

Original Paper

Geospatial Modeling of Deep Neural Visual Features for Predicting Obesity Prevalence in Missouri: Quantitative Study

Butros M Dahu, PhD; Solaiman Khan, MA; Imad Eddine Toubal, PhD; Mariam Alshehri, PhD; Carlos I Martinez-Villar, BSc; Olabode B Ogundele, PhD; Lincoln R Sheets, MD, PhD; Grant J Scott, PhD

University of Missouri, Institute for Data Science and Informatics, Columbia, MO, United States

Corresponding Author:

Butros M Dahu, PhD

University of Missouri

Institute for Data Science and Informatics

Columbia, MO

United States

Phone: 1 8325124825

Email: peterdahu@gmail.com

Abstract

Background: The global obesity epidemic demands innovative approaches to understand its complex environmental and social determinants. Spatial technologies, such as geographic information systems, remote sensing, and spatial machine learning, offer new insights into this health issue. This study uses deep learning and spatial modeling to predict obesity rates for census tracts in Missouri.

Objective: This study aims to develop a scalable method for predicting obesity prevalence using deep convolutional neural networks applied to satellite imagery and geospatial analysis, focusing on 1052 census tracts in Missouri.

Methods: Our analysis followed 3 steps. First, Sentinel-2 satellite images were processed using the Residual Network-50 model to extract environmental features from 63,592 image chips (224×224 pixels). Second, these features were merged with obesity rate data from the Centers for Disease Control and Prevention for Missouri census tracts. Third, a spatial lag model was used to predict obesity rates and analyze the association between deep neural visual features and obesity prevalence. Spatial autocorrelation was used to identify clusters of obesity rates.

Results: Substantial spatial clustering of obesity rates was found across Missouri, with a Moran I value of 0.68, indicating similar obesity rates among neighboring census tracts. The spatial lag model demonstrated strong predictive performance, with an R^2 of 0.93 and a spatial pseudo R^2 of 0.92, explaining 93% of the variation in obesity rates. Local indicators from a spatial association analysis revealed regions with distinct high and low clusters of obesity, which were visualized through choropleth maps.

Conclusions: This study highlights the effectiveness of integrating deep convolutional neural networks and spatial modeling to predict obesity prevalence based on environmental features from satellite imagery. The model's high accuracy and ability to capture spatial patterns offer valuable insights for public health interventions. Future work should expand the geographical scope and include socioeconomic data to further refine the model for broader applications in obesity research.

(JMIR AI 2024;3:e64362) doi: [10.2196/64362](https://doi.org/10.2196/64362)

KEYWORDS

geospatial modeling; deep convolutional neural network; DCNN; Residual Network-50; ResNet-50; satellite imagery; Moran I; local indicators of spatial association; LISA; spatial lag model; obesity rate; artificial intelligence; AI

Introduction

Overview

The prevalence of obesity has escalated to alarming levels globally, prompting an urgent need for innovative approaches to understand and combat this complex health issue. Spatial

technologies, such as geographic information systems (GISs), remote sensing (RS), spatial machine learning (ML), and spatial analysis, have emerged as powerful tools in obesity research, offering novel insights into the environmental and social determinants of this epidemic. This paper delves into the current applications of spatial technologies in obesity research,

highlighting how these tools can be used to unravel the intricate relationship between the built environment and obesity rates.

Furthermore, this paper explores the future promise of these technologies in advancing our understanding of obesity, guiding public health interventions, and shaping policies to create healthier and more equitable communities. Through a comprehensive examination of recent studies and advancements, this paper underscores the pivotal role of spatial technologies in transforming obesity research and ultimately contributing to the global fight against this pressing public health challenge.

Objectives

This research aims to use deep convolutional neural networks (DCNNs) to examine medium-resolution satellite imagery, with the goal of predicting obesity rates using deep neural visual features (DNVFs). Concentrating on the 1052 census tracts in Missouri, the project seeks to provide a scalable method for predicting obesity prevalence using DNVFs. This could improve the accuracy of public health initiatives and policy decisions.

Background

GISs, GPS, RS, and DCNNs such as Residual Network-50 (ResNet-50) are reshaping obesity research. GISs have been pivotal in depicting spatial distributions of obesity and its determinants, crafting metrics of obesity-promoting environments, and unveiling spatial patterns of obesity prevalence and obesogenic environmental attributes. GPS has predominantly been used to map individual movement patterns and, in conjunction with other instruments, to track behaviors contributing to obesity. The contribution of RS in supplying data on natural and built environments has been undervalued [1,2].

The integration of DCNNs, particularly ResNet-50, and the extraction of DNVFs have enhanced the precision of obesity predictions by analyzing satellite imagery and identifying subtle environmental features linked to obesity. When combined with spatial ML techniques, these advanced technologies are proving invaluable in providing more granular insights into obesogenic environments and individual exposure levels, thereby enriching our understanding of obesity's etiology and the impact of various interventions on obesity rates in Missouri [1,3,4].

By 2030, it is projected that 38% of the global population will be overweight, while another 20% will be obese [5,6]. This alarming trend toward a worldwide epidemic of "globesity" has raised concerns about the rapid and significant shift in global health patterns. The United States, and specifically the state of Missouri, is no exception to this trend, with an increasing burden of individuals who are overweight and obese.

Changes in dietary patterns, particularly the consumption of energy-dense, nutrient-poor foods, are closely linked to the rising prevalence of overweight and obesity and their associated health complications. Therefore, this study aims to explore the spatial patterns of overweight and obesity among adults in Missouri and predict the obesity rate for each census tract to increase public health awareness [7].

Significance

Earlier studies have shown links between different elements of the built environment and their effects on obesity and physical activity across various life stages [8-10]. Previous research has demonstrated a connection between obesity and various environmental factors, including the walkability of a region, land use patterns, urban sprawl, residential types, access to amenities such as recreational facilities and food outlets, socioeconomic deprivation, and perceived safety levels in a community [11-13]. In addition, proximity to and availability of natural spaces and sidewalks are linked to increased and more regular levels of physical activity, especially in urban settings [14,15].

While a relationship between obesity and the built environment is acknowledged, inconsistencies exist in the results of various studies and across different geographic regions concerning how specific features of the built environment influence obesity rates [11,16,17]. These discrepancies could arise from variations in the measurement methods and tools used in these studies, making it challenging to assess and compare the outcomes [18-20]. Furthermore, the measurement of these environmental features often requires substantial time and expenses and is susceptible to human error and bias [21,22]. There is a need for methodologies that provide standardized measurement criteria to enable comparisons between various studies [19,23,24]. Accurately assessing the impact of the built environment on obesity is crucial for designing and implementing successful community-based prevention and intervention programs [21,25,26].

In this paper, we introduce an innovative approach to extensively explore the relationship between adult obesity prevalence and different features of the spatial characteristics. This method uses a deep learning (DL) technique, using a DCNN to examine the physical characteristics of neighborhoods through medium-resolution (10 m) satellite imagery. Building on the foundational work of Maharana and Nsoesie [27] and Nguyen et al [28], who used DCNNs to analyze urban Google Street View (Google LLC) images focusing on predefined features, such as crosswalks, building types, and greenery, our study expands the scope to both urban and rural areas. Unlike Maharana and Nsoesie [27], who used a Visual Geometry Group-8 architecture and extracted features after training, we used the more advanced ResNet-50 architecture, using features extracted from a pretrained model to enhance efficiency and generalizability. Our approach provides a more comprehensive analysis of the DNVFs, identifying specific correlations with obesity rates at the census tract level across 1052 census tracts in Missouri. Moreover, our method is scalable, uses publicly available data and computing resources, allows for comparisons between various studies, and can be adapted to different geographic locations and regions.

RS Applications in Obesity Research

RS data, typically stored as raster data, offer a highly efficient means of spatially representing obesity risk. Raster data consist of regular square grids, with each grid cell aggregating information over a corresponding area on Earth's surface, such as high-resolution aerial photos and satellite images. These data

can be generated from radiation recorded by sensors or through GIS-based interpolation techniques from discrete data collected at observation stations [29-32].

RS technology, which acquires information through spaceborne satellites or airborne sensors without direct physical contact, records the intensity of radiation reflected or emitted by objects and converts it into various land surface (eg, vegetation) and meteorological (eg, temperature) properties for user analysis.

This technology has been instrumental in diverse applications, including environmental monitoring and climate-sensitive disease risk modeling, by providing extensive environmental data [1,29,33-35].

In obesity research, RS plays a pivotal role by offering detailed environmental data that can be used to identify obesogenic factors, such as the availability of green spaces or urban walkability. This spatial information, when combined with other data sources, can help researchers understand the complex interplay between the environment and obesity risk. Vector data, which include points, lines, and polygons with geographical coordinates, complement raster data by providing precise location information for features relevant to obesity studies [36-39].

Literature Review and Related Work

DL Applications in Urban and Environmental Health Research

DL techniques have become integral to understanding the complex dynamics between urban environments and public health outcomes. Through the application of convolutional neural networks (CNNs) and semantic segmentation architectures, these methodologies enable detailed analysis of spatial data for health-related urban studies.

CNNs in Urban Health Analysis

Pala et al [40] used transfer learning with a pretrained CNN to classify urban structures in high-resolution satellite imagery. This method facilitates the extraction of latent features indicative of various urban elements, critical for analyzing the impact of the urban environment on public health. The subsequent application of k-means clustering to categorize image tiles based on these features illustrates the synergy between CNNs and traditional data analysis methods, allowing the identification of meaningful correlations between urban characteristics and health outcomes.

Levy et al [41] explored the use of CNNs for predicting mortality rates from satellite images, demonstrating the models' capability to discern spatial patterns related to health outcomes. By comparing DL models with linear regression and a hybrid approach, the study delineated CNNs' efficiency in modeling complex, nonlinear relationships between environmental features and health indicators.

Guo et al [42] used Bayesian analysis and partial least squares regression in conjunction with the ResNet-50 architecture to examine obesity trends and the impact of environmental factors on obesity prevalence among Chinese children and adolescents. This integration of statistical methods with DL techniques

demonstrates a comprehensive approach to understanding the multifaceted influences on obesity, offering predictive insights into spatial and temporal variations of health outcomes.

Semantic Segmentation for Urban Environment Analysis

Han et al [43] deployed SegNet for semantic segmentation of Google Street View images to study urban environments' impact on psychological stress. SegNet, designed for pixel-wise segmentation, features an encoder-decoder architecture for contextual detail abstraction and precise urban element classification. This approach enabled detailed urban feature analysis, including buildings and green spaces, for assessing their influence on psychological stress. Using a combination of segmented imagery analysis within a human-machine adversarial framework and random forest classification, the study provided insights into visual urban characteristics' effects on stress, demonstrating DL's utility in urban health research.

Hong et al [44] used the U-Net architecture for semantic segmentation of unmanned aerial vehicle imagery to map green spaces and sidewalks, integrating DL with a GIS for urban and public health analysis. U-Net's encoder-decoder design, enhanced with skip connections, excels in extracting detailed features from high-resolution unmanned aerial vehicle images, crucial for accurate environmental mapping at the neighborhood level. This method optimizes the segmentation process, distinguishing between various urban elements and facilitating the comprehensive analysis of green spaces and sidewalks. By leveraging U-Net for precise pixel classification, the study demonstrates DL's efficacy in improving urban environment datasets, thus supporting targeted urban planning and public health initiatives.

In their analysis, Wang et al [45] applied the Fully Convolutional Network (FCN)-8s architecture to semantic segmentation of Tencent Street View (Tencent) images for studying the built environment's effects on health in older adults within Beijing's Haidian District. FCN-8s, a CNN that is fully convolutional and optimized for pixel-level segmentation, enabled the classification of urban features directly from images, facilitating the examination of variables, such as wealth, safety, and green space. The network uses convolution and deconvolution layers for processing spatial information, enhancing segmentation accuracy for complex urban scenes. This approach underscores the applicability of FCN-8s in urban analysis, providing a detailed assessment of environmental attributes and their correlations with health outcomes, thereby supporting data-driven urban planning and public health policy development.

Larkin et al [46] used GISs, satellite imagery, and a pyramid scene parsing network (PSPNet), designed for semantic segmentation, to analyze urban perceptions across 56 cities. PSPNet classifies each pixel in Google Street View images into urban elements (eg, trees, buildings, and roads) by leveraging a pyramid pooling module that aggregates context information at different scales, ensuring detailed scene parsing and global context comprehension. This method, combined with GIS-derived data on population density and green spaces, allowed for the quantitative linking of urban features to perceptions of safety, liveliness, and beauty. The study

showcases the integration of DL with GISs, demonstrating PSPNet's effectiveness in urban landscape analysis and the potential of merging image-derived metrics with environmental data for insights into the built environment's impact on human perceptions.

Discriminative and Generative Models in Urban Health Studies

Newton detailed both discriminative and generative models within the framework of CNNs for obesity studies in urban contexts, as detailed in the chapter "Deep Learning in Urban Analysis for Health" of the book *Artificial Intelligence in Urban Planning and Design* [47]. This examination reveals DL's capacity to identify subtle spatial patterns associated with health outcomes, reinforcing its value in enhancing urban planning and public health policy through advanced model applications.

