DOI: http://dx.doi.org/10.26483/ijarcs.v14i5.7027

Volume 14, No. 5, September-October 2023

ISSN No. 0976-5697

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

APPLICATION OF MULTIPLE MACHINE LEARNING TECHNIQUES IN CLASSIFYING OBESITY LEVEL USING MULTIVARIATE DATASET

Jayvee Ryan Banal
Batangas State University
Alangilan, Batangas, Philippines

Kenneth Dynielle Lawas
Laguna State Polytechnics University
San Pablo, Laguna, Philippines

Laurice M. Mariquina
City College of Calamba
Calamba City, Laguna, Philippines

Abstract**—** People's health is important. It must be preserved. There are a variety of issues that might contribute to a person's health, including their lifestyle. One of the most prominent concerns during this pandemic is a person's body weight, which is particularly significant these days. Obesity is one of those body diseases about which it is important to be aware of the reasons. This study intends to create different machine learning models to define what causes obesity, select the most appropriate model that did the best, and discuss how accurate it performed. It can also be used in determining the how the obesity impact in our daily lives. Through the use of different machine learning models such as KNN, Random Forest, Gradient Boosting and Ada Boost, the study be able to obtain the appropriate model. Despite being trained on unbalanced data, the classifiers utilized were able to predict the properties of the presented datasets that Random Forest has an accuracy of 83.6%.

Keywords**—**Obesity, Datasets, Machine Learning, Random Forest, Gradient Boosting, KNN

I. INTRODUCTION

One of the most important aspects of our lives that we must maintain is our health. The immune system of the body is still robust at a young age, but as the years pass, the body becomes weaker and weaker. Furthermore, people's lifestyles are changing, particularly in terms of food consumption and everyday tasks [1]. People eating habits have changed in modern times as a result of work and being away from home for many hours during the day. As a result, people are pushed to eat unhealthy foods that are high in fats and carbohydrates and are consumed quickly. Chemical processes are used in many foods nowadays, and many people do not exercise their bodies. As a result, a variety of disorders, such as obesity, have arisen. Having an atypical weight, such as being underweight or overweight, is one of the challenges people face these days. People developed an obesity problem because of the practices. According to the World Health Organization, the current health risk is abnormal or accumulated excessive fats, which can lead to overweight and obesity [2].

Obesity is one of the biggest health challenges in many nations affected by the pandemic. Throughout the lifecycle, it enhanced the mortality and morbidity of an independent risk factor [3]. Obesity in children and adolescents is one of the characteristics that tends to track and stabilize throughout adulthood [4]. As a result, childhood obesity must be addressed immediately in order to reduce the number of obesity-related comorbidities [5]. This implies that obesity can be risk in the health so people should know the causes and effect of having unbalanced lifestyle or having a weight abnormality.

BMI (Body Mass Index) is used to determine a person's weight status, which includes Insufficient weight, Normal weight, Obese, and Overweight [2][6]. When utilizing the BMI, which is calculated based on a person's height and weight, most adults are classified as obese or overweight. There are numerous elements that influence the different weight classes because physiological changes occur as people age. Puberty, age, and development rate in the fat loss and deposition are all factors to consider. As a result, incorporating overweight and obesity is extremely challenging for people of all ages.

The unhealthy person or being an obesity have many risks in the health such as in cardiovascular diseases, cancer and diabetes which can lead to a person in having critical condition [7]. So, many technologies or studies have been conducted to obtain on what is the treatment for the having weight problems.

In comparison to parametric methodologies, machine learning has the capacity to give efficient and effective classification performance of remotely sensed imagery. It has the ability to handle data with a high degree of dimensionality and to map classes with complex characteristics. Machine learning methods assist data in improving the performance of variable classification. It was also demonstrated that in comparison to traditional parametric classifiers, machine learning has a higher accuracy result when dealing with complex data with a high dimensional feature [8][9]. However, machine learning methods are not frequently employed these days due to several ambiguities about how to integrate and apply machine learning techniques efficiently. Some software in remote sensing is lacking in availability and optimization, which is one of the disadvantages of machine learning. Nonetheless, machine learning is still active in the field of remote sensing in terms of developing new algorithms that provide improved and enhanced remote sensing data classification capability. As a result, existing algorithms provide a powerful set of tools for extracting information from remotely sensed data and should be fully utilized [10][11][12]. With this, the researchers proved that machine learning could help to utilized datasets by applying the algorithms or methods.

