Project 2458 | July 2025

SJSU SAN JOSÉ STATE UNIVERSITY

MTI MINETA TRANSPORTATION INSTITUTE

Understanding Residential Development in a High-Quality Transit Area (HQTA): An Application of Deep Learning

Dohyung Kim, PhD

CSUTC California State University Transportation Consortium

CalPolyPomona

MINETA TRANSPORTATION INSTITUTE

Founded in 1991, the Mineta Transportation Institute (MTI), an organized research and training unit in partnership with the Lucas College and Graduate School of Business at San José State University (SJSU), increases mobility for all by improving the safety, efficiency, accessibility, and convenience of our nation's transportation system. Through research, education, workforce development, and technology transfer, we help create a connected world. MTI leads the California State University Transportation Consortium (CSUTC) funded by the State of California through Senate Bill 1 and the Climate Change and Extreme Events Training and Research (CCEETR) Program funded by the Federal Railroad Administration. MTI focuses on three primary responsibilities:

Research

MTI conducts multi-disciplinary research focused on surface transportation that contributes to effective decision making. Research areas include: active transportation; planning and policy; security and counterterrorism; sustainable transportation and land use; transit and passenger rail; transportation engineering; transportation finance; transportation technology; and workforce and labor. MTI research publications undergo expert peer review to ensure the quality of the research.

Education and Workforce Development

To ensure the efficient movement of people and goods, we must prepare the next generation of skilled transportation professionals who can lead a thriving, forward-thinking transportation industry for a more connected world. To help achieve this, MTI sponsors a suite of workforce development and education opportunities. The Institute supports educational programs offered by the Lucas Graduate School of Business: a Master of Science in Transportation Management, plus graduate certificates that include High-Speed and Intercity Rail Management and Transportation Security Management. These flexible programs offer live online classes so that working transportation professionals can pursue an advanced degree regardless of their location.

Information and Technology Transfer

MTI utilizes a diverse array of dissemination methods and media to ensure research results reach those responsible for managing change. These methods include publication, seminars, workshops, websites, social media, webinars, and other technology transfer mechanisms. Additionally, MTI promotes the availability of completed research to professional organizations and works to integrate the research findings into the graduate education program. MTI's extensive collection of transportation-related publications is integrated into San José State University's world-class Martin Luther King, Jr. Library.

Report 25-18

Understanding Residential Development in a High-Quality Transit Area (HQTA): An Application of Deep Learning

Dohyung Kim, PhD

July 2025

A publication of the
Mineta Transportation Institute
Created by Congress in 1991

College of Business
San José State University
San José, CA 95192-0219

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
25-18		

4. Title and Subtitle Understanding Residential Development in a High-Quality Transit Area (HQTA): An Application of Deep Learning	**5. Report Date** July 2025
	6. Performing Organization Code
7. Authors Dohyung Kim, PhD, orcid.org/0000-0001-5399-3710	**8. Performing Organization Report** CA-MTI-2458
9. Performing Organization Name and Address Mineta Transportation Institute College of Business San José State University San José, CA 95192-0219	**10. Work Unit No.**
	11. Contract or Grant No. SB1-SJAUX_2023-26
12. Sponsoring Agency Name and Address State of California SB1 2017/2018 Trustees of the California State University Sponsored Programs Administration 401 Golden Shore, 5th Floor Long Beach, CA 90802	**13. Type of Report and Period Covered**
	14. Sponsoring Agency Code

15. Supplemental Notes
10.31979/mti.2025.2458

16. Abstract
Considerable research addresses the dynamics between land use and public transportation, including the impacts of transit systems on urban development and vice versa. However, this research has underestimated the heterogeneity of contributing factors to urban developments near transit by property type: vacant lot and occupied property. That is, what drives development in vacant lots differs from what drives development in already-occupied property. A lack of comprehensive data that differentiates development records based on the type of property is one explanation for this research gap. This project fills this research gap by employing a deep learning algorithm and separately examining the contributing factors to vacant lot and occupied property development near transit. A deep learning model—a foundation model—detects and classifies residential development on vacant and occupied properties in high-quality transit areas (HQTAs) in Los Angeles County. Taking the classification as the dependent variable, this research constructs two multi-level logistic regression models: vacant and occupied models. The findings confirm the heterogeneity of the contributing factors to the development of vacant and occupied parcels. The findings also indicate that while the factors at the property level are more significant than the neighborhood-level factors in both models, the significance of the property characteristics in vacant lot development is more significant than in occupied lot development. While infill development on occupied lots in core cities is more likely to occur, the inner suburban cities tend to experience residential development on both vacant and occupied properties much less than the other areas. Another interesting contrast between vacant and occupied lot development is that occupied lot development is more likely to be associated with urban functions/amenities, while the vacant lot development tends to occur in areas with tranquility (where amenities are fewer) near transit. These insights highlight the need for strategies specific to property type in transit-oriented development planning and suggest that nuanced, data-driven approaches are essential to promoting efficient urban growth around transit hubs.

17. Key Words	18. Distribution Statement
High quality transit areas, residential development, infill development, densification of suburbs, deep learning.	No restrictions. This document is available to the public through The National Technical Information Service, Springfield, VA 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	80	

Form DOT F 1700.7 (8-72)

DOI: 10.31979/mti.2025.2458

Mineta Transportation Institute
College of Business
San José State University
San José, CA 95192-0219
Tel: (408) 924-7560
Fax: (408) 924-7565
Email: mineta-institute@sjsu.edu

transweb.sjsu.edu/research/2458

Acknowledgments

I thank my student assistant from Cal Poly Pomona, Omar Monsalvo, for assisting with the literature review, data collection, training data construction, and spatial analysis for variable measurement. I also thank Ava Hoffman at the California Department of Housing and Community Development and her colleagues who provided feedback. Additionally, I want to sincerely thank Dr. Hee-Jun Jung and Mr. Ji-Hwan Kim, who participated in this project as consultants. They provided excellent support for developing the deep learning algorithm and statistical models.

Contents

Acknowledgments vi

List of Figures ix

List of Tables x

Executive Summary 1

1. Introduction 3

2. Literature Review 5

    2.1 Residential Infill Development 5

    2.2 Upzoning and Downzoning for Residential Development 7

    2.3 Upzoning and Downzoning for Residential Development 8

    2.4 Public Transit and Development 9

    2.5 Detection of Residential Development 11

3. Research Methods 13

    3.1 Research Context 13

    3.2 Research Design.... 15

    3.3 The Identification of Residential Development.... 16

    3.4. The Identification of Contributing Factors to Residential Development 25

4. Vacant Model Results 34

    4.1 Vacant Model Outputs 34

    4.2. Findings from the Vacant Model Outputs.... 37

5. Occupied Model Results 42

5.1 Occupied Model Outputs 42

5.2 Findings from the Occupied Model Outputs 47

6. Conclusion 51

Appendix A: The Variables at the Property Level 53

Appendix B: The Variables at the Neighborhood Level 57

Bibliography 61

About the Author 69

List of Figures

Figure 1. Study Area: High-Quality Transit Areas (HQTAs) in Los Angeles County......... 14

Figure 2. Research Framework Diagram.. 16

Figure 3. Example of Longitudinal Aerial Image Acquisition 18

Figure 4. Example of Image Data Cleaning.. 18

Figure 5. Locations of the Labeled Samples by Type.. 19

Figure 6. Examples of Development Types.. 21

Figure 7. Architectures of Deep-Learning Models 22

Figure 8. Spatial Distribution of the Study Parcels by Development Type........................... 25

Figure 9. Job Centers in the Study Area.. 30

LIST OF TABLES

Table 1. Identification of Study Parcel 14

Table 2. Estimation Accuracy of the Models 23

Table 3. Confusion Matrix of Model 3 24

Table 4. Development Classification Results 24

Table 5. Independent Variables at the Property Level 28

Table 6. Independent Variables at the Neighborhood Level 31

Table 7. Descriptive Statistics of the Independent Variables (Vacant Model) 35

Table 8. Vacant Model Statistics 36

Table 9. Vacant Parcels by Level of Land Use Diversity 41

Table 10. Descriptive Statistics of the Independent Variables (Occupied Model) 44

Table 11. Occupied Model Statistics 46

Table 12. Parcel Type Comparison Between Core Cities and Other Areas 49

Executive Summary

Transportation and land use interactions mainly consider mutual relationships since transportation plays a role in connecting the focal points of economic, social, and recreational activities placed at separate locations and regulated by land use. Considering public transit as a sustainable mobility option, considerable research has addressed the dynamics between land use and transportation, including the impacts of transit systems on land use changes and urban development and vice versa. However, the research underestimated the heterogeneity of contributing factors to land use changes and urban developments near transit by property type: vacant lot and occupied property (with existing buildings). One of the reasons that this has been understudied is the lack of comprehensive data that systematically archives development records based on the types of property.

The purpose of this research project is to fill this research gap by employing a deep learning algorithm and separately examining the contributing factors to vacant lot and occupied property development near transit. For this empirical analysis, a deep learning model—a foundation model—was used to detect and classify residential development on vacant and occupied properties. The model compares and analyzes longitudinal satellite images at a parcel level in high-quality transit areas (HQTAs) in Los Angeles County. The foundation model analyzed 186,519 parcels in the study areas and detected 7,243 developed parcels. It classified them into four types with an accuracy of 94.3 percent: occupied unchanged (169,146 parcels), vacant unchanged (10,130 parcels), occupied developed (5,112 parcels), and vacant developed (2,131 parcels).

Taking the outputs of the classification as the dependent variable, I constructed two multi-level logistic regression models: one model for vacant parcels and another model for occupied parcels. Both models presented moderately strong fits, with Pseudo-R squared values of 0.520 for the vacant model and 0.539 for the occupied model. The findings confirm the heterogeneity of the contributing factors to the development of vacant and occupied parcels. While the factors at the property level were more significant than the neighborhood level factors for both vacant and occupied lot development, the variations in the significance of the factors at the property level were identified. The significance of the property characteristics in vacant lot development stands out more than occupied lot development does. While infill development on occupied lots in the core cities is more likely to occur, inner suburban cities tend to experience residential development on both vacant and occupied properties much less than other areas. Another interesting contrast between vacant and occupied lot development is that occupied lot development is more likely to be associated with the urban functions/amenities, while vacant lot development tends to occur in areas with tranquility near transit.

This research provides insight into adjusting local and regional housing development policies and developing new policies in ways that are customized for the unique characteristics of potential developable properties. This research especially emphasizes the heterogeneity of the contributing factors to the development of vacant and occupied parcels. This may provide a good starting point for local governments in developing customized sustainable development policies. Municipal governments may need to pay attention to how to make the vacant lots at less desirable locations more attractive, while reducing overall regulations for promoting occupied lot development. A regional collaboration between the core cities and suburban cities, particularly inner suburban cities, needs to be considered in terms of sharing and transferring the core cities' experiences with infill development.

Especially, it is important to incorporate developers into policy development since the likelihood of both vacant and occupied lot developments reflects developers' motivation for profitability. This research also found that people's location preferences were nuanced, balancing between tranquility and a preference for urban amenities. However, it is unclear whether the variations are truly due to people's location choices or the availability of developable properties in the areas with urban amenities. Future studies may need to clarify this topic by further investigating people's residential choices in high-quality transit areas. Nonetheless, this research contributes to deepening the discussion about the importance of land use and transportation integration for sustainable urban development.

1. Introduction

Transportation and land use interactions mainly consider mutual relationships since transportation plays a role in connecting the focal points of economic, social, and recreational activities placed at separate locations and regulated by land use. In planning theory, scholars define the relationship between the activity locations (to which land use is related) and accessibility (to which transportation is related) (Giuliano, 2017; Newman & Kenworthy, 1996). Thus, sustainable, smart urban development strategies have emphasized land use and transportation coordination that promotes the enhancement of public transportation services, the promotion of multi-modal transportation connections, and the accommodation of alternative transportation modes such as pedestrian, bike, and micro-mobility (Handy, 2005; Levine et al., 2019). In the State of California, transit-oriented infill development has been employed as a major strategy for reducing automobile dependency (e.g., vehicle miles traveled [VMT]). According to the Southern California Association of Governments (SCAG)'s 2024/2050 Regional Transportation Plan (RTP), for example, the region's growth strategies will accommodate 66 and 54 percent of forecasted household and employment growth in the areas with high quality transit services such as priority development areas (PDAs), neighborhood mobility Areas (NMAs), and transit priority areas (TPAs) (SCAG, 2024).

Considering public transit as a sustainable mobility option, considerable research has addressed the dynamics between land use and transportation. Extensive studies have provided important insights into the direct impact of transit systems and infrastructure on land use changes and housing development (Cervero & Landis, 1997; Kim et al., 2016b) and vice versa (Cervero & Kockleman, 1997; Ewing & Cervero, 2001; Kim et al., 2016a). In the urban context, as a smart growth strategy, infill development aims to provide new opportunities for underutilized properties in areas with existing transit and utility infrastructures (Jun et al., 2017; USEPA, 2014). Like their urban counterparts, municipalities in suburban areas also utilize transit accessibility as a key element for housing development and densification of their communities (Qian et al., 2024; Dong, 2016). The studies highlight that transit investments, such as metro rail transit, light rail transit, and bus rapid transit, can stimulate land development by improving accessibility and increasing property values, although the degree of impact varies (Cao & Pan, 2016).

However, the expected nature and magnitude of a transit system's impact on nearby infill development dynamics are relatively understudied (Kim et al., 2016b). Although prior research extensively discussed the potential impacts of transit systems on nearby land use changes, it often neglected the substantial systemic variations in land use outcomes by associated factors such as transit system types and local plans. More specifically, the research underestimated the heterogeneity of contributing factors to land use changes and urban developments near transit by property type: vacant lot and occupied property (with existing buildings). It is reasonable to hypothesize that vacant lots' attributes associated with development may differ from how occupied properties are attracted to development. It is also more effective for local municipalities to develop customized policies and strategies that attract development separately by type.

Nonetheless, there is a lack of research that addresses the heterogeneity in association with transit. One of the reasons that this has been understudied is the lack of comprehensive data that systematically archives development records based on the types of property (Jun et al., 2024; Kim et al., 2024). Although it is popular to detect urban development and land use changes employing government documents and remote sensing data with GIS technologies (Gabbe, 2018; Kim et al., 2022; Reba & Seto, 2019; Viana et al., 2019), these approaches hardly articulate development by parcel type in addition to other challenges such as discrepancies between the documents and reality and inconsistency at a regional level.

The purpose of this research is to fill this research gap by employing a deep-learning algorithm and separately examining the contributing factors to vacant lot and occupied property development near transit. For this empirical analysis, the research utilizes a deep learning model to detect and classify residential development on vacant and occupied properties. The model compares and analyzes longitudinal satellite images at a parcel level in high-quality transit areas (HQTAs) in Los Angeles County. Taking the outputs of the classification as the dependent variable, I constructed two multi-level logistic regression (MLR) models: one model for vacant parcels and another model for occupied parcels. Each model tests the influencing attributes at the property and neighborhood levels on vacant and occupied lot development. This approach enabled the identification of unique contributing factors to development in vacant and occupied lots. The results provide local municipalities with insight to adjust their housing development policies and develop new policies in ways that are customized for the unique characteristics of potential developable properties. It also contributes to deepening the discussion about the importance of land use and transportation integration for sustainable urban development.

2. Literature Review

Considerable research has addressed the systemic connections between residential/housing development and transportation. Considering public transit as a sustainable mobility option, extensive studies have provided important insights into how a change or intervention on the side of transit systems and infrastructure can affect land use changes, especially residential development. The spatial patterns and land use change dynamics of the development vary. As transit systems tend to form a regional network, the development occurs not only in urban areas but also in suburban areas. Residential development can be triggered by upzoning and downzoning. This literature review comprehensively explores the diverse aspects of residential development in association with public transit.

2.1 Residential Infill Development

Infill development broadly refers to urban development in already built-up areas, often near the center of cities. According to the definition by the American Planning Association (APA), infill development is (re)development that "increases density of development and the adaptive re-use of existing buildings, resulting in efficient utilization of land resources, more compact urban areas, and more efficient delivery of quality public services." (APA Policy Guide on Smart Growth, 2012) As a smart growth strategy, infill development aims to provide new opportunities for underutilized properties in areas with existing transportation and utility infrastructures (Jun et al., 2017; USEPA, 2014). In this way, infill development aims to limit excessive urban expansion and lead to sustainable development. Thus, cities widely accept infill development as an alternative to urban sprawl. Particularly, cities experiencing a shortage of housing supply have paid attention to infill development as a policy tool that can increase housing stock and densify their residential areas.

Although infill development is a popular urban development strategy, the definition is debatable. Farris (2001) applied an extremely broad definition of infill development that encompassed all residential development occurring within a central city. Wiley (2009) defined infill development more specifically as average lot size below a certain level of the development within the periphery of existing development. Another scholar argued that rehabilitation, replacement (with new structures at higher density), and expansion of existing buildings are considered as infill development, particularly if it increases the number of units built (Wheeler, 2001), while the literature separated redevelopment from infill development since the opportunities and underlying economic conditions of redevelopment and infill development are fundamentally different (Knaap, 2003).

Because the definition of infill development is loose and because infill development is complex and associated with multiple contributing factors, estimating the capacity of and the contributing factors to infill development is not straightforward (McConnell & Wiley, 2010). They include a wide range of parameters such as the characteristics of properties, neighborhood features, socioeconomic, demographic characteristics, environmental conditions, and many more. Kim et al.'s (2022) study found that infill development is more influenced by the physical characteristics of the property (e.g., built year and lot size) and neighborhood (e.g., land-use diversity and housing mix) than by the socioeconomic characteristics of the neighborhood (e.g., population density and crime rate). They also reported that transportation accessibility, such as the distance to rail transit, is important in infill development.

Employing deep learning and multilevel modeling, another study directly detected and analyzed residential infill development in the City of Los Angeles (Jun et al., 2024). The findings revealed that larger lot sizes, older properties, and greater bus transit accessibility were positively associated with infill development. In contrast, higher property values and floor area ratios (FAR) had a negative impact. At the neighborhood level, infill development was more prevalent in areas with more white residents and newer infrastructure. Contrary to conventional assumptions, infill development was not concentrated in central business districts but expanded outward.

In practice, infill development is often used interchangeably with brownfield development, redevelopment, revitalization, and/or mixed-use development. For example, local governments promote a revitalization policy that converts vacant properties to a mixed-use development in urban areas, aiming to play a key role in efficient land use, community revitalization, and economic development (Downs, 2005). This revitalization can be called infill development. Due to a lack of research on contributing factors to infill development and this mix-use in practice, it is beneficial to learn insights from redevelopment and brownfield development research.

Taking brownfield redevelopment in urban areas as a case, Green (2018) argued that contributing factors to the redevelopment include socio-economic factors (income levels), green development (sustainable building practices), and tax incentives. They also underscored the importance of integrating sustainability into redevelopment efforts and highlighted the need for targeted policy measures to support brownfield revitalization in lower-income areas. Bourne (1969) also examined factors influencing redevelopment patterns in central areas of Toronto, Canada, focusing on where new construction occurs and why. Findings indicated that redevelopment was concentrated in areas with larger lot sizes, proximity to transit, and established clusters of high-income housing. Similarly, Lewis (2012) examined the structural, neighborhood, and policy-related factors influencing renovation and redevelopment in Baltimore. The study found that structural characteristics, such as building age, size, and assessed value (older and larger properties), were strong predictors of renovation (older and larger properties being more likely to undergo improvements). Neighborhood factors such as higher vacancy rates and proximity to the central business district and the policies (e.g., federal historic districts, heritage areas, and community legacy zones) positively impact renovation. Surprisingly, however, properties closer to transit stations were less likely to be renovated.

While the growing housing demand in central cities can be met by residential infill development, various barriers, such as land assembly difficulties, higher infrastructure, and other regulatory costs prevent infill development (Farris, 2001). Additionally, significant challenges hinder the implementation of infill development, such as regulatory hurdles of restrictive zoning and permitting processes, increasing costs, and delaying projects (Wiley & McConnel, 2010). Strong opposition from residents, often driven by fears of congestion, decreased property values, and loss of open space, further complicates infill efforts. Furthermore, social issues, such as racial tension and lower school quality in the central city, can also act as barriers to infill development (Suchman et al., 1997).

At the same time, some negative impacts of infill development on social context in urban communities are well documented, such as gentrification and displacement (Myers & Gearin, 2001; Fainstein, 2005). Advocates of new urbanism and smart growth have insisted that this development promotes social and income diversity in varying degrees and conditions by alleviating spatial segregation exacerbated by urban sprawl (Duany et al., 2000; Talen, 2006). In a study on the spillover effects of new construction in San Francisco, Pennington (2021) concluded that new housing construction benefits incumbent tenants by reducing evictions and the risk of displacement, although gentrification occurred over time. However, there is growing consensus on the negative impacts of gentrification on social diversity triggered by housing development. They argue that high-density and mixed-use development does not guarantee the promotion of social diversity (Kim, 2015; Dale & Newman, 2009; Bramley & Power, 2009). Redevelopment and infill projects in urbanized areas tend to displace low-income households (Dale & Newman, 2009).

2.2 Upzoning and Downzoning for Residential Development

Land use changes can promote residential development in urban areas. For example, changing land uses, such as upzoning or downzoning, can provide physical spaces for residential development, although it is not exactly the same as infill development. Upzoning refers to development that requires changes in land use to a denser zoning type (e.g., single family residential zones to multifamily residential zones), whereas downzoning means the process of reducing the allowed intensity of development in a specific area (e.g., retail zones to residential zones). Examining how site-specific and locational factors influence land use change, Wilder (1985) indicated that land use succession, primarily between single-family and multi-family housing, was concentrated near the city core, whereas conversion mainly resulted in new single-family housing in the periphery. Parcels undergoing succession were typically older, located closer to the CBD, and exhibited extreme variations in size. Converted parcels, in contrast, were farther from the CBD and had smaller acreages than other vacant sites.

Gabbe (2018) examined the factors of upzoning. While upzoning does not exactly connote infill development, it is similar because it allows a higher development intensity in urban areas. The study finds that upzoning is related to various characteristics, including surrounding amenities (e.g., proximity to the beach and higher-performing schools), physical characteristics at the property level (e.g., slope and lot size), physical characteristics at the neighborhood level (e.g., population density), neighborhood housing characteristics (e.g., homeownership rate), and neighborhood socioeconomic characteristics (e.g., change in rent).

As an example of downzoning, Yang et al. (2022) addressed the adaptive reuse of declining malls. Economic feasibility remains dominant in decision-making, but legal barriers and regulatory hurdles present significant challenges. While malls have the potential to be repurposed for housing, developers are reluctant primarily to pursue reuse projects due to high renovation costs, rigid layouts, and uncertain returns. Similarly, a case study analyzed the potential of repurposing vacant malls (NARRG, 2020). The study recognized the main challenges of securing enough funding and cutting through bureaucratic barriers.

