

# Predicting Urban Heat Island severity on the census-tract level using Bayesian networks

Ghiwa Assaf, Xi Hu, Rayan H. Assaad\*

Smart Construction and Intelligent Infrastructure Systems (SCIIS) lab, John A. Reif, Jr. Department of Civil and Environmental Engineering, New Jersey Institute of Technology, Newark, NJ 07102, United States



## ARTICLE INFO

### Keywords:

UHI severity  
Urban climate  
Bayesian networks  
White-box machine learning  
Forecasting models  
Census-tract predictions

## ABSTRACT

Urban development and population growth have resulted in several critical impacts on the society, environment, and economy. One of the main impacts is the increase in temperature observed in urban areas, also known as the Urban Heat Island (UHI) effect. UHI has become the focus of several research studies due to the associated negative implications it causes. Despite that some research efforts examined different characteristics of the UHI phenomenon, there is a gap in the literature in terms of developing interpretable machine learning models that can accurately estimate or predict the severity of UHI (rather than air temperature or other UHI characteristics) as well as that can provide predictions on the census-tract level in the US. To that extent, this paper fills this gap by developing a knowledge-based white-box Bayesian network model that predicts UHI severity based on demographic, meteorological, and land use/land cover factors. First, a dataset for all census tracts in the state of New Jersey, USA was developed, which is comprised of 13 independent variables or factors affecting UHI severity. Second, expert knowledge was obtained from 10 UHI experts using the systematic three-round Delphi method to develop four different Bayesian network models. Third, the performance of the four Bayesian models was assessed and compared to choose the optimal model with the highest accuracy. Finally, sensitivity analysis was conducted to assess the influence of each key factor on the UHI severity. The results showed that the optimal model is a tree-augmented Bayesian network that can predict or estimate the UHI severity with an accuracy of 87.88%. The outcomes of this paper also reflected that the following 8 variables are the key factors that impact UHI severity: NDVI during winter season, NDVI during summer season, imperviousness, tree canopy, building area, population density, water body areas, and annual rainfall. The findings also identified the land use/land cover category as the major category affecting UHI severity compared to demographic- and meteorological-related factors. The proposed white-box Bayesian network model in this paper adds to the body of knowledge by allowing practitioners and researchers to perform micro-level UHI predictions and inferences about future or unknown UHI severity pertaining to individual communities or small/localized geographical areas. This enables more focused and targeted UHI mitigation actions to be planned, designed, and implemented in the most affected communities, which ultimately results in better decisions, actions, and outcomes to reduce or address the UHI effect.

## 1. Introduction

Urbanization has led to having numerous cities or urban clusters having high coverage of impervious or rough surfaces, excessive release of heat and pollution, and limited vegetation (Abbassi, Ahmadikia & Baniasadi, 2022; Assaf & Assaad, 2023c). These characteristics make cities more prone to heat waves and increased temperatures, also known as the Urban Heat Island (UHI) effect (Hidalgo, Masson, Baklanov, Piégeon & Gimeno, 2008). The UHI is an expression used to represent the

impact that mankind activities and urban development have on the air and surface temperatures in terms of having higher temperatures as compared to their rural surroundings (Assaf & Assaad, 2023d; Hou et al., 2022; Magee, Curtis & Wendler, 1999; Weng, Lu & Schubring, 2004). Other researchers defined the UHI effect as a high temperature dome covering urban areas and cities (Comarazamy, González, Luvall, Rickman & Picón, 2006).

The UHI can be quantified as the difference between the urban temperature and the rural temperature; where most studies defined the rural temperature as the temperature at a 5 km or 10 km buffer from the

\* Corresponding author.

E-mail address: [rayan.hassane.assaad@njit.edu](mailto:rayan.hassane.assaad@njit.edu) (R.H. Assaad).

Nomenclature	
UHI	Urban Heat Island
NDVI	Normalized Difference Vegetation Index
XGB	Extreme Gradient Boosting
LULC	Land Use Land Cover
LST	Land Surface Temperature
AUHI	Air Urban Heat Island
SUHI	Surface Urban Heat Island
LCZ	Local Climate Zones
ZTD	Zenith Tropospheric Delay
GNSS	Global Navigation Satellite System
MAT	Mean Annual Temperature
AR	Annual Rainfall
P	Population
PD	Population density
WBA	Waterbody Area
NVDI_a	Normalized Difference Vegetation Index for urbanized areas during the entire year
NVDI_as	Normalized Difference Vegetation Index for urbanized areas during the summer season of a year
NVDI_aw	Normalized Difference Vegetation Index for urbanized areas during the winter season of a year
BA	Building area
TRL	Total Road Length
BH	Building Height
I	Imperviousness
TC	Tree Canopy
MCD	Minimum Covariance Determinant
LOF	Local Outlier Factor
SVM	Support Vector Machine
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
STD	Standard Deviation
Q <sub>1</sub>	First Quartile
Q <sub>2</sub>	Second Quartile
Q <sub>3</sub>	Third Quartile
c	The percentage of anomalies in the raw dataset
n	The number of neighbors to a data sample with outliers
kernel	The kernel type of support vector machine classifier
m	The type of skewed data methods
nb	The number of intervals a variable is discretized into
SVF	Sky View Factor
NJ	New Jersey

urban area (Azhdari, Soltani & Alidadi, 2018). The UHI has several causes such as the anthropogenic heat or the heat generated from human activities which is considered as the major causal effect of UHI. Anthropogenic heat include heat emitted from vehicles, buildings, industrial facilities, and air conditioners (Kim, Yeo & Kim, 2022; Shah-mohamadi, Che-Ani, Maulud, Tawil & Abdullah, 2011). Another major cause of UHI is the reduced natural landscape in the cities such as green areas and water bodies which have the ability to cool the air (Jiang, 2021). The change in urban geometry (e.g., spacings and dimensions of buildings) is also one of the causes of UHI since it affects the wind flow and increases the surfaces absorption of solar energy (Maharjan et al., 2021). Finally, the adapted man-made materials used in urban development (e.g., asphalt and concrete) are one of the main causes of the UHI since they are characterized by low ability to reflect, and high ability to store, solar energy (Bohenstengel, Evans, Clark & Belcher, 2011).

The UHI phenomenon is expected to exacerbate and become more prevalent in the future (Kim & Brown, 2021; Lauwaet et al., 2016). The extreme diurnal UHI severity has several negative and severe implications on the environment, human health, and well-being. For instance, the extreme temperatures during the day and associated heat waves (i.e., diurnal UHI) could lead to numerous respiratory, cardiovascular, and other chronic diseases, especially for old and poor people (Jacob & Winner, 2009; Patz, Campbell-Lendrum, Holloway & Foley, 2005). In addition to that, high temperatures could also increase fatality, decrease worker's productivity, increase air pollution levels, and escalate cooling loads (Deschenes & Moretti, 2009; Guan et al., 2014; Kjellstrom, Holmer & Lemke, 2009; Lu, Fu, Dewan & Li, 2023). For instance, the impact of UHI on increasing electrical cooling loads has been estimated to be 1,000,000 \$/hour in the United States (Akbari, 2009). Moreover, a 2.0 °C increase in the temperature in China has led to the increase in the annual heat-related mortality from 32.1 per million people in the year 1986 to 81.3 per million people in the year 2005 (Li et al., 2022). Those severe impacts of the UHI have made it one of the most studied phenomena related to climate change since the increase in the UHI intensity and extreme heats lead to the exacerbation of heat vulnerabilities and risks in urban areas (Lu et al., 2023). Thus, exploring the UHI effect, the factors that affect it, as well as its mitigation measures to reduce heat risks as well as promote sustainable and resilient cities have become

indispensable for promoting sustainable societies (Li et al., 2022). Such an exploration will be essential to enable decision-makers and city planners gain an enhanced understanding of the UHI phenomena and thus manage the UHI based on informative decisions (Berger et al., 2017).

Most of the existing research, that was devoted to study different aspects of UHI, were primarily based on satellite images or field data collected from weather stations and sensors. Recently, integrating machine learning algorithms to quantify, estimate, or predict different characteristics related to the UHI effect has been gaining momentum because it helps identifying patterns and relationships related to the UHI phenomenon. In relation to that, most of the existing studies have used supervised black-box machine learning algorithms such as artificial neural networks, random forests, support vector regression, Extreme Gradient Boosting (XGB) regression, among others.

Starting with artificial neural networks, this algorithm was used in several studies to examine different characteristics of UHI (Mirzaei, 2015). For instance, the artificial neural networks algorithm was used to analyze the impact of factories' characteristics such as the shape, type, production stage, scale, and internal structure on the UHI effect and to reveal the optimal characteristics that minimize the UHI effect (Liu et al., 2021). In their study, Liu et al. (2021) concluded that the optimal factories' characteristics that reduce the UHI impact are having a scale of 1.13 km<sup>2</sup> and a shape index between 1.11 and 3.13. Also, artificial neural networks based on the long short-term memory model were used to gain insights and predict the highly fluctuating UHI effect based on the air temperature as a response variable (Yun et al., 2020). Artificial neural networks were also implemented to predict the spatial distribution of the UHI effect (Equerre, Mirzaei, Riffat & Wang, 2021) and the peak energy loads during extreme temperatures events to efficiently manage the supply and demand loads (O'Malley, Piroozfar, Farr & Pomponi, 2015). Moreover, neural networks were also used with other machine learning algorithms. For instance, Mohammad, Goswami, Chauhan and Nayak (2022) used artificial neural networks and XGB, where the artificial neural network using the cellular automata model was used to predict the Land Use Land Cover (LULC), and the XGB regression algorithm was implemented to predict the summer and winter Land Surface Temperature (LST) based on spatial and temporal data for the change of LULC and seasonal LST. Moreover, Vulova, Meier,

Fenner, Nouri and Kleinschmit (2020) attempted to estimate the nocturnal air Urban Heat Island (AUHI) by predicting the nocturnal air temperature using three machine learning algorithms: model averaged neural networks, random forests, and stochastic gradient boosting. By comparing the performance of the three algorithms, Vulova et al. (2020) found that random forest was the most performing model in predicting nocturnal air temperature.

In addition to artificial neural networks, the random forest algorithm was also one of the commonly used algorithms for studying UHI-related aspects. In relation to that, the random forest algorithm was used to explore the relation between the morphological spatial pattern of green spaces and UHI phenomenon (Lin, Qiu, Tan & Zhuang, 2023). Using the built random forests, Lin et al. (2023) concluded that the density of core green spaces is the most contributing factor to UHI mitigation followed by the density of perforation and loop. Moreover, Chen et al. (2022) also used random forests to predict the air temperature and the AUHI using geometric, physical, and anthropogenic parameters and found that, under calm wind conditions, the AUHI is most correlated with the anthropogenic heat flux and built-up area factors. The random forests algorithm was also used to predict the SUHI intensity and distribution in year 2030 based on the prediction of landscape patterns (Shen et al., 2022). Furthermore, Yoo (2018) implemented random forests to specify the urban physical and socioeconomic factors that significantly affect the UHI effect and found that the percent of impervious surfaces and NDVI are the most correlated factors with the UHI effect. Yoo (2018) also concluded that, in sub-urban dwelling districts, the age-related factor could be considered as one of the most correlated factors with the UHI effect; however, in urban dwelling districts, socioeconomic factors such as ethnicity, educational attainment, and household level economic status were found to be the major influential factors. Additionally, Oliveira, Lopes, Niza and Soares (2022) also used random forests to predict the nocturnal surface Urban Heat Island (SUHI) during a heat wave event based on the Local Climate Zones (LCZ) and geostationary satellite data. Oukawa, Krecl and Targino (2022) also utilized random forest models and multiple linear regression to analyze and predict the spatiotemporal evolution of the UHI effect using the air temperature as the response variable. Furthermore, the random forests algorithm was implemented to determine the importance of factors that influence park cooling intensity and identified that the park area and green areas evapotranspiration are the most influential factors (Gao, Zaitchik, Hou & Chen, 2022). The random forests were also used to identify the nonlinear daytime and nighttime impacts of urban and industrial forms on the SUHI and disclosed that the industrial factors contribute to 43.38% of the daytime SUHI variance while the urban factors contribute to 20.44% of the nighttime SUHI variance (Ming, Liu, Gu, Wang & Liu, 2023).

Support vector regressor and Ridge regression algorithms were also implemented to enhance the knowledge of different UHI-related aspects. In relation to that, Mendez-Astudillo et al. (2021) used support vector regressor to predict the air temperature from zenith tropospheric delay (ZTD) data collected from the global navigation satellite system (GNSS); where ZTD relies on some environmental factors such as water vapor pressure, temperature, and pressure collected along the signal path emitted from the satellite to the receptor. In their study, Mendez-Astudillo et al. (2021) trained the model using ZTD data and used the predicted air temperature to analyze the cycle of diurnal AUHI. Moreover, the Ridge regression model was also implemented to identify the impact of urban fabric such as building coverage ratio, sky view factor, vegetation index, surface/volume ratio, and the canyon geometry factor on the SUHI (Okumus & Terzi, 2021).

