Spring 2021

A Spatial Risk Prediction Model for Drug Overdose

Parisa Bozorgi

A SPATIAL RISK PREDICTION MODEL FOR DRUG OVERDOSE

by

Parisa Bozorgi

Bachelor of Science
School of Surveying Engineering, 2011

Master of Science
University of Maryland Baltimore County, 2015

Submitted in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in

Environmental Health Sciences

The Norman J. Arnold School of Public Health

University of South Carolina

2021

Accepted by:

Dwayne E. Porter, Major Professor

Jan M. Eberth, Committee Member

Geoffrey I. Scott, Committee Member

Jeannie P. Eidson, Committee Member

Tracey L. Weldon, Interim Vice Provost and Dean of the Graduate School

DEDICATION

I want to thank and dedicate this dissertation research to my family, who have supported me even from a far. I missed them through these years, and I could not have been here without their love and understanding.

ACKNOWLEDGEMENTS

It is a genuine pleasure to express my deep sense of thanks to my advisor, Dr. Dwayne Porter. I will be forever thankful for your guidance and support. I would like to acknowledge my supportive dissertation committee members Dr. Jan Eberth, Dr. Jeannie Eidson, and Dr. Geoffrey I. Scott for all of their support throughout my Ph.D. Your discussion, ideas, and feedback have been absolutely invaluable. I would especially like to thank Dr. Eberth for her financial support during my study. I am also thankful to my husband Dr. Amir Karami, for his constant encouragement throughout my research period. I'm extremely grateful to my Mom for her love, prayers, and educating me. I also thank Chris Finney from the SCRFA, who helped provide the data I needed. In addition, I thank my colleagues at SCDHEC for their support, encouragement, and collaboration.

ABSTRACT

Drug overdose is a leading cause of unintentional death in the United States and has contributed significantly to a decline in life expectancy from 2015 to 2018. Overdose deaths, especially from opioids, have also been recognized in recent years as a significant public health issue. To address this public health problem, this study sought to identify neighborhood-level (e.g., block group) factors associated with drug overdose and develop a spatial model using machine learning (ML) algorithms to predict the likelihood or risk of drug overdoses across South Carolina. This study included block group level socio-demographic factors and drug use variables which may influence the incidence of drug overdose. In particular, this study developed a new index of access to measure spatial access to treatment facilities and incorporated these variables to assess the relationship between drug overdose and accessibility to the treatment centers. We explored different ML algorithms (e.g., XGBoost, Random Forest) to identify optimum predictors in each category. The categories were combined into a final ensemble predictive model–that addressed spatial dependency. An evaluation was conducted to validate ~~that~~ the final model generalized well across the different datasets and geographical areas. Results of the study identified strong neighborhood-level predictors of a drug overdose, pinpointing the most critical neighborhood-level factor(s) that place a community at risk and protect communities from developing such problems. These factors included proportion of households receiving food stamps, households with income less than $35,000, high opioid prescription rates, smoking accessories expenditures, and low accessibility to opioid

treatment programs and hospitals. The generalized error of spatial models did not increase considerably in spatial cross-validation compared to the error estimated from normal cross-validation. Our model also outperformed the geographic weighted regression method. ~~Our~~ Results show that variables regarding socio-demographic factors, drug use variables, and protective resources can assist in spatial drug overdose prediction. Our finding highlights several specific pathways toward community-level intervention targeted to a vulnerable population facing potentially high burdens of drug abuse and overdose.

TABLE OF CONTENTS

Dedication iii

Acknowledgements iv

Abstract v

List of Tables ix

List of Figures x

Chapter 1: Introduction 1

1.1. Problem overview 2

1.2. Literature review 5

1.3. Gaps in the literature 10

1.4. Purpose statement 11

Chapter 2: Facility attractiveness and social vulnerability impacts on spatial accessibility to opioid treatment programs in South Carolina 13

2.1. Introduction 14

2.2. Study area 18

2.3. Data sources 19

2.4. Method 20

2.5. Analysis 21

2.6. Results 26

2.7. Discussion 30

2.8. Conclusion 33

Chapter 3: The development of a machine learning model for prediction of drug overdose in South Carolina 34

3.1. Introduction 35

3.2. Data and material 38

3.3. Methodology 51

3.4. Results 63

3.5. Discussion and conclusion 84

Chapter 4: Conclusion 90

References 94

LIST OF TABLES

Table 2.1. Distribution of spatial accessibility scores 28

Table 2.2. Spearman's Correlation between W2SFCA and 2SFCA 30

Table 3.1. Candidate explanatory variables 43

Table 3.2. Performance metrics of ML algorithms over different selection methods (risk factor domain) 67

Table 3.3. Performance metrics of ML algorithms over different selection methods (socio-demographic domain) 67

Table 3.4. Performance metrics of ML algorithms over different selection methods (protective resource domain) 67

Table 3.5. XGBoost Hyper-parameter values extracted with the use of grid search and parallel processing. 74

Table 3.6. Descriptive statistics of the prediction models 77

LIST OF FIGURES

Figure 2.1. Study area and spatial distribution of OTP facilities in SC 27

Figure 2.2. Geographic variation of spatial accessibility score (W2SFCA)..... 27

Figure 2.3. Geographic variation of spatial accessibility score includig SVI..... 28

Figure 2.4 Geographic variation of spatial accessibility score (2SFCA)..... 29

Figure 2.5. Hot spot and cold spot of spatial accessibility score (W2SFCA)..... 30

Figure 3.1. Descriptive statistics of the finalized variables 49

Figure 3.2. Descriptive statistics of the finalized variables 50

Figure 3.3. A machine learning-based framework to predict drug overdose..... 55

Figure 3.4. Ensembling process 55

Figure 3.5. Global Moran's I 63

Figure 3.6. Gi* statistic cluster map 64

Figure 3.7. Gi* statistic significance map..... 64

Figure 3.8. Local Moran's I cluster map..... 65

Figure 3.9. Local Moran's I significance map 65

Figure 3.10. Important predictors in risk factors domain 69

Figure 3.11. Important predictors in socio-demographic domain..... 69

Figure 3.12. Important predictors in protective resource domain..... 70

Figure 3.13. Tune max tree depth, subsample, and colsample parameters (risk factors) 70

Figure 3.14. Tune child weight, gamma and eta parameters (risk factors) 71

Figure 3.15. Tune max tree depth, subsample, and colsample parameters (socio-demographic) 71

Figure 3.16. Tune child weight, gamma and eta parameters (socio-demographic) 72

Figure 3.17. Tune max tree depth, subsample, and colsample parameters ((protective resources) 72

Figure 3.18. Tune child weight, gamma and eta parameters (protective resources) 73

Figure 3.19. Tune max tree depth, subsample, and colsample parameters (ensemble model) 73

Figure 3.20. Tune child weight, gamma and eta parameters (ensemble model) 74

Figure 3.21. Scatter plots of predicted versus observed drug overdose values 77

Figure 3.22. Map of actual drug overdose (top map) and prediction for each individual model and ensemble model at block group level 78

Figure 3.23. Performance evaluation results of models according to RMSE 81

Figure 3.24. Performance evaluation results of models across 5-folds 81

Figure 3.25. Distribution of residuals for individual and ensemble models 82

Figure 3.26. Moran's *I* scatter plot for ensemble model residual with 8 nearest neighbors 82

Figure 3.27. Moran's *I* scatter plot for ensemble model residual with 25 nearest neighbors 83

Figure 3.28. Moran's *I* scatter plot for ensemble model residual with 35 nearest neighbors 83

Figure 3.29. Moran's *I* scatter plot for ensemble model residual with queen contiguity neighbors 84

CHAPTER 1

INTRODUCTION

1.1. PROBLEM OVERVIEW

Drug overdose is a leading cause of unintentional death in the United States and has contributed significantly to a decline in life expectancy from 2015 to 2018 (Wilson, 2020). Overdose deaths, especially from opioids, have also been recognized in recent years as a significant public health burden (CDC, 2020b). This health crisis has emerged in three waves (CDC Injury Center, 2021). First, in the 1990s, pharmaceutical companies touted opioids as an effective and safe treatment for chronic pain, leading to a considerable rise in the numbers of opioid prescriptions. Then, in 2010, there was an increase in both the incidence of drug overdose and the use of heroin. The third wave hit in 2013 with significant increases in overdose deaths involving synthetic opioids such as fentanyl. Between 2010 and 2019, there were over 530,000 overdose deaths, and of those, more than half involved opioids (National Institute on Drug Abuse, 2021). The United States leads the world in opioid use, consuming about 80% of all opioids in the world. One out of every three adults in the country uses prescription opioids (Rummans et al., 2018). Among opioid-related deaths, the largest percentage were from the use of synthetic opioids other than methadone (this class includes illicitly manufactured fentanyl) with more than 36,000 deaths in 2019 with an age-adjusted death rate of 11.4 per 100,000 (National Institute on Drug Abuse, 2021). The prescription opioids category, which includes natural and semi-synthetic opioids (e.g., oxycodone and hydrocodone) and methadone, was the second most common cause of opioid deaths through 2017, with 17,029 deaths, or an age-adjusted death rate of 5.2 per 100,000 (National Institute on Drug Abuse, 2021). Since 2018, the rate of drug deaths due to prescription opioids fell slightly, perhaps reflecting tighter controls and changing practices that reduced the number of opioids prescribed. While deaths due to

prescription opioids have decreased, deaths due to heroin have increased (National Institute on Drug Abuse, 2021).

Concurrent with the rise in overdose deaths, there has been a rise in non-fatal overdoses and hospitalizations related to opioid abuse and misuse. Between 2005 and 2014, drug overdoses resulting in inpatient hospital admission and emergency department visits increased 64.1% and 99.4%, respectively (Weiss et al., 2020). Although considerably more attention has been devoted to the study of fatal overdoses, a non-fatal overdose is estimated to be between 20 to 30 times more common (Darke et al., 2003) and is associated with a range of harms. For instance, among injection drug users (IDU), non-fatal overdoses remain a critical determinant of morbidity and can lead to aspiration pneumonia, hypoxic brain injury, rhabdomyolysis, and renal failure (Darke & Hall, 2003). Repeated overdoses place a person at even greater risk of physical and cognitive impairment (Darke et al., 2007). Additionally, drug overdose can result in acute kidney injury due to dehydration, hypotension, and urinary retention. People who engage in drug use or high-risk behaviors associated with drug use are also at risk for acquiring and transmitting hepatitis B and C viral infections such as HIV (Mallappallil et al., 2017).

1.1.1. DRUG OVERDOSE IN SOUTH CAROLINA

The opioid epidemic has particularly affected South Carolina (SC), and deaths due to both drug overdose and opioids have been steadily increasing. In 2018, SC ranked 9th among states with the highest opioid prescription rate in the country at 793 per 1,000 residents (CDC, 2020c). The total number of drug overdose deaths was 1,131 in 2019, a 2.5% increase from 1,103 in 2018, and a 43.3% increase in total overdose deaths since

2015, jumping from 789 to 1,131. Further, of those 1,131 total overdose deaths in 2019, 923 involved prescription drugs (81.6%), 876 involved opioids (77.5%), 537 involved fentanyl (47.5%), 196 involved heroin (17.3%), and 230 involved cocaine (20.3%). Opioid deaths continue to rise in SC. In 2019, opioid-related overdose deaths increased by 7.3% from the preceding year. The five counties with the highest number of opioid-involved deaths were Horry (131); Charleston (107); Greenville (102); Spartanburg (55); and Richland (52). Horry County saw a 54% increase in opioid-related deaths, the highest in the state, going from 85 in 2018 to 131 in 2019 (*SC Drug Overdose Deaths*, n.d.).

Further evidence of this crisis comes from the administration of naloxone (Narcan or Evzio). Naloxone is an antidote medicine used to reverse and counter the effects of opioids in an overdose event. In 2019 there were 6,989 naloxone administrations by emergency medical technicians (EMTs), an approximate 11.2% increase from the 6,285 administered doses in 2018. Since 2015, there has been an approximate 41.6% increase in naloxone administrations, jumping from 4,933 to 6,989 (SC Department of Alcohol and Other Drug Abuse Services, n.d.).

Given increases in drug overdose deaths in SC, there is a need to develop more intervention and services to prevent drug overdose and overdose death. However, the first step to developing and implementing these services is to identify the factors that can predict drug overdose. In the following sections, we first review current factors identified in the literature that are related to the opioid and drug overdose epidemic. Then, we review methods and frameworks in past research that are commonly applied to predict drug-related outcome.

1.2. LITERATURE REVIEW

1.2.1. DRUG OVERDOSE FACTORS

Becoming involved in drug use and abuse may stem from a variety of factors, including genetic, biological, cognitive, family, and peer group factors (Scheier, 2010). Several individual characteristics that are linked to drug use, abuse, and overdose are identified by previous studies. For example, the non-Hispanic white individuals are more likely to have an overdose (Knowlton et al., 2013; Zedler et al., 2014). People who are divorced, separated, or not married are also at increased risk of fatal opioid overdose (Lanier et al., 2012). A systematic review also indicates that individual with lower incomes or insecure housing, without high school diploma, who are smokers, or who have been recently released from prison are at increased risk for drug use and overdose (Martins et al., 2015). Compared to people who own a house, those who rent are at increased risk for opioid overdose or abuse (CDC, 2015). While individual-level studies help identify people at risk of drug abuse and overdose, identifying contextual characteristics of a neighborhood environment that predict drug overdose is also important for community-based health intervention. Contextual characteristics of the neighborhood in relation with substance abuse have been theorized (Callahan, 2018; Galea et al., 2005). There are few studies that quantify these theories in relation with drug abuse and opioid overdose (Fite et al., 2009; Fuller et al., 2005; Frankenfeld and Leslie, 2019; Hembree et al., 2005). These studies show that a neighborhood disadvantage characterized by low income, poverty, low educational attainment, and high unemployment manifest greater risk regarding opioid overdose. Another study illustrates that residing in disadvantaged neighborhoods is associated with a higher overdose rate as residents misuse drugs to manage chronic stress resulting from

exposure to economic hardship and the associated experience of depression and anxiety (Boardman et al., 2001). Population density may affect substance use and overdose risk through a higher degree of collective socialization within dense urban areas, in which the norms and activities of a social network (peer pressure) influence individual behaviors (Galea et al., 2005; Latkin et al., 2003; Schroeder et al., 2001).

The impact of community socio-economic conditions on drug overdose rates may vary between rural and urban areas. Some studies indicate that rural residents are at a higher risk of opioid and drug overdose and overdose deaths than individuals living in urban areas (King et al., 2014). Also, studies show greater rates of opioid prescribing in rural areas (García et al., 2019; Keyes et al., 2014). Several factors may contribute to these findings. For example, rural residents are less likely to be administered naloxone during an overdose emergency than urban residents (Frank et al., 2016), and they often have fewer accessible treatment facilities than individuals living in urban areas (Dick et al., 2015; Kvamme et al., 2013).

Research has also shown that exposure to tobacco outlets, including convenience stores, gas stations, and other stores that typically sell tobacco products, is associated with increased rates of smoking among youth and young adults (Cantrell et al., 2015; Novak et al., 2006). Exposure to point-of-sale tobacco and alcohol advertisements, promotions, and marketing can also increase smoking rates as well as alcohol consumption among youth (Bryden et al., 2012; Paynter & Edwards, 2009). These behaviors are strong predictors of illicit and prescription drug abuse in young adults (Griffin et al., 2019).

Neighborhood protective resources within a community such as hospitals, opioid treatment programs, libraries, parks, exercise facilities, and learning centers can potentially decrease the risk of drug overdose. For example, exposure to green space has been associated with calming effects and reduced psychological stress (De Vries et al., 2013), thus countering, to a limited extent, the stressful conditions of economically disadvantaged neighborhoods (Mitchell & Popham, 2008) and consequent drug use as a coping behavior. Additionally, being physically active may significantly improve health outcomes by lowering an individual's risk for depression (Warburton et al., 2006). There are some discrepancies in the research about whether the presence of a fitness facility in a community promotes an active lifestyle (Ding et al., 2011; J. Feng et al., 2010); however, a dearth of these facilities in a neighborhood may provide residents with less of an opportunity to be physically active (Eriksson et al., 2012). Nevertheless, populations who perceive their communities as unsafe may be less likely to participate in outdoor fitness activities, including playing, walking, or running around the neighborhood (Molnar et al., 2004), and would benefit from improved access to exercise facilities.

Opioid use disorder medications are effective in overdose reduction and in promoting recovery (Krawczyk et al., 2020; Pitt et al., 2018). Studies illustrate that increased access to treatment decreases rates of drug overdose deaths, infectious disease transmission rates, and criminal activity. Additionally, increased access to treatment is associated with treatment retention and social functioning (Kakko et al., 2003; Schwartz et al., 2013). However, the effect of spatial access to these facilities, particularly to opioid treatment programs, has been understudied.

1.2.2. DATA ANALYSIS METHODS

Current studies apply many different methods, including spatial and non-spatial statistical approaches and intelligent algorithms to identify factors associated with overdose. Some studies have examined risk factors for opioid overdose using traditional statistical models designed to establish causation (Chichester et al., 2020; Frankenfeld & Leslie, 2019; Glanz et al., 2018; Seal et al., 2001; Thornton et al., 2018; Zedler et al., 2014). Chichester et al. (2020), for example, used multivariate regression along with principle component analysis (PCA) to identify an individual's risk factors for a fatal overdose. Geissert et al. (2018) employed logistic regression and ordinary least square (OLS) to predict opioid overdose as a linear combination of risk factors (Geissert et al., 2018). Similarly, Seal et al. (2001) used logistic regression to determine the risk factors for non-fatal overdose among street-recruited injection heroin users. Another study used negative binomial regression to compare county socioeconomic characteristics to death rates (Frankenfeld & Leslie, 2019).

Spatial approaches can be divided into hotspot methodologies, and spatial regression approaches. Hotspot methods such as Getis-Ord Gi* statistics and local Moran's *I* are exploratory spatial data analysis techniques that are commonly used to analyze drug overdoses spatially (Amram et al., 2019; Rossen et al., 2014; Stopka et al., 2019a). These methods identify and measure areas of local and global spatial association. Spatial regression methods, including OLS and geographic weighted regression (GWR) have been used to identify potential predictors of opioid misuse and those patients susceptible to abuse (X. Chen et al., 2017). To estimate the effect of proximity to various facilities (e.g., alcohol and tobacco stores, treatment centers) on drug overdose, research provides a measure of

accessibility using different approaches. The simplest way to compute access to facilities is to measure the number of providers or facilities within an administrative boundary or within a specified distance buffer (Eriksson et al., 2012). Other simple, distance-based measures can also be calculated using Geographic Information Systems (GIS), including the minimum distance between a point (e.g., patient address or population represented by a geographic centroid) and the closest facility, the true distance based on the actual facility used, and the average distance to all facilities with use potential (F. Wang, 2012). For instance, Amram et al. (2019) measured spatial access to methadone clinics by identifying the number of clinics within 20-minutes of walking time from patient addresses.

Furthermore, several studies provide a framework for the development and evaluation of a cumulative index based on socio-demographic characteristics of the population (Bohnert et al., 2011). These methods use established, including the use of equal weights for each variable, a sum of z-scores for selected variables, and the use of principal components analysis or factor analysis to estimate weights and build a composite index.

Among intelligent algorithms, machine learning (ML) techniques have been used widely to predict outcomes in a variety of healthcare applications. Machine learning is an efficient and effective approach to predicting and identifying hidden patterns in datasets with many variables. It can provide insight into modeling that is free from strict methodological assumptions required for traditional statistical approaches (Song et al., 2004; Wiemken & Kelley, 2020). However, machine learning applications in the context of drug overdose have been limited to identify individuals at risk of drug abuse or opioid

overdose using electronic medical records (Badger et al., 2019; Cochran et al., 2017; Ellis et al., 2019; Lo-Ciganic et al., 2019; Yang et al., 2015).

1.3. GAPS IN THE LITERATURE

The need to develop and assess strategies to combat the opioid epidemic warrant intense research activity. The research that has been carried out has studied different drug-related outcomes, including opioid overdose, drug overdose death, drug dependence, and abuse. A large body of work has been devoted to studying overdose mortality, though common wisdom and evidence suggest that non-fatal overdose events are much more common than fatal ones (Edwards, 2016). Although previous findings have identified potential risk factors for overdose, the studies have not investigated the scope of protective resource drivers such as access to treatment centers and recreational and green spaces that may affect the health and overdose risk for people who use opioids. Moreover, previous studies were conducted in specific populations (people who inject drugs, Medicaid recipients, veterans, and privately insured populations) that often do not generalize well to other US populations. Most research on predictors of drug-related overdose or overdose mortality has been devoted to identifying individual factors while ample evidence suggests that economic features of the populations' geographic contexts such as unemployment, poverty, and median household income can strongly influence drug use and abuse behaviors and overdose rates (Galea et al., 2003). Regarding the methods currently used in practice, most algorithms are based on traditional statistical approaches (e.g., OLS, GWR). These approaches have limited ability to handle nonlinear risk prediction and complex interactions among predictors. Machine learning algorithms often show better performance compared to traditional linear regression models (P. Feng et al., 2018) as they can handle

complex nonlinear relationships between the predictors and the responses and do not assume a specific shape of response function (e.g., linear or polynomial) (Shalev-Shwartz & Ben-David, 2014). Current studies that use machine learning methods for predicting drug overdose have not captured the spatial variation in their models and have not been compared to traditional regression. Moreover, although hotspot analysis can identify event concentrations with associated significance levels, they do not explain the factors contributing to these events. Essentially, hotspot methods only consider the dependent variable and are solely dependent on time or space to interpolate the events that occurred in the past.

1.4. PURPOSE STATEMENT

This study aims to identify neighborhood-level (e.g., block group) factors associated with drug overdose and develop a spatial model using machine learning algorithms to predict areas at most risk of drug overdoses across South Carolina. Identifying neighborhood characteristics that function either as potential protective factors or potential risk factors in association with drug overdose data can highlight specific pathways toward community-level intervention targeted to a vulnerable population. This study uses high-resolution spatial data at the block group level that can greatly enhance public health studies (Gabrysch et al., 2011) by improving context and decreasing spatial uncertainty (Murray et al., 2014) when compared with more aggregate units such as census tracts or ZIP Codes (Grubesic & Matisziw, 2006). This study includes socio-demographic factors and drug use variables which may influence the incidence of drug overdose. In particular, this study measures spatial access to treatment facilities and incorporates them as variables to assess the impact of access to these facilities on drug overdose. To the best

of our knowledge, no study has yet examined and compared different types of neighborhood-level factors related to drug overdose utilizing machine learning approaches.

This dissertation has been structured in the following manner: Chapter 1 provides an overview of the problem, past studies, and the purpose of the dissertation. Chapter 2 provides a detailed review of the literature regarding accessibility to protective resources/assets. We also present a spatial accessibility index that builds off of the two-step floating catchment area (2SFCA) method (W. Luo & Wang, 2003b) and which has three dimensions: a facility attractiveness index defined by services rendered and incorporated into the Huff Model (Dramowicz, 2005). A facility catchment area is defined as a function of facility attractiveness to account for variable catchment size, and a social vulnerability index (SVI) is incorporated to account for non-spatial factors that mitigate or compound the impacts of spatial access to care. The index guides the work in subsequent chapters and can be used as a model for future accessibility research.

Chapter 3 details the process of developing a spatial model to predict drug overdose across the state of SC at the block group level. We recognize the most critical neighborhood-level factors that place a community at risk of experiencing drug overdoses and factors that may help protect communities from developing such problems. Subsequently, we develop a robust spatial model using machine learning algorithms to predict drug overdose. An evaluation was conducted to validate that the final model generalized well across the different datasets and areas.

In Chapter 4, we emphasize significant study findings, discuss our work's strengths and limitations, and consider the public health implications.

CHAPTER 2

FACILITY ATTRACTIVENESS AND SOCIAL VULNERABILITY IMPACTS ON SPATIAL ACCESSIBILITY TO OPIOID TREATMENT PROGRAMS IN SOUTH CAROLINA[1]

---

1 Parisa Bozorgi, Jan M. Eberth, Jeannie P. Eidson, Dwayne E. Porter. Facility attractiveness and social vulnerability impacts on spatial accessibility to opioid treatment programs in South Carolina. Under review by International Journal of Environmental Research and Public Health

2.1. INTRODUCTION

In the United States, drug overdose deaths have more than tripled from 1999 to 2018. In 2018, opioid overdose was involved in almost 70% of these deaths (CDC, 2020a). In 2019, a total of 1,131 drug overdose deaths occurred in South Carolina, a 2.5% increase from 2018 with 77.4% involving an opioid. From 2018 to 2019, deaths involving all opioids, prescription opioids, and heroin increased by 7.4%, 7%, and 16%, respectively (SCDHEC, n.d.).

Three medications are currently approved by the Food and Drug Administration (FDA) to treat opioid dependence: methadone, buprenorphine, and naltrexone (National Academies of Sciences et al., 2019). Methadone can only be dispensed from the U.S. Substance Abuse and Mental Health Services Administration (SAMHSA)-certified OTPs and is the only safe option for pregnant and breastfeeding women. However, in 2017, over 70 percent of people who needed treatment for opioid use disorder (OUD) did not receive medications (Lipari, 2018). Of those who get access to specialty care, a minority (<30%) receive treatment with methadone or buprenorphine (Krawczyk et al., 2017). Among those in treatment, the numbers of people who receive evidence-based medications such as buprenorphine, methadone, and naltrexone are rising, but remain low (Beetham et al., 2019; Krawczyk et al., 2017; Shulman et al., 2019).

Previous studies identified obstacles to receiving treatment, including poor accessibility and availability, treatment cost, and lack of health insurance coverage, and lack of support services such as assistance with housing and transportation (Huskamp et al., 2018; Mancher et al., 2019). One study found patients traveled an average of 49 miles

to reach medication prescribers, and those traveling a mean distance greater than 45 miles to prescribers were less likely to regularly receive medications (Rosenblum et al., 2011a). Longer travel distances have also been associated with shorter length of stay in outpatient methadone clinics and lower probability of completion and aftercare utilization (Rosenblum et al., 2011a). Treatment retention is especially crucial among methadone-maintained patients because of the importance of continued medication often required to achieve and sustain treatment gains (Cooper et al., 2002; Meade et al., 2015). Further, traveling long distances for daily treatment like methadone adds a significant burden of transportation cost for most patients, especially for rural residents who need to travel a longer distance. Patients may also face a number of other challenges when seeking care such as difficulty finding child care and transportation (Chatterjee et al., 2018). The distance to an OTP has also been associated with the number of missed doses in the first month of treatment. Specifically, patients who lived more than 10 miles from the OTP were more likely to miss treatments compared to individuals who lived within 5 miles of the OTP (Amiri et al., 2018b).

