Anbar Journal of Engineering Science (AJES) **Vol. 14 , No. 1 (2023) , pp. 80 ~ 88**
P-ISSN: 1997-9428 ; E-ISSN: 2705-7440 DOI: http://doi.org/10.37649/aengs.2023.139350.1045

University of Anbar

Anbar Journal of Engineering Science©

journal homepage: https://ajes.uoanbar.edu.iq/

Detection of Obesity Stages Using Machine Learning Algorithms

Sukru Kitis[a], Hanife Goker[b]
[a] *Kütahya Dumlupınar University, Kütahya, TURKEY*
[b] *Gazi University, Ankara, TURKEY*

PAPER INFO

Paper history
Received: 28/03/2023
Revised: 20/04/2023
Accepted: 26/04/2023

Keywords:
Obesity
Machine Learning
Classification
Obesity Stages
Random Forest

ABSTRACT

Obesity is the excess body weight relative to the height above the desired level due to an excessive increase in body fat to lean mass. It causes many health problems due to its negative effects on body systems (cardiovascular, musculoskeletal, gastrointestinal, respiratory, skin, endocrine, genitourinary system) and psychosocial status. This study aims to effectively detect the eating and physical condition-based obesity stages using machine learning algorithms. The dataset contains data for estimating obesity stages in individuals from Mexico, Peru, and Colombia and is available as an open source. There are 2111 records and 17 attributes in the dataset. In the records, obesity stages were categorized as insufficient weight, normal weight, overweight level I, overweight level II, obesity type I, obesity type II and obesity type III. The 10-fold cross-validation method was used to validate the model, and the performances of the Support Vector Machine (SVM), Random Forest (RF), and Multilayer Perceptron (MLP) classification algorithms were compared. It has been determined that the highest performance among the algorithms whose performances are compared belongs to the RF Algorithm (95.78%). This paper's abstract has been presented at the International Conference on Computational Mathematics and Engineering Sciences held in Ordu (Turkey), / 20-22 May. 2022.

1. Introduction

Obesity is a chronic condition characterized by increased body fat relative to lean body mass, resulting from the fact that the energy taken into the body with food is more than the energy spent [1]. Obesity, diabetes, hypertension, and high blood lipid levels, which are related and cause cardiovascular diseases, are evaluated with a common name as Metabolic Syndrome since they are seen in the same source disease group. Obesity and overweight are major public health concerns, with more obese or overweight adults than underweight adults. 40% of women and 39% of men aged 18 and over, accounting for nearly 2 billion adults, were overweight, and 15% of women and 11% of men, more than half a billion, were obese worldwide. Both obesity and overweight have increased markedly over the past four decades [2].

* Corresponding author: Sukru Kitis; sukru.kitis@dpu.edu.tr ; +90-5453343233

Moreover, diabetes has become the fifth cause of death in parallel with the rapid increase in obesity globally [3]. For every 1 kg of body weight, obesity increases the incidence of diabetes by 5%. 200 million people have diabetes worldwide, which is predicted to double in the next 30 years [3] likely.

