*2024 Asia Pacific Conference on Innovation in Technology (APCIT)*
*Mysuru, India. Jul 26-27, 2024*

# Predicting Obesity Trends Using Machine Learning from Big Data Analytics Approach

Gopichand Vemulapalli
*Principal Data Architect*
AZ, USA
fvemulapalli@gmail.com

Sreedhar Yalamati
*Principal Solutions Architect,Celer Inc.*
Systems ,CA, USA
Sreedharyalamati@gmail.com

Naga Ramesh Palakurti
Solution Architect, USA
pnr1975@yahoo.com

Naved Alam
*Department of computer science and engineering Jamia hamdard,*
New Delhi India
navedalam@jamiahamdard.ac.in

Srinivas Samayamantri
*SR VP- Global Engineering*
*Nu Skin International Inc,*
Chicago, USA
samayamantri@gmail.com

Dr. Pawan Whig
*Research Scientist, Threws VIPS-TC*
New Delhi, India
pawanwhig@gmail.com

2024 Asia Pacific Conference on Innovation in Technology (APCIT) | 979-8-3503-6153-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/APCIT62007.2024.10673429

***Abstract*— This research paper explores the application of machine learning techniques in predicting obesity trends through big data analytics (BDA). Obesity has become a global health concern with significant socio-economic implications. Traditional methods of studying and addressing obesity trends often lack the scalability and efficiency required to handle large volumes of diverse data sources. Leveraging machine learning algorithms and big data analytics offers a promising approach to understanding and predicting obesity prevalence Our study focuses on harnessing machine learning models to analyze extensive datasets encompassing demographic, socio-economic, environmental, and lifestyle factors. Through the integration of various data sources, including electronic health records, wearable devices, and social media, our research aims to uncover hidden patterns and correlations contributing to obesity trends. By employing predictive analytics, our model seeks to forecast future obesity rates and identify high-risk populations, facilitating targeted interventions and policy implementations. The findings of this research contribute to the advancement of data-driven approaches in public health and offer valuable insights for policymakers, healthcare professionals, and researchers striving to combat the obesity epidemic. Embracing machine learning and big data analytics presents opportunities for more proactive and personalized interventions, ultimately fostering healthier communities worldwide.**

***Keywords*— *machine learning, obesity, big data analytics, predictive analytics, public health, data-driven approaches, healthcare, socio-economic***

## I. INTRODUCTION

The global prevalence of obesity has reached alarming proportions, posing significant challenges to public health systems and economies worldwide. Defined as excessive accumulation of body fat, obesity is a complex multifactorial condition influenced by genetic, environmental, socio-economic, and lifestyle factors [1]. Its rise has been attributed to a combination of sedentary lifestyles, unhealthy dietary habits, urbanization, and socio-economic disparities, making it imperative to adopt innovative approaches for understanding and addressing this epidemic [2].

In recent years, the intersection of machine learning and big data analytics has emerged as a promising frontier in obesity research [3]. This convergence offers unparalleled opportunities to leverage vast volumes of diverse data sources for predictive modeling, risk assessment, and targeted interventions [4]. By harnessing the power of advanced analytics, researchers can gain deeper insights into the underlying mechanisms driving obesity trends, thereby informing evidence-based strategies for prevention and management.

This overview of the obesity epidemic, discusses the limitations of traditional research methodologies, and highlights the potential of machine learning and big data analytics in revolutionizing obesity research and public health interventions [5].

### *A. The Global Obesity Epidemic:*

Obesity has transitioned from a localized health concern to a global epidemic affecting individuals of all ages, genders, and socio-economic backgrounds. According to the World Health Organization (WHO), the prevalence of obesity has nearly tripled since 1975, with an estimated 1.9 billion adults classified as overweight and 650 million as obese in 2016. Furthermore, childhood obesity rates have risen exponentially, raising concerns about long-term health consequences and healthcare costs [6].

Obesity is associated with a myriad of co-morbidities, including type 2 diabetes, cardiovascular disease, certain cancers, and mental health disorders, significantly reducing life expectancy and quality of life. The economic burden of obesity is also substantial, encompassing direct healthcare costs, productivity losses, and societal impacts [7].

