DOI : https://doi.org/10.56452/7-2-550

ISSN: 0974-5823 Vol. 7 No. 2 February, 2022

International Journal of Mechanical Engineering

Classification of Obesity Using Several Machine Learning Techniques

Jyoti Parsola
Asst. Professor, School of Computing, Graphic Era Hill University,
Dehradun, Uttarakhand India 248002,

Abstract:

Several potential dangers are linked to obesity. It's a major reason why so many people are becoming sick and dying from chronic diseases. There are several obstacles in the way of uncovering the causes and effects of obesity. The standard regression method assumes independence and linearity among variables and restricts the number of predictors that may be examined. If you're looking for an alternative to traditional approaches to data analysis on obesity, consider using Machine Learning (ML) techniques. Using an innovative strategy with sophisticated machine learning techniques for forecasting obesity as an attempt to go beyond traditional prediction models, this study aims to assess the ability of three different ML methods to identify the presence of obesity using freely accessible health data. These methods are Logistic Regression, Classification and Regression Trees (CART), and Nave Bayes. Meanwhile, the primary purpose of this research is to identify, from among the available variables, a collection of risk factors for adult obesity. In addition, we use the Synthetic Minority Oversampling Technique (SMOTE) to predict obesity status from the known risk variables in order to resolve data imbalance. Based on the results of this analysis, Logistic Regression seems to be the most effective technique. However, the kappa coefficients reveal a weak agreement between the projected and actual rates of obesity. Adult obesity can be predicted by a number of factors, including geographical location, marital status, age range, level of education, consumption of sugary drinks, fat content/oily foods, grilled foods, food that has been preserved, seasoning powders, soft/carbonated drinks, alcohol, medically diagnosed hypertension, mental/emotional disorders, lack of physical activity, smoking, and consumption of fruits and vegetables. Health officials might use this information to better manage chronic illnesses, particularly those linked to obesity, if they knew what risk factors to look out for. Furthermore, employing ML approaches on publically accessible health data, such as Indonesian Basic Health Research (RISKESDAS), is a viable way to bridge the gap for a more solid understanding of the correlations between numerous risk variables in predicting health outcomes.

Introduction

It is crucial in healthcare to assess a patient's degree of obesity. Type-2 diabetes, heart disease, and several malignancies are only few of the chronic conditions for which obesity is a risk factor. When you realise you have a weight problem, you may be more motivated to take action. Intentional weight control also has the added advantage of reducing the risk of illness and improving health. The World Health Organisation (WHO) has established criteria for obesity; nevertheless, Body Mass Index (BMI) alone is insufficient for correctly classifying obesity since it does not adequately capture body-type variables. Each location and individual has unique nutritional needs. Anthropometric data is necessary for accurately representing body type. However, conventional anthropometric techniques cannot be used in everyday life since they need the services of qualified professionals. Less intrusive

than standard anthropometric methods, research into the use of a 3D scanner for human body measuring is presently being pursued. The gold standard for assessing human body fat percent (bf%), Computed Tomography (CT) or Dual-energy X-ray absorptiometry (DXA), carries the danger of radiation exposure when measured often. The human body is not irradiated by a 3D scanner as it is with a CT or DXA machine26. In addition, there isn't a single, best way to assess or forecast health hazards; rather, several approaches might be used. For this reason, this research gathered paired 3D body scan and DXA data from Koreans to aid in the categorization of obesity based on anthropometric measures. Patients and their loved ones are not the only ones who suffer from the harmful impacts of obesity; society as a whole feels the repercussions. It's a double whammy that both undernourishment and obesity have become more common in Southeast Asia, making nutrition-related issues there particularly difficult to solve.

The use of ML methods for modelling epidemiological data is gaining traction in the academic literature. These techniques may help us learn more about the prevalence of diseases, how to spot them early on, what causes them, and where we may take action to prevent or treat them. Many machine learning (ML) techniques and algorithms have been tested on obesity-related data 8). In order to reduce morbidity and mortality caused by obesity, it is crucial to build an accurate data categorization to aid the process of determining predicted risk variables from the supplied data.