Obesity is calculated by taking weight and height values. For the calculation, the weight is divided by the square of the height value (kg/m^2) [4]. Body mass index (BMI) calculation is used as a standard worldwide to calculate obesity. Individuals is defined as following; less than 18.5 kg/m2: Weak individual, between 18.5– 24.9 kg/m2: Normal weight individual, between 25 – 29.9 kg/m2: Overweight individual, between 30 – 39.9 kg/m2: Obese individual, over 40 kg/m2: severely obese (morbidly obese) [5]. Machine learning methods are most frequently utilized in the field of health sciences for determination of post-disease complications, estimation, diagnosis, and it is aimed to provide better quality health care to patients by saving time and workload [6]. Woldaregaya et al. [7] proposed a patient-oriented and artificial intelligence-based patient follow-up system for patients with diabetes to ensure patient comfort in health management processes and improve treatment processes. In another study on the opportunities created by machine learning in healthcare, a model that classifies patients according to risk groups using hospital data was proposed, thus improving patient follow-up and related actions [8]. In another study, an artificial intelligence model was proposed to identify high-cost and risky patients to determine the necessary treatment and steps [9]. Another example of health management processes related to machine learning techniques solutions is that health insurance providers were improved and accelerated by integrating machine learning into their pre-approval processes [10]. A different perspective on hospital management, which is a part of health management, was presented by Rodriguez et al. [11]. In a study by Mercaldo et al. [12], the Hoeffding Tree algorithm was used to diagnose and classify diabetes. The study, as mentioned earlier, created a machine learning model over the costs of patients hospitalized for diabetes and similar chronic diseases. It aimed to estimate the costs of the disease with the model [13]. Today, artificial intelligence and machine learning have begun to affect doctors, patients, hospitals, and all health-related fields. Machine learning methods and artificial intelligence are currently used in areas such as computed diabetic retinopathy analysis, tomography, and heart attack risk determination by electrocardiograph (ECG). Since there is a lot of data in these fields, algorithms can be as successful as specialist doctors in making diagnoses [5]. The use of machine learning applications and artificial intelligence in health is carried out in many sub-activity areas, such as medical diagnosis and cost estimation, imaging analysis, disease tracking, resource planning and emergency management, and processing of unstructured data [14, 15, 16]. Artificial intelligence models, also used in functionalizing high-dimensional patient data, are active in increasing data reliability and quality [13,17,18]. Given the widespread impact of this disease, its early diagnosis is critical.

In this study, we proposed a machine learning model to classify obesity stages accurately and effectively. The main contributions of this study are as follows:

a) The performances of machine learning algorithms were compared in determining the best classification performance to assist expert decisions in the early detection of obesity stages.

b) Experiments have shown that RF algorithm is advantageous over machine learning methods in the multi-classification of obesity stages.

c) As a result of the performance analysis of the RF algorithm, an overall accuracy value of 95.78% was found.

2. Material and Method

2.1. Proposed Model

In this study, a machine learning model was proposed to be used in the early diagnosis of obesity disease stages. The flow diagram of the proposed model is given in Figure 1. In this study, the obesity disease dataset was used which is accessible as open source for data scientists and machine learning practitioners. The dataset contains data for the estimation of obesity stages in individuals from the countries of Mexico, Peru and Colombia [19]. The characteristic of the dataset is shown in Table 1.

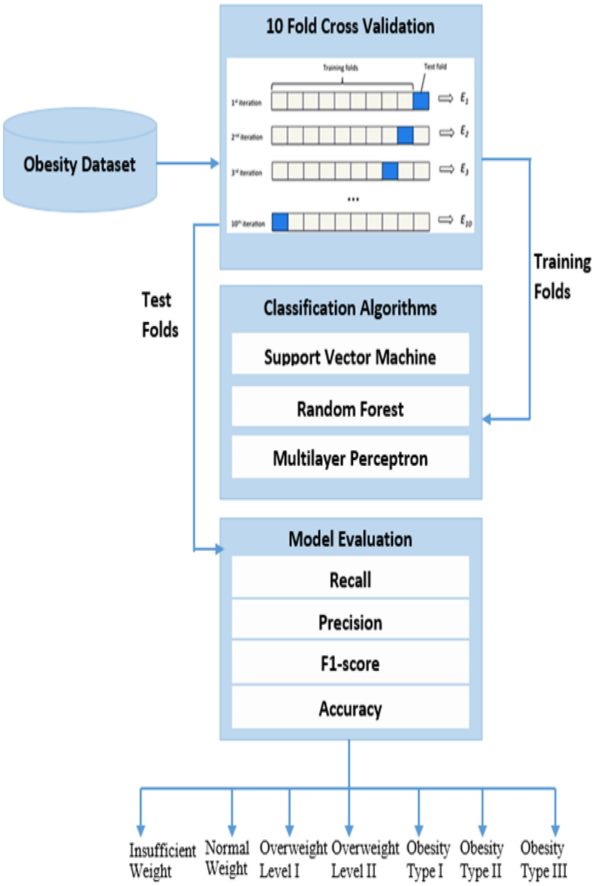

Figure 1. Flow diagram of the proposed model

2.2. Dataset

The dataset contains 17 attributes and 2111 records. In the records, obesity stages were categorized as insufficient weight, normal weight, overweight level I, overweight level II, obesity type I, obesity type II and obesity type III.

