Explainable GeoAI and statistical analysis reveal complementary insights about disparities of 311 help requests during the 2022 Buffalo blizzard

Ryan Zhenqi Zhou[1], Yingjie Hu[1,2*], Kai Sun[1], Ryan Muldoon[3], Susan Clark[4], and Kenneth Joseph[2]

[1] GeoAI Lab, Department of Geography, University at Buffalo, Buffalo, NY 14261, USA

[2] Department of Computer Science and Engineering, University at Buffalo, Buffalo, NY 14261, USA

[3] Department of Philosophy, University at Buffalo, Buffalo, NY 14261, USA

[4] Department of Environment and Sustainability, University at Buffalo, Buffalo, NY 14261, USA

Abstract:

The 2022 Buffalo blizzard was a catastrophic winter storm that struck Buffalo, New York in the week of Christmas in 2022. It claimed 47 lives and left much of the region stranded for the holiday week. In this disaster, the 311 call service was used by many residents to request help for issues due to the blizzard. This study examines these 311 help requests and their potential disparities across communities. Specifically, we aim to: (1) understand the spatial and temporal distributions of different types of 311 help requests; (2) identify the physical and social vulnerability factors, as well as human behavior factors, that are associated with the use of 311 calls. Methodologically, we leverage both explainable geospatial artificial intelligence (GeoAI) methods and statistical analysis to analyze 311 help requests and their associated factors. Our analysis shows significant spatial disparities in 311 help requests across communities. Results from explainable GeoAI and statistical analysis also reveal complementary insights on key factors associated with 311 help requests, such as historical 311 request behavior and percentage of minority population. These results could inform future disaster management decisions and help mitigate the negative impacts of winter storm disasters.

Keywords:

Winter storm; 311; disaster management; explainable AI; statistical analysis.

** This is a preprint. The official version is at:* https://doi.org/10.1016/j.ijdrr.2025.105635

1. Introduction

The 2022 Buffalo blizzard was a catastrophic winter storm that slammed the city of Buffalo and the surrounding region during the week of Christmas. It came in with hurricane-force winds and brought severe whiteout conditions, over 50 inches of snow, and wind chill temperatures 30 degrees below zero Fahrenheit [1]. Despite the fact that Buffalo is a city highly experienced with managing snow, this disaster claimed the lives of 47 people and left much of the region stranded during Christmas [2,3]. Understanding the impacts of this blizzard on local communities can inform future disaster management decisions and help mitigate the negative impacts of winter storm disasters.

In this disaster, the 311 call service was one of the major channels used by many residents to request help for issues due to the blizzard [1,4]. Examples of these 311 help requests include snow plowing, removing trees that were damaged or blown down by the hurricane-force winds, and delayed garbage pickups due to roads being impassable from snow. While issues reported in 311 calls are probably less severe as some other emergencies (e.g., deaths and hospitalizations), they reflect various disruptions caused by a disaster on people and their communities. Studying these 311-reported issues, therefore, can help improve our understanding of these disruptions and their potential inequities, and may inform future city investments to enhance community resilience. Thanks to the Open Data Initiatives in many U.S. cities [5], 311 call data are often made publicly available, including in the City of Buffalo. This study uses 311 help requests as a lens to understand the impacts of the 2022 blizzard on communities and the potential impact disparities.

Previous research has utilized 311 call data to study disasters, which has focused on three main topics. The first topic is assessing disaster impacts and community needs. For example, 311 call data has been used to assess the impacts of flooding in Hurricane Harvey [6], to identify short-term and long-term community needs following Hurricane Sandy [7], and to examine the impacts of burst water pipes during the Texas winter storm in 2021 [8]. The second topic is evaluating the effectiveness of government agencies in responding to disaster-related issues. Zobel et al. [9] developed a metric based on the number of 311 requests that have not been addressed each day for comparing the performance of different government agencies in New York City following several disasters, such as Hurricanes Irene and Sandy. The third topic is predicting future community needs based on historical 311 call data. Researchers used spatial and temporal patterns of historical 311 calls and also developed methods to predict potential future needs and to help government agencies better prepare for future disasters [10,11].

While 311 call data has been used in other previous disasters, this current study investigates the 311 help requests in the catastrophic Buffalo blizzard in 2022. More broadly, this study is among the smaller number of studies that examine the impacts of winter storm disasters. Winter storms have led to significant damages and disruptions in not only Buffalo but also many other geographic regions. For example, the Texas winter storm in 2021 was considered as the costliest natural disaster in the history of Texas [12,13]. The 2024 Iowa blizzard in January brought the coldest temperatures in over one hundred years in the region and stranded hundreds of thousands

of households in the state [14,15]. This study, therefore, contributes to winter storm disaster research by improving our understanding of winter storm impacts and the potential impact disparities across communities.

More specifically, we aim to answer two research questions (RQs) in this study:

RQ1: *What are the spatial and temporal distributions of the 311 help requests for different issues due to the blizzard?*

RQ2: *What are the factors, such as physical and social vulnerability factors and human behavior factors, that are associated with the 311 help requests?*

In addition to answering the two RQs, this study also explores the use of new methodologies for disaster research. In particular, we explore the combination of explainable geospatial artificial intelligence (GeoAI) methods and statistical analysis to identify factors associated with 311 help requests. There has been an increasing interest in using AI for disaster research [16–18]. GeoAI methods, i.e., methods that integrate geospatial principles and AI, are promising new approaches for analyzing the geospatial data that are frequently involved in disaster research [19,20]. In this study, we use GeoAI methods and statistical analysis in a complementary manner: GeoAI methods are used for identifying factors for predicting potential future 311 help requests, while statistical analysis methods are used for identifying factors associated with existing 311 help requests. An explainable AI framework is further employed to explain the results of the GeoAI methods.

The remainder of this paper is organized as follows. Section 2 describes the study area and data. Section 3 presents our study design as well as the GeoAI and statistical methods used for data analysis. Section 4 presents the results, including the spatial and temporal distributions of the 311 help requests and the complementary results from the GeoAI and statistical methods. Section 5 discusses the implications of the results from both a disaster management perspective and a methodological perspective. Finally, Section 6 concludes this work.

2. Study area and data

2.1. Study area and time period

Our study area is the City of Buffalo which was severely affected by the 2022 blizzard [1,3]. We use census block group (CBG) as the geographic unit of our analysis, which offers a relatively fine spatial resolution with socioeconomic and demographic data available from the U.S. Census Bureau. We use CBGs to approximate communities in this study. Figure 1 shows the boundaries of the City of Buffalo and the CBGs.

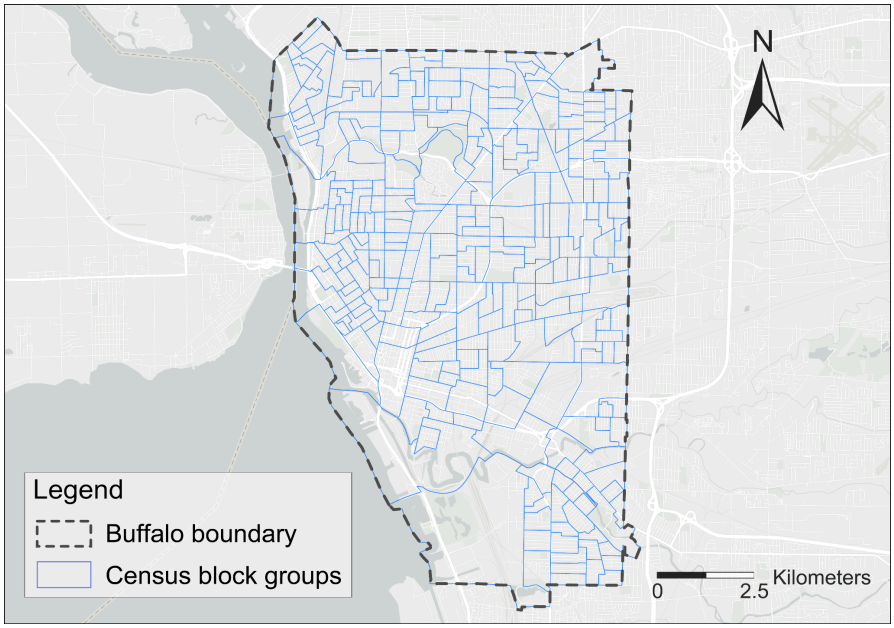

Figure 1. The boundaries of Buffalo and the census block groups in this study. Data were obtained from the 2022 TIGER/line Shapefile from the U.S. Census Bureau.

The time period of this study is from December 19, 2022 to January 15, 2023. We further divide this time period into three stages: *preparation stage* (December 19 to 22, 2022), *blizzard stage* (December 23 to 25, 2022), and *recovery stage* (December 26, 2022 to January 15, 2023). These three stages are determined based on the timeline of the blizzard and the response activities as documented by the reports from the City of Buffalo [1] and the New York State Division of Homeland Security and Emergency Services [3].

2.2. Data

We use the following data to study the impacts of the blizzard through the lens of 311 help requests and to examine the factors associated with these requests:

311 help request data. This dataset contains 311 help requests in the City of Buffalo since the year of 2008. We obtained this dataset from the Open Data Portal of Buffalo. The whole dataset is organized as a large comma-separated value (CSV) file. Each row is a data record representing a 311-reported issue that needs help, and each column represents an attribute of the issue. The attributes include the reference number of the 311 call, its open date, close date, reason of this 311 request (e.g., snow plowing, damaged trees, and delayed garbage pickup), the ID and address of the property having the issue, and the CBG where the property is located. The 311 data does not have demographic information about the callers. We focus on the data in the study time period, i.e., between December 19, 2022 and January 15, 2023.

Property assessment roll data. Because most 311 help requests are linked to specific properties, we use the property assessment roll data from the City of Buffalo to obtain more information about properties. This dataset provides detailed attributes about properties in the city,

including the ID and address of the property, property type (e.g., whether a property is residential, commercial, or manufacturing), year built, the assessed value of the property, and the CBG where the property is located. Values in this dataset (e.g., assessed property value) are based on a snapshot as of December 1st, 2022. We focus on residential properties only in this dataset, and derive three CBG-level property information: the total number of residential properties in a CBG, the median residential property value, and their median year built.

Physical vulnerability data. We use snow depth data from the Snow Data Assimilation System (SNODAS) provided by the National Snow and Ice Data Center to represent the physical vulnerability of different CBGs during this blizzard. This dataset provides estimates of snow depth with a temporal resolution of daily and a spatial resolution of 1 kilometer. In addition to snow depth, we have also considered using wind speed to capture another aspect of physical vulnerability related to this blizzard. However, there is a lack of high-resolution wind speed data for the study area. For example, the wind speed data from ERA5 and ERA5-Land have coarse spatial resolutions of 31 kilometers and 9 kilometers respectively, resulting in the entire study area being mostly covered by a single pixel. Given this lack of high-resolution wind speed data, we focus on using snow depth data to represent physical vulnerability. We note that physical vulnerability in this study refers to the exposure of communities to the disaster (e.g., heavy snow) as used in some previous studies [21,22]. Physical vulnerability here does not imply structural or systemic weaknesses in the infrastructures or buildings of the affected communities, which are also frequently used in the literature [23].

Social vulnerability data. For social vulnerability, we use CBG-level socioeconomic and demographic data from the American Community Survey (ACS) of the Census. Four categories of social vulnerability variables are selected in this study, which are: (1) socioeconomic status, (2) household composition and disability, (3) minority status and language, and (4) housing and transportation. These four categories of variables are selected largely based on the Social Vulnerability Index (SVI) of the Centers for Disease Control and Prevention (CDC) [24] and the Social Vulnerability Index (SoVI) developed by Cutter et al. [25]. These two indices focus on factors that can adversely affect communities during disasters by taking into account the multidimensional nature of vulnerability. We collect CBG-level data for variables in these four categories from the website of the U.S. Census Bureau.

Historical 311 request behavior data. The past behavior of people in using the 311 call service may affect their use of 311 during the blizzard. With this consideration, we also use help requests before the blizzard to represent the historical 311 request behaviors of different CBGs. We extract this historical data from the same 311 dataset from the City of Buffalo. Specifically, we use the 311 help requests from the previous snow season in 2021 (from November 1, 2021 to February 28, 2022) and also from the snow season of 2022 but before the blizzard (from November 1, 2022 to December 18, 2022). We extract the historical 311 request data for each CBG. We also extract the historical 311 request data in nearby CBGs (i.e., CBGs adjacent to the current CBG) based on the

consideration that the 311 call behavior of people in nearby CBGs might influence the behaviors of people in the current CBG as well.