Predictions of obesity risk have been made using ML (10) based on information encoding compliance with dietary guidelines and other characteristics. Further applications of ML include the use of electronic health records to foretell childhood obesity before the age of 2 (10), the foretelling of obesogenic settings for children (11), and the modelling of medication dosage responses using aggregated metabolomics, lipidomics, and other clinical data (12).

Literature Review

Yaren Celik et.al.,(2021) Since 1980, obesity has been a major social and public health concern that has demanded more focus. This is prompting a steady stream of research on the causes and effects of childhood obesity and methods for forecasting the onset of the problem. Several techniques of categorization were used to assess the extent of obesity in this research. Different machine learning approaches have their outcomes compared based on the assessment criteria. By using the Cubic SVM technique and carefully picking problem-specific features, we were able to achieve a 97.8% success rate.

A. Ramya et.al.,(2021) Data mining has become more crucial in the contemporary day. Through a knowledge discovery process including the collection of many forms of data, we are able to unearth previously concealed yet crucial information. This makes data mining crucial for getting at the useful hidden information. The use of machine learning algorithms in data mining has proven successful for surfacing relevant information in a hurry. The BMI is used to compare a small number of ML algorithms. When a person's body mass index (BMI) is 30 or more, they are considered obese. Depression, poorer work performance, and disability are just some of the ways in which obesity reduces quality of life. In this article, we used obesity data to test and evaluate many different categorization machine learning methods, including KNN, XGB, Logistic Regression, and DT.

Ala Othman Barzinji et.al.,(2021) There are health hazards linked with being overweight. To slow the rising tide of childhood obesity throughout the world, it is crucial to comprehend the patterns at play. The worldwide incidence of childhood and teenage obesity was predicted using many machine

learning algorithms. This research introduces a fresh method for applying machine learning to make obesity forecasts through 2030. This information originates from a 2015 study of people all across the world. The primary research was using machine learning algorithms to project global obesity rates for 2030, 2040, and 2050. In the second research, we used the SDI to more precisely determine the obesity prevalence rates. Model forecasting accuracies of up to 99% R2 for the primary study and up to 92% R2 for the SDI study are encouraging

Balbir Singh et.al.,(2019) Obese people are at increased risk for a number of health concerns, including type 2 diabetes, respiratory issues, heart disease, and stroke. Keeping up a routine of regular exercise and nutritious diet may be essential in preserving one's health. Therefore, it is crucial to identify cases of childhood obesity. The massive quantity of data made accessible by the Millennium Cohort Study was used in this work. Predicting adolescent BMI from older measurements has been tested using a number of different regression and artificial neural network models. Positive findings have been found, with a prediction accuracy of over 90% being attained. Data mining and the reliability of predictions are examined, along with a number of related concerns.

Kunal Rajput et.al. (2018) Finding health-related issues and their remedies is a major focus of modern medical research, which plays a significant role in community safety. Co-morbidity screening for early intervention When a disease is identified, it allows physicians and patients to take steps towards treating or curing the underlying condition. Examining patients' medical records is one way to spot the sickness, but it's a laborious, manual procedure that's prone to mistakes. Therefore, it has become essential to develop automated or semi-automatic methods for detecting the presence of co-morbidities or pre-existing diseases. In this research, we take i2b2 clinical datasets that are open to the public and apply machine learning and deep learning algorithms to them in order to identify the presence of chronic diseases like obesity. Positive outcomes from our trials.

Models for deep learning, or DL models, are artificial systems that draw inspiration from the human brain. Artificial Neural Networks (ANNs), Recurrent Neural Networks (RNNs), and Convolutional Neural Networks (CNNs) are the three kinds of deep learning models that have so far been applied to the problem of childhood and teenage obesity.

