ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20211119133806

Original scientific paper
Received: 2021-11-19, Accepted: 2022-08-26

Development of Early Stage Diabetes Prediction Model Based on Stacking Approach

Ilkay Cinar, Yavuz Selim Taspinar, Murat Koklu*

Abstract: Diabetes is a disease that may pose direct or indirect risks in terms of human health. Early diagnosis can minimize the potential harm of this disease to the body and reduce the probability of death. For this reason, laboratory tests are performed on diabetic patients. The analysis of these tests enables the diagnosis of diabetes. The aim of this study is so quickly diagnose diabetes by using data obtained from patients with machine learning methods. In order to diagnose the disease, k-nearest neighbor (k-NN), logistic regression (LR), random forest (RF) models and the stacking meta model which is created by combining these three models were used. The dataset used in the research includes test samples taken from 520 people. The dataset has 17 features, including 16 input features and 1 output feature. As a result of the classification through this dataset, different classification results were obtained from the models. The classification success of the models LR, k-NN, RF and stacking were found to be 91.3%, 91.7%, 97.9% and 99.6%, respectively. F-score, precision and recall performance metrics were utilized for a detailed analysis of the models' classification results. The obtained results revealed that the stacking model has a sufficient level to be used as a decision support system in the early diagnosis of diabetes.

Keywords: decision support system; diabetes; early stage; machine learning; stacking

1 INTRODUCTION

Diabetes is a disease in which insulin cannot be produced by the pancreas, or insulin sufficiently-produced in the pancreas cannot be used by the human body. Insulin, a type of hormone produced by the pancreas, plays a key role in transferring glucose from consumed nutrients to blood cells in the body and then converting it into the energy. When the body is unable to produce insulin, the level of glucose in the blood increase. High levels of glucose in the blood, on the other hand, can be detrimental to the viscera and lead to dysfunction in the tissue.

Diabetes is generally treated in three subheadings as type 1, type 2, and gestational diabetes. Gestational diabetes is a type of diabetes that occurs during pregnancy only because of the hormonal changes. Common symptoms of diabetes mellitus are polyuria, polydipsia, polyphagia, sudden weight loss, being underweight, obesity, pruritus, delayed recovery, blurred vision, genital thrush, nervousness, muscle stiffness, and etc. [3-5]. Early diagnosis of diabetes is essential for taking preventive measures. Besides, effective treatment at the first stage of the disease will always have additional benefits for patients [6].

Diagnosing diabetes through medical testing may not provide confident results due to the clinical complexity, time-consuming process, and high expenses. On the other hand, thanks to the machine learning algorithms, a disease such as diabetes can be predicted in a short time with lower costs [7]. Machine learning, a sub-branch of artificial intelligence (AI), relates to the development of algorithms and techniques that enable computers to learn based on the past experiences. In other words, the system can define and understand the input data and accordingly make decisions, predictions, and classifications [5, 8].

The contributions of this article can be summarized as follows:

- Studies have been carried out for the early diagnosis of diabetes by using the Early Stage Diabetes Risk Prediction dataset within The University of California, Irvine (UCI) repository of machine learning databases.
- For the early diagnosis of diabetes, transfer learning has been applied by utilizing deep learning architectures VGG16 and VGG19.
- The results obtained through transfer learning have been shared.
- The results obtained based on the literature studies carried out using the Early Stage Diabetes Risk Prediction dataset have been compared.

The remaining parts of the article have been organized as follows: The second chapter includes previous studies that focus on diabetes disease prediction via using machine learning algorithms and have significance in terms of the literature. The third chapter covers the description of the dataset, the research methods, and the explanations on performance metrics. The fourth chapter includes experimental results. And lastly, in the fifth chapter, the results and discussion are presented.

2 RELATED WORKS

Kandhasamy and Balamurali [3] compared the performances J48 decision tree (DT), k-nearest neighbors (K-NN), random forest (RF) and support vector machines (SVM) algorithms in order to classify diabetic patients. The results of the study indicated that the random forest algorithm has the higher classification accuracy compared to the other algorithms.

Perveen et al. [9] conducted a study on the prediction of the disease by using diabetes risk factors. The experimental results of the study in which J48 decision tree, Adaboost and Bagging algorithms were utilized to perform classification operations showed that Adaboost algorithm provides better results than J48 decision tree and Bagging algorithms.

