Linear Regression and K Nearest Neighbors Machine Learning Models for Person Fat Forecasting

Alshaimaa A. Tantawy[1*]

[1] Faculty of Computers and Informatics, Zagazig University, Sharqiyah, Egypt

Email: AlshaimaaTantawy@zu.edu.eg

Abstract

Predicting a person's person fat percentage is an important part of keeping tabs on their health and fitness. An accurate assessment of person fat allows for the development of individualized programmer for health and wellbeing, the promotion of illness prevention, and the evaluation of the efficacy of weight management initiatives. This study reviews the current state of the art in person fat prediction approaches, which includes the use of machine learning algorithms. Obesity is a chronic condition characterized by high levels of person fat and is linked to several health issues. Since several methods exist for estimating person fat percentage to evaluate obesity, these assessments are usually expensive and need specialized equipment. Therefore, determining obesity and its associated disorders requires an accurate estimate of person fat proportion according to readily available person measures. This paper presented a machine-learning model for forecasting person fat. This problem is a regression, so this paper used two regression models to deal with the regression dataset. This paper used linear regression (LR) and k nearest neighbors (KNN). The two models were applied to real datasets. The dataset has 252 records. The results showed the LR has the highest score than the KNN model.

Keywords: Machine Learning; Linear Regression; K Nearest Neighbors; person Fat; Prediction; Regression Problem.

1. Introduction

Over the past couple of centuries, obesity has become a worldwide epidemic. According to the most up-to-date assessments of global obesity rates, a minimum of 30% of men and 35% of women are obese in a number of regions throughout the globe[1], [2]. The percentage of obese adults in the United States has increased to 39.6%, from 33.7% in 2007-2008.5 Men have a 27% obesity rate, and women have a 25% obesity rate in Canada. Obesity rates have more than doubled since 1980 in more than 70 nations, and they are rising in the majority of the world's regions. From 1980 to 2013, the percentage of people having an internal mass index (BMI) of above 25 kg/m2 rose from 29 percent to 37 percent in males and from 30 percent to 38 percent in women[3]–[6].

The use of special equipment, such as underwater weighing, and dual-energy X-ray, is necessary for accurate and immediate estimation of person fat. This highlights the need for an accurate system that can forecast person fat with little effort and expense. The BMI is a common, low-cost, and easy-to-use tool for gauging obesity[7], [8]. The BMI does not, however, differentiate between fat and lean mass; it only measures nutritional condition. Person fat, on the other hand, may show the true physical status and is a more accurate indicator of obesity[9]–[12].

Doi: https://doi.org/10.54216/IJAACI.030204
Received: August 22, 2022 Revised: December 21, 2022 Accepted: February 19, 2023

Multiple studies have built forecasting models by considering fat mass as a regression issue and using readily observable characteristics like age, weight, length, and physical circumference.

To address the issue of over-fitting, the author in [13] proposed a weighted TSVR. They partition the plane into thirds and then penalize samples differently based on where they fall inside those thirds. In particular, they imposed harsher punishments on data found in Area 3 and less severe ones on those found in Region 2. Last but not least, the weighted TSVR produces less error in forecasting than SVR and TSVR. Using nine standard data sets as benchmarks, they demonstrated that our weighted TSVR significantly outperforms SVR and just slightly lags behind TSVR in both the linear and nonlinear cases.

In the article [14], an ANN-based software solution was proposed for forecasting BF%. There are a total of 2755 participants used in the training, verification, and testing of the ANNs, with ages ranging from 18 to 88 years and person mass indexes (BMIs) ranging from 16.60 to 64.60.

This paper proposed two machine learning models for person fat prediction. The two machine learning models are linear regression (LR), and K nearest neighbors (KNN).

2. Person Fat

Predicting a person's person fat percentage is crucial for gauging their health and fitness level. As a key indicator of general health, chronic disease risk, and weight control programmer efficacy, person fat percentage is an important metric in person composition analysis. In order to customize individual health regimens, monitor progress over time, and evaluate the efficacy of lifestyle treatments, an accurate measurement of person fat percentage is helpful.

Simple anthropometric measures all the way up to high-tech imaging techniques and machine learning algorithms have been used to make predictions about person fat. Skinfold thickness measures and other anthropometric measurements have been used for a long time as reliable and inexpensive ways to estimate person fat percentage. Better and more accurate person composition measurement may be obtained by the use of imaging methods including dual-energy X-ray absorptiometry (DXA), magnetic resonance imaging (MRI), and computed tomography (CT). Predicting person fat percentage from a variety of anthropometric and other characteristics is a common use of machine learning algorithms as of late.