• Artificial Neural Networks (ANN): The simplest kind of neural network is the artificial neural network. Multiple perceptrons or neurons may represent it at each layer. Since they may adapt to any non-linear function, Universal Function Approximators is another name for them. Activation functions are used to introduce this non-linearity. Multi-Layer Perceptron has been utilised for obesity prediction in studies [26], [27], [31], and [35].

• Recurrent Neural Networks (RNN): To analyse sequential data, a class of neural networks known as recurrent neural networks has been developed. When compared to networks without sequence-based specialisation, these sorts of networks are able to grow to significantly longer sequences. RNN was used in a study to predict obesity ([28]).

• Convolutional Neural Networks (CNN): The deep learning community has mostly settled on the use of Convolutional Neural Networks (CNNs). CNNs are constructed using "filters" or "kernels," which are learned automatically and without human involvement. The convolution procedure aids these kernels in their feature extraction from the input data. Similar to how RNNs use shared parameters, CNNs apply a single filter to several regions of the input to produce a feature map.

DOI : https://doi.org/10.56452/7-2-550

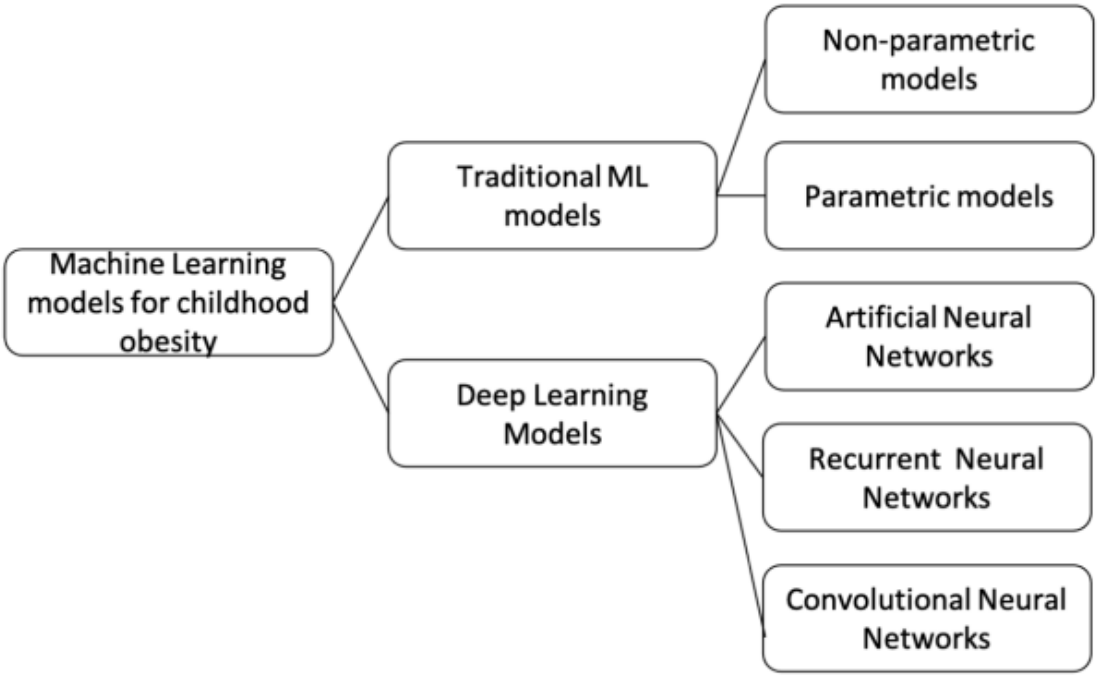

Figure 1. Machine learning models for child and adolescent obesity: a literature categorization by model type

