Anadolu Psikiyatri Dergisi || ISSN: 1302-6631 || Volume 25; Number 5
https://doi.org/10.5281/zenodo.13912888

Enhancing Obesity Prediction Accuracy Using Support Vector Machine

Dr.T.Jaya[#1], Aarthy Amsalekha Sivakami S [*2]

*[1] Professor, Department of Electronics and Communication Engineering, Saveetha Engineering College, Anna University, Chennai.

[*2]Assistant Professor Lourdes Mount College of Engineering, Anna University, Chennai.

Abstract --- **Obesity has become a critical global health concern, presenting substantial risks for the future. The most prevalent health challenges worldwide, it requires urgent attention. Predicting diseases early can help doctors and patients take proactive measures to reduce risks, stop the progression of the condition, and tackle its root causes, thereby decreasing the likelihood of complications and fatalities. Although reviewing medical records can identify diseases, this method is often time-consuming, prone to errors, and expensive when done manually. Consequently, the development of predictive models using automated scientific techniques has become increasingly important. The Support Vector Machine (SVM) model has shown significant promise, with a classifier achieving a remarkable accuracy of 99.05%, outperforming other models. Global obesity statistics highlight the severity of the issue. These numbers have nearly tripled over the past three decades.**

Keyword –- Obesity, Predicting, Support Vector Machine (SVM), Overweight, Fatalities.

I. INTRODUCTION

Obesity is a prevalent global health issue linked to numerous diseases, risks, and even death. It poses a significant threat to the future of global health. [1] According to studies, obesity has been a growing and dynamic global concern since the 1980s, affecting adults, youths, and children alike. This alarming rise has prompted new research focusing on detecting obesity levels in both young and old populations.

Obesity is typically characterized as the accumulation too much body fat can negatively impact a person's health. [2] Many people over the age of 16 struggle with weight issues, often due to high consumption of starchy, fatty foods and a lack of physical activity. [3] Obesity is a health issue that impacts all ages of people, from children, teenagers to adults.

Research indicates that obesity is a complex condition resulting from the excessive consumption of high-calorie and fatty foods, which leads to unregulated weight gain. [4] Biological factors, such as genetics, contribute to obesity and can be categorized into types like polygenic, monogenic, syndromic, and leptin-related. Additional risk factors include social influences, dietary habits, and psychological elements.[5]

Despite years of ongoing research, managing and preventing obesity remains a challenging task due to our limited understanding of the condition and the contributing factors, like environmental and biological elements, interact in complicated ways. To create a better machine learning model to forecast obesity using a local dataset that includes key features of obesity. The specific objectives are as follows:

1. Identify the optimal set of features for predicting obesity.
2. Develop a model capable of predicting individual obesity utilizing machine learning techniques with data sourced locally.
3. Use a supervised machine learning algorithm to better predict obesity levels in the future
4. Check how well the model performs by using suitable evaluation metrics.

II. LITERATURE REVIEW

[6] Multiple studies have explored the link between weight and activity, comparing various machine-learning and statistical models to predict obesity levels.

[7] Another study applied learning techniques such as regression, Neural Networks and Decision Trees to forecast obesity levels using sensors data, smartphones, and medical records.

[8] Subsequent research aimed to project future obesity rates among adults in various countries by employing: normal weight, overweight, moderate obesity, and severe obesity. Based on data from the Surveillance System survey, the prediction indicates that by 2030, 52.4% of adults will be obese, with 22.5% classified as critically obese, and obesity rates in 25 states exceeding 30%.

[9] Additionally, a study focused on patterns of BMI increase in children developed a system that identifies people at severe risk of becoming obese before it gets severe. The research found that the most significant BMI increase occurs between ages 2 and 4, with accurate predictions possible at ages 5 to 6.

[10] logistic regression model was employed to estimate BMI in children aged 2 to 17 in rural regions utilizing publicly available datasets. The findings indicated that small areas are crucial for developing interventions, with obesity rates ranging from 27% to 40%.

[11] In the realm of jungle computing, the integration of clusters and edges facilitates advanced techniques for predicting obesity. Federated learning, as applied to health data, allows for collaborative machine learning across various devices without centralizing sensitive patient information. This decentralized approach effectively identifies patterns and trends in health-related data, enabling early detection of risk factors associated with obesity. By continuously updating the predictive model with data from various sources, it is possible to track changes in health behaviors and environmental factors that contribute to obesity. This approach enhances the accuracy and scalability of obesity prediction systems while maintaining patient privacy, paving the way for personalized and timely interventions.

[12] The initial approach to identifying obesity using data collected from parents, guardians, and children. The authors examined risk factors, including obesity, parental education, children's behaviors, and environmental influence.