While findings from these studies were critical in advancing our understanding about the importance of a geographic perspective on access to OTPs, inequality in spatial accessibility to OTPs in South Carolina has not been studied. Determining and evaluating geographic variation in spatial access to OTPs may help explain why some areas have a higher rate of drug overdose or drug overdose death.

Access to care is a multidimensional concept influenced by both spatial and nonspatial factors that can be further categorized into potential and revealed accessibility. Revealed accessibility focuses on the actual use of health care services, whereas potential

accessibility considers the population as the potential users of health care providers (W. Luo & Wang, 2003c). Spatial access to health care is primarily dependent on three factors: supply, demand, and travel costs between supply and demand. The two-step floating catchment area (2SFCA) method is based on the gravity model (W. Luo & Wang, 2003a) that considers both supply and demand, as well as their interaction. First, it defines a catchment (service area) of 30 minutes drive time around the facility and the population-to-provider ratios (PPR). The second step identifies a catchment around the demand location and searches for all the facilities within the demand's catchment area. Each facility found in a resident's catchment area will have a corresponding PPR, calculated in step one. The spatial accessibility index is calculated by summing all the PPR of all facilities within the demand catchment. The final 2SFCA score is computed in a two-step process expressed as follows:

Step 1: Generate a 30 minute drive time zone (catchment) concerning the provider site and compute the provider-to-population ratio at each provider location:

$$R_j = \frac{S_j}{\sum_{i \in \{d_{ij} \leq d_0\}} P_i} \qquad (1)$$

Where:

- R_j is the provider-to-population ratio at physician location *j*;
- P_i is a population of block group i; and
- d_0 is a travel threshold; d_{ij} is travel time between i and j.

Step 2: Generate another 30 min drive time catchment concerning the population site and compute the spatial accessibility index (A_i) for each population site:

$$A_i = \sum_{j \in \{d_{ij} \leq d_0\}} R_j \qquad (2)$$

Where

- R_j is the provider-to-population ratio at physician location *j*;
- d_0 is a travel threshold;
- d_{ij} is travel time between i and j; and
- and A_i is a spatial accessibility index of each population site i.

Despite the popularity of 2SFCA, the method has a drawback that it does not consider distance decay and assumes all services within the catchment area are equally accessible. Also, it uses a fixed catchment size, which is more problematic for urban and rural areas which may have very different commuting behaviors (Shah et al., 2016; Cooper et al., 2002). Modifications to the basic form of 2SFCA include improvements in catchment size (W. Luo & Whippo, 2012a; McGrail & Humphreys, 2014; Ni et al., 2015), the inclusion of competitive effects among the facilities (J. Luo, 2014; Wan, Zou, et al., 2012) and nonspatial factors (Lin et al., 2018a; Mao & Nekorchuk, 2013), incorporating distance decay within catchments (W. Luo & Qi, 2009a) and implementing variable catchment sizes (W. Luo & Whippo, 2012b).

Spatial accessibility models have been widely used to measure access to different types of healthcare facilities and services, including inpatient health care, mammography, cancer screening, and primary care (Ranga & Panda, 2014; Donohoe et al., 2016; Stewart et al., 2020; Lin et al., 2018b). However, geographic variation in accessibility to OTPs remains primarily unknown. This research develops a spatial access model building off the two-step floating catchment area (2SFCA) method and accounting for nonspatial factors and facility attractiveness, providing a more reasonable pattern than the traditional 2SFCA method. Specifically, this research examines spatial accessibility to OTPs to identify low

and high spatial access areas in South Carolina. The findings provide a support for state and local governments to better allocate treatment resources where access to treatments is limited.

2.2. STUDY AREA

A spatial accessibility model was calculated for block groups in South Carolina, a state located in the southeastern region of the U.S. with a population of 5,148.714 over a 32,020 mi2 area and characterized by rural and urban landscapes (Wikipedia, 2020). South Carolina has 46 counties and 3,046 block groups. There are 21 OTPs statewide, with most clustered in urban areas and only 4 OTPs located in rural areas. From a demographic perspective, many counties (28 out of 46 counties) are classified as highly vulnerable populations based on the CDC SVI score, which accounts for almost 30% of the state's total population.

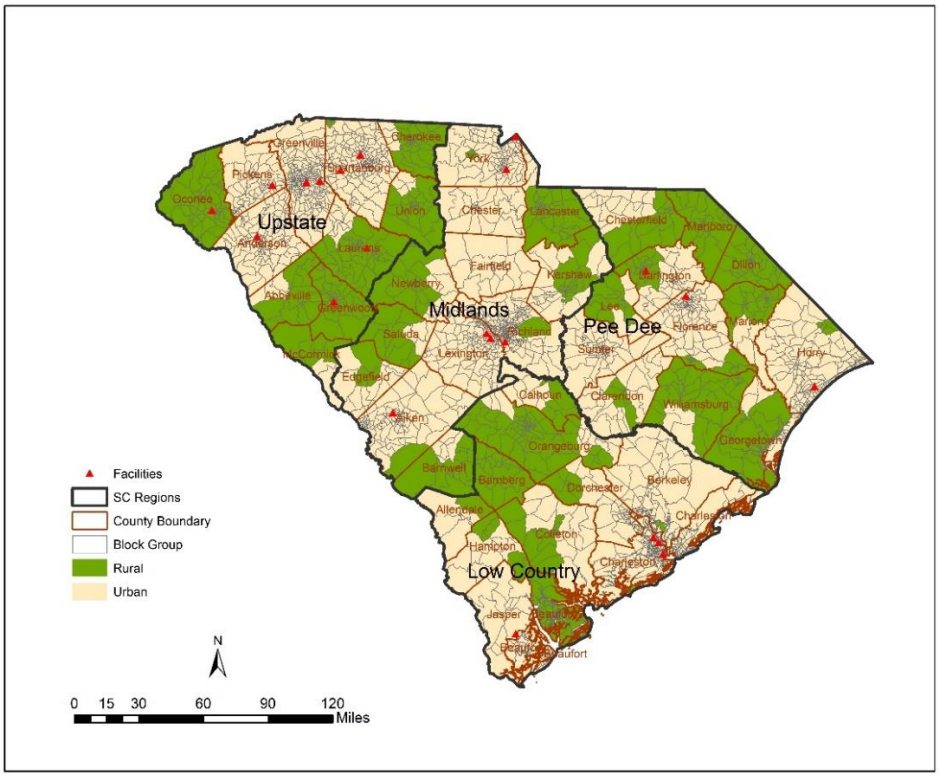

Figure 2.1. Study area and spatial distribution of OTP facilities in SC

2.3. DATA SOURCES

Information on OTPs was obtained from the publicly available Substance Abuse and Mental Health Services Administration (SAMHSA) data released in 2019. The data contain the location and services provided by facilities. The location of services was geocoded with the corresponding street addresses.

Population data were extracted at the block group level from the U.S. Census Bureau's Integrated Public Use Microdata Series (IPUMS), explicitly using the 2013-2017 American Community Survey. To represent population location more accurately, we calculated population-weighted block group centroids based on Census block population.

Distances between OTP service locations and demand locations were calculated based on the 2018 street network using Network Analysis of ArcGIS Pro (ESRI Inc., Redlands, CA, USA).

The Social Vulnerability Index (SVI) at the Census tract level was obtained from the 2017 Centers for Disease Control and Prevention (CDC) (CDC, 2018). The SVI was created to identify socially vulnerable populations and rank U.S. Census tracts based on the resident population's demographics. It ranks four domains (Socioeconomic Status, Household Composition & Disability, Minority Status & Language, and Housing & Transportation) based on 2-5 demographic indicators in addition to Overall Vulnerability, which aggregates all the indicators into a single summary rank. We assumed that all the block groups within the Census tract have the same overall ranking as their Census tract.

2.4. METHOD

2.4.1. OVERVIEW

This study estimates facility attractiveness and uses the Huff Model for quantifying the probability of a person's preference on an OTP site, accounting for factors including distance to and the attractiveness of the OTP site. A key feature of the proposed model, besides measuring attractiveness of the facility based on multiple attributes, is to integrate the Centers for Disease Control and Prevention (CDC) Social Vulnerability Index (SVI) to account for nonspatial factors. The facility catchment size is also determined as a function of facility attractiveness. We evaluate the relation between our model (i.e., weighted 2SFCA (W2SFCA)) and the 2SFCA model using the Spearman correlation coefficient and the Intraclass Correlation Coefficient (ICC). To assess whether high or low access score

cluster spatially, the optimized hot spot analysis with optimal distance band identified based on incremental spatial autocorrelation is used. Choropleth maps of the final accessibility indices highlight differences between the methods.

2.5. ANALYSIS

To address the limitation of previous accessibility models, our method focuses on enhancing the provider catchment size and applying nonspatial factors, in three steps.

In the first step, a facility catchment size was defined as a function of facility attractiveness. To determine facility attractiveness, a composite index of attractiveness was developed based on factors including the type of opioid treatment, ancillary services provided, payment/insurance types accepted, Medicare/Medicaid patient acceptance, and language services. A facility's service was given more weight if the facility is located within an area where the majority of the population are vulnerable due to lack of that service. For example, greater weight was allocated to the housing and transportation services provided by the facility if the site was located in an area where the majority of the population are classified in the highest vulnerability category for housing and transportation; otherwise, no weight is given to that service. Determination of the highest/lowest vulnerable population is based on the CDC SVI score (4 categories representing 0-25%, 25.01-50%, 50.01-75%, 75.01-100%). In this step, the facility attractiveness at a treatment facility j (C_j) was quantified as a sum of the weighted attributes mentioned earlier:

$$C_j = \sum_{k=1}^{n} W_k X_k \qquad (3)$$

Where:

- X_k is the *kth* attributes assigned for treatment facility *j*;
- W_k is the weight assigned to the attribute X_k.

A high score effectively increases the size of the population competing to access available services. Then, we used the Huff model to estimate the most likely population accessing the facility. For each block group, we measured and/or created:

- a population-weighted centroid to represent the location of the demand population.
- the travel time between each block group centroid and facility address using the origin-destination (O.D.) cost matrix function of ArcGIS Pro 2.3.
- an 80-minute drive-time catchment area around the demand location calculated using the closest facility function of ArcGIS Pro 2.3.
- the Huff model selection probability of a population location on each treatment facility within its catchment using Equation (4).

$$Prob_i = \frac{C_j e^{\frac{-d^2}{\beta}}}{\sum_{s \in D_0} C_s e^{\frac{-d^2}{\beta}}} \qquad (4)$$

Where:

- $Prob_i$ is the Huff model-based selection probability of population i at treatment facility *j*;
- C_j is the attractiveness of treatment facility *j* calculated from the previous step;
- d_{ij} is the shortest travel time from population i to treatment facility $j$and β is the distance impedance coefficient.

Calculation of the shortest travel time from population centroid to the OTPs showed that an 80-minute drive-time ensured each block group has access to at least one OTP within its catchment. The value of β was estimated using the Gaussian function (Equation 5). A value of 0.01 was considered a threshold value when the distance decay function approaches to 0 (Wan, Zhan, et al., 2012). The Gaussian function was adopted as the distance decay function because it has been proved superior to other functions in simulating the distance impedance effect (L. Wang, 2007).

$$f_d = e^{\frac{-d^2}{\beta}} \qquad \beta = -\frac{{d_0}^2}{\ln 0.01} \qquad (5)$$

In the second step, the facility catchment size (D) was defined as a function of the treatment facility attractiveness using the Gaussian function (Equation 6). To differentiate the facility catchment size in urban and rural areas, we determined the urban/rural status of the facilities using the 2013 urban-rural classification from USDA's Rural-Urban Commuting Area (RUCA) codes. RUCA codes classify U.S. Census tracts using measures of population density, urbanization, and daily commuting. A facility within a metropolitan area (codes 1-3) was defined as urban; all other facilities are labeled as rural (codes 4-10). Among facilities located in rural areas, the facility catchment size (D) was based on a threshold of 60-minute drive-time vs. 30-minute drive-time for facilities located in urban areas. Towards our goal of defining effective facility catchment sizes, these numbers were multiplied by the facility attractivness formulated using the Gaussian function. The facility catchment sizes ranged from 17.2 – 30 minutes in urban areas and from 32.5 – 46.2 minutes in rural areas.

$$\text{D} = e^{-(\frac{C_j - C_m}{C_m})^2} * 30 \text{ min} \quad C_j \leq C_m$$
$$\text{D} = e^{-(\frac{C_j - C_m}{C_m})^2} * 60 \text{ min} \quad C_j \leq C_m \qquad (6)$$

Where:

- D is the facility catchment size
- C_m is the maximum attractiveness score.

Then, provider-to-population ratio (R_j) were calculated using Equation 7.

$$R_j = \frac{C_j}{\sum_{(i \in D_0)} Prob_{ij} W_{ij} P_i} \qquad (7)$$

$$W_{ij} = e^{\frac{-d^2}{\beta}}$$

Where:

- R_j is a provider-to-population ratio at treatment facility j;
- P_i is a weighted population of block group i;
- D_0 is a travel threshold;
- W_{ij} is a travel impedance between i and j;

The numerator was weighted by the facility attractiveness because facilities offering more services are more attractive than others.

In the third step, an 80-minute drive-time catchment area was defined around the population-weighted block group centroid and the ratios were summed from all facility locations falling within this catchment area. However, to account for nonspatial factors, we consider the output by the CDC SVI index associated with each population location. A

high SVI score effectively reduces a population catchment size due to the higher social vulnerability and associated service needs of the population. Areas with higher scores for A_i are considered to have better spatial accessibility to OTPs. The accessibility score is expressed as:

$$A_i = \sum_{(j \in D_0)} R_j \, W_{ij} \, Prob_{ij} \, SVI^{-1} \qquad (8)$$

Where:

- A_i is the accessibility at population location *i*;

Using the same datasets, we compared our weighted 2SFCA (W2SFCA) model with the original 2SFCA model. Choropleth maps were also generated using ArcGIS Pro, allowing the visualization of our final accessibility index vs. the traditional 2SFCA method. We also conducted the hot spot analysis using the local Getis-Ord Gi* statistic for the spatial accessibility score. The method identifies statistically significant clusters of high values (hot spots) and low values (cold spots) within the framework of the conceptualized spatial relationship. The Gi* statistic consists of a ratio of the weighted average of the values in the neighboring locations to the sum of all values, including the value at the location (x_i) (Equation 9). To quantify spatial relationship among block groups, we generated a spatial weight matrix that related each location to its nearest eight neighboring locations.

$$G_i^* = \frac{\Sigma_j w_{ij} x_i}{\Sigma_j x_i} \qquad (9)$$

Where:

- w_{ij} is the spatial weight between districts *i* and *j*

2.6. RESULTS

Measures of central tendency and dispersion among the two accessibility scores is shown in Table 2.1. We tested the association between the two methods with data measured continuously using the Spearman correlation coefficient method. A positive relationship was found with a coefficient of 0.73 and a p-value of 0.003 (Table 2.2).

The ICC was measured by a single-rating and 2-way random-effects model with two methods across 3045 subjects (Table 2.2). Although the obtained ICC value was 0.71 (indicating moderate reliability), a 95% confidence interval ranges between 0.2 and 0.8, meaning that there is a 95% chance that the true ICC value lands on any point between 0.2 and 0.8. Therefore, the level of reliability can be interpreted as poor to moderate. The geographic patterns of accessibility index computed by the W2SFCA (before and after including SVI) and the traditional 2SFCA model are shown in Figures 2.2-2.4. The spatial distribution of accessibility by W2SFCA (Figure. 2.2) showed a relatively similar pattern to the traditional 2SFCA (Figure. 2.4). However, the range of the accessibility scores by W2SFCA was smaller than the range of 2SFCA. For spatial comparison of the two methods, quantile classification groups with four classes were used.

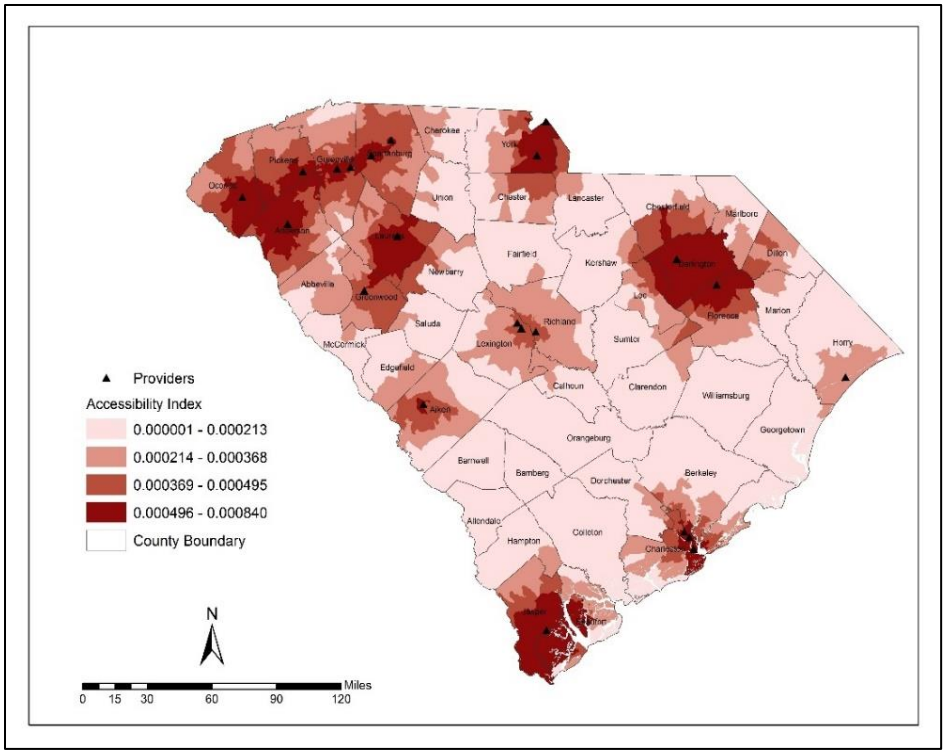

Figure 2.2. Geographic variation of spatial accessibility score (W2SFCA)

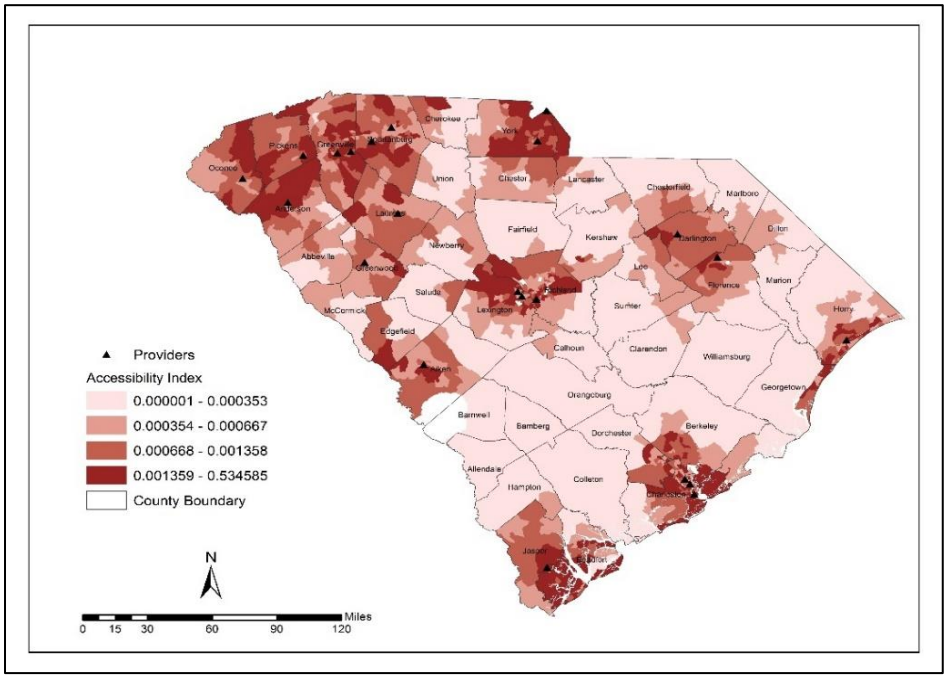

Figure 2.3. Geographic variation of spatial accessibility score (W2SFCA) including SVI

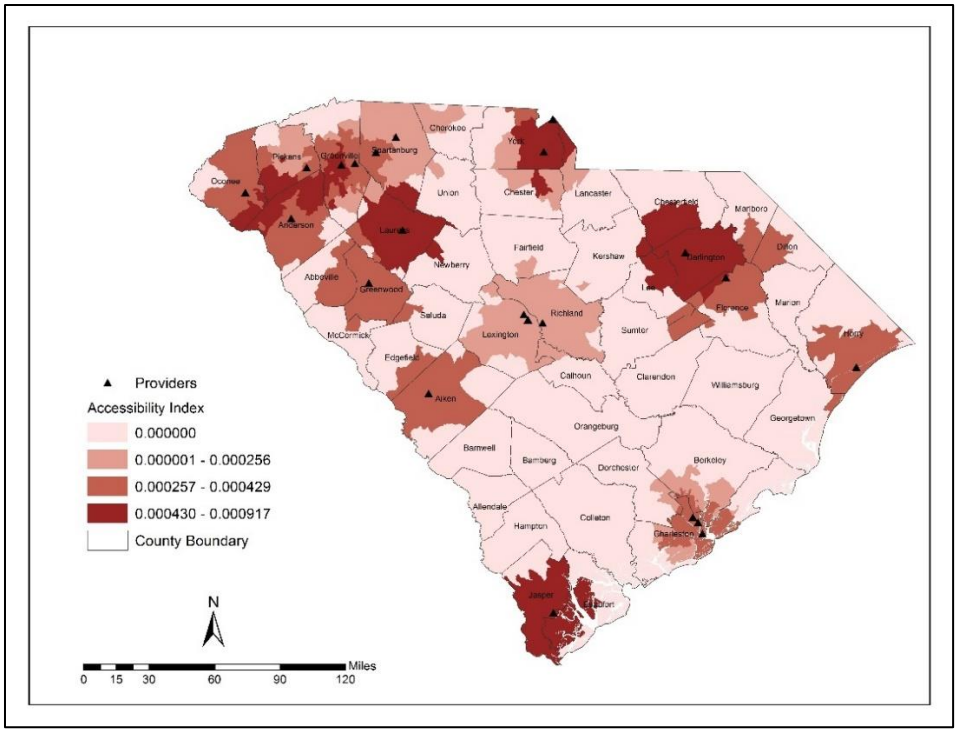

Figure 2.4. Geographic variation of spatial accessibility score (2SFCA)

Table 2.1. Distribution of spatial accessibility scores

Variable	Mean	Median	SD	IQR	Range
W2SFCA	0.00035	0.00036	0.00017	0.00028	0.00083
2SFCA	0.00024	0.00025	0.00020	0.00038	0.00091

According to the results obtained from W2SFCA shown in Figure 2.2, the spatial accessibility to OTPs is unevenly distributed. Areas with higher access were primarily located in the northern part of the state, with very few located in the south and southeast of the state. From the results of the accessibilities analysis with the proposed method, approximately 21% of the state's population lives in areas with low access, 23% live in areas identified as medium-low access, 26% live in areas identified as medium-high

access, and 30% live in high access areas. A majority of the population with low access (85%) live in areas with a moderate to a high level of social vulnerability.

In comparison with the 2SFCA, as expected, W2SFCA revealed more details of accessibility than 2SFCA. For example, in the vicinity of OTPs located in Richland and Lexington counties, the accessibility is underestimated by 2SFCA. The 2SFCA model detected all the block groups within these counties as areas with low accessibility while some of their block groups encompassed an OTP provider, and some were close to nearby OTP sites. This is due to the same catchment size regardless of the attractiveness of the OTP facilities. The weighted score by SVI revealed disparity in accessibility to OTPs relative to the socio-economic status of the population (Figure 2.3). As shown in Figure 2.3, some block groups adjacent to the OTP facility are identified as areas with low access within the Spartanburg city limits. People living in this area are ranked as a highly vulnerable population, and their socio-economic status can affect their accessibility to the OTPs. Some of these OTP facilities are among facilities with the lowest attractiveness index indicating they either do not accept Medicaid/Medicare patients or do not provide additional services that can be beneficial for vulnerable populations.

Table 2.2. Spearman's Correlation and ICC between W2SFCA and 2SFCA

		W2SFCA
2SFCA	Spearman's Correlation	0.73
	ICC (95% CI)	0.71(0.213-0.867)

Results of the hot spot analysis are shown in Figure 2.6. Cold spots with clusters of low accessibility were discovered in the much of the Midlands, Pee Dee, and Lowcountry regions (notable exceptions in Charleston, Beaufort, Darlington and Florence Counties). Hot spots with clusters of high accessibility were clustered in the Upstate region, as well as Aiken County, the border of York and Lancaster County, and the counties listed above. Many of these hot spots were clustered near metropolitan areas of the state, or bordering states.