3. Methods

3.1. Overview of the study design

We aim to answer two research questions in this study: RQ1: *What are the spatial and temporal distributions of the 311 help requests for different issues due to the blizzard?* RQ2: *What are the factors, such as physical and social vulnerability factors and human behavior factors, that are associated with the 311 help requests?* To answer these two RQs, we design this study into two parts, as shown in Figure 2. To answer RQ1, we perform spatial and temporal visualizations and analysis on the 311 call data to understand their spatial and temporal patterns and disparities across communities. To answer RQ2, we use statistical models and machine learning models to examine the factors associated with 311 help requests. We test three spatial statistical models which are Spatial Lag Model (SLM), Spatial Error Model (SEM), and Geographically Weighted Regression (GWR) [22,26]. We also test three machine learning models which are Support Vector Machine (SVM), Random Forest (RF), and Geographical Random Forests (GRF) [27,28]. Among the three machine learning models, the GRF model is considered as a GeoAI model since it integrates geospatial principles with an AI model [29]. The best statistical model and the best machine learning model are then selected for result interpretation. An explainable AI framework, SHAP (SHapley Additive exPlanations) [30], is employed to explain the result of the best machine learning model. We choose to use statistical models first in our analysis since they have been widely used in the literature [22,31,32], while machine learning and AI models have gained increasing attention in more recent years. Therefore, this analytical sequence allows us to first obtain results from more established statistical models with good interpretability, and then compare the results with those from machine learning models to understand their similarities and differences.

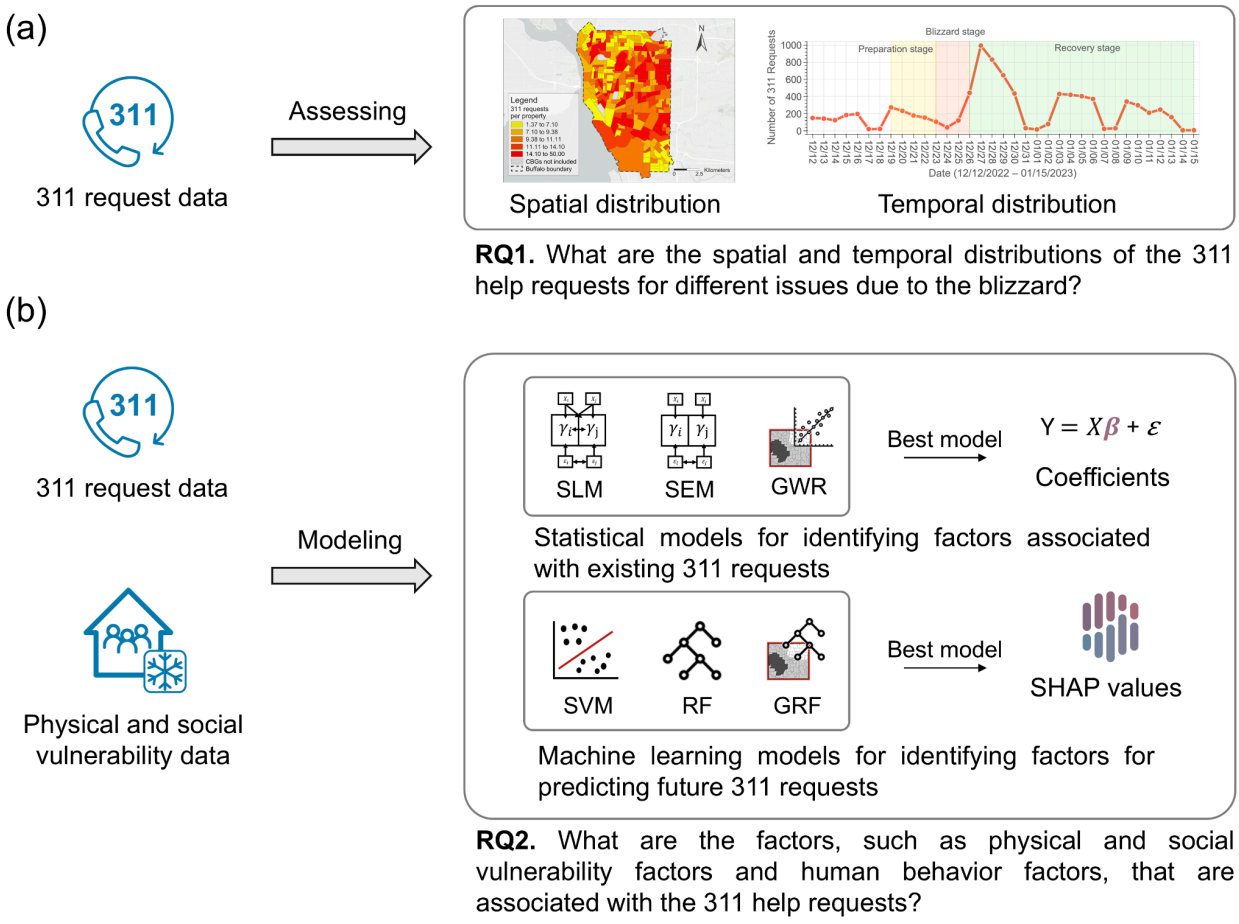

Figure 2. An overview of the study design. (a) RQ1; (b) RQ2.

3.2. Data preprocessing

3.2.1. Preprocessing 311 help request data

We perform two preprocessing steps for the 311 help request data. First, we remove duplicate 311 calls about the same issues from the data. These duplicate calls were likely made by some residents when their reported issues were not solved after some time. These calls share the pattern in that they were about the same issue and the same household address, and they were closed at exactly the same time (at the same seconds). We merge these duplicate calls to keep only one record per issue to reduce the potential influence of some residents who made multiple calls about the same issues. 326 duplicated calls are removed in this step. Second, we filter the 311 data to focus on requests for residential properties only. We focus on residential requests because most people were stuck at home during the blizzard, and they were likely to report issues affecting their residence. About 80% of the 311 requests during the study time period are linked to residential properties. While there also exist 311 requests about other types of properties (e.g., manufacturing facility, road intersection, and vacant land), the reporting of issues at these properties largely depends on whether a person happened to be there, and may not represent most issues at these

types of properties during the blizzard period. With this consideration, we focus on 311 help requests linked to residential properties. A total of 7559 requests are included in our analysis.

After these two data preprocessing steps, we aggregate the 311 requests to CBGs to obtain CBG-level requests. Since CBGs with more properties may have more 311-reported issues by chance, we further normalize the number of 311 requests by the number of residential properties in the CBG using Equation (1) to obtain requests per property during the blizzard:

$$rpp_j = \frac{r_j}{p_j} \times 100 \quad , \quad (1)$$

where rpp_j refers to requests per property in CBG j; r_j is the number of requests linked to residential properties in CBG j; and p_j is the number of residential properties in CBG j. To further increase the robustness of our analysis, we remove the CBGs that have fewer than 20 residential properties to reduce the possible biases that may be introduced by the small numbers of properties. In total, 4 CBGs are removed and 286 CBGs are included in this study.

In addition to calculating 311 requests per property during the 2022 blizzard period, we also calculate historical 311 requests per property in each CBG from the previous snow seasons in 2021 and 2022. Historical 311 requests per property in the adjacent CBGs of a target CBG are also calculated. These variables capture the historical 311 request behaviors of residents in the current and surrounding CBGs, which may affect their 311 requests during the 2022 blizzard.

3.2.2. Preprocessing physical and social vulnerability data

For physical vulnerability, the raw snow depth data from SNODAS is raster data with a temporal resolution of daily and a spatial resolution of 1 kilometer. To obtain CBG-level snow depth information, we first calculate the mean snow depth value for each pixel over the study time period. We then obtain the mean snow depth for each CBG using a weighted average approach. The snow depth of a CBG is calculated by averaging the mean snow depths of the pixels overlapping with the current CBG and weighted based on their overlapping areas. These operations are completed using the software ArcGIS Pro.

For social vulnerability, we use data values directly from the ACS of the US Census, or perform simple calculations to obtain percentage values when needed. For example, we calculate the percentage of individuals below poverty, the percentage of civilian unemployed, and the percentage of persons with no high school diploma based on the corresponding individual numbers in the ACS data. We calculate the median year built of residential properties in a CBG and the median value of residential properties in a CBG using the property assessment roll data.

3.2.3. Summary of prepared variables

Table 1 summarizes the variables prepared from the datasets. The dependent variable is residential *311 requests per property (pp)* during the studied blizzard time period, and the independent variables contain 22 variables covering physical vulnerability, social vulnerability, and historical

311 request behaviors. All variables are prepared at the CBG level. We also conduct tests to check the multicollinearity of the independent variables. We calculate the Variance Inflation Factor (VIF) values, and the VIF values of all 22 variables are below the typical cut-off value of 5, suggesting low multicollinearity. Therefore, we use all 22 variables as the independent variables.

Table 1. Notations and descriptions of the prepared variables.

Variable Notations	**Descriptions**
Dependent Variable	
311 requests per property (pp)	Number of residential 311 requests per property during the 2022 blizzard time period (from Dec. 19, 2022 to Jan. 15, 2023)
Independent Variables	
Physical vulnerability	
Snow depth	Mean snow depth during the blizzard time period
Social vulnerability	
(1) Socioeconomic status	
% poverty	Percentage of population below the federally defined poverty line
% unemployed	Percentage of unemployed civilian population age 16 or over
Income	Per capita income (in dollars)
% < highschool	Percentage of population age 25 or over without high school completion
(2) household composition and disability	
% age ≥ 65	Percentage of population equal or over age 65
% age < 18	Percentage of population below age 18
% disability	Percentage of households with 1 or more persons with a disability
% single parent	Percentage of households that are male or female householders with no spouse present and children under 18
(3) Minority status and language	
% minority	Percentage of non-White population
% not well English	Percentage of population age 5 or over who speak English “not well” or “not at all”
(4) Housing and transportation	

% single-unit structure	Percentage of detached single housing units
% multi-unit structure	Percentage of housing units with 10 or more units in structure
% mobile homes	Percentage of housing units that are mobile homes
% owner-occupied housing units	Percentage of owner-occupied housing units
% crowding housing units	Percentage of occupied housing units with more than one person per room
% in group quarters	Percentage of population in group quarters (e.g., correctional institutions, college dormitories, and military quarters)
% no vehicle available	Percentage of households with no vehicle available
Median year built	Median year built of the residential properties in a CBG based on the Buffalo property assessment roll data
Median property value	Median value of the residential properties in a CBG based on the Buffalo property assessment roll data
Historical 311 request behavior	
Historical 311 requests pp	Historical number of residential 311 requests per property during the previous snow seasons in 2021 and 2022 before the blizzard
Nearby-CBG historical 311 requests pp	Historical number of residential 311 requests per property of the adjacent CBGs. The number of 311 requests of the target CBG is not included

3.3. Statistical analysis

We use statistical analysis to identify important factors among the 22 independent variables that are associated with the dependent variable. The purpose of the statistical analysis here is inference, i.e., to infer the associations between the independent variables related to vulnerability and human behaviors and the dependent variable of 311 requests per property. Considering the spatial autocorrelations commonly existing in geographic data, we use three spatial statistical models, which are spatial lag model, spatial error model, and geographically weighted regression. These spatial statistical models have also been used in previous disaster research involving geographic data to accommodate the effects of spatial autocorrelation [22,32,31]. In the following, we briefly describe each model.

Spatial Lag Model (SLM): SLM is a variant of linear regression that takes into account spatial autocorrelation. In SLM, the dependent variable is influenced by not only the independent variables but also its own spatially lagged values. The SLM model used in this study is in the form of Equation (2):

$$\mathcal{Y} = \theta_0 + \rho W\mathcal{Y} + \theta_p P + \theta_s S + \theta_h H + \varepsilon \quad , \qquad (2)$$

where $\mathcal{Y}$ is the dependent variable, which is the number of 311 requests per property during the blizzard. θ_0 is the constant, W is the spatial weight matrix, ρ is the coefficient of spatial autoregressive term, θ_p, θ_s, and θ_h are the regression coefficients for the variables in physical vulnerability, social vulnerability, and historical 311 request behavior respectively, and ε is the error term. Note that each of θ_s and θ_h contains multiple coefficients for the multiple variables in that group. We use Python and the *spreg* library to implement SLM [33]. To configure the model, we apply the commonly used Queen's case contiguity for the spatial weight matrix W.

Spatial Error Model (SEM): Similar to SLM, SEM also takes spatial autocorrelation into account. Different from SLM, SEM models spatial dependencies in the error structure, rather than in the dependent variable itself. The SEM model used in this study is in the form of Equation (3):

$$\mathcal{Y} = \theta_0 + \theta_p P + \theta_s S + \theta_h H + (\lambda W \varepsilon + u) \quad , \tag{3}$$

where the $\mathcal{Y}$, θ_0, θ_p, θ_s, and θ_h have the same meaning as in SLM; λ is the spatial autoregressive parameter for the error component; W is the spatial weights matrix, ε is the error term, and u represents the vector of residuals. Like SLM, we use the *spreg* library to implement the SEM model, and use Queen's case contiguity for the spatial weight matrix W.

Geographically Weighted Regression (GWR): GWR extends the traditional linear regression model by fitting local linear regression models for each geographic unit [26]. In this study, each geographic unit is a CBG. The GWR model is in the form of Equation (4):

$$\mathcal{Y} = \theta_0(x_i, y_i) + \theta_p(x_i, y_i)P + \theta_s(x_i, y_i)S + \theta_h(x_i, y_i)H + \varepsilon_i \quad , \tag{4}$$

where (x_i, y_i) is the spatial coordinates of geographic unit i. The coefficients have the same meaning as in the SLM and SEM models, but can vary across geographic locations to capture the underlying local variations and processes. We use the *mgwr* library to implement the GWR model [34]. We configure the model using the golden search method to determine the optimal bandwidth by minimizing the Akaike Information Criterion (AIC) value.

