



A Spatial Machine Learning Approach for Analyzing Population-Level Factors Associated with Emergency Department Visits with Diabetes Diagnoses

Haley Mullen · Kathleen Stewart ·
Stephen B. Thomas

Received: 11 February 2026 / Accepted: 3 August 2026
© The Author(s) 2026

Abstract Emergency department (ED) visits for diabetes arise from complications related to both acute and chronic symptoms including, but not limited to, cardiovascular diseases, renal failure, neurological issues, and retinopathy. Research has shown that such complications disproportionately affect racial and ethnic minority populations. This is likely due to the complex pathogenesis of diabetes, which includes both individual-level factors such as race and family history and population-level drivers such as neighborhood walkability and access to healthy food outlets, which may be heavily influenced by residential segregation. We use spatial machine learning and statistical methods to estimate population-level factors associated with ED visits with diabetes diagnoses at the ZCTA level in Maryland as stratified by race and local measures of residential segregation. Using XGBoost and SHAP, we identified population-level

factors associated with diabetes-coded ED visits and used logistic regression to estimate odds ratios. Our findings show that ED visits were associated with factors including housing vacancy, heavy traffic, greenspace, and health behaviors like physical leisure time and smoking across all subgroups. Further, we found that the factors associated with diabetes-coded ED visits differ across racial groups and across communities with different patterns of residential segregation. Heavy traffic, a proxy for traffic-related air pollution and barriers to walkability, was consistently linked to higher odds of diabetes-coded ED visits, especially among Black patients.

Keywords Diabetes · Segregation · Chronic disease · Explainable machine learning · Health disparities · Environmental health

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11524-026-01152-5>.

H. Mullen (✉) · K. Stewart
Center for Geospatial Information Science, Department of Geographical Sciences, University of Maryland, College Park, MD 20742, USA
e-mail: mullenh@umd.edu

S. B. Thomas
Center for Health Equity, Department of Health Policy and Management, University of Maryland, College Park, MD 20742, USA

Introduction

According to the U.S. Center for Disease Control's National Diabetes Statistics Report, 38.1 million adults aged 18 years or older, or 14.7% of all US adults, had diabetes as of 2021. The healthcare costs associated with diabetes are significant; according to the American Diabetes Association, one in four healthcare dollars spent in the USA is attributable to patients with diagnosed diabetes. These costs encompass inpatient care, emergency department visits, outpatient care, insulin and related medical supplies,

and indirect costs associated with productivity loss, absenteeism, and premature mortality [1, 2].

Diabetes is associated with significant morbidity and mortality. Diabetes-related complications include cardiovascular disease, kidney disease, neuropathy, blindness, and lower extremity amputation, among other outcomes [3]. Each of these complications poses acute risks and, when combined with poor diabetes management, can result in emergency department (ED) visits. ED utilization among patients with diabetes is driven by social and medical factors that span both individual behaviors and neighborhood drivers, including urbanicity, lack of access to primary care, poor adherence to care, and complications [4–6].

Notably, diabetes burden is not evenly distributed across racial groups [7–10]. Specifically, Black-White disparities in diabetes have been reported in many studies, which have shown that compared to White patients, Black patients are between 1.4 and 2.2 times more likely to be diagnosed with diabetes, twice as likely to die from diabetes, and 2–4 times more likely to experience complications, including amputation, eye diseases, and renal diseases—a disparity that has remained largely unchanged for 20 years [7, 9].

Research has sought to explain these differences, and has primarily focused on socioeconomic pathways, including those relating to healthcare utilization and educational attainment [11, 12]. However, recent work has increasingly examined residential segregation as a structural driver of disparities in diabetes outcomes [7, 13, 14]. In urban areas, contemporary patterns of residential segregation reflect structural inequities embedded in local housing, land use, and transportation [15]. Following epidemiology’s Fundamental Cause Theory, segregation is a fundamental cause of racial disparities in diabetes because it shapes racial inequities in economic mobility and exposure to built environment features [13, 16, 17]. Residential segregation reflects longstanding structural racism in housing and lending, concentrating racially minoritized groups in neighborhoods with fewer health-promoting resources and greater exposure to stressors [17].

Existing work on the spatial relationships between diabetes and segregation is limited and shows mixed results. Some studies report little or no association between segregation and diabetes incidence, while others find positive associations with

outcomes such as mortality [7, 14]. These inconsistencies reflect differences in outcome definitions, segregation measures, and spatial scales, underscoring the need for research that explicitly examines spatial heterogeneity in these relationships. To address this, this study proposes three objectives: (1) to identify social and environmental factors that are most strongly associated with ED encounters containing a diabetes-specific diagnosis code, (2) to assess whether these associations differ across racial groups and across neighborhoods with different degrees of segregation, and (3) to describe the spatial distribution and effect sizes of key population-level factors that contribute to ED encounters with diabetes codes in Maryland ZCTAs to inform local public health interventions.

Materials and Methods

Study Design

This study adopts a cross-sectional analysis using encounter-level data from the Healthcare Cost and Utilization Project’s (HCUP) State Emergency Department Databases for Maryland for all encounters between 2015 and 2019. These encounters were linked to ZCTA-level measures of segregation, built environment, socioeconomics, and health-related behaviors derived from publicly available sources.

This study first uses the XGBoost (eXtreme Gradient Boosting) algorithm and SHAP (Shapley Additive Explanations) to understand the most influential drivers of ED visits, then leverages generalized estimating equations (GEE) and odds ratios (ORs) to isolate the effect size of modifiable drivers while accounting for ZCTA-level clustering. XGBoost is an explainable machine learning approach that was applied here to identify potential drivers of diabetes-coded ED visits and extract spatial effects. XGBoost is a tree-based method that builds decision trees and combines them to improve prediction accuracy [18]. Unlike traditional regression, XGBoost can flexibly model non-linear relationships between predictors and ED encounters without requiring strong assumptions about linearity or additivity.

