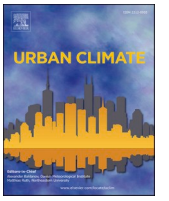
Urban Climate 52 (2023) 101731

Contents lists available at ScienceDirect

Urban Climate

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



Application of local climate zones combined with machine learning to predict the impact of urban structure patterns on thermal environment

Tsz-Kin Lau [a], Yu-Cheng Chen [b,*], Tzu-Ping Lin [a]

[a] Department of Architecture, National Cheng Kung University, 1 University Rd., East Dist., Tainan 701, Taiwan
[b] Department of Architecture, Nanhua University, 55, Sec. 1, Nanhua Rd., Dalin Township, Chiayi County 62249, Taiwan

ARTICLE INFO

Keywords:
Urban heat island
Convolutional neural network
Computer vision
Local climate zone
Urban structure combination

ABSTRACT

The urban heat island effect is a pertinent concern and estimating the thermal stress in cities is a crucial research goal. Most relevant studies have focused on the relationship between urban structures and high air temperatures, ignoring the effect of different combination of urban structures on high air temperature identifications. In this study, an efficient method was developed for identifying the thermal risk based on combination of urban structures to explore the relationship between urban environment and meteorological conditions. After the preprocessing of LCZ and meteorological data, data were input into the convolutional neural network (CNN) model constructed in this study. After the training process, the accuracy rate of the proposed model increased as the sampling rate increased, and the model achieved a high accuracy rate of 81.97% at 50% sampling. Furthermore, the model performed particularly well at identifying low and medium level of thermal risk. This means that the model could estimate the classes of thermal risk in most areas in Taipei City. The feasibility of using the proposed CNN model to identify the level of thermal risk was proven; the model is a convenient tool for estimating thermal stress in different areas with incomplete air temperature data.

1. Introduction

Taiwan has undergone rapid development, causing substantial areas of natural land to be replaced by infrastructure such as buildings and roads. High urbanization caused serious urban heat island (UHI) effects in Taiwan, especially in Taipei City because of its terrain condition. Different types of land cover and urban structures substantially affect the meteorological conditions in Taiwan (Zhang et al., 2020). Air temperatures in regions with a large area of impermeable surfaces are usually higher than those elsewhere (Chen et al., 2018). The relationship between air temperature and land cover type has been confirmed; built-up areas experience higher air temperatures than do areas with more natural land cover (Beck et al., 2018). Many shortcomings caused by high air temperature like worse thermal comfort, higher level of morbidity and mortality and other serious health problem (Matzarakis and Amelung, 2008; Singh et al., 2020; Kim et al., 2012). Therefore, method for estimating thermal risk in different areas is necessary.

According to the importance of land cover type to air temperature, large-scale observation of land surface type is needed. Satellite imagery is a convenient approach for observing land surface status. Compared with other satellite imaging methods such as the

* Corresponding author.
E-mail address: leo210801@nhu.edu.tw (Y.-C. Chen).

https://doi.org/10.1016/j.uclim.2023.101731
Received 8 January 2023; Received in revised form 6 June 2023; Accepted 22 October 2023

Moderate Resolution Imaging Spectroradiometer (MODIS) and Indian Remote-Sensing Satellite (IRS)-P6, Landsat has a higher spatial resolution (30 m). Furthermore, Landsat images of different historical periods are free to use, unlike images from commercial satellites such as SPOT (from French: Satellite pour l'Observation de la Terre) and IKONOS (Duro et al., 2012). Landsat imagery has been widely applied in studies involving land use/land cover classification. Landsat images have been used for investigating how land surface characteristics, such as land surface temperature (LST) and land use/land cover, changed over two decades (Mukherjee and Singh, 2020). Moreover, Landsat imagery has also been widely used for classifying local climate zones (LCZs) which is a scheme of classifying urban areas related to climate and characteristics of the land surface.

LCZ was widely applicated in the studies of climate and weather modeling and urban heat island (Stewart, 2011; Stewart and Oke, 2012). LCZ was used for evaluating the relationship between subjective perceptions and the thermal environment. The result of the LCZ classification provided the features of the physical and thermal environment for the study (Lau et al., 2019). Moreover, the relationship between land coverage type, surface structure, and local climate was also explored based on LCZ classification (Zhang et al., 2020). Satellite imagery and LCZ classification made it easier and more convenient to observe the characteristics of land surface and built-types. In the past, LCZ has been frequently used to investigate urban climate, especially in the context of the heat island effect (UHI) (Leconte et al., 2015; Chen et al., 2019; Zhou et al., 2020; Badaro-Saliba et al., 2021), because LCZ can specify the causes of urban warming through different urban forms. The related study includes the combination of LCZ and the Weather Research and Forecasting (WRF) to predict climate conditions in the urban canopy. Since LCZ can quantify various urban development factors, the urban canopy information corresponding to each LCZ type, such as building height, sky visibility, and material albedo and absorption rate, can be used to predict the microclimate by WRF (Molnár et al., 2019). The results show that LCZ can effectively evaluate the high temperature environment in urban areas and improve the microclimate differences due to the complex built environment, which cannot be presented by simply classifying cities into three categories based on development intensity (Chen et al., 2021). Given the benefits and achievements of LCZ, LCZ was used in this study to classify and perform the land coverage type.

Take the advancements in computer technology, machine learning became a great and convenient tool for data analysis. Neural Networks (NNs) is a common algorithm within machine learning and they have been applied in numerous research fields, such as object detection, machine translation, voice recognition, and autopilot system operation (Peng and Chen, 2019). A well-trained NN can detect and classify a target quickly and accurately. NNs operate in a manner similar to the human optical system and can recognize an object immediately after an image is received. Convolutional neural networks (CNNs) are commonly used for image classification or object recognition (Redmon et al., 2016). Furthermore, NNs have been applied to predict changes in land cover for modeling surface UHI intensity and predicting the intensity of the UHI effect (Nadizadeh Shorabeh et al., 2020; Gobakis et al., 2011). Due to the achievements of NNs in urban studies, CNN was used in this study to analyze the pattern of urban structure.