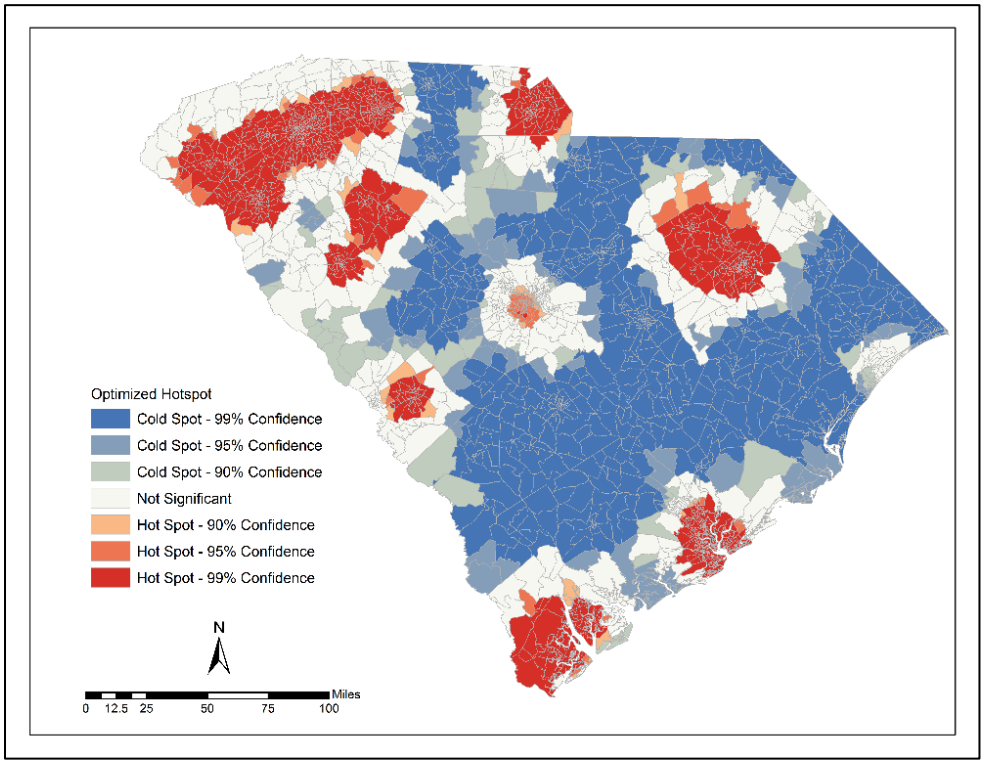

Figure 2.5. Hot spot and cold spot of spatial accessibility score (W2SFCA)

2.7. DISCUSSION

The primary goal of this study is to explore the geographic variation of spatial accessibility to OTPs and to identify areas with poor accessibility in South Carolina. This paper outlines a new index of access that integrates facility attractiveness and socio-

economic factors to the existing metrics. The facility attractiveness includes services offered by the facility that helps to measure each facility's attractiveness for opioid users. Most previous studies use a distance impedance coefficient β to create weights within the service catchment. These studies measure β by using the actual travel distance of patients who visited the treatment center. However, estimating β based on the empirical data is likely to be confounded with the existing distribution of facilities in a region instead of representing the patients' inclination to travel to a facility. We defined facility catchment size as a function of facility attractiveness formulated by the Gaussian function to moderate the effect on spatial access measure for different impedance coefficients (J. Luo, 2014). The SVI includes variables that help to identify populations who are more likely to have a lack of access to the OTPs. The integration of these factors makes this approach more realistic and provides a better fit for modeling access to OTPs. Additionally, this index has been designed to use data at the small geographic unit (block group), which identifies areas with poor access to OTPs at a much finer geographic scale than existing methods.

We compared our model with the 2SFCA methods. We found that spatial accessibility is underestimated in some areas using the 2SFCA method. This problem has been partially alleviated in the W2SFCA method by incorporating SVI and facility attractiveness into the model. We showed that not only being too far from the facility can result in decreased access to the facility, but also sociodemographic factors and lack of accommodation at the facility (e.g., not accepting certain insurance plans) can present an obstacle to access care at the facility. Our findings have several public health implications. It can be used for the identification of OTPs accessibility variations throughout the state and possibly improving access to OTPs. Specifically, the scale of analysis provides more

granularity to uncover local areas of spatial homogeneity and heterogeneity for community-based interventions. Moreover, results of cluster analysis (e.g., clusters of low access) can be overlaid with the clustering of a high rate of drug overdose to target interventions in areas where treatment programs are most needed. Our methodology is also deployable to other healthcare facilities such as HIV care providers and mental health services.

Despite this notable advantage of W2SFCA, several issues deserve attention when interpreting the results. Population locations used for this study are weighted block group centroids. The developed method, however, has the potential to further articulate the population selection behavior because the block group population is not necessarily a proper indicator of opioid treatment needs. This can be partially addressed in future development by incorporating the number of patients with a history of prescription opioid use or experienced opioid overdoes. This study also assumes that all patients traveled by car and don't consider different modes of transportation, such as public transportation, as it is somewhat limited in the state. Moreover, it is possible to adjust the weights used for estimating the attractiveness score. Different weighting scenarios can be implemented in the future study to assess sensitivity and robustness of the spatial accessibility score. Among treatments provided at OTP facilities, methadone currently needs to be taken under the supervision of a practitioner (*Methadone*, n.d.); however, patients can take the treatment at home for maintenance purposes if they meet certain criteria. Policies to make take-home treatment more accessible should be considered to minimize the impact of geographic distance on treatment utilization. The impact of these policies on accessibility could be an important future area of spatial accessibility research.

2.8. CONCLUSION

This study provided a new perspective for analyzing healthcare accessibility, including both spatial and nonspatial factors to define accessibility to OTPs in South Carolina. The results of this study indicated a significant variation in access to OTPs statewide. Cluster of low spatial access were mainly observed in the middle, south, and southeast of the state with exception in the metropolitan area of Columbia and Charleston. Rather than defined accessibility solely on distance to OTP facilities, we considered the role of facility attractiveness and social vulnerability of the potential demand populations. The traditional 2SFCA overestimates regional accessibility and the W2SFCA can provide a more realistic evaluation. Based on this study, policymakers and public-health officials should consider optimizing the allocation of existing healthcare resources or putting additional resources into low accessibility areas.

CHAPTER 3

THE DEVELOPMENT OF A MACHINE LEARNING MODEL FOR PREDICTION OF DRUG OVERDOSE IN SOUTH CAROLINA[2]

[2] Parisa Bozorgi, Dwayne E. Porter, Amir Karami, The Development of a Machine Learning Model For Prediction of Drug Overdose in South Carolina. Under review by Health & Place Journal

3.1. INTRODUCTION

Drug overdose is a leading cause of unintentional death in the United States and has contributed significantly to a decline in life expectancy from 2015 to 2018 (Wilson, 2020). Overdose deaths, especially from opioids, have also been recognized in recent years as a significant public health burden (CDC, 2020b). Prescription opioids have the highest levels of dependence, abuse, and poisoning (Hastings et al., 2020). Among opioid-involved deaths, the category of synthetic opioids other than methadone (illicitly manufactured fentanyl) was the most common with more than 36,000 deaths in 2019. The prescription opioids category, which includes natural and semi-synthetic opioids (e.g., oxycodone and hydrocodone) and methadone, was the second most common with 17,029 deaths with an age-adjusted mortality rate of 5.2 per 100,000 (National Institute on Drug Abuse, 2021). The opioid epidemic has particularly affected South Carolina (SC), and deaths due to both drug overdose and opioids have been steadily increasing. In 2018, South Carolina ranked 9th among states with the highest opioid prescription rate in the country at 793 per 1,000 residents. The total number of drug overdose deaths was 1,131 in 2019, a 2.5% increase from the 1,103 in 2018, and a 43.3% increase in total overdose deaths since 2015. Further, of those 1,131 total overdose deaths in 2019, 923 involved prescription drugs, 876 involved opioids, 537 involved fentanyl, 196 involved heroin, and 230 involved cocaine. Opioid deaths continue to rise in South Carolina. In 2019, opioid-related overdose deaths increased by 7.3% from the preceding year (*SC Drug Overdose Deaths*, n.d.). The five counties with the highest number of opioid-involved deaths were Horry (131), Charleston (107); Greenville (102); Spartanburg (55); and Richland (52). Horry County saw a 54% increase

in opioid-related deaths, the highest in the state, going from 85 in 2018 to 131 in 2019 (*SC Drug Overdose Deaths*, n.d.).

Given increases in drug overdoses, there is a need to develop more intervention and prevention services to prevent drug overdose and overdose death. The first step to developing and implementing these services is to identify a set of factors that best predict the location and magnitude of potential drug overdoses. Studies have typically examined drug-related outcomes at the level of the individual. While individual-level studies help to identify individuals at risk of drug abuse and overdose, identifying neighborhood-level factors that predict drug overdose is important for community-based intervention including policies and programs (Hembree et al., 2005). There is a lack of studies investigating the association of neighborhood characteristics with a drug overdose at the neighborhood level. Most neighborhood-level studies have examined the relationship between socio-demographic characteristics with opioid overdose and they have not examined different types of neighborhood-level factors such as drug-related risk factors and protective factors. Methodological approaches used to quantify these factors are also important. A large body of works have measured drug overdose risk by identifying overdose hot spots using geospatial techniques such as Getis-Ord Gi* and local Moran's *I* statistics (Dworkis et al., 2017; Hernandez et al., 2020; Saloner & Karthikeyan, 2015; Stopka et al., 2019b). These studies employed such methods to identify areas in need of overdose prevention and harm reduction resources, such as Narcan training, needle exchange facilities, or safe injection sites. However, developing policies based on these approaches can be ineffective as these methods rely on past events. Spatial regression methods, including OLS and geographic weighted regression (GWR), have been also used to identify potential predictors of opioid

misuse and patients susceptible to abuse (X. Chen et al., 2017). Furthermore, several studies used traditional statistical approaches to predict risk of opioid addiction and overdose or to examine the association of risk factors with opioid overdose (Chichester et al., 2020; Frankenfeld & Leslie, 2019; Glanz et al., 2018; Seal et al., 2001; Thornton et al., 2018; Zedler et al., 2014). These studies were conducted in specific populations (e.g., people who inject drugs, Medicaid recipients, veterans, and privately insured populations) that may not generalize well with other US populations.

In recent years, machine learning (ML) techniques have been used widely for predicting outcomes in a variety of healthcare applications. Machine learning algorithms often show better performance compared to traditional linear regression models (P. Feng et al., 2018) as they can handle complex nonlinear relationships between the predictors and the responses and do not assume a specific shape of response function (e.g., linear or polynomial) (Shalev-Shwartz & Ben-David, 2014). Machine learning applications in the context of drug overdoses have been limited to identifying individuals at risk of drug abuse or opioid overdose using electronic medical records (Badger et al., 2019; Cochran et al., 2017; Ellis et al., 2019; Lo-Ciganic et al., 2019; Yang et al., 2015). These studies have not captured geographic variation in their models.

To address the above limitations, this study aims to fill this research gap by building a spatial model to predict a location and magnitude of potential drug overdose using machine learning. We identify the most important block group level predictors of drug overdose using feature selection methods. Specifically, we build an ensemble model using three individual predictive models that are combined into a final predictive model. The three individual models include socio-demographic characteristics, drug related factors,

and protective resources. The ensemble model is built using a machine learning algorithm and GWR, which we then compare them in terms of R-squared and Root Mean Square Error (RMSE). Our models are unique in their inclusion of spatial dependency in the machine learning model to account for spatial autocorrelation. We use a high-resolution spatial data at the block group level that can greatly enhance public health studies (Gabrysch et al., 2011) by improving context and decreasing spatial uncertainty (Murray et al., 2014) when compared with larger, aggregate units such as census tracts or ZIP Codes (Grubesic & Matisziw, 2006). In the following section, we detail the process by which a spatial model is developed to predict drug overdoses in South Carolina.

Aim 1: Exploring advanced GIS and spatial statistical methods to examine spatial dependency and spatial pattern in drug overdose.

Hypothesis 1.1: There is a significant spatial autocorrelation in drug overdose across South Carolina.

Hypothesis 1.2: There is a local spatial association in drug overdose in South Carolina.

Aim 2: Identify top community protective resources, overdose-related risk factors, and socio-demographic factors and develop/validate a predictive spatial risk model of the top factors in each domain.

Hypothesis 2.1: risks factors can assist in drug overdose prediction.

Hypothesis 2.2: socio-demographic factors can assist in drug overdose prediction.

Hypothesis 2.3: community/neighborhood protective resource factors can assist in drug overdose prediction.

Aim 3: Develop and validate a predictive spatial risk model of drug overdose by ensembling of the three domains: community protective resources, overdose-related risk factors, and socio-demographic factors in South Carolina.

Hypothesis 3.1: Ensembling can improve the efficiency and effectiveness of the predictive spatial risk model in aim 2.

3.2. DATA AND MATERIAL

3.2.1. STUDY AREA

A spatial risk prediction model was calculated for all block groups (n= 3046) in South Carolina, a state located in the southeastern region of the U.S. with a population of 4,625,364 over a 30,060 mi^2 area. The state is characterized by both rural and urban landscapes and racial/ethnic diversity (US Census Bureau, n.d.).

South Carolina is divided into four Environmental Affairs (EA) regions that provide local support to the communities located within their boundaries: 1) Upstate, 2) Midlands, 3) Pee Dee, and 4) Lowcountry. The Upstate region covers the northwest quadrant of S.C., Midlands covers the center of the state from York to Barnwell counties, Lowcountry covers the south quadrant, and Pee Dee contains the northeast region of the state. South Carolina has 46 counties and 3,059 block groups, which are small subdivisions of Census tracts designed to be demographically homogeneous. They typically have a population between 600 and 3,000 people (US Census Bureau, n.d.). Block groups assigned to bodies of water and those with no residential population were excluded from the analysis, thereby reducing the total number of block groups to 3046.

3.2.2. DATA SOURCES

Variables were collected from various sources. The 2018 Naloxone administration data was obtained from the S.C. Department of Environmental and Health Control, Bureau of Emergency Medical Services (SCDHEC, EMS). This data set includes records of all patients who received naloxone by the EMS or law enforcement during 2018. This Naloxone administration dataset does not reflect private third-party administrations not made by EMS or law enforcement.

The 2018 inpatient and emergency department (ED) discharges related to drug overdose were also obtained from S.C. Revenue and Fiscal Affairs Office (SCRFA). This data set includes the unique number of individuals who experienced drug overdose defined by ICD-10 codes T36-T50 aggregated at the block group level. Block groups with less than 10 individuals with an overdose were suppressed by the SCRFA.

Opioid prescriptions were obtained from the South Carolina Prescription Drug Monitoring Program (SCPDMP), which is called South Carolina Reporting & Identification Prescription Tracking System (SCRIPTS). The SCRIPTS database includes all retail and outpatient hospital pharmacy dispensing of schedules II-IV controlled substances. It also consists of any controlled substance dispensing activity of those substances which occurs in the state of South Carolina, i.e., mail-orders pharmacies. The database does not include methadone clinics and emergency room/departments dispensing (less than a 48-hour supply). The rate of individuals who received at least one opioid prescription was obtained by dividing the number of individuals who received at least one

opioid prescription, during 2018, by the average population (older than 10) of each block group and multiplying the result by 100,000 to create a per capita rate.

Socio-demographic data were extracted at the neighborhood (e.g., block group) level from the U.S. Census Bureau's Integrated Public Use Microdata Series (IPUMS), explicitly using the 2014-2018 American Community Survey (IPUMS, n.d.).

The location of opioid treatment programs and buprenorphine providers were obtained from the Substance Abuse and Mental Health Services Administration (SAMHSA) released in 2019. The data contain the location and services provided by facilities. The location of services is geocoded with the corresponding street addresses (SAMHSA, n.d.) using ArcGIS Pro 2.6 (ESRI, 2020).

The urban-rural classification is derived from USDA's Rural-Urban Commuting Area (RUCA) codes. To determine the block groups' urban/rural status, we categorized them based on the rural-urban commuting area (RUCA) codes. RUCA codes classify U.S. Census tracts using measures of population density, urbanization, and daily commuting. For this analysis, a block group within a metropolitan defined Census tract (Code 1-3) is defined as urban. All other block groups are labeled as rural (Code 4-10) (USDA, n.d.).

The geocoded locations of off/on-premises alcohol retail stores, tobacco, library, and parks were obtained from ESRI Business Analyst by searching ~~for~~ businesses by North American Industry Classification System (NAICS) codes. To verify the accuracy of the business locations, we assessed the latitude and longitude coordinates using an in-house geocoder.

Block group level crime and smoking behaviors data were also collected from the 2018 ESRI market potential database (ESRI, n.d.). The market potential database is based on survey data from MRI-Simmons and measures the likely demand for a product or service in an area. The database includes an expected number of consumers for each product or service.

3.2.3. DATA PREPROCESSING

We manually reviewed EMS cases to identify drug-related overdose cases using a text search of chief complaint and to select for terms involving heroin, drug, and opioid and ICD-10 codes including T40.0 -- T40.6, heroin T40.1, methadone T40.3, cocaine T40.5. We then geocoded drug overdose and prescription data. All addresses were prepossessed to improve the geocoding quality, which has an impact on the derivation of the data at various geographic aggregation. Cases, including homeless and transient populations, persons who lacked a valid address, persons with only a P.O. box (n = 120), and addresses outside S.C. were removed. The data was reviewed for misspelled address information using Google Maps. Addresses were matched using a minimum match score of 85, spelling sensitivity of 80, and side offset of 10 feet, i.e., the default settings of ArcGIS. We then conducted interactive re-matching in ArcGIS, where addresses can be reviewed manually and corrected on a case by case basis as necessary. Addresses that couldn’t be geocoded to the exact location were removed from the dataset. We then aggregated EMS data at the block group level and generated rates for the following analyses.

Block group was selected as a unit of analysis, which allows for the determination of risk-levels at a granular level. For each block group, a rate of drug overdose per 1,000 persons (over the age of 10) was calculated and served as the dependent variable in the model. We defined an overdose event as a case involving the administration of Naloxone by EMS personnel for a nonfatal drug overdose involving a single dose or multiple doses for an individual patient, or an overdose discharged from the ED or hospital.

We investigated independent variables that were known or plausibly associated with a drug overdose based on prior research. The starting pool of independent variables was 115. As a result of consultation with experts, SCDHEC and the Department of Alcohol and Other Drug Abuse Services (DAODAS), variables that were not necessarily important in explaining variation in drug overdose were removed. The final data comprised 83 variables. A list of variables is summarized in Table 3.1. Descriptive statistics, including median, mean, first quartile, and third quartile, were calculated for all the finalized variables included in the model. (Figures 3.1 and 3.2).

Table 3.1. Candidate explanatory variables

Category	**Subcategory**	**Variables**	**Data Source**
Socio-demographic	Unemployment	Long term unemployed Population 16+	ACS (2014-2018)
	Population	Population Density	
		Women Population 10+	
		Men Population 10+	

		Population age 25-34	
		Population age 35-44	
		Population age 45-54	
		Population age 55-64	
		Population age 65+	
	Income	HH Income $15000−24999	
		HH Income $15000-24999	
		HH Income $25000-34999	
		HH Income $35000−49999	
		HH Income $50000−74999	
		HH Income $75000−99999	
		HH Income $100000−149999	
		HH Income $150000−199999	
		HH Income $200000+	
		HH Income less than $35000	
	Race	White Population	
		Black Population	

		Other Races	
	Marital status	Pop Age 15+ Widowed	
		Pop Age 15+ Married	
		Pop Age 15+ Never Married	
		Pop Age 15+ Divorced	
	Diversity	Diversity Index	ESRI 2018
	Education attainment	Education Pop Age 25+: < 9th Grade	ACS (2014-2018)
		Education: High School/No Diploma	
		Education: High School Diploma	
		Education: Some College/No Degree	
		Education: Associate Degree	
		Education: Bachelor's Degree	
		Education: Graduate Degree	
	Households	HHs w/No Retirement Income	ACS (2014-2018)
		Median Household Income	
		HHs: Inc Below Poverty Level	
		HHs w/Pop <18: Oth Fam/Fem HHr	

		HHs:Inc at/Above Poverty Level	
		HHs with Social Security Income	ESRI 2018
		HHs w/No Social Security Income	
		HHs w/Public Assist Income	
		HHs with Retirement Income	ACS (2014-2018)
		HHs w/Food Stamps/SNAP	ESRI 2018
		Owner Households	ACS (2014-2018)
		Renter Households	
		Owner HHs by Vehicles Avail: 0	
		Renter HHs by Vehicles Avail: 0	
		Median Year Householder Moved In	
	Insurance coverage	Pop <19: No Health Insurance	
	Urbanicity	Urban/Rural	USDA (RUCA Code)
	Housing	Housing Affordability Index	ESRI 2018
		Vacant Housing Units	
		Average Home Value	
		Housing: Mobile Homes	

Risk Factors	Smoking and Prescriptions	Smoking Accessories Average	ESRI 2018 SCDHEC
		Smoking Products Average	
		Smoked cigarette/vaporizer last 12 months	
		Smoked e-cigarette/vaporizer last 12 months	
		Smoked menthol cigarettes in last 12 months	
		Smoked non-menthol cigarettes in last 12 months	
		Drugs and Vitamins	
		Used prescription drug for anxiety/panic	
		Used prescription drug for backache/back pain	
		Used prescription drug for migraine headache	
		Used prescription drug for depression	
		Used prescription drug for sinus congestion/headache	
		Nonprescription Drugs	
		Prescription Drugs	
		Medicare Rx Drug Premium	
		Individuals with 1+ opioid prescription	SCDHEC
		Access to Tobacco Stores	

		Access to Liquor Stores	Facility location from SCDOR and ESRI 2018
	Crime	Total Crime Index	ESRI 2018
		Personal Crime Index	
		Property Crime Index	
Protective Resources	Accessibility measurements	Access to Parks	Facility location from ESRI 2018
		Access to Fitness	
		Access to Library	
		Access to Hospitals	Hospital location from SCDHEC
		Access to OTP	Facility location from SAMSHA 2019
		Access to Menta Health Facility	
		Access to Buprenorphine Providers	

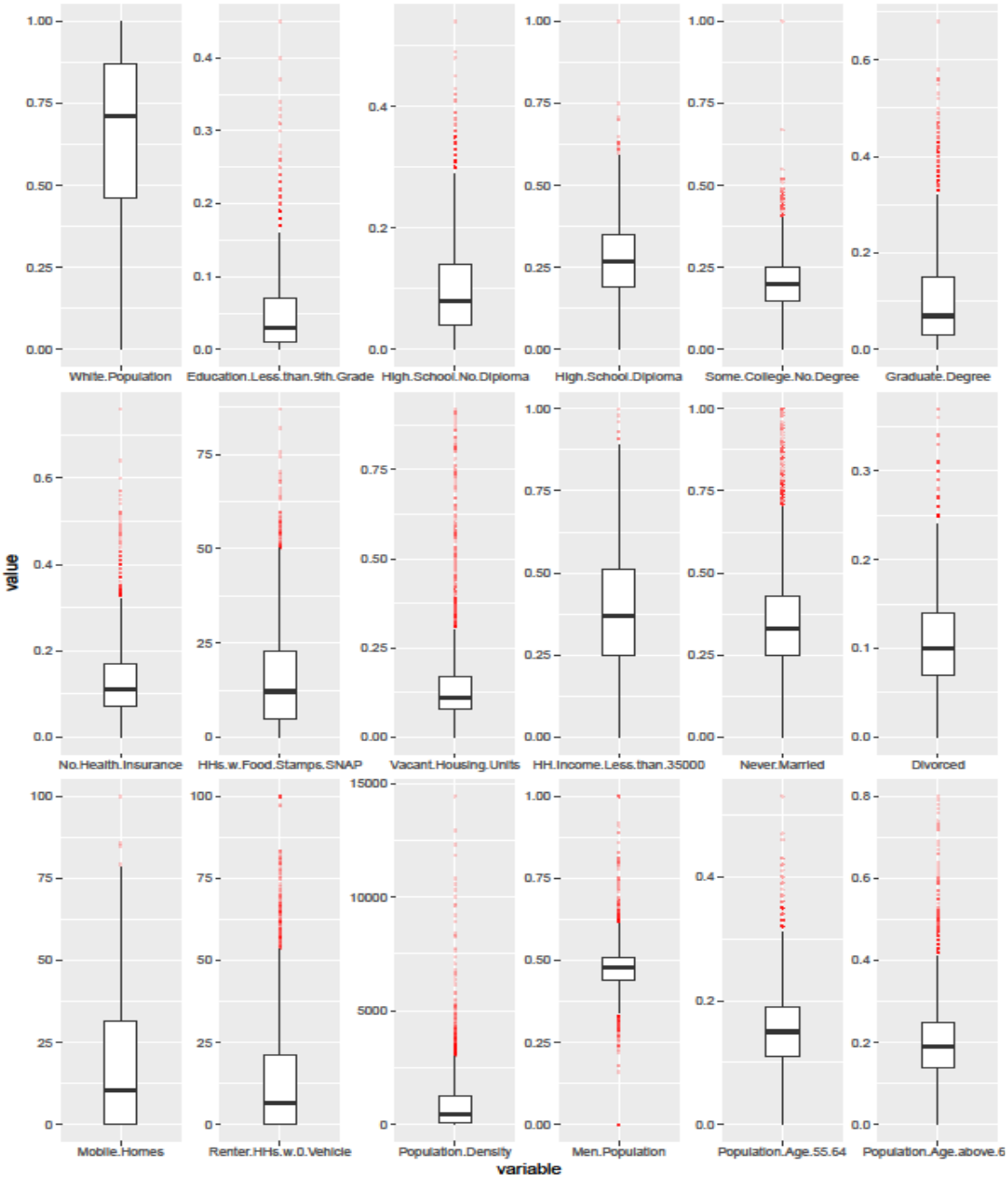

Figure 3.1. Descriptive statistics of the finalized socio-demographic variables

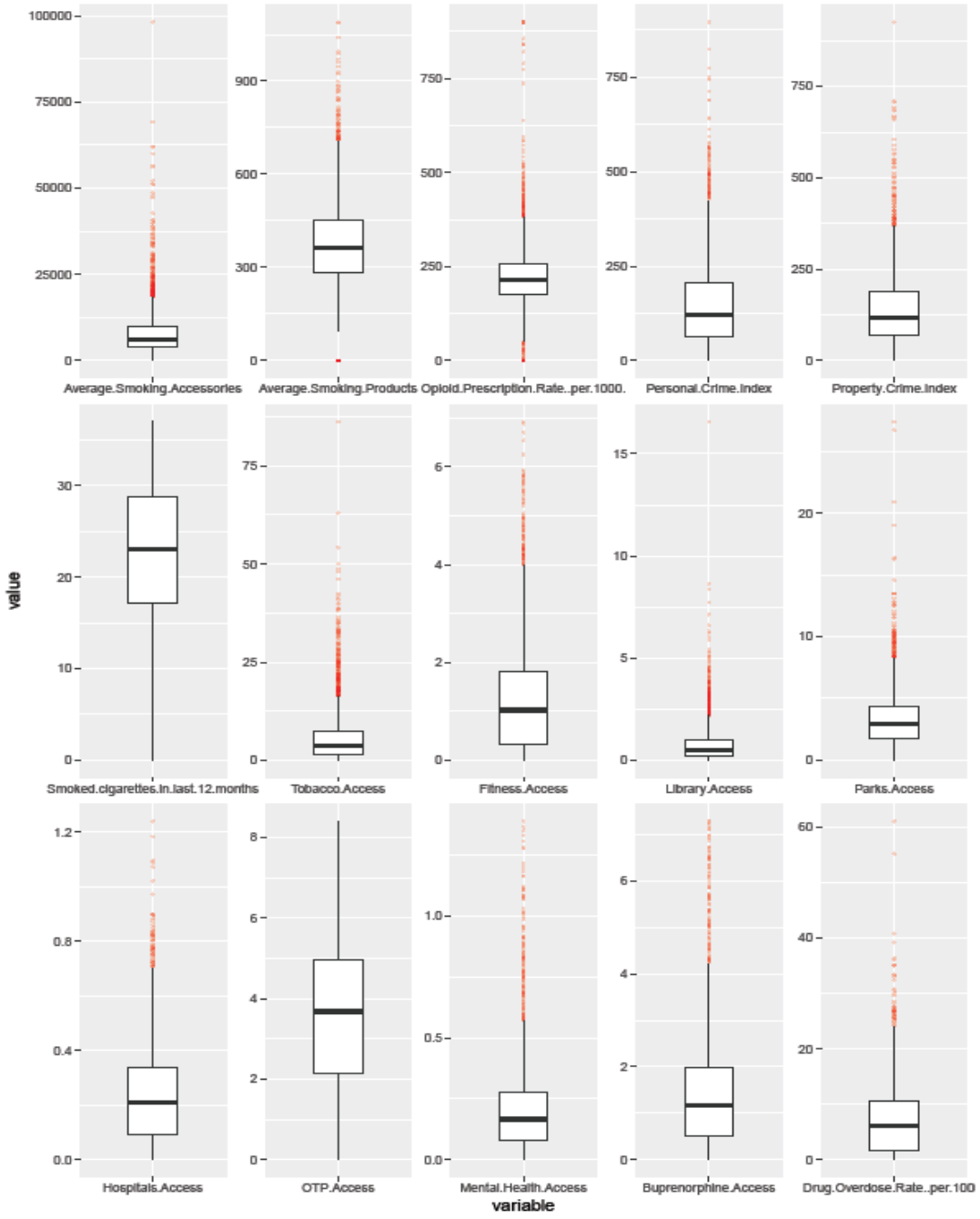

Figure 3.2. Descriptive statistics of the finalized drug risk factors and preventive sources variables

Spatial factors such as access to tobacco and liquor stores, parks, and libraries were calculated using Enhanced Two-Step Floating Catchment Area (E2SFCA) (W. Luo & Qi, 2009b). Accessibility to opioid treatment programs (OTPs), buprenorphine practitioners, mental health facilities, and hospitals were measured using our developed spatial accessibility model (see Chapter 2). Variables were divided into three domains (risk factors, protective factors, and socio-demographics) with the hypothesis that each domain would have a different relationship with a drug overdose.