We use two metrics to assess the goodness of fit of the statistical models, which are: root mean square error (RMSE) and R squared (R^2). RMSE measures the average difference between the fitted dependent variable values and the observed values (Equation (5)). The lower the RMSE, the better fit a model is.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\widehat{y_i} - y_i)^2} \quad , \tag{5}$$

where N is the total number of CBGs, $\widehat{y_i}$ is the fitted 311 request per property of the ith CBG, and y_i is the observed 311 request per property. The second metric R^2 is calculated using Equation (6):

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \widehat{y_i})^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2} \quad , \qquad (6)$$

which measures the overall consistency between the fitted 311 request values and the observed values. The higher the R^2 value, the better fit a model is. We use both metrics to assess the three statistical models, and use the model with the highest goodness of fit to identify factors associated with the 311 requests per property during the blizzard.

3.4. Machine learning

While statistical models are used to identify factors associated with observed 311 requests, we also use machine learning models to identify the factors important for predicting potential future 311 requests. The purpose of machine learning here is prediction, i.e., to predict potential future 311 requests based on the independent variables related to vulnerability and human behaviors. Different from statistical models that are typically fitted on the entire dataset, machine learning models are often trained and tested on separated training and test datasets to simulate the scenario of prediction [35]. Specifically, we use ten-fold cross-validation in this study in which the entire dataset is evenly divided into ten folds; each time, nine folds of data are used to train a machine learning model and one fold is reserved for prediction. This process is repeated ten times to ensure that each fold of data has been used for prediction, and the prediction results from the ten times are aggregated for evaluation. We use three machine learning models which are: support vector machine, random forest, and geographical random forests. These models have been used in previous studies on 311 data [27,28,36]. We have also considered other more complex models such as deep neural networks (DNNs); however, the small size of the data (i.e., 286 CBGs) is unlikely to train such complex models effectively [37,38]. Thus, we focus on these three simpler machine learning models. In the following, we briefly describe these models.

Support Vector Machine (SVM): SVM is a supervised machine learning model used for both classification and regression tasks. In this study, we use SVM for regression. This model aims to predict continuous variables by identifying an optimal hyperplane that minimizes errors within a defined margin of tolerance. It allows customization of kernel type, error tolerance, and margin. We implement SVM using the *scikit-learn* Python library, and perform grid search to identify the best values for the kernel type, error tolerance *C*, and margin *epsilon*. The search spaces for the kernel type, error tolerance *C*, and margin *epsilon* are set to: {*rbf, poly*}, {0.01, 0.1, 1, 10, 100, 1000}, and {0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1}.

Random Forest (RF): RF is an ensemble machine learning method that combines multiple decision trees to improve prediction accuracy. It can be used for both classification and regression tasks. In this study, we use RF for the regression task. We use Python and the *scikit-learn* library to implement the RF model, and also perform grid search to identify the best values for three hyperparameters: n_{tree}, which is the number of decision trees in the RF model; m_{try}, which determines the maximum number of features tried at each splitting node of a tree; max_{depth}, which

determines the maximum depth of the trees constructed during the training process. The search spaces for n_{tree}, m_{try}, and max_{depth} are set to: [10, 300] with an interval of 10, $\{S, \sqrt{S}, log_2 S, S/2, S/3\}$ where S is the number of input features, and [5, 20] with an interval of 5.

Geographical Random Forests (GRF): GRF extends the RF model by training multiple local RF models across different spatial locations [28,29]. GRF fits a local model for each geographic unit (i.e., CBG in this study) using only nearby observations, and further combines such a local model with a global RF model for prediction. This approach allows the GRF model to capture both the local variations and the global trend of the data. Compared with the RF model, GRF needs to determine the values of two additional hyperparameters, which are: *bandwidth*, which is the number of nearest neighbors used to train local models, and *local weight*, which determines the relative weights between the local and the global models. We use Python and the *PyGRF* library (Sun et al., 2024) to implement the GRF model. For the *bandwidth* and *local weight* hyperparameters, we determine their values via incremental spatial autocorrelation provided by the *PyGRF* library. The values of three remaining hyperparameters n_{tree}, m_{try}, and max_{depth} are determined using the grid search method, and their search spaces are set as the same as used for the RF model.

We assess the performance of the three machine learning models using the same RMSE and R squared metrics. However, both RMSE and R squared are calculated differently from the statistical models, i.e., the two metrics are calculated based on the ten-fold cross-validation process in which a model is trained on nine folds of data and makes predictions on the reserved one fold. The predictions from each of the ten folds are aggregated to calculate RMSE and R squared. The machine learning model with the best performance is then used for identifying the factors among the 22 independent variables that are important for predicting future 311 requests.

An explainable AI framework, SHAP [30], is used to interpret the result of the best machine learning model. SHAP is developed based on cooperative game theory by considering input features as players in a game and estimating the importance of each input feature based on its average marginal contributions across all possible feature coalitions. The output of the SHAP framework is SHAP values for each feature and each feature value in a data record. In this study, while the RF and GRF models could also output feature importance values, those importance values do not indicate whether an input feature has a positive or negative influence on the target variable value. In comparison, SHAP provides information about both the magnitude of feature importance and also the direction of influences. We use the *SHAP* Python package to implement the SHAP framework.

4. Results

4.1. Overview of the 311 help requests and their temporal distribution patterns

We first provide an overview of the number of residential 311 requests over the study time period. Figure 3(a) shows the total number of requests on each day over the three stages of the blizzard. The 311 request data includes the open and close times of the requests. In this study, we aggregate the requests to each day based on their open time to show a daily pattern of the reported issues, and the data could be aggregated to a finer temporal granularity (e.g., hourly) as well when needed. Here, we are looking at the total number of requests here, rather than requests per property, in order to understand the overall 311 request pattern of the entire city. In the preparation stage between Dec. 19 and Dec. 22 (background colored in yellow), the number of 311 requests gradually decreased, probably due to the blizzard preparation activities (e.g., running to grocery stores and gas stations to stock up food and gas) that diverted the attention of residents away from less urgent issues that one could make 311 calls under a normal situation. In the blizzard stage between Dec. 23 and Dec. 25 (background colored in red), the number of 311 requests first decreased but then increased, likely because residents were initially hoping to ride through the storm, but more issues started to emerge as the blizzard progressed which forced some residents to call 311. In the recovery stage between Dec. 26 and Jan. 15 (background colored in green), the 311 requests initially increased sharply and hit the highest number on Dec. 27 and then gradually decreased after Dec. 28. The high numbers of 311 requests between Dec. 26 and 30 were likely due to the recovery activities of residents immediately after the storm subsided, which revealed various issues that needed help. The number of 311 requests eventually returned to the value range before the blizzard at the end of the study time period. Note that there is also a weekly pattern in 311 requests with the number of requests rising on weekdays and falling on weekends, as the 311 service is not open normally during weekends (the service stayed open during the blizzard weekend on Dec. 24 and 25).

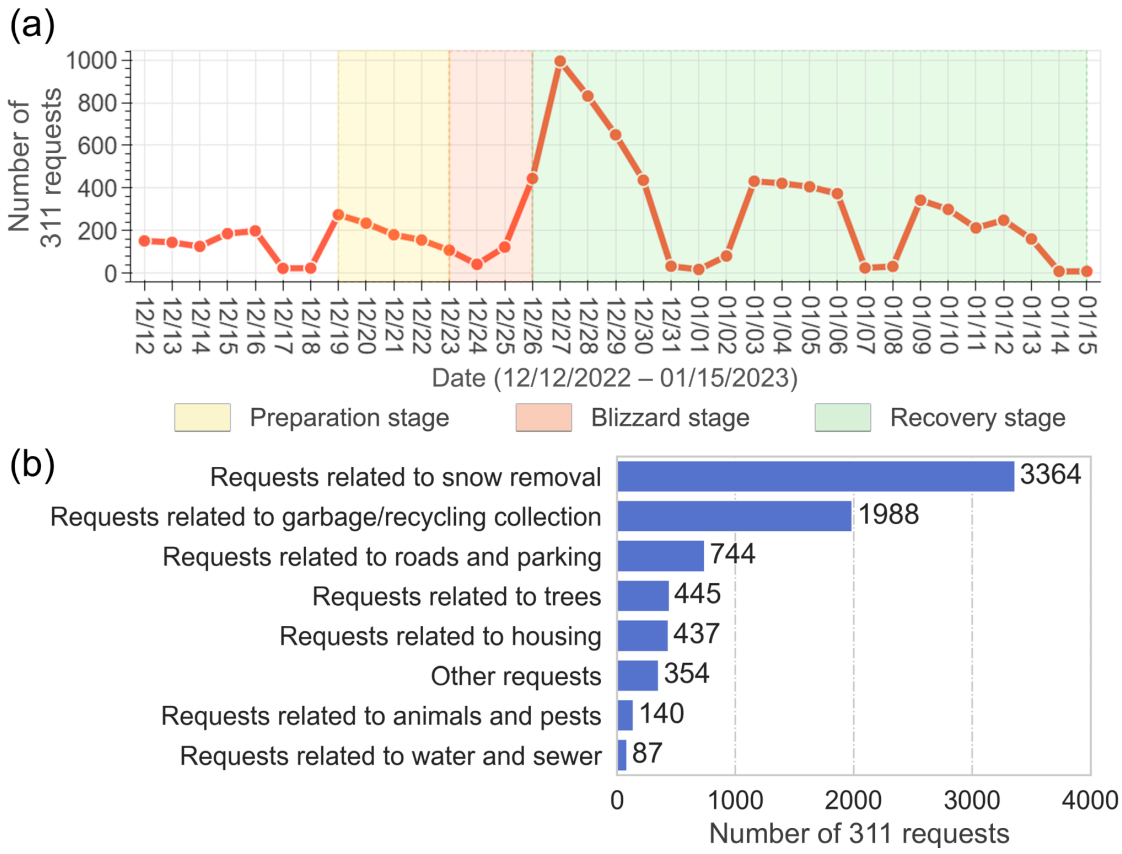

Figure 3. An overview of 311 requests in the study time period: (a) temporal distribution of total requests over the three stages of the blizzard; (b) the numbers of requests in eight categories.

We also examine the numbers of 311 requests in different categories based on their reported issues. The result is shown in Figure 3(b). We identify eight main categories from the 311 requests, which are: (1) *Requests related to snow removal* (e.g., requests for snow plowing and street salting, and issues related to vehicles blocking snow plows); (2) *Requests related to garbage/recycling collection* (e.g., delayed garbage/recycling pick up and replacement for damaged garbage totes); (3) *Requests related to roads and parking* (e.g., damaged streetlights, parking violations and abandoned vehicles, and potholes); (4) *Requests related to trees* (e.g., removing fallen trees and inspecting damaged trees); (5) *Requests related to housing* (e.g., basement flooding and electricity issues); (6) *Requests related to water and sewer* (e.g., sewer issues and water pipe issues); (7) *Requests related to animals and pests* (e.g., removing dead animals and seeking shelter for lost animals); and (8) *Other requests* (e.g., citizen service and information requests). As shown in Figure 3(b), requests related to snow removal have the highest number of 311 requests, followed by requests on garbage/recycling and then requests on parking and roads. These top three categories are likely all linked to the snow plowing activities: heavy snow accumulation required street snow plowing and led to delayed trash pick up; meanwhile, parking violations, stranded vehicles, and abandoned vehicles might have blocked some snow plowing activities. While the other categories of 311 requests have comparatively smaller

numbers, they reveal various other issues caused by the blizzard, such as damaged trees, basement flooding, dead animals, and sewer problems. To provide additional context, Supplementary Figure S1 shows the number of historical 311 requests across these eight categories during the previous snow seasons in 2021 and 2022. During relatively normal snow seasons without a severe blizzard, requests related to garbage/recycling collection are the most common type of requests, and there are much fewer tree-related requests than those related to housing issues.

We further investigate the temporal distributions of 311 requests in each of the eight categories. The result is shown in Figure 4. As can be seen, 311 requests show different temporal patterns across different categories. For requests related to snow removal (sub figure (a)), it shows a pattern largely similar to the overall pattern of all requests as shown in Figure 3(a), likely due to the large number of snow removal requests in all 311 calls. For requests related to trees (sub figure (d)), it shows a different temporal pattern in which the number of requests largely increased at the onset of the blizzard, likely due to the hurricane-force wind that blew down trees. There were also many tree-related requests in the recovery stage, and many of them were about inspecting or removing damaged trees. For requests related to garbage/recycling collection (sub figure (b)), roads and parking (sub figure (c)), housing (sub figure (e)), water and sewer (sub figure (f)), and animals and pests (sub figure (g)), they share a similar temporal pattern in that the peak numbers of these requests showed up in the recovery stage starting roughly from Jan. 2 (about one week after the blizzard subsided). This result partially reflects the situation in which residents were initially focusing on snow removal immediately after the blizzard; as streets were being cleared up and snow melted, other issues, such as delayed garbage collection, basement flooding, and displaced animals, started to show up.