Moving forward, a proactive approach is needed to make land use changes a viable solution for increasing the housing supply. Policymakers may need to consider implementing financial incentives and zoning flexibility to encourage rezoning projects, while planners and developers need to engage communities early in the process to ensure residential development on repurposed spaces aligns with the local context. Without financial incentives, leadership, and a market-driven approach, underutilized properties such as vacant malls will stay empty, dragging down surrounding property values.

2.3 Residential Densification in Suburbs

A large volume of research has paid attention to the spatial patterns of suburban housing development and the factors of the built environment in the development. Similar to their urban counterparts, the core strategies of suburban municipalities often involve increasing population and housing densities. Multi-family housing development can play a crucial role in densifying suburban communities. Despite the stereotypical suburban image of single-family detached housing units, multi-family housing has grown in U.S. suburbs. A Canadian study reported that the suburbanization of rental housing is a global phenomenon caused by the financialization of rental housing by rental corporations targeting displaced low-income renters from central urban areas to less affluent suburban neighborhoods (August & Walks, 2018).

Urban proximity has been a significant component of inclusive housing development as employment accessibility and other activity opportunities are critical for disadvantaged populations (Schultheiss et al., 2024). In the same vein, studies confirmed the positive roles of inner suburbs in high-density development (Lee & Leigh, 2005; Chakraborty et al., 2010). The studies suggested that urban amenities such as public transit, retail/commercial, and job accessibility increase housing development in suburban communities. Inner suburbs are more likely to be suburban areas equipped with urban amenities. The findings from Levin and Inam (2004) indicated that many developers recognize significant market interest in transit-oriented development, especially in inner suburban areas where accessibility to transit hubs is high. A study investigating the redevelopment of single-family homes in Chicago's inner-ring suburbs indicated that smaller homes, lower floor area-to-lot size ratios, and properties undervalued relative to their surroundings are more likely to be redeveloped (Charles, 2011). Proximity to transit and high-ranking school districts further increases the likelihood of redevelopment. However, racial disparities are evident; areas with higher proportions of Black and Hispanic residents are significantly less likely to see reinvestment.

Qian et al. (2024) emphasized transit accessibility as a key element for housing development in suburbs. Transit stations in suburban Portland, Oregon, positively influence housing growth and increased population density (Dong, 2016). It found that stations in operation since 2004 saw housing developments grow by 40%, compared to the regional average of 18%. The suburbanization trends in recent decades show that the migration of low-income households and people of color to the suburbs has been expedited (Kim & Kim, 2023; Howell & Timberlake, 2014). This may reflect the simultaneous increase in low-income population and transit access in the suburbs (Liu & Bardaka, 2021). Thus, multi-family housing development in suburban cities can play a critical role in accommodating the shifts in affordable housing demand (Qian et al., 2024). Another study suggested that lower-income households are being pushed to suburban, auto-dependent areas (Kramer, 2018). While transit-oriented development encourages density, rising housing costs near transit hubs often exclude those who need public transportation most. While lower-income and racially diverse populations often live near transit, affordable housing is increasingly found in areas with poor transit access.

Even though transportation accessibility is important for densifying suburbs, proximity to employment and activity locations is not always what people want. More people prefer tranquility to the proximity to employment and activity locations as well as transit access (Schultheiss et al., 2024). Neighborhood safety and attractive appearance are also significant elements associated with the satisfaction of neighborhoods among both urban and suburban residents (Lovejoy et al., 2010). For example, high-end condominiums are developed in high-amenity areas beyond transit accessibility (Atkinson-Palombo, 2010).

Although many suburban municipalities have paid attention to the applications of smart growth strategies to their jurisdictions, densification in suburban cities is not always welcome. Restrictive land use regulations and NIMBYism (Not in My Backyard) negatively affect housing development (Manville & Monkkonen, 2024). Although constructing high-density housing with the same levels of amenities and qualities as traditional suburban homes can be a strategy to overcome NIMBYism (Danielsen et al., 1999), this approach may work against affordable housing development as it can increase the prices of new housing.

2.4 Public Transit and Development

A rich volume of research has explored the impacts of transit systems on land development and housing growth. The research is not limited to the U.S. but is well conducted worldwide. For example, a study in a global context highlighted that transit investments, such as metro rail transit, light rail transit, and bus rapid transit, can stimulate land development by improving accessibility and increasing property values, although the degree of impact varies (Cao & Pan, 2016). It also found that rapid transit investments often encourage higher-density development and mixed-use projects. While areas with already high accessibility may experience limited development effects, transit expansion generally leads to increased building permits, adaptive reuse of old buildings, and new commercial opportunities. In some cities, such as Curitiba and Bogotá, bus rapid transit (BRT) has encouraged higher-density development, particularly when combined with supportive zoning and transit-oriented development policies (Stokenberga, 2014). In cities such as Seoul, BRT investments have intensified land use (e.g., converting single-family homes into higher-density condominiums). A study on a new rail corridor in Perth, Australia reported that real estate values in TOD areas tend to rise, which can push lower-income households further away from transit, making affordability a concern (Olaru et al., 2011). As TOD locations became more desirable, housing prices rose, sometimes limiting access for lower-income households.

In general, much of the research that focuses on the cities in the U.S. also points out the contribution of transportation and transit accessibility to infill development in urban contexts (Kim et al., 2022; Jun et al., 2024). A study provides insight into land use changes by the phases of transit development in Minneapolis (e.g., pre-construction, construction, and operational phase) (Hurst & West, 2014). The study found varying effects across different property types. Industrial parcels near light rail stations were the most likely to undergo land use changes, often redeveloped into multi-family housing or commercial uses, particularly in central areas closer to downtown Minneapolis. Single-family properties also experienced a slight increase in the likelihood of conversion near light rail stations, primarily into higher-density residential or mixed-use developments. Vacant land, often expected to be a key focus of redevelopment near transit, showed little significant transformation.

Mejias and Deakin (2005) examined the challenges and opportunities associated with urban arterial redevelopment, focusing on San Pablo Avenue from a developer's perspective. Their findings indicated that regulatory complexities, fragmented land ownership, and financial constraints often hinder redevelopment efforts. Developers cited zoning restrictions, lengthy approval processes, and community resistance as significant barriers. Thus, areas with clear planning guidelines, infrastructure investments, and public-private partnerships saw higher redevelopment success. Transit accessibility and mixed-use potential were also identified as key drivers of revitalization.

Public transit is a key element of housing development beyond urban areas (Liu & Bardaka, 2021; Qian et al., 2024). Many suburban municipalities employ strategies that densify their communities, especially along major transit corridors. The Portland study resonates with this point, confirming the positive influence of transit service on housing growth in suburban areas (Dong, 2016). Development was most successful near stations with high transit ridership, vacant land zoned for residential or mixed-use purposes, and a mix of residential and nonresidential land uses in the suburban areas of Portland. However, areas with existing multifamily housing saw slower density increases, likely because of limited developable land.

Similarly, a study investigated how rail transit investments influence land use changes in the suburban areas of Los Angeles, focusing on the Gold Line Phase I corridor in Los Angeles County (Pasadena to East Los Angeles) (Kim et al., 2016b). This research found that vacant parcels near transit stations were more likely to be developed for urban purposes, including residential and commercial uses. These effects were more pronounced around stations with higher ridership, where transit investments contributed to infill development and facilitate the redevelopment of industrial parcels into multifamily housing and open spaces.

Another group of researchers conducts analyses in a regional context that incorporates both urban and suburban areas (Kim & Li, 2021; Levin & Inam, 2004). Analyzing how transit-oriented development (TOD) influences the densification of residential land uses, a study explored upzoning, the conversion of single-family homes to multifamily housing, in transit-rich areas in Southern California (Kim & Li, 2021). It found that land use intensification was more likely to occur where transit services were available or planned. The research also suggests that high-quality transit areas (HQTAs) promote denser housing developments and influence local zoning decisions, creating a feedback loop between transit investment and land use changes. Another study explored developers' perceptions of transit-oriented and pedestrian-friendly development (Levin & Inam, 2004). The study suggests that zoning laws, parking mandates, and land-use restrictions actively prevent compact, mixed-use, and transit-supportive developments regardless of urban and suburban areas. Rather than a lack of consumer interest, the study highlights how zoning laws and planning policies force developments to remain low-density, ultimately limiting their potential to support robust transit networks.

Regarding regulations, the literature commonly points out local regulatory constraints and community resistance as the primary challenge for housing development around transit corridors (Mejias & Deakin, 2005; Kim et al., 2016b; Levin & Inam, 2004; Stokenberga, 2014). If regulatory barriers were lifted, developers would have greater freedom to create walkable, high-density, transit-integrated neighborhoods, leading to more sustainable and accessible urban environments. It highlights the need for policy reforms to streamline the development process and foster stakeholder collaboration. By addressing regulatory and financial barriers, cities can create more attractive conditions for investment, ultimately enhancing the economic and social vitality of transit corridors.

2.5 Detection of Residential Development

A large volume of research has attempted to detect urban development. A conventional approach is to utilize government documents, such as zoning codes and building permits, combined with other data, such as aerial images (Gabbe, 2018; Kim et al., 2022). Although this method allows the articulation of types of urban development at a parcel level, this approach includes multiple limitations, including discrepancies between documents and reality, inconsistencies between municipalities, and more. Thus, employing this conventional approach for multiple jurisdictions presents challenges.

Another popular approach is to detect the physical changes in natural and built environments, primarily employing remote sensing data with GIS technologies (Reba & Seto, 2019; Viana et al., 2019). This approach primarily analyzes longitudinal satellite data and detects the changes in the spatial patterns of built-up areas from rural and/or natural land to urban land types (Jensen, 2004; Yuan et al., 2005). With advances in satellite remote sensing technologies and extensive archives of satellite data, innovative methodologies and techniques for the classification and time series analysis of land resources were introduced, which led to reliable results in detection accuracy. Some international studies tracked infill development using a time-series analysis based on satellite imagery and maps (Rahimi, 2016; Abedini & Khalili, 2019). Some recent studies have applied urban informatics and big data approaches (e.g., a multi-dimensional spatial scan technique and a machine learning algorithm) to modeling urban growth and land-use changes (Pijanowski et al., 2014; Pan et al., 2020).

In recent years, deep learning and computer vision techniques have been increasingly used to analyze images to better understand cities. As an alternative to in-person observations, deep learning techniques that analyze images, such as satellite and Google Street View images, can be more time-efficient and less costly (Brownson et al., 2009). High-resolution satellite images at frequent intervals have become excellent assets for the applications of deep learning techniques. Some studies have used satellite images and deep learning to measure land-use classification (Liu et al., 2017; Romero et al., 2016) and the impact of transport infrastructure (Pavlovic et al., 2022). Two studies have applied deep learning to the detection of residential development at a parcel level (Jun et al., 2024; Kim et al., 2024). Both studies attempted to identify infill development by applying a deep learning algorithm that detects changes in longitudinal image datasets. Since applying deep learning to satellite image analysis can detect actual physical changes on the Earth's surface, it is expected to identify infill development that is insufficiently recorded in land-use plans and maps (Makhamreha & Almanasyeha, 2011).

3. Research Methods

3.1 Research Context

This research aims to develop a systemic framework that identifies residential development at a parcel level and estimates the likelihood of parcels to be developed based on contributing factors to the infill development. The research takes the high-quality transit areas (HQTAs) in the County of Los Angeles as the geographical study area. The HQTAs defined by the local metropolitan planning organization (MPO), SCAG, generally refers to the areas with good transit accessibility. More specifically, the HQTAs represent the areas within one-half mile from a "major transit stop" and a "high-quality transit corridor." A "major transit stop" is defined as a site containing an existing rail or bus rapid transit station, a ferry terminal served by either a bus or rail transit service, or the intersection of two or more major bus routes with a frequency of service interval of 15 minutes or less during the morning and afternoon peak commute periods. A "high-quality transit corridor" refers to a corridor with fixed route bus service with service intervals no longer than 15 minutes during peak commute hours.

The HQTAs include most of the City of Los Angeles and also cover the surrounding primary cities, such as Long Beach and Glendale, as well as inner suburban cities such as Pasadena, South Gate, and Montebello (Figure 1). The HQTAs are also stretched to the west border of the county along the corridors between Interstate 10 (I-10), Interstate 210 (I-210), and California 60 (SR-60).

Figure 1. Study Area: High-Quality Transit Areas (HQTAs) in Los Angeles County

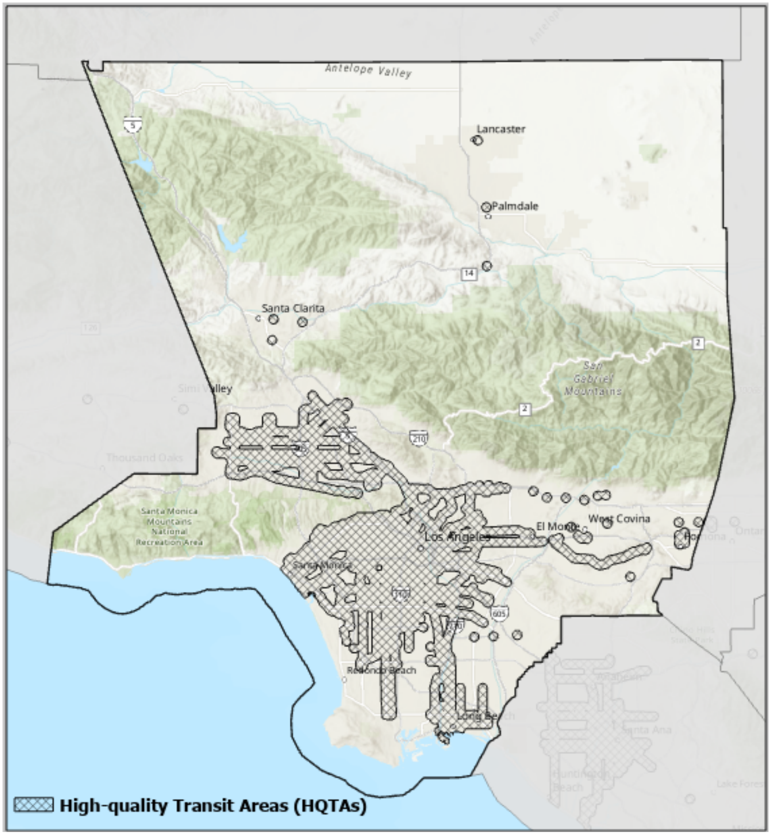


Table 1. Identification of Study Parcel

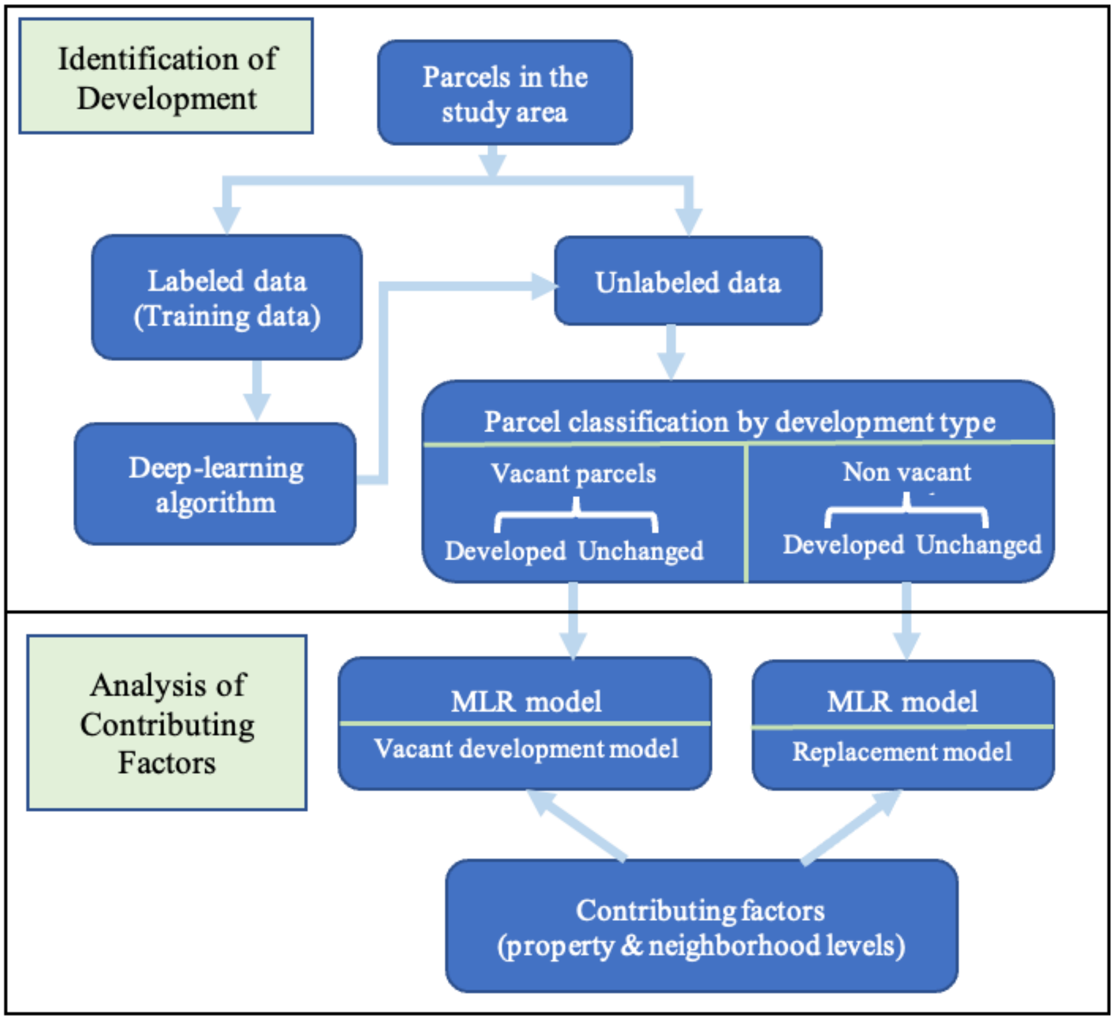
		Parcel Count
Total within HQTA		871,066
Land use	Multi-family	151,406
	Mobile home	5,838
	Mixed residential	15,768
	Mixed use	2,957
	Vacant	13,544
	Publicly owned	275
	Sub total	189,513
Data cleaning		3,269
Total Study Parcel		186,519

It empirically examines the residential development in the study area from 2015 to 2019. This time frame is selected to avoid the influence of the COVID-19 pandemic that may have directly or indirectly influenced housing development. The unit of analysis of this research is the parcel. In other words, it identifies and classifies the development at a property parcel level. Thus, the parcels where residential development can occur are the subject of this study. To conduct this study, the property parcel data was collected from SCAG. In total, there were 871,006 parcels in the HQTAs. Of them, 189,778 parcels were identified as residential developable parcels according to their land use codes (Table 1). The data cleaning process removed 3,269 parcels, resulting in 186,519 parcels as the study parcels.

3.2 Research Design

To achieve the research goal, this research broadly consisted of two stages: (1) identifying residential development and (2) analyzing contributing factors to the development (Figure 2). The identification of residential development primarily consisted of classifying the study parcels. For this identification, we developed a deep learning algorithm that detects the construction of new buildings on parcels by comparing longitudinal aerial images. With the training dataset, the deep-learning algorithm classified the parcels into four types in a nested structure. Based on the status in 2015, the study parcels were classified as either vacant or occupied lots. Depending on the development experience, each category was further divided into two types: parcels that have experienced development or parcels that remained unchanged.

Figure 2. Research Framework Diagram



The second stage involved identifying the properties and neighborhood attributes that influence the development. I constructed two multi-level logistic regression (MLR) models: the vacant and replacement models. The vacant model estimated the likelihood of the development case where a vacant lot has been developed. The replacement model estimated the developments that occurred in parcels with an existing primary building and that replaced the building with a new building construction. Taking the parcel classification as the dependent variable, the models explored both property (micro) and neighborhood (macro) factors influencing the development.

3.3 The Identification of Residential Development

This stage centers around developing the deep learning algorithm that detects the construction of new buildings on parcels by comparing longitudinal aerial images. Thus, this stage consists of three phases: Data Acquisition and Labeling, Deep Learning Algorithm Development, and Estimation of Development.

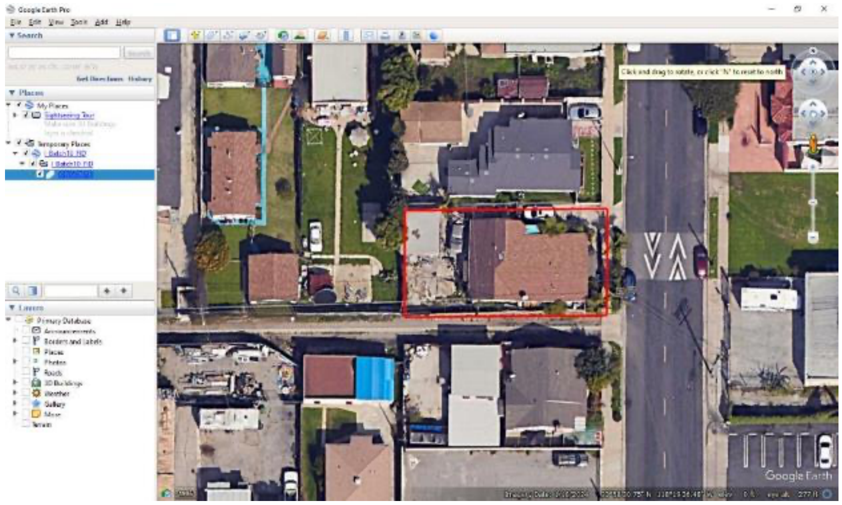
3.3.1 Data Acquisition and Data Labeling

The first phase of identifying residential development involved acquiring aerial images that portray the longitudinal changes over parcels. I collected longitudinal aerial images for each parcel using the Google Earth Engine Application Programming Interface (GEE-API). To capture the study period, I collected all the historical aerial images provided by GEE-API. They included the eight-year images from 2005 to 2020, specifically 2005, 2009, 2010, 2012, 2014, 2016, 2019, and 2020. In other words, 1,492,152 (186,519 8) aerial images were collected for this study. For the data acquisition, I created a Google Earth file (KMZ file) that captured each parcel's boundary. Using the boundary file, I clipped the aerial images within the boundary (Figure 3).

However, we only utilized the four years of the aerial images (2014, 2016, 2019, and 2020) to ensure the detection of changes during the study period (2015–2019). By leveraging this approach, we expected to improve the performance of deep-learning algorithms, as the analysis considered a sequence of images rather than a direct comparison of just two points in time (e.g., 2014 and 2020). This allowed the construction of imagery datasets that contain four imagery files per parcel representing longitudinal changes.