Review Studies on ML and DL Applications

Siddiqui et al [48] offered a comprehensive review of ML and DL models for predicting childhood and adolescent obesity. This survey integrated 39 studies, examining a variety of ML models (including artificial neural network, recurrent neural network, and CNN) and datasets (ranging from surveys and cohorts to electronic health records and image datasets). The literature was categorized based on methodologies (traditional ML vs DL), dataset types, and outcomes predicted (overweight, obesity, or both). The survey distinguished between studies aimed at identifying risk factors and those predicting obesity, underlining the significance of model interpretability. It identified research gaps and advocated for advancements through large-scale multimodal datasets and the development of interpretable models for obesity prediction.

Wirtz Baker et al [49] conducted a systematic review under PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines to evaluate the use of nontraditional data sources, such as geospatial data and social media, in studying the built environment's impact on obesity. Analyzing 53 studies, this review emphasized the value of geospatial data within GISs for assessing physical environment features and the insights social media data offer into sociocultural dynamics related to obesity. The review suggested future research avenues should leverage digital advancements to gain a comprehensive understanding of obesity's environmental determinants.

Zhou et al [50] provided a mini-review focusing on the application of open-source ML models in obesity research. This review selected 25 studies, categorizing ML models into supervised and unsupervised learning and covering applications in nutrition, the environment, genetics, and the microbiome. The review aimed at research reproducibility and efficiency, advocating for open-source tools and the integration of diverse data sources for a multifaceted analysis of obesity's causes.

Shaamala et al [51] presented a systematic review of ML applications in geospatial analysis, identifying key application areas including classification, detection, extraction, clustering, regression, modeling, prediction, and optimization. Techniques, such as random forests, support vector machines, and CNNs, were highlighted for their effectiveness in classification,

detection, and extraction tasks, with specific mention of architectures such as U-Net and You Only Look Once for semantic segmentation and object detection. The review also discussed the use of ensemble models and advanced neural networks in regression, modeling, and prediction tasks, along with optimization algorithms for spatial planning. This work aimed to map the current landscape and identify future research directions in ML applications for geospatial analysis.

Cross-Disciplinary Applications of Advanced ML

Recent advances in ML [52-54], highlight the potential of such technologies in both pandemic responses and chronic disease management like obesity. These studies used sophisticated algorithms— dilated efficient residual global attention, fuzzy ensemble models, and artificial intelligence techniques for minimal data analysis, to enhance diagnostic and predictive accuracies, akin to our use of DCNNs for analyzing environmental determinants of obesity from satellite imagery. The success of these methods in medical diagnostics underscores their applicability in spatial health studies, suggesting promising avenues for using similar approaches in obesity research to improve public health outcomes.

Spatial Analysis and Geolocation

Spatial analysis and geolocation are pivotal tools in obesity research, providing crucial insights into the complex interplay between environmental factors and obesity prevalence. By leveraging these techniques, researchers can map and analyze spatial patterns of obesity, identify obesogenic environments, and understand how individual interactions with these environments contribute to obesity risk [55,56]. Through the integration of GISs, GPS, and RS, spatial analysis and geolocation enable the examination of geographical variations in obesity rates, the assessment of accessibility to health-promoting resources, and the evaluation of the impact of urban planning on physical activity levels. Consequently, they play an indispensable role in advancing our understanding of the spatial determinants of obesity and informing targeted interventions and policies to combat this global health challenge [55,57].

Methods

Overview

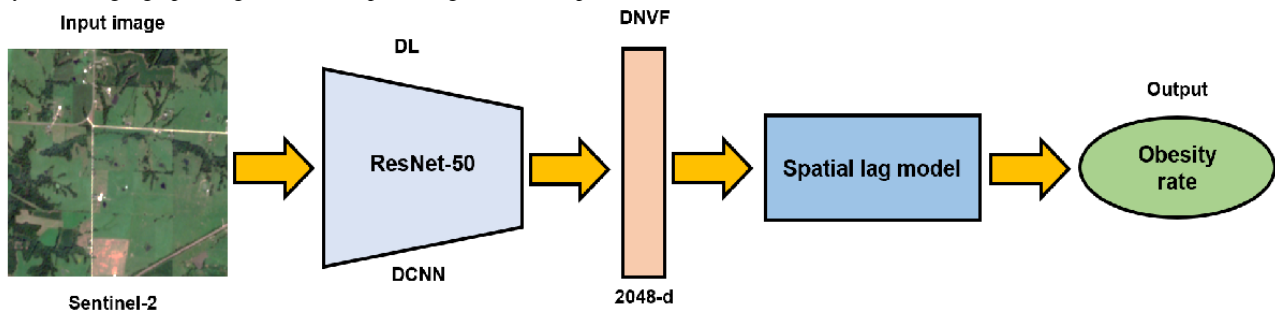
Our study analysis consisted of 4 main steps. First, we processed Sentinel-2 satellite images to extract features of the built environment using ResNet-50. Second, we merged the polygon's census tracts using the TIGER (topologically integrated geographic encoding and referencing)/Line shapefiles with the obesity rates for each census tract using the US Centers for Disease Control and Prevention (CDC) data [58]. Third, we conducted an exploratory spatial analysis (ESA), particularly spatial autocorrelation and local clustering, to determine the modeling approach and techniques. Fourth, as our ESA suggests that there is substantial spatial autocorrelation and spatial dependency in the obesity rates, we used a spatial regression technique, particularly the spatial lag model (SLM), to build a model to predict the obesity rate for each census tract and to

assess the association between the built environment and obesity prevalence.

Figure 1 illustrates an analytical workflow designed to predict obesity rates from satellite imagery. The process begins with a high-resolution input image from the Sentinel-2 satellite, capturing the intricate details of the Earth's surface. The image on the left in Figure 1 was raw data that was fed into a DL framework. Inside this framework, a ResNet-50 DCNN was

used to analyze and extract high-dimensional features from the image, known as DNVFs, which are 2048D vectors encapsulating the critical visual cues from the input data. Subsequently, the extracted features were used as the input of the SLM, which finally output the predicted obesity rate of the given geographic unit. This output is represented as a quantifiable measure, providing valuable insights for public health and resource allocation.

Figure 1. Flowchart illustrating the process of using Sentinel-2 satellite imagery processed through a Residual Network-50 (ResNet-50) deep convolutional neural network (DCNN) to extract deep neural visual features, which are then used as the inputs of the spatial lag model, ultimately yielding predicted obesity rates for geographic regions. DL: deep learning; DNVF: deep neural visual feature.



Obesity Prevalence Data

Figure 2 provides a choropleth map that serves as a visual representation of obesity prevalence among the various census tracts within the state of Missouri for the year 2022. Each census tract is color-coded to denote the obesity rate percentage for individuals residing in that tract, with the intent of offering a comprehensive geographic overview of the obesity landscape across the state. In the map, lighter shades of blue show lower obesity rates, and darker hues correspond to higher obesity rates. The scale itself delineates a range starting at 25%, represented by a light blue, and progresses to 50%, indicated by a dark blue color. Notably, the map reveals a significant variation in obesity rates across different regions, with some census tracts exhibiting markedly higher rates and others reflecting lower obesity rates.

We used 2022 estimates of annual crude obesity prevalence at the census tract level, derived from the 500 Cities project [59,60]. These estimates are based on data from the Behavioral Risk Factor Surveillance System, which surveys individuals aged ≥ 18 years [60]. Obesity is identified using a BMI threshold of 30, calculated as the individual's weight in kg divided by their height in m^2 [59]. Our study focused on the mid-Missouri region in the United States. The 1052 census tracts (Missouri State) covered in this study have an aggregate area of 69,707 square miles (180,540 km^2). They have a total population of 6.2 million (based on the 2020 census).

The CDC data list 1387 census tracts, 4506 block groups, and 343,565 census blocks. Given that the number of TIGER/Line census tract shapefiles in the state was 1654, polygon (census tract) IDs in the TIGER/Line data had to be aligned to census tract IDs in the CDC data. To fix the mismatching issue, census tracts with subdivisions (tract names with 2 trailing digits different from 0) in both datasets were joined into larger polygons.

First, we joined the polygons in the TIGER/Line dataset by removing all subdivisions.

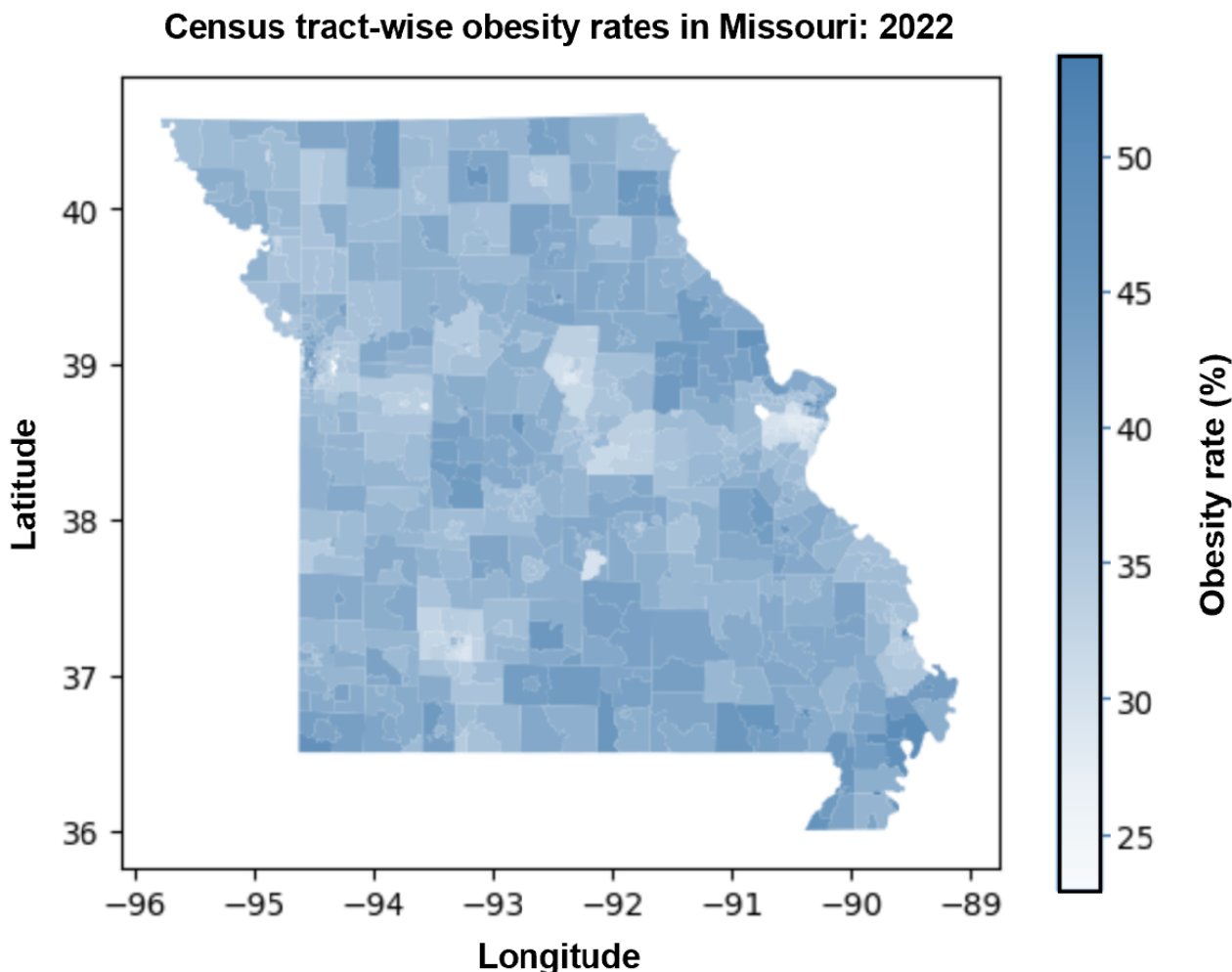
Of the initial 1654 polygons, 55.26% ($n=914$) had names following a naming convention of the type "XXXX.YY," where "YY" corresponded to the subdivision within a particular tract. The remaining 44.74% ($n=740$) had no subdivisions and names containing 2 trailing zeros (eg, "XXXX.00"). Setting the 2 trailing digits to 0 in these 914 names resulted in repeated names with 323 unique names. If a name was repeated, all repeated elements were joined to become a single polygon, which resulted in a set of 1063 polygons.

Similarly, the IDs of the census tracts in the CDC data consisted of a string of digits, with the last 2 digits corresponding to a set of subdivisions different from that of the TIGER/Line data. Of the 1387 tracts, 63.52% ($n=881$) had IDs with 2 trailing zeros (no subdivisions), while the remaining 36.48% ($n=506$) had subdivisions. We set the initial 506 string IDs to a single subdivision with repeated entries, which resulted in 178 unique IDs. Because each of these new unique IDs had multiple obesity rates, an average obesity rate was calculated as a proxy for the obesity rate of the newly joined area. The average obesity rate for these joint areas, weighted by census tract population, was calculated as follows:

$$\bar{w} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}$$

where \bar{w} is the new obesity rate for the joint area, n is the number of subdivisions (repeated entries) within the original census tract, w_i is the population in a subdivision i , and x_i is the obesity rate in i . The joined 1059 CDC entries were then matched to the joined 1063 TIGER/Line tracts, for a final overlapping set of 1055 census tract polygons with their corresponding obesity rates, which we used as the inputs to our models.

