*Article*

# Ecological Associations between Obesity Prevalence and Neighborhood Determinants Using Spatial Machine Learning in Chicago, Illinois, USA

Aynaz Lotfata [1,*], Stefanos Georganos [2], Stamatis Kalogirou [3] and Marco Helbich [4]

1 Geography Department, Chicago State University, Chicago, IL 60605, USA
2 Division of Geoinformatics, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden
3 Data and Technology for Audit (DATA), European Court of Auditors, 1615 Luxembourg, Luxembourg
4 Department of Human Geography and Spatial Planning, Faculty of Geosciences, Utrecht University, 3584 CS Utrecht, The Netherlands
* Correspondence: alotfata@csu.edu

**Abstract:** Some studies have established relationships between neighborhood conditions and health. However, they neither evaluate the relative importance of neighborhood components in increasing obesity nor, more crucially, how these neighborhood factors vary geographically. We use the geographical random forest to analyze each factor's spatial variation and contribution to explaining tract-level obesity prevalence in Chicago, Illinois, United States. According to our findings, the geographical random forest outperforms the typically used nonspatial random forest model in terms of the out-of-bag prediction accuracy. In the Chicago tracts, poverty is the most important factor, whereas biking is the least important. Crime is the most critical factor in explaining obesity prevalence in Chicago's south suburbs while poverty appears to be the most important predictor in the city's south. For policy planning and evidence-based decision-making, our results suggest that social and ecological patterns of neighborhood characteristics are associated with obesity prevalence. Consequently, interventions should be devised and implemented based on local circumstances rather than generic notions of prevention strategies and healthcare barriers that apply to Chicago.

**Keywords:** obesity; neighborhoods; spatial variation; spatial machine learning; geographical random forest; spatial analytics

## 1. Introduction

The United States ranks 12th in the world regarding the number of overweight people [1]. Obesity prevalence was 40.0% among individuals aged 20 to 39 in 2020, 44.8% among those aged 40 to 59, and 42.8% among those aged 60 and up. Obesity is related to heart disease, stroke, type 2 diabetes, and several cancer types [2]. Although an active lifestyle reduces the risk of obesity, most adults in developed and developing countries cannot meet the recommended levels of physical activity due to sedentary lifestyles, the use of passive modes of transportation, etc. [3,4].

While it is generally understood that individual-level factors such as genetic predisposition [5] and behavioral aspects (e.g., physical activity) [6] play a role in weight gain, more research into the influence of residential neighborhood characteristics is needed to provide a multidisciplinary understanding of obesity prevalence. Environmental characteristics such as urban form, neighborhood safety, socioeconomic capital, and food availability have received attention in the literature [7–9]. As suggested elsewhere [7–11], these determinants are likely related to a wide range of social and environmental traits such as urban mobility and crime. Evidence also suggests that the built environment can positively influence health behaviors or be a health stressor [12]. For instance, the risk of obesity rises with residential instability and unaffordability of rent [13]. Although housing costs and house

insecurity (i.e., homelessness) are a significant burden in the United States, little is known about their impact on obesity prevalence. Cost-burdened low-income households have limited resilience in economic crises or job loss, resulting in housing insecurity and other significant sacrifices that harm health [13]. It reduces a household's ability to pay for health-promoting necessities such as nutritious food, healthcare visits, energy, and home maintenance [14–16]. Additionally, urban amenities such as parks, bike lanes, and playgrounds encourage an active lifestyle and reduce the incidence of obesity [17–21]. While green spaces are important in reducing obesity concerns [22], there are inconsistent results concerning the association between green spaces and obesity [23].

To date, obesity is primarily investigated in public and behavioral health, and to a lesser extent, from a geographical science perspective. Spatial modeling could be a powerful means for urban health practitioners to grasp geographical patterns and dynamics that otherwise remain unnoticed. While most obesity research relies either on linear aspatial [24,25] or linear spatial models [26–31], nonlinear spatially explicit modeling to investigate the relationships between tract-level obesity prevalence and socioenvironmental factors is lacking, as we are aware of. For example, Ferdowsy et al. [32] used nonlinear random forest modeling to assess obesity risk by means of behavioral factors. Ghosh and Guha [33] employed latent Dirichlet allocation to investigate obesity-related themes in Twitter data. Further, obesity studies on the built environment are undertaken at the microscale utilizing geospatial technologies such as geographical positioning systems [34–36] and a focus on pediatric obesity [37,38]. However, obesity is context dependent and driven by an interplay of policy, social, economic, cultural, environmental, behavioral, and biological factors, as well as cross-sector and nonlinear interactions across these dimensions [39,40]. Collinearity across socioenvironmental linear models complicates the analyses in this way, necessitating the development of collinearity-aware models [41]. We investigate the spatial distribution of obesity prevalence in locals using the geographical random forest (GRF) model.

The GRF is a novel tree-based spatial machine-learning model [42,43]. It has the advantage of not presupposing local linearity and often outperforms an aspatial random forest model in predictive performance [44], but at the cost of greater computational complexity [45]. The GRF model is conceptually inspired by the geographically weighted regression [46], except that it is calibrated using a random forest rather than traditional least squares. GRF, unlike geographically weighted regression, does not need to account for multicollinearity and can evaluate all independent variables without the requirement for collinearity screening. It may also be used to examine local relationships between independent and dependent geographical variables while typically resulting in higher prediction accuracy than geographically weighted regression [42,44].

To our knowledge, no study used the GRF to examine spatial variation (i.e., nonstationarity) in the associations between tract-level obesity prevalence and socio-environmental neighborhood variables. Our goals were to (a) investigate local associations between obesity prevalence and tract-level variables in order to focus prevention and intervention efforts in high-risk areas and (b) compare GRF's prediction performance to that of a traditional random forest regression model. Of note, our goal was not to propose a holistic approach to model and map obesity prevalence by considering every available factor, but rather to present a new spatial approach to the community to understand obesity prevalence.

GRF was recently used to model socioeconomic circumstances in the European Union regions [47]. Furthermore, it was used to model the relative importance of 29 socioeconomic and health-related factors to the COVID-19 death rate, outperforming commonly used local and global regressions [44]. It was also applied to predict diabetes prevalence in the United States [42]. However, besides this limited number of applications, more evidence is needed.

## 2. Materials and Methods

### 2.1. Data

Obesity is a major problem in Chicago, where 61.2% of adults in the metropolitan area are overweight or obese [48]. Obesity was especially recognized as a contributing

cause of death in metropolitan cities such as Chicago, Illinois [48]. We obtained cross-sectional, ecological data for all 793 census tracts in Chicago from various sources. We obtained census tract polygon geometries as TIGER/Line Shapefiles from the United States Census Bureau to conduct spatial analysis and mapping using geographic information systems (GIS) [49]. With an average size of 0.28 square miles (standard deviation [SD] $\pm 0.39$), census tracts were deemed a suitable analytical scale for assessing area-level obesity prevalence. The 11-digit Federal Information Processing Standards codes were used to enrich the geometries with census tract-level variables.

We obtained estimates of obesity prevalence per census tract based on responses to the Behavioral Risk Factor Surveillance System survey and the Centers for Disease Control and Prevention's PLACES Project [50]. Obesity is defined as having a body mass index (BMI) of at least $30 \text{ kg/m}^2$ using self-reported height and weight data. The obesity prevalence estimates for the year 2020 served as our response variable (Table 1).