In order to identify factors associated with ED visits with a diabetes diagnosis, seven separate XGBoost models were implemented: one set of models was

stratified by racial group (Black, White, Latino, and all patient races), and another set was stratified by segregation (high Black segregation, low segregation overall, and high White segregation). The latter three models included ZCTAs divided into tertiles based on ICE (Index of Concentration at Extremes) values, with XGB 5 including ZCTAs with ICE values between -1 and 0.07 (a high concentration of Black ED visits), XGB 6 included ZCTAs with ICE values between 0.08 and 0.5 (low-to-moderate concentration of either groups' ED visits), and XGB 7 included ZCTAs with ICE values between 0.6 and 1 (high concentration of White ED visits). In addition to conventional residential segregation indices based on area level racial composition, this complementary measure reflects the concentration of ED encounters by race to capture how the burden of ED utilization is spatially distributed across racial groups. Each model included the same set of built environment, demographic, segregation, and spatial covariates (full list in Supplementary Materials). Scaled, population-adjusted rates of diabetes-coded ED visits were the outcome variable in each model (Supplementary Materials).

To interpret the fitted models, we used SHAP. SHAP assigns each predictor a contribution to the model's prediction for each observation, which can be interpreted as feature importance. SHAP values indicate how much each predictor increases or decreases the estimated rate of encounters. SHAP values were summarized across all ZCTAs to identify predictors with consistently large contributions to estimated encounter rates.

The results from SHAP were then used to guide the statistical modeling with GEE. Variables were eligible for inclusion in the regressions if they (1) had high importance in the XGBoost models via SHAP and (2) represented modifiable factors that could be targeted by public health interventions. Our GEE specification treats each ZCTA as a cluster and assumes within-cluster correlation to account for spatial dependencies among ED encounters from patients in the same ZCTAs. The resulting models yielded inferences about the overall patient population while accounting for ZCTA-level clustering. From the GEE, odds ratios (ORs) were generated to quantify how the odds of visiting an ED for diabetes changes as predictor values increase, holding all other variables constant [19].

Covariates

This study uses covariates relating to racial residential segregation, built environment, socioeconomic status, and comorbidities with diabetes. All data collected were ZCTA-level measures from publicly available sources, including the U.S. Census Bureau, EPA, NaNDa, and CDC Places. For the XGBoost models, covariates were used in their original units given the algorithm's flexibility with different scales, whereas for the GEE we re-scaled percent-based and continuous predictors (e.g., per 10 percentage-point or per 5-unit change) to yield interpretable odds ratios that were more reflective of substantive population-level changes in modifiable risk factors.

Diabetes-Coded Emergency Department Visits

To examine diabetes-coded ED visits, this study uses encounter-level data from HCUP. In order to capture encounters specifically related to diabetes, data includes only those encounters where any of the first ten diagnosis codes corresponded to one of five diabetes-related ICD-10 codes: E08 (Diabetes mellitus due to underlying condition), E09 (Drug or chemical induced diabetes mellitus), E10 (Type 1 diabetes mellitus), E11 (Type 2 diabetes mellitus), or E13 (Other specified diabetes mellitus). This definition excludes encounters where diagnosis codes were related to some diabetes-related complications (e.g., foot ulcers or hyperglycemia) but did not include any of the ICD-10 codes that explicitly indicate diabetes mellitus.

A summary of exclusion criteria applied to the ED data for use in the machine learning models is provided in the Supplementary Materials. In order to examine the associations between segregation, the built environment, and diabetes-coded visits, encounters were aggregated and adjusted for population. A scaling factor of 1000 was applied to estimate the prevalence of diabetes-related ED visits at a broader population level. The result is a population-adjusted rate of ED visits with diabetes-specific diagnostic codes per 1000 residents.

Our generalized estimating equations used the ED data as a binary outcome variable where the ED encounters that included a diabetes-specific ICD-10 code were coded as "1" and all other encounters were coded as "0." Here, diabetes-related encounters were defined using the same ICD-10 codes in any of

the first 10 diagnosis fields. After filtering for missing data and exclusion of any patients younger than 18 years of age, the dataset yielded 7,469,808 observations for the statistical analysis. Individual encounter data was then linked with ZCTA-level predictors using patient ZIP codes.

Results

Baseline Spatial Distribution of Diabetes-Coded ED Visits and Segregation

We mapped each racial groups' distribution of diabetes-coded ED visits per 1000 population to understand spatial patterns of ED visits by race (Fig. 1). The darkest polygons indicate ZCTAs where there are disproportionately high diabetes-related ED visits relative to baseline population levels. In terms of overall visits for all races, there were the highest rates of ED visits relative to the baseline population in the western and southern tips of Maryland and parts of the Eastern Shore (Fig. 1a). The only population group with disproportionately high rates of diabetes-related ED visits (i.e., > 1000) relative to their population

was Black patients, which can be shown in the dark green polygons (Fig. 1b).

Estimating the Effect of Population-Level Factors on Diabetes-Coded ED Visits by Racial Group and Degree of Segregation

The first model was designed to estimate the effect of segregation and the built environment on diabetes for all racial groups. This model yielded an R^2 of 0.51, suggesting a moderate association between covariates and the outcome. The second model tested the association for White patients specifically and yielded a slightly lower R^2 of 0.39. The third model, which tested the association for Black patients, yielded the highest R^2 of 0.63. The fourth model tested the association for Latino patients and yielded an R^2 of 0.41. These results suggest that out of all models, the strongest association was observed for Black patients.

SHAP was applied to learn the importance of each feature in contributing to these model predictions. The top 20 most important features per SHAP for each of the race-stratified models were examined for all races (Fig. 2a), White patients (Fig. 2b), Black patients (Fig. 2c), and Latino patients (Fig. 2d).