After collecting the imagery files, we performed data cleaning. This refers to the removal of the parts of the image that contain the outside of parcel boundaries. Since the collected images were in the JPG format, the image files are always rectangular. In the case where a parcel is not rectangular in shape, the image files naturally include the areas outside the parcel boundary. However, there was potential for the outside areas to create some noise when comparing images. For example, if there were some changes in the outside areas while there was no change in the parcel, there was a chance that the deep-learning algorithm would recognize the parcel as one that experienced development. To remove this possibility, we erased the parts of the images that capture outside the parcel boundaries (Figure 4).

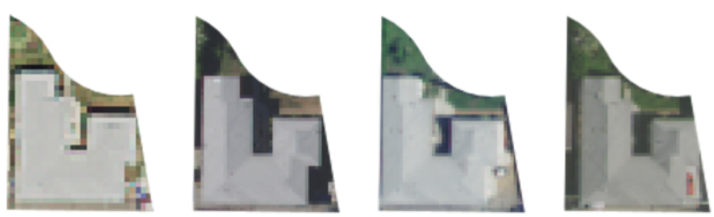
Figure 3. Example of Longitudinal Aerial Image Acquisition



May, 2005 | June, 2009 | April, 2010 | April, 2012

May, 2014 | July, 2016 | February, 2019 | May, 2020

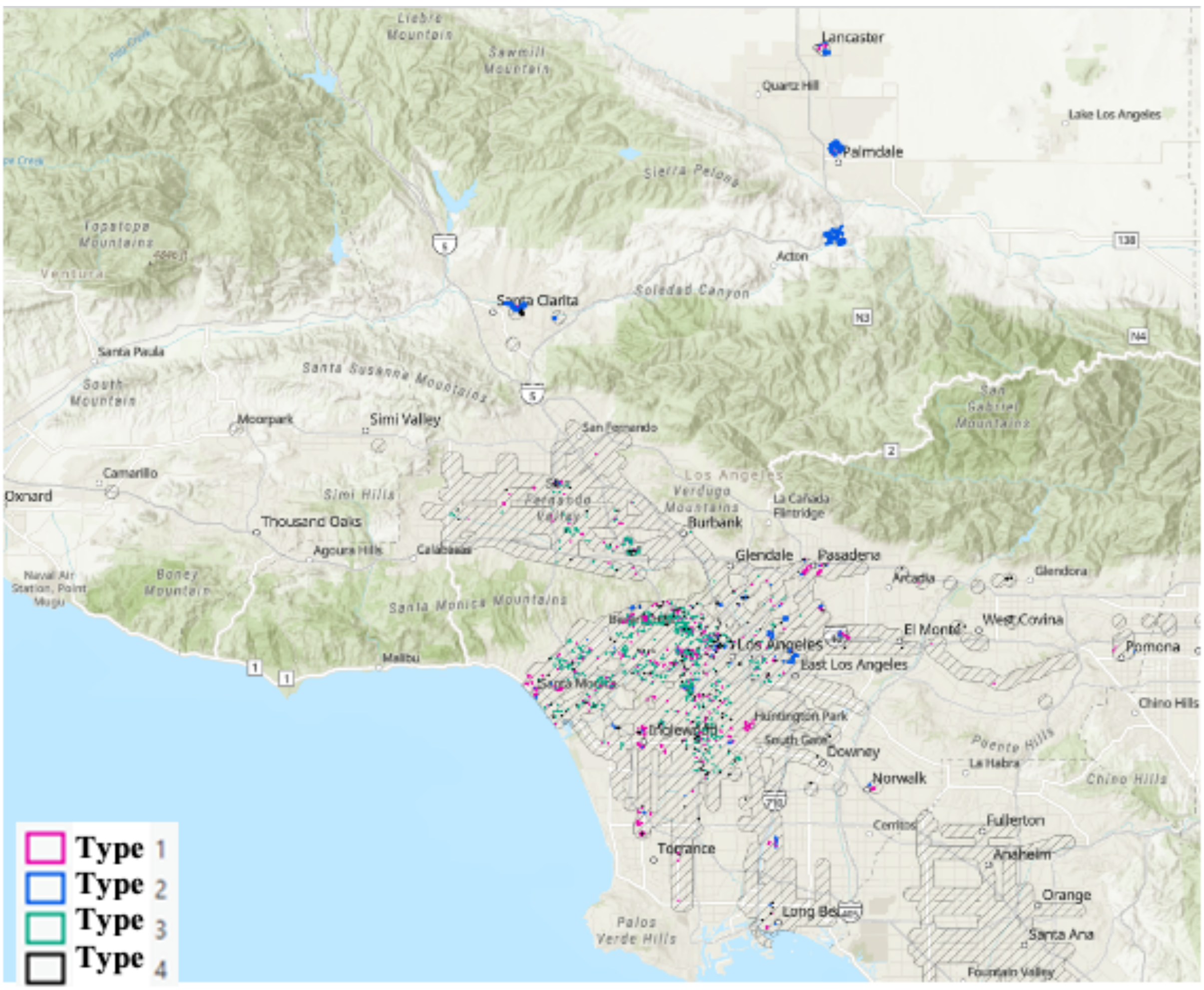
Figure 4. Example of Image Data Cleaning



2014 2016 2019 2020

After archiving the imagery datasets, we performed data labeling. Labeled data, also known as training data, is pivotal in training deep-learning algorithms. Typically, deep-learning algorithms analyze labeled data and produce an inferred function that maps unlabeled datasets. The labeled data refers to raw data annotated with meaningful labels that classify the data's outcomes. In this study, labeling refers to the process of manually identifying and tagging parcels that experienced development during the study period separately from parcels remaining unchanged.

Figure 5. Locations of the Labeled Samples by Type

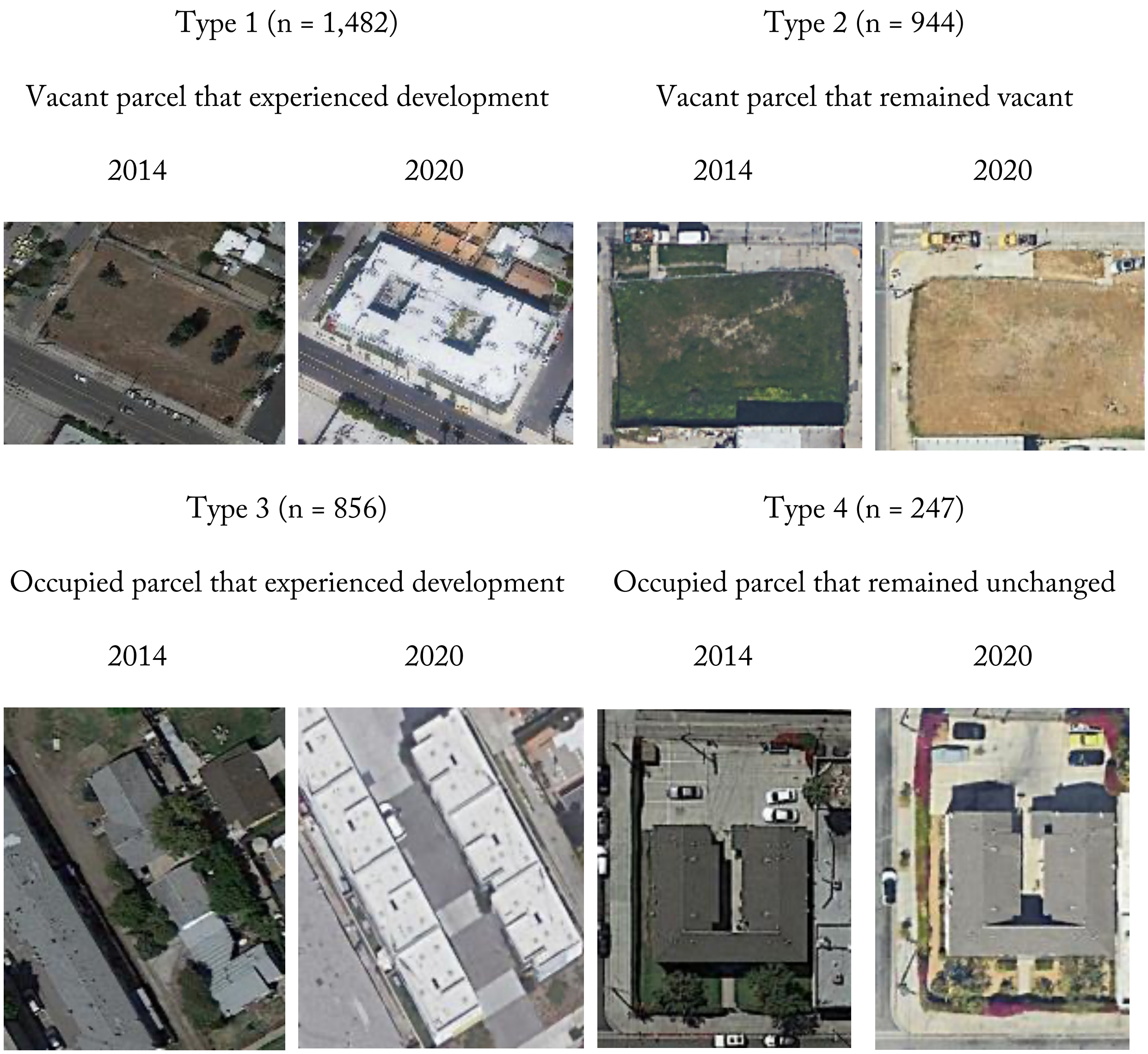


Note: Type 1 = Vacant parcel that experienced development, Type 2 = Vacant parcel that remained vacant, Type 3 = Occupied parcel that experienced development, and Type 4 = Occupied parcel that remained unchanged

The data labeling involved selecting sample parcels, comparing the longitudinal images of the parcels, and identifying the typology of the parcels. This process was carefully performed since the training data would become the reference for the deep-learning algorithms. To capture diverse building and development types, we selected parcel samples that are evenly distributed in the study area spatially. Moreover, we carefully compared the longitudinal images of the parcels and decided on the typology of the parcels. When we were unable to make a clear judgment based on the longitudinal images, we also reviewed the historical Google Street View images. Comparing the longitudinal aerial images to the Google Street View images led us to label the typology confidently.

Throughout this data labeling process, 3,529 parcels, corresponding to 14,116 images, were selected and labeled (Figure 5), while the remaining images were allocated to the test dataset. The training dataset was used explicitly for identifying patterns, whereas the test dataset was employed for model evaluation and validation (Lachenbruch & Mickey, 1968). We tagged one of the labels to the 3,529 parcels manually. The labels included the following: Type 1 (occupied parcels that remained unchanged), Type 2 (vacant parcels that remained vacant), Type 3 (occupied parcels that experienced development), and Type 4 (vacant parcels that experienced development). The occupied parcels refer to parcels with an existing primary building (Figure 6). It is noteworthy that the number of parcels was imbalanced by type. For example, the count of label 1 was much higher than the count of label 4. Although this imbalanced training data presented a challenge in developing the deep-learning algorithm, which will be explained in the next chapter, it reflects reality. In other words, a significant number of parcels remained unchanged compared to the developed parcels, while the number of vacant parcels was much smaller than the parcels occupied by buildings.

Figure 6. Examples of Development Types



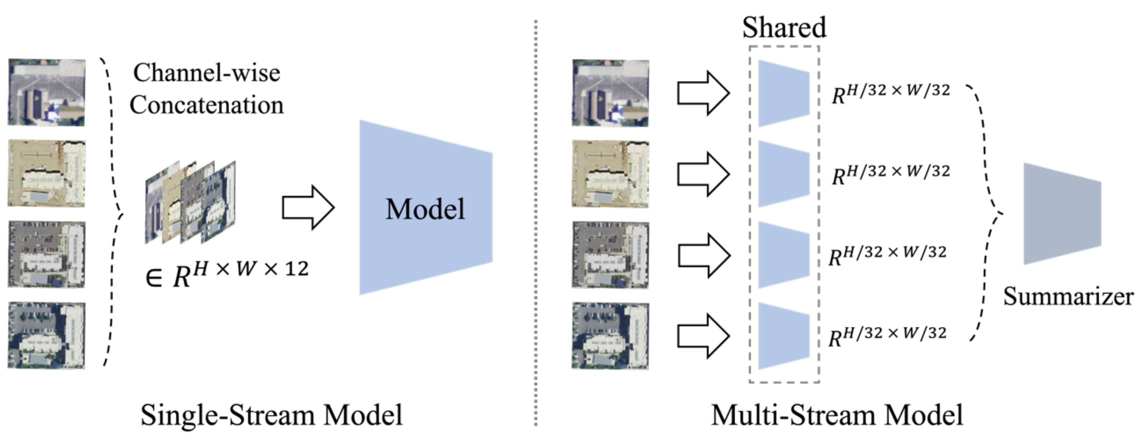
3.3.2 Developing the Deep-Learning Algorithm

This research developed multiple deep-learning algorithms, compared their performances, and identified the best algorithm. Primarily, two image classification deep-learning architectures, ResNet-50 and RemoteCLIP, were employed. ResNet-50 is one of the most popular convolutional neural network (CNN) backbone architectures. In CNN models, one or multiple convolution layers extract features from the input by executing convolution operations. Each layer is a set of nonlinear functions of weighted sums at different coordinates of spatially nearby subsets of outputs from the prior layer, which allows for the reuse of the weights. Applying various convolutional filters, CNN machine learning models can capture the high-level representation of the input data, making CNN techniques widely popular in computer vision tasks. Since a ResNet-50 model is a CNN that is 50 layers deep, it is used to denote the variant that can work with 50 neural network layers. Thus, it can effectively dissect a picture, identify its objects and scenes, and categorize them accordingly.

Given the relatively small dataset, we also adopted RemoteCLIP, a foundation model specialized in satellite imagery analysis. Rather than develop a deep-learning algorithm from scratch, a foundation model fine-tunes algorithms based on pre-trained models such as GPT-3, BERT, and DALL-E that have absorbed wide-ranging knowledge from massive datasets and have undergone extensive preliminary training. Thus, this approach excels at processing data sequences through attention mechanisms that dynamically evaluate the importance of different inputs. This design enables the models to generate coherent and contextually relevant outputs across various data types, including text and images.

Recent advancements in large-scale multimodal data training have demonstrated strong generalizability in various downstream tasks such as contrastive language-image pre-training (CLIP) (Radford, 2021). CLIP is designed to understand images in conjunction with textual descriptions. This multimodal model can perform tasks that require linking images with relevant text, making it exceptionally useful in applications that span both visual and textual data.

Figure 7. Architectures of Deep-Learning Models



To optimize the deep learning model for our analysis, we explored both single-stream and multi-stream architectures for handling extracted images (Figure 7). The single-stream model concatenates multiple time-series images into a single composite image with dimensions $R^{H \times W \times 12}$, where H and W denote image height and width, and 12 represents the combined channels from four images. This consolidated image serves as the model input. Conversely, the multi-stream model processes multiple images separately through a shared model, producing predictions by summarizing outputs from multiple branches. The single-stream model offers computational efficiency due to its simpler architecture, but may suffer from lower detection capabilities. In contrast, though computationally intensive, the multi-stream model offers greater flexibility and accuracy in identifying development patterns.

In summary, we tested three model configurations: (1) a single-stream ResNet-50 model trained from scratch (Model 1), (2) a multi-stream ResNet-50 model trained from scratch (Model 2), and (3) a pre-trained multi-stream model based on RemoteCLIP (Model 3). Since pre-trained models typically process RGB images, the single-stream architecture was incompatible with RemoteCLIP. Therefore, we employed RemoteCLIP within a two-stream architecture, replacing the model component while keeping other processing elements intact. It is noteworthy that RemoteCLIP itself was not retrained, ensuring that only the additional framework components underwent learning.

3.3.3 Deep-Learning Algorithm Testing and Selection

Each model was evaluated to determine its performance in classifying urban development within the training dataset. The best-performing model was subsequently used to classify all parcels in the study. Following model training, we assessed performance using the K-fold cross-validation protocol, an evaluation technique that divides data into K non-overlapping subsets: each subset served as a test dataset once, while the remaining subsets were used for training, ensuring a comprehensive evaluation (Lachenbruch & Mickey, 1968). We employed a 10-fold cross-validation approach, wherein nine subsets were used for training and one for testing in each iteration.

Table 2. Estimation Accuracy of the Models

Model	Overall Accuracy	Class-average Accuracy
Model 1	0.9110	83.022
Model 2	0.9071	82.555
Model 3	0.9434	91.584

Overall, all tested models accurately estimated urban development patterns (Table 3). The results include both overall accuracy and class-average accuracy. The former represents performance across all categories, while the latter evaluates accuracy within individual classes before averaging the results. Although the overall accuracy of each model showed minimal variance, Model 3 exhibited significantly superior class-average accuracy compared to the others.

Based on these findings, Model 3 was identified as the best-performing model. The confusion matrix for the first fold illustrates Model 3's performance, achieving an overall accuracy of 0.9434 (Table 4). While the model demonstrated exceptional accuracy in classifying “unchanged” parcels, including both Types 2 and 4, it showed relatively poorer performance in identifying the “developed” cases (Types 1 and 3). Additionally, the model sometimes misclassified the “occupied developed” cases (Type 3) as “occupied unchanged” (Type 4) at a slightly higher rate (0.09). Nonetheless, the results confirmed that the pre-trained multi-stream model (Model 3) offered the most reliable classification of urban development changes within Los Angeles residential parcels. Thus, we employed Model 3 for the rest of the research.

Table 3. Confusion Matrix of Model 3

		Predicted			
		Type 1	Type 2	Type 3	Type 4
Actual	Type 1	0.95	0.05	0.00	0.00
	Type 2	0.00	1.00	0.00	0.00
	Type 3	0.09	0.01	0.90	0.00
	Type 4	0.04	0.08	0.00	0.88

Note. Type 1 = Occupied unchanged, Type 2 = Vacant unchanged, Type 3 = Occupied developed, and Type 4 = Vacant developed

3.3.4. Estimation of Study Parcel Typology

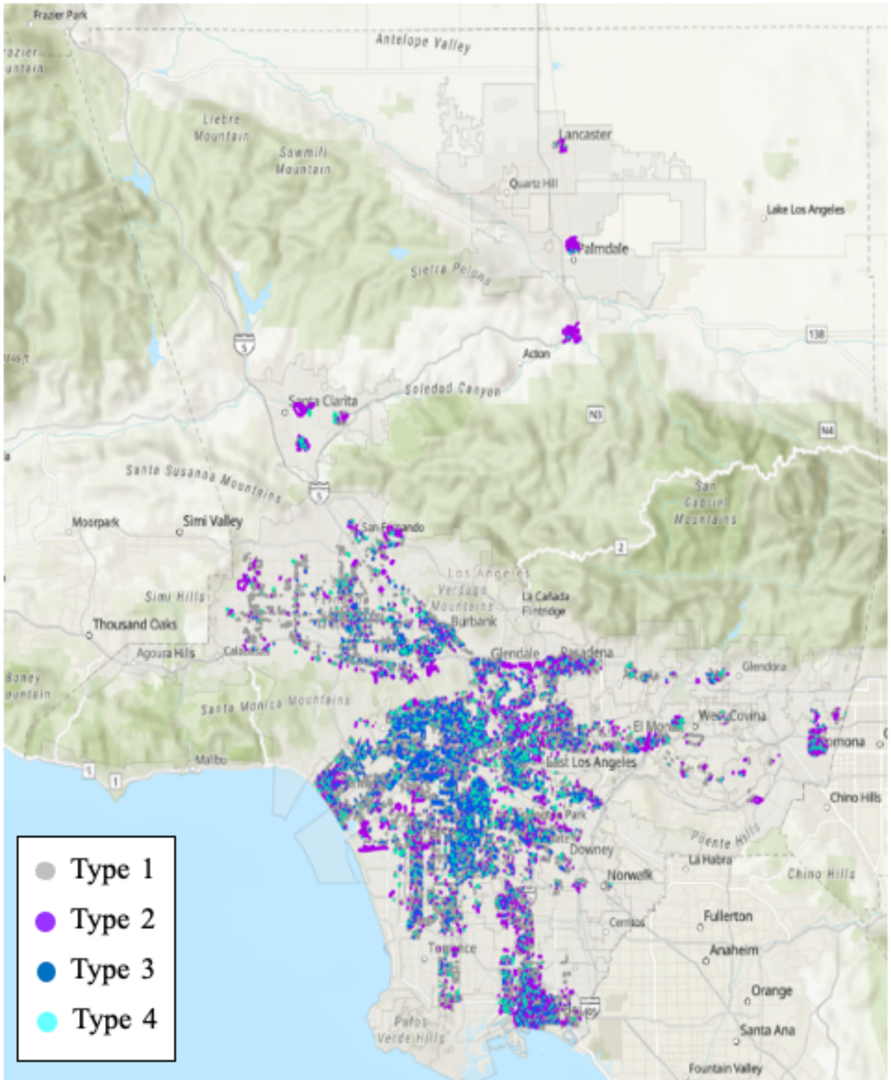
We employed Model 3 to classify the typology of the study parcels. The 186,519 study parcels were broadly divided into occupied (Types 1 and 3 = 174,258) or vacant (Types 2 and 4 = 12,261) parcels (Table 5). In other words, the study area was predominantly occupied by the parcels occupied by buildings (93.1%). About 3.8 percent of the study's parcels experienced residential development, including 5,112 occupied and 2,131 vacant parcels.

Table 4. Development Classification Results

Development Type	Count	%
Type 1: Occupied unchanged	169,146	90.7
Type 2: Vacant unchanged	10,130	5.4
Type 3: Occupied developed	5,112	2.7
Type 4: Vacant developed	2,131	1.1
Total	186,519	100.0

All the development types were scattered in the study area rather than showing a precise spatial distribution of a particular development type (Figure 8). However, there was a certain level of consistency between the development of occupied and vacant parcels. The aggregation of development was clear in downtown Los Angeles. After a gap, the development stretched to the west and south-west bounds, including Crenshaw, Inglewood, Culver City, Century City, and Hancock Park. The concentration of the vacant lot development was closer to downtown Los Angeles than the occupied development. This included areas such as Silver Lake, Echo Park, and Vermont Square. At a distance from downtown Los Angeles, it was also clear that the development of vacant and occupied lots occurred in the San Pedro area.

Figure 8. Spatial Distribution of the Study Parcels by Development Type



3.4 Identification of Contributing Factors to Residential Development

Taking the classification estimates made by Model 3 as the dependent variable, we employed multi-level logistic regression (MLR) modeling to explore the contributing factors of residential development. We employed the MLR models because of the hierarchical structure of the contributing factors. The factors affecting the residential development of each property would vary and were likely to have a nested relationship since some were unique features of each parcel, while others were the conditions of the neighborhood in which a property was located. Thus, for this data, neighborhood conditions are factors shared by multiple properties located within the same neighborhood. In other words, the property–neighborhood relationships are nested within a neighborhood.

Because of the nested data structure, the independence assumption of standard regression models is violated and, as a result, underestimates the standard errors of the regression coefficients. A multilevel model can partition the variance between the neighborhood level (Level 2) and the property level (Level 1) and use level-specific variables to explain the variance at each level. As shown below, the Level 2 equation (Eq. (2)) is nested in the intercept of the Level 1 equation (Eq. (1)), thereby creating a mixed model (Eq. (3)).

