Unraveling the Complexities of the Obesity Epidemic in the United States: A Comprehensive Research Exploration

Introduction

In the contemporary landscape of American society, the pervasive challenge of obesity has evolved into a multifaceted issue with profound implications, transcending mere health concerns to infiltrate various aspects of daily life and societal structures. This study aims to delve deeper into the underlying complexities of obesity by meticulously analyzing extensive datasets sourced from reputable institutions such as the National Health and Nutrition Examination Survey (NHANES), Behavioral Risk Factor Surveillance System (BRFSS), and the Centers for Disease Control and Prevention (CDC). Obesity isn't just about individuals being overweight; it's intricately linked to a myriad of chronic health conditions, placing a significant burden on healthcare systems and perpetuating health disparities. Moreover, its socio-economic ramifications ripple through communities, impeding productivity, and straining public resources. By scrutinizing factors ranging from genetics and environmental influences on socio-economic disparities and behavioral patterns, we endeavor to unravel the intricate web of causes contributing to the obesity epidemic. Our research problem centers on identifying the underlying factors contributing to the rise in obesity rates in the United States and proposing evidence-based strategies that can be implemented by policymakers, healthcare professionals, and other stakeholders to mitigate the obesity epidemic and its associated health and socioeconomic consequences.

Literature Review

X. Pang and team (2019) develops a machine learning model named XGBoost to understand the patterns in Obesity among children. In this research, authors considered most common factors that contribute to childhood obesity such as weight, height, race, and ethnicity. Through the developed machine learning model, they accurately predicted possible obesity among the children of age between 24-36 months with accuracy of 81%. This research is restricted to understand the obesity in children and does not consider the extra factors that could led to obesity in people with age. The interesting assumption in this research is that authors also included factors such as body temperature and respiratory rate which practically could be the best predictors for obesity even in people of any age.[1] X. Pang team's research is proven to analyze the obesity patterns in children with age between 24-36 months by developing the machine learning model with accuracy of 81%. This is achieved by considering factors such as weight, height, race, ethnicity along with uncommon but important factors such as body temperature and respiratory rate. But this research is limited to predicting the obesity possibilities and understanding patterns in children and does not consider the factors that could add-on with age contributing to obesity such as lifestyle, diet, and living environment. The research highlights this issue but did not consider which is included in our research to even understand some new factors that lead to obesity with age. Our research along with understanding the underlying factors for obesity, also filters the necessary suggestions to people, policy makers and drug developers after analyzing the patterns to eradicate obesity. In 2023, H. A. Obaid and their team worked on developing a framework that dynamically suggests personalized exercise training and improvements in physical activity for individuals suffering from obesity and aiming for weight loss. This framework considers various factors including age, BMI, and others. The research underscores the significance of monitoring caloric consumption, weight tracking, and engaging in physical activity, identifying different thresholds where obesity can adversely affect health. The developed framework has been shown to be 93% efficient in achieving its defined objectives. However, the research did not delve into analyzing the crucial underlying factors contributing to obesity or weight loss. Nonetheless, it incorporates fuzzy logic to provide dynamic suggestions to individuals based on their current condition, offering tailored exercise training and strategies for enhancing physical activity.[2] H. A. Obaid's team developed a framework, based on fuzzy logic, that provides suggestions for individuals with obesity based on their current conditions, offering suitable exercise training along with physical activity, achieving a proven efficiency of 93% with its objectives. However, this research is limited to suggestions through its fuzzy logic framework and does not give importance to understanding the underlying factors that lead people to this problem. Understanding these factors and providing suggestions based on them might increase the accuracy of the model even further, leading to more efficient suggestions for consumers. This issue is primary in the research we proposed, initially analyzing the underlying factors that lead to obesity across different age groups and providing dynamic suggestions to individuals, policymakers, and, most importantly, drug developers to achieve positive results with clinical trials aimed at eradicating obesity from its roots.

