

An In-Depth Analysis of Obesity Prediction through Machine Learning Methods

Dr.Kolluru Venkata Nagendra¹, Dr.Krishna Prasad²

PDF Research Scholar¹, Research Guide²

*Department of Computer Science and Engineering^{1,2},
Srinivasa University, Mangalore, Karnataka State, India.^{1,2}*

Abstract: Obesity represents a significant public health issue, as it is linked to various health conditions, including diabetes, cardiovascular diseases, and certain types of cancer. The Body Mass Index (BMI), which is calculated by dividing an individual's weight in kilograms by the square of their height in meters, is the primary method employed to diagnose obesity. Individuals with a BMI exceeding 30 kg/m² are categorized as obese, while those with a BMI over 25 kg/m² are classified as overweight. Despite its widespread use due to its simplicity and affordability, BMI has limitations, particularly in its inability to consider variations in muscle mass, bone density, and fat distribution. Consequently, alternative assessment methods such as skinfold thickness measurement, bioelectrical impedance analysis, and dual-energy X-ray absorptiometry (DXA) are utilized to evaluate body fat and muscle mass. The research indicated that the Random Forest model (RF) achieved the highest accuracy rate of 95.78%. Other algorithms MLP, SVM, RS, NB and DT also yielded commendable accuracy results. This study implies that the practical implementation of the model could assist healthcare professionals in identifying individuals who are overweight or obese, thereby facilitating the early detection, prevention, and management of obesity-related health issues.

Keywords: Obesity Detection, Obesity Prevention, Eating habits, Classification, Machine Learning, Body Mass Index, Rough Set, Support Vector Machines.

I. INTRODUCTION

The World Health Organization (WHO) characterizes overweight and obesity as conditions marked by an abnormal or excessive accumulation of fat in the body, which can adversely impact an individual's overall health. A widely used metric for evaluating these conditions is the Body Mass Index (BMI), which is calculated by dividing a person's weight in kilograms by the square of their height in meters. A BMI of 25 or higher indicates overweight, while a BMI exceeding 30 is classified as obesity.

Obesity represents a significant public health challenge and is linked to a heightened risk of various health issues, including diabetes, cardiovascular diseases, and certain types of cancer.[1]

Obesity is a multifaceted condition influenced by several factors, with the primary one being the imbalance between energy intake and energy expenditure. This imbalance disrupts numerous metabolic processes, leading to increased fat cell production and the storage of excess energy as triglycerides. Additionally, factors such as inadequate nutrition, insulin resistance, systemic inflammation, and genetic factors significantly contribute to the onset of obesity.[3]-[8].

II. CAUSES OF OBESITY

Obesity fundamentally occurs when an individual consumes more calories than their body can utilize. Several factors may contribute to why a person might ingest more food than necessary:

a) Certain medications: Some medications prescribed for various health conditions can lead to weight gain. Examples include antidepressants, steroids, anti-seizure drugs, diabetes medications, and beta-blockers.

b) Disability: Both adults and children with physical or learning disabilities are at a heightened risk for obesity. Physical limitations, along with insufficient access to specialized education and resources, can exacerbate this issue.

c) Eating habits: Consuming excessive calories, particularly from ultra-processed foods, sugary items, and those high in saturated fats, can lead to overweight conditions.

d) Genetics: Studies indicate that individuals with obesity may possess specific genes, known as obesity-susceptibility genes that influence appetite. However, it remains uncertain whether those who are overweight share the same genetic characteristics.

e) Lack of physical activity: Excessive screen time—such as watching television, playing video games, or using mobile devices can significantly reduce the time available for physical exercise.

f) Lack of sleep: Failing to obtain at least seven hours of sleep can disrupt the hormones responsible for regulating hunger.

g) Stress: The body's response to stress involves the production of hormones like cortisol, which influence hunger. During stressful periods, individuals are more inclined to consume high-fat and high-sugar comfort foods, which the body tends to store as additional fat.

III. THE COMPLICATIONS OF OBESITY

Obesity impacts the body in numerous ways, leading to metabolic alterations that heighten the likelihood of severe health conditions. It can exert both direct and indirect influences on your overall well-being.