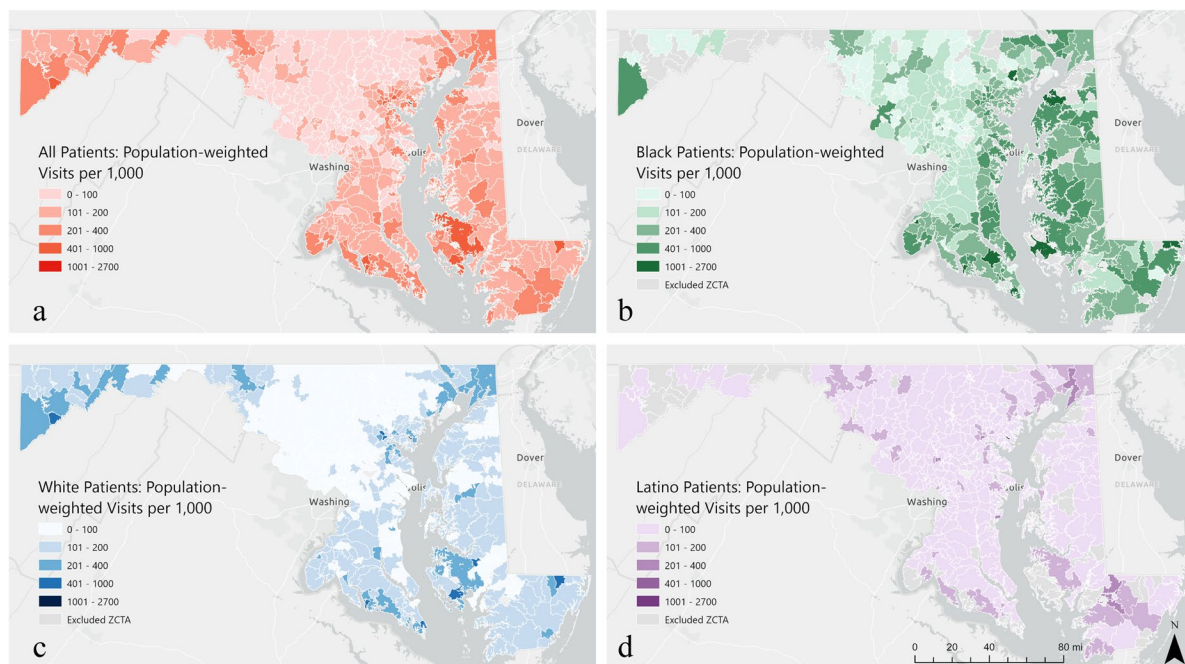


Fig. 1 Baseline distribution of diabetes-coded ED visits by race for Maryland hospitals

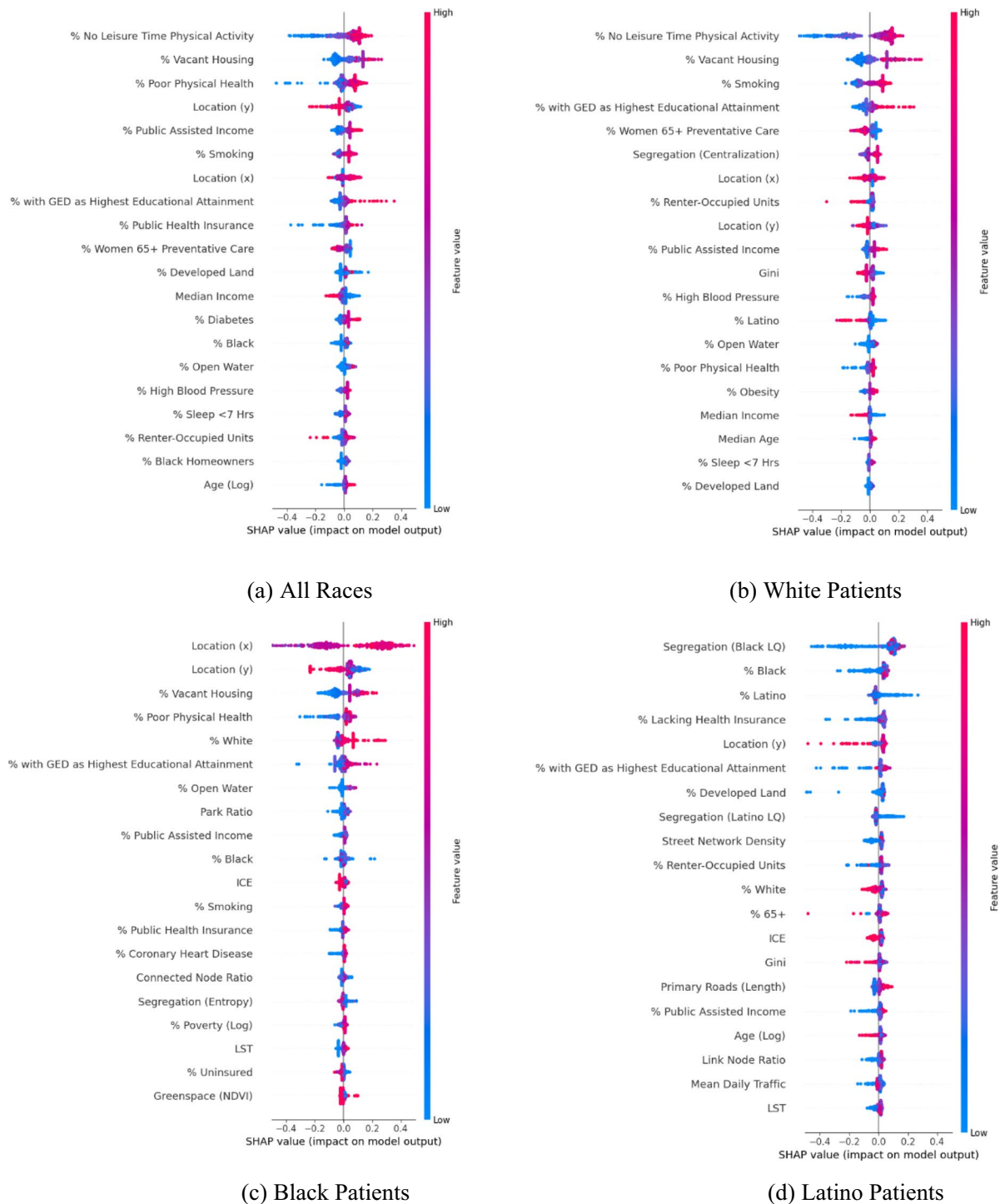


Fig. 2 SHAP summary plots for race-stratified XGBoost models. Each point represents an individual observation (i.e., ZCTA). For each predictor, the distribution of points indicates how that predictor's SHAP values are distributed across the dataset