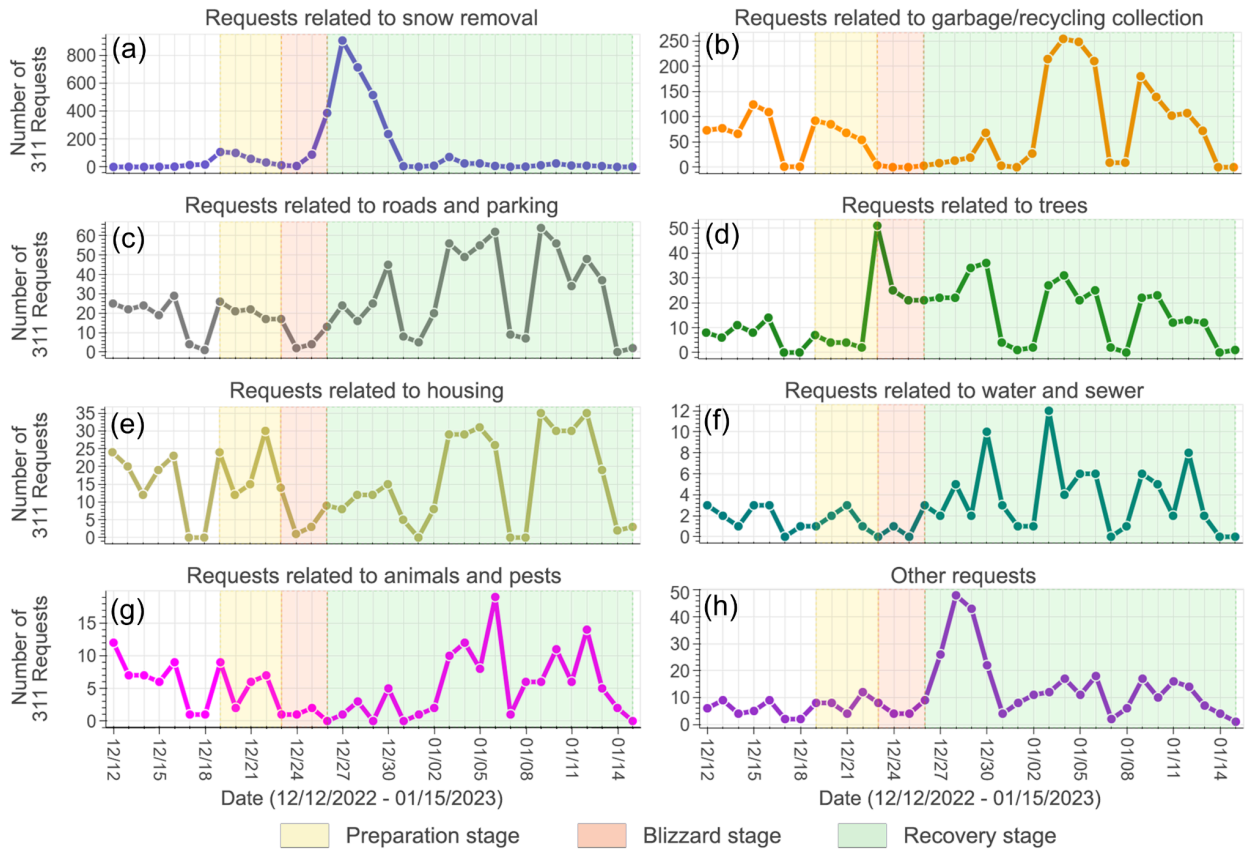

Figure 4. The temporal distributions of the number of 311 requests in different categories during the blizzard. Note that the scales of y-axis in different sub figures are in different value ranges.

4.2. Spatial distribution of the 311 requests per property

We further examine the spatial distribution of the 311 requests. Here, we look into 311 requests per property, since different CBGs have different numbers of residential properties. The result is visualized in Figure 5. We use Quantile Breaks for map visualization, and each color represents 20% of the CBGs. A visual examination of the figure suggests that CBGs with high 311 requests per property are clustered on the east side of the city, which has a high percentage of minority and low-income population. To further quantify the spatial pattern, we use the global Moran's I index to assess the spatial autocorrelation of the data. The value of Moran's I ranges from -1 to 1, where positive values indicate positive spatial autocorrelations (i.e., similar values are clustered), and negative values indicate negative spatial autocorrelations (i.e., dissimilar values are clustered). Here, the global Moran's I index is 0.2107 ($p < 0.001$), suggesting a significant and positive spatial autocorrelation, i.e., CBGs with similar 311 requests per property are spatially clustered. This result confirms our observed clustered pattern.

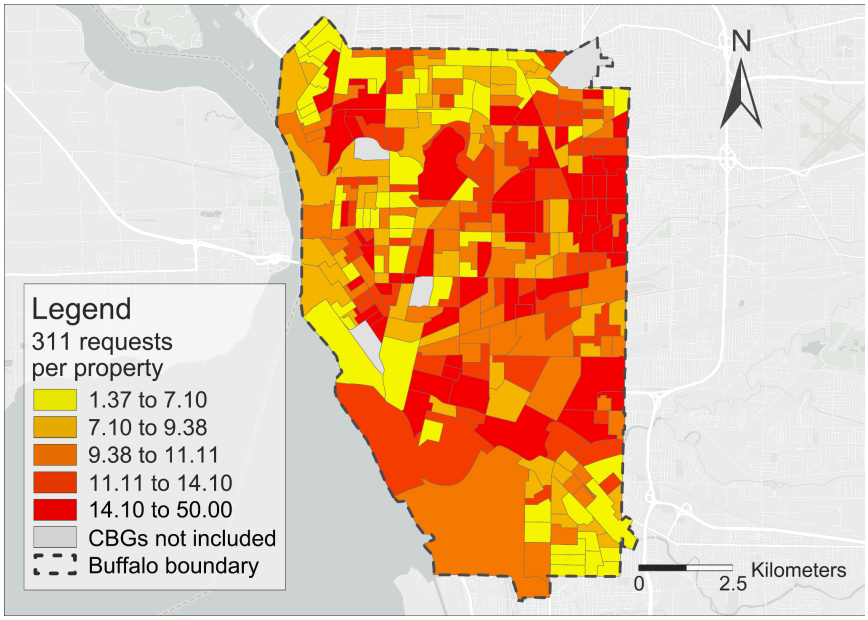

Figure 5. The spatial distribution of 311 requests per property during the study time period.

We also investigate the spatial distributions of the 311 requests per property across the eight different categories. The result is visualized in Figure 6. We also calculate the global Moran's I index for each category to assess the extent of spatial autocorrelation, and the results are included in the subfigures. For requests related to snow removal (subfigure (a)), they show a similar spatial pattern as the total 311 requests per property, in which CBGs with high numbers of requests are clustered on the east side of the city. The Moran's I of snow removal related requests suggests an even stronger positive spatial autocorrelation, with an index value of 0.3538 ($p < 0.001$). We further compare the spatial distribution of snow removal related requests with that of accumulated snow depth during the study time period in Supplementary Figure S2. The comparison shows that while high numbers of snow removal related requests are clustered on the east side of the city, the accumulated snow depth shows a general south-to-north snow pattern typically observed in this region (the southern areas are closer to Lake Erie and generally receive more lake-effect snow). The east side of the city was also severely affected, receiving an average of about 2.5 meters of accumulated snow depth in the study time period. This comparison suggests that snow removal related 311 requests are not solely determined by snow depth but are likely associated with additional socioeconomic factors. We also note that the snow depth data from SNODAS has a spatial resolution of 1 km and may not fully capture the spatial variation of snow depth at the finer CBG level. For requests related to garbage collection (subfigure (b)) and requests related to trees (subfigure (d)), they show weaker but still significant spatial autocorrelations, with index values of 0.1724 ($p < 0.001$) and 0.1022 ($0.001 < p < 0.05$) respectively. Garbage collection related requests also seem to cluster on the east side of the city. This result is understandable: if streets remain unplowed, then the collection of garbage will likely be delayed. Tree-related requests do not show a clear spatial pattern. The Moran's I indexes of the other categories of 311 requests are not significant, suggesting more or less random spatial patterns.

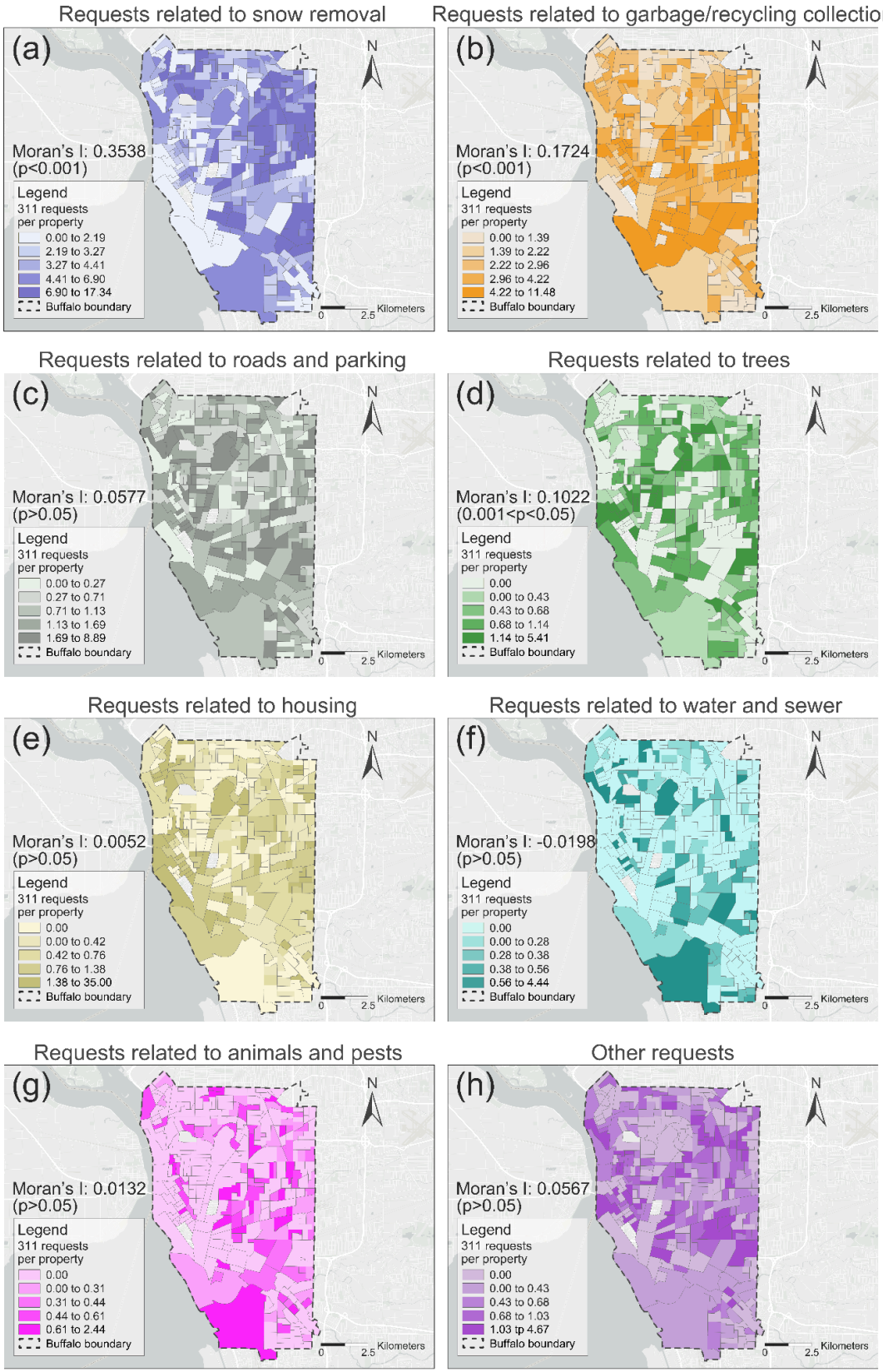

Figure 6. The spatial distributions of the 311 requests per property in eight categories.

4.3. Statistical analysis results

We fit three statistical models, i.e., SLM, SEM, and GWR, based on the prepared independent variables and the dependent variable (see Table 1). The goodness of fit of these three statistical models is shown in Table 2. The three statistical models have fairly close R^2 and RMSE values. We choose the GWR model to further identify the factors that are associated with existing 311 requests per property, since GWR achieves the highest R^2 of 0.5752 and the lowest RMSE of 3.4679. In addition, GWR provides local coefficient values at different locations, and allows us to further examine how the effects of the factors vary across different neighborhoods.

Table 2. Goodness of fit of the three statistical models.

Model	**R^2**	**RMSE**
SLM	0.5504	3.5681
SEM	0.5275	3.6580
GWR	**0.5752**	**3.4679**

Because GWR fits many local models rather than a single global model, each regression coefficient has a set of coefficient values and a set of t-values indicating the significance of coefficients in these local models [34,39]. We use a box plot to visualize the values of each coefficient, and use the percentage of significant t-values to understand its significance (insignificant t-values are filtered out based on a significance level of 0.05). Figure 7 shows the box plots of the regression coefficients from GWR. The top four variables have 100% significant t-values, and two variables, *% owner-occupied housing units* and *snow depth*, have 61.9% and 56.3% significant t-values respectively. The other variables have low or zero percent significant t-values.

