
- Breiman, L., 2001. Random forests. Machine Learn. 45, 5–32.
- Cai, M., Ren, C., Xu, Y., Lau, K.K.-L., 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 Clim. 24, 485–502.
- Chan, E.H., Wang, A., Lang, W., 2016. Comprehensive Evaluation Framework for Sustainable Land Use: Case Study of Hong Kong in 2000–2010. J. Urban Plann. Dev. 142, 05016007.
- Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Lu, M., 2015. Global land cover mapping at 30 m resolution: A POK-based operational approach. ISPRS J. Photogramm. 103, 7–27.
- Cohen, B., 2015. Urbanization, City growth, and the New United Nations development agenda. Cornerstone 3, 4–7.
- Danylo, O., See, L., Bechtel, B., Schepaschenko, D., Fritz, S., 2016. Contributing to WUDAPT: a local climate zone classification of two cities in Ukraine. IEEE J-Stars 9, 1841–1853.
- de Colstoun, E.C.B., Huang, C., Wang, P., Tilton, J.C., Tan, B., Phillips, J., Niemczura, S., Ling, P.-Y., Wolfe, R., 2017. Documentation for the Global Man-made Impervious Surface (GMIS) Dataset From Landsat.
- Demuzere, M., Bechtel, B., Mills, G., 2019. Global transferability of local climate zone models. Urban Clim. 27, 46–63.
- Ellickson, R.C., 2012. The law and economics of street layouts: How a grid pattern benefits a downtown. Ala. L. Rev. 64, 463.
- Fallmann, J., Forkel, R., Emeis, S., 2016. Secondary effects of urban heat island mitigation measures on air quality. Atmos. Environ. 125, 199–211.
- Founda, D., Santamouris, M., 2017. Synergies between urban heat island and heat waves in Athens (Greece), during an extremely hot summer (2012). Sci. Rep. 7, 10973.
- Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., Huang, X., 2010. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. Remote Sens. Environ. 114, 168–182.
- Fu, T., Ma, L., Li, M., Johnson, B.A., 2018. Using convolutional neural network to identify irregular segmentation objects from very high-resolution remote sensing imagery. J. Appl. Remote Sens. 12, 025010.
- Gilbertson, J.K., Kemp, J., Van Niekerk, A., 2017. Effect of pan-sharpening multi-temporal Landsat 8 imagery for crop type differentiation using different classification techniques. Comput. Electron. Agric. 134, 151–159.
- Giridharan, R., Emmanuel, R., 2018. The impact of urban compactness, comfort strategies and energy consumption on tropical urban heat island intensity: a review. Sustain. Cities Soc. 40, 677–687.
- Giridharan, R., Ganeshan, S., Lau, S., 2004. Daytime urban heat island effect in high-rise and high-density residential developments in Hong Kong. Energy Build. 36, 525–534.
- Glorot, X., Bordes, A., Bengio, Y., 2011. Deep sparse rectifier neural networks. In: Proceedings of the fourteenth international conference on artificial intelligence and statistics, pp. 315–323.
- Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y., 2016. Deep Learning. MIT Press Cambridge.
- Hamwood, J., Alonso-Caneiro, D., Read, S.A., Vincent, S.J., Collins, M.J., 2018. Effect of patch size and network architecture on a convolutional neural network approach for automatic segmentation of OCT retinal layers. Biomed. Opt. Express 9, 3049–3066.
- Han-qiu, X., Ben-qing, C., 2004. Remote sensing of the urban heat island and its changes in Xiamen City of SE China. J. Environ. Sci. 16, 276–281.
- Huang, C., Li, Y., Change Loy, C., Tang, X., 2016. Learning deep representation for imbalanced classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5375–5384.
- Jeatrakul, P., Wong, K.W., Fung, C.C., 2010. Classification of imbalanced data by combining the complementary neural network and SMOTE algorithm. In: International Conference on Neural Information Processing. Springer, pp. 152–159.