Cohort Study: Most of the early research on utilising machine learning models to predict childhood and teenage obesity makes use of previously collected cohort studies. Longitudinal studies, such as cohort studies, track the same group of people through time. The influence of a variable or risk factor is often researched by exposing a subset of the participants to it and then following their progress through time. Researchers may learn more about the variables that increase or decrease a person's risk of developing an illness by conducting a cohort study. A prospective cohort is a specific kind of cohort research. Attrition bias is exacerbated by the lengthy duration of these research since participants may lose interest and stop taking part. Participants in cohort studies may alter their behaviour if they are aware they are being watched and researched. The 'Hawthorne effect' [43] describes this kind of behaviour, which may have an impact on many different routines, including eating habits, cleanliness routines, etc. The retrospective cohort is another popular choice since its members have an established diagnosis or result. The follow-up phase of this cohort study is finished when the study itself begins. The difference in illness risk between exposed and non-exposed groups is explored by looking at historical or self-reported data. However, there are a few drawbacks to this approach as well. There is a greater likelihood of bias in retrospective cohort studies due to the nature of the sample used. The data may also be of low quality since it was not collected for the purpose of the research at hand, which is yet another drawback of collecting retrospective data. However, several of the cohorts used in the publications included in this review did include anthropometric, behavioural, demographic, and other factors that are known to be linked with obesity. Most research looking for correlations and trends have used cohorts.

Electronic Health Records: All of a patient's medical records, including past diagnoses, medications, allergies, procedures, lab results, radiological pictures, etc., are stored in an EHR. Electronic health record data is continuously updated, making it available for any kind of analysis, whether it

descriptive or predictive, at any moment. Structured and unstructured data are both present in EHR. Information that has been "organised into specific fields as part of a schema, with each field having a defined purpose" is what the Healthcare Information and Management Systems Society (HIMSS)9 calls "structured data." Name, contact details, demographics, lab results, etc., might all fall under this category. Data that "cannot be easily organised using pre-defined structures" is said to be unstructured. Unstructured text data processing is accomplished by means of Natural Language Processing.

Image Datasets: Since the advent of deep learning-based Convolutional Neural Networks, the use of pictures for illness prediction has steadily increased. Due to the sensitive nature of healthcare imaging data and the difficulty in acquiring it for research purposes. As a result, there is a dearth of research using visual information for forecasting adult obesity. There have been few studies using facial photos for obesity prediction or diagnosis in adults, but none have been conducted on children due to a lack of publically accessible data. MRI (Magnetic Resonance Imaging) image files are widely used in the study of childhood obesity and overweight,

Gender	Statistics	Age	Height (cm)	Weight (kg)	BMI (kg/m^2)	DXA BF(%)	BIA BF(%)
Male (87)	Average	24.07	178.10	77.92	24.64	20.13	18.45
	Std	4.25	5.37	13.52	4.00	8.70	7.42
	Min	20.00	165.55	54.30	16.22	5.50	5.40
	Max	39.00	191.83	120.90	37.61	37.40	37.00
Female (73)	Average	24.14	165.63	57.01	20.74	27.55	25.83
	Std	4.54	4.96	10.13	3.34	6.94	6.29
	Min	20.00	155.87	40.70	15.70	6.70	11.70
	Max	37.00	173.99	89.60	32.25	46.00	45.70

Table 1. Collected data statistics

Index	Sex	Underweight	Normal	Overweight	Obesity
BMI (kg/m^2)	Male/Female	BMI < 18.5	18.5 < BMI < 23	23 ≤ BMI < 25	25 ≤ BMI
BIA, DXA	Male	BF% < 10	10 ≤ BF% ≤ 20	20 < BF% ≤ 25	25 < BF%
	Female	BF% < 20	20 ≤ BF% ≤ 28	28 < BF% ≤ 35	35 < BF%

Table 2. Obesity class standard cutoff.