Metabolic Alterations: Metabolism refers to the process by which your body transforms calories into energy necessary for its functions. When calorie intake exceeds the body's energy requirements, the surplus is converted into lipids and stored as body fat. Once the storage capacity for lipids is reached, fat cells expand. These enlarged fat cells release hormones and other substances that trigger inflammation. This inflammation can result in insulin resistance, impairing the body's ability to utilize insulin effectively to reduce blood sugar and fat levels. Elevated blood sugar and fat levels can lead to hypertension, and together, these factors contribute to metabolic syndrome, a prevalent condition associated with obesity. Metabolic syndrome further elevates the risk of developing various diseases, including:

a) Cardiovascular Diseases: Obesity significantly raises the risk of cardiovascular issues, such as coronary artery disease, congestive heart failure, heart attacks, and strokes.

b) Fatty Liver Disease: Excess fats in the bloodstream can accumulate in the liver, which plays a crucial role in blood filtration. When the liver starts to store excess fat, it may result in chronic inflammation (hepatitis) and long-term damage (cirrhosis).

c) Gallstones: Elevated cholesterol levels in the blood can lead to cholesterol build-up in the gallbladder, increasing the risk of cholesterol gallstones and related gallbladder disorders.

d) Kidney Disease: Conditions such as high blood pressure, diabetes, and liver disease are significant contributors to chronic kidney disease.

e) Type 2 Diabetes: Obesity is a specific risk factor for the development of Type 2 diabetes.

Statistically, obesity is associated with an increased risk of premature mortality from various causes. Research indicates that even modest weight loss (between 5% and 10% of your current weight) can significantly reduce this risk.

Direct effects: Excessive body fat can compress the organs within the respiratory system, leading to additional stress on the musculoskeletal system. This can result in various health issues, including:

- ✓ Arthritis.
- ✓ Asthma.
- ✓ Back pain.
- ✓ Obesity hypoventilation syndrome.
- ✓ Sleep apnea.

Being obese heightens the likelihood of developing several conditions and diseases, such as:

- Certain types of cancer, including esophageal, pancreatic, colorectal, breast, uterine, and ovarian cancers.
- Depression and other mood disorders.
- Female infertility and complications during pregnancy.
- Cognitive and memory-related issues, with an increased risk of Alzheimer's disease and dementia.

IV. OBESITY -PREVENTION

Preventing obesity is significantly more manageable than addressing it after it has developed. This is due to the body's ability to regulate body mass by adjusting its responses to hunger signals in relation to the energy expended through daily activities. Once a new elevated "set point" is established, the body perceives this as the new normal weight, which may result in a higher position on the scale or BMI chart. By evaluating your current habits and implementing sensible modifications now, you can avert the onset of obesity in the future. Here are some suggestions:

a) Implement minor adjustments: If you have a routine of daily snacks or energy boosters, such as high-calorie sugary beverages, consider substituting them. An additional 150 calories consumed each day can accumulate to 10 extra pounds over the course of a year, which is comparable to a small bag of potato chips or just two double-stuffed cookies.

b) Increase physical activity: Alternatively, think about how you can expend an extra 150 calories daily by engaging in an activity that suits your fitness level.

c) Shop with intention: Fill your pantry with nutritious foods and reserve sweets and indulgent treats for special occasions when dining out.

d) Foster overall wellness: Limit your screen time, enjoy the outdoors, and ensure you manage stress effectively while getting sufficient sleep to maintain balanced hormone levels. Concentrate on making positive lifestyle changes and engaging in healthy activities rather than fixating on how these efforts impact your weight.

It is crucial to emphasize that the health risks associated with obesity are just a fraction of the many potential complications linked to this condition. Therefore, the treatment and management of obesity are vital to reduce the likelihood of these and other chronic health issues. [1],[8].

Efforts to combat obesity necessitate a comprehensive strategy that promotes a healthy lifestyle, which includes regular physical activity and a nutritious diet. Public policies, health education, and health promotion initiatives are vital in addressing the obesity epidemic. Accurately assessing the severity of an individual's obesity is essential for proper diagnosis and tailored treatment. In this regard, machine learning stands out as a powerful resource, capable of identifying individuals who may benefit from specific dietary plans or interventions. These algorithms are adept at extracting insights from data, and numerous studies have demonstrated their effectiveness in predicting and managing obesity. [9] [11].

