scientific reports

Check for updates

OPEN

An improved machine learning-based model for prediction of diurnal and spatially continuous near surface air temperature

Ibrahim Ademola Adeniran[1], Majid Nazeer[1,2], Man Sing Wong[1,2,3]✉ & Pak-Wai Chan[4]

Near-surface air temperature (Tair) is crucial for assessing urban thermal conditions and their impact on human health. Traditional Tair estimation methods, reliant on sparse weather stations, often miss spatial variability. This study proposes a novel framework using a federated learning artificial neural network (FLANN) for fine-scale Tair prediction. Leveraging spatially complete thermal data from Landsat 8/9, Sentinel 3, and Himawari 8/9 (105 acquisition days, 2013–2023), and data from automatic weather stations, 23 predictor variables were extracted. After rigorous selection processes, nine variables significantly correlated with Tair were identified. Comparative analysis against established machine learning and linear models, using cross-validation data, showed FLANN's superior performance with a Pearson correlation coefficient (r) of 0.98 and a root mean square error (RMSE) of 0.97 K, compared to r and RMSE of 0.85 and 1.09, respectively, for the linear model. FLANN showed greater improvements for urban stations with r and RMSE differences of 0.19 and −2.03 K. Application of FLANN to predict Tair in Hong Kong in July 2023 enabled detailed urban heat island (UHI) analysis, revealing dynamic spatial and temporal UHI patterns. This study highlights FLANN's potential for accurate Tair prediction and UHI analysis, enhancing urban thermal environment management.

Keywords Federated learning, Neural network, Near-surface temperature, Urban heat island, Remote sensing, Machine learning

In the wake of rapid urbanization and the escalating global population, our planet faces an increasing strain on its environment, with climate change emerging as a central concern[1]. This paradigm not only revolutionized our way of life but also opened frontier for new research. Among the myriad consequences of climate change, temperature-related challenges, such as heat stress, urban heat islands (UHI), and drought, have taken centre stage in the scientific community[2–4]. Traditionally, monitoring such conditions relied on near-surface air temperature (T_{air}) measurements, typically taken at about two meters above sea level[5]. These measurements can be obtained either by mounting thermal sensors on mobile platforms such as drones and vehicles or through stationary weather stations distributed systematically across study regions[6,7]. Weather stations, renowned for their high accuracy and long temporal resolution due to stringent quality assurance measures, have been widely utilized for thermal studies[8].

However, their limitation lies in their point-based nature, capturing the thermal characteristics of immediate surroundings without scientifically defined spatial extents[9]. Analysing complex spatiotemporal temperature patterns over extensive and diverse regions, especially those with intricate topography, presents significant challenges when relying solely on limited, non-uniformly distributed weather stations[10]. Furthermore, regions with harsh terrains, such as polar region, mountainous areas, dense forest among others, often suffer from an acute under sampling of atmospheric conditions resulting from difficulty in the installation of weather station in such region, limiting the insights provided by traditional weather stations. To address this limitation, many spatial interpolation methods have been proposed, including inverse distance-weighted interpolation[11], global interpolation[12], and kriging interpolation[13], Nevertheless, T_{air} obtained through spatial interpolations may still have limitations; namely, the T_{air} accuracy highly relies on the density of the stations and the characteristics of the target region[14].

[1]Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China. [2]Research Institute for Sustainable Urban Development, The Hong Kong Polytechnic University, Hong Kong, SAR, China. [3]Research Institute of Land and Space, The Hong Kong Polytechnic University, Hong Kong, SAR, China. [4]The Hong Kong Observatory, Hong Kong, SAR, China. ✉email: Ls.charles@polyu.edu.hk

In contrast to the point-based nature of measurement from weather stations, satellite sensors have the capability to continuously scan the earth's surface and atmosphere over vast areas at regular intervals and as such provide spatially continuous measurements of environmental phenomena[15,16]. For example, satellite sensors equipped with thermal infrared (TIR) sensors, such as Landsat 4–9, Sentinel 3, Himawari 8, and 9, can be used to retrieve land surface temperature (LST), relative humidity, column water vapour (CWV) among other quantities that can better be used to answer environmental research questions. Thus, to bridge the information gap resulting from spatial limitation of weather stations, researchers have turned to retrieve T_{air} from LST[17–19].

Although T_{air} and LST are generally correlated, the specific relationship between them varies both geographically and temporally[20]. Consequently, a simple linear relationship is often inadequate to calibrate LST to T_{air} accurately[21,22]. This challenge has driven the development of machine learning models to predict T_{air} from satellite-derived LST data together with other environmental parameters which include normalized difference vegetation index (NDVI), normalized difference building index (NDBI), normalized difference water index (NDWI), elevation data and location data among others. For example Zhang, et al.[18] employed the random forest (RF) model to predict hourly T_{air} from LST retrieved from the FY-4 A geostationary satellite, the model achieved a predictive performance of up to 1.65 K. Similarly, Che, et al.[7] reconstructed fine-scale T_{air} based on LST retried from Moderate Resolution Imaging Spectroradiometer (MODIS) Aqua and Terra satellite data using four different machine learning models which include RF, artificial neural network (ANN), support vector machine (SVM), and gaussian process regression (GPR), with ANN having the highest accuracy and less bias of -0.36 K. Despite the accuracy of these models, some limitation still exists in the estimation of T_{air} from LST. This includes the lack of scientific definition for the spatial extent of weather stations, which invariably determine the radius of aggregation of LST and other predictor variables for T_{air} prediction. Consequently, studies have employed varying radii to aggregate LSTs and other satellite-derived data for T_{air} prediction. For example Otgonbayar et al.[23] utilized the point-to-point (P2P) data extraction method, overlaying the coordinates of weather stations onto satellite data to extract predicting variables. Employing the Partial Least Squares (PLS) regression model, the study achieved an accuracy (coefficient of determination, R^2) of 0.87. Other studies, such as Deo and Şahin[24], and Benali, et al.[25] predicted T_{air} from MODIS Terra and Aqua LST (MOD11A2 and MYD11A2) using an ANN model. Employing the moving window technique, they overlaid the corresponding locations of weather stations onto satellite data, extracting predictor variables by estimating the average value of satellite data within the window size. Deo and Şahin[24], utilized a window size of 3 × 3 pixels, achieving a prediction accuracy (R^2) of 0.96, while Benali et al.[25] used a 5 × 5 window, resulting in an R^2 of 0.94.

Apart from the undefined spatial resolution of weather stations, other limitations in existing studies stem from the prediction of T_{air} from LST, including the trade-off between the spectral and temporal resolutions of the thermal sensors. This trade-off makes it challenging to estimate LST at both high spatial and temporal resolutions simultaneously, thus hindering the accurate estimation of T_{air} under such conditions. Additionally, satellite TIR-based estimation of LST is prone to significant interference from cloud contamination[17,26,27]. This leads to the estimated T_{air} lacking all-weather applicability due to the frequent occurrence of cloud coverage, making it unsuitable for applications under diverse weather conditions.

In view of these limitations, this study aims to (1) develop an improved model for predicting spatially continuous T_{air} from LST under all-weather conditions using Hong Kong as a case study and (2) predict diurnal T_{air} variations across the study area, facilitating the analysis of the surface UHI phenomenon.

Study area and data used

Study area

Hong Kong is a highly dense city located along the South China Sea coast. It has a rugged topography with about 70% of the study area not developable as a result of highland, steep slopes and water bodies (Fig. 1). Resulting in high concentration of urban development and human activities in the limited developable land especially around Kwun Tong, Wong Tai Sin and Yau Tsim Mong districts which are the three densely populated districts in the study area[28]. Hong Kong has four seasons which varies from winter, spring, summer, and autumn. With length of day and night varying across different season. Length of day is relatively short in winter and longer in the summer[29]. Given the rough topography of Hong Kong and the density of buildings in the developed region it becomes difficult to have weather station all over the city and the ones available cannot be representative of the area[30]. Thus, making Hong Kong a suitable location for our study and output of the study a useful component for environmental studies in Hong Kong and other cities with similar landscape or topography.

Data sources

For this study, the data were broadly categorized in two groups based on their spatial characteristics: station data and continuous spatial data.

Station data

This encompasses point-specific data obtained from the Automatic Weather Stations (AWSs) located around Hong Kong. Managed primarily by the Hong Kong Observatory (HKO), these data primarily pertain to meteorological parameters. The focal variable of interest in our study, near-surface temperature (T_{air}) was obtained from this source. Out of the 61 weather stations in Hong Kong, 50 stations provided air temperature data (Fig. 1). Additionally, predictor variables such as precipitation, solar radiation, wind speed, and wind direction were also acquired from this source.

Continuous spatial data

This category comprises spatially continuous datasets spanning the study area, encompassing environmental variables with known associations to air temperature as documented in the literature (see Table 1). The data for

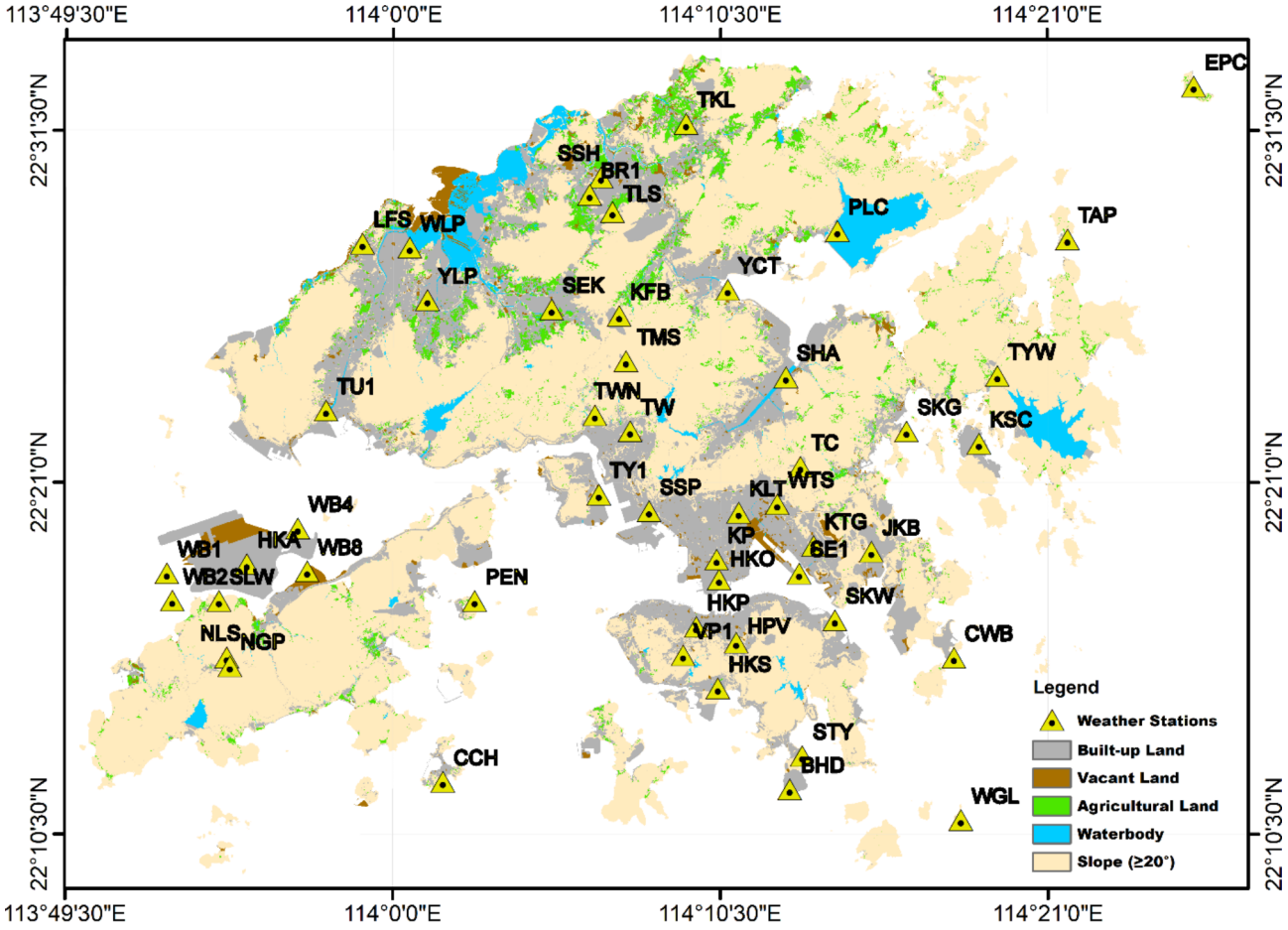

Fig. 1. Map of Hong Kong with spatial location of Weather Stations. Source: Hong Kong Planning Department[31].

