

AN INTELLIGENT DATA-PROCESSING AND MACHINE LEARNING WORKFLOW FOR URBAN HEAT ASSESSMENT IN SMART CITIES

Nikolay Nikolov

University of Architecture, Civil Engineering and Geodesy, Sofia
Bulgaria,

Corresponding Author, e-mail: n.nikolov_fhe@uacg.bg

Abstract: Smart city management relies on practical and repeatable ways to turn spatial data into information that planners can use. Sofia, Bulgaria, was selected as a test area for the proposed satellite-based urban heat assessment. The work uses 14,400 grid observations prepared in Google Earth Engine for June 1–September 15, 2024. The dataset combines Landsat 8/9 land surface temperature, NDVI, NDBI, MODIS albedo and geographic coordinates. These variables were examined through correlation analysis and used for regression model comparison with train-test validation, five-fold cross-validation and spatial block cross-validation. Gradient Boosting achieved the best result with $R^2 = 0.886$, mean cross-validation $R^2 = 0.890 \pm 0.004$ and spatial block validation $R^2 = 0.830 \pm 0.035$. An additional 2023 check confirmed the same best-performing model but with lower validation scores, showing that model performance is year-dependent. The workflow supports early-stage identification of priority areas for green infrastructure.

Key words: smart cities, machine learning, remote sensing, urban heat, decision support.

1. INTRODUCTION

Urban heat is an important problem for many contemporary cities. It affects public spaces, energy demand, outdoor comfort and the quality of urban life. The problem is usually stronger in dense built-up areas, where vegetation is limited and artificial surfaces dominate the urban structure. The Urban Heat Island effect describes the tendency of urban areas to be warmer than their surrounding rural or semi-natural zones. This effect is related to changes in the urban energy balance, surface materials, heat storage, reduced evapotranspiration and human activity. The physical basis of this phenomenon was explained in [1].

Remote sensing provides a practical way to study urban heat at city scale. Satellite thermal data make it possible to observe spatial differences in land surface temperature and to identify areas where heat accumulates. Land surface temperature (LST) is not the same as air temperature, but it gives useful information about the thermal behaviour of

urban surfaces [2]. Vegetation and built-up surfaces are among the most important factors in surface heat analysis. Previous studies have shown that land surface temperature is closely related to vegetation abundance and impervious surface patterns [3, 4]. Green infrastructure is one of the practical responses to urban heat and can support climate adaptation when it is planned in suitable locations [5, 6].

In the context of smart cities, urban heat assessment can be supported by information technologies. Remote sensing data, spatial datasets and machine learning models allow large urban areas to be analysed in a systematic way. These methods do not replace detailed field studies, but they can provide an early screening tool for identifying heat-prone zones. This paper focuses on Sofia, Bulgaria. The study uses a grid-based remote sensing dataset with LST, normalized difference vegetation index (NDVI), normalized difference built-up index (NDBI) and albedo as an indicator of surface reflectivity.

The aim of this paper is to develop and test an application-oriented data-processing and machine learning workflow for urban heat assessment. The individual components used in the study are well established. The novelty of the paper is their practical integration into one reproducible workflow that connects cloud-masked satellite data extraction, grid-based harmonization, regression model comparison, validation, feature importance interpretation and priority classification for smart city decision support. The workflow is tested for Sofia, Bulgaria, during one summer season and should therefore be interpreted as a city-scale and season-specific application.

2. RELATED WORK

Satellite-derived indicators can be used as numerical inputs in geospatial data-processing workflows for modelling and decision support. In this paper, LST, NDVI, NDBI and albedo are used as the main input variables.

NDVI is based on the contrast between red and near-infrared reflectance and is widely used for vegetation monitoring [7]. NDBI is used to represent built-up surface signal in satellite imagery [8]. Landsat-based land surface temperature retrieval provides information about the thermal behaviour of urban surfaces [9]. Surface albedo is also relevant because it describes the reflected part of incoming solar radiation and supports the interpretation of surface reflectivity effects [10]. For smart city applications, the value of these indicators increases when they are integrated into a reproducible data-processing pipeline. A single satellite indicator can describe one surface property, but it cannot provide a complete decision-support output. For this reason, spatial variables need to be harmonized, structured into a common dataset and analysed through statistical and machine learning methods.