However, there are benefits and drawbacks to every approach. Although anthropometric measures are easy to take, their precision and consistency may be suspect. However, imaging methods are expensive and sometimes need specialized hardware to provide accurate findings. The predictive power of machine learning algorithms might be greatly improved with more extensive training data and thorough validation.

Person fat distribution and estimate are affected by a number of characteristics, including age, gender, and ethnicity. To guarantee precise and unique outcomes, these considerations must be included into prediction models. Different cultures and ethnic groups have unique person compositions, which should be taken into consideration by person fat prediction methods.

3. Material and Methods

This section is divided into two sub-sections, the first sub-section presented the dataset and the second presented the forecasting algorithms.

3.1 Dataset

The dataset of person fat is collected from Kaggle. Table 1 shows the sample of dataset. The dataset has 252 records and 14 features. The 14 features of dataset are shown in first row in Table 1. Figures 1-4 show the distribution of features in dataset.

Doi: https://doi.org/10.54216/IJAACI.030204
Received: August 22, 2022 Revised: December 21, 2022 Accepted: February 19, 2023

Table 1: The sample of person fat dataset

	Density	Person Fat	Age	Weight	Height	Neck	Chest	Abdomen	Hip	Thigh	Knee	Ankle	Biceps	Forearm
BFP0	1.0708	12.3	23	154.25	67.75	36.2	93.1	85.2	94.5	59	37.3	21.9	32	27.4
BFP1	1.0853	6.1	22	173.25	72.25	38.5	93.6	83	98.7	58.7	37.3	23.4	30.5	28.9
BFP2	1.0414	25.3	22	154	66.25	34	95.8	87.9	99.2	59.6	38.9	24	28.8	25.2
BFP3	1.0751	10.4	26	184.75	72.25	37.4	101.8	86.4	101.2	60.1	37.3	22.8	32.4	29.4
BFP4	1.034	28.7	24	184.25	71.25	34.4	97.3	100	101.9	63.2	42.2	24	32.2	27.7
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
BFP247	1.0736	11	70	134.25	67	34.9	89.2	83.6	88.8	49.6	34.8	21.5	25.6	25.7
BFP248	1.0236	33.6	72	201	69.75	40.9	108.5	105	104.5	59.6	40.8	23.2	35.2	28.6
BFP249	1.0328	29.3	72	186.75	66	38.9	111.1	111.5	101.7	60.3	37.3	21.5	31.3	27.2
BFP250	1.0399	26	72	190.75	70.5	38.9	108.3	101.3	97.8	56	41.6	22.7	30.5	29.4