Level 1 residential development: $\ln(\frac{p_{ij}}{1-p_{ij}}) = {}_{0} + {}_{i}X_{ij} + e$ (1)

Level 2: ${}_{0} = \beta_{00} + \beta_{0j}W_{j} + r_{0}$ (2)

Mixed model: $\ln(\frac{p_{ij}}{1-p_{ij}}) = \beta_{00} + \beta_{0j}W_{j} + {}_{i}X_{ij} + r_{0} + e$ (3)

where, p = probability of residential development,

i = property,

j = neighborhood,

X = property characteristics,

W = neighborhood characteristics,

${}_{i}$ = Level 1 coefficients,

β_{0j} = Level 2 coefficients,

β_{00} = intercept,

e = Level 1 residual, and

r_0 = Level 2 residual.

We constructed two multi-level logistic regression (MLR) models, a vacant model and an occupied model. The vacant model consists of parcels classified into Types 2 (vacant, unchanged) and 4 (vacant, developed), while Types 1 (occupied, unchanged) and 3 (occupied, developed) were incorporated into the occupied model. I did this for theoretical and practical reasons. Theoretically, it is reasonable to distinguish between the features of vacant and occupied parcels due to the discrepancy in their characteristics. For example, while the features of existing buildings, such as building value and floor area ratio (FAR), do not apply to vacant parcels (those with no building), they may be important for occupied parcels. Separating them also has practical value, as the number of vacant parcels is significantly smaller than the number of occupied parcels. Under these circumstances, it will be challenging to identify the unique contributing factors to vacant parcel development. Therefore, it is reasonable to separate vacant lots from occupied lots and build two models separately. The dependent variable of the MLR models is a binary nominal variable (0 or 1) identified by the deep learning model. Residential properties that experienced development were coded as 1, while those that did not experience infill development were the reference group (coded as 0). In other words, in the occupied model, Types 1 and 3 parcels were coded as 0 and 1, respectively. In the vacant model, Types 2 and 4 parcels were coded as 0 and 1, respectively.

3.4.1. Measurement of the Independent Variables at the Property Level

The independent variables in the models represent the contributing factors to the development. They comprise the characteristics of properties (Level 1) and neighborhoods (Level 2) where the properties belong.

The Level 1 variables measure a wide range of attributes of the study parcels. The variables were categorized into four broad categories: Property Characteristics, Transportation Accessibility, Accessibility to Amenities, and Locational Characteristics (Table 6). The Property Characteristics category consisted of variables that represent physical characteristics of the properties. They included lot size, property age, building value, publicly owned property, property value, and floor area ratio (FAR). The variables were included in much previous research as important determinants of property development (Charles, 2011; Gabbe, 2018). The variables were extracted from the property appraisal office's parcel data. The parcel data was collected from SCAG. It is noteworthy that the Yr_Bult, Imp_Lnd, and Prt_Far variables were excluded from the vacant model. Since the variables are associated with the building in the parcels, they are irrelevant to vacant parcels. Thus, they were only considered in the occupied model.

Table 5. Independent Variables at the Property Level

Category	Name	Description
Property characteristics	Prt_Area	Lot size (in acres)
	Yr_Bult	Building age (2025 – the year the building was built)
	Imp_Lnd	The ratio of building value (building value/land value)
	Pbl_Prop	Dummy (1 if the parcel is owned by the public; 0 otherwise)
	Apr_Val	Property appraisal value (in dollars)
	Prt_Far	Floor area ratio (FAR)
Transportation accessibility	Bus_Trn	Dummy (1 if there is a bus station within 1/4 mile; 0 otherwise)
	Hwy_Acc	The inverse distance to the nearest highway ramp (in miles)
	Rail_Trn	Dummy (1 if there is a rail station within 1/2 mile; 0 otherwise)
Accessibility to amenities	Shp_Cntr	The inverse distances to the nearest shopping center (in miles)
	Prk_Acc	The inverse distance to the nearest public park (in miles)
	Cbd_Acc	The inverse distance to downtown Los Angeles (in miles)
	Job_Cntr	The inverse distance to the nearest job center (in miles)
Locational characteristics	Innr_Sub	Dummy (1 if the parcel is within the inner suburban city; 0 otherwise)
	Core_Cty	Dummy (1 if the parcel is within one of the Core cities; 0 otherwise)

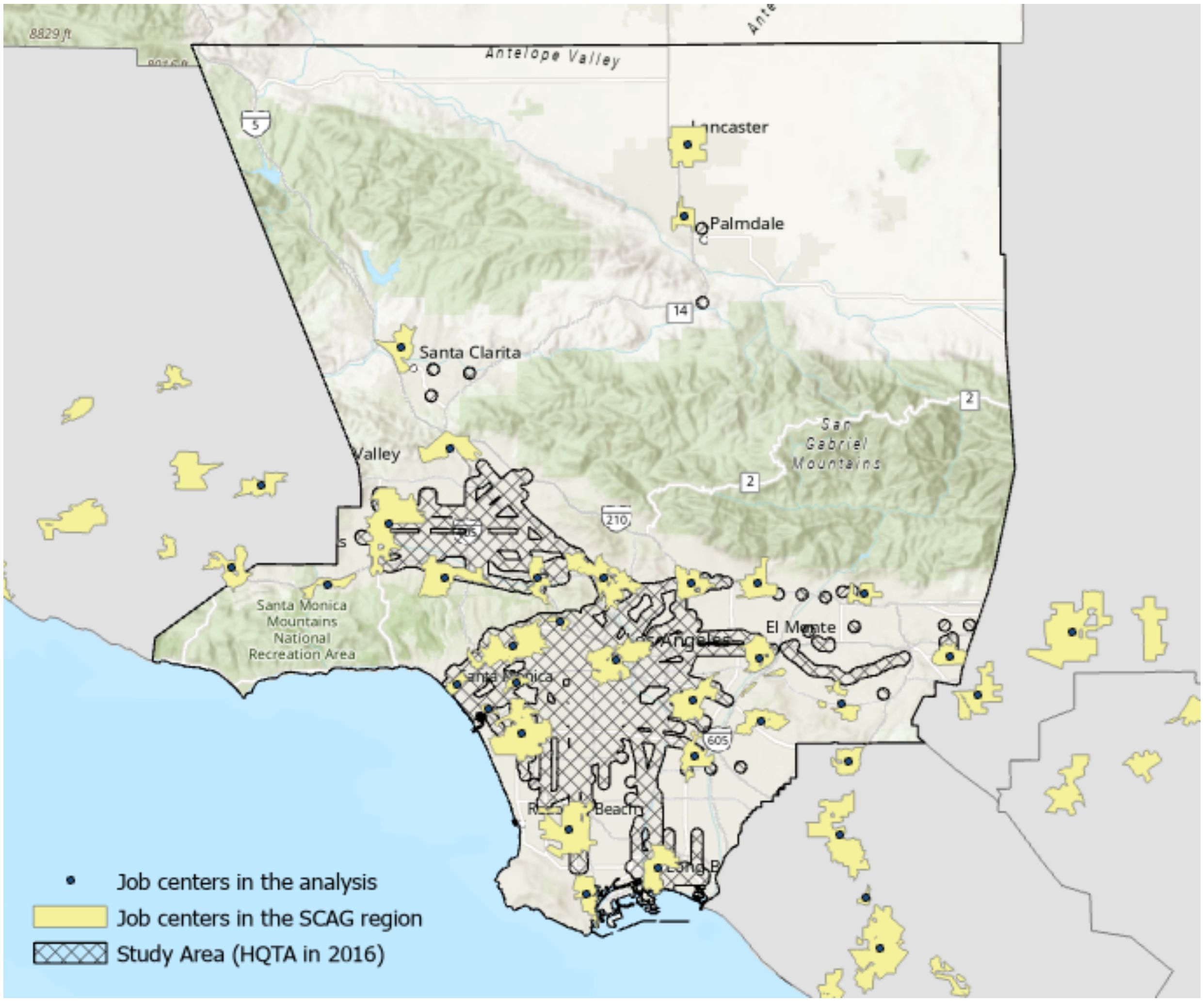
Given that transportation accessibility is also an important component of property characteristics, we measured highway, bus transit, and rail transit accessibility (Gabbe, 2018; Kim et al., 2022). Although the study area offers good transit accessibility overall, transportation accessibility is still important since the area are mixed with different types of transit services as well as automobile accessibility. We examined the variations of the transportation accessibility. Bus and rail transit accessibility variables were defined in a binary nominal form. If the property is located within a 0.25-mile and 0.5-mile buffer from a bus stop and rail station, respectively, the property was coded as 1. Otherwise, it was coded as 0. Automobile accessibility was measured with inverse distance to the nearest highway ramp.

For this analysis, we downloaded and utilized the state-wide transit stop data from the California Department of Transportation (Caltrans). For the highway accessibility variable, we calculated the inverse Euclidean distance between a property and its nearest highway ramp. In other words, the variable is 1/the Euclidean distance. This measurement represents the spatial relationship between the property and highway ramp more intuitively since the properties closer to the nearest ramp get a larger value that indicates better accessibility. For this measurement, we utilized the Census Topologically Integrated Geographic Encoding and Referencing system (TIGER) road file.

As urban amenities are important factors in residential location decisions (Frenkel et al., 2013; van Vliet et al., 2013; Rahimi, 2016; Sant´e et al., 2010) and can thus affect residential development, we included variables on accessibility to amenities such as shopping centers, parks, job centers, and downtown Los Angeles. The locations of the shopping centers and parks were retrieved from the parcel data according to the land use code. There were 28 job centers in Los Angeles County defined by SCAG in 2016 (Figure 9). In addition to the job centers, we included eight job centers located near the county boundary because the parcels near the county boundary can be closer to one of the job centers outside of the county. The job centers included in this study take about 79% of total jobs in the SCAG region. The same inverse Euclidean distance method was employed for all the variables. After generating the centroids of the amenities, we measured the distance between each parcel and the centroids.

The Locational Characteristics variables take into account spatial variations of residential development. Since the study area was primarily located within core cities like Los Angeles, Long Beach, and Glendale, the residential development in the study area could be defined as infill development. Infill development is generally defined as urban development that constructs new buildings or replaces primary buildings in already built-up areas (often near the center of cities and towns), further filling urban areas. As major cities widely accept infill development, the Core_Cty variable attempts to explore whether the residential development aligns with the infill development. Thus, we employed a binary nominal form to adopt the definition of core city by the U.S. Census (Los Angeles, Long Beach, and Glendale). Similarly, previous studies reported the positive roles of inner suburbs in high-density development, balancing between the metropolitan center and the urban fringes regarding jobs and housing. (Lee & Leigh, 2005; Chakraborty et al., 2010). The inner suburbs represent the built-up areas close to principal cities. Inner suburbs have played a pivotal role in increasing multi-family housing development as they are more likely to be the suburban areas equipped with urban amenities. This study divided and coded the study parcels into two categories: those within an inner suburban city (coded as 1) or not (coded as 0). As per Lichter et al. (2023), the inner suburban cities were defined as cities that share a border with one of the core cities. The maps illustrate the property level variables can be found in Appendix A.

Figure 9. Job Centers in the Study Area



3.4.2. Measurement of the Independent Variables at the Neighborhood Level

Level 2 represents the neighborhood to which each property belongs. The neighborhood unit in this research is a census block group that typically includes 600–3,000 people. Although the census tract was also considered as the neighborhood unit, we selected the census block group as the unit. After constructing the datasets based on the census tract and block group, we developed multiple MLR models and compared their performance. This comparison confirmed that the models with census block groups are more stable and reliable. This is probably because of the larger geographical scale of the census tract. A larger unit tends to generalize the attributes of the unit (and lose the details of the attributes) and make sample sizes smaller. For this reason, we selected the census block group as the neighborhood unit. Since housing consumption is collective, consuming neighborhood characteristics as well as housing stock (Leven et al., 1976), we included sociodemographic characteristics, housing characteristics, and land use/urban form variables at the neighborhood level (Level 2) as factors to infill development in addition to the property-level (Level 1) (Table 7).

Table 6. Independent Variables at the Neighborhood Level

Category	Name	Description
Socio/demographic characteristics	Pop_Den	Population density (Population/acres)
	Wht_Pop	Percent of non-Latino/Hispanic White population
	Md_Inc	Median household income
	Clg_Grd	Percent of population with a college or higher degree
	Md_Yrblt	2025 - Median year built
	Sch_Prf	School proficiency index
Housing characteristics	Md_Rnt	Median gross rent
	Md_Hvl	Median owner-occupied housing value
	Mv_In	2025 - Median year moved in
Land use/ Urban form	Bus_Den	Bus stop density (# of bus stops per acre)
	Retl_Prop	The proportion of retail land use
	LU_Dvrst	Land use diversity (land use entropy)
	Job_Prox	Regional job proximity index
	Walk_Inx	National walkability index

The sociodemographic characteristics variables primarily summarize the neighborhoods' social status and population composition. For the sociodemographic characteristics, we included population density, percentage of non-Hispanic whites (hereafter white), median household income, educational attainment, neighborhood age, and school proficiency (Charles, 2011; Gabbe, 2018; Kim et al., 2022). We used the 2015 American Community Survey (ACS) (5-year estimate) data for all the sociodemographic variables except for the Sch_Prf variable. The population density was calculated by dividing the population by the area of the census block group. The neighborhood age variable (Md_Yrblt) is the value subtracted from the median built year of the buildings from 2025. This allows a direct interpretation of how neighborhood age is associated with infill development.

The U.S. Department of Housing and Urban Development's (U.S. HUD) school proficiency index was utilized for the variable (Sch_Prf). The index describes which neighborhoods have high-performing elementary schools nearby and which are near lower-performing elementary schools based on school-level data about the performance of 4th-grade students on state exams. The school proficiency index is a function of the percentage of 4th grade students proficient in reading (r) and math (m) on state test scores for up to three schools (i=1,2,3) within 1.5 miles of the block-group centroid. The school proficiency index is calculated as follows:

$$School_i = \sum_{n=i}^{3} \left(\frac{s_i}{\sum_i^n s_i}\right) \times [\frac{1}{2} \times r_i + \frac{1}{2} \times m_i]$$

where, r = students proficient in reading,

m = students proficient in math, and

s = 4th grade school enrollment.

The housing characteristics variables primarily illustrate a neighborhood's housing value and monthly rent cost. Both represent the median values. Additionally, we calculated the period of homeowners' residence by subtracting the median year a household moved into an owner-occupied housing unit from 2025. All the variables were computed based on the census data (ASC 5-year estimate).

The land use and urban form variables represent built environment conditions that may attract residential development. We had two transportation factors in this category: bus stop density and walkability. The bus stop density variable was measured by dividing the number of bus stops in each census block group by the area of the block group. The walkability variable indicates the national walkability index provided by the U.S. Environmental Protection Agency (USEPA). The index measures the relative walkability of our nation's communities at the census block group level. The index is based on measures of the built environment that affect the probability of whether people walk as a mode of transportation: street intersection density, proximity to transit stops, and diversity of land uses.[1] The index scores are on a scale of 1 to 20. A higher index score indicates a more walkable neighborhood.

We also measured two land use characteristics: retail land use and land use diversity. The retail land use variable (Retl_Prop) represents a proportion of retail land uses in each census block group (acres of retail land uses/acres of the census block group). The land use diversity variable (LU_Drvst) was developed by calculating the land-use entropy, which is commonly used to measure land use diversity (Kim et al., 2018; Cervero & Kockelman, 1997). It reflects the uniformity of land use mixtures, ranging from 0 to 1. The value of 0 indicates homogenous land use (dominated by one land use type), while 1 represents the most diverse and balanced land use mix. The land use diversity is calculated as the following equation:

$$LU\_Drvst_i = \sum \frac{P_i \times LN(P_i)}{LN(i)}$$

where Pi = proportion of land-use category i within the census block group,

[1] For more information about the national walkability index, refer to

i = number of land-use categories, and

LN = natural logarithm of a number.

Additionally, each neighborhood's regional job proximity was also included. This variable is the jobs proximity index developed by the U.S. HUD. The jobs proximity index quantifies the accessibility of a given residential neighborhood as a function of its distance to all job locations within a core-based statistical area (CBSA), with distance to larger employment centers weighted more heavily. Specifically, a gravity model was used, where the accessibility (Ai) of a given residential block-group is a summary description of the distance to all job locations, with the distance from any single job location positively weighted by the size of employment (job opportunities) at that location and inversely weighted by the labor supply (competition) to that location. The index values are percentile ranked with values ranging from 0 to 100. The higher the index value, the better the access to employment opportunities for residents in a neighborhood. More formally, the model has the following specification:

$$A_i = \sum_{i=1}^{n} \left(\frac{E_i d_{ij}^{-2}}{\sum L_j}\right)$$

where i = residential locations,

j = job locations,

d = distance as "as the crow flies" between block-groups i and j,

E = the number of jobs in block-group j, and

L = the number of workers within a CBSA.

Appendix B contains maps illustrating the neighborhood-level variables. After compiling the dependent and independent variables, we constructed the two MLR models. During the data compilation and cleaning process, 292 of the study parcels (186,519) were removed from the dataset. Thus, the total number of parcels included in the models was 186,227. They are located in 3,559 census block groups.

4. Vacant Model Results

4.1 Vacant Model Outputs

The vacant model was constructed with 12,157 parcels on 2,146 census block groups. Of 12,157 parcels, 10,040 remained unchanged (coded as 0), and 2,117 parcels experienced residential development (coded as 1). To construct the vacant model, we tested the multicollinearity of the independent variables and checked their contributing significance. Throughout this process, we selected 20 variables for the model, including 10 (out of 15) at the property level and 10 (out of 14) at the neighborhood level.

Before constructing the vacant model, we computed the descriptive statistics of the variables and conducted an independent samples *t*-test with the variables at the property level (Table 8). The independent samples' *t*-test compared the means of the variables in two groups by the dependent variable: parcels remained unchanged and parcels developed. All the variables under the property characteristics category presented a significant mean difference between the parcels that remained unchanged and those that were developed. While the mean of the appraisal value (Apr_Val) of the developed parcels was significantly higher than the mean of the unchanged parcels, the sizes of the unchanged parcels and their probability of being publicly owned were much higher than those of the developed parcels. Similarly, the variables that presented significant differences in the means between the unchanged and developed parcels included rail transit accessibility and accessibility to shopping centers. Additionally, the proportion of developed properties in the core cities was higher than in other areas at a 90 percent confidence level. This implies that the properties in the core cities are more likely to experience residential development than those in the other areas. These variables tended to significantly contribute to the likelihood of a vacant parcel experiencing residential development. Overall, the variables at the neighborhood level presented consistent distributions indicated by relatively small standard deviation values. Of them, the standard deviations of the variables, population density (Pop_Den), school proficiency (Sch_Prf), and regional job proximity (Job_Prox), were relatively large. This suggests the spatial variations of the variables in the study area.

The vacant model presented a relatively reliable model fit (Table 9). The AIC (51,470.13) and BIC (51,477.31) of the vacant model were lower than the null model (60381.158 and 60388.563, respectively). AIC and BIC are commonly used model fit measures that penalize models with more independent variables. Smaller AIC and BIC values indicate a stronger model. The null model is a reference model customary to run an unconditional model without any independent variables, a one-way ANOVA model with random effects. Thus, the smaller AIC and BIC of the vacant model than those of the null model indicate a strong model fit of the vacant model. The pseudo-R-squared value was also promising: The Pseudo-R-squared value of 0.52 presents the moderate to strong goodness of the model.

Table 7. Descriptive Statistics of the Independent Variables (Vacant Model)

			Pooled (N=12,157)		Unchanged (=0) (N=10,040)		Developed (=1) (N=2,117)		*t*-test	
Level	Category	Variable	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	*t*	*p*-value
Property level (Level 1)	Property characteristics	Prt_Area	0.60	4.41	0.65	4.82	0.39	1.23	4.683	0.000
		Pbl_Prop	0.18	0.38	0.20	0.40	0.08	0.27	17.601	0.000
		Apr_Val (logged)	10.82	1.91	10.61	1.90	11.84	1.62	-29.590	<.0.001
	Transportation accessibility	Hwy_Acc	4.54	12.55	4.52	11.89	4.63	15.28	-0.320	0.749
		Rail_Trn	0.04	0.21	0.04	0.20	0.06	0.24	-3.372	0.001
	Accessibility to amenities	Shp_Cntr	1.16	1.49	1.14	1.47	1.29	1.57	-4.232	0.000
		Cbd_Acc	0.20	0.23	0.20	0.23	0.20	0.26	-0.024	0.981
		Job_Cntr	0.62	1.10	0.61	1.04	0.63	1.33	-0.506	0.613
	Locational characteristics	Innr_Sub	0.19	0.39	0.19	0.39	0.18	0.38	1.435	0.151
		Core_Cty	0.64	0.48	0.64	0.48	0.66	0.48	-1.788	0.074
			(N= 2,146)							
Neighborhood level (Level 2)	Socio/demographic characteristics	Pop_Den	26.8	20.2						
		Md_Inc (logged)	10.82	0.52						
		Md_Yrblt	56.4	23.8						
		Sch_Prf	43.6	26.1						
	Housing characteristics	Med_Rnt (logged)	7.01	0.31						
		Md_Hvl (logged)	12.91	0.54						
		Mv_In	23.1	10.1						
	Land use/ Urban form	LU_Dvrst	0.5	0.1						
		Job_Prox	54.7	29.0						
		Walk_Inx	14.4	2.7						

Table 8. Vacant Model Statistics

Level	Category	Variable	Coef.	Std. Error	*t*	Sig.
		Intercept	-3.276	1.9323	-1.695	0.090
Property level (Level 1)	Property characteristics	Prt_Area	-0.072	0.0297	-2.441	0.015
		Pbl_Prop	-1.376	0.1526	-9.017	0.000
		Apr_Val	0.401	0.0241	16.688	0.000
	Transportation accessibility	Hwy_Acc	-0.002	0.0035	-0.704	0.482
		Rail_Trn	0.114	0.1686	0.676	0.499
	Accessibility to amenities	Shp_Cntr	-0.057	0.0249	-2.301	0.021
		Cbd_Acc	-0.255	0.2435	-1.048	0.295
		Job_Cntr	0.005	0.0339	0.139	0.889
	Locational Characteristics	Innr_Sub	-0.344	0.1418	-2.428	0.015
		Core_Cty	-0.009	0.1226	-0.074	0.941
Neighborhood level (Level 2)	Socio/demographic characteristics	Pop_Den	0.004	0.0036	1.029	0.304
		Md_Inc	-0.301	0.1699	-1.770	0.077
		Md_Yrblt	0.001	0.0019	0.422	0.673
		Sch_Prf	0.001	0.0023	0.514	0.607
	Housing characteristics	Med_Rnt	0.382	0.2367	1.613	0.107
		Md_Hvl	-0.167	0.1382	-1.211	0.226
		Mv_In	-0.002	0.0064	-0.306	0.759
	Land use/ Urban form	LU_Dvrst	-1.791	0.3674	-4.875	0.000
		Job_Prox	0.001	0.0020	0.584	0.559
		Walk_Inx	0.071	0.0185	3.851	0.000
-2 log likelihood			51,468.13	60,379.16 (Null Model)		
AIC			51,470.13	60,381.16 (Null Model)		
BIC			51,477.31	60,388.56 (Null Model)		
Pseudo-R-squared			0.520			

The model outputs imply the significance of the variables at the property level compared to the neighborhood level. While five out of ten variables at the property level present a correlation with the dependent variable at a 95 percent confidence level, two out of ten variables only show the correlation at the neighborhood level. Of the contributing factors at the property level, the property characteristics are the most influential factors to the likelihood of a parcel being developed, as all the variables under this category show a correlation with the dependent variable at a statistically significant level. Additionally, the variables of Shp_Cntr (accessibility to shopping centers) and Innr_Sub (parcels located in inner suburbs) are also significant contributing factors to residential development. While the App_Val variable positively correlates with the dependent variable, the other four variables negatively correlate with the dependent variable. This suggests that the properties with higher appraisal values and the properties that are smaller and located further away from the shopping center are more likely to be developed. The properties publicly owned and located in the inner suburbs are less likely to be developed. At the neighborhood level, only two variables—land use diversity and walkability—under the land use and urban form category, present a correlation with the dependent variable at a 95 percent confidence level. While the walkability variable (Walk_Inx) positively correlates with the dependent variable, the land use diversity variable (LU_Dvrst) negatively correlates.

