Estimating Obesity Levels in Individuals Using Classification Methods

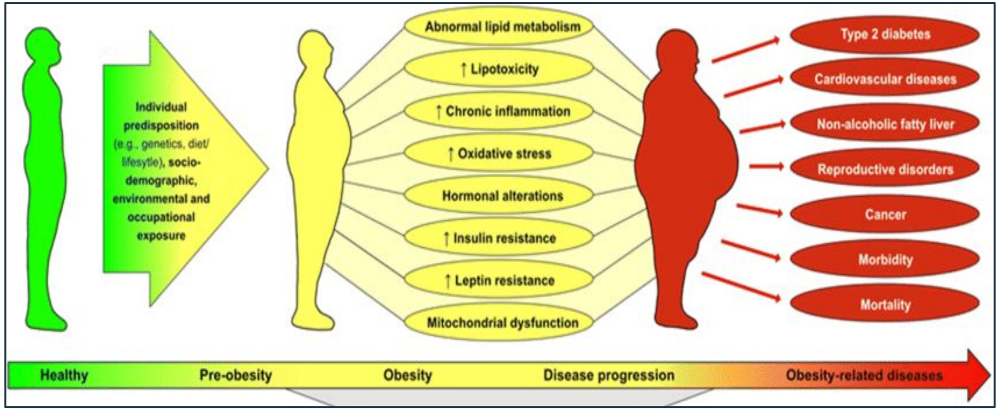
Abdinasir Ali, Amina Danlami, Ben Earl-Kinley, Ethan Sadler, Isaac Adeyeye, Sara Polley Farjoodi

Overview

- Introduction
- Machine Learning Benefits
- Related Work
- Approach
- Dataset
- Preprocessing
- Experimental Methods
- Baseline Methods
- Evaluation

Introduction

- 1 in 8 are obese worldwide
- Potential health complications
- Strain on healthcare systems



'"Omics" platforms in obesity research' (2024)

How can machine learning classification techniques aid in alleviating pressure on healthcare systems?

- Predict obesity of the patients through datasets
- Helps doctors analyze contributing factors to obesity
- Determine the most at-risk individuals
- Lessen burden on healthcare facilities

Related Work

Visualization obesity risk prediction system based on machine learning

Jinsong Du, Sijia Yang, Yijun Zeng, Chunhong Ye, Xiao Chang & Shan Wu

Scientific Reports **14**, Article number: 22424 (2024) | Cite this article

1465 Accesses | **3** Altmetric | Metrics

- A visualized obesity risk prediction system based on machine learning techniques
- Uses largely medical data and measurement as features
- They use multiple models which each score AUC values of around 0.95:
    - Random Forest, **XGBoost**, LGBoost, Gradient-Boosted Trees, Back Propagation Neural Network, Linear Regression
- Hip circumference, Chest circumference, Body fat mass, Diet and Triglycerides were the most important features in determining risk
- Creates a visualization of the risk for each patient
- Various works from other groups:
    - **"Age-specific risk factors for the prediction of obesity using a machine learning approach"** (Front Public Health. 2023 Jan 17;10:998782. doi: 10.3389/fpubh.2022.998782)
    - **"Machine learning model to predict obesity using gut metabolite and brain microstructure data"** (Osadchiy, V., Bal, R., Mayer, E.A. et al. Machine learning model to predict obesity using gut metabolite and brain microstructure data. Sci Rep 13, 5488 (2023). https://doi.org/10.1038/s41598-023-32713-2)

Du, J., Yang, S., Zeng, Y. *et al.* Visualization obesity risk prediction system based on machine learning. *Sci Rep* 14, 22424 (2024). https://doi.org/10.1038/s41598-024-73826-6

Du, J., Yang, S., Zeng, Y. *et al.* Visualization obesity risk prediction system based on machine learning. *Sci Rep* 14, 22424 (2024). https://doi.org/10.1038/s41598-024-73826-6

