

editor@ijprems.com

MACHINE LEARNING APPLICATIONS IN OBESITY DETECTION: A COMPREHENSIVE REVIEW

Ameena Shaik¹, Maragoni Swapna², Shilpa Rathod³, Remalli Rohan⁴

¹M.Sc.-Data Science - PG Scholar, Department of Computer Science, Telangana Social Welfare Residential Degree College for Women, Jagathgirigutta, Hyderabad, India.

^{2,3}Degree lecturer, Department of Computer Science, Telangana Social Welfare Residential Degree College for Women, Jagathgirigutta, Hyderabad, India.

⁴Researcher, Computer Science Educator, Hyderabad, Telangana, India.

DOI: https://www.doi.org/10.56726/IRJMETS37794

ABSTRACT

A complex health condition, obesity is influenced by genetic, environmental, and lifestyle variables. It is linked to a higher risk of developing diseases like cancer, diabetes, cardiovascular disease, metabolic syndrome, and hypertension. Obesity's negative health effects stem from the accumulation of metabolic and physical stress caused by extra body fat and dysregulated adipokine release, which causes inflammation and a number of chronic illnesses. By utilizing extensive datasets to pinpoint important risk factors and boost classification accuracy, recent developments in machine learning present intriguing approaches for improving obesity prediction, detection, and prevention. High accuracy in forecasting the risk of obesity is demonstrated by current models, which include algorithms like Random Forest, Cat Boost, and Support Vector Machines. This suggests the possibility of individualized health interventions. Additionally, research on factors specific to age and gender highlights the significance of adjusting tactics to individual features in order to improve the efficacy of obesity therapies. The most recent results from machine learning applications in obesity research are summarized in this study, which offers a thorough understanding of the contributing variables, assessment techniques, and possible therapies. The necessity for ongoing advancements in customized therapies and predictive modelling highlights how crucial machine learning is to the advancement of public health campaigns aimed at managing and reducing obesity

Keywords: Obesity, Prediction, Detection, Prevention, Health, Machine Learning, Data Analysis, Risk Factors, Digital Health

1. INTRODUCTION

The world is facing with many lifestyle disorders. These are being raised because of unbalanced highly refined food, sedentary lifestyle and stressful mental conditions. Over nutrition is a form of malnutrition in which the intake of nutrients exceeds the amount required for normal growth, development and metabolism. It can be of two types obesity and oversupplying a specific nutrient. Obesity is one of the commonest lifestyle disorders. Obesity is a medical condition in which excess body fat accumulates to the extent that it may have a negative effect on health, leading to reduced life expectancy and /or increased health problems. Latin word "OBESUS" meaning fat (or) which means to eat excessively

BMI >30.



Figure 1: Obesity at variance ages

1.1 Definition and Types of Obesity

Obesity, as defined by the World Health Organization (WHO), is an abnormal or excessive accumulation of body fat that poses a significant health risk. This condition is categorized using the Body Mass Index (BMI), where a BMI of 30 or above is considered obese. There are three main classes of obesity: Class I (BMI 30-34.9), Class II (BMI 35-39.9), and Class III (BMI \geq 40), which is often referred to as "severe" or "morbid" obesity as showed in figure 1. Other classification approaches consider fat distribution, such as central (abdominal) and peripheral obesity, which can impact health risks differently and inform specific interventions as per World Health Organization (WHO) [2].



1.2 Causes of Obesity

Obesity is a complex, multifactorial condition resulting from a combination of genetic, environmental, and behavioural factors. Key contributors include as shown in figure 2:

- > **Dietary habits:** High consumption of calorie-dense foods, processed snacks, and sugary beverages.
- > Physical inactivity: Sedentary lifestyles lead to energy imbalances, promoting weight gain.
- > Genetics: Genetic factors can predispose individuals to obesity.
- Socioeconomic and psychological factors: Lower socioeconomic status, limited access to healthy food options, and psychological issues such as stress and depression can elevate obesity risk.
- Emerging research in machine learning (ML) and the Internet of Things (IoT) also highlights potential for using predictive algorithms to identify lifestyle and physiological patterns associated with obesity, which can aid in early intervention and prevention [11].



