Received: 30 May 2023 | Revised: 1 August 2023 | Accepted: 25 September 2023 | Published online: 25 September 2023

Journal of Data Science and Intelligent Systems
2024, Vol. 2(4) 233–240
DOI: 10.47852/bonviewJDSIS32021149
BON VIEW PUBLISHING

RESEARCH ARTICLE

Random Forest Ensemble Machine Learning Model for Early Detection and Prediction of Weight Category

Samuel Iorhemen Ayua[1,*]

[1]*Mathematical Sciences, Taraba State University, Nigeria*

Abstract: The number of insurgents in our nation today is significantly rising each day, and the majority of those affected are living as internally displaced persons (IDP) in various IDP camps. These people experience a variety of health problems as a result of numerous factors. Due to financial difficulties and a lack of accessibility to healthcare facilities and medical professionals, these health risk factors may occasionally go undetected for long periods. BMI excesses, such as those in the underweight, overweight, and obese categories, are linked to several health issues, including low birth weight, poor quality of life, diabetes mellitus, cardiovascular diseases, and higher mortality. In the context of this paper, identifying the health status of IDPs depends critically on human body weight. Considering people living in IDP camps, early detection of the weight categories like underweight, overweight, and obese people is crucial because if not, they will be an early death or other health complications. To reduce mortality rates and other health complications that may result from improper and lately identifying underweight, overweight, and obese members in IDP camps, the researcher collected datasets from the IDP camps, trained, and developed a random forest (RF) ensemble model of supervised learning that will aids the medical practitioner in early detection and prediction of the weight category of IDPs. After hyper-parameter tuning and feature selection, the RF machine learning algorithms identify three significant parameters from the dataset's original 10 parameters to use as the model parameter. The highest accuracy obtained was 92% on the test dataset and 96% on the training dataset for the RF classifier using three features, while the accuracy of 83% was obtained on the test dataset and 87% on the training dataset for the RF classifier using ten features.

Keywords: internally displaced persons (IDP), random forest (RF), ensemble learning, machine learning (ML), weight category, prediction

1. Introduction

Machine learning (ML), according to Archana et al. [1], is a process in which computers learn from experience or instruction to understand concepts or skills. It aims to create computers that can learn on their own without the direct involvement of humans. To recognize patterns in data and make better decisions, the learning process begins with observation [2]. They go on to say that data entry and training of the algorithms to find patterns that can make future predictions based on new data are key components of ML.

Today, insurgency and terrorism are on the rise in our country, and the majority of those impacted live in internally displaced person (IDP) camps as IDPs. These people experience a variety of health problems as a result of numerous factors. Due to financial difficulties and a lack of accessibility to healthcare facilities and medical professionals, these health risk factors may occasionally go undetected for long periods.

Human weight specifically plays a very important key role in this paper, either positively or negatively. Considering factors that may contribute to depression of people living in IDP camps, such as inadequate nutrition, a high intake of carbohydrates, etc., which may cause premature death or other health complications. Early detection of the weight category of IDPs, especially the underweight, overweight, and obese, is crucial. The three main health issues in the world are underweight, overweight, and obesity. Normal energy consumption deficits lead to underweight conditions, whereas excessive energy intake causes overweight conditions.

Weight category (underweight, overweight, and obese) among IDPs is a significant long-standing challenge to the well-being and a public health concern that has negative effects on IDP members. Underweight, overweight, and obese individuals typically fall into the worst weight category. IDPs who are underweight may experience a variety of problems, such as decreased productivity at work and a higher risk of miscarriage, stillbirth, low birth weight, and infant mortality in women [3, 4]. Various maternal and fetal complications during pregnancy, delivery, and the postpartum period can be brought on by maternal obesity [5]. Diabetes, stroke, heart disease, cardiovascular disease, and respiratory issues are among the communicable diseases that are linked to being underweight or overweight [6–9].

Lack of early identification, in particular for the adverse weight category, increases the likelihood of experiencing health problems. An important task for preventing weight category-related difficulties

*Corresponding author: Samuel Iorhemen Ayua, Mathematical Sciences, Taraba State University, Nigeria. Email: ayuasamuel@tsuniversity.edu.ng

is an early detection of IDP's adverse weight category. Numerous plans have also been proposed to lessen this risk, but the current approaches are not yet fully sufficient to accomplish this [10–19]. It is now more important than ever to use an automated system to identify and predict the weight category of IDP early on. Today, ML-based algorithms are widely used as automated systems for early disease prediction that can be accurately predicted [20, 21]. Nevertheless, several ML algorithms have been used to predict diseases like hypertension [22], anemia [20, 23], diabetes [22, 24], low birth weight [25–27], and child mortality [28–30] using different demographic and health survey (DHS) datasets. Additionally, some ML-based studies on child underweight have been done [31–37], but no studies have been done in Nigeria, specifically the IDPs. In this study, an attempt was made for the first time to adapt ML algorithms for predicting obesity using a local hospital dataset in Nigeria [38].

The proposed system aimed to develop a random forest (RF) ensemble ML model of supervised learning that would aid the medical professional in the early detection and prediction of adverse weight category of members in IDP camps in the northeast of Nigeria, so that mortality rate and other health complications that arise as a result of inadvertent identification of underweight, overweight, and obese persons may be greatly reduced to the barest minimal.

And the particular objectives are:
1) Dataset collection, preprocessing, and determination of the best set of features for weight category (underweight, overweight, and obese) prediction;
2) Developed a model that predicts the weight category (underweight, overweight, and obese) of IDPs using RF ensemble ML techniques with the locally collected dataset;
3) Evaluate the model performance using GridSearchCV, Confusion Matrix, and performance accuracy metrics;
4) Validate the model performance using the test dataset;
5) Make live predictions from the model using the new data points.