Although LCZ and NNs were widely applied in urban studies, there are still some problems that should be discussed and solved. Many studies presented the relation of air temperature to land coverage types and LCZ classes. But most of the previous studies performed the relationship between air temperature and a single type of land coverage or LCZ class. Due to the cooling effect from a green space can influence the air temperature of an environment up to 400 m away (Doick et al., 2014), the pattern of urban structure within a mesoscale area should be considered, rather than only the proportion of vegetation coverage or built-up areas. Moreover, most

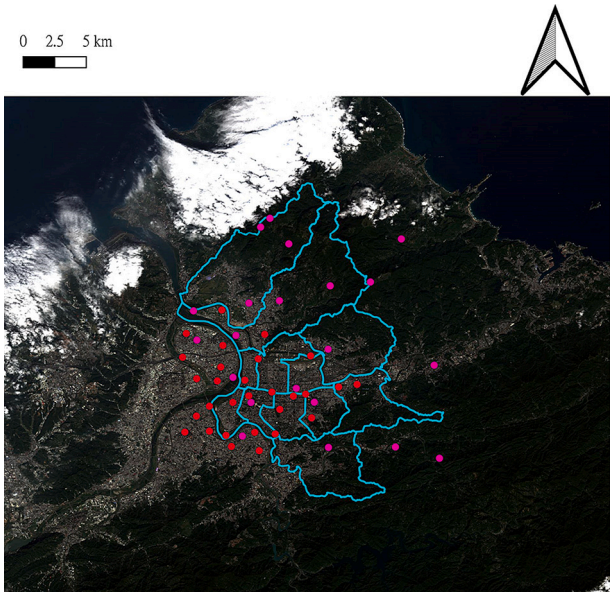


Fig. 1. Research area in this work: Map of Taipei City. (Data source: base-map from software ArcGIS) The Blue lines are the borders of districts in Taipei City. In addtion, measurement points in the environs of Taipei City selected for investigation were also presented. Red: HiSAN measurement points. Purple: CWB measurement points. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of the previous studies mapped air temperature with interpolation and performed fine. But the dense distribution of measurement points is required, otherwise, the smooth distribution produced by interpolation is difficult to describe the air temperature distribution in detail.

Therefore, a method to estimate the thermal risk based on the land coverage classification was performed in this article. The level of thermal risk was identified based on the statistics from the meteorological dataset to present the distribution of thermal risk within cities. CNN was used in this study to identify the thermal risk based on the pattern of LCZ. Through the process of convolution and pooling, the patterns of the urban structure were processed and analyzed. Finally, the map of the thermal risk distribution of cities was produced in detail by the well-trained CNN model to estimate the thermal risk in each area. This article provides an easier, more economical, and more convenient estimation of thermal risk in each area, and has lower requirements for measuring equipment.

2. Materials and methods

2.1. Study area

Taipei City in northern Taiwan was chosen as the research area. Geographically, Taipei City consists of the Taipei Basin (southwest), the Tatun volcanoes (northern), and the western foothills (southeast; Teng et al., 2001). The total area of Taipei City is approximately 271.8 km^2, and the population of Taipei was approximately 2.65 million in 2019. In most areas of Taipei City, buildings are densely distributed and rise 3–9 stories. Most high-rise buildings above 10 stories are located in prosperous areas such as Wanhua District and Xinyi District. Northern Taiwan has a humid subtropical climate with an average annual temperature of approximately 23 °C (Central Weather Bureau [CWB], 2011).

The UHI effect is substantial in northern Taiwan, with the Taipei Basin exacerbating this effect (Lin et al., 2008). From the terrain reclassification data of Iwahashi et al. (2018), they identified 15 terrain groups, 7 of which relate to Taipei City's terrain, namely rough steep mountains, smooth steep mountains, rough moderate mountains, smooth moderate mountains, hills that are rough in small and large scale, hills that are smooth in small scale and rough in large scale, and dissected terraces and moderate plateaus. Most of the urban terrain in Taipei City is a basin, and its topography has a great influence on UHI in Taipei City. The research area was shown in Fig. 1.

2.2. Classification of urban structure combination

The descriptions and definitions of various artificial and natural environments vary by country. Therefore, to establish a city classification method that different users can adapt, the systematic development of standard guidelines for climatic and environmental factors is necessary.

Stewart and Oke (2012) proposed the first LCZ system. They quantified and validated relevant information, including sky visibility, building height, street aspect ratio, projected building area, impervious ratio, average building height, and surface roughness. This classification method is termed LCZx, where x is a number from 1 to 10 for built-up areas, and letters ranging from A to G are used for classifying non-built-up areas (Stewart and Oke, 2012). LCZ classification can be used in combination with Landsat images as a convenient method of estimating land cover types. This method is free to use and can be applied in studies involving different time points and software versions.

2.2.1. Satellite image use

The Landsat 8 images used in this study were collected by the US Geological Survey (USGS), and these images were captured on May 4, 2015. Landsat 8 passed over Taiwan at 10:20 a.m. with an interval of 16 days, and it provided regular land surface data, including LST, during the daytime. Multispectral and thermal data from Landsat 8 were used for LCZ classification in the present study. Data from eight spectral bands were used, namely bands 1–7 and band 10. Thermal band images were resampled from a spatial resolution of 100 m to one of 30 m.

2.2.2. LCZ training

The LCZ classification process was based on that of the World Cities Database and Access Portal Tool (WUDAPT). The WUDAPT system uses Landsat 8 images as the primary reference layer, and Google Earth images and supervised classification methods were applied to artificially construct the demonstration area. In this study, at least five training samples are collected for LCZ sampling. If an area does not have an LCZ, the classification can be omitted. After training of the LCZ model is conducted, SagaGIS software performs and automates the final classification process. In this study, the resolution of each LCZ image was set to 100 m to prevent small-scale images and large-scale images from causing decision errors (Bechtel et al., 2015).

The accuracy of LCZ classification in 20 cities around the world has shown that the overall accuracy can be as high as 76% (Ren et al., 2019), while the LCZ trained in this study was evaluated at 79% by WUDAPT.

The classification process of Local Climate Zones (LCZ) is subject to variations due to the differences in urban perception among trainers, affecting the accuracy of the classification.

This study focused on the LCZ classification research in Taipei City. We collected 50 representative images of various urban typologies, including dense areas, high-rise buildings, and different levels of vegetated green spaces. These images comprised Google Earth satellite imagery and Google Maps street view photos. We conducted a questionnaire survey with 50 architecture students specializing in spatial planning and design to assess their perception of the representative areas' LCZ.

Subsequently, we used the most selected LCZ type for each representative area as a reference for classification during training. By employing this method in the LCZ training process, we enhanced the objective classification process, resulting in approximately a 6% improvement in accuracy compared to solely relying on individual subjective classification. This approach enables the utilization of primary LCZ data with higher correctness and applicability.

2.3. Microclimate data collection

Air temperature data were collected from the high-density street-level air temperature observation network (HiSAN) in Taipei city. To investigate the thermal risk within Taipei City, information on high air temperature should be collected and analyzed. Due to the statistics of meteorological datasets, most of the highest daily air temperatures appeared at 1 p.m. and it affected the thermal environment seriously. Therefore, the average daily air temperature at 1 p.m. in August 2020 was selected and analyzed in subsequent analysis. Estimating the level of thermal risk in various areas is the main purpose of this article and the data on high air temperature is necessary. Through the analysis of patterns grouped by LCZ classes and high air temperature, the relationship between urban structure combination and thermal risk could be identified and estimated.

2.3.1. HiSAN

In HiSAN, many measurement points are distributed in different districts for examining the relationship between urban development patterns and thermal conditions. The installation of HiSAN sensors has shown at Fig. 2. HiSAN sensors (LOGPRO TR-32 Tecpel Co., Ltd.) are energy efficient and lightweight, and they can store large amounts of data. Before installation, the air temperature accuracy of these sensors was confirmed (Chen et al., 2018).