Machine learning encompasses a variety of sophisticated algorithms that excel in analysing and interpreting large datasets. These algorithms can classify, model, learn, predict, and analyse data, thereby enhancing our comprehension of obesity and improving our predictive abilities. They can uncover patterns and trends within data that may be too intricate for human analysis, facilitating early diagnosis, treatment, and management of obesity. Additionally, machine learning can be utilized to forecast future outcomes, such as the risk of developing obesity-related health conditions, which assists healthcare providers in creating more effective intervention strategies. In summary, machine learning holds significant promise for deepening our understanding of obesity and advancing more effective prevention and treatment approaches. [9], [12]-[14].

V. METHODS

The primary objective of this research is to employ machine learning techniques to establish tailored treatment plans for patients. In this study, body mass index (BMI) serves as a key indicator of obesity, which is subsequently categorized into various classifications: underweight, normal weight, overweight level I, overweight level II, obese type I, obese type II, and obese type III. This categorization is derived from the following equation and adheres to the standards set by the World Health Organization (WHO). A detailed classification of BMI in accordance with WHO guidelines is presented in Table 1. The goal of this study is to create a predictive model that can accurately classify individuals into these defined obesity categories, potentially facilitating early detection and intervention for obesity. [15].

$$\text{Body Mass Index (BMI)} = \frac{\text{Weight (kg)}}{(\text{Height (m)})^2}$$

| | |
|--------------|--------------|
| Under weight | <18.5 |
| Normal | 18.5 to 24.9 |
| Over Weight | 25.0 to 29.9 |
| Obesity I | 30.0 to 34.9 |
| Obesity II | 35.0 to 39.9 |
| Obesity III | >40 |

Table 1: According to WHO - BMI Classification

5.1. Dataset

The dataset utilized in this research [15] offers insights that facilitate the evaluation of an individual's obesity level based on their eating patterns and physical health. Comprising 2111 instances (rows) and 17 variables (columns), the dataset features a mix of numeric, binary, and categorical input variables. It includes a singular target variable that indicates a person's obesity level. This dataset, named "NObesity," categorizes individuals into several classifications: Underweight, Normal Weight, Obesity Level I, Obesity Level II, Obesity Type I, Obesity Type II, and Obesity Type III. A detailed summary of the data variables, including their names, types, and definitions, is presented in Table 2. This dataset serves as the foundation for training and assessing the machine learning models created in this study, aimed at predicting individual obesity levels. Each model incorporates a distinct set of patient data alongside a specific target variable. The input variables for the model encompass demographic details, anthropometric data, and information regarding dietary habits and physical activity.

Table 2. Dataset description

| Attributes | Values |
|--|--|
| Gender | 1 = Female or 0 = Male |
| Age | Numeric |
| Height | Numeric |
| Weight | Numeric |
| Family with overweight / obesity | 1 = Yes/ 0 = No |
| FAVC (frequent consumption of high caloric food) | 0 = Yes/ 1 = No |
| FCVC (frequent consumption of vegetables) | 1,2 or 3 |
| NCP (number of main meals) | 1, 2, 3 or 4 |
| CAEC (consumption of food between meals) | (1 = No, 2 = Sometimes, 3 = Frequently or 4 = Always) |
| Smoke | 0 = Yes/ 1 = No |
| CH20 (Consumption of water daily) | 1, 2 or 3 |
| SCC (Calories consumption monitoring) | 0 = Yes/ 1 = No |
| FAF (Physical activity frequency) | 0, 1, 2 or 3 |
| TUE (Time using technology devices) | 0, 1 or 2 |
| CALC (Consumption of alcohol) | 1 = No, 2 = Sometimes, 3 = Frequently or 4 = Always |
| MTRANS (Transportation used) | Automobile, motorbike, bike, public transportation or walking |
| Obesity level | 1 = Insufficient_Weight, 2 = Normal_Weight, 3 = Overweight_Level_I, 4 = Overweight_Level_II, 5 = Obesity_Type_I, 6 = Obesity_Type_II, 7 = Obesity_Type_III |

5.2. Machine Learning Methods

Machine learning (ML) is a swiftly evolving field that has garnered significant interest in recent years. It encompasses the methodical examination of algorithms and statistical models that enable computer systems to execute tasks without the necessity of direct programming. The primary objective of machine learning is to empower computers to utilize data more effectively and efficiently through a technique known as reinforcement learning. This approach enables computers to learn from data and progressively enhance their performance. Tools have historically been employed to carry out various tasks with greater efficiency. As the volume of available data continues to grow, the demand for machine learning becomes increasingly pressing. This technology is utilized to extract insights from extensive datasets when conventional methods fall short in interpreting the information they contain.