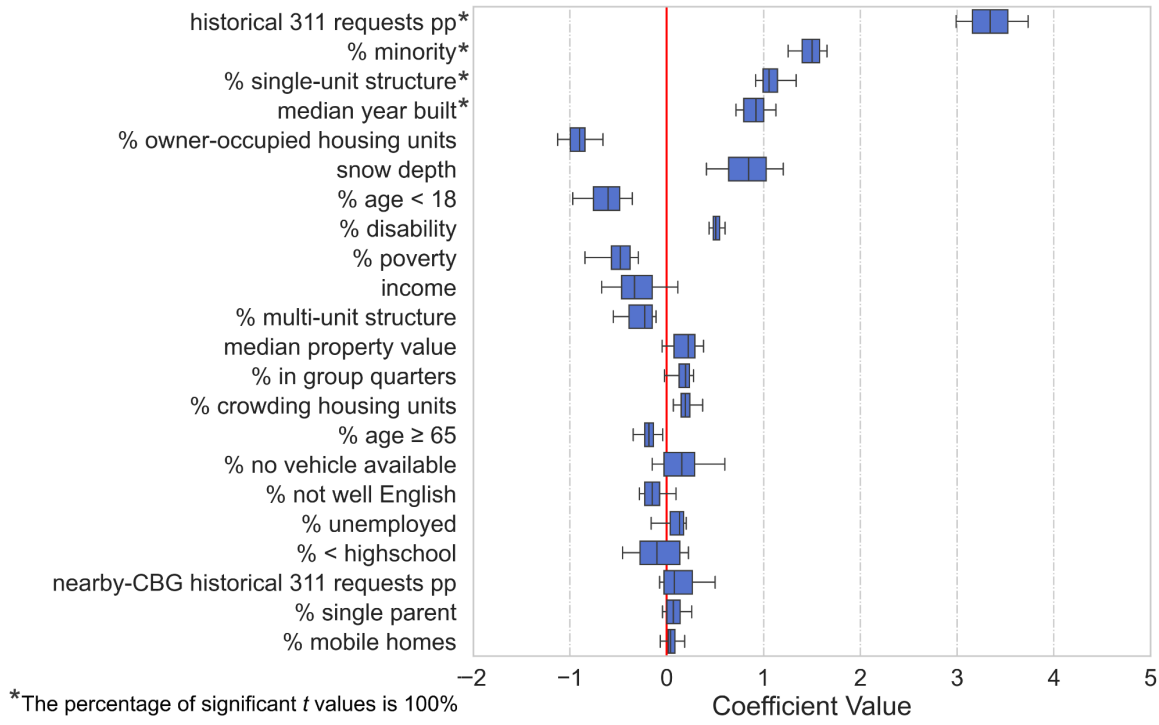

Figure 7. Regression coefficients of the independent variables output by the GWR model.

As can be seen in Figure 7, *historical 311 requests pp* shows the highest positive association with 311 requests per property during the blizzard, and this association is also significant across all CBGs. Note that the independent variables have been standardized before the regression analysis; therefore, the coefficients can indicate the relative importance of the independent variables. The high coefficient values of *historical 311 requests pp* suggest that CBGs with high historical requests are likely to have high 311 requests during the 2022 blizzard. *% minority* has the second highest and positive regression coefficients, suggesting that CBGs with high percentages of minority population are associated with high 311 requests per property during the blizzard. *% single-unit structure* and *median year built* both have positive and significant associations with 311 requests. This result suggests that CBGs with higher percentages of single-unit structures (e.g., single-family houses) and newer houses are, in fact, associated with more 311 requests per property during the blizzard. This result is surprising since neighborhoods with higher percentages of single-family houses and newer properties are often linked to lower vulnerability [24]. We will further investigate this result in our following analysis. The variable *% owner-occupied housing units* has a negative regression coefficient (significant in 61.9% CBGs), which suggests that a lower percentage of owner-occupied houses (or a higher percentage of renter-occupied houses) is linked to more 311 requests per property. The variable *snow depth* has a positive coefficient (significant in 56.3% CBGs), suggesting deeper snow is generally associated with more 311 requests. For the remaining variables, their regression coefficients are mostly insignificant, and their coefficient values are also close to 0.

One advantage of GWR is that it allows the regression coefficients to be visualized spatially to understand their spatial variations. Figure 8 provides such spatial visualizations of the top four most significant variables. For each row in Figure 8, the first map shows the data values of the variable, and the second map shows the regression coefficients. For row (a), the first map shows that the *historical 311 requests pp* has a more or less random pattern which is different from the more clustered pattern of the 311 requests during the blizzard; the regression coefficient values in the second map shows that CBGs on the east side of the city have higher coefficient values than those on the west side. This result suggests that CBGs on the east side tend to have more 311 requests per property during the blizzard if they had similar *historical 311 requests pp*. For row (b), the first map shows that CBGs with high percentage of minority population are clustered on the east side of the city; the second map shows that the regression coefficients of *% minority* have almost a reverse pattern in that lower coefficient values are observed in roughly the east side and higher coefficient values are observed in the other regions. This result suggests that in CBGs where the percentage of minority population is already high, a further increase in *% minority* does not lead to much increase in 311 request per property during the blizzard; however, for CBGs where the percentage of minority population is low, an increase in *% minority* can lead to a larger increase in 311 requests. For row (c), the first map shows that CBGs with high *% single-unit structure* are in the peripheral area of the city; the second map shows another reverse pattern in the coefficients suggesting that CBGs with low *% single-unit structure* (e.g., CBGs in the central area of the city) are linked to higher increases in 311 requests per property during the blizzard when their *% single-unit structure* increases. For row (d), the first map shows that most newer properties (which are properties built between 1900 and 1920) are in the peripheral regions of the city; the second map of coefficients shows that newer properties on the east side are associated with higher increases of 311 requests per property during the blizzard. While the positive coefficients of *% single-unit structure* and *median year built* are initially surprising, we think that these results are likely due to the special housing situation in the study area, i.e., these "newer" properties are in fact old single-family houses built around 100 years ago. Without sufficient financial capabilities to maintain these old houses, families are likely to encounter various issues during a disaster, as also documented in a recent study [40].

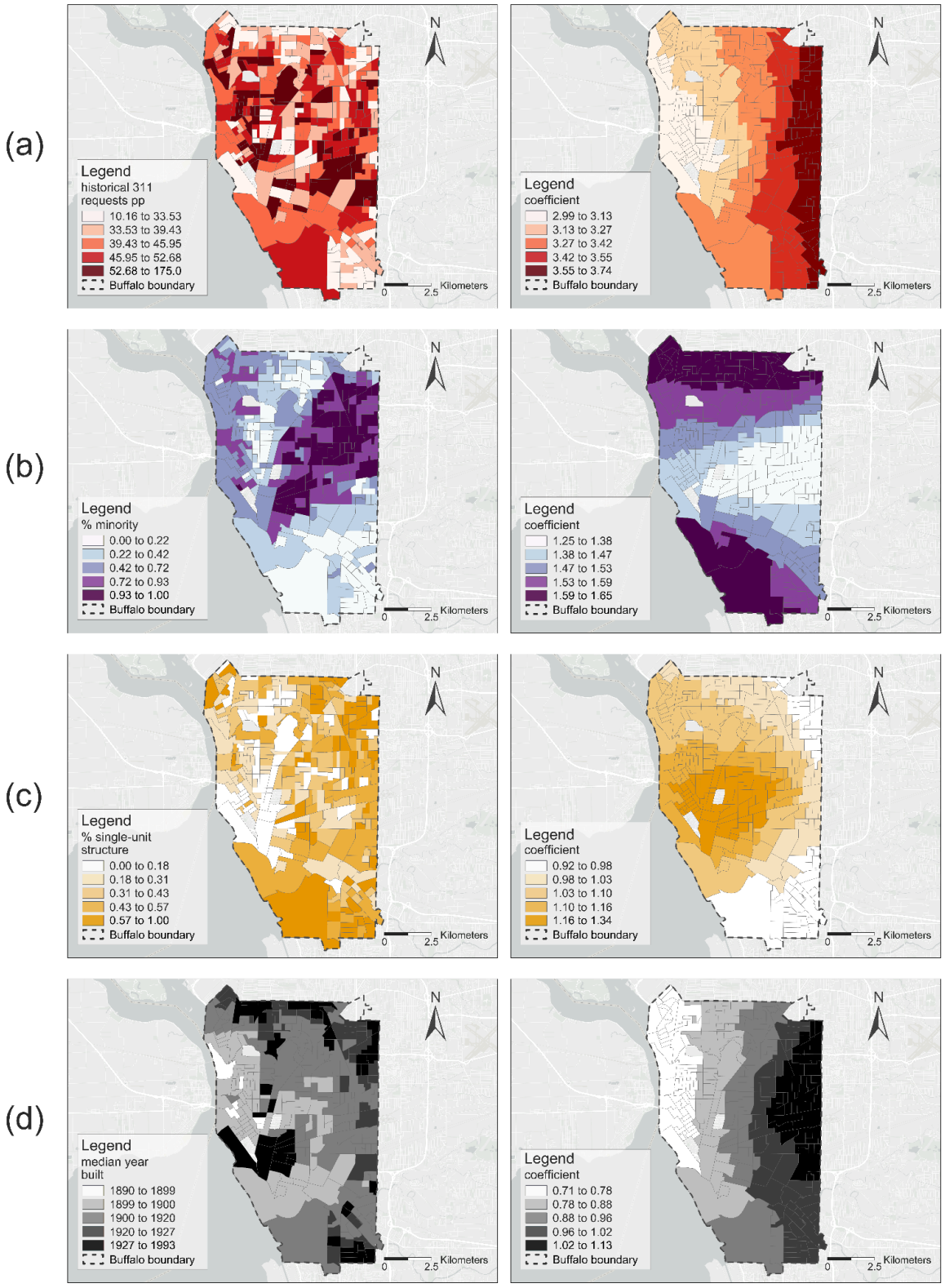

Figure 8. Regression coefficients of the top four most significant variables output by the GWR model: (a) *historical 311 requests pp*; (b) *% minority*; (c) *% single-unit structure*; (d) *median year built*

4.4. Machine learning results

Table 3 summarizes the R^2 and RMSE values of the three machine learning models. These R^2 and RMSE values are not directly comparable with those of the statistical models reported earlier. The evaluation processes of these two types of models are different: statistical models are evaluated based on a single fitting of the entire dataset, while machine learning models are evaluated via ten-fold cross-validation. From a temporal perspective, statistical models are retrospective in that the models are fitted on existing observations, while machine learning models are prospective in that the models are trained to predict new observations [41]. Machine learning models in this study therefore can provide additional predictive insights by helping identify the factors that are important for predicting potential future 311 requests. Among the three machine learning models, GRF achieves the highest R^2 of 0.4310 and the lowest RMSE of 4.0136. We note that statistical models can also be used for making predictions, although machine learning models typically have higher prediction accuracy given their ability to capture complex, non-linear relationships often existing in data. With curiosity, we further test the performance of the GWR model for making predictions in the same setting as the machine learning models (i.e., ten-fold cross-validation). The GWR model achieves an R^2 of 0.3661 and an RMSE of 4.2366. The Python packages of SLM and SEM models do not provide a prediction function, probably because they are mostly used for statistical inference. Given that the GRF model achieves the highest performance, we use it along with SHAP to understand the factors that are important for predicting potential future 311 requests.

Table 3. Prediction performance of the three machine learning models.

Model	**R^2**	**RMSE**
SVM	0.2474	4.6162
RF	0.4230	4.0419
GRF	**0.4310**	**4.0136**

Figure 9 shows the SHAP values of the 22 independent variables visualized in what is called a "beeswarm plot" in the SHAP framework. Note that the latitudes and longitudes of CBGs are also included in the plot as two additional variables, because the GRF model needs the latitude and longitude information to choose local models for making predictions. In the beeswarm plot, each row is an input feature (or an independent variable), and each dot represents a SHAP value for the corresponding input feature at a data record (i.e., a specific CBG in this study). The color of the dot indicates whether the corresponding input feature value is high (in red color) or low (in blue color). Dots on the right side of the zero vertical line represent positive SHAP values, and dots on the left side represent negative SHAP values. The SHAP framework provides explanations for a model by starting the model prediction from the average predicted value of the target variable and

then gradually adjusting the prediction based on the input feature values. A positive SHAP value means the input feature value will increase the predicted target variable value from the mean, and a negative SHAP value means the input feature value will decrease the predicted target variable value from the mean. In the beeswarm plot, an input feature is considered having positive influence on the target variable, if red dots are on the right side of the zero line and blue dots on the left side (i.e., higher feature values make the values of the target variable higher). By contrast, an input feature is considered having negative influence, if red dots are on the left side and blue dots are on the right side (i.e., higher feature values make the values of the target variable lower). The absolute SHAP values are aggregated to assess feature importance, and input features are ordered from top to bottom in the beeswarm plot based on their importance.

Figure 9. The beeswarm plot of the SHAP values based on the GRF model. The rankings are based on mean absolute SHAP values.