Machine learning methods are suitable for this type of task because urban surface temperature depends on several interacting variables. Regression models can be used to estimate land surface temperature from satellite-derived predictors and to compare the contribution of vegetation, built-up signal and albedo. Recent IJITS research confirms the relevance of statistical modelling and machine learning workflows for applied technical analysis and prediction tasks [11]. Hybrid statistical and machine learning modelling has also been used for predictive analysis in cyber-physical systems [12].

The reviewed works show that satellite-derived indicators and machine learning methods can support urban surface analysis. However, these elements are often presented as separate analytical steps. The gap addressed in this paper is the lack of a compact and reproducible workflow that links cloud-masked satellite extraction, grid-based data harmonization, regression model comparison, validation, feature importance and priority classification in one planning-oriented procedure. The proposed workflow is therefore incremental, but it adds practical value by connecting data processing and machine learning outputs with smart city decision support.

3. PROPOSED WORKFLOW AND DATASET

3.1. Study area

The workflow was tested using Sofia, Bulgaria, as an urban case study. The municipality covers 1,311 km² and includes dense urban districts, suburban settlements, agricultural land, forested areas, transport infrastructure and water bodies [13]. According to official statistical data, the population of Sofia Capital District was 1,295,931 inhabitants as of 31 December 2024 [14]. This demographic context supports the need for practical tools for smart city environmental monitoring.

3.2. Workflow structure

In this workflow, Google Earth Engine is used to extract the satellite-derived variables. The Landsat and MODIS layers are then combined in a common grid dataset and processed in Python. Statistical analysis and machine learning are applied to the prepared data. At the end of the procedure, the model outputs are converted into priority classes for decision support.

3.3. Empirical dataset

The empirical dataset comes from a Google Earth Engine (GEE) export for Sofia, Bulgaria. A rectangular area was used that encompasses the municipality of Sofia and the nearby suburban area. Its coordinates are [23.00, 42.45, 23.65, 42.90]. The export returned 14,400 grid-based point observations at 500 m spatial scale. In the final table, one row corresponds to one grid observation. These rows are analytical spatial records. They are not meteorological station measurements and they are not administrative units. The satellite observation period was from 1 June 2024 to 15 September 2024. This period was selected to represent the hot summer season, when surface urban heat effects are expected to be more pronounced. Landsat 8 and Landsat 9 Collection 2 Level-2 images were filtered by the study area, date range and cloud cover below 40%. Cloud, cloud shadow, cirrus, snow, fill and dilated cloud pixels were masked using the QA_PIXEL band. LST was derived from the Landsat ST_B10 band using the Collection 2 scale factor and offset and was converted from Kelvin to degrees Celsius [15]. NDVI was calculated from the red and near-infrared surface reflectance bands, while NDBI was calculated from the short-wave infrared and near-infrared bands. A median composite was then generated for LST_C, NDVI and NDBI over the selected summer period. The albedo layer was prepared from the MODIS MCD43A3 white-sky shortwave albedo product [16]. The values were averaged for the selected summer period, rescaled using

the 0.001 scale factor and integrated into the final grid-based data stack. Table 1 presents the variables included in the empirical dataset.

Table 1. Variables included in the empirical dataset

Variable	Unit / Range	Description
LST_C	°C	Median Landsat-derived land surface temperature for 1 June–15 September 2024
NDVI	Unitless, approx. -1 to +1	Vegetation greenness indicator calculated from NIR and red reflectance
NDBI	Unitless, approx. -1 to +1	Built-up surface indicator calculated from SWIR1 and NIR reflectance
Albedo	Unitless, 0 to 1	MODIS white-sky shortwave surface albedo
Latitude, Longitude	Decimal degrees	Geographic coordinates of the sample points

NDVI, NDBI and albedo were selected because they describe vegetation, built-up surfaces and surface reflectivity. These variables are available from open satellite products and can be extracted in Google Earth Engine in a reproducible way. The dataset does not include field air temperature, detailed land-cover classes, building morphology, tree canopy density or imperviousness density. For this reason, the workflow should be interpreted as an early-stage screening tool, not as a substitute for field measurements or detailed microclimate modelling. In this single-city and single-season study, qualitative climatic factors were not included as predictors because they are either constant for Sofia or require additional climatological data. Future multi-city or multi-season applications should combine such factors with the satellite-derived numerical variables.