Figure 2: Causes of Obesity

1.3 Global Prevalence of Obesity

In 2019, WHO reported that over 650 million adults globally were classified as obese, representing approximately 13% of the global population. The prevalence of obesity continues to increase, particularly in urban areas and among young people, making it a critical global health challenge. Childhood obesity is also rising, prompting a focus on early intervention and prevention. ML and IoT technologies play a growing role in identifying high-risk groups and customizing prevention strategies, particularly in children and adolescents [6] [5]. Technological advancements have enabled researchers to investigate obesity from various perspectives. For example, machine learning models are increasingly used to study socioeconomic, genetic, and environmental factors affecting obesity, resulting in personalized health recommendations. Advanced tools, such as agent-based modelling, neural networks, and synthetic data balancing techniques like SMOTE, have improved the accuracy of obesity risk predictions across diverse populations [13] [10] [24]. In conclusion, substantial progress has been made in understanding and addressing obesity through ML and IoT; however, challenges remain, including data quality, infrastructure needs, and demographic-specific requirements. Future research aims to refine these tools for better early detection, risk assessment, and targeted prevention strategies, supporting global health efforts to mitigate obesity's impact.

2. LITERATURE SURVEY

The literature review looks at current research on the use of machine learning for predicting and preventing obesity. It draws attention to important approaches that have increased the precision of detecting obesity risk factors, such as feature selection strategies and predictive models. In order to comprehend the complex nature of obesity and its preventative measures, the study also examines studies from both international and demographic sources as given in the table 1.

Scheinker et al. [1] investigated U.S. county-level determinants of obesity using both conventional epidemiology and ML techniques, revealing that factors like education, income, and area were significant contributors to obesity. ML models, such as gradient boosting, showed slightly better accuracy than traditional methods but lacked interpretability Meanwhile, Machorro-Cano et al. [2] proposed PISIoT, a smart health platform that uses IoT and machine learning to track health and give recommendations, particularly for obese older adults. In India, Pereira et al. [3] used machine learning to predict obesity based on various lifestyle factors like stress and sleep patterns, urging that better data quality could improve prediction accuracy. In a similar vein, Herts et al. [4] showed that obesity can affect liver lesion detection in CT scans, while Taghiyev et al. [5] developed a two-stage hybrid model to predict obesity causes in Turkey, achieving a high accuracy rate of 91.4%.

@International Journal Of Progressive Research In Engineering Management And Science



INTERNATIONAL JOURNAL OF PROGRESSIVE
RESEARCH IN ENGINEERING MANAGEMENT
AND SCIENCE (IJPREMS)e-ISSN :
2583-1062Impact
(Int Peer Reviewed Journal)Impact
Factor :
7.001

Colmenarejo [6] reviewed how machine learning is used to predict obesity in children and adolescents, noting its predictive power with large health datasets. Chatterjee et al. [7] discussed how machine learning models, using data from sources like UCI and Kaggle, could identify risk factors tied to urbanization and lifestyle changes. Runge et al. [8] explored the use of the Controlled Attenuation Parameter (CAP) for detecting liver fat in obese children, showing CAP's usefulness as a screening tool. Abdulrahman and Alnagar [9] used logistic regression and decision trees to assess obesity awareness in Saudi Arabia, highlighting the role of education in raising awareness of obesity risks. Ren et al. [10] proposed a high-order simulated annealing neural network to track obesity, achieving an impressive 98.7% accuracy. StephenClark et al. [11], who identified socio-environmental factors in the UK Bio Bank, emphasize how data balance and socio-environmental factors play crucial roles in predicting obesity.

Further research includes Erika Cheng's [12] use of LSTM to predict childhood obesity, Yu-Chi Lee et al.'s [13] study on genetics and nutrition interactions in obesity prediction, and Sonoda et al.'s [14] use of logistic regression for predicting obesity in Japanese children. Palmas et al. [15] demonstrated the potential of CT scans and AI in assessing body composition, while Jeon et al. [16] highlighted key obesity indicators in a study of South Korean medical data. Torres-Martos et al. [17] developed an ML model combining genetic and lifestyle data for childhood obesity prediction, demonstrating enhanced accuracy and interpretability. Other studies, such as those by Harjatmo et al. [18], Zhou et al. [19], and Lim et al. [20], examined how lifestyle, urban design, and dietary habits influence obesity prevalence in urban populations. Similarly, Jeon et al. [21] and Banal et al. [22] looked at the impacts of physical activity and sedentary behaviour in obesity risk. Kiss et al. [23] used wearable devices to track health data and identify risk factors for youth obesity, while Rajbhoj et al. [24] introduced "Obesity Guard," an ML system for early detection and prevention. Maulana et al. [25] used the Cat Boost model to predict obesity in Latin America, while Singh and Tawfik [26] emphasized the importance of feature selection in machine learning for accurate obesity prediction.