Figure 2. Choropleth map displaying the distribution of obesity rates for individuals across Missouri census tracts in 2022. The variations in the color intensity reflect the range of obesity prevalence, with darker blue indicating higher obesity rates. The color scale to the right quantifies the obesity rates corresponding to each color shade.



Acquiring Satellite Imagery

We acquired our imagery inputs by selecting the Sentinel-2 products intersecting our previously defined set of census tracts. These satellite products were downloaded from ESA’s Copernicus Dataspace Ecosystem. Because ESA’s OpenSearch application programming interface uses http requests to search for products, we defined a shortened geometry string that could fit in our search query. This was done by joining our previously defined census tract polygons into a state boundary that was further simplified into a closed polygon of 54 vertices using the implementation of the Douglas-Peucker algorithm included in GeoPandas.

Figure 3 presents our geospatial imagery coverage data analysis using 33 Sentinel-2 satellite images of Missouri from the year

2022, each with a resolution of 10,980×10,980 pixels at 10 m per pixel.

Our search resulted in 187 intersecting Sentinel-2 products between July 1, 2022, and August 31, 2022. Overlapping images were removed in 2 steps. First, products with completely overlapping geometries (corresponding to the same universal transverse mercator zone tile) were filtered by discarding all but the product with the largest area and the lowest cloud percentage. Second, 7 partially overlapping products (which also happened to have little state coverage) were discarded after visual inspection. This resulted in a set of 33 Sentinel-2 images that were used to define our model inputs. These products were downloaded from ESA’s Dataspace Ecosystem to Nautilus. All 33 Sentinel-2 image sizes were 10,980×10,980 pixels. The images were then normalized and tiled into 224×224 pixel chips. This created 82,500 three-band (red-green-blue) image chips.

Figure 3. A chart illustrating the collection of 33 Sentinel-2 satellite image tiles, each identified by a specific tile code and date of capture in 2022. The chart provides information on cloud cover across these images, with a mean cloud coverage of 0.769% and a maximum cloud coverage of 4.953%.

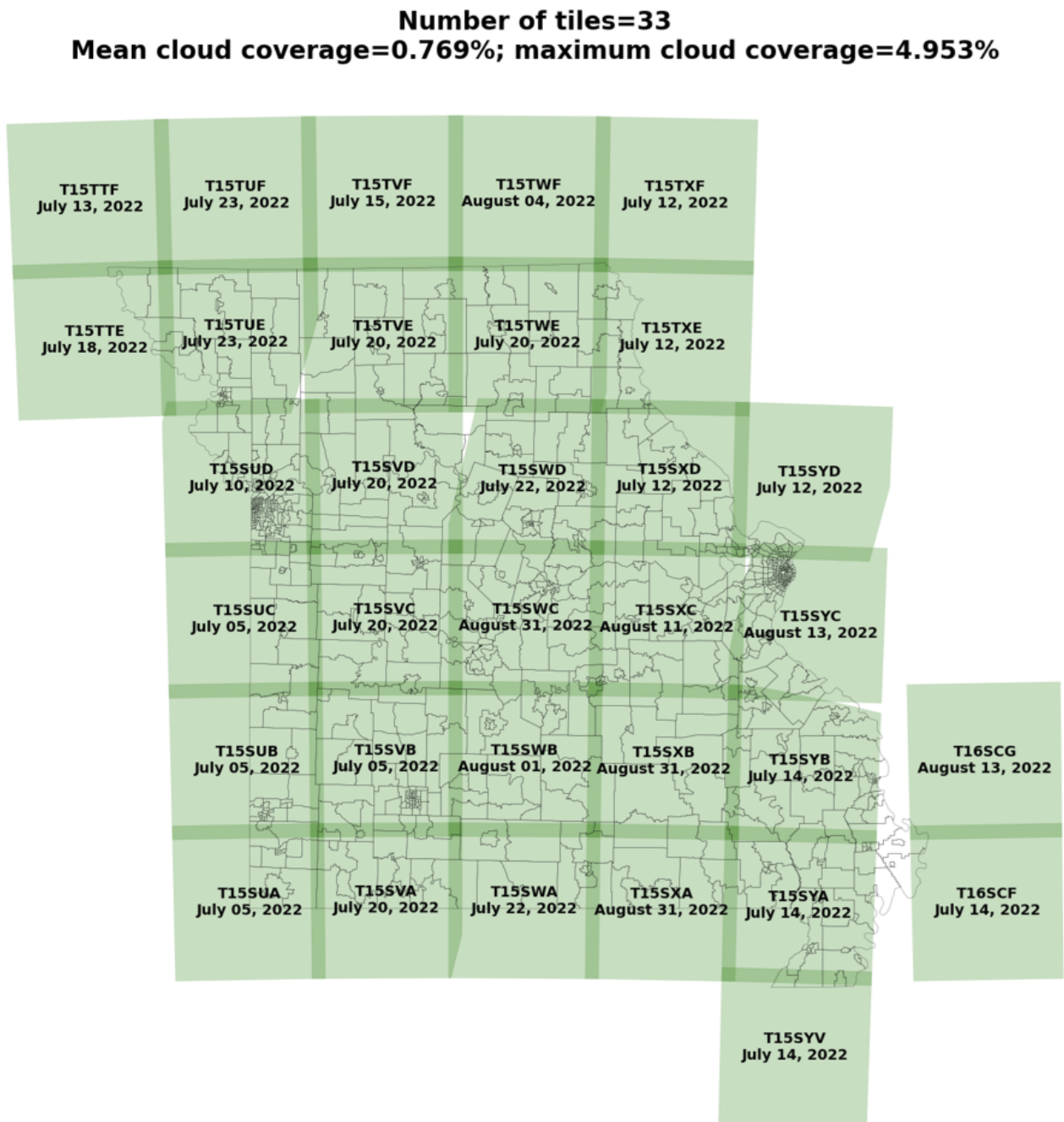


Image Processing

Overview

Building on acquiring Sentinel-2 imagery, our image-processing efforts used DCNNs to advance the analysis of these extensive datasets [61,62]. We used the DNVFs from a pretrained neural network to obtain the built area features of the 82,500 (63,592 intersected with Missouri census tracts) satellite image chips. We used the ResNet-50 network, which is composed of 50 layers (48 convolution layers along with 1 max pool and 1 average pool layer) and is trained on approximately 1.2 million images from the ImageNet database (a dataset of >14 million images used for large-scale visual recognition challenges) for

recognizing objects belonging to 1000 categories [63]. For each chip we passed through the network, we extracted the 2048 features from the last hidden layer of the network before the output layer [64]. Because each census tract in our dataset could have >1 image chip intersecting it, we calculated a corresponding weighted mean feature vector for a tract with the features of the intersecting chips [65]. For a census tract t , its mean feature vector F_t was calculated as follows:

$$F_t = \frac{\sum_{c \in C} w_c F_c}{\sum_{c \in C} w_c}$$

where C is the set of image chips intersecting the census tract, F_c is a ResNet-50 feature vector obtained from chip c , and w_c is the (scalar) number of pixels in c intersecting the census tract. To include the obesity rate of a given chip in our analysis, the inverse case was considered. When a chip intersected with >1 census tract, we calculated the obesity rate in such a chip as the average obesity rate of all the intersecting census tracts [65]. So, for a chip c , the weighted obesity rate is similarly calculated as follows:

$$o_c = \frac{\sum_{t \in T} w_t o_t}{\sum_{t \in T} w_t}$$

where T is the nonempty set of tracts intersecting a chip c , o_t is the obesity rate in a census tract, and w_t is the number of pixels in c corresponding to the t census tract. We do not link these features to specific elements in the built environment or the obesity prevalence. Rather, these DNVFs collectively represent an indicator to help predict the obesity prevalence for each census tract. The image chips were tiled for a size of 224×224 pixels or 2240 m (2.24 km) by 2240 m (2.24 km). The graphics processing unit (NVIDIA GeForce RTX 2080 Ti)-accelerated processing speed was approximately 77.51 tiles per second or 1064.25 seconds (17.73 min) of DCNN inference for our study area. This translates to a rate of 0.0129 seconds per tile.

Spatial Autocorrelation

Following image processing, we explored spatial autocorrelation to understand the interdependencies of obesity rates across different geographic locales. Spatial autocorrelation is a concept in spatial analysis that quantifies the degree of similarity between observations in neighboring geographic locations. It measures how much the value of a variable at one location is influenced by the values of the same variable at nearby locations. Positive spatial autocorrelation indicates that similar values are clustered together, while negative spatial autocorrelation suggests that dissimilar values are adjacent to each other. This concept is crucial in identifying patterns and relationships in geographic data, as it helps to determine whether the spatial distribution of a variable is random, clustered, or dispersed [66,67].

In obesity research, spatial autocorrelation is used to examine the spatial distribution of obesity rates or related factors, such as physical activity levels, access to healthy foods, or socioeconomic status. By assessing the degree of spatial autocorrelation, researchers can identify areas where obesity rates are higher or lower than expected based on the surrounding locations. This analysis can reveal clusters of high or low obesity prevalence, indicating potential hot spots or areas of concern. It also helps to understand the spatial dynamics of obesity, which can inform targeted public health interventions and policies [67,68].

The application of spatial autocorrelation in obesity research extends to exploring the environmental and social determinants of obesity. By analyzing the spatial autocorrelation of variables, such as access to parks, fast-food outlets, or income levels, researchers can investigate how these factors contribute to the

spatial variation in obesity rates. This can lead to a better understanding of the complex interactions between the built environment, socioeconomic factors, and obesity, ultimately guiding the development of more effective strategies for obesity prevention and management at the local and regional levels [67-69].

Global Moran I

To quantitatively measure the spatial autocorrelation observed, we applied global Moran I, a statistical tool that evaluates the correlation of a variable with its spatially lagged counterpart. It is a widely used tool in spatial analysis, especially in fields such as geography, ecology, and epidemiology [70,71].

Moran I is defined as the correlation coefficient between a variable and its spatially lagged counterpart. It measures the extent to which similar values of a variable are located near each other in space [70,72].

The global Moran I is calculated as follows:

$$I = \frac{N}{W} \times \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2}$$

where N is the number of spatial units and w_{ij} is the spatial weight between units i and j . x_i and x_j are the values of the variable of interest at units i and j , respectively. \bar{x} is the mean value of the variable across all spatial units. W is the sum of all spatial weights, $\sum_{i=1}^N \sum_{j=1}^N w_{ij}$.

Moran I is a crucial measure in spatial analysis, used to identify and quantify spatial patterns in a dataset. A positive Moran I indicates spatial clustering, where areas with similar values are geographically close to each other. For instance, regions with high values tend to be surrounded by other high-value regions, and the same applies to low values. In contrast, a negative Moran I suggests spatial dispersion, where areas with high values are typically surrounded by areas with low values, indicating a dissimilar distribution. When Moran I is close to 0, it implies a random spatial pattern, signifying no significant spatial autocorrelation among the observed values [70-72].

In our research, we focused exclusively on Moran I as our measure of spatial autocorrelation, opting not to use Geary C. While Geary C is sensitive to local variations and emphasizes the dissimilarity between neighboring observations, Moran I provides a more comprehensive view of the overall spatial pattern. To assess the statistical significance of Moran I, we used permutation tests or analytical methods to create a reference distribution under the null hypothesis of no spatial autocorrelation. By comparing the observed Moran I to this distribution, we can determine its significance and thereby gain valuable insights into the spatial relationships and processes at play in our study [70-72].

Local Indicators of Spatial Association

Further dissecting spatial patterns, we used local indicators of spatial association (LISA) statistics to pinpoint specific areas exhibiting significant spatial clustering or dispersion. Unlike global measures, such as Moran I, which provide a single

summary statistic for the entire study area, LISA statistics offer insights into the spatial patterns at individual locations, allowing for the detection of local clusters or outliers [75,76].

LISA statistics are designed to assess the degree of spatial autocorrelation for each observation within a dataset, considering the values of neighboring observations. They help to identify areas where the value of a variable is significantly different from or similar to its surrounding values, indicating the presence of spatial clusters or hot spots [75,76].

LISA statistics play a pivotal role in spatial analysis by providing a detailed examination of local spatial patterns within a dataset. These statistics are instrumental in detecting local clusters of high or low values, commonly referred to as hot spots or cold spots, which can reveal areas of concentrated phenomena, such as high crime rates or regions of environmental degradation [76,77].

In addition, LISA statistics are adept at identifying spatial outliers, where the value of an observation significantly deviates from the values of its neighboring observations, indicating anomalies or irregularities in the spatial distribution. By exploring the local spatial patterns and relationships, LISA statistics offer a nuanced understanding of the spatial dynamics at play, allowing researchers to uncover subtle variations and trends that may be obscured in a global analysis. This localized approach is essential for targeted interventions and policy making, as it provides a granular view of the spatial structure of the data, enabling more precise and effective responses to spatially varying challenges [74,75,77].