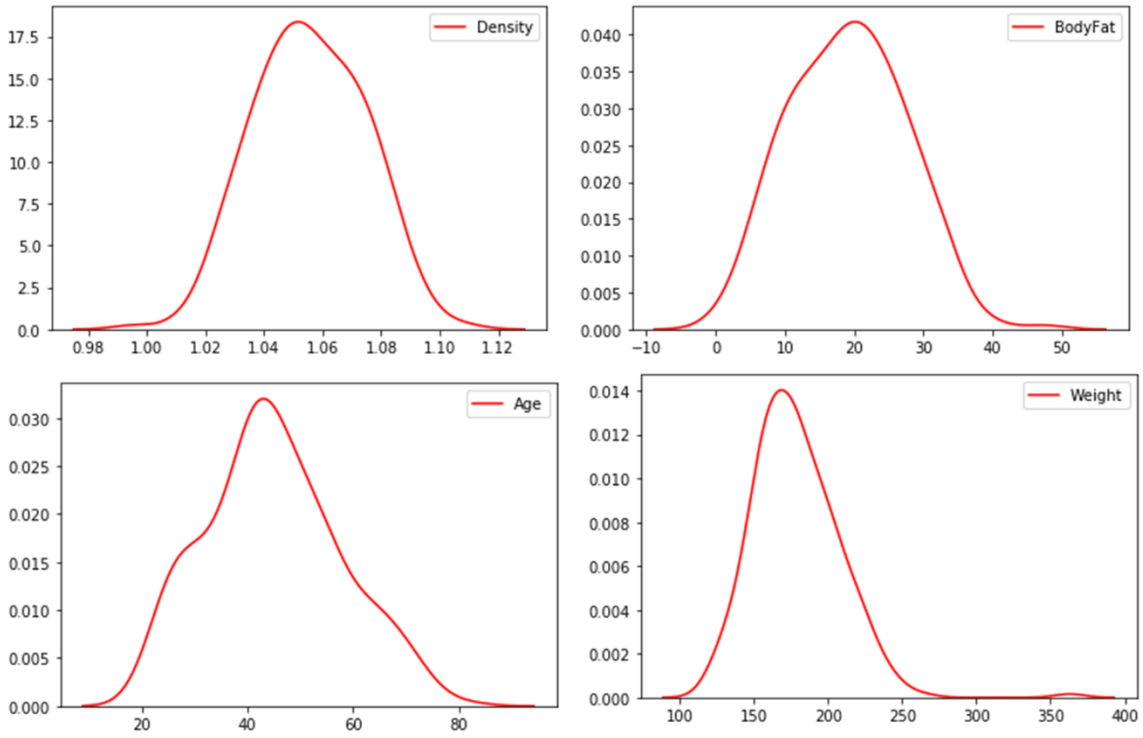

Figure 1: The density, age, person fat and weight features.

Doi: https://doi.org/10.54216/IJAACI.030204
Received: August 22, 2022 Revised: December 21, 2022 Accepted: February 19, 2023

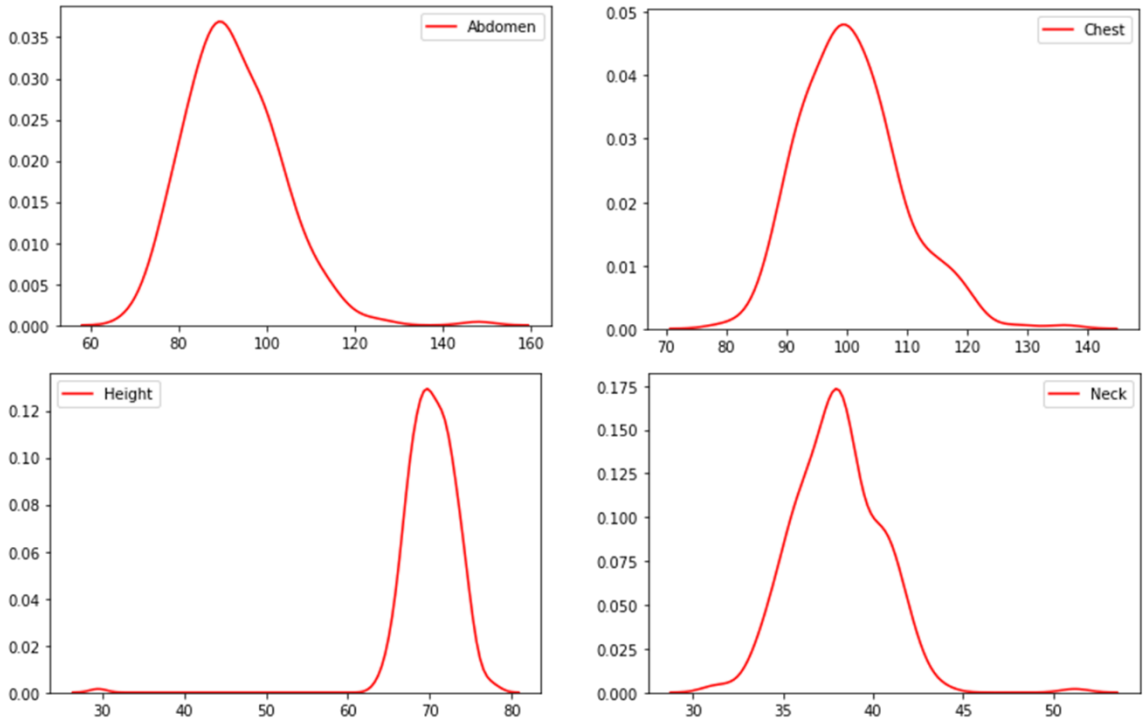

Figure 2: The abdomen, chest, Neck, and height features.

