

## Possibility Prediction Of Diabetes Mellitus At Early Stage Via Stacked Ensemble Deep Neural Network

Ahmet Haşim YURTTAKAL<sup>1</sup>, Hatice BAŞ<sup>2</sup>

<sup>1</sup>Afyon Kocatepe University, Faculty of Engineering, Department of Computer Engineering, Afyonkarahisar, Turkey

<sup>2</sup>Yozgat Bozok University, Faculty of Science and Letters, Department of Biology, Yozgat, Turkey

Corresponding author e-mail: ahyurttakal@aku.edu.tr ORCID ID: <http://orcid.org/0000-0001-5170-6466>

hatice.bas@bozok.edu.tr ORCID ID: <http://orcid.org/0000-0001-8296-0360>

Received Date: 01.06.2021

Accepted Date: 18.08.2021

### Abstract

Diabetes Mellitus is a chronic metabolic disease caused by the deficiency of insulin action or secretion, or both, one of the hormones that balance the blood glucose level. It is one of the health problems that negatively affect people's quality of life. If diabetes is not detected in the early stages, it can cause serious complications such as heart and renal diseases, retinopathy, stroke, digestive disorders, and amputation. Because of the presence of a long asymptomatic period, early detection of diabetes is not realised usually. For this reason, around 50% of diabetic patients are not received a treatment due to undiagnosed at early stages. This situation results other diseases mentioned above, which diabetes causes. On the other hand, ensemble learning is a machine learning model in which multiple models are trained to solve the same problem and combined to achieve better results. Deep neural networks are one of the machine learning algorithms and they are the multi-layered state of artificial neural networks developed inspired by the information processing method of the human brain. In this study, a stacked ensemble-based deep neural network approach is proposed for diabetes possibility assessment in the early stages. The proposed approach was tested on a dataset of 520 patients. As a result, the proposed method achieved the highest success rate with 99.36% accuracy and 99.19% AUC, although the test percentage was kept higher than the prediction studies conducted on the same dataset.

© Afyon Kocatepe University

### Keywords

Diabetes Mellitus;  
Likelihood prediction;  
Stacked ensemble;  
Deep neural network

## Yığılmış Topluluk Derin Sinir Ağı Aracılığıyla Erken Evrede Diyabet Olasılık Tahmini

### Öz

Diyabet, kan glikoz düzeyini dengeleyen hormonlardan birisi olan insülin etkisinin veya salgılanmasının ya da her ikisinin eksikliğinden kaynaklanan kronik metabolik bir hastalıktır. İnsanların yaşam kalitesini olumsuz etkileyen sağlık sorunlarından biridir. Şeker hastalığı erken evrelerde tespit edilmezse kalp ve böbrekhastalıkları, retinopati, felç, sindirim bozuklukları ve amputasyon gibi ciddi komplikasyonlara neden olabilir. Uzun bir asemptomatik dönemin varlığı nedeniyle, şeker hastalığının erken teşhisi genellikle yapılamamaktadır. Bu nedenle diyabet hastalarının yaklaşık %50'si erken evrede teşhis edilemediği için tedavi alamamaktadır. Bu durum diyabetin neden olduğu yukarıda saydığımız diğer hastalıkları da beraberinde getirir. Öte yandan, toplu öğrenme, aynı sorunu çözmek için birden fazla modelin eğitildiği ve daha iyi sonuçlar elde etmek için birleştirildiği bir makine öğrenimi modelidir. Derin sinir ağları, makine öğrenme algoritmalarından biridir ve insan beyninin bilgi işleme yönteminden esinlenerek geliştirilen yapay sinir ağlarının çok katmanlı halidir. Bu çalışmada, erken aşamalarda diyabet olasılık değerlendirmesi için yığılmış topluluk tabanlı bir derin sinir ağı yaklaşımı önerilmiştir. Önerilen yaklaşım, 520 hastadan oluşan bir veri seti üzerinde test edildi. Sonuç olarak önerilen yöntem, aynı veri seti üzerinde yapılan tahmin çalışmalarına göre test yüzdesi daha yüksek tutulmasına rağmen % 99.36 doğruluk ve % 99.19 AUC ile en yüksek başarı oranını elde etmiştir.

