Expert Systems With Applications 115 (2019) 557–564

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Classification and diagnosis of cervical cancer with stacked autoencoder and softmax classification

Kemal Adem[a,*], Serhat Kiliçarslan[a], Onur Cömert[b]

[a] Gaziosmanpasa University, Department of Informatics, Tokat 60250, Turkey
[b] Gaziosmanpasa University, Technical Sciences Vocational School, Tokat 60250, Turkey

ARTICLE INFO

Article history:
Received 4 April 2018
Revised 28 August 2018
Accepted 28 August 2018
Available online 29 August 2018

Keywords:
Cervical cancer
Machine learning
Deep learning
Stacked autoencoder

ABSTRACT

Cervical cancer is one of the most common cancer types in the world, which causes many people to lose their lives. Cancer research is of importance since early diagnosis of cancer facilitates clinical applications for patients. The aim of the study is the automatic diagnosis of cervical cancer. For this purpose, a data set containing 668 samples, 30 attributes and 4 target variables (Schiller, Citology, Biopsy and Hinselmann) from the UCI database was used in the training and testing phases. Softmax classification with stacked autoencoder, one of the deep learning methods, has been used to classify data sets. At first, by applying stacked autoencoder to the raw data set, a reduced dimension dataset is obtained. This data set has been trained for classification by applying the softmax layer. In this phase, 70% (468) of the data set was allocated for training and the remaining 30% (200) for testing. In order to compare the classification performance of the softmax classification with stacked autoencoder, decision tree, kNN, SVM, Feed Forward NN, Rotation Forest models, which are machine learning methods, are used. In the study, proposed models were applied separately to 4 target variables of the dataset, and their classification successes were compared. Finally, softmax classification with stacked autoencoder model, which was applied for the first time in the cervical cancer dataset, performed better than the other machine learning methods with a correct classification rate of 97.8%. Given the interest in machine learning methods in cancer research, new methods of diagnosis are presented in this study in terms of patient diagnostic support systems.

1. Introduction

Cervical cancer is the fourth most common cancer type among women around the world, causing death. Cervical cancer can be treated with early diagnosis using "Pap Smear" test. According to 2014 data, cervical cancer has declined from the fourth to the tenth rank as the importance attached to early diagnosis increases (Hacıkamiloğlu et al., 2017). According to a study by the World Health Organization (WHO), the cervical cancer is the second most common cancer type in women living in underdeveloped regions in 2012. Approximately 445,000 people have been diagnosed with this disease, and 270,000 people have lost their lives because of this disease (Bailey et al., 2016). The first stage symptoms of cervical cancer are usually not distinctive, but the most common symptom is the unusual vaginal bleeding (Goodman, 2000). These studies show that early diagnosis is an important factor in the treatment of cervical cancer. In this study, data mining techniques were used to classify numerical data of the “Pap Smear" test results. Different classification algorithms were used in this study to simplify doctors' decision making process and shorten the diagnostic times in the cervical cancer diagnosis.

In this study, stacked autoencoder model was used for dimension reduction. The reason for the use of this model is to determine the complex relationships on the dataset and to achieve better classification performance. It also simplifies the classification process by reducing the data dimension of each autoencoder in the model and moving the samples belonging to different classes together. A diagnostic support system was proposed by applying machine learning methods and a deep learning method to the digitized data set of “Pap Smear" test results to predict cervical cancer in accordance with environmental and genetic factors. Softmax classification with stacked autoencoder was used as the deep learning model, and SVM, kNN, Decision tree, Feed Forward NN, Rotation Forest were used as machine learning methods for the diagnosis of cervical cancer. Each model used in the study was applied on the dataset and performance comparisons were made.

The main contribution of our study to the literature is the first implementation of the proposed stacked autoencoder model on the

* Corresponding author.
E-mail addresses: kemal.adem@gop.edu.tr (K. Adem), serhat.kilicarslan@gop.edu.tr (S. Kiliçarslan), onur.comert@gop.edu.tr (O. Cömert).

https://doi.org/10.1016/j.eswa.2018.08.050
0957-4174/

digitized cervical cancer data set. The study aims to simplify the decision making process of doctors and shorten the time of cervical cancer diagnosis with the proposed model.

2. Literature review

There are numerous studies in the literature aimed at comparing the success of algorithms used on cervical cancer. Studies were categorized as machine learning and deep learning algorithms. Considering the studies on cervical cancer, it is observed that SVM, Random Forest, C5.0, and Logistic Regression models were used as machine learning methods.

Among the studies conducted with machine learning, the one with logistic regression algorithm has an 88.70% classification success rate over 710 samples with 12 attributes, 577 training and 133 test data (Ho, Jee, Lee, & Park, 2004) and another has an 89% success rate over 1728 samples and 133 attributes (Yamal et al., 2015). In one of the studies using the C5.0 algorithm, the success rate was 92.44% over 168 samples with 12 attributes, 118 training and 50 test data (Tseng, Lu, Chang, & Chen, 2014), and it was 67.5% in another study conducted over 237 samples with 10 attributes (Sharma, 2016). In a study using SVM algorithm, correct classification rate was 90.48% obtained by applying dimension reduction algorithms over 858 samples with 32 attributes and 4 class features (Wu & Zhou, 2017). And in a study conducted with the Random Forest method, experimental studies have been carried out over 75 samples, 38 attributes and 7 classes, and the dataset has been classified with success rate of 80.18% (Kurniawati, Permanasari, & Fauziati, 2016).

