scientific reports

OPEN

Machine learning-based obesity classification considering 3D body scanner measurements

Seungjin Jeon[1], Minji Kim[1], Jiwun Yoon[2], Sangyong Lee[2] & Sekyoung Youm[1]✉

Obesity can cause various diseases and is a serious health concern. BMI, which is currently the popular measure for judging obesity, does not accurately classify obesity; it reflects the height and weight but ignores the characteristics of an individual's body type. In order to overcome the limitations of classifying obesity using BMI, we considered 3-dimensional (3D) measurements of the human body. The scope of our study was limited to Korean subjects. In order to expand 3D body scan data clinically, 3D body scans, Dual-energy X-ray absorptiometry, and Bioelectrical Impedance Analysis data was collected pairwise for 160 Korean subjects. A machine learning-based obesity classification framework using 3D body scan data was designed, validated through Accuracy, Recall, Precision, and F1 score, and compared with BMI and BIA. In a test dataset of 40 people, BMI had the following values: Accuracy: 0.529, Recall: 0.472, Precision: 0.458, and F1 score: 0.462, while BIA had the following values: Accuracy: 0.752, Recall: 0.742, Precision: 0.751, and F1 score: 0.739. Our proposed model had the following values: Accuracy: 0.800, Recall: 0.767, Precision: 0.842, and F1 score: 0.792. Thus, our accuracy was higher than BMI as well as BIA. Our model can be used for obesity management through 3D body scans.

Grading a patient's obesity level is an important activity in healthcare[1]. Obesity acts as a risk factor for various diseases, including chronic diseases, type-2 diabetes, heart disease, and certain cancers[2–7]. Knowing that you are obese can motivate you to manage your weight[8,9]. In addition, intentional weight management can lower the risk of poor health and is associated with the benefit of a lower risk of disease[10]. However, Body Mass Index (BMI), which represents the criteria for obesity defined by the World Health Organization (WHO), does not accurately classify obesity as it does not sufficiently reflect body-type factors[11–13]. Nutrition varies between regions and body types[14,15]. Reflecting the body-type factor requires anthropometric measurements. However, traditional anthropometric methods are impractical because such measuring requires trained experts. As an alternative, research on using a 3D scanner for human body measurement is being actively conducted[16–27] as it is a less-invasive method than traditional anthropometric measurements. Computed Tomography (CT) or Dual-energy X-ray absorptiometry (DXA), which is the gold standard for measuring human body fat percent (bf%), involve the risk of exposure to radiation when frequent measurements are taken. Unlike CT or DXA, 3D scanners do not expose the human body to radiation[26]. In addition, health risks can be analyzed or predicted through various measures, not just a single one. Therefore, in this study, 3D body scan data and DXA data were collected in pairs for Koreans, and obesity classification was performed using anthropometric values obtained from the 3D body scan data.

Löffler-Wirth et al. classified the body types of residents in Leipzig, Germany using a 3D scanner[17]. Body measurements were extracted from 8499 people through a 3D scanner. After dividing this by height, the body types were grouped through Self-organizing Map (SOM). Clustering body types using a Machine Learning methodology through large-scale experiments is meaningful, but it is not scalable because it cannot be paired with clinical information such as DXA. Pleuss et al.[25] conducted a sampling study to supplement Löffler-Wirth et al. but the number of samples was quite small. In Ng et al. and Bennett et al. 3D scanner data and DXA showed a strong correlation through statistical analysis; the relationship can be explained but it is difficult to estimate bf% through various factors of the human body[16,27]. Some studies used machine learning; Harty et al.[21] developed a new bf% estimation equation based on 3D human body data and 4C model anthropometric data. Using a decision tree model is useful in that important factors and the criteria necessary for formulas can be distinguished. Lu et al.[20] proposed a methodology to predict bf% through a machine learning-based framework after extracting

[1]Department of Industrial and Systems Engineering, Dongguk University, 30 Pildong-ro 1-gil, Jung-gu, Seoul 04620, Republic of Korea. [2]Department of Physical Education, Korea National Sport University, Seoul, Republic of Korea. ✉email: skyoum@dgu.edu

features from 3D body data obtained using a 3D scanner. Although it showed higher accuracy than BMI and BOD POD, the experiment was conducted on relatively few subjects, i.e., 50 men. Although active research is underway on DXA and 3D body scan data, not many machine learning models have been studied.

Most obese people do not perceive themselves as obese[28,29], which may hinder public health initiatives[30]. Knowing one's obesity group can serve as a prerequisite for behavior change for health[31] and will be helpful for health management and disease prevention.

The purpose of our study is to develop a machine learning framework to classify obesity among Koreans based on the bf% of DXA—considered the gold standard—using body measurements extracted from 3D scanners. By selecting input features through a Genetic Algorithm, we not only improve the performance of the Machine Learning model but also observe the selected input features to help in healthcare.