The risk factor domain included access to liquor and tobacco stores, the rate of individuals who received at least one opioid prescription, smoking products expenditure, smoking behaviors, and crime data. The hypothesis was that areas with higher access to the liquor and tobacco stores and a higher rate of patients would experience a higher rate of overdose. The protective resource domain included accessibility measurements to facilities such as hospitals, libraries, parks, opioid treatment programs. The hypothesis was that areas with higher accessibility score would have a lower rate of drug overdose. The third category of data was related to the socio-demographic characteristics of the neighborhood. The relationship of each of these three domains with drug overdose events was explored in-depth in the following sections of this study.

3.3. METHODOLOGY

3.3.1. SPATIAL STATISTICS

The analysis of spatial data is complicated by a phenomenon known as spatial autocorrelation (SAC) that needs to be accounted for in machine learning approaches. Spatial autocorrelation occurs when the values of variables sampled at nearby locations are

not independent of each other (Tobler, 1970). To account for spatial autocorrelation, we first checked whether spatial autocorrelation was present in our data using Global Moran's *I*. The Global Moran's *I* values range approximately between −1 to 1. A Moran value near zero indicates no spatial pattern, no spatial autocorrelation (confirming the null hypothesis of spatial randomness). A negative spatial autocorrelation coefficient reflects neighboring areas with large inverse values–e.g., large values and small values are neighbors (i.e., dissimilarity). A positive spatial autocorrelation coefficient reflects neighboring areas with similarly high or low values (i.e., similarity). A pseudo p-value for the Global Moran's *I* was calculated via a Monte Carlo simulation consisting of 999 random replications.

To identify statistically significant clusters of overdoses in particular neighborhoods, local spatial variations were also examined. Two local measures of spatial association including Anselin Local Moran's *I* (i.e., LISA) and local Getis-Ord (Gi*) statistics were used to detect clusters or outliers and the most important type of spatial correlation. Anselin Local Moran's *I* was utilized to detect clusters and outliers of areas with extreme drug overdose values unexplained by random variation. Further, the Gi* statistic was applied to provide additional information indicating the intensity and stability of the hot spot and cold spot clusters. The Gi* statistic consists of a ratio of the weighted average of the values in the neighboring locations to the sum of all values, including the value at the location (x_i) (Equation 1). In contrast, the local Moran's *I* statistic includes only neighboring features and the value x_i is not included (Equation 2).

$$G_i^* = \frac{\Sigma_j w_{ij} x_i}{\Sigma_j x_i} \qquad \text{Equation 1}$$

$$I_i = \frac{\Sigma_j w_{ij} z_i z_j}{\Sigma_i z_i^2}$$ Equation 2

The statistical significance of a Z-score identifies the presence and intensity of local clusters of hot spots and cold spots of the event, relative to the hypothesis of spatial randomness. We quantify spatial relationships using K-Nearest Neighbor (KNN). To define number of neighbors (K), we followed a general rule of thumb which evaluates each neighbor in the context of a minimum of eight neighbors for hotspot analysis (ESRI, 2011).

3.3.2. MODELING PROCESS OVERVIEW

To create the spatial risk prediction model, we undertook a machine learning process to identify different types of predictive power from the variables. Machine learning is a field of computer science that uses computer algorithms to identify hidden patterns in datasets with a multitude of variables and can be used to predict various outcomes. Machine learning algorithms typically build a model from test inputs in order to make data-driven predictions or decisions. Machine learning can be divided into categories such as supervised and unsupervised (Osisanwo et al., 2017). In a supervised learning model, the algorithm learns on a labeled dataset. Supervised learning can be further categorized into classification and regression. Classification is predicting discrete class labels, while regression is the task of predicting a continuous quantity (Osisanwo et al., 2017). An unsupervised model, in contrast, provides unlabeled data that the algorithm discovers hidden patterns in data on its own. Unsupervised learning models are used for three main tasks: clustering, association and dimensionality reduction (Gentleman & Carey, 2008).

We conducted a feature selection process to recognize the best subset of variables that could provide better prediction performance. By identifying the best subset of

variables, we constructed different predictive models informed by supervised machine learning techniques. Several spatial risk prediction models for each domain were built. The predictive power of each of these domains was captured separately to explore how well each of these domains predicted overdose independently. All three models were also combined into a final ensemble predictive model. The predictive power of the three separate and the combined models was compared with each other. All models were trained and tested by the proportion of 80/20 percent of the data during each iteration process, which was repeated 1000 times. All models' parameters were determined using grid search approach with 5-fold cross-validation. We then conducted an evaluation process to validate the final model was generalizable across the different datasets and areas.

The following sections will outline the machine learning process of feature selection, model construction, and validation process. The systematic framework is shown in Figure 3.3. The diagram depicts the flow from raw data through the development of predictive models, and their evaluation towards identifying risk of drug overdose. Figure 3.3 shows the ensembling process that combines the output predictions of the three domains.

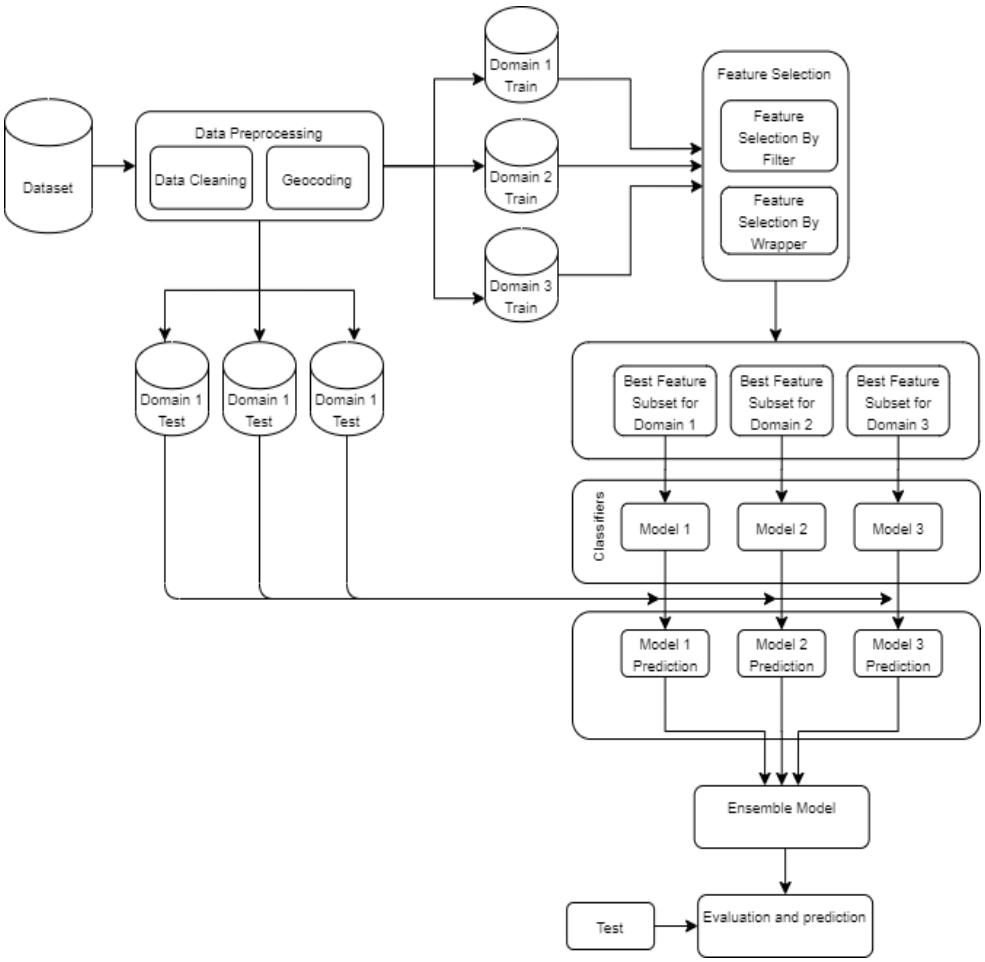

Figure 3.3. A machine learning-based framework to predict drug overdose

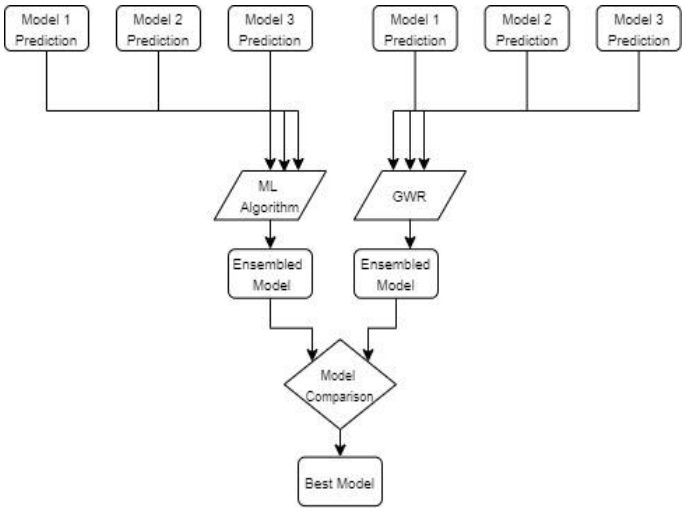

Figure 3.4. Ensembling process

3.3.3. DESIGN EXPERIMENTS

In designing experiments for the prediction model, a variety of well-established machine learning algorithms were used. These included the following algorithms: Linear Regression (LR), Sequential Minimal Optimization (SMO), Random Forest (RF), and Extra Gradient Boosting (XG-Boost). In the following paragraphs, the models used in this study are briefly described.

Linear Regression is a statistical model that finds the coefficients of the best fitting linear model in order to describe the relationship between a continuous dependent variable and one or more independent variables. Sequential Minimal Optimization (SMO) is based on nonlinear transformations of the variables into a higher-dimensional feature space (Vapnik, 2000).

Ensemble models synthesize the results of multiple learning algorithms to obtain better performance than individual algorithms and help decrease variance and bias and improve predictions. The ensemble models used in our study were random forests and gradient boosting. Random Forest is a tree-based ensemble model that develops multiple random decision trees through a bagging method (Ajit, 2016). The Random Forest algorithm works by generating a large number of independent classification or regression decision trees and then employing the majority of the vote (for classification) or averaging (for regression) to generate predictions. This reduces the drawback of the large variance in decision trees. Decision splits are made based on impurity and information gain. Extra Gradient Boosting (XGBoost) is also an ensemble prediction model based on decision trees (T. Chen & Guestrin, 2016).

Compared to other algorithms, XGBoost has higher interpretability, predictive accuracy, and computational speed (Ajit, 2016). In contrast to Random Forest, this model successively builds decision trees using gradient descent in order to minimize the error. A final prediction is made using a weighted majority vote of all of the decision trees (Ajit, 2016). Both Random Forest and XGBoost are robust against outliers. Because of bootstrap sampling, outliers appear in individual trees less often, and therefore, their influence is curtailed (Ajit, 2016). They can also recognize non-linear relationships in data, which is useful when modeling spatial relationships. They are not affected by co-linearity in the data. This is highly valuable as socio-demographic data can be highly correlated.

We split each category's dataset into training (80%) and test (20%) datasets. We performed a grid-search approach to tune parameters for each algorithm. The most important parameters for XGBoost are the number of trees (nrounds), the learning rate (eta), and the depth of each tree (depth). These parameters control the complexity and the fitness of the model. The rest of the parameters are complementary and help to avoid situations of overfitting and underfitting. For RF, an appropriate number of trees (ntrees) and the number of randomly selected predictors at each tree node (mtry) were specified. For SMO, a regularization or complexity parameter (C) and the radial kernel search parameter (gamma) that minimize cross-validation error were selected.

We then implemented our modeling process in two steps: feature selection and model building. Feature selection is one of the critical steps in the development of a prediction model, which aims at eliminating less important variables without losing much of the total information (Bagherzadeh-Khiabani et al., 2016). It is desirable to reduce the

number of input variables to reduce the computational cost of modeling, risk of model overfitting, training time, and to improve the performance of the model.

There are two main types of feature selection techniques: filter and wrapper methods (Liu et al., 2010). The filter method ranks the feature subset based on the correlation between the outcome variable and independent variables. Subsets that show high correlation with the outcome variable and less correlation with independent variables will be ranked at a higher value. This method doesn't involve any machine learning algorithm (Liu et al., 2010). Wrapper methods use a specific learning algorithm to select features. The method utilizes a search procedure in the space of possible features and then generates and evaluates various subsets in order to find the best one (Sánchez-Maroño et al., 2007).

We used the most common feature selection methods: Filter (i.e., correlation-based feature selection (CFS)) and wrapper subset evaluation methods to measure the effect of different feature selection methods on the model performance. To select the best subset of features in each category, each dataset was fed into each machine learning algorithm separately. The performance of the various algorithms was then investigated using the paired T-test to determine whether the performance was statistically significantly different among the algorithms. We considered statistical significance at the confidence level of 95%, associated with a p-value < 0.05. Additionally, we computed the SHapley Additive exPlanation (SHAP) to rank the features. SHAP is an additive feature attribution method, in which each prediction is explained by the contribution of the features of the dataset to the model output (Lim & Chi, 2019). More specifically, SHAP approximate Shapley values, an idea from game theory that is the solution for the problem of computing the

contribution to a model's prediction of every subset of features given a dataset with n features (Lim & Chi, 2019).

To build a model for each category, the model was trained on the training dataset using the best features. The trained model from each algorithm was then used to predict on the 20% test dataset. Further, an ensemble model was created by combining three models–each model derived from each category–into a robust fused prediction model aimed at reducing the overall error. To build the ensemble model, we used two approaches including, machine learning and geographic weighted regression (GWR) (Brunsdon et al., 1996). GWR is a local spatial statistical technique that assumes non-stationarity in relationships. That is the relationships between the dependent variable and the explanatory variables changes from location to location. GWR, unlike global statistics, generates an equation for every component (i.e., area) in the dataset by calibrating each one using the target feature and its neighbors. In this respect, nearby features produce a higher weight in the calibration than distant features. The prediction of three individual models was served as independent variables, and drug overdose rate served as a dependent variable. For GWR, an adaptive bi-square kernel type and a KNN search were used for bandwidth selection.

We incorporated spatial dependency into ML models using neighborhood matrices, which specify the relationship between each data location and those at a neighboring location. The neighborhood can be identified by the adjacency of block groups that share a common border, a distance-based weight matrix, or a specific number of neighbors. Regarding the first two approaches, since block group polygons are of widely varying sizes, there will be problems with the distribution of the neighbor cardinalities. In addition, there will be a potential problem with isolates when using a distance-based weight matrix. Thus,

to have the same number of neighbors for each location and avoid the problem of isolates, we defined the neighborhood relationship using KNN. To find a suitable number (k) of nearest neighbors, different k values range from 8 to 46 were examined, and the corresponding estimation errors were obtained. We followed the rule of thumb suggested by ESRI (ESRI, 2011) to determine the min/max for k. The k=35 that resulted in the minimum error was selected.

The machine learning experiments were implemented in the open-source Waikato Environment for Knowledge Analysis platform and Python 3.6 with computing libraries, which included Numpy 1.15.4, Pandas, Scikitlearn, and XGBoost. Spatial analysis was performed in GeoDa 1.10.0.8 (University of Chicago, n.d.) and ArcGIS Pro 2.6 (ESRI, 2020).

3.3.4. MODEL VALIDATION

We validated the model through several measurements. First, we assessed the performance of various algorithms in terms of Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). RMSE is a common measurement of the differences between regression model predicted values and observed values. It is formally defined as $RMSE = \sqrt{\frac{\sum_{j=1}^{n} (\hat{y}_j - y_j)^2}{n}}$, where $\hat{y}$ represents the prediction, and y represents the observed value at observation n. Lower RMSE scores are typically more desirable. An RMSE value of 0 would indicate a perfect fit for the data. RMSE can be difficult to interpret on its own; however, it is useful for comparing models with similar outcome variables. In our case, the outcome variables (drug overdose rate) are consistent across modeling datasets, and therefore can be reasonably compared using RMSE. MAE measures the average magnitude

of the errors in a set of predictions without considering their direction. It's the average over the test sample of the absolute differences between prediction and actual observation where all individual differences have equal weight. It is defined as $MAE = \frac{1}{n}\sum_{J=1}^{n}\left|\hat{y}_j - y_j\right|$, where $\hat{y}$ represents the prediction, and y represents the observed value at observation n. In order to compare algorithms, we established a zero-rule algorithm as a baseline by which to compare all evaluated algorithms. The zero-rule algorithm predicts the mean of the training dataset.

Second, we conducted both normal and spatial cross-validation. The goal of normal cross-validation is to test the model's ability to predict new data that was not used in estimating it in order to flag problems like overfitting or selection bias. It helps to verify that the model is generalizable across different subsets of the data, not just the initial test set. Normal cross-validation is based on partitioning the set of observations into equally sized subsets to train the classifier on all but one of these subsets and test it on the remaining one. We conducted n-fold cross-validation tests to ensure that the model is generalizable. This means that the initial data set was divided into n equal subsets, with n-1 subsets used to train the model and the remaining subsets used to test the model; this partitioning was repeated n times (folds). We then average errors measured on these test data sets (RMSE and MAE in our case) across n folds. The preferred number of folds in n-fold is suggested to be between 5 and 10 (Hastie et al., 2009; Kohavi, 1995). For this study, we implemented 5-fold cross-validation.

Cross-validation assumes that (pairs of) observations in different subsets of the partition are independent. In a spatial context, spatial autocorrelation causes the normal

random cross-validation techniques to underestimate the prediction error. When test data is randomly selected for cross-validation from the entire spatial domain, training and test data from nearby locations will be dependent (spatial autocorrelation). Consequently, if the objective is to predict outside the spatial structure of the training data, error estimates from random cross-validations will be overly optimistic. To provide a useful estimate for our model prediction performance without optimistic bias due to SAC, we performed spatial cross-validation. This effectively forces testing on more spatially distant features, thus decreasing spatial dependence and reducing optimism in error estimates (Trachsel & Telford, 2016). To implement the spatial cross-validation, we did not divide the data into subsets randomly, but instead, we spatially divided training and test datasets. A spatially segregated hold-out prevents spatial dependency between training and test datasets and thus makes the two datasets to be more likely independent (Townsend Peterson et al., 2007). Our test dataset included all block groups within a county and a county's immediate block group neighbors, and the training dataset included the rest of the block groups. This was done in an iterative process splitting was done until all 46 counties had been test dataset. The error was then averaged over the splits.

Lastly, we assessed the model residuals using Global Moran's *I*. The technique tests the model's spatial autocorrelation by calculating the residual Moran's *I*, where 0 indicates the weakest spatial autocorrelation model and the p-value > 0.05 represents no significant spatial autocorrelation exists.

3.4. RESULTS

3.4.1. SPATIAL STATISTICS

The result of Global Moran's *I* showed the presence of statistically significant (p-value < 0.0001) positive spatial autocorrelation (Global Moran's $I = 0.22$) in drug overdose, confirming the presence of spatial clustering (Figure 3.5). Significant clusters of block groups with high (hot spots) and low (cold spots) overdose rates, as assessed by the Getis-Ord Gi* tool are shown in Figure 3.6. Cold spots are mainly located in lower Midlands region, while hot spots located in the northeast. The Figure 3.7 shows the locations with significant local Gi* for various p-values. Anselin Local Moran's *I* confirmed the significant hot and cold spots identified by the Getis-Ord Gi * tool. The Anselin Local Moran's *I* showed core clustering of high drug overdose block groups next to high ones (HH) consistently located in the northeast and north of the state (Figure 3.8 and 3.4). The analysis also showed a core cold spot (L) located in the lower Midlands region, on the bottom of the I-95 corridor. Statistically significant spatial outliers (HL, LH clustering) were evident in the central and western parts of the state.