Current research is focused on the creation of autonomous learning robots that do not require explicit programming. This advancement could enable robots to adjust to changing environments and execute tasks with greater efficiency and effectiveness. Machine learning finds applications across numerous sectors, including healthcare, finance, transportation, and manufacturing. It holds the potential to transform our daily lives and work practices, presenting abundant opportunities for innovation and development.

(i). Multi-layer perceptron (MLP)

A multilayer perceptron (MLP) [16], [17] is a type of feedforward artificial neural network frequently employed in supervised learning applications, including classification and regression tasks. The architecture of an MLP comprises three primary layers: the input layer, the output layer, and one or more hidden layers situated between them. The input layer is tasked with receiving and processing the input data, which is subsequently transmitted through the network to the output layer. The output layer generates the final prediction or result based on the processed input data. The hidden layers, positioned between the input and output layers, contain neurons that execute various computational functions.

Within an MLP, neurons are interconnected through weighted links and utilize a nonlinear activation function to handle incoming data. During the training process, these connection weights are modified to enhance the network's performance. This adjustment process, known as backpropagation, functions as a supervised learning algorithm that employs gradient descent to reduce the discrepancy between the predicted and actual outputs. MLPs are particularly effective for addressing nonlinear problems, as they possess the capability to approximate any continuous function. Their applications span numerous domains, including image recognition, natural language processing, and speech recognition.

(ii). Support vector machine (SVM)

Support Vector Machines (SVMs) [18] are a category of supervised learning algorithms frequently employed for classification purposes. These algorithms function by establishing a boundary known as a "hyperplane" that distinguishes various classes of data points within a high-dimensional feature space. The main objective of SVMs is to construct a boundary that maximizes the distance to the nearest data points from each class, thereby ensuring an optimally separated hyperplane. SVMs excel in handling complex data sets of small to medium size and are particularly well-suited for cases where the data cannot be separated linearly, meaning it cannot be divided by a single straight line. In such instances, SVMs utilize a technique referred to as the kernel trick, which transforms the data into a higher-dimensional space to enable linear separation. These machines are applicable across numerous domains in machine learning, including text and image classification, as well as bioinformatics. They are grounded in a robust theoretical framework and have demonstrated outstanding performance in practical applications. Additionally, SVMs are effective for both classification and regression tasks, showcasing high efficiency in managing both linear and nonlinear data. [19]-[20].

(iii). Rough sets (RS)

Rough sets (RS) represent a mathematical framework for data processing and mining, initially introduced by Pawlak in the 1980s. This innovative approach enhances traditional set theory by enabling the expression of complex concepts that may be challenging to grasp through conventional reasoning methods. Within rough set theory, a subset of a universal set is characterized by ordered pairs of subsets and supersets that belong to the universal set itself. A key concept introduced by this theory is "approximate sets," which arise from the equivalence of two distinct numerical sets. These approximate sets are categorized into upper and lower sets, formed using equivalence classes. These sets comprise the union of all equivalence classes that act as subsets of a specified set.

Rough sets find applications across various domains, including data mining, knowledge discovery, and decision-making. Their capacity to address uncertainty and imprecision in data renders rough sets particularly beneficial for managing incomplete or inconsistent information. Additionally, rough set theory supports the theory of complete sets and has

played a crucial role in the creation of new algebraic structures. [21],[22].

(iv). Random forest (RF)

Random Forest is an ensemble learning approach that combines multiple decision trees to address various problems. In this framework, each tree is trained on a distinct random subset of the dataset, and the collective decision or consensus value from these trees determines the output of the Random Forest. This technique is particularly effective for handling high-dimensional and intricate datasets, as it mitigates the risk of overfitting and enhances the model's generalization capabilities.