One of the most commonly used LISA statistics is local Moran I. The local Moran I for observation i is calculated using the following equation:

$$I_i = \frac{(x_i - \bar{x})}{s^2} \sum_{j=1}^n w_{ij} (x_j - \bar{x})$$

where I_i is the local Moran I for observation i . x_i and x_j are the values of the variable for observations i and j , respectively. \bar{x} is the mean value of the variable across all observations. s^2 is the variance of the variable. w_{ij} is the spatial weight between observations i and j . n is the total number of observations.

The local Moran I statistic is used to assess the degree of spatial autocorrelation at the local level, identifying clusters of similar values and spatial outliers.

In assessing the significance of LISA statistics, permutation tests are commonly used. This involves randomly shuffling the observed values and recalculating the statistic multiple times to generate a reference distribution. By comparing the observed LISA values to this distribution, researchers can determine the statistical significance of the spatial patterns detected. This process is crucial for ensuring that the identified clusters and outliers are not due to random chance but are indicative of underlying spatial processes [74,77].

In the context of obesity research, LISA statistics provide invaluable insights into the local spatial dynamics of obesity prevalence and related factors. High positive values of Local Moran I can reveal clusters of areas with high obesity rates (high-high clusters), while high negative values can indicate regions where high obesity rates are surrounded by areas with lower rates (high-low outliers), and vice versa [73,76].

By visualizing these patterns on maps, researchers can identify specific neighborhoods or regions that may require targeted public health interventions. Moreover, the detailed local insights offered by LISA statistics can aid in understanding the spatial distribution of obesity and its association with environmental and socioeconomic factors, ultimately informing more effective strategies for addressing the obesity epidemic at a granular level [77].

Incorporating Spatial Dependencies With an SLM

Integrating the insights gained from both global and local spatial analyses, we used the SLM to incorporate the spatial dependencies identified among the observations into our regression analysis. It is used to account for the influence of neighboring observations on each other, which is a common phenomenon in spatial data [78].

In an SLM, the dependent variable in one location is assumed to be affected by the values of the dependent variable in neighboring locations. This spatial dependence is captured by including a spatially lagged dependent variable as an additional explanatory variable in the regression model [79].

SLMs are used in various fields, including economics, geography, and public health, to analyze spatial data where observations are not independent but influenced by nearby observations. It helps to provide more accurate estimates and inferences by accounting for spatial dependency [80].

In obesity research, SLMs can be used to study the spatial distribution of obesity rates and their determinants. For example, it can be used to examine how obesity rates in one area are influenced by the rates in adjacent areas, which might be due to shared environmental factors, social networks, or economic conditions. This can provide insights into the spatial diffusion of obesity and help in identifying areas that might benefit from targeted interventions [78-80].

The SLM is represented by the following equation:

$$y = \rho W y + X \beta + \varepsilon$$

where y is the vector of the dependent variable (eg, obesity rates). ρ is the spatial autoregressive parameter, representing the strength of the spatial dependence. W is the spatial weight matrix, defining the spatial relationship between observations. X is the matrix of independent variables (eg, socioeconomic factors and access to healthy food). β is the vector of coefficients for the independent variables. ε is the vector of error terms.

In this model, $\rho W y$ is the spatially lagged dependent variable, capturing the influence of neighboring observations on each observation in the dataset. In the SLM, the spatial weight matrix

W is essential for capturing the spatial correlations between nearby observations. We computed W using the contiguity method, more specifically the “queen” case, which is best suited for polygon features as it defines neighbors based on shared boundaries or vertices between spatial units (eg, census tracts). This approach identifies neighboring spatial units by determining which share at least 1 point of contact, either a border or a corner. Once the queen contiguity structure was defined, the weights were standardized so the sum of weights for each observation equals 1, ensuring comparable influence across all spatial units of analysis.

The estimation of the SLM typically requires specialized techniques, such as maximum likelihood or spatial 2-stage least squares, due to the presence of the spatially lagged dependent variable. The SLM is a powerful tool in spatial analysis, particularly valuable in fields, such as obesity research, where the spatial distribution of variables is of interest. By accounting for the influence of neighboring observations, the SLM provides a more accurate representation of spatial dependencies, leading to better-informed decisions and interventions.

Ethical Considerations

We did not seek further approval or exemption from an IRB because this study falls under Exemption 4 of the NIH Human

Subjects Research Exemptions. This exemption applies as our study involved secondary analysis of an existing, IRB-approved CDC dataset [60], which is deidentified and publicly available, hence not requiring informed consent.

Results

Table 1 presents a detailed summary of key metrics across various regions, such as population, area (in km²), obesity crude prevalence (%), and the number of chips. The data are systematically organized to illustrate the minimum, median, and maximum values for each metric, providing a clear understanding of the range and central tendencies within the dataset. For population, the minimum value recorded is 102, the median is 4058, and the maximum reaches 75,569, reflecting the diversity in population sizes across different areas. The area of the regions varies significantly, with the smallest being just 0.49 km², the median at 20.19 km², and the largest extending to 1787.47 km². In terms of obesity rates, the lowest rate observed is 23%, the median stands at 39.2%, and the highest rate is 53.7%, indicating varied health metrics across the regions. Finally, the number of chips ranges from a minimum of 1 to a maximum of 442, with a median of 14, highlighting different levels of chip distribution or consumption.

Table 1. A summary of key regional metrics, covering total population, area (km²), obesity crude prevalence, and number of chips intersected with each tract.

	Values, median (range)
Total population, n	4058 (102-75,569)
Area (km ²)	20.19 (0.49-1787.47)
Obesity crude prevalence (%)	39.2 (23-53.7)
Chips, n	14 (1-442)

Figure 4A details individual image chip boundaries, illustrating their overlap with 7 distinct census tracts (numbered for reference). **Figure 4B** further narrows down to census tract 0608, demonstrating the intersection with 150 specific image chips for granular analysis. The figure highlights the granularity and density of data distribution within the geographic study area.

Figure 4A displays a visualization of an individual image chip intersecting 7 distinct census tracts, enabling a detailed analysis of these specific overlaps. This image chip intersects with one of the highest numbers of census tracts (0002, 0003, 0005, 0009, 0010, 0021, and 0022), and it is located in Boone County. **Figure 4B** is a single census tract, specifically “Census Tract 0608”

with 150 intersecting image chips. This census tract is among the areas with one of the highest numbers of image chips. This level of granularity reveals a concentrated cluster of data points, potentially signifying a region of particular interest or higher measurement intensity.

The left map in **Figure 5** depicts the spatial distribution of feature 1112 across Missouri, with red circles highlighting areas of high values, which seems to correlate with (urban areas). The right map depicts actual obesity rates (%) across the state, with blue circles indicating regions with lower obesity prevalence. Notable discrepancies between feature values and obesity rates can be observed in several regions.

Figure 4. Multiscale analysis of satellite image chips and census tracts in Missouri.

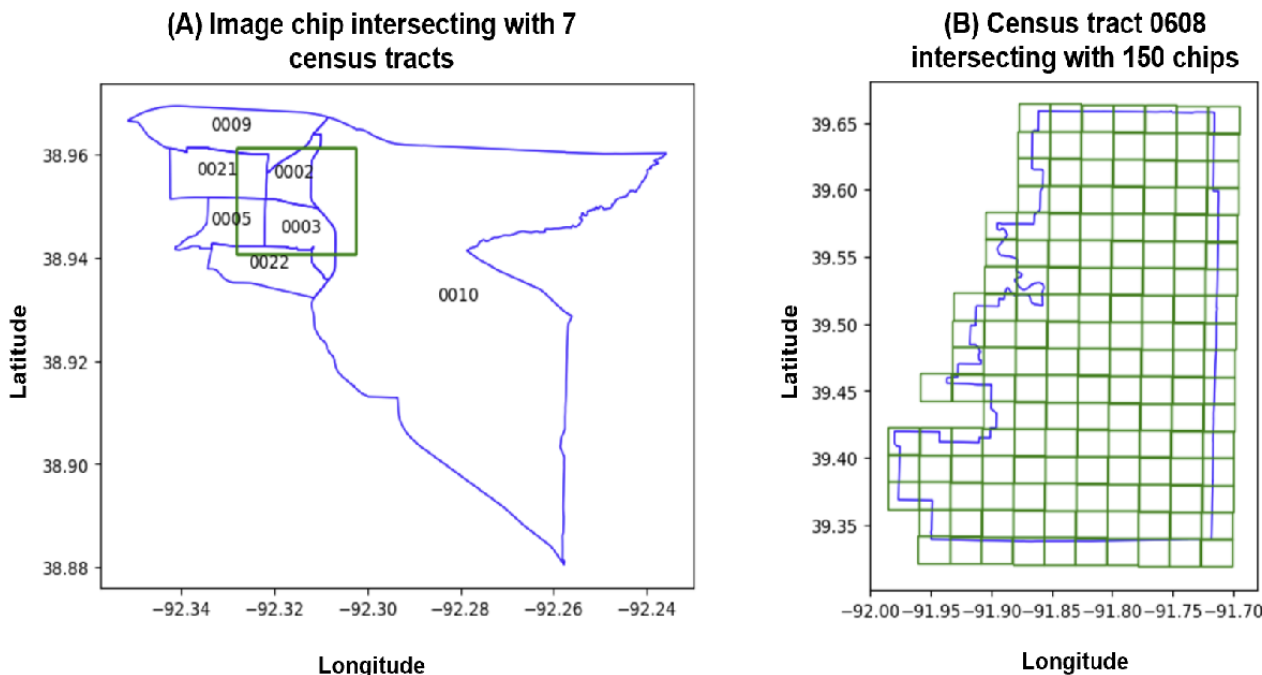


Figure 5. A map showing feature 1112 (left) values across census tracts in Missouri compared with the obesity rate (right).

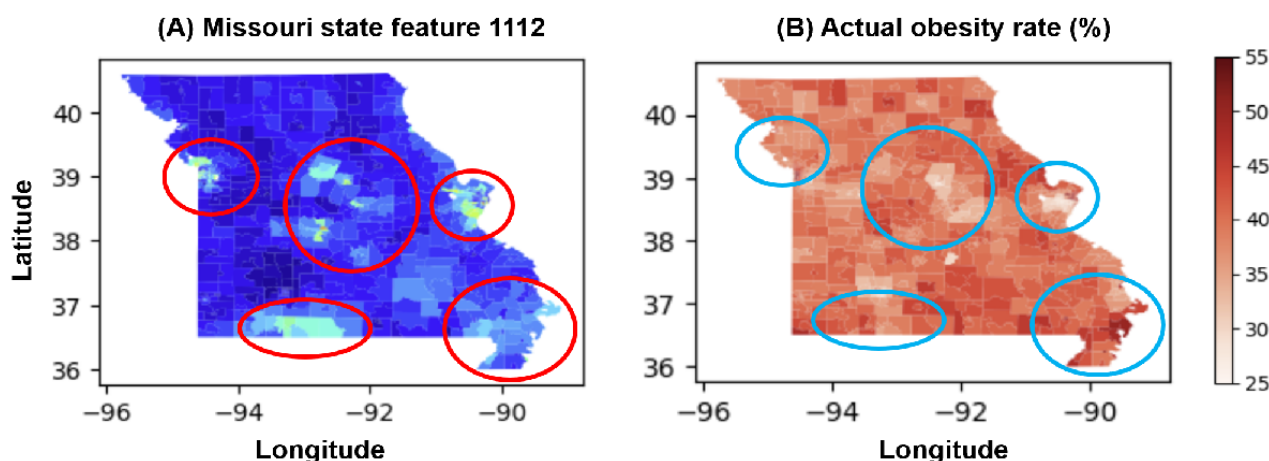
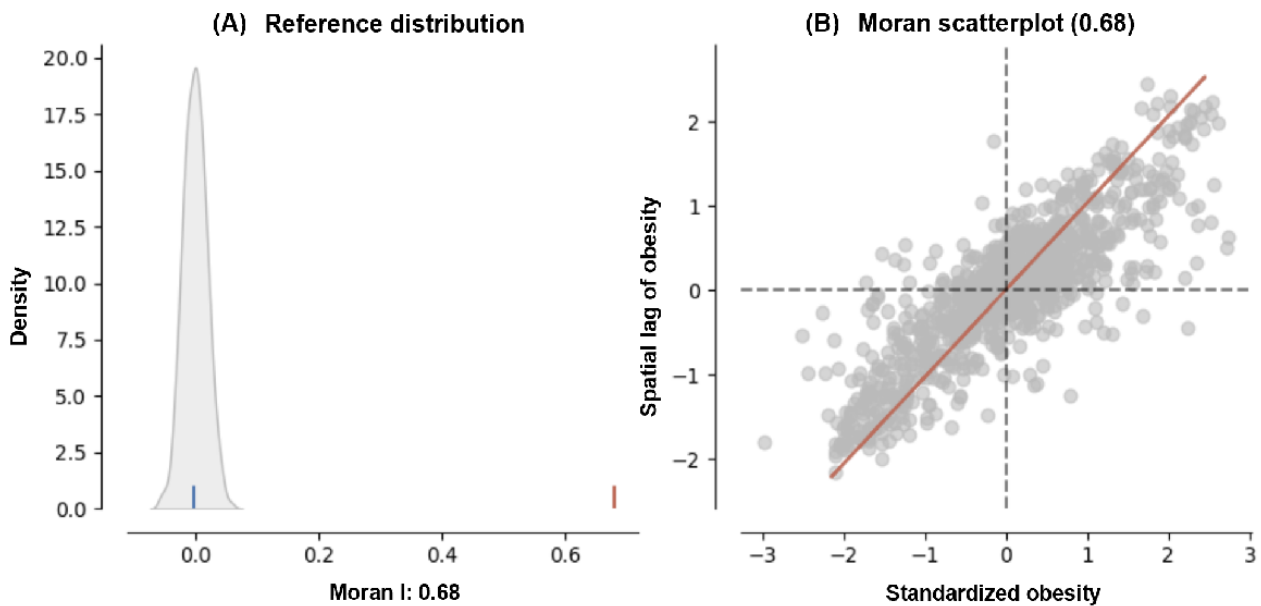