In one of the studies using image processing and machine learning 20 attributes were used by subtracting statistical features of 917 pap smear images. Random Forest algorithm was applied to the attribute set and it was classified with success rate of 94.44% (Sun, Li, Cao, & Lang, 2017). In the study performed on multispectral images, 3 different attributes of 123 images were extracted and 78% correct classification rate was obtained with random forest algorithm (Obukhova, Motyko, Kang, Bae, & Lee, 2017). In a different study, CNN (Convolution Neural Network) model, which is a deep learning approach, to 65 images was applied. In this model, 32×32 filters were applied to the images and trained for 300 epochs. Thus, the attributes of the images were removed and a correct classification rate of 77.25% was achieved with the neural network. (Almubarak et al., 2017), it was 98.3% conducted over 917 images with ConvNET (Zhang, Lu, Nogues, Summers & Yao, 2017), and it was classified 99.19% on 917 images with CNN in another study (Taha, Dias, & Werghi, 2017).

There is no significant difference between studies using CNN on pap smear images and our study for the diagnosis of cervical cancer in terms of correct classification performance. In addition, CNN model seems to be worse in terms of time performance than the stacked autoencoder we proposed because it reveals the visual features in the training phase.

In this study, the proposed stacked autoencoder model was applied for the first time on a digitized cervical cancer data set. First, by applying stacked autoencoder to the raw data set, a reduced dimension dataset is obtained. This data set has been trained for classification by applying the SoftMax layer. Tests were conducted on the dataset not included in the training and the results were reported. The results were compared by applying methods of the other machine learning to the same data sets.

3. Materials and methods

The dataset with 858 samples, 32 attributes and 4 target variables to be used for classifying the cervical cancer disease in the study with machine learning and deep learning models was obtained from '*UC Irvine Machine Learning Repository*' (Fernandes, Cardoso, & Fernandes, 2017). Of the 32 attributes in the study, the "Time since first diagnosis" and "Time since last diagnosis" attributes were removed from attributes list due to patients' personal information confidentiality and incomplete data reasons. Completion of the missing data with existing algorithms reduces the success performance because there are differences between samples in health data. Therefore, instead of completing the missing data in our study, it was decided to remove the related parts from the dataset. 190 samples from the data set have been removed due to missing data. Thus the remaining 668 samples and 30 attributes were used in the study. The names of the 4 target variable attributes in the data set are expressed as Schiller, Citology, Biopsy and Hinselmann. In Table 1, the distributions of the samples according to the target variables are given.

Table 1
The distributions of the samples according to the target variables.

Target variant	Positive		Negative	
	Number	Rate	Number	Rate
Schiller	63	9.43%	605	90.57%
Citology	39	5.84%	629	94.16%
Biopsy	45	6.74%	623	93.26%
Hinselmann	30	4.49%	638	95.51%

For more effective evaluation of machine learning algorithms, binary logistic regression (blogres) analysis was performed to show the statistical significance of each target variable in the data set. The results of the analysis are shown in Table 2.

Sig values for citology, biopsy, schiller target variables are 0.2%, 0.7% and 2.1%, respectively, as shown in Table 2. The chi-square value is significant because these values are less than 5%. That is, there is a relationship between attributes and target variables. In addition, in order to better evaluate the classification performance, the data set must be tested against noise. For this, an additive Gaussian noise is applied to the data set. The additive Gaussian noise is a widely used noise model, especially for real world data. With the addition of noise batched auto coding model can be tested the robustness against noise (Tao, Zhang, Yang, Wang, & Lu, 2015; Zhu & Wu, 2004). In the study, the StackedAutoEncoder_Model_2, which is the most successful model for the Schiller target variable, was subjected to Gaussian noise and the results were re-evaluated. For the Gaussian noise, sigma parameter values of 0.005 and 0.05 were applied. The results are shown in Table 3.

As shown in Table 3, It appears to have a low impact such as 1.6% on the classification accuracy of the Schiller target variable when the added noise is taken as 0.05. In this study, SVM, kNN, Decision tree, Feed Forward NN, Rotation Forest and Softmax classification with Stacked Autoencoder models were applied for each target variable separately, and the correct classification success rates were compared for each of them.

3.1. Decision trees

Among the machine learning methods, Decision tree is the first classification and estimation algorithm that comes to mind. It is widely used since it is easier to configure and understand than other machine learning algorithms (Das & Turkoglu, 2014; Rokach & Maimon, 2008). Decision trees present a very powerful approach in knowledge discovery and data mining applications. Decision trees, has a structure that consists of branches and leaves. In the Decision tree algorithm, the routes are taken into account when traversing from the root node to the leaf node (Geetha & Nasira, 2014). When creating the tree, depth and breadth are aimed

Table 2
The statistical significance of each target variable.

Omnibus tests of model coefficients (Citology)					Omnibus tests of model coefficients (Hinselmann)				
		Chi-square	df	Sig.			Chi-square	df	Sig.
Step 1	Step	52,850	27	,002	Step 1	Step	28,744	27	,373
	Block	52,850	27	,002		Block	28,744	27	,373
	Model	52,850	27	,002		Model	28,744	27	,373
Omnibus tests of model coefficients (Schiller)					Omnibus tests of model coefficients (Biopsy)				
		Chi-square	df	Sig.			Chi-square	df	Sig.
Step 1	Step	37,804	27	,021	Step 1	Step	48,455	27	,007
	Block	37,804	27	,021		Block	48,455	27	,007
	Model	37,804	27	,021		Model	48,455	27	,007

Table 3
Effect of adding Gaussian noise to the data set on classification performance.

Noise rate	Accuracy (%)
Real data	96.9
$\mu = 0.005$	96.1
$\mu = 0.05$	95.3

to be small. In order to make the size of Decision trees to be small, criteria such as information gain, chi-square statistics, and GINI index are used. Classification and estimation procedures were performed by using algorithms such as C4.5, Decision Stump, Logistic Model Trees, Random Forest and Random Tree in Decision trees.

3.2. kNN

k-nearest neighbor (kNN) is a classification algorithm, among the supervised learning methods, which can be successfully applied in large databases. In the kNN algorithm, classification is performed by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (Dudani, 1976). Euclidean, Manhattan, Minkowski and Chebyshev parameters are used to calculate distances (Lee, Wan, Rajkumar, & Isa, 2012). In the study, the classification process was performed using the distance parameter and $k = 5$ value.