this category were predominantly acquired from two primary sources: (i) from satellite sensors and (ii) from the Hong Kong open data website (https://data.gov.hk/). LST, being the primary predictor for T_{air},was acquired from satellite data. To maximize spatial and temporal resolution of sun-synchronous and geostationary satellite, LST data were sourced from both satellite sensors. This includes Landsat 8 and 9, Sentinel 3 Sea and Land Surface Temperature Radiometer (SLSTR), and Himawari 8 and 9. Landsat 8 and 9 provided a spatial resolution of 100 m and a temporal resolution of 8–16 days, while Sentinel 3 SLR offered a spatial resolution of 1000 m with a temporal resolution of 12 to 24 h. Geostationary data from Himawari 8 and 9 featured a spatial resolution of 2000 m with a temporal resolution ranging between 2.5 min and 10 min. In addition, Land Surface Emissivity data (LSE) was sourced from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) sensor (onboard Terra satellite). Auxiliary data like digital terrain model (DTM), building data, land utilization data, and population data were sourced from the Hong Kong Open data website (https://data.gov.hk/) as made available by different government departments.

Methodology

The methodology for this study was structured into three broad stages i.e., data pre-processing stage, modelling stage, and the model application stage where optimized model was employed for diurnal UHI analysis.

Data pre-processing

The first step in the data pre-processing was the optimization of hourly fine-scale LST for each Landsat 8/9 acquisition date, resulting in a seamless twenty-four-hour fine-scale dataset for each of the 105 Landsat 8 acquisition dates (Appendix 1). This was achieved using the diurnal Integrated Spatiotemporal Fusion Algorithm (ISFAT) developed by Adeniran et al.[32]. The diurnal ISFAT method predicts hourly fine-scale LST by integrating temporal information from moderate-scale LST datasets retrieved from Sentinel-3 SLTR ($T_{M,(d_p,t_m)}$ and $T_{M,(d_p,t_n)}$) and hourly coarse-scale LST data from Himawari 8 and 9 satellite sensors ($T_{c,(d_p,t_p)}$) with the spatial information from Landsat 8/9 data collected on each acquisition date. To ensure seamless integration, all input data were resampled to match the spatial resolution of the finest LST data (Landsat 8/9). The fine-scale LST at each timestamp was then predicted using a temporal and weighted scale function, as detailed in Eq. (1). This model was selected for its proven ability to generate seamless fine-scale LST by incorporating a masked weight

Predictor variable	Spatial resolution	Temporal resolution	Source
Precipitation (PR)	Point	1 h	HKO
Solar radiation (SR)	Point	1 h	HKO
Wind speed (WS)	Point	1 h	HKO
Wind direction (WD)	Point	1 h	HKO
Relative humidity (RH)	Point	1 h	HKO
Distance to water (DTW)	Point	1 year	Calculated using ArcGIS software
Distance to building (DTB)	Point	1 year	Calculated using ArcMap with reference to data building data from building department
Longitude	Point	No	Extracted from weather station
Latitude	Point	No	Extracted from weather station
Station class (SC)	No	No	Estimated using LCZ and Google Earth data
Population density (PD)	TCU	1 year	Census and Statistics Department
Elevation	5 m	No	Calculated using ArcMap using DTM
Slope	5 m	No	Calculated using ArcMap using DTM
Aspect	5 m	No	Calculated using ArcMap using DTM
Hillshade	5 m	No	Calculated using ArcMap using DTM
SVF	5 m	No	Calculated using ArcMap using DTM and Building height
LST	100 m	1 h	Satellite data
Land utilization (LU)	10 m	1 year	HK planning department
LSE	100 m	1 year	Satellite data
NDVI	100 m	1 day	Satellite data
NDWI	100 m	1 year	Estimated from annual average
NDBI	100 m	1 year	Estimated from annual average
Acquisition date (DOY)	point	No	Derived from satellite metadata
Acquisition time (TOD)	point	No	Derived from satellite metadata

Table 1. List of variables used in this study and the information about the data products associated with the variables.

function, which prevents data loss from missing pixels, and by effectively estimating and integrating model residuals into the predictions, as demonstrated by Adeniran et al.[32].

$$T_{F,(d_p,t_p)} = T_{F,(d_p,t_m)} - T_{M,(d_p,t_m)} + T_{M,(d_p,t_n)} - T_{c,(d_p,t_n)} + T_{C,(d_p,t_p)} \tag{1}$$

where d_p is the acquisition date of the base fine scale LST, t_m represents the time of the day when the fine-scale LST was collected, t_n correspond to another time stamp the medium scale LST is available on d_p, while t_p is the time of the day to be predicted. For this study, t_p ranges from $t_{00:00}$to $t_{23:00}$ at hourly intervals.

Additionally considering that the Day of Year (DOY) and Time of Day (TOD) inherently exhibit cyclic characteristics a cosine transformation was applied to both variables to accurately reflect this cyclicity in the modelling. This approach follows the methodology of Araghi et al.[33], as presented in Eqs. (2) and (3). Where t_n is the time of the day in hours (0–23), d_n is the ordinal day number of acquisition date within the year, and Y denotes the total number of days in the year (365 or 365 in a leap year). This constrains the DOY and TOD values between − 1 and 1, ensuring the continuity of the input variables.

$$TOD = \cos\left(\frac{2\pi\, t_n}{24}\right) \tag{2}$$

$$DOY = \cos\left(\frac{2\pi\, d_n}{Y}\right) \tag{3}$$

Following the optimization and transformation, resampling and projections procedure were applied to the optimized LST data along with other predictor variables retrieved from satellite data. This step ensured that all data could be easily grouped together, including ensuring their alignment in a uniform coordinate system (WGS 1984 as datum and UTM Zone 49 N).

Model development

Federated learning artificial neural network

In this study, we introduce an adaptive federated learning artificial neural network (FLANN) based framework for predicting air temperature (T_{air}) using satellite retrieved LST. The FLANN framework, developed by McMahan et al.[34], is an ANN technique that trains an algorithm through multiple independent sessions, each with its own dataset. This approach ensures that training occurs privately for different data classes, preventing

overfitting to a specific sub-class and maintaining data privacy across respective classes. To tailor this model for T_{air} prediction, we incorporated an initial classification layer for station and remotely sensed data classification and spatial analysis into the FLANN framework (Fig. 2). This layer includes various analyses designed to systematically classify station data into disaggregated classes of Urban, Urban Oasis, Suburban, and Rural, as explained in the next section.

Following the FLANN framework, a local ANN model was individually trained using data from each of these classes concurrently. After training, the model parameters (e.g., weights and biases) from the locally

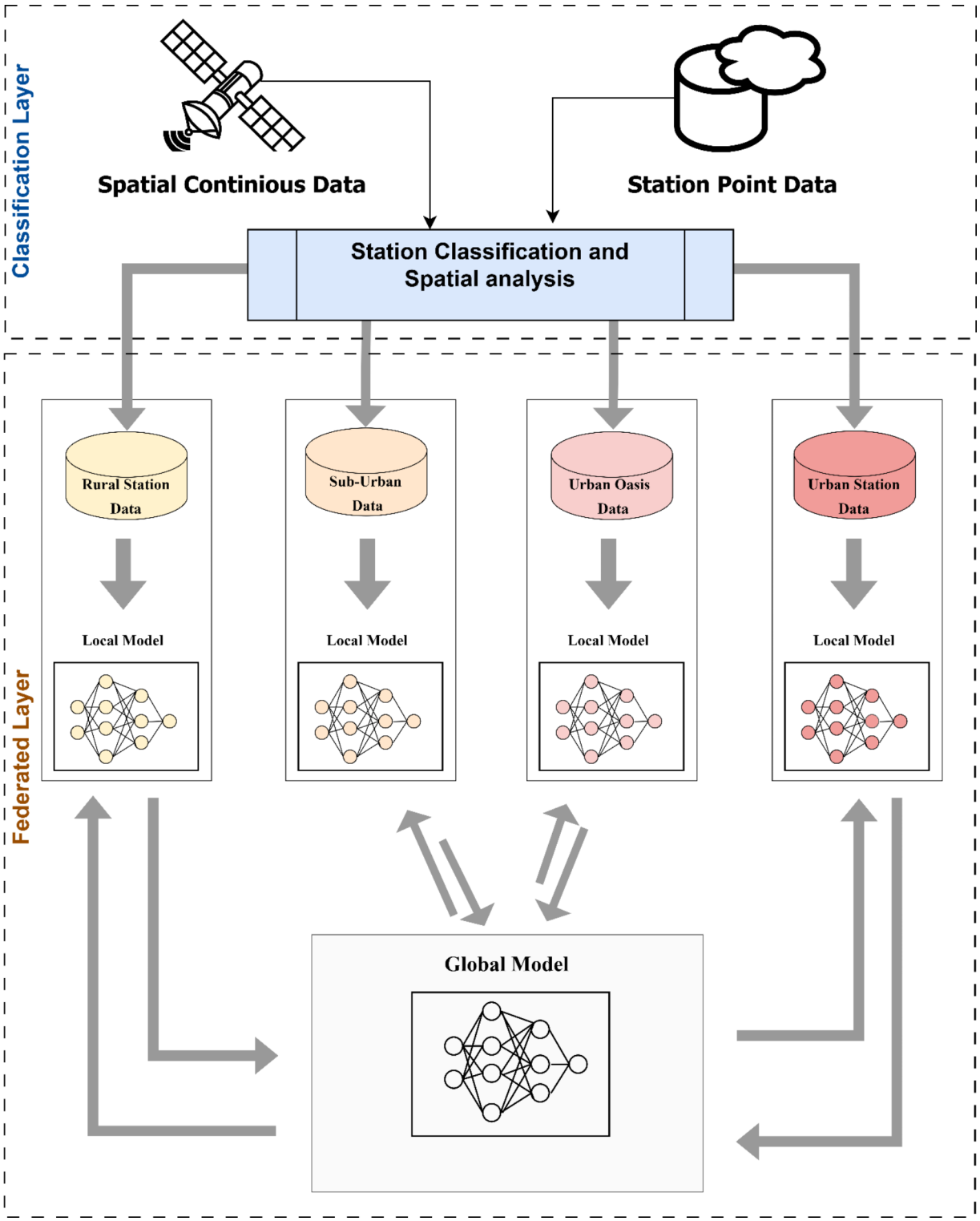

Fig. 2. Adaptive FLANN based near-surface temperature prediction framework.

trained models for each station class were aggregated and sent to a global model. This global model then used these parameters to train a unified ANN model that benefits from the collective knowledge of all local models while still preserving the distinct characteristics of each class. The global model, once updated, redistributed the learned parameters back to the local models, which were then refined through further training iterations. This iterative process continued until the model converged, ensuring that the final global model was both robust and generalizable across all station classes.