- Kaloustian, N., Bechtel, B., 2016. Local climatic zoning and urban heat island in Beirut. Procedia Eng. 169, 216–223.
- Khoshgoftaar, T.M., Golawala, M., Van Hulse, J., 2007. An empirical study of learning from imbalanced data using random forest. Tools with Artificial Intelligence, 2007. ICTAI 2007. In: 19th IEEE International Conference on. IEEE, pp. 310–317.

- Kim, M., Lee, J., Han, D., Shin, M., Im, J., Lee, J., Quackenbush, L.J., Gu, Z., 2018a. Convolutional neural network-based land cover classification using 2-D spectral reflectance curve graphs with multitemporal satellite imagery. *IEEE J-Stars* 11, 4604–4617.
- Kim, M., Lee, J., Im, J., 2018b. Deep learning-based monitoring of overshooting cloud tops from geostationary satellite data. *Gisci. Remote Sens.* 55, 763–792.
- Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
- Koppel, K., Zalite, K., Voormansik, K., Jagdhuber, T., 2017. Sensitivity of Sentinel-1 backscatter to characteristics of buildings. *Int. J. Remote Sens.* 38, 6298–6318.
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105.
- Kursa, M.B., 2014. Robustness of Random Forest-based gene selection methods. *BMC Bioinf.* 15, 8.
- Lauwaet, D., Hooyberghs, H., Maiheu, B., Lefebvre, W., Driesen, G., Van Looy, S., De Ridder, K., 2015. Detailed Urban Heat Island projections for cities worldwide: dynamical downscaling CMIP5 global climate models. *Climate* 3, 391–415.
- Lebedev, A., Westman, E., Van Westen, G., Kramberger, M., Lundervold, A., Aarsland, D., Soininen, H., Kłoszewska, I., Mecocci, P., Tsolaki, M., 2014. Random Forest ensembles for detection and prediction of Alzheimer's disease with a good between-cohort robustness. *NeuroImage: Clin.* 6, 115–125.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521, 436–444.
- Lee, J., Im, J., Kim, K., Quackenbush, L.J., 2018. Machine learning approaches for estimating forest stand height using plot-based observations and airborne LiDAR data. *Forests* 9, 268.
- Li, M., Im, J., Beier, C., 2013. Machine learning approaches for forest classification and change analysis using multi-temporal Landsat TM images over Huntington Wildlife Forest. *Gisci. Remote Sens.* 50, 361–384.
- Liu, M., Wang, M., Wang, J., Li, D., 2013. Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: application to the recognition of orange beverage and Chinese vinegar. *Sens. Actuat. B* 177, 970–980.
- Liu, T., Abd-Elrahman, A., Morton, J., Wilhelm, V.L., 2018a. Comparing fully convolutional networks, random forest, support vector machine, and patch-based deep convolutional neural networks for object-based wetland mapping using images from small unmanned aircraft system. *Gisci. Remote Sens.* 55, 243–264.
- Liu, Y., Fang, X., Xu, Y., Zhang, S., Luan, Q., 2018b. Assessment of surface urban heat island across China's three main urban agglomerations. *Theor. Appl. Climatol.* 133, 473–488.
- Marcos, D., Volpi, M., Kellenberger, B., Tuia, D., 2018. Land cover mapping at very high resolution with rotation equivariant CNNs: Towards small yet accurate models. *ISPRS J. Photogramm.* 145, 96–107.
- Mathew, A., Khandelwal, S., Kaul, N., 2018. Investigating spatio-temporal surface urban heat island growth over Jaipur city using geospatial techniques. *Sustain. Cities Soc.* 40, 484–500.
- Min, S., Lee, B., Yoon, S., 2017. Deep learning in bioinformatics. *Briefings Bioinf.* 18, 851–869.
- Mohammadianesh, F., Salehi, B., Mahdianpari, M., Gill, E., Molinier, M., 2019. A new fully convolutional neural network for semantic segmentation of polarimetric SAR imagery in complex land cover ecosystem. *ISPRS J. Photogramm. Remote Sens.* 151, 223–236.