When applying a 90 percent confidence level, median income (Med_Inc) at the neighborhood level showed a statistically significant correlation with the dependent variable. The negative correlation indicated the higher likelihood of residential development in neighborhoods with lower income levels. However, this study employed a 95 percent confidence level as the threshold for statistical significance. Thus, the model interpretation hereafter was based on a 95 percent confidence ($p < 0.05$), and so these two variables were not included in the list with statistical significance.

4.2. Discussion: Vacant Model

Revitalization policies that convert vacant properties to residential and mixed-use developments have gained popularity in many cities. They aim to fill in gaps in communities and play a key role in efficient land use, community revitalization, and economic development (Downs, 2005). The model outputs provide local governments with insight about residential development on vacant parcels in areas with sound transit accessibility.

Overall, the outputs suggest the significant impacts of the factors at both the property and neighborhood levels on vacant lot development. The vacant model also demonstrates that a considerable share of the variation in vacant lot development could be attributed to neighborhood-level differences. At the same time, the outputs confirm that the factors at the property level are more influential on development than the neighborhood factors. While half of the property-level variables present a statistically significant correlation with the dependent variable, two out of ten neighborhood-level variables only show the correlation. This finding is understandable and expected due to the direct influence of property attributes on developers' decisions. It is reasonable to say that the individual attributes of each property may be critical for development decisions, while neighborhood characteristics have contextual and indirect influence on the decisions.

Perhaps this is also the reflection of the distinctive study area. Unlike broader geographical areas such as a city or county, the study area primarily incorporates the areas along major transit corridors and arterials. The land use, sociodemographic, and housing characteristics of the neighborhoods in the study area tend to be relatively consistent and show fewer variations in comparison to city-wide characteristics, for example. The smaller standard deviations in the descriptive statistics also support this argument. This relative consistency may contribute to diminishing the impacts of the neighborhood-level variation on vacant lot development. Surprisingly, none of the sociodemographic and housing characteristics at the neighborhood level are significant contributing factors to vacant lot development.

Of the factors at the property level, property characteristics are the most significant contributors to development. All the variables under this category present a statistically significant correlation with the dependent variable. This finding aligns with the previous study that emphasized that the property characteristics are directly related to developers' profitability (Jun et al., 2024). Since developers are likely to look for properties that can generate higher returns when developed, the physical and financial features each property can offer will become an important concern. The outputs suggest that smaller-sized and higher-value vacant lots are more likely to be developed. These features are quite in contrast to the profitable features, such as larger size, lower value, and lower floor area ratio (FAR) in the study.

It is hard to straightforwardly compare the outputs with Jun et al. (2024) due to the context of the study, which took a city as the study area and did not distinguish vacant lot development from occupied lot development. However, this contrast portrays interesting findings. The contrast may suggest that developers' interests and willingness to invest in properties are higher in areas with excellent transit accessibility. Developers may not mind developing smaller, high-value properties, as they expect higher returns with denser, intensive development in the HQTAs. Perhaps this also implies their interest in developing properties in desirable areas. Because properties in desirable areas may have higher values, developers seem to prefer to invest in vacant properties at more desirable locations, although the properties are more expensive and smaller. This investment may be able to grant a higher return despite a higher up-front cost. Notably, this finding reinforces the unique contribution of this research that separates vacant parcels from occupied parcels. Vacant properties tend to be of lower value and FAR than occupied properties since there is no building in the vacant properties. As presented above, the number of vacant lots is much smaller than occupied parcels in urbanized areas. Thus, it is hard to identify this finding without separating vacant and occupied properties.

The negative correlation of the publicly owned property variable (Pbl_Prop) under the property characteristics category also raises important policy discussions. Publicly owned properties can be critical assets for housing development, especially in situations with a significant housing supply shortage. However, the outputs imply that the likelihood of the properties being developed is lower than that of those owned by the private sector. Much literature commonly points out regulatory constraints as the primary challenge for housing development (Mejias & Deakin, 2005; Kim et al., 2016b; Levin & Inam, 2004; Stokenberga, 2014). The complexity of the development process on publicly owned properties and/or regulations tied to the properties may make the properties less attractive to developers. A rezoning process can also be a challenge for developers. Another layer of complexity can be associated with the ownership of the properties. For example, while a consistent state-wide policy can guide the development of properties owned by the state, the policies associated with local jurisdictions' properties may vary.

Although future research may need to investigate this issue further, public agencies may need to work to identify and reduce regulations and red tape that discourage housing developments. For instance, in 2019, the State of California created an inventory of all state-owned parcels, identified parcels in excess of state agencies' foreseeable needs, and prioritized them for affordable housing development (Governor Newsom's Executive Order on excess state property [EO N-6-19]). To support the executive order, the California Department of Housing & Community Development (HCD) launched the Excess Sites Local Government Matching Grants Program (LGMG). LGMG aims to accelerate housing production on excess state sites and facilitate a spirit of collaboration between the state, local governments, and selected developers for affordable housing development. It provides grant-based funding to match certain local government funding for selected developers and local governments for predevelopment and development of affordable housing on excess state sites. Following the state's efforts, local jurisdictions may also pay attention to prioritizing locally-owned properties.

According to the outputs, the properties in the inner suburbs are less likely to experience residential development on vacant parcels. This finding contrasts with previous research that emphasized the positive roles of inner suburbs in high-density development (Lee & Leigh, 2005; Chakraborty et al., 2010). The studies suggest that urban amenities such as public transit, retail/commercial, and job accessibility increase residential development, especially multi-family housing development, in suburban communities. Inner suburbs are more likely to be suburban areas with relatively good public transit. As transit accessibility has been critical to affordable housing development (Qian et al., 2024), inner suburbs played a pivotal role in balancing the metropolitan center and the urban fringes regarding urban amenities and residential contexts. This means that residential development may come to inner suburbs that are more likely to be suburban areas equipped with urban amenities.

However, the finding is opposed to this conventional observation. This implies that many inner suburbs are probably built out, particularly in areas with good transit accessibility. It is reasonable to assume that the built environment of inner suburbs became similar to core cities after playing significant roles in absorbing the housing and jobs spilled over from core cities for a long time. Since transit infrastructure is an important amenity to facilitate the densification of suburbs (Newman & Kenworthy, 2006; Seltzer & Carbonell, 2011), major transit corridors in inner suburbs have already become the target areas for suburban densification. This would have left transit corridors in inner suburbs short of vacant lots for housing development, particularly in a metropolitan area such as Los Angeles County. About 19.1 percent (2,321) of the vacant parcels in the model were located in inner suburban cities, which took about 25.6 percent of the study area, while about 63.9 percent (7,772) of the vacant parcels were within the core cities, which took about 58.5 percent of the study area. This finding suggests that future studies may need to pay attention to these changing roles of inner suburbs in the context of sustainable urban development.

In other words, it is important to develop policy strategies that promote infill development and make vacant lots more attractive for housing development in inner suburbs.

In the same vein, it is surprising that there is no significant likelihood of vacant parcel development in the core cities. As a smart growth strategy, infill development has become a policy alternative to urban sprawl and is widely accepted by cities, particularly many metropolitan cities (Rahimi, 2016). Infill development is expected to provide new opportunities for underutilized properties by utilizing the amenities in already built-up areas such as the existing transportation (Jun et al., 2017; USEPA, 2014). However, the finding does not confirm infill development on vacant lots in the core cities. This finding is consistent with a study reporting that bus and rail transit do not align with the location of infill development in the City of Los Angeles (Kim et al., 2022). As discussed previously, it is important to lead housing development to the areas with existing transit service from the perspective of sustainable urban development and smart growth. Since about 64 percent of the vacant lots are within the core cities, as presented above, the cities have more opportunities to utilize the lots.

The negative correlations between land use diversity and shopping accessibility variables present the nuance of the properties experiencing development. Like the previous research reporting that people's residential location choice tends to associate with serenity rather than the proximity to urban functions such as activity locations (Schultheiss et al., 2024), this suggests that the development tends to occur in the areas with tranquility rather than the proximity to shopping and diverse land uses. In other words, development is more likely to spatially align with the areas with residential context rather than urban amenities. Urban amenities and functions are typically associated with vibrant and crowded environments. Since it is reasonable to assume that the study area already offers relatively high levels of urban amenities and intense and diverse land uses, the development tends to follow a balance between tranquility and proximity to urban functions by identifying the locations that offer tranquility in areas where urban functions are relatively high.

Another reason for the negative correlation can be the scarce vacant properties in the neighborhoods with high-quality urban amenities. In other words, the vacant lots in these neighborhoods are already built, and so not many vacant lots are left for development. For example, the number of vacant parcels significantly declines as the level of land use diversity increases (Table 10). Due to the small number of developable vacant properties in the neighborhoods with high-quality urban amenities, vacant lot development inevitably tends to occur in neighborhoods with lower amenities.

Table 9. Vacant Parcels by Level of Land Use Diversity

Land Use Diversity Percentile	Vacant Parcels in Total	Average Vacant Parcels in Census Block Group
1 (0.020–0.264)	2,040	9.533
2 (0.264–0.319)	1,291	6.005
3 (0.319–0.367)	1,200	5.581
4 (0.367–0.411)	1,280	6.009
5 (0.411–0.445)	1,484	6.870
6 (0.445–0.485)	1,296	6.028
7 (0.485–0.527)	1,344	6.280
8 (0.527–0.584)	862	4.009
9 (0.584–0.655)	752	3.498
10 (0.655–0.833)	608	2.841

Although research has emphasized the significance of transit accessibility as an amenity to facilitate residential development, particularly in suburbs (Newman & Kenworthy, 2006; Seltzer & Carbonell, 2011), the outputs present a statistically insignificant correlation. This finding was expected due to the context of the study area, which had good transit accessibility overall. This implies that the subtle differences in transit and highway accessibility do not become a significant attraction for residential development when achieving a certain level of accessibility.

5. Occupied Model Results

5.1 Occupied Model Outputs

Of the 186,227 parcels selected, 174,070 occupied parcels in 3,471 census block groups were included in the Occupied model. Of 174,070 parcels, 5,091 parcels (2.9%) experienced residential development (coded as 1), and 168,979 parcels (97.1%) remained unchanged (coded as 0). For comparison purposes, we kept a consistent set of independent variables in both models. The only difference in the independent variables was that two property characteristic variables were added to the occupied model: Yr_Bult and Imp_Lnd. They were only added to the occupied model since they are associated with buildings on properties. This made 22 variables in total, including 12 at the property level and 10 at the neighborhood level.

Before constructing the vacant model, we computed the descriptive statistics of the variables and conducted independent samples *t*-tests with the variables at the property level (Table 11). The independent samples' *t*-tests compared the means of the variables in two groups categorized by the dependent variable: parcels that remained unchanged and parcels that developed. All the variables except for two, highway accessibility (Hwy_Acc) and Accessibility to shopping centers (Shp_Cntr), presented a significant mean difference between the parcels that remained unchanged and those that were developed. The tests for the variables under the property characteristics category present mixed outputs. While the means of parcel size (Prt_Area), appraisal value (Apr_Val), and building/land value ratio (Imp_Lnd) of the undeveloped parcels were significantly higher than the means of the developed parcels, the means of the publicly owned ratio (Pbl_Prop) and buildings' year built (Yr_Bult) of the developed parcels were much higher than the means of unchanged parcels.

Similarly, the variables that presented significantly high means of the developed parcels include the rail transit proximity (Rail_Trn), accessibility to downtown Los Angeles (Cbd_Acc), and the proportion in the core cities (Core_Cty). In other words, the properties with higher values of the variables were more likely to experience residential development. These variables were more likely to significantly contribute to the likelihood of an occupied parcel experiencing residential development. Like the variables in the vacant model, the variables at the neighborhood level generally presented consistent distributions indicated by relatively smaller standard deviation values. The variables with relatively high standard deviations remained consistent with the ones in the vacant model. They included population density (Pop_Den), school proficiency (Sch_Prf), and regional job proximity (Job_Prox).

The occupied model presented a relatively reliable model fit (Table 12). According to commonly used model fit measures, such as -2 log likelihoods, AIC, and BIC values, the occupied model showed a stronger model fit than the null model (a reference model) as the measures of the occupied model were smaller than the null model. The pseudo-R-squared value was also promising. The Pseudo-R-squared value of 0.539 indicates relatively strong model fit.

Like the vacant model, the occupied model outputs imply the significance of the variables at the property level compared to the neighborhood level. While five out of 12 variables at the property level presented a correlation with the dependent variable at a 95 percent confidence level, only two out of ten variables at the neighborhood level showed a correlation at a 95 percent confidence level. At the property level, two out of five variables under the property characteristics category, one out of the accessibility to amenity category, and both variables under the locational characteristics were correlated with the dependent variables at a 95 percent confidence level. Unlike the vacant model, which showed the dominant influence of the property characteristics on the likelihood of a parcel being developed, the occupied model outputs did not present such dominance. Instead, the variables at a statistically significant level were evenly distributed in all the categories at the property level, except for the transportation accessibility category.

Table 10. Descriptive Statistics of the Independent Variables (Occupied Model)

			Pooled (N=174,070)		Unchanged (=0) (N=168,979)		Developed (=1) (N=5,091)		t-test	
Level	Category	Variable	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	t	p-value
Property level (Level 1)	Property characteristics	Prt_Area	0.214	0.473	0.215	0.469	0.199	0.601	1.908	0.000
		Pbl_Prop	0.008	0.087	0.008	0.086	0.013	0.112	-3.308	0.000
		Apr_Val (logged)	12.492	1.415	12.498	1.416	12.267	1.376	11.785	0.000
		Yr_Bult	72.460	21.552	72.354	21.403	75.973	25.766	-9.918	0.000
		Imp_Lnd (logged)	0.695	0.442	0.701	0.443	0.519	0.370	34.172	0.000
	Transportation accessibility	Hwy_Acc (logged)	1.094	0.640	1.093	0.640	1.124	0.642	-3.425	0.815
		Rail_Trn	0.037	0.189	0.037	0.188	0.045	0.207	-2.752	0.000
	Accessibility to amenities	Shp_Cntr (logged)	0.803	0.408	0.804	0.408	0.771	0.397	5.803	0.156
		Cbd_Acc (logged)	0.148	0.107	0.147	0.106	0.170	0.112	-14.401	0.000
		Job_Cntr (logged)	0.406	0.288	0.406	0.289	0.401	0.259	1.471	0.000
	Locational characteristics	Innr_Sub	0.232	0.422	0.235	0.424	0.129	0.335	22.207	0.000
		Core_Cty	0.603	0.489	0.599	0.490	0.755	0.430	-25.475	0.000
Neighborhood level (Level 2)	Socio/demographic characteristics	Pop_Den	31.147	18.171						
		Md_Inc (logged)	10.479	1.504						
		Md_Yrblt	56.051	24.912						
		Sch_Prf	42.940	27.019						
	Housing characteristics	Med_Rnt (logged)	7.017	0.704						
		Md_Hvl (logged)	11.700	3.815						
		Mv_In	23.684	9.502						
	Land use/ Urban form	LU_Dvrst	0.420	0.142						
		Job_Prox	55.393	28.548						
		Walk_Inx	14.602	2.533						

Two property characteristics variables, App_Val and Imp_Land, related to the financial cost of development, negatively correlated with the dependent variable at a 95 percent confidence level. This suggests that properties with lower appraisal values and a lower proportion of building values are more likely to be developed. Additionally, three variables—accessibility to downtown Los Angeles (Cbd_Acc), inner suburbs (Innr_Sub), and core cities (Core_Cty)—significantly contributed to residential development. While the Cbd_Acc and Core_cty variables positively correlated with the dependent variable, Innr_Sub negatively correlates with it. In other words, occupied properties closer to downtown Los Angeles and located in one of the core cities were more likely to be developed. In contrast, occupied lots in the inner suburbs were less likely to be developed.

Table 11. Occupied Model Statistics

Level	Category	Variable	Coef.	Std. Error	t	Sig.
		Intercept	-3.021	0.1906	-15.855	0.000
Property level (Level 1)	Property characteristics	Prt_Area	0.003	0.0244	0.106	0.916
		Pbl_Prop	-0.033	0.1211	-0.273	0.785
		Apr_Val	-0.038	0.0071	-5.267	0.000
		Yr_Bult	0.000	0.0006	0.489	0.625
		Imp_Lnd	-0.370	0.0294	-12.594	0.000
	Transportation accessibility	Hwy_Acc	-0.009	0.0189	-0.500	0.617
		Rail_Trn	0.038	0.0594	0.647	0.517
	Accessibility to amenities	Shp_Cntr	-0.029	0.0325	-0.878	0.380
		Cbd_Acc	0.282	0.1257	2.244	0.025
		Job_Cntr	-0.080	0.0478	-1.678	0.093
	Locational characteristics	Innr_Sub	-0.123	0.0416	-2.958	0.003
		Core_Cty	0.168	0.0354	4.750	0.000
Neighborhood level (Level 2)	Socio/demographic characteristics	Pop_Den	0.000	0.0007	-0.646	0.518
		Md_Inc	0.001	0.0080	0.152	0.879
		Md_Yrblt	0.001	0.0005	1.246	0.213
		Sch_Prf	0.000	0.0006	-0.534	0.593
	Housing characteristics	Med_Rnt	0.034	0.0186	1.856	0.063
		Md_Hvl	-0.002	0.0038	-0.640	0.522
		Mv_In	-0.002	0.0015	-1.214	0.225
	Land use/ Urban form	LU_Dvrst	0.001	0.0936	0.016	0.988
		Job_Prox	0.001	0.0006	2.209	0.027
		Walk_Inx	0.025	0.0054	4.655	0.000
-2 log likelihood		970,727.68	1,105,418.02 (Null Model)			
AIC		970,729.68	1,105,420.02 (Null Model)			
BIC		970,739.75	1,105,430.08 (Null Model)			
Pseudo-R-squared		0.539				

At the neighborhood level, only two variables, land use diversity and walkability, presented a correlation with the dependent variable at a 95 percent confidence level. While the walkability variable positively correlated with the dependent variable, the land use diversity variable negatively correlated with it.

When applying a 90 percent confidence level, two more variables showed a statistically significant correlation with the dependent variable: the proximity to the job centers (Job_Cntr) at the property level and the median rent (Med_Rnt) at the neighborhood level. While the Job_Cntr variable presented a negative correlation, the Med_Rnt variable presented a positive correlation. This indicates that the occupied properties located further away from the job centers and those in the neighborhoods with higher median rent prices were more likely to be developed. However, since this study employed a 95 percent confidence level as the threshold for statistical significance, these variables were not counted as statistically significant.

5.2 Discussion: Occupied Model

Previous research has discussed the relationship between residential development and transit service in two directions: the impacts of transit service on residential development and vice versa. However, few studies have directly addressed the heterogeneity of the properties developed, separating vacant parcel development from the development of underutilized lots occupied by existing buildings. The model outputs articulate the relationship between the transit system and occupied properties.

The variables at the property level (5 out of 12) were more significant contributing factors to residential development in occupied properties than the neighborhood level factors (2 out of 10). While the factors at both the property and neighborhood levels significantly influence the occupied lot development, the outputs confirm that the factors at the property level are more influential on the development than the neighborhood factors. This is also confirmed in the vacant model. Thus, regardless of development types, the features of properties more directly influence residential development than the characteristics of neighborhoods.

In comparison to vacant lot development, the contribution of the property characteristics to occupied lot development is less influential. Since two out of five statistically significant variables at the property level belongs to the property characteristics category, there is no question about the significance of the property characteristics. However, only two out of the six variables in the category of the occupied model show the correlation, while the vacant model presents that all the variables under the property characteristics category present a statistically significant correlation with the dependent variable. This finding probably reflects the complex and multi-layered property characteristics associated with residential development on occupied properties at a certain level. The property characteristics of occupied properties associated with the development will be more complex than those of vacant properties. This complexity and the endogenous dynamics among the variables in the category may hinder the model from articulating the contribution of the variables to the development. Future studies may need to investigate the relationship between property characteristics and residential development in depth.

The significant variables under the property characteristics category represent the features related to property values. They include appraisal value and the ratio of building value to land value. The negative correlations of the variables indicate that the properties with a higher value and a higher ratio of building value are less likely to be developed. The variables are directly associated with developers' financial portfolios. This implies that the properties that require lower land assembly costs (e.g., acquisition, relocation, demolition, clearance, and site preparation) are more attractive for residential development. This finding aligns with the previous literature (Kim et al, 2022; Jun et al., 2024) and is understandable from the developers' perspective on financial concerns and profitability.

In this respect, this finding raises an interesting contrast between occupied and vacant lot development. The occupied properties with lower values are more likely to be developed, while the vacant parcels with higher values are more likely to be developed. As explained above, developers may not mind investing in higher-value vacant properties and expecting a higher return, since the overall scale of financial investment on vacant lot development is relatively small. However, they seem more conservative in the cases of occupied lot development, which requires a substantial financial investment. For these, they tend to pay more attention to the occupied properties with lower values.