The application of machine learning is one of the many technologies that have been employed in determining the causes and treatments of having weight problems. Machine learning has aided transformation in a variety of fields, including health. It delivers an accurate result in demonstrating the classification difficulties' capabilities. It will supply a balanced dataset wherein reality on how to obtain the reasons of weight problems notably obesity via the usage of machine learning.

Obesity prevention is critical for the general public. It aids people in improving and maintaining their health. It also helps to avoid diseases like cancer. Obesity can also result in mortality. As a result, it is critical for individuals to understand the causes of obesity and how to prevent it in order to improve people's health lifestyles.

The researchers' goal is to determine the causes of obesity other than BMI. Furthermore, how will machine learning algorithms classify the causes of obesity and how accurate will it compute for unbalanced datasets on the elements influencing the probability of having weight problems. By resolving the issue, researchers will be able to discover why people become obese and anticipate whether they will develop weight problems.

The model utilized the following classifiers for training the model: AdaBoost, KNN, Gradient Boosting and Random Forest. The mentioned classifiers have proven their accuracy with different classification problems in the aspect of classification of the obesity.

This study is structured as discusses the related studies regarding the obesity and how machine learning techniques have been effectively applied to detect or classify the causes of obesity, chapter 3 evaluates the data and the machine learning techniques used for the classification of the obesity, chapter 4 examines the result of the classifiers used, and chapter 5 states the conclusion of the study.

II. RELATED WORKS

Obesity is the accumulation of fat in a person's body. The BMI, or Body Mass Index, is currently the most widely used method for determining fat mass. Because of its ease of use, it is widely utilized, although it does not guarantee that all measurements of body fat mass are accurate. As a result, several studies have been undertaken to evaluate whether a person's weight status is healthy. The study analyzes the performance of various approaches for identifying obesity classification. When compared to other techniques, the researchers discovered that the Fuzzy Rule-Based System (FRBS) is more accurate. It also means that FRBS's performance is precise [13].

The study points out that several obesity predictions models or intervention tools have been developed in recent years. Despite the numerous advancements in obesity prediction models, obesity remains a major problem for both children and adults today, particularly in the context of the pandemic [2]. As a result, people should be aware of these health issues because they can pose a major threat to one's health.

A study has shown how IoT (Internet of Things) may be used in the evaluation and prediction of BMI or obesity. Obesity can cause a variety of ailments, including cardiovascular disease. As a result, people want a simple and quick tool to monitor and control their BMI, particularly for the prediction of obesity. As a result, the study's researchers developed a device that can be powered by the Internet of Things. The data analytics methods were implemented in Arduino. This means that he or she will be able to quickly determine whether or not they need to control or monitor their health. Future studies will benefit from the findings, which show that by using certain techniques, people can reduce their risk of developing weight problems [14].

According to the study of Cui et al. (2021), through the use of the different machine learning algorithms the researchers predict the causes of the obesity. The machine learning that has been applied are K Nearest Neighbor, Random Forest and Decision Tree. Among of these three techniques the Decision Tree shows more accurate performance which can predict what are the causes of the obesity of the people. In addition, it presented how different aspects in the people lifestyle could affect on having weight problems.

Based on [15], the study demonstrates a framework that can aid in the reduction of obesity through the application of machine learning and the Internet of Things (Internet of Things). One of the machine learning approaches employed by the researchers in the framework model is the KNN and

Random Forest algorithm. On the study of Cui et al. (2021), KNN, or K-Nearest Neighbor, is a method for determining the similarities between the projected projection and the data points in the features. It signifies that the preparation set will be based on the points near coordinates. The Random Forest, on the other hand, examines the aspects of decision trees on diverse subsets that can increase the dataset's accuracy. It also classifies and combines problems into a single cycle.

One of the machine learning models that can assist in the use of datasets is the Gradient Boost Algorithm. It's also used to remove bias, noise, and variance from the data, making the prediction model more accurate. This method enables future models to learn from earlier models' errors. It's also utilized to classify data sets and run regressions on them in order to get the best results from the datasets we'll be training with [16].

Among the classifier ensemble models, AdaBoost is one of the best. In AdaBoost, heuristics that correctly categorize a webpage are given more weight, while heuristics that mislabel it are given less. AdaBoost is a supervised learning algorithm that was created by AdaBoost. The most significant disadvantage of AdaBoost has been the overfitting problem, which occurs when a model is extremely complex, such as when there are too many parameters relating to the number of observations, resulting in random errors [17].