**Table 1.** Source and Nature of Data.

| Dataset | Source | Data Nature |
| --- | --- | --- |
| Estimates of obesity prevalence per census tract | Centers for Disease Control | Ratio (percentage) |
| Census tract polygon geometries | United States Census Bureau | TIGER/Line Shapefiles |
| Crime ratio | Chicago Police Department | Ratio |
| Poverty | Chicago Health Atlas | Ratio (percentage) |
| Unemployment | Chicago Health Atlas | Ratio (percentage) |
| Eviction rate | Chicago Health Atlas | Ratio (percentage) |
| Biking | Divvy Bike share system | Ratio |
| Green space | Google Earth Engine | Index |
| Severe rent | Chicago Health Atlas | Ratio (percentage) |
| Vacant housing | Chicago Health Atlas | Ratio (percentage) |

Based on previous studies [51,52], we selected eight covariates. The years of the covariates match the obesity data in 2019. First, we added severe rent as a percentage of the population spending more than 50% of their income on housing rent. Households with high monthly expenses have little money left over for other necessities such as food, clothing, utilities, and health care. These experiences may negatively affect physical and mental health [13]. Second, the poverty percentage was quantified based on income that varies by family size and composition, accounts for the uneven distribution of income among a population; third, the unemployment percentage to the corresponding census tract population; fourth, the eviction percentage, eviction filings per 100 renter-occupied households; and fifth, percentage of vacant housing per census tract by the Chicago Health Atlas [52]. Sixth, we included available green space (e.g., parks, green open spaces, residential gardens) using the averaged Normalized Difference Vegetation Index per census tract. Remote sensing data were obtained from Landsat 8 NASA Earth Data [53,54]. Seventh, using data from the Divvy Bike share system, we calculated bicycle usage as the ratio of bike trips to, from, and within census tracts to the corresponding population [55]. Eighth, we considered crime extracted from the Chicago Police Department [56] as the ratio between the number of all types of crimes and the corresponding census tract population. The crime data refer to the locations of crime incidents (Table 1).

### 2.2. Analytical Approaches

Aspatial Random Forest

Since the repertoire of machine learning techniques is large [57], we selected a well-established and typically well-performing regression-based model, namely the random forest, to assess the association between obesity prevalence and neighborhood determinants. The standard random forest is a global machine learning approach assessing associations

uniformly across space, while the model can assess non-linearities and interactions between variables [44]. Moreover, there are no strict statistical assumptions [58].

Briefly, a random forest comprises a collection of separate decision trees for regression analyses [57]. Each decision tree is fitted from a given training dataset. First, a subset is created by randomly selecting samples with replacements from the original training set (usually 2/3 of the training set). The remaining data (usually the other 1/3) referring to the out-of-bag set is used for model evaluation (i.e., the assessment of the predictive accuracy). Additionally, a random subset of covariates is also picked for each node in each decision tree. The same approach is repeated for a large number of iterations, yielding a forest of trees trained with random subsets of training data. Finally, each tree's prediction error is computed, and all trees' final output is the average prediction value.

The out-of-bag accuracy is a robust independent measure frequently used to evaluate each variable's relevance and the overall model performance [57]. We used an increase in Mean Square Error (IncMSE) as a metric for determining the relevance of each variable [58]. The out-of-bag error is calculated by randomly permuting the values of each variable in the out-of-bag set. If the out-of-bag error rises, the variable is deemed to be important; and the greater the change, the more relevant the variable is in estimating the dependent variable [44,58]. However, the traditional random forest results in a single regression model assumed to be valid for the entire study area. As a result, the algorithm fails to account for geographic variations in the associations, which may lead to an inadequate representation of the associations.

### 2.3. Geographical Random Forest

To circumvent the restrictive assumption of a uniform association of the traditional random forest, we fitted a geographical random forest (GRF) with spatial weighting capable of modeling spatial non-stationarity [47]. Technically, GRF is calibrated locally using only nearby observations through a spatial kernel and a spatial weights matrix [59]. The main principle of GRF is similar to the geographically weighted regression [60,61], in which a moving window is applied to create local submodels. Each local random forest is evaluated for each site depending on the input data from surrounding observations. We used an adaptive spatial kernel for the GRF since it is widely used when data points are unevenly distributed spatially and when spatial autocorrelation is assumed to be present in the data [47,59]. We used the minimized out-of-bag error to determine an optimal bandwidth (BW). For a thorough discussion of the GRF, see Georganos et al. [47,59].

To determine the GRF's optimal hyperparameter (i.e., the number of trees and the proportion of randomly sampled features at each node), we used a random grid search as performed elsewhere [42,44]. From a set of possible hyperparameter combinations, we utilized 10-fold cross validation to determine the most suitable ones. We used a bandwidth of 30 observations, the number of trees was set to 1000, and the number of variables randomly sampled as candidates at each split was set to 5. Both the random forest and the GRF models were trained with these hyperparameters. We then compute performance metrics such as the mean square error (MSE), mean absolute error (MAE), root-mean-square error (RMSE), and coefficient of determination ($R^2$). Since significant residual spatial autocorrelation violates regression assumptions concerning the independence of observations, we used the well-established Moran's *I* statistic to investigate the degree of residual spatial autocorrelation. We also use the local Moran's *I* to track potential spatial residual clustering [60].

We used the permutation feature importance approach to evaluate the predictors' role in the random forest and GRF models. While the former is a global and aspatial model, GRF decomposes a random forest in local sub-models, considering, therefore, data-inherent nonstationarity and spatial autocorrelation. GRF yields local feature importance, local residuals, and local goodness of fit statistics for each predictor in each local random forest model [44]. Similar to the random forest, we ranked the variables' importance based on the percent change in the MSE [44]. Further, we mapped the local variable

importance to examine how each independent variable's effect on obesity prevalence varies geographically.

Partial dependence plots were used to characterize the nonlinear relationships between the obesity prevalence and the covariates. The partial dependence plots reveal whether the relationship between the target and a feature is linear, monotonic, curvilinear, or more complex by presenting the expected target response as a function of the input features of interest [62]. All our statistical analyses are conducted with the "SpatialML" package [63] in the R Statistical Computing Environment [64]. For cartography, we used ArcGIS 10.8.1.