Each point shows how a given predictor's SHAP value (i.e., feature importance) changes as the value of that predictor increases, with points on the right indicating positive contributions to higher ED visit rates and points on the left indicating negative contributions. Vacant housing units were ranked as a top factor across all models except for the model for Latino patients (Fig. 2). In urban health research, vacant housing units are frequently used as a proxy for adverse neighborhood outcomes and/or neighborhood disorder, representing residential environments that have been affected by significant economic and/or social decline [20, 21]. The distribution of this variable in the models for Black, White, and all patients suggests that ZCTAs with higher rates of vacant housing were associated with higher rates of ED visits, and ZCTAs with less vacant housing were associated with fewer ED visits. Across all models, location and age also exhibited strong associations with ED visits. Poverty, as measured by percent at or below the poverty line and percent receiving publicly assisted income/SNAP, and land use (open water and developed land) were similarly important across each model.

Differences in the relative importance of these factors across racial groups were also observed. Built environment indicators including open water, park ratio, developed land, and street network density were more important in the Black and Latino models compared to the White model, which was generally more influenced by socioeconomic factors. Differences were also noted in the relative importance of health insurance; the Latino model (Fig. 2d) yielded population lacking health insurance as the third most important factor via SHAP, while the Black model (Fig. 2c) yielded population using public health insurance (i.e., Medicare or Medicaid) as the thirteenth most important factor, neither of which were in the top 20 SHAP features for the White model (Fig. 2b).

SHAP was further applied to the models stratified by segregation (Fig. 3). The top 20 most important features per SHAP for each of the segregation-stratified models were examined for ZCTAs with high levels of Black segregation (Fig. 3a), low levels of both White and Black segregation (Fig. 3b), and high levels of White segregation (Fig. 3c).

Consistent with the race-stratified models, the SHAP analysis for the segregation-stratified models also yielded higher values for vacant housing, but

especially in the two models for highly segregated ZCTAs, for which it was the second most important factor (Fig. 3a and c). Notably, the prevalence of adults reporting no physical leisure time activity was an important predictor in the low segregation (Fig. 3b) and high White segregation (Fig. 3c) models, but not for the high Black segregation model, suggesting that it was more influential in determining diabetes-coded visits in ZCTAs with little segregation or higher concentrations of White populations relative to Black populations. Similarly, although prevalence of adults lacking health insurance was in the top 20 SHAP features across all strata, it was the most influential in the high Black segregation model (Fig. 3a), indicating a stronger association in ZCTAs with high levels of Black segregation.

The distribution of SHAP values for NDVI, or greenspace, shows bright blue polygons that represent those ZCTAs where low greenspace was associated with higher rates of diabetes-coded visits, which were concentrated in the southern part of the Eastern Shore and Western Maryland (Fig. 4b). The bright red polygons indicate ZCTAs where higher prevalence of heavy traffic was associated with higher rates of diabetes-coded visits, which were somewhat dispersed throughout the state but appeared in parts of Prince George's County, the Eastern Shore, and sections along the state's Western border (Fig. 4c). The SHAP values for LST (land surface temperature), where the bright red polygons represent ZCTAs in which higher LST was associated with higher ED visits, demonstrating the strongest associations between LST and ED visits that appeared in both rural and urban parts of Maryland, including in the northeastern and western parts of the state along with parts of Baltimore City and its surrounding neighborhoods (Fig. 4d).

Estimating the Effect of Population-Level Factors on Diabetes-Coded Emergency Department Visits

Based on the SHAP analysis across all models, we selected a subset of predictors to quantify the direction and magnitude of association with diabetes-coded ED visits as measured by GEE and resulting odds ratios (ORs). To focus on factors that were both identified as highly influential via SHAP and were modifiable via public health interventions, we selected vacant housing, heavy traffic, intersection density, prevalence of adult population reporting no