2. Literature Review

Anisat et al. [38] uses machine learning techniques on a publicly available clinical dataset to predict obesity status using different machine learning algorithms to solve the issues of validity of diagnosis, time-consuming factor, and also provide a reliable diagnosis system that can be used for all genders. The ML model was designed using the python programming language. Algorithms used to achieve the aim of their research were gradient boosting (GB) classifier, RF classifier, K-nearest neighbor (KNN), and support vector machine (SVM). The gradient boosting algorithm outperformed the other algorithms compared with an accuracy of 99.05%.

Cheng et al. [39] attempted to investigate the connection between physical activity and weight status in humans and to contrast some ML and conventional statistical models for predicting obesity levels. The National Health and Nutrition Examination Survey Dataset was used in their model, and 11 different algorithms, including the random subspace, logistic regression, decision table, Naive Bayes (NB), the radial basis function, K-nearest neighbor, classification via regression, J48, and multilayer perception, were used for their implementation and evaluation. The algorithms with the highest overall accuracy were the random subspace classifier algorithm. The evaluation metrics used were the Receiver Operating Characteristics (ROC) curve and area under the curve (AUC).

Artificial neural networks, deep learning, decision tree analysis, and other ML techniques are used by DeGregory et al. [40] to predict and/or classify obesity levels from a large dataset obtained from sensors, smartphones, and electronic medical health records. They concluded that ML will offer sophisticated tools to forecast, categorize, and describe risks associated with obesity and its consequences.

Ward et al. [41] attempted to predict the percentage of adults in the United States who will be obese shortly by fitting multinomial regression to estimate the prevalence of four body mass index (BMI) categories (less than 25 is normal weight, 25 to less than 30 is overweight, 30 to less than 35 is moderate obesity, and above 35 is severe obesity) on a self-reporting bias dataset from the behavioral risk factor With 48.9% of adults being obese and 24.2% being severely obese in 2030, they concluded from their analysis that adult obesity rates will continue to rise. In 25 states, the prevalence will be higher than 25%.

The BMI increase pattern in children was examined by Rossman et al. [42]. They also created a system that predicts which children are at a high risk of becoming obese before it reaches a critical point. They concluded from their analysis that the BMI increases most between the ages of 2 and 4 and that their accurate prediction occurs between 5 and 6 years of age.

To model data from an electronic health record, Pang et al. [43] developed a model that predicts early childhood obesity using XGBoost, the ID3 decision tree model, and the recurrent neural network ML algorithm. The XGBoost outperformed all other models, having the best AUC, a value of 0.81 (0.001) among them.

Using a publicly available dataset, Davila-Payan et al. [44] developed a logistic regression model to assess the likelihood of BMI in rural children between the ages of 2 and 17 using the data. As prevalence among censuses ranges from 27 to 40%, the result demonstrates the importance of estimates in creating effective involvements and assisting in the planning of potential solutions to the problem.

Manna and Jewkes [45] presented a model that makes use of the fuzzy signature to comprehend and manage the complexities of the children's obesity dataset and a solution that could handle the risk associated with early obesity and children's motor development.

Adnan et al. [46] developed a preliminary method for predicting obesity using data gathered from primary sources, including parents, caregivers, and the kids themselves. The authors of this paper make an effort to pinpoint risk factors like childhood obesity, parental education levels, children's habits and lifestyles, and environmental influences. The proposed framework makes use of the NBTree, a hybrid decision tree/NB algorithm.

Adnan et al. [47] used data mining to predict childhood obesity. The proposed survey aimed to offer the necessary data on the obesity issue. They were implemented using the neural network (NN), NB, and decision tree models.

Abdullah et al. [48] categorized obesity in grade 6 students from two distinct Malaysian districts. The information gathered was modeled using a classification technique. Decision trees, SVMs, NNs, and Bayesian networks are the ML classification models employed.

Using data gathered before the children turned 2 years old, Dugan et al. [49] published an article to predict childhood obesity. To do this, they looked at six different ML techniques, including ID3, random tree, RF, J48, NB, and Bayes train on CHIKA data. When using an ensemble ML approach to predict obesity in humans, the RF method [50] achieved an overall accuracy of 85% while taking into account factors like age and BMI (weight and height) of individuals. They employ ML techniques like the partial least squares, the RF, and the linear model; they obtained an accuracy of 89.68% from a RF.

Ward et al. [41] researched the prediction of the percentage of obesity levels in adults in the nearest future using fitting multinomial regression to estimate the prevalence of four BMI categories (less than 25 is normal weight, 25 to less than 30 is overweight, 30 to less than 35 is moderate obesity, and above 35 is severe obesity) on a self-reported bias dataset.

Data from an Afghan child malnutrition survey were examined by Momand et al. [34]. They used three ML-based algorithms to predict children's nutritional status: RF, PART rule induction, and NB. They found that RF and PART induction rule algorithms were the best and had marginally better prediction accuracy.

For their investigation, Bitew et al. [31] used Ethiopian DHS 2016 data. Ten thousand six hundred forty-one kids made up the dataset. They used three well-known ML-based algorithms, RF, linear regression (LR), and KNN, to predict the risk of child mortality. They suggested that the RF-based algorithm was superior based on the methodical evaluation of the performance parameters.