III. MATERIALS AND METHODOLOGY

In this section, the research methodology and techniques are presented. The machine learning algorithms also discussed.

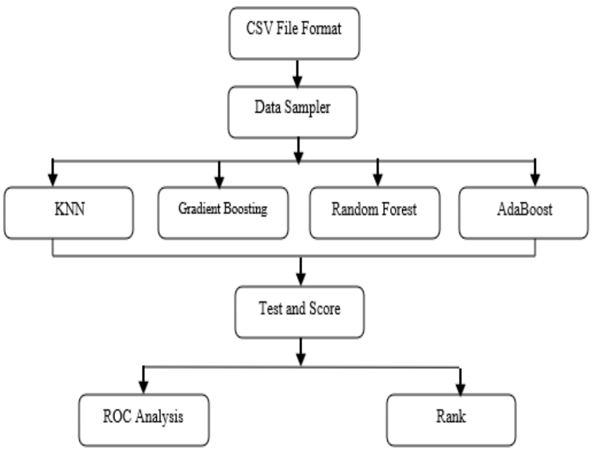


Fig. 1 Model Architecture

The model architecture of datasets in the Orange Data Mining is shown in Figure 1. It includes the datasets file, models, visualization, and evaluation of the data in the datasets, with a focus on obesity datasets. As a result, it was able to achieve good accuracy while simulating data.

A. The Dataset

The dataset used in this paper is from UCI ML repository which consists of 2111 rows and 17 columns that indicates the attributes of the data [6] [13] [18]. The data for the calculation of obesity levels in adults aged 14 to 61 from Mexico, Peru, and Colombia, with a variety of eating habits and physical conditions, was acquired utilizing an online platform with a survey [19].

TABLE I. LIST OF FEATURES, ITS ROLE, DESCRIPTION, AND TYPE

Attribute	**Role and Description**
Age	Feature-Age
Gender	Feature-Gender
Age	Feature-Age (yrs)
Height	Skip-Height (in m)
Weight	Skip-Weight (in kgs)
SMOKE	Feature-Smoking Habit
CALC	Feature-Consumption of alcohol
MTRANS	Feature-Transportation used
Family_history_with_overweight	Feature-Family member suffered from obesity
FAVC	Feature-Frequent consumption of high calorie food
FCVC	Feature-Frequency of consumption vegetables
NCP	Feature-Number of main meals
CAEC	Feature-Consumption of food between meals
CH2O	Feature-Consumption of water daily
SCC	Feature-Calorie’s consumption monitoring
FAF	Feature-Physical activity frequency
TUE	Feature-Time under technology devices
Nobeyesdad	Target-Obesity Levels

All the features in dataset have essential role in processing the model techniques. The height and weight are not included in the table above since predicting people's obesity levels should be based on their physical conditions and eating habits, and height and weight have a significant impact on the BMI index's labeling data [19]. Therefore, the datasets will have 15 features and 1 target that will process on the machine learning techniques to determine the obesity level.

B. Data Sampling

The qualities were chosen based on their function, but the weight and height were excluded. In addition, a subset of data instances comprising 70% of the training dataset and 30% of the testing dataset was chosen. To have an efficient result for the completability which can simulate the complexity of random data, replicable (deterministic) sampling was chosen [20].

C. Data Visualization

To fully understand the obesity dataset, for visualizing the data the study used the Scatter Plot for some of the attributes of the dataset.

Scatter Plot

For the obesity dataset, the scatter plot visualization tool creates a two-dimensional plot. The researchers used an intelligent data visualization technique to generate useful projections, which resulted in the top-scoring pair of variables, Vocabulary richness W/C and Access, in terms of average classification accuracy.