### Anahtar kelimeler

Diyabet; Olasılık tahmini; Yığılmış topluluk; Derin sinir ağları

© Afyon Kocatepe Üniversitesi

### 1. Introduction

Diabetes Mellitus (DM) is a common disease all over the world that causes increased morbidity and mortality(Deshpande *et al.* 2008). According to the International Diabetes Federation (IDF), Diabetes is estimated to be associated with 11.3% of global deaths among people aged 20-79 years. This means that every eight seconds, a patient dies from diabetes (IDF Atlas 2019). There are different types of diabetes (Nair 2007). Type 1 is caused by insulin deficiency as a result of pancreatic  $\beta$ -cell destruction. Type 2 occurs due to insulin resistance in the tissues (Coşansu 2015). In these cases, blood glucose is not carried to the cells and blood glucose level rises above acceptable range. This condition manifests itself with symptoms such as dry mouth, weight loss, drinking too much water and eating too much. Since diabetes invites other diseases such as cardiovascular diseases, hypertension, increase in plasma lipid ratio and fatty liver, early diagnosis is very important. If blood glucose level is not kept under control, it can cause irreversible organ damage on the kidneys, eyes and nervous system, too (Kitabchi *et al.* 2009). However, all these unwanted effects can be prevented by receiving appropriate treatments and changes in lifestyle that keep blood glucose level in balance. And this is only possible with early diagnosis of diabetes.

Correct diagnosis of diseases is a difficult and time-consuming task, even with many years of medical education. In recent years, with the use of artificial intelligence in the field of health, machine learning algorithms are frequently used in the automatic, fast and easy diagnosis of diseases(Davenport 2019). Machine learning is artificial intelligence applications that can automatically learn from a group of data and detect patterns. One of the key advantages of machine learning is its ability to make consistent and high-performance predictions using complex and nonlinear relationships between features (Hastie *et al.* 2009). Machine Learning algorithms have also achieved successful results in the field of health. Some of these studies are as follows. Diagnosing breast cancer (Hu *et al.* 2020),

classifying skin lesions (Jinnai *et al.* 2020), determining diabetic retinopathy (Zhao *et al.* 2020).

Machine learning algorithms have been used to predict diabetes as well as in many diseases. Some of the datasets used in diabetes studies in the literature are Pima Indians Diabetes Database (Smith *et al.* 1988), LARS Diabetes Dataset (Efron *et al.* 2004) and Abel Vikas’s Diabetes Dataset (IntRes. 1).

In this study, the estimation of the probability of diabetes was carried out in the early stages. In the classification stage, a stack ensemble model using 2 Deep Neural Networks (DNN) in the base learner stage and a different DNN in the Meta Learner stage was proposed. The recommended method achieved the highest success rate with 99% accuracy and 99% AUC, according to the studies on the same dataset.

### 2. Material and Methods

#### 2.1 Dataset

The dataset used in the study was created by a direct questionnaire with people with suspected diabetes at Sylhet Sylhet Diabetes Hospital in Bangladesh (Islam *et al.* 2020). This dataset has been posted publicly on the UCI Machine Learning Repository (Dua and Graff 2019). There are 200 negative classes and 320 positive classes in the dataset. Figure1 shows the numbers of the dataset consisting of 520 patients according to their class numbers.

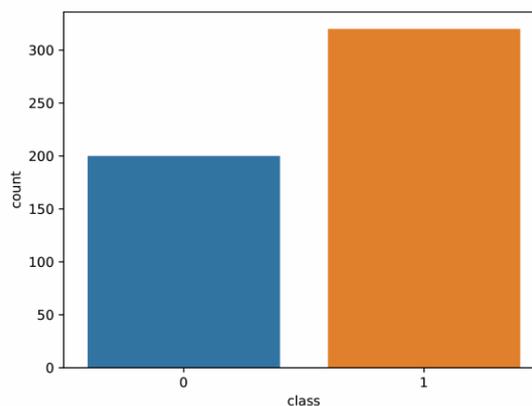


Figure 1. Count plot of dataset

In the class information, a positive status means that the patient has diabetes, while a negative status means that there is no diabetes. Table 1 shows the values of a dataset consisting of a total of 16 features, including age, gender, and 14 different symptom data.