## 3. Results

Figure 1 depicts the obesity prevalence and socioenvironmental determinants at the tract level. Obesity is more prevalent in Chicago's southern and western neighborhoods (tracts). The south and northwest have the greenest space, whereas the downtown and surrounding neighborhoods have a high percentage of people who bike. The share of low-income and unemployed persons is largest on the southern and western sides. Criminal activity is centered on the west and south sides, where there is a significant percentage of vacant homes, severe rents, and high eviction rates, increasing the vulnerability of unemployed and low-income people.
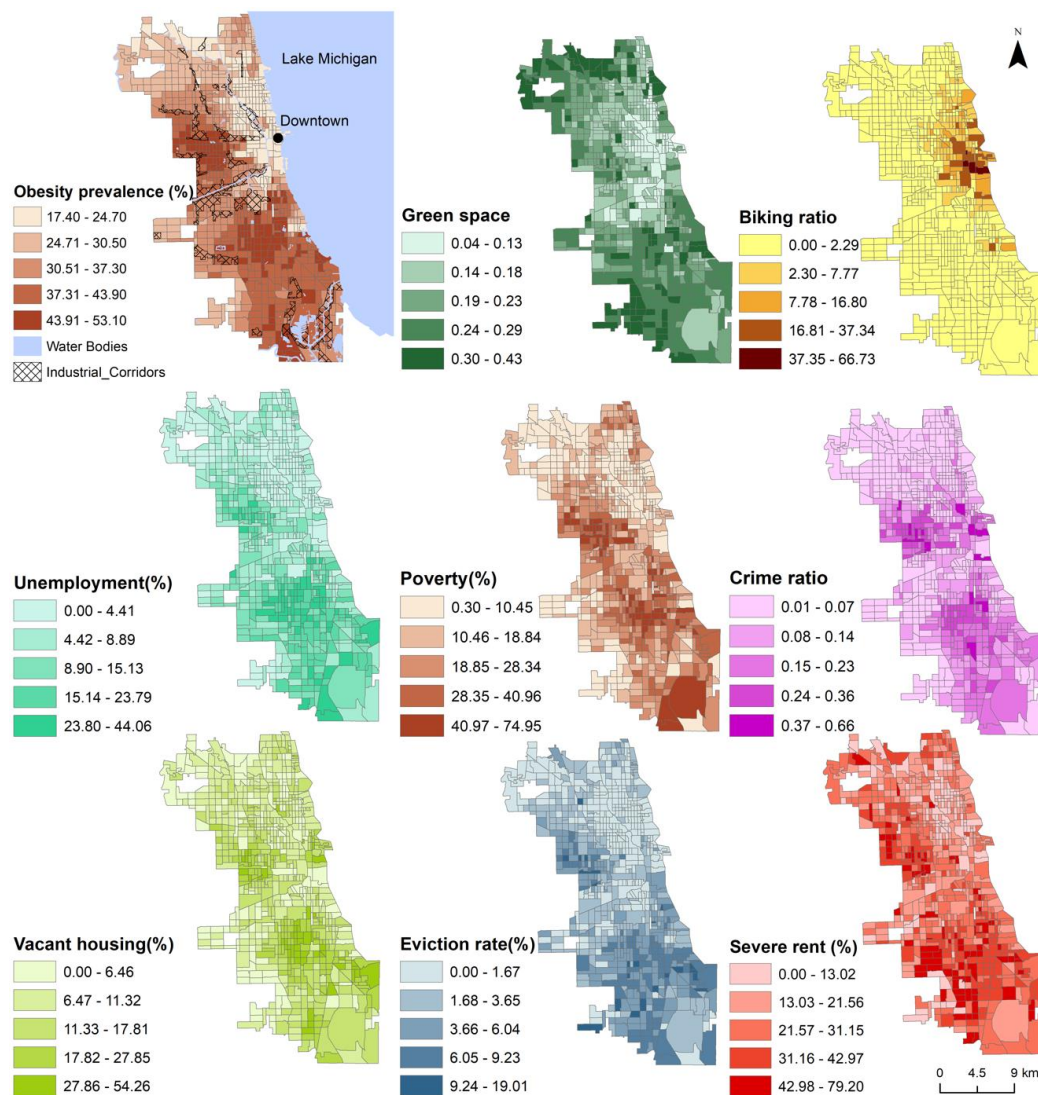


**Figure 1.** Tract-level obesity prevalence (%) and the socio-environmental determinants.

*Aspatial Random Forest and GRF Results*

In the out-of-bag set, the GRF model had a lower MSE, RMSE, and MAE and a higher $R^2$ than the global random forest model (Table 2). Overall, the associations show complex shapes, highlighting the need to utilize nonlinear models. Figure 2 depicts the nonlinear relationship between obesity prevalence and the covariates. The covariates biking (Figure 2b), crime (Figure 2c), unemployment (Figure 2d), and eviction rate (Figure 2e) were curvilinearly associated with obesity prevalence, while vacant housing (Figure 2f) and severe rent (Figure 2h) show a rather nonlinear relationship with obesity prevalence. While the overall associations indicate nonlinearities, linear correlations exist within specific ranges. There is, for example, a positive linear association between eviction rate and obesity prevalence in the range below 2%, but the effect remains stable thereafter (Figure 2e). Similarly, in the range between 15–27%, there is a linear positive association between severe rent and obesity prevalence, but the effect is neglectable after that (Figure 2h).

**Table 2.** Results of random forest and geographically random forest.

| | Random Forest | | Geographical Random Forest | | | | |
|---|---|---|---|---|---|---|---|
| | | | | Local Feature Importance (MSE) | | | |
| Rank | Variable | Global Feature Importance (MSE) | Variable | Min. | Max. | Mean | Std. Dev. |
| 1 | Crime ratio | 19.480 | Poverty | −5.948 | 63.174 | 4.842 | 7.561 |
| 2 | Poverty | 17.169 | Crime ratio | −2.702 | 83.830 | 4.236 | 8.772 |
| 3 | Unemployment | 13.408 | Unemployment | −2.808 | 100.967 | 3.265 | 10.735 |
| 4 | Eviction rate | 10.981 | Eviction rate | −4.412 | 48.119 | 2.238 | 5.076 |
| 5 | Biking | 8.443 | Vacant housing | −6.4041 | 22.261 | 1.046 | 2.493 |
| 6 | Green space | 2.841 | Green space | −3.849 | 24.330 | 1.007 | 2.719 |
| 7 | Severe rent | 1.574 | Severe rent | −5.697 | 12.340 | 0.593 | 1.492 |
| 8 | Vacant housing | 1.048 | Biking | −1.561 | 44.731 | 0.488 | 2.460 |
| | $R^2$ (out-of-bag) | 0.867 | 0.904 | | | | |
| | MSE (out-of-bag) | 10.401 | 7.466 | | | | |
| | RMSE (out-of-bag) | 2.786 | 2.658 | | | | |
| | MAE (out-of-bag) | 2.016 | 1.972 | | | | |

Table 2 compares the importance of these covariates in the random forest and GRF model. The crime ratio is the most important variable in the former, followed by poverty, unemployment, and eviction rate, according to the permutation-based feature importance, whereas poverty is the most important variable in the GRF, followed by the crime ratio, unemployment, and eviction rate. As shown in Table 2, the average positive MSE of the GRF model shows that most tracts have positive local covariate importance. Other determinants' importance ordering also varies from that of the global random forest model. Biking, for example, is placed fifth in the random forest model but last in the GRF model (Table 2). The difference in feature importance is likely because the random forest is a global model and does not take into spatial and local variations. In contrast, GRF assesses the spatial (i.e., local) variation of the predictor variables.