Table 1. Characteristics of the dataset

Dataset	Number of classes	Number of features	Number of samples
Obesity Stages	7	17	2111

2.3. Division of Data into Training and Test Sets

The model learns to use the training dataset, and the model's performance is validated with the testing dataset. The 10-fold cross-validation method was used to validate the model. In databases consisting of several thousand rows or less, the k-fold cross-validation method can be used to divide the data into k groups. The dataset is randomly divided into k groups. When we examine the literature, it is seen that the k value is generally chosen as 10 [20]. In this study, the k value was chosen as 10. In this method, the first group is used for testing and the other groups are used for learning. This process is continued with one group for testing and the other for learning purposes. The average of the ten resulting error rates will be the estimated error rate of the constructed model.

2.4. Model Evaluation

Model evaluation metrics can be derived from the confusion matrix. In the confusion matrix, the rows represent the actual numbers of the samples in the test set, and the columns represent the model's prediction [21]. Confusion matrix has True positive (TP), True negative (TN), False positive (FP) and False negative (FN) parameters. TP; the predicted value is positive and its positive. FP; the predicted value is positive, but it is negative. FN; the predicted value is negative, but its positive. TN; the predicted value is negative, and its negative. Various model performance measures are calculated using the confusion matrix's parameters. The most basic model performance criteria used for the performance evaluation of the model; recall, precision, f1 score, and accuracy [22]. The performance criteria of this model calculated with these parameters obtained from the confusion matrix are given below:

$$\text{Recall} = TP / (TP+FN) \tag{1}$$

$$\text{Precision} = TP / (TP + FP) \tag{2}$$

$$\text{F1- score} = \frac{(2 \times \text{Recall} \times \text{Precision})}{(\text{Recall} + \text{Precision})} \tag{3}$$

$$\text{Overall Accuracy} = \frac{\text{Correctly classified values}}{\text{Total number of values}} \tag{4}$$

2.5. Classification Algorithms

SVM is an algorithm introduced to the literature by Vapnik, which performs the classification by finding the optimal discriminating hyper-plane [23]. Obtaining this hyper-plane makes a distinctive and supervised binary classification possible. A hyper-plane is obtained with a training dataset applied to the algorithm. Here, classification is performed by labelling each input feature vector as binary. The SVM method chooses a special solution that separates the classes with a maximum margin [24].

RF is a classification algorithm introduced by Breiman [25]. The algorithm works as a combination of multiple decision trees. It is also known as one of many bagging-type ensemble classifiers. Combining several decision trees alleviates the concept between variance and bias, reducing the likelihood of overfitting. RF algorithm first generates a training set from the feature set. This set is applied to each tree and the achievements of the trees are revealed. Each leaf has a linear model in model trees to optimize the local subspace defined by that leaf. The tree variety is generated by randomly sampling the training data, and changing it for each tree. Then, when growing a tree, only a random subset of all attributes is considered at each node, calculating the best split for that subset instead of always calculating the best possible split for each node. Thus, simplifying the optimization procedure, random trees use this split selection to create a reasonably balanced tree where a single global setting for the ridge value works across all leaves [26].

MLP is a feed-forward neural network with an input layer, one or more hidden layers, and an output layer. MLP works in such a way that it takes features as input, multiplies those features with the initial weights in the hidden layers, and then sends the weighted features to an activation function that gives the output as a probability distribution. The instance with the highest probability is declared as a class of the input [27].

3. Experimental Results and Discussion

In this study, the performances of a machine learning classification algorithm were compared to classify the obesity stages with high performance accurately. Classification algorithms are trained with the training dataset and validated with the test dataset.