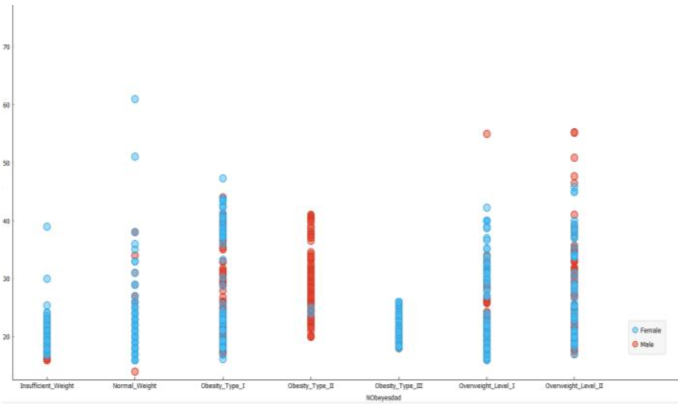


Fig. 2 Scatter Plot for Age, Gender and Obesity Level Attributes

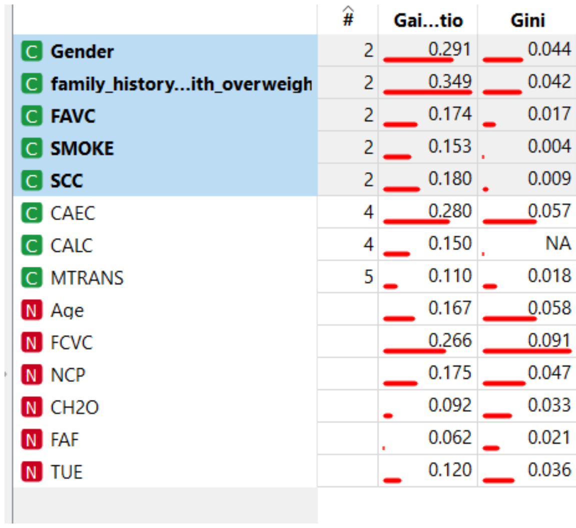
Figure 2 depicts the data in relation to determining which age and gender have a high degree of obesity. According to the graph, the female has a high risk of obesity, which is caused by a variety of factors. This means that at a specific age, a male is less likely than a female to gain weight.

D. Statistical Validation

The tool that used in order to visualize the attributes in the dataset is Rank. It was used to fully understand the given data in the dataset. To visualize the result, the ROC analysis was also used to determine the result of the model.

Rank

In data mining, one of the attribute classifications is rank. It describes the relationship between the scores of the variables and the numeric or discrete target variable. It can also be used to link and rank the attributes of datasets. The qualities will be ranked using the Information Gain Ratio and Gini Decrease computations based on the gathered scores. Moreover, two sorts of patterns are considered: frequent rankings and association rules, which represent relationships between such rankings. Algorithms for mining frequent ranks and frequent closed rankings are proposed and evaluated with synthetic and real data [21]. This implies that Rank can help to analyze the data which deals with the given attributes of the datasets.



	#	Gai...tio	Gini
C Gender	2	0.291	0.044
C family_history...ith_overweigh	2	0.349	0.042
C FAVC	2	0.174	0.017
C SMOKE	2	0.153	0.004
C SCC	2	0.180	0.009
C CAEC	4	0.280	0.057
C CALC	4	0.150	NA
C MTRANS	5	0.110	0.018
N Age		0.167	0.058
N FCVC		0.266	0.091
N NCP		0.175	0.047
N CH2O		0.092	0.033
N FAF		0.062	0.021
N TUE		0.120	0.036

Fig. 3 Rank of the feature-attributes

The figure above shows the order in which the attributes in the datasets were ranked. It depicts the top five factors that contribute to obesity, including gender, family history of obesity, frequent food consumption, calorie consumption tracking, and smoking habits. In addition, the figure also contains if the attributes are categorical or numerical. It is possible to make more accurate predictions in the supplied datasets using these computed scores.

E. Machine Learning Classifiers

AdaBoost

Adaptive Boosting is a meta-algorithm for combining weak learners that adjusts to the 'hardness' of each training sample. It is applicable to both regression and classification problems. AdaBoost removes all occurrences with uncertain target values, continues categorical variables, removes empty columns, and imputes missing values with mean values as part of its preprocessing.

TABLE II. PARAMETERS USED FOR THE ADABOOST CLASSIFIER

Parameters	Value
Number of Estimators	50
Learning Rate	1.00000
Boosting Method	SAMME.R
Regression Loss Function	Linear

There are four hyper-parameters in AdaBoost that should be optimized because they can affect model performance include the base estimator, max depth, number of estimators, and learning rate. For the AdaBoost model, decision trees were used as base estimators, and each decision tree-based model was built with a maximum leaf depth of 4 to avoid overfitting; the number of n estimator trees was set to 50. As a result, decision tree-based weak classifiers helped to generalize (increase) the performance of the strong classifier. To adjust the contribution of each model for a strong classifier, the learning rate was set to one [22]. This implies that through adjusting the value of the estimators, the better performance on the model.