K. Alam and their team (2023) developed an android application aimed at guiding people to track their calorie consumption and physical activity, thus promoting a healthy lifestyle, and preventing obesity. The research addresses the general problem of obesity, often resulting from an imbalance between calorie intake and physical activity. The team integrated a deep learning model to analyze food consumption and physical activity tracked by built-in mobile sensors, providing necessary steps to prevent obesity. An interesting finding is that the application achieves a 99% accuracy rate in fulfilling its objectives through the implication of the deep learning model. However, the developed model focuses solely on providing appropriate suggestions to individuals and healthcare professionals. It falls short in addressing the eradication of the problem by suggesting policies to policymakers. Such policies have the potential to eliminate this issue, which differs from the focus of our research.[3] The research highlights that obesity is mainly caused by an imbalance between food consumption and physical activity. However, this assumption is just one aspect considered in our research. In addition to this assumption, we broaden our analysis to explore other factors contributing to obesity. While the application developed by this team offers appropriate suggestions to individuals and healthcare professionals, like our research, we extend our recommendations to policymakers to address the problem at its core and to drug developers to create more efficient drugs. In summary, while this research acknowledges the imbalance between dietary habits and physical activity as a significant factor in obesity, our research primarily focuses on identifying patterns of various factors contributing to obesity. We provide filtered suggestions to individuals with obesity, healthcare professionals, drug developers, and policymakers, aiming to address the issue comprehensively.

George has put together a detailed study on how obesity can be a root cause of many life-threatening diseases and the categories by which these obesities are caused. He also stresses the various secretions in different parts of the human body caused by these types of fat accumulation. These secretions at different tissues are leading to various health diseases, both directly and indirectly. The main factor for this categorization is BMI and each value categorizes the risk and their health conditions. The overall focus of this research is on how all these together are reducing the life span of people and in some cases causing sudden deaths in humans recently. It also speaks about the benefits of weight loss from the medical aspects of how each step towards the reduction of weight helps improve their health. They talk based on previous proven evidence and what percentage of it was a success or failure.[4] This research will help us understand the different categories obesity is caused and give a better explanation of underlying factors that lead to obesity in people. This will be the primary step of our research to identify the different patterns that are causing obesity in different age groups of people. Through this research, the author tries to focus his study mostly on medical terms and how each enzyme shows changes based on the activities and drugs, but we will use this evidence of success to provide suggestions and assurance to people to motivate them to focus on the health aspects and that it over the years can be dangerous. This will also add up to the suggestions we can provide to policy makers and drug developers for the demand for more specific obesity drugs based on different categories.

N. A. Wahab and their team (2022) developed a web-based application named HealthyMe, which provides recommendations to obese patients regarding their food selection, exercise, and overall lifestyle. The authors' approach to providing appropriate suggestions is based on three principles: the principle of cause and effect, attractiveness, and tailoring. Through this method, the authors observed that consumers were attracted to the services, which helped them reduce their weight and adopt a healthier lifestyle. However, this application's services were limited to obese patients and did not filter suggestions that could be useful for policy makers or drug developers. Nonetheless, the interesting factor lies in the approach used to provide suggestions only after understanding the individual's food selection and lifestyle, making the service more efficient.[5] Through this research, the authors developed a web-based application that provides appropriate suggestions to obese patients based on their food consumption, exercise, and overall lifestyle. While current lifestyle is often the primary factor influencing an individual's health, there may be hidden factors such as genetics that contribute to obesity, warranting investigation and filtering suggestions accordingly. This is where our research initially identifies the factors responsible for obesity. The HealthyMe application is limited to providing result-oriented suggestions to obesity patients based on their food consumption and lifestyle. However, our research, after analyzing these factors, offers necessary suggestions to various segments of the population, including patients, drug developers, healthcare professionals, and policy makers.

The strengths of the studies include the introduction of physiological indicators to enhance obesity prediction in children, the utilization of fuzzy logic for dynamic, personalized exercise recommendations, the development of a practical tool for individuals to monitor health habits, a detailed exploration of obesity's medical risks based on BMI categories, and the creation of a web-based app emphasizing personalized lifestyle recommendations. However, weaknesses include a limited focus on early childhood without addressing obesity factors in older populations, overlooking underlying causes of obesity by focusing solely on physical activity, failure to address broader societal or policy interventions for obesity, lacking a focus on preventative strategies and broader determinants of obesity, and services not being extended to influence policy makers or drug development sectors.