[13] Data mining techniques were used to predict childhood obesity, with models such as Neural Networks, Naïve Bayes, and Decision Trees employed for implementation.

[14] A study employed learning models, Decision Trees, Support Vector Machines, and Bayesian Networks, that classify obesity in sixth-grade children across two districts.

[15] A published article examining obesity prediction using the random forest method demonstrated the highest accuracy, achieving an accuracy rate of 85%.

[16] Ensemble learning is used to identify obesity in people based on age and BMI. The researchers used algorithms like linear models, random forests, and partial least squares, with random forest achieving 89.68% accuracy.

In a study predicting future adult [17] obesity levels using multinomial regression, researchers noted a limitation: Only one method was used for handling multiple classes, and the dataset was based on self-reported information, which could introduce biases.BMI alone may not accurately identify obesity, as it can result in false positives for athletes whose high BMIs are due to muscle rather than fat.

[18] It's important to add more features to create better machine learning models instead of just using manual or statistical methods. This research aims to enhance an existing model for more efficient obesity prediction with minimal human intervention, utilizing a local dataset with principal features.

III. MATERIALS AND METHODS

A. Overview of the Existing System

In obesity research, data mining and machine learning are used to identify risk factors and create prediction models. While effective, some argue that overreliance on algorithms may lead to biased predictions and overlook important contextual factors.

B. Review of the Existing System

Traditionally, clinical procedures for predicting obesity have been time-consuming, as they require trained physicians to perform necessary diagnostic processes. Complications often arise due to delayed intervention, as manual methods are typically implemented after symptoms have already appeared in patients. Recently, numerous studies have explored obesity prediction using machine learning techniques, yielding varying accuracy levels depending on individual challenges.

To diagnose obesity, healthcare professionals perform physical exams and recommend tests. These typically include gathering patient health history, conducting a physical examination, calculating BMI, measuring waist size, looking at lifestyle and family history.

C. Problems with the Existing System

The current system faces several issues:

1. Inaccurate Diagnosis: *Patients may not be assessed by experts, leading to inaccurate obesity diagnoses.*

2. Limited Features: The reliance on a single feature, body measurement, may result in incorrect predictions.

3. Low Accuracy: The system is prone to low accuracy in predictions.

D. Description of the Proposed System

The proposed model will assist medical personnel in predicting obesity in patients more efficiently. It will also help determine the class or level of obesity, facilitating early intervention decisions. The system offers several advantages:

1. Increased Reliability: Enhances the accuracy of diagnostic results.

2. Reduced Diagnostic Burden: Minimizes the challenges associated with diagnosing obesity.

3. Real-Time Classification: Provides a system for real-time classification of obesity.

4. Time and Cost Savings: Reduces the time and cost associated with conducting multiple medical tests.

E. Viability Assessment of the Proposed System

The proposed model looks at how well it works, costs, user needs, benefits to the organization, and technical challenges during development. This analysis will assess whether the system can be successfully implemented based on several factors:

1. ***Schedule Feasibility:*** *This looks at how long the project will take and if it meets the expected needs. It will depend on the resources available, how relevant it is to the study, and the project timeline. A Gantt chart can help suggest a schedule for completing the project.*
2. ***Economic Feasibility:*** *The system is cost-effective. The analysis will show that the proposed model is realistic regarding its costs.*
3. ***Operational Feasibility:*** *This evaluates how effectively the model will interact with users. The model will feature a simple user interface to enhance operational ease and flexibility.*
4. ***Technical Feasibility:*** *This involves checking the model's actual setup and its technical needs, including technology and skills required.*

F. Overview of the Proposed System

The proposed system's SVM model consists of the information management module and the p analysis module. Former handles patient data, while the latter determines the level of obesity.