All variables were harmonized to a common spatial grid before modelling. This step was necessary because Landsat and MODIS products have different native spatial resolutions. The final grid dataset should therefore be interpreted as an analytical spatial layer rather than as parcel-level or street-level measurement data.

NDVI and NDBI were calculated as normalized difference indices:

$$\text{NDVI} = (\text{NIR} - \text{RED}) / (\text{NIR} + \text{RED})$$

(1)

$$\text{NDBI} = (\text{SWIR1} - \text{NIR}) / (\text{SWIR1} + \text{NIR})$$

(2)

where NIR is the near-infrared reflectance, RED is the red band reflectance, and SWIR1 is the short-wave infrared reflectance from Landsat surface reflectance bands.

4. DATA PROCESSING AND MACHINE LEARNING METHODOLOGY

4.1. Data preparation and analysis

The empirical dataset was prepared as a numerical modelling table. Before modelling, it was checked for missing values and inconsistent records. No missing values were found in the final table. For model training, 70% of the observations were assigned to the training subset, while the remaining 30% were kept for testing. The test subset was not involved in fitting the models. It was used only to check how well the trained models predicted land surface temperature for unseen grid observations.

Descriptive statistics were calculated for LST_C, NDVI, NDBI and albedo. Pearson correlation was used to examine linear relationships, while Spearman correlation was used to check whether the relationships remained stable when ranking values were compared. The data processing and machine learning analysis were performed in Python. The workflow used pandas and NumPy for data preparation, scikit-learn for regression modelling and validation, and Matplotlib for visualisation. The same environment was used for descriptive statistics, correlation analysis, train-test validation, five-fold cross-validation and feature importance calculation.

4.2. Predictive models

A comparison was made with three regression models: Multiple Linear Regression, Random Forest Regression and Gradient Boosting Regression. Multiple Linear Regression was used as a transparent baseline model. Random Forest Regression and Gradient Boosting Regression were used as nonlinear ensemble models. They were included because they can represent nonlinear relationships between the satellite-derived predictors and land surface temperature. The same input variables were used in all three models: NDVI, NDBI and albedo. This made the comparison consistent. The models were run with the standard scikit-learn settings, without an additional tuning stage. Hyperparameter tuning was not the main objective, because the study focuses on a transparent and reproducible workflow rather than maximum accuracy for one dataset. Future applications may include grid or randomized search for parameters such as tree depth, number of trees, learning rate and minimum leaf size.

4.3. Model validation

Model validation was carried out using three procedures. First, a 70/30 train-test split was applied, where the test subset was used only for evaluating predictions on unseen grid observations. Second, five-fold cross-validation was used as a stability check, with MAE, RMSE and R^2 reported as mean and standard deviation values. Third, spatial block cross-validation was applied because the dataset has an explicit spatial structure. Neighbouring grid observations may have similar land-cover and thermal characteristics, which can make random validation overly optimistic. For this procedure, the 14,400 observations were grouped into five spatial blocks using standardized latitude and longitude values. In each run, one block was used for testing and the remaining four blocks for training. This provided a more conservative estimate of model performance under spatial separation.

4.4. Priority index

After model evaluation, the results were transformed into a simple priority index. The index combines standardized land surface temperature, NDBI and NDVI. High temperature and high built-up signal increase the priority score. High vegetation signal decreases it.

$$\text{Priority} = \text{LST_z} + \text{NDBI_z} - \text{NDVI_z} \quad (3)$$

Albedo was not included in the priority index because the index was designed as a simple planning-oriented screening indicator focused on heat intensity, built-up surface

signal and vegetation deficit. Albedo was retained in the predictive modelling stage because it contributes to the interpretation of surface reflectivity effects.

An equal-weight index was used because the aim was to keep the priority score simple and understandable for planning use. In this formulation, LST_z reflects surface heat intensity, NDBI_z adds the built-up surface component, and NDVI_z reduces the score where vegetation is stronger. Other weighting schemes could also be tested, for example model-based weights or multi-criteria assessment, but they would require additional assumptions and input from local planning stakeholders.

The index is a screening tool. Cells with high priority values should be examined in more detail through field assessment, land ownership analysis and local planning constraints. Cloud masking, sensor noise, within-cell variation and seasonal differences may influence the exact ranking. Future work should test the stability of the priority classes with multi-year data and alternative thresholds.