Based on the review of the relevant literature, three major research gaps were identified. First, most of the studies implemented machine learning algorithms to predict different characteristics of UHI (such as air temperature, land surface temperature, land use land cover, consequences of factories' characteristics on UHI, effect of morphological spatial pattern of green spaces on UHI, and impacts of urban fabric of

UHI, etc.) rather than to particularly predict the UHI severity level itself. Furthermore, existing studies did not consider a comprehensive set of indicators or factors that could impact the UHI severity such as demographic, meteorological, and LULC factors. Second, little-to-no previous studies were conducted to estimate or predict UHI on the census-tract level, which is important to implement because it generally provides predictions on a finer (i.e., more detailed or concentrated) level. This will actually enable community-specific predictions or estimations of the UHI severity as well as a better understanding of the UHI severity, the associated factors that affect it, and the mitigation actions that could be implemented to alleviate its impacts. Third, although several previous studies have implemented machine learning algorithms to enhance the knowledge of UHI and its characteristics, most of those studies relied on black-box models. While the developed models by the existing studies are important and useful, their black-box models might hinder their practical applicability since they present the output without an understanding of the internal model logic or relationships or without an explanation of the estimation or prediction process (Yousefi et al., 2021). On the other hand, the significance of implementing white-box models lies in their transparent and interpretable nature that provides knowledge about the relations built and how they affect the model's output. Having a machine learning model that represents the relations is key for predicting UHI severity as it reflects the dependencies between different major factors affecting the UHI severity and thus it helps in developing appropriate mitigation approaches. Moreover, previous research efforts used machine-based algorithms without leveraging expert or human knowledge in the domain of UHI which is indispensable in such studies since it increases the practicality and reliability of any developed model, enhances the learning process, and develops interpretable and purposeful models. Thus, there is a need to develop white-box machine learning models that leverage expert knowledge to accurately predict UHI severity.

To that extent, this paper aims to fill these research gaps through developing a novel knowledge-based white-box Bayesian network machine learning model to predict the UHI severity on the census-tract level based on multiple demographic, meteorological, and LULC factors.

## 2. Methods and data

The research methodology in this paper is comprised of two parts: dataset development, and methods as shown in Fig. 1. First, dataset development was conducted. Then, the methods were performed to analyze the developed dataset as well as construct the Bayesian networks.

### 2.1. Dataset development

#### 2.1.1. Study area

The State of New Jersey (NJ) in the US was selected as the study area, which is located in the northeastern coastal area of the US, as shown in Fig. 2.

The State of New Jersey was chosen as the designated area of focus due to the severe UHI effect in its urbanized regions (Solecki et al., 2005). For instance, the city of Newark in New Jersey has been ranked as the city with the second-worst or highest UHI among 159 U.S. cities (Climate Central, 2022). Despite being the fourth smallest state in terms of land area, New Jersey is the eleventh most populous state and has the most densely populated urban agglomeration in the entire country with a population of 9.29 million residents (Nationonline, 2023). This high population along with the urbanized areas associated with high impermeability and high building heights are one of the major factors that led to the exacerbation of the UHI effect in New Jersey (Kaulessar, 2022). New Jersey also ranks among the most educated states in the US, with approximately 40% of the population aged 25 or older holding a bachelor's degree (U.S. Census Bureau, 2020a). Meanwhile, New Jersey has a heterogeneous natural environment such as forests, wetlands, rivers, and

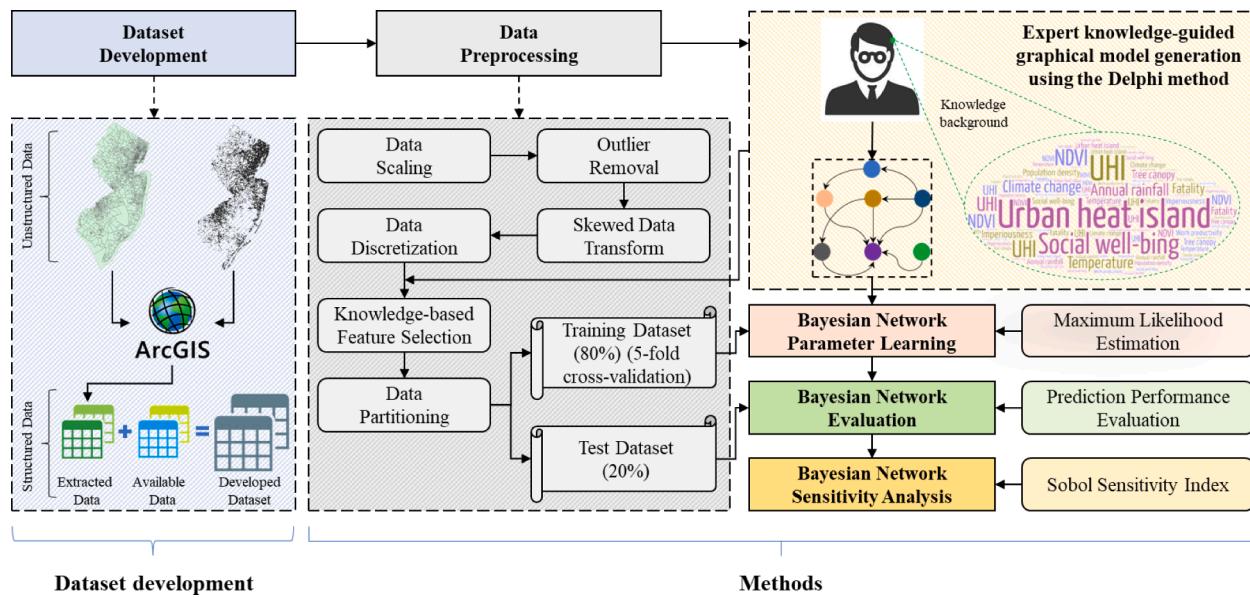


Fig. 1. Research methodology.

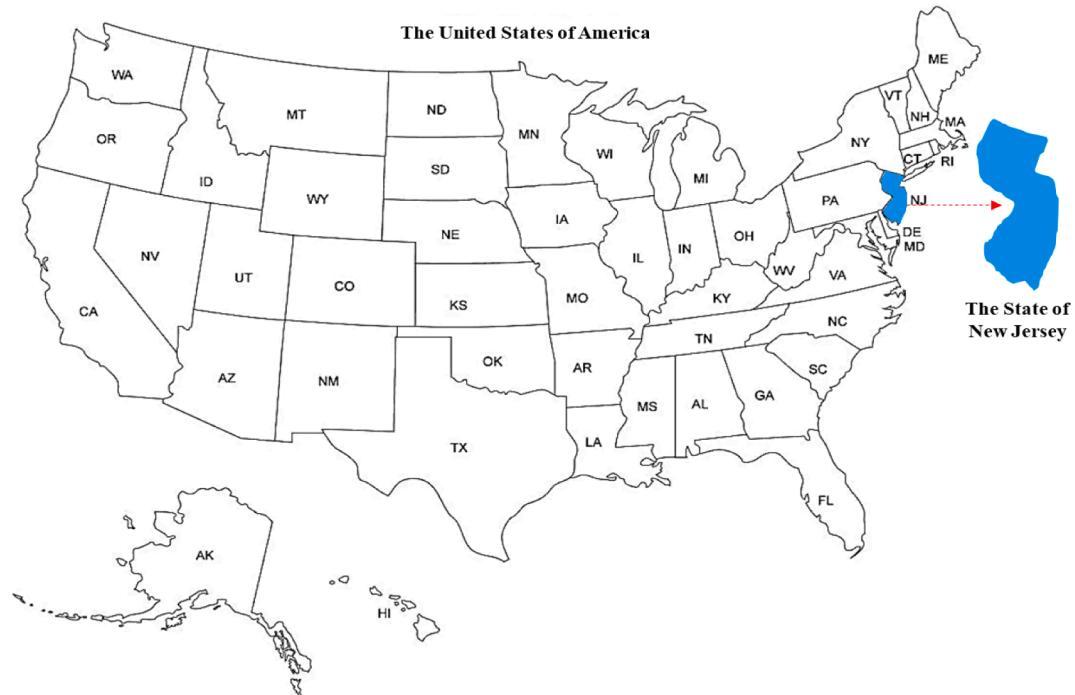


Fig. 2. The selected study area (The State of New Jersey in the United States).

coastline. Ultimately, the dense population and its diverse social, environmental, and demographic attributes underline the significance of studying the UHI severity in New Jersey.

#### 2.1.2. Influencing factors identification

To develop the Bayesian network model, a dataset for the census tracts in the State of NJ was collected and developed. A total of 13 influencing factors or variables that could impact the UHI severity were identified based on the existing literature as shown in Table 1. The basis for the selection of the different influencing factors includes variables that: 1) can capture the various aspects that could impact the UHI effect including meteorological factors, demographic factors, and land use/land cover factors (Antonopoulos, Trusty & Shandas, 2019; Ivajnšić &

Žiberna, 2019; Li et al., 2011), and 2) are reported in the literature to influence the UHI severity. More specifically, the basis for the selection of each one of the influencing factors is shown in the last column of Table 1.

As shown in Table 1, three NDVI variables were selected because previous studies have shown that, while seasonal changes might affect different variables, the NDVI is the variable that is greatly affected by such changes which have proved to have a great impact on the UHI effect as compared to other seasonable changes in other variables (Rasul, Balzter & Smith, 2016; Wu, Zhang & Zang, 2019; Xin et al., 2022; Yuan & Bauer, 2007).

While the proposed model in this paper was built based on an extensive list of factors affecting the UHI severity, the model could be

**Table. 1**  
Independent variables.

Data variable	Description			Unit	References supporting the impact of the variable on the UHI effect
Category	Name	Abbreviation			
Meteorological factors	Mean Annual Temperature	MAT	The average air temperature for the whole year in a specific census tract	°C	Schwarz and Manceur (2015)
	Annual Rainfall	AR	The summation of daily precipitation throughout the entire year	mm	
Demographic factors	Population	P	The total number of residents in a census tract	counts	Dewan et al. (2021)
	Population density	PD	A measurement of the number of residents per square-kilometer land area	counts/km <sup>2</sup>	Ramírez-Aguilar and Souza (2019)
Land use/Land cover	Waterbody Area	WBA	The total area of a region with a substantial accumulation of water	m <sup>2</sup>	Schwarz and Manceur (2015)
	NDVI_average	NVDI_a	Normalized Difference Vegetation Index for urbanized areas during the entire year	-	Mpkairi and Muvengwi (2019)
	NDVI_summer	NVDI_as	Normalized Difference Vegetation Index for urbanized areas during the summer season of a year	-	
	NDVI_winter	NVDI_aw	Normalized Difference Vegetation Index for urbanized areas during the winter season of a year	-	
	Building area	BA	The total floor area of all buildings in a census tract	m <sup>2</sup>	Kotharkar and Surawar (2016)
	Total Road Length	TRL	The total distance between the start and end of all road segments	m	Tang et al. (2017)
	Building Height	BH	The average vertical distance between the ground level up to the highest point of each building in a census tract	m	Zhou et al. (2022)
	Imperviousness	I	The percentage ratio between the total area preventing penetration or passage of water into the soil and the total land area in a census tract	%	Mathew, Khandelwal and Kaul (2016)
	Tree Canopy	TC	The percentage ratio between the total urban area shared by trees and the total area of land	%	Loughner et al. (2012)

further expanded by future studies by considering additional factors (e.g., meteorological factors such as wind, circulation, as well as others) that were not necessarily readily and accurately available at the census-tract level for the State of New Jersey but might be available or important for other national or international geographical locations or areas.

#### 2.1.3. Data sourcing, extraction, and fusion

The data for the different variables considered in this paper were obtained and/or extracted from various sources as shown in the “Data source” column of Table 2. Table 2 shows that the sourced data includes both structured and unstructured data. Structured data were in tabular form while unstructured data include TIFF file, Shapefile, and Disc image file, which carry rich information that cannot be directly used

without pre-processing it first. Therefore, the unstructured data was further converted to a structured/tabular form through data extraction using geographic information system software (i.e., ArcGIS). Ultimately, using the GEOFID for each census tract as a unique index, all data were fused into one single Excel file to form the final developed dataset. This dataset is comprised of 14 variables and 1313 complete data samples (i.e., without missing values).

The data for the dependent variable (i.e., UHI) was retrieved for each census tract in New Jersey from the database created by Chakraborty, Hsu, Sheriff and Manya (2020a) on the average surface heat island effect. It is worth mentioning that the UHI was calculated as the LST difference between urban and non-urban or rural areas; where the urban areas and rural areas were differentiated based on the US Census Bureau's definition of urbanized area; and hence, the rural areas were defined as all non-urban and non-water land areas (Chakraborty et al., 2020a). In relation to that, the UHI severity was determined based on the widely adopted classification scale for assessing UHI severity level based on 4 classes (non-existent, slight, moderate, and strong) (GE et al., 2019; Lin et al., 2021) as shown in Table 3.

## 2.2. Methods

### 2.2.1. Data preprocessing

Preprocessing of the developed dataset was conducted to transform the raw data into a clean dataset so that the Bayesian networks can better learn and recognize the underlying UHI-related patterns and improve prediction accuracy and reliability (Xie et al., 2021). In relation to that, the following six tasks were performed for preprocessing the raw data: data scaling/normalization, outlier removal, skewed data transformation, data discretization, knowledge-based feature extraction, and data partitioning.

