A COMPARATIVE ANALYSIS OF J48 AND SVM ALGORITHMS IN PREDICTING THE OBESITY OF HUMANS BASED ON DIETARY INTAKE

1st Arce, Azel Keren G.
School of Technology and Computer Studies, Biliran Province State University
azelarce14@gmail.com

2nd Nacua, Mark Kenneth A.
School of Technology and Computer Studies, Biliran Province State University
markynacua41@gmail.com

3rd Espejon, Atasha Mae
School of Technology and Computer Studies, Biliran Province State University
espejonatashamae@gmail.com

4th Aldaya, Marinel B.
School of Technology and Computer Studies, Biliran Province State University
marinelaldaya2124@gmail.com

Abstract

Obesity is a global issue for public health, representing the most significant contributor to chronic diseases such as diabetes, cardiovascular conditions, and cancer. This study discussed how machine learning techniques or methods, namely J48 and SVM can be used to predict obesity based on data from dietary intake. Using a dataset containing attributes such as age, meal frequency, and food intake. The researchers used a predictive design, which includes data preprocessing, modeling, and evaluation through 10-fold cross-validation. Key evaluation metrics such as accuracy, precision, recall, and F1-score were used to assess model performance. The best algorithm was the J48 decision tree, with an overall accuracy of 96.4%, and age and daily meal frequency were the most significant predictors of obesity. These results demonstrate the promise of ML in predicting obesity but also point out the problems of dealing with class imbalances and the need for further refinement of model performance. Future research should involve the using more diverse datasets and incorporating more lifestyle and genetic variables. Other modeling techniques can be applied to enhance predictive capabilities. In improving the accuracy and reliability of these models, this study contributes to the development of data-driven tools for obesity prevention and management, enabling more targeted public health interventions and personalized dietary recommendations.

Keywords: *Obesity, Algorithms, Machine Learning, Data Mining, Performance Metrics, J48, SVM, Decision Tree, Classification Algorithms, Dietary Intake, Health*

1. Introduction

Data mining is the process of discovering interesting patterns and knowledge from large amounts of data [1]. Data mining classifiers find wide applications in medical domain for diverse diagnosis [2]. Different data mining techniques are used for prediction and decision making for different kinds of diseases like cancer, heart disease, diabetes etc. [3]. There are many types of data mining classifiers which can be applied to predict diseases. In this study we are focusing on obesity prediction using Data Mining technique.

Obesity and its attendant conditions have become significant health problems worldwide, and obesity is currently ranked as the fifth most common leading cause of death globally. The World Health Organization (WHO) defines obesity as an “abnormal or excessive fat accumulation that may impair health,” further clarifying that “the fundamental cause of obesity and overweight is an energy imbalance between calories consumed and calories expended” [4], [5]. Multiple studies have demonstrated that obesity is not a simple problem but a complex health issue stemming from a combination of individual factors (genetics, learned behaviors) and substantial causes (unhealthy societal or cultural eating habits, food deserts) [6] [7]. Most researchers also agree that obesity is an “acquired” disease that heavily depends on lifestyle factors (i.e., personal choices), such as low rates of physical activity and chronic overeating, despite its genetic and epigenetic influences. Researchers have also noted that various forms of obesity, including abdominal obesity, are related to increased risk of several chronic conditions and diseases, which include asthma, cancer, diabetes, hypercholesterolemia, and cardiovascular diseases [8], [9].

Thus, while obesity is undoubtedly a condition, it also exacerbates pre-existing conditions and instigates new ones. More specifically, Bischoff et al. [10] Maintained that obesity can affect nearly every organ system, from the cardiovascular (CV) system to the endocrine system, central nervous system, and gastrointestinal (GI) system. In addition, obesity is associated with the growing prevalence of several CV conditions, from hypertension and coronary heart disease (CHD) to atrial fibrillation (AF) and even total heart failure [11]. The obesity epidemic has overgrown over the last few decades into a significant public health challenge in the United States and, increasingly, worldwide. Between

1970 and 2000, the percentage of obese Americans doubled to almost 30% [12], with two-thirds of Americans now overweight [13]. Similar obesity epidemics are underway across the globe. [14], [15], [16]. [17], [18], [19]. Worldwide, nearly half a billion were overweight or obese in 2002 [20].

Over the last several years, ML techniques have been used in several health applications, including disease recognition. However, limited research has focused on associations between ML approaches and obesity. Moreover, understanding the potential association of obesity and its chronic diseases with severe outcomes is vital but still often neglected in previous studies. This study aims to address these gaps by Locating the most significant works related to obesity, their causes, and their risk factors; Identifying the ML techniques used most often and productively to predict obesity; and Surveying associations between obesity and other risks, conditions, and diseases.