Method

Materials. The collection and use of data used in this study and ethical review were approved by the Institutional Review Board of Korea National Sport University (20220411-021). All methods were carried out in accordance with relevant guidelines and regulations. Informed consent was obtained from all subjects and/or legal guardians regarding including their information/images in an online open-access publication and paper. We confirmed that informed consent was obtained from all subjects (for participation). This experiment faithfully followed the strict regulations and guidelines of the Institutional Review Board. The dataset used in this study was collected between 2022–04–11 and 2022–06–30. Recruitment was based on BMI to have a BMI distribution similar to Size Korea. There were a total of 160 subjects: 73 women and 87 men. As shown in Fig. 1, the men wore tight-fitting bottoms and swimming caps and the women wore tight-fitting tops and bottoms and swimming caps. The subjects were measured using a 3D scanner, DXA, and BIA. Through this, a total of 160 3D body data, DXA data, and BIA data were obtained. Statistics for the collected data are presented in Table 1.

First, we verified whether our 160 data were a fair representation of general Korean anthropometric data. The 8th Anthropometric Data collected between 2020 and 2021 by Size Korea was used for validation[32]. As our main participants were in their 20 s and 30 s, we used the data of 1547 women and 1306 men in 20 s and 30 s from amongst the 8th Anthropometric Data for accurate comparison. In order to verify whether the anthropometric characteristics of Koreans can be viably represented, a Kolmogorov–Smirnov (KS) test was conducted to confirm that the two sample distributions were identical. This test is useful for determining the difference in variance between two samples. In the case of men, the p value of the KS test came out as 0.926, which is valid at the significance level of 0.01. In the case of women, the p value of the KS test was 0.052, which is valid at the significance level of 0.01. And a t-test was conducted to compare the average BMI within the data we collected with the average BMI within the Size Korea data. For men, the t-test result was a p value of 0.5807, at a significance level of 0.01, so we hypothesized that there is no difference between the average BMI of Korean men in 20 s and 30 s and the average BMI of men we collected. Similarly, for women, the t-test result was a p value of 0.0532, and at a significance level of 0.01, so we hypothesized that there is no difference between the average BMI of Korean
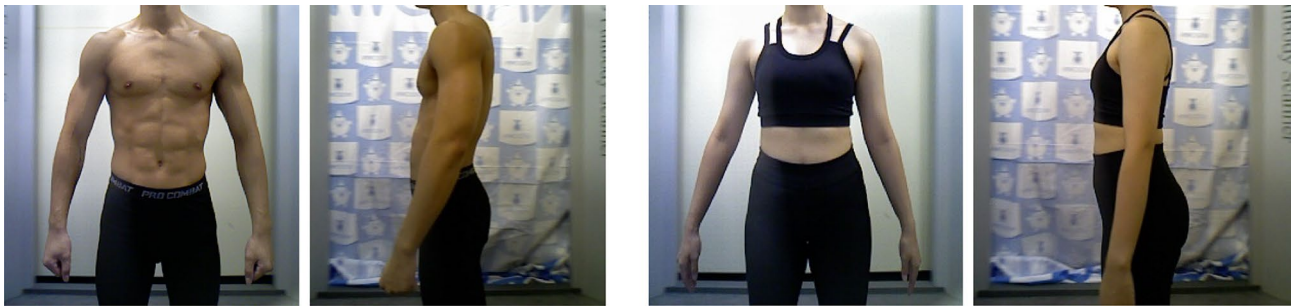

Figure 1. Subjects' attire and posture.

Gender	Statistics	Age	Height (cm)	Weight (kg)	BMI (kg/m²)	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.

2

women in 20 s and 30 s and the average BMI of women we collected. Thus, we verified that the data we collected was representative of Koreans in 20 s and 30 s.

3D scanner.   The 3D scanner was a PFS-304A model (PMT innovation company, Gyeonggi-do, Korea, PFS software ver. 1.3). Figure 2a shows the 3D scanner used in this study, wherein the camera module rotates 360° through a motor mounted on the top when it scans and subjects can be measured in a stationary position. The subjects' postures were measured by taking the A-pose recommended by ISO-7250[33]. If the A-pose is not taken (e.g., anthropometry is not performed when the arm is close to or attached to the body), a blind spot is formed and the correct mesh shape is not created. As shown in Fig. 2a, the measurements of the 3D scanner were taken in an indoor lighting environment, and they are summarized in Table 2.

Dual-energy X-ray absorptiometry (DXA).   Lunar (GE Healthcare, Madison Wisconsin, America, EnCore software ver. 13.60.03) in Fig. 2b was used for DXA. DXA is a body component-measuring instrument that measures body fat, lean body mass, and bone mass and has long been regarded as the gold standard for measuring body components. The DXA device was handled and measured by an expert; during measurement, the subject maintained an immobile posture while lying down, and the following body parts were examined: arm (left, right), leg (left, right), trunk (left, right), Android, and Gynoid. BMD, BMC, fat%, fat(g), and lean(g) were obtained through DXA for each body part. Here, BMD represents bone density and BMC represents bone mineral mass. Fat% represents fat percentage, fat (g) represents fat weight, and lean (g) represents weight excluding fat. This study classified obesity in individuals based on the total body fat% (bf%) according to DXA.