First, data normalization was performed to scale the values of each

**Table. 2**  
Data source and data type of variables.

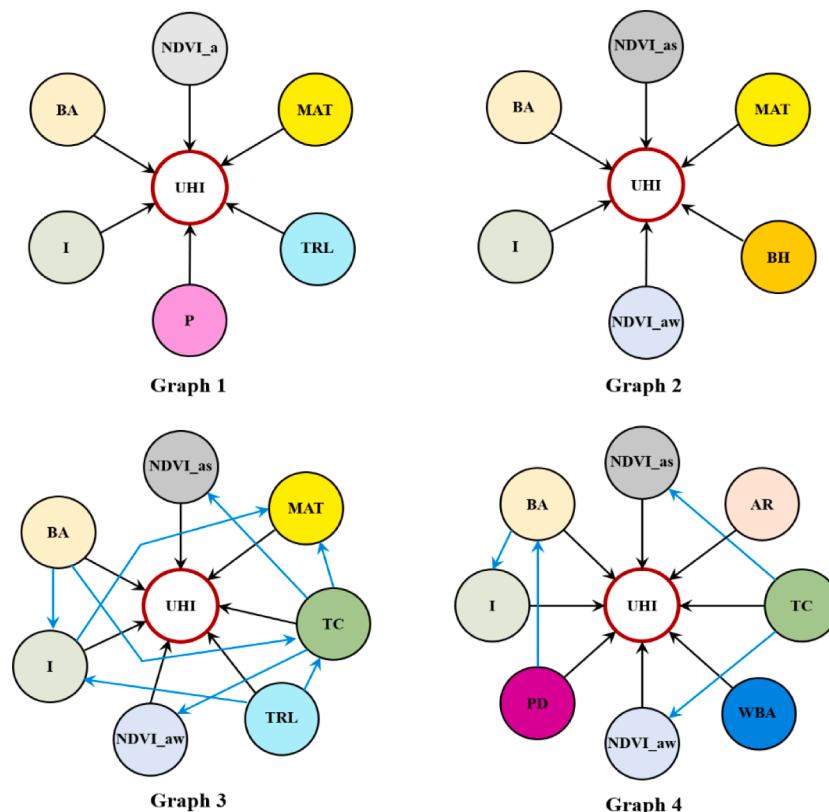
Data variable	Data source		Data type
Category	Abbreviation		
Meteorological factors	MAT, and AR	Global Climate Monitor (2023) U.S. Census Bureau (2020b)	Unstructured (Shapefile) Structured
Demographic factors	P, and PD	U.S. Census Bureau (2020b)	Structured
Land use/Land cover	WBA	Chakraborty, Hsu, Sheriff and Manya (2020a)	Structured
	NDVI_a, NDVI_as, and NDVI_aw	Heris, Foks, Bagstad and Troy (2020)	Unstructured (TIFF file)
	BA	New Jersey Office of GIS (2021)	Unstructured (Shapefile)
	TRL	Li, Gong and Zhou (2020)	Unstructured (TIFF file)
	BH	Multi-Resolution Land Characteristics Consortium (2019a)	Unstructured (Disc image file)
	I	Multi-Resolution Land Characteristics Consortium (2019b)	Unstructured (Disc image file)
	TC	Chakraborty, Hsu, Sheriff and Manya (2020a)	Structured
Dependent variable	UHI		

**Table. 3**  
UHI severity (i.e., dependent/output variable).

UHI severity level	UHI intensity (°C)
Non-existent	[−1, 1]
Slight	[1, 3]
Moderate	[3, 5]
Strong	≥ 5

data variable into a unified interval and the Min-Max normalization was applied in this study. Second, three different types of outlier removal methods were applied and compared to remove the abnormal values in the developed dataset. The used outlier removal methods include Minimum Covariance Determinant, Local Outlier Factor, and One-class Support Vector Machine. Third, skewed data transform was conducted to decrease the data variability resulting from data skewness using four methods: the sigmoid method (Zhang, Geisler, Ray & Xie, 2022), the log-transform method (Feng, Wang, Lu & Tu, 2013), the cube root method (Sinimaa, Spichakova, Belikov & Petlenkov, 2021), and the hyperbolic tangent method (Tsai, Liou, Simak & Cheng, 2017). Fourth, data discretization that segments the values of continuous variables into a finite set of bins/intervals is a significant operation for developing Bayesian networks. Equal-width discretization was selected due to its superior performance in developing Bayesian network models (Assaad, Hu, & Hussein, 2023; Zhu et al., 2015). Fifth, for feature selection, the expert knowledge-based feature selection method is adapted due to their “white-box” characteristics (i.e., high interpretability) (Chen, Chen, Hou, Hou & Xie, 2021; Velikova, Lucas, Samulski & Karssemeijer, 2013). In addition, the Delphi method was used to systematically and objectively gather expert knowledge to guide the feature selection process and identify the associated relationships and dependencies between the different variables. The gathered expert knowledge was represented using four graphical models as shown in Fig. 3. Sixth, the data was partitioned into 80% as the training set to develop the Bayesian network models, and 20% as the testing set to test the developed models.

From the perspectives of structure characteristics, the four graphical models shown in Fig. 3 could be grouped into two types: Naïve Bayesian networks (i.e., Graphs 1 and 2 in Fig. 3) and tree-augmented Bayesian networks (i.e., Graphs 3 and 4 in Fig. 3). A Naïve Bayesian network does not consider dependencies between input variables, whereas a tree-augmented Bayesian network captures possible dependencies between the influencing factors.



**Fig. 3.** Graphical models representing the dependencies between influencing factors and UHI severity.

### 2.2.2. Bayesian network parameter learning

Bayesian network parameter learning is a machine learning process to estimate the conditional probability distribution of a graphical model based on a given dataset, and then generate a Bayesian belief network as a probabilistic model for performing predictive analytics (Friedman et al., 1998). The conditional probability between the nodes in the graph quantitatively represents the strength of causality between the nodes and is often parameterized by a set of parameters ( $\Theta$ ). Using Bayes' theorem, the conditional probability can be further characterized by the joint probability between the nodes. Thus, the aim of Bayesian network parameter learning is to optimize  $\Theta$  by maximizing the likelihood or log-likelihood of the joint probability.

In order to optimize the performance of the proposed Bayesian networks in this paper, hyperparameter tuning was implemented using 5-fold cross-validation and grid search.

**Table 4** summarizes all relevant hyperparameters and associated values that were investigated in this study.

### 2.2.3. Bayesian network evaluation

As shown in **Table 2**, the output variable is the UHI severity which could take 4 different classes/levels: non-existent, slight, moderate, and strong. Thus, classification-based evaluation criteria for the proposed Bayesian networks were used to assess the models' prediction performances. More specifically, five metrics were used to evaluate the performance of the developed Bayesian networks, including: accuracy, precision, recall, F1 score, and AUC, as presented in Eqs. (2) to (6).

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

**Table 4**

Hyperparameters for Bayesian network parameter learning.

Data preprocessing	Method	Hyperparameter		
		Notation	Investigated values	Description
Outlier removal	Minimum Covariance Determinant (MCD)	c	{0.2, 0.3, 0.4, 0.5}	The percentage of anomalies in the raw dataset
	Local Outlier Factor (LOF)	n	{1, 2}	The number of neighbors to a data sample with outliers
		c	{0.2, 0.3, 0.4, 0.5}	The percentage of anomalies in the raw dataset
	One-class SVM	c	{0.2, 0.3, 0.4, 0.5}	The percentage of anomalies in the raw dataset
		kernel	{'linear', 'sigmoid'}	The kernel type of support vector machine classifier
Skewed data transform	-	m	{'sigmoid', 'log+1', 'cube root', 'hyperbolic tangent'}	The type of skewed data methods
Independent variables data discretization	Uniform discretization	nb	{3, 4}	The number of intervals a variable is discretized into

$$F1\ score = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \quad (5)$$

$$AUC = P(TP > FP) \quad (6)$$

Where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative, respectively; and  $P(TP > FP)$  denotes the probability of the Bayesian network having more confidence with randomly selecting a positive sample instead of a negative sample being positive.

#### 2.2.4. Bayesian network sensitivity analysis

Sensitivity analysis was performed on the best performing Bayesian network to quantify and understand the effect or impact of the different independent variables on the UHI severity. In relation to that, a higher sensitivity index indicates that the model output of the Bayesian networks (i.e., the UHI severity) is more sensitive to the change of the corresponding independent/input variable or factor (O’Neil & Niu, 2017; Zhang, Wu, Skibniewski, Zhong & Lu, 2014).

### 3. Analysis of results

This section presents and analyzes the results obtained in this study.

#### 3.1. Descriptive statistics of the developed dataset

As detailed in the “Methods and Data” section, data of 13 independent variables that could affect the UHI severity was collected. Table 5

presents a statistical summary of the collected data.

As shown in Table 5, there are some significant differences between the mean and median of the different independent variables such as population density (PD) and imperviousness (I). For example, the mean of the imperviousness data was 35.99%, which is much lower than its average of 40.84%, indicating a left skewness issue of the data. Such skewness issue can pose a great challenge for most machine learning models, including Bayesian networks, and can impact their capability in correctly and effectively recognizing the underlying patterns. These challenges are encountered since probabilistic models commonly require normally or approximately normally distributed data (Liu & Chen, 2020; Mishra, Mallick, Jena & Chae, 2020). Therefore, the results of the descriptive statistics shown in Table 5 show the need for implementing skewed data transformation in order to better facilitate the learning process of Bayesian networks and to reliably predict the UHI severity. Thus, skewed data transformation was implemented in this paper as detailed before in the “Methods and Data” section.

As for the dependent variable (i.e., the UHI severity), Fig. 4 shows the geospatial distribution of the UHI severity level among the different census tracts in NJ. It is to be noted that the distribution of UHI severity in NJ’s census-tracts was calculated according to the scale shown in Table 3 which has been used and recommended by different studies (GE et al., 2019; Lin et al., 2021). In other words, the “UHI is Non-Existent” level corresponds to UHI intensity between  $[-1^{\circ}\text{C}, 1^{\circ}\text{C}]$ , the “UHI is Slight” level corresponds to UHI intensity between  $[1^{\circ}\text{C}, 3^{\circ}\text{C}]$ , the “UHI is Moderate” level corresponds to UHI between  $[3^{\circ}\text{C}, 5^{\circ}\text{C}]$ , and the “UHI is Extreme” level corresponds to UHI  $\geq 5^{\circ}\text{C}$  (Lin et al., 2021). Also, it is worth mentioning that the UHI severity data for each census tract

**Table 5**

Descriptive statistics of the variables in the developed dataset.

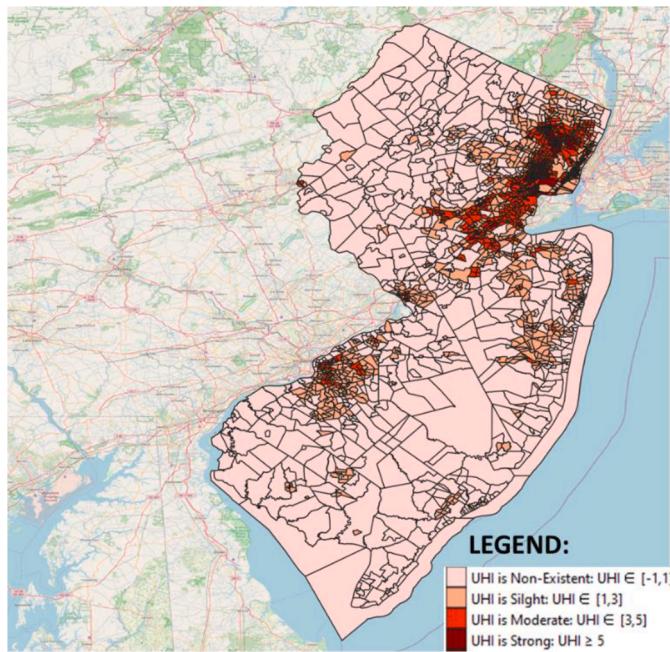
Variable	Unit	Mean	<sup>a</sup> STD	Minimum	Maximum	<sup>b</sup> Q1	<sup>c</sup> Q2	<sup>d</sup> Q3
MAT	°C	12.61	0.47	10.64	13.60	12.39	12.39	13.04
AR	mm	1095.51	84.61	893.50	1261.00	1057.70	1057.70	1057.70
P	counts	4212.33	1487.15	81.00	9359.00	3126.50	4124.00	5273.00
PD	counts/km <sup>2</sup>	3325.45	4143.58	11.09	34,656.36	921.30	1712.20	4037.96
WBA	m <sup>2</sup>	194,646.49	605,589.28	0.00	9,178,383.65	0.00	21,291.38	114,823.56
NDVI_a	-	0.38	0.12	0.01	0.64	0.28	0.41	0.48
NDVI_as	-	0.53	0.15	0.06	0.80	0.41	0.57	0.64
NDVI_aw	-	0.20	0.09	-0.03	0.54	0.12	0.20	0.26
BA	m <sup>2</sup>	291,285.83	216,793.30	19,788.70	3,258,781.00	141,498.65	241,826.60	384,395.25
TRL	m	27,660.86	22,894.92	918.50	193,652.14	11,873.96	22,205.04	36,160.98
BH	m	2.05	1.86	0.00	13.56	0.82	1.42	2.63
I	%	40.84	22.78	1.65	94.45	22.63	35.99	59.03
TC	%	20.74	15.91	0.00	74.49	6.55	18.22	32.49

<sup>a</sup> STD: standard deviation.