By training individual trees on different random data subsets, the correlation among the trees is minimized, thereby fostering greater diversity within the ensemble. This approach, commonly known as "bagging" or "bootstrap aggregation" of decision trees, involves developing a robust collection of related trees to tackle a specific issue. Furthermore, the Random Forest algorithm incorporates a randomization aspect, where each tree relies on values derived from an independently generated random vector. This characteristic bolsters the model's robustness and offers additional safeguards against overfitting. [23], [24], and [25].

(v). Naive Bayes (NB)

Naive Bayes is a fundamental learning approach rooted in the work of Thomas Bayes. It employs Bayes' theorem and operates under the premise of strong conditional independence among features concerning the class. Despite the fact that this independence assumption may not hold true in many real-world situations, Naive Bayes remains popular due to its computational efficiency and other advantageous characteristics.

This method is predicated on the idea that the features used to predict a value are independent of the value itself. For instance, when determining the species of a fish based on its length and weight, it is evident that the weight of a fish of a specific species typically depends on its length, and vice versa, which may challenge the independence assumption. Nevertheless, research has indicated that this assumption imposes fewer limitations on classification tasks involving categorical predictions than previously believed. In some instances, unbiased Bayesian learning has resulted in lower error rates

compared to more complex learning methods, such as the development of univariate decision trees. [27]. The quantitative aspect of a Bayesian network consists of three primary components: probability theory, Bayes' theorem, and conditional probability functions. Bayes' theorem operates on the principle that the conditional probability is directly proportional to the likelihood of an event taking place. This framework enables graphical models to effectively illustrate probability distributions through conditional dependencies or independencies. [28],[29].

(vi). Decision table (DT)

Decision tables serve as an effective instrument for examining and comprehending intricate decision-making processes. They offer a methodical and organized depiction of the interconnections among various variables and the associated actions dictated by those variables. This characteristic renders decision tables especially beneficial in domains such as economics, finance, and computer programming, where sound decision-making is crucial.

Moreover, decision tables can facilitate the creation of decision-making software and systems that automate the decision-making process in accordance with the rules outlined in the table. This efficient methodology enhances the consistency and effectiveness of decision-making across diverse applications and industries, leading to more reliable and favourable decision outcomes. [30], [31].

5.3. Results

The findings of the study indicate that the Random Forest (RF) classifier exhibited the highest level of accuracy, achieving a score of 95.78%. In comparison, the Naive Bayes (NB) classifier recorded an accuracy of 67.41%. The Support Vector Machine (SVM) classifier attained a result of 84.23%, while the Decision Table (DT) classifier reached 84.89%. The Rough Set (RS) classifier achieved an accuracy of 87.83%. Additionally, the multi-layer perceptron (MLP) classifier recorded a score of 94.36%. In summary, the Random Forest classifier proved to be the most effective tool for accurately classifying and identifying obesity.

VI. PERFORMANCE EVALUATION

The primary metrics for assessing classification tasks encompass accuracy, precision, recall, and the F1

score. Accuracy measures the proportion of correctly classified instances relative to the total number of instances. Precision indicates the ratio of true positive predictions to the total number of positive predictions made. Recall, also known as sensitivity or the "true-positive rate," reflects the proportion of correct positive predictions against the total number of actual positive instances. The F1 score serves as an indicator of a classifier's equilibrium between precision and recall, calculated as the harmonic mean of these two metrics.

Moreover, ROC (Receiver Operating Characteristic) curves and AUC (Area Under the Curve) are frequently employed to assess the efficacy of binary classification models. ROC curves depict the relationship between the rate of true positive predictions and the rate of false positive predictions across various classification thresholds, while the AUC offers a holistic measure of the classifier's performance by quantifying the area beneath the ROC curve. The following equations used in this study. For a detailed understanding of these formulae, please consult the relevant reference:

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN'} \\
 \text{Precision} &= \frac{TP}{TP + FP'} \\
 \text{Recall} &= \frac{TP}{TP + FN'} \\
 \text{Specision} &= \frac{TN}{TN + FP'} \\
 F - \text{Measure} &= \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}}
 \end{aligned}$$

The confusion matrix consists of four distinct categories: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). True positives (TP) denote the count of positive instances that have been accurately classified, whereas false positives (FP) signify the number of positive instances that have been misclassified. True negatives (TN) indicate the number of negative instances that have been correctly identified, while false negatives (FN) refer to the instances where negative classifications were made incorrectly.