3.3. Support vector machine

The purpose of the Support Vector Machine (SVM), which is a classifier based on statistical learning theory, is to estimate the most appropriate decision support function that can distinguish two or more classes (Cortes & Vapnik, 1995). An example support vector operation is shown in Fig. 1.

As shown in Fig. 1, by finding the decision function in SVM, the most suitable hyperplane is determined which can separate the dataset. The equations for the support vectors in SVM are given in Eq. (1) to be used in a binary classification problem that can be differentiated linearly.

$$wx + b = +1 \qquad \text{For } y = +1 \text{ class} \tag{1}$$

$$wx + b = -1 \qquad \text{For } y = -1 \text{ class}$$

where, y is the class labels, w is the weight vector, and b is the approximate value. To increase the optimum plane spacing, it is necessary to minimize the value of w as shown in Eq. (2).

$$m = \frac{2}{\sqrt{ww}} \qquad f_{min}(w) = \frac{ww}{2} \tag{2}$$

Based on Eq. (2),

$$y_i(wx_i + b) - 1 \geq 0 \tag{3}$$

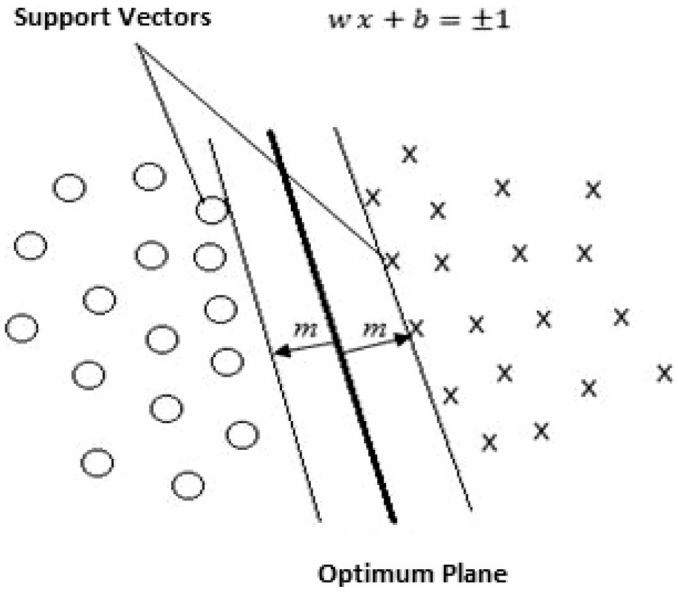

Fig. 1. An example optimal plane and support vectors (Chapelle & Vapnik, 2000).

is obtained. The solution of Eq. (3) with Lagrange optimization is calculated. The decision function of the support vector machine for a two-class problem is given in Eq. (4) (Chapelle & Vapnik, 2000; Vapnik, 1998).

$$f(x) = sign\left(\sum_{i=1}^{k} a_i y_i(x_i) + b\right) \tag{4}$$

The coefficients a_i are the solution of the problem. Detailed discussions on SVM refer to (Chapelle & Vapnik, 2000).

3.4. AutoEncoders

An autoencoder, which based on unsupervised learning running on artificial neural networks, is used to perform a kind of encoding and decoding without performing classification function (Baştürk, Yüksei, Badem, & Çalışkan, 2017; Kaynar, Yüksek, Görmez, & Işık, 2017). Fig. 2 shows a general autoencoder model.

As shown in Fig. 2, this model consists of three layers, as in artificial neural networks. These layers include the input, hidden and output layers (Le et al., 2011). Each layer consists of a specified number of neurons. Autoencoder operations are performed in the model we have determined as the hidden layer. Here, the number of hidden layers is one of the important factors affecting the performance of the created model. There are two important features of the autoencoder model: the input layer and the output layer being chosen equal, and the number of input layers often greater

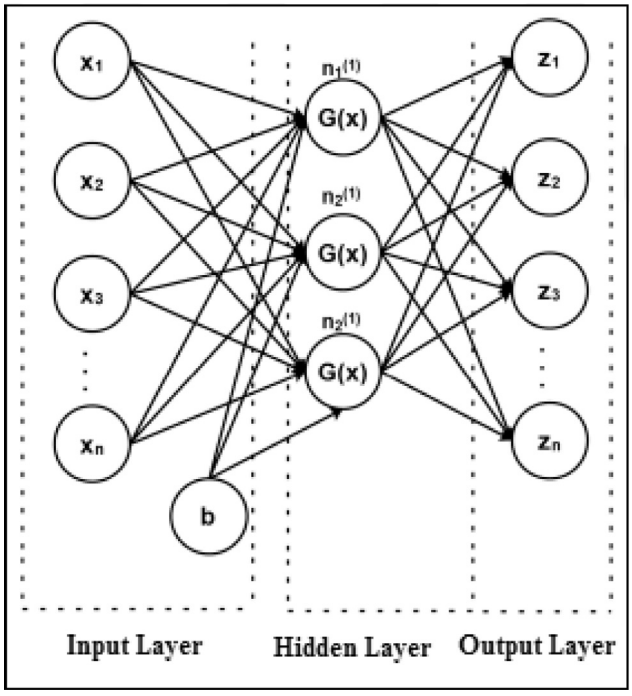

Fig. 2. A general autoencoder model.

than the number of hidden layers. Thanks to these two features, the autoencoder model performs better than the feed-forward neural networks.