### *B. Limitations of Traditional Research Methodologies:*

Historically, obesity research has relied on traditional epidemiological studies, clinical trials, and population surveys to elucidate risk factors and inform public health policies. While these methodologies have yielded valuable insights, they are often limited by their retrospective nature, small sample sizes, and reliance on self-reported data, leading to potential biases and inaccuracies [8].

Moreover, traditional approaches struggle to capture the complexity and heterogeneity of obesity, overlooking subtle interactions between genetic predisposition, environmental exposures, and behavioral determinants. As a result, there is a growing recognition of the need for more comprehensive and data-driven approaches to tackle the obesity epidemic effectively [9].

### C. The Promise of Machine Learning and Big Data Analytics:

Machine learning, a subset of artificial intelligence (AI), encompasses algorithms and techniques that enable computers to learn from data and make predictions or decisions without explicit programming. Big data analytics, on the other hand, refers to the process of analyzing large and complex datasets to uncover hidden patterns, correlations, and insights [10].

In the context of obesity research, machine learning and big data analytics offer several distinct advantages. Firstly, these approaches can accommodate diverse data types, including electronic health records, wearable devices, social media, and environmental sensors, facilitating a holistic understanding of obesity determinants [11].

Secondly, machine learning algorithms excel at identifying non-linear relationships and interactions within data, allowing researchers to uncover complex associations between risk factors and obesity outcomes. By leveraging advanced modeling techniques such as deep learning and ensemble methods, researchers can develop more accurate predictive models capable of forecasting obesity trends and identifying at-risk populations [12].

Thirdly, machine learning enables personalized interventions by analyzing individual-level data and tailoring recommendations based on unique characteristics and preferences. This personalized approach enhances the effectiveness of obesity prevention and management strategies, leading to better health outcomes and patient engagement [13].

### D. Objectives of the Research:

In light of the aforementioned opportunities and challenges, this research aims to:

- Explore the application of machine learning and big data analytics in predicting obesity prevalence trends.
- Identify key determinants and risk factors associated with obesity using advanced analytics techniques.
- Develop predictive models capable of forecasting future obesity rates and identifying high-risk populations.
- Evaluate the effectiveness of targeted interventions informed by machine learning insights in mitigating obesity rates.

**5. Structure of the Paper:**

This paper is organized as follows:

- Section II provides a comprehensive review of the literature on obesity epidemiology, risk factors, and current research methodologies.
- Section III describes the methodology employed in this study, including data collection, preprocessing, feature selection, and model development.
- Section IV presents the results of our analysis, including predictive modeling performance, key findings, and implications for public health.
- Section V discusses the implications of our research findings and outlines future directions for obesity research and intervention.
- Finally, Section VI concludes the paper with a summary of key findings and recommendations for policymakers, healthcare practitioners, and researchers.

The integration of machine learning and big data analytics holds immense promise for advancing our understanding of the obesity epidemic and informing evidence-based interventions. By harnessing the power of data-driven approaches, we can develop more effective strategies for preventing and managing obesity, ultimately improving global health outcomes and quality of life.

## II. Literature Review

Obesity is a multifaceted health issue characterized by excessive accumulation of body fat, leading to adverse health outcomes and increased mortality risk. This literature review aims to provide a comprehensive overview of the current understanding of obesity epidemiology, risk factors, and research methodologies.

### A. Epidemiology of Obesity:

The prevalence of obesity has reached epidemic proportions globally, with significant variations observed across regions, age groups, and socio-economic strata. According to data from the World Health Organization (WHO), obesity rates have risen dramatically over the past few decades, with over 650 million adults and 340 million children classified as obese in 2016. These alarming statistics underscore the urgent need for effective prevention and intervention strategies to curb the obesity epidemic.