This federated learning approach not only enhances the model's predictive accuracy by leveraging diverse datasets but also ensures that the distinct spatial and temporal characteristics inherent in the different classes are adequately captured. Additionally, by preventing the direct sharing of raw data between classes, this method safeguards the privacy and integrity of the data used in the study.

Station classification and spatial analysis

Station classification Studies on urban thermal environment such as UHI analysis, commonly categorize weather stations into two primary classes i.e., rural and urban[35]. This classification is typically based on factors such as population density[36], surrounding environment, and percentage of vegetation cover[35]. In this study, given the complexity of the urban environment and the heterogenous land cover surrounding the weather stations, classification of weather stations based on local climate zone (LCZ) was introduced. The study area was classified into two main types following the LCZ framework by Stewart and Oke[37]. The "Land Cover Type" (LCT) encompasses built-up areas, including LCZ classes 1 to 10, which range from dense high-rise buildings to sparsely built regions and heavy industrial zones. The "Vegetation Cover Type" (VCT) includes natural areas, covering LCZ classes A to G, which consist of dense and scattered trees, bushes, low plants, bare rock or paved surfaces, bare soil or sand, and water bodies. Following the LCZ classification, ring of influence was then designed around each weather station, with varying distances of 100, 300, 500, and 1000 m. These zones were then overlaid on the classified LCZ map. Guided by Google Earth data sourced from Google Earth Pro 2023, the stations were subsequently divided into four distinct classes based on specific conditions as presented below.

Urban stations: This class include stations that are located in highly developed urban areas. They are surrounded by LCT in all four buffers, with minimal to no vegetation in the immediate 100 m buffer. The LCT percentage in this buffer must exceed 70%, and a similar dominance of LCT is expected in the 300, 500, and 1000 m buffers. See Fig. 3a for sample station in this category.

Urban Oasis Stations: This class encompass stations that are also situated within the developed region of the study area. However, the 100 m buffer, contains a relatively large percentage (greater than 30%) of VCT. The 300, 500, and 1000 m buffers, in contrast, are predominantly dominated by various LCT, an example of station that fall to this category is presented in Fig. 3b.

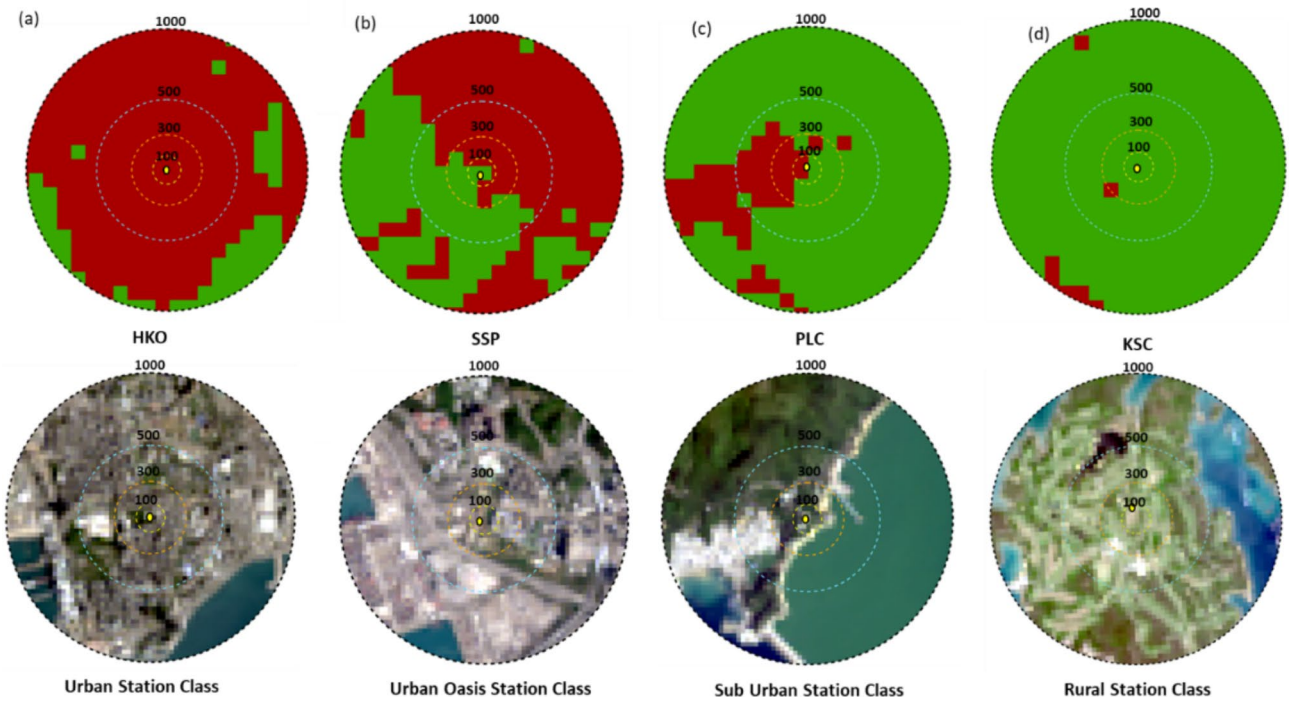

Fig. 3. Illustration of four buffer zones (i.e. 100, 300, 500 and 1000 m) around the weather stations with typical examples of stations classified as (**a**) urban, (**b**) urban oasis, (**c**) sub-urban and (**d**) rural. The map was generated with the ArcMap v10.7 software (https://www.esri.com/en-us/arcgis/products/arcgis-desktop) and LCZ and base map are generated from Landsat data retrieved from United States Geological Survey (https://earthexplorer.usgs.gov/).

Sub-urban stations: These stations are the ones located outside the highly developed urban area and are predominantly surrounded by VCT in most of the buffer rings. However, the total aggregate of LCT in the four buffers is greater than 30%. See Fig. 3c for sample station in this category.

Rural stations: This class represents stations which are located in the countryside and are largely surrounded by VCT. The percentage of LCT within the four buffer zones is less than 30 An example of a rural station is presented in Fig. 3d.

Spatial analysis Following the classification of weather stations into these four classes: Urban (ur), Urban Oasis (uo), Sub-Urban (su), and Rural (ru), an experiment was carried out to determine the optimum spatial aggregate radius for data in the respective classes. In this experiment continuous spatial variables were extracted at 50 m, 100 m, 300 m, 500 m, 700 m and 1000 m for each station class. After which correlation analysis was carried out between the aggregated data at different radius and the target variable (T_{air}) to determine the optimum radius for aggregating the different continuous spatial variable for the different station class.

Existing T_{air} prediction models

To enable a robust assessment of the performance of the newly introduced FLANN model, we employed existing models that have demonstrated acceptable results for predicting T_{air} in the study area. These models include the multiple linear regression model (MLR), the random forest model (RF), and the multilayer perceptron neural network model (MLPANN).

Variable selection and sensitivity analysis

After extracting spatially continuous and station-discrete variables, as presented in Table 1, we assembled a comprehensive set of 23 predictor variables to predict T_{air}. However, to ensure the robustness and reliability of the T_{air}predictor model, we conducted assessments for autocorrelation and multicollinearity. Variables exhibiting high collinearity were systematically eliminated through a series of iterations between multicollinearity testing and the removal of variables with Variance Inflation Factor (VIF) scores exceeding the threshold value of 5.0 following Benitez, et al.[38]. After each VIF analysis iteration, variables registering excessive VIF scores were systematically eliminated. Notably, among variables with heightened VIF scores, the one demonstrating the weakest correlation was prioritized for removal in each cycle. Finally, sensitivity analysis of the variables will be carried out to highlight the impact of individual variable on Tair prediction using the optimum model.

Model validation

In this study, following the optimization of fine scale hourly LST, spatial matching was carried out based on station point data and continuous spatial data from 2013 to 2023. Ultimately 1224 daily sample data points and 127,296 total points were collected over the study period. Given that the model is intended for T_{air} prediction in pixels without available weather station data, the leave-station-out validation method was employed. This approach is particularly appropriate due to the limited number of stations (50) in the study area with T_{air} data across the study period. In the leave-one-station-out validation, the model was iteratively trained on data from all but one station and then validated on data from the excluded station. This process was repeated for each of the 50 stations, ensuring that each station served as the validation set exactly once. The final model accuracy was computed by averaging the performance metrics obtained across all iterations. This method provides a robust evaluation of the model's ability to generalize and predict T_{air} in locations where station data are unavailable, effectively simulating real-world conditions.

Assessment of model accuracy

To assess the accuracy of each model developed in this study, the Pearson correlation coefficient (r), root mean square error (RMSE), and the mean absolute error (MAE) between the predicted T_{air} value and the test T_{air} value sampled from the station data will be computed.

Diurnal UHI analysis

Following the development of the robust model for the estimation of T_{air} prediction, hourly T_{air} was predicted from optimised hourly LST on June 2, 2023, over the study area and the resulting set of T_{air} was employed to analyse the pattern of UHI over the course of the day. UHI was estimated following the study of[39] by subtracting the average temperature of rural pixels ($\Delta T_{air_{RU(i,j)}}$) from other pixels ($T_{air(i,j)}$) in the study area as presented in Eq. (2).

$$UHI = T_{air(i,j)} - \Delta T_{air_{RU(i,j)}} \tag{4}$$

In addition, the thermal comfort in the study area was assessed by studying degree of temperature variance across the study area throughout the day using the Urban Thermal Field Variance Index (UTFVI, Eq. (3) (Naim and Kafy[40]. The study area was further classified into six different regions based on their UTFVI scores (Table 2) and the percentage change across the day was assessed to understand how the thermal environment responds to human activities throughout the day.

$$UTFVI = \frac{T_s - T_m}{T_m} \tag{5}$$

UTFVI range	UTFVI class	UTCL
< 0	None	Excellent
0–0.005	Weak	Good
0.005–0.01	Middle	Normal
0.01–0.015	Strong	Bad
0.015–0.02	Stronger	Worse
> 0.02	Strongest	Worst

Table 2. Threshold values of Urban Thermal Field Variance Index (UTFVI).