Residential development on occupied parcels exemplifies the features of conventional infill development. The properties in the core cities and closer to downtown Los Angeles are more likely to be developed. This aligns with the definition of conventional infill development. Infill development refers to urban development that constructs new buildings or replaces primary buildings in already built-up areas (often near the center of cities and towns), further filling urban areas. In this way, infill development utilizes the existing transportation and utility infrastructures and provides new opportunities for underutilized properties. Therefore, the occupied lot development in the HQTAs, particularly in the core cities and adjacent areas to downtown Los Angeles, probably reflects the cities' policy directions that promote sustainable urban development and smart growth strategies like infill development. Overall, this policy direction has gained popularity, particularly in California. For example, Governor Newsom issued a new executive order (N-2-24) on July 31, 2024 to accelerate and streamline infill development projects to transform undeveloped and underutilized properties statewide into livable and affordable housing for Californians.

An interesting contrast between vacant and occupied lot development in the core cities was identified. Although the distributions of the occupied and vacant properties in the core cities are very comparable (60.3% and 63.9%) (Table 13), the likelihood of occupied lot development in the core cities presents a statistically significant positive correlation. However, the correlation between vacant lot development in the cities and the dependent variable is not identified. This probably indicates that vacant lots in the core cities are less attractive for residential development, even considering the lower development cost. This raises an imperative policy question: how do you promote infill development on vacant parcels in core cities? Thus, it will be important for the cities to construct an inventory of the vacant parcels, understand the characteristics of the parcels, and develop policies and programs that promote the development of the parcels.

Table 12. Parcel Type Comparison Between Core Cities and Other Areas

Area	Occupied Parcels		Vacant Parcels	
	Count	%	Count	%
Core Cities	105,023	60.3%	7,772	63.9%
Other Areas	69,047	39.7%	4,835	39.8%
Grand Total	174,070	100.0%	12,157	100.0%

In the same vein, occupied parcel development is more likely to be associated with neighborhoods that offer urban functions, such as being closer to downtown Los Angeles, reasonable job proximity, and walkability. A large volume of research has reported the positive contribution of urban amenities and sustainable built environments to housing development (Kamal & Proma, 2017; Kim et al., 2022; Jun et al., 2024). The current planning directions, including smart growth strategies, transit-oriented development (TOD), and infill development, also emphasized the close connection between housing density and sustainable built environments such as mixed-use, transit-friendly, and so on (Ewing, 1997; Knaap & Talen, 2005; Rahimi, 2016). The development in the neighborhood with urban amenities will contribute to sustainable urban development and smart growth strategies. It is also expected to promote transit ridership and active transportation. This is another contrast between the vacant and occupied lot development, since the vacant lot development avoids urban functions and is more likely to be in areas with quality living environment , such as tranquility. This confirms that the occupied lot development presents the features of infill development in comparison to the vacant lot development. This contrast may also imply the availability of developable, underutilized occupied properties. Unlike the scarcity of vacant parcels in the neighborhoods with good urban amenities, developable occupied parcels exist in all the neighborhoods in the study area. In coordination with regional agencies such as MPOs and transit agencies, local governments may ponder this nuance and develop customized policies that promote vacant and occupied lot development separately.

The outputs clearly confirm shrinking development activities on occupied properties in inner suburbs. This finding is consistent with the vacant model. Conventionally, inner suburban communities were recognized for their significant role in high-density development (Lee & Leigh, 2005; Chakraborty et al., 2010). Unlike this conventional observation, however, the model results identify a lower likelihood of the occupied parcels in inner suburbs being developed. Many inner suburbs in the study area were probably built out, and their built environment became similar to core cities after playing significant roles in absorbing the housing and jobs spilled over from core cities for a long time. However, unlike the core cities, they have not proactively adopted smart growth strategies. This finding suggests that future policy directions may need to pay attention to these changing roles of inner suburbs in the context of sustainable urban development. Like their urban counterparts, inner suburban cities have also reinvented their communities to be more sustainable, and the core strategy is systemic densification by promoting infill development on underutilized properties, particularly along major transit and transportation corridors.

The walkability variable (Walk_Inx) was the only variable consistently positively correlated with the likelihood of development in both models. A high-quality pedestrian environment in the HQTAs may be closely associated with multimodality between walking and public transit, which will reduce automobile dependency. Thus, the positive correlations confirm the quality and roles of the HQTAs in sustainable development. According to SCAG's 2024/2050 Regional Transportation Plan (RTP), SCAG plans to accommodate 66 percent of forecasted household growth in the areas with high quality transit services (SCAG, 2024). Coordinating with the plan, local jurisdictions will also need to prepare programs and policies that can guide residential development to the HQTAs. This coordination will bring positive results in sustainable development in the region.

It is also noteworthy that the elements counted in the walkability index measurement are beyond transportation aspects. In measuring the index, the EPA took into consideration not only transportation factors such as intersection density and proximity to transit stops but also land use aspects such as employment mix (the mix of employment types such as retail, office, or industrial) and employment and housing mix (the mix of employment types and occupied housing) (USEPA, 2021). Thus, the index indicates walkable neighborhoods that make it easier to walk to stores, jobs, and other places, which encourages people to be more active. Residential development in the neighborhoods will ultimately contribute to residents' public health benefits (e.g., reduction of obesity and diabetes rates) and encourage social interaction, which engenders a sense of community and improves people's mental health (USEPA, 2021).

6. Conclusion

By employing a deep-learning algorithm that detects and classifies residential development by property type, this research proposes an innovative method for research focusing on land use and transportation dynamics. Unlike conventional methods that depend on government documents or geo-spatial analysis, the deep-learning method presents pronounced potential for detecting urban development at a micro level, such as a parcel level, in regional-scale studies. The high estimation accuracy of the deep-learning model also confirms its potential.

Based on the innovative method, this research separates developments on vacant and occupied lots near transit and examines the unique contributing factors to each development type. The findings confirm the heterogeneity of the contributing factors to the development of vacant and occupied parcels. While the factors at the property level are more significant than the neighborhood level factors for both vacant and occupied lot development, the variations in the significance of the factors at the property level were identified. The property characteristics play a much more significant role in vacant lot development than in occupied lot development. While infill development on occupied lots in the core cities is more likely to occur, the inner suburban cities tend to experience residential development on vacant and occupied properties much less than the other areas. Another interesting contrast between vacant and occupied lot development is that the occupied lot development is more likely to be associated with urban functions/amenities. In contrast, vacant lot development tends to occur in areas with tranquility near transit.

This research provides insight into adjusting local and regional housing development policies and developing new policies customized for the unique characteristics of potential developable properties. The research also contributes to deepening the discussion about the importance of land use and transportation integration for sustainable urban development. This research especially emphasizes the heterogeneity of the contributing factors to the development of vacant and occupied parcels. It is important to clearly understand the attributes of vacant lots that attract development and differ from those of occupied properties.

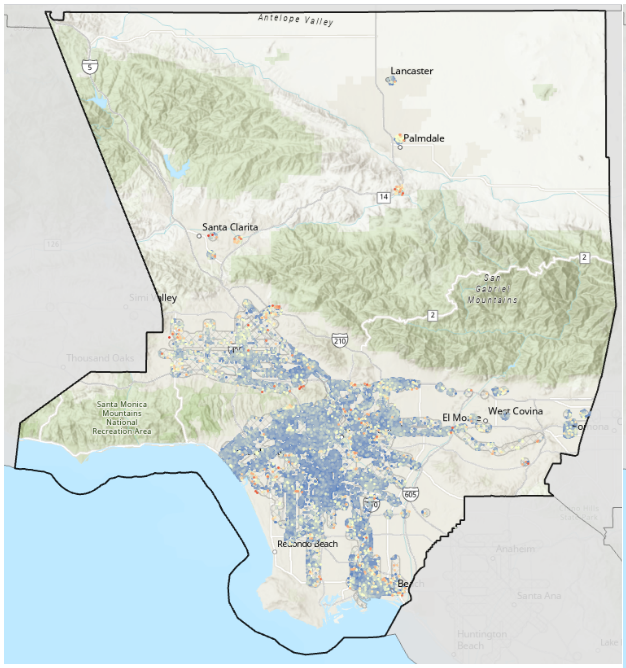
This may set a good starting point for local governments in developing customized sustainable development policies. For instance, it seems that a more significant number of vacant lots are left at less desirable locations after desirable vacant lots have already been developed. Thus, policy attention may need to be paid to making vacant lots more attractive. Municipal governments may promote the redevelopment of occupied lots by reducing overall regulations because the likelihood of redevelopment aligns with developers' general interests in profitability (e.g., less expensive properties) and because much literature points out regulations and community resistance as one of the most significant challenges to development.

As the core cities' development strategies such as infill development in the high-quality transit areas contribute to sustainable urban development, it is necessary to continue to promote this type of development. At the same time, a regional collaboration between the core and suburban cities, particularly inner suburban cities, needs to be considered in terms of sharing and transferring the core cities' experiences with infill development. Although many inner suburban cities may believe that they are built out, the cities have underutilized properties that have development potential, as their urban counterparts' cases show. Learning from the core cities, suburban municipalities may consider developing programs and strategies that identify and incentivize the underutilized properties near public transit. Speaking of collaboration, it is important to incorporate developers into the collaboration and policy development. This research suggests that the patterns of both vacant and occupied lot developments are significantly associated with developers' profitability.

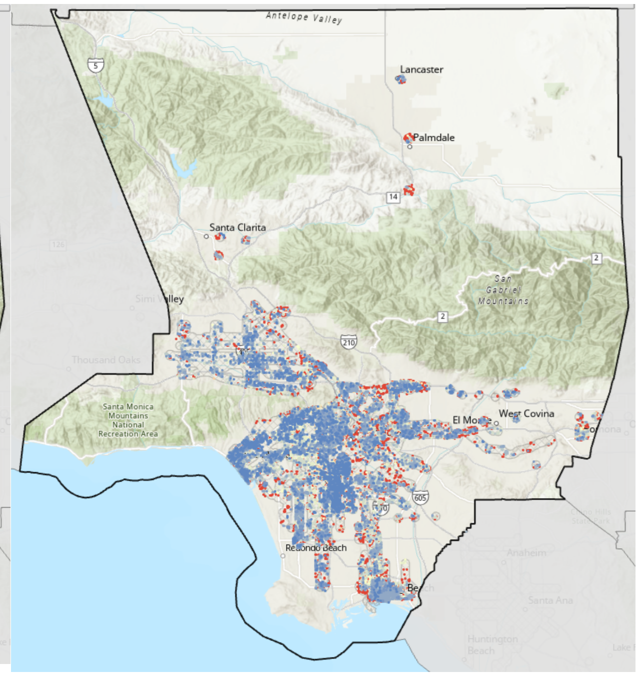
Ultimately, sustainable urban development can be achieved through strong public-private partnerships. Another important player in this topic is the residents. Thus, it is also important to clearly understand people's residential location choices and preferences. This research revealed a nuance in location preference, from balancing between tranquility and urban amenities to preference for urban amenities. However, it is unclear whether the nuance is truly due to people's location choices or the availability of developable properties in the areas with better urban amenities. Future studies may need to clarify this topic by deeply understanding people's residential choices in high-quality transit areas.

Appendix A: Variables at the Property Level

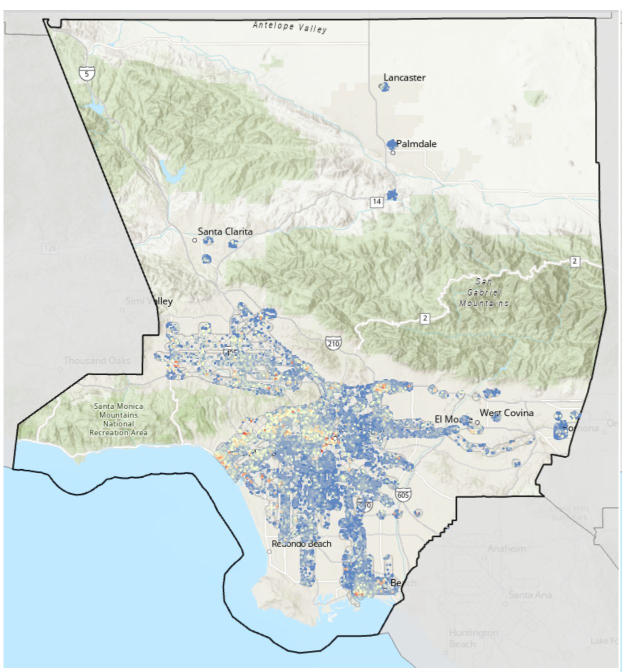
Lot Size



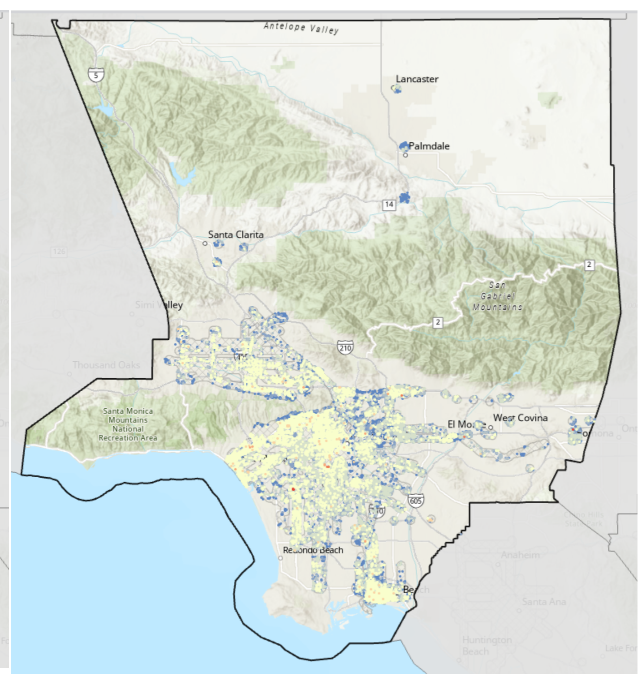
Building Age



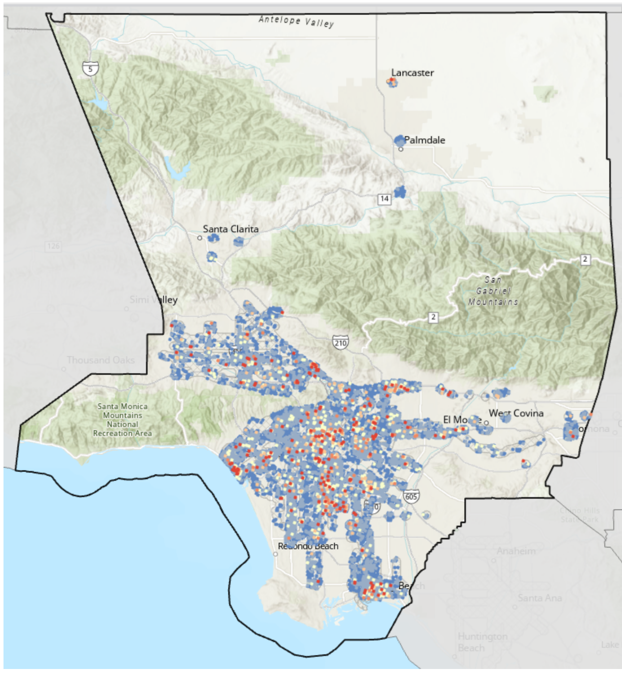
Property Appraisal Value



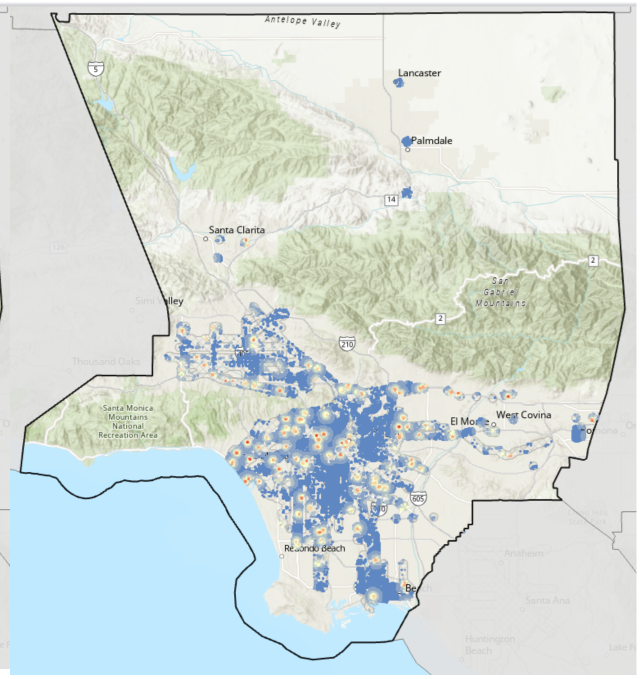
Floor Area Ratio (FAR)