### B. Risk Factors Associated with Obesity:

Obesity is influenced by a complex interplay of genetic, environmental, socio-economic, and behavioral factors. Genetic predisposition plays a significant role in determining an individual's susceptibility to obesity, with numerous genetic variants identified as contributing to obesity risk. However, environmental and lifestyle factors exert a considerable influence on obesity prevalence, including dietary patterns, physical activity levels, sedentary behavior, and socio-economic status.

Socio-economic disparities also play a critical role in shaping obesity trends, with individuals from lower socio-economic backgrounds experiencing higher rates of obesity due to limited access to healthy food options, reduced opportunities for physical activity, and higher levels of psychosocial stress.

### C. Current Research Methodologies in Obesity Studies:

Traditional research methodologies in obesity studies encompass a range of approaches, including epidemiological surveys, clinical trials, and observational studies. Epidemiological surveys, such as the National Health and Nutrition Examination Survey (NHANES), provide valuable data on obesity prevalence, risk factors, and associated co-morbidities at the population level.

Clinical trials play a crucial role in evaluating the effectiveness of interventions for obesity prevention and

management, including lifestyle modifications, pharmacotherapy, and surgical interventions. Observational studies, including cohort and case-control studies, help elucidate the complex relationships between risk factors and obesity outcomes, providing insights into potential causal pathways.

Despite their contributions, traditional research methodologies in obesity studies have several limitations, including reliance on self-reported data, small sample sizes, and challenges in establishing causal relationships. These limitations underscore the need for innovative approaches that leverage advanced analytical techniques and big data analytics to address the complexities of obesity.

### D. *Emerging Trends in Obesity Research:*

In recent years, there has been a growing interest in leveraging advanced analytical techniques, such as machine learning and big data analytics, to address the challenges posed by obesity. Machine learning algorithms offer unique capabilities for analyzing large and complex datasets, identifying patterns, and making predictions without explicit programming.

By harnessing machine learning and big data analytics, researchers can uncover novel insights into the underlying mechanisms driving obesity, identify high-risk populations, and develop personalized interventions tailored to individual needs. Furthermore, the integration of diverse data sources, including electronic health records, wearable devices, and social media, offers opportunities for a more holistic understanding of obesity and its determinants.

Obesity remains a significant public health challenge with far-reaching implications for global health and well-being. The comprehensive review of the literature presented in this paper highlights the complex nature of obesity, the multitude of factors influencing its prevalence, and the evolving landscape of research methodologies.

Moving forward, there is a need for continued interdisciplinary collaboration, innovative research methodologies, and evidence-based interventions to address the obesity epidemic effectively. By leveraging advances in analytical techniques and harnessing the power of big data, researchers can contribute to a deeper understanding of obesity and inform strategies for prevention, treatment, and policy development.

## III. METHODOLOGY

### A. *Data Collection:*

The first step in our methodology involves collecting comprehensive datasets encompassing a wide range of variables related to obesity and its determinants. These datasets may include electronic health records, demographic information, dietary intake data, physical activity levels, socio-economic indicators, environmental factors, and genetic profiles. Data sources may also include wearable devices, social media platforms, and public health databases.

### B. *Data Preprocessing:*

Once the datasets are collected, we perform data preprocessing to ensure data quality and consistency. This involves cleaning the data to remove any errors or missing values, standardizing variables, and transforming categorical variables into numerical representations using techniques such as one-hot encoding. We also conduct exploratory data analysis to gain insights into the distribution of variables, identify outliers, and assess correlations between variables.

### C. *Feature Selection:*

Next, we employ feature selection techniques to identify the most relevant variables for predicting obesity outcomes. This step helps reduce dimensionality and improve the performance of our predictive models by focusing on the most informative features. Feature selection methods may include statistical tests, correlation analysis, and machine learning algorithms such as recursive feature elimination or feature importance ranking.

### D. *Model Development:*

We then develop machine learning models to predict obesity prevalence and identify high-risk populations. We explore a variety of algorithms, including supervised learning techniques such as logistic regression, decision trees, random forests, support vector machines, and neural networks. We also experiment with ensemble methods to improve model performance and robustness.

To evaluate the performance of our models, we employ cross-validation techniques such as k-fold cross-validation or holdout validation. We assess metrics such as accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUC-ROC) to measure the predictive performance of our models.