Figure 6A is a visualization of a probability distribution related to the Moran I statistic, a measure used in spatial analysis. The bell-shaped curve represents the expected distribution of Moran I under the null hypothesis of spatial randomness. The x-axis shows the range of Moran I values, and the y-axis represents the density or the probability of those values. There is a vertical line marked on the plot that indicates the observed Moran I value, which here is 0.68. This is significantly higher than 0, suggesting that there is a significant positive spatial autocorrelation in the dataset. The peak of the distribution curve is centered very close to 0, with the tails tapering off smoothly on either side. A blue line marks the mean of the reference distribution, indicating where Moran I would fall if the null hypothesis were true (no spatial autocorrelation). A red line marks the actual observed Moran I, which falls far to the right

of the mean, indicating a stronger-than-expected positive spatial correlation.

Figure 6B depicts a scatterplot with a regression line. This scatterplot is used to visualize the spatial autocorrelation of a variable, here, obesity rates. The x-axis represents standardized values of obesity rates, meaning the data have been normalized to have a mean of 0 and an SD of 1. The y-axis represents the standardized spatial lag of obesity, which is essentially the average obesity rate of neighboring areas. The scatterplot is dense with points that represent the different locations (such as census tracts) within Missouri. The points are predominantly clustered around the regression line (red line), suggesting that areas tend to be similar to their neighbors' high values are next to high values, and low values are next to low values.

Figure 6. (A) Reference distribution for Moran I, confirming positive spatial autocorrelation with an observed value of 0.68, significantly deviating from the null hypothesis of spatial randomness (blue line). (B) Moran scatterplot, demonstrating the relationship between standardized obesity rates and their standardized spatial lag, with a clear positive slope indicating that regions with high obesity rates are typically surrounded by similar regions. These findings are instrumental for spatially informed public health strategies, pinpointing areas for targeted interventions and enhancing the precision of predictive models for obesity prevalence across the state.



In addition, the regression line slopes upward, reinforcing the indication of positive spatial autocorrelation, as higher standardized values of obesity are associated with higher spatial lags.

Figure 7A represents a local version of Moran I scatterplot, which examines the spatial autocorrelation of obesity rates at a localized level. The x-axis displays the standardized obesity rates, indicating the deviation from the mean rate of obesity. The y-axis shows the spatial lag of obesity, which represents

the average rate of obesity in neighboring locations. Points are color coded; blue represents low-low clusters where locations and their neighbors have lower than average obesity rates; red represents high-high clusters where locations and their neighbors have higher than average rates; tan points represent low-high or high-low outliers where a location's obesity rate significantly differs from its neighbors. The solid black line indicates the trend, and the presence of color-coded points away from the origin suggests spatial clusters and outliers.

Figure 7. Multifaceted spatial analysis of obesity rates in Missouri using deep learning and satellite imagery. (A) Moran local scatterplot, indicating spatial autocorrelation of obesity rates with clusters of high (high-high; HH) and low (low-low; LL) values, alongside substantial spatial outliers (low-high; LH and high-low; HL). (B) Map of the local indicators of spatial associations (LISA) clusters across Missouri, identifying regions with statistically substantial spatial associations of obesity rates. (C) Actual obesity prevalence (illustrated with a choropleth map), correlating geographic data with obesity percentages. Combined, these visual tools leverage deep neural visual features extracted via a Residual Network-50 (ResNet-50) model applied to Sentinel-2 satellite imagery, offering a robust framework for understanding and predicting obesity distribution at the community level. NS: not significant.

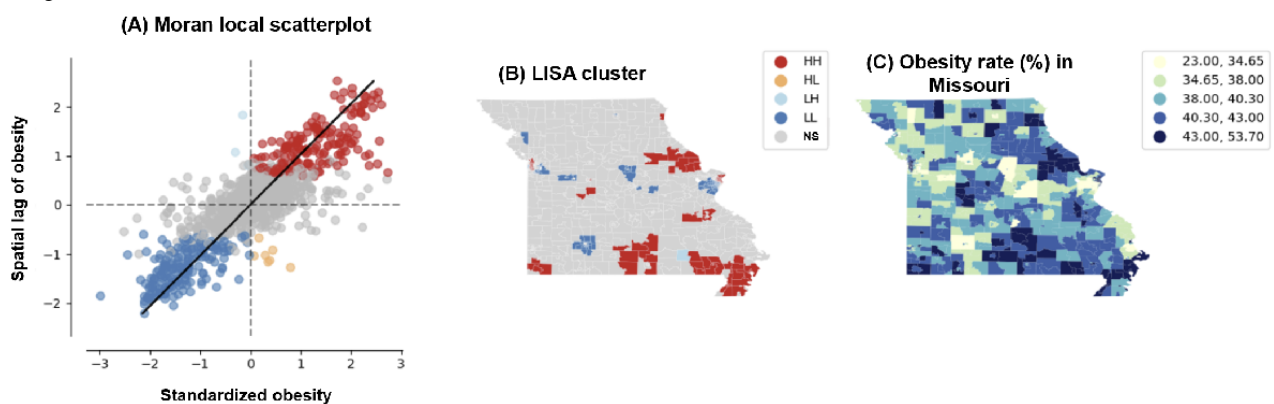


Figure 7B provides a spatial representation of LISA results, showing clusters and spatial outliers of obesity rates across Missouri. The map of Missouri is marked with areas color coded consistent with the legend in Figure 7A, with high-high clusters in red and low-low clusters in blue, among others. This

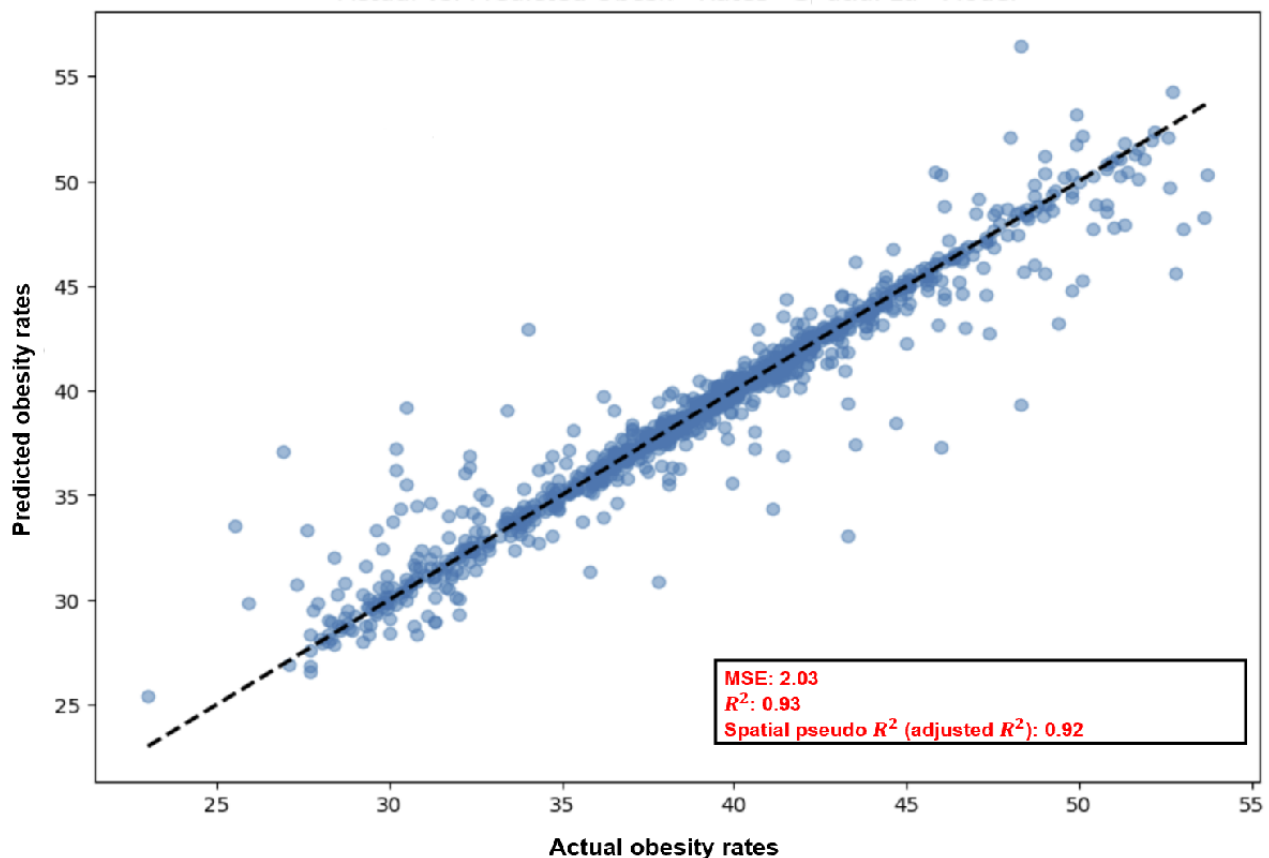
visualization allows for the geographical identification of areas with statistically significant local spatial autocorrelation, revealing regions where similar or dissimilar obesity rates cluster spatially.

Figure 7C depicts a choropleth map indicating the obesity rate by county or census tract across Missouri. The color gradient represents varying rates of obesity, with darker colors indicating higher rates. This visualizes the actual obesity prevalence geographically, which can be cross-referenced with the LISA results for a more nuanced understanding of spatial patterns.

As we delve deeper into the intricacies of spatial data, it is also essential to explore other spatial regression techniques, such as geographically weighted regression (GWR). GWR extends the capabilities of traditional spatial models by allowing relationships to vary across space, offering a nuanced understanding of local variations in the data. This progression from global to local models, from SLM to GWR, signifies a move toward more refined and location-specific analyses in spatial research.

Figure 8. A scatterplot comparing actual versus predicted obesity rates in Missouri, derived from a spatial lag model using Residual Network-50 (ResNet-50) deep neural features and Sentinel-2 imagery. Data points close to the dashed best-fit line, with an R^2 of 0.93 and spatial pseudo R^2 of 0.92, highlight the model's high accuracy in mapping the spatial distribution of obesity rates. MSE: mean squared error.

Actual versus predicted obesity rates: spatial lag model



The x-axis displays the actual obesity rates, expressed as percentages, observed within the dataset. The y-axis, similarly, shows the obesity rates as predicted by the SLM. Each point on the graph represents a specific location within Missouri, such as a county or census tract, where the coordinates correspond to the actual and predicted obesity rates. The scatter of points is densest along the line of best fit, which is depicted by a dashed line, indicating a strong correlation between predicted and actual values.

Beneath the scatterplot, 3 key performance indicators provide quantitative assessments of the model's accuracy. The mean

squared error, given as 2.03, is a measure of the average of the squares of the errors, the differences between predicted and actual rates. The lower the mean squared error, the better the model's predictions match the observed data. The R^2 value, noted as 0.93, represents the proportion of variance in the obesity rate that is predictable from the independent variables included in the model. An R^2 of 0.93 suggests that 93% of the variability in the actual obesity rates can be explained by the model, which is a very high level of explanatory power. The spatial pseudo R^2 , which is the adjusted R^2 for spatial regression

models, is 0.92, reinforcing the model's strong predictive capability while accounting for spatial dependencies.

Our study leveraged SLMs and DCNNs to predict obesity prevalence across 1052 census tracts in Missouri, achieving an R^2 value of 0.93 and a spatial pseudo R^2 of 0.92. These metrics indicate that our models explained 93% of the variability in obesity rates, demonstrating high predictive accuracy and robustness in modeling obesity prevalence using geospatial and DL methodologies.