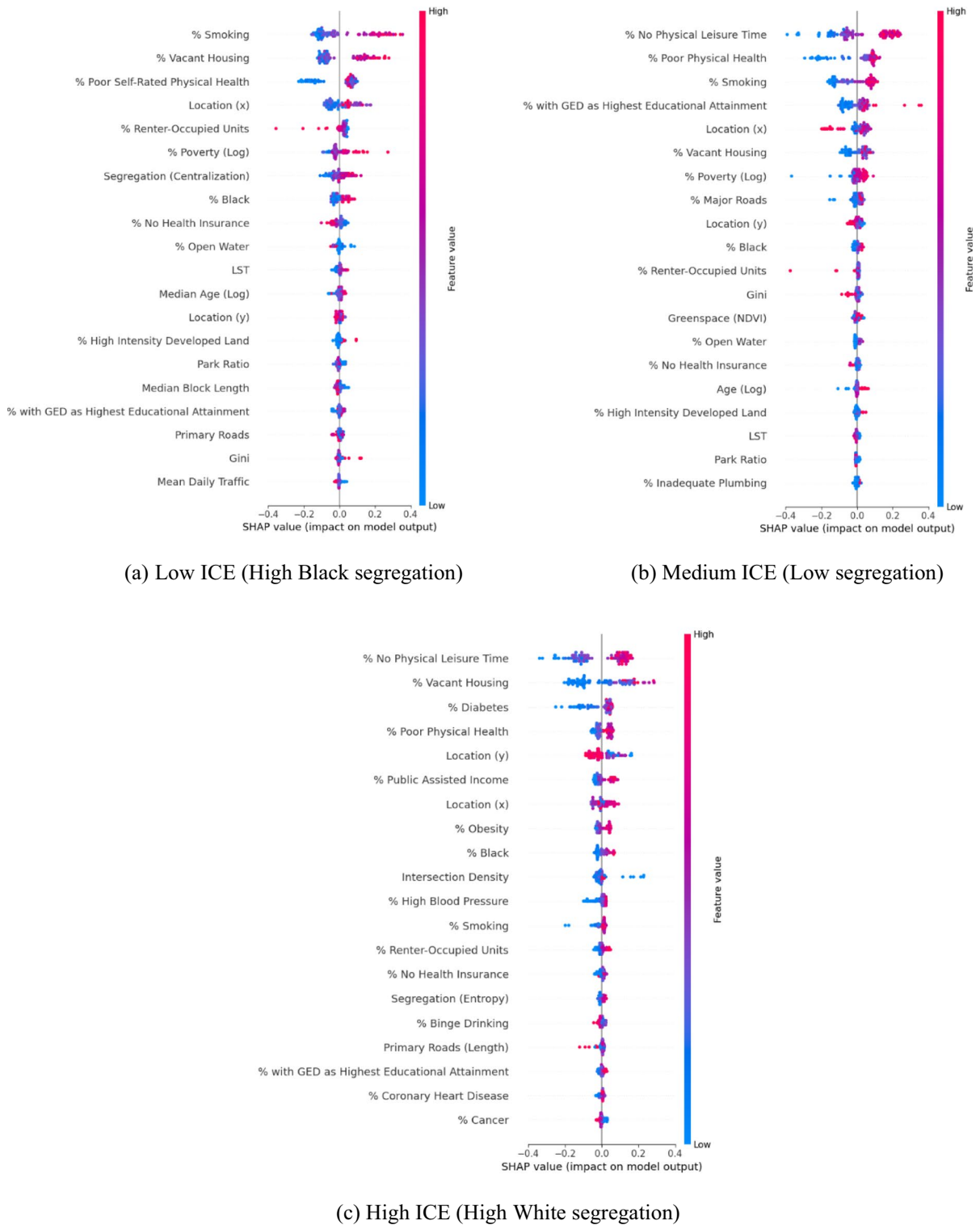


Fig. 3 SHAP summary plots for segregation-stratified models

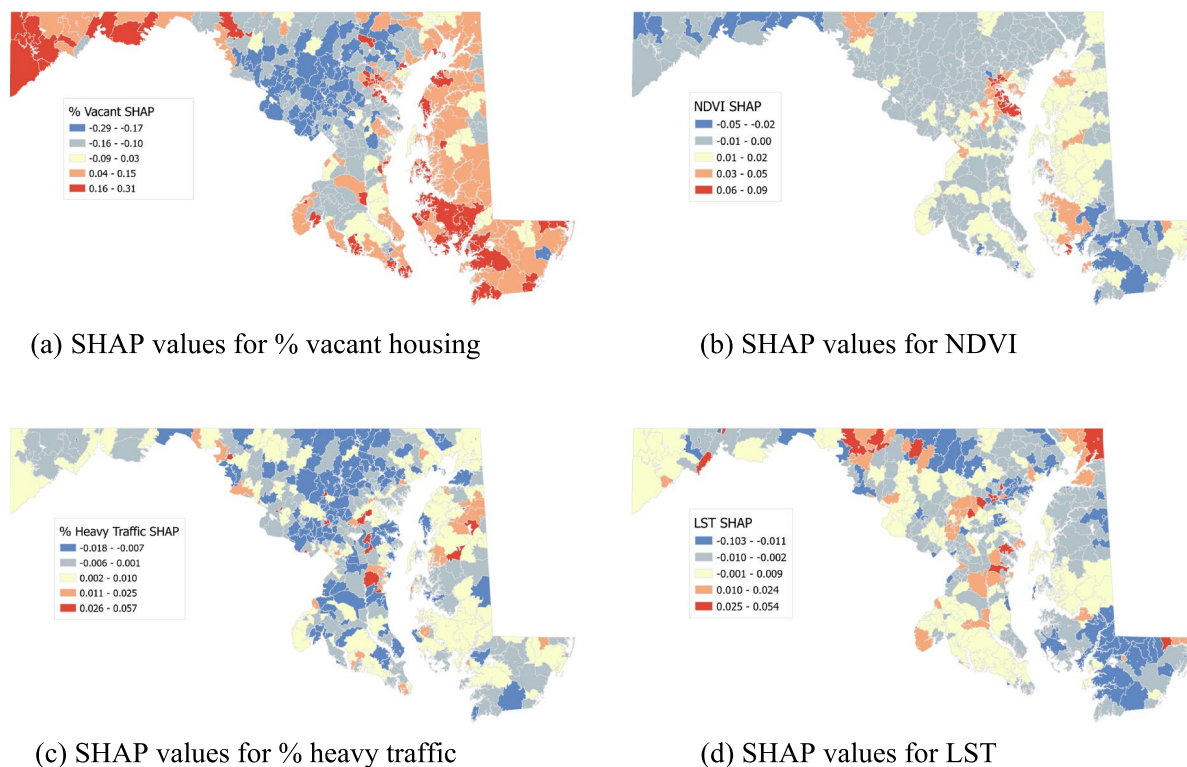


Fig. 4 Distribution of SHAP values for key predictor variables, including **a** vacant housing, **b** NDVI (greenspace), **c** heavy traffic, and **d** LST (land surface temperature), based on the XGB model for all races

physical leisure time activity, smoking, NDVI, and LST (Table 1).

Overall, heavy traffic, lacking physical leisure time, smoking, and LST all had statistically significant associations with diabetes-coded ED visits ($p < 0.05$). However, the strongest predictor was heavy traffic (OR = 1.78, CI = 1.23–2.59), which was found to nearly double the odds of visiting the ED for diabetes. Heavy traffic is likely indicative of several neighborhood characteristics and exposures that influence diabetes risk, including traffic-related air pollution and barriers to walkability and physical activity.

Generalized estimating equations and odds ratios (OR) were also constructed using the same stratifications employed in the machine learning approach. Three models were constructed stratified by racial group (i.e., Black, White, and Hispanic/Latino), and another three were stratified by level of segregation in patients' home ZCTAs (i.e. low, medium, and high ICE values) (Table 1).