Bioelectrical impedance analysis (BIA).   BIA is a method of measuring body composition through the impedance difference between body fat and lean body mass[34]. As shown in Fig. 2c, Inbody770 was used as the measuring device and the subjects were measured in the same clothing and immobile posture as in the previous experiments. Total Body Water (TBW), IntraCellular Water (ICW), ExtraCellular Water (ECW), protein, minerals, Body Fat percentage (bf%), and Fat-free Mass (FFM) were obtained through BIA. Among them, bf% was used for comparison using the methodology proposed in this study. The obesity group of BIA was divided using the bf% criteria presented in Table 3.

Data preprocessing.   Data obtained through DXA and BIA only used bf%. Based on the bf% obtained from the DXA of the test subjects in the experiment, the obesity class label was derived as per the cutoff in Table 3. The bf% standard cutoff for Obesity was set in accordance with the WHO[35] classification, and the standard cutoffs from Lobman et al. were used for the rest, namely underweight, normal, and overweight[36]. Currently, there is no clear obesity category for DXA[37], but the cutoff for obesity was based on 25%(bf%) for men and 35%(bf%) for women by the Korean Society for Obesity, and the rest of the groups were classified by reference to McArdle and Chang[38,39]. Therefore, subjects were labeled into four groups: underweight, normal, overweight, and obese. The distribution results of labeling based on bf% measured by DXA Table 3 are depicted in Table 4. As shown in Fig. 3, the data obtained from the 3D scanner extracted body measurements from the mesh data based on five landmarks: the back of the neck, the umbilicus, the groin, and the armpit (left, right).

Framework.   We propose a machine learning-based methodology to classify obesity groups. Figure 4 shows the proposed machine learning-based framework. Body measurements are first obtained from the 3D scanner, and then data preprocessing is performed by matching it with the bf% of DXA and labeling it. After that, the final model is selected through the process of "Choose ML model" and "Feature selection Genetic Algorithm."
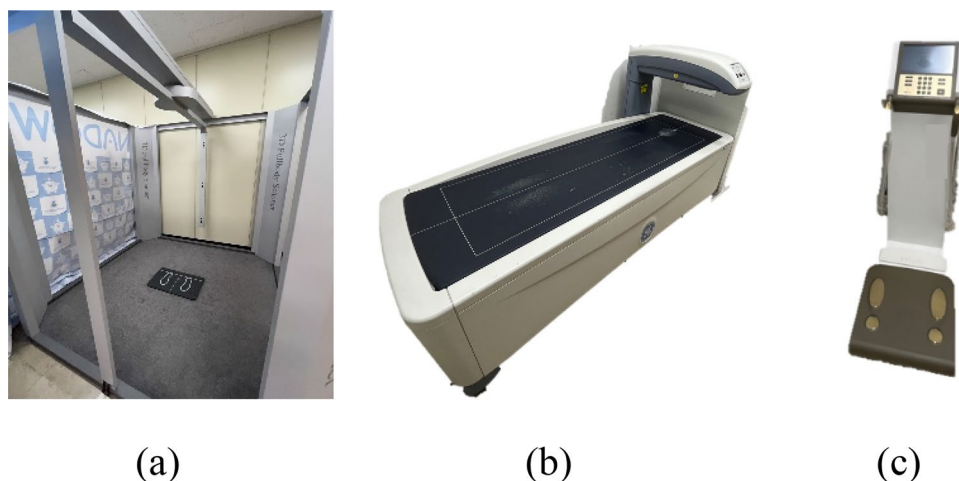

(a)   (b)   (c)

Figure 2. Equipment used in the experiment. (**a**) 3D scanner (PFS-304 of PMT), (**b**) DXA (Lunar of GE), (**c**) BIA (Inbody770).

Value	description	Value	description
WCR	Waist-to-chest ratio	Thick thing cir	Circumference of the thickest part of the innermost thigh
WHR	Waist-to-hip ratio	Middle thigh cir	Circumference of the middle of the thigh
WHtR	Waist-to-height ratio	Knee cir	Horizontal circumference through the midpoint of the kneecap
THR	Thigh-to-height ratio	Calf cir	Circumference of the most convex part of the calf
Height of the back of the neck	Height from floor to the back of the neck in a standing position	Arm cir	Circumference of the thickest part of the arm
Shoulder height	Height from floor to the shoulders in a standing position	CS. area of the back of the neck	Cross-sectional area of the back of the neck
Chest height	In a standing position, the height from floor to the top of the nipples and the beginning of the chest	Shoulder CS	Cross-sectional area of the shoulder
(breast) Chest height	Height from floor to nipple in standing position	Chest area	Cross-sectional area of the part passing through the midpoint of the sternum
Waist height	Height from floor to the point in front of the waist in a standing position	(breast) Chest area	Cross-sectional area of the part passing through the nipple point
Navel height	Height from the standing position to the navel	Waist CS	Cross-sectional area of the part passing through the front of the waist, the side of the waist, and the back of the waist
Height below the navel	The height from the standing position to the iliac bone below the navel	Navel waist CS	Cross-sectional area of the part passing through navel point, navel level, waist point, navel level, back point
Hip height	Height from floor to hip protrusion in standing position	Area below the navel	Cross-sectional area of the part that passes through iliac crest below navel
Groin height	Vertical height from floor to groin (actual leg length)	Hip CS	Cross-sectional area of the part passing through the buttock protrusion
Thick thigh height	Height from floor to the thickest part of the thigh	Groin area	Cross-sectional area of the groin
Midthigh height	Height from floor to mid-thigh	Thick thigh area	Cross-sectional area of the thickest part of the inner-most thigh
Knee height	Vertical height from floor to the top of the shinbone	Meddle thigh area	Cross-sectional area of the middle part of the thigh
Calf height	Height from floor to the point of the thickest part of the calf	Knee CS	Cross-sectional area passing through the midpoint of the kneecap
Neck cir	Circumference passing under the back of neck and under the shield cartilage	Calf CS	Cross-sectional area of the most convex part of the calf
Shoulder cir	Circumference from the end of the shoulder to the end of the shoulder opposite the back of the neck	Total volume	Volume of total body
Cir. of the chest	Horizontal circumference through the midpoint of the sternum	Shoulder volume	Volume of shoulder
(breast)Chest cir	Horizontal circumference through nipple point	Chest volume	Volume of chest
Waist cir	Horizontal circumference passing through the point in front of the waist, the point in the side of the waist, and the point in the back of the waist	Epigastric volume	Volume of epigastrium
Navel waist cir	Horizontal circumference passing through navel point, navel level, waist point, navel level, back point	Lower abdominal volume	Volume of lower abdomen
Below navel cir	Horizontal circumference through the iliac crest below the navel	Thigh volume	Volume of thigh
Hip cir	Horizontal circumference through the buttock protrusion	Calf volume	Volume of calf
Groin cir	Groin circumference	Abdominal volumetric	Volume of abdominal