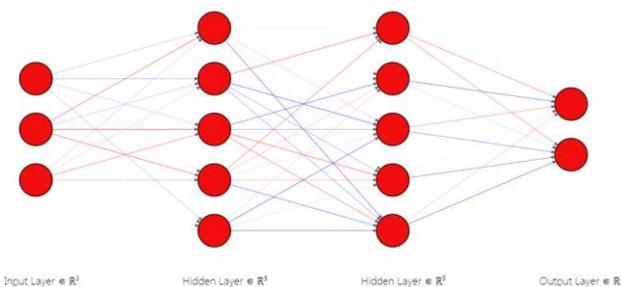
**Table 1.** Values of features (Islam *et al.* 2020)

Features	Values
Age	1.20–35, 2.36–45, 3.46–55, 4.56–65, 5.65+
Sex	1.Male, 2.Female
Polyuria	1.Yes, 2.No.
Polydipsia	1.Yes, 2.No.
Sudden weight loss	1.Yes, 2.No.
Weakness	1.Yes, 2.No.
Polyphagia	1.Yes, 2.No.
Genital thrush	1.Yes, 2.No.
Visual blurring	1.Yes, 2.No.
Itching	1.Yes, 2.No.
Irritability	1.Yes, 2.No.
Delayed healing	1.Yes, 2.No.
Partial paresis	1.Yes, 2.No.
Muscle stiffness	1.Yes, 2.No.
Alopecia	1.Yes, 2.No.
Obesity	1.Yes, 2.No.
Class	1.Positive, 2.Negative.

Age feature categorized under 5 different situations. All other features consist of 2 separate situations. In the status information, "Yes" means the symptom is present, and "No" means that there is no symptom.

**2.2 Proposed Model**

Deep Neural Networks are multi-layered versions of Artificial Neural Networks (ANN), which are inspired by the information processing method of the human brain (Awad and Khanna 2015). In the Figure 2, an example ANN is given with an input layer, two hidden layers and an output layer.



**Figure 2.** Sample ANN structure

The results produced by the network are taken from the output layer by processing the net input obtained as a result of adding the products of the information entered into the network with their own weights (W) with an activation function. Since the single layer perceptron, which was first developed and evaluated as the most primitive ANN, is not sufficient for the solution of nonlinear problems, multi-layer perceptron has been developed (Nelles 2020).

Ensemble learning is a artificial intelligence approach in which multiple machine learning algorithms are trained and combined to solve the same problem. The main idea is to reduce the variance by combining several such weak learners to create a strong learner with better performances. The most used ensemble learning approaches are bagging, boosting and stacking (Zhou 2009). The stacking ensemble unites heterogeneous weak learners by training them in a meta model (Wolpert 1992). In the study, multiple strong DNN networks in the weak learner stage were combined with another strong DNN network in the meta learner stage. This approach has been found to increase the performance of the model to a certain depth.

**2.3 Evaluation Metrics**

In order to evaluate the performance of the classification models used in machine learning, the error matrix, in which the predictions and actual values of the target attribute are compared, is frequently used. Figure 3 shows confusion matrix structure.

		Predicted labels	
		Negative (N)	Positive (P)
True labels	Negative (N)	TN	FP
	Positive (P)	FN	TP

**Figure 3.** Confusion matrix

Classification estimates have one of four evaluations called TN, TP, FN and FP. The formulas of the metrics used in the evaluation are given in the following equations (1-5).