<sup>b</sup> Q<sub>1</sub>: the first quartile.

<sup>c</sup> Q<sub>2</sub>: the second quartile.

<sup>d</sup> Q<sub>3</sub>: the third quartile.



**Fig. 4.** Distribution of UHI severity in NJ's census tracts.

was defined as the LST difference between urban areas and rural areas Chakraborty, Hsu, Manya and Sheriff (2020b).

### 3.2. Developed Bayesian networks: optimal hyperparameters and model performance

The following subsections present the optimal hyperparameters for the developed Bayesian networks and their associated performance.

#### 3.2.1. Optimal hyperparameters

The optimal set of hyperparameters were tuned and identified for each of the 4 Bayesian graphical network models. More specifically, using 5-fold cross validation and grid search, the set of hyperparameter

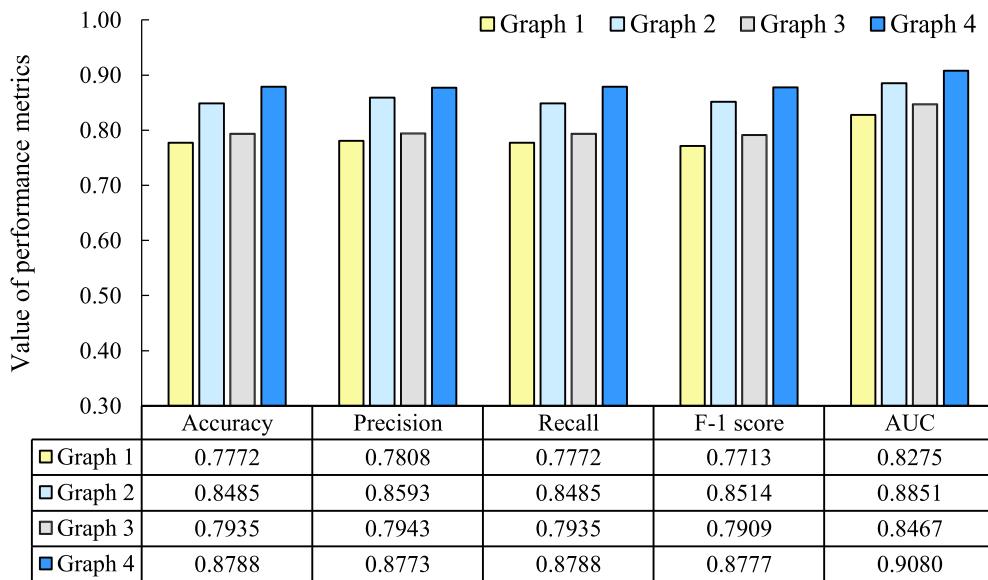
**Table. 6**  
Optimal hyperparameters for each of the developed Bayesian network.

Data preprocessing	Method	Optimal hyperparameter for each graphical model								
		Notation	Graph 1		Graph 2		Graph 3		Graph 4	
Independent variables data discretization	Uniform discretization	nb	3	4	3	4	3	4	3	4
Outlier removal	MCD	c	-	-	-	-	-	-	-	-
	LOF	n	-	-	-	-	-	-	-	2
		c	-	-	-	-	-	-	-	0.2
	One-class SVM	c	0.3	0.3	0.5	0.5	0.3	0.3	0.5	-
		kernel	linear	-						
Skewed data transform	-	m	sigmoid							

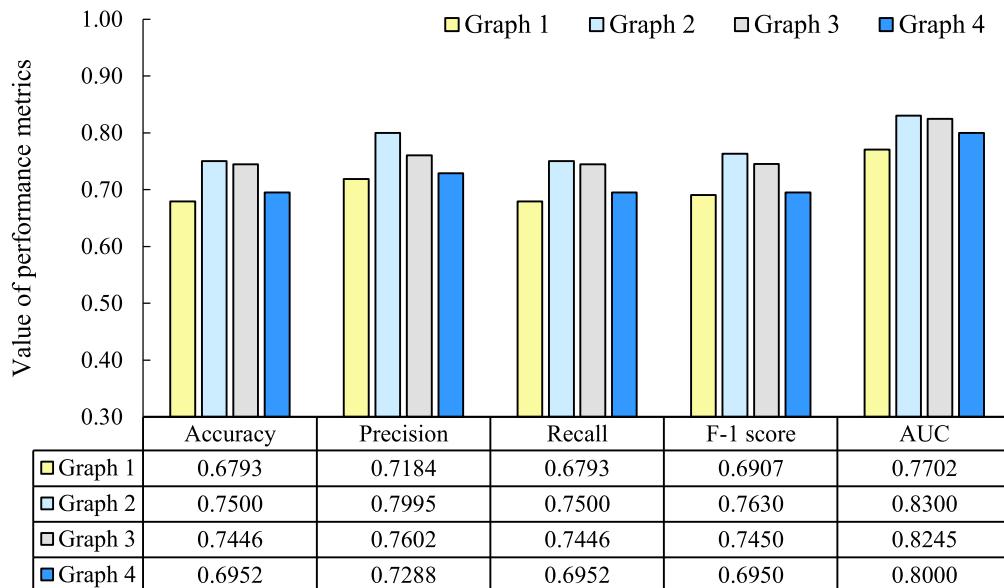
**Table. 7**  
Performance of each graphical model on the testing set (number of intervals = 3 vs. 4).

Number of discretization intervals	Graph number	Accuracy	Precision	Recall	F1 score	AUC
3	1	0.7772	0.7808	0.7772	0.7713	0.8275
	2	0.8485	0.8593	0.8485	0.8514	0.8851
	3	0.7935	0.7943	0.7935	0.7909	0.8467
	4	<b>0.8788*</b>	<b>0.8773*</b>	<b>0.8788*</b>	<b>0.8777*</b>	<b>0.9080*</b>
4	1	0.6793	0.7184	0.6793	0.6907	0.7702
	2	0.7500	0.7995	0.7500	0.7630	0.8300
	3	0.7446	0.7602	0.7446	0.7450	0.8245
	4	0.6952	0.7288	0.6952	0.6950	0.8000

\* Bold values denote the largest performance criteria for all graphical models and number of intervals.



**Fig. 5.** Performance of the developed Bayesian models (number of intervals = 3).



**Fig. 6.** Performance of the developed Bayesian models (number of intervals = 4).

87.88%. To better understand the generalizability performance of the recommended Bayesian Graph 4 model, Table 8 shows the confusion matrix of the model's prediction or performance on the unseen testing dataset for the different classes or category levels of UHI severity.

### 3.2.3. Internal structure of the optimal bayesian model

Leveraging the “white-box” nature of Bayesian networks, this paper further demonstrates the internal structure of the optimal model (i.e., Graph 4 with 3-discretization intervals). In relation to that, and in order to help practitioners and researchers in implementing the proposed Bayesian network in practice, Fig. 7 shows the internal architecture of the proposed model to estimate or predict the UHI severity along with the learned probability distributions for the different variables.

The limits of the various intervals shown in Fig. 7 for the different independent variables are shown in Table 9.

As shown in Fig. 7 and Table 9, each of the independent variables/factors in Graph 4 was discretized into three intervals (which was determined to be the best permitting Bayesian network model as detailed in the previous subsections).

Moreover, in Fig. 7, the estimated likelihood tables (shown in the

**Table. 8**  
Confusion matrix of the proposed Bayesian network model.

	Non-existent	Slight	Moderate	Strong
Non-existent	100.0%	0.0%	0.0%	0.0%
Slight	0.0%	91.7%	7.9%	0.0%
Moderate	0.0%	8.3%	81.6%	12.1%
Strong	0.0%	0.0%	10.5%	87.9%

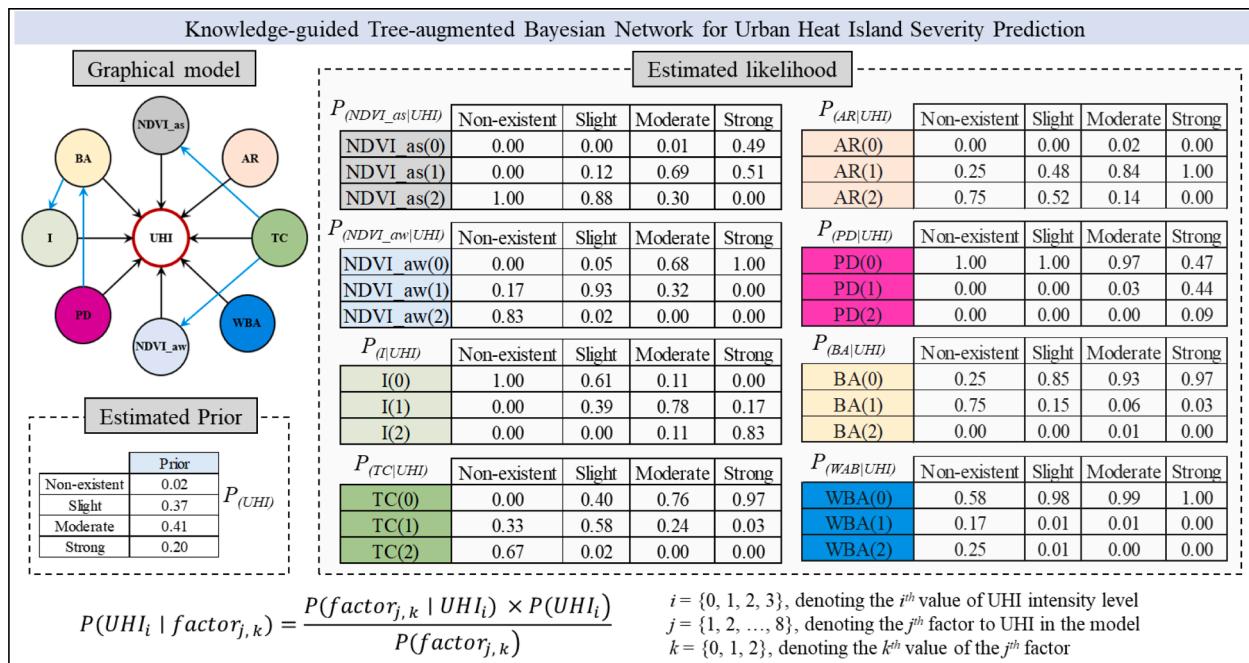


Fig. 7. The optimal graphical model and the learned conditional distributions.

**Table. 9**  
Discretization intervals for the independent factors.

Factor	Interval 0	Interval 1	Interval 2	Unit
AR	[894, 1016)	[1016, 1139)	[1139, 1261]	mm
BA	[0.29, 0.54)	[0.54, 1.05)	[1.05, 1.56]	km <sup>2</sup>
I	[1.65, 30.10)	[30.10, 58.56)	[58.56, 87.01]	%
NDVI_as	[0.14, 0.36)	[0.36, 0.57)	[0.57, 0.78]	-
NDVI_aw	[0.03, 0.20)	[0.20, 0.37)	[0.37, 0.54]	-
PD	[61, 5622)	[5622, 11,182)	[11,182, 16,743]	counts/km <sup>2</sup>
TC	[0, 24.83)	[24.83, 49.67)	[49.67, 74.49]	%
WBA	[0, 0.55)	[0.55, 1.10)	[1.10, 1.66]	km <sup>2</sup>

middle of Fig. 7) could be used to obtain the conditional probability that quantitatively indicates how likely it is to have each of the UHI severity levels given a certain observed data of the independent variables. The estimated prior of UHI (i.e.,  $P_{(UHI)}$ ) shown in the lower left part of Fig. 7 also probabilistically quantifies the chance of having the different UHI severity levels.

To offer insights into the use of the developed Bayesian network model in practice, the following paragraph provides a detailed description or instructions for practitioners, researchers, or users on how to easily and simply implement the proposed model in the real-world.

First, practitioners, researchers, or users need to specify/collect the values for the following 8 inputs or variables: AR (annual rainfall), BA (building area), I (imperviousness), NDVI\_as (NDVI in summer), NDVI\_aw (NDVI in winter), PD (population density), TC (tree canopy), and WBA (waterbody area). These are the nodes shown in the proposed or recommended Bayesian network Graph 4 (see Fig. 7) and thus are needed to estimate or predict the UHI severity. These variables are generally easily accessible by practitioners or researchers as they can be obtained from open-source datasets. Second, and after specifying the values of the 8 inputs, practitioners or users will simply map them into their associated discretization bins or rules shown in Table 9. Third, practitioners, researchers, or users can use the prior and likelihood conditional probability tables shown in Fig. 7 by mapping the value of each input variable to the corresponding cell in the tables to get the likelihood and prior values. These probability tables were provided to ease the process of implementing the proposed model in practice.