1.1 Objective of the Study

This study aims to predict the obesity of a human based on their dietary intake using the J48 and SVM Algorithm.

Specifically, it seeks to:

1. Gather datasets relevant to human obesity based on dietary intake.
2. Create a model using classification and clustering algorithms, including J48 and SVM.
3. Evaluate the model performance in terms of:
    a. Accuracy
    b. Precision
    c. Recall
    d. F1-Score

2. Review of Related Literature

This section presents the review of related literature and studies that are relevant to the topic of predicting obesity based on dietary intake. The review highlights previous research on dietary factors contributing to obesity, the role of predictive modeling in public health. Obesity is commonly recognized as a critical public health issue and has drawn significant interest across the health sciences. In addition to original research using traditional scientific methods, studies in this area have discussed prevention, treatment, and quality of life for those living with obesity, often through SLRs and novel techniques, such as ML. This section summarizes several related studies in preparation for comparisons with the current work and offers an overview of the current literature addressing obesity from several perspectives.

Simmonds et al. [21] A systematic review was conducted later, combined with a meta-analysis, to examine whether BMI and similar measures used to calculate childhood obesity could also predict adult obesity. Their review supported the conclusion that teenage obesity is a notable public health crisis because it often continues into adulthood. Accordingly, acting to reduce teen obesity can also reduce adult obesity. Early action is one of the most suitable approaches because once children have become overweight, this trend often persists through their adolescence and adulthood. Researchers have drawn from various techniques to build predictive and prognostic models for biomedical applications. In addition to the logistic and Cox regression models most often utilized [22], ML techniques have generally shown promising potential [23]. In particular, when used as an algorithmic framework, ML can provide insight into data collection, facilitate the development of inferences, and even derive knowledge from findings. Therefore, ML approaches have already been applied for various prognostic and diagnostic purposes [24]. ML algorithms have also been used across various health and healthcare domains to predict the development or presence of particular health conditions based on pre-determined characteristics [25]. ML has also been particularly utilized in obesity research [26], [27], [28], [29].

Some researchers caution that ML describes a wide-ranging variety of techniques, which can be characterized only broadly depending on whether their learning phases are supervised (i.e., whether a specific algorithm uses outcome data for training). Supervised ML methods include classifiers, unsupervised ones include clustering, and semi-supervised methods include options such as label propagation [30]. Felso et al. [31] also reported that the role of other supposed mediators (ghrelin, screen time, and leptin levels) remains uncertain. Overall, most of the literature on obesity focuses on exploring the potential parameters that cause, impact, and/or worsen obesity in adults, considering the representative samples. Over the last several years, ML techniques have also been used in several health applications, including disease recognition. However, limited research has focused on associations between ML approaches and obesity. Moreover, understanding the potential association of obesity and its chronic diseases with severe outcomes is vital but still often neglected in previous studies. Studies have also indicated infrequency or lack of exercise and individual education levels as determining factors of obesity [32], [33] , [34], [35], [36], [37].

Current research has already identified or verified certain unhealthy habits, including excess drinking, smoking, insufficient exercise, and overeating, as direct causes of obesity and other chronic illnesses [38]. Further studies by Kadouh and Acosta, [39], [40] They have also shown that obesity may be similar to a heterogeneous chronic condition in which numerous factors interact, producing an energy imbalance that leads to increased body weight. Thus, biological, environmental, and behavioral factors are all determinants of obesity. Accordingly, differences in the prevalence of obesity among different population groups could be influenced by various behavioral and environmental factors, predominantly increasing calorie consumption and reduced physical activity. [41], [42]. Another study conducted by Cheng et al. [33] Revealed that significant predictors of adult obesity, particularly around the age of 55 (women and men), include maternal smoking during pregnancy, childhood neurological functions, educational qualifications, trait conscientiousness, and physical exercise.

An earlier study by Cheng and Furnham [43] also identified many of the same traits and maintained that each was significantly, but also independently, associated with chances of adult obesity. Witten and Frank (2011) provide a comprehensive overview of the J48 algorithm in their book

Data Mining: Practical Machine Learning Tools and Techniques. They describe J48 as an implementation of the C4.5 algorithm, which is widely used for classification tasks. The authors emphasize the algorithm's ability to handle both categorical and continuous data, making it suitable for various applications, including health-related studies [44]. Bhatia and Singh (2018) investigate the prediction of obesity using the J48 algorithm, focusing on lifestyle factors such as diet and physical activity. Their study reveals that J48 can effectively identify key predictors of obesity, providing insights into the behavioral aspects of weight management. The authors emphasize the potential of decision tree algorithms in developing personalized health recommendations [45].