LOGPRO TR-32 sensors have a resolution of 0.1 °C and an accuracy of ±0.5 °C for estimating air temperature, and they have a resolution of 1% and an accuracy of ±5% for estimating relative humidity. Radiation covers were used to avoid the influence of radiation and to prevent sensor damage by rain and wind.

2.3.2. CWB

Taiwan's CWB began developing a global numerical weather forecast system in the early 1980s. It provides meteorological information such as air temperature, humidity, and wind speed (Liou et al., 1997). CWB provides various kinds of meteorological data for public in profit or non-profit uses, such as air temperature, relative humidity, and wind speed. For air temperature, ±60 °C can be observed using the sensor in CWB, and it has the resolution of 0.1 K and accuracy of ±0.1 K. In total, 52 measurement points were selected for assessing the air temperature of Taipei City; the related measurement data were provided by HiSAN and CWB. The distribution of measurement points has shown at Fig. 1.

Fig. 2. Equipment installed at a measurement point. (A) LOGPRO instrument. (B) Radiation shield. (C) LOGPRO instrument and radiation cover at a measurement point.

2.3.3. Temperature distribution

From the statistics of the micro-climate data, most of the highest air temperatures appeared at 1 p.m. To investigate the relationship between high air temperature and urban structure combination better, the average air temperature at 1 p.m. at the 52 measurement points in August 2020 in Taipei City was calculated using HiSAN and CWB data. Subsequently, inverse distance weighted (IDW) interpolation was used for calculating the air temperature distribution in the whole of Taipei City because the measurement points could not cover all of Taipei City. IDW interpolation is widely used in spatial analysis. In IDW interpolation, cell values are calculated using a linear-weighted combination set of sample points (Nistor et al., 2020). All IDW interpolation calculations were conducted using QGIS 3.20.2 software, and the interpolated maps with resolution of 50 m were generated.

2.3.4. Definition of thermal condition classes

After the calculating average air temperature at 1 pm in August 2020 in Taipei. To determine the different severe levels of thermal risk, three classes were defined: Low, Medium, and High levels of thermal risk. The areas with average air temperature lower than the upper quartile were defined as the class of Low level of thermal risk. The upper quartile represented the thermal condition in most areas in Taipei without bias from the extremum. The areas with average air temperature higher than the upper quartile were defined as the class of medium level of thermal risk which means the areas have the heating problem and they should be concerned. The areas with average air temperature higher than the top 90% of average air temperature distribution were defined as the class of High level of thermal risk.

2.4. CNN analysis

A general CNN has three parts: an input layer, hidden layers, and an output layer. The input layer receives images with a fixed size. Hidden layers are convolutional layers and pooling layers. The convolutional layer extracts feature from images by sliding a kernel with a fixed size on an image and outputs a feature map into the pooling layer. Subsequently, the pooling layer reduces the amount of data from the previous feature map through max pooling and outputs a smaller feature map into the next layer. The flattened layer inside the output layer then accepts the feature map from the last pooling layer and flattens the feature map into a one-dimensional array. Finally, the array is output into a fully connected layer inside the output layer, and the training or prediction process commences (LeCun et al., 1998).

The CNN model constructed in this study had one input layer, five hidden layers, and one output layer. The input layer was set to

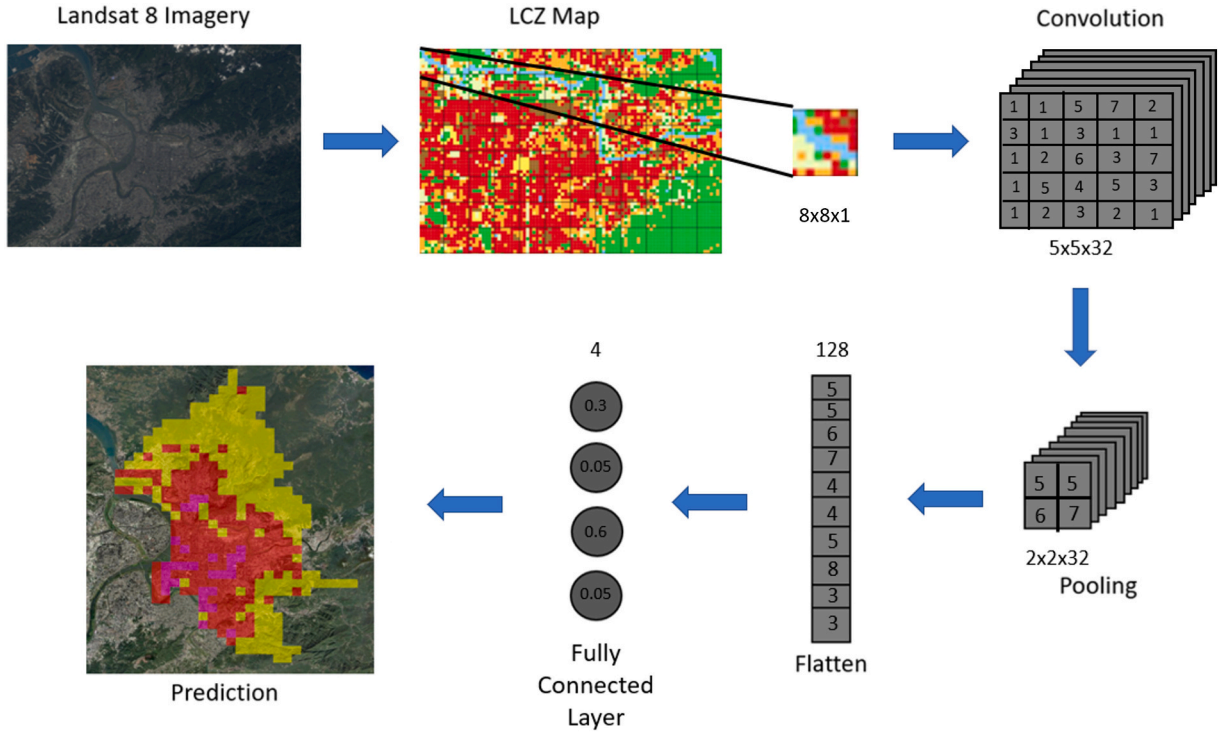


Fig. 3. Structure of the CNN model for estimating the level of thermal risk in areas with different pattern of LCZs. The images from the LCZ map were cropped to a grid size of 800 m and entered into the CNN model for training. Image features were extracted after the convolution processes, and the feature maps from the convolutional layers were sent into the pooling layer for dimensionality reduction. The feature maps from the pooling layers were input into the flatten layer for flattening. Finally, after flattening, the feature maps were sent to the fully connected layer for training and prediction.