The project is on employing machine learning algorithms for early obesity detection and prevention. To predict obesity, four supervised classifiers were tested: Support Vector Machine, Decision Tree, Random Forest, and Logistic Regression. Random Forest outperformed the rest of them with 100% accuracy, highlighting its potential for proactive healthcare solutions [27]. The study looked at machine learning methods for predicting overweight or obesity based on dietary patterns and physical condition data. After testing a number of classification methods, Random Forest, SVM, and Gradient Boosting were shown to perform the best (78% accuracy, 79% precision, 78% recall, and 78% F1-score). This demonstrates how machine learning models can be used to support healthcare decision-making [28]. The study examined cutting-edge machine learning methods to improve the prediction of obesity risk, with an emphasis on a novel ANN-PSO hybrid model. With an astounding accuracy of 92%, this model outperformed conventional regression techniques. Critical obesity risk factors were identified by feature importance analysis using SHAP, highlighting the potential of such techniques for individualised treatment [29].

Author	Dataset	Feature Extraction	Algorithms	Result
Taghiyev, A. et al. [5]	EHR from AksaraySultanhani Family Health Center	Best variable selection	Hybrid classification model	91.4% accuracy
Runge, J. H. et al. [8]	Data from 60 children with severe obesity	Imaging scores and evaluations	Controlled attenuation parameter (CAP)	Sensitivity of 75% for detecting steatosis
Abdulrahma n&Alnagar [9]	1369 respondents from KSA	Demographics (age, gender, education)	Logistic Regression, Decision Tree	High knowledge of risk factors (95.5%).
Ren et al. [10]	Medical big data	Potential target indicators selected	High-order Simulated Annealing Neural Network	Accuracy of obesity monitoring model reached 98.7%.
Cheng et al.	Longitudinal dataset	Medical history	Long short-term	MAE = 0.98

Table 1. Summary of Methodologies



INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

(Int Peer Reviewed Journal)

e-ISSN : 2583-1062

Impact Factor :

www.ijprems.com editor@ijprems.com

Vol. 04, Issue 12, Decembaer 2024, pp : 1234-1239

Factor 7.001

[12]	(children 0-4 years)	from clinical encounters	memory (LSTM)	
Lee et al. [13]	Framingham Offspring Study	Genetic SNPs, DNA methylation sites, dietary factors	Stochastic gradient boosting	70% accuracy
RisaSonoda et al. [14]	1,504 Japanese children	7 binary variables	Multivariable logistic regression	AUC = 0.803
Fiorella Palmas et al. [15]	70 patients (46 with obesity)	Tissue area and Hounsfield units from CT images	FocusedON-BC software	r = 0.926 (compared with DXA)
SeungjinJeo n et al. [16]	21,100 participants	Blood test and blood pressure data	Six machine learning classifiers	Over 70% accuracy for ages 19-39
JunhwiJeon et al. [21]	21,100 participants from KHNANES survey	Triglycerides, ALT, glycatedhemoglob in, uric acid	Various machine learning classifiers	Over 70% accuracy in 19-39 age group
Orsolya Kiss et al. [23]	2971 participants, Fitbit data from the ABCD Study	Sleep measures, physical activity, sociodemographic data	Glass box machine learning models	AUC Mean = 0.726
Aga Maulana et al. [25]	2111 individuals from Mexico, Peru, Colombia	Gender, age, height, weight, eating habits, physical activity	CatBoost, logistic regression, KNN, random forest, naive Bayes	Highest accuracy at 95.98%
Singh,B. H.[26]	Demographic, behavioral, and physiological datasets	Feature selection techniques to enhance model accuracy	Random Forest, Support Vector Machine	High accuracy achieved
Dr. S. M. Rajbhoj et al. [27]	Kaggle First dataset for text input is with 500 text snippets with information about obesity levels. The second dataset contains 2878	Age, Gender, Height, Weight,BMI Activity Level, Eating Habits	Support Vector Machine, Decision Tree, Random Forest, Logistic Regression	Random Forest achieved 100% accuracy, outperforming other models.
Elias Rodríguez et al. [28]	Data related to physical condition and eating habits	gender, age, height, weight and the existence of	Decision Tree, Support Vector Machines, K- Nearest Neighbors, Gaussian Naive Bayes,	Random Forest achieved the best performance: 78% accuracy, 79% precision, 78% recall,