As can be seen in Figure 9, *historical 311 requests pp*, *% minority,* and *median year built* are three most important variables for the GRF model to predict 311 requests per property during the blizzard. The mean absolute SHAP values of these three variables are 1.258, 0.857, and 0.210,

respectively, which are also shown in the mean absolute SHAP plot in Supplementary Figure S3. The three variables also have positive influences on the target variable, as they have most red dots on the right side and most blue dots on the left side. The importance of these three variables identified by GRF is largely similar to the result of GWR. Meanwhile, the variable *% single-unit structure*, which has high coefficient values in GWR, is considered much less important by the GRF model. The GRF model considers the variable *median property value* as the fourth most important variable for predicting 311 requests, and the mean absolute SHAP value of this variable is 0.208 (also shown in Supplementary Figure S3). *Median property value* has a negative influence on the target variable (as the blue dots are on the right side), which suggests that higher property values are linked to lower 311 requests per property. Two other variables, *snow depth* and *% owner-occupied housing units,* are also considered as fairly important input features, with positive and negative influences respectively on the target variable. These two variables are also identified as fairly important by the GWR model. The remaining variables are considered less important for predicting 311 requests by the GRF model.

Since a SHAP value is computed for each feature and in each data record, we can spatially plot out SHAP values across CBGs in the study area. Figure 10 shows the SHAP values of the top four most important variables. For each row, the first map shows the data values of the variable, and the second map shows the SHAP values of this variable across CBGs. It is worth noting that SHAP values are different from the coefficient values that we have seen previously from GWR: a coefficient value suggests how much the dependent variable will change given a unit change in an independent variable, while a SHAP value suggests how much the dependent variable will move away from its mean value given a specific independent variable value. As can be seen in rows (a), (b), and (c) in Figure 10, the SHAP values of *historical 311 requests pp*, *% minority,* and *median year built* show largely similar spatial patterns as those of the data values of these three variables. This is because these three variables all have positive influences on the target variable; thus, higher data values of these three variables also mean higher SHAP values. In particular, the SHAP values of *historical 311 requests pp* and *% minority* show almost identical spatial patterns as their data values, likely due to their strong and positive influences on the target variable, i.e., higher *historical 311 requests pp* and higher *% minority* will most likely lead to *311 requests pp* values that are higher from the mean. The SHAP values of *median property value* (in row (d)) show a reverse pattern compared with the spatial pattern of its data values, due to the negative influence of this variable, i.e., higher *median property value* will move the target variable *311 requests pp* toward the lower end away from the mean value.

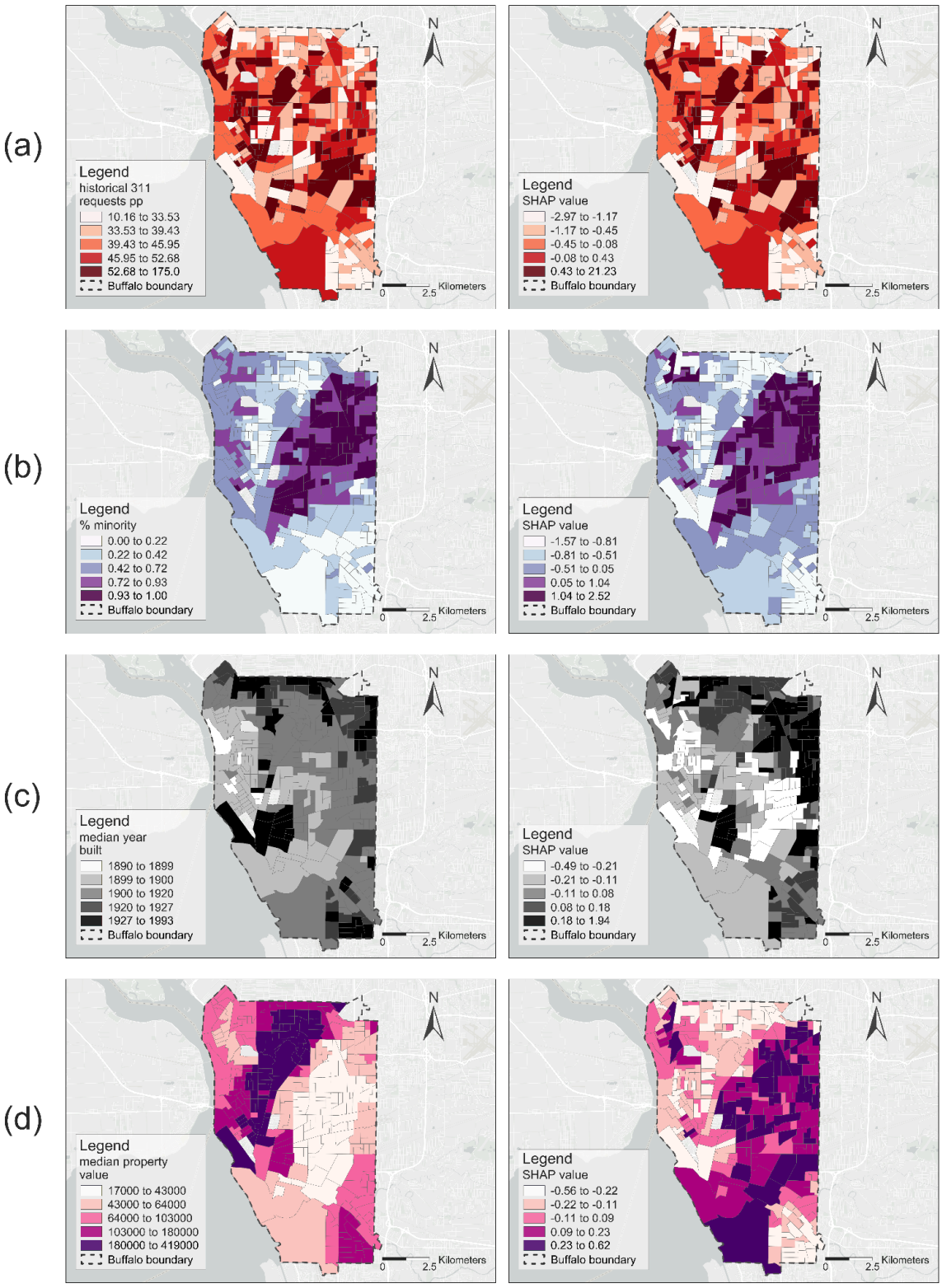

Figure 10. The SHAP values of the top four most important independent variables output by the GRF model: (a) *historical 311 requests pp*; (b) *% minority*; (c) *median year built*; (d) *median property value*.

5. Discussion

5.1. Answering the research questions

This study sets out to answer two research questions focusing on 311 help requests to understand the impacts of the 2022 Buffalo blizzard. For RQ1: *what are the spatial and temporal distributions of the 311 help requests for different issues due to the blizzard,* our study reveals a statistically significant spatial clustering pattern for 311 requests per property during the studied blizzard time period (Moran's I = 0.2107 and $p < 0.001$), and high values are observed on the east side of the city. From a temporal perspective, a large number of 311 requests started to show up when the storm subsided after Dec. 25; meanwhile, requests for different issues showed different temporal patterns. For example, requests related to trees had a sharp increase at the onset of the blizzard, likely due to the hurricane-force wind blowing down trees; requests related to snow removal peaked immediately after the blizzard and gradually decreased in the following days; and requests for other issues, such as delayed garbage collection, basement flooding, and displaced animals, showed up roughly one week after the blizzard. For RQ2: *what are the factors, such as physical and social vulnerability factors and human behavioral factors, that are associated with the 311 help requests,* we use both statistical analysis and machine learning to identify these factors in a complementary manner. The best statistical model, GWR, suggests that *historical 311 requests pp, % minority*, *% single-unit structure* and *median year built* are the top four variables significantly associated with existing 311 requests. The best machine learning model GRF is a GeoAI model, which suggests that *historical 311 requests pp, % minority*, *median year built*, and *median property value* are the top four variables important for predicting potential future 311 requests. Two other variables, *snow depth* and *% owner-occupied housing units,* also show moderate importance in both statistical analysis and machine learning.

5.2. Implications for disaster management

This study contributes to the disaster research literature by increasing our understanding of community impacts of winter storm disasters. Compared with other natural disasters such as hurricanes, earthquakes, and wildfires, winter storms and blizzards have been studied less frequently. Meanwhile, existing research suggests that climate change will likely increase weather variability and lead to more extreme weather events, including more extreme winter storms [42,43]. The severe impacts of more recent winter storms, such as Winter Storm Uri in 2021 and Winter Storm Elliott in 2022, have already motivated new winter storm research [8,44–47]. Our study, therefore, adds to this increasing size of winter storm literature and could help increase community resilience to similar future disasters.

From a disaster management perspective, this study reveals disparities in 311 help requests per property in the studied blizzard time period. The disparities are shown both geographically and socially. Geographically, CBGs with higher 311 requests per property are mostly clustered on the east side of the city. Socially, CBGs with higher percentages of minority population are linked to higher 311 help requests per property during the blizzard, and this significant association is

suggested by both statistical analysis and machine learning which utilize largely different models. While one could argue that an alternative explanation is that minority population groups may be more likely to make 311 calls or make more repeating calls for the same issues, our modeling approaches have explicitly included historical 311 requests as an independent variable and our preprocessing steps have removed duplicative calls. Despite these modeling and preprocessing steps, the variable *% minority* is still identified as the second most important variable associated with 311 requests per property during the blizzard. While previous research has shown that disadvantaged population groups are often disproportionately affected by disasters [25,48,45,49], this study suggests that *% minority* is the most important social vulnerability factor in this blizzard compared with other factors included in the models, such as *% poverty* and *% < highschool*. We note that this study is conducted at the CBG level and we do not know the percentage of calls actually made by minority individuals, since the 311 data do not contain demographic information. Nevertheless, this result likely suggests that minority populations in this region face various challenges during winter storms and possibly other disasters.

The results of this study could inform future winter storm management in both the Buffalo region and other regions in the country. For the Buffalo region, emergency management services could consider resilience strategies to better prepare the east side of the city when a severe winter storm is anticipated. These strategies could include pre-positioning snow plowing equipment, communicating with local residents to remove parked vehicles from the streets, setting up temporary service centers, and equipping emergency repair teams with snow-agile vehicles to access this area under heavy-snow conditions. Additional efforts could also be made to reach out to minority populations in the city beyond the east side to understand the challenges and needs faced by minority populations during winter storms. Such understanding could inform the allocation of winter storm preparation resources and help increase community resilience in future winter storms. The results of this study could also inform winter storm management in other geographic regions. For example, the temporal patterns of 311 calls suggest a possible sequence of issues that may be caused by a winter storm, especially a blizzard: from early on, trees may be blown down by strong winds; the heavy snow can create low-visibility conditions and block roads for emergency vehicles; once snow stops, snow removal is an immediate need of most communities; as streets are being cleared up and snow melts, other issues begin to surface, such as delayed trash collection, basement flooding, sewer issues, and displaced animals. While this sequence of issues may not be new to residents highly experienced with snow, recent winter storms have increasingly affected regions where snow is less frequent, such as regions in the American South [12,50]. Thus, these possible issues and their sequence could be communicated to emergency responders and residents in these regions when a winter storm is forecasted.

5.3. Methodological implications

Methodologically, this study explores the use of statistical analysis and machine learning in a complementary manner. While statistical models have been widely used in the literature, there has been an increasing interest in using machine learning and AI methods in disaster research

[16,51,52]. These two types of models have some methodological and conceptual overlap. Statistical models are frequently used to understand variable associations and draw inferences, but can also be used to make predictions in a machine learning fashion. While machine learning models often focus on predictions, feature importance and explainable AI frameworks also provide some interpretability that could be compared to the results of statistical models to some extent. In this study, we use three spatial statistical models to explore the associations between a variety of factors and the 311 help requests (by fitting models on the entire dataset). We also use three machine learning models, including a GeoAI model, to predict potential future 311 requests and examine feature importance (by training and testing models via ten-fold cross-validation). We then choose the best models from these two approaches and compare their results. These two approaches provide complementary insights on the disparities of 311 help requests. Two variables, *historical 311 requests pp* and *% minority*, are identified as the most important factors by both approaches, and the variable *median year built* is also identified as among the top most important factors by both approaches. Meanwhile, the results of these two approaches also have their differences: the statistical approach based on the GWR model considers the variable *% single-unit structure* as an important factor associated with 311 help requests, while the machine learning approach based on the GRF model considers the variable *median property value* as an important factor for predicting 311 requests. Two other variables, *snow depth* and *% owner-occupied housing units,* are also considered as fairly important by both approaches, although their statistical significance or feature importance is comparatively lower. With results from these two approaches, disaster managers could choose to further investigate the most critical variables that are shown to be overlapping in both approaches (e.g., *historical 311 requests pp* and *% minority*), or could select all important variables suggested by both approaches for further examination.

This study also explores the use of the SHAP framework to explain results from AI models. While machine learning and AI models have shown good performance in prediction tasks, their lack of explanation is a major drawback that limits their use in potential applications [53]. The SHAP framework is a general explainable AI framework that can be used to interpret input feature importance of any AI models [30,54]. In this study, the best machine learning model is the GRF model, and SHAP is used in combination with the GRF model to identify the important factors for predicting 311 requests. It is worth noting that the GRF model itself also has the ability to output feature importance, which is inherited from the RF model. We plot out the feature importance of the GRF model in Supplementary Figure S4, and the relative feature importance closely aligns with the importance output by the SHAP framework. The default feature importance of GRF, however, does not show whether an input feature positively or negatively influences the target variable to be predicted; such directional information is provided by SHAP. In addition, SHAP supports the explanation of AI models that do not provide feature importance themselves. In our study, had the best machine learning model been SVM, we would not have been able to interpret feature importance without SHAP. Overall, this study shows that SHAP is a promising framework that can increase our ability to interpret the results of AI models in disaster research.