One of the main reasons why SVM is frequently used today is its success in modeling complex irregular data with appropriate kernel function selection. This means that nonlinear relationships in a high-dimensional feature space are expressed as a linear line [36,37]. MLP is a typical non-parametric neural network classifier and powerful classifiers that can perform better than other classifiers [38]. In this method, parameters are set with the help of a validation set or cross validation techniques [39]. “Complex, high-dimensional data, mixed categorical and numerical variables, non-Gaussian statistical distributions, non-linear relationships, etc.” the hardware of the RF classifier in the face of problems explains the intense interest in the literatüre [40,41,42,43]. Therefore SVM, RF and MLP algorithms used in this article.

The results of the SVM, RF, and MLP classification algorithms were compared according to the model performance criteria. The results of the SVM algorithm are shown in Table 2 according to the model performance criteria.

Table 2. Performance results of the SVM classification algorithm

Obesity Stages	Precision	Recall	F1-Score	TP	Overall Accuracy
1	0.96	0.88	0.92	1894	89.72%
2	0.70	0.81	0.75		
3	0.81	0.84	0.82		
4	0.85	0.83	0.84		
5	0.96	0.91	0.93		
6	0.99	0.98	0.98		
7	0.99	0.99	0.99		

1: Insufficient Weight, 2: Normal Weight, 3: Overweight Level 1, 4: Overweight Level 2, 5: Obesity Type 1, 6: Obesity Type 2, 7: Obesity Type 3

Precision, recall, f1-score, and overall accuracy values were calculated by using the TP, FP, FN and TN values in the confusion matrix.

Table 3. Performance results of the MLP classification algorithm

Obesity Stages	Precision	Recall	F1-Score	TP	Overall Accuracy
1	0.95	0.92	0.94	1992	94.36%
2	0.88	0.89	0.89		
3	0.87	0.91	0.89		
4	0.94	0.90	0.92		
5	0.98	0.97	0.97		
6	0.98	0.99	0.98		
7	0.99	1.00	1.00		

The performances of classification algorithms are tested by using these values, especially accuracy and f1-score values. These values should be close to 1. In the performance analysis results of the SVM algorithm, the overall accuracy value is 89.72%.

The results of the MLP algorithm are shown in Table 3 according to the model performance criteria.

In the performance analysis results of the MLP algorithm, the overall accuracy value is 94.36%. It can be seen that the precision, f1-score, and recall values of the MLP algorithm are close to 1. The fact that these criteria are close to 1 indicates that the model performs well and does not make a random prediction. The results of the RF algorithm are shown in Table 4.

Table 4. Performance results of the RF classification algorithm

Obesity Stages	Precision	Recall	F1-Score	TP	Overall Accuracy
1	0.94	0.99	0.97		
2	0.96	0.84	0.90		
3	0.90	0.96	0.93		
4	0.94	0.95	0.95	2022	95.78%
5	0.97	0.98	0.97		
6	0.99	0.99	0.99		
7	1.00	1.00	1.00		

The performances of the SVM, RF, and MLP classification algorithms were compared. It has been determined that the highest performance among the algorithms whose performances are compared belongs to the RF Algorithm (95.78%). Precision, recall and f1 score values are over 95% in almost all labels. These results prove that the model's performance is quite high and the proposed model does not produce a random result. Performance evaluation metrics of machine learning classifiers are given in Figure 2.

Compared classifiers are ranked as RF, MLP and SVM, respectively, according to their performance. The RF algorithm performs best when the mean of precision, f1 score, recall, and overall accuracy values are examined.

The performance analysis results of the RF algorithm were found to be 95.7% precision, 95.85% recall, 95.85% f1-score, and 95.78% overall accuracy. The performance analysis results of the MLP algorithm were found to be 94.14% precision, 94% recall, 94.14% f1-score, and 94.36% overall accuracy. The performance analysis results of the SVM algorithm were found to be 89.42% precision, 89.14% recall, 89% f1-score, and 89.72% overall accuracy. Comparative analysis, the results of this study and the relevant literature for the diagnosis of obesity are presented in Table 5.