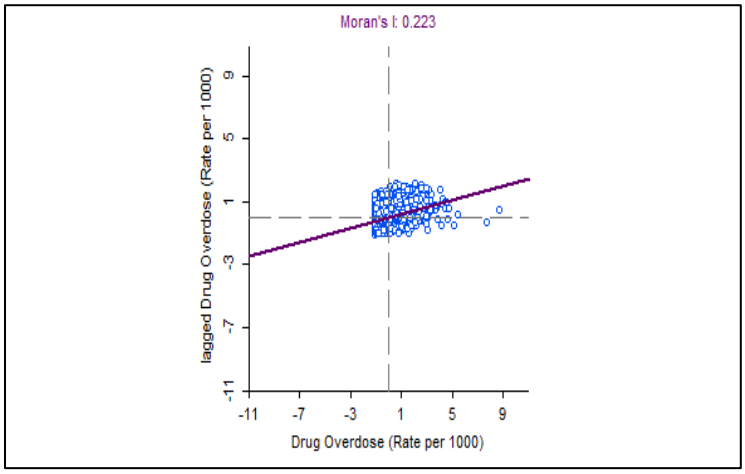

Figure 3.5. Global Moran's I statistic for drug overdose

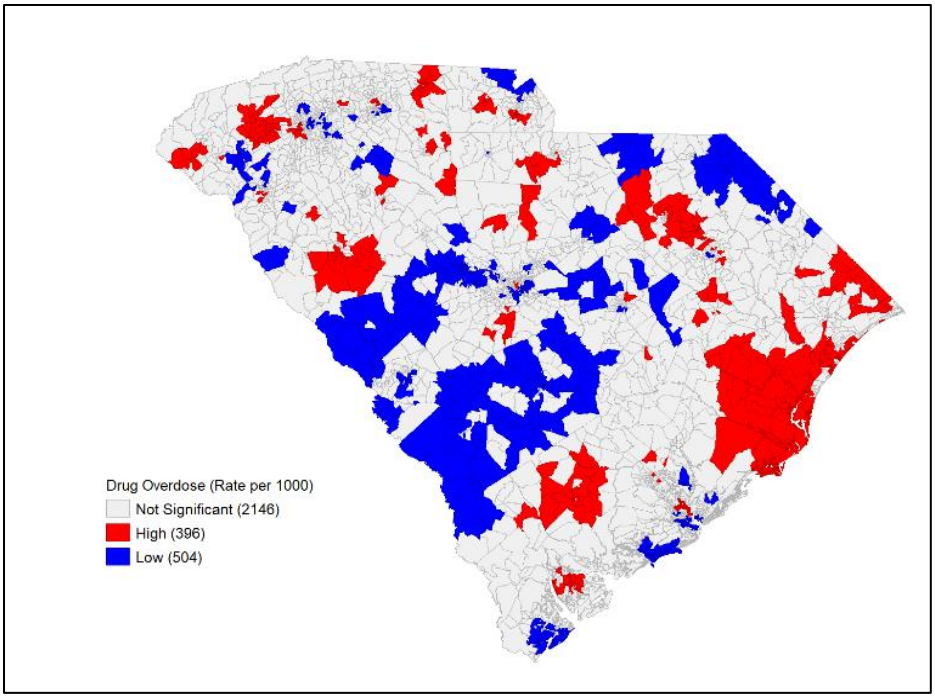

Figure 3.6. Gi* statistic cluster map of the drug overdose

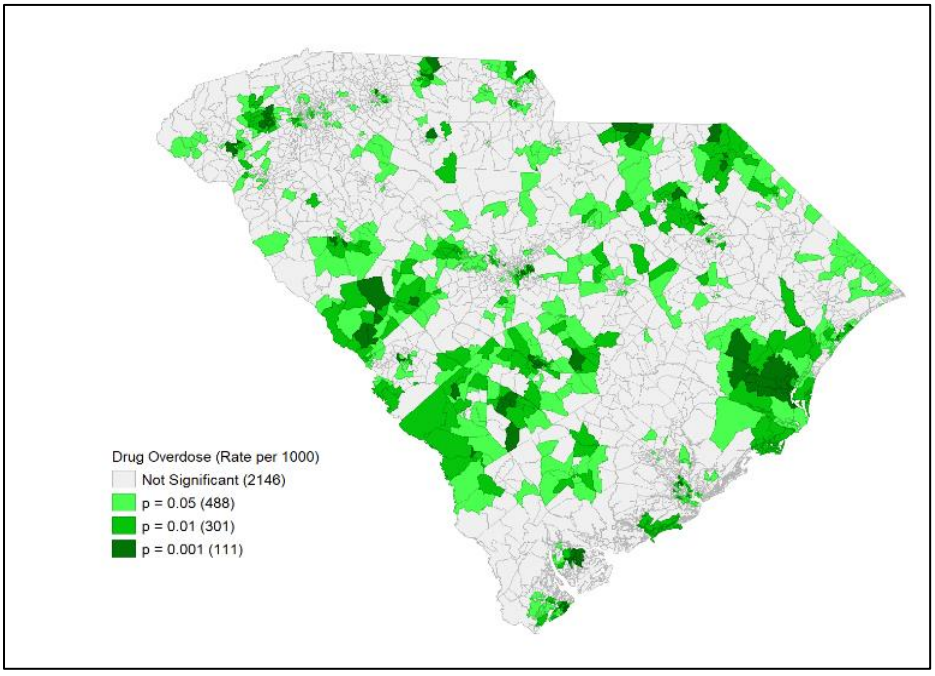

Figure 3.7. Gi* statistical significance map of the drug overdose

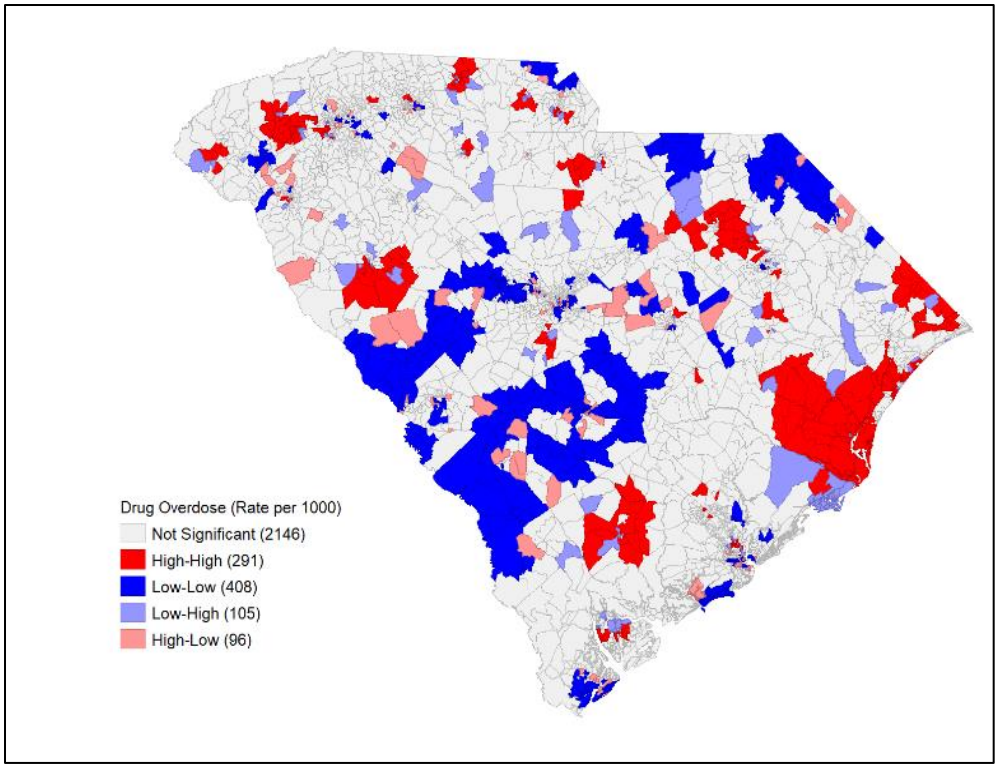

Figure 3.8. Local Moran's I cluster map of the drug overdose

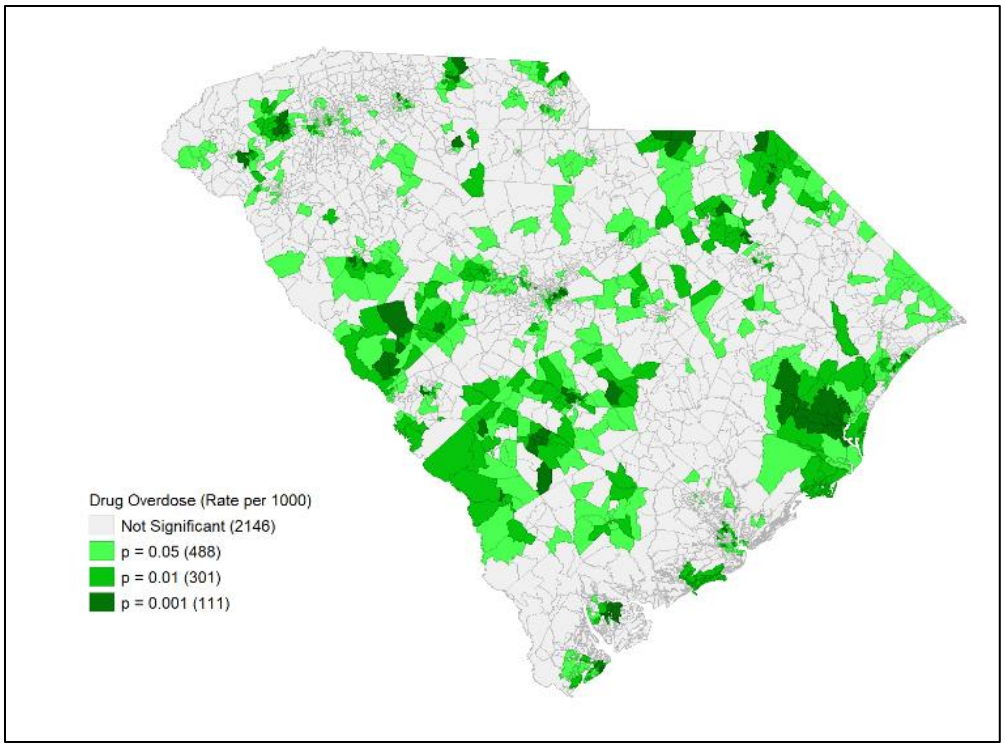

Figure 3.9. Local Moran's I statistical significance map of the drug overdose

3.4.2. PREDICTION PERFORMANCE OF MACHINE LEARNING ALGORITHMS

All the experiments were performed on each domain by splitting each into 80% training and 20% testing datasets. Tables 3.2 through 3.4 describes the comparative evaluation scores of different models across different feature selection methods after hyperparameter optimization. The feature selection methods mentioned earlier provided a different set of important features. However, there were features that were presented as the most important ones by all algorithms. The most important features were then ranked using SHAP. To determine the performance of different models on a varying number of variables, the MAE and RMSE were calculated. The higher the MAE and RMSE the model had, the worse was the performance of the model. In the risk factor category, the lowest error was obtained by the XGBoost method using wrapper RF feature subset selection. Wrapper LR algorithm in feature selection yielded the most accurate subset of features for the XGBoost method in the socio-demographic category. Regarding the protective resource category, the wrapper RF algorithm for the XGBoost method yielded the most accurate subset of features.

Table 3.2. Performance metrics of ML algorithms over different feature selection methods (risk factor domain)

Attribute evaluator		RF			LR			SMO			XGBoost			ZeroR	
		R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	MAE	RMSE
CFS		0.28	4.68	6.16	0.30	4.64	5..86	0.34	4.63	5.89	0.29	3.91	5.07	5.01	6.37
Wrapper	LR	0.38	4.28	5.84	0.40	4.45	5.73	0.4	4.43	5.76	0.42	3.61	4.83	5.01	6.37
Wrapper	RF	0.41	4.28	5.69	0.38	4.51	5.79	0.38	4.47	5.84	**0.44**	**3.18**	**4.18**	5.01	6.37
Wrapper	SMO	0.39	4.28	5.80	0.40	4.45	5.73	0.41	4.42	5.75	0.42	3.29	4.28	5.01	6.37
Normalized		0.39	4.25	5.80	0.40	4.45	5.73	0.40	4.41	5.75	0.41	3.36	4.29	5.01	6.37
Standardized		0.39	4.25	5.80	0.40	4.45	5.73	0.40	4.41	5.75	0.41	3.36	4.29	5.01	6.37

Table 3.3. Performance metrics of ML algorithms over different feature selection methods (socio-demographic domain)

Attribute evaluator		RF			LR			SMO			XGBoost			ZeroR	
		R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	MAE	RMSE
CFS		0.28	4.68	6.04	0.31	4.63	5..95	0.30	4.59	5.98	0.28	4.60	5.87	5.01	6.37
Wrapper	LR	0.32	4.57	5.92	0.35	4.54	5.86	0.33	4.51	5.90	**0.38**	**3.41**	**4.52**	5.01	6.37
Wrapper	RF	0.25	4.93	6.24	0.22	4.92	6.29	0.21	4.92	6.28	0.31	3.85	4.92	5.01	6.37
Wrapper	SMO	0.32	4.57	5.92	0.35	4.54	5.87	0.34	4.51	5.88	0.32	3.51	4.78	5.01	6.37
Normalized		0.32	4.58	5.92	0.32	4.57	5.92	0.31	4.53	5.97	0.28	3.36	4.90	5.01	6.37
Standardized		0.31	4.57	5.92	0.33	4.55	5.90	0.31	4.53	5.97	0.28	3.36	4.90	5.01	6.37

Table 3.4. Performance metrics of ML algorithms over different feature selection methods (protective resource domain)

Attribute evaluator		RF			LR			SMO			XGBoost			ZeroR	
		R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	R2	MAE	RMSE	MAE	RMSE
CFS		0.10	4.78	6.23	0.09	4.93	6.23	0.06	4.90	6.32	0.10	4.60	6.22	5.01	6.37
Wrapper	LR	0.11	4.72	6.12	0.11	4.89	6.22	0.08	4.84	6.32	0.12	4.56	5.91	5.01	6.37
Wrapper	RF	0.12	4.60	6.01	0.09	4.91	6.23	0.08	4.87	6.31	**0.12**	**4.41**	**5.56**	5.01	6.37
Wrapper	SMO	0.12	4.92	6.43	0.10	4.92	6.23	0.08	4.91	6.27	0.13	4.74	5.80	5.01	6.37
Normalized		0.10	4.61	6.02	0.11	4.89	6.22	0.08	4.84	6.33	0.13	4.67	5.82	5.01	6.37
Standardized		0.10	4.58	5.98	0.11	4.89	6.22	0.08	4.84	6.33	0.13	4.67	5.82	5.01	6.37

Regarding the important features, in the risk factors category, the opioid prescription rate provided the strongest predictive power in all the selection methods. Access to liquor and tobacco retail stores were selected as important variables through all feature selection experiments. In the final model, the opioid prescription rate and average smoking accessories expenditures provided the strongest predictive power. In the socio-demographic domain, population with income less than $35,000, percent widowed, percent divorced, and population density were selected as important features throughout all feature selection experiments. Furthermore, urban/rural status, unemployment rate, and vacant housing units were selected by the RF and SMO algorithms. In the final model, households with food stamps provided the strongest predictive power, followed by a population with income less than $35,000 and population density. Access to OTP facilities calculated by weighted 2SFCA and access to fitness were selected as important variables in all the selection methods in the protective resource category. In the final model, access to OTP facilities provided the strongest predictive power within this category of variables, followed by access to the hospitals. Figures 3.10, 3.11 and 3.12 show the variables included in each model in order of their relative predictive importance in the model. The x-axis is essentially the average magnitude change in model output calculated by SHAP values. The XGBoost model's parameters tuning experiments for each domain and ensemble model are shown through Figures 3.13 – 3.20. The final optimized parameters for each domain are depicted in Table 3.5.

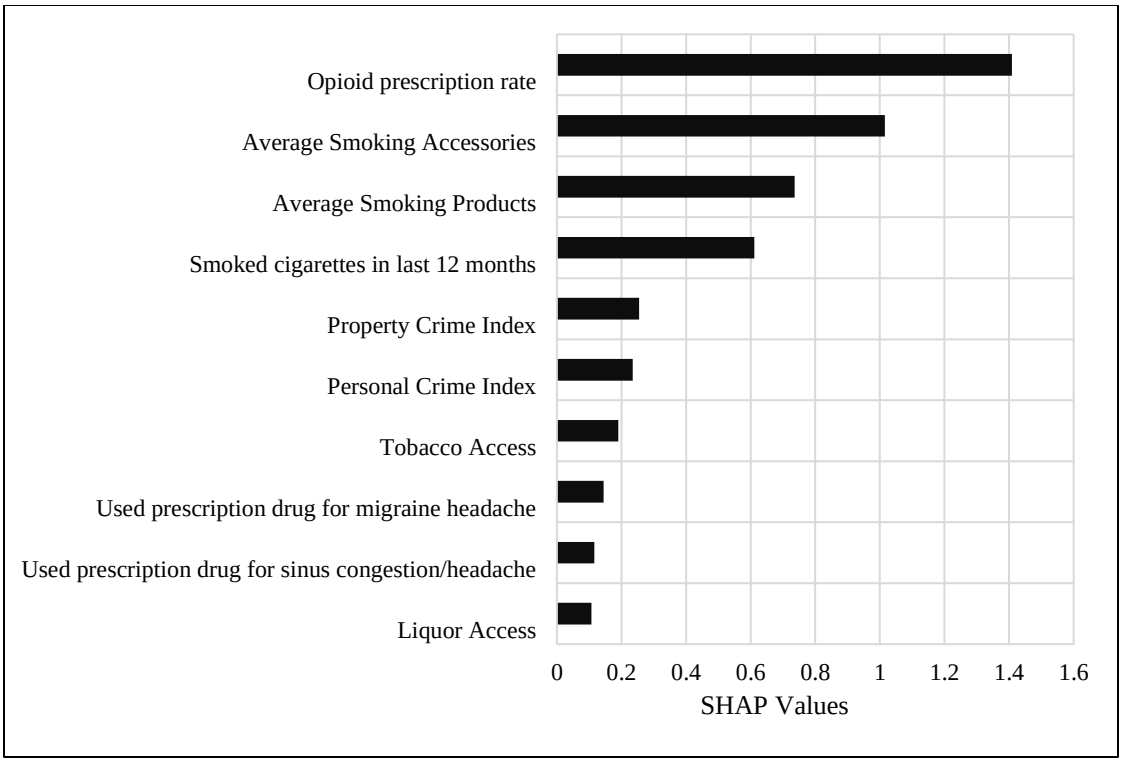

Figure 3.10. Important predictors in risk factors domain

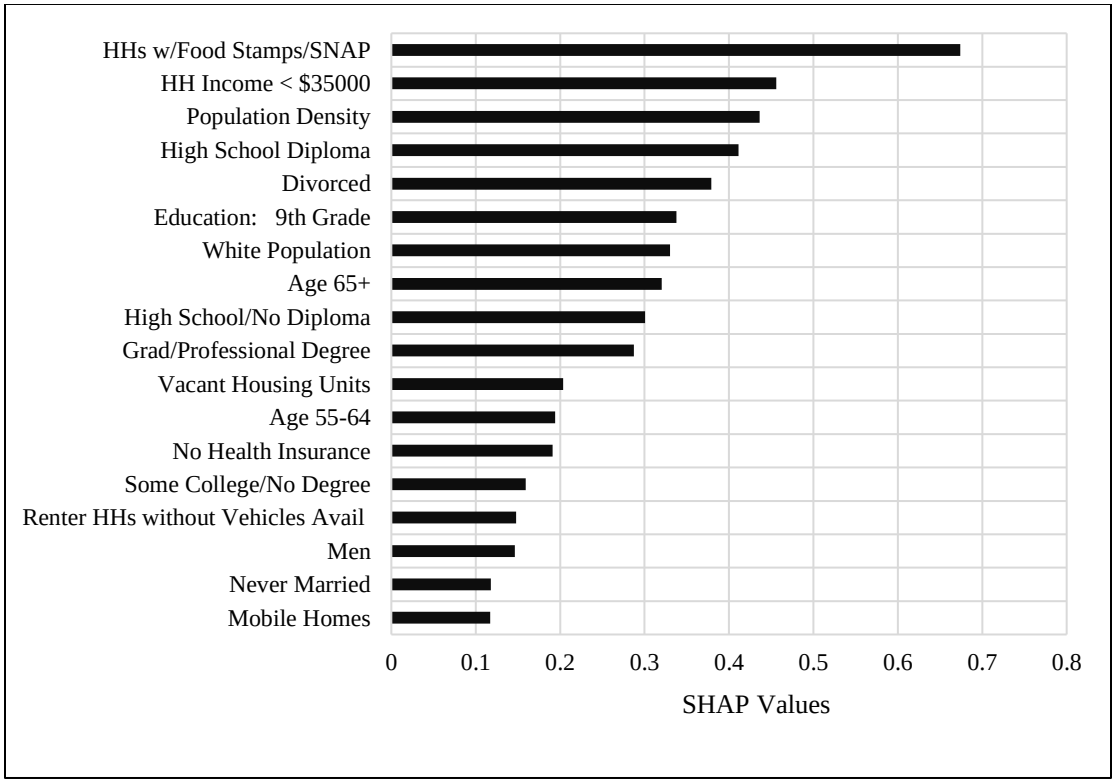

Figure 3.11. Important predictors in socio-demographic domain

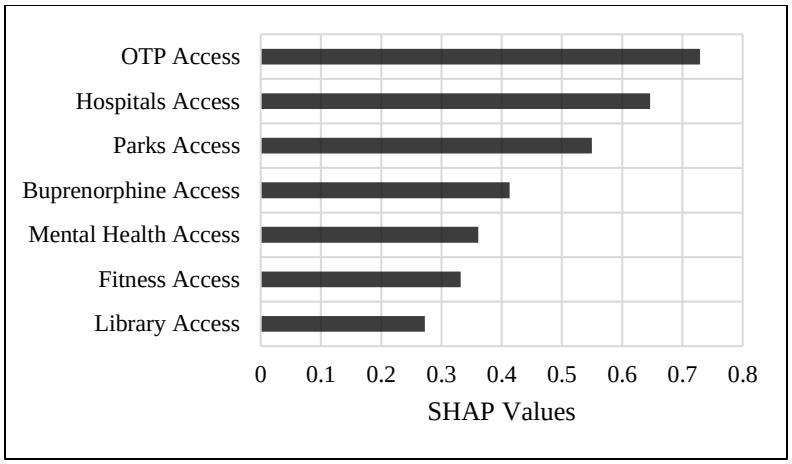

Figure 3.12. Important predictors in protective resource domain

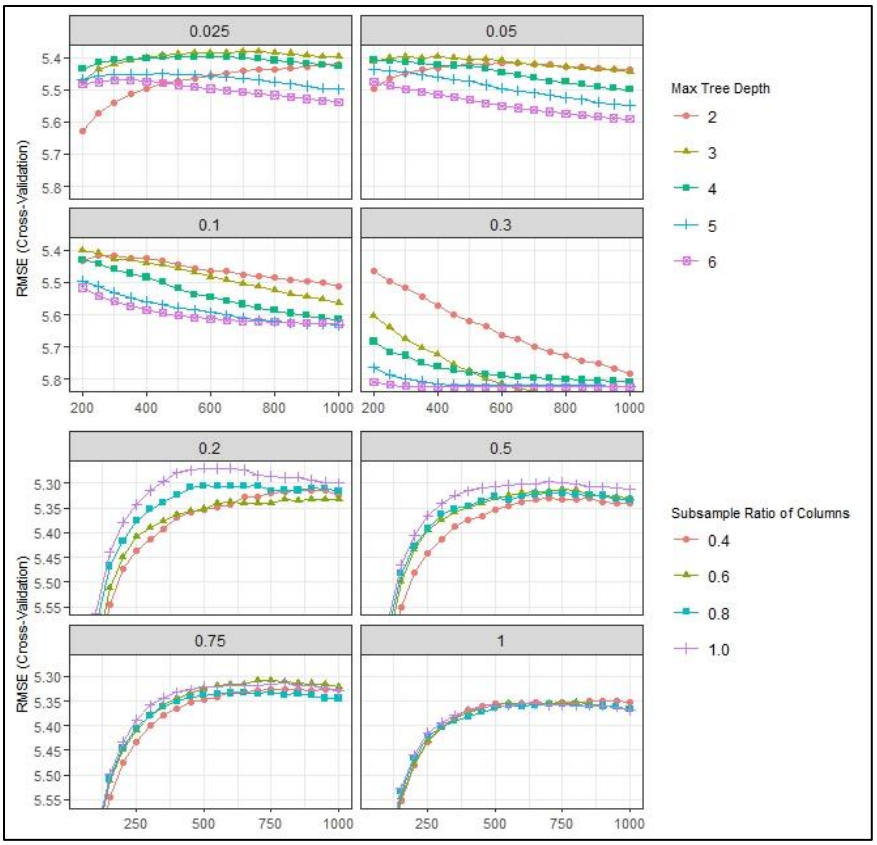

Figure 3.13. Tune max tree depth, subsample, and colsample parameters (risk factors)

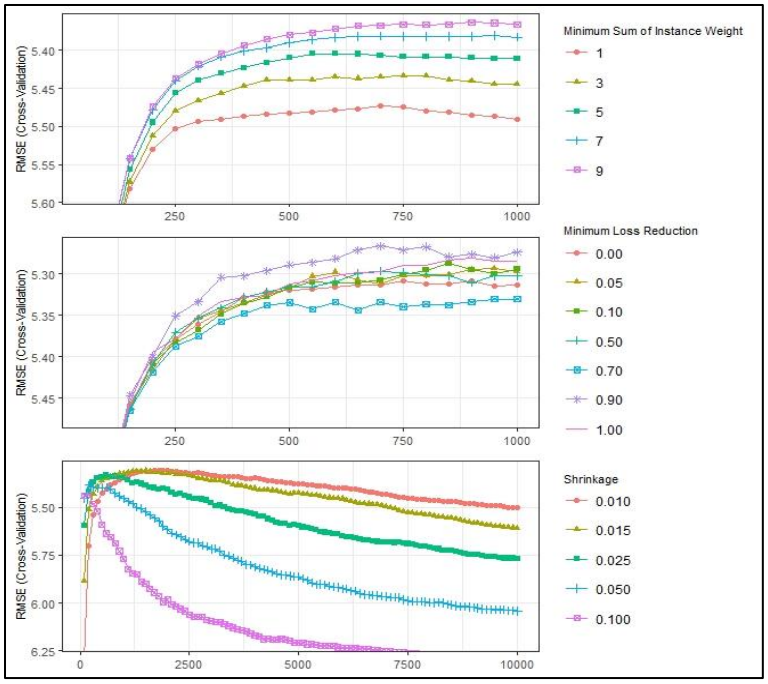

Figure 3.14. Tune child weight, gamma and eta parameters (risk factors)

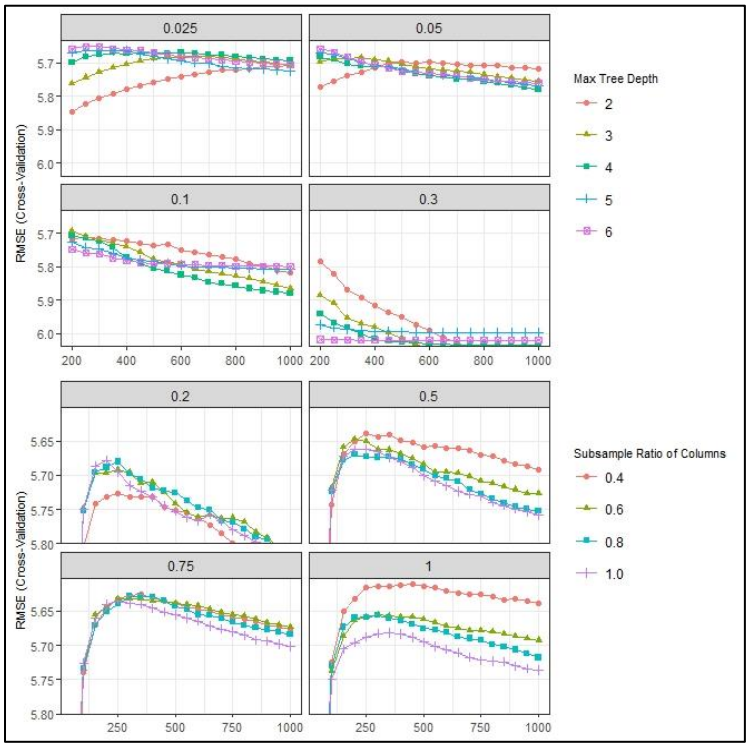

Figure 3.15. Tune max tree depth, subsample, and colsample parameters (socio-demographic)

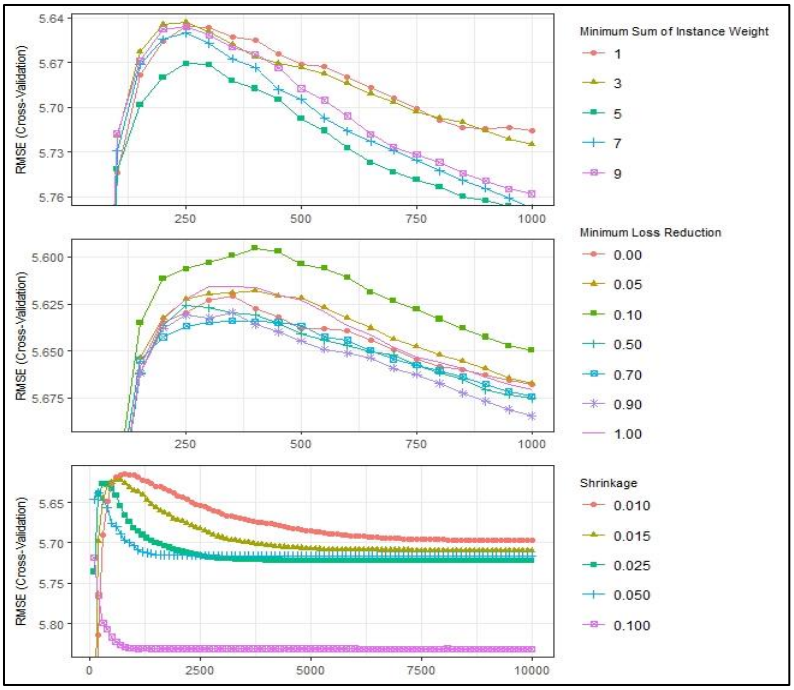

Figure 3.16. Tune child weight, gamma and eta parameters (socio-demographic)