In the study conducted by Husain and Khan [10], the distinctive performances of the ensemble learning model were investigated, for prediction of diabetes at an early stage. An ensemble model has been developed by combining these algorithms to improve the overall prediction accuracy by using different machine learning algorithms. 0.75 AUC and

an accuracy of 96% has been ensured by the model developed after the classification process.

Sisodia and Sisodia [2] aimed at developing a model which can predict the probability of diabetes in patients, with the maximum accuracy. Within the scope of their study, three machine learning classification algorithms were used: DT, SVM and Naive Bayes (NB). As a conclusion, it was figured out that the Naive Bayes has a better performance compared to other algorithms with an accuracy rate of 76.30%.

Alehegn et al. [11] proposed an ensemble method for predicting diabetes. The results obtained with the proposed method were compared with the results of the most common machine learning methods. The proposed method was found to have the highest classification accuracy with a rate of 90.36%.

In their studies, Choudhury and Gupta [12] aimed to detect diabetes by using machine learning techniques such as RF, NB, K-NN, SVM, LR, DT and ANN. The results of the study revealed that the highest classification accuracy (77.61%) among all algorithms belongs to LR.

Alam et al. [13] conducted studies to distinguish the most important qualities for the predictions of diabetes disease and make classification of the disease. In order to determine important qualities, principal component analysis (PCA) method was used in the study. Moreover, artificial neural network (ANN), RF, and K-means clustering methods were utilized for the classification processes. As a result of the classification, it is determined that ANN has the highest accuracy value with a rate of 75.7%.

For the diagnosis of diabetes, Challa and Chinnaiyan [14] used the classification algorithms of DT, SVM, K-NN and RF within the scope of their studies. At a rate of 78.25%, the highest classification accuracy was obtained with the DT.

Rajni and Amandeep [4], by using RB-Bayes, proposed a model to determine whether the person has diabetes or not. Furthermore, they performed classification through SVM, NB and K-NN algorithms and compared with the model they proposed. The results of the classification processes showed that the highest classification accuracy belongs to the proposed RB-Bayes model with the rate of 72.9%.

Kowsher et al. [5], in order to detect early diabetic patients, utilized deep neural networks (ANN) together with seven machine learning algorithms and compared their results. Deep ANN has achieved the highest classification accuracy in detecting diabetic patients. As a result of the classification, an accuracy of 95.14% was obtained.

In their studies, Ayon and Islam [15] used deep ANN in order to effectively detect diabetic patients. They compared the results using 5-fold and 10-fold cross-validation during the training of the neural network. The classification accuracy obtained after 5-fold cross-validation was 98.35%, while the accuracy was found to be 97.11% after 10-fold cross-validation. Following the experimental results, it is figured out that the proposed system provides promising results with 5-fold cross-validation.

Le Minh et al. [16] proposed a model to predict the early onset of diabetes disease. They used Multi-Layer Perceptron (MLP) to reduce the number of input features, while they use Gary Wolf Optimizer (GWO) and Adaptive Particle Swarm Optimization (APSO) to optimize the number of input features. The proposed method provided 97% accuracy for APGWO-MLP.

Harz et al. [17] used artificial neural networks to predict whether a person has diabetes or not. They achieved a prediction accuracy of 98.73% as a result of the study.

Yucelbas and Yucelbas [18] aimed to determine which features effective in the early diagnosis of diabetes, according to gender. It is indicated that the weakness feature was not effective on 108 current research results and that the classification accuracy obtained before the selection of male subjects was found to be 97.86%, with 13 features.

In the study conducted by Ridwan [1], an accuracy of 90.20% and an AUC value of 0.95 were obtained by using the Naive Bayes (NB) classification method.

Hana [19], used neural network and linear discriminant analysis (LDA) algorithm to analyze diabetic patients. While an accuracy of 90.38% was achieved with the LDA algorithm, a classification accuracy of 95.19% was achieved with the neural network.

Kaur and Kumari [8], in the study they conducted, used the R data manipulation tool to develop trends and identify risk factors and patterns. SVM, Radial-Basis Function (RBF) Kernel Support Vector Machine, k-NN, ANN and Multi-factor Dimensionality Reduction (MDR) algorithms was used in order to classify patients as diabetic and non-diabetic. As a result, the highest classification accuracy was achieved in Linear Kernel SVM with 89%.