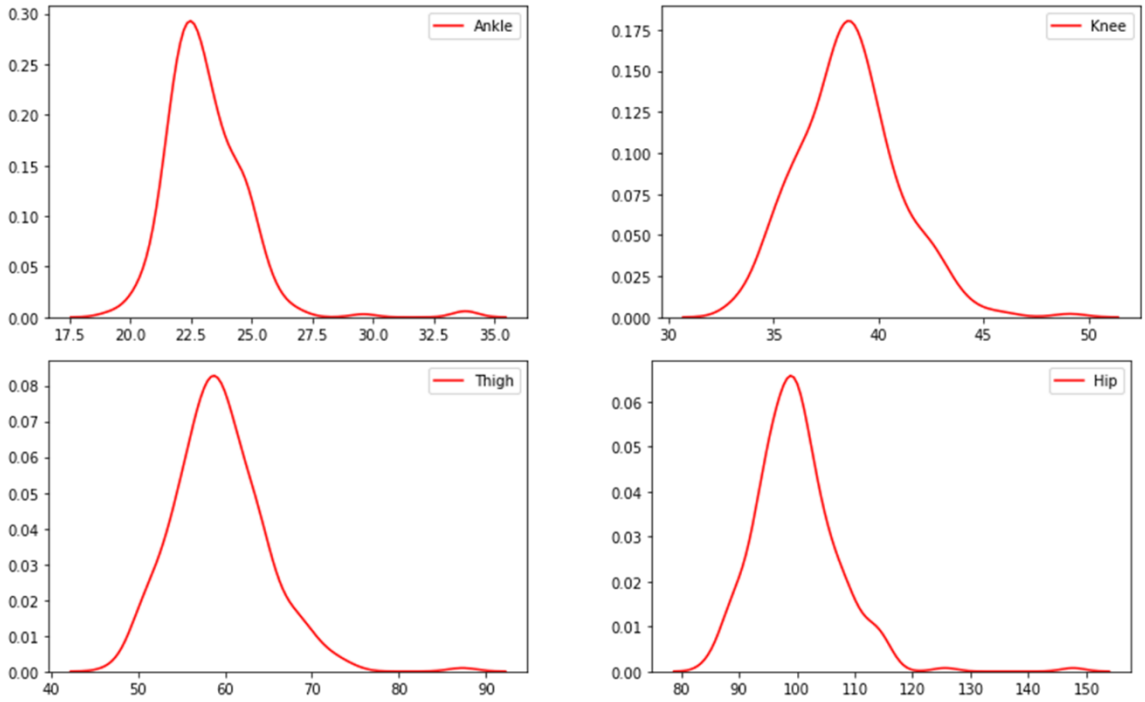

Figure 3: The ankle, thigh, knee, and hip features.

Doi: https://doi.org/10.54216/IJAACI.030204
Received: August 22, 2022 Revised: December 21, 2022 Accepted: February 19, 2023

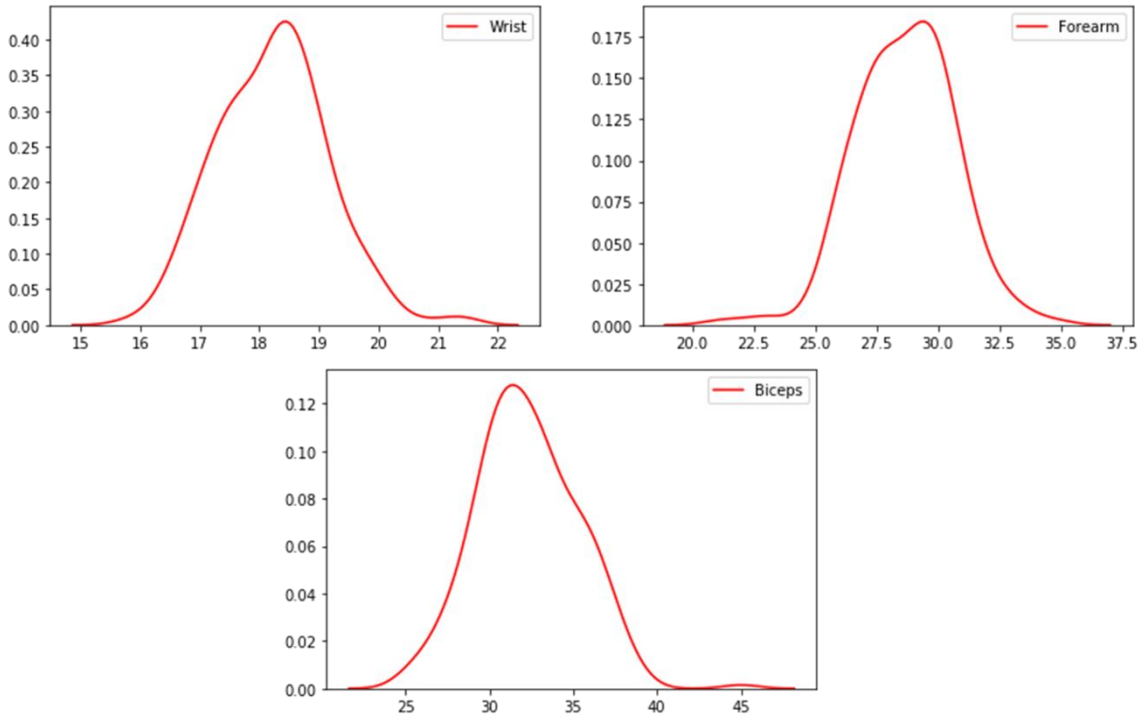

Figure 4: The wrist, biceps, forearm features.