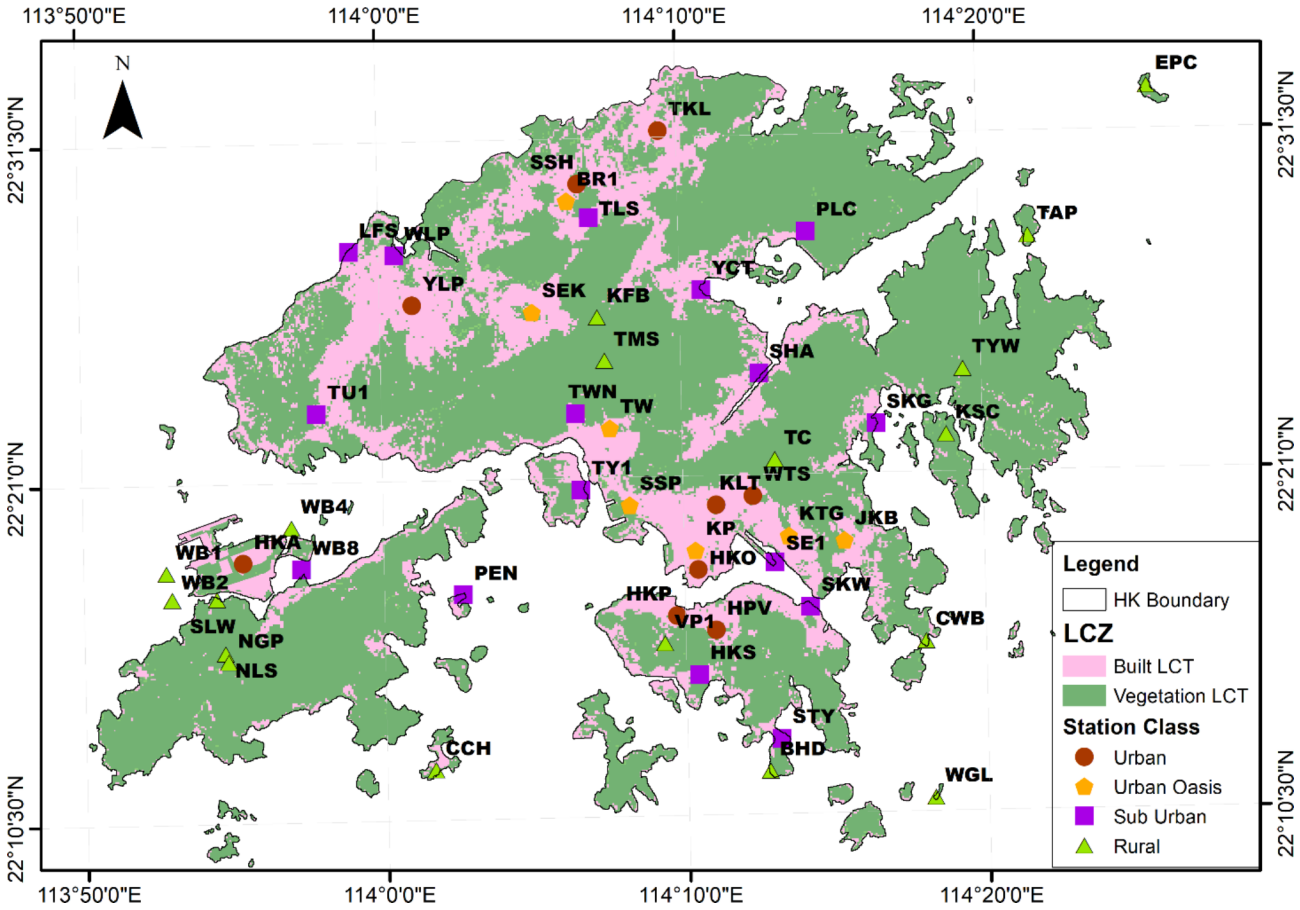

Fig. 4. Station classification over study area.

Results

Station classification and spatial analysis

Station classification

Result of station classification as presented in Fig. 4 revealed that the study area is dominated by station classified as rural station (18) followed by station in sub urban class (16). Urban and Urban Oasis account for nine and seven stations respectively.

Spatial analysis

Result of spatial analysis as presented in Fig. 5 revealed that, for the Urban class (ur), the variables LST, NDVI, NDBI, and NDWI which are positively correlated with T_{air} each exhibited their highest correlation coefficients at 50 m radius. However, land utilization and topographic variables such as slope and elevation, despite being positively correlated, showed an optimal aggregation radius at 500 m. Conversely, SVF and hill shade, with an inverse relationship, had an optimal aggregation radius of 1000 m. In the Urban Oasis class, variables such as LST, NDBI, and NDWI showed the highest correlation coefficients at 300 m radius. Notably, NDVI maintained an optimal aggregation radius of 50 m, similar to the Urban class. For the Sub-Urban class, the variables LST, NDVI, NDBI, and NDWI demonstrated the highest correlation coefficients (0.63, 0.19, 0.17, 0.098 respectively) at a 500 m radius. Interestingly, the variables slope, elevation, and hill shade exhibited a negative correlation and

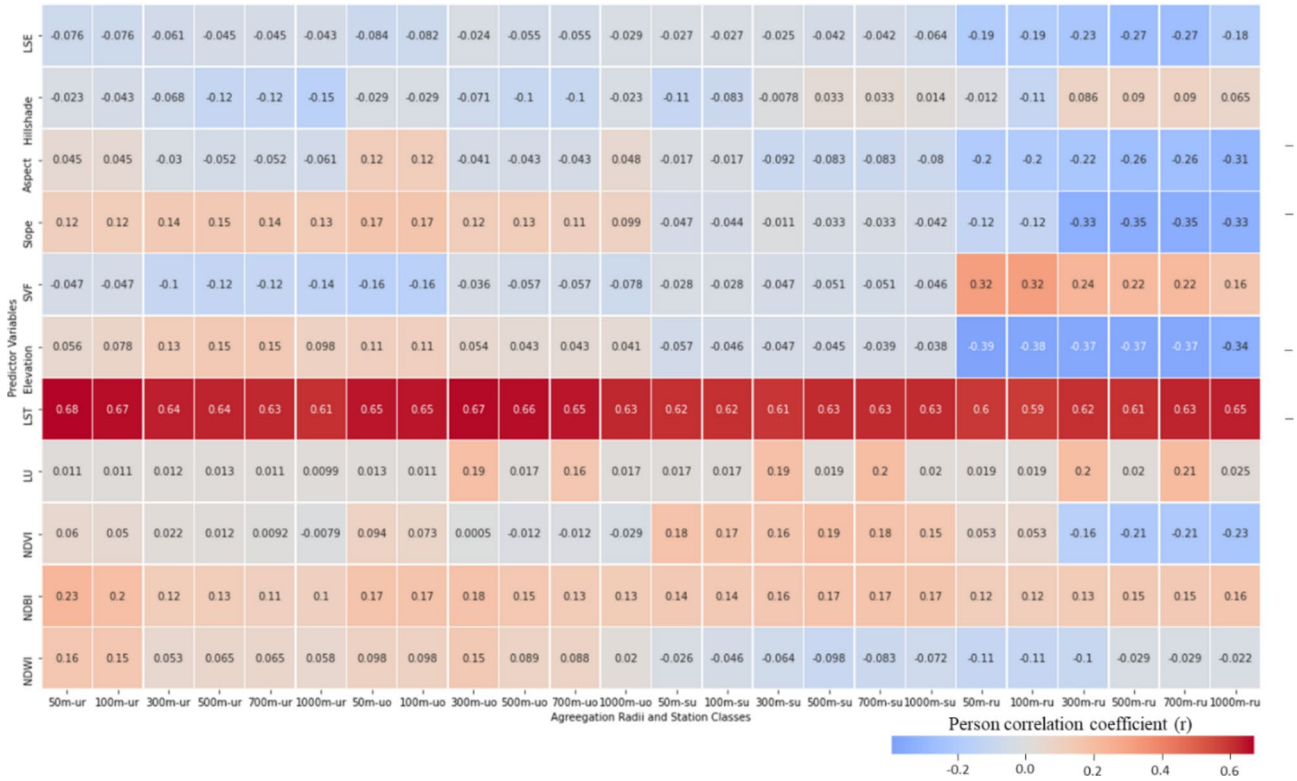

Fig. 5. Relationships between continuous spatial variables aggregated at 50 m, 100 m, 300 m, 500 m, and 1000 m radius across Urban (ur), Urban Oasis (uo), Sub-Urban (su), and Rural (ru) classes.

had an optimal aggregation radius of 50 m. In the Rural class, variables such as LST, LU, and NDBI presented the highest correlation coefficients at 1000 m radius. However, for variables which showed a negative correlation, elevation, aspect, and NDWI had optimum radius of 50 m while for slope and LSE optimal aggregation radius was 700 m.

Overall, the optimal aggregation radius for variables positively correlated with T_{air} was notably lower (50–300 m) in the Urban and Urban Oasis classes but increased significantly in the Sub-Urban and Rural classes. This trend can be attributed to the presence of tall buildings in the Urban and Urban Oasis classes, influencing the variables. It is also observed that variables like slope, elevation, and NDWI which are positively correlated to T_{air} in the urban and urban oasis classes become negatively correlated in the rural class. While SVF and hill shade that are negatively correlated in urban and urban oasis class became positively correlated in the rural class. This underscores the high impact of the surrounding environment on the predictor variables and the need to consider the difference in variable aggregation.

Comparing the Pearson corelation coefficient (r) of using the optimum spatial extent (OSE) for each station class with the conventional pixel-to-pixel (P2P) and 3×3 pixel aggregation methods commonly used in the literature (Table 3), the optimal aggregation radius method demonstrated improved r value across all variables. Notably, the improvements in the r values were most pronounced for slope variable, with r scores increased by up to 0.28 compared to traditional P2P aggregation method.

Variable selection

After a series of iterations between multicollinearity testing and the removal of variables with VIF scores above the threshold (VIF > 5), we were left with a total of 9 variables (Table 4). These resulting variables were then employed for the development of the T_{air} prediction model.

Comparative analysis of predictive models

The performance of four distinct models, including Multiple Linear Regression (MLR), Random Forest (RF), Multi-Layer Perceptron (MLPNN), and Federated Learning (FLANN), was evaluated for predicting air temperature (T_{air}). Notably, all machine learning (ML) models (RF, MLPNN, and FLANN) demonstrated outstanding performance, with r values exceeding 0.90 (see Fig. 6). Conversely, the linear model (MLR) registered a lower r value of 0.85, indicating a relatively lower accuracy. This discrepancy however can be attributed to the intricate landscape complexity of Hong Kong, leading to better performance of the ML models due to their ability to capturing non-linear relationships more effectively.

Variable	OSE	P2P	3 × 3
LSE	− 0.20	− 0.15	− 0.19
Hillshade	0.09	− 0.04	− 0.07
Aspect	− 0.30	− 0.15	− 0.25
Slope	− 0.32	− 0.04	− 0.14
SVF	0.22	0.13	0.18
Elevation	− 0.21	− 0.03	− 0.19
LST	0.68	0.60	0.66
LU	− 0.07	− 0.06	− 0.01
NDVI	− 0.24	− 0.11	− 0.24
NDBI	0.15	0.13	0.09
NDWI	0.06	0.05	0.06

Table 3. Person correlation coefficients of continuous spatial variables aggregated using optimal spatial extent (OSE), P2P, and 3 × 3 window size methods.