Table 3. Confusion matrix

| | | Actual value | |
|------------------|----------|---------------------|---------------------|
| | | Positive | Negative |
| Predicted values | Positive | TP (True Positive) | FN (False Negative) |
| | Negative | FP (False Positive) | TN (True Negative) |

| Algorithm | Accuracy |
|-----------|----------|
| MLP | 94.16% |
| SVM | 85.24% |
| RS | 88.93% |
| RF | 95.78% |
| NB | 68.45% |
| DT | 84.90% |

Table 5 Accuracy of all ML Methods

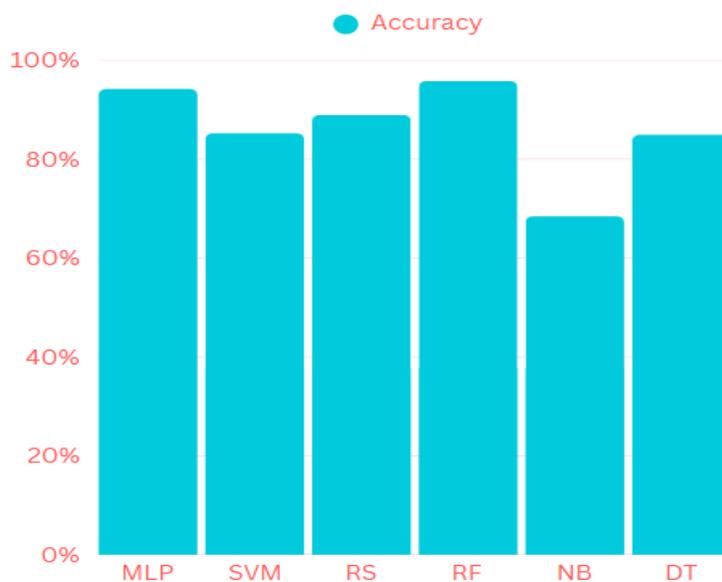


Fig: Accuracy of all ML Methods

The passage above describes a study in which machine learning algorithms were used to classify features associated with obesity and evaluate their performance. Several algorithms were used in the study, including MLP, SVM, RS, RF, NB and DT, which are well established in the field of categorization and have a solid performance history. The models were subjected to ten-fold cross-validation, and their results were presented in tables and graphs. The performance of the models was assessed in terms of accuracy, as shown in Table 5.

VII. CONCLUSIONS

The study proposed to use machine learning classification methods to estimate the prevalence of obesity in people from Andhra Pradesh, using accessible data on their eating habits and physical condition. The study highlights the importance of researching the extent of obesity, as it is associated with a wide range of diseases and affects people of all ages and genders. In the study, Six Machine learning models – MLP, SVM, RS, RF, NB and DT were used to build an intelligent model for detecting overweight or obese individuals, which can help professionals in the field in their decision-making. The results of the study show that RF achieves the highest accuracy of 95.78 % in performance measures. The study concludes that machine learning is an effective tool in medicine that can be used to make timely treatment decisions for people at risk of obesity.

REFERENCES

- [1] S. Hiel et al., "Link between gut microbiota and health outcomes in inulin – treated obese patients: Lessons from the Food4Gut multicenter randomized placebo-controlled trial," *Clinical Nutrition*, Vol. 39, No. 12, pp. 3618–3628, Dec. 2020, <https://doi.org/10.1016/j.clnu.2020.04.005>
- [2] WHO, Obesity and Overweight, World Health Organization, 2020. Available from: <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>.
- [3] R. Legrand et al., "Commensal *Hafnia alvei* strain reduces food intake and fat mass in obese mice—a new potential probiotic for appetite and body weight management," *International Journal of Obesity*, Vol. 44, No. 5, pp. 1041–1051, May 2020, <https://doi.org/10.1038/s41366-019-0515-9>
- [4] K. Prokopidis, E. Chambers, M. Ni Lochlainn, and O. C. Witard, "Mechanisms linking the gut-muscle axis with muscle protein metabolism and anabolic resistance: implications for older adults at risk of sarcopenia," *Frontiers in Physiology*, Vol. 12, p. 770455, Oct. 2021, <https://doi.org/10.3389/fphys.2021.770455>
- [5] J. Xu et al., "Structural modulation of gut microbiota during alleviation of type 2 diabetes with a Chinese herbal formula," *The ISME Journal*, Vol. 9, No. 3, pp. 552–562, Mar. 2015, <https://doi.org/10.1038/ismej.2014.177>