5.4. Limitations

This study is not without limitations. First, as noted earlier, this study is conducted at the CBG level not the individual level, and the 311 call data do not contain demographic information of the individuals who made the phone calls. While we can connect 311 call data with neighborhood-level demographic data from the U.S. Census, individual-level demographic data about the callers could allow us to pinpoint the issues faced by specific population groups during this blizzard. Second, while this research has examined the community impacts of the 2022 blizzard from the lens of 311-reported issues, other aspects of the 311 data and other datasets could also be studied to achieve a more comprehensive understanding of this disaster. For example, we could compare the open and close time of the 311 calls to understand the response time taken to address issues in this blizzard, and whether the response time varies across different communities. We could also leverage additional datasets, such as human mobility data, to examine the disruptions of the blizzard on the life of residents which may not necessarily result in 311 calls but manifest as decreased human movements. These and other directions could be further explored in the near future.

6. Conclusions

The 2022 Buffalo blizzard was a catastrophic winter storm disaster that severely affected communities in the city of Buffalo. Given that many residents used the 311 call service to request help for issues caused by the blizzard, this study examines these 311 help requests and their potential disparities across communities. We find a spatial clustering pattern of high 311 requests per property on the east side of Buffalo, and varied temporal patterns of different categories of 311 requests. Statistical analysis and machine learning are utilized in a complementary manner to identify factors associated with 311 requests per property during the blizzard. We test three statistical models (SLM, SEM, and GWR) and three machine learning models (SVM, RF, and GRF), and choose the best models for result interpretation. An explainable AI framework SHAP is further employed to interpret the result of machine learning. The results suggest that *historical 311 requests pp* and *% minority* are two important factors associated with 311 help requests, which are identified by both statistical and machine learning approaches. While this study is not without limitations, we hope that it makes a modest contribution by adding to the winter storm research literature, supporting disaster management decisions related to winter storms, and informing future disaster research interested in using machine learning and explainable AI methods.

Abbreviations

GeoAI: Geospatial Artificial Intelligence
CBG: Census Block Group
CSV: Comma-Separated Value

SNODAS: Snow Data Assimilation System
ACS: American Community Survey
SVI: Social Vulnerability Index
SLM: Spatial Lag Model
SEM: Spatial Error Model
GWR: Geographically Weighted Regression
SVM: Support Vector Machine
RF: Random Forest
GRF: Geographical Random Forests
SHAP: SHapley Additive exPlanations
VIF: Variance Inflation Factor
AIC: Akaike Information Criterion
RMSE: Root Mean Square Error
DNNs: Deep Neural Networks

Declaration of competing interest

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

Data availability

Data related to the analysis results from this study are available on GitHub: https://github.com/geoai-lab/Buffalo-Blizzard-311-Requests. The original 311 help request data are publicly available from the Open Data Portal of Buffalo: https://data.buffalony.gov/Quality-of-Life/311-Service-Requests-July-2008-May-2024-/whkc-e5vr/about_data. The property assessment roll data are also publicly available from the Open Data Portal of Buffalo: https://data.buffalony.gov/Economic-Neighborhood-Development/2023-2024-Assessment-Roll/drey-kz4e/about_data. The snow depth data used in this study are derived from the Snow Data Assimilation System (SNODAS) and can be accessed at: https://nsidc.org/data/g02158/versions/1. The socioeconomic and demographic data are obtained from the American Community Survey (ACS) of the U.S. Census Bureau and are publicly available at: https://data.census.gov/cedsci/.

Acknowledgments

This work is supported by the U.S. National Science Foundation under Grant No. BCS-2416886. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Reference

[1] S.M. Kaufman, R. Zimmerman, K. Ozbay, A. Smith, S.H. Lambson, C. Curry, E. Jeng, J. Gao, E. Kaval, Lessons Learned from the Buffalo Blizzard: Recommendations for Strengthening Preparedness and Recovery Efforts, (2023).

[2] S. Watson, Death toll in Buffalo blizzard rises to 47 people, Buffalo News (2023). https://buffalonews.com/news/local/death-toll-in-buffalo-blizzard-rises-to-47-people/article_04c578e0-9814-11ed-b391-dbf7d2370f3d.html (accessed July 24, 2023).

[3] Guidehouse, New York State Preparedness and Response Efforts: Blizzard of 2022 After-Action Review, (2023).

[4] A. Reyes, Buffalo Common Council approves emergency manager, fleet director positions as city releases blizzard report, WKBW News (2023). https://www.wkbw.com/news/local-news/blizzard-of-22/buffalo-common-council-approves-emergency-manager-fleet-director-positions-as-city-releases-blizzard-report.

[5] J. Thorsby, G.N. Stowers, K. Wolslegel, E. Tumbuan, Understanding the content and features of open data portals in American cities, Gov. Inf. Q. 34 (2017) 53–61.

[6] J.L. Rainey, K. Pandian, L. Sterns, K. Atoba, W. Mobley, W. Highfield, R. Blessing, S.D. Brody, Using 311-Call data to Measure Flood Risk and Impacts: The Case of Harris Country TX, Inst. Disaster Resilient Tex. Galvest. TX USA 22 (2021).

[7] A. Eugene, N. Alpert, W. Lieberman-Cribbin, E. Taioli, Using NYC 311 call center data to assess short-and long-term needs following Hurricane Sandy, Disaster Med. Public Health Prep. 16 (2022) 1447–1451.

[8] C.-C. Lee, M. Maron, A. Mostafavi, Community-scale big data reveals disparate impacts of the Texas winter storm of 2021 and its managed power outage, Humanit. Soc. Sci. Commun. 9 (2022) 1–12.

[9] C.W. Zobel, M. Baghersad, Y. Zhang, Calling 311: evaluating the performance of municipal services after disasters., in: Proc. 14th ISCRAM Conf., Albi, France, 2017: pp. 164–172.

[10] L. Xu, M.-P. Kwan, S. McLafferty, S. Wang, Predicting demand for 311 non-emergency municipal services: An adaptive space-time kernel approach, Appl. Geogr. 89 (2017) 133–141. https://doi.org/10.1016/j.apgeog.2017.10.012.

[11] N. Madkour, B. Tokgoz, Predicting Non-Emergency 311 Calls for an Effective Resource Allocation After a Disaster, in: Proc. 2020 IISE Annu. Conf., New Orleans, USA, 2020: pp. 1–6.

[12] M. Ferman, Winter storm could cost Texas more money than any disaster in state history, Tex. Trib. (2021). https://www.texastribune.org/2021/02/25/texas-winter-storm-cost-budget/.

[13] L. Metzger, The Texas Freeze: Timeline of events, 2022. https://environmentamerica.org/texas/center/articles/the-texas-freeze-timeline-of-events.

[14] V. Reyna-Rodriguez, J. Mendiola, P. Joens, What were the lows around Iowa as arctic cold settles over the state?, Moines Regist. (2024). https://www.desmoinesregister.com/story/weather/2024/01/11/winter-storm-iowa-blizzard-live-updates-be-prepared-parking-511-national-weather-service-forecast/72189570007/.

[15] H. Yan, R. Shackelford, 'Life-threatening' blizzard will slam Iowa as record-low temps or tornado-spawning storms could hit the South and East Coast, CNN (2024). https://www.cnn.com/2024/01/12/weather/winter-weather-alerts-blizzard-floods/index.html.

[16] M.M. Kuglitsch, I. Pelivan, S. Ceola, M. Menon, E. Xoplaki, Facilitating adoption of AI in natural disaster management through collaboration, Nat. Commun. 13 (2022) 1–3.

[17] A. Masrur, M. Yu, P. Mitra, D. Peuquet, A. Taylor, Interpretable machine learning for

analysing heterogeneous drivers of geographic events in space-time, Int. J. Geogr. Inf. Sci. 36 (2022) 692–719.
[18] R. Suwaileh, T. Elsayed, M. Imran, H. Sajjad, When a disaster happens, we are ready: Location Mention Recognition from crisis tweets, Int. J. Disaster Risk Reduct. (2022) 103107.
[19] S. Gao, Y. Hu, W. Li, Handbook of geospatial artificial intelligence, CRC Press, 2023.
[20] L. Zou, A. Mostafavi, B. Zhou, B. Lin, D. Mandal, M. Yang, J. Abedin, H. Cai, GeoAI for Disaster Response, in: Handb. Geospatial Artif. Intell., CRC Press, 2023: pp. 287–304. https://www.taylorfrancis.com/chapters/edit/10.1201/9781003308423-14/geoai-disaster-response-lei-zou-ali-mostafavi-bing-zhou-binbin-lin-debayan-mandal-mingzheng-yang-joynal-abedin-heng-cai (accessed December 19, 2024).
[21] S. Fuchs, T. Frazier, L. Siebeneck, Physical vulnerability, Vulnerability Resil. Nat. Hazards (2018) 32–52.
[22] Z. Wang, N.S.N. Lam, N. Obradovich, X. Ye, Are vulnerable communities digitally left behind in social responses to natural disasters? An evidence from Hurricane Sandy with Twitter data, Appl. Geogr. 108 (2019) 1–8. https://doi.org/10.1016/j.apgeog.2019.05.001.
[23] J. Birkmann, O.D. Cardona, M.L. Carreño, A.H. Barbat, M. Pelling, S. Schneiderbauer, S. Kienberger, M. Keiler, D. Alexander, P. Zeil, T. Welle, Framing vulnerability, risk and societal responses: the MOVE framework, Nat. Hazards 67 (2013) 193–211. https://doi.org/10.1007/s11069-013-0558-5.
[24] B.E. Flanagan, E.W. Gregory, E.J. Hallisey, J.L. Heitgerd, B. Lewis, A social vulnerability index for disaster management, J. Homel. Secur. Emerg. Manag. 8 (2011) 0000102202154773551792.
[25] S.L. Cutter, B.J. Boruff, W.L. Shirley, Social vulnerability to environmental hazards, Soc. Sci. Q. 84 (2003) 242–261.
[26] C. Brunsdon, A.S. Fotheringham, M.E. Charlton, Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity, Geogr. Anal. 28 (1996) 281–298. https://doi.org/10.1111/j.1538-4632.1996.tb00936.x.
[27] S. Cheng, S. Ganapati, G. Narasimhan, F.B. Yusuf, A machine learning‐based analysis of 311 requests in the Miami‐Dade County, Growth Change 53 (2022) 1627–1645.
[28] K. Sun, R.Z. Zhou, J. Kim, Y. Hu, PyGRF: An Improved Python Geographical Random Forest Model and Case Studies in Public Health and Natural Disasters, Trans. GIS (2024) tgis.13248. https://doi.org/10.1111/tgis.13248.
[29] S. Georganos, T. Grippa, A. Niang Gadiaga, C. Linard, M. Lennert, S. Vanhuysse, N. Mboga, E. Wolff, S. Kalogirou, Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling, Geocarto Int. 36 (2021) 121–136. https://doi.org/10.1080/10106049.2019.1595177.
[30] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Adv. Neural Inf. Process. Syst. 30 (2017).
[31] C.-H. Sung, S.-C. Liaw, A GIS-based approach for assessing social vulnerability to flood and debris flow hazards, Int. J. Disaster Risk Reduct. 46 (2020) 101531.
[32] A. Mollalo, B. Vahedi, K.M. Rivera, GIS-based spatial modeling of COVID-19 incidence rate in the continental United States, Sci. Total Environ. 728 (2020) 138884.
[33] L. Anselin, S.J. Rey, Open Source Software for Spatial Data Science, Geogr. Anal. 54 (2022) 429–438. https://doi.org/10.1111/gean.12339.
[34] T. Oshan, Z. Li, W. Kang, L. Wolf, A. Fotheringham, mgwr: A Python Implementation of