Additionally, we mapped the determinants to understand better the spatial distribution of the local variable importance (IncMSE) (Figure 3). It is important to note that values above zero have importance on obesity prevalence, while higher tract values suggest greater importance. Green spaces seem to be of minor importance across the city (Figure 3a). It is the same for biking, with the exception of a few places southwest of downtown (Figure 3b). Unemployment, on the other hand, is most important in the area south of downtown (Figure 3c). Poverty is most prevalent in various areas to the south and southwest of downtown (Figure 3d). The crime rate is most important in southern tracts (Figure 3e). Except for severe rent, which has the highest importance in the south, vacant housing (Figure 3f), eviction rate (Figure 3g), and severe rent (Figure 3h) have the same level of importance across Chicago.
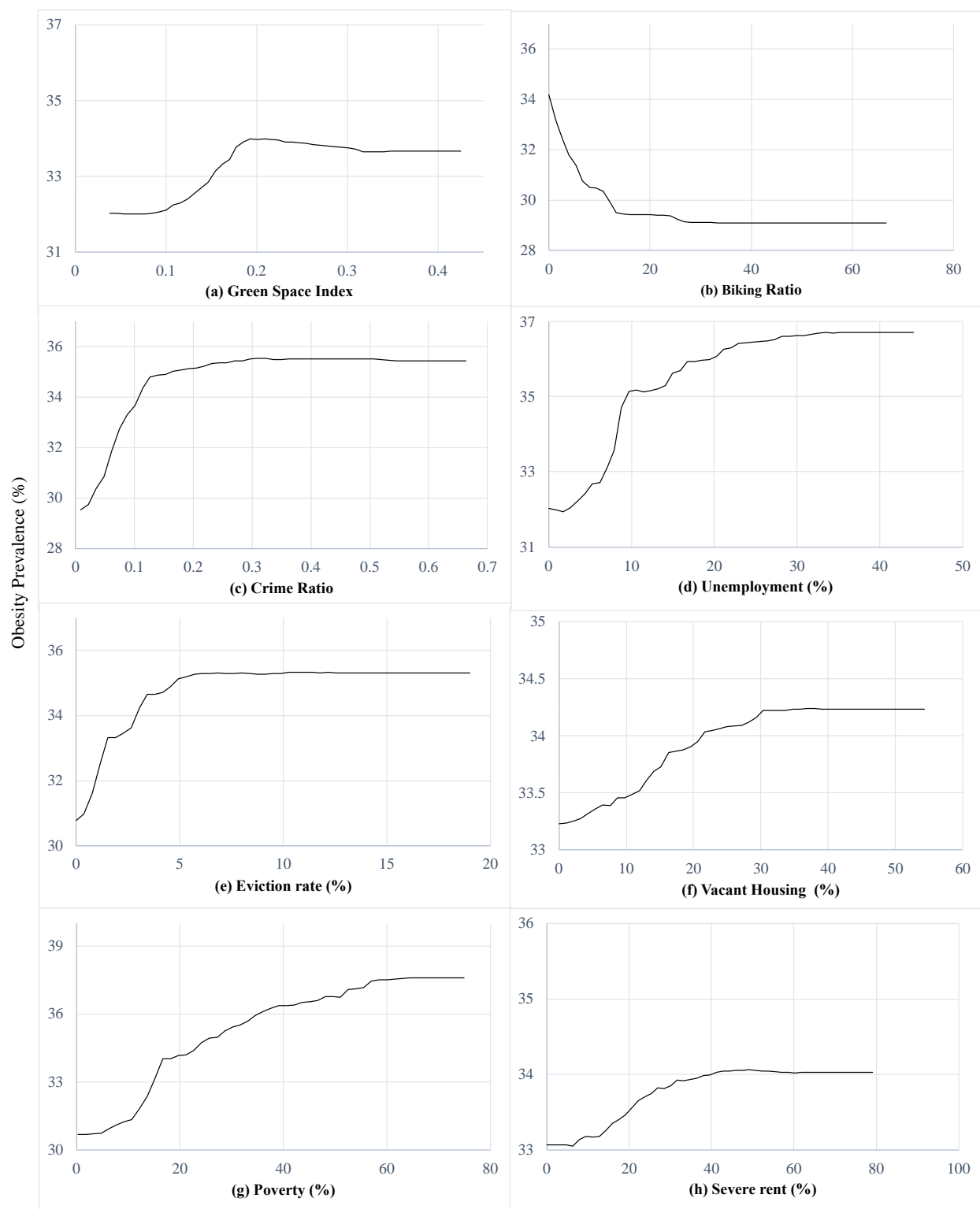
**Figure 2.** Partial dependency plots of the determinants of the global random forest.
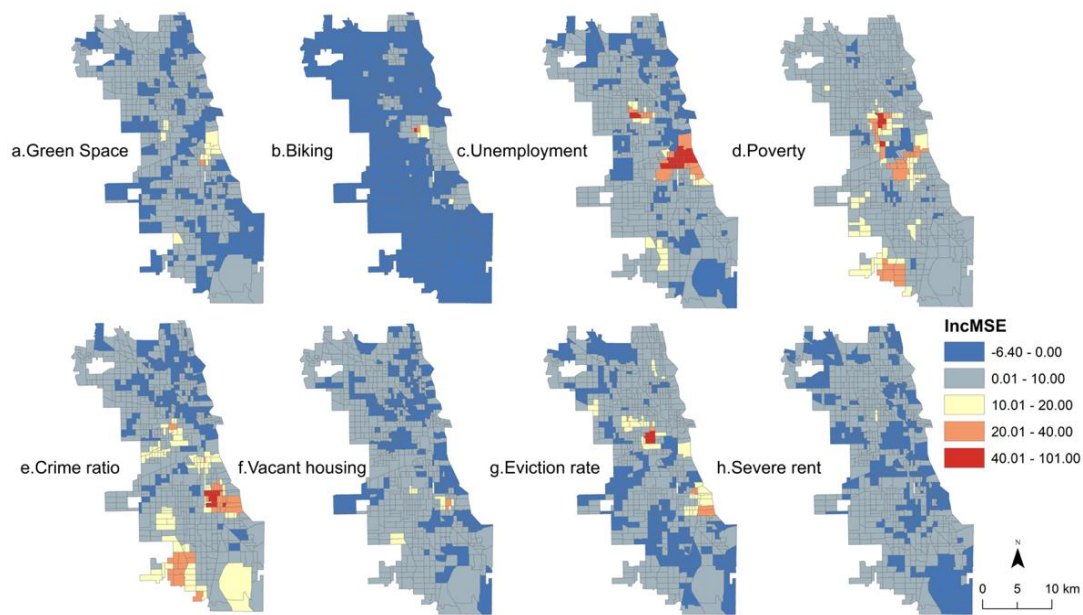
**Figure 3.** Spatial variation of local feature importance (IncMSE) of green spaces, biking, crime ratio, poverty, vacant housing, unemployment, severe rent, and eviction rate using the geographically random forest model. Negative values indicate features with no predictive contribution.

Figure 4a depicts the local $R^2$ of the GRF. The model fit varied across space, ranging between 35% to 60%. Predominantly tracts in the north of the city, the $R^2$ was below 0.5 (Figure 4a). The global Moran's $I$ test ($I = -0.01$, $p = 0.18$) confirms there is no spatial residual autocorrelation. Additionally, the local Moran's $I$ also reveal that there is no geographical clustering of residuals in most locations and that the residuals are randomly distributed (Figure 4c). These results confirm that the GRF models the data well.
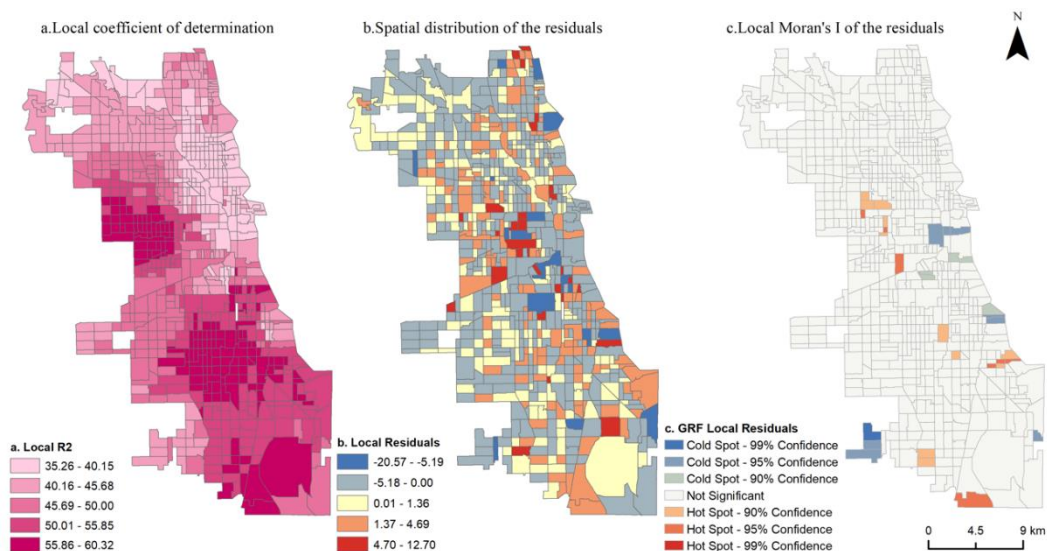


**Figure 4.** (**a**). Local coefficient of determination ($R^2$). (**b**). Spatial distribution of the standardized local residuals. (**c**). Local Moran's $I$ of standardized residuals.