Fourth, and once the likelihood and prior values associated with the input variables are obtained, the posterior probability for each UHI severity level is calculated using Bayes' theorem as shown in Eq. (7). Finally, the estimated or predicted UHI severity corresponds to that having the highest posterior probability.

$$P(UHI_i | factor_{j,k}) = \frac{P(factor_{j,k} | UHI_i) \times P(UHI_i)}{P(factor_{j,k})} \quad (7)$$

where  $P(UHI_i | factor_{j,k})$  is the posterior probability used to predict the chance of having each UHI  $i^{th}$  severity level ( $i = \{0, 1, 2, 3\}$ ) based on the  $k^{th}$  ( $k = \{0, 1, 2\}$ ) interval of each one of the  $j = \{1, 2, \dots, 8\}$  input variables or factors;  $P(factor_{j,k} | UHI_i)$  is the likelihood probability;  $P(UHI_i)$  represents the prior probability for each UHI  $i^{th}$  severity level; and  $P(factor_{j,k})$  is the normalizing summation.

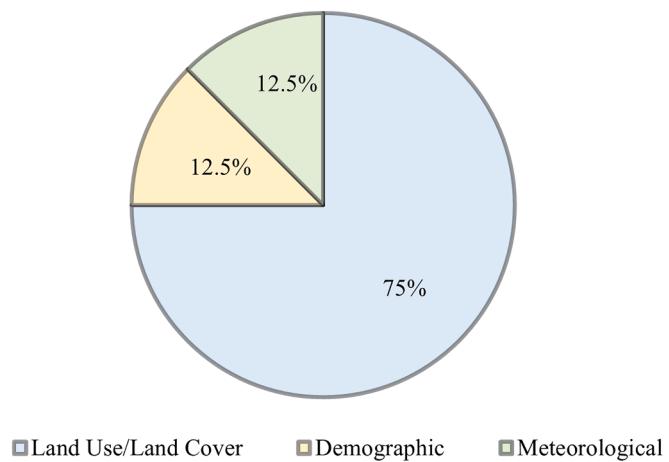
## 4. Discussion

This section provides an in-depth discussion of the results obtained in this paper in terms of: the key factors affecting UHI severity (which are those represented in the best performing Graph 4 Bayesian model), the dependencies between the key factors, and the influence or impact of each factor on the UHI severity.

### 4.1. Key factors affecting UHI severity

The key factors affecting the UHI severity are the variables that were identified in the optimal graphical Bayesian network model (i.e., Graph 4). These factors include the following 8 variables: AR (annual rainfall), BA (building area), I (imperviousness), NDVI\_as (NDVI in summer), NDVI\_aw (NDVI in winter), PD (population density), TC (tree canopy), and WBA (waterbody area).

As shown in Table 1, each one of these variables belong to one of the following three categories: land use/land cover (includes the following 6 variables: BA, I, NDVI\_as, NDVI\_aw, TC, and WBA), demographic (i.e., includes the following 1 variable: PD), and meteorological factors (i.e., includes the following 1 variable: AR). Fig. 8 shows the percentage fraction of each factor category out of the total number of key variables identified (i.e., 8 factors).



**Fig. 8.** Distribution of the categories of key factors affecting the UHI severity.

**Fig. 8** shows that some factor categories were presented more frequently than the others in relation to the key factors affecting UHI severity.

Starting with the land use/land cover category (which is the category that has the largest number of key factors affecting UHI severity, with a percentage of 75% as shown in **Fig. 8**), the key factors that fall in this category are: imperviousness, tree canopy, building area, NDVI<sub>as</sub>, NDVI<sub>aw</sub>, and water body area. The next paragraphs discuss each one of these variables.

Imperviousness refers to the fraction of impervious surfaces that absorb and emit heat and are characterized by low albedo coefficients, increased surface roughness, and raising surfaces sensible heat fluxes. These characteristics of impervious surfaces lead to higher temperatures and ultimately to more severe UHI effect (Ma & Peng, 2021), which explains having the ‘imperviousness’ variable as one of the key factors affecting the UHI severity as found in this paper.

Building area can be considered to impact the UHI severity in two ways by its effect: on the building material and on the Sky View Factor (SVF). Building area impacts the quantity of building material used in an area. This material is characterized by its high heat storage, imperviousness, and low albedo which consequently lead to higher temperatures (Aflaki et al., 2017; Lee, Moon, Choi & Yoon, 2018). Moreover, as the building area increases, the SVF decreases (Hu, White & Ding, 2016), where SVF is a measure of open canyon space or visible sky, and thus lower values of SVF correspond to lower cooling rates of an area (Gal, Lindberg & Unger, 2009). This explains having the ‘building area’ variable as one of the key factors affecting the UHI severity as found in this paper.

Moreover, the optimal model identified two significant UHI mitigation or cooling practices: green spaces (represented by the NDVI and tree canopy) and waterbodies (represented by the WBA). Green spaces have been identified as effective cooling practices and include green roofs, green walls, parks, street vegetation, and woodlands or forests (Li et al., 2022). The NDVI, which represents vegetation layers and their corresponding health in general, is identified as the key factor affecting the UHI severity in its two seasonal measures: summer and winter. The results showed that representing the NDVI through its winter (i.e., NDVI<sub>aw</sub>) and summer (NDVI<sub>as</sub>) values has a greater impact on UHI severity compared to its annual average (i.e., the NDVI<sub>a</sub> variable). This shows that the NDVI significantly varies during different seasons, and that this variation has an important impact on the UHI severity. Specifically, in winter, the growth rate of vegetation is slower and sometimes null due to the low temperature and extreme weather conditions. On the other hand, vegetation shows significant blossoming during the summer compared to the winter season. This explains the difference in vegetation or NDVI between winter and summer. Also, NDVI has a

crucial relationship with the changes in temperature due to the cooling ability of vegetation through evapotranspiration caused by photosynthesis (Lindén, 2011). This explains having ‘NDVI<sub>aw</sub>’ and ‘NDVI<sub>as</sub>’ as key factors affecting the UHI severity as found in this paper. Furthermore, tree canopy indicates the fraction of an area that is shaded by trees. Tree canopy is perceived to affect the UHI effect in two manners: by the shading effect of trees and by the cooling abilities of trees. In relation to that, the shading effect can reduce the solar radiation and consequently the heat convection and storage of urban surfaces. This effect can directly lower the LST in urban areas (Yu et al., 2020). On the other hand, trees can also cool their surrounding through evapotranspiration and the release of moisture into the atmosphere (Tamaskani Esfahanikalateh, Ngarame & Yun, 2021). This explains having the ‘tree canopy’ variable as one of the key factors affecting the UHI severity as found in this paper.

Furthermore, although green spaces have significant cooling impact, this impact varies depending on the characteristics of the green spaces such as: their size or area, their composition, and their configuration and also depending on external factors such as the background temperature (Berger et al., 2017; Li et al., 2022). Several efforts have been devoted to determine how the UHI mitigation impacts of green spaces can be enhanced to ensure effective cooling of the cities. For instance, the optimal size of green spaces in 3 megacities in Asia was determined and it was identified that the optimal size changes depending on the study area; where it was found to be 0.37 ha in Dhaka, 0.77 ha in Kolkata, and 0.42 ha in Bangkok (Li et al., 2022). Moreover, the optimal size of parks and woody vegetation should be greater than 0.6 ha to ensure effective cooling of the surroundings (Lu et al., 2023). In fact, some studies concluded that parks with an area smaller than 0.5 ha have a contradicting effect as they were 0.42 °C hotter than their surroundings (Chang, Li & Chang, 2007). It was also identified that green spaces with an area less than 0.5 ha have a negligible cooling efficiency that is less than 1 °C and close to 0 °C (Fan et al., 2019). Furthermore, and to enhance the cooling efficiency of green spaces, efforts to combine green spaces with other cooling practices have been implemented. For instance, the combination of green spaces with water bodies, known as green-blue spaces, has a mean air temperature reduction of 3.3 °C higher than the sum of the cooling effects of green spaces alone and water bodies alone (Shi et al., 2020). Green spaces were also combined with cool roofs, or roofs with high albedo coefficient (Yuan, Emura & Farnham, 2017). Thus, the identification of the optimal characteristics of green spaces in the studied area is essential for effective and efficient cooling of the city.

Water bodies are known as key mitigation approaches for the UHI effect. In fact, the UHI reduction capability of water bodies depends mainly on the area of the water (Assaf & Assaad, 2023a). Water bodies tend to decrease the surrounding temperature through transforming the latent and sensible heat fluxes (Sun & Chen, 2012). Moreover, water bodies form urban cooling islands around them that could extend up till 100 m buffer (Assaf & Assaad, 2023a). Furthermore, water bodies cool the surrounding area through their evaporation and their corresponding “oasis effects”; where oasis effect is defined as the creation of an environment that is cooler than its surroundings due to evaporation (Gupta, Mathew & Khandelwal, 2019). This explains having the ‘water body area’ variable as one of the key factors affecting the UHI severity as found in this paper. Furthermore, the cooling effect of water bodies is impacted by its size, geometry, type, and distance from the downtown. For instance, an increase of 10% in the WBA can reduce the UHI intensity by 11.33% (Lin et al., 2020) and the decrease of land surface temperature by 0.43 °C (Wang, Zhan & Ouyang, 2019). Also, water bodies that are far from the center of the city have a lower cooling impact compared to water bodies closer to the city center (Peng et al., 2020). That being said, having large water bodies that are close to the city center would enhance the cooling of the city. Finally, it is worth mentioning that although the model identified only two cooling practices (i.e., green spaces and waterbodies), several other UHI mitigation

initiatives could also be implemented to cool cities and urban areas. Such initiatives include permeable pavement, cool pavement, pavement watering, and optimal urban planning such as optimal street orientations (Assaf and Assaad, 2023b).

Moving to the demographic category that was only represented by 12.5% of the key factors affecting UHI severity (as shown in Fig. 8) through the population density variable. Population density, or the population per unit land area, is a measurement for the concentration of human in a certain geographic area. Moreover, population density and its variation with time is considered a direct representative of urbanization, which is one of the main phenomena that exacerbates the UHI effect (Qizhi, Ying & Kang, 2016; Zhang, Shou & Dickerson, 2009). Moreover, increased population density corresponds to increased temperatures and consequently increased UHI severity (Kotharkar & Surawar, 2016). This is mainly due to the increase in the release of anthropogenic heat, urbanization, and modernization when the population intensity increases. To that extent, it can be concluded that as population density increases UHI severity increases while as population density decreases UHI severity will decrease as well, due to the decrease of the anthropogenic heat release. This explains having the 'population density' variable as one of the key factors affecting the UHI severity as found in this paper.

Furthermore, as shown in Fig. 8, the meteorological category corresponds to only 12.5% of the key factors affecting UHI severity where the meteorological key factor was found to be the annual rainfall. It is perceived that the evaporation effect of rainwater reduces latent and sensible heat fluxes (Li et al., 2011). Moreover, rainfall increases soil's moisture and heat storage (Yuan et al., 2022) as well as reduces the heat absorbed and stored by urban surfaces (Yang, Ren & Hou, 2019). That being said, precipitation decreases solar radiation, heat storage, temperature and consequently reduces the UHI severity (Chen et al., 2021). In fact, rainfall has been proven to mitigate the UHI effect and decrease its severity; where the rainfall's mitigation capabilities depend on the intensities and duration of the rainfalls (He, 2018). This explains having the 'annual rainfall' variable as one of the key factors affecting the UHI severity as found in this paper.

#### 4.2. Dependencies among key factors affecting UHI severity

The results showed that the dependencies or relationships among the key factors also play a significant role in UHI severity classification. In fact, the optimal model (i.e., Graph 4) has revealed four dependencies among the key factors affecting UHI severity. Those dependencies were formed between demographic factors and LULC factors as described in

the paragraph below.

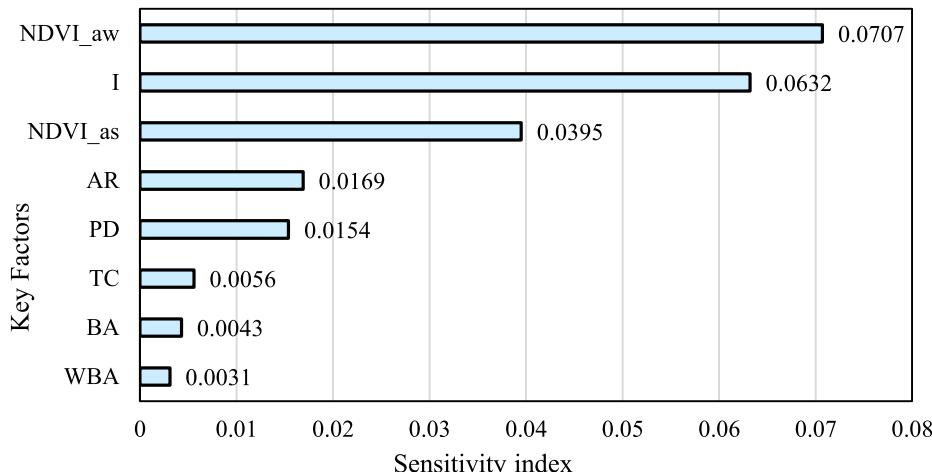
The dependencies in Graph 4 showed that as population density varies, building area varies as well (see Fig. 7). This relationship could be explained through the fact that as population density increases, building area will increase to accommodate the population need for shelter (Emilsson, 2021). Another dependency is the one between building area and imperviousness (see Fig. 7), which could be explained by the fact that as building area increases building material increases as well, and since buildings in urban areas are generally composed of impervious material, then consequently the imperviousness fraction increases with the increase of building area and it decreases with the decrease of building area. Finally, other identified dependencies in this paper demonstrated that tree canopy impacts both NDVI<sub>as</sub> and NDVI<sub>aw</sub> (see Fig. 7). Since tree canopy is a representation of the extent of presence of trees in a certain area; then as the tree canopy fraction changes, the NDVI would change as well. This is explained by the definition of the NDVI that reflects the quantification of vegetation greenness of the studied area.