The existing literature on obesity offers a multifaceted understanding of the epidemic, with diverse approaches to prediction, intervention, and prevention. While some studies emphasize the importance of considering physiological indicators in childhood obesity prediction, others focus on personalized exercise recommendations using fuzzy logic. However, notable gaps exist, such as the oversight of broader lifestyle and environmental factors in childhood obesity studies and neglecting biological, genetic, and socio-economic determinants in intervention frameworks. Additionally, some studies miss opportunities to address systemic issues through policy-level interventions, while others fail to translate findings into preventive strategies or broader health policies. Despite these shortcomings, the literature provides valuable insights into obesity management, highlighting the need for comprehensive approaches that consider both individual and systemic factors. Competing arguments within this literature may revolve around the efficacy of different intervention strategies and the balance between individualized and systemic approaches to obesity prevention.

Hypotheses

Our hypothesis posits that the surge in obesity rates within the United States is not a simple phenomenon but rather the result of a complex interplay of various underlying factors. We contend that unraveling these multifaceted determinants is paramount for devising effective strategies to prevent and manage obesity. Specifically, our proposal suggests that factors such as lifestyle choices, dietary habits, genetic predispositions, socioeconomic status, and environmental influences collectively contribute to the escalating obesity epidemic spanning diverse age groups. By delving deeply into these interconnected factors, our aim is to uncover novel insights and provide evidence-based recommendations to address the obesity crisis comprehensively, considering its profound implications on both health outcomes and socioeconomic well-being. Through this comprehensive analysis, we seek to inform targeted interventions that can mitigate the detrimental effects of obesity and promote healthier living across populations.

Our study hypothesizes several key factors contributing to the escalating obesity rates in the United States. First, we posit that unhealthy dietary habits, characterized by increased consumption of processed foods, high-calorie beverages, and sugary snacks, positively correlate with obesity rates. Additionally, sedentary lifestyles, marked by low levels of physical activity due to factors like reduced physical education in schools and heightened screen time, are expected to exacerbate the obesity epidemic. Socioeconomic factors, including income level and access to healthcare and nutritious food, are anticipated to play a crucial role, with individuals from lower socioeconomic backgrounds facing a higher risk of obesity due to limited resources for healthy living. Genetic predispositions are also expected to influence individual susceptibility to obesity, interacting with environmental factors. Environmental factors such as neighborhood walkability and food availability are projected to contribute to regional disparities in obesity rates. Furthermore, psychological factors like stress levels and societal and cultural attitudes toward body image and food consumption are hypothesized to impact obesity rates. Our research aims to fill a gap in the existing literature by integrating these factors into a holistic understanding of obesity trends across different age groups and proposing evidence-based recommendations for obesity prevention and management targeting individuals, policymakers, healthcare professionals, and drug developers, thus offering novel solutions for combating this pressing public health issue.

Data

The dataset being analyzed in this study offers a detailed perspective on the Body Mass Index (BMI) of individuals across different states within the United States. It serves as a valuable resource for understanding the prevalence of obesity, as individuals with a BMI surpassing 35 are identified as obese, providing a clear and stringent classification criterion. Moreover, the dataset goes beyond simple BMI measurements by meticulously organizing individuals according to various demographic variables. These include gender, household income, age, education level, and ethnicity, among others. This level of granularity allows researchers to conduct a thorough examination of obesity prevalence and its associations with different demographic factors. By categorizing individuals based on these parameters, the dataset facilitates a nuanced exploration of the complex interplay between obesity and various social and economic factors, shedding light on disparities and trends across different population groups within the United States.

The unit of analysis in this dataset is individual BMI measurements across different states in the USA. Everyone is categorized based on their BMI, with those surpassing the threshold of 35 classified as obese. The dataset allows for the examination of obesity prevalence and its correlates within different demographic groups, offering insights into the multifaceted nature of obesity across diverse populations.