To predict the malnutrition status of children under the age of five in Bangladesh, Talukder and Ahammed [36] used five classifiers, including LR, SVM, Linear Discriminant Analysis (2021) LDA, KNN, and RF. The models were built using 6853 pieces of data, and they discovered that the RF-based algorithm had the highest accuracy (68.5%).

Rahman et al. [51] recently conducted a study on Bangladesh DHS 2014 data, which included 7079 cases. They used three ML-based algorithms SVM, RF, and LR to identify potential risk factors before putting them into practice. The overall performance showed that stunting, wasting, and being underweight were the three conditions for which LR-RF had the highest predictive performance scores.

The main flaw in this research is that it only considers one weight category when solving multiclass problems. Using the IDP dataset from northeast Nigeria, the model build will be able to determine an accurate weight category of IDPs. To develop an advanced prediction model using a RF ensemble ML approach and implement it for use in actual situations, it is imperative to gather and use more features from the northeast IDP centers. To develop a model that is more accurate and involves fewer human errors in predicting an individual's weight status, this study will concentrate on the IDP dataset that was gathered during the model's construction, training, and testing. Based on these considerations, this paper looked into using ML methods from RF ensemble to create a prediction model for the weight categories (underweight, overweight, and obese) in IDP centers. The research gap comes from these.

3. Research Methodology

This discusses the methodology employed to achieve the aim and objectives of the study. It helps to plan, organize, and implement ML projects from one step to another. The Python programming language, ML algorithms, and other libraries designed specifically for ML research were used for this work.

3.1. Dataset population and description

The dataset size used for this research was 500, from 3 IDP centers in North East Nigeria. The dataset was acquired in the form of measurement, questionnaire, and interview and was recorded in Excel and converted into comma-separated value (CSV) format. The features considered include age, gender, height, weight, religion, region, drinking water, toilet facility, cooking fuel, and food consumed as independent features and weight category as dependent features.

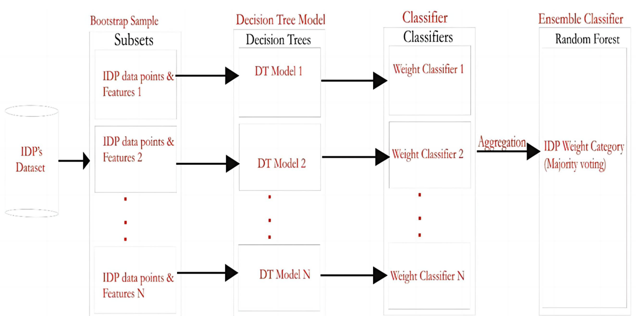
3.2. Method and development tools

The study makes use of a RF ensemble ML technique with an N subset (i.e. bootstrap sample) and N decision trees in Jupyter notebook on the ANNACONDA Navigator toolkit. Python packages like sklearn, numpy, matplotlib, etc. were used, along with some libraries (GridSearchCV, Confusion Matrix, etc.), to build the proposed model using the Python programming language.

3.3. The proposed model architecture

This research describes the processes involved in building the proposed model from the dataset acquisition and preprocessing to the simulation of the model as shown in Figure 1.

Figure 1
Random forest ensemble architectural model for weight category prediction



Steps in implementing the IDP's weight category predictions and detection using random forest ensemble technique are shown below:

Step 1: The dataset collected from IDP's camps (i.e. IDP's dataset) was preprocessed and important features were selected for use in the model.
Step 2: The original dataset (IDP's Dataset) was bootstrap into N training subsets (bootstrap sample) with a replacement known as row sampling.
Step 3: The training subsets of IDP's data points and features from the bootstrap sample were used in training the N decision trees.

Table 1
Weight category raw dataset uploaded into Jupyter notebook

Age	Gender	Height	Weight	Religion	Region	Drinking water	Toilet facility	Cooking fuel	Food consumed	Weight category
48	Male	178	96	Christian	Takum	Improved	Non-hygienic	Improved	Balance diet	Overweight
8	Male	189	87	Muslim	Wukari	Non-improved	Hygienic	Non-hygienic	Balance diet	Obesity
55	Female	185	110	Christian	Donga	Improved	Non-hygienic	Improved	Balance Diet	Overweight
52	Female	195	104	Christian	Tela	Improved	Hygienic	Non-hygienic	Non-balance diet	Normal weight
31	Male	149	61	Christian	Donga	Non-improved	Non-hygienic	Improved	Non-balance diet	Normal weight

Step 4: Each decision tree generates a weight classifier, such that for N decision trees N weight classifier was obtained.
Step 5: The N weight classifiers are aggregated to generate output based on majority voting.
Step 6: The weight category of IDPs was obtained using RF ensemble classifier method-based ensemble learning.

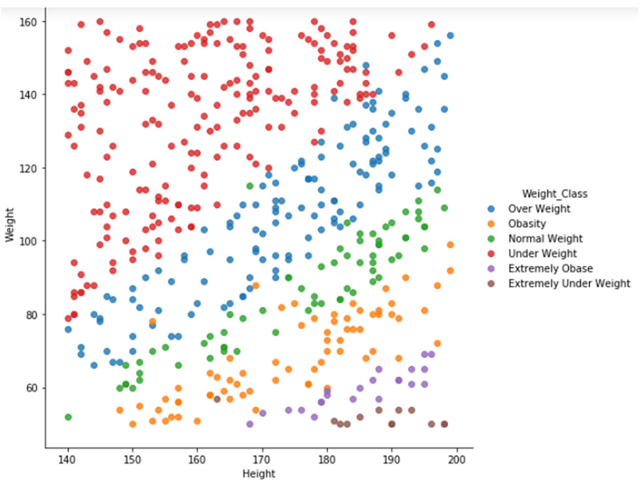
4. Results and Discussion

The dataset acquired was in the form of measurement, questionnaire, and interview, which was recorded in Excel and converted into CSV format. The features considered include age, gender, height, weight, religion, region, drinking water, toilet facility, cooking fuel, etc., as independent features and weight category as dependent features. With the help of data preprocessing (replacement of missing value, feature selection, and feature reductions of values), the processed dataset obtained and used for the building and training of the model is discussed below.