Gradient Boosting

The gradient boosting is one of the uncommon models in the data mining, but it can use to predict datasets. The data instances in the attributes of each feature will reduced when there is an estimation in gain information. This also associated with the value of the training error. It can serve also as filtration technique for the adopted training in which the data instances with large gradients will be kept while the small gradients are randomly discarded [23].

TABLE III. PARAMETERS USED FOR THE GRADIENT BOOSTING CLASSIFIER

Parameters	Value
Number of trees	50
Learning rate	0.300
Limit depth of individual trees:	3
Do not split subsets smaller than:	2
Fraction of training instances:	1.00

Some of the other methods have their own parameters, as seen in the table above. It solves an infinite-dimensional optimization problem sequentially to build a model in the form of linear combinations of decision trees. It has a lower number of trees, making the parameters and output predictors less sensitive while yet maintaining good performance [24].

KNN (K-Nearest Neighbor)

K-Nearest Neighbor, often known as KNN, is a simple machine learning technique or model. It determines which examples in the class are unlabeled. The training and test datasets are compiled from the given data's observations. It also has a variable k that determines how many neighbors to choose when using the KNN algorithm. This reduces the variance of the random mistake, resulting in a random error. In the training dataset, the value of the parameter k will be visible [25][26][27].

TABLE IV. PARAMETERS USED FOR THE KNN CLASSIFIER

Parameters	Value
Number of Neighbors	1
Metric	Euclidean
Weight	Uniform

The non-parametric algorithm in the table above classifies new and similar entries in the training datasets. The Euclidean distance, which defines the similarity of distance between records, is the most popular parameter in this method. KNN estimated density will flow and speed its input based on the flow and speed of the value. The trained data points are represented by blue triangles, which are positioned in the feature space based on their input values [28]. As a result, the procedure can produce an excellent result.

Random Forest

Random Forests is an optimization-based classification and regression strategy. The Random Forest classifier generates numerous decision trees during training and returns a class that is the mode of the individual trees' classification classes. Random Forest classification outperforms all other decision tree algorithms because it uses a forest of classification trees to judge [29]. After being trained with the dataset and classifier parametric feature values, the Random Forest Classification Model produces more trustworthy results. The model's performance is assessed using the confusion matrix derived from the sample and predicted values [6].

TABLE V. PARAMETERS USED FOR THE RANDOM FOREST CLASSIFIER

Parameters	Value
Number of Trees	100
Min. number of subset division	5

The Table V presented the parameters used in classifying the Random Forest. The number of trees used in the simulation aids in simulating the performance of the variables. Thousands of trees must be trained in order to obtain a stable evaluation of the variable importance. To test for stability, train several random forests with a set number of trees and see if the variables' importance rankings fluctuate between the forests. While the minimum number of the subset division, provides also good results but it depends on the improvement of the tuning performance [30].

IV. Results and Discussion

The results of the four proposed machine learning algorithms, AdaBoost, Gradient Boosting, KNN, and Random Forest, were presented in this section. It was used in attributes obtained from a dataset of 2111 rows and 17 columns from the UCI ML repository, which included obesity causes.

Cui et al. (2021) published a study that gave an estimation of obesity levels based on physical circumstances, eating habits, and other aspects that influence obesity level determination. The study's findings suggest that the Random Forest has a high level of accuracy, however the data has been categorized and balanced, with some data being eliminated. In addition, Celik et al. (2021) conducted a comparison of different clustering algorithms for classifying obesity. The researchers processed the machine learning methods using height and weight as features [1]. With this, the SVM provides the researchers with a high level of accuracy. As a result, the researchers' study differed from others in that the dataset was unbalanced, and height and weight were not included in the features because they had a significant impact on assessing obesity status using machine learning methods.

The machine learning classifications had the highest precision of 83.6 percent when it came to predicting the fractions of obesity levels based on specific parameters. The Random Forest has the highest accuracy, F1 score, and recall, with an overall score of 83.6 percent. Gradient Boosting has an overall score of 81.3 percent. Then there's KNN, which has an accuracy of 80.7 percent. Finally, AdaBoost's performance is 73.7 percent, with poor scores across the board, making it the least accurate machine learning technique. Even though the classifiers' accuracy is poor, it still reveals that they have a high overall accuracy.