3.2 Forecasting Algorithms

Linear Regression

Multiple linear regression (LR) is a parametric technique used to determine the existence of linear relationships among predictor and predicted parameters[15]–[18].

$$E_r = C_0 + \sum_{i=1}^{m} C_i y_{ir} \quad (1)$$

Where E is an estimated variable and C is a coefficient of predictors.

KNN

Without specifying a parametric link among the predictor and forecasted parameters beforehand, K-NN regression allows for real-time forecasting of the quantity of the predicted value using information gleaned from the data being collected[19]–[22].

This technique is based on using the Euclidean distance function (S_{rt}) to determine the proximity (neighborhood) between the number of indicators for each historical assessment $Y_t = y_{1t}, y_{2t}, y_{3t}, \dots\dots\dots\dots y_{mt}$ and the number of indicators for every current assessment $Y_r = y_{1r}, y_{2r}, y_{3r}, \dots\dots\dots\dots y_{mr}$

$$S_{rt} = \sqrt{\sum_{i=1}^{m} w_i (y_{ir} - y_{it})^2} \quad (2)$$

The forecasted value can be computed by using function of probabilistic as:

Doi: https://doi.org/10.54216/IJAACI.030204
Received: August 22, 2022 Revised: December 21, 2022 Accepted: February 19, 2023

$$P_t = \sum_{j=1}^{K} q(S_{rj}) \times T_j \quad (3)$$

The kernel function can be computed as:

$$q(S_{rj}) = \frac{\frac{1}{S_{rj}}}{\sum_{j=1}^{K} \frac{1}{S_{rj}}} \quad (4)$$

4. Results

This section presented the results of applying the two models in the person fat dataset. Table 2 shows the descriptive statistics on the dataset. The dataset has 252 cases and 15 features. Feature 4 has the highest average, standard deviation, and maximum value. Feature 2 has the minimum value.

Table 2: The statistics analysis of dataset.

	Amount	Average	Standard deviation	minimum	25%	50%	75%	Maximum
BFC1	252	1.055574	0.019031	0.995	1.0414	1.0549	1.0704	1.1089
BFC2	252	19.15079	8.36874	**0**	12.475	19.2	25.3	47.5
BFC3	252	44.88492	12.60204	22	35.75	43	54	81
BFC4	252	**178.9244**	**29.38916**	118.5	159	176.5	197	**363.15**
BFC5	252	70.14881	3.662856	29.5	68.25	70	72.25	77.75
BFC6	252	37.99206	2.430913	31.1	36.4	38	39.425	51.2
BFC7	252	100.8242	8.430476	79.3	94.35	99.65	105.375	136.2
BFC8	252	92.55595	10.78308	69.4	84.575	90.95	99.325	148.1
BFC9	252	99.90476	7.164058	85	95.5	99.3	103.525	147.7
BFC10	252	59.40595	5.249952	47.2	56	59	62.35	87.3
BFC11	252	38.59048	2.411805	33	36.975	38.5	39.925	49.1
BFC12	252	23.10238	1.694893	19.1	22	22.8	24	33.9
BFC13	252	32.27341	3.021274	24.8	30.2	32.05	34.325	45
BFC14	252	28.66389	2.020691	21	27.3	28.7	30	34.9
BFC15	252	18.22976	0.933585	15.8	17.6	18.3	18.8	21.4

This paper built preprocessing on the dataset. In weight features, this paper selected weights greater than 250. In the height feature, this paper selected a height less than 30. Figures 5 and 6 show the weight and height after preprocessing dataset.

Doi: https://doi.org/10.54216/IJAACI.030204
Received: August 22, 2022 Revised: December 21, 2022 Accepted: February 19, 2023