Dehghani and Hossain (2019) apply the J48 decision tree algorithm to predict obesity in adults. Their research highlights the algorithm's ability to analyze various health indicators, such as body mass index (BMI), dietary patterns, and exercise habits. The study demonstrates that J48 can serve as a practical tool for healthcare professionals in assessing obesity risk and implementing preventive measures [46]. Uçar et al. [47] estimated individual percentages of body fat using hybrid machine learning algorithms, such as MLFFNN, support vector machine regression model (SVM), and decision tree (DT) regression. This study also used real data sets, which comprised 13 anthropometric measurements of actual individuals. Dunstan et al. [48] implemented three different ML algorithms (SVM, RF and extreme gradient boosting) for nonlinear regression in predicting obesity at the national level. Their method was validated considering its absolute prediction error and the proportion of surveyed countries, in which the prevalence of obesity was predicted satisfactorily. Their study forecasted that flours and baked goods, dairy-product cheeses, and sugar-sweetened carbonated drinks were the food categories that most closely and accurately predicted the prevalence of obesity.

From the literature review, it is made clear that obesity, the most complicated public health issue, is brought on by biological, behavioral, and environmental causes of influence. Machine learning-another good tool to predict and study the risk of obesity. It shows that using machine-learning techniques like J48, SVM, and decision trees can evaluate and analyze dietary patterns, lifestyle factors, and various health indicators to predict the prevention and management of obesity. This research highlighted the impact of ML in transforming public health strategies by strengthening the awareness of important signs, empowering early interventions, and providing Research-based solutions to obesity.

3. Methodology

This study uses the J48 and SVM algorithms to develop predictive models for obesity based on dietary intake data. J48 and SVM, decision tree-based methods, are effective for handling categorical and continuous data. Accuracy, precision, recall, and F1-score will be used to evaluate model performance and assess their capacity to predict obesity outcomes. The dataset comprises elements such as age and daily meal count for predicting obesity based on dietary intake. These metrics will provide an overall evaluation of the model's success.

3.1. Research Design

This study uses a predictive research design to predict human obesity based on dietary intake by using the J48 and SVM algorithms. The primary focus is on developing predictive models that classify the level of obesity, leveraging the machine learning algorithms in order to analyze and predict the relationship between dietary factors and obesity.

3.2. Research Instrument

In this study, R was used as a software tool to set up data in terms of models and for assessment. The data was prepared in R with proper tools such as transformation, normalization, and encoding. Training and evaluation of models were also carried out in R, making use of functions in R for executing several machine learning algorithms, e.g. J48 or SVM and evaluation methods for model assessment and hyperparameter tuning. R also has advanced visualization features that provide extra understanding and presentation of results derived from the models.

3.3 Data Collection

The data utilized in this study was sourced from a secondary source: Koklu, N., & Sulak, S.A. (2024). The dataset has 15 attributes, which are the following; Sex, Age, Height, Obese Families, Fast Food Intake, Vegetable Intake, Daily Meal Count, Snacking, Smoking, Liquid Intake Daily, Calorie Calculation, Physical Exercise, Hours Using Technology, Transportation Type Used, and Class. The data had been improved by including labeled outcomes for supervised learning, and sampling techniques were utilized to guarantee a varied representation of eating habits and physical activity levels. The target label for the dataset is Class.

Class	
1	Underweight
2	Normal
3	Overweight
4	Obese

Table 1. Target Label of the Dataset

3.4. Data Preprocessing

The dataset has undergone several preparation procedures that improve data quality and compatibility with machine learning methods. Initially, data is cleaned to address empty values and correct conflicts. The dataset involves normalization, standardizing dietary variables to a uniform scale for comparison. Feature selection methods are utilized to determine dietary factors most significantly associated with obesity, improving model precision.

3.5. Data Modeling and Analysis

This study utilizes two classification algorithms—J48 and SVM—to predict obesity based on dietary intake and other factors. J48: This decision tree algorithm understands important nutritional attributes and produces a categorization tree, offering a clear and understandable context. SVM: A support vector machine (SVM) is a type of supervised learning algorithm used in machine learning to solve classification and regression tasks. SVMs are particularly good at solving binary classification problems, which require classifying the elements of a data set into two groups.