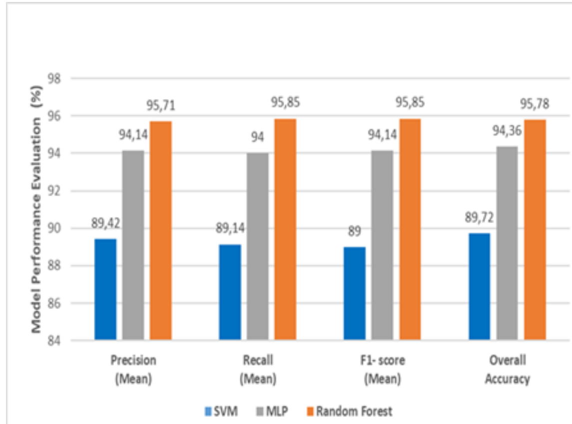

Figure 2. Performance results of classification algorithms

In recent studies, the use of machine learning algorithms for disease prediction is increasing therefore, we have explored some of the recent works for the prediction of obesity's disease. Kivrak's study [28], the data set on the open-source access website was used for the prediction of obesity stages and consists of patient records of 17 variables created by the machine learning repository. Kivrak applied CNN algorithms and accuracy value of this algorithm 82%. The dataset used by Singh & Tawfik [29] is young adults (14 years old and older). They used Lineer SVM, quadratic SVM, MLPFFA, NN algorithms, and MLPFFA algorithm the highest prediction accuracy with 93.4%. Daud et al. [30] used a Malaysian grocery dataset containing demographic and anthropometric data (8273 records). The accuracy value of the J48 algorithm was 89.41%. Researchers constructed their studies using the same dataset which 17 attributes and 2111 records. For example, Alqahtani et al. [31] used RF, MLP algorithms and MLP algorithm the highest prediction accuracy with 95.06%. Huang [32] used RF, LR, SVM. RF algorithm is best and highest prediction accuracy with 95.6%. Cui et al. [33] used LR, SVM, k nearest neighbor, ID3, CART, RF, XGBoost, GBDT, LightGBM, and Aggregate Prediction algorithms.

Table 5. A comparative analysis with relevant literature studies

Research	Algorithms	Dataset	Best Classifier	Accuracy
Kivrak (2021) [28]	CNN	2111 records and 17 attributes	CNN	82%
Singh & Tawfik (2019) [29].	Lineer SVM, quadratic SVM, MLPFFANN	Young adult (14 years old and older)	MLPFFANN	93.40%
Daud et al. (2018) [30].	J48 decision tree	Malaysian grocery data, demographic data and anthropometric data (8273 records)	J48	89.41%
Qahtani et al. (2021) [31].	RF, MLP	2111 records and 17 attributes	MLP	95.06%
Huang (2022) [32].	RF, Logistic Regression (LR), SVM	2111 records and 17 attributes	RF	95.60%
Cui et al. (2021) [33].	LR, SVM, KNN, ID3 decision tree, CART, RF, XGBoost, GBDT, LightGBM, Aggregate Prediction	2111 records and 17 attributes	Aggregate Prediction	86.29%
Jindal et al. (2018) [34].	RF	2111 records and 17 attributes	RF	89.68%
Lopes et al. (2021) [35].	SVM linear, RF	2111 records and 17 attributes	RF	95.20%
The proposed model	SVM, RF, MLP	2111 records and 17 attributes	RF	95.78%

CNN: Convolutional neural network, MLPFFANN: Multilayer perceptron feed forward artificial neural network, LR: Logistic regression, KNN: k-Nearest Neighbor, CART: Classification and regression tree, XGBoost: Extreme gradient boosting, GBDT: Gradient-boosted decision trees.