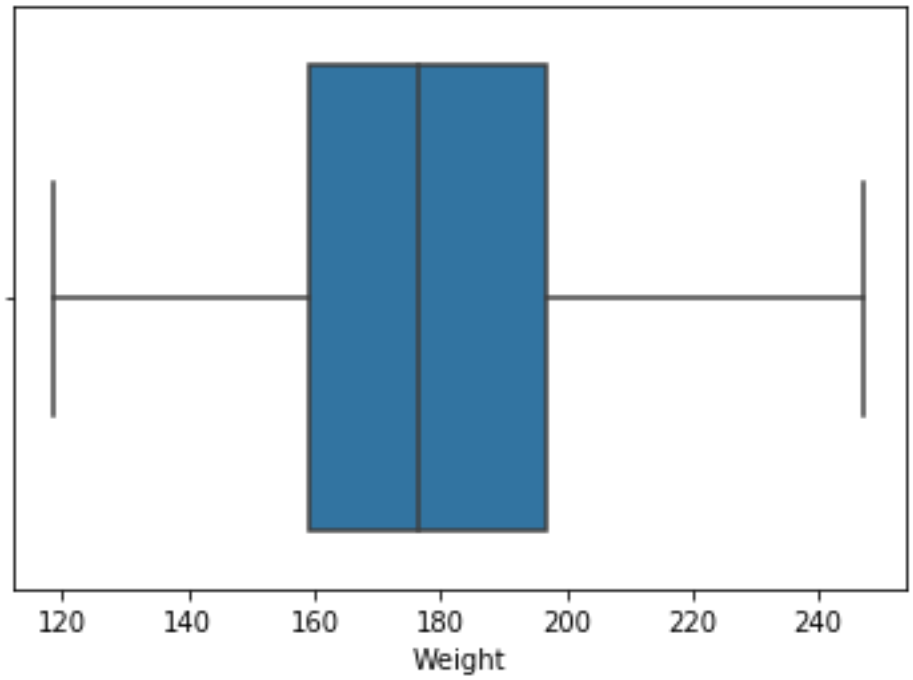

Figure 5: The weight feature after preprocessing dataset.

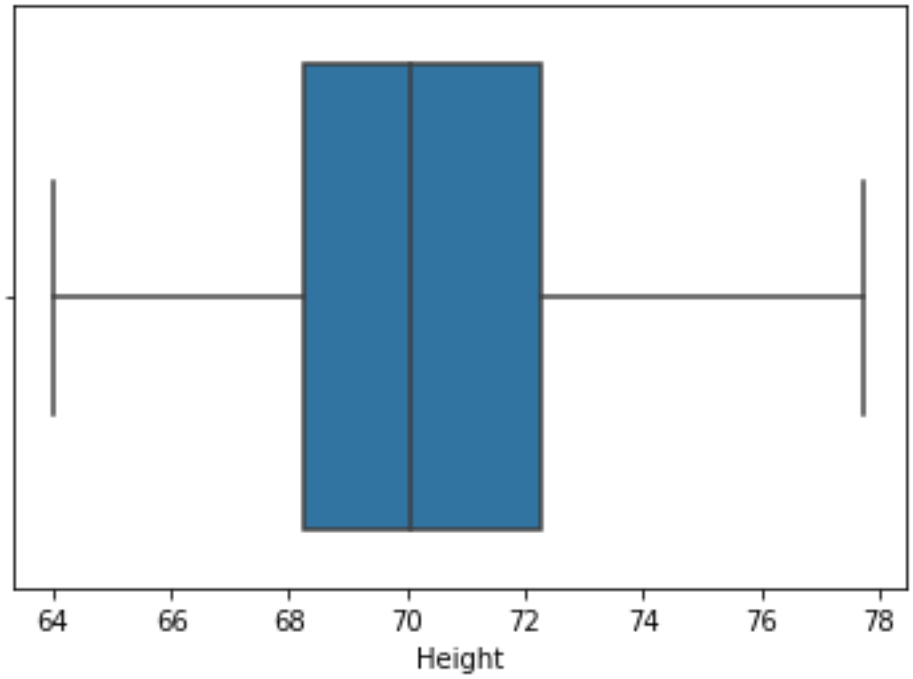

Figure 6: The height feature after preprocessing dataset.

Doi: https://doi.org/10.54216/IJAACI.030204
Received: August 22, 2022 Revised: December 21, 2022 Accepted: February 19, 2023

Figure 7 shows the score of two models into dataset. The LR has the highest score 0.91 and the KNN is the lowest score 0.545. So, the LR is better than the KNN model in this dataset.