In contrast, a similar study by Hales et al [25] used Visual Geometry Group-CNN-F models to analyze obesity across 1695 census tracts in 6 cities, achieving varied results across the different regions. Their overall R^2 for the combined cities was lower, explaining 64.8% of the variation in obesity prevalence with a root mean square error of 4.3. Notably, the highest individual city R^2 was 73.3% for Memphis, which is significantly lower than our study's performance. Their study also highlighted a weaker performance when using point of interest data alone, which explained only 42.4% of the variation in obesity prevalence, with a root mean square error of 4.3 across all census tracts.

Our study's superior performance is due to our use of a SLM to handle spatial dependencies, using ResNet-50 for advanced feature extraction from Sentinel-2 imagery, and focusing on Missouri for a more tailored geographic analysis.

Discussion

Principal Findings

The results in Figure 6 and the Moran I scatterplot can help spatially explain and predict obesity rates in Missouri using the DNVFs by identifying clusters of areas with either high or low rates of obesity. The positive Moran I value (0.68) suggests that similar values are located near each other. This means that if a certain census tract has a high obesity rate, it is likely that the neighboring tracts also have high obesity rates and vice versa for low rates. This pattern of spatial clustering is crucial for public health planning because it can help identify hot spots where interventions might be more necessary. For predictive modeling, spatial autocorrelation needs to be considered to improve the accuracy of predictions for obesity rates. Spatial models, such as SLMs or GWR, can use this autocorrelation to better understand and predict how obesity rates vary across Missouri. These models can incorporate not just the obesity rate of one area but also the context provided by surrounding areas, which can significantly influence health outcomes.

The subparts shown in Figure 6 help explain and predict obesity rates in Missouri. The Moran I scatterplot in Figure 7A identifies not only regions with high or low obesity rates but also those that deviate from the surrounding trend, which are crucial for targeted public health interventions. The LISA map in Figure 7B provides an immediate visual understanding of the geographic clustering of obesity rates, highlighting areas where policy interventions or further research might be needed. Finally, the choropleth map in Figure 7C allows researchers to observe

the actual prevalence of obesity and how it correlates with the clusters identified through Moran I and LISA analyses.

These results, when combined with DNVFs extracted from the ResNet-50 model using Sentinel-2 satellite imaging, can enhance predictions and explanations of obesity rates. The ResNet-50 model, pretrained on ImageNet, can detect and analyze environmental and physical features that correlate with obesity rates. By integrating these features with spatial statistics, researchers can develop more sophisticated models that account for both the physical characteristics of the environment as captured in satellite imagery and the spatial relationships of obesity rates across the state. This multimodal approach allows for a deeper understanding of the drivers behind obesity patterns, which can inform more targeted and effective public health strategies.

The high R^2 and adjusted R^2 values in Figure 7 indicate that the SLM, which includes deep neural features extracted from the ResNet-50 model using Sentinel-2 satellite imaging, is highly effective in both explaining and predicting obesity rates. By incorporating the spatial lag of obesity rates, essentially considering not just the individual obesity rate of each area but also the influence of adjacent areas, the model captures the spatial autocorrelation inherent in the data. This consideration is crucial because the prevalence of obesity can be influenced by both location-specific factors and the characteristics of neighboring regions.

The use of DNVFs from ResNet-50 means that the model is leveraging complex, high-level features extracted from satellite imagery that may correlate with environmental factors affecting obesity rates, such as the availability of green spaces or the walkability of neighborhoods. The model's success, as indicated by the scatterplot, demonstrates the potential for these advanced ML techniques to capture and use subtle spatial patterns and characteristics that contribute to public health outcomes.

Overall, the results depicted in this figure underscore the model's potential as a tool for public health officials and policy makers. The ability to accurately predict obesity rates at a granular spatial level can facilitate targeted interventions and resource allocation, contributing to more effective public health strategies and better health outcomes for communities across Missouri.

Limitations

Despite the promising results, our obesity research study has some limitations. First, the estimates of obesity prevalence from the Behavioral Risk Factor Surveillance System rely on self-reported measurements of height and weight, which are subject to bias and often result in an underestimation of the true rate of obesity [20,22]. Moreover, BMI fails to directly measure body fat, which can differ based on gender, age, race, and ethnicity. In addition, the risks of mortality and morbidity at a given BMI may not be the same across various racial and ethnic groups [23,24]. Variations in the timing between when the obesity data and the satellite images are collected can also introduce biases into our analysis.

One of the primary limitations of our study pertains to the dataset's size and geographical coverage. The research was confined to 1052 census tracts within the entire state of Missouri, limiting the generalizability of the findings. Although these tracts were selected to represent a diverse range of urban and rural areas, they do not encompass other neighboring states or the entire country's varied demographic and geographic profiles. Furthermore, the limited number of census tracts might not provide a sufficiently robust dataset for more complex ML models. The use of Sentinel-2 imagery, while innovative, was also constrained by the number and resolution of available images. This limitation potentially affects the accuracy and granularity of the features extracted for obesity rate prediction, particularly in areas where medium-resolution satellite imagery was not available or was of lower quality.

Another significant limitation arises from the DCNN and ResNet-50 features used in our study. While these methods are state of the art in image analysis and feature extraction, their effectiveness is inherently tied to the quantity and quality of the input data. The pretrained ResNet-50 is not fully optimized for the specific nuances of satellite image analysis related to obesity rate prediction. Therefore, this study's findings must be interpreted with caution, acknowledging that the models used, although advanced, might not capture the complete range of factors influencing obesity rates as discernible from satellite imagery.

Conclusions

Our study advances the field of spatial regression modeling by integrating DNVFs with traditional models to analyze the geographical distribution of obesity rates across Missouri. This

innovative approach, evidenced by high R^2 values, not only underscores the efficacy of incorporating ML in public health analytics but also provides a methodological framework for similar studies.

While our findings contribute significantly to understanding the spatial dynamics of obesity, they are tempered by limitations, such as the reliance on self-reported BMI data, which may underestimate true obesity rates due to reporting biases. Furthermore, the study's focus on Missouri restricts its broader applicability, suggesting the need for expanded geographic research that includes more diverse populations and environmental settings.

Future research should aim to incorporate longitudinal and multiregional analyses, integrating additional variables, such as socioeconomic status, access to health care, and urbanization levels. These expansions will enable a more comprehensive assessment of the factors influencing obesity and facilitate the development of targeted, effective public health interventions. In addition, we plan to expand our use of RS data by incorporating high-resolution imagery and time-series data in future studies. This approach will enable a more detailed analysis of the dynamic changes in the built environment and their impact on obesity rates, enhancing the precision and applicability of our findings for public health strategies.

By addressing these limitations and exploring these future directions, subsequent research can build upon our findings to enhance the predictive accuracy of obesity prevalence models and ultimately support more nuanced and effective public health strategies.

Acknowledgments

Computational resources for this research have been supported by the National Science Foundation National Research Platform, as part of GP-ENGINE (award OAC 2322218).

Data Availability

The datasets generated and analyzed during this study are not publicly available but are available from the corresponding author on reasonable request.

Authors' Contributions

BMD contributed to the conception and design of the manuscript; acquisition, analysis, and interpretation of the data; and drafted and critically revised the manuscript. SK contributed to the design; acquisition, analysis, and interpretation of the data; and drafted and critically revised the manuscript. IET contributed to the analysis and interpretation of the data and drafted and critically revised the manuscript. MA contributed to the design, data acquisition, and data analysis and drafted and critically revised the manuscript. CIM-V contributed to the conception and design of the manuscript, acquisition and interpretation of the data, and critically revised the manuscript. OBO contributed to the conception and design of the manuscript, interpretation of the data, and critically revised the manuscript. LRS contributed to the conception and design of the manuscript, acquisition and interpretation of the data, and critically revised the manuscript. GJS contributed to the conception and design of the manuscript, acquisition and interpretation of the data, facilitated the computing and provided funding, and critically revised the manuscript. All authors provided final approval for the final manuscript and agreed to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

References

1. Jia P, Xue H, Yin L, Stein A, Wang M, Wang Y. Spatial technologies in obesity research: current applications and future promise. *Trends Endocrinol Metab.* Mar 2019;30(3):211-223. [doi: [10.1016/j.tem.2018.12.003](https://doi.org/10.1016/j.tem.2018.12.003)] [Medline: [30712979](https://pubmed.ncbi.nlm.nih.gov/30712979/)]
2. Ortega Hinojosa AM, Davies MM, Jarjour S, Burnett RT, Mann JK, Hughes E, et al. Developing small-area predictions for smoking and obesity prevalence in the United States for use in environmental public health tracking. *Environ Res.* Oct 2014;134:435-452. [doi: [10.1016/j.envres.2014.07.029](https://doi.org/10.1016/j.envres.2014.07.029)] [Medline: [25261951](https://pubmed.ncbi.nlm.nih.gov/25261951/)]
3. Ng M, Fleming T, Robinson M, Thomson B, Graetz N, Margono C, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet.* Aug 30, 2014;384(9945):766-781. [FREE Full text] [doi: [10.1016/S0140-6736\(14\)60460-8](https://doi.org/10.1016/S0140-6736(14)60460-8)] [Medline: [24880830](https://pubmed.ncbi.nlm.nih.gov/24880830/)]
4. Jia P, Cheng X, Xue H, Wang Y. Applications of geographic information systems (GIS) data and methods in obesity-related research. *Obes Rev.* Apr 2017;18(4):400-411. [doi: [10.1111/obr.12495](https://doi.org/10.1111/obr.12495)] [Medline: [28165656](https://pubmed.ncbi.nlm.nih.gov/28165656/)]
5. Anekwe CV, Jarrell AR, Townsend MJ, Gaudier GI, Hiserodt JM, Stanford FC. Socioeconomics of obesity. *Curr Obes Rep.* Sep 2020;9(3):272-279. [FREE Full text] [doi: [10.1007/s13679-020-00398-7](https://doi.org/10.1007/s13679-020-00398-7)] [Medline: [32627133](https://pubmed.ncbi.nlm.nih.gov/32627133/)]
6. Hruby A, Hu FB. The epidemiology of obesity: a big picture. *Pharmacoeconomics.* Jul 2015;33(7):673-689. [FREE Full text] [doi: [10.1007/s40273-014-0243-x](https://doi.org/10.1007/s40273-014-0243-x)] [Medline: [25471927](https://pubmed.ncbi.nlm.nih.gov/25471927/)]
7. Kundu S, Sharma P, Singh S, Kumar P. District-level heterogeneity in overweight or obesity among women of reproductive age: a geo-spatial analysis in India. *PLoS One.* Aug 17, 2023;18(8):e0290020. [FREE Full text] [doi: [10.1371/journal.pone.0290020](https://doi.org/10.1371/journal.pone.0290020)] [Medline: [37590188](https://pubmed.ncbi.nlm.nih.gov/37590188/)]
8. Giles-Corti B, Macintyre S, Clarkson JP, Pikora T, Donovan RJ. Environmental and lifestyle factors associated with overweight and obesity in Perth, Australia. *Am J Health Promot.* 2003;18(1):93-102. [doi: [10.4278/0890-1171-18.1.93](https://doi.org/10.4278/0890-1171-18.1.93)] [Medline: [13677967](https://pubmed.ncbi.nlm.nih.gov/13677967/)]
9. Saelens BE, Sallis JF, Black JB, Chen D. Neighborhood-based differences in physical activity: an environment scale evaluation. *Am J Public Health.* Sep 2003;93(9):1552-1558. [doi: [10.2105/ajph.93.9.1552](https://doi.org/10.2105/ajph.93.9.1552)] [Medline: [12948979](https://pubmed.ncbi.nlm.nih.gov/12948979/)]
10. Burdette HL, Whitaker RC. Neighborhood playgrounds, fast food restaurants, and crime: relationships to overweight in low-income preschool children. *Prev Med.* Jan 2004;38(1):57-63. [doi: [10.1016/j.ypmed.2003.09.029](https://doi.org/10.1016/j.ypmed.2003.09.029)] [Medline: [14672642](https://pubmed.ncbi.nlm.nih.gov/14672642/)]
11. Black JL, Macinko J. Neighborhoods and obesity. *Nutr Rev.* Jan 2008;66(1):2-20. [doi: [10.1111/j.1753-4887.2007.00001.x](https://doi.org/10.1111/j.1753-4887.2007.00001.x)] [Medline: [18254880](https://pubmed.ncbi.nlm.nih.gov/18254880/)]
12. Inagami S, Cohen DA, Finch BK, Asch SM. You are where you shop: grocery store locations, weight, and neighborhoods. *Am J Prev Med.* Jul 2006;31(1):10-17. [doi: [10.1016/j.amepre.2006.03.019](https://doi.org/10.1016/j.amepre.2006.03.019)] [Medline: [16777537](https://pubmed.ncbi.nlm.nih.gov/16777537/)]
13. Rundle A, Diez Roux AV, Free LM, Miller D, Neckerman KM, Weiss CC. The urban built environment and obesity in New York City: a multilevel analysis. *Am J Health Promot.* 2007;21(4 Suppl):326-334. [doi: [10.4278/0890-1171-21.4s.326](https://doi.org/10.4278/0890-1171-21.4s.326)] [Medline: [17465178](https://pubmed.ncbi.nlm.nih.gov/17465178/)]
14. Rutt CD, Coleman KJ. Examining the relationships among built environment, physical activity, and body mass index in El Paso, TX. *Prev Med.* Jun 2005;40(6):831-841. [doi: [10.1016/j.ypmed.2004.09.035](https://doi.org/10.1016/j.ypmed.2004.09.035)] [Medline: [15850885](https://pubmed.ncbi.nlm.nih.gov/15850885/)]
15. Li F, Harmer P, Cardinal BJ, Bosworth M, Johnson-Shelton D, Moore JM, et al. Built environment and 1-year change in weight and waist circumference in middle-aged and older adults: Portland neighborhood environment and health study. *Am J Epidemiol.* Feb 15, 2009;169(4):401-408. [FREE Full text] [doi: [10.1093/aje/kwn398](https://doi.org/10.1093/aje/kwn398)] [Medline: [19153214](https://pubmed.ncbi.nlm.nih.gov/19153214/)]
16. Rundle A, Neckerman KM, Freeman L, Lovasi GS, Purciel M, Quinn J, et al. Neighborhood food environment and walkability predict obesity in New York City. *Environ Health Perspect.* Mar 2009;117(3):442-447. [FREE Full text] [doi: [10.1289/ehp.11590](https://doi.org/10.1289/ehp.11590)] [Medline: [19337520](https://pubmed.ncbi.nlm.nih.gov/19337520/)]
17. Li F, Harmer PA, Cardinal BJ, Bosworth M, Acock A, Johnson-Shelton D, et al. Built environment, adiposity, and physical activity in adults aged 50-75. *Am J Prev Med.* Jul 2008;35(1):38-46. [FREE Full text] [doi: [10.1016/j.amepre.2008.03.021](https://doi.org/10.1016/j.amepre.2008.03.021)] [Medline: [18541175](https://pubmed.ncbi.nlm.nih.gov/18541175/)]
18. Jeffery RW, Baxter J, McGuire M, Linde J. Are fast food restaurants an environmental risk factor for obesity? *Int J Behav Nutr Phys Act.* Jan 25, 2006;3:2. [FREE Full text] [doi: [10.1186/1479-5868-3-2](https://doi.org/10.1186/1479-5868-3-2)] [Medline: [16436207](https://pubmed.ncbi.nlm.nih.gov/16436207/)]
19. Lovasi GS, Hutson MA, Guerra M, Neckerman KM. Built environments and obesity in disadvantaged populations. *Epidemiol Rev.* 2009;31(1):7-20. [doi: [10.1093/epirev/mxp005](https://doi.org/10.1093/epirev/mxp005)] [Medline: [19589839](https://pubmed.ncbi.nlm.nih.gov/19589839/)]
20. Kuczmarski MF, Kuczmarski RJ, Najjar M. Effects of age on validity of self-reported height, weight, and body mass index: findings from the Third National Health and Nutrition Examination Survey, 1988-1994. *J Am Diet Assoc.* Jan 2001;101(1):28-34; quiz 35-6. [doi: [10.1016/S0002-8223\(01\)00008-6](https://doi.org/10.1016/S0002-8223(01)00008-6)] [Medline: [11209581](https://pubmed.ncbi.nlm.nih.gov/11209581/)]
21. Papas MA, Alberg AJ, Ewing R, Helzlsouer KJ, Gary TL, Klassen AC. The built environment and obesity. *Epidemiol Rev.* 2007;29:129-143. [doi: [10.1093/epirev/mxm009](https://doi.org/10.1093/epirev/mxm009)] [Medline: [17533172](https://pubmed.ncbi.nlm.nih.gov/17533172/)]
22. Merrill RM, Richardson JS. Validity of self-reported height, weight, and body mass index: findings from the National Health and Nutrition Examination Survey, 2001-2006. *Prev Chronic Dis.* Oct 2009;6(4):A121. [FREE Full text] [Medline: [19754997](https://pubmed.ncbi.nlm.nih.gov/19754997/)]
23. Zheng W, McLerran DF, Rolland B, Zhang X, Inoue M, Matsuo K, et al. Association between body-mass index and risk of death in more than 1 million Asians. *N Engl J Med.* Feb 24, 2011;364(8):719-729. [FREE Full text] [doi: [10.1056/NEJMoa1010679](https://doi.org/10.1056/NEJMoa1010679)] [Medline: [21345101](https://pubmed.ncbi.nlm.nih.gov/21345101/)]