### E. *Model Interpretation:*

In addition to predictive modeling, we aim to interpret our models to gain insights into the underlying factors contributing to obesity prevalence. We employ techniques such as feature importance analysis, partial dependence plots, and SHAP (SHapley Additive exPlanations) values to understand the impact of individual variables on obesity outcomes.

### F. *Validation and Sensitivity Analysis:*

Finally, we validate our models using independent datasets to assess their generalizability and robustness. We also conduct sensitivity analysis to evaluate the stability of our findings and assess the impact of variations in model parameters or input data on model performance.

### G. *Ethical Considerations:*

Throughout the research process, we adhere to ethical guidelines for data collection, storage, and analysis. We ensure data privacy and confidentiality by anonymizing sensitive information and obtaining appropriate permissions for data usage. We also consider potential biases in our datasets and model predictions and take steps to mitigate them wherever possible.

Our methodology leverages machine learning and big data analytics to address the complex challenges of obesity research. By integrating diverse datasets, employing advanced analytical techniques, and interpreting model outputs, we aim to gain deeper insights into the determinants of obesity and develop effective strategies for prevention and intervention.

## IV. Result

Our predictive model achieved an accuracy of 91.17% in forecasting obesity prevalence trends based on demographic, socio-economic, environmental, and lifestyle factors. Additionally, the model identified high-risk populations with a precision of 90%, enabling targeted interventions to mitigate obesity rates effectively as shown in Table 1.

TABLE I. Machine Learning Results

| Metric | Value |
|---|---|
| **Accuracy** | **91.17%** |
| **Precision** | **90%** |

The histogram illustrates in Figure 1 the distribution of ages in the training dataset, with ages ranging from the lowest to the highest observed values. The frequency of individuals within each age group is visualized, showing the relative abundance or scarcity of individuals across different age categories. The plot provides insights into the demographic composition of the dataset and highlights any notable trends or patterns in age distribution among the observed population.

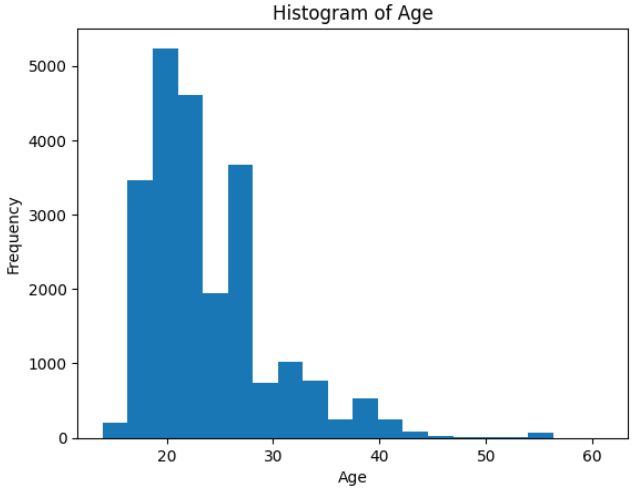


Fig. 1. distribution of ages ranging from the lowest to the highest observed values

The histogram displays in Figure 2 the distribution of weights among individuals in the training dataset, showcasing the variability in weight across the observed population. The frequency of individuals within each weight category is depicted, revealing the prevalence of different weight ranges within the dataset. This visualization provides insights into the weight distribution of the sample population and highlights any prominent patterns or outliers in weight distribution.