Variable	VIF	*r*
LST	4.873	0.68
DTW	4.335	0.12
DTB	4.740	0.21
PD	1.982	0.09
NDBI	1.360	0.15
NDVI	1.360	− 0.24
TD	1.364	0.12
TOD	2.098	0.03
SC	3.908	0.32

Table 4. Predictor variables selected post-multicollinearity assessment.

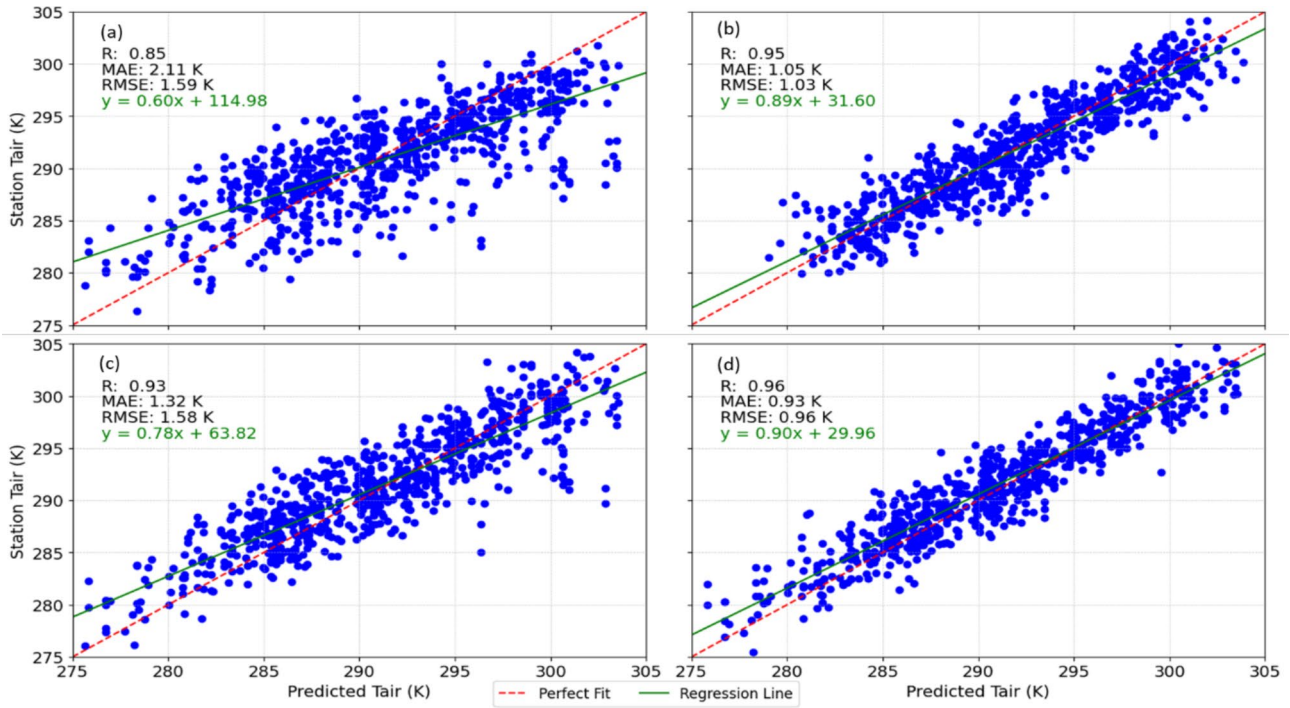

Fig. 6. Comparison of predictive models using (**a**) Multiple linear regression (MLR), (**b**) Random Forest (RF), (**c**) Multiple layer perceptron, and (**d**) Federated Learning (FL) framework.

Assessment of model predictive performance over station points

Spatial assessment of the distribution of RMSE across the study area based on station class revealed that stations situated within urban areas, including both urban and urban oasis classes, consistently exhibited lower performance across all four T_{air} estimations (Fig. 7). Conversely, rural and sub-urban stations, dominated by water bodies and vegetative land use consistently outperformed their urban counterparts. Notably for T_{air} predictions using MLP, RF, MLPNN and FLANN, 41%, 77%, 82%, and 82% of the rural station prediction recorded a mean RMSE between 0.0 and 1.0 K. This contrasts with urban stations where none achieved an RMSE in that range for MLPANN prediction. Similarly, in RF predictions, only 11% of urban stations reached an average RMSE in this range, while in FLANN predictions, the percentage was 22%. The superior performance of all models in rural station class can be attributed to a denser network of station points, enhancing data reliability. Additionally, the homogeneity of land use and the peripheral location of rural station points in the study area, where human activities are minimal, which further enhance the accuracy. Even within specialized training methods like the FLANN model, rural stations showed improved accuracy over other station classes. In contrast, urban stations, with their dense human activities and heterogeneous land use, showed the poorest accuracy. The limited number of station points in urban areas, coupled with the variability introduced by human activities, challenges the accuracy of predictions. Sub-urban stations, with a higher density of station points than urban areas but less variability in land use, followed rural stations in accuracy. Urban oasis stations, despite being better than urban stations, still demonstrated relatively lower accuracy, likely due to their proximity to urban areas and less station point density.

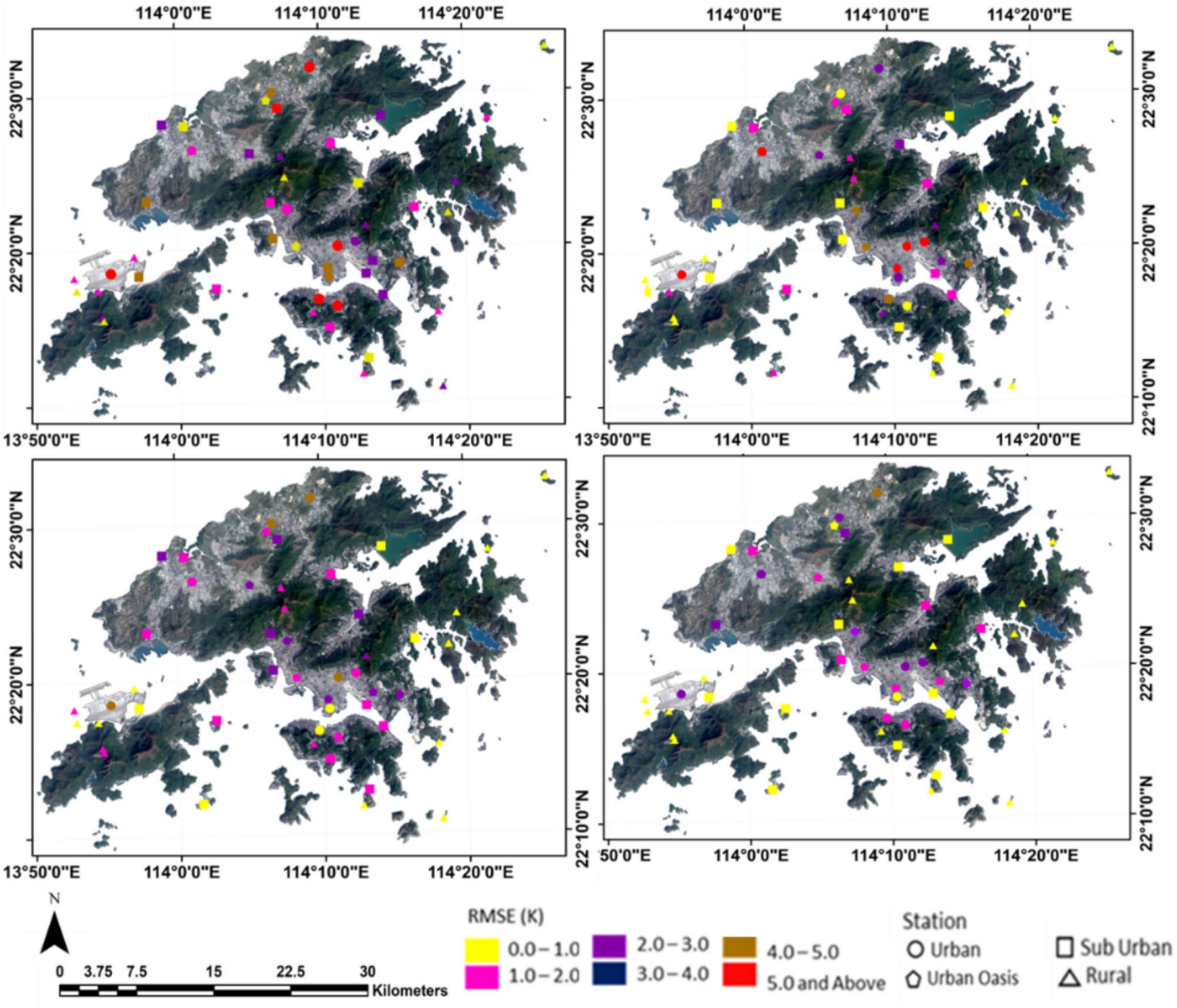

Fig. 7. Spatial distribution illustrating the station specific RMSE across different model predictions: (**a**) Multiple Linear Regression (MLR), (**b**) Random Forest (RF), (**c**) Multiple Layer Perceptron (MLPANN), and (**d**) Federated Learning (FLANN).

Model performance across varying level of environmental development

To comprehensively assess model performance across diverse environmental contexts, we leveraged cross-validation (CV) data for each station class, thus ensuring a rigorous evaluation of predictive capabilities. Table 5 delineates the performance metrics, highlighting the strengths and weaknesses of each model across different station classes. FLANN based model emerged as the highly accurate model, outperforming other models in three out of four station classes including Urban, Urban Oasis, and Sub-Urban classes. However, in the Rural class, RF displayed comparable performance, given its adaptability to high-density station data within this category. The FLANN framework's superior performance can be ascribed to its segmented training approach. This method allows for localized processing and training of data specific to each station class, mitigating overfitting risks associated with an imbalanced density of training data or unique relationships inherent to classes. The MLR model however is the least accurate of all the model across the four stations classes. This highlights the inability of linear model to effectively capture the relationship between T_{air} and LST or other predictor variables.

Sensitivity analysis of key variables in T_{air} prediction

Sensitivity analysis was conducted using the FLANN framework to assess the impact of key predictor variables on the accuracy of T_{air} predictions. The results as presented in Fig. 8 underscore that LST is the most influential variable, with its removal causing a significant drop in model performance across all station classes. The R decreased by at least 0.17 across the board, with the Rural class experiencing the largest decline of 0.23. The RMSE correspondingly increased by at least 1.43 K, peaking at 1.94 K in the Rural class, confirming LST's critical role in T_{air} prediction.

While TOD exhibited the least overall impact, its removal still led to a notable increase in RMSE (0.13 K) and a slight reduction in R (0.03) in the Urban class. This suggests that, in urban environments where the UHI effect is significant, TOD contributes to model accuracy by capturing temperature variations not fully accounted for by LST alone.

Despite LST being the most influential predictor, this analysis highlights the importance of the contributory strength of all input variables used in the study. Even TOD, which might be considered less impactful overall, plays a critical role in enhancing prediction accuracy in urban stations where UHI effects can alter the relationship between LST and T_{air}. This underscores the necessity of a multi-variable approach in accurately modeling complex environmental phenomena like air temperature.