3.6. Evaluation Metrics

The effectiveness of each model is measured using four key evaluation metrics:

a. Accuracy: The percentage of correct predictions among all predictions. It measures the overall correctness of the model by calculating the proportion of correct predictions (both positive and negative) out of all predictions.

$$Accuracy = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Where:
TP = True Positives (correctly predicted positive cases)
TN = True Negatives (correctly predicted negative cases)
FP = False Positives (incorrectly predicted as positive)
FN = False Negatives (incorrectly predicted as unfavorable)

b. Precision: The proportion of proper positive classifications among all optimistic predictions. Indicates how many of the predicted positive cases were correct.

$$Precision = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Where:
TP = True Positives
FP = False Positives

c. Recall: The ratio of correctly identified positive cases to all positive cases. Measures how well the model identifies actual positive cases.

$$Recall = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Where:
TP = True Positives
FN = False Negatives

d. F1-Score: The harmonic means of precision and recall balances the two metrics, offering a measure of model performance that is particularly useful when dealing with class imbalances.

$$F1 - Score = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

3.7. Interpretation and Visualization

The model results are analyzed to determine the dietary factors most impacting obesity predictions. To demonstrate the connections between nutritional variables and obesity risk, visualization methods-along with decision trees, decision boundary (for J48 and SVM) and feature importance plots-were developed. Visualization techniques like the ggplot2 in R. These techniques help interpret the dietary inputs on obesity and key findings effectively. This process uses a disciplined approach to ensure data quality, model reliability, and clarity while carrying out the research processes. These methods and tools have deliberately been chosen to enhance the predictive accuracy and usability of obesity prediction models based on their nutritionally referred dietary intake.

3.8. Cross Validation

This study applies cross-validation, a method used for evaluating machine learning models by training them on subsets of the available data and testing them on the complementary subset. Cross-validation helps detect overfitting, which occurs when a model fails to generalize patterns to new, unseen data. In 10-fold cross-validation, the dataset is randomly divided into ten equal parts. During each iteration, 10% of the data is collected for testing, while 90% is used for training. This process is repeated ten times, each subset serving as the test set once. The mean accuracy of the algorithm is calculated across all iterations.

4. Results and Discussion

This section presents the outcomes of the analysis and provides a discussion of the results aligning with the objectives of the study.

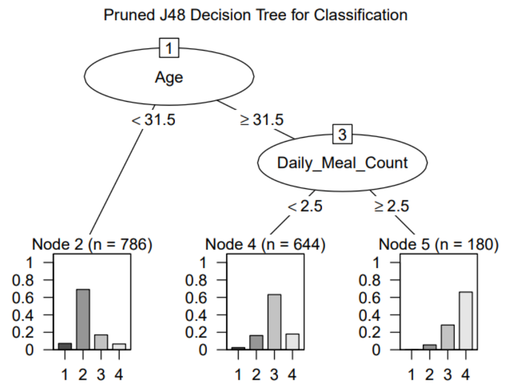

Figure 1. J48 Decision Tree

Figure 1 presents the J48 decision tree categorizes data based on two principal attributes: Age and Daily Meal Count. The first split in the decision tree is based on Age (31.5 years). Individuals under 31.5 years are directed to Node 2, while those aged 31.5 and above are classified further based on their Daily Meal Count. Node 2, which contains 786 instances, mostly classified as '*Normal*'. This indicates that being under 31.5 years old is a strong determinant for belonging to *Normal*. For individuals aged 31.5 years and above, Node 3 splits them further based on their Daily Meal Count, resulting in two subcategories. Node 4 represents individuals who consume fewer than 2.5 meals per day and includes 644 instances, and are mostly classified as '*Overweight*'. The bar graph for this node shows a dominant class, suggesting that having fewer than 2.5 meals per day strongly aligns to *Overweight* Node 5 contains individuals who consume 2.5 meals or more per day, with 180 instances. The bar graph here also indicates a predominant class, demonstrating that eating 2.5 or more meals per day are mostly classified as *Obese*.

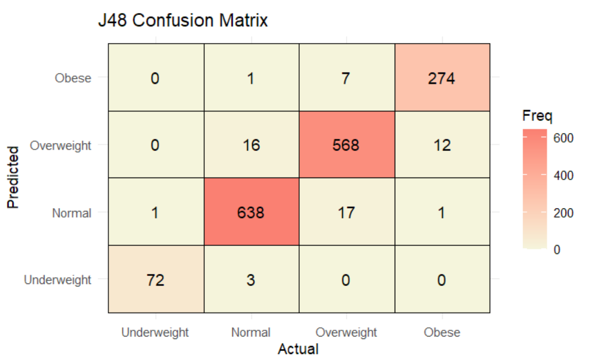

Figure 2. J48 Confusion Matrix Heatmap

Figure 2 illustrates the confusion matrix for the J48 model's predictions, where the diagonal cells represent correct predictions, and the off-diagonal cells represent misclassifications. For *Underweight*, the model correctly predicted 72 instances. For *Normal*, the model accurately predicted 638 instances. For *Overweight*, the model correctly predicted 568 instances. For *Obese*, the model correctly predicted 274 instances.