TABLE VI. MODEL COMPARISON - F1 SCORE, PRECISION, AND RECALL

Classifiers	F1	Precision	Recall
KNN	0.794	0.800	0.807
Random Forest	0.835	0.838	0.836
Gradient Boosting	0.812	0.814	0.813
AdaBoost	0.737	0.737	0.737

The table above compares the F1 score, precision, and recall of four classifiers: KNN, Random Forest, Gradient Boosting, and AdaBoost, all of which achieved a satisfactory result in terms of obesity level classification accuracy. Random Forest has the most precision or accuracy, and the other classifiers, particularly Gradient Boosting, have the potential to achieve high accuracy. Even though, AdaBoost has the lowest accuracy, it nevertheless generated a good result of 73.7 percent. This means that all four classifiers can be used to classify the unbalanced dataset with a high degree of accuracy.

TABLE VII. MODEL COMPARISON - ACCURACY

Classifiers	Accuracy
KNN	80.70%
Random Forest	83.60%
Gradient Boosting	81.40%
AdaBoost	73.70%

Random Forest is the most used classification model for classifying the obesity level, as evidenced by the above statistics. It also has the highest true positive rates and produces the fewest false positives when compared to the other classifiers, making it the reliable classifier in the study.

ROC Analysis

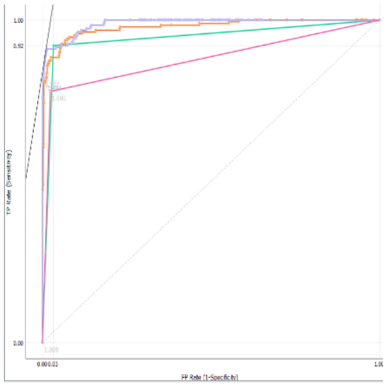


Fig. 3 ROC Analysis of Insufficient Weight

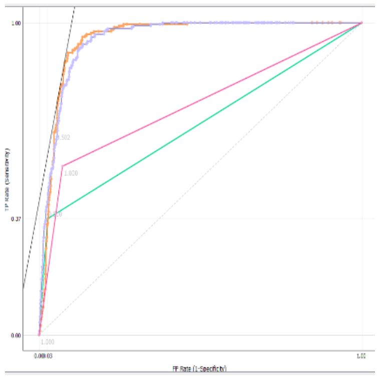


Fig. 4 ROC Analysis of Normal Weight

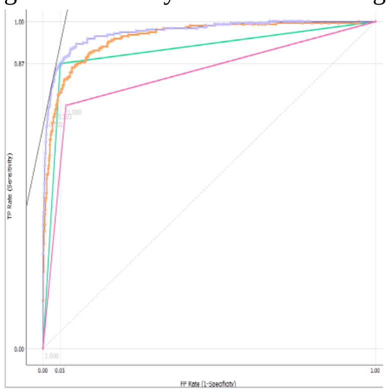


Fig. 5 ROC Analysis of Obesity Type I

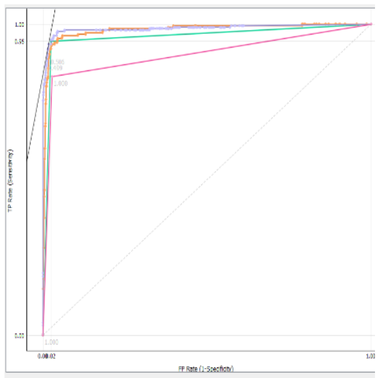


Fig. 6 ROC Analysis of Obesity Type II

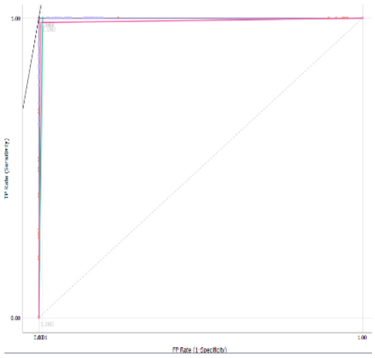


Fig. 7 ROC Analysis of Obesity Type III

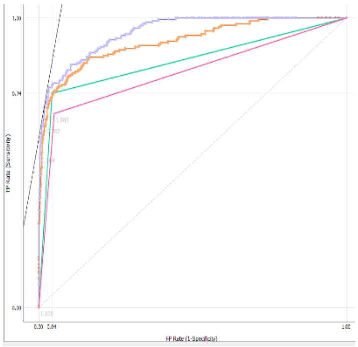


Fig. 8 ROC Analysis of Overweight Level I

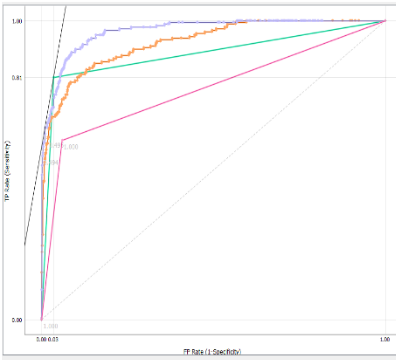
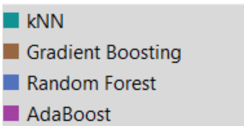