#### 4.3. Influence of key factors on UHI severity

To quantify and better understand the influence or impact of each one of the key variables on the UHI severity, sensitivity analysis was performed. In relation to that, Fig. 9 shows the sensitivity results for each one of the 8 key factors impacting UHI severity.

As shown in Fig. 9, three land use/land cover factors (i.e., NDVI<sub>aw</sub>, imperviousness (I), and NDVI<sub>as</sub>) were identified as the top three factors to which the UHI severity is most sensitive. These three factors are the key contributors that characterize the landscape features of an urbanized region since they can significantly affect the absorption and/or re-emission of heat from the sun (US EPA, 2022). Moreover, it was determined that UHI severity is more sensitive to NDVI<sub>aw</sub> than NDVI<sub>as</sub> (see Fig. 9). Although these two indices seem similar, their values are significantly different where NDVI<sub>aw</sub> generally has much lower values than NDVI<sub>as</sub>. This can be also concluded from Table 5 that shows the mean value of NDVI<sub>as</sub> is 0.53 while that of NDVI<sub>aw</sub> is 0.20. More specifically, the results of the sensitivity analysis showed that NDVI<sub>aw</sub> is the most influential factor on UHI severity. This demonstrates that UHI severity is affected by the reduction in the vegetation layers, represented by the decrease of the NDVI values in winter season, more than it being sensitive to the increase in the vegetation and the NDVI values in the summer season.

Furthermore, although water bodies are considered as one of the mitigation approaches for the UHI effect, the sensitivity analysis results



**Fig. 9.** Sensitivity index of key factors affecting UHI.

show that WBA has a low sensitivity index to the UHI severity compared to other key factors (see Fig. 9). This could be explained by the fact that since the developed model studies the UHI on a small scale (i.e., census-tract level), some census tracts might not include a waterbody due to their small areas. Thus, several census tracts might be associated with a WBA value of 0. This explains the results of the sensitivity analysis that shows a lower sensitivity index for WBA compared to other key factors (see Fig. 9). Moreover, urban changes and/or developments greatly affect vegetation and imperviousness; whereas, water bodies and their corresponding areas (i.e., the WBA factor) are the least affected by urban changes (i.e., it is not very common to have land reclamation or land fill of water bodies). Thus, and since the UHI effect is a consequence of urban changes, this explains why UHI is less sensitive to WBA and more sensitive to the change in the other factors such as the LULC factors (represented by the NDVI, imperviousness, and tree canopy) and urban growth factors (represented by population density and building areas).

#### 4.4. Guidelines for directing future UHI-Related research using Bayesian parameter learning

Fig. 10 summarizes the proposed model construction process in this study that could be used by future studies to predict the UHI severity using Bayesian networks or to perform UHI-related research.

Despite the fact that the proposed workflow achieved promising results in accurately predicting the UHI severity, there are two main aspects that need to be carefully considered as they might impact the performance of the model (i.e., error/accuracy), including the outlier rate and the number of discretization intervals.

The effect of outlier rate on the model prediction performance was investigated for the identified optimal Bayesian network in this paper (i.e., Graph 4 with 3-interval discretization and an outlier rate of 0.5). The obtained results are shown in Fig. 11.

Fig. 11 shows that the model prediction performance would decrease as the outlier rate decreases. For example, an outlier rate of 0.20 may result in a drop of model prediction accuracy by approximately 6% (see Fig. 11) during model training process. The insights can be attributed to two reasons. First, a higher outlier rate specified for the outlier removal method will drop more potential outliers, thus producing a cleaner dataset for developing a more reliable and robust Bayesian network model. Second, a higher outlier rate also indicates that more data samples will be disregarded. As a result, there would be less data samples in the clean dataset. Previous studies commonly revealed that Bayesian network models often work better with small dataset because Bayesian algorithms can naturally incorporate prior knowledge to compensate for the limited amount of data (MacAllister, Kohl & Winer,

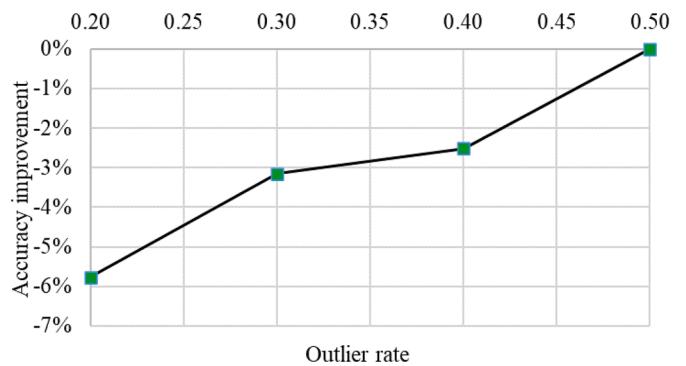


Fig. 11. Effect of outlier rate on prediction performance of the optimal model.

2020). Therefore, Bayesian models are less prone to overfitting when smaller datasets are used.

The number of data discretization intervals can also affect the prediction performance of Bayesian network models. Using Graph 4 as an example, the results shown in Table 7 reflected that the testing accuracy was 18.36% higher when 3 intervals were selected compared to selecting 4 intervals. The reason behind this is that a smaller number of discretization intervals can reduce the data complexity and hence could achieve more stable and accurate predictions. In addition, when the data is discretized into a smaller number of intervals, the impacts of the noises and outliers will be reduced, which also increases the robustness of the Bayesian model, resulting in better prediction performance.

## 5. Conclusion

This paper presented a novel knowledge-based white-box Bayesian network model that estimates or predicts UHI severity on the census-tract level based on 13 demographic, meteorological, and land use/land cover factors. This paper: 1) leveraged expert knowledge to develop networks that reflect factors and relations that impact the UHI severity, 2) identified key factors and relations that explain the UHI severity, and 3) quantified the influence of each of the key factors on the UHI severity. Based on the results obtained from the four investigated Bayesian networks, it was shown that the tree-augmented Graph 4 is the optimal network for predicting the UHI severity; where this network includes 8 nodes (i.e., factors) and 12 edges (i.e., relationships). More precisely, the optimal network identified the following eight key factors that impact the UHI severity: NDVI\_aw, NDVI\_as, imperviousness, tree canopy, building area, water body area, population density, and annual rainfall.

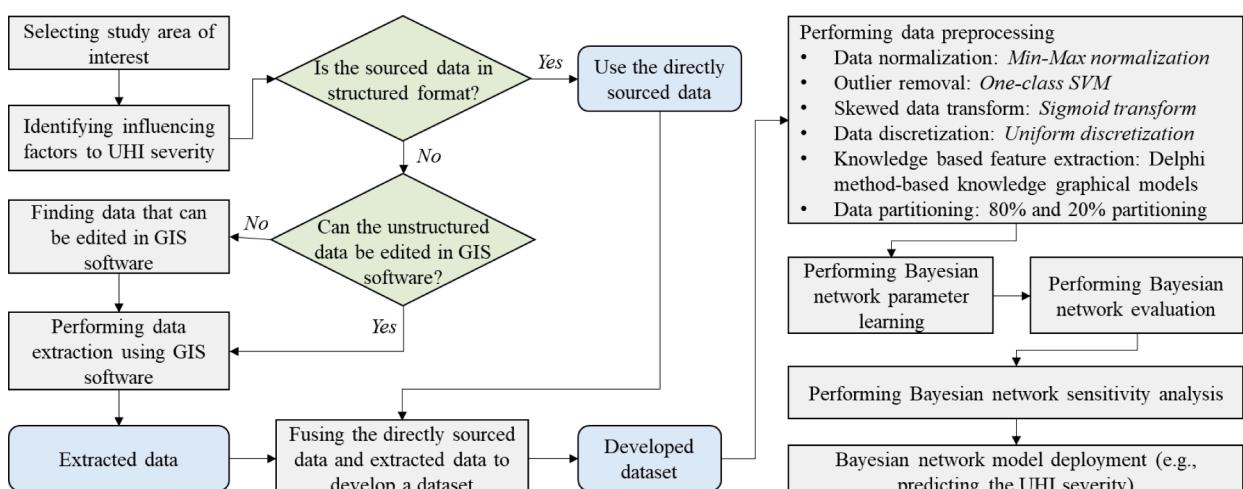


Fig. 10. Construction process of the Bayesian network model for UHI severity prediction.

The optimal model had an accuracy of 87.88%, a precision of 87.735%, a recall of 87.88%, F1 score of 87.77%, and an AUC of 90.80%. Furthermore, the optimal network prediction efficiency was also demonstrated in the confusion matrix (see Table 8) that shows that 100% of UHI non-existent level intensities were accurately predicted, 91.7% of the slight UHI intensities were accurately predicted, 81.6% of the moderate UHI level intensities were accurately predicted, and 87.9% of the strong UHI level intensities were accurately predicted. Thus, one of the main significances of the developed model is its ability to accurately, precisely, and efficiently model and predict the UHI severity with high accuracy measures.

Moreover, it was shown that 75% of the key factors were related to land use/land cover while the rest of the key factors were distributed among the demographic and meteorological categories. Finally, the results highlighted that land use/land cover factors are major factors that impact the UHI severity whether being the contributors to the UHI effect (such as imperviousness and building area) or being the mitigation approaches for the UHI effect (such as vegetation represented by NDVI, tree canopy, and water body areas). Also, this paper determined that the representation of NDVI by two measures/indicators (i.e., the winter and summer NDVI values) is more representative than using the annual average of NDVI for predicting UHI severity. This can be related to the major changes of the NDVI value between the winter and summer seasons, which makes the annual average less representative than the two separate seasonal indicators.

The results also showed that dependencies between factors play a significant role in UHI severity classification. In fact, this paper identified four key relationships; where it was shown that the variation of population density affects the building area, and that the building area variation consequently changes the imperviousness fraction. On the other side, changes in tree canopy affect both measures of NDVI (i.e., during winter and summer seasons). Furthermore, the sensitivity analysis showed that UHI severity is most sensitive to three land use/ land cover factors: NDVI\_aw, imperviousness, and NDVI\_as.

The findings of this paper are valuable in UHI management, mitigation, and adaptive measures. The identification of the key factors affecting UHI is needed for recognizing which aspects or phenomena greatly impact the UHI severity. Moreover, the results shed lights on some of the crucial UHI mitigation initiatives and their degree of impact on UHI severity; where vegetation, trees, water bodies, and rainfall have been identified as key factors impacting the UHI severity. On the other hand, this paper also identified impervious surfaces, population density, and building area as major factors that exacerbate the UHI severity. This identification can help project planners and decision-makers in knowing which aspect they have to focus on when developing UHI management plans. For instance, this paper showed the need to reduce building areas and impervious surfaces, which could be achieved through implementing more green initiatives such as green roofs or green facades and replacing the conventional asphalt pavements by permeable pavements to reduce imperviousness. Moreover, authorities need to limit population densities in urban areas by reducing migration from rural to urban areas through providing additional opportunities in the rural surroundings.

Ultimately, this study contributes to the body of knowledge by paving the way for having intrinsically interpretable (i.e., white-box), simple, and robust Bayesian network-based models for supporting UHI severity prediction on the census-tract level. First, the white-box or interpretability attribute of the proposed approach in this research gives experts and practitioners a higher confidence in the model's outputs since the model proposed in the paper provides practitioners with the internal structure used to estimate the UHI severity by graphically showing the causal relationships between the variables. In fact, the outputs or predictions from the proposed Bayesian network model are easy to interpret as every UHI severity prediction or estimation provided by the model can be traced and checked (i.e., using the dependencies reflected in the network's architecture), which is a very important

feature that is lacking in existing black-box machine learning models.

Second, and to the best knowledge of the authors, this is the first research effort that attempts to predict the UHI severity (rather than air temperature or other UHI-related characteristics) on the census-tract level in the US. This is particularly important as it allows practitioners and researchers to perform micro-level UHI predictions and inferences about future or unknown UHI severity pertaining to individual communities or small/localized geographical areas. For instance, the proposed model could be used to quantify or understand the impact of urban development on the UHI severity for a community of interest through varying specific urban growth or development factors (such as population density, building area, or imperviousness) and predicting the associated UHI severity, which consequently helps identifying the needed intervention or actions that have to be implemented to reduce the UHI severity.

Third, the approach proposed in this paper provides a finer (rather than a coarse) level of granularity in relation to the predicted UHI severity and thus allows more focused and targeted UHI mitigation actions to be planned, designed, and implemented in the most affected communities. This ultimately results in better decisions, actions, and outcomes to reduce or address the UHI effect.