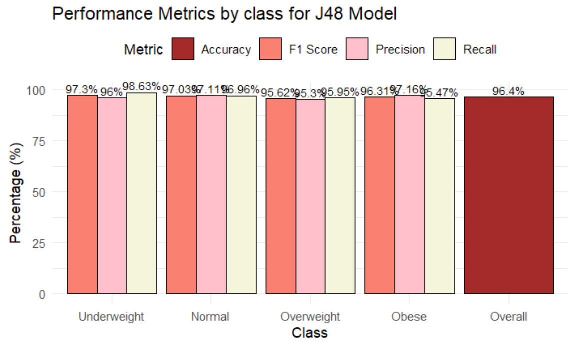

Figure 3. J48 Performance Metrics by Class

Figure 3 presents the J48 performance metrics, which shows an overall accuracy of 96.4%. *Underweight* has the highest precision at 98.63% and the highest recall at 96%. *Normal* follows closely with a precision of 96.96% and a recall of 97.03%. *Overweight* has a precision of 95.95% and a recall of 95.62%, while Obese achieves a precision of 97.16% and a recall of 95.47%. The F1 scores are relatively consistent across classes, with the highest at 97.11% for *Normal* and the lowest at 95.38% for *Overweight*.

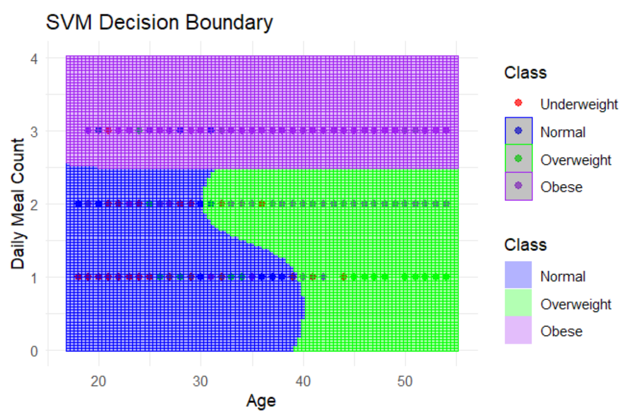

Figure 4. SVM Decision Boundary

Figure 4 visualizes the decision boundaries of an SVM, where the X-axis represents the Age of individuals and the Y-axis represents their Daily Meal Count. Different colored regions in the plot correspond to the predicted classes by the SVM, with purple indicating *Overweight* and green indicating *Obese*. The points scattered across the plot represent the actual data points, with their colors and shapes denoting the true class of each observation. The curvy lines separating the colored regions are the decision boundaries calculated by the SVM to best differentiate between the classes based on the features of Age and Daily Meal Count. The plot effectively shows how the SVM classifier makes predictions by assigning new data points to the regions defined by these boundaries.

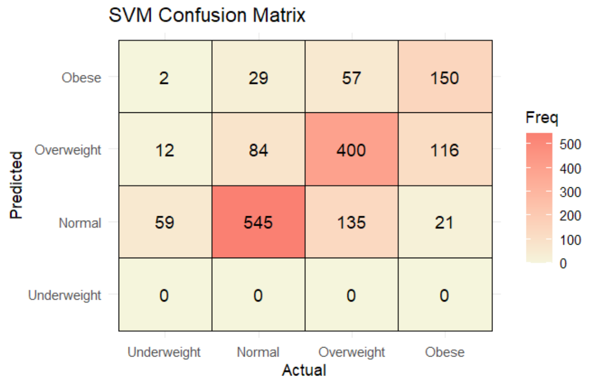

Figure 5. SVM Confusion Matrix Heatmap

The confusion matrix shows the SVM model's performance, with correct predictions along the diagonal (0, 545, 400, 150). *Normal* and *Overweight* classes had the best accuracy, with 545 and 400 correct classifications, respectively. Misclassifications for *Underweight* were significant, with most instances predicted as *Obese* (150) or *Overweight* (57). For Obese, 150 were correctly predicted, but 57 were misclassified as *Overweight*. *Normal* showed strong performance despite 59 misclassified as *Underweight* and 135 as *Overweight*. Misclassifications reflect overlap between adjacent categories, highlighting areas for improvement.

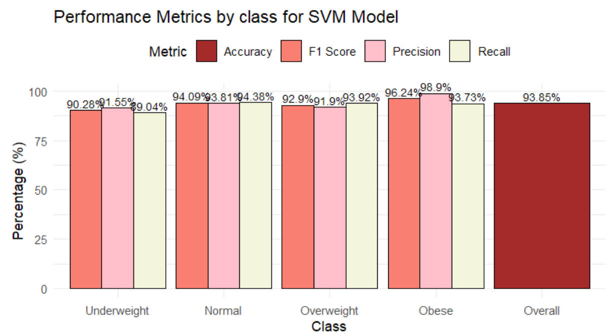

Figure 6. SVM Performance Metrics by Class

Figure 6 presents the SVM Performance Metrics, achieving an overall accuracy of 93.98%. *Underweight* Class has an F1 score of 90.28%, with a precision of 91.55% and a recall of 89.04%. *Normal* Class has an F1 score of 94.01%, a precision of 93.8%, and a recall of 94.22%. *Overweight* Class achieves an F1 score of 93.05%, with a precision of 92.21% and a recall of 93.92%. *Obese* Class has the highest F1 score at 96.8%, with a precision of 98.91% and a recall of 94.77%.