Abd Rahman et al. [20] aimed to develop a prediction model using three different machine learning algorithms to classify Type 2 diabetes mellitus (T2DM) of the Malaysian population. DT, SVM and NB were used as classification algorithms and as a result, in terms of accuracy (0.87), sensitivity (0.9), specificity (0.8), sensitivity (0.9), F1 score (0.9), and AUC value (0.93), the best overall prediction performance was achieved with the random forest algorithm.

Tripathi and Kumar [21] carried out a study to predict diabetes at an early stage, by utilizing LDA, K-NN, SVM and RF machine learning algorithms. It was observed that the RF algorithm, which achieved a maximum accuracy of 87.66% after the classification processes, performed better than the other algorithms used.

Naz and Ahuja [22] presented a methodology aimed at diabetes prediction using machine learning algorithms for the early diagnosis of diabetes. In the study, the classification processes was performed by using the NB, DT, ANN and Deep Learning (DT) algorithms. As a result, DL achieved the highest classification accuracy of 98.07%.

Hana [23] performed the classification process by the C 4.5 decision tree algorithm to detect diabetes. As a result, a classification accuracy of 93.02% was achieved.

3 MATERIAL AND METHODS

3.1 Database

Within the scope of this study, early-stage diabetes risk estimation dataset was used. This data set consists of a total of 520 people, 320 of whom are diabetic and 200 of whom are non-diabetic [24]. The dataset has a total of 17 features, including 16 input features and 1 output feature. In order for the data to be processed more easily, changes have been made in the values belonging to the features in a way that not

affect the classification result. The features and values the dataset includes are given in Tab. 1.

Table 1 Features and values of dataset

Features	Values	Conversion
Age	16 - 90	
Gender	Male, Female	Male=1, Female=0
Polyuria	Yes, No	Yes=1, No=0
Polydipsia	Yes, No	Yes=1, No=0
Sudden Weight Loss	Yes, No	Yes=1, No=0
Weakness	Yes, No	Yes=1, No=0
Polyphagia	Yes, No	Yes=1, No=0
Genital Thrush	Yes, No	Yes=1, No=0
Visual Blurring	Yes, No	Yes=1, No=0
Itching	Yes, No	Yes=1, No=0
Irritability	Yes, No	Yes=1, No=0
Delayed Healing	Yes, No	Yes=1, No=0
Partial Paresis	Yes, No	Yes=1, No=0
Muscle Stiffness	Yes, No	Yes=1, No=0
Alopecia	Yes, No	Yes=1, No=0
Obesity	Yes, No	Yes=1, No=0
Class	Positive, Negative	Positive=1, Negative=0

While the trainings were carried out, 16 features were given as input and 1 output feature was given as output to the models. Data pre-processing was not performed since there is no missing data in the dataset that would affect the classification success of the models.

3.2 k-Nearest Neighbor

k-NN is a machine learning method based on calculating the distances between data in the dataset [25]. The distance of an object to its neighbors is calculated according to a specified parameter which is called "*k*" and indicates the number of neighbors. Objects are divided into classes according to the specified number of neighbors. Different distance determination methods are used in determining the distance between objects [26]. The k value was determined as 5 in this study, while Euclidean distance was used to determine the distance between objects.

3.3 Logistic Regression

LR algorithm is a machine learning method that can be used in classification problems [27]. There is a linear relationship between the dependent and independent variables in linear regression, hence it is used in the solution of single input and single output problems. Due to this limitation of linear regression, logistic regression algorithm is used. In logistic regression, many independent variables are used to predict the dependent variable, which is the output variable. It is not necessary for the independent variables, i.e. input variables, to be evenly distributed [28].

3.4 Random Forest

RF algorithm is an ensemble machine learning method that contains many decision trees. Each decision tree it contains performs a query on objects randomly taken from the dataset, and the object is placed on a node. Results from each decision tree are voted. The most suitable class for the object is determined as a result of voting the estimates from the trees [29].

3.5 Stacking

Stacking is an ensemble machine learning method that can classify data with results from different classifiers and a training result within itself. The model emerged by the results from different models and the result of training within itself is called the meta-model. The meta-model is expected to provide more successful results than the classifiers that comprise it. However, sometimes, it may give lower results since the classification ability is impacted by many parameters [30]. In the study, the Stacking meta-model was created through from the results of k-NN, LR and RF methods and the result of the training within the model.

