# LEVERAGING SOCIAL MEDIA TO BETTER UNDERSTAND PEOPLE'S OBESITY RELATED BEHAVIORS

by

Chuqin Li

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Software and information System

Charlotte

2021

Approved by:

Dr. Albert Park

Dr. Yaorong Ge

Dr. Siddharth Krishnan

Dr. Jennifer Webb

# ABSTRACT

CHUQIN LI. Leveraging social media to better understand people's obesity related behaviors. (Under the direction of DR. ALBERT PARK)

With the worldwide prevalence of obesity doubling over the past 30 years across a wide range of demographic and socioeconomic groups, the obesity epidemic is a major public health challenge today. Excessive food consumption and lack of exercise are two major contributors to obesity. In recent years, these two activities have been integrated into our daily lives together with social media. A growing body of literature delineated how social networks impact people's health-related behaviors and suggested social media has a role in affect people's obesity-related behaviors. This study aims to gain an insightful understanding of which online social factors are impacting users' obesogenic behaviors and explore computational methods to examine those behaviors using social media data.

Our work consists of three overall research aims. In the first aim, a systematic review was conducted to examine online social factors concerning obesity. A total of 1,608 studies that related to obesity and social media were collected from the three most popular electronic databases for the field. After close inspection, 50 studies were further examined and ten types of online social factors were identified within four-level social-ecological model, which was used to explain each factors' potential impact on an individual from varying levels of online social structure to user's connection to the real world. In the second aim, we learned how the local food environment influences state level obesity rate using social media data. Publicly available Yelp and MyFitnessPal data were collected via a novel approach to characterize the local food environment. Statistically significant Pearson correlation coefficient between the state's food environment and state obesity rate was observed. We further built a computational model to predict the state-level obesity rate using aforementioned data, in which we achieved a Pearson correlation of 0.791 across US states and the District of Columbia. In the third aim, we studied how a major social disrupting event, COVID-19 shutdown, affects users' dietary behavior using social media data. Tweets relating to people's dietary behavior with images from April to June of 2019, 2020, and 2021 were collected. An observational study of behavior patterns was conducted by using image classification models, visualization tools, and text analysis methods. People are found eating more healthier food during complete and partial shutdowns than before Covid-19. Results of this dissertation could help the public health agencies, policymakers, organizations and health researchers to better utilize social media to carry out obesity-related education, obesity surveillance, and develop public health policy to address this challenge.

# ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor Dr.Albert Park, and Dr.Yaorong Ge for their continuous help, invaluable advice, and infinite patience to me in all these years. Thank you Dr.Park for all instructions and encouragement on my research. I'm so glad to work with you.

I would also like to thank my committee members: Dr.Siddharth Krishnan and Dr.Jennifer Webb. Thank you for your time, suggestions, and valuable comments to help me finish this dissertation.

Special thanks are given to program coordinator Dr.Mohamed Shehab, and Ms. Sandra Krasuse. Thank you for all the caring you give to students and for always being on my side.

Furthermore, I would like to express my gratitude to my parents and my husband. Without tremendous support from my families, it's impossible for me to finish this challenging journey.

In addition, I'm very grateful to my friends and research collaborators who make my study and life at UNCC a wonderful time: Ganquan Weng,Dr. Bin Kong, Dr.Jinyue Xia, Dr.Xiongwei Xie, Dr.Boshu Ru, Dr.Jingang Liu, Dr.Nasheen Nur, Dr.Junjie Shan, Adesoji Ademiluyi, Alexis Jordan, all my colleagues works in the Health Informatics Lab, and many others.

In the end, I want to acknowledge Dr.Park for giving me financial support that sponsored my research. Meanwhile, I want to thank the Graduate Assistant Support Plan (GASP) at the University of North Carolina at Charlotte.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# CHAPTER 1: Introduction

One third of global population, over 2 billion people is overweight or obese in the world.[1] With the worldwide prevalence of obesity doubling over the past 30 years and its association implication to the public health make obesity as one of the most common, serious, and costly disease in the world.[2] Obesity has been suggested to cause chronic diseases, such as hypertension, type 2 diabetes, cardiovascular disease, and some cancers,[3] and associated with stigma,[4] higher health costs for individual and government,[5] life expectancy decrease and lower life quality.[6] Global efforts are required to understand the contributors to obesity and control obesity pandemic.

In the past couple of years, social media becomes an essential part of people's life. In this dissertation, we use the broad social media concept, which include various platforms that allowing users to profiles, make connections with other users, and view contents made by others in the system.[7] Survey on U.S. adults in 2018 found the social media landscape shows a long-standing trends and newly emerging narratives, with more than three-quarters of American adults go online on a daily basis, and a quarter of Americans were reported to be almost constantly online [8] Increasing number of people's social interaction is publicly shared online and social networks are found impact people's obesity-related attitudes and behaviors.[9]

Traditionally, health-related studies collected data through face-to-face or survey-based methods. A major limitation of these traditional data collections is that they are cost-ineffective and labor-intensive and the methods can only gather a limited number of samples and the significant delay between collecting and reporting. However, social media data is organic, continuously updated, and generally free for large scale collection. Researchers use social media data to study public health, especially the "lifestyle disease" like obesity.[10] In this dissertation, we focused on the field that leveraging social media to better understand obesity-related behaviors. We identified gaps in this field and filled it with three separate studies.

This dissertation follows the guidelines for three-paper dissertation, which includes five chapters. In addition to this introduction and the conclusion chapter, other three chapters include three separate papers. Chapter 1 (paper 1) is a systematic review paper, in which we systematically reviewed 50 papers that utilizing social media to study obesity-related behaviors. In this study, we identified ten types of online social factors that impact an individual's obesity-related online behaviors and explain their potential impact on an individual. This study gives us an overall picture of how social media has been used in obesity-related studies and their limitations. Chapter 2 (paper 2) illustrates a novel approach to learn local food environment using social media data. In this study, we used the Yelp and MyFitnessPal data to characterize the local food environment and predict state level obesity rates. Different from previous studies, this study is the first one to use Yelp data to learn local food environment. Chapter 3 (paper 3) used the social media data to understand how the unexpected social interrupting event, COVID-19 shutdown, affect users' dietary behaviors. In this study, we used the Twitter data to understand user's dietary healthiness and its relationship to state level obesity rate. This study shows the potential of social media to understand users' health-related behaviors during health crisis.

## 1.1 Overview of Paper 1

Evidence in the literature suggests social factors have a substantial role in the spread of obesity. Close social tie with an obese friend increases the probability of becoming obese. However, the role of social factors that exist in social media is underexplored in obesity research. With the rapid proliferation of social media over the past few years, individuals socialize on social media and share their health-related daily routines, including dieting and exercising. Thus, it is timely and imperative to review previous studies focused on social factors in social media and obesity. This study aimed to examine online social factors in relation to obesity research.

A systematic review was conducted. We searched PubMed, ACM, and ScienceDirect for articles published by July 5, 2019. A total of 1,608 studies were identified from the selected databases. Of these, 50 studies met eligibility criteria. Ten types of online social factors were identified, and a social-ecological model was adopted to explain their potential impact on an individual from varying levels of online social structure to social media users' connection to the real world.

We found four levels of interaction found on social media. Gender is the only factor found at the individual level that affects user's obesity-related online behaviors. Social support is the most predominant factor among identified factors, which benefits users for their weight loss journey at the interpersonal level. Some factors, such as stigma, are also found associated with a healthy online social environment. Understanding the effectiveness of these factors is essential to help users create and maintain a healthy lifestyle.

## 1.2 Overview of Paper 2

The food environment is associated with resident's obesity outcomes. Social media data from Yelp and MyFitnessPal (MFP) were used to learn about the relationship between food environments and rates of obesity at the state level. We first compared the differences in food category availability between two states with lowest and highest obesity rates: Colorado and Mississippi. Using the popular dishes for food category from Yelp.com and the nutrition information for each popular dish from MFP, we characterized the local food environment by averaging calorie for each category and the weighted score for each state. The Pearson correlation coefficient between state's food environment and state obesity rate is statistically significant. Dimensions from the concept of access were adopted to built computational models to predict state-level obesity rate. We achieved a Pearson correlation of 0.791 across US states and the District of Columbia.

## 1.3 Overview of Paper 3

The pandemic of COVID-19 has a great impact on people's lifestyle. Understanding how COVID-19 shutdown affect people's dietary healthiness is important at this moment. This study firstly explored the feasibility of using Twitter image data to examine the healthiness of people's dietary. Transfer learning for deep learning were used to determine the healthiness level of dieting images. Results showed a significantly correlation between the healthiness of dieting images and state obesity rate in 2019. Using image classification models, text analysis methods, and visualization tools, we studied how COVID-19 shutdown affect people's dietary. Overall, although higher percent of negative emotions were found associated to dietary-related tweets, people's dietary were found to be healthier during and partially after the shutdown. This study could be used as a preliminary work for further analysis.

# CHAPTER 2: Using Social Media to Understand Online Social Factors Concerning Obesity: a Systematic Review

## 2.1 Introduction

### 2.1.1 Background

The obesity epidemic is a significant public health challenge in today's society. The growing prevalence of obesity and its implications in public health make obesity one of the most common, dangerous, and costly diseases in the world.[1] One-third of the global population, over 2 billion people, are overweight or obese. Obesity rates reached 39.8% among adults and 18.5% among youth in the U.S. in 2016, which is a significant increase in these age groups from 1999.[2]

Obesity is recognized as an major risk factor for population health4 due to its association with social stigma[3], chronic diseases,[4] medical complications,[5] reduced life expectancy,[6] lower quality of life,[7] and higher health costs for individuals[8] and government.[9] The World Health Organization (WHO) suggests that obesity is likely the cause of chronic diseases like hypertension, type 2 diabetes, cardiovascular diseases, and some cancers.[10] Women with obesity are more vulnerable to infertility, miscarriage, and other child-bearing related complications.[5] Obesity in childhood also increases the risk of other diseases, and may even be carried through adolescence to adulthood.[4] Obesity also increases individual and fiscal expenditure. The annual medical cost of obesity in the U.S. was $147 billion US dollars in 2008, and the annual medical cost for an individual with obesity was estimated to be $1,429 higher than individuals with a healthy weight.[9] Obesity may affect human development,[6] suggesting the prevalence of obesity could be a detriment to human life expectancy in the 21st century. This could reverse the increase in the life expectancy seen in the 20th century. Global efforts are paramount to control the obesity pandemic.

### 2.1.2 Social factors as important drivers of obesity pandemic

Recent developments in research have identified two main factors, exercise, and diet[11, 12]; however, there are other factors associated with obesity.[13] According to a study, developed society is the leading cause of the current obesity pandemic in that it creates an obesogenic environment.[14] An obesogenic environment is defined as easy access to inexpensive and calorie-dense food, excessive food intake, insufficient physical activity, and inexpensive non-physical entertainment.[5, 15] The obesity epidemic with interconnected social factors could result in an obesity pandemic.

Social factors are defined as factors that affect an individual's lifestyle.[16] These influences have a significant effect on people's health-related behaviors.[11, 17] Thus, social factors play an important role in the spread of obesity. For example, a study conducted by Christakis & Fowler[18] tracked a densely interconnected social network of 12,067 people for 32 years. It showed that a person's chances of becoming obese increased by 57% if he or she has a close relationship with someone who is obese. Furthermore, the self-perception of weight can also be influenced by their peers. Previous researches indicate that children and adolescents who are surrounded by many overweight peers might have inaccurate perceptions of their weight and underestimate it.[19, 20]

With the recent ubiquity of social technologies, these peer effects are expanding to the general public. A recent study on a large-scale social network shows social influences also affect collective public health behaviors, such as obesity and tobacco usage.[21] Similarly, examining user interactions on social media has been proven useful in understanding public attitudes and perceptions surrounding health topics.[22] As a result, it is timely and imperative to understand online social factors that exist in an online social environment for counteracting the obesity epidemic.

### 2.1.3 Social media for understanding obesity

Social network sites serve as web-based services that allow individuals to build social profiles, form connections with other users, and view other profiles in the system.[23] The popular social network sites, like Facebook, Twitter, and Reddit, attract millions of users since they were first introduced in 2004, 2006, 2005, respectively. A 2018 survey on U.S. adults found that the social media landscape shows a long-standing trend of continuous usage throughout the day and newly emerging narratives (e.g., posts, tweets, images).[16] For example, 69% of U.S. adults use at least one social media site; 74% of Facebook users and 46% of Twitter users access the site daily.[24] Internet integration may offer possibilities for accessing obesity-related information, including weight loss, obesity diagnosis, and weight management.

Social media platforms show the potential to change users' health behaviors. An increasing number of users' social interaction is shared publicly online. This makes social media a vital data source for studying public health, especially "lifestyle disease" like obesity.[22] Chang et al.[25] systematically reviewed the use and impact of social media in online weight management and demonstrated social media plays a role in retaining and engaging participants for weight management.

### 2.1.4 Review aim

The primary aim of this review is to extend the knowledge on the influences of online social factors concerning obesity-related behavior to better inform future studies in examining interventions utilizing social factors. This is the first study to systematically review online social factors' effects of obesity-related behaviors in the online social media environment. Other related systematic reviews had examined obesity and social media usage and the effectiveness of social media in intervention studies for obesity prevention.[26, 27]

## 2.2 Methods

### 2.2.1 Data sources and search strategy

Three most popular electronic databases, PubMed, Association for Computing Machinery (ACM), and ScienceDirect were used in this review. PubMed is known as a comprehensive database in biomedical research;[28] The ACM database is maintained by the world's largest scientific and educational computing society;[29] and ScienceDirect provide us access to an extensive database of scientific and medical research.[30] The search strategy in ACM and ScienceDirect was designed by combining search terms: social media and obesity. The full search string in ACM and ScienceDirect is '"Social media" AND obesity'. Medical Subject Heading (MeSH) terms "social media"[31] and "obesity" were utilized in the PubMed search. The full search string in PubMed is '"Social Media" [Mesh]' AND "Obesity" [Mesh]'. All searches were completed on July 5, 2019.

### 2.2.2 Study selection and screening

Our study aims to review studies that utilize social media with elements of social factors for obesity research. To meet the review aim, we define the social media in this study as an internet-based platform allowing individual users to create and exchange content (e.g. blogs, online discussion board, Twitter) based on the previous study by Kaplan & Haenlein.[32]

All studies that meet the inclusion criteria are included in this review. Inclusion criteria are defined as follows: (a) obesity is the primary study topic; (b) social media serves as the main platform; (c) social interactions are incorporated; (d) publish in the peer-reviewed literature; (e) in the English language. To understand how online social factors can influence people in understanding or improving weight-management outcomes, we set either organic or encouraged social interaction as an inclusion criterion. Other types of scholarly articles are excluded: comments, systematic reviews, conference reports, and letters. Moreover, design studies that only suggest the use of social media, such as a randomized controlled trial study design by Willis et al.[33], are excluded.

Two independent reviewers first screened all articles based on the title and abstract. All articles were categorized into (a) include, if this paper meets the inclusion criteria; (b) exclude, if this paper does not meet the inclusion criteria; and (c) need full-text review, if the abstract cannot provide enough information or the abstract is not available. A paper was then excluded at the screening stage if two reviewers both agree to exclude it based on title and abstract. Except for the excluded articles, all articles were moved to the eligibility stage, which requires two reviewers to do a full-text review. At the eligibility stage, any disagreement was discussed to form a consensus. The third reviewer, a tiebreaker, was introduced if the consent could not be reached.

## 2.3 Results

In total, 1608 studies were identified from our selected database and search strategy; of those, 16 were duplicated (Figure 2.1). After removing the duplicates and assessing the title and abstract, 1507 articles were excluded, and 85 remained for full-text reading. Full-text examination excluded 35 articles. In total, 50 articles met our inclusion criteria and were included in this systematic review. We summarized the online social factors and their corresponding effectiveness in the following section. Further, we examined how different social media platforms were utilized in prior studies.

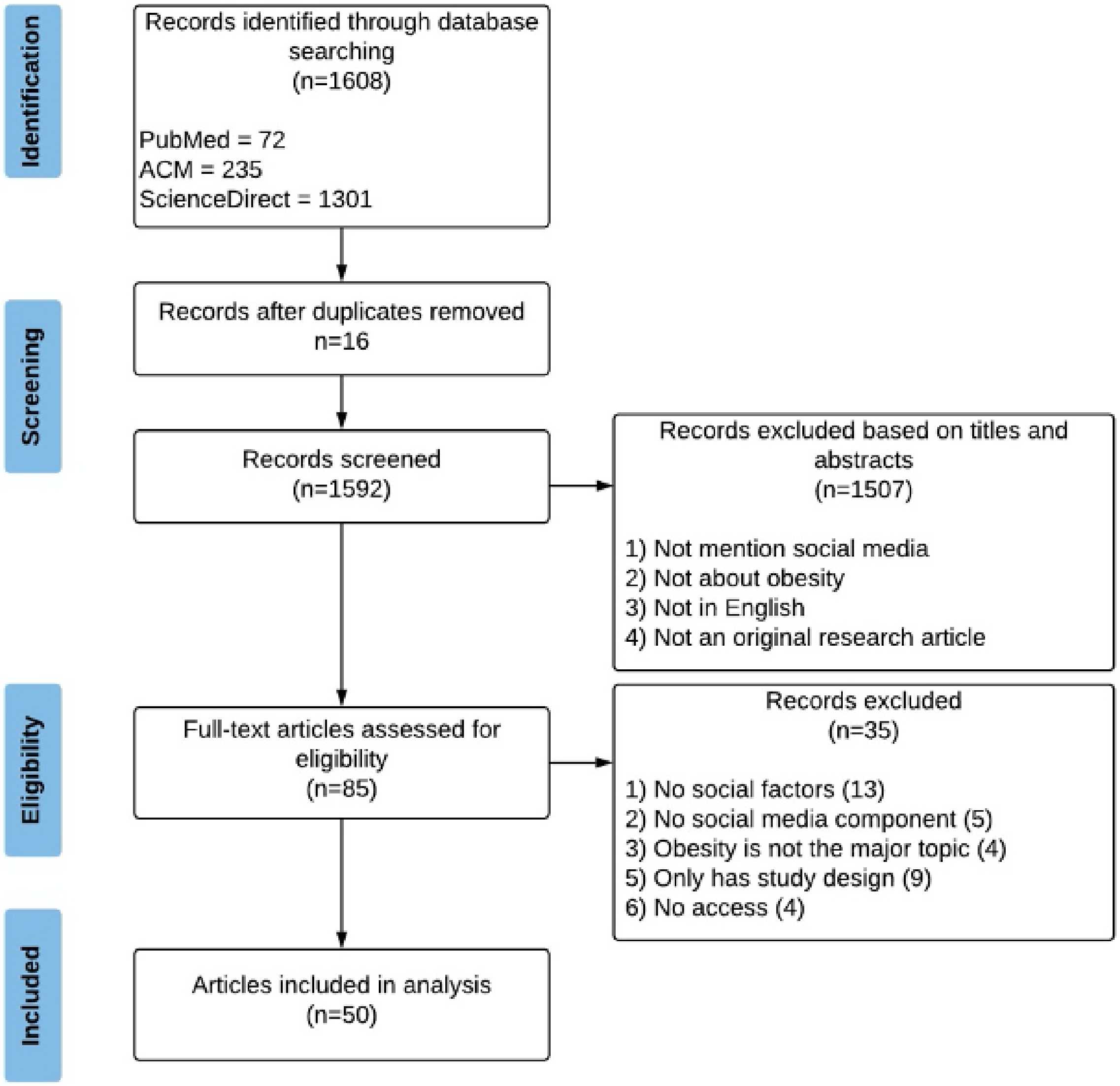



Figure 2.1: Study flow.

### 2.3.1 Online social factors

Traditionally, obesity is linked to behavior at the individual level, such as overeating and lack of exercise. However, new studies shed more light on social factors that are contributing to obesity-related behaviors. We studied online social factors in this study. There is no universal definition of online social factors in the literature. Here, we defined online social factors as social factors that exist in online social environments and have the potential to affect users' behaviors. We focused on identifying online social factors and understanding their potential effect on user's obesity-related behaviors. We found ten different online social factors that were used and mentioned in previous studies. In the order of the most common to the least common online social factors were: social support and social ties, gender, geo-cultural factors, stigma, obesogenic environment, source credibility, school environments, social movements, policy, and social sharing behaviors. We will first discuss the most frequently mentioned social factor, social support and social ties.

#### 2.3.1.1 Social Support and Social Ties

Social support is emotional comfort and material resources that are given to you by peers connected to you in a social network, is the most frequently mentioned factor in previous studies related to our study aim. Users of social media platforms can exchange social support. They view social media as a good place for finding/receiving social support and locating information platforms for those interested in changing their lifestyle and eating habits.[34] A previous study found users who tried to use Twitter to record their weight-loss journey report that they receive more social support from the virtual environment than from their real families and friends.[35]

In online communities, we identified two types of social support: informational support and emotional support. Informational support includes sharing resources and providing professional feedback through social networks.[36, 37, 38] Emotional support primarily came from peers. Social culture and the concept of social media encourage users to be active with other users. In some intervention studies[37], program participants were instructed to discuss progress, issues, and goals with other participants using social media platforms. Through various platforms/programs, peer encouragement[39], peer support[40, 41], and peer pressure[38] are found.

Moreover, two platforms were found as the major conduits for social support: blogs and Facebook. Savolainen[42] asserts that the main strength of interactive blogs is that they can provide emotionally supportive forums for sharing opinions. From a blogger's perspective, Leggatt-Cook & Chamberlain[40] learned bloggers hope to create and build a community that will support them and their attempt to lose weight. Facebook's private group was widely used in intervention studies. It was created to share resources and serve as a platform for participants to communicate.[37] In Waring et al.'s study[38], first-time mothers were often found to seek out other mother's advice and support from Facebook groups. Twitter is primarily utilized to collect public opinions, but it was also found to offer the opportunity to create a supportive network.[36]

Social support is suggested to be very important for users who try to lose weight. Lack of support showed a negative effect on weight loss. Pappa et al.[41] found the peers searching for support were inversely associated with weight loss. Even in the anonymous platform Reddit, the authors observed an increasing number of users returning to the community, and a greater weight loss was reported from users if the users received support in the community.[39] The effectiveness of social support in weight loss was reported to be positive. He et al.[43] found social support positively correlates with weight loss. Jane et al.[34] also found individuals have better health outcomes if they were supported by professionals. Chomutare et al.[44] mentioned that their study found a positive correlation between online participation and weight loss by analyzing the data on older women with obesity who were active in an online community. A study conducted on individuals with mental illness also found that weight loss was associated with perceived peer group support because the subjects felt compelled to pursue weight loss goals.[45] We also found one detracting study on social support influence on BMI reduction. An experiment on college students showed students in the Facebook support group did not show a significant difference in BMI with the controlled group at the end of 24 months. However, the author reported a significantly greater increase in the number of appropriate weight-control strategies than those students who are not in the support group.[46] Similarly, a meta-analysis by Merchant et al.[37] suggests interventions that give participants professional support during their diets and physical activities are more effective than those that do not.

Social support is understood through social ties, the connection among peers. Social tie theory concludes that the probability of a person becoming obese increases if friends with obesity surround them.[18] Phan et al.'s study[47] in this systematic review has adopted the social tie theory in their experimental study confirming that individuals tend to perform similar lifestyle behaviors as their friends from the online environment. Social ties also affect received social support. Social support has a negative correlation with weak ties (i.e., unfamiliar individuals) in the context of weight control. Chen et al.'s study[48] reported that participants found community competition and support from strong ties (e.g., couples, parents) were motivating, while support from unfamiliar participants was demotivating.