5.Conclusion

The results of the J48 and SVM classifiers demonstrate the strengths and weaknesses of each model in predicting obesity based on dietary intake. The J48 decision tree achieves a high overall accuracy of 96.4%, effectively identifying key factors such as age and meal frequency. The model shows strong performance, but there are notable misclassifications between different Classes. Precision and recall values for all classes indicate that J48 performed well overall, with *Underweight* having the highest precision and recall, suggesting it is the most reliably predicted.

The SVM model, with an accuracy of 93.98%, provides strong decision boundaries and achieves high F1 scores. However, it struggles with misclassifications, particularly for *Underweight* and *Normal*. Despite these challenges, the SVM classifier performs well in predicting *Overweight* Class, which shows the highest correct classification rate. The performance metrics reveal areas for improvement, especially in handling misclassifications between closely related classes.

In summary, while both models show promising results, J48 outperforms SVM in terms of overall accuracy and consistency. The SVM model excels in predicting *Obese* Class but requires further refinement to handle misclassifications in other classes. Therefore, J48 is the more reliable model for predicting obesity based on dietary intake.

Acknowledgement

A traveler cannot reach their destination without a studious reflection of those who endeavored to make their journey complete. The researchers take this opportunity to thank all those people for their help and support.

Reymon Santianez, DIT, for his unwavering support, guidance, and encouragement throughout this research journey. His expertise and insights were invaluable to the completion of this study.

To the researchers' parents and guardians, sisters and brothers, who served to be the shield and comfort throughout their entire academic career, without them the researchers would not have made it this far. The hard work and dedication set by their parents throughout life continue to inspire the researchers to aim for the top. Their constant enthusiasm and support will be treasured forever.

To the researchers’ classmates and friends who have been a rescue every time the researchers needed help, their love, wishes, and prayers will be treasured forever.

References

[1] O. J. O. S. Olalekan Akinola, "Accuracies and Training Times of Data Mining Classification Algorithms: An Empirical Comparative Study," *Journal of Software Engineering and Applications,* vol. 8, no. 9, pp. 470-477, 2012.

[2] A. H. M. S. U. Marjia Sultana, "Analysis of data mining techniques for heart disease prediction," *2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT,* pp. 1-5, 2016.

[3] "A Survey of Data Mining Techniques on Medical Data for Finding Locally Frequent Diseases," *" Int. J. Adv. Res. Comput. Sci. Softw. Eng.,* vol. 3, no. 8, pp. 2277-128, 2013.

[4] A. R. Salvador Camacho, "Is the calorie concept a real solution to the obesity epidemic?," *Glob. Health Action,* vol. 10, no. 1, 2017.

[5] Z. A. A. N. M. T. B. S. S. A. H. A. J. M. L. S. A. K. H. A. U. S. N. S. Sadaf Ibrahim, "Overweight and obesity prevalence and predictors in people living in Karachi," *J. Pharmaceut. Res. Int.,* vol. 33, no. 31, p. 194–202, 2021.

[6] M. M. K. W. P. M. D. Ellen P. Williams, " Overweight and Obesity: Prevalence, Consequences, and Causes of a Growing Public Health Problem," *National Center for Biotechnology Information ,* vol. 4, no. 3, pp. 363-70, 2015.

[7] A. H. A. G. M. S. M. S. C. N. T. K. Syahrul Sazliyana Shaharir, "Steroid-induced diabetes mellitus in systemic lupus erythematosus patients: analysis from a Malaysian multi-ethnic lupus cohort," *Int. J. Rheum. Dis,* vol. 18, no. 5, p. 541–547, 2015.

[8] X. H. C. Y. J. L. K. H. P. L. Y. W. Q. W. Z. W. R. G. H. B. X. C. Lihua Hu, "Prevalence of overweight, obesity, abdominal obesity and obesity-related risk factors in southern China," *PloS One,* vol. 12, no. 9, 2017.

[9] ,. I. K. A. R. G. R. H. P. D. Natharnia Young, "Obesity: how much does it matter for female pelvic organ prolapse?," *Int. Urogynecol. J.,* vol. 29, no. 8, p. 1129–1134, 2018.