4. Conclusion

In this study, the obesity disease dataset was used which is accessible as open source. The model learns to use the training dataset and the performance of the model is validated with the testing dataset. The 10-fold cross-validation method was used to validate the model. In this method, the first group is used for testing and the other groups are used for learning. In this study, the performances of a machine learning classification algorithm were compared to classify the obesity stages with high performance accurately. Classification algorithms are trained with the training dataset and validated with the test dataset. The performances of the SVM, RF, and MLP classification algorithms were compared. It has been determined that the highest performance among the algorithms whose performances are compared belongs to the RF Algorithm (95.78%). Precision, recall and f1 score values are over 95% in almost all labels. These results prove that the model's performance is quite high, and the proposed model does not produce a random result. Comparative analysis, the results of this study, and the relevant literature for the diagnosis of obesity are presented in Table 5; the proposed model has the highest score.

This study, used BMI for detection of obesity stages. Future research can use WHR (Waist Hip Ratio) and RFM (Relative Fat Mass index). Additionally, this proposed model can use the detection of cholesterol value.

References

[1] B. Tozoğlu, Ş. Polat, H. Celik, B. Mazlumoglu, “Examination of obesity awareness levels of university students in terms of different variables,” Journal of Health and Medical Sciences, 5(2), 133-142. 2022, DOI: 10.31014/aior.1994.05.02.223

[2] WHO, “Obesity and overweight,” https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight (accessed on july 2022).

[3] S. Birtane, E. Canayaz, Z. A. Altikardes, H. Korkmaz, “Mamdani tip bulanık mantık kümeleri kullanılarak metabolik sendrom risk tespit karar destek sisteminin geliştirilmesi,” https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8238035 (accessed on july 2022).

[4] Acıbadem, “Vücut kitle endeksi hesaplama,” https://www.acibadem.com.tr/ilgi-alani/vucut-kitle-indeksi-hesaplama/ (accessed on july 2022).

[5] E. Karal, M. Turan, “Hekime tanı koymada yardımcı, yapay zeka destekli hastalık tespit uzmanı,” European Journal of Science and Technology Special Issue 26, pp. 100-116, 2021, DOI: 10.31590/ejosat.945518

[6] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, I. Chouvarda, “Machine learning and data mining methods in diabetes research,” Computational and Structural Biotechnology, 15. 104–116. 2017, https://doi.org/10.1016/j.csbj.2016.12.005

[7] A. Z. Woldaregaya, E. Arsand, S. Walderhaug, D. Albers, L. Mamykina, T. Botsis, G. Hartvigsen, “Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes,” Artificial Intelligence in Medicine. Volume 98, Pages 109-134, 2019, DOI: 10.1016/j.artmed.2019.07.007.

[8] R. B. Parikh, M. Kakad, D. W. Bates, “Integrating predictive analytics into high-value care: the dawn of precision delivery,” JAMA, 315, 651-652, 2016, DOI: 10.1001/jama.2015.19417.

[9] D. W. Bates, S. Saria, L. Ohno-Machado, A. Shah, G. Escobar, “Big data in healthcare: using analytics to identify and manage high-risk and high-cost patients,” Health Aff, 33, 1123-1131, 2014, DOI: 10.1377/hlthaff.2014.0041

[10] F. H. D. Araújo, A. M. Santana, P. A. S. Neto, “Using machine learning to support healthcare professionals in making pre authorization decisions,” International Journal of Medical Informatics, 94:1–7, 2016, DOI:10.1016/j.i.jmedinf.2016.06.007

[11] G. Rodriguez, O. L. P. Espitia, C. Franco, V. Augusto, “Predicting healthcare costs of diabetes using machine learning models,” Elsevier Inc, 2019 DOI: 10.1016/j.jval.2019.09.903

[12] F. Mercaldo, V. Nardone, A. Santone, “Diabetes mellitus affected patients classification and diagnosis through machine learning techniques,” Procedia Computer Science, 112, 2519-228, 2017 https://doi.org/10.1016/j.procs.2017.08.193

[13] Ü. Veranyurt, A. F. Deveci, M. F. Esen, O. Veranyurt, “Makine öğrenmesi teknikleriyle hastalık sınıflandırması: random forest, k-nearest neighbour ve adaboost algoritmaları uygulaması,” Usaysad Derg, 6(2), 275-286, 2020.