$$Sensitivity = \frac{TP}{TP+FN} \tag{1}$$

$$Specificity = \frac{TN}{FP+TN} \tag{2}$$

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

$$Accuracy = \frac{TP+TN}{P+N} \tag{4}$$

$$F1\ Score = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{5}$$

### 3. Experimental Results

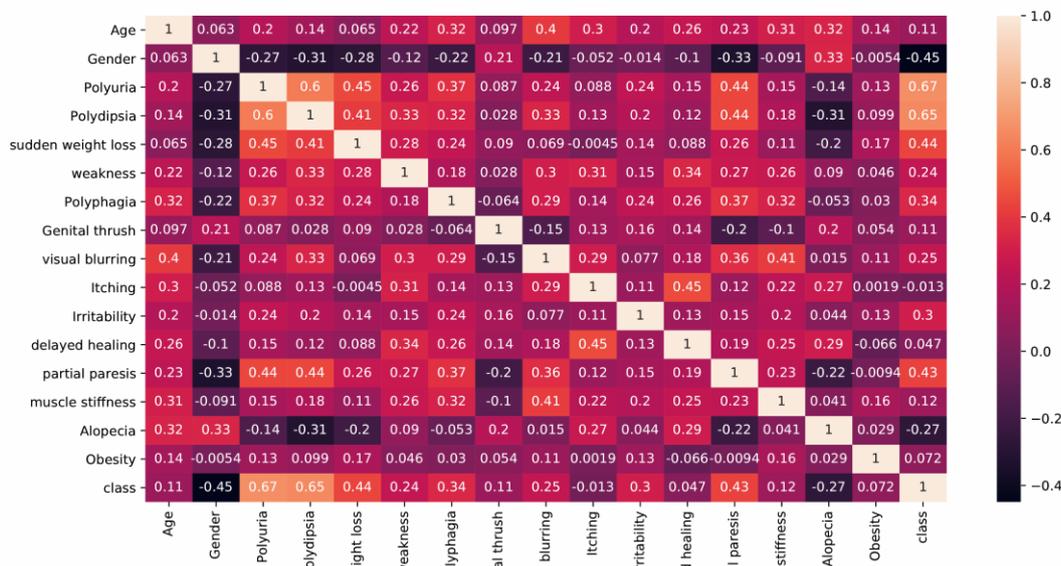
In order to prevent overfitting in training phase, the dataset is divided into two parts, 70% of which is training set and 30% test set, just before the

training. In this way, the data in the test set were not used in the training phase.

**Table 2.** Training and test set

Class	Training Set	Test Set	Total
Positive	226	94	320
Negative	138	62	200
<b>Total</b>	<b>364</b>	<b>156</b>	<b>520</b>

Correlation indicates the direction and strength of the linear relationship between two random variables in probability theory and statistics. The values of two variables with positive correlation increase or decrease together. While the value of one of the two variables with negative correlation increases, the value of the other decreases (Godfrey 1980). In Figure 4, the correlation between features was examined. According to the figure, there was a positive correlation between polyuria and polydipsia.



**Figure 4.** Correlation matrix

The number of trainable parameters of a DNN used in the study was 725 and its structure was given in Table 3.

**Table 3.** A DNN structure

Layer	Output Shape	Param
dense (Dense)	(None, 25)	425
dense_1 (Dense)	(None, 10)	260
dense_2 (Dense)	(None, 5)	55
dense_3 (Dense)	(None, 2)	12

The output layers of the two DNNs were first combined in the layer with the activation function ReLu, the output shape value of 10. Then the model was terminated to the output layer. The structure of the entire model was given in Table 4.

**Table 4.** Proposed stacked ensemble model complexity

Layer	Output Shape	Param	Connected to
dense_input (InputLayer)	[(None, 16)]	0	
dense_4_input	[(None, 10)]	0	

(InputLayer)	16))		
ensemble_1_dense (Dense)	(None, 25)	425	dense_input[0][0]
ensemble_2_dense_4 (Dense)	(None, 25)	425	dense_4_input[0][0]
ensemble_1_dense_1 (Dense)	(None, 10)	260	ensemble_1_dense[0][0]
ensemble_2_dense_5 (Dense)	(None, 10)	260	ensemble_2_dense_4[0][0]
ensemble_1_dense_2 (Dense)	(None, 5)	55	ensemble_1_dense_1[0][0]
ensemble_2_dense_6 (Dense)	(None, 5)	55	ensemble_2_dense_5[0][0]
ensemble_1_dense_3 (Dense)	(None, 2)	12	ensemble_1_dense_2[0][0]
ensemble_2_dense_7 (Dense)	(None, 2)	12	ensemble_2_dense_6[0][0]
concatenate (Concatenate)	(None, 4)	0	ensemble_1_dense_3[0][0] ensemble_2_dense_7[0][0]
dense_8 (Dense)	(None, 10)	50	concatenate[0][0]
dense_9 (Dense)	(None, 2)	22	dense_8[0][0]

The model had a total of 1576 parameters, including 72 Trainable params and 1.504 non-trainable params. The graph of the model was given in Figure 5.