#### 2.3.1.2 Gender

We identify gender as one online social factor because we found six previous studies[36, 38, 49, 50, 51, 52] indicating that women and men have different online social behaviors regarding obesity. Abbar et al.[49] showed that women often are more willing to share information online, such as preparing low-calorie food, which was also supported by other related studies.[53, 54] Women would like to share their family members' weight management experiences on social media. Only mothers of a childhood weight loss camp were willing to use social media to receive informational support and post their children's progress.[36] In a Twitter-delivered weight loss program, Waring et al.[38] found that a great proportion of women read each other's progress. These participants were reading other people's tweets more than posting their own progress. Women were critical when self-evaluating their weight. In a study by Kuebler et al.[50], Yahoo Answers data found that most women, when asked their self-perspective on their weight, tended to overestimate. Online social norms show the characteristics of women and men that are socially constructed.

#### 2.3.1.3 Geo-cultural factors

Users' health behaviors occur in a setting composed of online as well as social and cultural environments. Geo-cultural factors explain how user's online behaviors are affected by physical surroundings. Several studies found that social media data could provide insight into the health conditions of United States residents. Three studies by Gore et al.[55], Culotta[56], and Abbar et al.[49] used Twitter data to predict county-level health. Abbar et al.[49] discovered the calories of food mentioned in tweets correlated to the county's obesity rate. Gore et al.[55] found that the tweets in areas with lower obesity rates have three features: 1) tweets have more positive sentiments; 2) more tweets mention fruits and vegetables; 3) physical activities are more frequently mentioned of any intensity. Culotta[56] has similar findings that negative emotions have been found in tweets from the high obesity rate area. Garimella[57] further validate the feasibility of using image data to track public health concern. They found that user and machine-generated tags of images on Instagram could be used to forecast the county's obesity rate. By analyzing pictures on Instagram, Mejova et al.[58] found the number of fast-food restaurants in a county in the United States have a positive correlation with the local obesity rates. They further revealed locally owned restaurants with dietary and nutritional alternatives were more popular in areas with lower obesity rates. Another interesting finding from Weber  Mejova[59] showed the percentage of profiles with a valid profile picture was higher in areas with a higher obesity rate. Another branch of the geo-cultural factor is involved in culture and religion. After a weight loss camp in Qatar, the study concluded that the religious month and cultural orientation, which should be considered, was critical to the outcome, affecting user's online recording behavior.[51] These findings suggest that users' online obesity-related behaviors are related to location-specific environments.

#### 2.3.1.4 Stigma

Stigmatization and associated discrimination, sometimes referred to as weight bias, affect the subject's mental and physical health, as well as social behavior. Studies on western culture highlight the stigmatization of individuals with obesity that individuals with obesity were stigmatized and associated with lazy, low self-control, and moral laxity.[53] Social media is used to propagate social stigmatism, mainly in the form of fat-shaming, a practice of humiliating and criticizing overweight individuals on social media.[60] Mejova et al.[60] found up to 27.6% of the non-URL tweets mentioning obesity were fat-shaming, with some self-hate messages. In a recent study by Karami et al.[61], Twitter users often coupled exercise-related terms with obesity. This could also indicate that individuals associate exercise and self-regulation (or lack of) as the main cause and solution for obesity. Although the social stigmatism of obesity is widespread online, individuals were also pushing back social stigmatism using social media platforms. People expressed anger caused by the obesity stigma on Twitter in a retaliatory manner to address widespread stigmatization against overweight and obese.[62]

Stigma has been found to undermine a user's mental health, but its effects on user's online interaction remain inconclusive. A person's mental health status revealed by social media data indicates a user's mental health is impacted by their social surroundings. A study by Kuebler et al.'s[50] suggests that people with obesity residing in counties with higher levels of BMI have better physical and mental health compared to people with obesity living in regions with a low level of obesity rate. Another two studies investigated the impact of weight stigma by comparing users' online behavior between users with normal weight and obesity. May et al.[63] did not find weight bias in their study since the weight status had no effect on the rate of interactions and follow backs. However, a different study by Weber et al.[59] found users who are labeled as overweight had fewer followers and fewer directed tweets.

Although stigmatization may not affect the user's online behavior, the widespread stigmatization on social media will diminish a user's mental health.

#### 2.3.1.5 Obesogenic environment

Obesogenic environment refers to an environment that promotes high energy intake and sedentary behavior.[64] By analyzing the content users post to social media, we can find that the obesogenic environment is one of the major causes of the obesity pandemic. A content analysis of frequent retweets about obesity by So et al.[62] revealed that four major societal determiners of obesity are discussed on Twitter: cheap and unhealthy food, school food systems, portion size, and dysfunctional food systems. Among these determiners, easy access to cheap, calorie-dense foods received the highest tweeting rates. This finding suggests online information environment is changed by physical obesogenic environment by informing user's behaviors.

#### 2.3.1.6 Source credibility

Credible health information sources are persuasive,[65] however, some social media obesity-related information was found incomplete or inaccurate. The low-credibility source could exert a negative influence on user's obesity-related behaviors. The primary reason revealed by the previous studies is that the information from professionals is lacking. Online information from professionals about obesity proved more accurate than from other users. Erdem & Sisik[66] analyzed the content of 300 YouTube clips on bariatric surgery, also known as weight-loss surgery, and suggest the content from professional accounts tends to be more accurate. In another study, Basch et al.[67] analyzed the top 100 most widely viewed weight loss videos on YouTube and found only one professional video; consumer-created videos dominated the domain. Mejova[60] examined 1.5 million tweets mentioning obesity and diabetes and found only 23% of the content came from verified users (i.e., Twitter accounts that are associated with a governmental or academic institution). Similarly, more individuals than organizations tweeted about childhoodobesity.[65]

Misinformation in content could also harm users. YouTube advertisements for rapid weight-loss products and commercial videos focused too much on workouts instead of the importance of maintaining a balanced diet.[67] Top cited domains relating to obesity and diabetes on Twitter are not affiliated with guidelines provided by governmental or academic institutions.[22] The discrepant information from user-generated content can lead to a drop-in trust for these platforms. Messages presented in traditional social media platforms, like blogs, were seen as a more reliable source than other newer social media platforms, such as Facebook.[68] Meitz et al.[68] compared the source credibility perceptions among different platforms and found messages on Facebook were perceived as significantly less relevant than messages presented in blogs. Together, these studies reinforce the importance of source credibility on conducting user's health behaviors.

##### 2.3.1.7 School Environment

School, serving as a center of childhood development, has an influential role in a child's early behavioral development. Findings from social media content analysis suggest that school environment plays an important role in affecting children's obesity-related behaviors. From one aspect, the school decides students' daily routine and food selection. So et al.[62] argued against excess homework and Harris et al.[65] noticed that most public schools do not regulate access to junk food. Additionally, teachers' participation in combating childhood obesity is critical. In a preschool obesity prevention curriculum parents showed a strong desire to see more engagement from their classroom teachers.[69] Changing school environment was the most common strategy to combat childhood obesity mentioned on Twitter. As people tweets: "Americans expect schools to lead in preventing childhood obesity."[65] User's attitudes towards school environments are shaping the online discourse on childhood obesity.

#### 2.3.1.8 Social movements

Social movements are defined as an organized effort by a group of people to bring out or impede social or cultural change.[70] Social media offers us a new possibility to explore social movements efficiently. In recent years, there have been several distinct online trends regards obesity was found in social media: 'body positivity', 'thinspiration', 'fitspiration', and 'HAES'. The term 'body positive', originally came from the 1960s feminist movement and resurfaced in the fat acceptance movement. A content analysis of 'body positive' images on Instagram showed that this movement seeks to challenge beauty standards while rejecting the inaccessible body image and promoting acceptance of all body types and appearances.[53] However, another prevailing trend called 'thinspiration' surfaced with the intent of spreading thin body imagery and inspiring weight loss. Content analytics showed body-positive images on Instagram drew a broad range of body sizes,[53] while thinspiration images on Twitter tend to depict ultra-thin and scantily clad women.[71] Exposure to guilt-inducing and body objectifying messages has been found to increase body dissatisfaction and negative moods. The study also showed that the more times individuals view thinspiration content, the higher the probability they will report eating disorder symptoms.[67] Afterward, a new trend called 'Fatspiration', supporting fat acceptance, has come to prominence in the mainstream,[54] was found online. Another trend 'HAES,' 'Health at every size,' promoting wellness over weight-loss, was also identified. Notwithstanding, discrimination against obesity had not been resolved. Fat stigmatization content within the 'HAES' tag and 'Fatspiration' tag was found in a content analysis of the Instagram images.[72] Social media shapes the online information environment and helps us to fully understand the context of social movements. Without understanding the full scope there is potential for a negative shift in social norms.

#### 2.3.1.9 Policy

Governments tried to combat obesity by establishing new policies. Three studies[65, 73, 74] use social media to study the public's attitude and reaction toward government policy. In 2016, the UK also published a plan to reduce England's rate of childhood obesity within the next ten years, "Childhood obesity: a plan for action." Most comments to the related online newspaper articles were found to be negative in Gregg et al.'s study[74]. Later, in 2017, Kang et al.[73] collected relevant tweets to investigate the public's opinion on a new school meal policy. They found 70% of tweets were neutral, though the number of negative tweets was still higher than positive tweets. Negative tweets express interest and concern about the policy and suspicion of the effectiveness of the campaign. Instead of worrying, some users also using social media to support the announcement and execution of the policies. Harris[65] found 'Bye Bye Junk Food', a USDA rule that requires healthier snacks for kids and adoption of the physical education classes as a core subject in schools, was a prominent movement in communication about childhood obesity on Twitter. Online information can shape individual's attitude about certain policy. With social media, policymakers could better disseminate policy and raise public awareness.

#### 2.3.1.10 Social sharing behaviors

Two psychology theories, social sharing of emotions and cognitive dissonance theory relating to sharing behaviors have been found to be related to the online sharing behaviors in literature. Social media makes it easier for users to share their opinion, and user's online obesity-related behaviors have been found to be guided by these theories. In 1995, Rime[75] proposed that emotion is a critical motivator in social sharing. The social sharing of emotions has shown that people have the innate need to tell others when they experience an emotionally impactful event.[62] According to So et al.[62], a content analysis of frequent retweets about obesity on Twitter, the emotionally evocative tweets, specifically evoking amusement, were the most frequently retweeted. Another theory, cognitive dissonance theory, posits that we are experiencing psychological discomfort when we encounter beliefs that are inconsistent with our own. As a result, people try to reduce this discomfort by exposing themselves to information that helps them to resolve the cognitive conflict. A study on Instagram pictures showed the people who reside in high obesity areas are more willing to post food-related photos than people in low obesity areas.[58] Understanding the underlying mechanisms of these phenomena, we can better manage obesity-related behavior.

### 2.3.2 Study year and region

All the included studies were published between 2011 â 2019, the number of each study by year is shown in (Figure 2.2). This number started to increase from 2014 (n=7), and it reached a peak in 2017 (n=14). However, this is partially due to the fact our search was done in July of 2019.

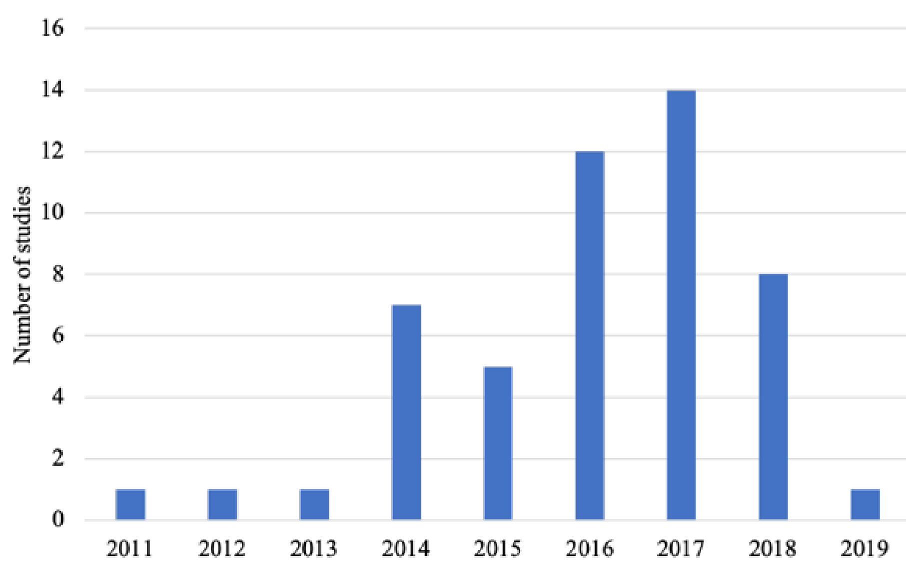



Figure 2.2: The number of studies published in each year.

In total, nine study regions were mentioned, with the majority of the studies coming from the United States (n=35). Except for China, Australia, and Qatar, all other regions were located in Europe. The number of studies from other countries was shown in (Figure 2.3). Several groups of users were studied. Children, women of child-bearing age (e.g., pregnant[52], postpartum women,[38] mothers with newborn[76]), adults with other relevant illnesses (e.g., diabetes[48]) were three leading types of user groups that were studied. Only one study by Aschbrenner et al.[45] focused on adults with obesity and severe mental illness.

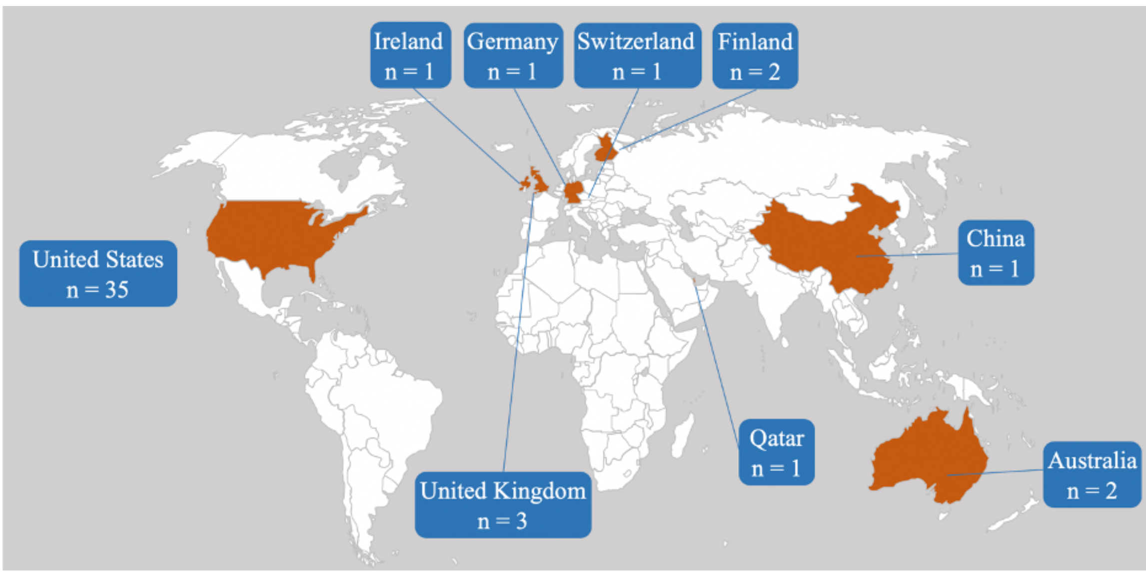



Figure 2.3: The distribution of study region.

### 2.3.3 Social media platforms and their roles

We identified three different roles that social media serves in each study. Inspired by Leroux et al.'s study[77], three potential roles of social media in obesity-related studies were identified: data collection, intervention pathway, and ancillary resources. Data collection is used to define when social media platforms were only used to collect data used in the study. The intervention pathway defines when social media usage as a comprehensive channel in an intervention study. The role of social media includes delivering the message and serving as an online communication platform for participants in weight management/loss intervention studies. We define ancillary resources as the role when social media is used as an experiment platform. Data are also collected when serving as ancillary resources. The major difference between the data collection and ancillary resources is the source of data. If the data is collected for the purpose of analyzing and understanding the data, we define it as data collection; and if the data is generated from the experiment, we define it as the ancillary resource. The type of study can be used to distinguish the intervention pathway and ancillary resources. The social media platforms only serve as an intervention pathway in an intervention study. The categories are mutually exclusive.

We also categorized social media platforms into six different types: microblogging, social network, weblog, photo/video sharing, online forum, and messaging. Facebook, the largest online social network, was the most frequently used (n = 14). Facebook was used the most for intervention studies. Facebook was used as a message delivery channel, in which private Facebook groups were introduced as a smaller online support group for participants to share goal-related resources, individual signs of progress, and messages. Only one study collected data from Facebook by accessing public Facebook posts.[22] Similar to Facebook, a Chinese social network platform, WeChat, was used in He et al.'s study[43] as the intervention pathway. Twitter allows users to communicate with others from all over the world. Twitter, the microblogging platform, was used to collect data to understand the public's opinion (n = 11). One study also used Twitter as an ancillary resource (n = 1) to conduct their experiment. For example, May et al.[63] created four Twitter accounts portraying women (two obese, two normal/overweight) who were interested in weight loss and pretended to behave as regular users. Later, they examined the interaction with other users and mimic users (e.g., follow back rate). Another study used Twitter as the intervention pathway to involve participants.[36] Not many studies investigated Blogs (n = 4), mainly due to the difficulty in extracting meaningful insights from large pieces of text. Similarly, only two studies analyzed clips from YouTube due to its complexity in analyzing videos. Instagram, one of the most prominent photo-sharing platforms, was used in 7 studies. Two studies used Instagram as an intervention pathway, in which the participants were asked to upload their meals. In other studies, researchers analyzed public photos to understand public health,[57] social movements,[53],[72] and social sharing behavior.[58] A small number of studies analyzed Twitter and Instagram data with extended data points. For example, researchers combined demographic information and geolocation information to better predict the obesity rate of the regions.[56] Reddit has been used in a lot of health-related studies[78]; however, we only found two studies using Reddit in our search. Another type of social media is a forum. Three were identified to investigate a forum, Yahoo Answers, a Self-developed application (i.e., HealthTogether), and another platform that was not specified in the literature. WhatsApp was chosen as the intervention pathway in one study, to serve as the alternative to the traditional text message method. In this study, group-chat rooms were formed to deliver information and allow participants to communicate virtually. Please see (Table 2.1) for more comprehensive information. The amount of data used in the study is concluded in Data and the major study object is summarized in Group. The distribution of the three roles the social media serves is shown in (Figure 2.4).

Table 2.1: A descriptive overview of social media platforms.

| Type | Platforms | No. of studies | Data collection | Intervention pathway | Ancillary resource | Subtotal |
|---|---|---|---|---|---|---|
| **Microblog** | Twitter | 14 | 11 | 2 | 1 | N=14 |
| **Social network** | WeChat | 1 | - | 1 | - | N=15 |
| | Facebook | 14 | 1 | 13 | - | |
| **Weblog** | Blogs | 4 | 4 | - | - | N=4 |
| **Photo/Video sharing** | Instagram | 7 | 5 | 2 | - | N=11 |
| | Pinterest | 1 | 1 | - | - | |
| | YouTube | 3 | 3 | - | - | |
| **Online forum** | Reddit | 2 | 2 | - | - | N=11 |
| | Yahoo Answer | 1 | 1 | - | - | |
| | Self-developed application | 2 | - | 2 | - | |
| | Unknow community | 6 | 3 | 2 | 1 | |
| **Messaging** | WhatsApp | 1 | - | 1 | - | N=1 |

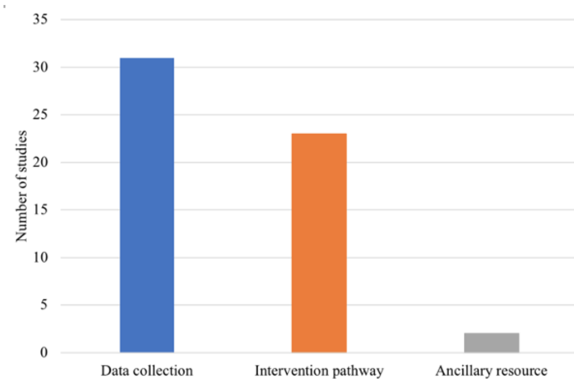
Table 2.2: Summary of Individual study.