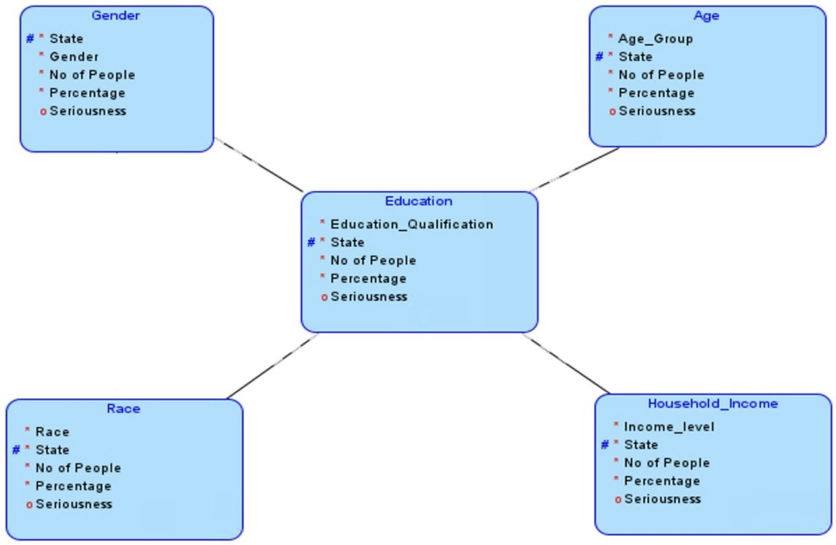

Fig 1. Data Modular Diagram

The data utilized in this research is sourced from the Behavioral Risk Factor Surveillance System (BRFSS), a widely recognized system of health-related telephone surveys in the United States. BRFSS collects state-level data on various health-related risk behaviors, chronic health conditions, and preventive service utilization among U.S. residents. Operated under the guidance of the Centers for Disease Control and Prevention (CDC), BRFSS ensures adherence to stringent data collection guidelines to maintain accuracy and reliability.

Strengths of the dataset include its comprehensive coverage of health-related behaviors and conditions across different states in the USA, as well as its adherence to CDC guidelines, ensuring high standards of data quality and reliability. The meticulous organization of individuals based on various demographic parameters enables detailed exploration of obesity prevalence and its correlates. However, potential weaknesses include the generalization of results due to data aggregation and the exclusive reliance on BMI as a metric for obesity classification. To mitigate these limitations, robust statistical methods and sampling algorithms are employed, enhancing the validity and reliability of the analysis. Overall, while the dataset offers valuable insights into obesity prevalence and its correlates, researchers should be cautious of its limitations and exercise prudence in interpretation.