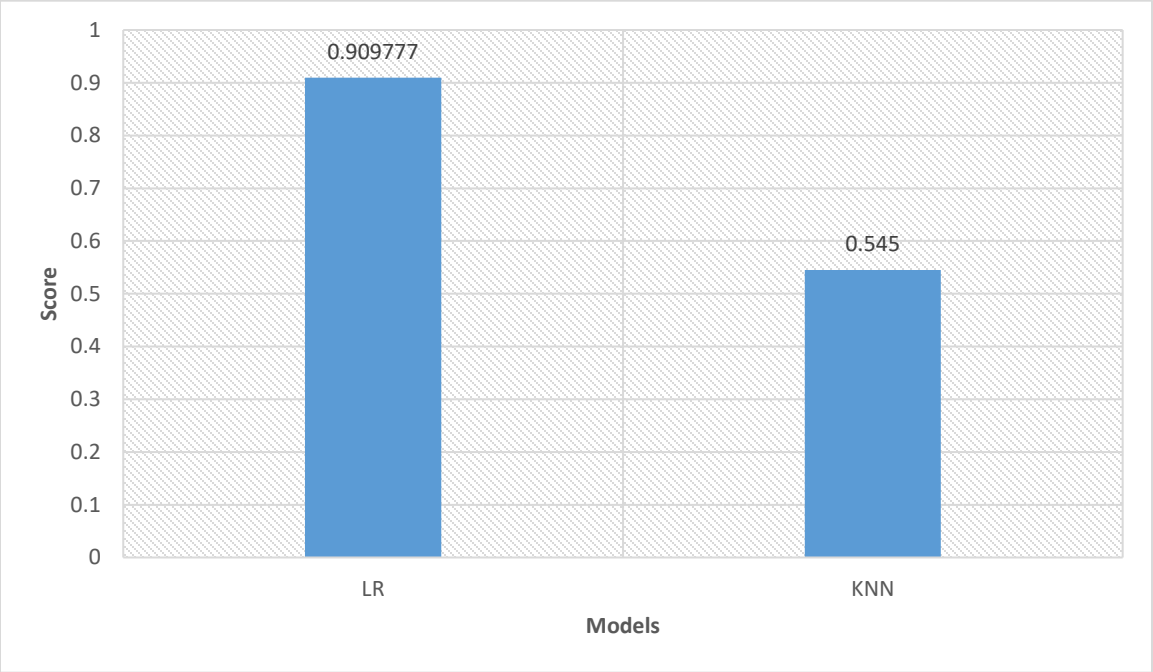

Figure 7: The comparison score of LR and KNN models.

5. Conclusion

A reliable estimate of body fat percentage is crucial for monitoring and controlling health. Personalized health programmer, early diagnosis of health problems, and efficient weight control measures are all made possible by the availability of trustworthy and accessible prediction technologies. However, in order to increase accuracy and reliability, body fat prediction algorithms need ongoing study and development. Body fat prediction has been aided by anthropometrics, imaging technology, and machine learning algorithms, all of which have their own advantages and disadvantages. More reliable and precise forecasts may be achieved by combining different methods and capitalizing on their respective strengths. Personalized body fat measurement requires consideration of individual characteristics such as age, gender, and race. Prediction models may be more inclusive and reliable if they account for differences in body composition among populations and ethnic groups.

There are multiple manners in which obesity is detrimental to cardiovascular health. It is pertinent to note that different types of obesity have different relationships with CVD. Specialists seeing a wide variety of fat phenotypes continue to find these complicated concerns associated with obesity to be their greatest difficulty. Because obesity has taken on epidemic proportions, doctors need to know how to spot the most common kinds of dangerous obesity, such as stomach obesity and severe overweight. So, in this study, we described a method for predicting person fat using machine learning. We conducted experiments on two different machine learning algorithms, all aimed at predicting person fat percentage, and found that our suggested solution, the linear regression, outperformed the others.

References

[1] P. Costa-Urrutia *et al.*, "Obesity measured as percent person fat, relationship with person mass index, and percentile curves for Mexican pediatric population," *PLoS One*, vol. 14, no. 2, p. e0212792, 2019.

[2] L. A. Fowler, J. R. Fernández, S. E. Deemer, and B. A. Gower, "Genetic risk score prediction of leg fat and insulin sensitivity differs by race/ethnicity in early pubertal children," *Pediatr. Obes.*, vol. 16, no. 12, p.

Doi: https://doi.org/10.54216/IJAACI.030204
Received: August 22, 2022 Revised: December 21, 2022 Accepted: February 19, 2023

e12828, 2021.

[3] A. M. Barberio *et al.*, "Central person fatness is a stronger predictor of cancer risk than overall person size," *Nat. Commun.*, vol. 10, no. 1, p. 383, 2019.

[4] O. O. Woolcott and R. N. Bergman, "Relative Fat Mass as an estimator of whole-person fat percentage among children and adolescents: A cross-sectional study using NHANES," *Sci. Rep.*, vol. 9, no. 1, p. 15279, 2019.

[5] S. M. Alwash, H. D. McIntyre, and A. Mamun, "The association of general obesity, central obesity and visceral person fat with the risk of gestational diabetes mellitus: Evidence from a systematic review and meta-analysis," *Obes. Res. Clin. Pract.*, vol. 15, no. 5, pp. 425–430, 2021.