Table 2. Measurement of 3D mesh data.

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 3. Obesity class standard cutoff.

Choose ML model. In data preprocessing, the data in the state of finished preprocessing the 3D body measurements and Sex are used as input values for the Logistic Regression[40], Decision Tree[41], Random Forest[42], Support Vector Machine (SVM)[43], Gradient Boosting[44], and AdaBoost[45] and fivefold cross validation is performed. It is divided into 120 training sets and 40 test sets. Among these models, Accuracy, F1, Recall, and Precision values were compared, and a model with good performance was selected and used as a classifier in the "Feature

Train set			
Class	Male	Female	Total
Underweight	8	4	12
Normal	23	23	46
Overweight	12	12	24
Obese	25	13	38
Test set			
Class	Male	Female	Total
Underweight	3	3	6
Normal	6	8	14
Overweight	5	6	11
Obese	5	4	9

Table 4. Subject class distribution.
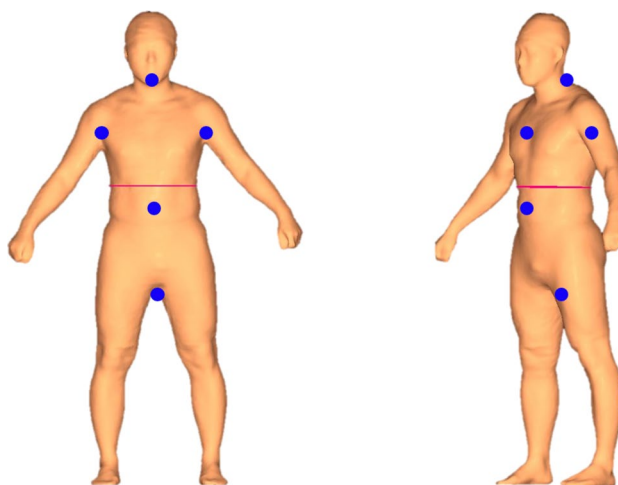
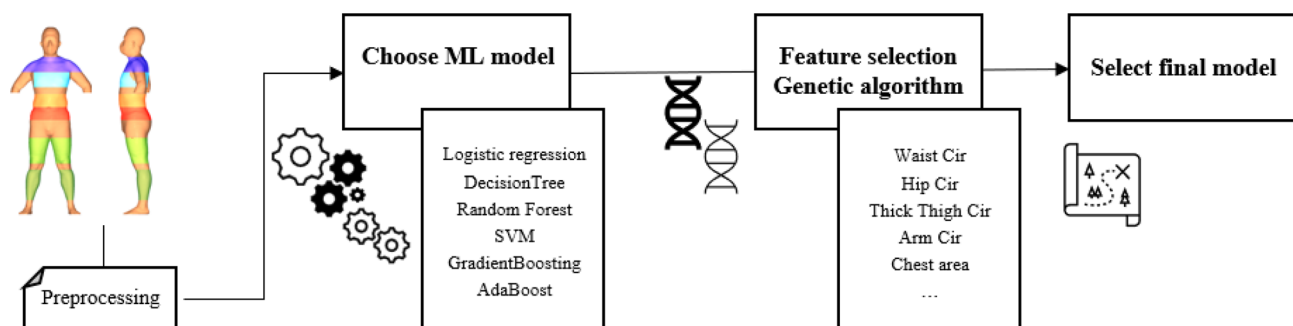**Figure 3.** Sample of the 3D mesh data and standard landmarks for measurement.