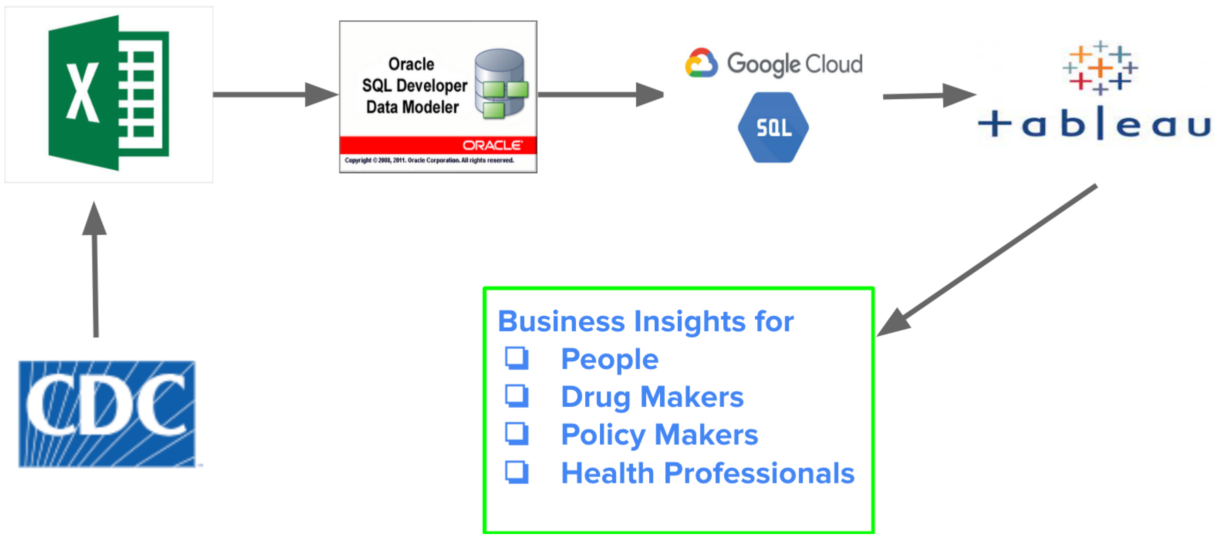

Fig 2. The flow of the project

This project flow outlines a data processing and analysis pipeline. It begins with collecting the data from CDC official website about the obesity rates of the USA and then scraping them to an Excel workbook, moves through Oracle SQL to create tables to have a CICD pipeline setup for zero touch automation, Developer Data Modeler for database design, then to Google Cloud for SQL-based data management, and finally into Tableau for visualization. The aim is to generate business insights for four key stakeholder groups: people (likely consumers or the public), drug makers, policy makers, and health professionals, indicating a focus on pharmaceutical or public health data.

Variables and econometric model

The dependent variable in this study is the percentage of obese individuals aggregated for every unique value of each factor. These factors include gender, race, education level, age group, and household income. The percentage of obese individuals serves as an indicator of obesity prevalence within each subgroup defined by these factors. By examining variations in obesity rates across different demographic categories, researchers can discern patterns and trends in obesity prevalence and identify disparities among various population groups.

The key independent variables in this study are gender, race, education level, age group, and household income. These variables represent different demographic characteristics of the population under study and are hypothesized to influence obesity prevalence. Each independent variable has multiple unique values or categories, such as 'Male' and 'Female' for gender, and 'White', 'Black', 'Asian', etc., for race. To group age and household as these are measured columns, we have considered $2^k \geq n$, which divides age into six groups: 18-24, 24-34, 35-44, 45-54, 55-64, and 65+. Household income has seven groups: ‘less than $15,000’, ‘$15,000-$24,999’, ‘$25,000-$34,999’, ‘$35,000-$49,999’, ‘$50,000-$99,999’, ‘$100,000-$199,999’, and ‘$200,000’.To identify patterns, understand the common characteristics and habits of obese people, groups are created for every factor based on BMI data collected from individuals state wise in the USA. All factors are equally important in identifying patterns and understanding the characteristics of obese people. The percentage of obese people aggregated for every unique value of factor is dependent variable on those different values of variables. Different factors contributing to obesity are independent variables. These variables are crucial for understanding how demographic factors contribute to variations in obesity rates across different population subgroups. In this study, there are no explicitly defined control independent variables. However, the model may include additional factors or covariates that are controlled for to isolate the effects of the key independent variables on the dependent variable (obesity prevalence). These control variables could include factors such as geographic location, access to healthcare, dietary habits, physical activity levels, and genetic predispositions. By controlling for these variables, researchers aim to minimize confounding effects and ensure that any observed associations between the key independent variables and obesity prevalence are robust and reliable.

The model employed in this study is a multivariate regression model, where the percentage of obese individuals for each unique value of the independent variables (gender, race, education level, age group, and household income) serves as the dependent variable. The model aims to identify and quantify the relationship between these independent variables and obesity prevalence while controlling for potential confounding factors. By analyzing variations in obesity rates across different demographic categories, the model elucidates the impact of various demographic factors on obesity prevalence, providing valuable insights for public health interventions and policy-making efforts aimed at reducing obesity disparities.