3.6 Confusion Matrix

Confusion matrix is a table which is used to see the classification numbers of data samples in the solution of classification problems with machine learning methods. The correctly and incorrectly classified data belong the classes can be reached by using the data in this table [31]. In the table, four data exist as true positive (*TP*), true negative (*TN*), false positive (*FP*) and false negative (*FN*). Accordingly,

- True Positive (*TP*) represents the true classified positive samples,
- True Negative (*TN*) represents the true classified negative samples,
- False Positive (*FP*) represents the false classified positive samples,
- False Negative (*FN*); represents the false negative samples [32].

Tab. 2 shows the placement of these values on the matrix.

Table 2 Confusion matrix

		PREDICTED	
		Negative	Positive
ACTUAL	Negative	*TN*	*FP*
	Positive	*FN*	*TP*

3.7 Performance Evaluation Metrics

For the detailed performance evaluation of the models trained with the data in the dataset, there are also different metrics other than the classification success [33]. F-score, precision, and recall metrics are the other metrics utilized for evaluation of the success of the model [34]. The F-score is a measurement metric which will include all error cost, not just

misclassified samples. The class variable is used to evaluate the performance of models in datasets that are not evenly distributed. It is calculated by taking the harmonic average of precision and recall values [35]. Precision is a metric used to see how many samples classified as true and false positives are actually positive. Recall, on the other hand, is the metric showing how many of the samples that should be predicted positively were classified as positive [36]. The four performance metrics used in the study are calculated by the formulas in Tab. 3.

Table 3 Performance metrics equations

Metrics	Equation
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
F-score	$2\times\frac{Precision\times Recall}{Precision+Recall}$
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$

These four metric values represent the success of the model. The higher the value, the higher the success of the model [37].

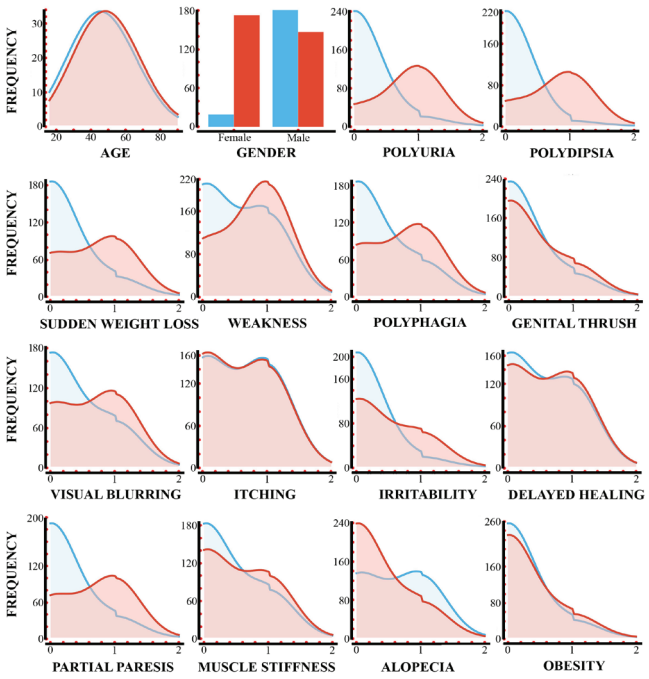


Figure 1 The data distribution charts

4 EXPERIMENTAL RESULTS

Thanks to the data distribution charts, preliminary information can be obtained about their classification success before the models are trained. The distribution ratio of class values also enables to make decision about which kind of testing procedures to be performed on models. The repetition status of each feature in the dataset, that is, the data distribution graphs created according to its frequency are shown in Fig. 1. The result value indicated in blue color indicates negative (0), and the result value indicated in red color indicates positive (1). The output demonstrated in blue color indicates that the value is negative (0), and the output demonstrated in red color indicates that the value positive (1).