The main purpose of using the autoencoder is to obtain new data by eliminating the noise in the data. In the encoder phase, Eq. (5) is applied to reduce the dimension of the data from the input layer and feed them to the hidden layer. In the decoder phase, the reduced data in the hidden layer are decoded by Eq. (6) to obtain the data closer to the input data. After these two phases, the backpropagation algorithm given in Eq. (7) is used in order to make the new values closer to the data in the input layer. The purpose of this process, which is performed via hidden layer, is to try to reveal important data in the dataset (Baştürk et al., 2017; Bengio, 2009; Chen, Gou, Wang, Li, & Jiao, 2018; Kaynar, Yüksek, et al., 2017; Kaynar, Aydın, & Görmez, 2017; Le et al., 2011).

$$y = s\left(W^t x + b\right) \tag{5}$$

$$z = s\left(W y + b'\right) \tag{6}$$

$$\min \sum_{i=1}^{n} (z - x)^2 \tag{7}$$

In Eq. (5) and Eq. (6), s refers to gauss, sigmoid and tanh activation functions used (Bengio, 2009). w represents the weights between the input and the hidden layer, and b represents the bias value (Baştürk et al., 2017). The value z indicates the reconstructed state of the input values using y values of the hidden layer, obtained from the input value x (Chen et al., 2018).

The overfitting problem in artificial neural networks is also present in the autoencoder structure. The regularization method is used to overcome this problem. The operations performed in this method are shown in Eq. (8).

$$\min \left[\left(\sum_{i=1}^{n} \left(x' - x\right)^2 \right) + \gamma L(w) \right] \tag{8}$$

In Eq. (8), γ is the regularization parameter, $L(w)$ is the weight adjustment parameter. As seen in Eq. (8), overfitting is avoided by multiplying the error term and weighting factor by the backpropagation algorithm. Here, the parameters after the sigma are not constant, but are determined by trial and error in order to obtain the best result (Baştürk et al., 2017).

Since complex networks and their classification require a network structure with more hidden layers, multiple autoencoders are connected in succession to obtain stacked autoencoder model as shown in Fig. 3.

As shown in Fig. 3, in stacked autoencoder structure, the success rate is reduced due to the reduction of the input data dimension. The softmax layer is applied to overcome this problem. The Softmax layer is a probability-based linear classifier used in cases where there are two or more classes. This layer increases the classification performance by using the attributes received from the stacked autoencoder structure. L2WeightRegularization and SparsityRegularization parameters are used in this model to reduce overfitting during training. By using these parameters, it is possible to eliminate the overfitting and to get better classification outcomes. The SparsityProportion parameter is used to control the sparseness of the data in the hidden neurons.

In the Stacked AutoEncoder method, non-linear maps (activation functions) are used, as in neural networks, when the dimension reduction process is performed. According to the PCA model, this situtation negatively affects the time performance even though it allows for more effective attributes to be foreground in data sets with complex relationships (Vincent, Larochelle, Lajoie, Bengio, & Manzagol, 2010).

Experiments were carried out as shown in Fig. 4 using three components of the deep neural network consisting of the data set used in the study, the autoencoder and the softmax layers.

As shown in Fig. 4, a classifier was created by adding two hidden autoencoders and a softmax layer at the end of the second autoencoder. The input data set has been implemented with softmax classification and stacked autoencoder by the following steps.

Step 1. The input dataset was given to the first autoencoder and the values were trained according to Eq. (7).
Step 2. The hidden layer of the first autoencoder is given as input to the second autoencoder as in Fig. 4. The second autoencoder is trained as in Step 1. (As a result of the experimental studies carried out, it was found that the best case was provided with two autoencoders.)
Step 3. Output values of stacked autoencoder is given to the SoftMax layer and classified as in Fig. 4.

The Stacked Autoencoder with softmax can provide better classification performance by detecting more complex relations on the dataset. In addition, with each of the autoencoders included, it reduces the attributes of the input data and facilitates the classification process by seperating data belonging to different classes. However, in linearly seperable data sets, simpler methods such as PCA provide better solutions.

At the same time, the stacked autoencoder is also disadvantageous in terms of time performance due to it examines the details according to the PCA model (Almotiri, Elleithy, & Elleithy, 2017). In Table 4, computational complexity values were compared with other machine learning methods of the proposed method using asymptotic measures (Cormen, Leiserson, & Rivest, 1990; Murphy, 2012).

As shown in Table 4, the complexity performance in training phase of the proposed method is lower than that of other machine learning methods. However, due to the dimension reduction of the proposed method, it shows higher performance in the test phase than the other methods.