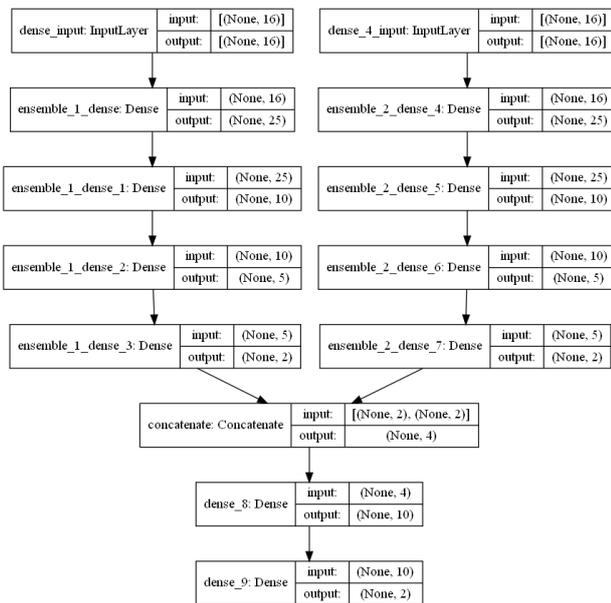


Figure 5. Proposed stacked ensemble model

Among the hyper parameters of the model, Adam was used as the optimizer, ReLu was used as the activation function in the hidden layers, and SoftMax was used as the activation function in the output layers. Figure 6 shows the accuracy value in training and test groups for 300 epochs. According to the figure, there was no overfitting during the training.

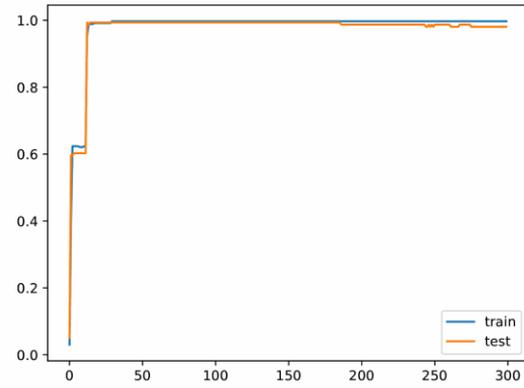


Figure 6. Training process

The success of the model was evaluated with a confusion matrix. The confusion matrix of the test set is given in Figure 7. According to the figure, out of 156 patients who were encountered for the first time after the training, the condition of 155 was estimated correctly, and only 1 patient was classified incorrectly.

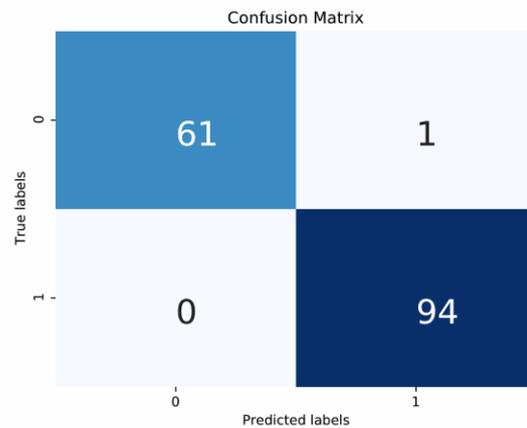


Figure 7. Confusion Matrix

One of the two DNN models had an accuracy of 0.974, while the second had an accuracy of 0.987. In Table 5, The values of the accuracy, precision, sensitivity, F1 measurement values of the stacked ensemble model created by combining the two DNN models were 0.993, 0.989, 1.000 and 0.995, respectively.

Table 5. Performance metrics

Measure	Value
Sensitivity	1.0000
Specificity	0.9839
Precision	0.9895
Accuracy	0.9936
F1 Score	0.9947

Another criterion that explains how well the model is at predicting is the ROC curve. Figure 8 shows ROC Curve. The AUC (Area Under Curve) value of the model is 0.9919.

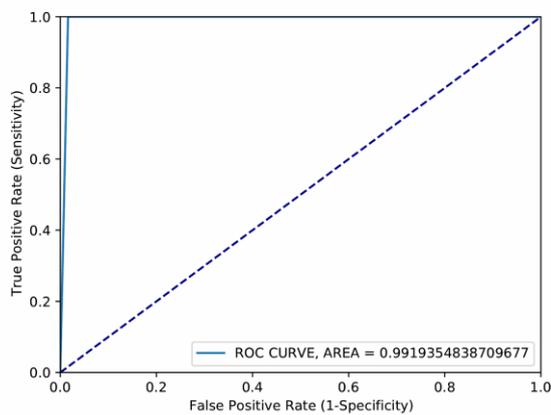


Figure 8. ROC Curve

In the proposed stacked ensemble model, attempts were made to combine 2, 3, 4, 5 DNNs, respectively. The most successful result was achieved in the combination of 2 DNNs. The values obtained in other trials are given in the Table 6.