To comprehensively analyze the relationship between obesity rates and the identified predictors, we have employed a multilevel logistic regression model. This analytical framework is well-suited to our study due to the binary nature of our dependent variable (obesity status: obese or not obese) and the hierarchical structure of our dataset, with individuals nested within states. By utilizing this approach, we can effectively capture variability at both the individual and state levels, providing a more accurate depiction of the factors influencing obesity across different regions.

In our logistic regression model, we estimate the probability of obesity as a function of several key predictors, including access to recreational facilities, socioeconomic factors, dietary habits, and control variables. Incorporating state as a random effect enables us to account for unobserved heterogeneity attributed to regional differences, such as climate, state policies, and cultural norms related to diet and physical activity. Interpreting the coefficients derived from this model will allow us to discern the strength and direction of the relationship between each independent variable and the likelihood of obesity, while controlling for other factors included in the model. Positive coefficients indicate that an increase in the predictor variable is associated with higher odds of obesity, whereas negative coefficients suggest protective factors that reduce the likelihood of obesity. This robust analytical approach offers valuable insights into how various environmental, economic, and personal factors interact to influence obesity rates, thereby informing the development of targeted interventions aimed at effectively reducing obesity prevalence across diverse community settings.

Empirical analysis and findings

The key variables under examination in this study include gender, race, education level, age group, and household income. These variables are essential factors contributing to obesity in individuals within the research data. To analyze the severity of obesity problems across different states in the USA, mean values of the percentage of obese individuals aggregated based on different groups of these factors are considered. For instance, states with a percentage of obese individuals greater than 33% for gender, 34% for education, 33% for age, 20% for race, and 34% for income are classified as facing serious obesity problems. Descriptive statistics, such as means, standard deviations, and percentiles, are calculated for these variables to provide a comprehensive overview of obesity prevalence and its correlates across different demographic categories within the dataset. The discrete choice model employed in this study facilitates the identification of patterns associated with higher BMI in individuals and provides insights crucial for formulating recommendations to address obesity effectively. By analyzing the aggregated percentage of obese individuals across different demographic groups and states in the USA, the model offers a comprehensive understanding of the severity of obesity problems and the underlying factors contributing to them. The model's ability to discern patterns and associations between demographic variables and obesity prevalence enhances its explanatory power, enabling researchers to identify potential interventions and policy measures to combat obesity at both individual and population levels.