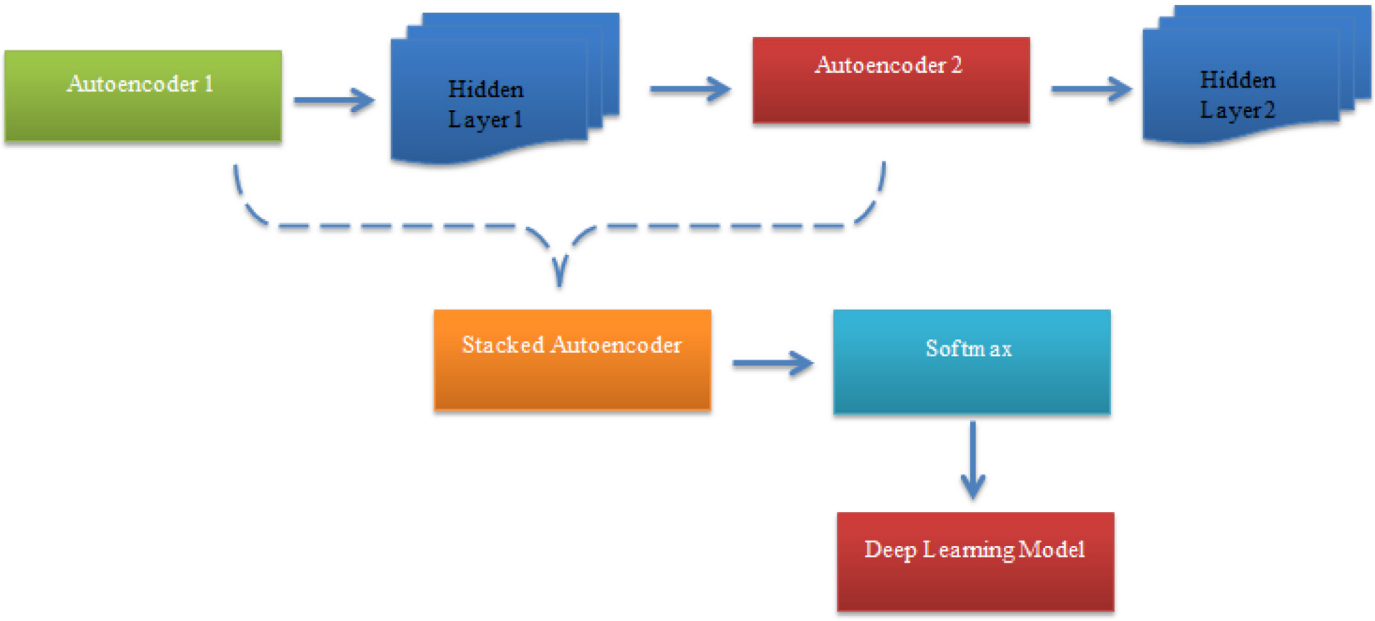

Fig. 3. A general stacked autoencoder model.

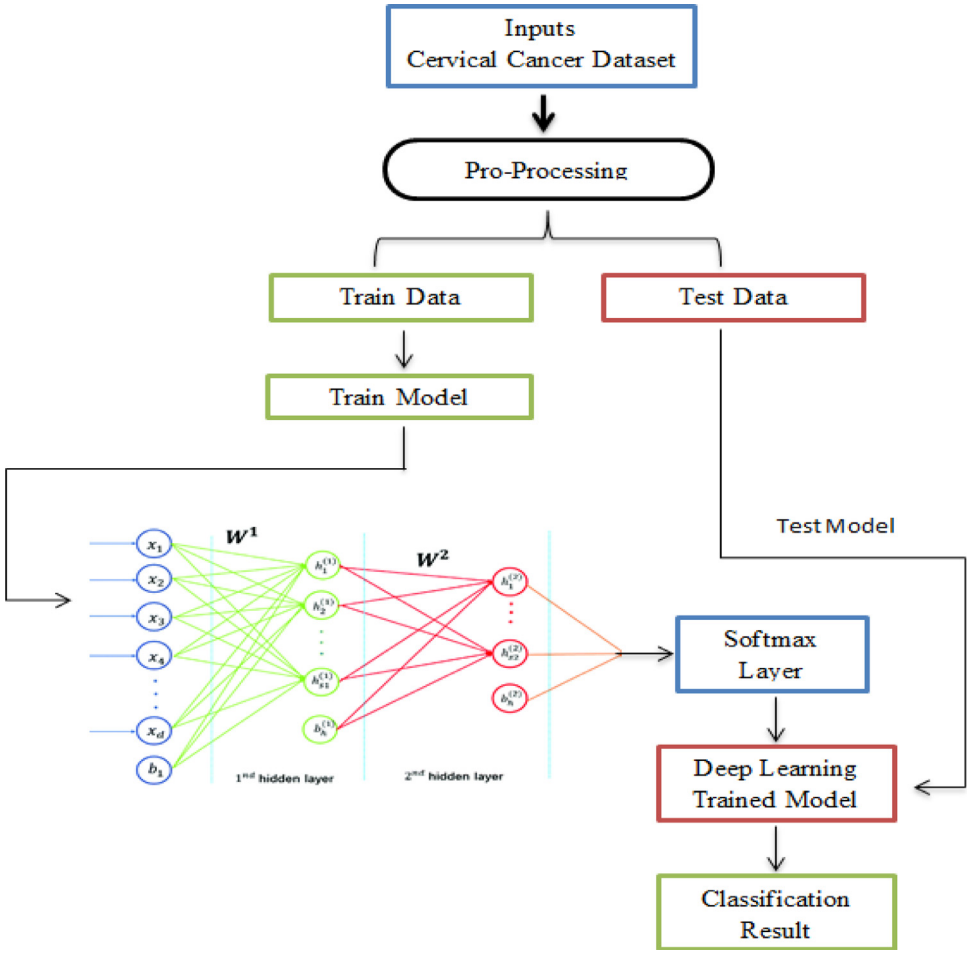

Fig. 4. Stacked autoencoder model used in the study.

Table 4
Computational complexities of the methods used in the study.

Method	Computational complexity	Parameter
Nearest neighbors search	$O(MLog(k)NLog(N))$	M: number of features
Support vector machine (*) using Linear	$O(N^2)$	k: number of neighbors
Support vector machine (*) using Gaussian	$O(N^2M)$	N: number of observations
Decision tree	$O(H)$	H: height of tree
Random Tree	$O(MNlog(N))$	h: number hidden neurons
Rotation Forest	$O(MNlog(N)) + O(\min(M^3,N^3))$	K: Sum number hidden
Feed Forward NN	$O(hNM)$	neurons of AE
Proposed Method	$O(NK + MhN)$	

Table 5
Stacked autoencoder models and parameters.

Models	AE1 hidden neuron	AE2 hidden neuron	Iteration	L2 Reg	Sparse Reg	Scale data	
						AE1	AE2
Model1	30	20	100	0.01	4	F	F
Model2						F	T
Model3						T	F
Model4						T	T

4. Results and discussion

Experiments were carried out as shown in Fig. 4 using three components of the deep neural network consisting of the data set used in the study, the autoencoder and the softmax layers. Stacked autoencoder structure was used as a deep neural network model. The reasons why this model is preferred to the known deep learning methods are: In the Stacked Autoencoder method, learning is performed more successfully than Restricted Boltzman Machine (RBM) because of inter-layer communication (Wang, Raj, & Xing, 2017). The Stacked autoencoder method is used because it shows better performance over noisy data sets (Gaussian noise) than the Deep Belief Network (DBN) algorithm (Kasun, Zhou, Huang, & Vong, 2013; Wang et al., 2017). Variational autoencoder (VAE) is usually used on larger and more complex data sets such as microarray, bigdata (Sønderby, Raiko, Maaløe, Sønderby, & Winther, 2016). The study was performed on the Matlab R2017b platform. The parameters used in the stacked autoencoder model are shown in Table 5.