Table 6. Other obtained values

DNNs	DNNs Acc	Number of Params	Accuracy	AUC Value
2	1.DNN Acc 0.974	1,576	0.99	0.99
	2.DNN Acc 0.987			
3	1.DNN Acc 0.936	2,348	0.98	0.98
	2.DNN Acc 0.962			
	3.DNN Acc 0.981			
4	1.DNN Acc 0.981	3,120	0.98	0.98
	2.DNN Acc 0.955			
	3.DNN Acc 0.968			
	4.DNN Acc 0.955			
5	1.DNN Acc 0.936	3,892	0.97	0.97
	2.DNN Acc 0.981			
	3.DNN Acc 0.929			
	4.DNN Acc 0.981			
	5.DNN Acc 0.974			

The accuracy values obtained with 10 fold cross validation of the proposed algorithm were given in Table 7.

Table 7. Cross Validation Results

	Loss	Accuracy (%)
Fold 1	0.016	99.423
Fold 2	0.018	99.230
Fold 3	0.010	99.807
Fold 4	0.028	98.846
Fold 5	0.015	99.615
Fold 6	0.105	96.538
Fold 7	0.032	99.230
Fold 8	0.047	98.269
Fold 9	0.019	99.423
Fold 10	0.081	96.730
Average	0.037	98.711

#### 4. Discussion and Conclusion

Diagnosing diabetes in the early stages is very important in determining the treatment processes. The use of machine learning algorithms at this stage has played an important role in the early diagnosis of diabetes.

Islam *et al.* (2019) developed an early stage diabetes diagnosis system with 16 features they determined on 520 patients, which are also used in this study. They used Naive Bayes (NB), Logistic Regression (LR), J48 and Random Forest (RF) algorithms in the classification stage and evaluated the accuracy of the model with 10k cross validation and percentage split (80:20). They obtained the highest success rate with RF in percentage split verification with 99% accuracy. Hossain *et al.* (2020) proved that ignoring a specific symptom of the rule-based ontology plan in their proposed system could affect the performance of classifiers Özer (2020) achieved an average F1 score of 98.9% as a result of 10-fold cross validation with the Long Short Term Memory (LSTM) network. The accuracy value was 98.65%. Alpan and İlgi (2020) compared the performance of 7 different machine learning algorithms using X Weka. It achieved the highest accuracy rate in K-Nearest Neighbors(K-NN) algorithm with 98.07% accuracy. Le *et al.* (2021) proposed a new meta-heuristic model named Adaptive Particle - Grey Wolf Optimizer to optimize Multilayer Perceptron (APGWO-MLP). They achieved 97% accuracy with percentage split (80:20) as a success rate.

In this study, stacked ensembles were used in prediction of diabetes at early stage. A comparison of the literature with studies using the same dataset was given in Table 8.

Table 8. Literature comparison

Study	Method	Accuracy	
		10k Cross Validation	Percentage Split
Islam et al.	RF	97.4%	%99 (80:20)
Özer	LSTM	98.65	-
Le et. al	APGWO-MLP	-	97% (80:20)
Alpan and İlgi	K-NN	98.07%	-
<b>This study</b>	<b>Stacked DNN</b>	<b>98.71%</b>	<b>99.36% (70:30)</b>

According to the results obtained, the proposed method reached 98.71% accuracy with 10 fold

cross validation. It also showed higher performance with an accuracy rate of 99.36%, although the percentage split (70:30) ratio was kept higher compared to the studies conducted on the same dataset. To our knowledge, this is the successful result obtained on the dataset. Only 1 of 156 patients allocated for testing were misclassified. Evaluating that patient as False Positive is also important in terms of health. As a result, the proposed method supports the use of artificial intelligence in the medical field. It helps physicians in the early diagnosis of diabetes quickly and accurately, independent of the user.

### Acknowledgment

We thank the M. M. Faniqul Islam, Rahatara Ferdousi, Sadikur Rahman and Humayra Yasmin Bushra for their valuable contribution to the execution of the study by providing the dataset free of charge.