| Article | Platform | Region / Group | Data | Online Social factors | Primary findings | Conceptual |
|---|---|---|---|---|---|---|
| Kang et al, 2017[73] | Twitter | U.S. / UNK. | 14,317 related tweets | Policy | More negative tweets about the school meal policy have been detected. The main target negative opinions were campaign and food. | Data collection |
| Fernandez-Luque et al, 2017 [51] | WhatsApp Instagram | Qatar / Children | UNK. (intervention study) | Gender, Geo-cultural factors | 1. More active users tend to have a better health outcome. 2. Females' engagement with social media is higher. 3. Nutritional advice in weight loss campaign has to consider religious and cultural traditions. | Intervention pathway |
| Lingetun et al, 2017[52] | Blogs | U.S. / Pregnant women with obesity / overweight | 13 Internet blogs | Gender | Three main themes of overweight pregnant women's blogs were identified: Pregnancy as an excuse; perspectives on the pregnant body, and becoming a mother. | Data collection |
| May et al, 2017[63] | Twitter | U.S. / Adults | UNK. (experiment study) | Social support, Gender, Stigma | 1. Investigated follow back rates. The number of interactions and organic follows did not differ by weight status. 2. Peers interacted more with each other than with professionals. 3. Women need five weeks to build an interactive weight loss community on Twitter. | Ancillary resource |
| Gore et al, 2015[55] | Twitter | U.S. / UNK. | Over 25 million tweets | Geo-cultural factors | Geological areas with lower obesity rates: (1) have happier tweets, and (2) more frequently discuss food, particularly fruits, vegetables, and physical activities. | Data collection |
| So et al, 2016[62] | Twitter | U.S. / UNK. | 200,000 tweets | Social sharing, School environment, Obesogenic environment, Stigma | Tweets that are emotionally evocative or humorous and express individual-level concerns for obesity were more frequently retweeted than their counterparts. | Data collection |
| Kent et al, 2016[22] | Facebook Twitter | U.S./ UNK. | 291 Facebook posts; 1091 tweets | Obesogenic environment | This study aims to understand the connection between obesity and cancer from Facebook and Twitter. They found: (1) The majority tweets focused on an associative or causal link between obesity and cancer. (2) Twitter tweets contain more negative sentiment than Facebook posts. | Data collection |
| Harris et al, 2014[65] | Twitter | U.S. / Children with obesity | 1110 tweets | Source credibility, Policy, School environment | This study investigated the communication about obesity on Twitter, and they found: (1) More tweets focus on individual behavior than policy or environment, (2) The government or educational tweets attract more attention, but the number of these tweets is less. | Data collection |
| Kuebler er al, 2013[50] | Yahoo Answer | U.S. / Adults | 3,926 user's questions; 300 bullying questions | Gender, Stigma, Geo-cultural factors | 1. Most women asking whether they were fat/obese were not fat/obese. 2. Users with obesity were significantly more likely to ask for advice about bullying than thinner users. 3. People with obesity who reside in counties with higher BMI may have better physical and mental health than people with obesity who live in counties with lower BMI. | Data collection |
| Leggatt-Cook & Chamberlain, 2012[40] | Blogs | U.S. / Adults | Ten blogs | Social support | Weight-loss bloggers typically write about daily success and failures, report calorie consumption and exercise output, and post photographs of their changing bodies. | Data collection |

| | | | | | | |
|---|---|---|---|---|---|---|
| Mejova, 2018[60] | Twitter | U.S. / UNK. | 1.5 million tweets | Source credibility, Stigma | 1. Tweets afflicted with government or institution are likely to be retweeted more. 2. The need to address the quality control of health information on social media is proposed. | Data collection |
| Munk et al, 2016[13] | Instagram | UK / UNK. | 82,449 geotagged posts | Obesogenic environment | 1. Sunday night is a good time to post Instagram. 2. There is no clear difference between thematic communities between high and low BMI areas. | Data collection |
| Garimella et al, 2016[57] | Instagram | U.S. / UNK. | 200,000 images | Geo-cultural | Both user-provided and machine-generated image tags provide information that can be used to infer a county's health statistics. | Data collection |
| Culotta, 2014[56] | Twitter | U.S. / UNK. | 1.4M user-profiles and 4.3M tweets | Geo-cultural | 1. 6 of 27 health statistics show a significant correlation with the linguistic analysis of the Twitter activity in the top 100 most populous counties in the U.S. 2. Twitter information, together with demographic information, improves the model's performance. | Data collection |
| Abbar et al, 2015[49] | Twitter | U.S. / UNK. | 892,000 tweets | Geo-cultural | The caloric values of the foods mentioned in the tweets were analyzed in relation to the statewide obesity rate. | Data collection |
| Weber & Mejova, 2016[59] | Twitter | U.S. / Overweight adults | 1,339 profile images | Geo-cultural, Stigma | User profile pictures could be used to obtain the user's weight information. | Data collection |
| Pappa et al, 2017[41] | Reddit | U.S. / UNK. | posts and comments of 107,886 unique users | Social support, Gender | The ten most-discussed semantic topics on posts in the LoseIt Reddit community were related to healthy food, clothing, calorie counting, workouts, looks, habits, support, and unhealthy food. | Data collection |
| Loh et al, 2018[79] | Facebook Instagram Twitter | U.S. / Children | UNK. (intervention study) | Social support | The study showed that social media and text messaging were innovative tools that should be included to increase the reach of multilevel community intervention. | Intervention pathway |
| Ling et al, 2018[80] | Facebook | U.S. / Children | UNK. (intervention study) | Social support | Participants in the survey mentioned that they enjoyed the Facebook platform because it provided new recipe and activity ideas and an opportunity to interact with other participants. | Intervention pathway |
| He et al, 2017[43] | WeChat | China / Adults | UNK. (intervention study) | Social support | 1. An intervention based on WeChat platform was effective on weight loss only for males. 2. Females show more activities on WeChat, but they lost less weight during the study. | Intervention pathway |
| Erdem & Sisik, 2018[66] | YouTube | U.S. / Adults | 175 videos | Source credibility | 1. There are no significant associations between the number of 'likes', 'dislikes' or 'views' and usefulness score. 2. Videos uploaded by medical professionals typically contain more useful information. | Data collection |
| Jane et al, 2017[34] | Facebook | Australia / Adults with obesity / overweight | UNK. (intervention study) | Social support | 1. This study shows that participants don't rely on each other in the same way that they would typically rely on their offline social connections. 2. The Facebook group reported the greatest reductions in initial weight compared to the control group, which had no social media components. | Intervention pathway |
| Fiks et al, 2017[76] | Facebook | U.S. / Low-come mothers with newborn | UNK. (intervention study) | Social support | Mothers of the intervention group were significantly less likely to pressure infants to finish food or give cereal in the bottle. | Intervention pathway |

| | | | | | | |
|---|---|---|---|---|---|---|
| Mejova et al, 2015[58] | Instagram | U.S. / UNK. | 20,848,190 posts | Obesogenic environment, Social sharing | 1. There is a link between obesity and the density of fast-food restaurants. 2. Food sharing behavior is higher for high obesity areas. | Data collection |
| Cunha et al, 2017[39] | Reddit | U.S. / UNK. | 70,949 posts and 922,245 comments | Social support | 1. Users receiving feedback on their posts have a higher probability of returning to the community. 2. Returning users who received comments on their posts reported losing more weight. | Data collection |
| Waring et al, 2016[36] | Twitter | U.S. / Women of child-bearing age | UNK. (intervention study) | Gender, Social support | Women of child-bearing age are interested in a weight loss program that was delivered entirely via Twitter. | Intervention pathway |
| Chomutare et al, 2016[44] | UNK | U.S. / Women with obesity | 140 Women with obesity in an Internet group | Gender, Social support | Women with high online participation levels lost more weight than women with low participation levels. | Data collection |
| West et al, 2016[81] | Facebook | U.S. / Adults | UNK. (intervention study) | Social support | Students maintained their weight, with no significant difference between weight gain prevention intervention group and control group over nine weeks. | Intervention pathway |
| Aschbrenner et al, 2016[45] | Facebook | U.S. / Adults with serious mental illness | UNK. (intervention study) | Social support | This study shows that weight loss was significantly associated with perceived peer-group support. | Intervention pathway |
| Merchant et al, 2014[37] | Facebook | U.S. / Adults | UNK. (intervention study) | Social support | In a Facebook involved weight-loss controlled trial: 1. Polls are the most popular posts followed by photos. 2. Participants visibly engaged with posts less over time. 3. 40% of participants reported passively engaging with the Facebook page. | Intervention pathway |
| Chen et al, 2016[48] | Health-Together | Switzerland /Adults | UNK. (intervention study) | Social support | Collaborating with buddies to compete in achieving fitness goals in a group was reported as motivating for dyads with strong ties. | Intervention pathway |
| Phan et al, 2015[47] | Online social network | U.S. / Adults | UNK. (experiment study) | Obesogenic environment | By incorporating all the human behavior determinants and environmental events, the proposed novel deep learning model achieves more accurate results in predicting the future activity levels of users. | Ancillary resource |
| Savolainen, 2011[42] | Blogs | Finland / UNK. | 50 blogs | Social support | Blogs provide an emotionally supportive forum that mainly serves to share opinions and information; they were seldom used for seeking information. | Data collection |
| Church et al, 2018[82] | Facebook | UNK. / Adults | UNK. (intervention study) | Social support | Participants lose weight during the six-week online clinical, emotional freedom techniques course and continue to lose weight in the following year, which indicates the long-term effects. | Intervention pathway |
| Turner-McGrievy et al, 2014[83] | Facebook | U.S. / vegan women with polycystic ovary syndrome | UNK. (intervention study) | Social support | The study result suggests that engagement with social media may be effective for short-term weight loss among vegan women with PCOS. | Intervention pathway |
| Lytle et al, 2017[46] | Social support website | U.S. / 2-year college students | UNK. (intervention study) | Social support, School environment | The social networking encouraged intervention group, and the control group doesn't have a significant difference in BMI at the end of the 24-month intervention study. | Intervention pathway |

| | | | | | | |
|---|---|---|---|---|---|---|
| Waring et al, 2016[36] | Facebook | U.S. / Postpartum women with overweight / obesity | UNK. (intervention study) | Social support | Facebook-based intervention is feasible for overweight and postpartum women with obesity in weight loss. However, research is further needed to determine how to engage participants in social networks better. | Intervention pathway |
| Basch et al, 2017[67] | YouTube | U.S. / UNK. | 98 weight loss videos | Source credibility | The number of videos about weight loss on YouTube from professionals is lacking. | Data collection |
| Webb et al, 2017[72] | Instagram | U.S. / UNK. | 400 images | Social movement | Health at every size tagged posts contain more physically-active portrayals and weight stigma than posts from fitspiration-tagged images. | Data collections |
| Taiminen, 2016[84] | Facebook Online forum | Finland / UNK. | UNK. (intervention study) | Social support | Active participants in the online community showed a more positive perception of achieving their goals, followed instructions more precisely, and perceived to receive more emotional support than participants who are not active in the online community. | Intervention pathway |
| Hales et al, 2016[85] | Social POD | U.S. / overweight adults | UNK. (intervention study) | Social support | The experiment group using a weight-loss mobile app lost significantly more weight than the comparison group. | Intervention pathway |
| Meitz et al, 2016[68] | Facebook | Germany / children | UNK. (intervention study) | Source credibility | In an online media-embedded health campaign against childhood obesity: (1) Participant's self-relevance varies based on different source credibility perceptions. (2) Provocative messages in the campaign may result in negative persuasion effects. | Intervention pathway |
| Ghaznavi & Taylor, 2015[71] | Twitter Pinterest | UNK. / UNK. | 300 images | Social movements | The study suggests thinspiration content promotes an objectified, sexual, extremely thin depiction of the thin ideal. Exposure to these contents has the potential harmful effects. | Data collection |
| Appleton et al, 2014[86] | Online forums | Australia / UNK. | 34 discussion threads | Social support | Four major themes were detected in parents' online discussion forums about children obesity: seeking advice, sharing advice, social support, and making a judgment. | Data collection |
| Karami et al, 2018[61] | Twitter | U.S. / UNK. | 4.5 million tweets | Social movement | Exercise and obesity, diabetes and obesity, diet, and obesity have a strong correlation with each other. The strongest correlation was found between exercise and obesity. | Data collection |
| Swindle et al, 2018[69] | Facebook | U.S. / parents | UNK. (intervention study) | School environment | Facebook is a feasible platform to provide nutrition education and facilitate parent's engagement. | Intervention pathway |
| De et al, 2014[3] | Online message boards | Ireland / UNK. | 2872 obesity-relevant comments | Stigma | The study analyzed obesity-related comments from multi-topic online message boards and determined that obesity stigma is pervasive, and the discussion of the issue is highly acceptable. | Data collection |
| Gregg et al, 2017[74] | Online forums | UK / UNK. | 1,704 comments | Policy | The study analyzed associated comments to the UK government about childhood obesity strategy and determined the comments are largely negative. | Data collection |
| Atanasova, 2018[87] | Blogs | UK / UNK. | 343 posts from 6 obesity blogs | Social support | The content of blogs highlighted the conclusion that there are no one-size-fits-all solutions to obesity that work for everyone. | Data collection |
| Cohen et al, 2019[53] | Instagram | UNK. / UNK. | 630 posts | Social movements | Body positive posts depicted a broad range of body sizes and appearances. | Data collection |



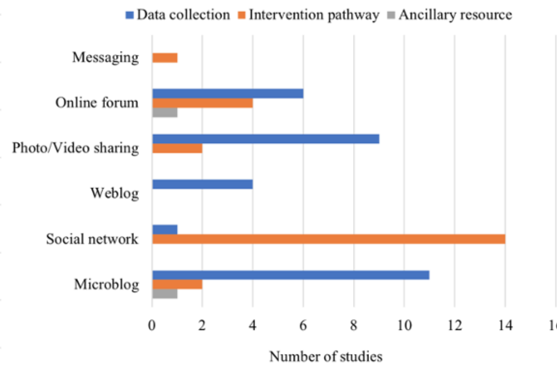




Figure 2.4: The distribution of social media platform's three roles.

## 2.4 Discussion

In our systematic review, we categorized and identified related the effects of online social factors on users' obesity-related behaviors and evaluated the role of social media platforms. Here, we showed the adopted social-ecological model to explain identified online social factors from different levels. Further, we discussed strategies for preventing obesity by utilizing the social-ecological model. We concluded the drawbacks we found in literature studies and provide suggestions for future studies.

### 2.4.1 Social-ecological model

Socio-ecological models were developed to further the understanding of the dynamic interrelations among various personal and environmental factors.[88] Revised by Bronfenbrenner, Bronfenbrenner's ecological theory applies social-ecological models to human development. The ecological framework identifies five environmental systems with which an individual interacts: microsystem, mesosystem, exosystem, macrosystem, and chronosystem.[89] Since its publication in 1979, Bronfenbrenner's major statement on the theory, the of the Ecology of Human Development, has shown widespread influence on the way psychologists and others approach the study of human beings and their environments. This social-ecological model here is proposed to understand how online social factors affect behaviors and provide guidance for developing a successful program through online social environments.

We classified those online social factors into four levels based on their effects: individual, interpersonal, online social environment, and connection to the real world. The proposed social-ecological model is shown in (Figure 2.5).

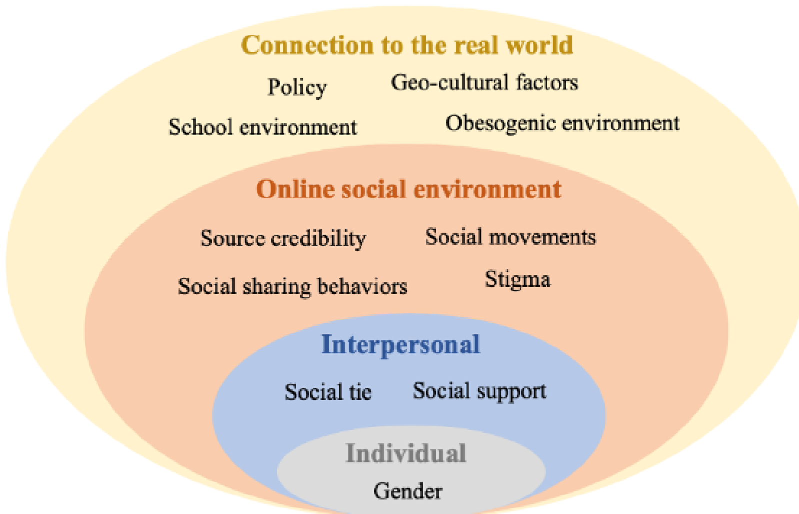



Figure 2.5: Social-ecological model.

Our aim is to identify online social factors, which can affect users' obesity-related behaviors. For the online environment, we use a four-level social-ecological model to better understand obesity and the effect of potential factors to help users combat obesity. This model considers the complex reciprocity between individual, interpersonal, online community, social media platforms, and connection to the real world. It allows us to understand the range of factors that potentially affect the user's online behaviors related to obesity. The overlapping rings in the model illustrate how factors at one level influence factors at another level. Besides helping to clarify the effectiveness of these factors, the model also suggests that in order to intervene in a user's behavior, it's necessary to act across multiple levels of the model at the same time.

The first level identifies personal factors that affect a person's online behaviors. The factor identified here is gender. From a biological perspective, women's bodies are more vulnerable to obesity due to the fact that women are more likely to store fat because of reproduction.[90] From the online environment, female users were found more attentive to their shape and figure than men and are more likely to search and share health information online through social media than men.

The second level explores relationships that may increase or reduce the risk of obesity. People's close social connections or family members influences influence their behaviors and contribute to their habits. The factors we discovered at this level are social support and online social ties. Social support from social media sites has been suggested to be very effective in users' weight loss experiences and online social tie has been proved to influence person's lifestyle behaviors.

The third level examines the online social environment in which social relationships occur and characteristics are associated with users' obesity-related behaviors. Factors that contribute to the online environment include source credibility, social movements, social sharing behaviors, and stigma. Social movements, stigma, and source credibility shape user's behavior by changing the online environment. Exposure to content with obesity-associated stigma has been proved to have negative effects on user's mental health. Exposure to content coming from unreliable resources may harm users and further damage user's trust in social media platforms. Moreover, exposure to some social movements may negatively affect users' behaviors. For example, exposure to content about 'thinspiration' will increase a person's body dissatisfaction and negative moods. Social sharing behaviors change the online environment by changing the user's sharing behavior directly. Users show a preference for sharing emotionally evocative content, especially when they are consistent with their beliefs.

The fourth level explores the broader societal factors that connect the virtual online environment to the real world. Factors in this level include policy, school environment, geo-cultural factors, and obesogenic environment. The physical environment can inform virtual online environment. For example, the number of fast-food restaurants and the calories of food mentioned in tweets correlated to the county's obesity rate.

User's unsatisfactory opinions of school environments and to government policies are found through social media data.

### 2.4.2 Preventing obesity

Reducing the obesity rate requires understanding the factors that influence obesity-related behaviors. This social-ecological model helps practitioners to develop effective prevention strategies.

Preventative strategies at an individual level promote attitudes, beliefs, and behaviors that combat obesity. The approach may include advocating males to care more about their weight status and be aware of the importance of having a healthy lifestyle. Female opinion leaders could be encouraged to share more healthy lifestyle tips to help women's maintaining a healthy lifestyle.

A prevention strategy at the interpersonal level may include designing family-focused weight loss programs, promoting users who want to combat obesity to join related online groups, share their weight loss/weight management experience online, and give encouragement to other users.A prevention strategy at the interpersonal level may include designing family-focused weight loss programs, promoting users who want to combat obesity to join related online groups, share their weight loss/weight management experience online, and give encouragement to other users.

Preventative strategies at the social environment level make a point of creating a healthier online social environment. For example, improving the source credibility level by monitoring illegal advertisement and advocating professional organizations to post strategies with combating obesity; coping with stigma issues by adopting state-of-art natural language processing technologies to remove stigma posts from social media; supervising the online environment by detecting major social movements that intersect with obesity; building users' healthy life beliefs by encouraging positive social sharing on social media. Emotionally evocative posts are more easily accepted depending on a user's social sharing behaviors.

Preventative strategies related to the real-world level accentuates the importance of building a healthy societal environment and set up a good connection between the online environment and real society. Social policies from the government that leading a healthy lifestyle, a good school environment that providing children with a balanced diet, and an anti-obesogenic environment can help to maintain a healthy societal environment.

### 2.4.3 Data Variety

We have observed some drawbacks in the included studies. The data variety needs to be expanded in future studies. First, the limitations of each social media platform should be considered. Although Twitter is one of the biggest online social network platforms, Twitter may not serve as the best channel to collect data and study obesity-related topics. Twitter is a platform that makes data publicly available. Due to privacy and stigma concerns, some users may refuse to share their confidential data on Twitter.[79] Comparing three major social media platforms (Facebook, Instagram, and Twitter) in a childhood obesity prevention intervention by Loh et al.[79], Facebook, allowing more comprehensive communication, longer and more frequent posts than the other two social media platforms, was found to have the highest fidelity and engagement. On the contrary, Twitter shows the least engagement and fidelity.[79] Second, a large amount of image data is an emerging resource. We discovered that textual data was the leading type in prior studies, and a large amount of media-syncretic image data was dismissed. There is one study from Mejova et al.[58] that analyzed picture tag together with the images. Textual data, together with its associated features (e.g., image, link, user profile), could give us more insight. Third, regions of studies were too limited. The prevalence of the obesity rate is also high in other areas in the world, like Mexico[91]. Given the cultural differences, it is meaningful to understand social elements of other areas. Another finding is the data collected through the experiment is not analyzed. Almost all intervention studies encourage participants to interact with others on the social media platform; the efficacy of those social components did not receive adequate study. Only two studies performed quantitative analysis on user's interaction behavior (e.g., how many posts the user submits every day)[43],[48] in an online support group. No qualitative analysis was employed on the textual data collected in the study, which could give us a clue towards how factors affect people's weight-loss experience.

Lastly, the data quantity in many studies was not sufficient, considering the large group of people with obesity. For example, two studies using blog data to do qualitative analysis used just ten[40] and thirteen[52] blogs. Manually conducting a thematic analysis is indeed labor-intensive, however, with the development of deep learning, pre-trained language models could be effectively employed to analyze large amounts of data.

### 2.4.4 Limitation

Our systematic review has some limitations. First, we only used three databases (PubMed, ACM, and ScienceDirect) in our study. If other databases, such as PsycINFO, EMBASE, and Scopus, were included in the study, we might have additional and possibly different findings. Second, the MeSH term "social media" was created in 2012[31]; thus, our study did not include studies published before 2012. This could have skewed the results. Future studies may consider a broader search strategy for more comprehensive results. Third, many studies didn't include the particular type of social factor and how those factors affect users; thus, the analysis of social factors is not sufficient. Finally, further discussion on the quality of study design, types of bias, and other limitations of investigated studies can bolster our findings.

## 2.5 Conclusion

We provided a comprehensive review of social media in relation to understanding obesity and isolating online social factors, including the platforms, data, and study results. We proposed a four-level social-ecological model to explain the dynamic interrelationships among users' obesity-related behaviors, personal characteristics, user's interpersonal connections, online social environments, and the real world. Understanding the potential and role of these factors will benefit us in several aspects: understanding users' online social behaviors concerning obesity, calibrating online social factors for weight management intervention studies and disseminating educational information to the public.

## REFERENCES


[1] "Adult Obesity Facts.." https://web.archive.org/web/20201005135332/https://www.cdc.gov/obesity/data/adult.html. [Online; Accessed: 2020-10-05].

[2] C. M. Hales, M. D. Carroll, C. D. Fryar, and C. L. Ogden, "Prevalence of obesity among adults and youth: United states, 2015–2016," 2017.

[3] A. De Brún, M. McCarthy, K. McKenzie, and A. McGloin, "Weight stigma and narrative resistance evident in online discussions of obesity," *Appetite*, vol. 72, pp. 73–81, 2014.

[4] D. S. Freedman, W. H. Dietz, S. R. Srinivasan, and G. S. Berenson, "The relation of overweight to cardiovascular risk factors among children and adolescents: the bogalusa heart study," *Pediatrics*, vol. 103, no. 6, pp. 1175–1182, 1999.

[5] D. R. Meldrum, M. A. Morris, and J. C. Gambone, "Obesity pandemic: causes, consequences, and solutionsâbut do we have the will?," *Fertility and sterility*, vol. 107, no. 4, pp. 833–839, 2017.

[6] S. J. Olshansky, D. J. Passaro, R. C. Hershow, J. Layden, B. A. Carnes, J. Brody, L. Hayflick, R. N. Butler, D. B. Allison, and D. S. Ludwig, "A potential decline in life expectancy in the united states in the 21st century," *New England Journal of Medicine*, vol. 352, no. 11, pp. 1138–1145, 2005.

[7] J. B. Schwimmer, T. M. Burwinkle, and J. W. Varni, "Health-related quality of life of severely obese children and adolescents," *Jama*, vol. 289, no. 14, pp. 1813–1819, 2003.

[8] S. J. Woolford, A. Gebremariam, S. J. Clark, and M. M. Davis, "Incremental hospital charges associated with obesity as a secondary diagnosis in children," *Obesity*, vol. 15, no. 7, pp. 1895–1901, 2007.

[9] E. A. Finkelstein, J. G. Trogdon, J. W. Cohen, and W. Dietz, "Annual medical spending attributable to obesity: Payer-and service-specific estimates: Amid calls for health reform, real cost savings are more likely to be achieved through reducing obesity and related risk factors.," *Health affairs*, vol. 28, no. Suppl1, pp. w822–w831, 2009.

[10] W. H. Organization, "Obesity: preventing and managing the global epidemic," 2000.

[11] F. C. Pampel, P. M. Krueger, and J. T. Denney, "Socioeconomic disparities in health behaviors," *Annual review of sociology*, vol. 36, pp. 349–370, 2010.

[12] S. Easton, K. Morton, Z. Tappy, D. Francis, L. Dennison, *et al.*, "Young peopleâs experiences of viewing the fitspiration social media trend: Qualitative study," *Journal of medical Internet research*, vol. 20, no. 6, p. e9156, 2018.

[13] A. K. Munk, M. S. Abildgaard, A. Birkbak, and M. K. Petersen, "(re-) appropriating instagram for social research: Three methods for studying obesogenic environments," in *Proceedings of the 7th 2016 International Conference on Social Media & Society*, pp. 1–10, 2016.

[14] J. A. Corsica and M. M. Hood, "Eating disorders in an obesogenic environment," *Journal of the American Dietetic Association*, vol. 111, no. 7, pp. 996–1000, 2011.

[15] J. O. Hill and J. C. Peters, "Environmental contributions to the obesity epidemic," *Science*, vol. 280, no. 5368, pp. 1371–1374, 1998.

[16] K. P. Smith and N. A. Christakis, "Social networks and health," *Annu. Rev. Sociol*, vol. 34, pp. 405–429, 2008.

[17] D. Umberson, R. Crosnoe, and C. Reczek, "Social relationships and health behavior across the life course," *Annual review of sociology*, vol. 36, pp. 139–157, 2010.