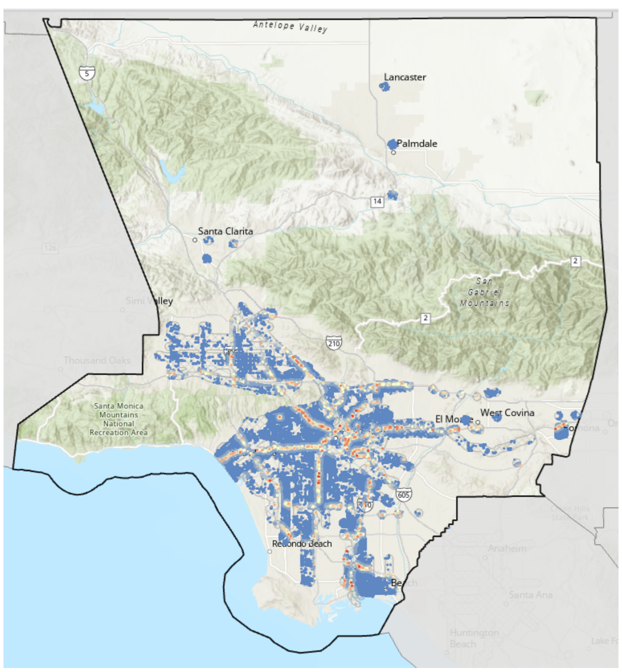
Building Value Ratio



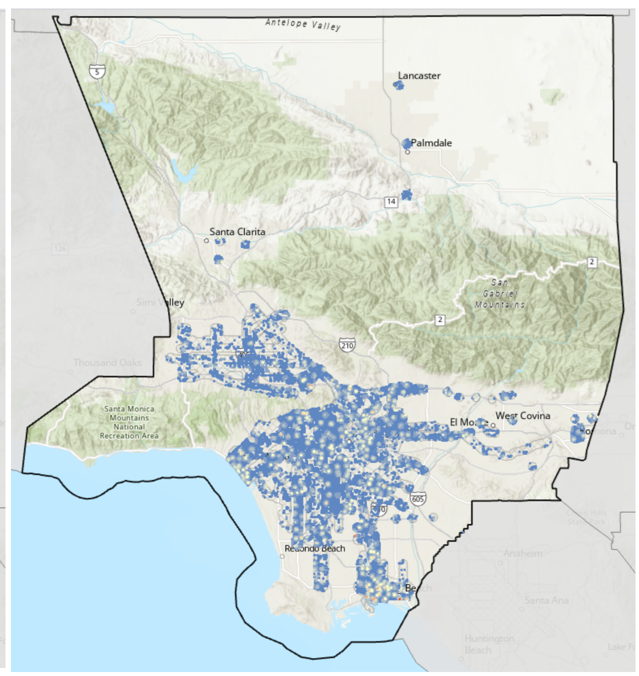
Shopping Accessibility



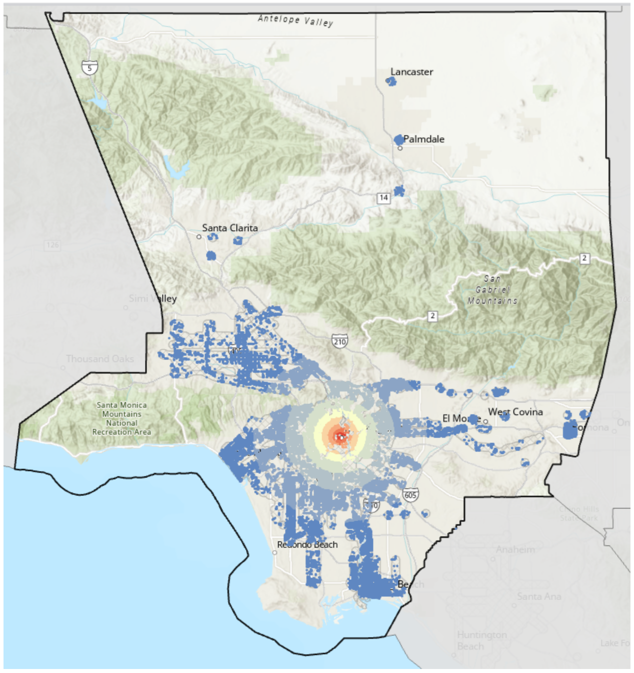
Highway Accessibility



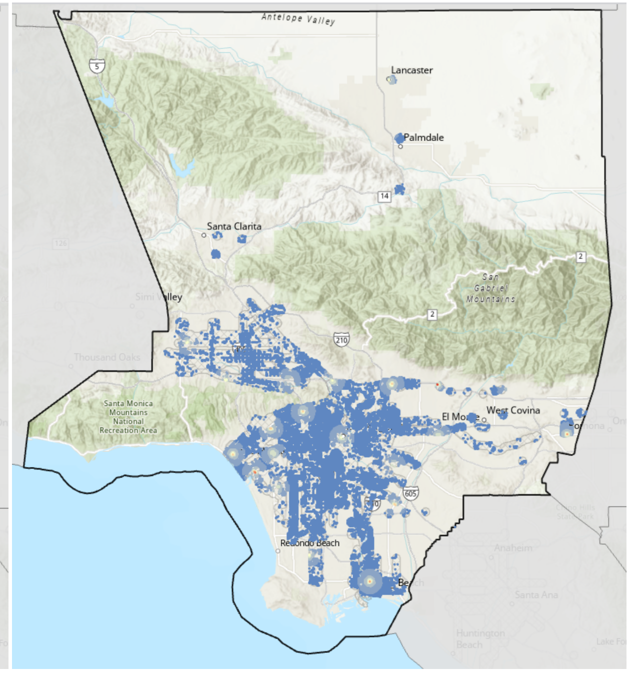
Park Accessibility



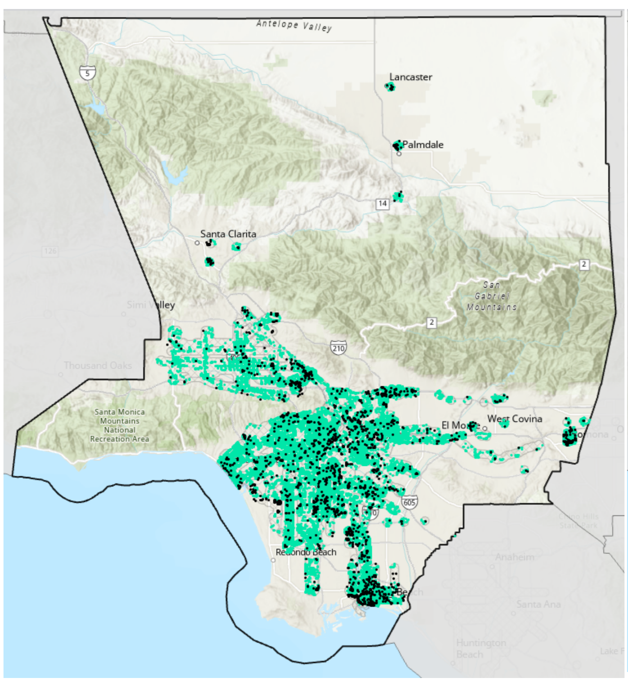
Proximity to Downtown Los Angeles



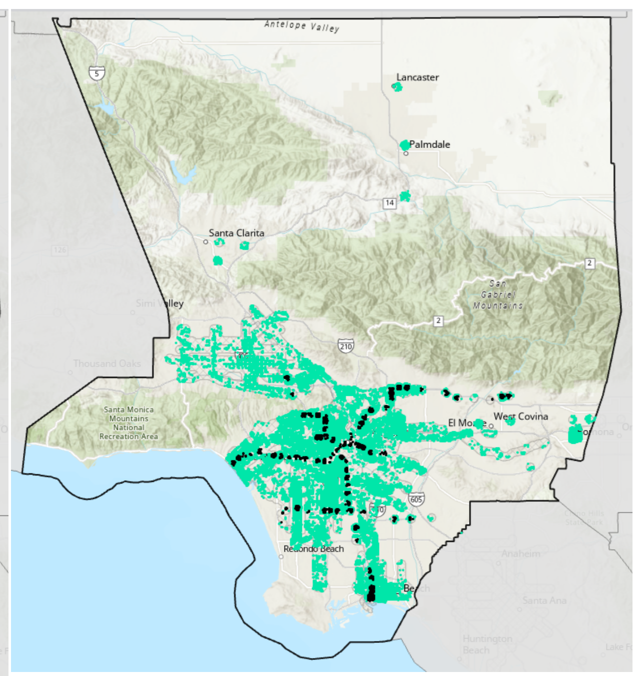
Proximity to Job Center



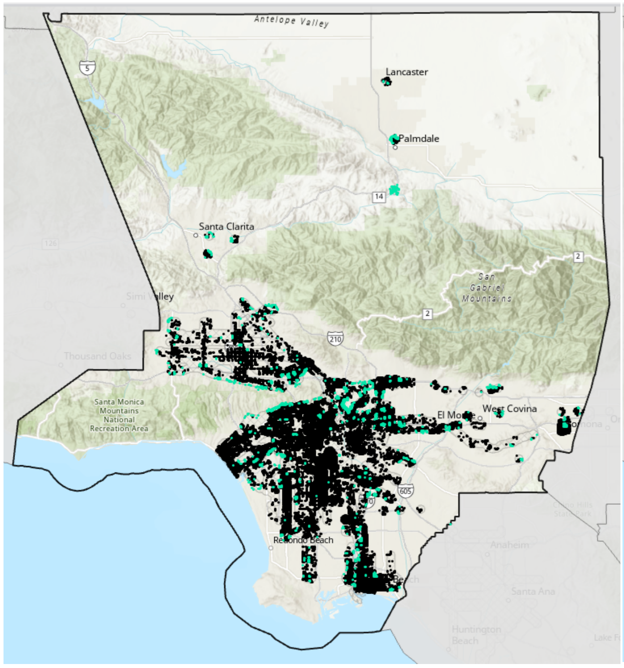
Publicly Owned Property



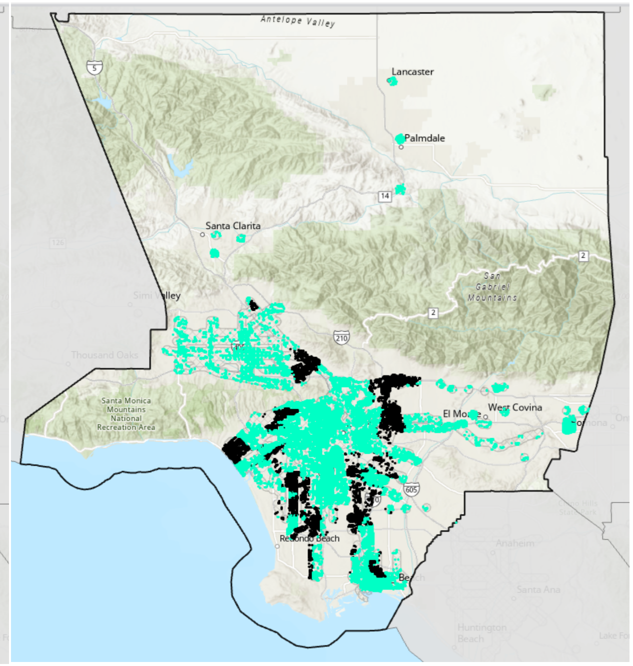
Rail Station Accessibility



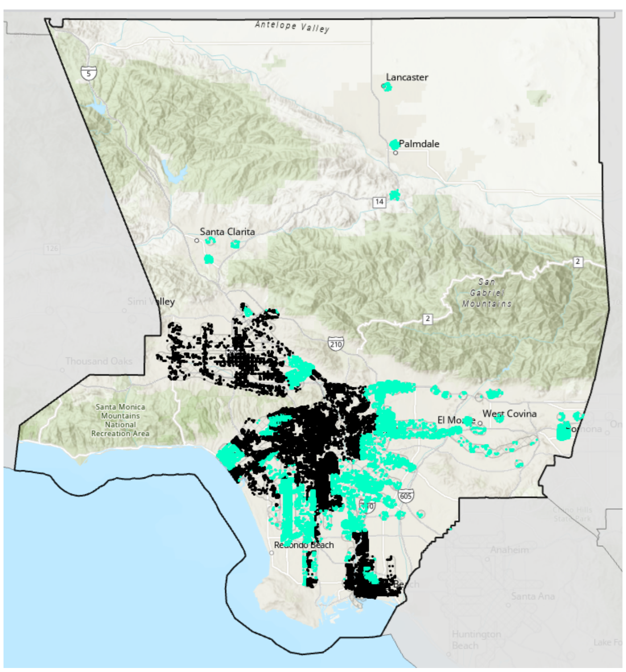
Bus Transit Accessibility



Property in Inner Suburbs



Property in Core Cities



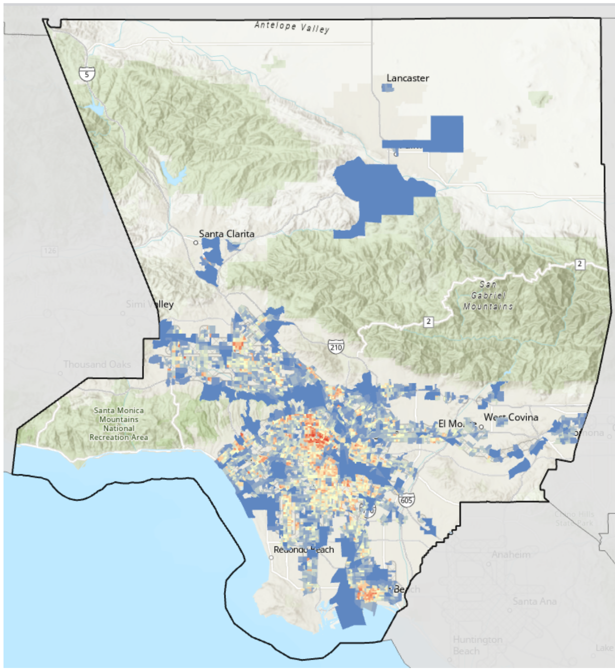
Legend

Low Medium High

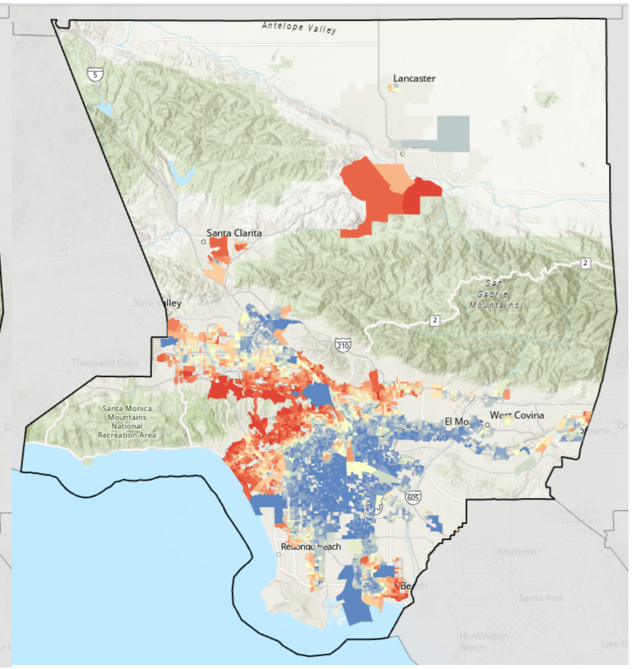
No Yes

Appendix B: Variables at the Neighborhood Level

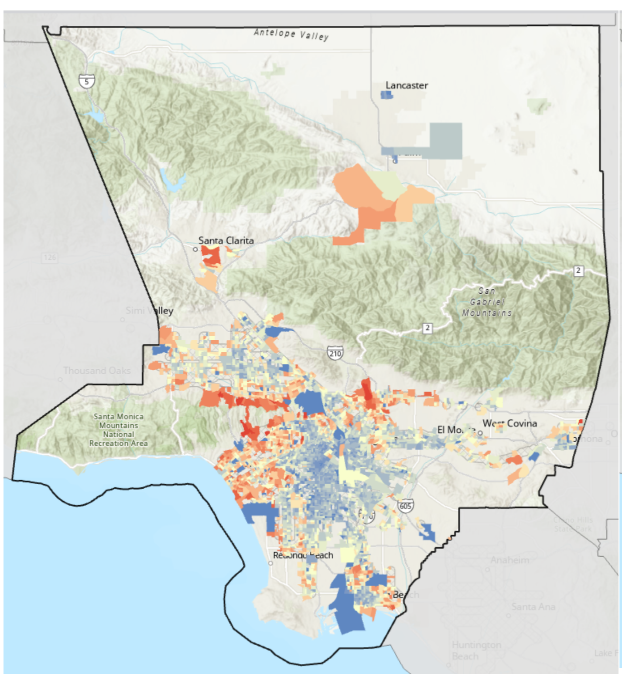
Population Density



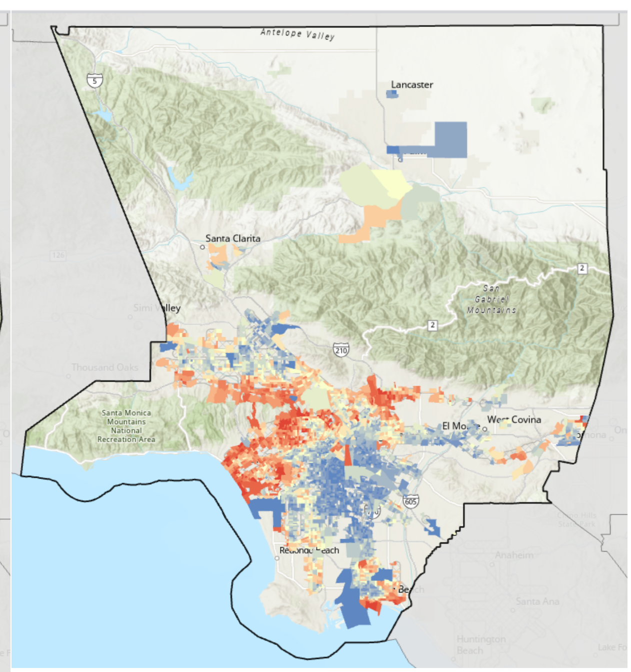
Ratio of White Population



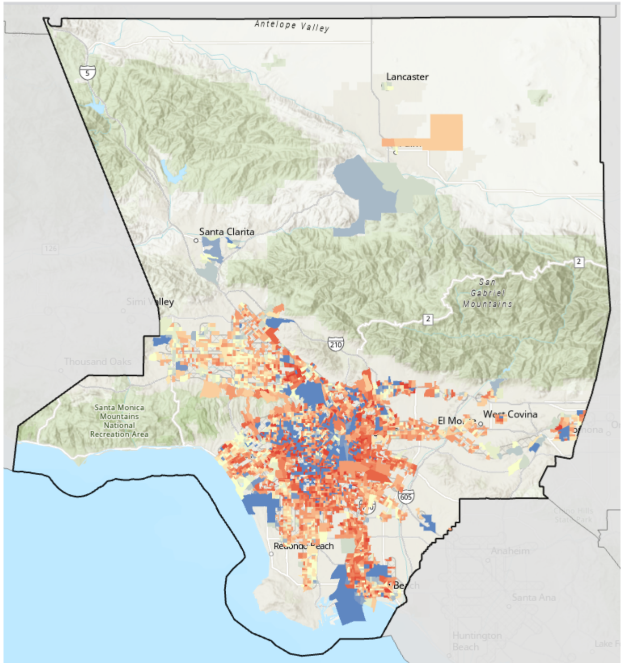
Household Median Income



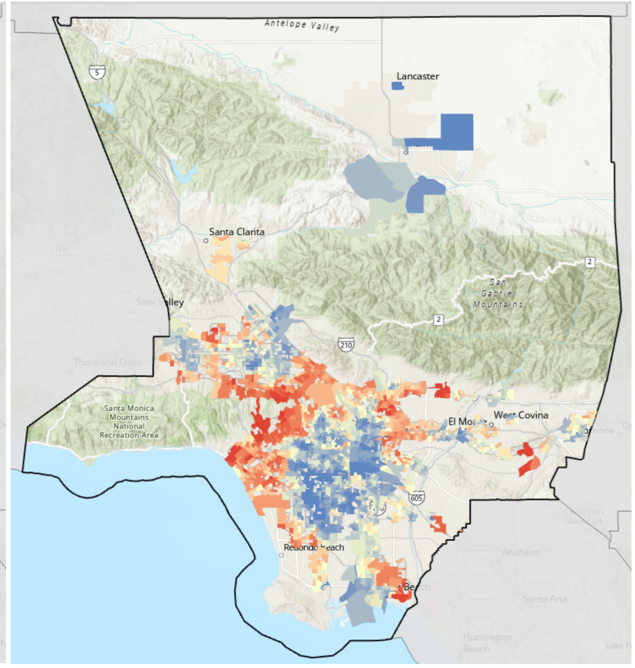
Ratio of College Graduate and Higher



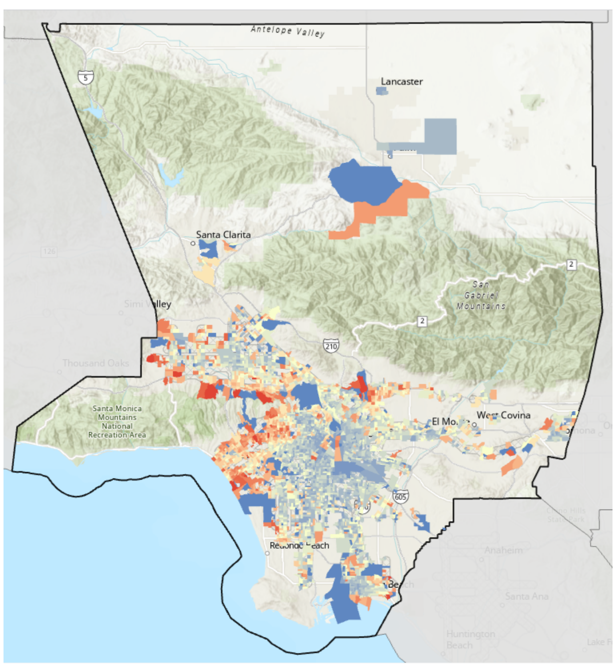
Neighborhood Age



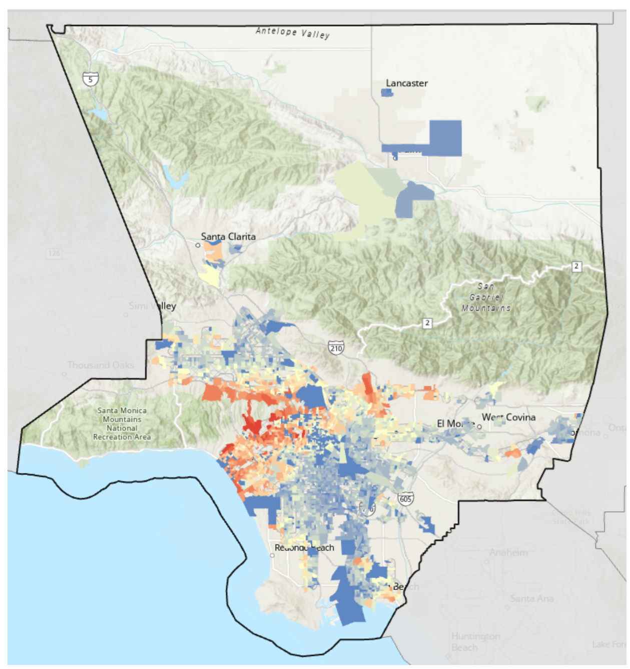
School Proficiency



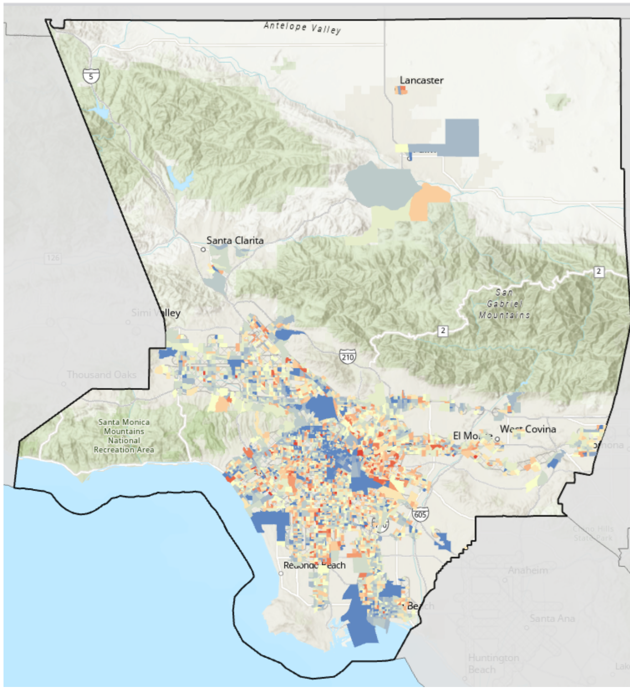
Median Rent Cost



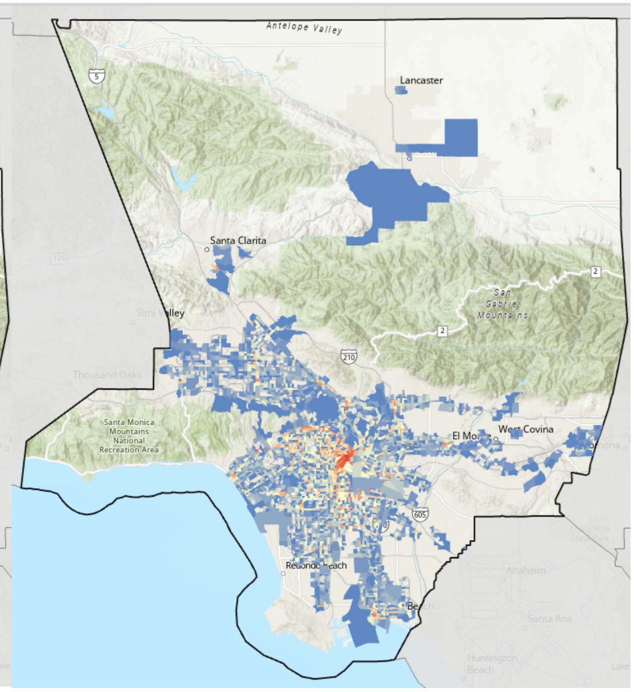
Median House Value



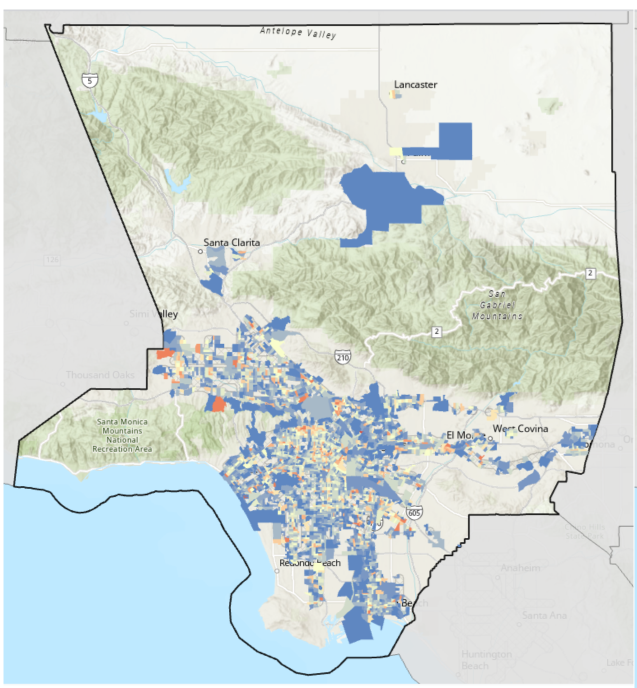
Median Year of Residency



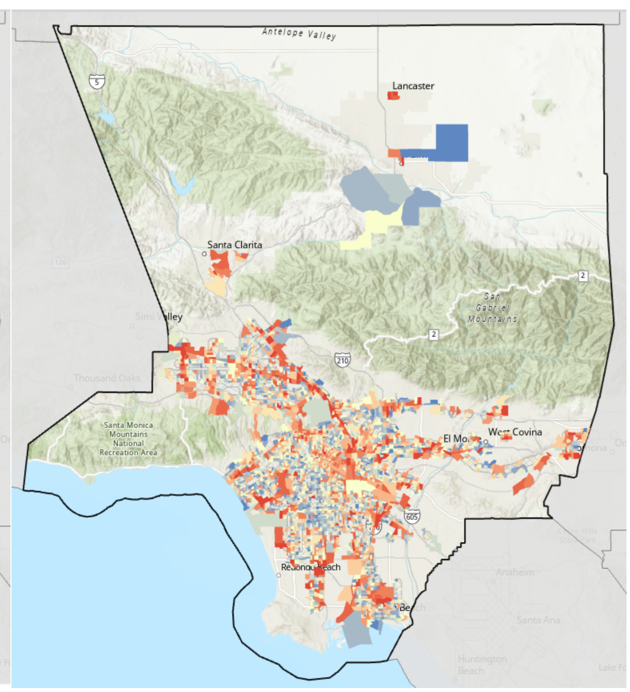
Bus Stop Density



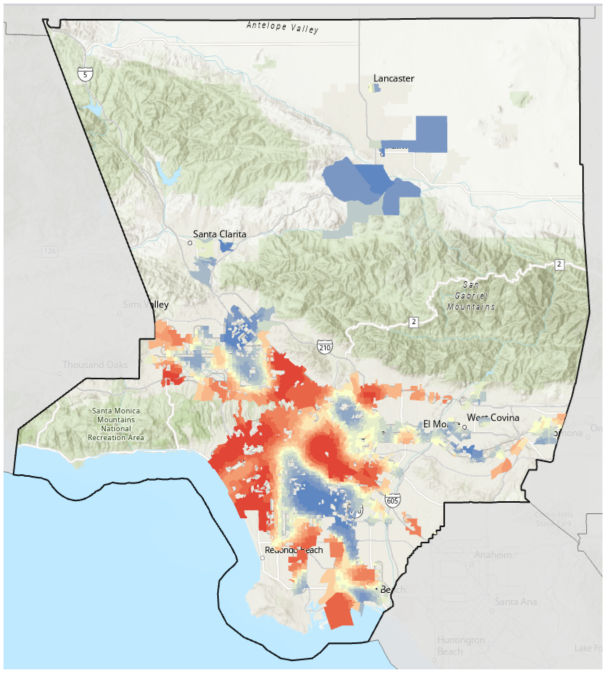
Ratio of Retail Land Use



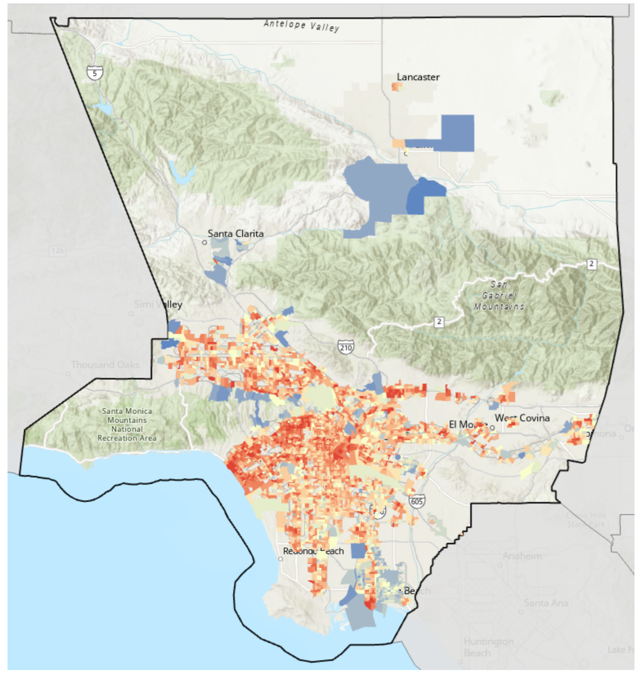
Land Use Diversity



Job Proximity



Walkability Index



Legend

Low Medium High

Bibliography

Abedini, A., & Khalili, A. (2019). Determining the capacity infill development in growing metropolitans: A case study of Urmia city. *Journal of Urban Management*, *8*(2), 316–327. https://doi.org/10.1016/j.jum.2019.04.001