5. RESULTS AND DISCUSSION

5.1. Descriptive statistics and correlations

The empirical analysis was based on 14,400 grid observations. Table 2 presents the descriptive statistics of the main variables.

Table 2. Descriptive statistics and correlations with LST

Variable	Mean	Median	Min	Max	Std	Pearson r	Spearman rho
LST_C, °C	34.50	34.84	20.51	54.42	5.58	-	-
NDVI	0.639	0.665	-0.058	0.892	0.182	-0.814	-0.871
NDBI	-0.205	-0.212	-0.518	0.267	0.137	0.887	0.891
Albedo	0.158	0.155	0.022	0.216	0.024	0.724	0.749

The descriptive statistics show a wide variation in land surface temperature across the grid. The mean value of 34.50 °C should be interpreted as a surface temperature indicator for the analysed satellite dataset, not as air temperature for Sofia. LST should not be treated as a direct substitute for near-surface air temperature. It describes the thermal behaviour of urban surfaces at the time of satellite observation, while air temperature is also affected by wind, humidity, shading, urban geometry and measurement height. Therefore, the workflow is mainly suitable for detecting surface heat exposure and should be combined with meteorological or field data for operational heat-health applications.

NDVI has a mean value of 0.639, which indicates that many grid cells have moderate or strong vegetation signal. NDBI has a negative mean value of -0.205 because the dataset includes both urban and non-urban surfaces. Albedo has a mean value of 0.158.

NDVI has a strong negative relationship with LST. NDBI has a strong positive relationship with LST. Albedo also has a positive correlation in this dataset. This result should be interpreted carefully because bright or exposed artificial surfaces may still become hot when they are dry, unshaded or strongly built-up.

All correlation coefficients reach statistical significance at $p < 0.001$. In this dataset, NDBI shows the strongest relationship with LST, followed by NDVI. This result points to a clear contrast between built-up and vegetated areas in Sofia and shows why both indicators are important for interpreting the surface heat pattern.

5.2. Proxy comparison and model performance

A proxy comparison was used to estimate the thermal contrast between urban-like and reference-like cells. Urban-like cells were defined as cells with $NDBI > 0$ or $NDVI < 0.30$. Reference-like green cells were defined as cells with $NDVI > 0.75$ and $NDBI < -0.25$. The urban-like group included 1,299 cells with mean $LST = 41.13\text{ }^{\circ}\text{C}$, while the reference-like green group included 5,230 cells with mean $LST = 28.77\text{ }^{\circ}\text{C}$. The resulting proxy surface heat difference was $12.36\text{ }^{\circ}\text{C}$. This value should not be interpreted as an official SUHI value because the classification is based on spectral thresholds, but it shows a clear thermal contrast between built-up and green areas.

Three regression models were tested on a 30% test subset. The results are presented in Table 3.

Table 3. Model performance comparison

Model	MAE, $^{\circ}\text{C}$	RMSE, $^{\circ}\text{C}$	R^2	Interpretation
Multiple Linear Regression	1.54	2.04	0.864	Strong baseline
Random Forest Regression	1.42	1.90	0.883	Improved nonlinear model
Gradient Boosting Regression	1.40	1.87	0.886	Best result

The linear model already explains a large part of the variation in land surface temperature. The two nonlinear models provide a moderate improvement, with Gradient Boosting Regression achieving the best test result: $MAE = 1.40\text{ }^{\circ}\text{C}$, $RMSE = 1.87\text{ }^{\circ}\text{C}$ and $R^2 = 0.886$. Table 4 shows the results of five-fold cross-validation.

Table 4. Five-fold cross-validation results

Model	MAE mean $\pm$ std	RMSE mean $\pm$ std	R^2 mean $\pm$ std
Multiple Linear Regression	1.549 ± 0.013	2.035 ± 0.028	0.866 ± 0.004
Random Forest Regression	1.397 ± 0.021	1.872 ± 0.035	0.886 ± 0.005
Gradient Boosting Regression	1.378 ± 0.017	1.839 ± 0.031	0.890 ± 0.004

The five-fold cross-validation results confirm the same model ranking. Gradient Boosting Regression achieved the best average performance, followed by Random Forest Regression and Multiple Linear Regression. The low standard deviations show that the ranking is stable and does not depend only on one random train-test split. Table 5 shows the results of spatial block cross-validation.