Approach

- **6 Classifier Models:**
  - Weighted k-Nearest Neighbours (kNN)
  - Decision Tree
  - Logistic Regression
  - Gaussian Naive Bayes
  - Random Forest
  - Support Vector Machine (SVM)
- Measured through Accuracy and Precision scores
- Ensure that the most accurate model is used
- Implementation: sklearn API and libraries

Implementation Overview

Dataset

Continuous	Categorical	Binary	Integer
• Age • Height • Weight • How many main meals do you eat daily? • How much water do you drink daily? • How often do you have physical activity?	• Gender • Do you eat any food between meals? • How often do you drink alcohol? • Which transportation do you usually use?	• Has a family member suffered or suffers from being overweight? • Do you eat high caloric food frequently? • Do you smoke? • Do you monitor the calories you eat daily?	• Do you usually eat vegetables in your meals? • How much time do you use technological devices, such as phones or computers?

Target:

Insufficient Weight	Normal Weight	Overweight Level I	Overweight Level II	Obesity Type I	Obesity Type II	Obesity Type III

Preprocessing

Data Preprocessing Pipeline

- Confirmed no missing values in the dataset
- Detected and removed outliers in numerical features (Age, Height, Weight) using the IQR method

Encoding & Transformation

- Applied one-hot encoding to categorical variables (e.g., family obesity history)
- Simplified target variable to binary classification (obesity: true/false)

Feature Selection

- Identified key predictive features: Height, Weight, Physical Activity Frequency, Transportation Modes, and Encoded Binary Features

Experimental Methods

- **Models Implemented**
  - kNN algorithm and Decision Tree Classifier
- **Evaluation Metrics**
  - **kNN**:
    - Evaluated with Manhattan & Euclidean distances using 10-fold cross-validation
    - Accuracy: **94.6%**, Precision: **94.7%**
  - **Decision Tree Classifier**:
    - Trained on 80-20 train-test split
    - Accuracy: **96.9%**, Precision: **99.4%**
- **Reliability Indicators**
  - Confusion matrix visualization highlighted Decision Tree's robustness in handling mixed features and complex data relationships
  - Models effectively classified individuals into "Obese" and "Not Obese" categories, validating their reliability for the task

Baseline Methods

Logistic Regression

- **Pros**: Simple, interpretable for feature relationships.
- **Cons**: Struggles with complex feature-target patterns.

Naive Bayes

- **Pros**: Efficient with categorical data.
- **Cons**: Assumes unrealistic feature independence.

Random Forest

- **Pros**: Handles mixed data types, resists overfitting.
- **Cons**: Computationally intensive with multiple trees.

Baseline Methods

Support Vector Machines (SVM)

- **Pros**: Effective for clear class boundaries, models non-linear relationships
- **Cons**: Can underperform with large datasets

Weighted kNN

- **Pros**: Emphasizes closer neighbors for improved classification

Evaluation Metrics

The following metrics were used:

- **Accuracy:** Percentage of correctly classified obesity levels
- **Precision:** Proportion of true positive predictions out of all positive predictions
- **Recall:** Proportion of true positive predictions out of all actual positive cases
- **F1-Score:** Harmonic mean of precision and recall, balancing false positives and false negatives
- **ROC-AUC:** Measures model ability to distinguish between classes

Model	Accuracy	Precision
K-Nearest Neighbors (kNN)	0.9375	0.9521
Decision Tree	0.9643	0.9937
Random Forest	0.9509	0.9873
Support Vector Machines (SVM)	0.9464	0.9581
Logistic Regression	0.9732	0.9877
Naive Bayes	0.8661	0.9592

References

Front Public Health. 2023 Jan 17;10:998782. doi: 10.3389/fpubh.2022.998782

Obesity and overweight. World Health Organization. (2024, March 1). https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

Omics Biomarkers in Obesity: Novel Etiological Insights and Targets for Precision Prevention - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Omics-platforms-in-obesity-research_fig1_342504751 [accessed 1 Dec 2024]

Osadchiy, V., Bal, R., Mayer, E.A. et al. Machine learning model to predict obesity using gut metabolite and brain microstructure data. Sci Rep 13, 5488 (2023). https://doi.org/10.1038/s41598-023-32713-2

Q&A