DOI : https://doi.org/10.56452/7-2-550

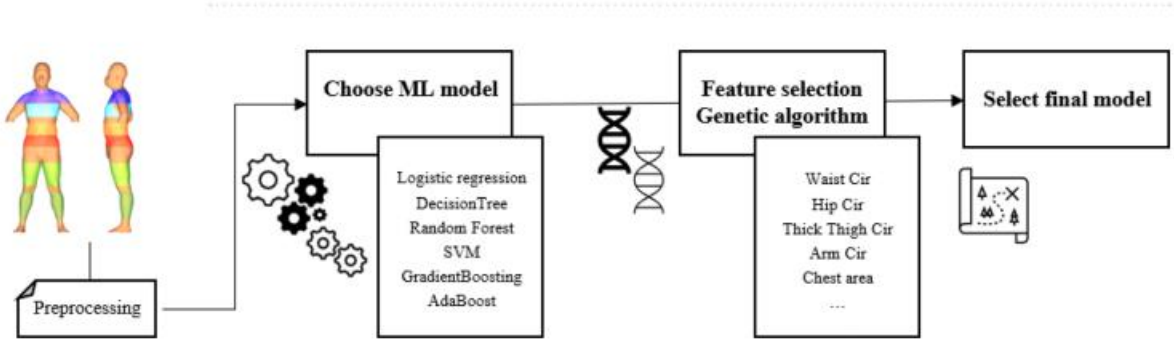

Figure 2. Overall framework of this study.

Choice of Indicators Evolutionary Computing. This method use a Genetic Algorithm to decide upon the machine learning model's input characteristics. The performance of the machine learning model may be improved by carefully selecting the input characteristics, and it can be determined whether or not a certain value among the 3D body measures in Table 2 affects the classification of obesity. Finding the Global Optimum when picking input characteristics by comparing all potential sets of input features is quite difficult. Therefore, a meta-heuristic method was adopted to find a good enough alternative to the Global Optimum. The Genetic Algorithm has been shown to outperform other meta-heuristic algorithms46,47 when it comes to selecting relevant variables. The GA was used to select features for analysis in this research. Like Charles Darwin's idea of natural selection and mammalian reproduction, GA uses a meta-heuristic method to solve complicated issues via efficient trial and error48. Through generational iteration, GA seeks to identify the most optimal input feature for this research. There are six stages to GA. The first step is to specify the settings and initialise the chromosomal combination. Population size (the total number of chromosomes in a generation) and mutation rate (the proportion of chromosomes lost due to mutations) are two such metrics. Input feature selection is represented by a ratio, and the mutation ratio is the proportion of mutated genes to the total number of chromosomes. We used a mutation rate of 20% and a population size of 100. The second step includes using a Random Forest to learn each input characteristic. Third, the ftness function was utilised to assess precision after evaluating each input feature's chromosomes. The fourth step included the accurate selection of great chromosomes from the present generation. The best 80% of chromosomes were chosen for this analysis. In the fifth stage, chromosomes were created by crossing across and mutating. In this scenario, chromosomes from both adoptive parents are combined.