The dataset used in the study was divided into training and test data. In all methods, the number of iterations was set to 100 and the k = 5 fold validity method was used to make a standard evaluation. As a result of experimental studies, parameters used in the other machine learning methods are k = 5 neighborhood and euclidean metric for distance measure in the kNN. In the SVM, the kernel scale is set to 0.5 for the linear kernel function and 5.5 for the Gaussian kernel function. For Feed Forward NN, the number of hidden neurons is set to 10, and the activation function is set to sigmoid.

The parameters used in the stacked autoencoder model were determined as the result of experimental studies. First, 30 hidden neurons and parameter values shown in Table 5 were applied to the training structure in the first autoencoder (AE1) to obtain new hidden layer data. This hidden layer is connected to the second autoencoder (AE2) as shown in Fig. 4. In the second autoencoder, 20 hidden neurons were identified and the parameter values shown in Table 5 were applied to the model. The number of hidden neurons expressing the number of attributes in AE1 and AE2 were determined as the result of experimental studies. Softmax layer was applied to the new data obtained from AE2. In order to obtain a deep neural network, classification was performed by stacking the softmax layer with autoencoders. In this study, four different models were developed to investigate the effect of scaling the data of AE1 and AE2 autoencoders. Each encoder in the model was trained independent from each other. Sparse Reg is a parameter that sets the regularization value of the objective function required for the optimization algorithm and is usually selected between 0.5 and 5. L2 Reg is a way of preventing overfitting. Thanks to this method, the precise changes that cause overfitting are stretched. This flexibility determines the L2 Reg parameter (Badem, Caliskan, Basturk, & Yuksel, 2016). In our study, Sparse Reg value was set at 4 and L2 Reg value was set at 0.01. Table 6 shows the results of the experimental studies.

Table 6 shows that the correct classification results obtained in stacked autoencoder model, which is one of the deep neural network methods, are more successful than the other machine learning methods. Looking at the machine learning methods, it is seen that Feed Forward NN is more successful than other methods. It was observed that sigmoid activation function gives more successful results since the dataset have a nonlinear distribution. After applying the machine learning methods to the dataset, we compared the proposed model (StackedAutoEncoder) with by applying the kPCA dimension reduction algorithm to the two methods (SVM-Gaussian and Feed Forward NN) which have the highest classification performance. As result of comparison, although the kPCA method seems to increase the success rate of the machine learning methods, the stacked autoencoder method, which is the model we propose in the study, seems to have achieved higher success rates. This shows that the stacked autoencoder model performs more effective dimension reduction than the kPCA method on the cervical data set. The study also investigated the effect of the scaling process on the classification performance of the dataset fed into the encoders in the stacked autoencoder models. When the results of 4 different models created for this purpose are examined, it is seen that the scale applied to the hidden layers reduces the success rate. It was observed that the non-scaled StackedAutoEncoder-Model_2 gives more successful results than the other stacked autoencoder models. In addition, Table 7 presents training-time performances of the algorithms used in experimental studies.

As seen in Table 7, the other machine learning models seems to be better in terms of time performance than the stacked autoencoder we proposed because there is no dimension reduction in the other machine learning models. The comparative results of the proposed machine learning and deeper learning methods with the studies on the same subject in the literature are shown in Table 8.

As shown in Table 8, the study conducted with deep learning method gave more successful results than the other studies on the same subject and same dataset in the literature. In Table 6, it is

Table 6
Experimental results for cervical cancer classification.

Dimension reduction method	Method	Schiller Acc. (%)	Citology Acc. (%)	Biopsy Acc.(%)	Hinselmann Acc.(%)	Mean Acc.
None	KNN	87.4	94.2	93.3	92.5	91.85
	Decision Tree	85.9	92.4	91.8	94	91.03
	SVM-Linear	90	93.3	92.5	95.2	92.75
	SVM-Gaussian	90.6	94.2	93.3	95.5	93.40
	Feed Forward NN	91.3	94.5	93.6	95.9	93.83
	Random Tree	85.93	89.72	88.32	90.11	88.52
	Rotation Forest	90.10	93.05	92.51	94.87	92.63
kPCA	SVM-Gaussian	90.7	94.4	93.5	95.6	93.55
	Feed Forward NN	91.5	94.6	93.9	96.1	94.03
StackedAutoEncoder-Model_1	Softmax	96.6	97.5	97.6	96.1	96.95
StackedAutoEncoder -Model_2		96.9	97.8	97.6	96.7	97.25
StackedAutoEncoder -Model_3		96.7	96.6	97.3	95.5	96.53
StackedAutoEncoder -Model_4		97.1	97.5	97.3	94.5	96.60

Table 7
Time performance results for cervical cancer classification.

Method	Schiller Time (s)	Citology Time (s)	Biopsy Time (s)	Hinselmann Time (s)
KNN	0.502	0.547	1.372	0.608
Decision Tree (Gini)	0.522	0.225	4.835	0.581
SVM-Linear	0.863	0.651	2.572	0.963
SVM-Gaussian	0.449	0.401	0.524	0.417
Feed Forward NN	1.127	1.013	2.450	1.184
Random Tree	0.505	0.312	3.936	0.529
Rotation Forest	0.621	0.297	4.330	0.564
kPCA + SVM-Gaussian	0.632	0.512	0.725	0.612
kPCA + Feed Forward NN	1.345	1.227	3.050	1.452
StackedAutoEncoder -Model_1	18.238	17.773	17.462	18.106
StackedAutoEncoder -Model_2	18.595	17.678	18.340	18.286
StackedAutoEncoder -Model_3	19.246	17.316	17.518	18.467
StackedAutoEncoder -Model_4	19.635	18.957	19.841	18.858

Table 8
Recommended method and literature comparison for cervical cancer classification.