The results of the stratified odds ratios revealed associations with diabetes-coded visits across racial

subgroups and levels of residential segregation. The strongest and most consistent predictor of ED visits across all subgroups was heavy traffic, for which the odds ranged from ~1.42 to ~4.31; all but one subgroup (White patients) were significant ($p < 0.05$). The odds were the highest for Black patients, for which a one-unit increase in heavy traffic was associated with over 4.3 times higher odds of visiting an ED for diabetes. Behavioral factors, including smoking and prevalence of adults who reported lacking physical leisure time, also showed consistent associations across subgroups; lacking physical leisure time was associated with increased odds of ED visits across all strata ($p < 0.05$), and smoking was associated with slightly decreased odds.

Discussion

This study examined associations between population-level predictors and ED visits with diabetes-specific diagnostic codes stratified by race and

Table 1 Overall and stratified odds ratios (ORs) and 95% confidence intervals are presented for each predictor. * denotes $p < 0.05$, ** denotes $p < 0.01$, and *** denotes $p < 0.001$

Predictor	Overall	Black	White	Hispanic	Low ICE	Medium ICE	High ICE
Vacant housing	1.01 (0.99–1.03)	1.03 (0.99–1.08)	0.99 (0.96–1.02)	1.01 (0.96–1.07)	1.03 (0.98–1.08)	0.96 (0.91–1.01)	1.01 (0.97–1.05)
Heavy traffic	1.78 (1.23–2.59)**	4.31 (2.20–8.46)*	1.42 (0.94–2.15)	3.94 (1.82–8.53)*	2.45 (1.28–4.70)**	2.74 (1.46–5.15)**	2.54 (1.05–6.13)*
Lacking leisure time	1.52 (1.38–1.69)*	1.24 (1.09–1.40)*	1.64 (1.44–1.87)*	1.15 (1.00–1.31)*	1.36 (1.20–1.55)*	1.58 (1.26–1.97)*	1.43 (1.12–1.83)**
Smoking	0.80 (0.72–0.89)*	0.86 (0.75–0.99)*	0.78 (0.68–0.89)*	0.94 (0.80–1.11)	0.76 (0.65–0.89)*	0.78 (0.63–0.97)*	0.86 (0.67–1.10)
Intersection density	1.00 (0.998–1.00)	0.998 (0.995–1.00)	0.998 (0.995–1.00)	1.00 (0.998–1.01)	1.00 (0.998–1.00)	1.00 (0.996–1.01)	0.999 (0.993–1.01)
LST	0.97 (0.95–0.98)*	0.99 (0.95–1.04)	0.97 (0.95–0.99)**	1.01 (0.96–1.06)	1.00 (0.95–1.05)	0.96 (0.92–1.00)	0.96 (0.90–1.03)
NDVI	1.18 (0.92–1.50)	1.27 (0.78–2.06)	0.90 (0.68–1.19)	1.26 (0.69–2.30)	1.12 (0.66–1.92)	0.79 (0.41–1.53)	1.51 (0.63–3.61)

Several predictors were re-scaled into substantively meaningful units: Percentage-based indicators, including lacking leisure time, smoking, and vacant housing, are modeled per 10 percentage-point increase in the corresponding ZCTA-level proportion. Heavy traffic is modeled per 0.1 increase in the ratio of primary and secondary roads to all roads in a ZCTA. Intersection density is modeled per 5 unit increase in intersections per ZCTA area. Land surface temperature (LST) is modeled per 5 °F increase, and NDVI is modeled per 0.1 unit increase in average NDVI per ZCTA

segregation. We apply an explainable spatial machine learning framework to identify population-level drivers of diabetes-coded ED encounters. Substantively, we show that population-level social and environmental factors are associated with diabetes-coded ED utilization and differ across racial groups and levels of residential segregation. Diabetes-related ED visits have so far been understudied with respect to segregation and environmental health disparities, and even fewer studies have leveraged spatial machine learning to capture spatially heterogeneous relationships [22]. Diabetes has been previously linked to demographic and behavioral risk factors, but far fewer studies have explored the spatial, environmental, and social drivers, particularly as they intersect with spatial patterns of structural drivers like segregation [8, 13, 23].

Prior work on segregation and diabetes has also yielded mixed results. In a study by Mayne et al., authors observed no significant association between segregation and diabetes incidence/management for participants in four major US cities [14]. Conversely, Rosenstock et al. reported a positive association between segregation and age-adjusted diabetes mortality for patients in the 50 most populous US cities [7]. Complicating interpretation of these findings is significant heterogeneity in study design and metrics used. Studies vary significantly both in how

segregation is operationalized (formal measures vs. racial composition) and diabetes outcome reporting (prevalence vs. mortality vs. ED visits vs. inpatient admissions, etc.). Existing work also fails to address the spatial dynamics of these relationships, leading to a lack of spatial knowledge about racial health disparities in diabetes and segregation, which could provide valuable insights into local variations of these relationships. Our findings show that associations between segregation, neighborhood conditions, and ED encounters in which diabetes is coded vary across racial groups and levels of segregation, suggesting that spatial context and methodological choices are central to understanding why prior studies have reached different conclusions.

Several of the population-level factors examined in our study, particularly vacant housing and heavy traffic prevalence, have previously been linked to negative metabolic outcomes and health behaviors in prior studies [20, 24, 25]. For example, Zhu et al. investigated the association between vacant housing and non-communicable diseases, including diabetes, and found that short-term vacant housing had a negative association with diabetes, while longer term vacant housing had a positive association [20]. However, vacant housing had a consistent positive association with ED visits per the SHAP analysis

in our study, which may be attributable to the difference in dependent variables (ED visits vs. prevalence). Our finding that lacking physical leisure time was significantly associated with ED visits was also consistent with prior literature that has documented a dose-response relationship between physical activity and diabetes risk [26]. Other research has also confirmed associations between neighborhood context and exercise and smoking, wherein residents of more disadvantaged neighborhoods may be more likely to smoke and less likely to engage in physical leisure [27, 28].