[6] I.-E. Jurca-Simina *et al.*, "What if person fat percentage association with FINDRISC score leads to a better prediction of type 2 diabetes mellitus," *Rom J Morphol Embryol*, vol. 60, no. 1, pp. 205–210, 2019.

[7] D. G. Whitney, F. Miller, R. T. Pohlig, and C. M. Modlesky, "BMI does not capture the high fat mass index and low fat-free mass index in children with cerebral palsy and proposed statistical models that improve this accuracy," *Int. J. Obes.*, vol. 43, no. 1, pp. 82–90, 2019.

[8] S. Delle Monache *et al.*, "Person mass index represents a good predictor of vitamin D status in women independently from age," *Clin. Nutr.*, vol. 38, no. 2, pp. 829–834, 2019.

[9] M. T. Hudda *et al.*, "Development and validation of a prediction model for fat mass in children and adolescents: meta-analysis using individual participant data," *bmj*, vol. 366, 2019.

[10] N. Stefan, "Causes, consequences, and treatment of metabolically unhealthy fat distribution," *lancet Diabetes Endocrinol.*, vol. 8, no. 7, pp. 616–627, 2020.

[11] K.-A. Shin and Y.-J. Kim, "Usefulness of surrogate markers of person fat distribution for predicting metabolic syndrome in middle-aged and older Korean populations," *Diabetes, Metab. Syndr. Obes. targets Ther.*, pp. 2251–2259, 2019.

[12] C.-L. Hsu *et al.*, "Role of fatty liver index and metabolic factors in the prediction of nonalcoholic fatty liver disease in a lean population receiving health checkup," *Clin. Transl. Gastroenterol.*, vol. 10, no. 5, 2019.

[13] Y. Xu and L. Wang, "A weighted twin support vector regression," *Knowledge-Based Syst.*, vol. 33, pp. 92–101, 2012.

[14] A. Kupusinac, E. Stokić, and R. Doroslovački, "Predicting person fat percentage based on gender, age and BMI by using artificial neural networks," *Comput. Methods Programs Biomed.*, vol. 113, no. 2, pp. 610–619, 2014.

[15] Y. Wang, W. Pang, and Z. Jiao, "An adaptive mutual K-nearest neighbors clustering algorithm based on maximizing mutual information," *Pattern Recognit.*, vol. 137, p. 109273, 2023.

[16] T. Adithiyaa, D. Chandramohan, and T. Sathish, "Optimal prediction of process parameters by GWO-KNN in stirring-squeeze casting of AA2219 reinforced metal matrix composites," *Mater. Today Proc.*, vol. 21, pp. 1000–1007, 2020.

[17] S. N. Betgeri, S. R. Vadyala, J. C. Matthews, M. Madadi, and G. Vladeanu, "Wastewater pipe condition rating model using K-nearest neighbors," *Tunn. Undergr. Sp. Technol.*, vol. 132, p. 104921, 2023.

[18] D. N. Cosenza *et al.*, "Comparison of linear regression, k-nearest neighbour and random forest methods in airborne laser-scanning-based prediction of growing stock," *For. An Int. J. For. Res.*, vol. 94, no. 2, pp. 311–323, 2021.

[19] C. Feng, B. Zhao, X. Zhou, X. Ding, and Z. Shan, "An Enhanced Quantum K-Nearest Neighbor Classification Algorithm Based on Polar Distance," *Entropy*, vol. 25, no. 1, p. 127, 2023.

[20] H. I. Dino and M. B. Abdulrazzaq, "Facial expression classification based on SVM, KNN and MLP

Doi: https://doi.org/10.54216/IJAACI.030204
Received: August 22, 2022 Revised: December 21, 2022 Accepted: February 19, 2023

classifiers," in *2019 International Conference on Advanced Science and Engineering (ICOASE)*, IEEE, 2019, pp. 70–75.

[21] Mona Mohamed, Intelligent Fat Predictor: Leveraging Linear Regression and K Nearest Neighbors in Obesity diseases., International Journal of Advances in Applied Computational Intelligence, Vol. 3 , No. 1 , (2023) : 08-18 (Doi : https://doi.org/10.54216/IJAACI.030101)

[22] N. D. Mu'azu and S. O. Olatunji, "K-nearest neighbor based computational intelligence and RSM predictive models for extraction of Cadmium from contaminated soil," *Ain Shams Eng. J.*, vol. 14, no. 4, p. 101944, 2023.

Doi: https://doi.org/10.54216/IJAACI.030204
Received: August 22, 2022 Revised: December 21, 2022 Accepted: February 19, 2023