Figure 4. Overall framework of this study.

selection Genetic Algorithm" process. In Table 4, as the quantity of data is small and imbalances exist, the model is selected by referring to Precision, Recall, and F1 score values rather than simply using Accuracy. Accuracy is the ratio of correctly predicted numbers to the total number. Precision is the sum of true positives and false positives and the ratio of true positives. Recall is the sum of true positives and false negatives and the ratio of true positives. F1 score is the harmonic mean of Precision and Recall.

Logistic regression predicts the probability of occurrence by using a linear combination between variables of input data, and the result is classified into a specific class. This study used multiclass logistic regression with a cross-entropy function. Decision Tree is the most preferred machine learning model as an explanatory model. It outputs a class in which input data is classified based on input variables through a tree structure; it is a way to perform a query on a node and branching out. Its performance is not as good as other models, and it is vulnerable

to overfitting. Random Forest is a machine learning model that uses a bundle of basic decision trees and averages them to compensate for performance. Through this, the performance can be generalized and made more robust against overfitting compared to a decision tree. Support Vector Machine determines the hyperplane to maximize the margin between support vectors; its purpose is to maximize the distance between various classes and to find a hyperplane that has a large difference from the training data to which the hyperplane is closest. The main idea of Gradient Boosting is to connect multiple non-deep decision trees, that is, weak learners. As basic trees can classify some data well, performance improves when trees are added. The loss function is defined and gradient descent is used to supplement the value to be classified by the next tree. AdaBoost stands for Adaptive Boosting. Unlike Gradient Boosting, this model is trained by adding weights to the classified samples. At this time, the learning model is created by adding weights to the next model in the sample that is poorly classified.

Feature selection Genetic Algorithm.   This process selects the input features of the previously selected machine learning model through a Genetic Algorithm. Selectively choosing the input features of the machine learning model not only improves the model's performance but also identifies whether a specific value among the 3D body measurements in Table 2 affects the classification of obesity.

While selecting input features, finding the Global Optimum by comparing all sets of input features combinations is practically impossible. Therefore, a meta-heuristic algorithm approach was chosen to find an optimal solution close enough to the Global Optimum. Previous studies have demonstrated that the Genetic Algorithm is superior to other meta-heuristic algorithms in variable selection[46,47]. In this study, the Genetic Algorithm (GA) was used as a feature selection method. GA takes a meta-heuristic approach to solving complex problems through efficient trial and error[48], hence mimicking Charles Darwin's theory of natural selection and mammalian reproduction. In this study, GA aims to find the best input feature through repeated generation reproduction. GA involves six steps. In Step 1, it initializes the combination of chromosomes, i.e., the initial input features, and sets the parameters. These parameters include population and mutation ratio, where population refers to the number of chromosomes in each generation, i.e., the number of combinations of input features. The mutation ratio refers to the ratio of gene mutations among all chromosomes; this corresponds to the ratio of selection of input features. We set the population to 100 and the mutation ratio to 20%. Step 2 involves learning each input feature in a Random Forest. In Step 3, fitness was evaluated for the chromosomes of each input feature and the fitness function was used to determine the accuracy. In Step 4, out of the current generation and current chromosomes, we selected excellent chromosomes with Accuracy. In this study, the top 80% were selected as excellent chromosomes. Step 5 involved generating next-generation chromosomes through crossover and mutation. In this case, crossover means mixing the selected adoptive parent chromosomes in half. We set this to stop when the 100th generation was passed, and until then, it was set to return to Step 2 and repeat all intervening steps. In Step 6, we selected the final model, picking the model that generated the highest Accuracy.

A total of 100 generations were generated and the input feature of the generation with the highest Accuracy, Recall, Precision, and F1 score was selected. Among the 100 generations, the generation with Accuracy, Recall, Precision, and F1 score of 0.8, 0.767, 0.842, and 0.792 was the highest and the corresponding input feature was selected as the final input feature. Accuracy reached 80% in the 50th epoch, after which it converged or even decreased. Figure 5 shows the flow of accuracy by generation. Table 5 presents the final selected features.

Results

Accuracy, Recall, Precision, and F1 score were calculated using DXA as the Ground Truth for the reference group classification. BIA classified obesity based on the bf% obtained through Inbody770, and BMI classified obesity according to the WHO Asian standard cutoff[13]. The cutoff for BIA and BMI is listed in Table 3.

Table 6 shows the results of classification using the above models; among them, as its Accuracy, Recall, Precision, and F1 score were 0.725, 0.692, 0.661, and 0.78, i.e., all higher than those of other models, Random Forest was selected as the machine learning model in the process.
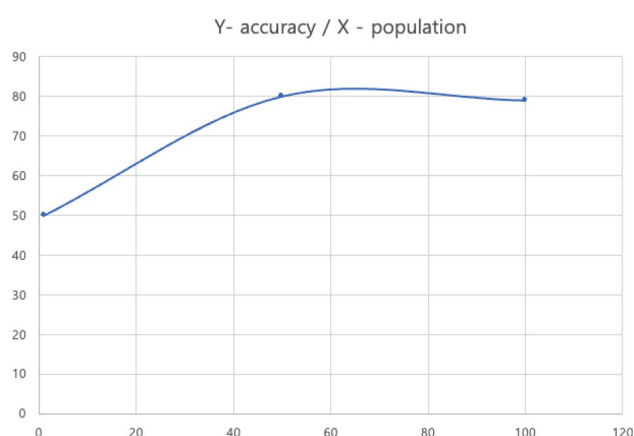

Figure 5. Accuracy flowchart by generation.