[18] N. A. Christakis and J. H. Fowler, "The spread of obesity in a large social network over 32 years," *New England journal of medicine*, vol. 357, no. 4, pp. 370–379, 2007.

[19] K. Maximova, J. J. McGrath, T. Barnett, J. O'Loughlin, G. Paradis, and M. Lambert, "Do you see what i see? weight status misperception and exposure to obesity among children and adolescents," *International journal of obesity*, vol. 32, no. 6, pp. 1008–1015, 2008.

[20] M. M. Ali, A. Amialchuk, and F. Renna, "Social network and weight misperception among adolescents," *Southern Economic Journal*, vol. 77, no. 4, pp. 827–842, 2011.

[21] D. Centola, "Social media as a tool in medicine," *Circulation*, vol. 127, pp. 2135–2144, 2013.

[22] E. E. Kent, A. Prestin, A. Gaysynsky, K. Galica, R. Rinker, K. Graff, and W.-Y. S. Chou, ""obesity is the new major cause of cancer": connections between obesity and cancer on facebook and twitter," *Journal of Cancer Education*, vol. 31, no. 3, pp. 453–459, 2016.

[23] D. M. Boyd and N. B. Ellison, "Social network sites: Definition, history, and scholarship," *Journal of computer-mediated Communication*, vol. 13, no. 1, pp. 210–230, 2007.

[24] P. R. Center, "10 facts about americans and facebook," 2019.

[25] T. Chang, V. Chopra, C. Zhang, and S. J. Woolford, "The role of social media in online weight management: systematic review," *Journal of medical Internet research*, vol. 15, no. 11, p. e262, 2013.

[26] L. Laranjo, A. Arguel, A. L. Neves, A. M. Gallagher, R. Kaplan, N. Mortimer, G. A. Mendes, and A. Y. Lau, "The influence of social networking sites on health behavior change: a systematic review and meta-analysis," *Journal of the American Medical Informatics Association*, vol. 22, no. 1, pp. 243–256, 2015.

[27] M. E. Waring, D. E. Jake-Schoffman, M. M. Holovatska, C. Mejia, J. C. Williams, and S. L. Pagoto, "Social media and obesity in adults: a review of recent research and future directions," *Current diabetes reports*, vol. 18, no. 6, pp. 1–9, 2018.

[28] M. E. Falagas, E. I. Pitsouni, G. A. Malietzis, and G. Pappas, "Comparison of pubmed, scopus, web of science, and google scholar: strengths and weaknesses," *The FASEB journal*, vol. 22, no. 2, pp. 338–342, 2008.

[29] "Association for Computing Machinery." `https://web.archive.org/web/20201006091320/https://www.acm.org/`. [Online; Accessed: 2020-10-06].

[30] "ScienceDirect." `https://web.archive.org/web/20201006025225/https://www.sciencedirect.com/`. [Online; Accessed: 2020-10-06].

[31] "Social Media MeSH Descriptor Data 2020." `https://meshb.nlm.nih.gov/record/ui?ui=D061108`. [Online; Accessed: 2020-10-06].

[32] A. M. Kaplan and M. Haenlein, "Users of the world, unite! the challenges and opportunities of social media," *Business horizons*, vol. 53, no. 1, pp. 59–68, 2010.

[33] E. A. Willis, A. N. Szabo-Reed, L. T. Ptomey, F. L. Steger, J. J. Honas, E. M. Al-Hihi, R. Lee, L. Vansaghi, R. A. Washburn, and J. E. Donnelly, "Distance learning strategies for weight management utilizing social media: A comparison of phone conference call versus social media platform. rationale and design for a randomized study," *Contemporary clinical trials*, vol. 47, pp. 282–288, 2016.

[34] M. Jane, M. Hagger, J. Foster, S. Ho, R. Kane, and S. Pal, "Effects of a weight management program delivered by social media on weight and metabolic syndrome risk factors in overweight and obese adults: a randomised controlled trial," *PLoS One*, vol. 12, no. 6, p. e0178326, 2017.

[35] K. O. Hwang, A. J. Ottenbacher, A. P. Green, M. R. Cannon-Diehl, O. Richardson, E. V. Bernstam, and E. J. Thomas, "Social support in an internet weight loss community," *International journal of medical informatics*, vol. 79, no. 1, pp. 5–13, 2010.

[36] M. E. Waring, K. L. Schneider, B. M. Appelhans, T. A. M. Simas, R. S. Xiao, M. C. Whited, A. M. Busch, M. M. Evans, and S. L. Pagoto, "Interest in a twitter-delivered weight loss program among women of childbearing age," *Translational behavioral medicine*, vol. 6, no. 2, pp. 277–284, 2016.

[37] G. Merchant, N. Weibel, K. Patrick, J. H. Fowler, G. J. Norman, A. Gupta, C. Servetas, K. Calfas, K. Raste, L. Pina, *et al.*, "Click "like" to change your behavior: a mixed methods study of college studentsâ exposure to and engagement with facebook content designed for weight loss," *Journal of medical Internet research*, vol. 16, no. 6, p. e3267, 2014.

[38] M. E. Waring, T. A. M. Simas, J. Oleski, R. S. Xiao, J. A. Mulcahy, C. N. May, and S. L. Pagoto, "Feasibility and acceptability of delivering a postpartum weight loss intervention via facebook: a pilot study," *Journal of nutrition education and behavior*, vol. 50, no. 1, pp. 70–74, 2018.

[39] T. Cunha, I. Weber, and G. Pappa, "A warm welcome matters! the link between social feedback and weight loss in/r/loseit," in *Proceedings of the 26th International Conference on World Wide Web Companion*, pp. 1063–1072, 2017.

[40] C. Leggatt-Cook and K. Chamberlain, "Blogging for weight loss: personal accountability, writing selves, and the weight-loss blogosphere," *Sociology of health & illness*, vol. 34, no. 7, pp. 963–977, 2012.

[41] G. L. Pappa, T. O. Cunha, P. V. Bicalho, A. Ribeiro, A. P. C. Silva, W. Meira Jr, and A. M. R. Beleigoli, "Factors associated with weight change in online weight management communities: a case study in the loseit reddit community," *Journal of medical Internet research*, vol. 19, no. 1, p. e17, 2017.

[42] R. Savolainen, "Asking and sharing information in the blogosphere: The case of slimming blogs,"

[43] C. He, S. Wu, Y. Zhao, Z. Li, Y. Zhang, J. Le, L. Wang, S. Wan, C. Li, Y. Li, *et al.*, "Social media–promoted weight loss among an occupational population: Cohort study using a wechat mobile phone app-based campaign," *Journal of medical Internet research*, vol. 19, no. 10, p. e357, 2017.

[44] T. Chomutare, E. Årsand, and G. Hartvigsen, "Effectiveness of an internet community for severely obese women," in *Nursing Informatics 2016*, pp. 597–601, IOS Press, 2016.

[45] K. A. Aschbrenner, J. A. Naslund, M. Shevenell, E. Kinney, and S. J. Bartels, "A pilot study of a peer-group lifestyle intervention enhanced with mhealth technology and social media for adults with serious mental illness," *The Journal of nervous and mental disease*, vol. 204, no. 6, p. 483, 2016.

[46] L. A. Lytle, M. N. Laska, J. A. Linde, S. G. Moe, M. S. Nanney, P. J. Hannan, and D. J. Erickson, "Weight-gain reduction among 2-year college students: the

choices rct," *American journal of preventive medicine*, vol. 52, no. 2, pp. 183–191, 2017.

[47] N. Phan, D. Dou, B. Piniewski, and D. Kil, "Social restricted boltzmann machine: Human behavior prediction in health social networks," in *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, pp. 424–431, 2015.

[48] Y. Chen, M. Randriambelonoro, A. Geissbuhler, and P. Pu, "Social incentives in pervasive fitness apps for obese and diabetic patients," in *Proceedings of the 19th ACM conference on computer supported cooperative work and social computing companion*, pp. 245–248, 2016.

[49] S. Abbar, Y. Mejova, and I. Weber, "You tweet what you eat: Studying food consumption through twitter," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 3197–3206, 2015.

[50] M. Kuebler, E. Yom-Tov, D. Pelleg, R. M. Puhl, and P. Muennig, "When overweight is the normal weight: an examination of obesity using a social media internet database," *PLoS One*, vol. 8, no. 9, p. e73479, 2013.

[51] L. Fernandez-Luque, M. Singh, F. Ofli, Y. A. Mejova, I. Weber, M. Aupetit, S. K. Jreige, A. Elmagarmid, J. Srivastava, and M. Ahmedna, "Implementing 360 quantified self for childhood obesity: feasibility study and experiences from a weight loss camp in qatar," *BMC medical informatics and decision making*, vol. 17, no. 1, pp. 1–13, 2017.

[52] L. Lingetun, M. Fungbrant, M. Claesson, and C. Baggens, "âi just want to be normalâ–a qualitative study of pregnant women's blogs who present themselves as overweight or obese," *Midwifery*, vol. 49, pp. 65–71, 2017.

[53] R. Cohen, L. Irwin, T. Newton-John, and A. Slater, "bodypositivity: A content analysis of body positive accounts on instagram," *Body image*, vol. 29, pp. 47–57, 2019.

[54] H. Brown, *Body of truth: How science, history, and culture drive our obsession with weight–and what we can do about it.* Da Capo Lifelong Books, 2015.

[55] R. J. Gore, S. Diallo, and J. Padilla, "You are what you tweet: connecting the geographic variation in americaâs obesity rate to twitter content," *PloS one*, vol. 10, no. 9, p. e0133505, 2015.

[56] A. Culotta, "Estimating county health statistics with twitter," in *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1335–1344, 2014.

[57] V. R. K. Garimella, A. Alfayad, and I. Weber, "Social media image analysis for public health," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 5543–5547, 2016.

[58] Y. Mejova, H. Haddadi, A. Noulas, and I. Weber, "foodporn: Obesity patterns in culinary interactions," in *Proceedings of the 5th International Conference on Digital Health 2015*, pp. 51–58, 2015.

[59] I. Weber and Y. Mejova, "Crowdsourcing health labels: inferring body weight from profile pictures," in *Proceedings of the 6th international conference on digital health conference*, pp. 105–109, 2016.

[60] Y. Mejova, "Information sources and needs in the obesity and diabetes twitter discourse," in *Proceedings of the 2018 international conference on digital health*, pp. 21–29, 2018.

[61] A. Karami, A. A. Dahl, G. Turner-McGrievy, H. Kharrazi, and G. Shaw Jr, "Characterizing diabetes, diet, exercise, and obesity comments on twitter," *International Journal of Information Management*, vol. 38, no. 1, pp. 1–6, 2018.

[62] J. So, A. Prestin, L. Lee, Y. Wang, J. Yen, and W.-Y. S. Chou, "What do people like to "share" about obesity? a content analysis of frequent retweets about obesity on twitter," *Health communication*, vol. 31, no. 2, pp. 193–206, 2016.

[63] C. N. May, M. E. Waring, S. Rodrigues, J. L. Oleski, E. Olendzki, M. Evans, J. Carey, and S. L. Pagoto, "Weight loss support seeking on twitter: the impact of weight on follow back rates and interactions," *Translational behavioral medicine*, vol. 7, no. 1, pp. 84–91, 2017.

[64] A. Lake and T. Townshend, "Obesogenic environments: exploring the built and food environments," *The Journal of the Royal society for the Promotion of Health*, vol. 126, no. 6, pp. 262–267, 2006.

[65] J. K. Harris, S. Moreland-Russell, R. G. Tabak, L. R. Ruhr, and R. C. Maier, "Communication about childhood obesity on twitter," *American journal of public health*, vol. 104, no. 7, pp. e62–e69, 2014.

[66] H. Erdem and A. Sisik, "The reliability of bariatric surgery videos in youtube platform," *Obesity surgery*, vol. 28, no. 3, pp. 712–716, 2018.

[67] C. H. Basch, I.-H. Fung, A. Menafro, C. Mo, and J. Yin, "An exploratory assessment of weight loss videos on youtubeâ¢," *Public health*, vol. 151, pp. 31–38, 2017.

[68] T. G. Meitz, A. Ort, A. Kalch, S. Zipfel, and G. Zurstiege, "Source does matter: Contextual effects on online media-embedded health campaigns against childhood obesity," *Computers in Human Behavior*, vol. 60, pp. 565–574, 2016.

[69] T. M. Swindle, W. L. Ward, and L. Whiteside-Mansell, "Facebook: the use of social media to engage parents in a preschool obesity prevention curriculum," *Journal of nutrition education and behavior*, vol. 50, no. 1, pp. 4–10, 2018.

[70] "Social movements definition." `https://web.archive.org/web/20201006123653/https://open.lib.umn.edu/sociology/chapter/21-3-social-movements/`. [Online; Accessed: 2020-10-06].

[71] J. Ghaznavi and L. D. Taylor, "Bones, body parts, and sex appeal: An analysis of# thinspiration images on popular social media," *Body image*, vol. 14, pp. 54–61, 2015.

[72] J. B. Webb, E. R. Vinoski, A. S. Bonar, A. E. Davies, and L. Etzel, "Fat is fashionable and fit: A comparative content analysis of fatspiration and health at every size® instagram images," *Body image*, vol. 22, pp. 53–64, 2017.

[73] Y. Kang, Y. Wang, D. Zhang, and L. Zhou, "The publicâs opinions on a new school meals policy for childhood obesity prevention in the us: A social media analytics approach," *International journal of medical informatics*, vol. 103, pp. 83–88, 2017.

[74] R. Gregg, A. Patel, S. Patel, and L. OâConnor, "Public reaction to the uk government strategy on childhood obesity in england: A qualitative and quantitative summary of online reaction to media reports," *Health Policy*, vol. 121, no. 4, pp. 450–457, 2017.

[75] B. Rimé, C. Finkenauer, O. Luminet, E. Zech, and P. Philippot, "Social sharing of emotion: New evidence and new questions," *European review of social psychology*, vol. 9, no. 1, pp. 145–189, 1998.

[76] G. FiksAlexander, S. GruverRachel, T. Bishop-GilyardChanelle, W. SuhAndrew, K. KalraGurpreet, A. DeRussoPatricia, A. ElovitzMichal, I. BerkowitzRobert, J. PowerThomas, *et al.*, "A social media peer group for mothers to prevent obesity from infancy: the grow2gether randomized trial," *Childhood Obesity*, 2017.

[77] J. S. Leroux, S. Moore, and L. Dubé, "Beyond the "i" in the obesity epidemic: a review of social relational and network interventions on obesity," *Journal of obesity*, vol. 2013, 2013.

[78] A. Park, M. Conway, and A. T. Chen, "Examining thematic similarity, difference, and membership in three online mental health communities from reddit: a text mining and visualization approach," *Computers in human behavior*, vol. 78, pp. 98–112, 2018.

[79] I. H. Loh, T. Schwendler, A. C. Trude, E. T. Anderson Steeves, L. J. Cheskin, S. Lange, and J. Gittelsohn, "Implementation of text-messaging and social media strategies in a multilevel childhood obesity prevention intervention: process evaluation results," *INQUIRY: The Journal of Health Care Organization, Provision, and Financing*, vol. 55, p. 0046958018779189, 2018.

[80] J. Ling, L. B. Robbins, N. Zhang, J. M. Kerver, H. Lyons, N. Wieber, and M. Zhang, "Using facebook in a healthy lifestyle intervention: Feasibility and preliminary efficacy," *Western journal of nursing research*, vol. 40, no. 12, pp. 1818–1842, 2018.

[81] D. S. West, C. M. Monroe, G. M. Turner-McGrievy, B. Sundstrom, C. Larsen, K. Magradey, S. Wilcox, and H. M. Brandt, "A technology-mediated behavioral weight gain prevention intervention for college students: controlled, quasi-experimental study," *Journal of medical Internet research*, vol. 18, no. 6, p. e133, 2016.

[82] D. Church, P. Stapleton, L. Sheppard, and B. Carter, "Naturally thin you: Weight loss and psychological symptoms after a six-week online clinical eft (emotional freedom techniques) course," *Explore*, vol. 14, no. 2, pp. 131–136, 2018.

[83] G. M. Turner-McGrievy, C. R. Davidson, E. E. Wingard, and D. L. Billings, "Low glycemic index vegan or low-calorie weight loss diets for women with polycystic ovary syndrome: a randomized controlled feasibility study," *Nutrition research*, vol. 34, no. 6, pp. 552–558, 2014.

[84] H. Taiminen, "How do online communities matter? comparison between active and non-active participants in an online behavioral weight loss program," *Computers in Human Behavior*, vol. 63, pp. 787–795, 2016.

[85] S. Hales, G. M. Turner-McGrievy, S. Wilcox, A. Fahim, R. E. Davis, M. Huhns, and H. Valafar, "Social networks for improving healthy weight loss behaviors for overweight and obese adults: a randomized clinical trial of the social pounds off digitally (social pod) mobile app," *International journal of medical informatics*, vol. 94, pp. 81–90, 2016.

[86] J. Appleton, C. Fowler, and N. Brown, "Friend or foe? an exploratory study of australian parentsâ use of asynchronous discussion boards in childhood obesity," *Collegian*, vol. 21, no. 2, pp. 151–158, 2014.

[87] D. Atanasova, ""keep moving forward. left right left": A critical metaphor analysis and addressivity analysis of personal and professional obesity blogs," *Discourse, context & media*, vol. 25, pp. 5–12, 2018.

[88] U. Bronfenbrenner, *Ecological systems theory.* Jessica Kingsley Publishers, 1992.

[89] U. Bronfenbrenner, "The biological theory of human development," *Making human beings human: Biological perspectives on human development*, pp. 3–15, 2005.

[90] Y. Lee, "Slender women and overweight men: gender differences in the educational gradient in body weight in south korea," *International journal for equity in health*, vol. 16, no. 1, pp. 1–18, 2017.

[91] K. Rtveladze, T. Marsh, S. Barquera, L. M. S. Romero, D. Levy, G. Melendez, L. Webber, F. Kilpi, K. McPherson, and M. Brown, "Obesity prevalence in mexico: impact on health and economic burden," *Public health nutrition*, vol. 17, no. 1, pp. 233–239, 2014.

# CHAPTER 3: A Novel Approach to characterize State Level food environment and Predict Obesity Rate Using Social Media Data

## 3.1 Introduction

The current obesity epidemic poses critical public health challenges in today's society. Obesity is a major risk factor for other chronic diseases, such as cardiovascular disease, cancer, diabetes, and respiratory disorders, which account for 60% of deaths worldwide.[1] Excessive body weight has resulting in $100 billion per year in medical expenditure.[2, 3] In 2017-2018, the prevalence of obesity in the US in adults was 42.4%.[4] This number has more than tripled since the 1960s. In the years 1960-1962, obesity rate was 13.4%.[5]

Environmental factors, including types of available food, have been identified as one of the main drivers of obesity.[6, 7, 3] American adults have developed a preference for dining out with friends as opposed to cooking at home.[8] A market research survey conducted in 2017 found that those who often frequent fast food restaurants are more concerned about value for money spent and service speed than the actual healthiness of the food offered.[8] This indication that perceived food availability tends to affect dietary outcomes has only been furthered in a literature review conducted by Caspi et al.[9] Those who live in areas highly saturated with high fat food items, tend to have health issues. Additionally, those who live in lower income areas are more likely to have at least one dietary related health issue.[9] In the US, people tend to eat what is affordable and available to them. Environments littered with low-cost, high fat foods have the tendency to be obesogenic. With food expenditures for dining out increasing in recent years, [3, 10] understanding the food Environmental factors are critical in preventing obesity pandemic and understanding related human behavior.

This notion is supported by recent studies that suggest the local food environment can be measured by the degree of food accessibility.[6, 11] These studies measured food accessibility with survey data,[12] yellow pages phone books,[13, 14] and local business directories.[15] A limited number of samples and a significant delay between collecting and reporting are major limitations of these traditional methods.[9] With the proliferation of social media, data from social media is organic, continuously updated, and generally free for large scale collection. We leverage a large-scale social media datasets to measure the food environment at the state level to predict state-level obesity rates. The obesity rate was obtained from the Behavioral Risk Factor Surveillance System (BRFSS), the nation's premier system to collect data to improve public health.[16] The primary aim of this study is to understand the impact of a food environment on obesity with three specific research questions.

- **RQ1**: Are there food availability difference in states with low and high prevalence of obesity? Is there a difference between states with low and high prevalence of obesity, when it comes to the availability of food options for state inhabitants?
- **RQ2**: How to quantify state-level food environment using calorie information?
- **RQ3**: Can we predict state-level obesity rate using publicly available social media data?

We reported our novel approaches and findings. To date, our study is the first to combine information from Yelp and MyFitnessPal (MFP) to learn about the local food environment and then to predict state level obesity rate.

## 3.2 Related work

### 3.2.1 Calorie with obesity

An increase in daily calorie consumption is regarded to be a major cause of the pandemic.[7] An increase in daily calorie intake rose by more than 500 calories in adults and more than 150 calories in children between 1977 and 2006 [17, 18] as did the portion size in restaurants.[19]

Exposure to a larger portion size could increase the risk of increasing calorie intake and weight gain.[20] Calorie intake is also affected by the access to local dining options. The prevalence of obesity was lower in areas that had supermarkets and higher in area with small grocery stores or an increased number of fast food restaurants.[12]

Analysis of data on environmental changes has identified the changes as a potential cause for increase in caloric intake. The enormous growth in dining out, particularly at "fast food" outlets, is one trend that has gotten a lot of attention. Fast food outlets have increased about 30,000 in 1970 to more than 233,000 locations in the USA in 2004.[3] Fast food can contribute to increasing obesity rate, because it generally provides food that is poor in micro-nutrients, low in fiber, high in glycemic load, and excessive in portion size and calorie.[21, 22]

### 3.2.2 How to characterize/quantify local food environment

Measures of food environment can be generally grouped into three categories: (i) the community food environment; (ii) the organizational nutrition environment; and (iii) the consumer food environment.[23] One way of conceptualizing food access dimensions is by adapting *a concept of access* proposed by Penchansky and Thomas [24]. The *concept of access* uses 5 dimensions to conceptualize the local food environment, including availability, accessibility, affordability, acceptability, and accommodation.[9] Availability refers to the relationship of the number and the type of food suppliers available to the customer. Accessibility refers to the relationship between the location of food suppliers and the location of customers, which is more geographically inherent than availability. Accessibility could be measured by travel time and distance between food supplier and customers. Affordability refers to the price customers need to pay for the food. Acceptability refers to the customers' attitude towards a business. Accommodation is another dimension of access, which assesses whether local businesses accept and adapt to local customers' needs.

A variety of approaches have been utilized to learn about local food environments by measuring the degree of food access. Approaches typically fall into two categories. The first category consists of methods that capture the food environment by relying on respondent-based data. The accessibility of food stores was asked in surveys or questionnaires. The methods in the second category uses the Geographic Information System (GIS) technology. It measures the buffer distance to food stores or the density of food stores in an area.[14, 15, 12, 13] By 2007, GIS-based measures of food environment outnumbered respondent-based measures, and the trends of using GIS measures continue.[9, 25, 26] The GIS data used in previous studies primarily use publicly available datasets, such as the U.S. Yellow pages phone book,[13, 14] published data from the local Departments of Environmental Health and state Departments of Agriculture,[12, 27] and local business directories.[15] A major limitation of these traditional data collections is that they are cost-ineffective and labor-intensive and the methods can only gather a limited number of samples and the significant delay between collecting and reporting.[9]

### 3.2.3 Using social media data to learn food environment or predict the obesity rate

Social media can be used to create indicators of the food environment. Nguyen et al.[28] characterized the food environment by calculating the calorie density of foods mentioned in tweets and the percent of each food theme out of all food-related Yelp entries for that state. They found that Twitter and Yelp characteristics were indicative of higher caloric foods were related to higher mortality, higher prevalence of chronic conditions, and worse self-rated health. [28] Social media can also be used to predict obesity rate. Fried et al.[29] presented the predicted power behind the language of food on social media. They collected the food-related tweets that contained meal-related hashtags: dinner, breakfast, lunch, brunch, snack, meal, and supper.