Fig. 9 ROC Analysis of Overweight Level II



As shown in the tables above, ROC Analysis was used to better visualize different targets such as Insufficient Weight, Normal Weight, Obesity Type I, Obesity Type II, Obesity Type III, Overweight Level I, and Overweight Level II. This technique plotted the true positive rate or TP against the false positive rate or FP in the training data models. It enters the model's result or performance based on the test data. This provides as a visual representation of how well the classifiers predicted obesity levels. Random forest, AdaBoost, Gradient Boosting, and KNN all outperformed random forest in terms of accuracy.

V. CONCLUSION

Obesity must be avoided by maintaining control and awareness. Even while people are aware of the problem, many of them do not change lifestyles, particularly in terms of eating and working. With this, many tools have been created to determine a person's weight condition. There are many studies have been developed specifically through the use of machine learning techniques as a result of this. Because machine learning approaches are relevant and have been demonstrated to categorize a specific dataset, the researcher employed this approach to classify the obesity levels of unbalanced datasets using a machine learning technique. The development of approaches for given datasets that make the precision or accuracy of the classifier in the given data more dependable. Random Forest, Gradient Boosting, KNN, and AdaBoost are the machine learning algorithms that were used. The Random Forest model achieved the most accuracy of 83.60 percent, making it more reliable than the other models. Although the Random Forest has the best precision, the other strategies have the potential to produce higher forecasts or computed accuracy because the majority of them do not go below 60%. However, the models need to be modified in order to increase their overall accuracy. The study can be utilized as a reference by researchers in selecting the appropriate classifier when designing solutions for comparable industries in the future.

VI. REFERENCES

[1] Konstantakopoulos, F., Georga, E. I., & Fotiadis, D. I. (2021). 3D Reconstruction and Volume Estimation of Food using Stereo Vision Techniques. In 2021 IEEE 21st International Conference on Bioinformatics and Bioengineering (BIBE) (pp. 1-4). IEEE.

[2] Siddiqui, H., Rattani, A., Woods, N. K., Cure, L., Lewis, R., Twomey, J., ... & Hill, T. (2021). A Survey on Machine and Deep Learning Models for Childhood and Adolescent Obesity. IEEE Access.

[3] Deng, Y., Dolog, P., Gass, J. M., & Denecke, K. (2019). Obesity entity extraction from real outpatient records: When learning-based methods meet small imbalanced medical data sets. In 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS) (pp. 411-416). IEEE.

[4] Ball, C., Novotny, E., Ahn, S. J., Hahn, L., Schmidt, M., Rathbun, S., & Johnsen, K. (2021). Scaling the Virtual Fitness Buddy Ecosystem as a School Based Physical Activity Intervention for Children. IEEE Computer Graphics and Applications.

[5] Pang, X., Forrest, C. B., Lê-Scherban, F., & Masino, A. J. (2019). Understanding early childhood obesity via interpretation of machine learning model predictions. In 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA) (pp. 1438-1443). IEEE.

[6] Cui, T., Chen, Y., Wang, J., Deng, H., & Huang, Y. (2021). Estimation of Obesity Levels Based on Decision Trees. In 2021 International Symposium on Artificial Intelligence and its Application on Media (ISAIAM) (pp. 160-165). IEEE.

[7] Shobana, G., & Bushra, S. N. (2021). Prediction of Cardiovascular Disease using Multiple Machine Learning Platforms. In 2021 International Conference on Innovative Computing, Intelligent Communication and

Smart Electrical Systems (ICSES) (pp. 1-7). IEEE.

[8] Macatangay, L. H., & Hernandez, R. M. (2020). A Deep Learning-Based Prediction and Simulator of Harmful Air Pollutants: A Case from the Philippines. In 2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC) (pp. 381-386). IEEE.

[9] Hernandez, R. M., & Hernandez, A. A. (2019). Classification of Nile Tilapia using convolutional neural network. In 2019 IEEE 9th International Conference on System Engineering and Technology (ICSET) (pp. 126-131). IEEE.

[10] Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. International Journal of Remote Sensing, 39(9), 2784-2817.