Consistent with emerging evidence, we also observed a robust association between heavy traffic and diabetes-coded ED visits, which was strongest among Black patients (OR = 4.31, CI = 2.20–8.46). Prior studies have identified proximity to heavy traffic and major roads as a significant environmental health risk for its role in mediating exposures to air pollution, noise, and reduced walkability, all of which are implicated in diabetes morbidity [24, 25, 29]. These exposures may be especially pervasive in racially segregated neighborhoods that can exacerbate disparities in disease risk and healthcare utilization [8, 23].

Despite these important contributions, several limitations should be acknowledged. The spatial and temporal resolutions may have obscured finer neighborhood scale patterns. ZCTAs may have included heterogeneous neighborhoods, potentially masking important local variations in the distribution of environmental characteristics and residential segregation. This study also used aggregated data from 2015 to 2019, which possibly obscured certain temporal changes within this time frame. Further, while our study highlights important associations between population-level factors and ED utilization for diabetes, it did not explicitly measure causal relationships. In addition, our outcome definition relies on ED encounters in which a diabetes-specific ICD-10 code appears in one of the first ten diagnosis fields; although sensitivity analyses indicated that this captures the vast majority of diabetes-coded encounters, it may underestimate the full set of encounters related to diabetes and its complications. As a result, our findings should be interpreted as characterizing spatial and population-level patterns among ED encounters with documented diabetes codes, rather than all possible diabetes-related encounters.

Conclusions

This paper investigated the spatial patterns of population-level drivers of diabetes-coded ED visits and examined disparities across racial subgroups and areas exhibiting different levels of segregation. This work highlights the importance of spatially explicit analytical techniques for understanding racialized health disparities, especially as relates to targeted public health interventions and their applicability across different population subgroups. We observed strong associations between ZCTAs with heavy traffic and diabetes-coded ED visits, particularly in highly segregated and racially marginalized population groups, which underscores how environmental exposures reify existing racial inequities. We also found a consistent positive association between behavioral factors, including smoking and prevalence of adults who reported lacking physical leisure time, and diabetes-related ED visits across all subgroups.

Our findings should be interpreted within the broader context of structural racism. The higher burden of diabetes-coded ED encounters in more segregated, disadvantaged neighborhoods reflects the cumulative effects of discriminatory housing and infrastructure policies. These policies have concentrated health risks and constrained access to primary care in Black and other historically minoritized communities. Because of this, interventions that incorporate environmental justice principles are essential. Local planners and policymakers should prioritize measures that increase walkability and access to transit, reduce traffic congestion near residential areas, and develop public greenspaces especially in segregated and disadvantaged communities. Pursuing such strategies will not only address the uneven distribution of environmental benefits and burdens but also has implications for reducing racial disparities in chronic health outcomes like diabetes.

Our findings have significant implications for future research, which should consider additional geographic contexts such as other states or localities. Our study identified locally specific associations between population-level drivers and diabetes, but additional work is needed to establish external validity with respect to such drivers at different granularities, e.g., cross-state or national level studies. Future work should also address longitudinal changes in the spatial distribution of population-level drivers of

diabetes, which might improve scholarly understanding of the evolution of the dynamics of diabetes-related ED visits over time.

Acknowledgements This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability The predictor variables used in this study are from publicly available datasets provided by the sources as listed in Table S1. The emergency department data were obtained from AHRQ's Healthcare Cost and Utilization Project (HCUP) and are not publicly available because of the HCUP Data Use Agreement and HIPAA limited data set restrictions.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Taha MB, Valero-Elizondo J, Yahya T, Caraballo C, Khera R, Patel KV, Ali HJR, Sharma G, Mossialos E, Cainzos-Achirica M, Nasir K. Cost-related medication nonadherence in adults with diabetes in the United States: the National Health Interview Survey 2013–2018. *Diabetes Care*. 2022;45(3):594–603. <https://doi.org/10.2337/dc21-1757>.
2. Parker ED, Lin J, Mahoney T, Ume N, Yang G, Gabbay RA, ElSayed NA, Bannuru RR. Economic costs of diabetes in the U.S. in 2022. *Diabetes Care*. 2023;47(1):26–43. <https://doi.org/10.2337/dci23-0085>.
3. Deshpande AD, Harris-Hayes M, Schootman M. Epidemiology of diabetes and diabetes-related complications. *Phys Ther*. 2008;88(11):1254–64. <https://doi.org/10.2522/ptj.20080020>.
4. Washington RE, Andrews RM, Mutter R. Emergency department visits for adults with diabetes, 2010. In: Healthcare Cost and Utilization Project (HCUP) Statistical Briefs. Agency for Healthcare Research and Quality (US). 2006. <http://www.ncbi.nlm.nih.gov/books/NBK179290/>. Accessed 20 Dec 2025.
5. Liu J, Yang C, Liu B, Kim R, Philis-Tsimikas A, Isasi C, Wolfe D, Levy C, Moon J-Y, Li L. Factors associated with progression to type 2 diabetes among women with gestational diabetes. *J Urban Health*. 2025;102(4):877–82. <https://doi.org/10.1007/s11524-025-01013-7>.
6. Uppal TS, Chehal PK, Fernandes G, Haw JS, Shah M, Turbow S, Rajpathak S, Narayan KVM, Ali MK. Trends and variations in emergency department use associated with diabetes in the US by sociodemographic factors, 2008–2017. *JAMA Netw Open*. 2022;5(5): e2213867. <https://doi.org/10.1001/jamanetworkopen.2022.13867>.
7. Rosenstock S, Whitman S, West JF, Balkin M. Racial disparities in diabetes mortality in the 50 most populous US cities. *J Urban Health*. 2014;91(5):873–85. <https://doi.org/10.1007/s11524-013-9861-4>.
8. Yedjou CG, Sims JN, Njiki S, Chitoh AM, Joseph M, Cherkos AS, Tchounwou PB. Health and racial disparities in diabetes mellitus prevalence, management, policies, and outcomes in the United States. *J Community Med Public Health*. 2024;8(3):460.
9. Hassan S, Gujral UP, Quarells RC, Rhodes EC, Shah MK, Obi J, Lee WH, Shamambo L, Weber MB, Narayan KVM. Disparities in diabetes prevalence and management by race and ethnicity in the USA: defining a path forward. *Lancet Diabetes Endocrinol*. 2023;11(7):509–24. [https://doi.org/10.1016/S22138587\(23\)00129-8](https://doi.org/10.1016/S22138587(23)00129-8).
10. Thomas SB, Quinn SC, Butler J, Fryer CS, Garza MA. Toward a fourth generation of disparities research to achieve health equity. Annual Review of Public Health. 2011;32:399–416. <https://doi.org/10.1146/annurevpublhealth-031210-101136>.
11. Agardh E, Allebeck P, Hallqvist J, Moradi T, Sidorchuk A. Type 2 diabetes incidence and socio-economic position: a systematic review and meta-analysis. *Int J Epidemiol*. 2011;40(3):804–18. <https://doi.org/10.1093/ije/dyr029>.
12. LaVeist TA, Thorpe RJ, Galarraga JE, Bower KM, Gary-Webb TL. Environmental and socio-economic factors as contributors to racial disparities in diabetes prevalence. *J Gen Intern Med*. 2009;24(10):1144–8. <https://doi.org/10.1007/s11606-009-1085-7>.
13. Kershaw KN, Pender AE. Racial/ethnic residential segregation, obesity, and diabetes mellitus. *Curr Diab Rep*. 2016;16(11):108. <https://doi.org/10.1007/s11892-016-0800-0>.
14. Mayne SL, Loizzo L, Bancks MP, Carnethon MR, Barber S, Gordon-Larsen P, Carson AP, Schreiner PJ, Bantle AE, Whitaker KM, Kershaw KN. Racial residential segregation, racial discrimination, and diabetes: the Coronary Artery Risk Development in Young Adults study. *Health Place*. 2020;62: 102286. <https://doi.org/10.1016/j.healthplace.2020.102286>.
15. Rothstein, R. The color of law: a forgotten history of how our government segregated America. *Liveright*. 2017. <https://www.epi.org/publication/the-color-of-law-a-forgotten-history-of-how-our-government-segregated-america/>. Accessed 16 Dec 2025.
16. Link BG, Phelan J. Social conditions as fundamental causes of disease. *J Health Soc Behav*. 1995. <https://doi.org/10.2307/2626958>.