have a fixed size for inputting images (8 × 8 × 1); Each input image contained 64 pixels. Each hidden layer comprised a convolutional layers and pooling layers. The activation function and padding methods used in the convolutional layers were ReLU and zero padding, respectively. ReLU, which can improve a trained model's generalization performance, is calculated as follows (Ide and Kurita, 2017):

$$ReLU(x) = max(0, x)$$

Zero padding considerably reduce the model training time but does not affect the model's accuracy or the size of the feature map generated; the equation for zero padding is as follows (Hashemi, 2019):

$$\mathrm{F}\left\{\frac{\left[\left(Img_{size} + 2 \times \left(pad_{size}\right) - Ker_{size}\right]}{Str} + 1\right\} = \mathrm{output}_{\mathrm{size}}$$

where Imgsize = input feature size (8 × 8)
Padsize = padding size (zero padding = 1)
Kersize = kernel size (4 × 4)
Str = stride(s)
Outputsize = size of the output feature map (8 × 8)

The kernels in the convolutional layers were set to have a size of 4 × 4. The pooling method selected was MaxPooling, and the filter size for pooling was set as 2 × 2. The output layer had a flatten layer and a fully connected layer. The flatten layer flattened the feature maps from the last hidden layer (multidimensional array) into a one-dimensional array. After the flattening process, arrays were sent into the fully connected layer through the softmax activation function for training and prediction.

Finally, the model built in this study was trained for 500 epochs, and the learning rate was set as 0.0001. The optimizer and loss function were set as adaptive moment estimation (Adam) and binary cross entropy, respectively. Adam is a common algorithm for training NNs and it has been used in many machine learning frameworks (Bock and Weiß, 2019). The model structure is presented in Fig. 3. The inputs of the model were the LCZ classification results and air temperature data from HiSAN and the CWB.

3. Results

3.1. LCZ classification map

The LCZ map of Taipei City was generated and presented at Fig. 4. LCZA comprises LCZs at the north and south sides of Taipei City and well-developed corridors running through the city from east to west. The main built-up areas are LCZ2, LCZ6, and LCZ9. The LCZs with the most natural land cover are LCZA and LCZG.

Classified images revealed urban development trends in Taipei City. Built-up areas are concentrated in the central and western

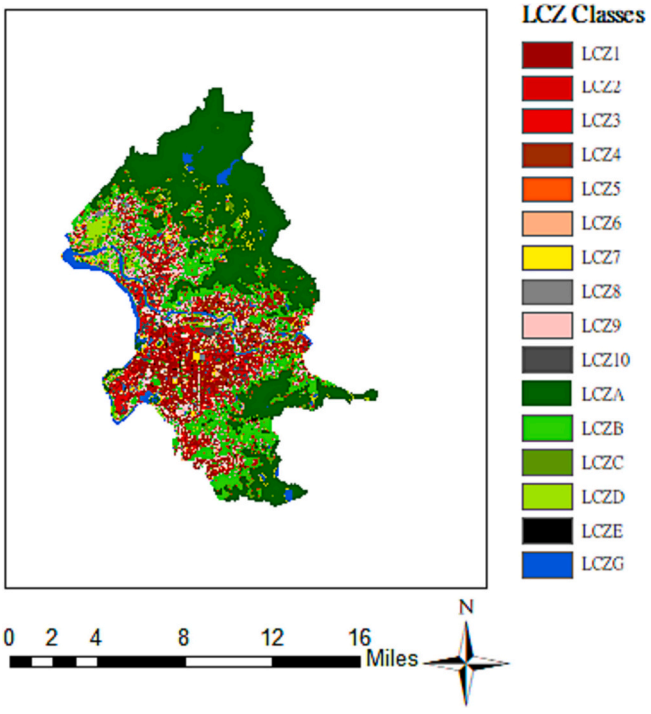


Fig. 4. LCZ map of Taipei City.

regions, whereas the northeastern and southeastern parts are primarily mountainous and suburban. The built environment near the mountainous parts is mostly an open built-up area. By contrast, the linear east–west axis area with concentrated development is primarily a densely built-up area. In general, cities have few large green spaces. Most such green spaces are scattered small parks. Rivers constitute the main water bodies in Taipei.

3.2. Thermal condition distribution

The aforementioned 52 measurement points were subjected to IDW interpolation to construct a complete air temperature map of Taipei City, and the results are presented in Fig. 5. The average air temperature during August 2020 in Taipei ranged from 24.66 °C to 33.79 °C. The highest and lowest air temperatures were noted at Shilin Night Market and the northern mountains of Taipei, respectively. Air temperatures over 33.5 °C were noted in prosperous districts such as Xinyi District and Zhongzheng District. Most of the green areas inside the city had an average air temperature of approximately 32.5 °C, which was at least 1 °C lower than the highest air temperature recorded in prosperous districts.

The air temperature distributions in each LCZ suggest that the air temperature of LCZ1 was higher than that of similarly built-up areas because the compact building barriers in LCZ1 lead to poor ventilation and increase anthropogenic heat storage. High air temperature blocks are located in LCZ1–3 and LCZ8. The high air temperature of LCZ1–3 was caused by high building density. LCZ8 and LCZ10 are industrial blocks and hotspots. Additionally, LCZA had the lowest air temperature of all areas. The area with the second lowest air temperature was LCZB-D, which has a natural environment.

The IDW interpolation results were resampled from a spatial resolution of 100 m to one of 800 m. After resampling, the average air temperature range was 25.31 °C to 33.71 °C.

Then, the interpolated map was used to define areas with three classes of thermal risk. The areas with average air temperature lower than the upper quartile were defined as the class of Low level of thermal risk. The upper quartile represented the thermal condition in most areas in Taipei without the bias from the extremum. The areas with average air temperature higher than the upper quartile were defined as the class of medium level of thermal risk which means the areas have the heating problem and they should be concerned. The areas with average air temperature higher than the top 90% of average air temperature distribution were defined as the class of high level of thermal risk. According to previous research, the heat discomfort threshold in Taiwan is about 33.5 °C (Kántor et al., 2014). The areas with average air temperature approaching the heat discomfort threshold have serious heat problem that need urgent improvement.

After the defining work of different levels of thermal risk, three classes were identified and assigned. The first class (Low level) is ranging from 25.31 °C to 32.29 °C which was acceptable in Taiwan. The second class (Medium level) is ranging from 32.30 °C to 32.91 °C which is related to relatively high air temperatures for Taipei City. Although it doesn't have very serious thermal risk, more consideration is needed in the future. The Third class (High level) is ranging from 32.92 °C to 33.71 °C which is close to the discomfort threshold in Taiwan, and it needs urgent improvement.

The classes were defined based on the statistics of average air temperature distribution. Although the values of the three classes are close, areas with different levels of thermal risk were still identified fine which is the main purpose of this work. While the lack of measuring points, IDW interpolation performed the distribution of thermal risk fine and it can be the reference for CNN to learn. The first, second, and third classes had 208, 165, and 43 areas, respectively (total: 416 areas). The images of areas including patterns of LCZs in these categories were used for subsequent analysis.