- [6] A. T. Wilkins and R. A. Reimer, "Obesity, early life gut microbiota, and antibiotics," *Microorganisms*, Vol. 9, No. 2, p. 413, Feb. 2021, <https://doi.org/10.3390/microorganisms9020413>.
- [7] T. Klancic and R. A. Reimer, "Gut microbiota and obesity: Impact of antibiotics and prebiotics and potential for musculoskeletal health," *Journal of Sport and Health Science*, Vol. 9, No. 2, pp. 110–118, Mar. 2020, <https://doi.org/10.1016/j.jshs.2019.04.004>
- [8] M. Blüher, "Obesity: global epidemiology and pathogenesis," *Nature Reviews Endocrinology*, Vol. 15, No. 5, pp. 288–298, May 2019, <https://doi.org/10.1038/s41574-019-0176-8>
- [9] K. W. Degregory et al., "A review of machine learning in obesity," *Obesity Reviews*, Vol. 19, No. 5, pp. 668–685, May 2018, <https://doi.org/10.1111/obr.12667>
- [10] R. C. Cervantes and U. M. Palacio, "Estimation of obesity levels based on computational intelligence," *Informatics in Medicine Unlocked*, Vol. 21, p. 100472, 2020, <https://doi.org/10.1016/j.imu.2020.100472>
- [11] E. De-La-Hoz-Correa, F. E. Mendoza-Palechor, A. De-La-Hoz-Manotas, R. C. Morales-Ortega, and S. H. Beatriz Adriana, "Obesity Level Estimation Software based on Decision Trees," *Journal of Computer Science*, Vol. 15, No. 1, pp. 67–77, Jan. 2019, <https://doi.org/10.3844/jcssp.2019.67.77>
- [12] B. J. Lee, K. H. Kim, B. Ku, J.-S. Jang, and J. Y. Kim, "Prediction of body mass index status from voice signals based on machine learning for automated medical applications," *Artificial Intelligence in Medicine*, Vol. 58, No. 1, pp. 51–61, May 2013, <https://doi.org/10.1016/j.artmed.2013.02.001>
- [13] X. Pang, C. B. Forrest, F. Lê-Scherban, and A. J. Masino, "Prediction of early childhood obesity with machine learning and electronic health record data," *International Journal of Medical Informatics*, Vol. 150, p. 104454, Jun. 2021, <https://doi.org/10.1016/j.ijmedinf.2021.104454>
- [14] R. E. Abdel-Aal and A. M. Mangoud, "Modeling obesity using abductive networks," *Computers and Biomedical Research*, Vol. 30, No. 6, pp. 451–471, Dec. 1997, <https://doi.org/10.1006/cbmr.1997.1460>.
- [15] F. M. Palechor and A. L. H. Manotas, "Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from Colombia, Peru and Mexico," *Data in Brief*, Vol. 25, p. 104344, Aug. 2019, <https://doi.org/10.1016/j.dib.2019.104344>.
- [16] E. Disse et al., "An artificial neural network to predict resting energy expenditure in obesity," *Clinical Nutrition*, Vol. 37, No. 5, pp. 1661–1669, Oct. 2018, <https://doi.org/10.1016/j.clnu.2017.07.017>
- [17] M. W. Gardner and S. R. Dorling, "Artificial neural networks (the multilayer perceptron)-a review of applications in the atmospheric sciences," *Atmospheric Environment*, Vol. 32, No. 14-15, pp. 2627–2636, Aug. 1998, [https://doi.org/10.1016/s1352-2310\(97\)00447-0](https://doi.org/10.1016/s1352-2310(97)00447-0).
- [18] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernelbased Learning Methods*. Cambridge University Press, 2000, <https://doi.org/10.1017/cbo9780511801389>.
- [19] L. Auria and R. A. Moro, "Support vector machines (SVM) as a technique for solvency analysis," *SSRN Electronic Journal*, 2008, <https://doi.org/10.2139/ssrn.1424949>
- [20] P. Rivas-Perea, J. Cota-Ruiz, D. G. Chaparro, J. A. P. Venzor, A. Q. Carreón, and J. G. Rosiles, "Support vector machines for regression: a succinct review of large-scale and linear programming formulations," *International Journal of Intelligence Science*, Vol. 3, No. 1, pp. 5–14, 2013, <https://doi.org/10.4236/ijis.2013.31002>.
- [21] A. Skowron and S. Dutta, "Rough sets: past, present, and future," *Natural Computing*, Vol. 17, No. 4, pp. 855–876, Dec. 2018, <https://doi.org/10.1007/s11047-018-9700-3>
- [22] Z. Pawlak, "Rough set theory and its applications to data analysis," *Cybernetics and Systems*, Vol. 29, No. 7, pp. 661–688, Oct. 1998, <https://doi.org/10.1080/019697298125470>
- [23] L. Breiman, "Random forests," *Machine Learning*, Vol. 45, No. 1, pp. 5–32, 2001, <https://doi.org/10.1023/a:1010933404324>
- [24] Gérard Biau, "Analysis of a random forests model," *The Journal of Machine Learning Research*, Vol. 13, pp. 1063–1095, Apr. 2012.
- [25] N. Horning, "Random forests: An algorithm for image classification and generation of continuous fields data sets," in *International Conference on Geoinformatics for Spatial*