Authors	Year	Data number	Method	Accuracy (%)
Wu and Zhou (2017)	2017	858	SVM	92.75
Fatlawi (2017)	2017	858	Decision Tree	60.90
Ceylan and Pekel (2017)	2017	858	Random Forest	89.00
Proposed method-1	2018	858	SVM-Gaussian	93.40
Proposed method-2	2018	858	Stacked Autoencoder	97.25

seen that Feed Forward NN and StackedAutoEncoder-Model_2 give better results in the success rates according to 4 target variables. In Table 8, the best of the achievement results of these 4 targets are given. In addition, we have achieved high success rates in the two methods we proposed, compared to other studies using the same dataset. It was observed that the Feed Forward NN, which is our proposed model of machine learning, increases the success rate using the Sigmoid activation function in the literature. It should be taken into consideration that the number of samples is small in the study conducted with SVM and Random Forest, which has the closest success rate to the one we have achieved in the study.

5. Conclusion

Cervical cancer can hide itself for a long time, but shortly after the treatment, it may be helpful if the doctor can predict the recurrence of disease based on several variables. Doctors are always expected to make the right decision. Many years of clinical work and experiments have to be done, to be able to make the right decision, but still faults are always possible. To better predict the disease, many investigators have tried to identify risk factors such as tumor size, depth, and the likelihood of regrowth affecting the disease. In this study, It was determined the disease to the data set by applied dimension reduction.

In this study, in order to classify cervical cancer differently from the studies in the literature, the other machine learning methods as well as the stacked autoencoder, which are deep learning solutions, were used for the first time. One of the major disadvantages of many machine learning methods is the difficulty in dimension reduction; however, this problem is easily eliminated with the deep learning. The stacked autoencoder model we used in the study works as a classifier with high accuracy by eliminating unnecessary attributes to reduce data dimension. It is seen that the proposed model in the study achieves better success rates in classifying cervical cancer data compared to the other machine learning methods and kPCA dimension reduction methods. The training time performance of the proposed model in the study is worse than the other methods due to the large amount of time spent in reducing the dimension and the number of samples used in the training. This situation is ignored because a single sample will be given to the proposed model during the test. For this reason, the stacked autoencoder model can be used as an alternative method in healthcare decision support systems.

Author contributions

Kemal Adem – Conceptualization; Formal analysis; Methodology; Software; Validation; Writing - original draft; Writing - review & editing.

Serhat Kiliçarslan – Conceptualization; Investigation; Software; Writing - review & editing.

Onur Cömert – Formal analysis; Methodology; Validation; Visualization & Writing - original draft.

References

Almotiri, J., Elleithy, K., & Elleithy, A. (2017). Comparison of autoencoder and Principal Component Analysis followed by neural network for e-learning using handwritten recognition. In *Systems, applications and technology conference (LISAT), 2017 IEEE long island* (pp. 1–5). IEEE.

Almubarak, H. A., Stanley, R. J., Long, R., Antani, S., Thoma, G., Zuna, R., et al. (2017). Convolutional neural network based localized classification of uterine cervical cancer digital histology images. *Procedia Computer Science, 114*, 281–287.

Badem, H., Caliskan, A., Basturk, A., & Yuksel, M. E. (2016). Yıgınlanmıs Özdevinimli kodlayici ileinsan aktivitelerinin siniflandirilmasi classification of human activity by using a stacked autoencoder. *TIPTEKNO2016, Antalya*.

Bailey, H. H., Chuang, L. T., DuPont, N. C., Eng, C., Foxhall, L. E., Merrill, J. K., et al. (2016). American society of clinical oncology statement: Human papillomavirus vaccination for cancer prevention. *Journal of Clinical Oncology, 34*(15), 1803–1812. http://ascopubs.org/doi/pdf/10.1200/JCO.2016.67.2014.

Baştürk, A., Yüksei, M. E., Badem, H., & Çalışkan, A. (2017). Deep neural network based diagnosis system for melanoma skin cancer. *Signal Processing and Communications Applications Conference (SIU), 25*, 1–4. http://ieeexplore.ieee.org/document/7960563.

Bengio, Y. (2009). Learning deep architectures for AI. *Foundations and Trends in Machine Learning, 2*(1), 1–127.

Ceylan, Z., & Pekel, E. (2017). Comparison of multi-label classification methods for prediagnosis of cervical cancer. *International Journal of Intelligent Systems and Applications in Engineering, 5*(4), 232–236.

Chapelle, O., & Vapnik, V. (2000). Model selection for support vector machines. *Advances in Neural Information Processing Systems*, 230–236.

Chen, W., Gou, S., Wang, X., Li, X., & Jiao, L. (2018). Classification of PolSAR images using multilayer autoencoders and a self-paced learning approach. *Remote Sensing, 10*(1), 1–17. doi:10.3390/rs10010110.

Cormen, T., Leiserson, C., & Rivest, R. (1990). *Introduction to algorithms*. MIT Press.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning, 20*(3), 273–297.

Daş, B., & Türkoğlu, İ. (2014). DNA dizilimlerinin sınıflandırılmasında karar ağacı algoritmalarının karşılaştırılması. *Eleco, 2014*, 381–383.

Dudani, S. A. (1976). The distance-weighted k-nearest-neighbor rule. *IEEE Transactions on Systems, Man, and Cybernetics, 6*(4), 325–327. doi:10.1109/TSMC.1976.5408784.

Fatlawi, H. K. (2017). Enhanced classification model for cervical cancer data set based on cost sensitive classifier. *International Journal of Computer Techniques, 4*(4), 115–120.

Fernandes, K., Cardoso, J. S., & Fernandes, J. (2017). Transfer learning with partial observability applied to cervical cancer screening. *Iberian Conference on Pattern Recognition and Image Analysis, 10255*, 243–250. doi:10.1007/978-3-319-58838-4_27.