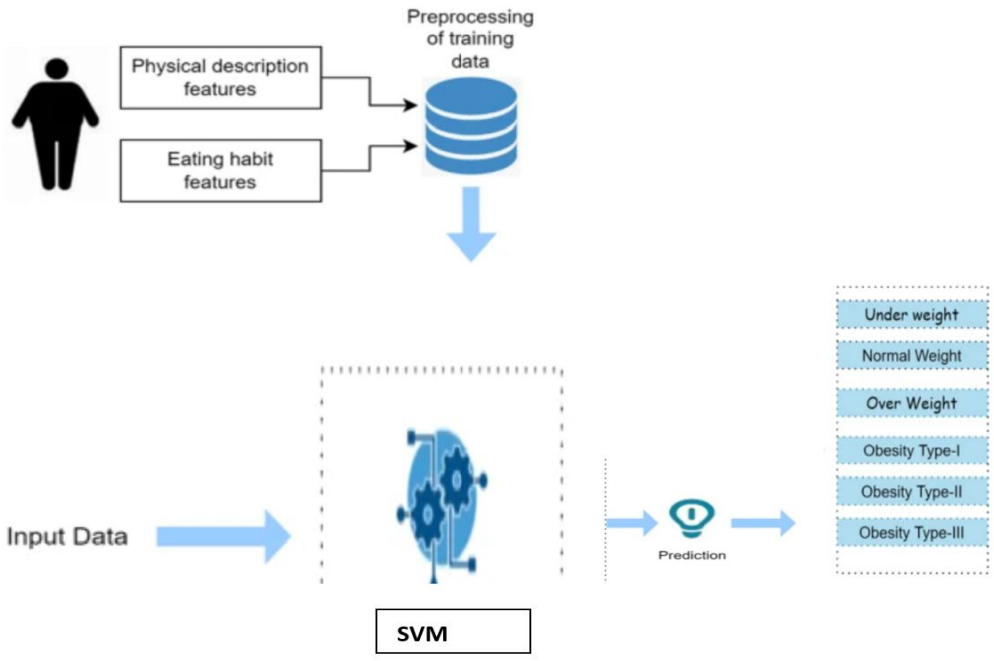

Fig. 1 SVM Model for Obesity Prediction

Prediction Module: This main part of the model uses the patient's details to make predictions and then sends the results to the analysis module.

IV. METHODOLOGY

Section explains how model's goals will be achieved. It uses a well-known method in machine learning projects to plan, organize, and carry out the project. Python, along with machine learning algorithms and libraries, will be used for this research.

A. Rationale and Details for the Chosen Methodology

The six-phase methodology is a repeating process that helps assess and review progress. It starts with understanding the research concept and goes through modeling and deployment. Since the data for this project comes from a healthcare center database, this method is a good fit for analyzing and building the proposed model.The methodology phases include:

1. *Business Understanding: Engage with experts to gain insights into obesity, its causes, symptoms, and current prediction methods.*
2. *Understanding the Data: Familiarize with the data by evaluating its quality and the relationships between attributes.*
3. *Data Preparation: nvolves data cleaning (standardizing formats), data selection (picking relevant features), data reduction (removing unnecessary data), and data transformation.*
4. *Modeling: The model is trained and tested. Machine learning algorithms used in this research include:*

    - Gradient Boosting Classifier: An ensemble technique used for regression and classification that enhances predictions by sequentially adding trees to correct errors made by earlier models.
    - Random Forest Classifier: A versatile algorithm that creates an ensemble of decision trees through bagging, which boosts model performance.
    - Decision Tree Classifier: A supervised learning method that addresses classification and regression tasks using a tree structure to map out decisions and their potential outcomes.
    - Evaluation: Assess the model's performance and accuracy.
    - Deployment: Implement the model in a practical environment.

Support Vector Machine (SVM): Support Vector Machine (SVM) is a powerful supervised learning method often used for classification and regression. In obesity research, SVM can identify patterns and risk factors related to the condition, aiding in early detection and intervention. A recent study in Indonesia demonstrated the potential of machine learning in analyzing obesity data. The researchers found that logistic regression had the highest accuracy in predicting obesity among the methods they tested. Another study [19] compared the performance of Support Vector Machine and Modified Balanced Random Forest in detecting diabetes, a common comorbidity of obesity.

One of the key advantVges of the Support Vector Machine algorithm is its ability to handle high-dimensional data and complex non-linear relationships, making it well-suited for the analysis of obesity data. (DeGregory et al., 2018) By identifying key risk factors and developing predictive models, healthcare professionals can implement targeted interventions to address the obesity epidemic.

The research in this field is ongoing, and the application of machine learning algorithms, such as Support Vector Machine, continues to provide valuable insights into the multifaceted nature of obesity.

V. RESULT

The prediction model was checked for accuracy to see how well it meets the goal. The Support Vector Machine achieved an accuracy of 97.14%.

Result for Support Vector Machine.

```
In [35]: from sklearn.svm import SVC
         classifier = SVC(kernel = 'linear', random_state = 0)
         classifier.fit(x_train, y_train)

Out[35]: SVC(kernel='linear', random_state=0)

In [36]: y_predict = classifier.predict(x_test)

In [37]: accuracy = metrics.accuracy_score(y_test, y_predict)
         accuracy

Out[37]: 0.9716312056737588
```

VI. CONCLUSION

Diagnosing obesity presents significant challenges due to the complex and variable nature of the disease. Enhancing obesity diagnosis within the healthcare sector is crucial to minimizing associated risks and complications. Traditionally, assessing a patient's obesity status requires physical examinations and analysis of test results, often subject to a physician's interpretation. This project addresses these issues by developing an obesity prediction model designed to improve diagnostic accuracy, reduce time consumption, and provide a reliable tool applicable to all genders.