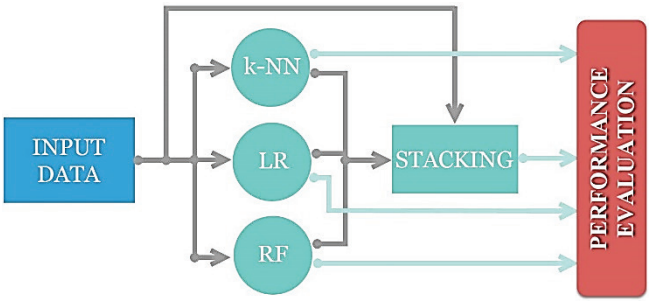
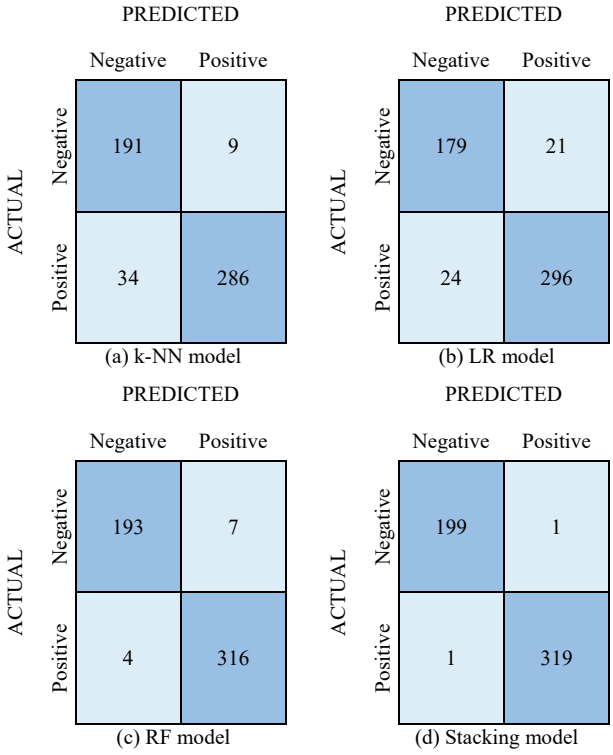


Figure 2 Flow chart of the performed processes

Output values are 200 negative (0), 320 positive (1). After analyzing the data distributions, the training of the models was carried out. The classification processes were performed with the k-NN, LR, RF models, and the Stacking model created by combining these models. The flow chart of the processes performed with the models is given in Fig. 2.

Table 4 Confusion matrix of all models



In order to make a comparison between the accuracies of different classification models, in the study, performance measurements were carried out on the dataset used. The cross-validation method was utilized to obtain a standard in classification and to get rid of the subjectivity of the classification methods performed with the train-test distinction. In cross-validation method, the dataset is divided

into *k* parts. Each part is utilized as a validation set. The remaining *k*–1 part is used for training the algorithm. The average of the success rates obtained as a result of these processes performed k times gives the classification success of the algorithm. In this way, classification success can be measured objectively. The value of k was determined as 10 in the training with a dataset containing 520 lines of data. Confusion Matrices of all methods used are given in order. The performances of the classifiers were compared by using confusion matrix values. Confusion Matrices obtained as a result of the classification with all models are shown in Tab. 4.

In Tab. 4 (a), with the k-NN model, 43 data in total were classified incorrectly. In Tab. 4(b), with the LR model, 45 data in total were classified incorrectly, while 475 data were classified correctly. In Tab. 4(c), 11 data were classified incorrectly and 509 data were classified correctly with the RF model. In Tab. 4(d), with the Stacking meta model based on k-NN, LR and RF models, 2 data were classified incorrectly, while 518 data were correctly classified. The results obtained as a result of the statistical calculations by the data of Confusion Matrix are included in Tab. 5.

Table 5 Performance metrics of models

	F-Score	Precision	Recall
k-NN	0.918	0.923	0.917
LR	0.914	0.914	0.913
RF	0.982	0.973	0.987
Stacking	**0.996**	**0.996**	**0.996**

Classification success rates of LR, k-NN, RF models and the Stacking meta model which is a combination of these models are shown in Tab. 6.

Table 6 Classification success rates of models

	LR	k-NN	RF	Stacking
Accuracy	91.3%	91.7%	97.9%	**99.6%**

According to Tab. 6, LR model has the lowest classification success with a rate of 91.3% while the highest classification success belongs to the Stacking Meta Model with 99.6%. The Stacking Model has the highest F-score value with 0.966, precision 0.996 and recall 0.996 values. The lowest F-score value, on the other hand, belongs to the LR Model with 0.914, precision 0.914, recall 0.913 values. Fig. 3 gives the success rates of the models.