Table 1 shows the raw IDP dataset for weight category prediction uploaded into Jupyter notebook for data preprocessing.

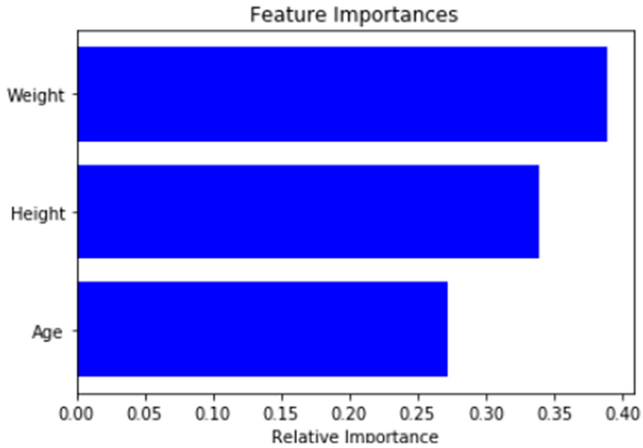
Figure 2 shows the plot of the raw IDP dataset of height against weight, which was used for training and testing of the model after the best parameter were obtained.

Figure 2
Scattered plot of height against weight for weight category classification



According to Figure 3, the bar chart shows the important features according to their scores, which were selected to use in the model.

Figure 3
Bar chart showing important features according to their scores



After parameter tuning, the model obtained important hyper-parameters as shown in Table 2, which was used in the building of the model for weight prediction.

Table 2
Important hyper-parameters used in the model after parameter tuning

Hyper-parameters	Values
Bootstrap	True
criterion	entropy
max_depth	16
n_jobs	−1
max_features	sqrt
min_samples_leaf	1
min_samples_split	2
n_estimators	50

From the study, our experimental results showed that the RF-ensemble system considering three features using 3-fold cross-validation on model testing realizes the accuracy of 92% as shown in Table 3 and the confusion matrix as shown in Figure 4. This summary in matrix form shows how the prediction is correct and incorrect per classes, as we can see the model has a very good confusion matrix having less confused classes.

Table 3
Classification report for model testing using three features

	Precision	Recall	F1-score	Support
Extremely obese	0.83	1.00	0.91	5
Extremely under weight	1.00	1.00	1.00	1
Normal weight	0.92	0.92	0.92	13
Obesity	0.94	0.88	0.91	17
Overweight	0.92	0.88	0.91	26
Underweight	0.92	0.95	0.93	37
Accuracy	–	–	**0.92**	99
Macro avg	0.92	0.94	0.93	99
Weighted avg	0.92	0.92	0.92	99

Table 4
Classification report for model testing using ten features

	Precision	Recall	F1-score	Support
Extremely obese	0.50	0.67	0.57	3
Extremely under weight	1.00	1.00	1.00	2
Normal weight	0.71	0.71	0.71	14
Obesity	0.92	0.73	0.81	15
Overweight	0.70	0.88	0.78	24
Underweight	0.97	0.88	0.93	42
Accuracy	–	–	**0.83**	100
Macro avg	0.80	0.81	0.80	100
Weighted avg	0.85	0.83	0.83	100

Figure 4
Confusion matrix for model testing using three features

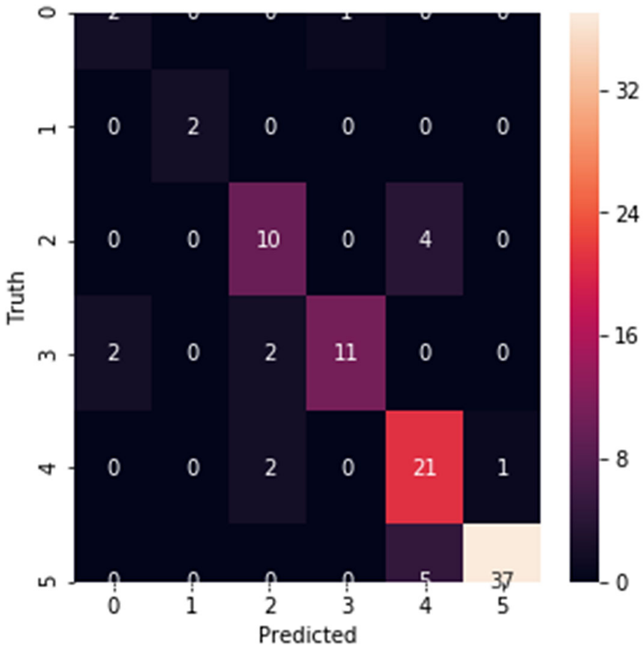
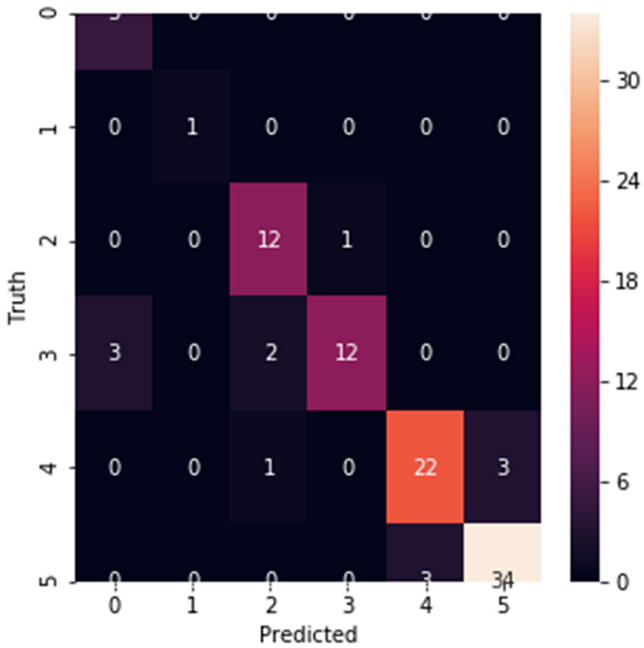