## 4. Discussion

This cross-sectional ecological study looks at the prevalence of obesity at the tract level in Chicago using the geographical random forest (GRF), an innovative spatial machine learning approach. According to our findings, the GRF model outperformed the typically used aspatial random forest model in terms of prediction accuracy. This suggests that

GRF considers spatial heterogeneity in the associations and identifies the factors that trigger local variations in obesity prevalence rates that contribute to developing place-based interventions to control obesity. Our finding corroborates the results of a few other studies using the GRF model [23,44,47]. We also found three of the top four most important local factors (i.e., poverty, crime, and unemployment) refer to neighborhood determinants, and one is associated with housing (eviction rate). The ramifications of the findings are described below.

### 4.1. Poverty

Poverty has the greatest proportionate importance per census tract in the west and southwest downtown neighborhoods, while it has the least importance in the rest of Chicago (Figure 3d). In addition, the overall positive association between obesity and poverty in Figure 2d demonstrates the importance of poverty. Poverty creates an obesogenic environment in which people may lack access to affordable, healthful foods [65], lack funds for sports equipment and physical activity participation, and are exposed to psychological stress [66], as well as live in overcrowded houses with poor sleep quality [67]. However, the relationship between high wealth and increased physical activity is not well established and requires further research [68]. Since poverty is associated with an increased risk of obesity, policymakers and planners should assess the consequences of neighborhood poverty on health outcomes [69]. Deprivation promotes the formation of harmful habits and cultures, which are then passed down through generations. Obesity risk in emerging adulthood is significantly increased by cumulative exposure to neighborhood poverty. The deterioration of neighborhood socioeconomic conditions is also a significant obesity risk factor [69]. Our study suggests that multidisciplinary policies and organizations (such as the USDA Food and Nutrition Service and housing authorities) work together to control and reduce obesity.

### 4.2. Crime

The importance of crime per census (i.e., a variety of illegal behaviors) tract is greatest in the southern districts, although it is not widespread in Chicago's northern neighborhoods (Figure 3e). Criminality is strongly associated with obesity prevalence along southern Lake Michigan, where a large percentage of people face severe rent, housing insecurity, poverty, and unemployment (Figure 1). The community and individual level factors may be associated with neighborhood crime. Business withdrawal, population outmigration, physical deterioration, declining community resources, and crumbling public infrastructure are witnessed at the community level in high-crime zones that are unsafe for physical exercise [70,71]. In addition, individuals' perceptions of unsafety, anxiety, and stress are influenced by neighborhood crime, which restricts participation in physical activities [72]. While dealing with crime issues at the neighborhood level, crime types should be treated separately because related institutions and policies should regulate crimes (home burglary vs. robbery). Well-designed neighborhoods with well-maintained socioeconomic capitals (i.e., sidewalks) encourage healthy behaviors and inhibit illegal behaviors.

### 4.3. Unemployment

The importance of unemployment per census tract is highest in the southern downtown neighborhoods along Lake Michigan (Figure 3c). Similarly, unemployment is strongly associated with obesity in southwest downtown south districts, where green spaces are abundant, but eviction rates are high (Figure 1). Unemployment is well known to be associated with an increased risk of illness. This association could be partly explained by the negative health-related behaviors—particularly smoking, diet, exercise, and alcohol consumption—caused by the lower income, altered daily routine, and psychological stress that typically accompany job loss [73]. However, Hughes et al. [51] found that job seekers were less likely to be overweight than never-unemployed people, implying that unemployment may vary with BMI for different populations. Ruhm [73] finds that a 1% increase in the state unemployment rate is associated with a 2% decrease in daily fat

intake between 1987 and 1995; also he finds that during economic downturns, body weight falls among the severely obese and exercise increases among the previously physically inactive. Cutler et al. [74] find that a higher unemployment rate at graduation is associated with lower income and greater obesity later in life. Deb et al. [75] explore the impact of business closures on obesity. They find that job separation increases the likelihood of being obese more significantly for females, lower incomes, those least educated, and the middle aged, compared to elderly individuals. Unemployment should be studied primarily at the household and individual levels regarding obesity prevalence.

### 4.4. Eviction Rate

The importance of the eviction rate per census tract is greatest in the south, although it is also high in some western neighborhoods (Figure 3g). Along Lake Michigan, where unemployed people dwell in poverty, the eviction rate is highly associated with obesity (Figure 1). The association between eviction of renter-occupied properties and health is not well understood [76], while housing affordability and eviction are inextricably associated. However, little is known about how policy interventions, such as supply-side housing subsidy programs designed to increase affordable housing, affect local eviction dynamics [77]. Housing insecurity and unaffordability can lead to stress, worry, and despair, as well as change metabolism and raise the risk of obesity [78–80]. During the COVID-19 global pandemic, when tenants had lost their employment, research on eviction rates at the neighborhood level is highlighted. Our findings draw attention to the need for localities to respond quickly in order to protect public health from the obesity epidemic through current measures such as increasing rental assistance and extending the eviction moratorium [81].

### 4.5. Other Factors

Other factors such as green spaces, severe rent, and vacant housing had minimal importance on the prevalence of obesity in Chicago. However, the literature [82] highlights the importance of green spaces as a valuable resource for physical activity and hence has the potential to contribute to reducing obesity and improving health. It piques the interest of planners and policymakers researching green space functionalities in terms of accessibility, availability, and visibility in Chicago, primarily focused on the tourism industry, rather than using green infrastructure to promote health. Similarly, biking infrastructure concentrated in the city center serves to promote tourism rather than physical activity [83]. A series of connected bike routes allows neighbors to quickly get to all places by bike, primarily in the city center. Integrating walking and cycling routes with green space is critical for creating a built environment that promotes physical activity. Additionally, unoccupied properties, such as underutilized urban spaces, cause urban blights, foster crime in neighborhoods, and might indirectly affect an active lifestyle [84].

### 4.6. Limitations

The tract-level estimates, which are assessed in several prior studies [85,86], have certain limitations. The outcome variable was collected by means of a telephone survey which likely faces problems due to recall bias and social desirability bias [87,88]. We cannot exclude that the population-level bias in self-reported weight and height is larger in telephone interviews than in in-person interviews to measure BMI. In addition to the data limitation, locally weighted models have some drawbacks. For example, we used an adaptive kernel bandwidth to choose the optimal number of tracts to train the GRF that accounts for differences in tract size. The tracts with different sizes varying across a geographic area may result in spillover effects of the dependent variable in neighboring tracts or spatial autocorrelation of the residuals. Not unexpectedly, the local $R^2$ in the GRF varied across space. For most tracts in the north of the city, the $R^2$ was below 0.5. These results indicate that the included variables only explained a limited fraction of the variance of the outcome variable and alternative variables should be included to improve

the performance of the local models in these regions. While our results may be sensitive to the underlying analytical scale and zoning, causal inference is also hampered by the cross-sectional and ecological nature of the data [89].