[14] S. Narula, K. Shameer, A. M. S. Omar, J. T. Dudley, P. P. Sengupta, “Reply: Deep learning with unsupervised feature in echocardiographic imaging,” J Am Coll Cardiol, 69:2101 2, 2017, DOI: 10.1016/j.jacc.2017.01.062

[15] D. H. Alonso, M. N. Wernick, Y. Yang, G. Germano, D. S. Berman, P. Slmoka, ”Prediction of cardiac death after adenosine myocardial perfusion SPECT based on machine learning,” J Nucl Cardiol, 2018, doi: 10.1007/s12350-017-0924-x.

[16] A. Esteva, B. Kupre, R. A. Novoa, “Dermatologist-level classification of skin cancer with deep neural networks,” Nature, 542:115–8, 2017, https://doi.org/10.1038/nature21056

[17] B. X. Tran, C. A. Latkin, V. T. Giang, "The current research landscape of the application of artificial intelligence in managing cerebrovascular and heart diseases: A bibliometric and content analysis," Int. J. Environ. Res. Public Health, 16:2699, 2019, DOI: 10.3390/ijerph16152699

[18] S. L. Cichosz, M. D. Johansen, O. Hejlesen, "Toward big data analytics: review of predictive models in management of es and its complications," J es Sci Technol, 10(1), 27-34, 2015, DOI: 10.1177/1932296815611680

[19] F. M. Palechor, A. D. L. H. Manotas, "Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from Colombia, Peru and Mexico," Data in Brief, 104344, 2019.

[20] H. Akpınar, "Veri tabanlarında bilgi keşfi ve veri madenciliği," İstanbul Üniversitesi İşletme Fakültesi Dergisi, 29(1), 1-22. 2000.

[21] G. Zeng, "On the confusion matrix in credit scoring and its analytical properties," Communications in Statistics-Theory and Methods, 49(9), 2080-2093, 2020, https://doi.org/10.1080/03610926.2019.1568485

[22] E. Başaran, "Classification of white blood cells with SVM by selecting SqueezeNet and LIME properties by mRMR method," Signal, Image and Video Processing, 1-9, 2022, https://doi.org/10.1007/s11760-022-02141-2

[23] V. Vapnik, R. Izmailov, "Reinforced SVM method and memorization mechanisms," Pattern Recognition, 119, 108018, 2021, https://doi.org/10.1016/j.patcog.2021.108018

[24] M. Omidvar, A. Zahedi, H. Bakhshi, "EEG signal processing for epilepsy seizure detection using 5-level Db4 discrete wavelet transform, GA-based feature selection and ANN/SVM classifiers," Computer Science Journal of Ambient Intelligence and Humanized Computing, 10395–10403, 2021, https://doi.org/10.1007/s12652-020-02837-8

[25] L. Breiman, "Random forests," Machine learning, 45(1), 5-32, 2001.

[26] J. L. M. Kumar, M. Rashid, R. M. Musa, M. A. M. Razman, N. Sulaiman, R. Jailani, A. P. A. Majeed, "The classification of EEG-based winking signals: a transfer learning and random forest pipeline," PeerJ, 9, e11182, 2021, DOI: 10.7717/peerj.11182

[27] U. Azmat, Y. Y. Ghadi, T. A. Shloul, S. A. Alsuhibany, A. Jalal, J. Park, "Smartphone sensor-based human locomotion surveillance system using multilayer perceptron," Special Issue Artificial Intelligence and Beyond in Medical and Healthcare Engineering, 12(5), 2550, 2022, https://doi.org/10.3390/app12052550.