17. Williams DR, Collins C. Racial residential segregation: a fundamental cause of racial disparities in health. *Public Health Rep.* 2001;116(5):404–16.
18. Chen T, Guestrin C XGBoost: a scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* 2016;785–794. <https://doi.org/10.1145/2939672.2939785>.
19. Sperandei S. Understanding logistic regression analysis. *Biochem Med.* 2014;24(1):12–8. <https://doi.org/10.11613/BM.2014.003>.
20. Zhu R, Newman G, Li D. The spatial relationship between long-term vacant housing and non-communicable diseases in U.S. shrinking and growing metropolitan areas. *Cities.* 2024;145: 104718. <https://doi.org/10.1016/j.cities.2023.104718>.
21. Garvin E, Branas C, Keddem S, Sellman J, Cannuscio C. More than just an eyesore: local insights and solutions on vacant land and urban health. *J Urban Health.* 2013;90(3):412–26. <https://doi.org/10.1007/s11524-012-9782-7>.
22. Abdel Magid HS, Desjardins MR, Hu Y. Opportunities and shortcomings of AI for spatial epidemiology and health disparities research on aging and the life course. *Health Place.* 2024;89: 103323. <https://doi.org/10.1016/j.healthplace.2024.103323>.
23. Morello-Frosch R, Lopez R. The riskscape and the color line: examining the role of segregation in environmental health disparities. *Environ Res.* 2006;102(2):181–96. <https://doi.org/10.1016/j.envres.2006.05.007>.
24. Rigolon A, Browning MHEM, McAnirlin O, Yoon H(V. Green space and health equity: a systematic review on the potential of green space to reduce health disparities. *Int J Environ Res Public Health.* 2021;18(5):2563. <https://doi.org/10.3390/ijerph18052563>.
25. Chen KL, Zimmerman FJ, Ong PM, Jones A, Nuckols TK, Ponce NA, Choi KR, Li XX, Padilla A, Nau CL. Association of neighborhood gentrification and residential moves with hypertension and diabetes control in Los Angeles County, 2014–2019: a retrospective cohort study. *Health Place.* 2023;83: 103109. <https://doi.org/10.1016/j.healthplace.2023.103109>.
26. Wing RR, Goldstein MG, Acton KJ, Birch LL, Jakicic JM, Sallis JF Jr., et al. Behavioral science research in diabetes: lifestyle changes related to obesity, eating behavior, and physical activity. *Diabetes Care.* 2001;24(1):117–23. <https://doi.org/10.2337/diacare.24.1.117>.
27. Ross CE. Walking, exercising, and smoking: does neighborhood matter? *Soc Sci Med.* 2000;51(2):265–74. [https://doi.org/10.1016/S0277-9536\(99\)00451-7](https://doi.org/10.1016/S0277-9536(99)00451-7).
28. Rose SW, Rayens MK, Azadi M, Kierstead EC, Kong AY, Horn K. Examining the relationship between racial, ethnic, and economic residential segregation and cigarette smoking among a nationally representative sample of young adults. *Nicotine Tob Res.* 2025;27(6):1016–24. <https://doi.org/10.1093/ntr/ntae277>.
29. Feyissa TR, Wood SM, Vakil K, Mc Namara K, Coffee NT, Alsharrah S, Daniel M, Versace VL. The built environment and its association with type 2 diabetes mellitus incidence: a systematic review and meta-analysis of longitudinal studies. *Soc Sci Med.* 2024;361: 117372. <https://doi.org/10.1016/j.socscimed.2024.117372>.

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