## 5. Conclusions

This study is the first to use the GRF model with spatial weights to assess geographic variations of obesity prevalence at Chicago's tract level in response to the determinants associated with neighborhood conditions. The GRF outperforms the typically used nonspatial random forest model in terms of out-of-bag prediction accuracy. Poverty is the most important factor in Chicago tracts, while biking is the least important. While poverty is the most important predictor in Chicago's south suburbs, crime is the most important factor in explaining obesity's prevalence. Future research should look at other aspects of household quality (such as mobile homes, homelessness, and ecological factors), as well as the spatial behavior of the obesity epidemic from a neighborhood-household standpoint as a whole. Understanding the spatial heterogeneity of obesity-determinant correlations could support place-based policy developments to address the spatially varying obesity determinant.

**Author Contributions:** Aynaz Lotfata, Stefanos Georganos: Conceptualization, Methodology, Software. Aynaz Lotfata: Data curation, Writing—Original draft preparation. Aynaz Lotfata: Visualization, Investigation. Stefanos Georganos, Stamatis Kalogirou, Marco Helbich: Supervision. Aynaz Lotfata, Stefanos Georganos: Software, Validation. Aynaz Lotfata, Stefanos Georganos, Stamatis Kalogirou, Marco Helbich: Writing—Reviewing and Editing. All authors have read and agreed to the published version of the manuscript.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this paper are openly available at https://www.cdc.gov/obesity/data/prevalence-maps.html (accessed on 20 May 2022).

## References

1. World Health Organization (WHO). 2021. Available online: https://www.who.int/news-room/facts-in-pictures/detail/6-facts-on-obesity (accessed on 11 January 2022).
2. U.S. Center for Disease Control and Prevention (CDC). Adult Obesity Facts. 2021. Available online: https://www.cdc.gov/obsity/data/adult.html#:~:text=From%201999%20%E2%80%932000%20through%202017,and%20certain%20types%20of%20cancer (accessed on 12 December 2021).
3. Smith, J.D.; Fu, E.; Kobayashi, M.A. Prevention and Management of Childhood Obesity and Its Psychological and Health Comorbidities. *Annu. Rev. Clin. Psychol.* **2020**, *16*, 351–378. [CrossRef] [PubMed]
4. World Health Organization (WHO). Physical Inactivity: A Global Public Health Problem. 2018. Available online: http://www.who.int/dietphysicalactivity/factsheet_inactivity/en/ (accessed on 11 January 2022).
5. Loos, R.J.F.; Yeo, G.S.H. The genetics of obesity: From discovery to biology. *Nat. Rev. Genet.* **2021**, *23*, 120–133. [CrossRef] [PubMed]
6. Zhang, Y.; Yang, J.; Hou, W.; Arcan, C. Obesity Trends and Associations with Types of Physical Activity and Sedentary Behavior in US Adults: National Health and Nutrition Examination Survey, 2007–2016. *Obesity* **2021**, *29*, 240–250. [CrossRef] [PubMed]
7. Adlakha, D.; Brownson, R.C.; Hipp, J.A. Built environment correlates of overweight and obesity among adults in Chennai, India. *Cities Health* **2021**, *5*, 289–297. [CrossRef]
8. An, R.; Shen, J.; Yang, Q.; Yang, Y. Impact of built environment on physical activity and obesity among children and adolescents in China: A narrative systematic review. *J. Sport Health Sci.* **2019**, *8*, 153–169. [CrossRef]
9. Xu, Y.; Wen, M.; Wang, F. Multilevel built environment features and individual odds of overweight and obesity in Utah. *Appl. Geogr.* **2015**, *60*, 197–203. [CrossRef]
10. Lenhart, C.M.; Wiemken, A.; Hanlon, A.; Perkett, M.; Patterson, F. Perceived neighborhood safety related to physical activity but not recreational screen-based sedentary behavior in adolescents. *BMC Public Health* **2017**, *17*, 722. [CrossRef]

11. Unger, E.; Diez-Roux, A.V.; Lloyd-Jones, D.M.; Mujahid, M.S.; Nettleton, J.A.; Bertoni, A.; Badon, S.E.; Ning, H.; Allen, N.B. Association of Neighborhood Characteristics With Cardiovascular Health in the Multi-Ethnic Study of Atherosclerosis. *Circ. Cardiovasc. Qual. Outcomes* **2014**, *7*, 524–531. [CrossRef]

12. Frank, L.D.; Engelke, P. Multiple Impacts of the Built Environment on Public Health: Walkable Places and the Exposure to Air Pollution. *Int. Reg. Sci. Rev.* **2005**, *28*, 193–216. [CrossRef]

13. Rodgers, J.; Briesacher, B.A.; Wallace, R.B.; Kawachi, I.; Baum, C.F.; Kim, D. County-level housing affordability in relation to risk factors for cardiovascular disease among middle-aged adults: The National Longitudinal Survey of Youths 1979. *Health Place* **2019**, *59*, 102194. [CrossRef]

14. Almardood, M.A.; Maghelal, P. Enhancing the use of transit in arid regions: Case of Abu Dhabi. *Int. J. Sustain. Transp.* **2019**, *14*, 375–388. [CrossRef]

15. Wang, L.; Qi, J. Association between Family Structure and Physical Activity of Chinese Adolescents. *BioMed Res. Int.* **2016**, *2016*, 4278682. [CrossRef]

16. Wang, N. Downsizing: How Micro Apartments are the Step Toward Better Affordable Housing in New York 2014. Available online: https://bit.ly/2RBCWJo (accessed on 10 December 2020).

17. Xiao, Y.; Miao, S.; Zhang, Y.; Xie, B.; Wu, W. Exploring the associations between neighborhood greenness and level of physical activity of older adults in shanghai. *J. Transp. Health* **2022**, *24*, 101312. [CrossRef]

18. Peng, B.; Erdogan, S.; Nasri, A.N.; Zou, Z. Towards a health-conscious transportation planning: A framework for estimating health impacts of active transportation at local level. *J. Transp. Health* **2021**, *22*, 101231. [CrossRef]

19. Passi-Solar, Á.; Margozzini, P.; Cortinez-O'Ryan, A.; Muñoz, J.C.; Mindell, J.S. Nutritional and metabolic benefits associated with active and public transport: Results from the Chilean National Health Survey, ENS 2016–2017. *J. Transp. Health* **2020**, *17*, 100819. [CrossRef]

20. Feng, X.; Astell-Burt, T. Can green space quantity and quality help prevent postpartum weight gain? A longitudinal study. *J. Epidemiol. Community Health* **2019**, *73*, 295–302. [CrossRef]

21. Andersen, L.; Riiser, A.; Rutter, H.; Goenka, S.; Nordengen, S.; Solbraa, A.K. Trends in cycling and cycle related injuries and a calculation of prevented morbidity and mortality. *J. Transp. Health* **2018**, *9*, 217–225. [CrossRef]

22. Ghimire, R.; Ferreira, S.; Green, G.T.; Poudyal, N.C.; Cordell, H.K.; Thapa, J.R. Green Space and Adult Obesity in the United States. *Ecol. Econ.* **2017**, *136*, 201–212. [CrossRef]

23. Luo, Y.; Yan, J.; McClure, S. Distribution of the environmental and socioeconomic risk factors on COVID-19 death rate across continental USA: A spatial nonlinear analysis. *Environ. Sci. Pollut. Res.* **2020**, *28*, 6587–6599. [CrossRef]

24. Frank, L.D.; Adhikari, B.; White, K.R.; Dummer, T.; Sandhu, J.; Demlow, E.; Hu, Y.; Hong, A.; Bosch, M.V.D. Chronic disease and where you live: Built and natural environment relationships with physical activity, obesity, and diabetes. *Environ. Int.* **2022**, *158*, 106959. [CrossRef]