www.ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

(Int Peer Reviewed Journal)

e-ISSN : 2583-1062 Impact Factor :

7.001

editor@ijprems.com

Vol. 04, Issue 12, Decembaer 2024, pp : 1234-1239

		overweight relatives (c_FMOW).	Multilayer Perceptron, Random Forest, Gradient Boosting, Extreme Gradient Boosting	and 78% F1-score. Demonstrated potential for healthcare.
Zarindokht Helforoush et al. [29]	Colombia, Peru, Mexico 14 to 61 years Online survey (anonymous participants) 17 features (mix of numeric, binary, and categorical variables)	SHAP (SHapley Additive exPlanations) was used for feature importance analysis	Artificial Neural Network (ANN), ANN- PSO (ANN with Particle Swarm Optimization hybrid model)	ANN-PSO achieved 92% accuracy, outperforming traditional regression models. Provided detailed insights into obesity risk factors.

3. CONCLUSION AND FUTURE SCOPE

Significant progress has been made in addressing obesity and associated medical issues including diabetes and cardiovascular disease between 2019 and 2024 with the use of machine learning (ML), the Internet of Things (IoT), and data analytics. Research shows that ML algorithms, including Naïve Bayes, J48, Random Forest, and decision trees, may use health data, including age, Cholesterol, physical activity, and psychological characteristics, to predict diseases connected to obesity. Real-time health monitoring is now possible with the combination of ML and IoT, allowing for more proactive treatment and quicker interventions. But there are still issues, like the requirement for more varied data and scalable models. Larger, more comprehensive datasets that take geographical and socioeconomic variances into account will improve model accuracy in future research. IoT-enabled real-time tracking of vital health metrics (such as blood sugar and physical activity) can yield ongoing, individualized insights that help doctors customize care. Predictive accuracy could be further increased by developments in deep learning and hybrid ML models. Adding behavioral data, such as sleep, stress, and lifestyle decisions, may also allow for a more comprehensive approach. As obesity rates rise, ML and IoT present intriguing tools for policy modelling and prevention-focused healthcare, opening the door to more effective global health initiatives.

4. REFERENCES

- [1] Scheinker, D.et al.(2019). County-level factors associated with obesity in the United States: Application of epidemiology and machine learning methods. Journal of Public Health, 41(4), 778–785.
- [2] Machorro-Cano, et al. (2019). PISIoT: A machine learning and IoT-based smart health platform for overweight and obesity control. Sensors, 19(16), 3469.
- [3] Pereira, N. et al. (2019). Obesity-related disease prediction from healthcare communities using machine learning. 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, pp. 1-7.
- [4] Herts, B. et al. (2020). Impact of obesity on the detection of low-contrast liver lesions using a phantom model. European Radiology, 30(4), 2085–2092.
- [5] Taghiyev, et al. (2020). A hybrid approach based on machine learning to identify the causes of obesity. CEAI, 22(2), 56-66.
- [6] Colmenarejo, G. (2020). Machine learning models to predict childhood and adolescent obesity: A review. Biostatistics and Bioinformatics Unit, IMDEA Food, CEI UAM+CSIC, Madrid, Spain.
- [7] Chatterjee, A.et al. (2020). Identification of risk factors associated with obesity and overweight—a machine learning overview. University of Agder, Norway.
- [8] Runge, J. H., et al. (2020). Accuracy of controlled attenuation parameter compared with ultrasound for detecting hepatic steatosis in children with severe obesity. European Radiology, 31,1588-1596
- [9] Abdulrahman, A. et al. (2021). Data Analysis and Computational Methods for Assessing Knowledge of Obesity Risk Factors among Saudi Citizens. Mathematical Methods in Medicine, 2021, Article ID 1371336.
- [10] Ren, L et al. (2021). Obesity Mass Monitoring in Medical Big Data Based on High-Order Simulated Annealing Neural Network Algorithm. Scientific Programming, 2021, Article ID 7052729.