- Paoletti, M., Haut, J., Plaza, J., Plaza, A., 2018. A new deep convolutional neural network for fast hyperspectral image classification. *ISPRS J. Photogramm.* 145, 120–147.
- Park, S., Im, J., Park, S., Yoo, C., Han, H., Rhee, J., 2018. Classification and mapping of paddy rice by combining landsat and SAR time series data. *Remote Sens.-Basel* 10, 447.
- Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Koppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.* 11, 1633–1644.
- Qiu, C., Mou, L., Schmitt, M., Zhu, X.X., 2019. Local climate zone-based urban land cover classification from multi-seasonal Sentinel-2 images with a recurrent residual network. *ISPRS J. Photogramm. Remote Sens.* 154, 151–162.
- Qiu, C., Schmitt, M., Mou, L., Ghamisi, P., Zhu, X., 2018. Feature importance analysis for local climate zone classification using a residual convolutional neural network with multi-source datasets. *Remote Sens.-Basel* 10, 1572.
- Rahaman, K.R., Hassan, Q.K., Ahmed, M.R., 2017. Pan-sharpening of Landsat-8 images and its application in calculating vegetation greenness and canopy water contents. *ISPRS Int. J. Geo-Inf.* 6, 168.
- Richardson, H.J., Hill, D.J., Denesiuk, D.R., Fraser, L.H., 2017. A comparison of geographic datasets and field measurements to model soil carbon using random forests and stepwise regressions (British Columbia, Canada). *Gisci. Remote Sens.* 54, 573–591.
- Rizwan, A.M., Dennis, L.Y., Chunho, L., 2008. A review on the generation, determination and mitigation of Urban Heat Island. *J. Environ. Sci.* 20, 120–128.
- Rumerhar, D., 1986. Learning representation by back-propagating errors. *Nature* 323, 533–536.
- Salata, F., Golasi, I., Petitti, D., de Lieto Vollaro, E., Coppi, M., de Lieto Vollaro, A., 2017. Relating microclimate, human thermal comfort and health during heat waves: an analysis of heat island mitigation strategies through a case study in an urban outdoor environment. *Sustain. Cities Soc.* 30, 79–96.
- Schmidhuber, J., 2015. Deep learning in neural networks: an overview. *Neural networks* 61, 85–117.
- Sim, S., Im, J., Park, S., Park, H., Ahn, M.H., Chan, P.W., 2018. Icing detection over East Asia from geostationary satellite data using machine learning approaches. *Remote Sens.-Basel* 10, 631.
- Sokolova, M., Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. *Inform. Process. Manage.* 45, 427–437.
- Soltan, H., Saon, G., Sainath, T.N., 2014. Joint training of convolutional and non-convolutional neural networks. *ICASSP 5572–5576.*
- Stewart, I.D., Oke, T.R., 2012. Local climate zones for urban temperature studies. *Bull. Am. Meteorol. Soc.* 93, 1879–1900.
- Sukhanov, S., Tankoyeu, I., Louradour, J., Heremans, R., Trofimova, D., Debes, C., 2017. Multilevel ensembling for local climate zones classification. In: 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, pp. 1201–1204.
- Tuia, D., Moser, G., Le Saux, B., Bechtel, B., See, L., 2017. 2017 IEEE GRSS data fusion contest: open data for global multimodal land use classification [Technical Committees]. *IEEE Geosci. Remote Sens. Mag.* 5, 70–73.
- Vedaldi, A., Lenc, K., 2015. Matconvnet: convolutional neural networks for matlab. In: Proceedings of the 23rd ACM International Conference on Multimedia. ACM, pp. 689–692.
- Verdonck, M.-L., Okujeni, A., van der Linden, S., Demuzere, M., De Wulf, R., Van Coillie, F., 2017. Influence of neighbourhood information on 'local climate zone'mapping in heterogeneous cities. *Int. J. Appl. Earth Obs. Geoinf.* 62, 102–113.