Table 5. Spatial block cross-validation results

Model	MAE mean $\pm$ std	RMSE mean $\pm$ std	R ² mean $\pm$ std
Multiple Linear Regression	1.647 $\pm$ 0.279	2.168 $\pm$ 0.368	0.823 $\pm$ 0.027
Random Forest Regression	1.664 $\pm$ 0.208	2.223 $\pm$ 0.340	0.812 $\pm$ 0.031
Gradient Boosting Regression	1.583 $\pm$ 0.225	2.114 $\pm$ 0.374	0.830 $\pm$ 0.035

Compared with standard five-fold cross-validation, spatial block validation produced more conservative results. Gradient Boosting Regression remained the best model, with MAE = 1.583 °C, RMSE = 2.114 °C and R² = 0.830 $\pm$ 0.035. The lower performance confirms that spatial autocorrelation can influence random validation results, but also shows that the selected satellite-derived predictors remain useful under spatial separation.

To partially address the temporal limitation of the study, the same workflow was repeated for Sofia for the summer period from 1 June to 15 September 2023. In both years, Gradient Boosting Regression was the best-performing model. The 2023 results were lower than those for 2024, which shows that model performance is year-dependent and should be checked with multi-year data. Table 6 presents temporal robustness check for Sofia.

Table 6. Temporal robustness check for Sofia

Period	Observations	Test R ²	Random CV R ²	Spatial block CV R ²
1 Jun–15 Sep 2024	14,400	0.886	0.890 $\pm$ 0.004	0.830 $\pm$ 0.035
1 Jun–15 Sep 2023	14,400	0.840	0.847 $\pm$ 0.006	0.665 $\pm$ 0.135

5.3. Priority classes and interpretation

The final step was to transform the empirical results into a planning-oriented indicator. Cells with high temperature, high built-up signal and low vegetation signal receive higher priority scores. The priority index was divided into four classes: low priority, 50% of cells; medium priority, 25%; high priority, 15%; and very high priority, 10%. The very high priority class includes 1,440 cells and marks the areas that should be examined first in a detailed planning review.

The very high priority cells should not be treated as automatic locations for intervention. Possible actions may include tree planting, shading measures, small green spaces, cool surfaces or combined design solutions. The chosen method would need to be tailored to local circumstances, ownership of the land, and underground utilities. High-priority cells require additional checks related to field conditions, land ownership, local planning rules and underground infrastructure. Cloud masking, sensor noise, within-cell variation and seasonal differences may influence the exact ranking, so the stability of the priority classes should be tested with multi-year data and alternative thresholds.

The workflow can be used as a lightweight analytical module in smart city monitoring systems or urban digital twins. It relies on open satellite products and a grid-

based structure, but direct transfer to other cities should not be assumed without local recalibration and validation. Transferability depends on cloud-free observations, land-cover structure, urban morphology, seasonal conditions and complementary urban or meteorological data.

6. CONCLUSION

This paper presented a reproducible data-processing and machine learning workflow for satellite-derived urban heat assessment in smart cities. The workflow was tested for Sofia, Bulgaria, using 14,400 grid observations with land surface temperature, NDVI, NDBI, albedo and geographic coordinates.

The results show a clear relationship between surface heat, vegetation and built-up surface signal. NDVI has a strong negative relationship with LST, while NDBI has a strong positive relationship. Gradient Boosting Regression achieved the best predictive performance, with $R^2 = 0.886$ on the test subset and $R^2 = 0.890 \pm 0.004$ in five-fold cross-validation. Under the stricter spatial block validation, the same model remained the best performer, with $R^2 = 0.830 \pm 0.035$.

The main contribution is the integration of satellite-derived indicators, machine learning validation and a planning-oriented priority index into one application-oriented workflow. The results support early-stage smart city analysis and green infrastructure prioritization, but they should be interpreted as a city-scale and season-specific validation. The additional 2023 robustness check showed the same best-performing model, but lower validation scores, indicating that the workflow is applicable across years while model performance remains sensitive to annual conditions. Future work should include multi-year data, additional urban layers and comparisons across different cities and climatic contexts.