DOI : https://doi.org/10.56452/7-2-550

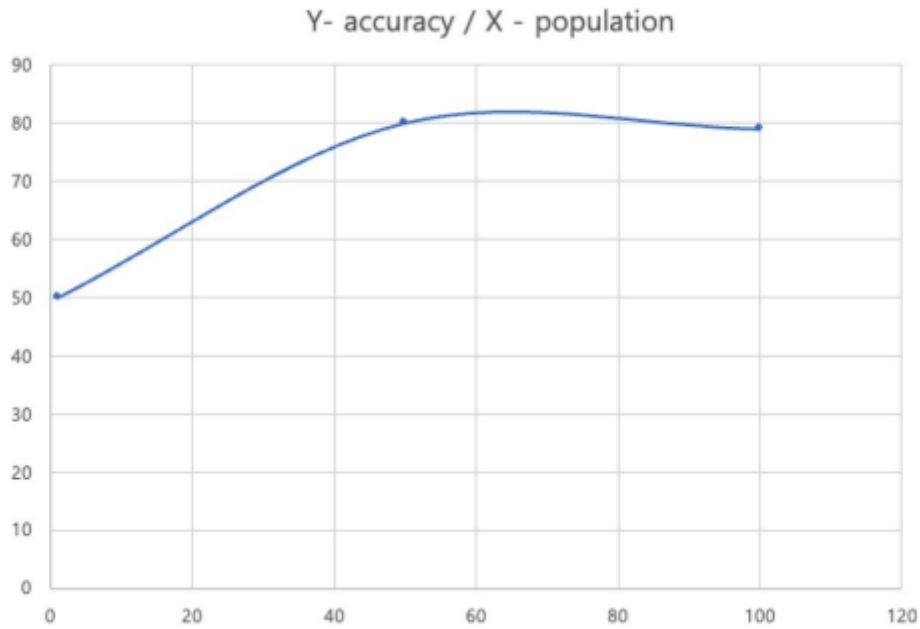

Figure 3. Accuracy flowchart by generation.

Rank	Classifier	Accuracy	Recall	Precision	F1 score
1	Random Forest	**0.725**	**0.661**	**0.780**	**0.692**
2	GradientBoosting	0.700	0.599	0.655	0.609
3	Logistic	0.500	0.515	0.443	0.462
4	SVM	0.650	0.477	0.383	0.412
5	DecisionTree	0.475	0.484	0.478	0.465
6	AdaBoost	0.425	0.358	0.431	0.356

Table 3. Results of the choosing the ML model

Conclusion

The risk of childhood obesity is rising as the availability of unhealthy fast food continues to expand. Promoting health is crucial to the development of every society. Obesity is a chronic condition that affects a lot of individuals. The World Health Organisation (WHO) estimates that over a billion individuals throughout the globe are overweight. Sometimes death is the end result of a lifetime of health problems brought on by obesity. Obesity is a condition that can be fought and ultimately eliminated if caught early enough. But early diagnosis is not as simple as it looks; there are no qualities that the person is aware of, and thus it is extremely unlikely that the person would make attempts to identify them, and neither is it practical to have an expert available on a 24-hour basis. In this research, we propose a basic idea for cloud-based diagnostics to detect possible obesity by using machine learning strategies. The most accurate of many machine learning algorithms was utilised to make diagnoses of the condition. The predictive algorithm was fine-tuned using previously collected data on overweight and obese individuals, yielding very precise predictions. In addition to machine learning, an IoT sensor network was employed to make an instantaneous diagnostic of the patient,

allowing for earlier identification and the subsequent elimination or at least reduction of obesity-related complications.

References

1. Y. Celik, S. Guney and B. Dengiz, "Obesity Level Estimation based on Machine Learning Methods and Artificial Neural Networks," *2021 44th International Conference on Telecommunications and Signal Processing (TSP)*, Brno, Czech Republic, 2021, pp. 329-332, doi: 10.1109/TSP52935.2021.9522628.

2. Ramya and K. Rohini, "Comparative evaluation of machine learning classifiers with Obesity dataset," *2021 International Conference on Computing Sciences (ICCS)*, Phagwara, India, 2021, pp. 38-41, doi: 10.1109/ICCS54944.2021.00016.

3. O. Barzinji, C. Ma, W. Du and J. Ma, "A Machine Learning Approach to Predict the Trend of Obesity Prevalence at a Global Level," *2021 IEEE/ACIS 6th International Conference on Big Data, Cloud Computing, and Data Science (BCD)*, Zhuhai, China, 2021, pp. 25-30, doi: 10.1109/BCD51206.2021.9581579.

4. Singh and H. Tawfik, "A Machine Learning Approach for Predicting Weight Gain Risks in Young Adults," *2019 10th International Conference on Dependable Systems, Services and Technologies (DESSERT)*, Leeds, UK, 2019, pp. 231-234, doi: 10.1109/DESSERT.2019.8770016.