Application of predicted T_{air} for urban thermal analysis

Diurnal Urban Heat Island (UHI) analysis

The FLANN model, which demonstrated superior accuracy in the model comparison, was selected as the optimal choice for predicting T_{air} from optimized LST across Hong Kong on a typical summer day (June 2, 2023) to assess hourly UHI pattern in the area. As presented in Fig. 9, the average *r* and RMSE of the FLANN predicted T_{air}across the four station classes in the study area, when compared with T_{air} from AWS, ranges between 0.94 and 0.98 and 0.41 K to 1.43 K, respectively, making it suitable for UHI assessment in the study area.

The spatial map of UHI analysis (Fig. 10) revealed that, at 12:00 am (00:00), a conspicuous pattern of Sink Island (SI) was evident at the centre of the study area (Lam Tsuen and Tai Mo Shan Country Park), as well as on Islands (Hong Kong Island and Lantau Island) at the bottom right and left corners. The spatial extent of the SI, particularly the one at the centre, continued to expand throughout the night. By 6:00 am (06:00), as the sun rose, the spatial extent of the SI diminished rapidly, and the region transitioned to exhibit a pattern of Heat Island (HI). Subsequently, later in the evening (17:00), as the sun began to set, the spatial pattern of the SI

Station class	Model	*R*	MAE (K)	RMSE (K)
Urban	MLR	0.79	2.23	1.89
	RF	0.92	1.09	1.27
	MLPANN	0.92	0.35	0.55
	FLANN	0.94	0.30	0.57
Urban oasis	MLR	0.84	2.00	2.15
	RF	0.93	0.97	1.15
	MLPANN	0.91	1.57	1.73
	FLANN	0.95	0.33	0.98
Sub-urban	MLR	0.83	2.68	2.31
	RF	0.92	0.62	1.22
	MLPANN	0.94	0.37	0.49
	FLANN	0.95	0.35	0.79
Rural	MLR	0.87	2.26	1.56
	RF	0.97	0.61	0.35
	MLPANN	0.92	1.21	1.27
	FLANN	0.96	0.74	0.97

Table 5. Accuracy assessment of predictive models based on station class test data. Where the bold values represent the best result in each class.

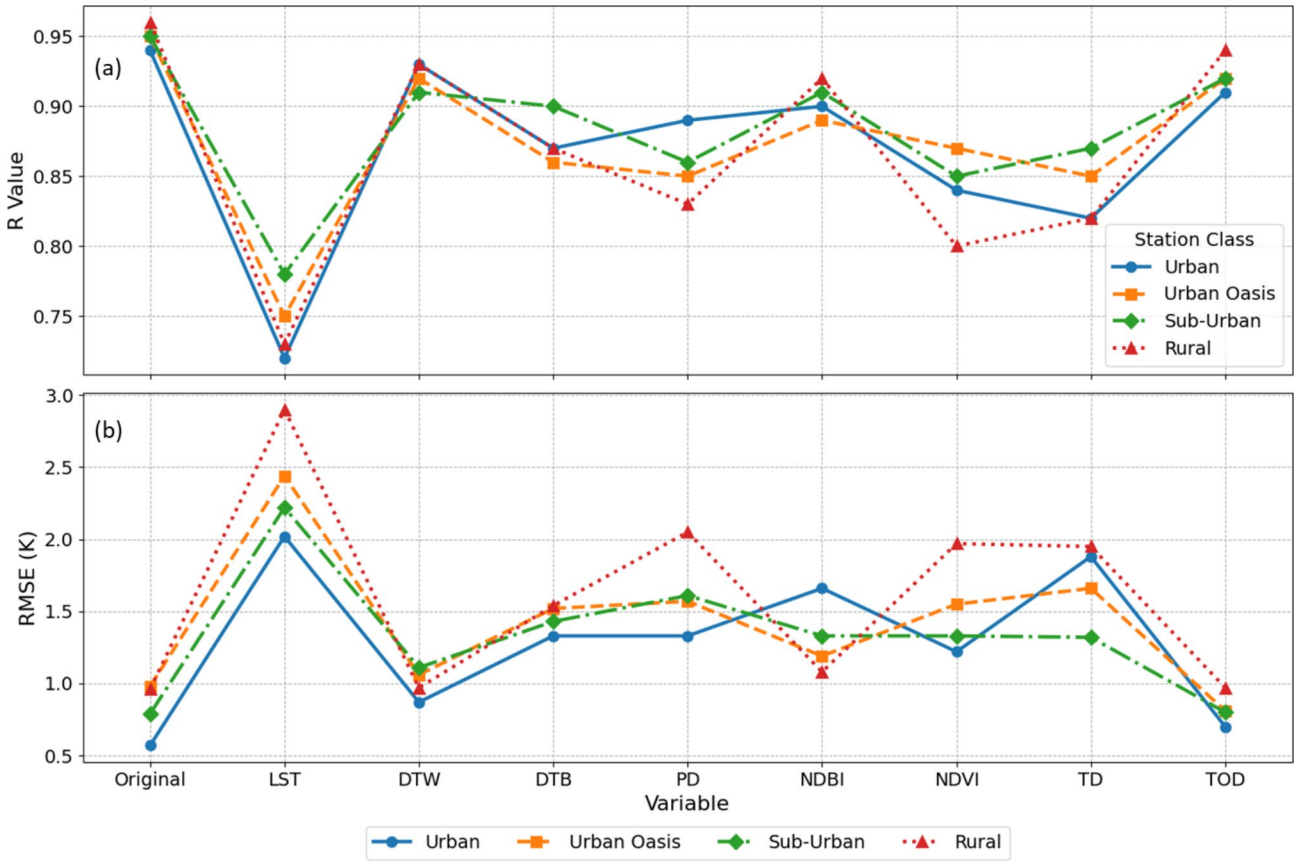

Fig. 8. Comparison of model accuracy with all predictor variables included versus accuracy with the removal of individual variables, evaluated using (**a**) the coefficient of determination (R) and (**b**) the root mean square error (RMSE).

started to increase once again. This diurnal analysis provides valuable insights into the dynamic nature of UHIs, highlighting the temporal variations in local temperature patterns within the urban landscape. Despite both rural and urban surfaces experiencing solar radiation at the same intensity, the prevalence of hard surfaces and sparse vegetation in the urban area contributes to a notable HI effect. This phenomenon is particularly evident in areas such as Sham Shui Po, Kowloon, and Hong Kong Airport, with the latter exhibiting the largest extent of the HI in the study area.

Urban Thermal Field Index (UTFI) analysis

Following the UHI analysis, the UTFI was computed to evaluate the thermal environment's quality and comfort levels for human life throughout the day in the study area (Fig. 11). The percentage of the region experiencing bad, worse, and worst thermal conditions (Table 4) increased from a combined 10% at 12:00 am (0:00) to 30% at 7:00 pm (19:00), at the expense of regions in normal, good, and excellent conditions, which decreased from a combined 85% at 12:00 am (00:00). The observed pattern exhibited a sinusoidal trend, with the percentage of pixels classified as bad, worse, and worst conditions increasing as the sun rose and decreasing as the sun set. Conversely, spatial extent of pixels classified as normal, good, and excellent conditions increased at sunset and decreased as the sun rose. This pattern can be attributed to the intricate interplay of land cover and urban development in the study in agreement with the study of Ullah, et al.[41]. Regions characterized by natural Land Use and Cover (LUC), featuring dense vegetation cover with minimal human development, exhibited normal, good, and excellent environmental conditions. In contrast, areas where vegetation cover has been replaced by hard surfaces (road construction, development of housing and other urban development) with scanty vegetation fell into regions characterized by bad, worse, and worst thermal conditions. This is due to the high emissivity of the hard surfaces, retaining solar radiations and contributing to the heating of the area. Anthropogenic activities, such as industrial processes, combustion of fossil fuels through driving, and heating, further exacerbate poor thermal conditions, especially during peak daytime hours.

Discussion

In this study, we developed an innovative framework that integrates remote sensing and machine learning for diurnal near-surface air temperature (T_{air}) prediction at a fine spatial resolution (100 m). This framework uniquely combines continuous spatial data with point station data, culminating in the development of a novel Artificial Neural Network model based on Federated Learning (FLANN). The FLANN model demonstrated

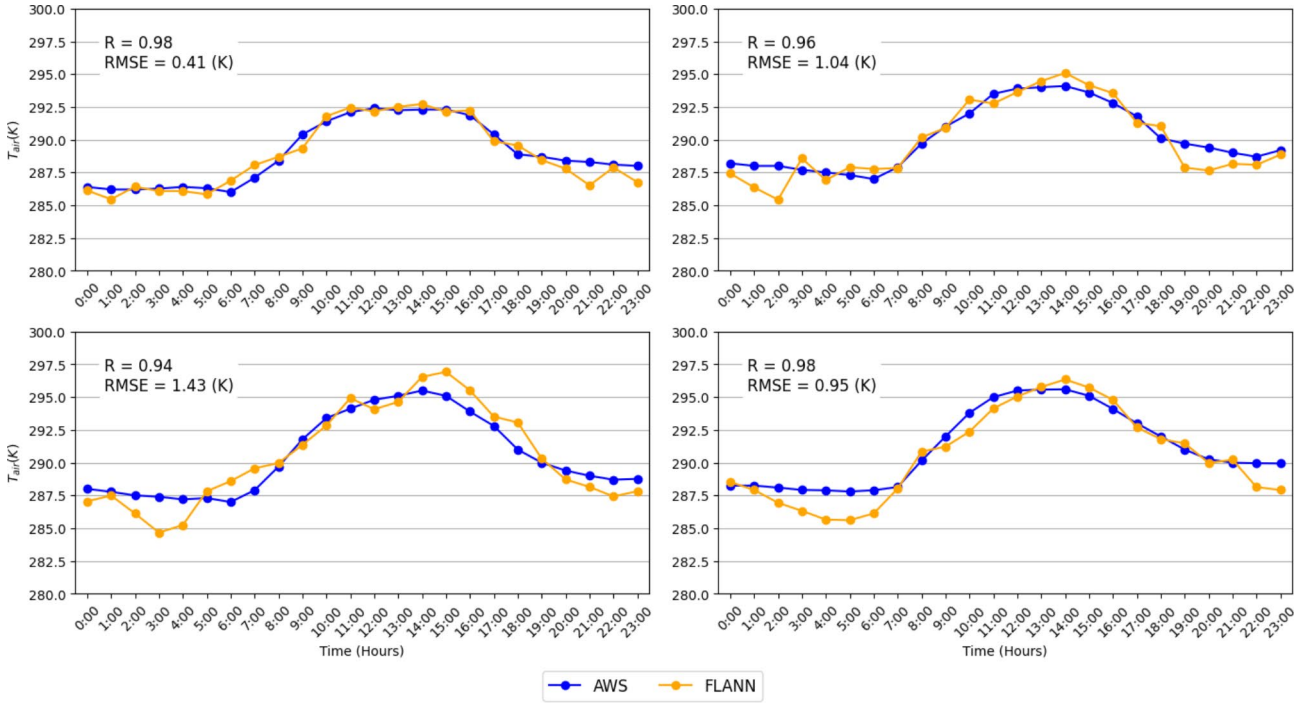

Fig. 9. Comparison between in-situ T_{air} from automatic weather stations (AWS) (blue color) and T_{air} predicted using FLANN (orange color) across (**a**) rural, (**b**) suburban, (**c**) urban oasis, and (**d**) urban stations on June 2, 2023.