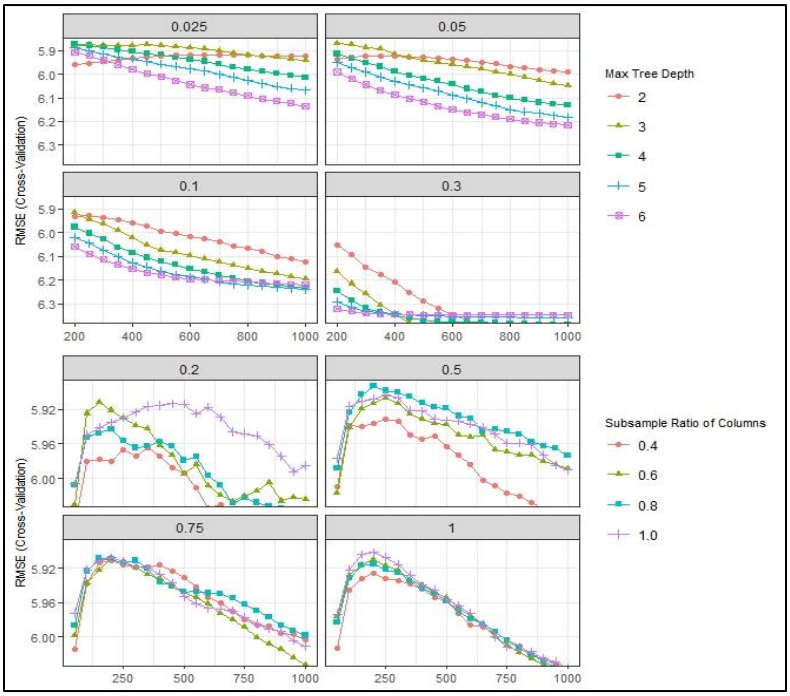

Figure 3.17**.** Tune max tree depth, subsample, and colsample parameters ((protective resources)

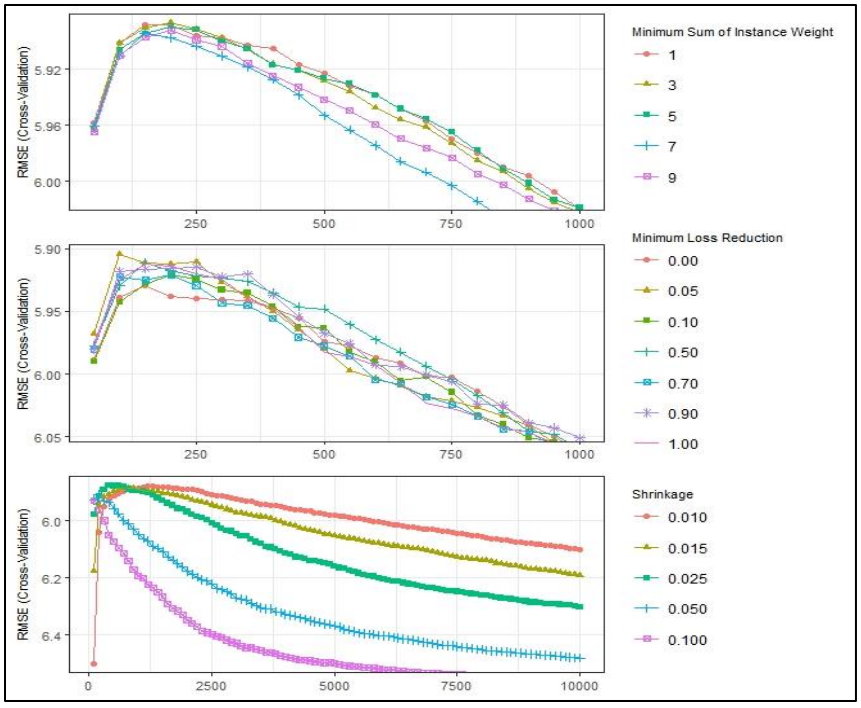

Figure 3.18. Tune child weight, gamma and eta parameters (protective resources)

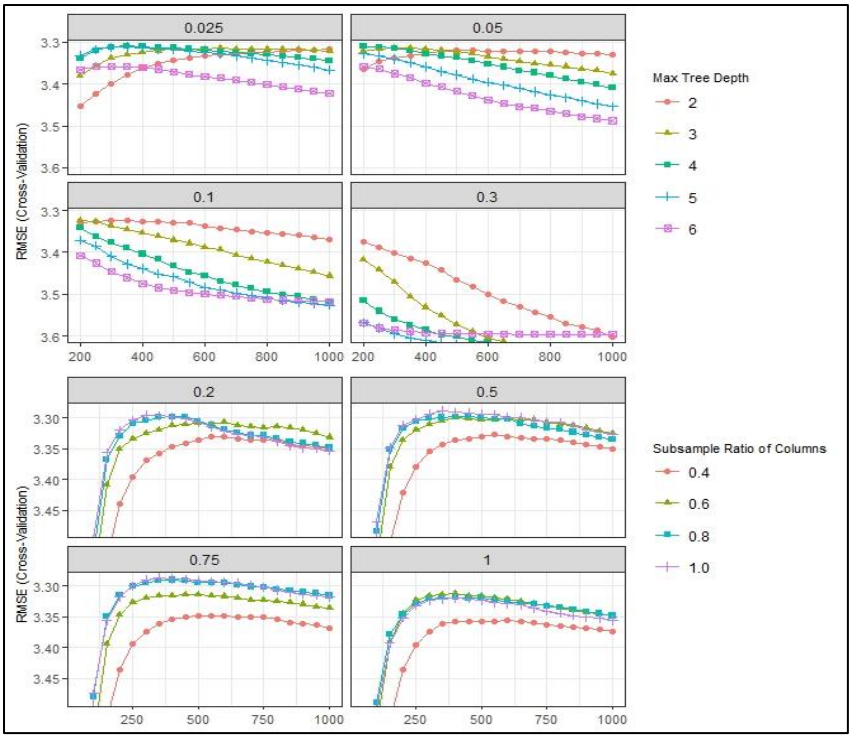

Figure 3.19. Tune max tree depth, subsample, and colsample parameters (ensemble model)

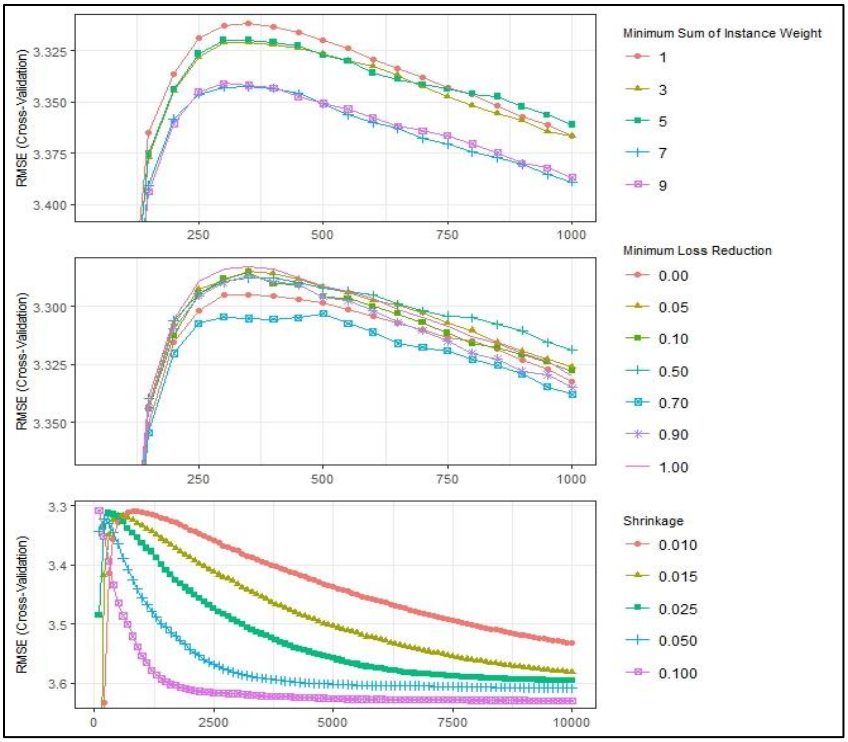

Figure 3.20. Tune child weight, gamma and eta parameters (ensemble model)

Table 3.5. XGBoost Hyper-parameter values extracted with the use of grid search and parallel processing.

Domain	***Parameters Name***	***Optimized Value***
Risk Factors Domain	nrounds	1800
	eta	0.01
	gamma	0.9
	depth	3
	min child weight	9
	subsample	0.2
	column sample	1
	lambda	2
	alpha	1

Socio- demographic Domain	nrounds	800
	eta	0.01
	gamma	0.1
	depth	6
	min child weight	3
	subsample	1
	column sample	0.4
	lambda	2
	alpha	1
Protective Resource Domain	nrounds	400
	eta	0.02
	gamma	0.05
	depth	3
	min child weight	3
	subsample	0.5
	column sample	0.8
	lambda	2
	alpha	1
Ensemble Model	nrounds	100
	eta	0.1
	gamma	1
	depth	4
	min child weight	1
	subsample	0.75

	column sample	1
	lambda	2
	alpha	1

We ensembled the three aforementioned domains (socio-demographic, risk, and protective resources) to one final model using the XGBoost and GWR. The prediction results from each of the domains served as independent variables with the aim of reducing the model's overall error. The prediction range and R-squared for each of the individuals and ensemble models are provided in Table 3.6. The results showed that the ensemble model by XGBoost achieved higher R-squared improved by 0.46-0.62 over individual models. The results also indicated that ensemble machine learning by XGBoost outperformed GWR. The Figure 3.4 shows how correlated the predictions were with the observed values. The plots indicate that all the models perform better when predicting lower values. Moving from the left to the right along the x-axis (from lower to higher values), the predicted and observed values become less correlated with each other. The actual drug overdose and prediction maps for each domain and ensemble model is shown in Figure 3.22. In the prediction maps, it is clear how the variables in each model strongly informed its predictions. As shown in the map, the risk factors prediction map shows more accuracy than the other two categories, particularly in northeast, such as block groups in Georgetown and Horry Counties. Protective factor and socio-demographic categories also predicted well in Georgetown County.

Table 3.6. Descriptive statistics of the prediction models

	Socio-demographic Model	Risk Factor Model	Protective Resource Model	Ensemble Model by XGBoost	Ensemble Model by GWR	Drug Overdose Rate (per 1000)
Mean	6.998	6.986	6.888	6.994	6.188	6.992
Std. Deviation	3.308	3.499	2.247	5.443	5.698	6.266
Minimum	0.017	0.016	1.403	0.001	0	0
Maximum	34.751	26.968	18.678	53.73	68.012	61.111
R2	0.19	0.27	0.11	0.73	0.71	-

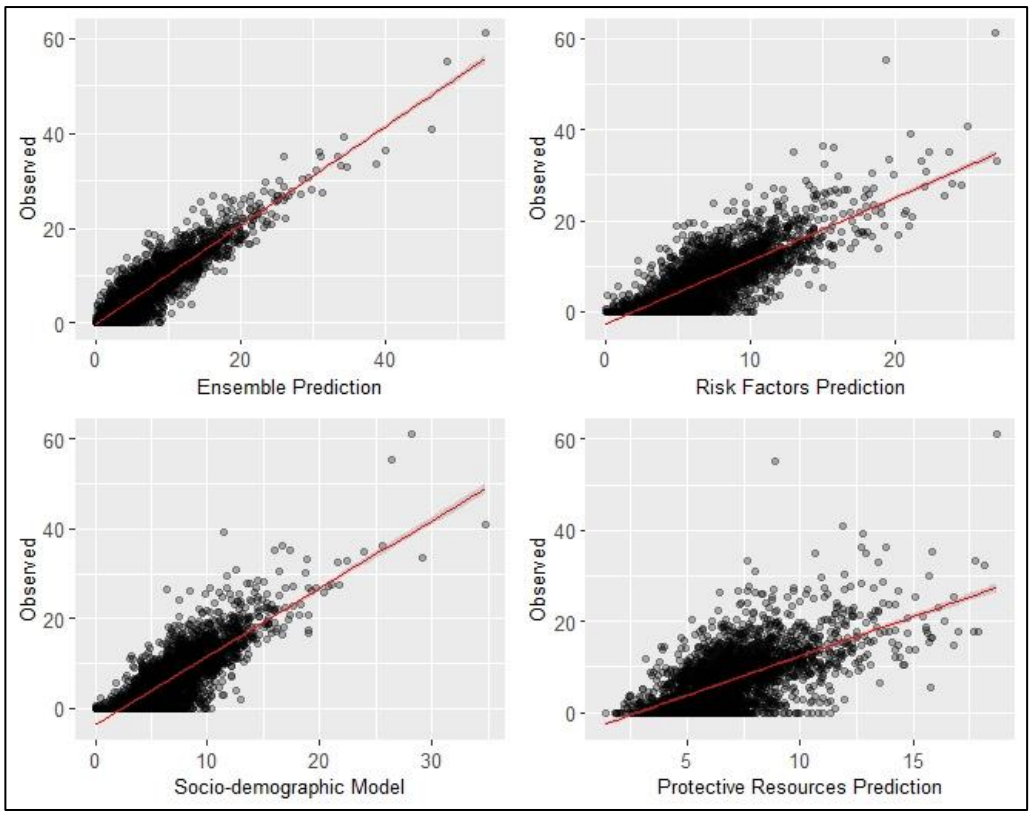

Figure 3.21. Scatter plots of predicted versus observed drug overdose values

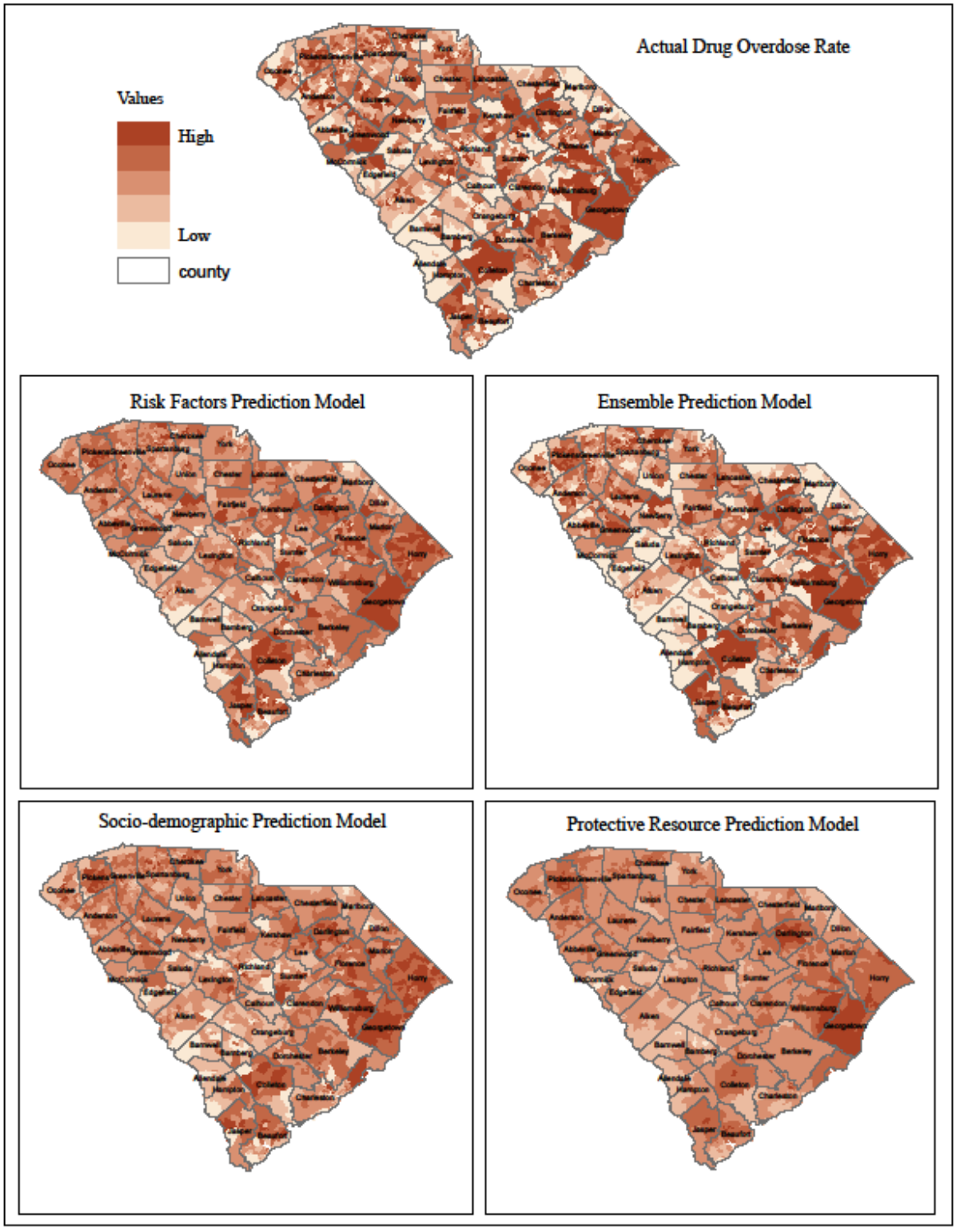

Figure 3.22. Map of actual drug overdose (top map) and prediction for each individual model and ensemble model at block group level

3.4.3. MODEL EVALUATION AND VALIDATION

The RMSE of prediction using 5-fold cross-validation is shown in Figure 3.23. The results suggested that the model was robust and predicted well for random subsamples. The MAE and RMSE for the risk factors model were 3.18 and 4.18, respectively. The MAE and RMSE for the socio-demographic model were 3.56 and 4.61, respectively, which was statistically significantly slightly higher than risk factors' model error at the significant level of 0.05. The protective factors model showed an MAE of 4.34 and RMSE of 5.46, statistically significantly higher than the socio-demographic and risk factor models' errors at the significant level of p-value = 0.05. The risk factors model provided the strongest predictive power to the final model, followed by the socio-demographic and protective resources models. The protective model contributed the least predictive power to the model, which was consistent with the higher level of errors that the model contained on its own.

The MAE and RMSE of the ensemble model using the XGBoost were 2.06 and 2.69, respectively. In the ensemble by GWR, the ensemble predictions had MAE and RMSE of 2.48 and 3.34, respectively. As with the distribution of errors shown in Figure 3.4 for each model, there were a few significantly larger outliers, likely accounting for the discrepancy between the prediction values and the areas with a higher rate of overdoses. The value of its outlier errors was lower than that of the other models. The distribution of errors indicated that across many subsets of the block groups, the ensemble model performed with the least amount of error. The normal cross-validation result indicated that the model built in the training set had minimal overfitting features and generalized well across the different datasets (Figure 3.24). The RMSE obtained from spatial cross-

validation was 3.39 which did not increase considerably compared to the error estimated from normal cross-validation (RMSE difference of 0.7).

Moreover, a spatial autocorrelation test for residuals was performed for each individual model and the ensemble model (Figure 3.25). To map the residuals, the standard deviation classification method with an interval of 1 standard deviation from the mean was used. The blue color emphasizes values above the mean, and the red color shows values below the mean. As it is shown in the map, the distribution of the residuals for each individual model indicates clustering of over and under predictions in some areas. The map of residuals for the ensemble model showed that no pattern exists; instead, the model's residuals exhibited a random noise meaning that there was no clustering of over and under predictions in the model.

The result was further confirmed statistically by applying a spatial autocorrelation statistic (Global Moran's *I*) on residuals. This detects significant clustering or random pattern in the residuals. The Moran's *I* report revealed ~~that~~ the pattern of the residuals was significantly clustered, with a Moran's *I* value of 0.048, 0.065, and 0.045 for risk factors, socio-demographic, and protective resource, respectively. Figures 3.26 through 3.29 shows scatter plots of the results obtained from the ensemble model residual analysis according to the different spatial relationship conceptualization methods. The pattern of the residuals was significantly different from random, with a Moran's index 0.009, -0.004, -0.003, and 0.008 for KNN = 8, KNN = 25, KNN = 35, and queen contiguity, respectively. The residuals had no statistically significant spatial autocorrelation.

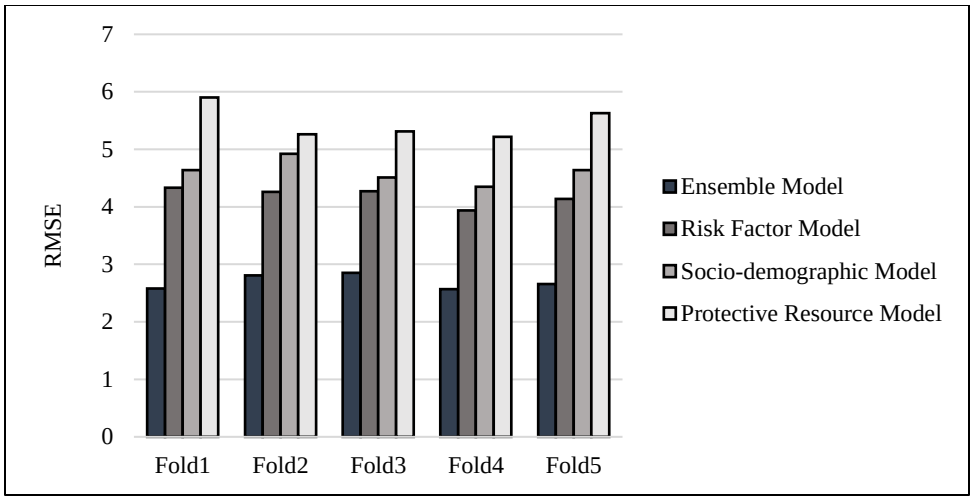

Figure 3.23. Performance evaluation results of models across 5-folds

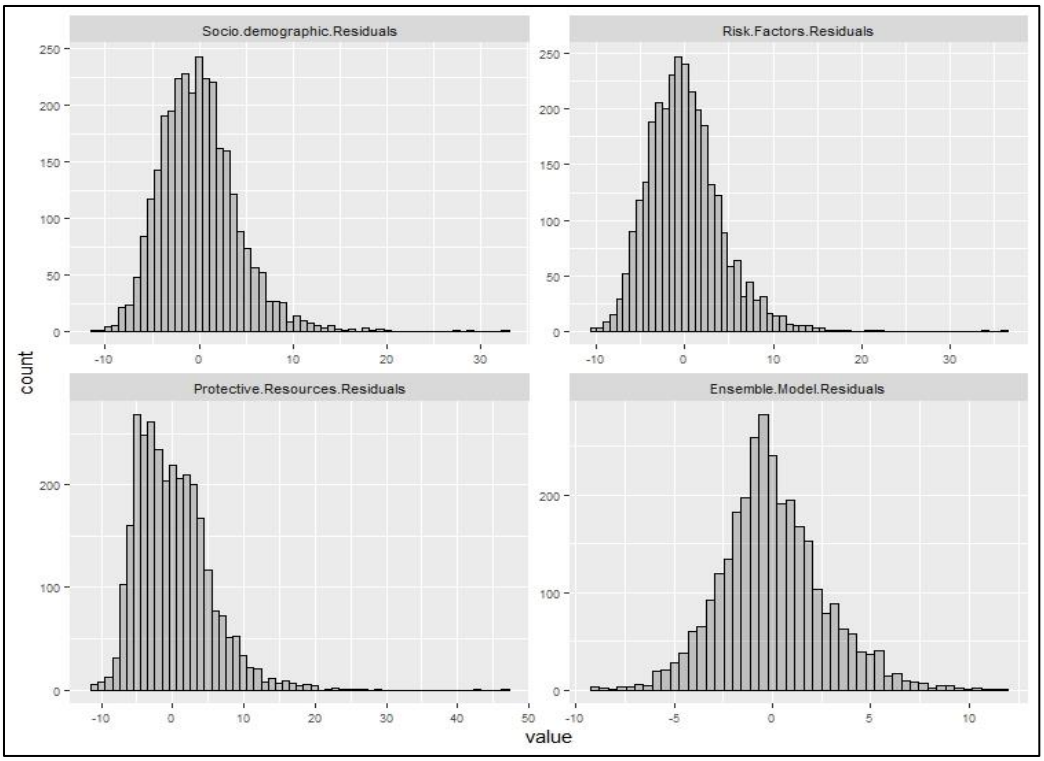

Figure 3.24. Performance evaluation results of models according to RMSE

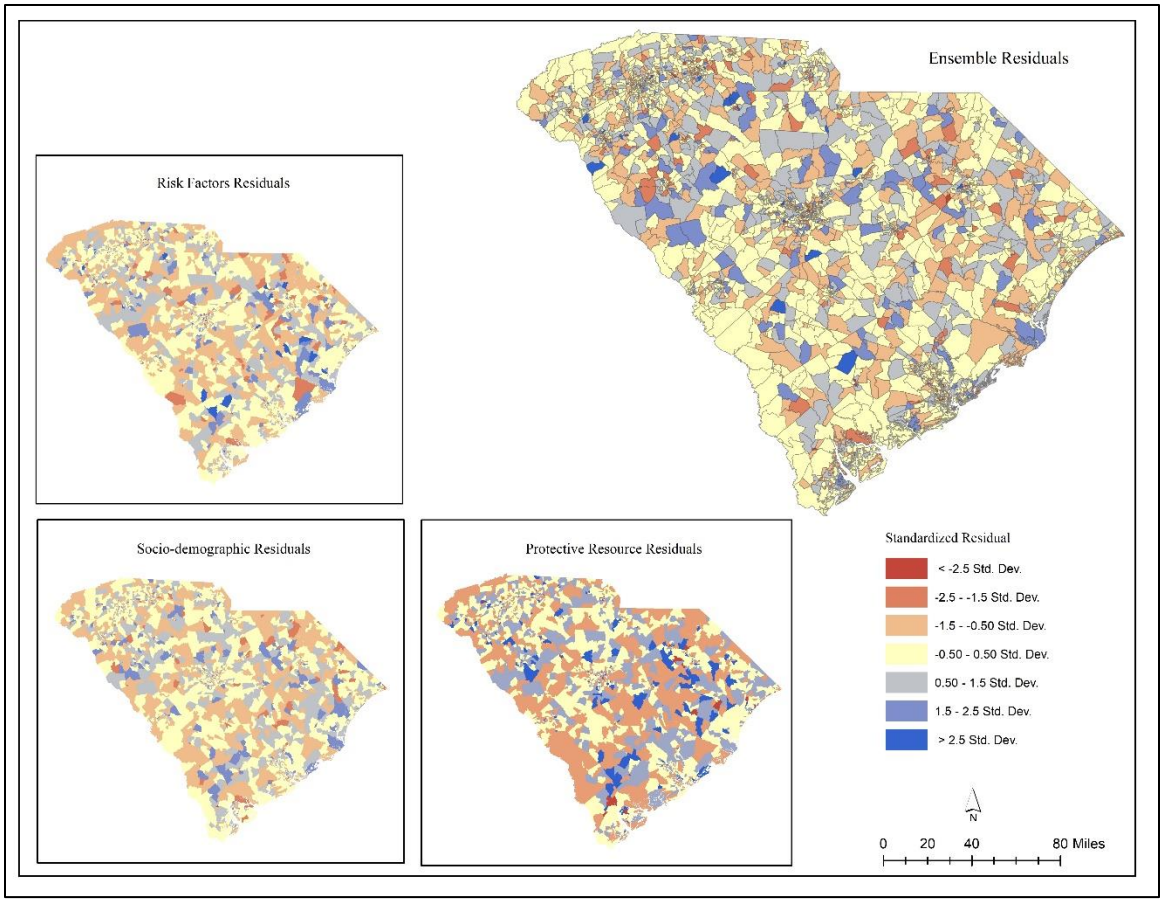

Figure 3.25. Distribution of residuals for individual and ensemble models

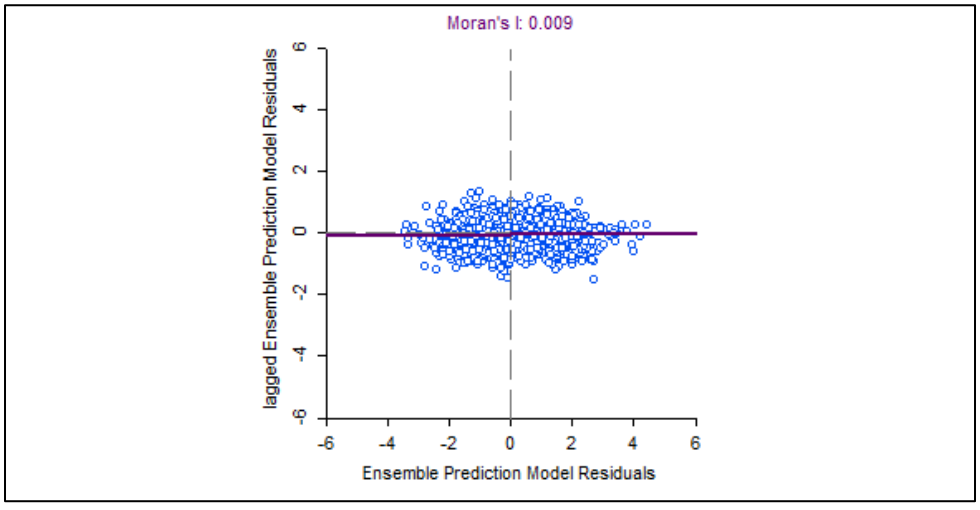

Figure 3.26. Moran's *I* scatter plot for ensemble model residual with 8 nearest neighbors