Value	Value	Value
WCR	Hip cir	Hip CS
Groin height	Thick thing cir	Groin area
Thick thigh height	Middle thigh cir	Thick thigh area
Midthigh height	Arm cir	Calf CS
Knee height	Shoulder CS	Total volume
Calf height	Chest area	Shoulder volume
Shoulder cir	Navel waist CS	Thigh volume
(breast) Chest cir	Area below the navel	Sex

Table 5. Selected features of the "Feature selection Genetic Algorithm".

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 6. Results of the choosing the ML model. Significant values are in bold.

The performance of the proposed approach is better than that using BMI. In Table 7, there is a difference of 0.271 in Accuracy from 0.529 to 0.8, 0.295 in Recall, 0.384 in Precision, and 0.33 in F1 score. This means that obesity can be classified more comprehensively by reflecting the various dimensions of the human body considering the body type rather than just the BMI, which classifies obesity through simple height and weight. Although BIA is widely used in body composition studies, concerns about its accuracy still exist[49,50]. As BIA assumes that the percentage of body water is approximately 73% and estimates it accordingly, low accuracy may ensue when the percentage of body water of an individual does not meet these conditions[51]. Furthermore, depending on the statistical model derived from a specific population, BIA may have differences in gender, age, ethnicity, and so on[50]. The Pearson's correlation between bf% of DXA and bf% of BIA in our collected data was 0.95. When evaluated after the obesity classification presented in Table 3, there was a misclassification. BIA scored 0.752, 0.742, 0.751, and 0.739 in Accuracy, Recall, Precision, and F1 scores, respectively. The proposed obesity classification showed an improvement of 0.048, 0.025, 0.091, and 0.053 in Accuracy, Recall, Precision, and F1 scores, respectively, compared to BIA. This shows that the Accuracy, Recall, Precision, and F1 score are 0.075, 0.1, 0.106, and 0.062 higher, respectively, than in the Random Forest model without feature selection. The selected features affect the classification of obesity more than the unselected features. We also compared with the model provided by Tian et al.[52]. Denotes, Tian's model is a regression model, not a classification model, and we obtained results by dividing it by the criteria of the Table 3. The R^2 and RMSE values of Tian's model were 0.56 and 5.74, respectively.

Conclusion

We collected 3D body scans, DXA, and BIA data pairwise for Korean subjects and used this data to classify obesity in individuals. By using not only 3D body data but also DXA and BIA, we developed a technique for clinically clear obesity judgment that is expandable in terms of healthcare.

This study proposes a methodology for classifying obesity using various body measurements through a 3D scanner, unlike the BMI, which classifies obesity solely based on height and weight. The present study specifically

Classifier	Accuracy	Recall	Precision	F1 score	R²/RMSE
Ours	**0.800**	**0.767**	**0.842**	**0.792**	
w/o GA	0.725	0.661	0.780	0.692	
BIA	0.752	0.742	0.751	0.739	
BMI	0.529	0.472	0.458	0.462	
Tian et.al	0.472	0.379	0.448	0.321	0.56/5.74

Table 7. Comparison between our method, BMI, BIA and Tian et al. within the test set. *w/o* without, *GA* genetic algorithm. Significant values are in bold.

considers Korean body types. The proposed methodology showed better performance in classifying obesity than the BMI through the machine learning methodology. It also showed better performance than BIA.

Pleuss et al. and Harty et al. conducted analysis using one or two machine learning models, but in this study, six machine learning models were compared to select a model suitable for obesity classification[21,25]. We performed feature selection through a Genetic Algorithm to identify the measurements of the human body that have an impact on determining obesity.

The proposed system showed superior performance over BMI and BIA. It can be used for long-term obesity healthcare monitoring by measuring one's body with a 3D scanner. Furthermore, as future healthcare is predicted to be Predictive, Preventive, Personalized, Participatory (4P)[53], a system that classifies obesity as an indicator of health can be utilized as a new healthcare service that satisfies the 4P.

Our study has certain limitations. First, there are limits on data. The collected data included males and females in their 20 s and 30 s. As such, it is not common to all age groups. It does not reflect the adult group over 40 years of age. Furthermore, obesity was not evenly distributed, and there was a limited amount of data. Given that the measurement values were extracted and used from the mesh data generated by the 3D human body scanner, it is difficult to ensure that the 3D information was fully used. Second, there are spatial and cost limitations associated with 3D scanners that make them unpopular. Third, it provides users with categorical information rather than continuous information. It does not reflect the continuous variation of human.

We intend to collect continuous pair data as future work and conduct experiments in various age groups as well as compare between men and women in their 20 s and 30 s. Furthermore, we intend to conduct experiments on patients with specific diseases or disorders. By referring to previous studies[54], we intend to consider using a three-dimensional representation such as curvature. In the Genetic Algorithm selection, selection methods such as elitism and tournament can be introduced to reduce computation complexity and optimize future research[55,56]. In order to reflect the continuous variability of human beings, we intend to introduce the regression methodology as a future study. We would like to observe the consistency of data obtained from different scanner devices and DXA as future research. We need to confirm this and do future work to validate the data and other models.