Figure 5
Confusion matrix for model testing using ten features



The RF classifier considering 10 features using 3-fold cross-validation on model testing realizes the accuracy of 83% as shown in Table 4 and the confusion matrix as shown in Figure 5. This summary in matrix form shows how the prediction is correct and incorrect per classes, as we can see the model has some classified classes as shown in the confusion matrix. Based on the result, the researcher concluded that adding more features that have zero impact on the model will only slow the performance of the model as many input features will be required at runtime.

Below is the bar chart showing the population of each weight category identified in the dataset collected as described in Figure 6.

Table 5 summarizes some major related literatures and states the contribution of the current research to the body of knowledge based on the final model.

Figure 6
Plot showing the total number of each weight category from the dataset

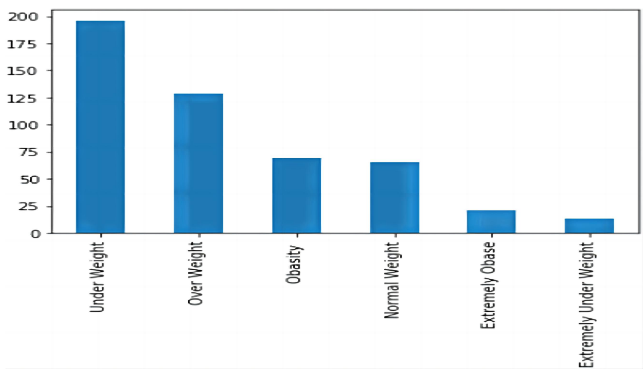


Table 5
Summary of the difference between our research and previous research published

Author	Year	Data size	Country	Classifier types	Final model
Shahriar et al.	2019	6995	Bangladesh	ANN, SVM, RF, NB, DT	RF used for the prediction of stunted, wasted, and underweight with the accuracy of 67.3%, 86.0%, and 70.0%, respectively
Talukder and Ahammed	2020	6868	Bangladesh	LDA, KNN, SVM, RF, LR	RF for the prediction of underweight with the accuracy of 68.5%
Rahman et al.	2021	7079	Bangladesh	SVM, RF, LR	RF used for the prediction of stunted, wasted, and underweight with the accuracy of 88.3%, 87.7%, and 85.7%, respectively
Musa et al.	2022	–	Nigeria	GB, RF, DT KNN,SVM	GB used for the prediction of obesity with an accuracy of 99.05%
Current study	2023	500	Nigeria	RF	RF used for the prediction of weight category (underweight, overweight, and obese) with an accuracy of 96% considering three features and 87% considering ten features

In this research, RF ML algorithm was used for the prediction of weight category (underweight, overweight, and obese) of the IDPs in the Northeast of Nigeria since it was proven to have better accuracy score and recommended by the many previous researchers. As compared to other research, in which their work was on the prediction of one or two of the weight category but not the three or all categories as the current research is. And the value of the accuracy obtained in the current research is 96% higher as compared to the accuracy of the previous research during the training and testing of the model. In future, we would to expand and get more dataset across all IDPs centers in Nigeria. Based on the result, the researcher also observed that adding more features that have zero impact on the model will only slow the performance of the model and cause over-fitting during model training.

5. Conclusion

Underweight, overweight, and obese are important public health concerns in all developing countries. This paper presents a comprehensive study for the detection and prediction of the IDP weight category using the FR classifier. From the dataset collected, our findings considered 10 risk factors for underweight, overweight, and obese prediction. These selected factors were used as input features in RF ensemble classifiers for the prediction of the weight category of IDPs. The RF classifier considering three features realizes the highest accuracy of 92% using the test dataset and 96% using the training dataset on the built model, while the RF classifier using 10 features realizes an accuracy of 83% using the test dataset and 87% using the training dataset on the built model. Based on the result, the researcher concluded that adding more features that have zero impact on the model will only slow the performance of the model as many input features will be required at runtime and may cause over-fitting during model training.

Our proposed RF system will be able to identify underweight, overweight, and obese at a lower cost and in less time, which if not early identify may result in different kinds of health-related issues and diseases. This study will help the government, healthcare providers, and policymakers with quick decision making and implementing required intervention as well as quality care practice to avoid severe complications on IDPs due to weight category-related issues.

This work suggests RF algorithm to be used on larger dataset, which may accurately classified and predict underweight, overweight, obese and yield higher accuracy when processes like data modeling, data processing, etc. are duly observed.

Recommendation

Considering the negative impact of these weight categories on IDPs, the researcher recommended the following:

1) The system should be implemented in all the Nigerian IDP camps to help remediate all weight-category health-related complications due to lack of early identification, unavailability of nearby healthcare centers, and medical practitioners.