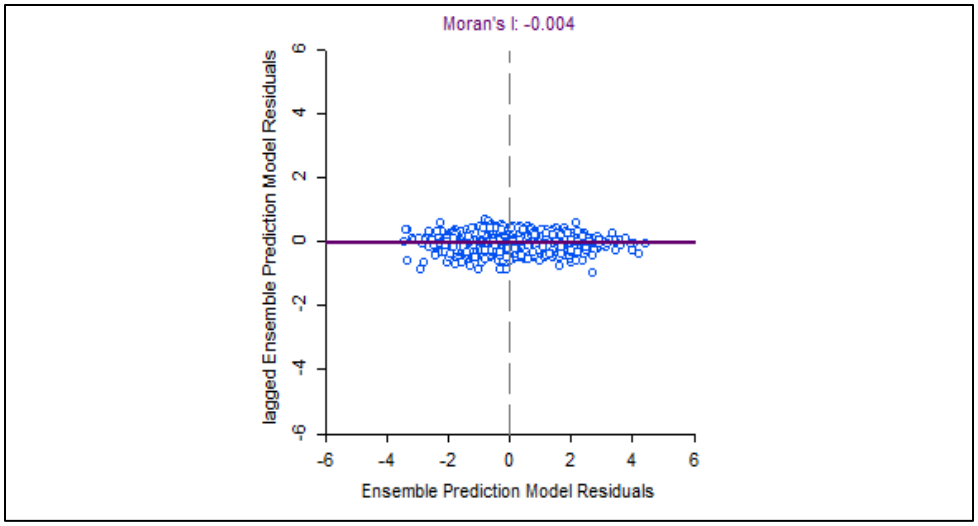

Figure 3.27. Moran's *I* scatter plot for ensemble model residual with 25 nearest neighbors

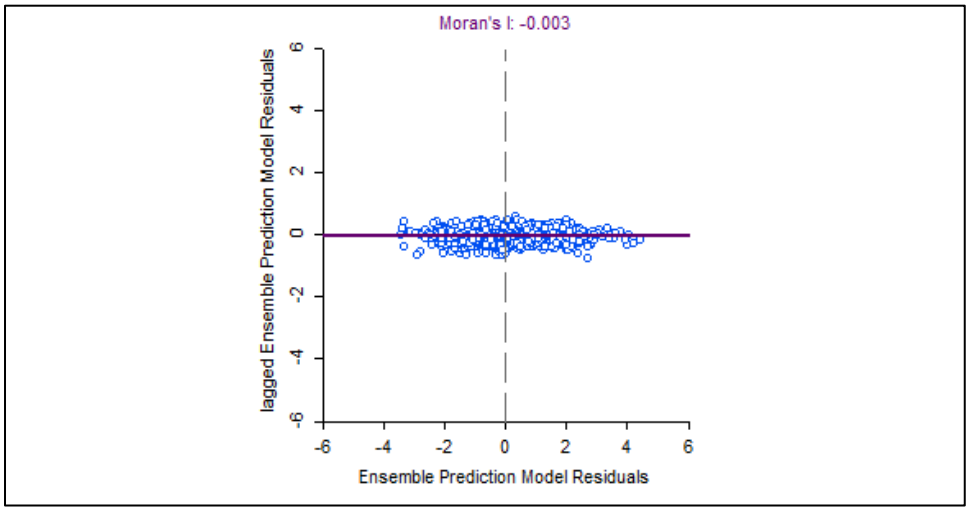

Figure 3.28. Moran's *I* scatter plot for ensemble model residual with 35 nearest neighbors

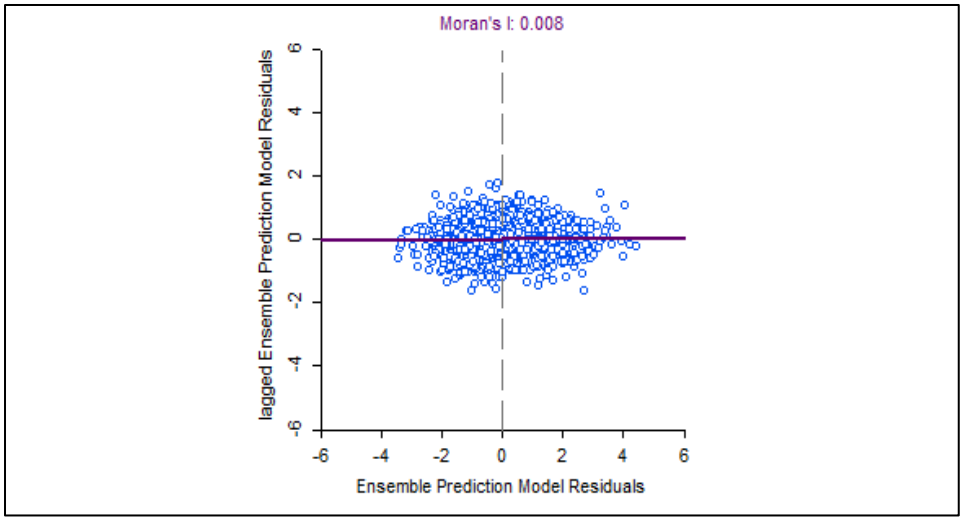

Figure 3.29. Moran's *I* scatter plot for ensemble model residual with queen contiguity neighbors

3.5. DISCUSSION AND CONCLUSION

We identified the most important features that contributed to a drug overdose and developed a spatial risk model to predict drug overdose at the neighborhood level across South Carolina. The data comprise 83 variables categorized into three domains: socio-demographic, risk factors, and protective resources. We used different feature selection techniques - including Linear Regression (LR), Sequential Minimal Optimization (SMO), Random Forest (RF), Extra Gradient Boosting (XGBoost), and SHAP to assess the best subset of features. Throughout the modeling process, the three individual and ensemble models were trained on random subsets of 80% of the block groups and tested on the remaining 20% of the block groups. We compared the models' performance on the test data using RMSE and MAE as the goodness of fit metrics. For each model, a grid-search approach with parallelized performance evaluation for model parameters tuning was used to generate the best model parameters. Using the 5-fold cross-validation technique, we assessed each of the model's performance to ensure the model's generalized well across

the different datasets. We also implemented spatial cross-validation by counties to validate the model across different areas. Finally, we used GWR and XGBoost to ensemble the three individual models.

All the methods produced a better performance with a reduced feature set than full features. There were no significant changes across the evaluated algorithms from standardizing or normalizing the data. The wrapper method was demonstrated to be superior compared to the filter-based method from the feature selection methods. Performance comparison results showed the XGBoost was the top-performing model in each domain. Analysis of the feature importance showed features including the individual with at least one opioid prescription, households with food stamps, and accessibility to opioid treatment facilities were the most important features contributing towards the prediction of a drug overdose. Within the protective resources, access to parks also contributed substantially (more than 0.5) to the model. While we considered this feature as a protective variable, greater access to the park was associated with greater drug overdoses.

The ensembled model achieved higher performance than each of the individual models. Importantly, enameling using XGBoost outperformed the more conventional spatial model technique (GWR). The ensemble model using XGBoost showed that the error decreased markedly, lowering the MAE and RMSE to 2.06 and 2.69, respectively, compared with MAE of 2.48 and RMSE of 3.34 obtained by GWR. Both models had similar R-squared values with very slight differences (R-squared = 0.73 for ensemble by XGBoost and R-squared = 0.71 for ensemble by GWR). A map of the predictions for the ensemble model showed that the combined model captured more of the nuance of the drug overdose risk, specifically in the coastal counties central to the state's shoreline. The range

of predictions was from 0.001-53.73, which was larger than the range of individual models. Our findings also suggested that the risk factor category, carrying the strong predictors of opioid prescription rate, played a crucial role in determining the course of the drug overdose epidemic. The opioid prescription rate predictor seems to correlate with the need for enhancing access to OTP, the most important predictor in protective resource category.

Analysis of the residuals for the ensemble model showed that the spatial variation was well captured by the model and there was no spatial autocorrelation in the residuals. Also, the ensemble model obtained relatively smaller maximum values of the residuals compared with individual models. There were overestimations and underestimations of drug overdose for the smaller and larger values, respectively, but most of the residuals fell into their confidence intervals. Individual models also produced similar distributions of the residuals with only slight overestimations and underestimations for the smaller and larger values, respectively. That was only a few of the residuals outside the range of their confidence intervals.

Normal cross-validation ensured that our model performed similarly when the data was trained on different subsets of our initial dataset. In spatial modeling, normal cross-validation generally returns a lower error, which indicates a potential over-optimistic estimate. However, the error of our prediction model did not increase considerably in spatial cross-validation compared to the error estimated from normal cross-validation that indicates the inclusion of the spatial dependency using KNN method were able to account for spatial autocorrelation in our model.

There are several notable strengths that distinguish this study from previous studies. First, this study is the first to use supervised machine learning methods that account for spatial dependency to predict neighborhoods at high risk of drug overdoses in South Carolina using various datasets. Second, the block group analysis provided more granularity to uncover local areas of spatial homogeneity and heterogeneity. Third, our model not only studied contextual aspects of the neighborhood (e.g., crime, socioeconomic status) and drug-related factors but also examined the effect of protective factors (i.e., adequate access to treatment centers) that may reduce the rate of drug overdoses. Forth, we measured accessibility to the facilities using floating catchment area methods (e.g., E2SFCA and W2SFCA) which is superior to the density-based methods used in past studies (Cantrell et al., 2015; Novak et al., 2006). Fifth, we revealed the effectiveness of spatial features in capturing spatial dependency and provided insights on the usage of spatial cross-validation in performance estimation. Sixth, this research showed that machine learning had a better performance compared with the traditional geographically weighted regression (GWR).

The ability of opioid treatment accessibility to predict drug overdose is in line with literature suggesting that enhancing spatial accessibility to treatment is associated with opioid-related mortality and treatment retention (Amiri et al., 2018a; Haley et al., 2019; Rosenblum et al., 2011b). In addition, variables describing access to tobacco and liquor stores lends empirical evidence to the theory that exposure to tobacco outlets and alcohol is associated with smoking and alcohol consumption (Bryden et al., 2012; Paynter & Edwards, 2009), which are known predictors of illicit and prescription drug abuse (Griffin et al., 2019). The high ranking of variables describing income less than $35,000 and use of

food stamps affirm the degree to which drug overdose is linked to economic characteristics of a neighborhood. Divorced and not married variables were also important predictors that were consistent with theory suggesting that family fragmentation or living alone may influence analgesic overdose through a social mechanism in a neighborhood (Cerdá et al., 2013). Education also may have immediate impacts on the drug overdose rate through the economic opportunities it engenders.

Despite previous research indicating ~~that the~~ urban/rural status of the neighborhood is associated with overdose (García et al., 2019; Keyes et al., 2014; King et al., 2014), urbanicity defined by RUCA codes wasn't helpful in drug overdose prediction. However, population density (which is an element often used to define urbanicity) was found to be an important predictor, consistent with literature indicating higher rate of drug overdose in dense areas (Galea et al., 2005; Latkin et al., 2003; Schroeder et al., 2001).

In addition, renter households with no vehicle may be a key demographic for targeted support by healthcare planners when allocating resources. Other important predictor included mobile homes. Mobile homes are more affordable than other housing types and primarily occupied by low income population (Boehm & Schlottmann, 2004). A possible explanation for this finding could be that it's a retirement option for elder people when they no longer have an income outside of social security. However, to better understand the possible relationship between drug overdose and mobile homes further exploration is needed.

Our model may be used in a decision-making capacity to prioritize the needs of specific communities based on the individual assessment of the predictors in each domain

before they were ensembled into one model. For example, the protective resource category may identify neighborhoods with lower risk more accurately than the risk factors and socio-demographic categories, helping policymakers to avoid expanding access to treatments where they would be less useful.

While we were successful in predicting drug overdose in South Carolina, there are some limitations that we plan to address in the future study when possible. First, the list of candidate predictors did not encompass all of the important risk factors of a drug overdose, such as exposure to a natural disaster that is known to predispose people toward using or abusing drugs as a coping mechanism (Cerdá et al., 2013). Similarly, we didn't have access to the block group level drug-related crime data. Second, we used the E2SFCA method to measure access to some facilities; however, alternative methods can be implemented to improve the accuracy of the accessibility measurements, which may have an impact on our final prediction. For example, measuring access to liquor and tobacco stores, parks, and libraries could be improved by defining more accurate catchment areas. Third, we only included a one-year estimate of the drug overdose; including time series overdose data in the model may result in better prediction.

In conclusion, we were able to identify strong neighborhood-level predictors of a drug overdose. Our findings may explain the spatial variability of a drug overdose and can complement existing policies by providing an opportunity to predict high-risk areas based on their community characteristics. This is supplemental to existing efforts and could make use of the infrastructure already in place. In the future, this model can be improved through the inclusion of more outcome and potential covariate data. The findings of this study must be interpreted with respect to these important strengths and limitations.

CHAPTER 4

CONCLUSION

This chapter provides an overview of our study findings and highlights the strengths and limitations of each chapter. In Chapter 1, we give a brief overview of why drug overdose and the opioid epidemic matter and certain aspects of one's environment may negatively or positively impact the drug overdose rate. We further discuss the populations that are most vulnerable to drug overdose and the gaps in the literature, examining the relationships between socio-demographic, protective resources, and risk factors with drug overdoses.

In Chapter 2, we introduce a new measure to quantify the spatial accessibility to opioid treatment programs (OTP). We use the measure in a case study to highlight the need to improve spatial disparity in accessibility to OTPs in South Carlina. The proposed method incorporates facility attractiveness and uses the Huff Model for quantifying the probability of a person's preference on an OTP site. We also used the social vulnerability index (SVI) to account for nonspatial factors that mitigate or compound the impacts of spatial access to care. Results of the study indicate a significant variation in access to OTPs statewide. Spatial access to OTPs is low across the entire state except for a limited number of metropolitan areas. Approximately 21% of the state's population lives in areas with low access, 23% live in areas identified as medium-low access, 26% live in areas identified as medium-high access, and 30% live in high access areas. A majority of the population with low access (85%) live in areas with a moderate to a high level of social vulnerability. Results provide more realistic estimates of access to care to assist policymakers in better targeting disadvantaged areas for OTP program expansion and resource allocation. In Chapter 3, we demonstrate how spatial access can be incorporated into a model using machine learning algorithms to predict potential risk of drug overdose. Despite the notable

advantage of our method, several issues deserve attention when interpreting the results. Population locations used for this study are weighted block group centroids. The developed method, however, has the potential to further articulate the population selection behavior because the block group population is not necessarily a proper indicator of opioid treatment needs. This can be partially addressed in future development by incorporating the number of patients with a history of prescription opioid use or experienced opioid overdoes. This study also assumes that all patients traveled by car and don't consider different modes of transportation, such as public transportation, as it is somewhat limited in the state.

For the spatial model, we identify the most important neighborhood-level (e.g., block group) factors associated with a drug overdose. Using these factors, we developed a model using machine learning algorithms to predict the likelihood or risk of drug overdoses across South Carolina. We also investigated whether the machine learning model captures spatial patterns better than conventional spatial techniques such as geographic weighted regression (GWR). Our model includes contextual aspects of the neighborhood (e.g., crime, socio-economic and demographic status) and drug-related factors to predict drug overdose.

Our results show that features including the prescription opioid rate and average smoking accessories representing expenditures for a product such as a cigar, tobacco, and pipe, are the most important predictors within the risk factors category. Within the socio-demographic domain, households with food stamps and income less than $3,500 have the strongest prediction power. Accessibility to opioid treatment facilities and hospitals are the most important features contributing towards the prediction of a drug overdose among

protective resource factors. Our results also indicate that protective factors like access to treatment centers may positively influence reducing drug overdose rates in some neighborhoods. For example, a neighborhood with greater access to OTPs and hospitals is less likely to experience a drug overdose. The results demonstrate that machine learning has the better performance results using various metrics compared with GWR.

While we were successful in predicting areas at high risk of overdose in South Carolina, there are some limitations that could be addressed in the future. First, the list of candidate predictors did not encompass all of the important risk factors of a drug overdose, such as exposure to a natural disaster that is known to impact people toward using or abusing drugs as a coping mechanism (Sinha, 2008). Similarly, we didn't have access to the block group level drug-related crime data. Second, we used the E2SFCA method to measure access to some facilities; however, alternative methods (e.g., V2SFCA) (W. Luo & Whippo, 2012b) can be implemented to measure accessibility. Third, we only included a one-year estimate of drug overdose; including time series overdose data in the model may result in better prediction.

Our overarching goal is to detect and prevent overdose before it occurs. This is complementary to existing efforts such as safe injection programs and prescription drug monitoring. Public health practitioners and other officials may use findings to inform decisions related to the development and implementation of drug overdose prevention efforts by facilitating better targeting of resources towards neighborhoods with greatest need.

REFERENCES

Ajit, P. (2016). Prediction of employee turnover in organizations using machine learning algorithms. *Algorithms*, *4*(5), C5.

Amiri, S., Lutz, R., Socías, M. E., McDonell, M. G., Roll, J. M., & Amram, O. (2018a). Increased distance was associated with lower daily attendance to an opioid treatment program in Spokane County Washington. *Journal of Substance Abuse Treatment*, *93*, 26–30. https://doi.org/10.1016/j.jsat.2018.07.006

Amiri, S., Lutz, R., Socías, M. E., McDonell, M. G., Roll, J. M., & Amram, O. (2018b). Increased distance was associated with lower daily attendance to an opioid treatment program in Spokane County Washington. *Journal of Substance Abuse Treatment*, *93*, 26–30. https://doi.org/10.1016/J.JSAT.2018.07.006

Amram, O., Socías, E., Nosova, E., Kerr, T., Wood, E., DeBeck, K., Hayashi, K., Fairbairn, N., Montaner, J., & Milloy, M. J. (2019). Density of low-barrier opioid agonist clinics and risk of non-fatal overdose during a community-wide overdose crisis: A spatial analysis. *Spatial and Spatio-Temporal Epidemiology*, *30*, 100288.

Badger, J., LaRose, E., Mayer, J., Bashiri, F., Page, D., & Peissig, P. (2019). Machine learning for phenotyping opioid overdose events. *Journal of Biomedical Informatics*, *94*, 103185.

Bagherzadeh-Khiabani, F., Ramezankhani, A., Azizi, F., Hadaegh, F., Steyerberg, E. W., & Khalili, D. (2016). A tutorial on variable selection for clinical prediction models: Feature selection methods in data mining could improve the results. *Journal of Clinical Epidemiology*, *71*, 76–85.

Beetham, T., Saloner, B., Wakeman, S. E., Gaye, M., & Barnett, M. L. (2019). Access to office-based buprenorphine treatment in areas with high rates of opioid-related mortality: An audit study. *Annals of Internal Medicine*, *171*(1), 1–9.

Boardman, J. D., Finch, B. K., Ellison, C. G., Williams, D. R., & Jackson, J. S. (2001). Neighborhood disadvantage, stress, and drug use among adults. *Journal of Health and Social Behavior*, 151–165.

Boehm, T. P., & Schlottmann, A. (2004). Is manufactured housing a good alternative for low-income families. *Evidence from the American Housing Survey*.

Bohnert, A. S., Nandi, A., Tracy, M., Cerdá, M., Tardiff, K. J., Vlahov, D., & Galea, S. (2011). Policing and risk of overdose mortality in urban neighborhoods. *Drug and Alcohol Dependence*, *113*(1), 62–68.

Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Analysis*, *28*(4), 281–298.

Bryden, A., Roberts, B., McKee, M., & Petticrew, M. (2012). A systematic review of the influence on alcohol use of community level availability and marketing of alcohol. *Health & Place*, *18*(2), 349–357.

Callahan, S. L. (2018). *Contextual perspectives on heroin addiction and recovery: Classic and contemporary theories*.

Cantrell, J., Pearson, J. L., Anesetti-Rothermel, A., Xiao, H., Kirchner, T. R., & Vallone, D. (2015). Tobacco retail outlet density and young adult tobacco initiation. *Nicotine & Tobacco Research*, *18*(2), 130–137.

CDC. (2015, July 7). *Today's Heroin Epidemic*. Centers for Disease Control and Prevention. https://www.cdc.gov/vitalsigns/heroin/index.html

CDC. (2018, September 10). *The Social Vulnerability Index*. data-and-tools-download.html

CDC. (2020a, March 19). *Understanding the Epidemic*. https://www.cdc.gov/drugoverdose/epidemic/index.html

CDC, C. for D. C. and P. (2020b, March 19). *Understanding the Epidemic | Drug Overdose*. https://www.cdc.gov/drugoverdose/epidemic/index.html

CDC, C. for D. C. and P. (2020c, December 7). *U.S. Opioid Dispensing Rate Maps*. https://www.cdc.gov/drugoverdose/maps/rxrate-maps.html

CDC Injury Center. (2021). *Drug Overdose*. https://www.cdc.gov/drugoverdose/data/analysis.html

Cerdá, M., Ransome, Y., Keyes, K. M., Koenen, K. C., Tardiff, K., Vlahov, D., & Galea, S. (2013). Revisiting the role of the urban environment in substance use: The case of analgesic overdose fatalities. *American Journal of Public Health*, *103*(12), 2252–2260.

Chatterjee, A., Yu, E. J., & Tishberg, L. (2018). Exploring opioid use disorder, its impact, and treatment among individuals experiencing homelessness as part of a family. *Drug and Alcohol Dependence*, *188*, 161–168.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 785–794.

Chen, X., Wang, Y., Yu, X., Schoenfeld, E., Saltz, M., Saltz, J., & Wang, F. (2017). Large-scale analysis of opioid poisoning related hospital visits in New York state. *AMIA Annual Symposium Proceedings*, *2017*, 545.

Chichester, K., Drawve, G., Giménez-Santana, A., Sisson, M., McCleskey, B., Dye, D. W., Walker, J., Mrug, S., & Cropsey, K. (2020). Pharmacies and features of the built environment associated with opioid overdose: A geospatial comparison of rural and urban regions in Alabama, USA. *International Journal of Drug Policy*, *79*, 102736.

Cochran, G., Gordon, A. J., Lo-Ciganic, W.-H., Gellad, W. F., Frazier, W., Lobo, C., Chang, J., Zheng, P., & Donohue, J. M. (2017). An examination of claims-based predictors of overdose from a large Medicaid program. *Medical Care*, *55*(3), 291.

Cooper, L. A., Hill, M. N., & Powe, N. R. (2002). Designing and evaluating interventions to eliminate racial and ethnic disparities in health care. *Journal of General Internal Medicine*, *17*(6), 477–486. https://doi.org/10.1046/j.1525-1497.2002.10633.x

Darke, S., & Hall, W. (2003). Heroin overdose: Research and evidence-based intervention. *Journal of Urban Health*, *80*(2), 189–200.

Darke, S., Mattick, R. P., & Degenhardt, L. (2003). The ratio of non-fatal to fatal heroin overdose. *Addiction*, *98*(8), 1169–1171.

Darke, S., Williamson, A., Ross, J., Mills, K. L., Havard, A., & Teesson, M. (2007). Patterns of nonfatal heroin overdose over a 3-year period: Findings from the Australian treatment outcome study. *Journal of Urban Health*, *84*(2), 283–291.