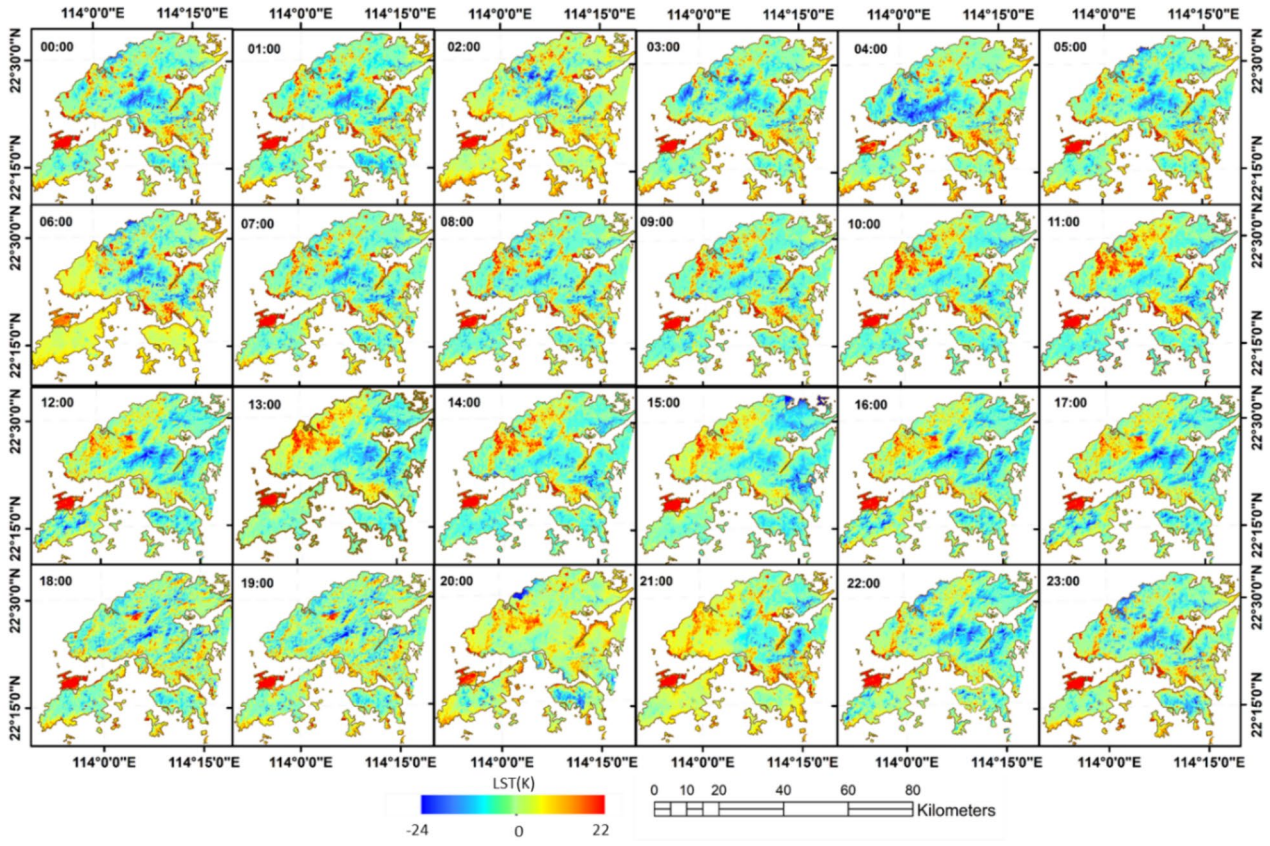

Fig. 10. Diurnal pattern of UHI in Hong Kong on June 2, 2023, across 24 h.

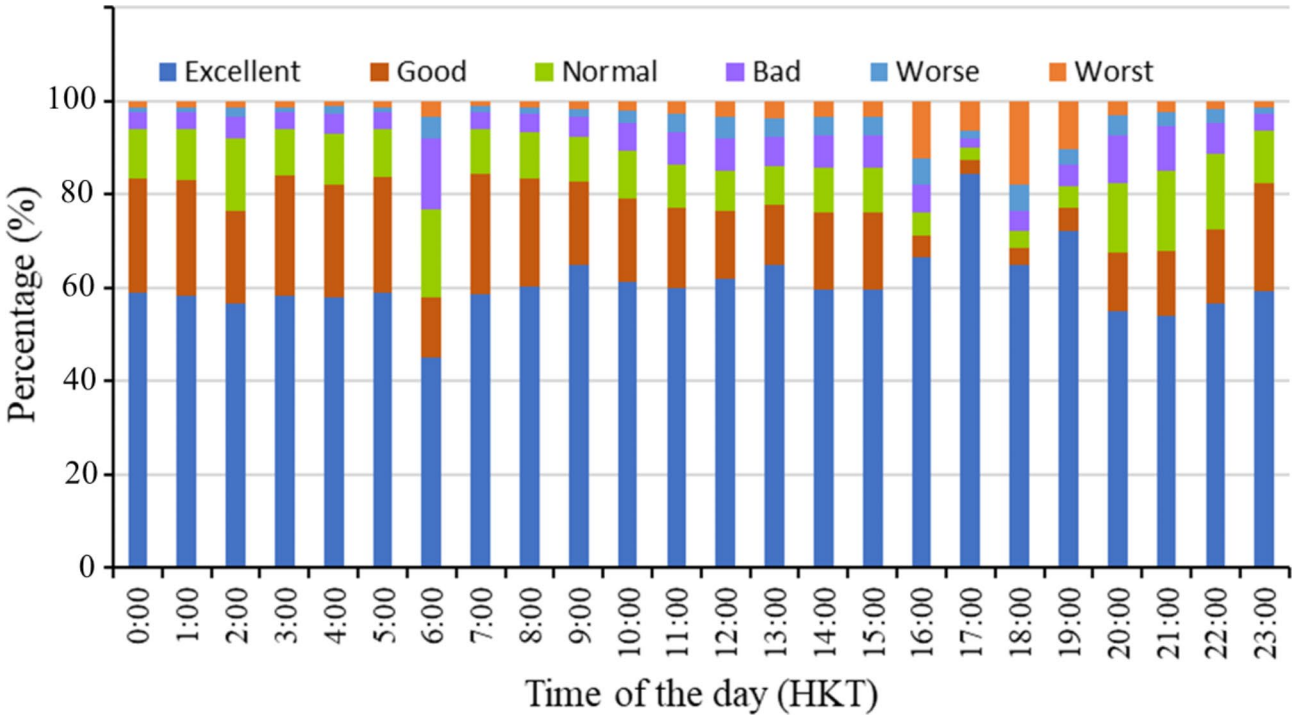

Fig. 11. Diurnal Urban Thermal Comfort over Hong Kong on a typical summer day (June 2, 2023).

superior performance compared to established methods, such as Multi-Layer Perceptron Artificial Neural Networks (MLPANN) and Random Forest (RF) models. Notably, in the urban oasis station class, the FLANN model achieved an R improvement of +0.04 over MLPANN and a 0.02 advantage over the RF model. This enhanced performance can be attributed to the segmented data processing and training intrinsic to the FL framework, which reduces the standard deviation of variables and optimizes model outcomes. Moreover, the segmented training structure ensures uniform model fitting across different station classes, regardless of variations in data density or unique station attributes that could otherwise lead to overfitting.

This study represents a significant advancement in T_{air} prediction methodologies by pioneering the use of the Federated Learning framework in this domain. Unlike traditional approaches, our framework introduces a novel classification layer before the federated learning process, where data extraction and preprocessing are conducted. This classification layer features an innovative data classification methodology based on the surrounding environmental conditions of each weather station. By classifying stations and optimizing data extraction methods for each class, the proposed framework accounts for unique environmental characteristics during model development. This stands in contrast to existing methods that often overlook the influence of the surrounding environment on station data, potentially leading to overfitting in stations that are overrepresented and underfitting in those that are underrepresented.

Our approach also advances the field by achieving fine-scale T_{air} prediction at an hourly temporal resolution through the optimization of Land Surface Temperature (LST) data fusion. By considering environmental variability across station classes and optimizing data extraction based on the most correlated buffer radius for each spatial variable, we have effectively addressed the variability in spatial resolution of weather stations due to surrounding environmental conditions. This comprehensive approach not only improves prediction accuracy but also ensures that the distinct spatial and temporal characteristics inherent in the different classes are adequately captured.

Employing the FLANN model for predicting T_{air} in Hong Kong provided significant insights into the diurnal patterns of UHI and urban thermal comfort. The model, tested by comparing T_{air} predictions with data collected from automatic weather station, demonstrated high accuracy (r ranging from 0.94 to 0.98 and RMSE between 0.41 and 1.43), making it highly effective for UHI assessment. Additionally, since the model was developed using the leave-station-out validation method, it is capable of generating T_{air} with similar accuracy for areas without weather stations. The continuous T_{air} values generated by the FLANN model for Hong Kong were used to analyze UHI patterns, revealing a dynamic landscape. Prominent Sink Island (SI) effects were observed in natural areas such as Lam Tsuen and Tai Mo Shan Country Park during nighttime, while Heat Island (HI) effects emerged in urban regions like Sham Shui Po, Kowloon, and Hong Kong Airport during the day. Furthermore, the Urban Thermal Field Index (UTFI) analysis highlighted a significant increase in regions experiencing poor thermal conditions as the day progressed, particularly in areas with dense urban development and minimal vegetation.

The implications of this study are profound, particularly in the context of urban temperature modelling. The ability to predict T_{air} at such a fine scale and high temporal resolution is crucial for urban climate studies, public

health, and infrastructure planning. The FLANN model's capacity to incorporate the distinct characteristics of various urban environments makes it particularly useful for cities with complex landscapes, such as Hong Kong. The study also underscores the importance of considering environmental factors in model development, which could lead to more accurate and reliable predictions in diverse urban settings.

While the study presents significant advancements, it also opens avenues for future research. The reliance on LST optimization from polar (Landsat-8/9, Sentinel-3) and geostationary (Himawari-8/9) satellites to generate hourly data highlights a limitation due to the lack of satellite sensors offering both fine temporal and spatial resolution. Future research could focus on optimizing other related variables (such as NDVI, NDBI, and NDWI) for diurnal variability to improve model accuracy. Additionally, the limited number of weather stations with T_{air} records in Hong Kong suggests that incorporating additional data sources, like the Community Weather Information Network (COWIN), could further enhance the model. Expanding the Federated Learning framework to other environmental variables or different regions could also provide broader validation of the methodology. Lastly, exploring alternative classification criteria for weather stations could offer new insights and further improve model performance.