Finally, while the proposed approach in this paper was applied to the census-tracts in NJ, a similar approach could be easily developed for other census tracts in different states in the US, especially since the needed data is publicly available. Thus, future studies are recommended to implement the developed model to other states since the impacts of the factors are dependent on the characteristics of the studied area and might change from one state to another. Moreover, while the developed model was validated and tested on unseen data and provided high accuracy of 87.88%, the proposed model could benefit from an in-situ validation of the predictions. This could be performed by future studies by installing weather stations or sensors that capture different characteristics of urban and rural areas (such as temperature, rainfall, etc.) at multiple locations in the study area. Then, the UHI can be calculated from the data collected from those weather stations as the LST difference between urban and rural areas and compared to that predicted by the proposed model in this paper.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data used is included in the manuscript

## References

- Abbassi, Y., Ahmadikia, H., & Baniasadi, E. (2022). Impact of wind speed on urban heat and pollution islands. *Urban Climate*, 44, Article 101200.
- Afklari, A., Mirnezhad, M., Ghaffarianhoseini, A., Ghaffarianhoseini, A., Omrany, H., Wang, Z. H., et al. (2017). Urban heat island mitigation strategies: A state-of-the-art review on Kuala Lumpur, Singapore and Hong Kong. *Cities (London, England)*, 62, 131–145.
- Akbari, H. (2009). Cooling our communities. *A guidebook on tree planting and light-colored surfacing*.
- Antonopoulos, C., Trusty, A., & Shandas, V. (2019). The role of building characteristics, demographics, and Urban Heat Islands in shaping residential energy use. *City and Environment Interactions*, 3, Article 100021.
- Assaad, R., Hu, X., & Hussein, M. (2023). Expert knowledge-guided Bayesian belief networks for estimating bridge pile capacity using parametric learning. *ASCE Journal of Bridge Engineering*, 28, Article 04023058.
- Assaf, G., & Assaad, R. (2023a). Quantification of the benefits of different natural Urban Heat Island mitigation initiatives: A systematic review and directions for future research. *Journal of Urban Climate*.

- Assaf, G., & Assaad, R. (2023b). Quantification of the benefits of different engineered Urban Heat Island mitigation initiatives: A systematic review and directions for future research. *Journal of Urban Climate*.
- Assaf, G., & Assaad, R. (2023d). Assessing the vulnerability of communities exposed to climate change-related challenges caused by the Urban Heat Island effect using machine learning. *Computing in Civil Engineering*, 2023, 787–794.
- Assaf, G., & Assaad, R. H. (2023c). Optimal preventive maintenance, repair, and replacement program for catch basins to reduce urban flooding: Integrating agent-based modeling and Monte Carlo simulation. *Sustainability*, 15, 8527. <https://doi.org/10.3390/su15118527>, 2023.
- Azhdari, A., Soltani, A., & Alidadi, M. (2018). Urban morphology and landscape structure effect on land surface temperature: Evidence from Shiraz, a semi-arid city. *Sustainable cities and society*, 41, 853–864.
- Berger, C., Rosentreter, J., Voltersen, M., Baumgart, C., Schmullius, C., & Hese, S. (2017). Spatio-temporal analysis of the relationship between 2D/3D urban site characteristics and land surface temperature. *Remote sensing of environment*, 193, 225–243.
- Bohnenstengel, S. I., Evans, S., Clark, P. A., & Belcher, S. (2011). Simulations of the London Urban Heat Island. *Quarterly Journal of the Royal Meteorological Society*, 137 (659), 1625–1640.
- Chakraborty, T., Hsu, A., Manya, D., & Sheriff, G. (2020b). A spatially explicit surface Urban Heat Island database for the United States: Characterization, uncertainties, and possible applications. *ISPRS Journal of Photogrammetry and Remote Sensing*, 168, 74–88.
- Chakraborty, T. C., Hsu, Angel, Sheriff, Glenn&, & Manya, Diego (2020a). United States surface Urban Heat Island database. *Mendeley Data*, V3. <https://doi.org/10.17632/x9mv4krrnm2.3>
- Chang, C. R., Li, M. H., & Chang, S. D. (2007). A preliminary study on the local cool-island intensity of Taipei city parks. *Landscape and urban planning*, 80(4), 386–395.
- Chen, S., Yang, Y., Deng, F., Zhang, Y., Liu, D., Liu, C., et al. (2022). A high-resolution monitoring approach of canopy Urban Heat Island using a random forest model and multi-platform observations. *Atmospheric Measurement Techniques*, 15(3), 735–756.
- Chen, Y., Chen, R., Hou, J., Hou, M., & Xie, X. (2021). Research on users' participation mechanisms in virtual tourism communities by Bayesian network. *Knowledge-Based Systems*, 226, Article 107161.
- Climate Central. (2022). *Urban Heat Islands*. <https://www.climatecentral.org/climate-e-matters/urban-heat-islands>. Accessed May 22, 2023.
- Comarazamy, D. E., González, J. E., Luvall, J., Rickman, D., & Picón, A. (2006). A validation study of the Urban Heat Island in the tropical coastal city of San Juan, Puerto Rico. *NASA Global Hydrology and Climate Center (GHCC)*, Huntsville, AL.
- Deschenes, O., & Moretti, E. (2009). Extreme weather events, mortality, and migration. *The Review of Economics and Statistics*, 91(4), 659–681.
- Dewan, A., Kiselev, G., Botje, D., Mahmud, G. I., Bhuiyan, M. H., & Hassan, Q. K. (2021). Surface Urban Heat Island intensity in five major cities of Bangladesh: Patterns, drivers and trends. *Sustainable Cities and Society*, 71, Article 102926.
- Emilsson, T. (2021). Urban life and climate change. *The impacts of climate change* (pp. 453–462). Elsevier.
- Equerre, V., Mirzaei, P. A., Riffat, S., & Wang, Y. (2021). Integration of topological aspect of city terrains to predict the spatial distribution of Urban Heat Island using GIS and ANN. *Sustainable Cities and Society*, 69, Article 102825.
- Fan, H., Yu, Z., Yang, G., Liu, T. Y., Liu, T. Y., Hung, C. H., et al. (2019). How to cool hot-humid (Asian) cities with urban trees? An optimal landscape size perspective. *Agricultural and Forest Meteorology*, 265, 338–348.
- Feng, C., Wang, H., Lu, N., & Tu, X. M. (2013). Log transformation: Application and interpretation in biomedical research. *Statistics in medicine*, 32(2), 230–239.
- Friedman, N., & Goldszmidt, M. (1998). Learning Bayesian networks with local structure. *Learning in graphical models* (pp. 421–459). Dordrecht: Springer.
- Gal, T., Lindberg, F., & Unger, J. (2009). Computing continuous sky view factors using 3D urban raster and vector databases: Comparison and application to urban climate. *Theoretical and applied climatology*, 95, 111–123.
- Gao, Z., Zaichik, B. F., Hou, Y., & Chen, W. (2022). Toward park design optimization to mitigate the urban heat island: Assessment of the cooling effect in five US cities. *Sustainable Cities and Society*, 81, Article 103870.
- GE, R. F., XU, K. P., ZHANG, L. X., WANG, X. H., CHI, Y. Y., & WANG, J. J. (2019). On monitoring and identification of hot spots of Urban Heat Island effect—A case study of the sixth-ring zone of Beijing. *Journal of Southwest China Normal University (Natural Science Edition)*, 44(9), 109–117.
- Global Climate Monitor. (2023). Accessed on November 16, 2022. <https://www.globalclimatedata.org/>.
- Guan, H., Soebarto, V., Bennett, J., Clay, R., Andrew, R., Guo, Y., et al. (2014). Response of office building electricity consumption to urban weather in Adelaide, South Australia. *Urban Climate*, 10, 42–55.
- Gupta, N., Mathew, A., & Khandelwal, S. (2019). Analysis of cooling effect of water bodies on land surface temperature in nearby region: A case study of Ahmedabad and Chandigarh cities in India. *The Egyptian Journal of Remote Sensing and Space Science*, 22(1), 81–93.
- He, B. J. (2018). Potentials of meteorological characteristics and synoptic conditions to mitigate Urban Heat Island effects. *Urban climate*, 24, 26–33.
- Heris, M.P., .Foks, N., Bagstad, K., & Troy, A. (2020)., A national dataset of rasterized building footprints for the U.S.: U.S. Geological Survey data release, 10.5066/P9J2Y1WG.
- Hidalgo, J., Masson, V., Baklanov, A., Pigeon, G., & Gimeno, L. (2008). Advances in urban climate modeling. *Annals of the New York Academy of Sciences*, 1146(1), 354–374.
- Hou, H., Su, H., Liu, K., Li, X., Chen, S., Wang, W., et al. (2022). Driving forces of UHI changes in China's major cities from the perspective of land surface energy balance. *Science of the Total Environment*, 829, Article 154710.
- Hu, Y., White, M., & Ding, W. (2016). An urban form experiment on Urban Heat Island effect in high density area. *Procedia Engineering*, 169, 166–174.
- Ivajnić, D., & Žiberna, I. (2019). The effect of weather patterns on winter small city Urban Heat Islands. *Meteorological Applications*, 26(2), 195–203.
- Jacob, D. J., & Winner, D. A. (2009). Effect of climate change on air quality. *Atmospheric environment*, 43(1), 51–63.
- JIANG, Y. (2021). The LA River Fish Passage.
- Kalessar, R. (2022). Newark is one of worst heat islands in the country. One community has nearly no trees. <https://www.northjersey.com/story/news/2022/07/27/urban-heat-island-effect-newark-nj-irondbound-section-one-of-worst-in-country/7629528001/>. Accessed May 22, 2023.
- Kim, S. W., & Brown, R. D. (2021). Urban heat island (UHI) intensity and magnitude estimations: A systematic literature review. *Science of the Total Environment*, 779, Article 146389.
- Kim, Y., Yeo, H., & Kim, Y. (2022). Estimating urban spatial temperatures considering anthropogenic heat release factors focusing on the mobility characteristics. *Sustainable Cities and Society*, 85, Article 104073.
- Kjellstrom, T., Holmer, I., & Lemke, B. (2009). Workplace heat stress, health and productivity—an increasing challenge for low and middle-income countries during climate change. *Global health action*, 2(1), 2047.
- Kotharkar, R., & Surawar, M. (2016). Land use, land cover, and population density impact on the formation of canopy Urban Heat Islands through traverse survey in the Nagpur urban area, India. *Journal of Urban Planning and Development*, 142(1), Article 04015003.
- Lauwaet, D., De Ridder, K., Saeed, S., Brisson, E., Chatterjee, F., Van Lipzig, N. P. M., et al. (2016). Assessing the current and future Urban Heat Island of Brussels, 15 pp.
- Lee, S., Moon, H., Choi, Y., & Yoon, D. K. (2018). Analyzing thermal characteristics of urban streets using a thermal imaging camera: A case study on commercial streets in Seoul, Korea. *Sustainability*, 10(2), 519.
- Li, C., Lu, L., Fu, Z., Sun, R., Pan, L., Han, L., et al. (2022). Diverse cooling effects of green space on Urban Heat Island in tropical megacities. *Frontiers in Environmental Science*, 10, 2400.
- Li, J., Song, C., Cao, L., Zhu, F., Meng, X., & Wu, J. (2011). Impacts of landscape structure on surface Urban Heat Islands: A case study of Shanghai, China. *Remote sensing of environment*, 115(12), 3249–3263.
- Li, X., Gong, P., Zhou, Y., et al. (2020). Mapping global urban boundaries from the global artificial impervious area (GAIA) data. *Environmental Research Letters*, 15.
- Lin, J., Qiu, S., Tan, X., & Zhuang, Y. (2023). Measuring the relationship between morphological spatial pattern of green space and Urban Heat Island using machine learning methods. *Building and Environment*, 228, Article 109910.
- Lin, M., Dong, J., Jones, L., Liu, J., Lin, T., Zuo, J., et al. (2021). Modeling green roofs' cooling effect in high-density urban areas based on law of diminishing marginal utility of the cooling efficiency: A case study of Xiamen Island. *China. Journal of Cleaner Production*, 316, Article 128277.
- Lin, Y., Wang, Z., Jim, C. Y., Li, J., Deng, J., & Liu, J. (2020). Water as an urban heat sink: Blue infrastructure alleviates Urban Heat Island effect in mega-city agglomeration. *Journal of Cleaner Production*, 262, Article 121411.
- Lindén, J. (2011). Nocturnal Cool Island in the Sahelian city of Ouagadougou, Burkina Faso. *International Journal of Climatology*, 31(4), 605–620.
- Liu, S., Zhang, J., Li, J., Li, Y., Zhang, J., & Wu, X. (2021). Simulating and mitigating extreme Urban Heat Island effects in a factory area based on machine learning. *Building and Environment*, 202, Article 108051.
- Liuy, Y., & Chen, Q. (2020). Bayesian inference of finite population quantiles for skewed survey data using skew-normal penalized spline regression. *Journal of Survey Statistics and Methodology*, 8(4), 792–816.
- Loughner, C. P., Allen, D. J., Zhang, D. L., Pickering, K. E., Dickerson, R. R., & Landry, L. (2012). Roles of urban tree canopy and buildings in Urban Heat Island effects: Parameterization and preliminary results. *Journal of Applied Meteorology and Climatology*, 51(10), 1775–1793.
- Lu, L., Fu, P., Dewan, A., & Li, Q. (2023). Contrasting determinants of land surface temperature in three megacities: Implications to cool tropical metropolitan regions. *Sustainable Cities and Society*, 92, Article 104505.
- Ma, X., & Peng, S. (2021). Assessing the quantitative relationships between the impervious surface area and surface heat island effect during urban expansion. *Pearl J*, 9, e11854.
- MacAllister, A., Kohl, A., & Winer, E. (2020). Using high-fidelity meta-models to improve performance of small dataset trained Bayesian Networks. *Expert Systems with Applications*, 139, Article 112830.
- Magee, N., Curtis, J., & Wendler, G. (1999). The Urban Heat Island effect at Fairbanks, Alaska. *Theoretical and applied climatology*, 64(1), 39–47.
- Maharjan, M., Aryal, A., Man Shakya, B., Talchabhadel, R., Thapa, B. R., & Kumar, S. (2021). Evaluation of Urban Heat Island (UHI) using satellite images in densely populated cities of South Asia. *Earth*, 2(1), 86–110.
- Mathew, A., Khandelwal, S., & Kaul, N. (2016). Spatial and temporal variations of Urban Heat Island effect and the effect of percentage impervious surface area and elevation on land surface temperature: Study of Chandigarh city. *India. Sustainable Cities and Society*, 26, 264–277.
- Mendez-Astudillo, J., & Mendez-Astudillo, M. (2021). A machine learning approach to monitoring the UHI From GNSS data. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–11.