24. Jafar TH, Islam M, Poulter N, Hatcher J, Schmid CH, Levey AS, et al. Children in South Asia have higher body mass-adjusted blood pressure levels than white children in the United States: a comparative study. *Circulation*. Mar 15, 2005;111(10):1291-1297. [doi: [10.1161/01.CIR.0000157699.87728.F1](https://doi.org/10.1161/01.CIR.0000157699.87728.F1)] [Medline: [15769771](https://pubmed.ncbi.nlm.nih.gov/15769771/)]
25. Hales CM, Carroll MD, Fryar CD, Ogden CL. Prevalence of obesity among adults and youth: United States, 2015-2016. *NCHS Data Brief*. Oct 2017;(288):1-8. [FREE Full text] [Medline: [29155689](https://pubmed.ncbi.nlm.nih.gov/29155689/)]
26. Casagrande SS, Whitt-Glover MC, Lancaster KJ, Odoms-Young AM, Gary TL. Built environment and health behaviors among African Americans: a systematic review. *Am J Prev Med*. Feb 2009;36(2):174-181. [doi: [10.1016/j.amepre.2008.09.037](https://doi.org/10.1016/j.amepre.2008.09.037)] [Medline: [19135908](https://pubmed.ncbi.nlm.nih.gov/19135908/)]
27. Maharana A, Nsoesie EO. Use of deep learning to examine the association of the built environment with prevalence of neighborhood adult obesity. *JAMA Netw Open*. Aug 03, 2018;1(4):e181535. [FREE Full text] [doi: [10.1001/jamanetworkopen.2018.1535](https://doi.org/10.1001/jamanetworkopen.2018.1535)] [Medline: [30646134](https://pubmed.ncbi.nlm.nih.gov/30646134/)]
28. Nguyen QC, Sajjadi M, McCullough M, Pham M, Nguyen TT, Yu W, et al. Neighbourhood looking glass: 360° automated characterisation of the built environment for neighbourhood effects research. *J Epidemiol Community Health*. Mar 2018;72(3):260-266. [FREE Full text] [doi: [10.1136/jech-2017-209456](https://doi.org/10.1136/jech-2017-209456)] [Medline: [29335255](https://pubmed.ncbi.nlm.nih.gov/29335255/)]
29. Dahu BM, Alaboud K, Nowbuth AA, Puckett HM, Scott GJ, Sheets LR. The role of remote sensing and geospatial analysis for understanding COVID-19 population severity: a systematic review. *Int J Environ Res Public Health*. Feb 28, 2023;20(5):4298. [FREE Full text] [doi: [10.3390/ijerph20054298](https://doi.org/10.3390/ijerph20054298)] [Medline: [36901308](https://pubmed.ncbi.nlm.nih.gov/36901308/)]
30. Bethlehem JR, Mackenbach JD, Ben-Rebah M, Compennolle S, Glonti K, Bárdos H, et al. The SPOTLIGHT virtual audit tool: a valid and reliable tool to assess obesogenic characteristics of the built environment. *Int J Health Geogr*. Dec 16, 2014;13:52. [FREE Full text] [doi: [10.1186/1476-072X-13-52](https://doi.org/10.1186/1476-072X-13-52)] [Medline: [25515179](https://pubmed.ncbi.nlm.nih.gov/25515179/)]
31. Yin L, Wang Z. Measuring visual enclosure for street walkability: using machine learning algorithms and Google Street View imagery. *Appl Geogr*. Nov 2016;76:147-153. [doi: [10.1016/j.apgeog.2016.09.024](https://doi.org/10.1016/j.apgeog.2016.09.024)]
32. Yin L, Cheng Q, Wang Z, Shao Z. 'Big data' for pedestrian volume: exploring the use of Google Street View images for pedestrian counts. *Appl Geogr*. Sep 2015;63:337-345. [doi: [10.1016/j.apgeog.2015.07.010](https://doi.org/10.1016/j.apgeog.2015.07.010)]
33. Lin G, Spann S, Hyman D, Pavlik V. Climate amenity and BMI. *Obesity (Silver Spring)*. Aug 2007;15(8):2120-2127. [FREE Full text] [doi: [10.1038/oby.2007.252](https://doi.org/10.1038/oby.2007.252)] [Medline: [17712131](https://pubmed.ncbi.nlm.nih.gov/17712131/)]
34. Alaboud K, Toubal IE, Dahu BM, Daken AA, Salman AA, Alaji N, et al. The quality application of deep learning in clinical outcome predictions using electronic health record data: a systematic review. *South East Eur J Public Health*. 2023;XXI:9-23. [doi: [10.70135/seejph.vi.427](https://doi.org/10.70135/seejph.vi.427)]
35. Dahu BM, Khan S, Salman AA, Andraws YM, Daken AA, Aburayya A. Epidemiological analysis of vaccination strategies and demographic patterns in COVID-19 cases in the Midwest region of the United States. *Nat J Community Med*. Jan 01, 2024;15(01):62-71. [doi: [10.55489/njcm.150120243461](https://doi.org/10.55489/njcm.150120243461)]
36. Michimi A, Wimberly MC. Natural environments, obesity, and physical activity in nonmetropolitan areas of the United States. *J Rural Health*. 2012;28(4):398-407. [doi: [10.1111/j.1748-0361.2012.00413.x](https://doi.org/10.1111/j.1748-0361.2012.00413.x)] [Medline: [23083086](https://pubmed.ncbi.nlm.nih.gov/23083086/)]
37. Alshehri M, Ouadou A, Scott GJ. Deep transformer-based network deforestation detection in the Brazilian Amazon using sentinel-2 imagery. *IEEE Geosci Remote Sens Lett*. Jan 17, 2024;21:1-5. [doi: [10.1109/lgrs.2024.3355104](https://doi.org/10.1109/lgrs.2024.3355104)]
38. Khan S, Dahu BM, Scott GJ. A spatio-temporal study of changes in air quality from pre-COVID era to post-COVID era in Chicago, USA. *Aerosol Air Qual Res*. 2022;22(8):220053. [doi: [10.4209/aaqr.220053](https://doi.org/10.4209/aaqr.220053)]
39. Mariam A, Ouadou A, Scott G. Deforestation detection in the Brazilian Amazon using transformer-based networks. In: *Proceedings of the IEEE Conference on Artificial Intelligence*. 2023. Presented at: CAI 2023; June 5-6, 2023; Santa Clara, CA. [doi: [10.1109/cai54212.2023.00130](https://doi.org/10.1109/cai54212.2023.00130)]
40. Pala D, Calderone AA, Franzini M, Malovini A, Larizza C, Casella V, et al. Deep learning to unveil correlations between urban landscape and population health. *Sensors (Basel)*. Apr 08, 2020;20(7):2105. [FREE Full text] [doi: [10.3390/s20072105](https://doi.org/10.3390/s20072105)] [Medline: [32276488](https://pubmed.ncbi.nlm.nih.gov/32276488/)]
41. Levy JJ, Lebeaux RM, Hoen AG, Christensen BC, Vaickus LJ, MacKenzie TA. Using satellite images and deep learning to identify associations between county-level mortality and residential neighborhood features proximal to schools: a cross-sectional study. *Front Public Health*. Nov 5, 2021;9:766707. [FREE Full text] [doi: [10.3389/fpubh.2021.766707](https://doi.org/10.3389/fpubh.2021.766707)] [Medline: [34805078](https://pubmed.ncbi.nlm.nih.gov/34805078/)]
42. Guo C, Wang H, Feng G, Li J, Su C, Zhang J, et al. Spatiotemporal predictions of obesity prevalence in Chinese children and adolescents: based on analyses of obesogenic environmental variability and Bayesian model. *Int J Obes (Lond)*. Jul 2019;43(7):1380-1390. [FREE Full text] [doi: [10.1038/s41366-018-0301-0](https://doi.org/10.1038/s41366-018-0301-0)] [Medline: [30568273](https://pubmed.ncbi.nlm.nih.gov/30568273/)]
43. Han X, Wang L, Seo SH, He J, Jung T. Measuring perceived psychological stress in urban built environments using Google street view and deep learning. *Front Public Health*. May 11, 2022;10:891736. [FREE Full text] [doi: [10.3389/fpubh.2022.891736](https://doi.org/10.3389/fpubh.2022.891736)] [Medline: [35646775](https://pubmed.ncbi.nlm.nih.gov/35646775/)]
44. Hong X, Sheridan S, Li D. Mapping built environments from UAV imagery: a tutorial on mixed methods of deep learning and GIS. *Comput Urban Sci*. Jun 02, 2022;2(1):12. [doi: [10.1007/s43762-022-00039-w](https://doi.org/10.1007/s43762-022-00039-w)]
45. Wang R, Liu Y, Lu Y, Zhang J, Liu P, Yao Y, et al. Perceptions of built environment and health outcomes for older Chinese in Beijing: a big data approach with street view images and deep learning technique. *Comput Environ Urban Syst*. Nov 2019;78:101386. [doi: [10.1016/j.compenvurbsys.2019.101386](https://doi.org/10.1016/j.compenvurbsys.2019.101386)]