De Vries, S., Van Dillen, S. M., Groenewegen, P. P., & Spreeuwenberg, P. (2013). Streetscape greenery and health: Stress, social cohesion and physical activity as mediators. *Social Science & Medicine*, *94*, 26–33.

Dick, A. W., Pacula, R. L., Gordon, A. J., Sorbero, M., Burns, R. M., Leslie, D. L., & Stein, B. D. (2015). Increasing potential access to opioid agonist treatment in US treatment shortage areas. *Health Affairs (Project Hope)*, *34*(6), 1028.

Ding, D., Sallis, J. F., Kerr, J., Lee, S., & Rosenberg, D. E. (2011). Neighborhood environment and physical activity among youth: A review. *American Journal of Preventive Medicine*, *41*(4), 442–455.

Donohoe, J., Marshall, V., Tan, X., Camacho, F. T., Anderson, R., & Balkrishnan, R. (2016). Evaluating and Comparing Methods for Measuring Spatial Access to

Mammography Centers in Appalachia (Re-Revised). *Health Services & Outcomes Research Methodology*, *16*(1), 22–40. https://doi.org/10.1007/s10742-016-0143-y

Dramowicz, E. (2005). Retail trade area analysis using the Huff model. *Directions Magazine, Jul*, *2*.

Dworkis, D. A., Taylor, L. A., Peak, D. A., & Bearnot, B. (2017). Geospatial analysis of emergency department visits for targeting community-based responses to the opioid epidemic. *Plos One*, *12*(3), e0175115.

Edwards, E. S. (2016). Patient characteristics and outcomes in unintentional, non-fatal prescription opioid overdoses: A systematic review. *Pain Physician*, *19*, 215–228.

Ellis, R. J., Wang, Z., Genes, N., & Ma'ayan, A. (2019). Predicting opioid dependence from electronic health records with machine learning. *BioData Mining*, *12*(1), 3.

Eriksson, U., Arvidsson, D., & Sundquist, K. (2012). Availability of exercise facilities and physical activity in 2,037 adults: Cross-sectional results from the Swedish neighborhood and physical activity (SNAP) study. *BMC Public Health*, *12*(1), 607.

ESRI. (n.d.). *Business Analyst Data*. Retrieved March 10, 2021, from https://www.esri.com/en-us/arcgis/products/arcgis-business-analyst/data-infographics

ESRI, R. (2011). ArcGIS desktop: Release 10. *Environmental Systems Research Institute, CA*.

ESRI, R., CA, USA. (2020). *About ArcGIS Pro—ArcGIS Pro | Documentation*. https://pro.arcgis.com/en/pro-app/latest/get-started/get-started.htm

Feng, J., Glass, T. A., Curriero, F. C., Stewart, W. F., & Schwartz, B. S. (2010). The built environment and obesity: A systematic review of the epidemiologic evidence. *Health & Place*, *16*(2), 175–190.

Feng, P., Wang, B., Li Liu, D., Xing, H., Ji, F., Macadam, I., Ruan, H., & Yu, Q. (2018). Impacts of rainfall extremes on wheat yield in semi-arid cropping systems in eastern Australia. *Climatic Change*, *147*(3), 555–569.

Fite, P. J., Lochman, J. E., & Wells, K. C. (2009). The influence of neighborhood disadvantage and perceived disapproval on early substance use initiation. *Addictive Behaviors*, *34*(9), 769–771.

Frank, J. W., Levy, C., Calcaterra, S. L., Hoppe, J. A., & Binswanger, I. A. (2016). Naloxone administration in US emergency departments, 2000–2011. *Journal of Medical Toxicology*, *12*(2), 148–156.

Frankenfeld, C. L., & Leslie, T. F. (2019). County-level socioeconomic factors and residential racial, Hispanic, poverty, and unemployment segregation associated with drug overdose deaths in the United States, 2013–2017. *Annals of Epidemiology*, *35*, 12–19.

Friedman, S. R., Tempalski, B., Brady, J. E., West, B. S., Pouget, E. R., Williams, L. D., Des Jarlais, D. C., & Cooper, H. L. (2016). Income inequality, drug-related arrests, and the health of people who inject drugs: Reflections on seventeen years of research. *International Journal of Drug Policy*, *32*, 11–16.

Fuller, C. M., Borrell, L. N., Latkin, C. A., Galea, S., Ompad, D. C., Strathdee, S. A., & Vlahov, D. (2005). Effects of race, neighborhood, and social network on age at initiation of injection drug use. *American Journal of Public Health*, *95*(4), 689–695.

Gabrysch, S., Cousens, S., Cox, J., & Campbell, O. M. (2011). The influence of distance and level of care on delivery place in rural Zambia: A study of linked national data in a geographic information system. *PLoS Med*, *8*(1), e1000394.

Galea, S., Ahern, J., & Vlahov, D. (2003). Contextual determinants of drug use risk behavior: A theoretic framework. *Journal of Urban Health*, *80*(3), iii50–iii58.

Galea, S., Rudenstine, S., & Vlahov, D. (2005). Drug use, misuse, and the urban environment. *Drug and Alcohol Review*, *24*(2), 127–136.

García, M. C., Heilig, C. M., Lee, S. H., Faul, M., Guy, G., Iademarco, M. F., Hempstead, K., Raymond, D., & Gray, J. (2019). Opioid prescribing rates in nonmetropolitan and metropolitan counties among primary care providers using an electronic health record system—United States, 2014–2017. *Morbidity and Mortality Weekly Report*, *68*(2), 25.

Geissert, P., Hallvik, S., Van Otterloo, J., O'Kane, N., Alley, L., Carson, J., Leichtling, G., Hildebran III, C., Wakeland, W., & Deyo, R. A. (2018). High Risk Prescribing and Opioid Overdose: Prospects for Prescription Drug Monitoring Program Based Proactive Alerts. *Pain*, *159*(1), 150.

Gentleman, R., & Carey, V. J. (2008). Unsupervised machine learning. In *Bioconductor case studies* (pp. 137–157). Springer.

Glanz, J. M., Narwaney, K. J., Mueller, S. R., Gardner, E. M., Calcaterra, S. L., Xu, S., Breslin, K., & Binswanger, I. A. (2018). Prediction model for two-year risk of opioid overdose among patients prescribed chronic opioid therapy. *Journal of General Internal Medicine*, *33*(10), 1646–1653.

Griffin, K. W., Lowe, S. R., Botvin, C., & Acevedo, B. P. (2019). Patterns of adolescent tobacco and alcohol use as predictors of illicit and prescription drug abuse in minority young adults. *Journal of Prevention & Intervention in the Community*, *47*(3), 228–242.

Grubesic, T. H., & Matisziw, T. C. (2006). On the use of ZIP codes and ZIP code tabulation areas (ZCTAs) for the spatial analysis of epidemiological data. *International Journal of Health Geographics*, *5*(1), 58.

Haley, S. J., Maroko, A. R., Wyka, K., & Baker, M. R. (2019). The association between county-level safety net treatment access and opioid hospitalizations and mortality in New York. *Journal of Substance Abuse Treatment*, *100*, 52–58.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer Science & Business Media.

Hastings, J. S., Howison, M., & Inman, S. E. (2020). Predicting high-risk opioid prescriptions before they are given. *Proceedings of the National Academy of Sciences*, *117*(4), 1917–1923.

Hembree, C., Galea, S., Ahern, J., Tracy, M., Piper, T. M., Miller, J., Vlahov, D., & Tardiff, K. J. (2005). The urban built environment and overdose mortality in New York City neighborhoods. *Health & Place*, *11*(2), 147–156.

Hernandez, A., Branscum, A. J., Li, J., MacKinnon, N. J., Hincapie, A. L., & Cuadros, D. F. (2020). Epidemiological and geospatial profile of the prescription opioid crisis in Ohio, United States. *Scientific Reports*, *10*(1), 1–10.

Huskamp, H. A., Riedel, L. E., Barry, C. L., & Busch, A. B. (2018). Coverage of medications that treat opioid use disorder and opioids for pain management in marketplace plans, 2017. *Medical Care*, *56*(6), 505.

IPUMS, U. of M. (n.d.). *IPUMS USA* (11.0) [Data set]. University of Minnesota. https://doi.org/10.18128/D010.V11.0

Jensen, M., Chassin, L., & Gonzales, N. A. (2017). Neighborhood moderation of sensation seeking effects on adolescent substance use initiation. *Journal of Youth and Adolescence*, *46*(9), 1953–1967.

Kakko, J., Svanborg, K. D., Kreek, M. J., & Heilig, M. (2003). 1-year retention and social function after buprenorphine-assisted relapse prevention treatment for heroin dependence in Sweden: A randomised, placebo-controlled trial. *Lancet (London, England)*, *361*(9358), 662–668. https://doi.org/10.1016/S0140-6736(03)12600-1

Keyes, K. M., Cerdá, M., Brady, J. E., Havens, J. R., & Galea, S. (2014). Understanding the rural–urban differences in nonmedical prescription opioid use and abuse in the United States. *American Journal of Public Health*, *104*(2), e52–e59.

King, N. B., Fraser, V., Boikos, C., Richardson, R., & Harper, S. (2014). Determinants of increased opioid-related mortality in the United States and Canada, 1990–2013: A systematic review. *American Journal of Public Health*, *104*(8), e32–e42.

Knowlton, A., Weir, B. W., Hazzard, F., Olsen, Y., McWilliams, J., Fields, J., & Gaasch, W. (2013). EMS runs for suspected opioid overdose: Implications for surveillance and prevention. *Prehospital Emergency Care*, *17*(3), 317–329.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Ijcai*, *14*(2), 1137–1145.

Krawczyk, N., Feder, K. A., Fingerhood, M. I., & Saloner, B. (2017). Racial and ethnic differences in opioid agonist treatment for opioid use disorder in a US national sample. *Drug and Alcohol Dependence*, *178*, 512–518.

Krawczyk, N., Mojtabai, R., Stuart, E. A., Fingerhood, M. I., Agus, D., Lyons, B. C., Weiner, J. P., & Saloner, B. (2020). Opioid agonist treatment is highly protective against overdose death among a US statewide population of justice-involved adults. *The American Journal of Drug and Alcohol Abuse*, 1–10.

Kvamme, E., Catlin, M., Banta-Green, C., Roll, J., & Rosenblatt, R. (2013). Who prescribes buprenorphine for rural patients? The impact of specialty, location and practice type in Washington State. *Journal of Substance Abuse Treatment*, *44*(3), 355–360.

Lanier, W. A., Johnson, E. M., Rolfs, R. T., Friedrichs, M. D., & Grey, T. C. (2012). Risk factors for prescription opioid-related death, Utah, 2008–2009. *Pain Medicine*, *13*(12), 1580–1589.

Latkin, C. A., Forman, V., Knowlton, A., & Sherman, S. (2003). Norms, social networks, and HIV-related risk behaviors among urban disadvantaged drug users. *Social Science & Medicine*, *56*(3), 465–476.

Lim, S., & Chi, S. (2019). Xgboost application on bridge management systems for proactive damage estimation. *Advanced Engineering Informatics*, *41*, 100922.

Lin, Y., Wan, N., Sheets, S., Gong, X., & Davies, A. (2018a). A multi-modal relative spatial access assessment approach to measure spatial accessibility to primary care providers. *International Journal of Health Geographics*, *17*(1), 33.

Lin, Y., Wan, N., Sheets, S., Gong, X., & Davies, A. (2018b). A multi-modal relative spatial access assessment approach to measure spatial accessibility to primary care

providers. *International Journal of Health Geographics*, *17*(1), 33. https://doi.org/10.1186/s12942-018-0153-9

Lipari, R. N. (2018). *Key Substance Use and Mental Health Indicators in the United States: Results from the 2018 National Survey on Drug Use and Health*. 82.

Liu, H., Motoda, H., Setiono, R., & Zhao, Z. (2010). Feature selection: An ever evolving frontier in data mining. *Feature Selection in Data Mining*, 4–13.

Lo-Ciganic, W.-H., Huang, J. L., Zhang, H. H., Weiss, J. C., Wu, Y., Kwoh, C. K., Donohue, J. M., Cochran, G., Gordon, A. J., & Malone, D. C. (2019). Evaluation of machine-learning algorithms for predicting opioid overdose risk among medicare beneficiaries with opioid prescriptions. *JAMA Network Open*, *2*(3), e190968–e190968.

Luo, J. (2014). Integrating the Huff model and floating catchment area methods to analyze spatial access to healthcare services. *Transactions in GIS*, *18*(3), 436–448.

Luo, W., & Qi, Y. (2009a). An enhanced two-step floating catchment area (E2SFCA) method for measuring spatial accessibility to primary care physicians. *Health & Place*, *15*(4), 1100–1107.

Luo, W., & Qi, Y. (2009b). An enhanced two-step floating catchment area (E2SFCA) method for measuring spatial accessibility to primary care physicians. *Health & Place*, *15*(4), 1100–1107. https://doi.org/10.1016/j.healthplace.2009.06.002

Luo, W., & Wang, F. (2003a). *Measures of spatial accessibility to health care in a GIS environment: Synthesis and a case study in the Chicago region*. *30*, 865–884. https://doi.org/10.1068/b29120

Luo, W., & Wang, F. (2003b). Spatial accessibility to primary care and physician shortage area designation: A case study in Illinois with GIS approaches. In *Geographic information systems and health applications* (pp. 261–279). IGI Global.

Luo, W., & Wang, F. (2003c). Measures of Spatial Accessibility to Health Care in a GIS Environment: Synthesis and a Case Study in the Chicago Region. *Environment and Planning B: Planning and Design*, *30*(6), 865–884. https://doi.org/10.1068/b29120

Luo, W., & Whippo, T. (2012a). Variable catchment sizes for the two-step floating catchment area (2SFCA) method. *Health & Place*, *18*(4), 789–795.

Luo, W., & Whippo, T. (2012b). Variable catchment sizes for the two-step floating catchment area (2SFCA) method. *Health & Place*, *18*(4), 789–795. https://doi.org/10.1016/j.healthplace.2012.04.002

Mallappallil, M., Sabu, J., Friedman, E. A., & Salifu, M. (2017). What do we know about opioids and the kidney? *International Journal of Molecular Sciences*, *18*(1), 223.

Mancher, M., Leshner, A. I., National Academies of Sciences, & Medicine. (2019). Barriers to Broader Use of Medications to Treat Opioid Use Disorder. In *Medications for Opioid Use Disorder Save Lives*. National Academies Press (US).

Mao, L., & Nekorchuk, D. (2013). Measuring spatial accessibility to primary health care services. *Health & Place*, *24*, 115–122.

Martins, S. S., Sampson, L., Cerdá, M., & Galea, S. (2015). Worldwide prevalence and trends in unintentional drug overdose: A systematic review of the literature. *American Journal of Public Health*, *105*(11), e29–e49.

McGrail, M. R., & Humphreys, J. S. (2014). Measuring spatial accessibility to primary health care services: Utilising dynamic catchment sizes. *Applied Geography*, *54*, 182–188.

Meade, M. A., Mahmoudi, E., & Lee, S. Y. (2015). The intersection of disability and healthcare disparities: A conceptual framework. *Disability and Rehabilitation*, *37*(7), 632–641. https://doi.org/10.3109/09638288.2014.938176

Methadone. (n.d.). Samhsa.Gov. Retrieved April 12, 2021, from https://www.samhsa.gov/medication-assisted-treatment/medications-counseling-related-conditions/methadone

Mitchell, R., & Popham, F. (2008). Effect of exposure to natural environment on health inequalities: An observational population study. *The Lancet*, *372*(9650), 1655–1660.

Molnar, B. E., Gortmaker, S. L., Bull, F. C., & Buka, S. L. (2004). Unsafe to play? Neighborhood disorder and lack of safety predict reduced physical activity among urban children and adolescents. *American Journal of Health Promotion*, *18*(5), 378–386.

Murray, A. T., Wei, R., & Grubesic, T. H. (2014). An approach for examining alternatives attributable to locational uncertainty. *Environment and Planning B: Planning and Design*, *41*(1), 93–109.

National Academies of Sciences, E., Division, H. and M., Policy, B. on H. S., Disorder, C. on M.-A. T. for O. U., Mancher, M., & Leshner, A. I. (2019). The Effectiveness of Medication-Based Treatment for Opioid Use Disorder. In *Medications for Opioid Use Disorder Save Lives*. National Academies Press (US). https://www.ncbi.nlm.nih.gov/books/NBK541393/

National Institute on Drug. (2021, January 29). *Overdose Death Rates*. National Institute on Drug Abuse. https://www.drugabuse.gov/drug-topics/trends-statistics/overdose-death-rates

Ni, J., Wang, J., Rui, Y., Qian, T., & Wang, J. (2015). An enhanced variable two-step floating catchment area method for measuring spatial accessibility to residential care facilities in Nanjing. *International Journal of Environmental Research and Public Health, 12*(11), 14490–14504.

Novak, S. P., Reardon, S. F., Raudenbush, S. W., & Buka, S. L. (2006). Retail tobacco outlet density and youth cigarette smoking: A propensity-modeling approach. *American Journal of Public Health, 96*(4), 670–676.

Osisanwo, F. Y., Akinsola, J. E. T., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., & Akinjobi, J. (2017). Supervised machine learning algorithms: Classification and comparison. *International Journal of Computer Trends and Technology (IJCTT), 48*(3), 128–138.

Paynter, J., & Edwards, R. (2009). The impact of tobacco promotion at the point of sale: A systematic review. *Nicotine and Tobacco Research, 11*(1), 25–35.

Pitt, A. L., Humphreys, K., & Brandeau, M. L. (2018). Modeling health benefits and harms of public policy responses to the US opioid epidemic. *American Journal of Public Health, 108*(10), 1394–1400.

Ranga, V., & Panda, P. (2014). Spatial access to inpatient health care in northern rural India. *Geospatial Health, 8*(2), 545–556. https://doi.org/10.4081/gh.2014.44

Rosenblum, A., Cleland, C. M., Fong, C., Kayman, D. J., Tempalski, B., & Parrino, M. (2011a). Distance traveled and cross-state Commuting to opioid treatment Programs in the united states. *Journal of Environmental and Public Health, 2011*. https://doi.org/10.1155/2011/948789

Rosenblum, A., Cleland, C. M., Fong, C., Kayman, D. J., Tempalski, B., & Parrino, M. (2011b). Distance Traveled and Cross-State Commuting to Opioid Treatment Programs in the United States. *Journal of Environmental and Public Health, 2011*. https://doi.org/10.1155/2011/948789

Rossen, L. M., Khan, D., & Warner, M. (2014). Hot spots in mortality from drug poisoning in the United States, 2007–2009. *Health & Place, 26*, 14–20.

Rummans, T. A., Burton, M. C., & Dawson, N. L. (2018). How good intentions contributed to bad outcomes: The opioid crisis. *Mayo Clinic Proceedings, 93*(3), 344–350.

Saloner, B., & Karthikeyan, S. (2015). Changes in substance abuse treatment use among individuals with opioid use disorders in the United States, 2004-2013. *Jama*, *314*(14), 1515–1517.

SAMHSA. (n.d.). *Substance Use Treatment Locator*. Retrieved March 10, 2021, from https://www.samhsa.gov/find-treatment

Sánchez-Maroño, N., Alonso-Betanzos, A., & Tombilla-Sanromán, M. (2007). Filter methods for feature selection–a comparative study. *International Conference on Intelligent Data Engineering and Automated Learning*, 178–187.

SC Department of Alcohol and Other Drug Abuse Services. (n.d.). Opioid Data. *Just Plain Killers*. Retrieved February 5, 2021, from http://justplainkillers.com/data/

SC Drug Overdose Deaths. (n.d.). Retrieved February 5, 2021, from https://scdhec.gov/sites/default/files/media/document/2019_South_Carolina_Drug_Overdose_Deaths.pdf

SCDHEC. (n.d.). *Opioid Epidemic*. Retrieved October 16, 2020, from https://scdhec.gov/opioid-epidemic

Scheier, L. E. (2010). *Handbook of drug use etiology: Theory, methods, and empirical findings.* American Psychological Association.

Schroeder, J. R., Latkin, C. A., Hoover, D. R., Curry, A. D., Knowlton, A. R., & Celentano, D. D. (2001). Illicit drug use in one's social network and in one's neighborhood predicts individual heroin and cocaine use. *Annals of Epidemiology*, *11*(6), 389–394.

Schwartz, R. P., Gryczynski, J., O'Grady, K. E., Sharfstein, J. M., Warren, G., Olsen, Y., Mitchell, S. G., & Jaffe, J. H. (2013). Opioid agonist treatments and heroin overdose deaths in Baltimore, maryland, 1995-2009. *American Journal of Public Health*, *103*(5), 917–922. https://doi.org/10.2105/AJPH.2012.301049

Seal, K. H., Kral, A. H., Gee, L., Moore, L. D., Bluthenthal, R. N., Lorvick, J., & Edlin, B. R. (2001). Predictors and prevention of nonfatal overdose among street-recruited injection heroin users in the San Francisco Bay Area, 1998–1999. *American Journal of Public Health*, *91*(11), 1842–1846.

Shah, T. I., Bell, S., & Wilson, K. (2016). Spatial Accessibility to Health Care Services: Identifying under-Serviced Neighbourhoods in Canadian Urban Areas. *PloS One*, *11*(12), e0168208. https://doi.org/10.1371/journal.pone.0168208

Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge university press.

Shulman, M., Wai, J. M., & Nunes, E. V. (2019). Buprenorphine treatment for opioid use disorder: An overview. *CNS Drugs*, 1–14.

Sinha, R. (2008). Chronic stress, drug use, and vulnerability to addiction. *Annals of the New York Academy of Sciences*, *1141*, 105.

Song, X., Mitnitski, A., Cox, J., & Rockwood, K. (2004). Comparison of machine learning techniques with classical statistical models in predicting health outcomes. *Medinfo*, 736–740.

Stewart, K., Li, M., Xia, Z., Adewole, S. A., Adeyemo, O., & Adebamowo, C. (2020). Modeling spatial access to cervical cancer screening services in Ondo State, Nigeria. *International Journal of Health Geographics*, *19*(1), 28. https://doi.org/10.1186/s12942-020-00222-4

Stopka, T. J., Amaravadi, H., Kaplan, A. R., Hoh, R., Bernson, D., Chui, K. K., Land, T., Walley, A. Y., LaRochelle, M. R., & Rose, A. J. (2019a). Opioid overdose deaths and potentially inappropriate opioid prescribing practices (PIP): A spatial epidemiological study. *International Journal of Drug Policy*, *68*, 37–45.

Stopka, T. J., Amaravadi, H., Kaplan, A. R., Hoh, R., Bernson, D., Chui, K. K., Land, T., Walley, A. Y., LaRochelle, M. R., & Rose, A. J. (2019b). Opioid overdose deaths and potentially inappropriate opioid prescribing practices (PIP): A spatial epidemiological study. *International Journal of Drug Policy*, *68*, 37–45.

Thornton, J. D., Dwibedi, N., Scott, V., Ponte, C. D., Ziedonis, D., Sambamoorthi, N., & Sambamoorthi, U. (2018). Predictors of transitioning to incident chronic opioid therapy among working-age adults in the United States. *American Health & Drug Benefits*, *11*(1), 12.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, *46*(sup1), 234–240.

Townsend Peterson, A., Papeş, M., & Eaton, M. (2007). Transferability and model evaluation in ecological niche modeling: A comparison of GARP and Maxent. *Ecography*, *30*(4), 550–560.

Trachsel, M., & Telford, R. J. (2016). Estimating unbiased transfer-function performances in spatially structured environments. *Climate of the Past*, *12*(5).

University of Chicago. (n.d.). *GeoDa*. Retrieved March 12, 2021, from https://geodacenter.github.io/

US Census Bureau. (n.d.). *Glossary*. The United States Census Bureau. Retrieved November 25, 2020, from https://www.census.gov/programs-surveys/geography/about/glossary.html

US Census Bureau. (n.d.). *South Carolina*. The United States Census Bureau. Retrieved April 14, 2021, from https://www.census.gov/geographies/reference-files/2010/geo/state-local-geo-guides-2010/south-carolina.html

USDA. (n.d.). *USDA - Rural-Urban Commuting Area Codes*. Retrieved March 10, 2021, from https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/

Vapnik, V. (2000). Statistics for engineering and information science. *The Nature of Statistical Learning Theory*.

Wan, N., Zhan, F. B., Zou, B., & Chow, E. (2012). A relative spatial access assessment approach for analyzing potential spatial access to colorectal cancer services in Texas. *Applied Geography*, *32*(2), 291–299.

Wan, N., Zou, B., & Sternberg, T. (2012). A three-step floating catchment area method for analyzing spatial access to health services. *International Journal of Geographical Information Science*, *26*(6), 1073–1089.

Wang, F. (2012). Measurement, optimization, and impact of health care accessibility: A methodological review. *Annals of the Association of American Geographers*, *102*(5), 1104–1112.

Wang, L. (2007). Immigration, ethnicity, and accessibility to culturally diverse family physicians. *Health & Place*, *13*(3), 656–671. https://doi.org/10.1016/j.healthplace.2006.10.001

Warburton, D. E., Nicol, C. W., & Bredin, S. S. (2006). Health benefits of physical activity: The evidence. *Cmaj*, *174*(6), 801–809.

Weiss, A. J., Heslin, K. C., Stocks, C., & Owens, P. L. (2020). *Hospital Inpatient Stays Related to Opioid Use Disorder and Endocarditis, 2016: Statistical Brief# 256*.

Wiemken, T. L., & Kelley, R. R. (2020). Machine learning in epidemiology and health outcomes research. *Annual Review of Public Health*, *41*, 21–36.

Wikipedia. (2020). South Carolina. In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=South_Carolina&oldid=989936925

Wilson, N. (2020). Drug and opioid-involved overdose deaths—United States, 2017–2018. *MMWR. Morbidity and Mortality Weekly Report*, *69*.

Yang, Z., Wilsey, B., Bohm, M., Weyrich, M., Roy, K., Ritley, D., Jones, C., & Melnikow, J. (2015). Defining risk of prescription opioid overdose: Pharmacy shopping and overlapping prescriptions among long-term opioid users in medicaid. *The Journal of Pain*, *16*(5), 445–453.

Zedler, B., Xie, L., Wang, L., Joyce, A., Vick, C., Kariburyo, F., Rajan, P., Baser, O., & Murrelle, L. (2014). Risk factors for serious prescription opioid-related toxicity or overdose among Veterans Health Administration patients. *Pain Medicine*, *15*(11), 1911–1929.