3.3. CNN results

After training, the proposed CNN model could be used to estimate the level of thermal risk in each area in Taipei. The model was trained for 500 epochs, and the learning rate was 0.0001. At 50% sampling, the best overall accuracy of the CNN model was 81.97%. The accuracy levels of the CNN model are presented in Table. 1.

In this study, sampling rates from 5% to 50% were tested, and the results were compared. Table 1 reveals that the overall accuracy increased with sampling percentage; the accuracy rate increased from 64.90% to 81.97%. More samples provided more features for the

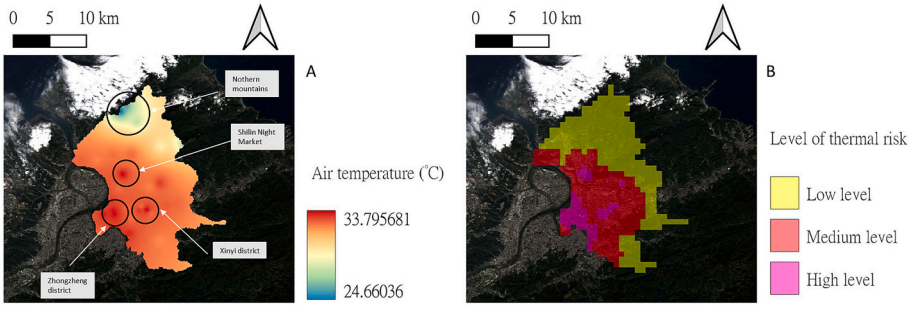


Fig. 5. Thermal distribution map of Taipei City. (A) Map constructed through IDW interpolation involving measurement points. (B) Map depicting thermal risk level based on Map A.

Table 1
Accuracy rates the CNN model at different case sampling percentages.

ACC\sample	5%	10%	20%	30%	40%	50%
OA	64.90%	70.67%	73.32%	73.80%	74.52%	81.97%
A1	79.81%	84.62%	89.90%	78.37%	82.21%	89.42%
A2	52.73%	66.67%	58.18%	79.39%	71.52%	80%
A3	39.53%	18.60%	51.16%	30.32%	48.84%	53.49%

OA: Overall accuracy. A1: Accuracy rate under the first category (low level of thermal risk). A2: Accuracy rate under the second category (medium level of thermal risk). A3: Accuracy rate under the third category (high level of thermal risk).

model to learn and helped it determine the characteristics of urban structures pattern that may cause high air temperatures. Table 1 also presents the accuracy rate of the model for each sampling category.

In each sampling situation, the model had high accuracy in the first category because the first category had the most samples and its characteristics were relatively uncomplicated. Moreover, for the second category, the accuracy of the model increased with the sampling percentage. Because random sampling was performed, the performance levels of the model identification were slightly different.

However, random sampling removes researcher bias, and related models are expected to output the best solutions to problems using the relevant sampling method (Kesemen et al., 2021). The model performance in the third category was lower than that in the first and second categories. Most of the patterns of the urban structure in the third category were generated with many LCZ1, LCZ2, LCZ3, LCZ4, LCZ6, and LCZ10. The lowest accuracy rate was achieved in each sampling scenario because the third category had the lowest number of samples for sampling. Additionally, the pattern of the urban structure in the third category was complicated and similar to those in the second category. The lack of training samples and low separability with the second class caused the relatively poor performance in identifying the area with a high level of thermal risk (Third category).

As previously indicated, the model with 50% sampling was selected for further investigations because it had the best overall accuracy and achieved the best identifications in all three categories. The relevant results are presented in Fig. 6. As indicated in this figure, identifications of the distribution of thermal risk made by the CNN model in this work were similar to the distribution determined through IDW interpolation.

The aforementioned CNN model with 50% sampling and well-training could classify most areas with a low level of thermal risk, such as suburbs and mountainous areas; the accuracy rate was 89.42%. The model also reached an accuracy rate of 80% for medium thermal risk level identifications related to the second category. For the second category, the thermal risk distribution identified by the model was similar to the air temperature map produced through IDW interpolation; areas with a medium level of thermal risk were concentrated in southwest Taipei City.

The model made some misjudgments; for example, some suburbs and mountainous areas were assessed as having a medium level of thermal risk. Finally, the model accuracy rate for identifying the high level of thermal risk in the third category was only 53.49%. Although the model had poor performance at detecting areas with a high level of thermal risk for the third category, the distribution of a high level of thermal risk identified by the model was still similar to the distribution produced through IDW interpolation, meaning the model could still detect areas with a high level of thermal risk.

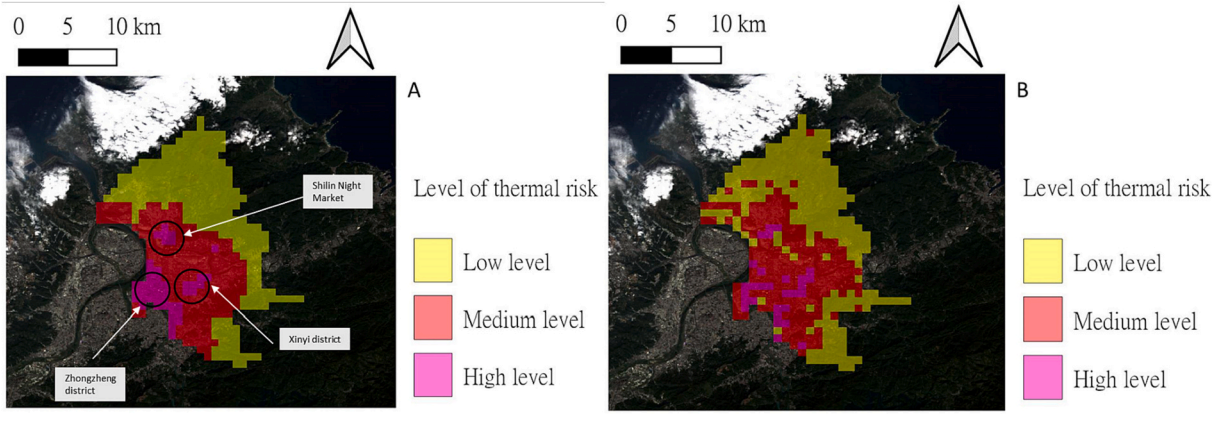


Fig. 6. Map showing areas with low, medium, and high level of thermal risk, as identified by the CNN model under 50% sampling. (A) Map of areas with low, medium, and high level of thermal risk in Taipei City. (B) Map showing areas with low, medium, and high level of thermal risk in Taipei City, as identified by the trained CNN model.

4. Discussion

4.1. Subjective cognition of urban structure combination

The subjective cognition approach in classifying Local Climate Zones (LCZ) is associated with several limitations and shortages. Firstly, due to the reliance on individual perceptions and interpretations, subjective cognition introduces a significant level of subjectivity into the classification process. This subjectivity can lead to inconsistencies and variations in the classification results, as individuals may have different understandings and interpretations of urban typologies.