[11] Lomboy, K. E. M. R., & Hernandez, R. M. (2021). A comparative performance of breast cancer classification using hyper-parameterized machine learning models. International Journal of Advanced Technology and Engineering Exploration, 8(82), 1080.

[12] Pita, M. S. U., Alon, A. S., Melo, P. M. B., Hernandez, R. M., & Magboo, A. I. (2021). Indoor Human Fall Detection Using Data Augmentation-Assisted Transfer Learning in an Aging Population for Smart Homecare: A Deep Convolutional Neural Network Approach. In 2021 IEEE 19th Student Conference on Research and Development (SCOReD) (pp. 64-69). IEEE.

[13] Ahn, S. H., Wang, C., Shin, G. W., Park, D., Kang, Y. H., Joibi, J. C., & Yun, M. H. (2018). Comparison of Clustering Methods for Obesity Classification. In 2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM) (pp. 1821-1825). IEEE.

[14] Obasi, C., Ndu, I., & Iloanusi, O. (2020). A Framework for Internet of Things-Based Body Mass Index Estimation and Obesity Prediction. In 2020 International Conference on e-Health and Bioengineering (EHB) (pp. 1-4). IEEE.

[15] Garg, S., & Pundir, P. (2021). MOFit: A Framework to reduce Obesity using Machine learning and IoT. In 2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO) (pp. 1733-1740). IEEE.

[16] Felix, A. Y., & Sasipraba, T. (2019). Flood detection using gradient boost machine learning approach. In 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE) (pp. 779-783). IEEE.

[17] Kazi, R. (2016). Evaluation of Machine Learning Classifiers for Phishing Detection.

[18] Celik, Y., Guney, S., & Dengiz, B. (2021). Obesity Level Estimation based on Machine Learning Methods and Artificial Neural Networks. In 2021 44th International Conference on Telecommunications and Signal Processing (TSP) (pp. 329-332). IEEE.

[19] Palechor, F. M., & de la Hoz Manotas, A. (2019). Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from Colombia, Peru and Mexico. Data in brief, 25, 104344.

[20] Pimentel-Alarcón, D. L., Boston, N., & Nowak, R. D. (2016). A characterization of deterministic sampling patterns for low-rank matrix completion. IEEE Journal of Selected Topics in Signal Processing, 10(4), 623-636.

[21] Henzgen, S., & Hüllermeier, E. (2019). Mining rank data. ACM Transactions on Knowledge Discovery from Data (TKDD), 13(6), 1-36.

[22] Kim, D. H., & Hong, S. J. (2021). Use of Plasma Information in Machine-Learning-Based Fault Detection and Classification for Advanced Equipment Control. IEEE Transactions on Semiconductor Manufacturing, 34(3), 408-419.

[23] Kumar, D., Abhinav, R., & Pindoriya, N. (2020). An ensemble model for short-term wind power forecasting using deep learning and gradient boosting algorithms. In 2020 21st National Power Systems Conference (NPSC) (pp. 1-6). IEEE.

[24] Biau, G., Cadre, B., & Rouvìère, L. (2019). Accelerated gradient boosting. Machine Learning, 108(6), 971-99

[25] Bu, A., & Wang, L. (2016). Research on the Rule of Acupuncture and Moxibustion for Treatment of Obesity Based on Data Mining. In 2016 International Conference on Smart City and Systems Engineering (ICSCSE) (pp. 602-605). IEEE.

[26] Zhang, Z. (2016). Introduction to machine learning: k-nearest neighbors. Annals of translational medicine, 4(11).

[27] Mendoza, A. M., & Hernandez, R. M. (2021, October). Application of Data Mining Techniques in Diagnosing Various Thyroid Ailments: A Review. In 2021 13th International Conference on Information & Communication Technology and System (ICTS) (pp. 207-212). IEEE.

[28] Raj, J., Bahuleyan, H., & Vanajakshi, L. D. (2016). Application of data mining techniques for traffic density estimation and prediction. Transportation Research Procedia, 17, 321-330.

[29] Muppavarapu, V., Rajendran, A., & Vasudevan, S. K. (2018). Phishing detection using RDF and random forests. Int. Arab J. Inf. Technol., 15(5), 817-824.

[30] Probst, P., Wright, M. N., & Boulesteix, A. L. (2019). Hyperparameters and tuning strategies for random forest. Wiley Interdisciplinary Reviews: data mining and knowledge discovery, 9(3), e1301.