[10] Y. B. T. C. M. C. C. C. N. M. D. N. E. D. D. F. L. G. C. G. B. K. M. L.-S. R. S. J. S. P. S. N. tephan C. Bischoff, "towards a multidisciplinary approach to understand and manage obesity and related diseases," *Clin. Nutr.,* vol. 36, no. 4, p. 917–938, 2017.

[11] P. A. M. T. S. C. R. V. M. S. N. B. Carl J. Lavie, "Obesity and cardiovascular diseases: implications regarding fitness, fatness, and severity in the obesity paradox," *J. Am. Coll. Cardiol.,* vol. 63, no. 14, p. 1345–1354, 2014.

[12] C. B. George A. Bray, "Handbook of obesity," *Clinical Applications, Third Edition,* p. 680, 2003.

[13] C. M. C. L. M. M. T. C. F. K. Ogden CL, "Prevalence of overweight and obesity in the United States," *JAMA,* vol. 295, no. 13, pp. 1549-55, 2006.

[14] J. S. Rennie KL, "Prevalence of obesity in Great Britain.," *Obes Rev,* vol. 6, no. 1, pp. 11-20, 2005.

[15] W. T. Z. P. D. D. O. N. S. J. e. a. Cameron AJ, "Overweight and obesity in Australia: the 1999-2000 Australian Diabetes, Obesity and Lifestyle Study (AusDiab)," *Med J Aust,* vol. 178, no. 9, pp. 427-32, 2003.

[16] Y. VD, "Prevalence of obesity in Turkey," *Obes Rev,* vol. 6, no. 1, pp. 9-10, 2005.

[17] R. A. K. M. E. M. K. N. Mohammadpour-Ahranjania B, "Prevalence of overweight and obesity in adolescent Tehrani students, 2000-2001: an epidemic health problem.," *Public Health Nutr,* vol. 7, no. 5, pp. 645-8, 2004.

[18] U. R. V. F. A. C. Kain J, "Trends in overweight and obesity prevalence in Chilean children: comparison of three definitions.," *Eur J Clin Nutr,* vol. 56, no. 3, pp. 200-4, 2002.

[19] V. F. K. J. U. R. Albala C, "Nutrition transition in Chile: determinants and consequences.," *Public Health Nutr,* vol. 5, no. 1, pp. 123-8, 2002.

[20] R. S., "Obesity: the disease of the twenty-first century.," *Int J Obes Relat Metab Disord,* vol. 26, no. 4, pp. 2-4, 2002.

[21] A. L. C. O. N. W. M. Simmonds, "Predicting adult obesity from childhood obesity: a systematic review and meta-analysis," *Obes. Rev. ,* vol. 17, no. 2, p. 2016, 95–107.

[22] L. M. P. E. C. T. Y. V. D. S. R. P. S. M. A. W. S. L. D. A. V. D. G. K. M. M. G. N. S. J. B. J. W. B. Abbasi Ali, "Prediction models for risk of developing type 2 diabetes: systematic literature search and independent external validation study," *BMJ 345,* p. e5900–e5900, 2012.

[23] K. Q. Y. L. D. Y. Y. J. H. T. Quan Zou, "Predicting diabetes mellitus with machine learning techniques," *Front. Genet.,* vol. 9, p. 515, 2018.

[24] J. E. E. Ziad Obermeyer, "Predicting the future - big data, machine learning, and clinical medicine," *N. Engl. J. Med. ,* vol. 375, no. 13, p. 1216–1219, 2016.

[25] H. T. Balbir Singh, "A machine learning approach for predicting weight gain risks in young adults, in: Conference Proceedings of 2019 10th International Conference on Dependable Systems, Services and Technologies," *DESSERT,* vol. 2019, p. 231–234, 2019.

[26] Z. A. J. A. W. E. S. J. L. G. Animesh Acharjee, "Integration of metabolomics, lipidomics and clinical data using a machine learning method," *BMC Bioinf. ,* vol. 17, no. 15, p. 440, 2016.

[27] S. M. A. C. S. D. Tamara M. Dugan, "Machine learning techniques for prediction of early childhood obesity," *Appl. Clin. Inf. ,* vol. 6, no. 3, pp. 506-520, 2015.

[28] J. K. S. G. J. S. G. L. Katherine Ellis, " Hip and wrist accelerometer algorithms for free-living behavior classification," *Med. Sci. Sports Exerc. ,* vol. 48, no. 5, pp. 933-940, 2016.

[29] E. P. A. A. C. L. R. N. D. O. S. A. d. S. A. L. F. C. M. I. F. S. E. W. E. A. L. H. G. a. L. Andreas Triantafyllidis, "Computerized decision support and machine learning applications for the prevention and treatment of childhood obesity: a systematic review of the literature,," *Artif. Intell. Med,* vol. 104, p. 101844, 2020.