And then use the lexical feature from the bag-of-words model and topic features obtained from Latent Dirichlet Allocation (LDA) to predict whether a state's rate of obesity is above or below the national median. Their best model reaches an accuracy of 80.39% in predicting overweight. Culotta [30] use the linguistic variables (LIWC and PERMA) from tweets and demographic variables to predict health-related statistics for 100 most populous counties in the U.S. The Pearson correlation for obesity between the predicted rate and the real rate is 0.64. Abbar et al.[31] did a similar study to Culotta. Abbar et al.[31] use the linguistic variables (LIWC), food features, average calorie per serving for food, and demographic variables from the food-related tweets to predict county-wide obesity rate, achieving a correlation of 0.775 for obesity. Public posts about food and eating behaviors may spread through Social networks[32]. These studies demonstrated a successful application of Twitter data in predicting state health outcomes. Although Yelp data together with Twitter data has been used to characterize food environment by Nguyen et al.[28], no previous study has been found to use Yelp and MFP data to predict state obesity. It also remains unclear, whether we can characterize state level food environment from perspectives of concepts of access and predict obesity rate according to the perspective using publicly available social media data.

## 3.3 Methods

### 3.3.1 Data Collection

Our study used three data sources: (i) Yelp, (ii) MFP, and (iii) BRFSS. The data used in this study to describe the state level food environment is collected via Yelp API (Application Programming Interface)[33] and from MFP using web scraping tool.

Yelp is a leading crowd-sourced reviews site in the U.S. that allows users to search restaurants and local businesses.[34] Users can post reviews and upload photos concerning a business' foods and services, which makes Yelp a location-based social media platform. To date, Yelp.com ranks 52th in the United States and 231st worldwide

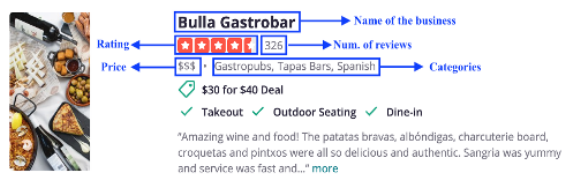



Figure 3.1: Example of the Yelp business list page.

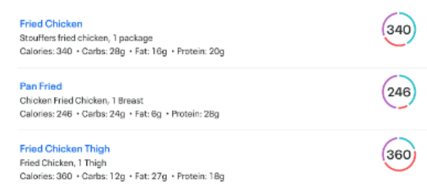



Figure 3.2: Example of the MFP nutrition fact list page.

based on internet traffic and engagement.[35]

The Yelp API allows users to search and query Yelp for over 50,000,000 businesses in 32 countries.[33] In order to obtain the data for this study, we converted 5-digit U.S. zip codes to latitude and longitude coordinates. We queried the detailed business content via Yelp API by searching the nearby businesses to the provided locations. The data is collected in Sep 2020 and it consists of profiles for 353,431 businesses in the United States.

An example of a restaurant's listing on Yelp.com is shown in Figure 3.1. As shown in Figure 3.1, the profile for each business includes the name of the business, average rating, the number of reviews, price level, and categories. Each business could choose up to three terms to describe their services and offerings. The queried business profile returned by the Yelp API not only contains the mentioned fields but also includes other details of the business, such as the business id, address, URL to the business's home page on Yelp.com, photos, and hours of operation. It's worth noting that chain businesses could have the same name, but each location has its unique business id.

Yelp publishes crowd-sourced reviews on many services businesses, such as restaurants, hospitals, and recreational activities. We removed businesses that are not related to the food industry in this study. To do this, each selected category was judged by two independent reviewers based on its relevance to the food field. Two judgments reached 100% agreement with kappa values equal to one. A total of 226 categories were picked from 332 categories. Categories not related to food, such as Hardware Stores, Horseback Riding, and Medical Centers, were removed. The summary of the collected Yelp data is listed in Table 3.1.

Table 3.1: Descriptive statistics of Yelp Data.

| # businesses | # categories | Rating (mean(std.)) | # Reviews (mean(std.)) | Price (mean(std.)) |
|---|---|---|---|---|
| 353,431 | 226 | 4.00 (0.75) | 99.16 (260.32) | 1.60$ (0.56$) |

To understand and objectively compare these categories, we further collect data on each category's most popular 100 restaurants nationwide and their most popular dishes to use as a proxy to estimate the caloric density of each category. Nutrition information (i.e., calories) of each popular dish was collected from MFP using the scraping package BeautifulSoup[36]. MFP is one of the most popular calorie-tracking smartphone apps worldwide with more than 10 million users.[37] Food nutrition information is collected by searching the food name in MFP's nutrition database. Figure 3.2 is the example of search result page when we search the term "fried chicken". The summary of the MFP data is listed in Table 3.2.

Table 3.2: Descriptive statistics of MFP Data.

| # categories | # popular dishes | # nutrition records |
|---|---|---|
| 226 | 37,295 | 3,110,744 |

The state-level obesity rate data was obtained from Behavioral Risk Factor Surveillance System (BRFSS), the nations' state-based health surveillance system that tracks U.S. residents' behavioral risk factors.[16] BRFSS provided the ground truth for the prevalence of obesity via self-reported obesity data among U.S. adults by state and territory in 2019. We collected the obesity rate for 49 states and D.C., because BRFSS did not have sufficient data on obesity rate for New Jersey in 2019.[38]

### 3.3.2 RQ1 Methods

We utilized the category information for each business in Yelp to calculate the degree of availability. The degree of availability of one category of food is defined as the market occupancy of one kind of food in a given area. We calculate the degree of availability of one category at the state level by dividing the total number of businesses for that category by the total number of businesses in that state. For example, the degree of availability of "Mexican" in a given area will be equal to one if all of the restaurants in that area sell Mexican food.

We compared the availability of different kinds of food to areas with low and high prevalence of obesity. We performed a comparison between two states. The two states we selected were Colorado and Mississippi. In 2019, Mississippi (MS) had the highest obesity rate (40.8%), while Colorado (CO) had the lowest obesity rate 23.8%.[38] We first calculated the availability of each category in the two pre-selected locations and further analyzed what categories of restaurants are more available in locations with the high or low obesity rate. The category with the biggest availability difference was further compared by adopting dimensions from *a concept of access*.

The affordability and acceptability of the categories were compared. Affordability refers to the food price the customer needs to pay. Price may affect users' choice of food. Low-income populations have a high risk of living in poor food environments and bear much of the burden of obesity and chronic disease.[14] We estimate the affordability using the price category data for each business. Here, we converted the price categories to numeric numbers for future analysis. For example, $ is converted to 1, and $$$$ is converted to 4. Acceptability refers to the client's attitude to the service provider. We use the average customer's rating and the total number of reviews of a business to measure the customer's attitude concerning a business. Studies showed that the consumers' preference increases with the number of reviews,[39] and consumer-generated ratings about restaurants are positively associated with the online popularity of restaurants.[40] The business with a higher rating and more reviews are considered more likely to be accepted by customers than businesses with poor ratings and a limited number of reviews.

### 3.3.3 RQ2 Methods

Because calorie intake is one of the major contributors to obesity, it's critical to understand nutritional content of food in order to evaluate its effect on obesity. We evaluate the state-level food environment quantitatively using the nutrition information, specifically calorie information, collected from MFP. Categories were turned into average calories per gram for known popular dishes for representative restaurants. The caloric density of each food category weighted by the availability of each category in a state became the weighted score of caloric density of the state.

To calculate the caloric density for category, we firstly collected popular dishes for each category. We chose the top 100 restaurants with the highest number of reviews for each category nationwide and used the web scraping tool, BeautifulSoup, to collect the popular dishes from the Yelp.com pages. Later, those popular dishes are searched in the MFP food nutrition database. We calculated the mean calorie of a popular dish by averaging the calorie per gram for all records returned from MFP for that dish. It should be noted that the nutrition database of MFP contains a combination of foods added by MFP and foods that are added by users and various units of measures (e.g., g, gram, package, breast, oz, piece, and slices) are used. We selected gram as the unified measuring unit for comparison. We included all records that use "gram" or variations of "gram" (e.g., "g", "gr", "grams") as their measuring unit.

### 3.3.4 RQ3 Methods

Based on the results of RQ1 and RQ2, we create features from the availability, affordability, and acceptability to food categories and state weighted score for caloric density for the state-level food environment to describe the local food environment.

We classified these features into three sets: (i) category availability: degree of availability to each category at the state level; (ii) category affordability and acceptability: average price, average rating, and an average number of reviews for each category at the state level; (iii) state weighted score for caloric density: calculated weighted score for caloric density for each state. We apply a combination of different feature sets and employ several popular machine learning models (i.e., Random Forest regression, SVM regression, XGBoost regression) for the prediction. We didn't use the state-of-the-art deep learning models (e.g., CNN regression) in this study, because we have a limited number of samples. Deep learning models would need a large sample size to outperform traditional machine learning techniques.[41] Because we were predicting the obesity rate at the state level, we use the Leave-One-Out Cross-Validation (LOOCV). LOOCV is an extreme version of k-fold cross-validation where k is set to N. N is the number of observations in the dataset. For N times, a model is created and trained on all the data except for the one point and a prediction is made for that point. Thus, we used information from the D.C. and 49 states to predict the other state. Then, we repeated this 50 times, while changing the predicting location. We evaluate our approach by calculating the Pearson correlation between the real obesity rate and the predicted obesity rate.

## 3.4 Results

### 3.4.1 RQ1 Results

We extracted business profile data of the food-related businesses located in the four pre-selected areas from the collected Yelp data. The summary of the data is presented in Table 3.3.

Table 3.3: The summary of the collected data for Colorado and Mississippi.

| **Region** | **Colorado** | **Mississippi** |
|---|---|---|
| # business | 7,109 | 3,845 |
| # business categories | 215 | 142 |
| Rating (mean (std.)) | 4.02 (0.74) | 3.83 (0.96) |
| # reviews (mean (std.)) | 106.59 (197.71) | 22.05 (50.14) |
| Price (mean (std.)) | 1.66$ (0.57$) | 1.50$ (0.55$) |

We firstly calculated the availability of each category in the given areas. In Mississippi, the categories with high availability included "Fast food", "Burgers", "Seafood" and "Sandwiches". In Colorado, categories with high availability were "Mexican", "Breakfast & Brunch", "Sandwiches", and "Burgers". The "Sandwiches" and "Burgers" categories had high availability in both Mississippi and Colorado. We further explored differences in availability for each category in order to understand the state-level food environment in both low obesity state and obesity prevalent state. This was also done in an effort to highlight the importance of access to different types of food. We used net value to measure availability differences between two different locations. Net differences were used to rank the categories in descending order. Results for net difference can be found in Table 3.4. A larger net value indicates a bigger difference. The * indicates that the net difference for a category is significantly different by the z-test (**** p≤0.0001; *** p≤0.001; ** p≤0.01; * p≤0.05). Shaded cells specify that a particular category is more available to customers living in an area with a higher obesity rate area, Mississippi. We found that 59 out of 138 categories showed significant differences between the two states.

Table 3.4: The 40 categories with the highest availability difference between Colorado and Mississippi.

| Category | Net value | Category | Net value | Category | Net value | Category | Net value |
|---|---|---|---|---|---|---|---|
| Fast Food | 0.0844**** | Pizza | 0.0278**** | Soul Food | 0.0170**** | Wine Bars | 0.0094**** |
| Seafood | 0.0824**** | Food Trucks | 0.0275**** | Vietnamese | 0.0156**** | Ramen | 0.0089**** |
| Breakfast & Brunch | 0.0679**** | Breweries | 0.0235**** | Restaurants | 0.0149**** | Pubs | 0.0085**** |
| Burgers | 0.0493**** | Buffets | 0.0227**** | Italian | 0.0115** | Juice Bars & Smoothies | 0.0082**** |
| Southern | 0.0470**** | Coffee & Tea | 0.0216**** | Beer Bar | 0.0111**** | Tex-Mex | 0.0081**** |
| Mexican | 0.0423**** | Cajun/Creole | 0.0204**** | Thai | 0.0108**** | Donuts | 0.0079**** |
| Bars | 0.0415**** | Cafes | 0.0184**** | Bakeries | 0.0105** | Indian | 0.0078**** |
| Chicken Wings | 0.0364**** | Cocktail Bars | 0.0177**** | Asian Fusion | 0.0103**** | Beer, Wine & Spirits | 0.0075** |
| American (New) | 0.0353**** | Convenience Stores | 0.0175**** | Chinese | 0.0098** | Diners | 0.0075** |
| Steakhouses | 0.0298**** | Barbeque | 0.0170**** | Japanese | 0.0097**** | Ice Cream & Frozen Yogurt | 0.0071* |

Note: The shaded cell indicate this category is more available in Mississippi.

As shown in Table 3.4, 16 out of 40 categories were more significantly available in Mississippi than in Colorado at the p $\leq$ 0.0001 level, including "Fast Food", "Buffets", "Donuts", etc. "Diners" and "Chinese" were more significantly available in Mississippi than in Colorado at the $p \leq 0.01$ level. "Ice Cream & Frozen Yogurt" was also found to be more available in Mississippi; however, the difference is not as significant as the aforementioned categories based on p value scores.

Alcohol related businesses, including "Breweries", "Cocktail Bars", "Beer Bar", "Wine Bars", "Pubs", were found to be significantly more available in Colorado. Moreover, "Breakfast & Brunch", "Coffee & Tea", "Mexican", "American (New)", "Pizza", "Food Truck", "Vietnamese", "Thai", "Asian Fusion", "Ramen", "Juicy Bars & Smoothies", "Indian", and "Cafes" were also found to be more available in Colorado than in Mississippi at the $p \leq 0.0001$ level. "Bakeries" and "Beer, Wine, & Spirits" were more available in Colorado than in Mississippi at the $p \leq 0.01$ level.

"Fast Food" was found to have the biggest availability difference between Colorado and Mississippi. We further explored the category with the biggest difference, "Fast food", in availability to fully understand the state-level food environment and the importance of access to different types of food. The availability of "Fast Food" in Mississippi was 13.49%, while the availability of "Fast Food" was 5.03% in Colorado. Due to the biggest prevalent difference in availability, we investigate the relationship between the availability of fast-food restaurants and the state-level obesity rate.

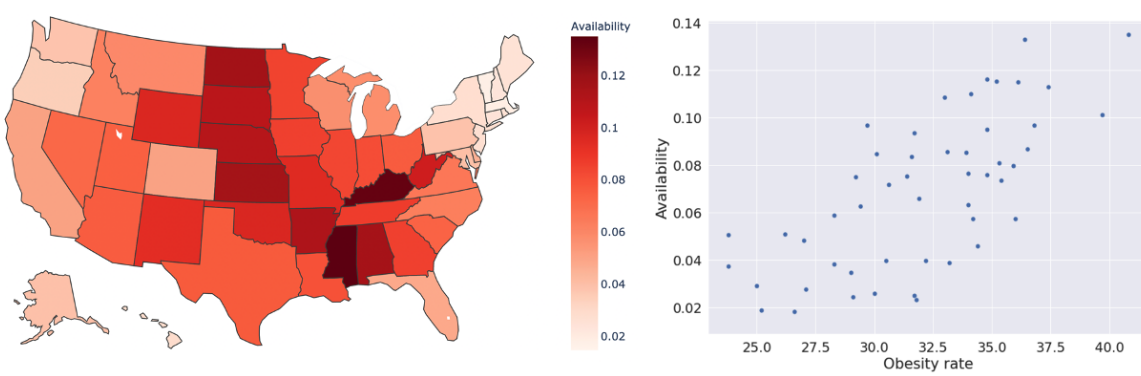



Figure 3.3: The relationship between the availability of fast-food restaurants and the state-level obesity rate.

We visualized the availability of fast-food restaurants map (Figure 3.3. left) and the scatter plot to show the relationship between the availability of fast-food restaurants and the state-level obesity prevalence (Figure 3.3. right). We found that the availability of fast-food restaurants was positively correlated with the obesity rate at the state level, with a resulting Pearson correlation of 0.676. From the heatmap, we also found that the Northeast had the lowest availability of fast food, and the Midwest and South have a higher availability of fast food than the West. We further adopted dimension from *a concept of access* to compare the fast-food restaurants with other restaurants.

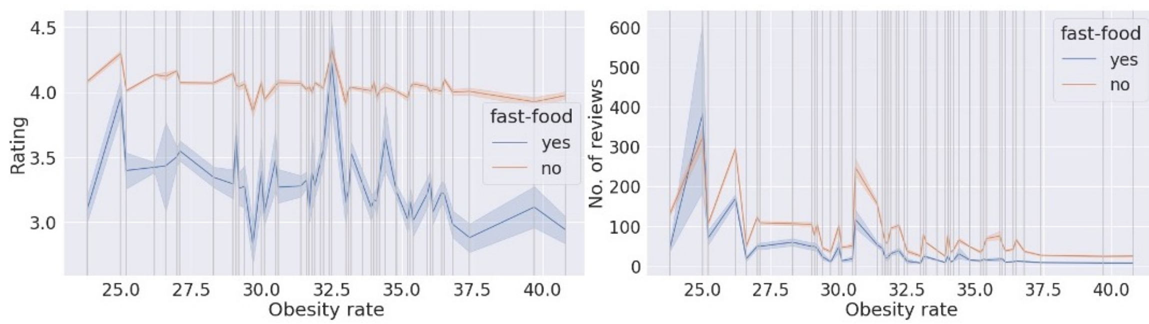



Figure 3.4: The relationship between acceptability (Rating and Number of reviews) to fast-food and local restaurants and the state-level obesity rate.

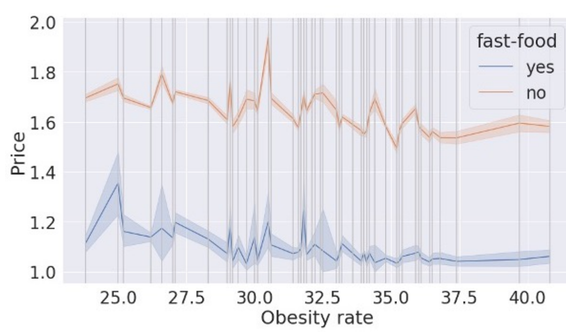



Figure 3.5: The relationship between affordability (price) to fast-food and local restaurants and the state-level obesity rate.

We compared the acceptability (Figure 3.4) and affordability (Figure 3.5) between fast-food and other restaurants. In Figure 3.4 and Figure 3.5, each vertical line represents one state with its corresponding obesity rate. Results showed that the acceptability of fast-food restaurants is lower than other restaurants, irrespective of the prevalence of obesity. We found that the average rating of fast-food restaurants shows a negative relationship with the obesity rate at the state-level. The residents in high obesity rate areas give fast-food restaurants a lower rating than residents in low obesity areas. We also found the range of number of reviews also shows a negative relationship with obesity rate. Results on affordability showed the price level of fast-food restaurants is lower than other restaurants. Other than that, the price of the fast-food restaurants and other restaurants has similar trends, which indicates the price of fast-food restaurants is affected by the local price indices.

### 3.4.2 RQ2 Results

The first step to quantify the food environment was to collect popular dishes for each category. Popular dishes for food categories give us an idea of why some categories are more popular in high obesity areas. We generated word cloud overviews to show the popular dishes for categories that we found in RQ1 that are more popular in Mississippi (Figure 3.6. left) and in Colorado (Figure 3.6. right). Fried food in Colorado is not as popular as in Mississippi. We collected 12,316 popular dishes for categories that are more available in Mississippi, of these 1.2% are fried chicken. For categories that are more available in Colorado, 0.44% popular dishes are fried chicken. The statistical test showed that the difference in proportions between that of fried chicken in Mississippi to that of fried chicken in Colorado is significant with p-value less than the significant level of 0.0001. Similarly, the percentage of other fried food, such as fried catfish, fried shrimp, chicken fried steak, and fried oysters are significantly higher in Mississippi than in Colorado. This finding is consistent with literature studies that the intake of fried food is associated with obesity.[42]

The second step is to calculate the caloric density for each category based on calorie information for all of the available popular dishes. On average there were 166 popular dishes per category. Table 3.5 shows the five most popular dishes per category along with the caloric density for each dish and for each category. We collected up to 100


Figure 3.6: Word cloud overviews of popular dishes for categories more available in Mississippi (left) and categories more available in Colorado (right).

most popular (i.e., highest number of reviews) restaurants in each category. The five most popular dishes for sample categories and its caloric density are shown in 3.5. The complete table for caloric density for category is shown in Appendix A.1.

Table 3.5: The example of top 5 popular dishes and their caloric density for selected categories.

| Category | Popular dish 1 (caloric density) | Popular dish 2 (caloric density) | Popular dish 3 (caloric density) | Popular dish 4 (caloric density) | Popular dish 5 (caloric density) | Caloric density for category |
|---|---|---|---|---|---|---|
| **Chicken Wings** | Fried Chicken (2.240) | Boneless Wings (1.836) | Buffalo Wings (2.02) | Kimchi Fried Rice (3.271) | Chicken Strips (2.108) | 17.31 |
| **Diners** | French Toast (2.545) | Eggs Benedict (2.208) | Chicken Fried Steak (2.665) | Huevos Rancheros (1.147) | Scrambled Eggs (1.649) | 7.289 |
| **Soul Food** | Fried Chicken (2.240) | Fried Catfish (3.283) | Sweet Potato Pie (2.525) | Red Beans and Rice (1.880) | Chicken Breast (1.453) | 6.337 |
| **Patisserie / Cake Shop** | Almond Croissant (4.102) | Chocolate Croissant (3.926) | French Toast (2.545) | Eggs Benedict (2.208) | Tiramisu (3.034) | 6.298 |
| **Southern** | Fried Chicken (2.240) | Fried Catfish (3.283) | Pecan Pie (4.749) | Pork Chop (1.590) | French Toast (2.545) | 5.667 |
| **Smokehouse** | Pulled Pork Sandwich (2.452) | Baby Back Ribs (2.301) | Beef Brisket (2.043) | Brisket Sandwich (2.698) | Pulled Pork (2.112) | 5.51 |
| **American (New)** | French Toast (2.545) | Eggs Benedict (2.208) | Poached Egg (1.414) | Fish Tacos (1.498) | Beet Salad (0.845) | 5.047 |
| **Brasseries** | French Onion Soup (0.808) | Pork Chop (1.590) | Steak Frites (2.465) | Duck Confit (2.646) | Beef Tartare (2.698) | 4.78 |
| **Poke** | Poke Bowl (1.482) | Seaweed Salad (3.510) | Spicy Tuna (1.955) | Octopus (1.838) | Fresh Fish (1.188) | 4.716 |
| **Dim Sum** | Shrimp Dumplings (1.620) | Peking Duck (8.847) | BBQ Pork Buns (2.505) | Har Gow (1.741) | Xiao Long Bao (2.419) | 4.215 |

We further calculate the caloric density for each popular dish. The caloric density for dishes after range from 0.556 to 62.383 with a median value at 2.399. Bakery food has a relative high caloric density. For example, the caloric density for almond croissant and pecan pie are higher than 4. Fatty meat also has a high caloric density. The caloric density for Peking duck reaches 8.847, which is even higher than fried chicken. Cooking method also affect the caloric density. For example, the caloric density for poached egg is 1.414, for scrambled egg is 1.649, and for egg benedict is 2.208. Another example is the calorie per gram for fried catfish is 3.283 and for fresh fish is 1.188. Salad and soup are found with low caloric density. The calorie per gram for beet salad and French onion soup is lower than 1 based on our calculation.