Acknowledgments

I am grateful to God for granting me success in this paper and for His help in my times of need. In addition, I am grateful to all that worked on actualizing this article's success.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by the author.

Conflicts of Interest

The author declares that he has no conflicts of interest to this work.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Author Contribution Statement

Samuel Iorhemen Ayua: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration.

References

[1] Archana, C., Raj, K., & Savita, K. (2013). Machine learning classification techniques: A comparative study. *International Journal on Advanced Computer Theory and Engineering*, *2*(4), 2319–2526.

[2] Deisenroth, M. P., Faisal, A. A., & Ong, C. S. (2020). *Mathematics for machine learning*. UK: Cambridge University Press.

[3] Boutari, C., Pappas, P. D., Mintziori, G., Nigdelis, M. P., Athanasiadis, L., Goulis, D. G., & Mantzoros, C. S. (2020). The effect of underweight on female and male reproduction. *Metabolism*, *107*, 154229. https://doi.org/10.1016/j.metabol.2020.154229

[4] Khan, M. N., Rahman, M. M., Shariff, A. A., Rahman, M. M., Rahman, M. S., & Rahman, M. A. (2017). Maternal undernutrition and excessive body weight and risk of birth and health outcomes. *Archives of Public Health*, *75*, 12. https://doi.org/10.1186/s13690-017-0181-0

[5] Melchor, I., Burgos, J., del Campo, A., Aiartzaguena, A., Gutiérrez, J., & Melchor, J. C. (2019). Effect of maternal obesity on pregnancy outcomes in women delivering singleton babies: A historical cohort study. *Journal of Perinatal Medicine*, *47*(6), 625–630. https://doi.org/10.1515/jpm-2019-0103

[6] Kc, B. (2019). Factors responsible for non-communicable diseases among Bangladeshi adults. *Biomedical Journal of Scientific & Technical Research*, *20*(1), 14742–14748.

[7] Nyberg, S. T., Batty, G. D., Pentti, J., Virtanen, M., Alfredsson, L., Fransson, E. I., . . . , & Kivimäki, M. (2018). Obesity and loss of disease-free years owing to major non-communicable diseases: A multicohort study. *The Lancet Public Health*, *3*(10), e490–e497. https://doi.org/10.1016/S2468-2667(18)30139-7

[8] Rahman, A., & Sathi, N. J. (2021). Sociodemographic risk factors of being underweight among ever-married Bangladeshi women of reproductive age: A multilevel analysis. *Asia Pacific Journal of Public Health*, *33*(2–3), 220–226. https://doi.org/10.1177/1010539520979924

[9] Rawal, L. B., Kanda, K., Mahumud, R. A., Joshi, D., Mehata, S., Shrestha, N., . . . , & Renzaho, A. (2018). Prevalence of underweight, overweight and obesity and their associated risk factors in Nepalese adults: Data from a nationwide survey. *PLoS ONE*, *13*(11), e0205912. https://doi.org/10.1371/journal.pone.0205912

[10] Abedin, M. M., Haque, M. E., Sabiruzzaman, M., Al Mamun, A. S. M., & Hossain, M. G. (2019). Multinomial logistic regression analysis of factors influencing malnutrition of non-pregnant married women in Bangladesh: Evidence from Bangladesh Demographic and Health Survey-2014. In *7th International Conference on Data Science and SDGs: Challenges, Opportunities and Realities*, 233–240.

[11] Ayele, E., Gebreayezgi, G., Mariye, T., Bahrey, D., Aregawi, G., & Kidanemariam, G. (2020). Prevalence of undernutrition and associated factors among pregnant women in a public general hospital, Tigray, Northern Ethiopia: A cross-sectional study design. *Journal of Nutrition and Metabolism*, *2020*(1), 2736536. https://doi.org/10.1155/2020/2736536

[12] Hasan, M., Sutradhar, I., Shahabuddin, A. S. M., & Sarker, M. (2017). Double burden of malnutrition among Bangladeshi women: A literature review. *Cureus*, *9*(12), e1986. https://doi.org/10.7759%2Fcureus.1986

[13] Hossain, M. M., Islam, M. R., Sarkar, A. S. R., Ali Khan, M. M., & Taneepanichskul, S. (2018). Prevalence and determinants risk factors of underweight and overweight among women in Bangladesh. *Obesity Medicine*, *11*, 1–5. https://doi.org/10.1016/j.obmed.2018.05.002

[14] Ilyas, U., & Parveen, K. (2019). Malnutrition and its associated risk factors among women of reproductive age in rural community of Lahore. *International Journal of Medical Research & Health Sciences*, *8*(3), 173–178.

[15] Khanam, R., Lee, A. S. C., Ram, M., Quaiyum, M. A., Begum, N., Choudhury, A., . . . , & Baqui, A. H. (2018). Levels and correlates of nutritional status of women of childbearing age in rural Bangladesh. *Public Health Nutrition*, *21*(16), 3037–3047. https://doi.org/10.1017/S1368980018001970

[16] Mustari, S., Hossain, I., Khatun, K., Ali, M. S., Rahman, H., Mondal, A., . . . , & Islam, S. (2017). Prevalence and determinants of malnutrition among poor women and children in the South-West Region of Bangladesh. *IOSR Journal of Pharmacy and Biological Sciences*, *13*(4), 62–69.