### 5. References

- Alpan, K., and İlgi, G. S. 2020. Classification of Diabetes Dataset with Data Mining Techniques by Using WEKA Approach. 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), IEEE, 1-7.
- Awad M., Khanna R., 2015. Deep Neural Networks. Efficient Learning Machines. Apress, Berkeley, 127-147.
- Coşansu, G., 2015. Diyabet: Küresel bir salgın hastalık. *Okmeydanı Tıp Dergisi*, **31(ek sayı)**, 1-6.
- Davenport, T., and Kalakota, R. 2019. The potential for artificial intelligence in healthcare. *Future healthcare journal*, **6(2)**, 94-98.
- Deshpande, A. D., Harris-Hayes, M., and Schootman, M. (2008). Epidemiology of diabetes and diabetes-related complications. *Physical therapy*, **88(11)**, 1254-1264.
- Dua, D. and Graff, C., 2019. UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.
- Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. 2004. Least angle regression. *Annals of statistics*, **32(2)**, 407-499.
- Federation, International Diabetes, 2019. IDF diabetes atlas. ninth edition, *Dunia: IDF*, 1-168.
- Godfrey, K. R., 1980. Correlation methods. *Automatica*, **16(5)**, 527-534.
- Hastie, T., Tibshirani, R. and Friedman, J., 2009. The elements of statistical learning: data mining, inference, and prediction, Springer, 1-764.
- Hossain, M. A., Ferdousi, R., and Alhamid, M. F. 2020. Knowledge-driven machine learning based framework for early-stage disease risk prediction in edge environment. *Journal of Parallel and Distributed Computing*, **146**, 25-34.
- Hu, Q., Whitney, H. M., and Giger, M. L. 2020. A deep learning methodology for improved breast cancer diagnosis using multiparametric MRI. *Scientific reports*, **10(1)**, 1-11.
- Islam, M. F., Ferdousi, R., Rahman, S., and 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, Springer, 113-125.
- Jinnai, S., Yamazaki, N., Hirano, Y., Sugawara, Y., Ohe, Y., and Hamamoto, R. 2020. The development of a skin cancer classification system for pigmented skin lesions using deep learning. *Biomolecules*, **10(8)**, 1123, 1-13.
- Kitabchi, A. E., Umpierrez, G.E., Miles, J.M., and Fisher, J.N., 2009. Hyperglycemic crises in adult patients with diabetes. *Diabetes care*, **32(7)**, 1335-1343.
- Le, T. M., Vo, T. M., Pham, T. N., & Dao, S. V. T. 2021. A Novel Wrapper-Based Feature Selection for Early Diabetes Prediction Enhanced With a Metaheuristic. *IEEE Access*, **9**, 7869-7884.
- Nair, M., 2007. Diabetes mellitus, part 1: physiology and complications. *British journal of nursing*, **16(3)**, 184-188.
- Nelles O., 2020 Neural Networks. Nonlinear System Identification. Springer, Cham. 239-297.
- Özer, İ., 2020 Uzun Kısa Dönem Bellek Ağlarını Kullanarak Erken Aşama Diyabet Tahmini. *Mühendislik Bilimleri ve Araştırmaları Dergisi*, **2(2)**, 50-57.
- Smith, J. W., Everhart, J. E., Dickson, W. C., Knowler, W. C., and Johannes, R. S., 1988. Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. *Proceedings of the annual symposium on*

*computer application in medical care*, American Medical Informatics Association, 261-265.

Wolpert, D. H. 1992. Stacked generalization. *Neural networks*, **5(2)**, 241-259.

Zhao, L., Ren, H., Zhang, J., Cao, Y., Wang, Y., Meng, D., Wu, Y., Zhang, R., Zou, Y., Xu, H., Li, L., Zhang, J., Cooper, M.E., Tong, N., Liu, F., 2020. Diabetic retinopathy, classified using the lesion-aware deep learning system, predicts diabetic end-stage renal disease in Chinese patients. *Endocrine Practice*, **26(4)**, 429-443.

Zhou Z.H., 2009. Ensemble Learning, Li S.Z., Jain A. (eds) *Encyclopedia of Biometrics*. Springer, Boston, 411-416.

#### **Internet Resources**

1-<https://data.world/abelvikas/diabetes-type-dataset>(30.05.2021)