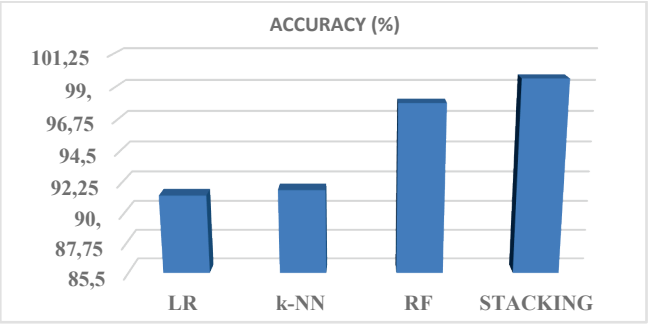


Figure 3 Accuracy of classification models

5 CONCLUSION

The changes in people's diet and lifestyle, may bring about an increase also in the diseases caused by diabetes besides the increase in diabetes disease in the society. Early diagnosis ensures to start treatment of diseases earlier and to halt the diseases' progression. With the classification models formed via using the early diagnosis diabetes dataset created for this purpose, it can be possible to detect diabetes at an early stage. Within the scope of this study, classification processes have been completed for the early diagnosis of diabetes. k-NN, LR, RF and the Stacking meta model created by combining these 3 models were used in classification. The classification successes obtained with these models are 91.7%, 91.3%, 97.9% and 99.4%, respectively. When the success rates are examined, it is understood that the Stacking Model has the highest classification success. The fact that the Stacking Model has a higher success compared to other methods is due to the models that make up this model classify the data correctly and incorrectly. In addition, models become able to classify the data more accurately when they are combined. It was also observed that the classification success of the Stacking Meta-Model, which was created using the models with different *FN* and *FP* classification numbers, is higher. A higher classification success was obtained compared to the results obtained in literature studies using the same data set. The comparison of the proposed model with the models in other studies is given in Tab. 7.

Table 7 Studies conducted by using the Early Stage Diabetes Risk Prediction dataset

Method	Accuracy (%)	References
NB	90.20	[1]
C4.5 DT	93.02	[23]
ANN	95.19	[19]
APGWO-MLP	97.00	[16]
k-NN	97.86	[18]
ANN	98.73	[17]
Proposed Stacking Model	**99.6%**	

The Stacking Model, which has the highest classification success, can be used as a decision support system in the early diagnosis of diabetes. Achieving 100% success in the field of health is always a desirable conclusion. It is thought that the success of classification can be increased via different machine learning approaches.

6 REFERENCES

[1] Ridwan, A. (2020). Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus. *Jurnal SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan), 4*(1), 15-21. https://doi.org/10.47970/siskom-kb.v4i1.169

[2] Sisodia, D. & Sisodia, D. S. (2018). Prediction of diabetes using classification algorithms. *Procedia computer science*, 132, 1578-1585. https://doi.org/10.1016/j.procs.2018.05.122

[3] Kandhasamy, J. P. & Balamurali, S. (2015). Performance analysis of classifier models to predict diabetes mellitus. *Procedia Computer Science*, 47, 45-51. https://doi.org/10.1016/j.procs.2015.03.182

[4] Rajni, R. & Amandeep, A. (2019). RB-Bayes algorithm for the prediction of diabetic in Pima Indian dataset. *International Journal of Electrical Computer Engineering, 9*(6), 4866. https://doi.org/10.11591/ijece.v9i6.pp4866-4872

[5] Kowsher, M., Turaba, M. Y., Sajed, T., & Rahman, M. M. (2019, December). Prognosis and treatment prediction of type-2 diabetes using deep neural network and machine learning classifiers. *The 22nd International Conference on Computer and Information Technology (ICCIT),* 1-6, IEEE. https://doi.org/10.1109/ICCIT48885.2019.9038574

[6] Jain, D. & Singh, V. (2018). Feature selection and classification systems for chronic disease prediction: A review. *Egyptian Informatics Journal, 19*(3), 179-189. https://doi.org/10.1016/j.eij.2018.03.002

[7] Kahramanli, H. & Allahverdi, N. (2008). Design of a hybrid system for the diabetes and heart diseases. *Expert systems with applications, 35*(1-2), 82-89. https://doi.org/10.1016/j.eswa.2007.06.004

[8] Kaur, H. & Kumari, V. (2020). Predictive modelling and analytics for diabetes using a machine learning approach. *Applied computing and informatics.* https://doi.org/10.1016/j.aci.2018.12.004