[17] Rahman, M. S., Mushfiquee, M., Masud, M. S., & Howlader, T. (2019). Association between malnutrition and anemia in under-five children and women of reproductive age: Evidence from Bangladesh Demographic and Health Survey 2011. *PLoS ONE*, *14*(7), e0219170. https://doi.org/10.1371/journal.pone.0219170

[18] Rahman, M., & Sultana, P. (2019). Distribution and risk factors of child malnutrition in Bangladesh, based on Bangladesh Demographic and Health Survey-2014 data. *Journal of Biometrics & Biostatistics*, *10*(1), 1000425.

[19] Tanwi, T. S., Chakrabarty, S., & Hasanuzzaman, S. (2019). Double burden of malnutrition among ever-married women in Bangladesh: A pooled analysis. *BMC Women's Health*, *19*(1), 24. https://doi.org/10.1186/s12905-019-0725-2

[20] Islam, M. M., Rahman, M. J., Roy, D. C., Islam, M. M., Tawabunnahar, M., Ahmed, N. A. M. F., & Maniruzzaman, M. (2022). Risk factors identification and prediction of anemia among women in Bangladesh using machine learning techniques. *Current Women's Health Review*, *18*(1), 118–133. https://doi.org/10.2174/1573404817666210215161108

[21] Ngiam, K. Y., & Khor, I. W. (2019). Big data and machine learning algorithms for health-care delivery. *The Lancet Oncology*, *20*(5), e262–e273. https://doi.org/10.1016/S1470-2045(19)30149-4

[22] Islam, M. M., Rahman, M. J., Roy, D. C., Tawabunnahar, M., Jahan, R., Ahmed, N. A. M. F., & Maniruzzaman, M. (2021). Machine learning algorithm for characterizing risks of hypertension, at an early stage in Bangladesh. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, *15*(3), 877–884. https://doi.org/10.1016/j.dsx.2021.03.035

[23] Jaiswal, M., Srivastava, A., & Siddiqui, T. J. (2019). Machine learning algorithms for anemia disease prediction. In *Recent Trends in Communication, Computing, and Electronics: Select Proceedings of IC3E 2018*, 463–469. https://doi.org/10.1007/978-981-13-2685-1_44

[24] Maniruzzaman, M., Kumar, N., Menhazul Abedin, M., Shaykhul Islam, M., Suri, H. S., El-Baz, A. S., & Suri, J. S. (2017). Comparative approaches for classification of diabetes mellitus data: Machine learning paradigm. *Computer Methods and Programs in Biomedicine*, *152*, 23–34. https://doi.org/10.1016/j.cmpb.2017.09.004

[25] Borson, N. S., Kabir, M. R., Zamal, Z., & Rahman, R. M. (2020). Correlation analysis of demographic factors on low birth weight and prediction modeling using machine learning techniques. In *Fourth World Conference on Smart Trends in Systems, Security and Sustainability,* 169–173. https://doi.org/10.1109/WorldS450073.2020.9210338

[26] Eliyati, N., Faruk, A., Kresnawati, E. S., & Arifieni, I. (2019). Support vector machines for classification of low birth weight in Indonesia. *Journal of Physics: Conference Series*, *1282*, 012010. https://doi.org/10.1088/1742-6596/1282/1/012010

[27] Faruk, A., & Cahyono, E. S. (2018). Prediction and classification of low birth weight data using machine learning techniques. *Indonesian Journal of Science & Technology*, *3*(1), 18–28. https://doi.org/10.17509/ijost.v3i1.10799

[28] Alves, L. C., Beluzo, C. E., Arruda, N. M., Bresan, R. C., & Carvalho, T. (2020). Assessing the performance of machine learning models to predict neonatal mortality risk in Brazil, 2000–2016. *medRxiv Preprint: 2020.05.22.20109165*. https://doi.org/10.1101/2020.05.22.20109165

[29] Jaskari, J., Myllärinen, J., Leskinen, M., Rad, A. B., Hollmén, J., Andersson, S., & Särkkä, S. (2020). Machine learning methods for neonatal mortality and morbidity classification. *IEEE Access*, *8*, 123347–123358. https://doi.org/10.1109/ACCESS.2020.3006710

[30] Mboya, I. B., Mahande, M. J., Mohammed, M., Obure, J., & Mwambi, H. G. (2020). Prediction of perinatal death using machine learning models: A birth registry-based cohort study in northern Tanzania. *BMJ Open*, *10*(10), e040132. https://doi.org/10.1136/bmjopen-2020-040132

[31] Bitew, F. H., Nyarko, S. H., Potter, L., & Sparks, C. S. (2020). Machine learning approach for predicting under-five mortality determinants in Ethiopia: Evidence from the 2016 Ethiopian Demographic and Health Survey. *Genus*, *76*(1), 37. https://doi.org/10.1186/s41118-020-00106-2

[32] Khare, S., Kavyashree, S., Gupta, D., & Jyotishi, A. (2017). Investigation of nutritional status of children based on machine learning techniques using Indian demographic and health survey data. *Procedia Computer Science*, *115*, 338–349. https://doi.org/10.1016/j.procs.2017.09.087

[33] Markos, Z., Doyore, F., Yifiru, M., & Haidar, J. (2014). Predicting under nutrition status of under-five children using data mining techniques: The case of 2011 Ethiopian Demographic and Health Survey. *Journal of Health & Medical Informatics*, *5*(2), 1000152. https://doi.org/10.4172/2157-7420.1000152

[34] Momand, Z., Mongkolnam, P., Kositpanthavong, P., & Chan, J. H. (2020). Data mining based prediction of malnutrition in Afghan children. In *12th International Conference on Knowledge and Smart Technology*, 12–17. https://doi.org/10.1109/KST48564.2020.9059388

[35] Shahriar, M. M., Iqubal, M. S., Mitra, S., & Das, A. K. (2019). A deep learning approach to predict malnutrition status of 0-59 month's older children in Bangladesh. In *IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology*, 145–149. https://doi.org/10.1109/ICIAICT.2019.8784823

[36] Talukder, A., & Ahammed, B. (2020). Machine learning algorithms for predicting malnutrition among under-five children in Bangladesh. *Nutrition*, *78*, 110861. https://doi.org/10.1016/j.nut.2020.110861

[37] Thangamani, D., & Sudha, P. (2014). Identification of malnutrition with use of supervised data mining techniques – Decision trees and artificial neural networks. *International Journal of Engineering and Computer Science*, *3*(9), 8236–8241.