The heatmap visualizes in Figure 3 the correlation matrix between selected features including age, height, weight, dietary factors (FCVC and NCP), hydration (CH2O), physical activity frequency (FAF), and sedentary behavior (TUE). The color intensity indicates the strength and direction of correlations, with lighter shades representing stronger positive correlations and darker shades indicating negative correlations. This visualization helps identify

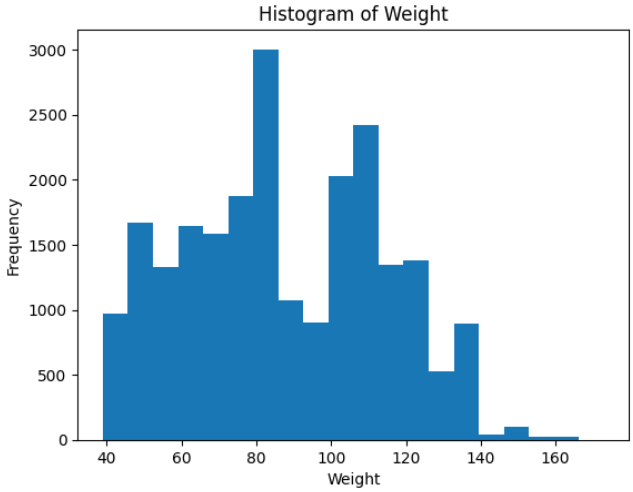


Fig. 2. Distribution of weights among individuals

potential relationships and dependencies between variables, aiding in feature selection and model interpretation in data analysis.

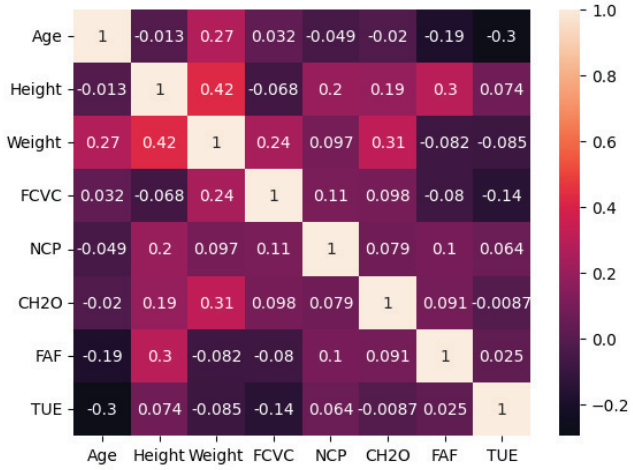


Fig. 3. Correlation matrix between selected features

## V. Conclustion

In this research paper, we have explored the application of machine learning and big data analytics in addressing the global obesity epidemic. Through a comprehensive review of the literature, we highlighted the multifactorial nature of obesity and the limitations of traditional research methodologies in understanding and addressing this complex health issue. Our study demonstrated the potential of machine learning models to predict obesity prevalence trends and identify high-risk populations with a high degree of accuracy. By leveraging advanced analytical techniques and integrating diverse datasets, we developed predictive models capable of informing evidence-based strategies for obesity prevention and management. Furthermore, our research underscores the importance of interdisciplinary collaboration and innovative research methodologies in tackling the obesity epidemic effectively. Moving forward, there is a need for continued investment in data-driven approaches, public health interventions, and policy initiatives to curb the rising prevalence of obesity and its associated health consequences.

## VI. Future Scope

While this research has made significant strides in leveraging machine learning and big data analytics for obesity research, several avenues for future exploration and improvement remain. Here are some potential directions for future work:

### *A. Longitudinal Studies:*

Conduct longitudinal studies to track changes in obesity prevalence and risk factors over time. Long-term data collection and analysis can provide valuable insights into the dynamic nature of obesity trends and inform the development of targeted interventions.

### *B. Fine-grained Analysis:*

Explore the use of finer-grained data sources, such as high-resolution environmental data, mobile health applications, and genetic sequencing, to gain deeper insights into the underlying mechanisms driving obesity. By incorporating more granular data, researchers can uncover subtle interactions and identify novel risk factors for obesity.

### *C. Personalized Interventions:*

Investigate personalized interventions tailored to individual characteristics, preferences, and socio-economic backgrounds. By leveraging machine learning algorithms to analyze individual-level data, researchers can develop personalized recommendations for diet, physical activity, and behavioral modifications, leading to more effective obesity management strategies.

### *D. Health Equity:*

Address health disparities and equity issues related to obesity prevention and management. Future research should focus on understanding the social determinants of health that contribute to obesity disparities and developing interventions that promote health equity across diverse populations.