[9] Perveen, S., Shahbaz, M., Guergachi, A., & Keshavjee, K. (2016). Performance analysis of data mining classification techniques to predict diabetes. *Procedia Computer Science*, 82, 115-121. https://doi.org/10.1016/j.procs.2016.04.016

[10] Husain, A. & Khan, M. H. (2018). Early diabetes prediction using voting based ensemble learning. In *International conference on advances in computing and data sciences*. 95-103, Springer. https://doi.org/10.1007/978-981-13-1810-8_10

[11] Alehegn, M., Joshi, R., & Mulay, P. (2018). Analysis and prediction of diabetes mellitus using machine learning algorithm. *International Journal of Pure Applied Mathematics, 118*(9), 871-878.

[12] Choudhury, A. & Gupta, D. (2019). A survey on medical diagnosis of diabetes using machine learning techniques, in *Recent developments in machine learning and data analytic,* 67-78, Springer. https://doi.org/10.1007/978-981-13-1280-9_6

[13] Alam, T. M., Iqbal, M. A., Ali, Y., Wahab, A., Ijaz, S., Baig, T. I. ... & Abbas, Z. (2019). A model for early prediction of diabetes. *Informatics in Medicine Unlocked*, 16, 100204. https://doi.org/10.1016/j.imu.2019.100204

[14] Challa, M. & Chinnaiyan, R. (2019, September). Optimized Machine Learning Approach for the Prediction of Diabetes-Mellitus. In *International Conference on Computational Vision and Bio Inspired Computing,* 321-328, Springer. https://doi.org/10.1007/978-3-030-37218-7_37

[15] Ayon, S. I. & Islam, M. (2019). Diabetes Prediction: A Deep Learning Approach. *International Journal of Information Engineering Electronic Business, 11*(2). https://doi.org/10.5815/ijieeb.2019.02.03

[16] Le, T. M., Vo, T. M., Pham, T. N., & Dao, S. V. T. (2020). A Novel Wrapper–Based Feature Selection for Early Diabetes Prediction Enhanced with a Metaheuristic. *IEEE Access*, 9, 7869-7884. https://doi.org/10.1109/ACCESS.2020.3047942

[17] Harz, H. H., Rafi, A. O., Hijazi, M. O., & Abu-Naser, S. S. (2020). Artificial Neural Network for Predicting Diabetes Using JNN. *International Journal of Academic Engineering Research (IJAER), 4*(10).

[18] Yucelbas, C. & Yucelbas, S. (2020). Determination of Effective Attributes in the Early Diagnosis of Gender-Based Diabetes. *Current Researches in Architecture Engineering Sciences,* 97.

[19] Hana, F. M. (2020). Perbandingan Algoritma Neural Network dengan linier discriminant analysis (LDA) pada klasifikasi penyakit diabetes. *Jurnal Bisnis Digital Dan Sistem Informasi*, *1*(1), 21-29.

[20] Abd Rahman, M. H. F., Salim, W. W. A. W., & Abd Wahab, M. F. (2020). Risk prediction analysis for classifying type 2 diabetes occurrence using local dataset. *Biological and Natural Resources Engineering Journal, 3*(1), 48-61.

[21] Tripathi, G. & Kumar, R. (2020, June). Early prediction of diabetes mellitus using machine learning. *In 2020 8th International Conference on Reliability, INFOCOM Technologies and Optimization (Trends and Future Directions) (ICRITO),* 1009-1014, IEEE. https://doi.org/10.1109/ICRITO48877.2020.9197832

[22] Naz, H. & Ahuja, S. (2020). Deep learning approach for diabetes prediction using PIMA Indian dataset. *Journal of Diabetes & Metabolic Disorders, 19*(1), 391-403. https://doi.org/10.1007/s40200-020-00520-5

[23] Hana, F. M. (2020). Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4. 5. *Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan), 4*(1), 32-39. https://doi.org/10.47970/siskom-kb.v4i1.173

[24] Islam, M. F., Ferdousi, R., Rahman, S., & Bushra, H. Y. (2020). Likelihood prediction of diabetes at early stage using data mining techniques. *In Computer Vision and Machine Intelligence in Medical Image Analysis,* 113-125. Springer, Singapore. https://doi.org/10.1007/978-981-13-8798-2_12