- Volpi, M., Tuia, D., 2016. Dense semantic labeling of subdecimeter resolution images with convolutional neural networks. *IEEE Trans. Geosci. Remote Sens.* 55, 881–893.
- Wang, C., Middel, A., Myint, S.W., Kaplan, S., Brazel, A.J., Lukasczyk, J., 2018. Assessing local climate zones in arid cities: the case of Phoenix, Arizona and Las Vegas, Nevada. *ISPRS J. Photogramm.* 141, 59–71.
- Wang, T., Wu, D.J., Coates, A., Ng, A.Y., 2012. End-to-end text recognition with convolutional neural networks. In: Pattern Recognition (ICPR), 2012 21st International Conference on. IEEE, pp. 3304–3308.
- Wurm, M., Stark, T., Zhu, X.X., Weigand, M., Taubenböck, H., 2019. Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. *ISPRS J. Photogramm. Remote Sens.* 150, 59–69.
- Xing, Y., Wang, M., Yang, S., Jiao, L., 2018. Pan-sharpening via deep metric learning. *ISPRS J. Photogramm.* 145, 165–183.
- Xu, Z., Guan, K., Casler, N., Peng, B., Wang, S., 2018. A 3D convolutional neural network method for land cover classification using LiDAR and multi-temporal Landsat imagery. *ISPRS J. Photogramm.* 144, 423–434.
- Yadav, N., Sharma, C., Peshin, S., Masiwal, R., 2017. Study of intra-city urban heat island intensity and its influence on atmospheric chemistry and energy consumption in Delhi. *Sustain. Cities Soc.* 32, 202–211.
- Yokoya, N., Ghamisi, P., Xia, J., Sukhanov, S., Heremans, R., Tankoyeu, I., Bechtel, B., Le Saux, B., Moser, G., Tuia, D., 2018. Open data for global multimodal land use classification: outcome of the 2017 IEEE GRSS Data Fusion Contest. *IEEE J-Stars* 11, 1363–1377.
- Yoo, C., Im, J., Park, S., Quackenbush, L.J., 2018. Estimation of daily maximum and minimum air temperatures in urban landscapes using MODIS time series satellite data. *ISPRS J. Photogramm.* 137, 149–162.
- Yu, X.R., Wu, X.M., Luo, C.B., Ren, P., 2017. Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework. *Gisci. Remote Sens.* 54, 741–758.
- Yue, K., Yang, L., Li, R., Hu, W., Zhang, F., Li, W., 2019. TreeUNet: Adaptive Tree convolutional neural networks for subdecimeter aerial image segmentation. *ISPRS J. Photogramm. Remote Sens.* 156, 1–13.
- Zhang, C., Pan, X., Li, H., Gardiner, A., Sargent, I., Hare, J., Atkinson, P.M., 2018. A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification. *ISPRS J. Photogramm.* 140, 133–144.
- Zhang, T., Tang, H., 2019. A Comprehensive Evaluation of Approaches for Built-Up Area Extraction from Landsat OLI Images Using Massive Samples. *Remote Sens.-Basel* 11, 2.
- Zhen, Z., Quackenbush, L.J., Stehman, S.V., Zhang, L., 2013. Impact of training and validation sample selection on classification accuracy and accuracy assessment when using reference polygons in object-based classification. *Int. J. Remote Sens.* 34, 6914–6930.
- Zhou, Y.-T., Chellappa, R., 1988. Computation of optical flow using a neural network. In: IEEE International Conference on Neural Networks, pp. 71–78.
- Zhou, Z.H., Liu, X.Y., 2006. Training cost-sensitive neural networks with methods addressing the class imbalance problem. *IEEE Trans. Knowl. Data Eng.* 18, 63–77.
- Ziaul, S., Pal, S., 2018. Analyzing control of respiratory particulate matter on Land Surface Temperature in local climatic zones of English Bazar Municipality and Surroundings. *Urban Clim.* 24, 34–50.