With the calorie information for these popular dishes, we calculate the caloric density for each category by averaging the caloric density for all popular dishes. The caloric density for a category varies from 1.941 to 23.452, with a median 5.473. The "Cheesesteaks" is the category with the highest caloric density, followed by the "Fried Chicken" with an caloric density of 17.310. The "Fruits & Veggies", "Food Tours", "Shaved Snow", "Gay Bars", and "Honey" are categories with the lowest caloric density among all food categories, with caloric density lower than 4.

Finally, we turned the caloric density for each category into the weighted score for caloric density for each state. The estimated weighted score for caloric density for states ranges from 5.786 to 6.430. Washington was the state with the lowest estimated weighted score for caloric density, while Georgia had the highest estimated weighted score for caloric density among all states. Colorado's score was 5.955, and Mississippi's score was 6.305. We performed a two-sample z-test between these two states. The result showed a significant difference with a z-value equal to 12.759 and a p-value smaller than 0.0001. The relationship between the state estimated weighted score for caloric density and state obesity rate is shown in Figure 3.7. The estimated weighted score for caloric density of states calculated using our approach showed a strong positive correlation (Pearson correlation coefficient equals 0.671 with p-value less than 0.0001) with the state-level obesity rate. The higher estimated weighted score for caloric density of a state indicates the state-level food environment is more obesity prone by serving high calorie density food. Moreover, the estimated caloric density weighted score for Southern food is higher than other areas in the United States, especially in Georgia, Alabama, and Mississippi.

### 3.4.3 RQ3 Results

We generated three sets of features for the prediction. The feature set was as follows: (i) category availability and (ii) category affordability and acceptability and (iii)

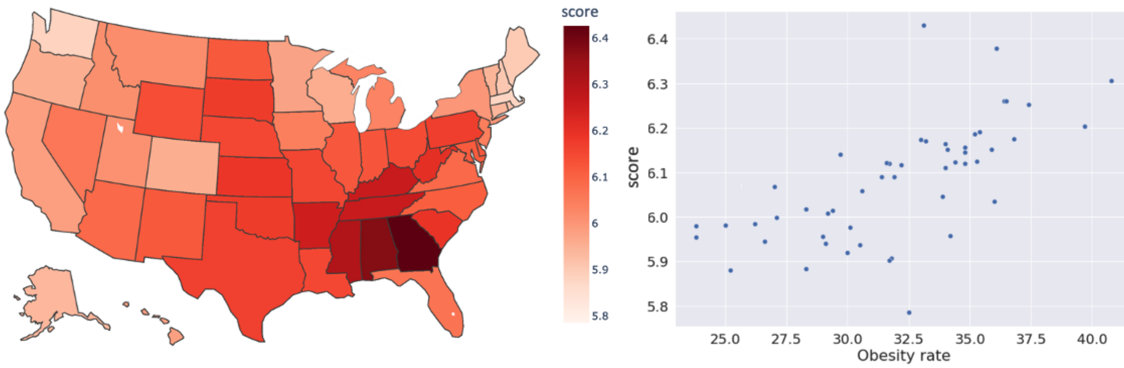



Figure 3.7: The weighted score for caloric density of each state.

weighted score for caloric density. Affordability and acceptability were created at the state level for the identified 226 categories. Estimated state weighted score for caloric density was created in RQ2. Since each state has only one estimated weighted score for caloric density, prediction models other than linear regression are not applicable for prediction using this set of feature. For categories that do not exist in a state, we used 0 to fill in the missing values for the category availability, affordability and acceptability. Table 3.6. shows the results of comparing different prediction models with the different combinations of input using the Pearson correlation coefficient to evaluate it.

Table 3.6: Pearson correlation coefficients for different combinations of input for prediction.

| **Features** | **Linear regression** | **Random Forest regression** | **SVM regression** | **XGBoost regression** |
|---|---|---|---|---|
| (i) category availability | 0.407 | 0.763 | 0.712 | 0.742 |
| (ii) category affordability and acceptability | 0.402 | 0.776 | 0.593 | 0.743 |
| (iii) state weighted score for caloric density | 0.622 | - | - | - |
| (i) + (ii) | 0.403 | 0.791 | 0.642 | 0.731 |
| (i) + (iii) | 0.336 | 0.771 | 0.714 | 0.710 |
| (i) + (ii) + (iii) | 0.402 | **0.796** | 0.643 | 0.708 |

The random forest model with all three sets of features performed the best. Additionally, the Pearson correlation coefficient between the predicted obesity rate and the real obesity rate is 0.796, which indicates the predicted value is correlated with the real value. The affordability and acceptability were more effective in predicting the state obesity rate compared to other two feature sets.

## 3.5 Discussion

### 3.5.1 Principal Findings

In this study, we characterized the food environments using the data from Yelp and MFP with an innovative data collection and processing methods. We characterize food environment using Yelp and MFP and predicted state level obesity rates. Additionally, our study contributes a new method to calculate food environment and data to estimate food calories for different popular dishes and categories of restaurants for future studies.

Our results showed the disparity in available food category between Colorado and Mississippi. "Fast food" are found to be more available in Mississippi than in Colorado. Fast food consumption has been found strongly associated with weight gain and obesity.[3] Individual-level diet and weight outcomes are thought to improve in neighborhoods that have access to high-quality food.[43] Comparing the state-level food availability difference, we found abundant access to fast-food options may contribute to a negative group level health outcome. Although fast food restaurants are notoriously for serving high calorie, low nutritional food,[21, 22] such as hamburgers, french fries, and fish and chips, [44] some differences have been found. By comparing the popularity of fast-food restaurants with other restaurants in Figure 3.4, we found the fast-food restaurants always have a lower number of reviews than other restaurants. However, in the District of Columbia, the average number of reviews of fast-food restaurants is higher than other restaurants. It may be because more alternative fast-food are available in cities, like sushi and poke, which are considered light and healthy.[45]

Other than using the available food category to characterize the state food environment, we also use the popular dish and nutrition content for popular dishes to quantify state food environment. Our study is the first one study that conducted a large scale analysis of popular dishes. We compared the popular dishes for Colorado and Mississippi. We found fried food are more popular in Mississippi. This finding is consistent with literature that the intake of fried food is associated with obesity.[42] With the collected popular dish, we calculated the weighted score for caloric density for each state. Similar studies exist. For examples, Nguyen et al.[28] quantify state food environment by calculating the caloric density of food mentions in geo-tagged tweets. They use a list of over 1,430 popular foods and beverages from the US Department of Agriculture's National Nutrient Database and calculated calories per 100g for each food item.[28] Abbar et al.[31] calculated the average calorie by checking the calorie per serving for the selected 500 food keyword. Different from these two studies, we used the MyFitnessPal, the biggest food database available,[37] to get the nutrition data. We collected nutrition data for 37,295 dishes, which can allow for an effective utilization of data points. In our study, Pearson correlation of weighted score for caloric density of states to state obesity rates is 0.671, which outperforms aforementioned one previous study[31], in which Pearson correlation of tweet caloric value to state obesity rates is 0.629.

Our prediction model is the first one to use Yelp and MFP data to predict state obesity rate. Constrast to previous studies that using Twitter data to predict obesity rate,[29, 31, 30] our model that using Yelp and MFP data has less selection bias. First, Twitter users are younger than general public [46], however the user group of Yelp is more evenly distributed by age with 33% of users at 55+ age group.[47] Second, the previous studies using Twitter data for prediction only use sampled data due to the massive amount data of Twitter. Despite these study had the same data source, their collection methods were different, which could have skewed the results.

### 3.5.2 Public Health Implication

Our study helps us understand the impact of food environment and related human behavior by showing the correlation between state level food environment and their obesity rate. Because of the pervasive usage of smartphones and social media apps like Yelp across the country, researchers could use social media data to gain an understanding of food environments in any part of America and other countries as well. In sum, our model has the potential to evaluate the food environments.

Not only does our model map out a landscape of available food, but it allows us to monitor and project public health within the area. The copious amounts of information on social media allow public health practitioners to monitor changes in food availability and population over time, and use this information to predict changes in the state obesity levels. Similarly, the computational methods could be used to inform dieting habits at the individual level. This allows for an early intervention in areas or individuals that face the greatest risk of increasing obesity rates or becoming obese.

Our study has reiterated a few fundamental findings that relates to the important of environment.[31, 30, 9] Our findings suggest that those who live in areas with a considerable availability to high calorie, fast-foods are more likely to be obese. This alludes to the idea that people eat what is readily available to them. Politicians and city planners could potentially use this information to develop an infrastructure of healthy food options in areas that have been traditionally concentrated with fast food restaurants. This sort of environmental intervention could potentially influence community behavior and lead to better health outcomes.

### 3.5.3 Limitation and Future Direction

The first limitation for our study lies in the data collection. Yelp provides us substantial data for local business; however, Yelp API results are restricted to 1,000 results for each query. We could collect up to 1,000 business data for each zip code center up to a distance of 40 km (about 25 miles). For urban environments, one zip code may have more than 1,000 businesses. To address this issue, we run several rounds for each zip code and remove the duplicates. Despite this effort, missing data may skew our results especially for the urban areas. We found second limitation when collecting nutrition data from MFP. For each search query, MFP returns 10 pages with 10 records on each page. Some popular dishes do not have an exact match, in which case MFP returns a partial matching dishes. Therefore, some caloric information may not be accurate. We averaged all the results to reduce effects of inaccurate information. Another limitation is not capturing the actual consumption. In this study, we calculate the caloric density for popular dishes. We found high caloric density food is correlated with obesity rate, consistent with previous study that was conducted at the individual level.[48] Last but not least, a similar analysis should be replicated at the zip code level to better inform local food environment. We used the state-level food environment in this study, because BRFSS provides state level obesity rate. More granular analysis will provide a better insight into how socio-economic status and their local food environment may be correlated with obesity [49, 50, 51, 14].

## 3.6 Conclusion

This study uses social media data to characterize state level food environment. State-level food environment shows disparity of available food between states with different obesity rate, suggesting the importance of food environment. Using availability of different categories of food along with affordability and acceptability data captured in social media, we created a state-level obesity rate prediction model with 0.796 correlation. Using our proposed method, public health practitioners could monitor the change of areas that face the greatest risk of increasing obesity rates to counter obesity pandemic.

## REFERENCES


[1] W. H. Organization *et al.*, "2008-2013 action plan for the global strategy for the prevention and control of noncommunicable diseases: prevent and control cardiovascular diseases, cancers, chronic respiratory diseases and diabetes," 2009.

[2] D. Thompson, J. Edelsberg, G. A. Colditz, A. P. Bird, and G. Oster, "Lifetime health and economic consequences of obesity," *Archives of internal medicine*, vol. 159, no. 18, pp. 2177–2183, 1999.

[3] R. Rosenheck, "Fast food consumption and increased caloric intake: a systematic review of a trajectory towards weight gain and obesity risk," *Obesity reviews*, vol. 9, no. 6, pp. 535–547, 2008.

[4] Craig M. Hales and Margaret D. Carroll and Cheryl D. Fryar and Cynthia L. Ogden, "Prevalence of Obesity and Severe Obesity Among Adults: United States, 2017â2018."

[5] heryl D. Fryar and Margaret D. Carroll and Cynthia L. Ogden, "Prevalence of Overweight, Obesity, and Extreme Obesity Among Adults: United States, Trends 1960â1962 Through 2009â2010."

[6] K. D. Hall, "Did the food environment cause the obesity epidemic?," *Obesity*, vol. 26, no. 1, pp. 11–13, 2018.

[7] S. N. Bleich, D. Cutler, C. Murray, and A. Adams, "Why is the developed world obese?," *Annu. Rev. Public Health*, vol. 29, pp. 273–295, 2008.

[8] V. I. Kraak, "Comprehensive restaurant-sector changes are essential to reduce obesity risk for all americans," *American journal of public health*, vol. 108, no. 2, p. 158, 2018.

[9] C. E. Caspi, G. Sorensen, S. Subramanian, and I. Kawachi, "The local food environment and diet: a systematic review," *Health & place*, vol. 18, no. 5, pp. 1172–1187, 2012.

[10] "U.S. Food Expenditures at Home and Abroad." `https://web.archive.org/web/20210609032840/https://www.fb.org/market-intel/u.s.-food-expenditures-at-home-and-abroad`. [Online; Accessed: 2021-06-07].

[11] L. K. Cobb, L. J. Appel, M. Franco, J. C. Jones-Smith, A. Nur, and C. A. Anderson, "The relationship of the local food environment with obesity: a systematic review of methods, study quality, and results," *Obesity*, vol. 23, no. 7, pp. 1331–1344, 2015.

[12] K. B. Morland and K. R. Evenson, "Obesity prevalence and the local food environment," *Health & place*, vol. 15, no. 2, pp. 491–495, 2009.

[13] J. Maddock, "The relationship between obesity and the prevalence of fast food restaurants: state-level analysis," *American journal of health promotion*, vol. 19, no. 2, pp. 137–143, 2004.

[14] J. P. Block, R. A. Scribner, and K. B. DeSalvo, "Fast food, race/ethnicity, and income: a geographic analysis," *American journal of preventive medicine*, vol. 27, no. 3, pp. 211–217, 2004.

[15] J. C. Spence, N. Cutumisu, J. Edwards, K. D. Raine, and K. Smoyer-Tomic, "Relation between local food environments and obesity among adults," *BMC public health*, vol. 9, no. 1, pp. 1–6, 2009.

[16] "BRFSS." `https://web.archive.org/web/20210430063347/https://www.cdc.gov/chronicdisease/resources/publications/factsheets/brfss.htm`. [Online; Accessed: 2021-06-07].

[17] K. J. Duffey and B. M. Popkin, "Energy density, portion size, and eating occasions: contributions to increased energy intake in the united states, 1977–2006," *PLoS Med*, vol. 8, no. 6, p. e1001050, 2011.

[18] J. M. Poti and B. M. Popkin, "Trends in energy intake among us children by eating location and food source, 1977-2006," *Journal of the American Dietetic Association*, vol. 111, no. 8, pp. 1156–1164, 2011.

[19] L. R. Young and M. Nestle, "The contribution of expanding portion sizes to the us obesity epidemic," *American journal of public health*, vol. 92, no. 2, pp. 246–249, 2002.

[20] M. B. E. Livingstone and L. K. Pourshahidi, "Portion size and obesity," *Advances in nutrition*, vol. 5, no. 6, pp. 829–834, 2014.

[21] M. A. Pereira, A. I. Kartashov, C. B. Ebbeling, L. Van Horn, M. L. Slattery, D. R. Jacobs Jr, and D. S. Ludwig, "Fast-food habits, weight gain, and insulin resistance (the cardia study): 15-year prospective analysis," *The lancet*, vol. 365, no. 9453, pp. 36–42, 2005.

[22] E. Isganaitis and R. H. Lustig, "Fast food, central nervous system insulin resistance, and obesity," *Arteriosclerosis, thrombosis, and vascular biology*, vol. 25, no. 12, pp. 2451–2462, 2005.

[23] B. Kelly, V. M. Flood, and H. Yeatman, "Measuring local food environments: an overview of available methods and measures," *Health & place*, vol. 17, no. 6, pp. 1284–1293, 2011.

[24] R. Penchansky and J. W. Thomas, "The concept of access: definition and relationship to consumer satisfaction," *Medical care*, pp. 127–140, 1981.

[25] H. Charreire, R. Casey, P. Salze, C. Simon, B. Chaix, A. Banos, D. Badariotti, C. Weber, and J.-M. Oppert, "Measuring the food environment using geographical information systems: a methodological review," *Public health nutrition*, vol. 13, no. 11, pp. 1773–1785, 2010.

[26] R. A. McKinnon, J. Reedy, M. A. Morrissette, L. A. Lytle, and A. L. Yaroch, "Measures of the food environment: a compilation of the literature, 1990–2007," *American journal of preventive medicine*, vol. 36, no. 4, pp. S124–S133, 2009.

[27] M. Mazidi and J. R. Speakman, "Higher densities of fast-food and full-service restaurants are not associated with obesity prevalence," *The American journal of clinical nutrition*, vol. 106, no. 2, pp. 603–613, 2017.

[28] Q. Nguyen, H. Meng, D. Li, S. Kath, M. McCullough, D. Paul, P. Kanokvimankul, T. Nguyen, and F. Li, "Social media indicators of the food environment and state health outcomes," *Public Health*, vol. 148, pp. 120–128, 2017.

[29] D. Fried, M. Surdeanu, S. Kobourov, M. Hingle, and D. Bell, "Analyzing the language of food on social media," in *2014 IEEE International Conference on Big Data (Big Data)*, pp. 778–783, IEEE, 2014.

[30] A. Culotta, "Estimating county health statistics with twitter," in *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1335–1344, 2014.

[31] S. Abbar, Y. Mejova, and I. Weber, "You tweet what you eat: Studying food consumption through twitter," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 3197–3206, 2015.

[32] M. A. Pachucki, P. F. Jacques, and N. A. Christakis, "Social network concordance in food choice among spouses, friends, and siblings," *American journal of public health*, vol. 101, no. 11, pp. 2170–2177, 2011.

[33] "YelpAPI." `https://web.archive.org/web/20210609105605/https://www.yelp.com/dataset/documentation/main/`. [Online; Accessed: 2021-06-07].

[34] "Yelp." `https://en.wikipedia.org/wiki/Yelp`. [Online; Accessed: 2021-06-07].

[35] "Yelp's Alexa Rank." `https://web.archive.org/web/20210611025251/https://www.alexa.com/siteinfo/yelp.com`. [Online; Accessed: 2021-06-10).

[36] "Beautifulsoup." https://pypi.org/project/beautifulsoup4/. [Online; Accessed: 2021-06-10).

[37] "MyFitnessPal Database." https://web.archive.org/web/20210211213534/https://play.google.com/store/apps/details?id=com.myfitnesspal.android&hl=en_US&gl=US. [Online; Accessed: 2021-06-10].

[38] "BRFSS Obesity Rate (2019)." https://web.archive.org/web/20210611091817/https://www.cdc.gov/obesity/data/prevalence-maps.html. [Online; Accessed: 2021-06-10).

[39] G. Viglia, R. Furlan, and A. Ladron-de Guevara, "Please, talk about it! when hotel popularity boosts preferences," *International Journal of Hospitality Management*, vol. 42, pp. 155–164, 2014.

[40] Z. Zhang, Q. Ye, R. Law, and Y. Li, "The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews," *International Journal of Hospitality Management*, vol. 29, no. 4, pp. 694–700, 2010.

[41] B. Liu, Y. Wei, Y. Zhang, and Q. Yang, "Deep neural networks for high dimension, low sample size data.," in *IJCAI*, pp. 2287–2293, 2017.

[42] T. V. Gadiraju, Y. Patel, J. M. Gaziano, and L. Djoussé, "Fried food consumption and cardiovascular health: a review of current evidence," *Nutrients*, vol. 7, no. 10, pp. 8424–8430, 2015.

[43] P. Gordon-Larsen, "Food availability/convenience and obesity," *Advances in nutrition*, vol. 5, no. 6, pp. 809–817, 2014.

[44] "Fast food wiki." https://web.archive.org/web/20210621164302/https://en.wikipedia.org/wiki/Fast_food. [Online; Accessed: 2021-06-10].

[45] "Poke is the biggest fast casual food trend." https://web.archive.org/web/20210621163951/https://www.businessinsider.com/poke-is-the-biggest-fast-casual-food-trend-2017-6. [Online; Accessed: 2021-06-10].

[46] Wojcik, Stefan and Hughes, Adam, "Sizing Up Twitter Users."

[47] "Yelp Fast Facts." https://web.archive.org/web/20210707191209/https://www.yelp-press.com/company/fast-facts/default.aspx. [Online; Accessed: 2021-07-01].

[48] Bassett, Mary T and Perl, Sarah, "Obesity: the public health challenge of our time," 2004.

[49] V. W. Chang and D. S. Lauderdale, "Income disparities in body mass index and obesity in the united states, 1971-2002," *Archives of internal medicine*, vol. 165, no. 18, pp. 2122–2128, 2005.

[50] L. R. Mobley, E. D. Root, E. A. Finkelstein, O. Khavjou, R. P. Farris, and J. C. Will, “Environment, obesity, and cardiovascular disease risk in low-income women,” *American journal of preventive medicine*, vol. 30, no. 4, pp. 327–332, 2006.

[51] S. E. Fleischhacker, K. R. Evenson, D. A. Rodriguez, and A. S. Ammerman, “A systematic review of fast food access studies,” *Obesity reviews*, vol. 12, no. 5, pp. e460–e471, 2011.

# CHAPTER 4: Leveraging social media data to understand the impact of COVID-19 on residents' dietary behavior

## 4.1 Introduction

The COVID-19 pandemic is the greatest public health challenge we have faced since World War Two. The World Health Organization (WHO) declared COVID-19 a global pandemic on 11 March 2020. To date, more than 200 million people have been infected and more than 4 million people have died from COVID-19 worldwide. COVID-19 was first identified in 2019 in Wuhan, China. As an infectious disease caused by a newly discovered coronavirus, COVID-19 infections range from symptoms similar to the common to cold to serious and life-threatening disease. Older people and those with underlying medical problems like cardiovascular disease, diabetes, and cancer have a higher probability to develop into a serious illness if infected with the COVID-19 virus.[1]

COVID-19 virus primarily spreads through droplets of saliva or discharge from the nose when an infected person coughs or sneezes. Because of this, social distancing could decrease viral spread. Following a significant surge in COVID-19, on 13 March 2020, former President Trump declared COVID-19 a new national emergency. Each state started to issue 'stay-at-home' orders, close public schools, announce mask mandates, and limit restaurant capacities. On March 19, California was the first to issue a statewide 'stay-at-home' order. As part of the order's provisions, residents were only permitted to leave home if they had an essential job or were shopping for essential needs.

At the same time, governments, researchers, and pharmaceutical companies around the world began working on developing and distributing a COVID-19 vaccine. On

March 2, 2021, President Joe Biden projected that the government would have enough vaccine doses for every American adult by the end of May. On the same day, Texas and Mississippi repealed their restrictions, and Gov. Greg Abbott tweeted: "Texas is OPEN 100%. EVERYTHING.". With the news of anticipated widespread roll-out of vaccine, repeals in state-wide mandates and optimism from politicians, it seemed that life would return to normal.

Efforts combating the spread of COVID-19 had a huge impact on people's lifestyles as well as dietary behavior. During shutdown, people's access to restaurants was restricted. People purchased more food during each shopping trip to limit the number of grocery store visits.[2, 3] They also began to prepare more meals at home. Other factors related to the COVID-19 pandemic may have also affected users' dietary behaviors. The fear of COVID-19, loss of job, and lack of socialization [4] caused an increase in high stress levels in individuals world-wide. Psychological stress has been associated with increased consumption of food, especially high-fat foods.[5] Additionally, the pandemic caused food insecurity in low-income families. This will likely increase the probability of consuming highly palatable foods that are high in fat and sodium.[6] Consuming high-fat foods has been suggested to play a large role in the development of obesity. Because obesity has been discovered as a risk factor for increased COVID-19 symptom severity, it is important to understand how the COVID-19 shutdown has affected people's dietary behavior.