Moreover, subjective cognition is prone to biases and errors. Trainers or individuals involved in the classification may have personal preferences or preconceived notions about certain urban features, which can influence their categorization decisions. These biases can introduce inaccuracies and distort the accurate representation of LCZs.

Additionally, subjective cognition may be limited by the trainers' or individuals' cognitive capacity and expertise. The accuracy and reliability of the classification heavily depend on their knowledge and understanding of urban morphology and climatic factors. Inexperienced or uninformed individuals may need to possess the necessary expertise to categorize LCZs, leading to erroneous classifications accurately.

Objective and standardized approaches, such as remote sensing data, satellite imagery, and machine learning algorithms, have been increasingly employed in LCZ classification to address these shortages. These methods provide more consistent and reproducible results, reducing subjectivity and improving the accuracy of the classification process.

4.2. CNN model performance

As indicated previously, the highest accuracy of the model was 81.97%, and this was achieved at 50% sampling. The model still required improvement, but a model accuracy rate of 70% is deemed acceptable (Liu et al., 2007). The accuracy rate for the first category (low level of thermal risk) was the highest at 89.42%. The model had accuracy rates of 80% and 53.49% for the second (medium level of thermal risk) and third (high level of thermal risk) categories, respectively. The classification results are presented in Table 2.

The confusion matrix in Table 2 shows the model's misjudgments. The second category, with 22 areas, was identified by the model to be the first category. The first category, with 27 areas, was identified to be the second category by the model. In addition to the low separability of these areas, the normal distribution of IDW interpolation–derived data affected the classification results. Despite the limitations related to measurement point establishment, IDW interpolation provided useful data for subsequent analysis. The model's accuracy was the lowest in the third category which was only 53.49%. A lack of samples caused this poor accuracy level. Additionally, the similarity between the second and third categories introduced confusion to the model training process.

In addition to the misjudgment of the CNN model, the prediction of the CNN model also showed the relationship between the LCZ combination and the thermal risk level. First, most areas with low thermal risk have LCZ combinations of LCZ A, B, and D, showing the cooling effect of green spaces. Then, areas with medium and high heat risk are close and similar. However, Fig. 6 shows that areas that include green spaces (LCZ A, B, and D) have a lower thermal risk than those that have only LCZs combined with LCZs 1, 2, and 3. According to the results shown in Fig. 6, the CNN model shows the importance of natural landscapes (LCZ A, B, C, D) in urban cooling, and that most of the high heat risk occurs in areas with dense and complex built environments, including LCZ 1, 2 and 3. Similar findings have been found in previous studies, but most of them considered the relationship between individual LCZs and air temperature (Shih et al., 2020). In this study, considering the combination of LCZs, the level of thermal risk may not be affected by only one LCZ type, indicating the importance of natural landscapes (LCZ A, B, C and D) in reducing thermal risk. The results also show that the combination of dense and complex built environments (LCZ 1, 2 and 3) needs improvement, which have higher thermal risk level.

4.3. Mapping advancement

Although the CNN model made some misjudgments, the results presented the thermal risk distribution fine, especially showing the cooling effects from green spaces. From Fig. 7, the thermal risk distribution made by the CNN model was presented. It has a similar distribution with IDW interpolation but is more detailed. Apart from the smooth distribution made by interpolation, CNN model identified the thermal risk in different areas based on LCZs' pattern which was not only the statistics of the proportion of LCZ classes. From the distribution made by interpolation, there are very serious levels of thermal risk in the whole of western Taipei City. But the CNN model performed a similar distribution but confirmed the contribution from green spaces like 228 Peace Park, Shuanyuan riverside Park and green space near Ming Chuan University.

Table 2
Confusion matrix of CNN model predictions under 50% sampling.

	L	M	H
L	186	27	4
M	22	132	16
H	0	6	23

L: Low level. M: Medium level. H: High level.

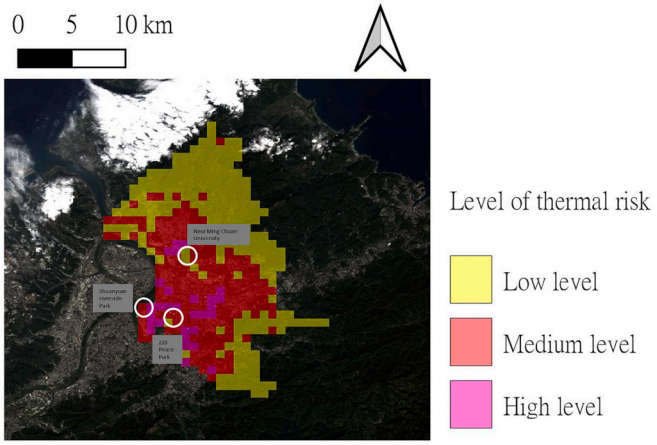


Fig. 7. Map of thermal risk distribution in Taipei City. Cooling effect from greens paces presented by the CNN model's estimation.

According to the green spaces, the thermal risk in these areas is reduced from High to Medium level. Similar phenomenon can be found in previous studies. Temperature distributions made by interpolation are common in urban studies because of the lack of measuring points. And regression and simulation were also applied in some previous studies for estimating the temperature distribution (Hsu et al., 2017; Taheri-Shahraiyni and Sodoudi, 2017). The temperature distributions generated by interpolation have smoother distribution and the values of each area are similar with neighborhood.

Although the temperature distributions made by NNs, non-linear regression and simulation were more complicated, they showed the strong relationship between temperature and urban's characteristics. As the comparison above, although interpolations are helpful and convenient in mapping temperature distribution, well-trained NNs and a well-sampled regression model can present the temperature distribution in more detail and the performance can be increased.

4.4. CNN model application

Owing to the CNN model's success, it was further applied in this study. The best model (overall accuracy: 81.97%) was used to identify the level of thermal risk in northern Taiwan. Because of differences related to image cropping, some discrepancies in Taipei identifications were noted. However, the results for Taipei city were approximately the same and they had a similar distribution.

The model was also able to estimate the level of thermal risk for areas with LCZs' pattern. In this work, areas of northern Taiwan, including Taipei City, New Taipei City, and parts of Taoyuan and Keelung, were subjected to LCZ classification using the same satellite images mentioned previously. Details on LCZ classification related to northern Taiwan is presented in Fig. 8. After LCZ classification, the relevant data were input into the CNN model, and the level of thermal risk in each area of northern Taiwan was estimated. The results are presented in Fig. 8.