Data availability

The datasets generated and/or analyzed during the current study are not publicly available because they contain personal information but are available from the corresponding author upon reasonable request.

References

1. Zelenytė, V. et al. Body size perception, knowledge about obesity and factors associated with lifestyle change among patients, health care professionals and public health experts. *BMC Fam. Pract.* **22**, 1–13 (2021).
2. Gade, W., Schmit, J., Collins, M. & Gade, J. Beyond obesity: The diagnosis and pathophysiology of metabolic syndrome. *Am. Soc. Clin. Lab. Sci.* **23**, 51–61 (2010).
3. Donohoe, C., Pidgeon, G., Lysaght, J. & Reynolds, J. Obesity and gastrointestinal cancer. *Br. J. Surg.* **97**, 628–642 (2010).
4. Després, J.-P. Body fat distribution and risk of cardiovascular disease: An update. *Circulation* **126**, 1301–1313 (2012).
5. Krauss, R. M., Winston, M., Fletcher, B. J. & Grundy, S. M. Obesity: Impact on cardiovascular disease. *Circulation* **98**, 1472–1476 (1998).
6. Ortega, F. B., Lavie, C. J. & Blair, S. N. Obesity and cardiovascular disease. *Circ. Res.* **118**, 1752–1770 (2016).
7. Grundy, S. M. Obesity, metabolic syndrome, and cardiovascular disease. *J. Clin. Endocrinol. Metab.* **89**, 2595–2600 (2004).
8. Burkbauer, L. et al. Awareness of link between obesity and breast cancer risk is associated with willingness to participate in weight loss intervention. *Breast Cancer Res. Treat.* **194.3**, 541–550 (2022).
9. Alick, C. L., et al. Motivating weight loss among black adults in relationships: Recommendations for weight loss interventions. *Health Educ. Behav.* 10901981221129182 (2022).
10. Haase, C. L. et al. Weight loss and risk reduction of obesity-related outcomes in 0.5 million people: Evidence from a UK primary care database. *Int. J. Obes.* **45**(6), 1249–1258 (2021).
11. Bergman, R. N. et al. A better index of body adiposity. *Obesity* **19**, 1083–1089 (2011).
12. Belarmino, G. et al. A new anthropometric index for body fat estimation in patients with severe obesity. *BMC Obesity* **5**, 1–8 (2018).
13. World Health Organization. The Asia-Pacific perspective: redefining obesity and its treatment (2000).
14. Visser, M. et al. One-and two-year change in body composition as measured by DXA in a population-based cohort of older men and women. *J. Appl. Physiol.* **94**, 2368–2374 (2003).
15. Gallagher, D. et al. Appendicular skeletal muscle mass: Effects of age, gender, and ethnicity. *J. Appl. Physiol.* **83**, 229–239 (1997).
16. Bennett, J. P. et al. Assessment of clinical measures of total and regional body composition from a commercial 3-dimensional optical body scanner. *Clin. Nutr.* **41**, 211–218 (2022).
17. Löffler-Wirth, H. et al. Novel anthropometry based on 3D-bodyscans applied to a large population based cohort. *PLoS ONE* **11**, e0159887 (2016).
18. Ng, B. K. et al. Detailed 3-dimensional body shape features predict body composition, blood metabolites, and functional strength: the Shape Up! studies. *Am. J. Clin. Nutr.* **110**, 1316–1326 (2019).
19. Wong, M. C. et al. Children and adolescents' anthropometrics body composition from 3-D optical surface scans. *Obesity* **27**, 1738–1749 (2019).
20. Lu, Y., McQuade, S. & Hahn, J. K. in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 3999–4002 (IEEE).
21. Harty, P. S. et al. Novel body fat estimation using machine learning and 3-dimensional optical imaging. *Eur. J. Clin. Nutr.* **74**, 842–845 (2020).
22. Heymsfield, S. B. et al. Digital anthropometry: A critical review. *Eur. J. Clin. Nutr.* **72**, 680–687 (2018).
23. Kennedy, S. et al. Digital anthropometric evaluation of young children: Comparison to results acquired with conventional anthropometry. *Eur. J. Clin. Nutr.* **76**, 251–260 (2022).
24. Medina-Inojosa, J., Somers, V. K., Ngwa, T., Hinshaw, L. & Lopez-Jimenez, F. Reliability of a 3D body scanner for anthropometric measurements of central obesity. *Obesity, open access* **2** (2016).