Studies examining various COVID-19-related life-style changes have been conducted worldwide. Cosgrove and Wharton [7] used survey data to examine perceived changes in dietary healthfulness of the general US population during the COVID-19 pandemic. Although surveys provide high-quality and targeted data, a survey has a limited sample size and requires money and labor. Compared to traditional surveys, social media is widely used in the United States,[8] and data from social media are generally free to collect. Social media data provides insights more rapidly and comprehensively.[9] Additionally, the usage of social media platforms increased 61% during the pandemic.[10]

Social media has emerged as a practical data source for large-scale public health studies. This is especially true for lifestyle diseases, such as obesity, since an increasing number of people's life are shared online.[11] In the last couple of years, social media users have shifted from text only to multimedia posts.[12] Social media image data could contain additional information that isn't provided in the textual data, since uploading images is easier and faster with a smart mobile device. Given the fact that Twitter has limited contextual information due to tweet length restrictions and the increasing number of multimedia postings, our study uses text and image data from social media to learn peoples' dietary behaviors.

We collected Twitter data from the months of March to June in 2019, 2020, and 2021 to investigate residents' food behavior before, during, and partially after the shutdown. We investigate how the COVID-19 shutdown impacted resident's dietary behavior and the relationships between residents' food behavior and the state-level obesity rate. Image classification models, text analysis methods, and visualization tools are used to answer two research questions in this study:

- **RQ1**: Does the state obesity rate correlate with the diet images on Twitter?
- **RQ2**: How has the COVID-19 shutdown affected other patterns of people's dietary behavior?

## 4.2 Methods

### 4.2.1 Data Collection

Twitter has a broad demographic user across ethnicity, gender, age, and income.[13] It is a good source of data to learn user's dietary behaviors on a large scale. We used the Twitter API to collect tweets about food for posts with images containing hashtags related to meals (Table 4.1) based on Fred et al.'s[14] study. We approximately collected 200,000 tweets from May, June, and July of 2019, 2020, and 2021. Tweets are short posts limited to 140 characters. In our collection, the average length of tweets was 13.3 words, after filtering out href links, non-alphanumeric characters, special characters, and punctuation. The tweet collection contains around 220,000 unique words.

Table 4.1: Number of tweets with geolocation info and images per hashtag.

| **Hashtags** | **2019** | **2020** | **2021** |
|---|---|---|---|
| #Breakfast | 17,203 (20.49%) | 12,956 (18.93%) | 10,728 (18.77%) |
| #Lunch | 25,157 (29.96%) | 16,855 (24.63%) | 15,629 (27.35%) |
| #Dinner | 25,742 (30.66%) | 23,431 (34.24%) | 18,597 (32.54%) |
| #Brunch | 5,956 (7.09%) | 3,471 (5.07%) | 4,376 (7.66%) |
| #Snack | 3,207 (3.82%) | 3,247 (4.75%) | 2,426 (4.24%) |
| #Meal | 5,631(6.71%) | 7,421 (10.84%) | 4,684 (8.20%) |
| #Supper | 863 (1.03%) | 913 (1.33%) | 615 (1.07%) |
| Total | 83,970 | 68,429 | 57,151 |

### 4.2.2 RQ1 Methods

We used image classification techniques to examine healthiness of food choices by classifying food images attached to tweets into four different health categories: Definitely Unhealthy, Unhealthy, Healthy, and Definitely Healthy. This classification schema comes from Vydiswaran et al.'s[15] study, in which public health nutrition assigned food-related keywords into four different health categories. Not all images contain food items. People not only tweet images about what they eat but also who they eat with and where they eat. Because of this we designed a workflow to finish the health category classification task. Examples of non-food images are shown in Figure I. Our workflow has two steps: 1) classify if an image contains food items (*food classification*); 2) if the images contain food items, classify the food into different healthy levels (*health category classification*).


Figure 4.1: Example of non-food images.

For the two proposed classification tasks, we used transfer learning technologies. A pre-trained ResNet, trained on more than a million images from the ImageNet database, is adopted to extract image features. ResNet is a residual convolutional neural network designed for image classification tasks.[16] We choose the 101 layers ResNet network because 101-layer ResNet has similar performance to the 152-layers ResNet on the ImageNet database, but is more computationally efficient. Image features learned from the pre-trained ResNet-101 are transferred to a SoftMax layer to do final classification tasks. To feed image data into the deep learning model, we pre-processed our image data by cropping each image to $224 \times 224$. We used the Adam optimizer and categorical-cross entropy as the loss function.

We used the publicly available Food-5K dataset as the training data to train the food classification model. This dataset contains two categories, food and non_food, each with 2.5K images. To train the health category classification model, following Vydiswaran et al.'s study, we collected images for food items listed as representative of each food category (TABLE II). We used the Google image search to collect the training dataset by searching and downloading images for each food item. The total number of images we collected is 23,151. We split two training datasets 80-20 for training and validation. We fine-tuned the parameters of pre-trained ResNet-101 in the training process. The fine-tuned classification models with the lowest loss is saved to classify images we collected from tweets to different health categories.

We used the percentage of healthy/unhealthy food of dieting images on Twitter at the state level to study the relationship between the state obesity rate and dieting images on Twitter.

Table 4.2: Food items for each category.

| **Def. Healthy** | **Healthy** | **Unhealthy** | **Def. Unhealthy** |
|---|---|---|---|
| Sushi | Coffee | Pizza | Starbucks |
| Apple | Tea | Grill | Ice cream |
| Fish | Coconut | Taco bell | Chocolate |
| Salad | Rice | Fries | Cake |
| Pumpkin | Turkey | Tacos | Bacon |
| Pineapple | Potatoes | Sauce | Cookie |
| Fruit | Chili | Steak | Icing |
| Eggs | Protein | Taco | McDonalds |
| Orange | Roasting | Oil | Coney Island |
| Oyster | baking | Chipotle | candy |

### 4.2.3 RQ2 Methods

Fear-related stress and the constant news from the media about COVID-19 may push one to consume so-called "comfort food" (high in sugar and fat) or bring about a greater consumption of snacks between meals.[17] In our RQ2, we examine how the COVID-19 shutdown impacted people's emotions, and thus eating behaviors. We also examined how COVID-19 affected dietary behavior for factors such as the time/frequency for each meal and the how often healthy/non-healthy foods were consumed.

Because tweets are short text and food-related tweets contain some food-specific sentiment words, the traditional dictionary-based sentiment analysis method, like LIWC may not be applicable.[15] We used Valence Aware Dictionary for Sentiment Reasoning (VADER) to peform the sentiment analysis. VADER is a parsimonious rule-based model for sentiment analysis, which is especially attuned to social media contexts.[18] VADER catches slang and emojis and appropriately adjusts the intensity of sentiment score by combination of signals. A positive, neutral, negative, and a compound sentiment score was given to each tweet. The compound score is computed by summing the valence scores of each word in the lexicon, adjusted according to the rules, and then normalized to be between -1 and 1. The -1 represents most extreme negative and +1 means the most extreme positive. In this study, we use the compound score of each tweet to do the sentiment analysis. We use the threshold values 0 to classify tweets into different sentiment category, if the compound score of a tweet is smaller than 0, we classify this tweet as negative, otherwise this tweet will be classified as positive.

Temporal histograms allow us to see how popular terms change over time. To do this, we first converted all tweets' post time to its local time based on location. We visualized the popularity of 'breakfast', 'lunch', 'dinner', 'snack' at the hour of day granularity. Additionally, we also visualize the hourly change of distribution of the four food health categories of food in a day.

## 4.3 Results

### 4.3.1 RQ1 Results

We used the saved classification models to classify tweets' images. The accuracy for the food classification model on the validation data is 95.36% and the accuracy for the health category classification model on the validation data is 85.21%. We further evaluate the saved two models on our Twitter image data. We manually selected 200 images and manually labeled them. The accuracy for food classification model on Twitter image data is 90.00% and the performance of food classification models is shown in Table 4.3. For images that contain food items, the model has higher recall than precision. The primary reason an image containing food items is incorrectly predicted as an image without food items is that we only corp the center of an image, but the food is at the corner of the picture. Two images that have Mcdonald and Burger King icons are classified as images containing food items by the model.

Table 4.3: Image classification model performance

| Class | Precision | Recall | F1-score | Num. of occurrence |
|---|---|---|---|---|
| **No Food** | 0.88 | 0.97 | 0.92 | 118 |
| **Contain Food** | 0.94 | 0.80 | 0.87 | 82 |

The accuracy for health category classification model on Twitter image data is 63.64% and the performance of food classification model for all the classes individually is shown in Table 4.4. The model's performance on Twitter data is not as good as on training data. By manually analyzing errors, we found photos from Twitter always contain various kinds of food. However, approximately 90% of the images we collected from Google only have one type of food. It's challenging for the model to predict when the food at different healthy categories is mixed up. We also could improve the model's performance by including images of more types of food in our training dataset. Our classification schema only contains forty kinds of food, and some popular food items are not included, such as cheese.

Table 4.4: Health category classification model performance

| Class | Precision | Recall | F1-score | Num. of occurrence | Predicted Def. Healthy | Predicted Healthy | Predicted Unhealthy | Predicted Def. Unhealthy |
|---|---|---|---|---|---|---|---|---|
| **Def. Healthy** | 0.64 | 0.60 | 0.62 | 15 | 9 | 2 | 2 | 2 |
| **Healthy** | 0.58 | 0.54 | 0.56 | 13 | 2 | 7 | 1 | 3 |
| **Unhealthy** | 0.87 | 0.63 | 0.73 | 32 | 3 | 3 | 20 | 6 |
| **Def. Unhealthy** | 0.35 | 1 | 0.52 | 6 | 0 | 0 | 0 | 6 |

The Figure 4.2 shows the prediction of food images in their respective classes. The top left images in Figure 4.2 contains images of salad, fruit, sushi, cheese, and poached egg, which are classified to represent the category of Def. Healthy. The food images classified to represent the category of Def. Unhealthy are shown in the bottom left. These including images for cake, pie, ice cream, chocolate, mac-n-cheese, and some meat. The top right is the food images classified as Healthy. This class includes images of roast chicken, egg, pancake, mashed potato, etc. Images on the bottom right are classified to represent the Unhealthy category. This includes images of burger, ramen, and spaghetti.

Table 4.5: Result of image classification models.

| | **Food** | **Non-food** | **Def. Healthy** | **Healthy** | **Unhealthy** | **Def.Unhealthy** |
|---|---|---|---|---|---|---|
| **2019** | 36,168 (29.11%) | 88,042 (70.88%) | 5,937 (16.42%) | 7,694 (21.27%) | 11,727 (32.42%) | 10,810 (29.89%) |
| **2020** | 35,184 (37.24%) | 59,287 (62.76%) | 5,870 (16.68%) | 8,170 (23.22%) | 10,928 (31.05%) | 10,216 (29.03%) |
| **2021** | 31,049 (36.77%) | 53,401 (63.23%) | 5,360 (17.26%) | 6,850 (22.06%) | 10,164 (32.74%) | 8,675 (27.94%) |

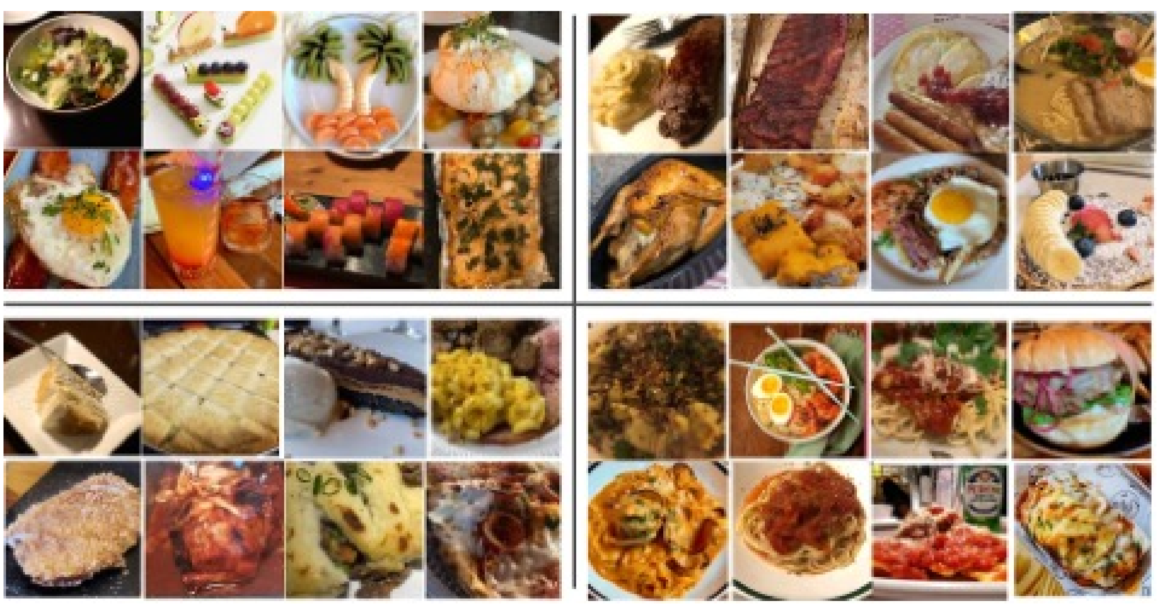


Figure 4.2: Food in top left images are predicted as Def. Healthy. Food in top right images are predicted as Healthy. Food in bottom left images are predicted as Def. Unhealthy. Food in bottom right images are predicted as Unhealthy.

We obtained the state obesity rate data from Behavioral Risk Factor Surveillance

System (BRFSS). BRFSS provided the ground truth for the prevalence of obesity via self-reported obesity data among U.S. adults by state and territory in 2019. In Figure 4.3, we classified 49 states into four different obesity levels based on the state obesity rate. New Jersey is excluded in this study, because the obesity rate data from BRFSS for New Jersey is missing due to lack of sufficient data. As shown in Figure 4.3, we used the box plot to visualize the relationship between state obesity rate and the four healthy categories of dieting images.

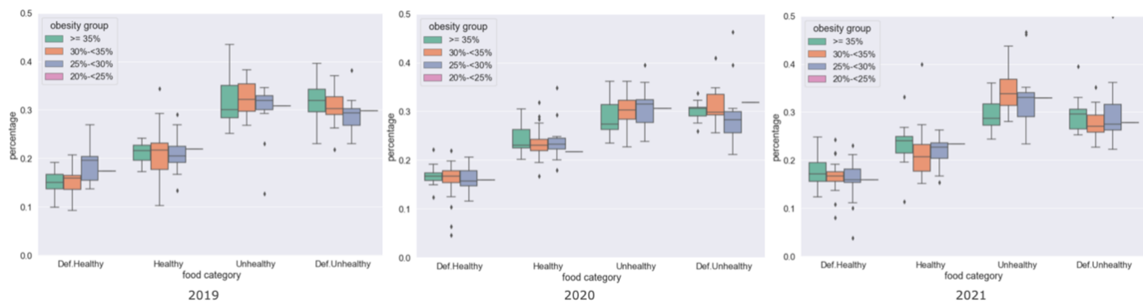



Figure 4.3: The relationship between state obesity rate and dieting images.

We found the percentage of Def. Healthy and Def. Unhealthy food of a state in 2019 is significantly correlated with the state obesity rate by Pearson correlation. The percentage of Def. Healthy food is negatively correlated with the state obesity rate (r = -0.360, p <= 0.05), and the percentage of Def. Unhealthy food is positively correlated with the state obesity (r = 0.306, p<=0.05). In other words, users in a state with a higher obesity rate post less Def. Healthy food and more Def. Unhealthy food on Twitter than users in a state with a lower obesity rate. No significant differences have been found between the state obesity rate and the percentage of Healthy and Unhealthy food images from tweets of a state in 2019.

### 4.3.2 RQ2 Results

We examined how COVID-19 shutdown affect people's dietary behaviors from the following aspects. The first change we found is that the percentage of non-food images decreased in 2020 and 2021. We can find from Table 4.5 ,the percentage of non-food images in 2019 is more than 70% and this number drops to around 63% in 2020 and 2021. This change is significant based on z-test with p value smaller than 0.0001. This could because some restaurants closed dine-in options and people gathered less due to the social distancing policies during and partially after COVID. During and partially after shutdown, the percentage of healthy (Healthy and Def. Healthy) food has been shown to have significantly increased by z test ($p < 0.0001$). The percentage of healthy food in 2019 is 37.69%. This number increases to 39.99% in 2020 and drops a little to 39.32% in 2021. We also used the Pearson correlation to examine the relationship between the state-level dieting images in 2020 and in 2021 and the state obesity rate of 2019, because BRFSS didn't publish obesity data for 2020 and 2021 and we assume that the state-level obesity rate will not sharply change in a short period. However, no significant correlations were found at the $p = 0.05$ level.

Another change we saw from the Figure 4.4 is that users showed a significantly more positive sentiment when posting dietary behaviors on Twitter in 2019 than in 2020 and 2021, irrespective of the healthy level of the food. The mean sentiment score for 2019 was 0.395 (Stdev=0.425), for 2020 was 0.349 (Stdev=0.432), and for 2021 was 0.348 (Stdev=0.431). Another interesting finding from this graph is there are no significant differences were found between the sentiment to healthy and unhealthy food. Users didn't show a more positive sentiment when eating unhealthy food.

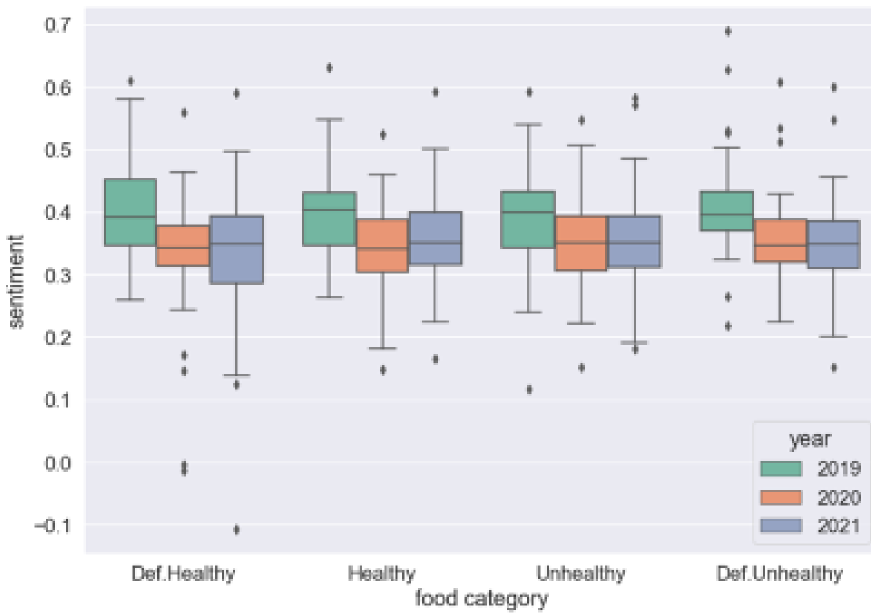



Figure 4.4: The sentiment to different health categories of food.

We later check the popularity change of different terms. The popularity change of breakfast, lunch, dinner and snack of a day are showing in the Figure 4.5. The COVID-19 shutdown didn't change the peak hour of three meals: breakfast peaks around 8 AM, lunch peaks around 11 AM, and dinner peaks around 6 PM. Under and partially after COVID-19 shutdown, users showed a trend to have breakfast and lunch later than before. In 2019, the most common time for breakfast is 7 AM & 8 AM, it shifts to 8 AM & 9 AM in 2020 and 2021. Interestingly, users showed a trend to have dinner earlier during and after shutdown than before. We also found people showed a trend to have snack more often during and after shutdown than before. In the 2019, two peak hours for snack is 2 PM & 8 PM. However, in 2020 and 2021, no peak hours for snack have been found.

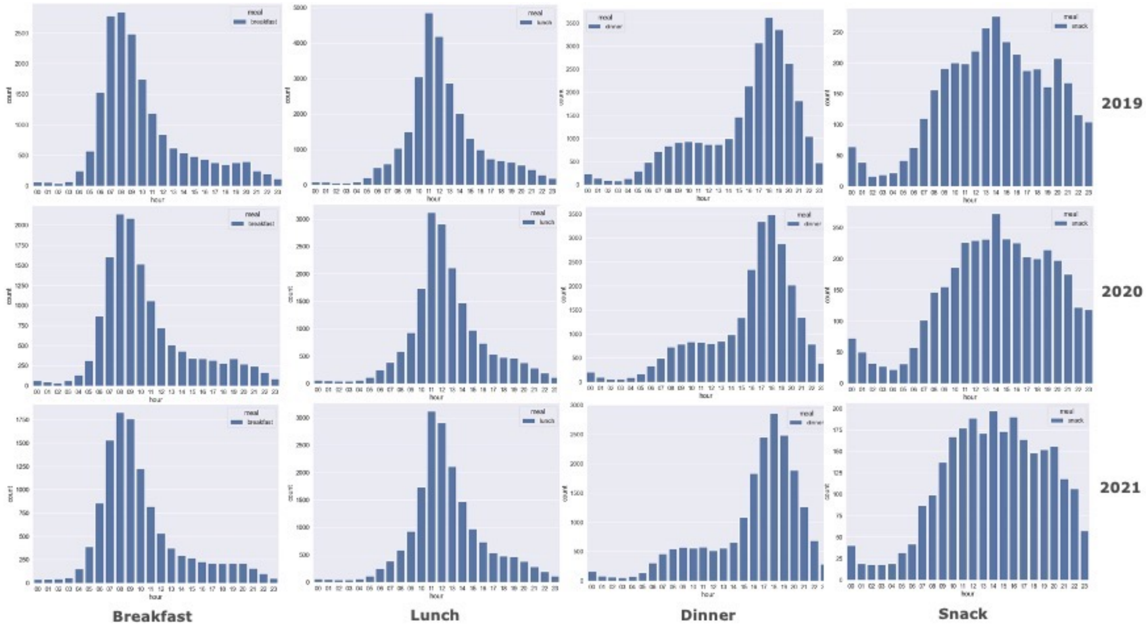



Figure 4.5: Temporal histograms showing the popularity of breakfast, lunch, dinner, snack by hour of the day. X-axis of each graph is the hour of a day (0-23). Y-axis the number of tweets contain that term.

Comparing to before COVID-19, the percentage of Def. Healthy food increased for breakfast, lunch and dinner; the percentage of Def. Unhealthy food decreased for breakfast, and dinner. People are also found to have more healthy snacks during and after COVID-19 shutdown. The percentage of Def. Healthy snack increased from 16.6% in 2019 to 19.3% in 2020 and 2021, and the percentage of Healthy snack increased from 21% in 2019 increased to 24.5% in 2020 and 23.1% in 2021. The percentage of Unhealthy snack decreased significantly after 2019, from 26% in 2019 to around 22% in 2020 and 2021. Other than that, we also found for all three years, people post more Def. Unhealthy food for snack time and at late night (1 AM - 3 AM) and more Def. Healthy food for lunch. All in all, people are found to have more Def. Healthy food, such as, vegetable and fresh food, during and after COVID-19 through the day.

### 4.3.3 Discussion & Limitations

In this paper, we present how the COVID-19 shutdown affected people's dietary behavior by examining Twitter images collected from six different hashtags related to people's dietary behaviors. Using computational methods, we present the relationship between state obesity rate and dieting images, and insights on how COVID-19 shutdown affect people's dietary behavior from the dieting healthiness, sentiment associated to dining, and eating patterns.