The research successfully created a time-efficient and accurate obesity prediction model using Python. The model incorporates various machine learning algorithms, including Gradient Boosting Classifier, Random Forest Classifier, K-Nearest Neighbor, and Support Vector Machine. Among these, the Gradient Boosting algorithm achieved the highest accuracy of 99.05%.

This prediction model is highly recommended for use in hospitals, clinics, diagnostic centers, and the broader healthcare sector. It offers a reliable method for predicting obesity status before it reaches an advanced stage, enhancing early intervention and patient care.

REFERENCES

[1] Eduado, D., Fabio, E., & Mendoza, P. (2019). Obesity Level Estimation Software based on Decision Tree. Jornal of Computer Science, 67-77.

[2] WHO. (2021, June 23). Obesity Related Diseases. Retrieved from World Health Organisation

[3] Guterrez, H. M. (2010). Diez problems de la poblacion de jalisco. Una perspective sociodeografica, 25-30.

[4] Hernandez, J. (2011). Obesity and its causes. International Journal for Medical Image.

[5] Del Cisne, P., & Zhingre, O. (2015). Factors that Influence Obesity. 10-13.

[6] Xiaolu, C., Shuo-yu, L., Jin, L., Shiyong , L., Jun, Z., Peng, N., et al. (2021). Does Physical Activity Predict Obesity-A Machine Learning and Statistical Method-Based Analysis. International Journal of environmental research and public Health.

[7] DeGregory, K. W., Patrick , K., DeSilvio, T., & Pleuss, J. D. (2018). A review of machine learning in obesity: machine learning in obesity reaserch. ResearchGate.

[8] Zachary, J., Ward, M. P., Sara, N. B., Angie, L., Cradock, Jessical, L., et al. (2019). Projected US State Level Prevalence of Adult Obesity and Severe Obesity. The New England Journal of Medicine, 2440-2450.

[9] Hagai, R., Smadar, S., Shiri, B.-H., Becca, F., Aron, W., & Eran, S. (2021). Prediction of Childhood Obesity from Nationwide Health Records. The Journal of Pediatrics, 132-140.

[10] Xuegin, P., Christopher, B. F., Felice, L.-S., & Aaron, J. M. (2021). Prediction of Early Childhood Obesity with Machine Learning and Electronic Health Record Data. International Journal of Medical Informatics.

[11] Intrusion Detection Using Federated Learning for Computing R. S. Aashmi1, and T. Jaya.

[12] Manna , S., & Jewkes, A. M. (2014). Understanding early childhood obesity risks: An empirical study using fuzzy signatures. IEEE international confluence (pp. 1333-1339). Beijing China : Xplore press.

[13] Adnan, M. H. (2011). A framework for childhood obesity classifications and predictions using NBtree. Proceedings of the 7th International Conference on Information Technology in Asia (pp. 1-6). Kuching, Sarawak, Malaysia: IEEE Xplore Press.

[14] Adnan, M. H., Damanhoori, F. D., & Husain, W. (2010). A survey on the usefulness of data mining for childhood obesity prediction. Proceedings of the 8th AsiaPacific Symposium on Information and Telecommunication Technologies (pp. 1-6). Kuching, Malaysia: IEEE Xplore Press.

[15] Abdullah, F. S., Manan, N. S., Ahmad, A., Wafa, S. W., & Shahril, M. R. (2016). Data miningtechniques for classification of childhood obecity among year 6 school children. proceeding of the international conference on soft computing and data mining (pp. 465-474). Springer: IEEE Xplore press.

[16] Tamara, M. D., Mukhopadhyay, S., Aaron , C., & Stephen, M. D. (2015). Machine lerning techniques for prediction on early childhood obesity. Researchgate.

[17] Kapil, J., Niyati, B., & Prashant, S. R. (2018). Obesity prediction using ensemble machine learning approaches . Researchgate.

[18]Perbandingan Support Vector Machine dan Modified Balanced Random Forest dalam Deteksi Pasien Penyakit Diabetes Purbolaksono, M. D., Irvan Tantowi, M., Imam Hidayat, A., & Adiwijaya, A. (2021).

[19] A review of machine learning in obesity. KW DeGregory, P Kuiper, T DeSilvio, JD Pleuss… - Obesity …, 2018.