25. Pleuss, J. D. et al. A machine learning approach relating 3D body scans to body composition in humans. Eur. J. Clin. Nutr. **73**, 200–208 (2019).
26. Wang, Q., Lu, Y., Zhang, X. & Hahn, J. K. in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 1729–1732 (IEEE).
27. Ng, B., Hinton, B., Fan, B., Kanaya, A. & Shepherd, J. Clinical anthropometrics and body composition from 3D whole-body surface scans. Eur. J. Clin. Nutr. **70**, 1265–1270 (2016).
28. Johnson, F., Beeken, R. J., Croker, H. & Wardle, J. Do weight perceptions among obese adults in Great Britain match clinical definitions? Analysis of cross-sectional surveys from 2007 and 2012. BMJ Open https://doi.org/10.1136/bmjopen-2014-005561 (2014).
29. Robinson, E. & Oldham, M. Weight status misperceptions among UK adults: The use of self-reported vs. measured BMI. BMC Obes. https://doi.org/10.1186/s40608-016-0102-8 (2016).
30. Ikeda, J. P., Crawford, P. B. & Woodward-Lopez, G. BMI screening in schools: Helpful or harmful. Health Educ. Res. **21**, 761–769 (2006).
31. Duncan, D. T. et al. Does perception equal reality? Weight misperception in relation to weight-related attitudes and behaviors among overweight and obese US adults. Int. J. Behav. Nutr. Phys. Act. **8**, 20. https://doi.org/10.1186/1479-5868-8-20 (2011).
32. Anthropometric report. Size Korea Available at: https://sizekorea.kr/human-info/meas-report?measDegree=8. (Accessed: 14th November 2022)
33. Standard, B. & ISO, B. Basic human body measurements for technological design.
34. AbuKhaled, M. et al. Electrical impedance in assessing human body composition: The BIA method. Am. J. Clin. Nutr. **47**, 789–792 (1988).
35. Committee, W. E. Physical status: The use and interpretation of anthropometry. http://helid.digicollection.Org/en/d/Jh0211e/ (1995).
36. Lohman, T., Wang, Z. & Going, S. B. *Human Body Composition* Vol. 918 (Human Kinetics, 2005).
37. Physical status: the use and interpretation of anthropometry. Report of a WHO expert committee. *World Health Organ Tech Rep Ser* **854**, 1–452 (1995).
38. McArdle, W. D., Katch, F. I. & Katch, V. L. *Essentials of Exercise Physiology* 4th edn. (Lippincott Williams Wilkins, 2006).
39. Chang, H.-S. A study of nutrient intakes, blood lipids and bone mineral density according to obesity degree by percentage of body fat and age between male and female teacher in Jeonbuk Province, Korea. *Korean J. Commun. Nutr.* **17**(1), 49–68 (2012).
40. Cox, D. R. The regression analysis of binary sequences. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **20**, 215–232 (1958).
41. Safavian, S. R. & Landgrebe, D. A survey of decision tree classifier methodology. *IEEE Trans. Syst. Man Cybern. Syst.* **21**, 660–674 (1991).
42. Breiman, L. Bagging predictors. *Mach. Learn.* **24**, 123–140 (1996).
43. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn.* **20**, 273–297 (1995).
44. Mason, L., Baxter, J., Bartlett, P. & Frean, M. Boosting algorithms as gradient descent. *Adv. Neural Inf. Process. Syst.* **12** (1999).
45. Freund, Y. & Schapire, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **55**, 119–139 (1997).
46. Raymer, M. L., Punch, W. F., Goodman, E. D., Kuhn, L. A. & Jain, A. K. Dimensionality reduction using genetic algorithms. *IEEE Trans. Evol. Comput.* **4**, 164–171 (2000).
47. Siedlecki, W. O. J. C. I. E. C. H. & Sklansky, J. A. C. K. On automatic feature selection. *Int. J. Pattern Recognit. Artif. Intell.* **2**, 197–220 (1988).
48. Holland, J. H. Genetic algorithms. *Sci. Am.* **267**, 66–73 (1992).
49. Achamrah, N. et al. Comparison of body composition assessment by DXA and BIA according to the body mass index: A retrospective study on 3655 measures. *PLoS ONE* **13**, e0200465 (2018).
50. Kim, H. et al. External cross-validation of bioelectrical impedance analysis for the assessment of body composition in Korean adults. *Nurs. Res. Pract.* **5**, 246 (2011).
51. Segal, K. R., Gutin, B., Presta, E., Wang, J. & Van Itallie, T. B. Estimation of human body composition by electrical impedance methods: A comparative study. *J. Appl. Physiol.* **58**, 1565–1571 (1985).
52. Tian, I. Y. et al. A device-agnostic shape model for automated body composition estimates from 3D optical scans. *Med. Phys.* **49**(10), 6395–6409 (2022).
53. Sagner, M. et al. The P4 health spectrum–a predictive, preventive, personalized and participatory continuum for promoting healthspan. *Prog. Cardiovasc. Dis.* **59**, 506–521 (2017).
54. Wang, Q., Lu, Y., Zhang, X. & Hahn, J. Region of interest selection for functional features. *Neurocomputing* **422**, 235–244 (2021).
55. Du, H., Wang, Z., Zhan, W. & Guo, J. Elitism and distance strategy for selection of evolutionary algorithms. *IEEE Access* **6**, 44531–44541 (2018).
56. Katoch, S., Chauhan, S. S. & Kumar, V. A review on genetic algorithm: past, present, and future. *Multimed. Tools Appl.* **80**, 8091–8126 (2021).

Acknowledgements

Author contributions

S.J. contributed to methodology, analysis, data acquisition and critically revised the manuscript. S.Y. contributed to methodology, review, and editing. M.K. contributed to the conception and review. J.Y contributed to the review and editing. S.L contributed to the review and editing. All authors have given their approval and agree with the published version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.