According to a comparison between the identified results in northern Taiwan and the actual air temperature, good identification

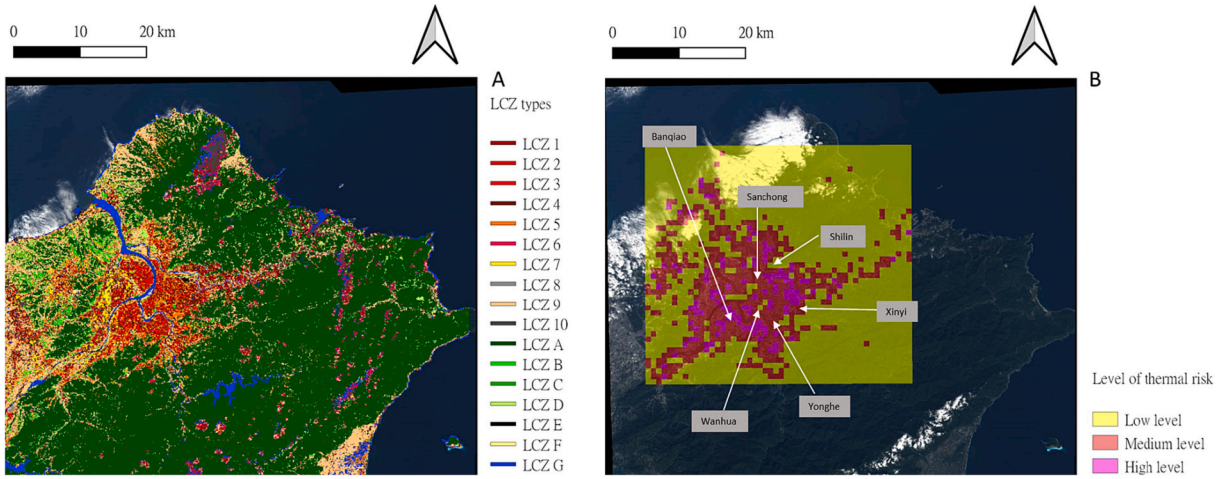


Fig. 8. (Left) LCZ map of northern Taiwan. (Right) Map of areas with level of thermal risk in northern Taiwan, as predicted by the CNN model.

results were achieved for hotspots such as Wanhua District, Shilin District, and Xinyi District in Taipei City and Yonghe District, Banqiao District, Xinzhuang District, and Sanchong District in New Taipei City; these identifications accurately reflected the high air temperature phenomenon in built environments. In areas with a low air temperature distribution, such as areas with vegetation and rivers and the mountainous regions around Taipei City, this model yielded accurate identification results. However, the air temperatures on the north side of Daan District and the south side of Songshan District may have been overestimated because of the inability to consider the environment outside the grid.

5. Conclusion

Climate change is a pertinent issue, and mesoscale studies examining the relationship between air temperature and urban structures pattern are lacking. For this reason, a method for understanding the relationship between air temperature and urban structures pattern for identifying the level of thermal risk was proposed. Additionally, the method could help the works of identifying the level of thermal risk in the regions without measuring instrument.

Landsat 8 images were used for LCZ classification and data on air temperature during August 2020 were collected by HiSAN and the CWB; these data were used in this study for analyzing the air temperature distribution in Taipei City, and machine learning was subsequently applied.

We took advantage of advances in machine learning and computer vision, and a CNN was used to analyze the level of thermal risk. Convolutional layers and pooling layers enabled the determination of urban structure combination characteristics. Furthermore, the relationship between high air temperature and urban structures pattern was determined using the fully connected layer in the proposed CNN model.

The feasibility of identifying the level of thermal risk using the proposed model and LCZ classification was confirmed. The relationship between urban structures pattern and thermal risk distribution was also discerned. Most high thermal risk level were noted in areas including more LCZ1, LCZ2, and LCZ3. Areas with more natural land cover, such as LCZA, LCZB, and LCZD, had the lower level of thermal risk in Taipei City. The results for identifying the level of thermal risk using the CNN model and LCZ classification were presented.

The accuracy of the CNN model increased with sampling percentage, and the best performance was noted for 50% sampling (accuracy: 81.97%). The CNN model was used to identify the level of thermal risk in northern Taiwan, including Taipei City and New Taipei City. The results confirmed that the CNN model could identify the level of thermal risk in areas subjected to LCZ classification.

Despite the achievements of this study, further improvements are required. A lack of measurement points and samples of urban structures were the main limitations of this work. More on-site measurements can be conducted in future studies to provide more realistic air temperature distributions. Additionally, more air temperature distribution maps and LCZ maps of other cities can be combined in further work to provide more urban structures pattern for the proposed CNN model. This would help the CNN model perform computations involving more diverse urban structures pattern to identify the level of thermal risk in different areas.

CRediT authorship contribution statement

Tsz-Kin Lau: Conceptualization, Methodology, Writing – original draft. **Yu-Cheng Chen:** Conceptualization, Methodology, Writing – review & editing. **Tzu-Ping Lin:** Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare no conflict of interest, and 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.

Acknowledgements

The authors would like to thank the National Science and Technology Council of Taiwan for financially supporting this research under Contract No. 111-2221-E-006-053-MY3, No. 112-2119-M-865-004- and No. 111-2221-E-343-001-MY3. Thanks to the additional support of The Taiwan Climate Change Projection Information and Adaptation Knowledge Platform (TCCIP).

References

Badaro-Saliba, Nada, Adjizian-Gerard, Jocelyne, Zaarour, Rita, Najjar, Georges, 2021. LCZ scheme for assessing urban heat island intensity in a complex urban area (Beirut, Lebanon). Urban Clim. 37, 100846.

Bechtel, B., Alexander, P., Böhner, J., Ching, J., Conrad, O., Feddema, J., et al., 2015. Mapping local climate zones for a worldwide database of the form and function of cities. ISPRS Int. J. Geo Inf. 4, 199–219.

Beck, C., Straub, A., Breitner, S., Cyrys, J., Philipp, A., Rathmann, J., Jacobeit, 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.