25. Hu, M.D.; Lawrence, K.G.; Bodkin, M.R.; Kwok, R.K.; Engel, L.S.; Sandler, D.P. Neighborhood Deprivation, Obesity, and Diabetes in Residents of the US Gulf Coast. *Am. J. Epidemiol.* **2021**, *190*, 295–304. [CrossRef] [PubMed]

26. Ha, H.; Xu, Y. An ecological study on the spatially varying association between adult obesity rates and altitude in the United States: Using geographically weighted regression. *Int. J. Environ. Health Res.* **2022**, *32*, 1030–1042. [CrossRef] [PubMed]

27. Oshan, T.M.; Smith, J.; Fotheringham, A.S. Targeting the spatial context of obesity determinants via multiscale geographically weighted regression. *Int. J. Health Geogr.* **2020**, *19*, 11. [CrossRef] [PubMed]

28. Lee, K.H.; Heo, J.; Jayaraman, R.; Dawson, S. Proximity to parks and natural areas as an environmental determinant to spatial disparities in obesity prevalence. *Appl. Geogr.* **2019**, *112*, 102074. [CrossRef]

29. Wende, D. Spatial risk adjustment between health insurances: Using GWR in risk adjustment models to conserve incentives for service optimisation and reduce MAUP. *Eur. J. Health Econ.* **2019**, *20*, 1079–1091. [CrossRef]

30. Yu, Q.; Scribner, R.A.; Leonardi, C.; Zhang, L.; Park, C.; Chen, L.; Simonsen, N.R. Exploring racial disparity in obesity: A mediation analysis considering geo-coded environmental factors. *Spat. Spatio-Temporal Epidemiol.* **2017**, *21*, 13–23. [CrossRef]

31. Slack, T.; Myers, C.A.; Martin, C.K.; Heymsfield, S.B. The geographic concentration of us adult obesity prevalence and associated social, economic, and environmental factors. *Obesity* **2013**, *22*, 868–874. [CrossRef]

32. Ferdowsy, F.; Alam Rahi, K.S.; Jabiullah, M.I.; Habib, M.T. A machine learning approach for obesity risk prediction. *Curr. Res. Behav. Sci.* **2021**, *2*, 100053. [CrossRef]

33. Ghosh, D.; Guha, R. What are we 'tweeting' about obesity? Mapping tweets with topic modeling and Geographic Information System. *Cartogr. Geogr. Inf. Sci.* **2013**, *40*, 90–102. [CrossRef]

34. Gailey, S.; McElroy, S.; Benmarhnia, T.; Bruckner, T.A. Green mobility and obesity risk: A longitudinal analysis in California. *Health Place* **2021**, *68*, 102503. [CrossRef]

35. Jia, P.; Xue, H.; Yin, L.; Stein, A.; Wang, M.; Wang, Y. Spatial Technologies in Obesity Research: Current Applications and Future Promise. *Trends Endocrinol. Metab.* **2019**, *30*, 211–223. [CrossRef]

36. Jia, P.; Cheng, X.; Xue, H.; Wang, Y. Applications of geographic information systems (GIS) data and methods in obesity-related research. *Obes. Rev.* **2017**, *18*, 400–411. [CrossRef]

37. Mtambo, O.P.L.; Debusho, L.K. Analysis of childhood overweight and obesity in Namibia using spatio-temporal quantile interval models. *J. Health Popul. Nutr.* **2021**, *40*, 51. [CrossRef]

38. Chalkias, C.; Papadopoulos, A.G.; Kalogeropoulos, K.; Tambalis, K.; Psarra, G.; Sidossis, L. Geographical heterogeneity of the relationship between childhood obesity and socio-environmental status: Empirical evidence from Athens, Greece. *Appl. Geogr.* **2012**, *37*, 34–43. [CrossRef]

39. Sallis, J.F.; Floyd, M.F.; Rodríguez, D.A.; Saelens, B.E. Role of Built Environments in Physical Activity, Obesity, and Cardiovascular Disease. *Circulation* **2012**, *125*, 729–737. [CrossRef]

40. Xue, H.; Slivka, L.; Igusa, T.; Huang, T.T.; Wang, Y. Applications of systems modelling in obesity research. *Obes. Rev.* **2018**, *19*, 1293–1308. [CrossRef]

41. Lu, B.; Harris, P.; Charlton, M.; Brunsdon, C. The GWmodel R package: Further topics for exploring spatial heterogeneity using geographically weighted models. *Geo-Spat. Inf. Sci.* **2014**, *17*, 85–101. [CrossRef]

42. Quiñones, S.; Goyal, A.; Ahmed, Z.U. Geographically weighted machine learning model for untangling spatial heterogeneity of type 2 diabetes mellitus (T2D) prevalence in the USA. *Sci. Rep.* **2021**, *11*, 6955. [CrossRef]

43. El-Rahman, S.A.; Alluhaidan, A.S.; AlRashed, R.A.; AlZunaytan, D.N. Chronic diseases monitoring and diagnosis system based on features selection and machine learning predictive models. *Soft Comput.* **2022**, *26*, 6175–6199. [CrossRef]

44. Grekousis, G.; Feng, Z.; Marakakis, I.; Lu, Y.; Wang, R. Ranking the importance of demographic, socioeconomic, and underlying health factors on US COVID-19 deaths: A geographical random forest approach. *Health Place* **2022**, *74*, 102744. [CrossRef]

45. Thakur, D.; Biswas, S. An Integration of feature extraction and Guided Regularized Random Forest feature selection for Smartphone based Human Activity Recognition. *J. Netw. Comput. Appl.* **2022**, *204*, 103417. [CrossRef]

46. Fotheringham, A.S.; Crespo, R.; Yao, J. Geographical and Temporal Weighted Regression (GTWR). *Geogr. Anal.* **2015**, *47*, 431–452. [CrossRef]

47. Georganos, S.; Kalogirou, S. A Forest of Forests: A Spatially Weighted and Computationally Efficient Formulation of Geographical Random Forests. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 471. [CrossRef]

48. U.S. Center for Disease Control and Prevention (CDC). Overweight and Obesity. 2021. Available online: https://www.cdc.gov/obesity/data/index.html (accessed on 10 December 2021).

49. U.S. American Community Survey (ACS). TIGER/Line with Selected Demographic and Economic Data. 2019. Available online: https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-data.html (accessed on 9 March 2022).

50. Selçuk, I.; Köktaş, A.M.; Toygar, A. Socioeconomic factors affecting the probability of obesity: Evidence from a nationwide survey in Turkey. *Qual. Quant.* **2022**, 1–17. [CrossRef]

51. Hughes, A.; Kumari, M. Unemployment, underweight, and obesity: Findings from Understanding Society (UKHLS). *Prev. Med.* **2017**, *97*, 19–25. [CrossRef]

52. Chicago Health Atlas (CHA). Available online: https://chicagohealthatlas.org/ (accessed on 10 December 2021).

53. Google Earth Engine, Landsat 8 Collection 1 Tier 1 Annual NDVI Composite. Available online: https://explorer.earthengine.google.com/#detail/LANDSAT%2FLC08%2FC01%2FT1_ANNUAL_NDV (accessed on 20 December 2021).

54. Chander, G.; Markham, B.L.; Helder, D.L. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sens. Environ.* **2009**, *113*, 893–903. [CrossRef]

55. Divvy System Data. Available online: https://ride.divvybikes.com/system-data (accessed on 15 December 2021).

56. U.S. Chicago Police Department. Available online: https://www.chicago.gov/city/en/depts/cpd.html (accessed on 15 January 2022).