REFERENCES

- [1] Oke, T.R. The energetic basis of the urban heat island. *Quarterly Journal of the Royal Meteorological Society*, vol.108, no.455, 1982, pp. 1-24. DOI: 10.1002/qj.49710845502.
- [2] Voogt, J.A., Oke, T.R. Thermal remote sensing of urban climates. *Remote Sensing of Environment*, vol.86, no.3, 2003, pp. 370-384. DOI: 10.1016/S0034-4257(03)00079-8.
- [3] Weng, Q., Lu, D., Schubring, J. Estimation of land surface temperature-vegetation abundance relationship for urban heat island studies. *Remote Sensing of Environment*, vol.89, no.4, 2004, pp. 467-483. DOI: 10.1016/j.rse.2003.11.005.
- [4] Yuan, F., Bauer, M.E. Comparison of impervious surface area and normalized difference vegetation index as indicators of surface urban heat island effects. *Remote Sensing of Environment*, vol.106, no.3, 2007, pp. 375-386. DOI: 10.1016/j.rse.2006.09.003.
- [5] Gill, S.E., Handley, J.F., Ennos, A.R., Pauleit, S. Adapting cities for climate change: the role of the green infrastructure. *Built Environment*, vol.33, no.1, 2007, pp. 115-133. DOI: 10.2148/benv.33.1.115.
- [6] Bowler, D.E., Buyung-Ali, L., Knight, T.M., Pullin, A.S. Urban greening to cool towns and cities: a systematic review of the empirical evidence. *Landscape and Urban Planning*, vol.97, no.3, 2010, pp. 147-155. DOI: 10.1016/j.landurbplan.2010.05.006.

- [7] Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, vol.8, no.2, 1979, pp. 127-150. DOI: 10.1016/0034-4257(79)90013-0.
- [8] Zha, Y., Gao, J., Ni, S. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing*, vol.24, no.3, 2003, pp. 583-594. DOI: 10.1080/01431160304987.
- [9] Sobrino, J.A., Jimenez-Munoz, J.C., Paolini, L. Land surface temperature retrieval from LANDSAT TM 5. *Remote Sensing of Environment*, vol.90, no.4, 2004, pp. 434-440. DOI: 10.1016/j.rse.2004.02.003.
- [10] Akbari, H., Pomerantz, M., Taha, H. Cool surfaces and shade trees to reduce energy use and improve air quality in urban areas. *Solar Energy*, vol.70, no.3, 2001, pp. 295-310. DOI: 10.1016/S0038-092X(00)00089-X.
- [11] Stanchev, P., Hinov, N. Statistical and AI-based reliability assessment of photovoltaic inverters using statistical modeling and machine learning. *International Journal on Information Technologies and Security*, vol.17, no.3, 2025, pp. 37-48. DOI: 10.59035/AGCX7287.
- [12] Vasilev, I. A hybrid “ARIMA–ML regression” model for enhanced predictive analysis in cyber-physical systems: Conceptual framework and simulation evaluation. *International Journal of Current Science Research and Review*, vol.9, 2026, pp. 2232-2243. DOI: 10.47191/ijcsrr/V9-i5-02.
- [13] Sofia Municipality. Geographical characteristics of Sofia Municipality. Official municipal information page. URL: <https://www.sofia.bg/en/web/sofia-municipality/geographical-feature>. Visited on 10.05.2026.
- [14] National Statistical Institute of Bulgaria. District Sofia (Stolitsa): Regional statistical indicators, population as of 31.12.2024. National Statistical Institute, 2025. URL: <https://www.nsi.bg/en/statistical-data/46/172>. Visited on 10.05.2026.
- [15] Earth Resources Observation and Science Center. Landsat 8-9 Operational Land Imager / Thermal Infrared Sensor Level-2, Collection 2. U.S. Geological Survey, 2020. DOI: 10.5066/P9OGBGM6.
- [16] Schaaf, C., Wang, Z. MODIS/Terra+Aqua BRDF/Albedo Daily L3 Global - 500m V061. NASA Land Processes Distributed Active Archive Center, 2021. DOI: 10.5067/MODIS/MCD43A3.061.

Information about the author:

Nikolay Nikolov - PhD, Assistant Professor at the University of Architecture, Civil Engineering and Geodesy, Sofia, Bulgaria. His research interests include smart cities, intelligent systems, sustainable urban environments, engineering education and data-driven modelling.

Manuscript received on 29 May 2026