- Infrastructure Development in Earth and Allied Sciences, 2010.
- [26] D. R. Bellhouse, “The reverend Thomas Bayes, FRS: A biography to celebrate the tercentenary of his birth,” *Statistical Science*, Vol. 19, No. 1, pp. 3–43, Feb. 2004, <https://doi.org/10.1214/088342304000000189>.
- [27] P. Domingos and M. Pazzani, “On the optimality of the simple Bayesian classifier under zero-one loss,” *Machine Learning*, Vol. 29, No. 2/3, pp. 103–130, 1997, <https://doi.org/10.1023/a:1007413511361>.
- [28] F. Itoo, Meenakshi, and S. Singh, “Comparison and analysis of logistic regression, Naïve Bayes and KNN machine learning algorithms for credit card fraud detection,” *International Journal of Information Technology*, Vol. 13, No. 4, pp. 1503–1511, Aug. 2021, <https://doi.org/10.1007/s41870-020-00430-y>
- [29] E. Frank, L. Trigg, G. Holmes, and I. H. Witten, “Technical note: naive bayes for regression,” *Machine Learning*, Vol. 41, No. 1, pp. 5–25, 2000, <https://doi.org/10.1023/a:1007670802811>
- [30] R. R. Holland, “Decision Tables,” *JAMA*, Vol. 233, No. 5, p. 455, Aug. 1975, <https://doi.org/10.1001/jama.1975.03260050061028>
- [31] R. Kohavi, “The power of decision tables,” in *Lecture Notes in Computer Science*, Vol. 912, pp. 174–189, 1995, https://doi.org/10.1007/3-540-59286-5_57
- [32] J. Shen, M. S. Obin, and L. Zhao, “The gut microbiota, obesity and insulin resistance,” *Molecular Aspects of Medicine*, Vol. 34, No. 1, pp. 39–58, Feb. 2013, <https://doi.org/10.1016/j.mam.2012.11.001>
- [33] N. Zmora, J. Suez, and E. Elinav, “You are what you eat: diet, health and the gut microbiota,” *Nature Reviews Gastroenterology and Hepatology*, Vol. 16, No. 1, pp. 35–56, Jan. 2019, <https://doi.org/10.1038/s41575-018-0061-2>
- [34] H. Lin, Y. An, F. Hao, Y. Wang, and H. Tang, “Correlations of fecal metabonomic and microbiomic changes induced by high-fat diet in the pre-obesity state,” *Scientific Reports*, Vol. 6, No. 1, pp. 1–14, Feb. 2016, <https://doi.org/10.1038/srep21618>
- [35] D.-H. Kim, D. Jeong, I.-B. Kang, H.-W. Lim, Y. Cho, and K.-H. Seo, “Modulation of the intestinal microbiota of dogs by kefir as a functional dairy product,” *Journal of Dairy Science*, Vol. 102, No. 5, pp. 3903–3911, May 2019, <https://doi.org/10.3168/jds.2018-15639>.