[38] Anisat, M. F., Basaky, F. D., & Osaghae, E. O. (2022). Obesity prediction using machine learning techniques. *Journal of Applied Artificial Intelligence*, *3*(1), 24–33. https://doi.org/10.48185/jaai.v3i1.470

[39] Cheng, X., Lin, S. Y., Liu, J., Liu, S., Zhang, J., Nie, P., …, & Xue, H. (2021). Does physical activity predict obesity—A machine learning and statistical method-based analysis. *International Journal of Environmental Research and Public Health, 18*(8), 3966. https://doi.org/10.3390/ijerph18083966

[40] DeGregory, K. W., Kuiper, P., DeSilvio, T., Pleuss, J. D., Miller, R., Roginski, J. W., …, & Thomas, D. M. (2018). A review of machine learning in obesity. *Obesity Reviews*, *19*(5), 668–685. https://doi.org/10.1111/obr.12667

[41] Ward, Z. J., Bleich, S. N., Cradock, A. L., Barrett, J. L., Giles, C. M., Flax, C., …, & Gortmaker, S. L. (2019). Projected U.S. state-level prevalence of adult obesity and severe obesity. *The New England Journal of Medicine*, *381*(25), 2440–2450. https://doi.org/10.1056/NEJMsa1909301

[42] Rossman, H., Shilo, S., Barbash-Hazan, S., Artzi, N. S., Hadar, E., Balicer, R. D., …, & Segal, E. (2021). Prediction of childhood obesity from nationwide health records. *The Journal of Pediatrics*, *233*, 132–140.e1. https://doi.org/10.1016/j.jpeds.2021.02.010

[43] Pang, X., Forrest, C. B., Lê-Scherban, F., & Masino, A. J. (2021). Prediction of early childhood obesity with machine learning and electronic health record data. *International Journal of Medical Informatics*, *150*, 104454. https://doi.org/10.1016/j.ijmedinf.2021.104454

[44] Davila-Payan, C., DeGuzman, M., Johnson, K., Serban, N., & Swann, J. (2015). Estimating prevalence of overweight or obese children and adolescents in small geographic areas using publicly available data. *Preventing Chronic Disease*, *12*(12), E32. https://doi.org/10.5888/pcd12.140229

[45] Manna, S., & Jewkes, A. M. (2014). Understanding early childhood obesity risks: An empirical study using fuzzy signatures. In *IEEE International Conference on Fuzzy Systems,* 1333–1339. https://doi.org/10.1109/FUZZ-IEEE.2014.6891838

[46] Adnan, M. H. M., Husain, W., & Rashid, N. A. A. (2011). A framework for childhood obesity classifications and predictions using NBtree. In *7th International Conference on Information Technology in Asia*, 1–6. https://doi.org/10.1109/CITA.2011.5999502

[47] Adnan, M. H. B. M., Husain, W., & Damanhoori, F. (2010). A survey on utilization of data mining for childhood obesity prediction. In *8th Asia-Pacific Symposium on Information and Telecommunication Technologies*, 1–6.

[48] Abdullah, F. S., Manan, N. S. A., Ahmad, A., Wafa, S. W., Shahril, M. R., Zulaily, N., …, & Ahmed, A. (2017). Data mining techniques for classification of childhood obesity among year 6 school children. In *Recent Advances on Soft Computing and Data Mining: The Second International Conference on Soft Computing and Data Mining*, 465–474. https://doi.org/10.1007/978-3-319-51281-5_47

[49] Dugan, T. M., Mukhopadhyay, S., Carroll, A., & Downs, S. (2015). Machine learning techniques for prediction of early childhood obesity. *Applied Clinical Informatics*, *6*(03), 506–520. https://doi.org/10.4338/ACI-2015-03-RA-0036

[50] Jindal, K., Baliyan, N., & Rana, P. S. (2018). Obesity prediction using ensemble machine learning approaches. In *Recent Findings in Intelligent Computing Techniques*: *Proceedings of the 5th ICACNI 2017, 2*, 355–362. https://doi.org/10.1007/978-981-10-8636-6_37

[51] Rahman, S. M. J., Ahmed, N. A. M. F., Abedin, M. M., Ahammed, B., Ali, M., Rahman, M. J., & Maniruzzaman, M. (2021). Investigate the risk factors of stunting, wasting, and underweight among under-five Bangladeshi children and its prediction based on machine learning approach. *PLoS ONE*, *16*(6), e0253172. https://doi.org/10.1371/journal.pone.0253172

How to Cite: Ayua, S. I. (2024). Random Forest Ensemble Machine Learning Model for Early Detection and Prediction of Weight Category. *Journal of Data Science and Intelligent Systems*, *2*(4), 233–240. https://doi.org/10.47852/bonviewJDSIS32021149