[30] S. a. D. A. S. Arielle, "Machine learning for the classification of obesity from dietary and physical activity patterns, in: Smart Innovation, Systems and Technologies," *Springer Science and Business Media Deutschland GmbH,* vol. 93, pp. 77-79, 2018.

[31] ´. S. L. K. H. ´. E. D. M. a. R. Felso, "Relationship between sleep duration and childhood obesity: systematic review including the potential underlying mechanisms," *Nutr. Metabol. Cardiovasc. Dis.,* vol. 27, no. 9, p. 751–761, 2017.

[32] Y. L. O. M. E. I. Akira Ishida, "Factors affecting adult overweight and obesity in urban China," *Pertanika J. Soc. Sci. Human,* vol. 28, no. 1, p. 503–513, 2020.

[33] M. S. A. G. A. F. Helen Cheng, "Biomedical, psychological, environmental and behavioural factors associated with adult obesity in a nationally representative sample," *J. Publ. Health ,* vol. 42, no. 3, pp. 570-578, 2020.

[34] J. N. T. A. T. S. J. R. N. P. J. H.-S. Simeon-Pierre Choukem, "Overweight/obesity and associated cardiovascular risk factors in sub-Saharan African children and adolescents: a scoping review," *Int. J. Pediatr. Endocrinol.,* vol. 1, no. 2020, p. 6, 2020.

[35] G.-L. R. R. N. C. A. M. ˜. C. G.-L. Amelia Ma, "Obesity and asthma: a missing link," *Int. J. Mol. Sci. ,* vol. 18, no. 7, p. 1490, 2017.

[36] M. S. G. B. N. P. Balaji Meriga, "Phytocompounds as potential agents to treat obesity-cardiovascular ailments, Cardiovasc.," *Hematol. Agents Med. Chem. ,* vol. 15, no. 2, p. 104–120, 2018.

[37] R. R. K. M. V. K. V. N. R. K. N. M. T. G. H. A. Abdul Gafoor Abdul Mubarak, "Patient and Procedure Selection for Bariatric and Metabolic Surgery in Malaysia- the Malaysian Consensus.," *Technical Report,* vol. 76, no. 2, 2021.

[38] F. J. C. K. C. L. L. Y. C. L. Cheong Kim, "Predicting factors affecting adolescent obesity using general bayesian network and what-if analysis," *Int. J. Environ. Res. Publ. Health ,* vol. 16, no. 23, p. 4684, 2019.

[39] A. A. Hoda C. Kadouh, " Current paradigms in the etiology of obesity," *Tech. Gastrointest. Endosc.,* vol. 19, no. 1, p. 2–11, 2017.

[40] M. A. R. G. L. H. A. R. M. H. G. M. C. A. A. A. M. J. A. B. A. A. M. A. e. a. Frank Qian, "Mendelian randomisation study of height and body mass index as modifiers of ovarian cancer risk in 22,588 brca1 and brca2 mutation carriers," *Br. J. Canc. ,* vol. 121, no. 2, p. 180–192, 2019.

[41] L. J. V. M. M. L. C. K. H. Sartorius Benn, "Determinants of obesity and associated population attributability, South Africa: empirical evidence from a national panel survey, 2008-2012," *PloS One,* vol. 10, no. 6, 2020.

[42] S. W. X. S. Y. Sun, "Estimating neighbourhood-level prevalence of adult obesity by socio-economic, behavioural and built environment factors in New York City," *Publ. Health ,* vol. 186, pp. 57-62, 2020.

[43] A. F. Helen Cheng, "Personality traits, education, physical exercise, and childhood neurological function as independent predictors of adult obesity," *PloS One ,* vol. 8, no. 11, p. e79586, 2013.

[44] I. H. &. F. E. Witten, "Data Mining: Practical Machine Learning Tools and Techniques," *BioMedical Engineering OnLine,* pp. 1-2, 2005.

[45] M. S. &. S. S. Bhatia, "A study on prediction of obesity using J48 algorithm," *International Journal of Advanced Research in Computer Science,* vol. 9, no. 5, pp. 1-5.

[46] M. &. H. M. S. Dehghani, "Using J48 decision tree algorithm for predicting obesity in adults," *International Journal of Information Technology and Computer Science,* vol. 11, no. 2, pp. 1-8, 2019.

[47] S. a. D. A. S. Arielle, "Machine learning for the classification of obesity from dietary and physical activity patterns, in: Smart Innovation, Systems," *Springer Science and Business Media Deutschland GmbH,* vol. 93, pp. 77-97, 2018.

[48] A. A. A. A. &. C. S. Taghiyev, "A hybrid approach based on machine learning to identify the causes of obesity," *Journal of Control Engineering and Applied Informatics,* vol. 22, no. 2, pp. 56-66, 2020.