[25] Tasdemir, S., Saritas, I., Ciniviz, M., & Allahverdi, N. (2011). Artificial neural network and fuzzy expert system comparison for prediction of performance and emission parameters on a gasoline engine. *Expert Systems with Applications, 38*(11), 13912-13923. https://doi.org/10.1016/j.eswa.2011.04.198

[26] Aslan, M. F., Durdu, A., & Sabanci, K. (2020). Human action recognition with bag of visual words using different machine learning methods and hyperparameter optimization. *Neural Computing and Applications, 32*(12), 8585-8597. https://doi.org/10.1007/s00521-019-04365-9

[27] Cruyff, M. J., Böckenholt, U., Van Der Heijden, P. G., & Frank, L. E. (2016). A review of regression procedures for randomized response data, including univariate and multivariate logistic regression, the proportional odds model and item response model, and self-protective responses. *Handbook of Statistics,* 34, 287-315. https://doi.org/10.1016/bs.host.2016.01.016

[28] Kalantar, B., Pradhan, B., Naghibi, S. A., Motevalli, A., & Mansor, S. (2018). Assessment of the effects of training data selection on the landslide susceptibility mapping: a comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN). *Geomatics, Natural Hazards and Risk, 9*(1), 49-69. https://doi.org/10.1080/19475705.2017.1407368

[29] Speiser, J. L., Miller, M. E., Tooze, J., & Ip, E. (2019). A comparison of random forest variable selection methods for classification prediction modeling. *Expert systems with applications,* 134, 93-101. https://doi.org/10.1016/j.eswa.2019.05.028

[30] Ghalejoogh, G. S., Kordy, H. M., & Ebrahimi, F. (2020). A hierarchical structure based on stacking approach for skin lesion classification. *Expert Systems with Applications,* 145, 113127. https://doi.org/10.1016/j.eswa.2019.113127

[31] Ozkan, I. A. (2020). A Novel Basketball Result Prediction Model Using a Concurrent Neuro-Fuzzy System. *Applied Artificial Intelligence, 34*(13), 1038-1054. https://doi.org/10.1080/08839514.2020.1804229

[32] Saritas, M. M. & Yasar, A. (2019). Performance analysis of ANN and Naive Bayes classification algorithm for data

classification. *International Journal of Intelligent Systems and Applications in Engineering, 7*(2), 88-91. https://doi.org/10.18201//ijisae.2019252786

[33] Kocer, S. & Tumer, A. E. (2017). Classifying neuromuscular diseases using artificial neural networks with applied Autoregressive and Cepstral analysis. *Neural Computing and Applications, 28*(1), 945-952. https://doi.org/10.1007/s00521-016-2383-8

[34] Saura, J. R. (2021). Using data sciences in digital marketing: Framework, methods, and performance metrics. *Journal of Innovation & Knowledge, 6*(2), 92-102. https://doi.org/10.1016/j.jik.2020.08.001

[35] Chicco, D. & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics, 21*(1), 1-13. https://doi.org/10.1186/s12864-019-6413-7

[36] Cinarer, G., Emiroglu, B. G., & Yurttakal, A. H. (2020). Prediction of Glioma Grades Using Deep Learning with Wavelet Radiomic Features. *Applied Sciences, 10*(18), 6296. https://doi.org/10.3390/app10186296

[37] Saritas, I. (2012). Prediction of breast cancer using artificial neural networks. *Journal of Medical Systems, 36*(5), 2901-2907. https://doi.org/10.1007/s10916-011-9768-0

Authors' contacts:

Ilkay Cinar
Department of Computer Engineering, Selcuk University,
Alaaddin Keykubat Campus, 42075 Konya, Turkey
ilkay.cinar@selcuk.edu.tr
ORCID: 0000-0003-0611-3316

Yavuz Selim Taspinar
Doganhisar Vocational School, Selcuk University
Alaaddin Keykubat Campus, 42075 Konya, Turkey
ytaspinar@selcuk.edu.tr
ORCID: 0000-0002-7278-4241

Murat Koklu
(Corresponding author)
Department of Computer Engineering, Selcuk University
Alaaddin Keykubat Campus, 42075 Konya, Turkey
Tel: +90 332 223 33 48, mkoklu@selcuk.edu.tr
ORCID: 0000-0002-2737-2360