Conclusion

This study applied a federated learning neural network (FLANN) framework to model near-surface temperature (T_{air}) from a group of continuous spatial data and point data. An innovative method was introduced for aggregating continuous data across different data points based on surrounding environmental conditions as modelled by local climate zone (LCZ) data. A total of 23 variables that correlate with T_{air} and have been used for its prediction in previous studies were subjected to autocorrelation and multicollinearity tests to select nine variables, which include land surface temperature (LST), distance to water (DTW), distance to building (DTB), population density (PD), normalized difference built-up index (NDBI), normalized difference vegetation index (NDVI), temperature difference (TD), day of year (DOY), and station class (SC), that were used for T_air prediction in this study. The performance of the FLANN model was compared to established machine learning (random forest (RF) and multilayer perceptron (MLPANN)) and linear (multiple linear regression (MLR)) models, and the FLANN model outperformed them, especially when the models were validated using station class-specific data. Following its improved performance, the FLANN model was thereafter employed for T_{air} prediction over Hong Kong on a typical summer day (June 2, 2023), where the resulting T_{air} data was employed for diurnal urban heat island (UHI) analysis. The dynamic spatial patterns of sink island (SI) and heat island (HI) underscored the temporal variability influenced by the solar cycle, land use, and human activities. Observations from the study highlighted the importance of considering diurnal factors in urban temperature studies. The evaluation of the urban thermal field index (UTFI) provided further insights into the thermal environment's quality. The transition from normal, good, and excellent conditions during the day to bad, worse, and worst conditions in the evening revealed the complex interplay between natural land cover and artificial surfaces. The diurnal analyses emphasized the temporal dynamics of UHI patterns, stressing the need for holistic approaches in urban climate studies.

Data availability

The satellite data which served as sources of the continuous data utilized in this study are openly accessible and sources have been thoroughly described in the method section while the station data sourced from government agency that are not publicly available are available from the corresponding author on reasonable request.

Received: 30 July 2024; Accepted: 30 October 2024
Published online: 09 November 2024

References

1. Mukherjee, F. & Singh, D. 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**, 385–407 (2020).
2. Kearl, Z. & Vogel, J. Urban extreme heat, climate change, and saving lives: lessons from Washington state. *Urban Clim.* **47**, 101392 (2023).
3. Abrahms, B. et al. Climate change as a global amplifier of human–wildlife conflict. *Nat. Clim. Change*. **13**, 224–234 (2023).
4. Dodman, D., Archer, D. & Satterthwaite, D. vol. 31, 3–12 (SAGE Publications Sage UK, 2019).
5. Lin, X. & Hubbard, K. Sensor and electronic biases/errors in air temperature measurements in common weather station networks. *J. Atmos. Ocean. Technol.* **21**, 1025–1032 (2004).
6. Vargas, J., Alsweiss, S., Toker, O., Razdan, R. & Santos, J. An overview of autonomous vehicles sensors and their vulnerability to weather conditions. *Sensors*. **21**, 5397 (2021).
7. Che, J. et al. Reconstruction of near-surface air temperature over the greenland ice sheet based on MODIS data and machine learning approaches. *Remote Sens.* **14**, 5775 (2022).
8. Coney, J. et al. How useful are crowdsourced air temperature observations? An assessment of Netatmo stations and quality control schemes over the United Kingdom. *Meteorol. Appl.* **29**, e2075 (2022).
9. Colaninno, N. & Morello, E. Towards an operational model for estimating day and night instantaneous near-surface air temperature for urban heat island studies: outline and assessment. *Urban Clim.* **46**, 101320 (2022).
10. Hooker, J., Duveiller, G. & Cescatti, A. A global dataset of air temperature derived from satellite remote sensing and weather stations. *Sci. Data*. **5**, 1–11 (2018).
11. Musashi, J. P., Pramoedyo, H. & Fitriani, R. Comparison of inverse distance weighted and natural neighbor interpolation method at air temperature data in Malang region. *CAUCHY: Jurnal Matematika Murni Dan. Aplikasi*. **5**, 48–54 (2018).
12. Ninyerola, M., Pons, X. & Roure, J. M. Objective air temperature mapping for the Iberian Peninsula using spatial interpolation and GIS. *Int. J. Climatol. J. R. Meteorol. Soc.* **27**, 1231–1242 (2007).
13. Zhao, W. & Duan, S. B. Reconstruction of daytime land surface temperatures under cloud-covered conditions using integrated MODIS/Terra land products and MSG geostationary satellite data. *Remote Sens. Environ.* **247**, 111931 (2020).

14. Li, X., Cheng, G. & Lu, L. Spatial analysis of air temperature in the Qinghai-Tibet Plateau. *Arct. Antarct. Alp. Res.* **37**, 246–252 (2005).
15. Li, Z. L. et al. Satellite-derived land surface temperature: current status and perspectives. *Remote Sens. Environ.* **131**, 14–37 (2013).
16. Krishnan, P. et al. Intercomparison of in situ sensors for ground-based land surface temperature measurements. *Sensors*. **20**, 5268 (2020).
17. Zhang, T., Zhou, Y., Wang, L., Zhao, K. & Zhu, Z. Estimating 1 km gridded daily air temperature using a spatially varying coefficient model with sign preservation. *Remote Sens. Environ.* **277**, 113072 (2022).
18. Zhang, Z., Liang, Y., Zhang, G. & Liang, C. Large-scale estimation of hourly surface air temperature based on observations from the FY-4A geostationary satellite. *Remote Sens.* **15**, 1753 (2023).
19. Yao, R. et al. Global seamless and high-resolution temperature dataset (GSHTD), 2001–2020. *Remote Sens. Environ.* **286**, 113422 (2023).
20. Naserikia, M. et al. Land surface and air temperature dynamics: the role of urban form and seasonality. *Sci. Total Environ.* **905**, 167306 (2023).
21. Shi, L. et al. Estimating daily air temperature across the Southeastern United States using high-resolution satellite data: a statistical modeling study. *Environ. Res.* **146**, 51–58 (2016).
22. Peng, X. et al. Correlation analysis of land surface temperature and topographic elements in Hangzhou, China. *Sci. Rep.* **10**, 10451 (2020).
23. Otgonbayar, M., Atzberger, C., Mattiuzzi, M. & Erdenedalai, A. Estimation of climatologies of average monthly air temperature over mongolia using MODIS land surface temperature (LST) time series and machine learning techniques. *Remote Sens.* **11**, 2588 (2019).
24. Deo, R. C. & Şahin, M. Forecasting long-term global solar radiation with an ANN algorithm coupled with satellite-derived (MODIS) land surface temperature (LST) for regional locations in Queensland. *Renew. Sustain. Energy Rev.* **72**, 828–848 (2017).
25. Benali, A., Carvalho, A., Nunes, J., Carvalhais, N. & Santos, A. Estimating air surface temperature in Portugal using MODIS LST data. *Remote Sens. Environ.* **124**, 108–121 (2012).
26. Zhao, W., Wen, F., Wang, Q., Sanchez, N. & Piles, M. Seamless downscaling of the ESA CCI soil moisture data at the daily scale with MODIS land products. *J. Hydrol.* **603**, 126930 (2021).
27. Adeniran, I. A., Zhu, R., Yang, J., Zhu, X. & Wong, M. S. Cross-comparison between sun-synchronized and geostationary satellite-derived land surface temperature: a case study in Hong Kong. *Remote Sens.* **14**. https://doi.org/10.3390/rs14184444 (2022).
28. Hong Kong Planning Department. Planning and Urban Design for a Liveable High-Density City. (2016).
29. Zhou, C., Zhang, D., Cao, Y., Wang, Y. & Zhang, G. Spatio–temporal evolution and factors of climate comfort for urban human settlements in the Guangdong–Hong Kong–Macau Greater Bay Area. *Front. Environ. Sci.* **10**, 1001064 (2022).
30. Zheng, Y. et al. Investigating the impact of weather conditions on urban heat island development in the subtropical city of Hong Kong. *Atmosphere*. **14**, 257 (2023).
31. Hong Kong Planning Department. *Land Utilization in Hong Kong* (2022). https://www.pland.gov.hk/pland_en/info_serv/open_data/landu/.
32. Adeniran, I. A. et al. Improved fusion model for generating hourly fine scale land surface temperature data under all-weather condition. *Int. J. Appl. Earth Obs. Geoinf.* **131**, 103981 (2024).
33. Araghi, A., Mousavi-Baygi, M., Adamowski, J., Martinez, C. & van der Ploeg, M. Forecasting soil temperature based on surface air temperature using a wavelet artificial neural network. *Meteorol. Appl.* **24**, 603–611 (2017).
34. McMahan, B., Moore, E., Ramage, D. & Hampson, S. & Y Arcas, B. A. In *Artificial Intelligence and Statistics* 1273–1282 (PMLR).
35. Lu, Y., He, T., Xu, X. & Qiao, Z. Investigation the robustness of standard classification methods for defining urban heat islands. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **14**, 11386–11394 (2021).
36. Hanberry, B. B. Imposing consistent global definitions of urban populations with gridded population density models: irreconcilable differences at the national scale. *Landsc. Urban Plann.* **226**, 104493 (2022).
37. Stewart, I. D. & Oke, T. R. Local climate zones for urban temperature studies. *Bull. Am. Meteorol. Soc.* **93**, 1879–1900. https://doi.org/10.1175/bams-d-11-00019.1 (2012).
38. Benitez, J., Henseler, J., Castillo, A. & Schuberth, F. How to perform and report an impactful analysis using partial least squares: guidelines for confirmatory and explanatory IS research. *Inf. Manag.* **57**, 103168 (2020).
39. Mohsin, T. & Gough, W. A. Characterization and estimation of urban heat island at Toronto: impact of the choice of rural sites. *Theoret. Appl. Climatol.* **108**, 105–117 (2012).
40. Naim, M. N. H. & Kafy, A. A. Assessment of urban thermal field variance index and defining the relationship between land cover and surface temperature in Chattogram city: a remote sensing and statistical approach. *Environ. Challenges*. **4**, 100107 (2021).
41. Ullah, N. et al. Spatiotemporal Impact of Urbanization on Urban Heat Island and Urban Thermal Field Variance Index of Tianjin City, China. *Buildings*. **12**, 399 (2022).

Acknowledgements

The authors would like to thank the Hong Kong Observatory (HKO) department for providing the hourly air temperature data from automatic weather stations used for his study. We also extend our appreciation to the United States Geological Survey (USGS), European Space Agency (ESA), and Japan Meteorological Agency (JMA) for the collection and free distribution of Landsat, Sentinel and Himawari 8 data, respectively. We would also like to thank the editors and anonymous reviewers who provided constructive comments on the manuscript.

Author contributions

I.A.A. performed the formal analysis and wrote the first draft of the manuscript, M.N. edited and reviewed the manuscript. M.S.W. supervised, acquired funding and reviewed the manuscript. P.-W.C. provided the weather stations data and reviewed the manuscript.

Funding

This research was supported by the General Research Fund (Grant Nos. 15603920 and 15609421), Collaborative Research Fund (Grant No. C5062-21GF) and Young Collaborative Research Fund (Grant No. C6003-22Y) from the Research Grants Council of Hong Kong, Research Institute for Sustainable Urban Development at the Hong Kong Polytechnic University (Project ID: 1-BBG2), Research Institute for Land and Space at the Hong Kong Polytechnic University (Project ID: 1-CD81).

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-78349-8.

Correspondence and requests for materials should be addressed to M.S.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.