Multiscale Geographically Weighted Regression for Investigating Process Spatial Heterogeneity and Scale, ISPRS Int. J. Geo-Inf. 8 (2019) 269. https://doi.org/10.3390/ijgi8060269.
[35] S. Athey, The impact of machine learning on economics, in: Econ. Artif. Intell. Agenda, University of Chicago Press, 2018: pp. 507–547. https://www.nber.org/system/files/chapters/c14009/c14009.pdf (accessed May 15, 2025).
[36] C. Agonafir, T. Lakhankar, R. Khanbilvardi, N. Krakauer, D. Radell, N. Devineni, A machine learning approach to evaluate the spatial variability of New York City's 311 street flooding complaints, Comput. Environ. Urban Syst. 97 (2022) 101854.
[37] A. Géron, Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems, O'Reilly Media, 2019.
[38] H.S.A. Magid, M.R. Desjardins, Y. Hu, Opportunities and shortcomings of AI for spatial epidemiology and health disparities research on aging and the life course, Health Place 89 (2024) 103323.
[39] H. Yu, A.S. Fotheringham, Z. Li, T. Oshan, W. Kang, L.J. Wolf, Inference in Multiscale Geographically Weighted Regression, Geogr. Anal. 52 (2020) 87–106. https://doi.org/10.1111/gean.12189.
[40] Z.A. Hamstead, Thermal insecurity: Violence of heat and cold in the urban climate refuge, Urban Stud. 61 (2024) 531–548. https://doi.org/10.1177/00420980231184466.
[41] G. Shmueli, To explain or to predict?, (2010). https://projecteuclid.org/journals/statistical-science/volume-25/issue-3/To-Explain-or-to-Predict/10.1214/10-STS330.short (accessed May 16, 2025).
[42] J. Cohen, L. Agel, M. Barlow, C.I. Garfinkel, I. White, Linking Arctic variability and change with extreme winter weather in the United States, Science 373 (2021) 1116–1121. https://doi.org/10.1126/science.abi9167.
[43] USGCRP, Fifth National Climate Assessment, U.S. Global Change Research Program (USGCRP), Washington, D.C., 2023. https://nca2023.globalchange.gov/.
[44] A. Nejat, L. Solitare, E. Pettitt, H. Mohsenian-Rad, Equitable community resilience: The case of Winter Storm Uri in Texas, Int. J. Disaster Risk Reduct. 77 (2022) 103070. https://doi.org/10.1016/j.ijdrr.2022.103070.
[45] J. Xu, Y. Qiang, H. Cai, L. Zou, Power outage and environmental justice in Winter Storm Uri: an analytical workflow based on nighttime light remote sensing, Int. J. Digit. Earth 16 (2023) 2259–2278. https://doi.org/10.1080/17538947.2023.2224087.
[46] P. Chen, W. Zhai, X. Yang, Enhancing resilience and mobility services for vulnerable groups facing extreme weather: lessons learned from Snowstorm Uri in Harris County, Texas, Nat. Hazards (2023) 1–22.
[47] R.Z. Zhou, Y. Hu, L. Zou, H. Cai, B. Zhou, Understanding the disparate impacts of the 2021 Texas winter storm and power outages through mobile phone location data and nighttime light images, Int. J. Disaster Risk Reduct. 103 (2024) 104339.
[48] S.S. Clark, M.L. Miles, Assessing the Integration of Environmental Justice and Sustainability in Practice: A Review of the Literature, Sustainability 13 (2021) 11238.
[49] Z. Liu, C. Liu, A. Mostafavi, Beyond residence: A mobility-based approach for improved evaluation of human exposure to environmental hazards, Environ. Sci. Technol. 57 (2023) 15511–15522.
[50] M. Deliso, M. Golembo, Massive winter storm continues to dump snow across the South, ABC News (2025).

[51] S. Ghaffarian, F.R. Taghikhah, H.R. Maier, Explainable artificial intelligence in disaster risk management: Achievements and prospective futures, Int. J. Disaster Risk Reduct. 98 (2023) 104123.
[52] Z. Liu, T. Felton, A. Mostafavi, Interpretable machine learning for predicting urban flash flood hotspots using intertwined land and built-environment features, Comput. Environ. Urban Syst. 110 (2024) 102096.
[53] M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, F. Prabhat, Deep learning and process understanding for data-driven Earth system science, Nature 566 (2019) 195–204.
[54] Z. Li, Extracting spatial effects from machine learning model using local interpretation method: An example of SHAP and XGBoost, Comput. Environ. Urban Syst. 96 (2022) 101845.

Supplementary

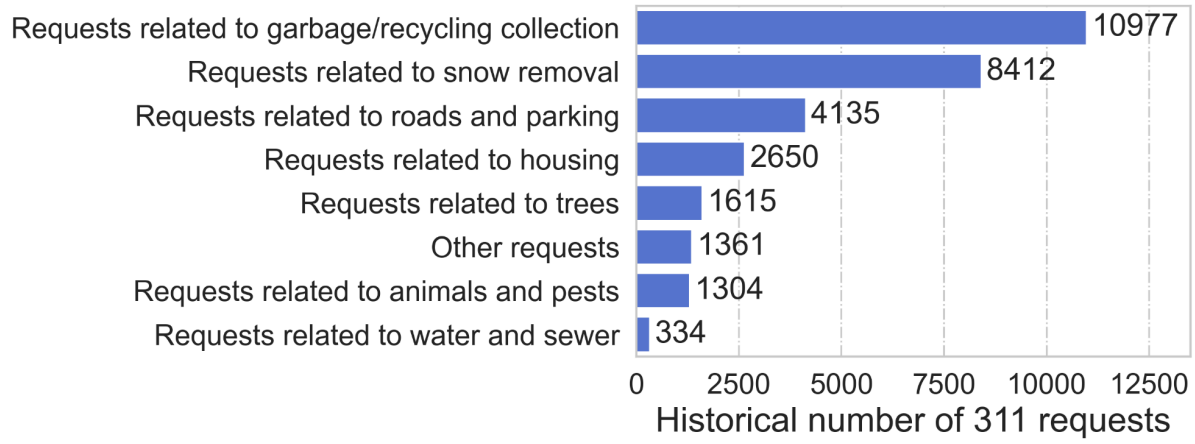

Figure S1. Historical 311 requests in eight categories during the previous snow seasons in 2021 and 2022 before the blizzard.

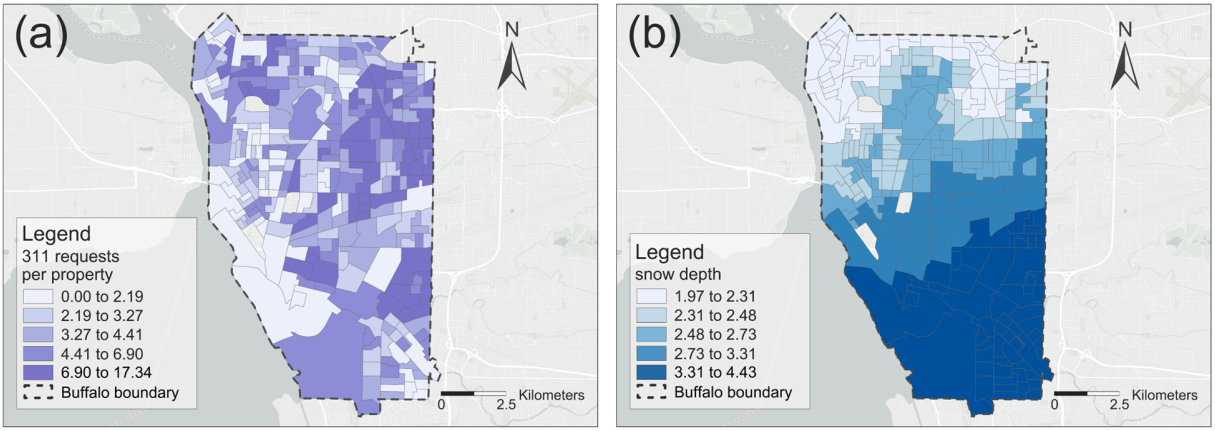

Figure S2. The spatial distributions of (a) 311 requests per property related to snow removal, and (b) accumulated snow depth (in meters) during the study period.

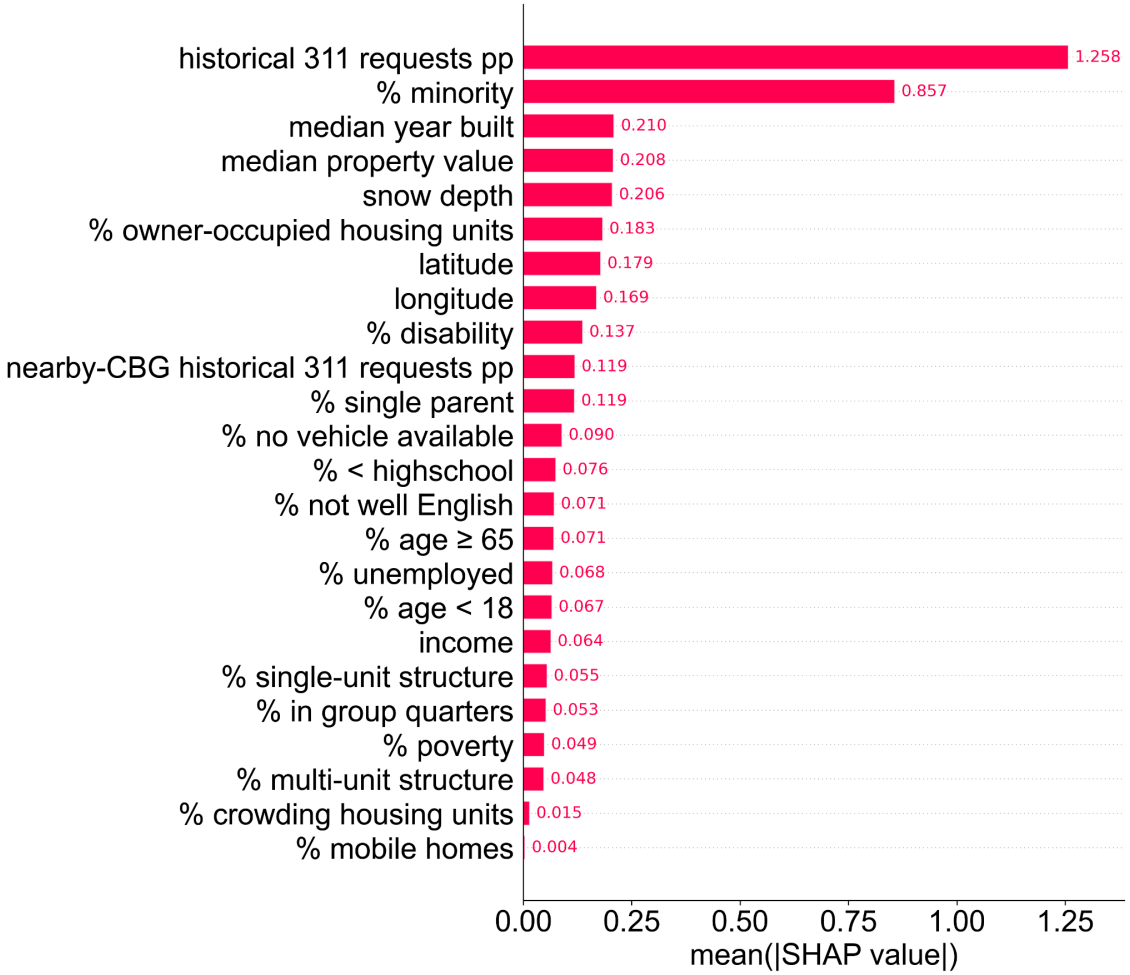

Figure S3. The mean absolute SHAP values based on the GRF model.

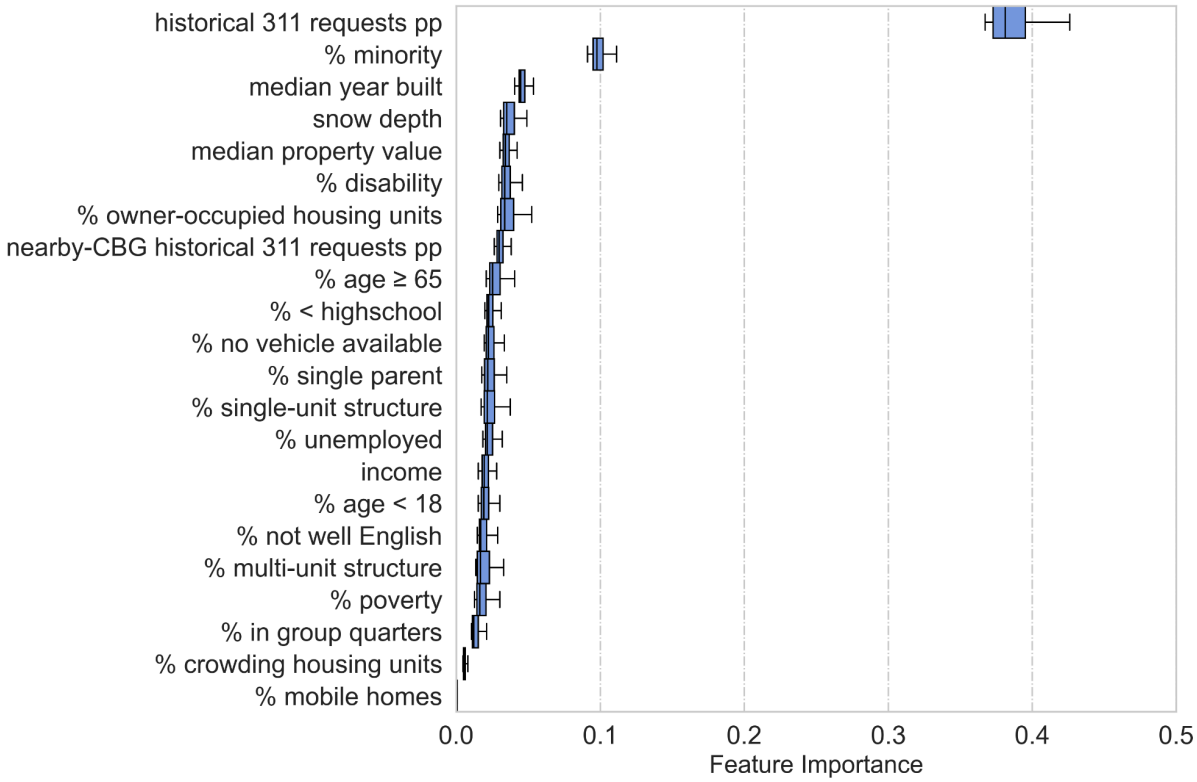

Figure S4. Feature importance of the independent variables output by the GRF model.