Geetha, A., & Nasira, G. M. (2014). Data mining for meteorological applications: Decision trees for modeling rainfall prediction. *IEEE International Conference on Computational Intelligence and Computing Research, 2*(8), 2181–2189. doi:10.1109/ICCIC.2014.7238481.

Goodman, A. (2000). Abnormal genital tract bleeding. *Clinical Cornerstone, 3*(1), 25–35. doi:10.1016/S1098-3597(00)90019-X.

Hacıkamiloğlu, E., Gültekin, M., Boztaş, G., Dündar, S., Utku, E. Ş., Ergün, A. K., et al. (2017). Cancer statistics Turkey. *Turkish Public Health Institution*. http://kanser.gov.tr/Dosya/ca_istatistik/2014-RAPOR._uzun.pdf.

Ho, S. H., Jee, S. H., Lee, J. E., & Park, J. S. (2004). Analysis on risk factors for cervical cancer using induction technique. *Expert Systems with Applications, 27*(1), 97–105. doi:10.1016/j.eswa.2003.12.005.

Kasun, L. L. C., Zhou, H., Huang, G. B., & Vong, C. M. (2013). Representational learning with extreme learning machine for big data. *IEEE Intelligent Systems, 28*(6), 31–34.

Kaynar, O., Aydın, Z., & Görmez, Y. (2017a). Sentiment analizinde öznitelik düşürme yöntemlerinin oto kodlayicili derin öğrenme makinalari ile karşilaştırılması. *Bilişim Teknolojileri Dergisi, 10*(3), 319–326. doi:10.17671/gazibtd.331046.

Kaynar, O., Yüksek, A. G., Görmez, Y., & Işık, Y. E. (2017b). Intrusion detection with autoencoder based deep learning machine. *Signal processing and communications applications conference (SIU)*. doi:10.1109/SIU.2017.7960180.

Kurniawati, Y. E., Permanasari, A. E., & Fauziati, S. (2016). Comparative study on data mining classification methods for cervical cancer prediction using pap smear results. In *Biomedical engineering (IBIOMED), international conference* (pp. 1–5). http://ieeexplore.ieee.org/document/7869827.

Le, Q. V., Ngiam, J., Coates, A., Lahiri, A., Prochnow, B., & Ng, A. Y. (2011). On optimization methods for deep learning. In *Proceedings of the 28th international conference on international conference on machine learning* (pp. 265–272).

Lee, L. H., Wan, C. H., Rajkumar, R., & Isa, D. (2012). An enhanced support vector machine classification framework by using Euclidean distance function for text document categorization. *Applied Intelligence, 37*(1), 80–99. https://link.springer.com/article/10.1007/s10489-011-0314-z.

Murphy, K. (2012). *Machine learning: A probabilistic perspective*. MIT Press.

Obukhova, N. A., Motyko, A. A., Kang, U., Bae, S. J., & Lee, D. S. (2017). Automated image analysis in multispectral system for cervical cancer diagnostic. In *Open innovations association (FRUCT), 2017 20th conference of* (pp. 345–351). IEEE.

Rokach, L., & Maimon, O. (2008). *Data mining with decision trees: Theory and applications*. Singapore: Hackensack.

Sharma, S. (2016). Cervical cancer stage prediction using decision tree approach of machine learning. *International Journal of Advanced Research in Computer and Communication Engineering, 5*(4), 345–348. doi:10.17148/IJARCCE.2016.5488.

Sønderby, C. K., Raiko, T., Maaløe, L., Sønderby, S. K., & Winther, O. (2016). Ladder variational autoencoders. *Advances in Neural Information Processing Systems*, 3738–3746.

Sun, G., Li, S., Cao, Y., & Lang, F. (2017). Cervical cancer diagnosis based on random forest. *International Journal of Performability Engineering, 13*, 446–457.

Taha, B., Dias, J., & Werghi, N. (2017). Classification of cervical-cancer using pap-smear images: A Convolutional neural network approach. In *Annual conference on medical image understanding and analysis* (pp. 261–272). Cham: Springer.

Tao, S., Zhang, T., Yang, J., Wang, X., & Lu, W. (2015). Bearing fault diagnosis method based on stacked autoencoder and softmax regression. In *Control conference (CCC), 2015 34th Chinese* (pp. 6331–6335). IEEE.

Tseng, C. J., Lu, C. J., Chang, C. C., & Chen, G. D. (2014). Application of machine learning to predict the recurrence-proneness for cervical cancer. *Neural Computing and Applications, 24*(6), 1311–1316. https://link.springer.com/article/10.1007/s00521-013-1359-1.

Vapnik, V. (1998). *Statistical learning theory*: 3. New York: Wiley 1998.

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., & Manzagol, P. A. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research, 11*(Dec), 3371–3408.

Wang H., Raj B., & Xing E.P. (2017). On the Origin of Deep Learning, abs/1702.07800, Access Date: 29.08.2018 Web.

Wu, W., & Zhou, H. (2017). Data-driven diagnosis of cervical cancer with support vector machine-based approaches. *IEEE Access, 5*, 25189–25195. doi:10.1109/ACCESS.2017.2763984.

Yamal, J. M., Guillaud, M., Atkinson, E. N., Follen, M., MacAulay, C., Cantor, S. B., et al. (2015). Prediction using hierarchical data: Applications for automated detection of cervical cancer. *Statistical Analysis and Data Mining: The ASA Data Science Journal, 8*(2), 65–74. doi:10.1002/sam.11261.

Zhu, X., & Wu, X. (2004). Class noise vs. attribute noise: A quantitative study of their impacts. *Artificial Intelligence Review, 22*, 177–210.

Zhang, L., Lu, L., Nogues, I., Summers, R. M., Liu, S., & Yao, J. (2017). Deeppap: Deep convolutional networks for cervical cell classification. *IEEE Journal of Biomedical and Health Informatics, 21*(6), 1633–1643.