Atkinson-Palombo, C. (2010). New housing construction in Phoenix: Evidence of "new suburbanism"? *Cities*, *27*(2), 77–86. https://doi.org/10.1016/j.cities.2009.10.001

August, M., & Walks, A. (2018). Gentrification, suburban decline, and the financialization of multi-family rental housing: The case of Toronto. *Geoforum*, *89*, 124–136. https://doi.org/10.1016/j.geoforum.2017.04.011

Bourne, L. S. (1969). Location factors in the redevelopment process: A model of residential change. *Land Economics*, *45*(2), 183-193. https://www.jstor.org/stable/3145124

Bramley, G., & Power, S. (2009). Urban form and social sustainability: The role of density and housing type. *Environment and Planning B*, *36*, 30–48. https://doi.org/10.1068/b33129

Brownson, R. C., Hoehner, C. M., Day, K., Forsyth, A., & Sallis, J. F. (2009). Measuring the built environment for physical activity: State of the science. *American Journal of Preventive Medicine*, *36*(4), S99–S123 (e112). https://doi.org/10.1016/j.amepre.2009.01.005

Byun, G., & Kim, Y. (2022). A street-view based method to detect urban growth and decline: A case study of midtown in Detroit, Michigan, USA. *PLoS One*, *17*(2), Article e0263775. https://doi.org/10.1371/journal.pone.0263775

Cao, X., & Pan, Q. (2016). Rapid transit and land development in a diverse world. *Transport Policy*, *51*, 1–3. https://doi.org/10.1016/j.tranpol.2016.06.001

Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment*, *23*, 199219. https://doi.org/10.1016/S1361-9209(97)00009-6

Cervero, R., & Landis, J. (1997). Twenty years of the Bay Area Rapid Transit system: Land use and development impacts. *Transportation Research Part A: Policy and Practice*, *314*, 309333. https://doi.org/10.1016/S0965-8564(96)00027-4

Charles, S. L. (2011). *Suburban Gentrification: Understanding the Determinants of Single-family Residential Redevelopment, A Case Study of the Inner-Ring Suburbs of Chicago, IL, 2000–2010*. W11-1. Joint Center for Housing Studies, Havard University. https://www.jchs.harvard.edu/sites/default/files/media/imp/w11-1_charles.pdf

Chakraborty, A., Knapp, G., Nguyen, D., & Shin, J.H. (2010). The effects of high-density zoning on multifamily housing construction in the suburbs of six US metropolitan areas. *Urban Studies, 47*(2), 437–451. https://www.jstor.org/stable/43084690

Danielsen, K.A., Lang, R.E., & Fulton, W. (1999). Retracting suburbia: Smart growth and the future of housing. *Housing Policy Debate*, *10*(3), 513–40. https://doi.org/10.1080/10511482.1999.9521341

Dale, A., & Newman, L. L. (2009). Sustainable development for some: Green urban development and affordability. *Local Environment*, *14*, 669–681. https://doi.org/10.1080/13549830903089283

Dong, H. (2016). If you build rail transit in suburbs, will development come? *Journal of the American Planning Association*, *82*(4), 316–326. https://doi.org/10.1080/01944363.2016.1215258

Downs, A. (2005). Smart growth: Why we discuss it more than we do it. *Journal of the American Planning Association*, *71*(4), 367–378. https://doi.org/10.1080/01944360508976707

Duany, A., Plater-Zyberk, E., & Speck, J. (2000). *Suburban nation: The rise of sprawl and the decline of the American dream*. North Point Press.

Ewing, R. (1997). Is Los Angeles-style sprawl desirable? *Journal of the American Planning Association*, *63*(1), 107–126. https://doi.org/10.1080/01944369708975728

Ewing, R., & Cervero, R. (2001). Travel and the built environment: A synthesis. *Transportation Research Record: Journal of the Transportation Research Board*, *1780*, 87114. https://doi.org/10.3141/1780-10

Fainstein, S. (2005). Cities and diversity: Should we want it? Can we plan for it? *Urban Affairs Review*, *41*(1), 3–19. https://doi.org/10.1177/1078087405278968

Farris, T. J. (2001). The barriers to using urban infill development to achieve smart growth. *Housing Policy Debate*, *12*(1), 1–30. https://doi.org/10.1080/10511482.2001.9521395

Frenkel, A., Bendit, E., & Kaplan, S. (2013). Residential location choice of knowledge-workers: The role of amenities, workplace and lifestyle. *Cities*, *35*, 33–41. https://doi.org/10.1016/j.cities.2013.06.005

Gabbe, C. J. (2017). Why are regulations changed? A parcel analysis of upzoning in Los Angeles. *Journal of Planning Education and Research*, 38(3), 289–300. https://doi.org/10.1177/0739456X17696034

Giuliano, G. (2017). Land use impacts of transportation investments highway and transit. In S. Hanson & G. Giuliano (Eds.), *The geography of urban transportation* (4th ed.). The Guilford Press.

Green, T. L. (2018). Evaluating predictors for brownfield redevelopment. *Land Use Policy*, *73*, 299–319. https://doi.org/10.1016/j.landusepol.2018.01.008

Handy, S. (2005). Smart growth and the transportation-land use connection: What does the research tell us? *International Regional Science Review*, *282*, 146167. https://doi.org/10.1177/0160017604273626

Howell, A.J., & Timberlake, J. M. (2014). Racial and ethnic trends in the suburbanization of poverty in U.S. metropolitan areas, 1980–2010. *Journal of Urban Affairs*, *36*(1), 79–98. https://doi.org/10.1111/juaf.12030

Hurst, N. B., & West, S.E. (2014). Public transit and urban redevelopment: The effect of light rail transit on land use in Minneapolis, Minnesota. *Regional Science and Urban Economics*, *46*, 57-72. http://dx.doi.org/10.1016/j.regsciurbeco.2014.02.002

Jensen, J. R. (2004). *Introductory digital image processing: A remote sensing perspective* (3rd ed.). Prentice Hall.

Jun, M.J., Kim, J. I., Kim, H. J., Yeo, C. H., & Hyun, J. Y. (2017). Effects of two urban development strategies on changes in the land surface temperature: Infill versus suburban new town development. *Journal of Urban Planning and Development*, *143*(3), 04017010. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000396

Jun, H., Kim, D., Kim, J., & Heo, J. (2024). Detection of infill development and contributing factors using deep learning and multilevel modeling. *Cities*. *150*, 105019. https://doi.org/10.1016/j.cities.2024.105019

Kamal, A., & Proma, N. (2017). Spatial modeling for sustainable residential infill in Texas peri-urban communities. *Journal of Urban Planning and Development, 143*(2), 04016036. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000372

Kim, D., Ahn, Y., Choi, S., & Kim, K. (2016a). Sustainable mobility: Longitudinal analysis of built environment on transit ridership. *Sustainability*, *8*(10), 1016–1030. https://doi.org/10.3390/su8101016

Kim, D., Shim, J., Park, J., Cho, J., & Kumar, S. (2022). Supervised machine learning approaches to modeling residential infill development in the City of Los Angeles. *Journal of Urban Planning and Development*, *148*(1), 04021060. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000787

Kim, J. (2015). Achieving mixed income communities through infill? The effect of infill housing on neighborhood income diversity. *Journal of Urban Affairs, 38*(2), 280–297. https://doi.org/10.1111/juaf.12207

Kim, J., Houston, H., Cho, J., Lo, A., Shi, X., Hoff, A., Branic, N., Ly, H. C., Walker, A. L., & Zhang, Y. (2016b). *Infill dynamics in rail transit corridors: Challenges and prospects for integrating transportation and land use planning*. CA16 2641. Institute of Transportation Studies. https://dot.ca.gov/-/media/dot-media/programs/research-innovation-system-information/documents/f0016803-ca16-2641-finalreport.pdf

Kim, J., Kim, D., Jun, H., & Huh, J. (2024). The detection of residential developments in urban areas: Exploring the potentials of deep-learning algorithms. *Computer, Environment, and Urban Systems*, *107*, 102053. https://doi.org/10.1016/j.compenvurbsys.2023.102053

Kim, J., & Li, X. (2021). Building more housing near transit: A spatial analysis of residential densification dynamics. *Transport Policy*, *114*, 15–24. https://doi.org/10.1016/j.tranpol.2021.08.018

Kim, S., & Kim, D. (2024). Climate change and cooling equity: Spatial dynamics of vulnerable populations. *Growth and Change*, *55*(1), 1–14. https://doi.org/10.1111/grow.12701

Knaap, G. (2003). Land supply and infrastructure capacity monitoring for smart urban growth. National Center for Smart Growth. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=8857761badcfdbbef8a0af1202cf520cda28a188

Kramer, A. (2018). The unaffordable city: Housing and transit in North American cities. *Cities*, *83*, 1–10. https://doi.org/10.1016/j.cities.2018.05.013

Lachenbruch, P. A., & Mickey, M. R. (1968). Estimation of error rates in discriminant analysis. *Technometrics*, *10*(1), 1–11. https://doi.org/10.1080/%2000401706.1968.10490530

Lee, S., & Leigh, N. G. (2005). The role of inner ring suburbs in metropolitan smart growth strategies. *Journal of Planning Literature*, *19*(3), 330–346. https://doi.org/10.1177/0885412204271878

Leven, C. L., Little, J. T., Nourse, H. O., & Read, R. B. (1976). *Neighborhood change: Lessons in the dynamics of urban decay*. Praeger.

Levin, J., & Inam, A. (2004). The market for transportation-land use integration: Do developers want smarter growth than regulations allow? *Transportation*, *31*, 409–427. https://doi.org/10.1023/B:PORT.0000037086.33893.9f

Lewis, R. (2012). The determinants of renovation and redevelopment in Baltimore City. *Regional Science Policy & Practice*, *4*(4), 335–355. https://doi.org/10.1111/j.1757-7802.2012.01080.x

Lichter, D. T., Thiede, B. C., & Brooks, M. M. (2023). Racial diversity and segregation: Comparing principal cities, inner-ring suburbs, outlying suburbs, and the suburban fringe. *The Russell Sage Foundation Journal of The Social Sciences*, *9*(1), 26–51. https://doi.org/10.7758/RSF.2023.9.1.02

Liu, C., & Bardaka, E. (2021). The suburbanization of poverty and changes in access to public transportation in the Triangle Region, NC. *Journal of Transport Geography*, *90*, 102930. https://doi.org/10.1016/j.jtrangeo.2020.102930

Liu, Q., Hang, R., Song, H., & Li, Z. (2017). Learning multiscale deep features for high-resolution satellite image scene classification. *IEEE Transactions on Geoscience and Remote Sensing*, *56*(1), 117–126. https://doi.org/10.1109/TGRS.2017.2743243

Lovejoy, K., Handy, S., & Mokhtarian, P. (2010). Neighborhood satisfaction in suburban versus traditional environments: An evaluation of contributing characteristics in eight California neighborhoods. *Landscape and Urban Planning*, *97*(1), 37–48. https://doi.org/10.1016/j.landurbplan.2010.04.010

Makhamreha, Z., & Almanasyeha, N. (2011). Analyzing the state and pattern of urban growth and city planning in Amman using satellite images and GIS. *European Journal of Social Sciences*, *24*(2), 225–264. https://eacademic.ju.edu.jo/z.makhamreh/Lists/Published%20Research/Attachments/11/EJSS_24_2_11.pdf

Manville, M., & Monkkonen, P. (2024). Unwanted housing: Localism and politics of housing development. *Journal of Planning Education and Research*, *44*(2), 685–700. https://doi.org/10.1177/0739456X21997903

McConnell, V., & Wiley, K. (2010). *Infill development: Perspectives and evidence from economics and planning*. Discussion Paper, RFF DP 10-13. Resources for the Future. https://media.rff.org/documents/RFF-DP-10-13.pdf

Mejias, L., & Deakin, E. (2005). Redevelopment and revitalization along urban arterials: Case study of San Pablo Avenue, California, from the developers' perspective. *Transportation Research Record: Journal of the Transportation Research Board*, *1902*(1), 26–34. https://doi.org/10.1177/0361198105190200104

Myers, D., & Gearin, E. (2001). Current preferences and future demand for denser residential environments. *Housing Policy Debate*, *12*, 633–659. https://doi.org/10.1080/10511482.2001.9521422

Naik, N., Kominers, S. D., Raskarb, R., Glaesere, E. L., & Hidalgo, C. A. (2017). Computer vision uncovers predictors of physical urban change. *PNAS*, *114*(29), 7571–7576. https://doi.org/10.1073/pnas.1619003114

Olaru, D., Smith, B., & Taplin, J. H. E. (2011). Residential location and transit-oriented development in a new rail corridor. *Transportation Research Part A: Policy and Practice*, *45*(3), 219–237. https://doi.org/10.1016/j.tra.2010.12.007

Qian, X., Xiao, R., Joby, R., & Jaller, M. (2024). Optimizing bikeshare service to connect affordable housing units with transit service. *Journal of Transport Geography*, *117*, 103894. https://doi.org/10.1016/j.jtrangeo.2024.103894

National Association of Realtors Research Group (NARRG). (2020). *Case studies on repurposing vacant retail malls*. National Association of Realtors. https://cms.nar.realtor/sites/default/files/documents/2020-case-studies-on-repurposing-vacant-retail-malls-05-08-2020.pdf?_gl=1*1c7zhfh*_gcl_au*MjEyNTQ4NDY4Ny4xNzQ0NzU2MTQy

Newman, P. W. G., & Kenworthy, J. R. (2006). Urban design to reduce automobile dependence. *Opolis: An International Journal of Suburban and Metropolitan Studies*, *2*(l), 35–52. https://escholarship.org/uc/item/2b76f089

Pan, H., S. Chen, Y. Gao, B. Deal, & Liu, J. (2020). An urban informatics approach to understanding residential mobility in Metro Chicago. *Environment and Planning B: Urban Analytics and City Science*, *47*(8), 1456–1473. https://doi.org/10.1177/2399808320924437

Pijanowski, B. C., Tayyebi, A., Doucette, J., Pekin, B. K., Braun, D., & Plourde, J. (2014). A big data urban growth simulation at a national scale: Configuring the GIS and neural network based land transformation model to run in a high performance computing (HPC) environment. *Environment Modeling & Software*, *51*, 250–268. https://doi.org/10.1016/j.envsoft.2013.09.015

Pavlovic, M., Ilic, S., Antonic, N., & Culibrk, D. (2022). Monitoring the impact of large transport infrastructure on land use and environment using deep learning and satellite imagery. *Remote Sensing*, *14*(10), 2494. https://www.mdpi.com/2072-4292%20/14/10/2494

Pennington, K. (2021). Does building new housing cause displacement? The supply and demand effects of construction in San Francisco. *SSRN Electronic Journal*, 1–67. http://dx.doi.org/10.2139/ssrn.3867764

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. *Proceedings of the International Conference on Machine Learning*, 8748–8763. https://proceedings.mlr.press/v139/radford21a/radford21a.pdf

Rahimi, A. (2016). A methodological approach to urban land use change modeling using infill development pattern—A case study in Tabriz, Iran. *Ecology Processes*, *5*(1), 1–15. https://doi.org/10.1186/s13717-016-0044-6

Reba, M., & Seto, K. C. (2019). A systematic review and assessment of algorithms to detect, characterize and monitor urban land change. *Remote Sensing of Environment*, *242,* 111739. https://doi.org/10.1016/j.rse.2020.111739

Romero, A., Gatta, C., & Camps-Valls, G. (2016). Unsupervised deep feature extraction for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(3), 1349–1362. https://doi.org/10.1109/TGRS.2015.2478379

Southern California Association of Governments (SCAG). (2024). *Connect SoCal*. SCAG. https://scag.ca.gov/sites/default/files/2024-05/23-2987-connect-socal-2024-final-complete-040424.pdf

Sant´e, I., García, A. M., Miranda, D., & Crecente, R. (2010). Cellular automata models for the simulation of real-world urban processes: A review and analysis. *Landscape and Urban Planning*, *96*(2), 108–122. https://doi.org/10.1016/j.landurbplan.2010.03.001

Schultheiss, M., Pattaroni, L., & Kaufmann, V. (2024). Planning urban proximities: An empirical analysis of how residential preferences conflict with the urban morphologies and residential practices. *Cities*, *152*, 105215. https://doi.org/10.1016/j.cities.2024.105215

Seltzer, E., & Carbonell, A. (2011). *Regional Planning in America: Practice and Prospect*. Lincoln Institute of Land Policy. https://www.lincolninst.edu/app/uploads/legacy-files/pubfiles/regional-planning-in-america-chp.pdf

Steinacker, A. (2003). Infill development and affordable housing: Patterns from 1996 to 2000. *Urban Affair Review*, *38*(4), 492–509. https://doi.org/10.1177/1078087402250357

Stokenberga, A. (2014). Does bus rapid transit influence urban land development and property values: A review of the literature. *Transport Reviews*, *34*(3), 276-296. https://doi.org/10.1080/01441647.2014.902404

Suchman, D. R., & Sowell, M. B. (1997). *Developing Infill Housing in Inner-City Neighborhoods: Opportunities and Strategies*. Urban Land Institute.

Talen, E. (2006). Neighborhood-level social diversity: Insights from Chicago. *Journal of the American Planning Association*, *72*(4), 431–446. https://doi.org/10.1080/01944360608976764

Taylor, M. (2015). *California's high housing costs: Causes and consequences*. California Legislative Analyst's Office. https://lao.ca.gov/reports/2015/finance/housing-costs/housing-costs.pdf

U. S. Environmental Protection Agency (USEPA). (2014). *Smart growth and economic success: Investing in infill development.* USEPA. https://www.epa.gov/sites/default/files/2014-06/documents/developer-infill-paper-508b.pdf

U. S. Environmental Protection Agency (USEPA). (2021). National walkability index: Methodology and user guide. US EPA. https://www.epa.gov/sites/default/files/2021-06/documents/national_walkability_index_methodology_and_user_guide_june2021.pdf

van Vliet, J., Naus, N., van Lammeren, R. J., Bregt, A. K., Hurkens, J., & van Delden, H. (2013). Measuring the neighbourhood effect to calibrate land use models. *Computers, Environment and Urban Systems*, *41*, 55–64. https://doi.org/10.1016/j.compenvurbsys.2013.03.006

Viana, C. M., Oliveira, S., Oliveira, S. C., & Rocha, J. (2019). Land use/land cover change detection and urban sprawl analysis. In H. R. Pourghasemi & C. Gokceoglu (Eds.), *Spatial modeling in GIS and R for earth and environmental science*. Elsevier.

Wheeler, S. (2002). *Smart infill*. Greenbelt Alliance.

Wilder, M.G. (1985). Site and situation determinants of land use change: An empirical example. *Economic Geography*, *16*(4), 332–344. https://www.jstor.org/stable/144053

Wiley, K. (2009). *An exploration of the impact of infill on neighborhood property values* [Doctoral dissertation, University of Maryland].

Yang, E., Hong, S., & Kim, Y. (2022). Factors influencing adaptive reuse of declining shopping malls in the U.S.: A multi-stakeholder view. *Journal of Green Building*, *17*(2), 83–108. https://doi.org/10.3992/jgb.17.2.83

Yuan, F., Sawaya, K. E., Loeffelholz, B. C., & Bauer, M. E. (2005). Land cover classification and change analysis of the Twin Cities (Minnesota) metropolitan area by multitemporal Landsat remote sensing. *Remote Sensing of Environment*, *98*(2–3), 317–328. https://doi.org/10.1016/j.rse.2005.08.006

About the Author

Dohyung Kim, PhD

Dr. Kim is a Professor in the Department of Urban and Regional Planning at California State Polytechnic University, Pomona. Since 2009, he has taught GIS applications in planning, local transportation planning, transportation methods and analysis, and planning research methods. Dr. Kim's research and practice focus on GIS applications in sustainable transportation planning, land use and transportation dynamics, and transportation safety. His recent research includes urban/suburban housing development in conjunction with transportation.

MTI FOUNDER

Hon. Norman Y. Mineta

MTI BOARD OF TRUSTEES

Directors