46. Larkin A, Gu X, Chen L, Hystad P. Predicting perceptions of the built environment using GIS, satellite and street view image approaches. *Landsc Urban Plan.* Dec 2021;216:104257. [FREE Full text] [doi: [10.1016/j.landurbplan.2021.104257](https://doi.org/10.1016/j.landurbplan.2021.104257)] [Medline: [34629575](https://pubmed.ncbi.nlm.nih.gov/34629575/)]
47. Newton DW. Deep learning in urban analysis for health. In: *Artificial Intelligence in Urban Planning and Design: Technologies, Implementation, and Impacts.* Amsterdam, The Netherlands. Elsevier Science; 2022.
48. Siddiqui H, Rattani A, Woods NK, Cure L, Lewis RK, Twomey J, et al. A survey on machine and deep learning models for childhood and adolescent obesity. *IEEE Access.* Nov 25, 2021;9:157337-157360. [doi: [10.1109/ACCESS.2021.3131128](https://doi.org/10.1109/ACCESS.2021.3131128)]
49. Wirtz Baker JM, Pou SA, Niclis C, Haluszka E, Aballay LR. Non-traditional data sources in obesity research: a systematic review of their use in the study of obesogenic environments. *Int J Obes (Lond).* Aug 2023;47(8):686-696. [doi: [10.1038/s41366-023-01331-3](https://doi.org/10.1038/s41366-023-01331-3)] [Medline: [37393408](https://pubmed.ncbi.nlm.nih.gov/37393408/)]
50. Zhou X, Chen L, Liu HX. Applications of machine learning models to predict and prevent obesity: a mini-review. *Front Nutr.* Jul 5, 2022;9:933130. [FREE Full text] [doi: [10.3389/fnut.2022.933130](https://doi.org/10.3389/fnut.2022.933130)] [Medline: [35866076](https://pubmed.ncbi.nlm.nih.gov/35866076/)]
51. Shaamala A, Yigitcanlar T, Nili A, Nyandega D. State-of-the-art machine learning models for geospatial analysis: a systematic review of urban and environmental studies. SSRN. Preprint posted online on February 16, 2024. [FREE Full text]
52. Mahanty C, Kumar R, Asteris PG, Gandomi AH. COVID-19 patient detection based on fusion of transfer learning and fuzzy ensemble models using CXR images. *Appl Sci.* Dec 02, 2021;11(23):11423. [doi: [10.3390/app112311423](https://doi.org/10.3390/app112311423)]
53. Asteris PG, Kokoris S, Gavriilaki E, Tsoukalas MZ, Houpas P, Paneta M, et al. Early prediction of COVID-19 outcome using artificial intelligence techniques and only five laboratory indices. *Clin Immunol.* Jan 2023;246:109218. [FREE Full text] [doi: [10.1016/j.clim.2022.109218](https://doi.org/10.1016/j.clim.2022.109218)] [Medline: [36586431](https://pubmed.ncbi.nlm.nih.gov/36586431/)]
54. Asteris PG, Gandomi AH, Armaghani DJ, Kokoris S, Papandreadi AT, Roumelioti A, et al. Prognosis of COVID-19 severity using DERGA, a novel machine learning algorithm. *Eur J Intern Med.* Jul 2024;125:67-73. [doi: [10.1016/j.ejim.2024.02.037](https://doi.org/10.1016/j.ejim.2024.02.037)] [Medline: [38458880](https://pubmed.ncbi.nlm.nih.gov/38458880/)]
55. Janssen V. Understanding coordinate systems, datums and transformations in Australia. In: *Proceedings of Surveying and Spatial Sciences Institute Biennial International Conference.* 2009. Presented at: SSSI 2009; September 28-October 2, 2009; Adelaide, Australia.
56. Kopejkin SM. Celestial coordinate reference systems in curved space-time. *Celest Mech.* Mar 1988;44:87-115. [doi: [10.1007/bf01230709](https://doi.org/10.1007/bf01230709)]
57. Seeger H. Geographical information systems: spatial referencing and coordinate systems. University of Edinburgh. URL: https://www.geos.ed.ac.uk/~gisteac/gis_book_abridged/files/ch30.pdf [accessed 2024-12-09]
58. TIGER/Line shapefiles. United States Census Bureau. URL: <https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html> [accessed 2024-06-02]
59. Adult obesity facts. Centers for Disease Control and Prevention. URL: <https://www.cdc.gov/obesity/php/data-research/adult-obesity-facts.html> [accessed 2024-06-02]
60. PLACES: local data for better health. Centers for Disease Control and Prevention. URL: <https://www.cdc.gov/places/index.html> [accessed 2024-06-02]
61. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* Feb 02, 2017;542(7639):115-118. [FREE Full text] [doi: [10.1038/nature21056](https://doi.org/10.1038/nature21056)] [Medline: [28117445](https://pubmed.ncbi.nlm.nih.gov/28117445/)]
62. Jean N, Burke M, Xie M, Davis WM, Lobell DB, Ermon S. Combining satellite imagery and machine learning to predict poverty. *Science.* Aug 19, 2016;353(6301):790-794. [doi: [10.1126/science.aaf7894](https://doi.org/10.1126/science.aaf7894)] [Medline: [27540167](https://pubmed.ncbi.nlm.nih.gov/27540167/)]
63. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet large scale visual recognition challenge. *Int J Comput Vis.* Apr 11, 2015;115(3):211-252. [FREE Full text] [doi: [10.1007/s11263-015-0816-y](https://doi.org/10.1007/s11263-015-0816-y)]
64. Mascarenhas S, Agarwal M. A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for image classification. In: *Proceedings of the International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications.* 2021. Presented at: CENTCON 2021; November 19-21, 2021; Bengaluru, India. [doi: [10.1109/centcon52345.2021.9687944](https://doi.org/10.1109/centcon52345.2021.9687944)]
65. Rathi P, Kuwar Gupta R, Agarwal S, Shukla A. Sign language recognition using ResNet50 deep neural network architecture. *SSRN Electron J.* Jan 2020. [doi: [10.2139/ssrn.3545064](https://doi.org/10.2139/ssrn.3545064)]
66. Legendre P. Spatial autocorrelation: trouble or new paradigm? *Ecology.* Sep 1993;74(6):1659-1673. [doi: [10.2307/1939924](https://doi.org/10.2307/1939924)]
67. Getis A. Spatial autocorrelation. In: Fischer M, Getis A, editors. *Handbook of Applied Spatial Analysis.* Berlin, Germany. Springer; 2010.
68. Dubin RA. Spatial autocorrelation: a primer. *J Hous Econ.* Dec 1998;7(4):304-327. [doi: [10.1006/jhec.1998.0236](https://doi.org/10.1006/jhec.1998.0236)]
69. Cliff AD, Ord K. Spatial autocorrelation: a review of existing and new measures with applications. *Econ Geogr.* Jun 1970;46:269-292. [doi: [10.2307/143144](https://doi.org/10.2307/143144)]
70. Poulidou T, Elliott SJ. An exploratory spatial analysis of overweight and obesity in Canada. *Prev Med.* Apr 2009;48(4):362-367. [doi: [10.1016/j.ypmed.2009.01.017](https://doi.org/10.1016/j.ypmed.2009.01.017)] [Medline: [19463485](https://pubmed.ncbi.nlm.nih.gov/19463485/)]

71. Schuurman N, Peters PA, Oliver LN. Are obesity and physical activity clustered? A spatial analysis linked to residential density. *Obesity (Silver Spring)*. Dec 06, 2009;17(12):2202-2209. [FREE Full text] [doi: [10.1038/oby.2009.119](https://doi.org/10.1038/oby.2009.119)] [Medline: [19390521](https://pubmed.ncbi.nlm.nih.gov/19390521/)]
72. Penney TL, Rainham DG, Dummer TJ, Kirk SF. A spatial analysis of community level overweight and obesity. *J Hum Nutr Diet*. Apr 24, 2014;27 Suppl 2(s2):65-74. [doi: [10.1111/jhn.12055](https://doi.org/10.1111/jhn.12055)] [Medline: [23701323](https://pubmed.ncbi.nlm.nih.gov/23701323/)]
73. Anselin L. Local Indicators of Spatial Association—LISA. *Geogr Anal*. Sep 03, 2010;27(2):93-115. [doi: [10.1111/j.1538-4632.1995.tb00338.x](https://doi.org/10.1111/j.1538-4632.1995.tb00338.x)]
74. Parra-Amaya ME, Puerta-Yepes ME, Lizarralde-Bejarano DP, Arboleda-Sánchez S. Early detection for dengue using local indicator of spatial association (LISA) analysis. *Diseases*. Mar 29, 2016;4(2):16. [FREE Full text] [doi: [10.3390/diseases4020016](https://doi.org/10.3390/diseases4020016)] [Medline: [28933396](https://pubmed.ncbi.nlm.nih.gov/28933396/)]
75. Jesri N, Saghafipour A, Koohpaei A, Farzinnia B, Jooshin MK, Abolkheirian S, et al. Mapping and spatial pattern analysis of COVID-19 in central Iran using the local indicators of spatial association (LISA). *BMC Public Health*. Dec 08, 2021;21(1):2227. [FREE Full text] [doi: [10.1186/s12889-021-12267-6](https://doi.org/10.1186/s12889-021-12267-6)] [Medline: [34876066](https://pubmed.ncbi.nlm.nih.gov/34876066/)]
76. Çelebi oğlu F. Regional disparity and clusters in Turkey: a LISA (local indicators of spatial association) analysis. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*. Dec 2010;28:35-48. [FREE Full text]
77. Martínez Batlle JR, van der Hoek Y. Clusters of high abundance of plants detected from local indicators of spatial association (LISA) in a semi-deciduous tropical forest. *PLoS One*. Dec 13, 2018;13(12):e0208780. [FREE Full text] [doi: [10.1371/journal.pone.0208780](https://doi.org/10.1371/journal.pone.0208780)] [Medline: [30543650](https://pubmed.ncbi.nlm.nih.gov/30543650/)]
78. Fang D, Thomsen MR, Nayga RMJ, Goudie A. Association of neighborhood geographic spatial factors with rates of childhood obesity. *JAMA Netw Open*. Aug 03, 2018;1(4):e180954. [FREE Full text] [doi: [10.1001/jamanetworkopen.2018.0954](https://doi.org/10.1001/jamanetworkopen.2018.0954)] [Medline: [30646092](https://pubmed.ncbi.nlm.nih.gov/30646092/)]
79. Hajizadeh M, Campbell MK, Sarma S. A spatial econometric analysis of adult obesity: evidence from Canada. *Appl Spatial Analysis*. Apr 3, 2015;9(3):329-363. [doi: [10.1007/s12061-015-9151-5](https://doi.org/10.1007/s12061-015-9151-5)]
80. Asosega KA, Adebajji AO, Abdul IW. Spatial analysis of the prevalence of obesity and overweight among women in Ghana. *BMJ Open*. Jan 17, 2021;11(1):e041659. [FREE Full text] [doi: [10.1136/bmjopen-2020-041659](https://doi.org/10.1136/bmjopen-2020-041659)] [Medline: [33455932](https://pubmed.ncbi.nlm.nih.gov/33455932/)]

Abbreviations

- CDC:** Centers for Disease Control and Prevention
- CNN:** convolutional neural network
- DCNN:** deep convolutional neural network
- DL:** deep learning
- DNVF:** deep neural visual feature
- ESA:** exploratory spatial analysis
- FCN:** Fully Convolutional Network
- GIS:** geographic information system
- GWR:** geographically weighted regression
- LISA:** local indicators of spatial association
- ML:** machine learning
- PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analyses
- PSPNet:** pyramid scene parsing network
- ResNet-50:** Residual Network-50
- RS:** remote sensing
- TIGER:** topologically integrated geographic encoding and referencing

Edited by C Xiao; submitted 16.07.24; peer-reviewed by A Maharan, PG Asteris, H Gangavarapu; comments to author 23.09.24; revised version received 18.10.24; accepted 15.11.24; published 17.12.24

Please cite as:

Dahu BM, Khan S, Toubal IE, Alshehri M, Martinez-Villar CI, Ogundele OB, Sheets LR, Scott GJ
Geospatial Modeling of Deep Neural Visual Features for Predicting Obesity Prevalence in Missouri: Quantitative Study
JMIR AI 2024;3:e64362
URL: <https://ai.jmir.org/2024/1/e64362>
doi: [10.2196/64362](https://doi.org/10.2196/64362)
PMID: [39688897](https://pubmed.ncbi.nlm.nih.gov/39688897/)

©Butros M Dahu, Solaiman Khan, Imad Eddine Toubal, Mariam Alshehri, Carlos I Martinez-Villar, Olabode B Ogundele, Lincoln R Sheets, Grant J Scott. Originally published in JMIR AI (<https://ai.jmir.org>), 17.12.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR AI, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.ai.jmir.org/>, as well as this copyright and license information must be included.