Bock, S., Weiß, M., 2019, July. A proof of local convergence for the Adam optimizer. In: In 2019 International Joint Conference on Neural Networks (IJCNN). IEEE, pp. 1–8.
Chen, Y.C., Yao, C.K., Honjo, T., Lin, T.P., 2018. The application of a high-density street-level air temperature observation network (HiSAN): dynamic variation characteristics of urban heat island in Tainan, Taiwan. Sci. Total Environ. 626, 555–566.
Chen, Y.C., Lo, T.W., Shih, W.Y., Lin, T.P., 2019. Interpreting air temperature generated from urban climatic map by urban morphology in Taipei. Theor. Appl. Climatol. 137 (3), 2657–2662.
Central Weather Bureau of Taiwan, 2011. Monthly report on climate system. http://www.cwb.gov.tw/V7/forecast/long/long_season.htm.
Chen, Yu-Cheng, Cheng, Fang-Yi, Yang, Cheng-Pei, Lin, Tzu-Ping, 2021. Explore the accuracy of the pedestrian level temperature estimated by the combination of LCZ with WRF urban canopy model through the microclimate measurement network. Environ. Sci. Proc. 8 (1), 14.
Doick, K.J., Peace, A., Hutchings, T.R., 2014. The role of one large greenspace in mitigating London's nocturnal urban heat island. Sci. Total Environ. 493, 662–671.
Duro, D.C., Franklin, S.E., Dubé, M.G., 2012. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. Remote Sens. Environ. 118, 259–272.
Gobakis, K., Kolokotsa, D., Synnefa, A., Saliari, M., Giannopoulou, K., Santamouris, M., 2011. Development of a model for urban heat island prediction using neural network techniques. Sustain. Cities Soc. 1, 104–115.
Hashemi, M., 2019. Enlarging smaller images before inputting into convolutional neural network: zero-padding vs. interpolation. J. Big Data 6 (1), 1–13.
Hsu, S., Mavrogianni, A., Hamilton, I., 2017. Comparing spatial interpolation techniques of local urban temperature for heat-related health risk estimation in a subtropical city. Procedia Eng. 198, 354–365.
Ide, H., Kurita, T., 2017, May. Improvement of learning for CNN with ReLU activation by sparse regularization. In: In 2017 International Joint Conference on Neural Networks (IJCNN). IEEE, pp. 2684–2691.
Iwahashi, J., Kamiya, I., Matsuoka, M., Yamazaki, D., 2018. Global terrain classification using 280 m DEMs: segmentation, clustering, and reclassification. Prog. Earth Planet Sci. 5 (1), 1–31.
Kántor, N., Tsai, K.T., Égerházi, L., Lin, T.-P., 2014. Outdoor thermal comfort requirements of Taiwanese and Hungarians in the warm months. In: Presentation and Extended Abstract on the 20th International Congress of Biometeorology, 28 September–2 October 2014, Cleveland, Ohio. Paper number: 5C.3.
Kesemen, O., Tiryaki, B.K., Tezel, Ö., Özkul, E., Naz, E., 2021. Random sampling with fuzzy replacement. Expert Syst. Appl. 185.
Kim, Y.-M., Kim, S., Cheong, H.-K., Ahn, B., Choi, K., January 01, 2012. Effects of heat wave on body temperature and blood pressure in the poor and elderly. Environ. Health Toxicol. 27.
Lau, K.K.L., Chung, S.C., Ren, C., 2019. Outdoor thermal comfort in different urban settings of sub-tropical high-density cities: an approach of adopting local climate zone (LCZ) classification. Build. Environ. 154, 227–238.
Leconte, Francois, Bouyer, Julien, Claverie, Rémy, Pétrissans, Mathieu, 2015. Using local climate zone scheme for UHI assessment: evaluation of the method using mobile measurements. Build. Environ. 83, 39–49.
LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86 (11), 2278–2324.
Lin, C.Y., Chen, F., Huang, J.C., Chen, W.C., Liou, Y.A., Chen, W.N., Liu, S.C., 2008. Urban heat island effect and its impact on boundary layer development and land–sea circulation over northern Taiwan. Atmos. Environ. 42 (22), 5635–5649.
Liou, C.S., Chen, J.H., Terng, C.T., Wang, F.J., Fong, C.T., Rosmond, T.E., Cheng, M.D., 1997. The second–generation global forecast system at the central weather bureau in Taiwan. Weather Forecast. 12 (3), 653–663.
Liu, C., Frazier, P., Kumar, L., 2007. Comparative assessment of the measures of thematic classification accuracy. Remote Sens. Environ. 107, 606–616.
Matzarakis, A., Amelung, B., 2008. Physiological equivalent temperature as indicator for impacts of climate change on thermal comfort of humans. In: Seasonal Forecasts, Climatic Change and Human Health. Springer, Dordrecht, pp. 161–172.
Molnár, G., Gyöngyösi, A.Z., Gál, T., 2019. Integration of an LCZ-based classification into WRF to assess the intra-urban temperature pattern under a heatwave period in Szeged, Hungary. Theor. Appl. Climatol. 138, 1139–1158.
Mukherjee, F., Singh, D., 2020. Assessing land use–land cover change and its impact on land surface temperature using LANDSAT data: a comparison of two urban areas in India. Earth Syst. Environ. 4 (2), 385–407.
Nadizadeh Shorabeh, S., Hamzeh, S., Zanganeh Shahraki, S., Firozjaei, M.K., Jokar, Arsanjani J., 2020. Modelling the intensity of surface urban heat island and predicting the emerging patterns: Landsat multi-temporal images and Tehran as case study. Int. J. Remote Sens. 41, 7400–7426.
Nistor, M.M., Rahardjo, H., Satyanaga, A., Hao, K.Z., Xiaosheng, Q., Sham, A.W.L., 2020. Investigation of groundwater table distribution using borehole piezometer data interpolation: case study of Singapore. Eng. Geol. 271, 105590.
Peng, H., Chen, S., 2019. BDNN: binary convolution neural networks for fast object detection. Pattern Recogn. Lett. 125, 91–97.
Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779–788.
Ren, C., Cai, M., Li, X., Zhang, L., Wang, R., Xu, Y., Ng, E., 2019. Assessment of local climate zone classification maps of cities in China and feasible refinements. Sci. Rep. 9 (1), 1–11.
Shih, W.Y., Ahmad, S., Chen, Y.C., Lin, T.P., Mabon, L., 2020. Spatial relationship between land development pattern and intra-urban thermal variations in Taipei. Sustain. Cities Soc. 62, 102415.
Singh, N., Singh, S., Mall, R.K., 2020. Urban ecology and human health: Implications of urban heat island, air pollution and climate change nexus. In: Urban Ecology. Elsevier, pp. 317–334.
Stewart, I.D., 2011. A systematic review and scientific critique of methodology in modern urban heat island literature. Int. J. Climatol. 31 (2), 200–217.
Stewart, I.D., Oke, T.R., 2012. Local climate zones for urban temperature studies. Bull. Am. Meteorol. Soc. 93 (12), 1879–1900.
Taheri-Shahraiyni, H., Sodoudi, S., 2017. High-resolution air temperature mapping in urban areas: a review on different modelling techniques. Therm. Sci. 21 (6 Part A), 2267–2286.
Teng, L.S., Lee, C.T., Peng, C.H., Chen, W.F., Chu, C.J., 2001. Origin and geological evolution of the Taipei basin, northern Taiwan. Western Pacific Earth Sci. 1 (2), 115–142.
Zhang, Y., Zhang, J., Zhang, X., Zhou, D., Gu, Z., 2020. Analyzing the characteristics of UHI (urban heat island) in summer daytime based on observations on 50 sites in 11 LCZ (local climate zone) types in Xi'an, China. Sustainability 13 (1), 83.
Zhang, Y., Zhang, J., Zhang, X., Zhou, D., Gu, Z., 2020. Analyzing the characteristics of UHI (Urban heat island) in summer daytime based on observations on 50 sites in 11 LCZ (local climate zone) types in Xi'an, China. Sustainability 13 (1), 83.
Zhou, Xilin, Okaze, Tsubasa, Ren, Chao, Cai, Meng, Ishida, Yasuyuki, Watanabe, Hironori, Mochida, Akashi, 2020. Evaluation of urban heat islands using local climate zones and the influence of sea-land breeze. Sustain. Cities Soc. 55, 102060.