57. Fernández-Delgado, M.; Cernadas, E.; Barro, S.; Amorim, D. Do we need hundreds of classifiers to solve real world classification problems? *J. Mach. Learn. Res.* **2014**, *15*, 3133–3181.

58. Breiman, L. Random forests Mach. *Learn* **2001**, *45*, 5–32.

59. Georganos, S.; Grippa, T.; Gadiaga, A.N.; Linard, C.; Lennert, M.; VanHuysse, S.; Mboga, N.; Wolff, E.; Kalogirou, S. Geographical random forests: A spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. *Geocarto Int.* **2021**, *36*, 121–136. [CrossRef]

60. Anselin, L. The Local Indicators of Spatial Association—LISA. *Geogr. Anal.* **1995**, *27*, 93–115. [CrossRef]

61. Fotheringham, A.S.; Brunsdon, C.; Charlton, M. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*; Wiley: Hoboken, NJ, USA, 2002.

62. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]

63. Kalogirou, S.; Georganos, S. Spatial Machine Learning. R Package. Available online: https://cran.r-project.org/web/packages/SpatialML/SpatialML.pdf (accessed on 10 January 2022).

64. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020. Available online: https://www.R-project.org/ (accessed on 1 June 2021).

65. Stanley, K.C.; Harrigan, P.B.; Serrano, E.L.; Kraak, V.I. A systematic scoping review of the literacy literature to develop a digital food and nutrition literacy model for low-income adults to make healthy choices in the online food retail ecosystem to reduce obesity risk. *Obes. Rev.* **2022**, *23*, e13414. [CrossRef]

66. Kwarteng, J.L.; Schulz, A.J.; Mentz, G.B.; Israel, B.A.; Perkins, D.W. Independent Effects of Neighborhood Poverty and Psychosocial Stress on Obesity Over Time. *J. Urban Health Bull. N. Y. Acad. Med.* **2017**, *94*, 791–802. [CrossRef]

67. Duraccio, K.M.; Krietsch, K.N.; Chardon, M.L.; Van Dyk, T.R.; Beebe, D.W. Poor sleep and adolescent obesity risk: A narrative review of potential mechanisms. *Adolesc. Health Med. Ther.* **2019**, *10*, 117–130. [CrossRef]

68.    Armstrong, S.; Wong, C.A.; Perrin, E.; Page, S.; Sibley, L.; Skinner, A. Association of Physical Activity with Income, Race/Ethnicity, and Sex Among Adolescents and Young Adults in the United States: Findings from the National Health and Nutrition Examination Survey, 2007–2016. *JAMA Pediatr.* **2018**, *172*, 732–740. [CrossRef]

69.    Huang, Y.; Sparks, P.J. Longitudinal exposure to neighborhood poverty and obesity risk in emerging adulthood. *Soc. Sci. Res.* **2022**, 102796. [CrossRef]

70.    Feuillet, T.; Valette, J.; Charreire, H.; Kesse-Guyot, E.; Julia, C.; Vernez-Moudon, A.; Hercberg, S.; Touvier, M.; Oppert, J. Influence of the urban context on the relationship between neighbourhood deprivation and obesity. *Soc. Sci. Med.* **2020**, *265*, 113537. [CrossRef]

71.    Lacoe, J.; Bostic, R.W.; Acolin, A. Crime and private investment in urban neighborhoods. *J. Urban Econ.* **2018**, *108*, 154–169. [CrossRef]

72.    Park, Y.; Garcia, M. Pedestrian safety perception and urban street settings. *Int. J. Sustain. Transp.* **2020**, *14*, 860–871. [CrossRef]

73.    Ruhm, C.J. Healthy living in hard times. *J. Health Econ.* **2005**, *24*, 341–363. [CrossRef]

74.    Cutler, D.M.; Huang, W.; Lleras-Muney, A. When does education matter? The protective effect of education for cohorts graduating in bad times. *Soc. Sci. Med.* **2015**, *127*, 63–73. [CrossRef] [PubMed]

75.    Deb, P.; Gallo, W.T.; Ayyagari, P.; Fletcher, J.M.; Sindelar, J.L. The effect of job loss on overweight and drinking. *J. Health Econ.* **2011**, *30*, 317–327. [CrossRef] [PubMed]

76.    Jin, R.L.; Shah, C.P.; Svoboda, T.J. The impact of unemployment on health: A review of the evidence. *J. Public Health Policy* **1997**, *18*, 275–301, reprinted in *Can. Med. Assoc. J.* **1995**, *153*, 529–540. [CrossRef]

77.    Hazekamp, C.; Yousuf, S.; Khare, M.; MacDowell, M. Unhealthy behaviours in urban Illinois communities affected by eviction: A descriptive analysis. *Health Soc. Care Community* **2021**, *29*, 867–875. [CrossRef]

78.    Preston, G.; Reina, V.J. Sheltered from Eviction? A Framework for Understanding the Relationship between Subsidized Housing Programs and Eviction. *Hous. Policy Debate* **2021**, *31*, 785–817. [CrossRef]

79.    Lotfata, A.; Tomal, M. Exploring Housing Determinants of Obesity Prevalence Using Multiscale Geographically Weighted Regression in Chicago, Illinois. *Prof. Geogr.* **2022**, 1–10. [CrossRef]

80.    Chambers, E.C.; Fuster, D. Housing as an obesity-mediating environment. *Int. J. Public Health* **2012**, *57*, 453–454. [CrossRef]

81.    Chicago Eviction. 2022. Available online: https://eviction.lcbh.org/reports/eviction-unemployment-covid-19 (accessed on 15 April 2022).

82.    Lachowycz, K.; Jones, A.P. Greenspace and obesity: A systematic review of the evidence. *Obes. Rev. Off. J. Int. Assoc. Study Obes.* **2011**, *12*, e183–e189. [CrossRef]

83.    Hohl, A.; Lotfata, A. Modeling spatiotemporal associations of obesity prevalence with biking, housing cost and green spaces in Chicago, IL, USA, 2015–2017. *J. Transp. Health* **2022**, *26*, 101412. [CrossRef]

84.    He, Q.; Li, J. The roles of built environment and social disadvantage on the geography of property crime. *Cities* **2021**, *121*, 103471. [CrossRef]

85.    Huang, H. Food Environment Inequalities and Moderating Effects of Obesity on Their Relationships with COVID-19 in Chicago. *Sustainability* **2022**, *14*, 6498. [CrossRef]

86.    Boutari, C.; Mantzoros, C.S. A 2022 update on the epidemiology of obesity and a call to action: As its twin COVID-19 pandemic appears to be receding, the obesity and dysmetabolism pandemic continues to rage on. *Metabolism* **2022**, *133*, 155217. [CrossRef] [PubMed]

87.    Cofie, L.E.; Cuevas, A.G. Length of Residency in the United States and Obesity Across Race/Ethnicity. *J. Immigr. Minor. Health* **2022**, 1–5. [CrossRef] [PubMed]

88.    Sanjeevi, N.; Freeland-Graves, J.H. Low diet quality is associated with adverse levels of metabolic health markers and clustering of risk factors in adults with type 2 diabetes. *J. Hum. Nutr. Diet.* **2022**. [CrossRef]

89.    Helbich, M.; Browning, M.H.M.; Kwan, M.-P. Time to address the spatiotemporal uncertainties in COVID-19 research: Concerns and challenges. *Sci. Total Environ.* **2021**, *764*, 142866. [CrossRef]