5. K. Rajput, G. Chetty and R. Davey, "Obesity and Co-Morbidity Detection in Clinical Text Using Deep Learning and Machine Learning Techniques," *2018 5th Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE)*, Nadi, Fiji, 2018, pp. 51-56, doi: 10.1109/APWConCSE.2018.00017.

6. M. M. Rahman, R. Amin, M. N. Khan Liton and N. Hossain, "Disha: An Implementation of Machine Learning Based Bangla Healthcare Chatbot," *2019 22nd International Conference on Computer and Information Technology (ICCIT)*, Dhaka, Bangladesh, 2019, pp. 1-6, doi: 10.1109/ICCIT48885.2019.9038579.

7. X. Pang, C. B. Forrest, F. Lê-Scherban and A. J. Masino, "Understanding Early Childhood Obesity via Interpretation of Machine Learning Model Predictions," *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, Boca Raton, FL, USA, 2019, pp. 1438-1443, doi: 10.1109/ICMLA.2019.00235.

8. N. C. Pereira, J. D'souza, P. Rana and S. Solaskar, "Obesity Related Disease Prediction from Healthcare Communities Using Machine Learning," *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kanpur, India, 2019, pp. 1-7, doi: 10.1109/ICCCNT45670.2019.8944798.

9. M. A. Subhi, S. H. Ali and M. A. Mohammed, "Vision-Based Approaches for Automatic Food Recognition and Dietary Assessment: A Survey," in *IEEE Access*, vol. 7, pp. 35370-35381, 2019, doi: 10.1109/ACCESS.2019.2904519.

10. C. Silva, M. Saraee and M. Saraee, "Predictive Modelling in Mental Health: A Data Science Approach," *2019 IEEE Conference on Sustainable Utilization and Development in Engineering and Technologies (CSUDET)*, Penang, Malaysia, 2019, pp. 6-11, doi: 10.1109/CSUDET47057.2019.9214626.

11. Q. Xue, X. Wang, S. Meehan, J. Kuang, J. A. Gao and M. C. Chuah, "Recurrent Neural Networks Based Obesity Status Prediction Using Activity Data," *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Orlando, FL, USA, 2018, pp. 865-870, doi: 10.1109/ICMLA.2018.00139.

12. C. A. C. Montaez, P. Fergus, A. C. Montaez, A. Hussain, D. Al-Jumeily and C. Chalmers, "Deep Learning Classification of Polygenic Obesity using Genome Wide Association Study SNPs," *2018 International Joint Conference on Neural Networks (IJCNN)*, Rio de Janeiro, Brazil, 2018, pp. 1-8, doi: 10.1109/IJCNN.2018.8489048.

13. Z. Zheng and K. Ruggiero, "Using machine learning to predict obesity in high school students," *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Kansas City, MO, USA, 2017, pp. 2132-2138, doi: 10.1109/BIBM.2017.8217988.

14. U. Khalil, O. A. Malik, D. Lai and O. S. King, "Identifying sub-groups of the obese from national health and nutritional status survey data using machine learning techniques," *7th Brunei International Conference on Engineering and Technology 2018 (BICET 2018)*, Bandar Seri Begawan, Brunei, 2018, pp. 1-4.

15. N. S. Rajliwall, G. Chetty and R. Davey, "Chronic disease risk monitoring based on an innovative predictive modelling framework," *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, Honolulu, HI, USA, 2017, pp. 1-8, doi: 10.1109/SSCI.2017.8285257.

16. I. D. Addo, S. I. Ahamed and W. C. Chu, "Toward Collective Intelligence for Fighting Obesity," *2013 IEEE 37th Annual Computer Software and Applications Conference*, Kyoto, Japan, 2013, pp. 690-695, doi: 10.1109/COMPSAC.2013.109.