[28] M. Kivrak, "Deep learning-based prediction of obesity levels according to eating habits and physical Condition," The Journal of Cognitive Systems (JCS), Vol. 6. No.1, 2021, https://doi.org/10.52876/jcs.939875

[29] B. Singh, H. Tawfik, "A Machine Learning Approach for Predicting Weight Gain Risks in Young Adults," Conf. Proc. 10th Int. Conf. Dependable Syst. Serv. Technol, pp. 231–234, 2019, DOI: 10.1109/DESSERT.2019.8770016.

[30] N. Daud, N. L. M. Noor, S. A. Aljunid, N. Noordin, N. I. M. Fahmi Teng, "Predictive Analytics: The Application of J48 Algorithm on Grocery Data to Predict Obesity," IEEE Conf. Big Data Anal, pp. 1–6, 2019, DOI: 10.1109/ICBDAA.2018.8629623.

[31] A. Alqahtani, F. Albuainin, R. Alrayes, N. A. Muhanna, E. Alyahyan, E. Aldahasi, "Obesity level prediction based on data mining techniques," IJCSNS International Journal of Computer Science and Network Security, Vol.21 No.3, 2021, https://doi.org/10.22937/IJCSNS.2021.21.3.14

[32] S. Huang, "Obesity prediction based on logistic regression, random forest and support vector machine," A thesis submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Science in the

Department of Biostatistics in the Gillings School of Global Public Health, 2022.

[33] T. Cui, Y. Chen, J. Wang, H. Deng, Y. Huang, “Estimation of Obosity Levels Based on Decision Trees,” International Symposium on Artificial Intelligence and Its Application on Media, 2021.

[34] K. Jindal, N. Baliyan, P. S. Rana, “Obesity prediction using ensemble machine learning approaches,” Recent Findings in Intelligent Computing Techniques. Springer, pp. 355–362, 2018, DOI: 10.1007/978-981-10-8636-6_37

[35] L. F. Lopes, A. C. Oliveira, R. X. Brito, S. A. F. Oliveira, L. T. R. Neto, “Estimativa Dos N´ıveis de Obesidade Com Base em Habitos Alimentares e Condic´ao F˜´ısica Atraves de ´Tecnicas de Aprendizado de M´aquina,” Anais Estendidos Do XXXIV Conference on Graphics, Patterns and Images, 2021.

[36] A. Subaşı, M. I. Gürsoy, “EEG Signal classification using PCA, ICA, LDA and support vector machines,” Expert Systems with Applications, 37: 8659-8666, 2010, https://doi.org/10.1016/j.eswa.2010.06.065

[37] K. S. Abakar, C. Yu, “Performance of SVM based on PUK kernel in comparison to SVM based on RBF kernel in prediction of yarn tenacity,” Indian Journal of Fibre&tectile Research, 39: 55-59, 2014.

[38] C. Zhang, X. Pan, H. Li, A. Gardiner, I. Sargent, J. Hare, P. M. Atkinson, “A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification,” ISPRS J. Photogramm. Remote Sens 140: 133-44, 2018, https://doi.org/10.1016/j.isprsjprs.2017.07.014

[39] T. Windeatt, “Ensemble MLP classifier design,” Computational Intelligence Paradigms, 2008: 133-147, Springer, Berlin, Heidelberg, 2008

[40] A. Liaw, M. Wiener, “Classification and Regression by Random Forest,” R News, 2 (December), 18–22, 2002, https://doi.org/10.1177/154405910408300516

[41] E. C. Gunther, R. W. Gerwien, M. P. Heyes, P. Bento, D. J. Stone, ”Prediction of clinical drug efficacy by classification of drug-induced genomic expression profiles in vitro,” Proceedings of the National Academy of Sciences, 100(16), 9608-9613, 2003, https://doi.org/10.1073/pnas.1632587100

[42] M. R. Segal, “Machine Learning Benchmarks and Random Forest Regression,” Center for Bioinformatics and Molecular Biostatistics, UC San Francisco, 18(3), 1-14, 2003.

[43] M. Pal, “Random forest classifier for remote sensing classification,” International Journal of Remote Sensing, 26(1), 217–222, 2005, https://doi.org/10.1080/01431160412331269698