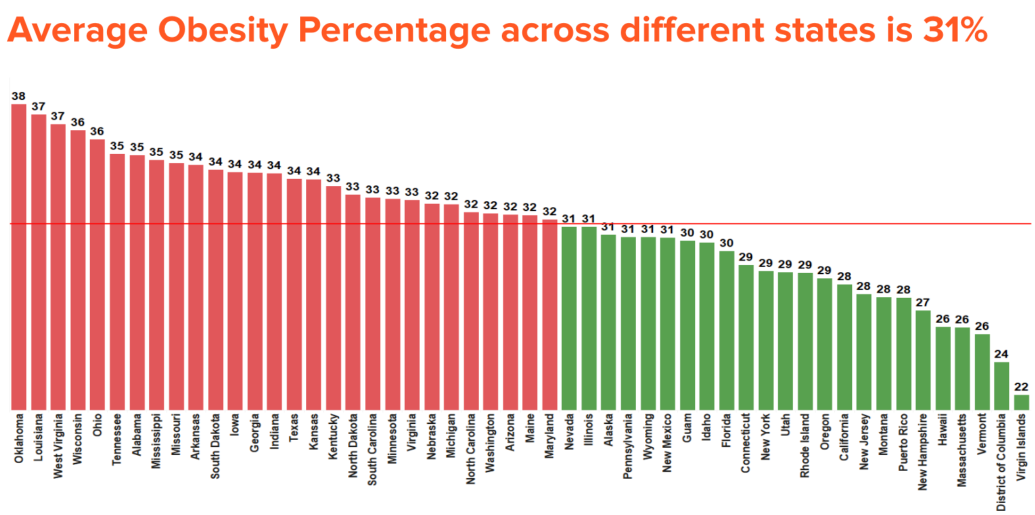

Fig 3. Average Obesity Percentage

The average threshold for categorizing states as obese was set at 31%, based on an analysis of obesity rates across the USA. This standardized measure provides a practical tool for identifying regions with significant obesity issues, guiding targeted interventions and resource allocation. Policymakers and healthcare professionals can utilize this threshold to prioritize efforts and monitor progress in addressing the obesity epidemic. The results obtained from the discrete choice model support the hypothesis that various demographic factors, including gender, race, education level, age group, and household income, significantly influence obesity prevalence. States with higher percentages of obese individuals across these demographic categories are identified as facing serious obesity problems, highlighting the importance of addressing disparities in obesity rates among different population groups. The model's findings provide empirical evidence supporting the hypothesis that demographic factors play a crucial role in shaping obesity prevalence, underscoring the need for targeted interventions and policy recommendations to address obesity effectively.

Based on the findings of the model, policy recommendations can be formulated to address obesity disparities and promote healthier lifestyles among various population groups. These recommendations may include targeted interventions such as community-based obesity prevention programs, initiatives to improve access to healthy food options and recreational facilities, and educational campaigns aimed at raising awareness about the importance of maintaining a healthy weight. Additionally, policies addressing socioeconomic factors such as income inequality and education attainment may also be recommended to address underlying determinants of obesity prevalence. By implementing evidence-based policies informed by the findings of the model, policymakers can work towards reducing obesity rates and improving overall population health outcomes.

Conclusion

In conclusion, our research offers comprehensive insights into the complex landscape of obesity prevalence across diverse demographic groups in the United States. Leveraging extensive data from the Behavioral Risk Factor Surveillance System (BRFSS) and the Centers for Disease Control and Prevention (CDC), we identified significant disparities in obesity rates, with socioeconomic and environmental factors playing pivotal roles. Our findings support the hypothesis that improved access to recreational facilities correlates with lower obesity rates, highlighting the importance of public infrastructure in promoting health. Additionally, socioeconomic factors like income and education level showed strong correlations with obesity, underscoring the need for policies addressing economic disparities. Dietary habits, particularly the consumption of high-calorie, processed foods, emerged as significant predictors of obesity, emphasizing the importance of nutrition in public health initiatives. The multifaceted nature of obesity, influenced by demographic factors, suggests the necessity for targeted public health strategies addressing supportive environments, socioeconomic conditions, and healthy dietary practices to combat the obesity epidemic effectively. Our recommendations span various sectors, including education, income support, pharmaceuticals, individual actions, and healthcare professionals, aiming to address obesity comprehensively and mitigate its adverse health impacts across diverse population groups. Further longitudinal research is recommended to delve deeper into causal mechanisms and intervention effects, enabling more tailored and effective obesity prevention and management strategies in the future.

References

[1] X. Pang, C. B. Forrest, F. Lê-Scherban and A. J. Masino, "Understanding Early Childhood Obesity via Interpretation of Machine Learning Model Predictions," 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA), Boca Raton, FL, USA, 2019, pp. 1438-1443.

[2] H. A. Obaid, S. Sheel, N. Ali Hussien, S. Rahim Alatba, M. Saleem and H. Warush, "Analyzing the physical activity and exercise management for obesity and weight loss based on fuzzy logic," 2023 International Conference on Emerging Research in Computational Science (ICERCS), Coimbatore, India, 2023, pp. 1-6.

[3] K. Alam, D. Chowdhury and P. Haque, "An Intelligent System for Obesity Detection through Human Activity Recognition and BMI Prediction," 2023 26th International Conference on Computer and Information Technology (ICCIT), Cox's Bazar, Bangladesh, 2023, pp. 1-6.

[4] Bray, G. A. (2004). Medical consequences of obesity. The Journal of clinical endocrinology & metabolism, 89(6), 2583-2589.

[5] N. A. Wahab, N. N. Norhisham, A. H. Mohammed, A. Osman, N. Ibrahim and S. S. M. Ilyas, "HealthyMe: A Persuasive Obesity Care Management Application," 2022 2nd International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS), Yogyakarta, Indonesia, 2022, pp. 213-218.