Our analysis proves image analyses could be used as a substantial resource to learn diet-related daily activities of the general public.[19] According to our analyses, people's dietary healthiness is related to state obesity rate in 2019. Users in higher obesity rate states post more high caloric Def. Unhealthy food than Def. Healthy food. It's critical to learn diet healthiness, since a leading cause of deaths and increased healthcare expenditures. It is also related to diet-related chronic diseases.[20] Improving diet quality is increasingly important for US during this pandemic, because nutrition intake impact our body's ability to resist infection.[21]

The COVID-19 pandemic dramatically impacted people's daily life. We observed from Twitter images that the COVID-19 shutdown has some health improvements in dietary healthiness during and partially after pandemic period. This finding is consistent with previous research conducted by Cosgrove & Wharton,[7] who did a survey to identify predictors of perceived dietary healthfulness changes over pandemic period in the US. Other than positive changes, certain adverse effects have also been found. The shutdown had a negative effect on people's emotions with respect to their dietary behaviors in this study. Consistent with previous studies,[15, 22] we found the average sentiment to dietary-related tweets is positive. However, the percentage of negative tweets increased by 25% during the pandemic. Studies also reported that participants experienced moderate to high stress levels during the pandemic period.[7, 4, 23] Increased psychological stress is commonly associated with binge eating disorder[24] and reduced diet quality[25]. These findings highlight the need of further explorations.

To figure out factors that affect user's emotions related to dietary-related behaviors, we used the keyword extraction algorithms, term frequency - inverse document frequency (TF-IDF), to summarize the topics of posts with different sentiment. We treat each tweet as a document and each word as a potential topic. Compared to before pandemic, we found the most common topics to negative tweets during pandemic are "miss" and "cook". Here is a modified example, "Even as an introvert this quarantine isn't easy ... i miss my friends and hood rat adventures... that said to hold myself accountable I'm commiting to 1 bar night and one small gathering a month!" Those negative emotions come from a general desire to return to previous life instead of diet-related concerns.

We also found some changes in people's eating patterns, including have breakfast & lunch later, dinner earlier and snack more often. Changes on eating pattern and dietary healthiness could be explained by COVID-19-related lifestyle changes. Transitioning from work in the office to working from home allows people to have more time in the morning for breakfast and prepare dinner earlier at home. People are reported prepare meals more at home in previous studies. Reduced exposure to current obesogenic environment helps people to have healthier diet choices.

Our study has some limitations. First, the image classification model could be improved from several aspects. We only collected the most popular ten types of food images for each health category. From Figure 4.2, we can find the health category model's performance on the Healthy and Unhealthy categories are not as satisfying as its performance on predicting the Def. Healthy and Def. Unhealthy food. Ramen has been classified as Healthy and Unhealthy. In our future study, more image data could be collected to train the image classification model. Other than that, our model may not do a good job in deciding the overall healthfulness of a plate of food. In our classification mode, we only crop the center of image to do the classification, however, more than one food could be found in one plate. Another limitation is our study has some potential bias considering the group of users of Twitter. Only 7.3% of Twitter users was reported younger than 18 years old. Although we found the COVID-19 shutdown has some positive impact on people's dietary healthfulness, another study reported its negative effect on children's dietary healthfulness. A longitudinal study on obese children in Italy found during the COVID-19 lockdown, potato chip, red meat, and sugary drink intakes increased significantly during the lockdown.[26] Future study should take these limitations into consideration. A more specific user group on social media could be targeted, like parents of kids, to learn health-related behaviors of those non-users.

#### 4.3.4 Conclusion

In this study, we used the social media data to understand how COVID-19 shutdown affected people's dietary behaviors. After using deep learning image classification procedures and text analysis methods, we found although the COVID-19 shutdown depresses user's emotions, diet healthiness has been found to improve during this pandemic. People have consumed healthier food throughout the day during and partially after shutdown than before, indicating people are making effort to improve their health during and after this unexpected health crisis. This study showed social media images could be used as a substantial resource to learn user's health-related behaviors and could be used as a preliminary work to further learn factors that impacting user's dietary behaviors or learn how COVID-19 affect users' healthy outcomes using social media data.

## REFERENCES


[1] Z. Shahid, R. Kalayanamitra, B. McClafferty, D. Kepko, D. Ramgobin, R. Patel, C. S. Aggarwal, R. Vunnam, N. Sahu, D. Bhatt, *et al.*, "Covid-19 and older adults: what we know," *Journal of the American Geriatrics Society*, vol. 68, no. 5, pp. 926–929, 2020.

[2] M. Górnicka, M. E. Drywień, M. A. Zielinska, and J. Hamułka, "Dietary and lifestyle changes during covid-19 and the subsequent lockdowns among polish adults: a cross-sectional online survey plifecovid-19 study," *Nutrients*, vol. 12, no. 8, p. 2324, 2020.

[3] N. Carroll, A. Sadowski, A. Laila, V. Hruska, M. Nixon, D. W. Ma, J. Haines, *et al.*, "The impact of covid-19 on health behavior, stress, financial and food security among middle to high income canadian families with young children," *Nutrients*, vol. 12, no. 8, p. 2352, 2020.

[4] C. L. Park, B. S. Russell, M. Fendrich, L. Finkelstein-Fox, M. Hutchison, and J. Becker, "Americansâ covid-19 stress, coping, and adherence to cdc guidelines," *Journal of general internal medicine*, vol. 35, no. 8, pp. 2296–2303, 2020.

[5] R. Sims, S. Gordon, W. Garcia, E. Clark, D. Monye, C. Callender, and A. Campbell, "Perceived stress and eating behaviors in a community-based sample of african americans," *Eating behaviors*, vol. 9, no. 2, pp. 137–142, 2008.

[6] C. W. Leung, E. S. Epel, L. D. Ritchie, P. B. Crawford, and B. A. Laraia, "Food insecurity is inversely associated with diet quality of lower-income adults," *Journal of the Academy of Nutrition and Dietetics*, vol. 114, no. 12, pp. 1943–1953, 2014.

[7] K. Cosgrove and C. Wharton, "Predictors of covid-19-related perceived improvements in dietary health: Results from a us cross-sectional study," *Nutrients*, vol. 13, no. 6, p. 2097, 2021.

[8] A. Perrin and J. Jiang, "About a quarter of us adults say they are âalmost constantlyâonline," *Pew Research Center*, vol. 14, 2018.

[9] C. Buntain, E. McGrath, J. Golbeck, and G. LaFree, "Comparing social media and traditional surveys around the boston marathon bombing.," in # *Microposts*, pp. 34–41, 2016.

[10] T. Nabity-Grover, C. M. Cheung, and J. B. Thatcher, "Inside out and outside in: How the covid-19 pandemic affects self-disclosure on social media," *International Journal of Information Management*, vol. 55, p. 102188, 2020.

[11] V. R. K. Garimella, A. Alfayad, and I. Weber, "Social media image analysis for public health," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 5543–5547, 2016.

[12] S. Abdullah, E. L. Murnane, J. M. Costa, and T. Choudhury, "Collective smile: Measuring societal happiness from geolocated images," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pp. 361–374, 2015.

[13] D. Murthy, A. Gross, and A. Pensavalle, "Urban social media demographics: An exploration of twitter use in major american cities," *Journal of Computer-Mediated Communication*, vol. 21, no. 1, pp. 33–49, 2016.

[14] D. Fried, M. Surdeanu, S. Kobourov, M. Hingle, and D. Bell, "Analyzing the language of food on social media," in *2014 IEEE International Conference on Big Data (Big Data)*, pp. 778–783, IEEE, 2014.

[15] V. Vydiswaran, D. Romero, X. Zhao, D. Yu, I. Gomez-Lopez, J. Lu, B. Iott, A. Baylin, P. Clarke, V. Berrocal, *et al.*, "" bacon bacon bacon": Food-related tweets and sentiment in metro detroit," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 12, 2018.

[16] S. Targ, D. Almeida, and K. Lyman, "Resnet in resnet: Generalizing residual architectures," *arXiv preprint arXiv:1603.08029*, 2016.

[17] H. Alfawaz, S. M. Yakout, K. Wani, G. A. Aljumah, M. G. Ansari, M. N. Khattak, S. D. Hussain, and N. M. Al-Daghri, "Dietary intake and mental health among saudi adults during covid-19 lockdown," *International Journal of Environmental Research and Public Health*, vol. 18, no. 4, p. 1653, 2021.

[18] C. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 8, 2014.

[19] A. Park, C. Li, J. Bowling, Y. Ge, and M. Dulin, "Diet, weight loss, fitness, and health related image sharing using twitter: an observation study," in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 2049–2053, IEEE, 2020.

[20] W. Raghupathi and V. Raghupathi, "An empirical study of chronic diseases in the united states: a visual analytics approach to public health," *International journal of environmental research and public health*, vol. 15, no. 3, p. 431, 2018.

[21] E. Prompetchara, C. Ketloy, and T. Palaga, "Immune responses in covid-19 and potential vaccines: Lessons learned from sars and mers epidemic," *Asian Pacific journal of allergy and immunology*, vol. 38, no. 1, pp. 1–9, 2020.

[22] A. Park and D. S. West, "Understanding socio-cultural factors related to obesity: Sentiment analysis on related tweets," *Online Journal of Public Health Informatics*, vol. 11, no. 1, 2019.

[23] M. Song, "Psychological stress responses to covid-19 and adaptive strategies in china," *World development*, vol. 136, p. 105107, 2020.

[24] A. Dingemans, U. Danner, and M. Parks, "Emotion regulation in binge eating disorder: A review," *Nutrients*, vol. 9, no. 11, p. 1274, 2017.

[25] K. Khaled, F. Tsofliou, V. Hundley, R. Helmreich, and O. Almilaji, "Perceived stress and diet quality in women of reproductive age: A systematic review and meta-analysis," *Nutrition journal*, vol. 19, no. 1, pp. 1–15, 2020.

[26] A. Pietrobelli, L. Pecoraro, A. Ferruzzi, M. Heo, M. Faith, T. Zoller, F. Antoniazzi, G. Piacentini, S. N. Fearnbach, and S. B. Heymsfield, "Effects of covid-19 lockdown on lifestyle behaviors in children with obesity living in verona, italy: a longitudinal study," *Obesity*, vol. 28, no. 8, pp. 1382–1385, 2020.

# CHAPTER 5: CONCLUSIONS

## 5.1 Conclusion

This dissertation is motivated by the fact that obesity is a major public health issues in today's society and the fact that social media has a role in affecting people's health-related behaviors and could be successfully use for health-related studies. In this dissertation, we provide an insightful understanding of how social media could be used in obesity-related studies, what and how online social factors affect people's obesity-related behaviors and computational methods to utilize social media data to learn those behaviors.

By systematically reviewing studies that using social media data to learn obesity, we found three different roles that social media served: data collection, intervention pathway, and ancillary resources. Ten online social factors were identified relating to behaviors: gender, social tie, social support, source credibility, social movements, social sharing behaviors, stigma, policy, geo-cultural factors, school environment, and obesogenic environment. Those factors affect users' obesity-related behaviors from four levels: individual, interpersonal, online social environment and connection to the real world.

Adopting social media data, we also showed the feasibility of using social media to study obesity-related behaviors. Yelp and MFP data were used to characterize state level food (obesogenic) environment. The importance of food environment to residents' dietary behaviors was stated by showing disparities of available food between states with different obesity rate. We further built a state-level obesity rate prediction model using the aforementioned data. The Pearson correlation coefficient between the predicted obesity rate and the real obesity rate reaches 0.796. The findings on people's dietary behaviors change over COVID-19 shutdown showed the social media images could also be used as a substantial resource to learn people's dietary behaviors during health crisis. People are found consumed healthier food throughout the day during and partially after the shutdown.

Social media has a growing potential to improve public health outcomes. To conclude this study, our findings are beneficial to future efforts to tackle the obesity pandemic from the following aspects. First, the social-ecological model provides guidelines for health practitioners to design obesity-related intervention studies. As each online social factor has a potential effect on users' obesity-related behaviors, a better understanding of each factor's role helps researchers better utilize social media to help participants do weight management. Second, our findings can help public health agencies to improve health literacy through online social networks. Considering online social factors (source credibility, social movements, social sharing behaviors, and stigma) at the online social environment level, public health agencies should maintain a healthy online social environment by leading a non-stigma and healthy weight social movements. When using online social media to publish obesity-related education information, public health agencies should develop a well-defined process to address users' concerns of source credibility [11] and use users' social sharing behaviors to disseminate information better. Third, our study demonstrated the approach to effectively and efficiently using social media data to predict local obesity rates. Public health practitioners could build a pipeline based on our suggested method to monitor and forecast the obesity rate change of areas that face the risk of increasing obesity rate and make an early warning or take proactive actions. Fourth, our study used the image on Twitter to study users' obesity-related lifestyles. This study filled the gap that various types of social media (e.g., text, image, video) on electronic health literacy are understudied [11] and showed the potential of using multiple types of social media data to do health-related studies in the future.

## 5.2 Future Study

However, certain limitations exist in this study. First, we only used three databases (PubMed, ACM, and ScienceDirect) to examine how social media is used in obesity-related studies. We believe if other related database, like Scopus, could be included in the study, we might have additional findings. Second, in our model that use social media data to predict state obesity rate, the method of how to estimate caloric density of each popular dish could be improved. In our study, we didn't capture the actual amount of food, which is also important for us to calculate the calorie of a dish. What's more, our analysis that use social media data to learn people's obesity-related behaviors are conducted at state level. We didn't include other factors in our analysis. Other factors, such as social economic status has been found related to users' obesity-related behaviors.

To address the limitations mentioned earlier and existing gaps in this field, we propose to do our future study from the following directions. First, recognizing the association between socioeconomic status and obesity, we could further our analysis by combining demographic data and social media data with learning users' obesity-related behaviors at the zip code level. A more granular analysis could give us a better insight into how the local food environment and social-economic status affect users' obesity-related behaviors. In another study, we could do to learn users' obesity-related behaviors using Reddit data. Reddit is a popular social networking and online gathering platform. Features of Reddit (e.g., anonymous and throwaway accounts) make a well-utilized platform for learning stigmatized illness.[12] Using data from different social media platforms could give us a more comprehensive understanding of users' online obesity-related behaviors. Last but not least, we could extend our study by using other types of social media data to study users' obesity-related behaviors. From text to the short video, ongoing trends have been found when users share messages through social media. Analyzing these mixed media, we can comprehensively understand how to use social media data to understand users' health behaviors.

Continued efforts are needed to counter the obesity pandemic. As the prolonged use of social media, we believe this study benefits public health practitioners by providing guidelines and solutions to strengthen obesity-related education and improve public health outcomes.

# REFERENCES


[1] C. J. Murray, M. Ng, "Nearly one-third of the world's population is obese or overweight, new data show", Archived at:https://web.archive.org/web/20210830185155/http://www.healthdata.org/news-release/nearly-one-third-world\%E2\%80\%99s-population-obese-or-overweight-new-data-show.

[2] "Adult Obesity Facts", https://www.cdc.gov/obesity/data/adult.html, Archived at:https://web.archive.org/web/20210917071618/https://www.cdc.gov/obesity/data/adult.html.

[3] C. E. Caspi, G. Sorensen, S. Subramanian, and I. Kawachi, "The local food-environment and diet: a systematic review," *Health & place*, vol. 18, no. 5,pp. 1172-1187, 2012.

[4] R. M. Puhl and C. A. Heuer, "The stigma of obesity: a review and update," *Obesity*, vol. 17, no. 5, pp. 941-964, 2009.

[5] E. A. Finkelstein, J. G. Trogdon, J. W. Cohen, and W. Dietz, "Annual Medical Spending Attributable To Obesity: Payer-And Service-Specific Estimates," *Health Aff.*, vol. 28, no. 5, pp. w822-w831, Sep. 2009.

[6] Maher, E, "Healthrelated quality of life of severely obese children and adolescents," *Child: Care, Health and Development*, 30(1), pp.94-95, 2004.

[7] D. M. Boyd and N. B. Ellison, "Social Network Sites: Definition, History, and Scholarship," *J. Comput. Commun.*, vol. 13, no. 1, pp. 210-230, 2007.

[8] A. Perrin and J. Jiang, "About a quarter of US adults say they are 'almost constantly'online," *Pew Research Center*, 14, 2018

[9] A. K. Munk, M. S. Abildgaard, A.; Birkbak, and M. K. Petersen, "(Re-)Appropriating Instagram for Social Research: Three methods for studying obesogenic environments," in *Proceedings of the 7th 2016 International Conference on Social Media  Society*, 2016, p. 19

[10] E. E. Kent, A. Prestin, A. Gaysynsky, K. Galica, R. Rinker, K. Graff, and W.-Y. S. Chou, ""obesity is the new major cause of cancer": connections be-tween obesity and cancer on facebook and twitter," *Journal of Cancer Education*, vol. 31, no. 3, pp. 453-459, 2016.

[11] A. Park, J. Bowling, S. George, C. Li, and S. Chen, "Adopting social media for improving health: opportunities and challenges," *North Carolina Medical Journal*, vol. 80, no. 4, pp. 240-243, 2019.

[12] A. Park, M. Conway, and A.T. Chen, "Examining thematic similarity, difference, and membership in three online mental health communities from Reddit: a text mining and visualization approach," *Computers in human behavior*, 78, pp.98-112, 2018.

# APPENDIX A: The table of caloric density for category

Table A.1: The caloric density for category.

| category | Caloric density | category | Caloric density | category | Caloric density | category | Caloric density |
|---|---|---|---|---|---|---|---|
| Cheesesteaks | 23.45 | Beer Bar | 6.12 | Lebanese | 5.47 | Szechuan | 4.88 |
| Chicken Wings | 17.31 | Polynesian | 6.11 | Cuban | 5.47 | Teppanyaki | 4.87 |
| Wraps | 16.45 | Tex-Mex | 6.07 | Caterers | 5.46 | Conveyor Belt Sushi | 4.86 |
| Food Stands | 15.6 | Butcher | 6.05 | Grocery | 5.45 | Organic Stores | 4.85 |
| Herbs & Spices | 11.13 | Hungarian | 6.03 | Tea Rooms | 5.43 | Portuguese | 4.85 |
| Cooking Schools | 10.73 | Wine Bars | 6.02 | Venezuelan | 5.43 | Buffets | 4.85 |
| Bed & Breakfast | 10.49 | Convenience Stores | 6.01 | Distilleries | 5.42 | Honduran | 4.85 |
| Wine Tasting Classes | 10.39 | Japanese | 6.01 | Salad | 5.42 | Beverage Store | 4.85 |
| Pretzels | 9.9 | Coffee & Tea | 5.96 | Comfort Food | 5.42 | International Grocery | 4.83 |
| Speakeasies | 9.4 | Mediterranean | 5.95 | Halal | 5.41 | Ramen | 4.82 |
| Empanadas | 9.17 | Desserts | 5.93 | Falafel | 5.4 | Asian Fusion | 4.81 |
| Wine Tasting Room | 7.81 | Polish | 5.89 | New Mexican Cuisine | 5.39 | Irish | 4.8 |
| Piano Bars | 7.63 | Street Vendors | 5.87 | Beer Gardens | 5.39 | Kosher | 4.78 |
| Donuts | 7.47 | Bakeries | 5.87 | Caribbean | 5.38 | Olive Oil | 4.78 |
| Korean | 7.37 | Hot Dogs | 5.83 | Pop-Up Restaurants | 5.36 | Indonesian | 4.78 |
| Shaved Ice | 7.37 | Izakaya | 5.83 | Spanish | 5.34 | Brasseries | 4.77 |
| Vegan | 7.29 | Supper Clubs | 5.83 | Sports Bars | 5.34 | Afghan | 4.77 |
| Diners | 7.28 | Guamanian | 5.81 | Local Flavor | 5.32 | Bars | 4.77 |
| Wineries | 7.25 | Latin American | 5.81 | Puerto Rican | 5.32 | Kombucha | 4.77 |
| Candy Stores | 7.24 | Armenian | 5.8 | Breakfast & Brunch | 5.28 | Mongolian | 4.74 |
| Game Meat | 7.21 | Ice Cream & Frozen Yogurt | 5.8 | Poutineries | 5.27 | Cantonese | 4.73 |
| Argentine | 7.21 | Dinner Theater | 5.8 | Vegetarian | 5.25 | Poke | 4.71 |
| Popcorn Shops | 7.11 | Waffles | 5.78 | Filipino | 5.25 | Laotian | 4.69 |
| Barbeque | 7.06 | Brazilian | 5.77 | Whiskey Bars | 5.23 | Australian | 4.68 |
| Acai Bowls | 6.82 | Gastropubs | 5.76 | African | 5.21 | Turkish | 4.58 |
| Macarons | 6.81 | Sandwiches | 5.75 | Ethiopian | 5.2 | Kebab | 4.54 |
| Middle Eastern | 6.79 | Tiki Bars | 5.73 | Vietnamese | 5.2 | Bubble Tea | 4.54 |
| Pubs | 6.77 | Pancakes | 5.72 | Hot Pot | 5.2 | Japanese Curry | 4.53 |
| Chocolatiers & Shops | 6.73 | Meat Shops | 5.72 | Nicaraguan | 5.19 | Seafood Markets | 4.52 |
| Persian/Iranian | 6.71 | Seafood | 5.71 | Salvadoran | 5.18 | Taiwanese | 4.48 |
| Basque | 6.6 | Irish Pub | 5.71 | Tapas Bars | 5.16 | Chinese | 4.47 |
| Pakistani | 6.6 | Gelato | 5.69 | Peruvian | 5.15 | Tacos | 4.45 |
| Hawaiian | 6.59 | Gluten-Free | 5.67 | Breweries | 5.14 | Singaporean | 4.43 |
| Eritrean | 6.59 | Southern | 5.66 | Noodles | 5.14 | Imported Food | 4.42 |
| Cafes | 6.56 | Food Delivery Services | 5.65 | Burmese | 5.1 | Himalayan/Nepalese | 4.4 |
| Internet Cafes | 6.55 | Specialty Food | 5.65 | South African | 5.1 | Cooking Classes | 4.36 |
| Burgers | 6.53 | Syrian | 5.64 | Fish & Chips | 5.1 | Indian | 4.36 |
| Czech | 6.52 | Pizza | 5.63 | Juice Bars & Smoothies | 5.09 | Public Markets | 4.33 |
| Italian | 6.51 | Live/Raw Food | 5.63 | Food | 5.06 | Cupcakes | 4.29 |
| American (Traditional) | 6.47 | Drive-Thru Bars | 5.62 | Themed Cafes | 5.05 | Hong Kong Style Cafe | 4.28 |
| Fast Food | 6.46 | Scandinavian | 5.61 | Colombian | 5.05 | Moroccan | 4.26 |
| Coffee Roasteries | 6.41 | Dominican | 5.61 | American (New) | 5.04 | Cambodian | 4.24 |
| Soul Food | 6.33 | Sushi Bars | 5.6 | French | 5.04 | Malaysian | 4.24 |
| Chicken Shop | 6.32 | Restaurants | 5.58 | Cafeteria | 5.03 | Brewpubs | 4.23 |
| Patisserie/Cake Shop | 6.29 | Cocktail Bars | 5.58 | Tapas/Small Plates | 5.02 | Dim Sum | 4.21 |
| Delis | 6.27 | Bagels | 5.57 | Cheese Shops | 5.02 | Shanghainese | 4.08 |
| Mexican | 6.24 | Food Court | 5.57 | Pan Asian | 4.97 | Thai | 4.08 |
| Russian | 6.19 | Health Markets | 5.57 | Cideries | 4.97 | Fruits & Veggies | 3.93 |
| Personal Chefs | 6.19 | Food Trucks | 5.52 | Creperies | 4.95 | Hainan | 3.92 |
| British | 6.18 | Dive Bars | 5.51 | Cajun/Creole | 4.94 | Austrian | 3.88 |
| Arabian | 6.18 | Smokehouse | 5.5 | Sicilian | 4.93 | Food Tours | 3.82 |
| Beer, Wine & Spirits | 6.17 | Modern European | 5.49 | Fondue | 4.91 | Shaved Snow | 3.68 |
| Pasta Shops | 6.15 | Farmers Market | 5.48 | Greek | 4.9 | Gay Bars | 2.05 |
| German | 6.13 | Do-It-Yourself Food | 5.47 | Soup | 4.89 | Honey | 1.94 |