www.ijprems.com

editor@ijprems.com

- [11] Clark, S., et al. (2022). A foresight whole systems obesity classification for the English UK Biobank cohort. BMC Public Health, 22(1), 349.
- [12] Cheng, E. R., et al. (2022). Predicting childhood obesity using machine learning: Practical considerations. Biomedinformatics, 2(1), 184–203.
- [13] Lee, Y. C., et al. (2022). Using machine learning to predict obesity based on genome-wide and epigenomewide gene–gene and gene–diet interactions. Frontiers in Genetics, 12, Article 783845.
- [14] Sonoda, R., et al. (2023). A point system to predict the future risk of obesity in 10-year-old children. Environmental Health and Preventive Medicine, 28(25).
- [15] Palmas, F., et al. (2023). Comparison of computed tomography and dual-energy X-ray absorptiometry in the evaluation of body composition in patients with obesity. Frontiers in Endocrinology, 14, Article 1161116.
- [16] Jeon, J., et al. (2023). Age-specific risk factors for the prediction of obesity using a machine learning approach. Frontiers in Public Health, 10, Article 998782.
- [17] Torres-Martos, Á., et al. (2023). A machine learning framework for obesity prediction in children based on genetic and lifestyle factors. Journal of Biomedical Informatics, 139, Article 104331.
- [18] Harjatmo, T.et al. (2023). Socioeconomic and lifestyle influences on obesity prevalence in urban communities. Journal of Urban Health, 30(2), 215-225
- [19] Zhou, Y., et al. (2023). Urban environment and obesity: The role of spatial planning in mitigating health risks. International Journal of Environmental Health Research, 33(1), 45-60.
- [20] Lim, H., et al. (2023). Influence of dietary patterns on obesity across age groups in South Korea. Journal of Nutrition and Health, 52(2), 123-137.
- [21] Jeon, S., et al. (2023). Predicting obesity risk based on physical activity and sedentary behavior using machine learning in South Korean adults. Frontiers in Public Health, 11, 223198.
- [22] Banal, et al. (2023). Dietary and lifestyle predictors of obesity in Filipino adolescents: A machine learning approach. Frontiers in Public Health, 11, 223267.
- [23] Kiss, O., et al. (2024). Using explainable machine learning and Fitbit data to investigate predictors of adolescent obesity. International Journal of Adolescent Health.
- [24] Rajbhoj, S. M., et al. (2024). Obesity Guard: Machine Learning for Early Detection & Prevention. International Research Journal on Advanced Engineering Hub, 2(07), 2041-2051.
- [25] Maulana, A., et al. (2024). Predicting Obesity Levels with High Accuracy: Insights from a CatBoost Machine Learning Model. Journal of Public Health Informatics.
- [26] Singh, B., & Tawfik, H. (2024). Predicting obesity using machine learning: A systematic approach. International Research Journal on Advanced Engineering Hub, 2(7), 2041-2051.
- [27] Rajbhoj, S. M., Shivale, S., Mulik, V. P., Shirke, S., & Yadav, A. P. (2024). Obesity Guard: Machine Learning for Early Detection & Prevention. Department of E&TC, Bharati Vidyapeeth's College of Engineering for Women, Pune, India.
- [28] Rodríguez, E., Rodríguez, E., Nascimento, L., Silva, A., & Marins, F. (2024). Machine Learning Techniques to Predict Overweight or Obesity. São Paulo State University (UNESP), Brazil.
- [29] Helforoush, Z., & Sayyad, H. (2024). Prediction and Classification of Obesity Risk Based on a Hybrid Metaheuristic Machine Learning Approach. Department of Mathematics and Systems Engineering, Florida Institute of Technology, Melbourne, FL, United States.