- Ming, Y., Liu, Y., Gu, J., Wang, J., & Liu, X. (2023). Nonlinear effects of urban and industrial forms on surface Urban Heat Island: Evidence from 162 Chinese prefecture-level cities. *Sustainable Cities and Society*, 89, Article 104350.
- Mirzaei, P. A. (2015). Recent challenges in modeling of Urban Heat Island. *Sustainable cities and society*, 19, 200–206.
- Mishra, S., Mallick, P. K., Jena, L., & Chae, G. S. (2020). Optimization of skewed data using sampling-based preprocessing approach. *Frontiers in Public Health*, 8, 274.
- Mohammad, P., Goswami, A., Chauhan, S., & Nayak, S. (2022). Machine learning algorithm based prediction of land use land cover and land surface temperature changes to characterize the surface Urban Heat Island phenomena over Ahmedabad city, 42. India: Urban Climate, Article 101116.
- Mpkairi, K. S., & Muvengwi, J. (2019). Night-time lights and their influence on summer night land surface temperature in two urban cities of Zimbabwe: A geospatial perspective, 29. Urban Climate, Article 100468.
- Multi-Resolution Land Characteristics Consortium (MLRC). (2019a). Dataset Type: Urban Imperviousness. Accessed on November 15, 2022. <https://www.mrlc.gov/data?f%5B0%5D=category%3ATree%20Canopy&f%5B1%5D=region%3Aconus&f%5B2%5D=year%3A2016>.
- Multi-Resolution Land Characteristics Consortium (MLRC). (2019b). Dataset Type: Tree Canopy. Accessed on November 15, 2022. [https://www.mrlc.gov/data?f%5B0%5D=\\_region%3Aconus&f%5B1%5D=year%3A2016](https://www.mrlc.gov/data?f%5B0%5D=_region%3Aconus&f%5B1%5D=year%3A2016).
- Nationsonline. (2023). "Population of the US States and principal US territories." Accessed on April 16, 2023. <https://www.nationsonline.org/oneworld/US-states-population.htm>.
- New Jersey Office of GIS (NJOBGIS) (2021). Road centerlines of NJ - Next Gen 911 (shp). Accessed on November 15, 2022. <https://www.arcgis.com/home/item.html?id=2b11e237a6e149c0a94027780c22c9c4>.
- Okumus, D. E., & Terzi, F. (2021). Evaluating the role of urban fabric on surface Urban Heat Island: The case of Istanbul. *Sustainable Cities and Society*, 73, Article 103128.
- Oliveira, A., Lopes, A., Niza, S., & Soares, A. (2022). An urban energy balance-guided machine learning approach for synthetic nocturnal surface Urban Heat Island prediction: A heatwave event in Naples. *Science of the total environment*, 805, Article 150130.
- O'Malley, C., Piroozfar, P., Farr, E. R., & Pomponi, F. (2015). Urban Heat Island (UHI) mitigating strategies: A case-based comparative analysis. *Sustainable cities and society*, 19, 222–235.
- O'Neill, Z., & Niu, F. (2017). Uncertainty and sensitivity analysis of spatio-temporal occupant behaviors on residential building energy usage utilizing Karhunen-Loeve expansion. *Building and Environment*, 115, 157–172.
- Okawara, G. Y., Krecl, P., & Targino, A. C. (2022). Fine-scale modeling of the Urban Heat Island: A comparison of multiple linear regression and random forest approaches. *Science of the total environment*, 815, Article 152836.
- Patz, J. A., Campbell-Lendrum, D., Holloway, T., & Foley, J. A. (2005). Impact of regional climate change on human health. *Nature*, 438(7066), 310–317.
- Peng, J., Liu, Q., Xu, Z., Lyu, D., Du, Y., Qiao, R., et al. (2020). How to effectively mitigate Urban Heat Island effect? A perspective of waterbody patch size threshold. *Landscape and Urban Planning*, 202.
- Qizhi, M., Ying, L., & Kang, W. (2016). Spatio-temporal changes of population density and urbanization pattern in China (2000–2010). *China City Planning Review*, 25(4).
- Ramírez-Aguilar, E. A., & Souza, L. C. L. (2019). *Urban form and population density: Influences on Urban Heat Island intensities in Bogotá*, 29. Colombia: Urban Climate, Article 100497.
- Rasul, A., Balzter, H., & Smith, C. (2016). Diurnal and seasonal variation of surface urban cool and heat islands in the semi-arid city of Erbil, Iraq. *Climate*, 4(3), 42.
- Schwarz, N., & Manceur, A. M. (2015). Analyzing the influence of urban forms on surface Urban Heat Islands in Europe. *Journal of Urban Planning and Development*, 141(3), Article A4014003.
- Shahmohamadi, P., Che-Ani, A. I., Maulud, K. N. A., Tawil, N. M., & Abdullah, N. A. G. (2011). The impact of anthropogenic heat on formation of Urban Heat Island and energy consumption balance. *Urban Studies Research*, 2011.
- Shen, C., Hou, H., Zheng, Y., Murayama, Y., Wang, R., & Hu, T. (2022). Prediction of the future Urban Heat Island intensity and distribution based on landscape composition and configuration: A case study in Hangzhou. *Sustainable Cities and Society*, 83, Article 103992.
- Shi, D., Song, J., Huang, J., Zhuang, C., Guo, R., & Gao, Y. (2020). Synergistic cooling effects (SCEs) of urban green-blue spaces on local thermal environment: A case study in Chongqing, China. *Sustainable Cities and Society*, 55, Article 102065.
- Sinimaa, M., Spichakova, M., Belikov, J., & Petlenkov, E. (2021). Feature engineering of weather data for short-term energy consumption forecast. 2021 IEEE Madrid powertech (pp. 1–6). IEEE.
- Solecki, W. D., Rosenzweig, C., Marshall, L., Pope, G., Clark, M., Cox, J., et al. (2005). Mitigation of the heat island effect in urban New Jersey. *Global Environmental Change Part B: Environmental Hazards*, 6(1), 39–49.
- Sun, R., & Chen, L. (2012). How can urban water bodies be designed for climate adaptation? *Landscape and Urban Planning*, 105(1–2), 27–33.
- Tamaskani Esfahanlalateh, A., Ngarambe, J., & Yun, G. Y. (2021). Influence of tree canopy coverage and leaf area density on Urban Heat Island mitigation. *Sustainability*, 13(13), 7496.
- Tang, J., Di, L., Xiao, J., Lu, D., & Zhou, Y. (2017). Impacts of land use and socioeconomic patterns on urban heat island. *International Journal of Remote Sensing*, 38(11), 3445–3465.
- Tsai, A. C., Liou, M., Simak, M., & Cheng, P. E. (2017). On hyperbolic transformations to normality. *Computational Statistics & Data Analysis*, 115, 250–266.
- U.S. Census Bureau. (2020a). NEW JERSEY: 2020 Census. Accessed on April 16, 2023. <https://www.census.gov/library/stories/state-by-state/new-jersey-population-change-between-census-decade.html>.
- U.S. Census Bureau. (2020b). Planning Database. Accessed on November 16, 2022. <https://www.census.gov/topics/research/guidance/planning-databases.2015.html#list-tab-8RWUXE4SFOMS3LQY8P>.
- U.S. Environmental EPA. Environmental Protection Agency. (2022). Learned about heat islands. Accessed on February 13, 2023. <https://www.epa.gov/heatislands/learn-about-heat-islands>.
- Velikova, M., Lucas, P. J., Samulski, M., & Karssemeijer, N. (2013). On the interplay of machine learning and background knowledge in image interpretation by Bayesian networks. *Artificial intelligence in medicine*, 57(1), 73–86.
- Vulova, S., Meier, F., Fenner, D., Nouri, H., & Kleinschmit, B. (2020). Summer nights in Berlin, Germany: Modeling air temperature spatially with remote sensing, crowdsourced weather data, and machine learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 5074–5087.
- Wang, Y., Zhan, Q., & Ouyang, W. (2019). How to quantify the relationship between spatial distribution of urban waterbodies and land surface temperature? *Science of The Total Environment*, 671, 1–9.
- Weng, Q., Lu, D., & Schubring, J. (2004). Estimation of land surface temperature–vegetation abundance relationship for Urban Heat Island studies. *Remote sensing of Environment*, 89(4), 467–483.
- Wu, X., Zhang, L., & Zang, S. (2019). Examining seasonal effect of Urban Heat Island in a coastal city. *PLoS One*, 14(6), Article e0217850.
- Xie, X., Zhang, D. M., Huang, H. W., Zhou, M. L., Lacasse, S., & Liu, Z. Q. (2021). Data fusion-based dynamic diagnosis for structural defects of shield tunnel. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 7(2), Article 04021019.
- Xin, J., Yang, J., Wang, L. E., Jin, C., Xiao, X., & Xia, J. C. (2022). Seasonal differences in the dominant factors of surface Urban Heat Islands along the urban-rural gradient. *Frontiers in Environmental Science*, 1681.
- Yang, P., Ren, G., & Hou, W. (2019). Impact of daytime precipitation duration on Urban Heat Island intensity over Beijing city. *Urban climate*, 28, Article 100463.
- Yoo, S. (2018). Investigating important urban characteristics in the formation of Urban Heat Islands: A machine learning approach. *Journal of Big Data*, 5(1), 1–24.
- Yousefi, L., Swift, S., Arzoky, M., Saachi, L., Chiovato, L., & Tucker, A. (2021). Opening the black box: Personalizing type 2 diabetes patients based on their latent phenotype and temporal associated complication rules. *Computational Intelligence*, 37(4), 1460–1498.
- Yu, Q., Ji, W., Pu, R., Landry, S., Acheampong, M., O'Neil-Dunne, J., et al. (2020). A preliminary exploration of the cooling effect of tree shade in urban landscapes. *International Journal of Applied Earth Observation and Geoinformation*, 92, Article 102161.
- Yuan, F., & Bauer, M. E. (2007). Comparison of impervious surface area and normalized difference vegetation index as indicators of surface Urban Heat Island effects in Landsat imagery. *Remote Sensing of environment*, 106(3), 375–386.
- Yuan, J., Emura, K., & Farnham, C. (2017). Is urban albedo or urban green covering more effective for urban microclimate improvement?: A simulation for Osaka. *Sustainable Cities and Society*, 32, 78–86.
- Yuan, Y., Li, C., Geng, X., Yu, Z., Fan, Z., & Wang, X. (2022). Natural-anthropogenic environment interactively causes the surface Urban Heat Island intensity variations in global climate zones. *Environment international*, 170, Article 107574.
- Yun, G. Y., Ngarambe, J., Duhirwe, P. N., Ulpiani, G., Paolini, R., Haddad, S., et al. (2020). Predicting the magnitude and the characteristics of the Urban Heat Island in coastal cities in the proximity of desert landforms. The case of Sydney. *Science of The Total Environment*, 709, Article 136068.
- Zhang, D. L., Shou, Y. X., & Dickerson, R. R. (2009). Upstream urbanization exacerbates Urban Heat Island effects. *Geophysical research letters*, 24, 36.
- Zhang, L., Geisler, T., Ray, H., & Xie, Y. (2022). Improving logistic regression on the imbalanced data by a novel penalized log-likelihood function. *Journal of Applied Statistics*, 49(13), 3257–3277.
- Zhang, L., Wu, X., Skibniewski, M. J., Zhong, J., & Lu, Y. (2014). Bayesian-network-based safety risk analysis in construction projects. *Reliability Engineering & System Safety*, 131, 29–39.
- Zhou, Y., Li, X., Chen, W., Meng, L., Wu, Q., Gong, P., et al. (2022). Satellite mapping of urban built-up heights reveals extreme infrastructure gaps and inequalities in the Global South. *Proceedings of the National Academy of Sciences*, 119(46), Article e2214813119.
- Zhu, J., & Collette, M. (2015). A dynamic discretization method for reliability inference in Dynamic Bayesian Networks. *Reliability Engineering & System Safety*, 138, 242–252.