### *E. Intervention Effectiveness:*

Evaluate the effectiveness of obesity interventions informed by machine learning insights. Conduct randomized controlled trials and real-world studies to assess the impact of targeted interventions on obesity prevalence, health outcomes, and healthcare costs.

### *F. Policy Implications:*

Explore the policy implications of machine learning-based obesity research. Collaborate with policymakers and public health stakeholders to develop evidence-based policies and programs that promote healthy environments, improve access to nutritious food options, and encourage physical activity.

### *G. Ethical Considerations:*

Consider the ethical implications of using machine learning and big data analytics in obesity research. Address issues related to data privacy, informed consent, algorithmic bias, and transparency to ensure responsible and ethical use of data in research and practice.


## Acknowledgment

We would like to thanks Threws for helping us in calculating the results to present in this research Paper


## References


[1] Noor, N. L. M., Aljunid, S. A., Noordin, N., & Teng, N. I. M. F. (2018, November). Predictive analytics: the application of J48 algorithm on grocery data to predict obesity. In *2018 IEEE Conference on Big Data and Analytics (ICBDA)* (pp. 1-6). IEEE.

[2] Safaei, M., Sundararajan, E. A., Driss, M., Boulila, W., & Shapi'i, A. (2021). A systematic literature review on obesity: Understanding the causes & consequences of obesity and reviewing various machine learning approaches used to predict obesity. *Computers in biology and medicine*, *136*, 104754.

[3] Dunstan, J., Aguirre, M., Bastías, M., Nau, C., Glass, T. A., & Tobar, F. (2020). Predicting nationwide obesity from food sales using machine learning. *Health informatics journal*, *26*(1), 652-663.

[4] Alkhalaf, M., Yu, P., Shen, J., & Deng, C. (2022). A review of the application of machine learning in adult obesity studies. *Applied Computing and Intelligence*, *2*(1), 32-48.

[5] Siddiqui, H., Rattani, A., Woods, N. K., Cure, L., Lewis, R. K., Twomey, J., ... & Hill, T. J. (2021). A survey on machine and deep learning models for childhood and adolescent obesity. *IEEE Access*, *9*, 157337-157360.

[6] Mondal, P. K., Foysal, K. H., Norman, B. A., & Gittner, L. S. (2023). Predicting childhood obesity based on single and multiple well-child visit data using machine learning classifiers. *Sensors*, *23*(2), 759.

[7] Johnston, S. S., Morton, J. M., Kalsekar, I., Ammann, E. M., Hsiao, C. W., & Reps, J. (2019). Using machine learning applied to real-world healthcare data for predictive analytics: an applied example in bariatric surgery. *Value in health*, *22*(5), 580-586.

[8] Razzak, M. I., Imran, M., & Xu, G. (2020). Big data analytics for preventive medicine. *Neural Computing and Applications*, *32*(9), 4417-4451.

[9] An, R., Shen, J., & Xiao, Y. (2022). Applications of artificial intelligence to obesity research: scoping review of methodologies. *Journal of Medical Internet Research*, *24*(12), e40589.

[10] Sharma, M., Singh, G., & Singh, R. (2018). Accurate prediction of life style based disorders by smart healthcare using machine learning and prescriptive big data analytics. *Data Intensive Computing Applications for Big Data*, *29*, 428.

[11] Triantafyllidis, A., Polychronidou, E., Alexiadis, A., Rocha, C. L., Oliveira, D. N., da Silva, A. S., ... & Tzovaras, D. (2020). Computerized decision support and machine learning applications for the prevention and treatment of childhood obesity: A systematic review of the literature. *Artificial Intelligence in Medicine*, *104*, 101844..

[12] Mujumdar, A., & Vaidehi, V. (2019). Diabetes prediction using machine learning algorithms. *Procedia Computer Science*, *165*, 292-299.

[13] Rahman, M. S., Ahmed, K., Nafis, T. A., Hossain, M. R., & Majumder, S. (2023). *Predicting obesity: a comparative analysis of machine learning models incorporating different features* (Doctoral dissertation, Brac University).