Journal of Computational Science 63 (2022) 101775

Contents lists available at ScienceDirect

Journal of Computational Science

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

Manifold learning methods for the diagnosis of ovarian cancer

Bartu Yesilkaya [a], Matjaž Perc [b,c,d,e], Yalcin Isler [a,f,*]

[a] Izmir Katip Celebi University, Department of Biomedical Engineering, Balatcik Campus, Cigli, Izmir, Turkey
[b] Faculty of Natural Sciences and Mathematics, University of Maribor, Koroška cesta 160, 2000 Maribor, Slovenia
[c] Department of Medical Research, China Medical University Hospital, China Medical University, Taichung 404332, Taiwan
[d] Alma Mater Europaea, Slovenska ulica 17, 2000 Maribor, Slovenia
[e] Complexity Science Hub Vienna, Josefstädterstraße 39, 1080 Vienna, Austria
[f] Islerya Medical and Information Technologies, Bornova, Izmir, Turkey

ARTICLE INFO

Keywords:
Ovarian cancer
Feature reduction
Manifold learning
Principal component analysis
Machine learning

ABSTRACT

Early detection of ovarian cancer is crucial for a good outlook. Different machine learning methods have already proven useful to that effect, but using many features and samples often yields a complex structure of classifier algorithms. This study investigates the effect of four different manifold learning methods prior to well-known classification algorithms to reduce the number of features and compares the achieved results with the well-known principal component analysis method. The NCI PBSII dataset, which consists of 253 samples with 15154 features, is used in this study. We tested nine distinct classifiers: k-nearest neighbors, decision tree, support vector machines, stochastic gradient descent, random forest, multi-layer perceptron, Naive Bayes, logistic regression, and AdaBoost. Among these classifiers, the logistic regression gives a maximum of 99.2% accuracy using these features. These classifiers were rerun for five distinct reduced feature sets obtained using principal component analysis, Multidimensional Scaling, Locally Linear Embedding, Isometric Feature Mapping, and t-Distributed Stochastic Neighbor Embedding methods. Among these feature reduction methods, Locally Linear Embedding hit the maximum classifier performance five times (of nine classifiers) with an average of 15.4 components. Both the logistic regression classifier with 28 Multidimensional Scaling components and the stochastic gradient descent classifier with 30 Locally Linear Embedding components achieved the maximum accuracies of 99.8%. On the other hand, the commonly used principal component analysis resulted in a maximum of 99.7% accuracy using stochastic gradient descent with 30 principles. In conclusion, although principal component analysis is the most commonly used feature reduction method, the Locally Linear Embedding (a manifold learning method) may give higher classifier performances with fewer components in the diagnosis of ovarian cancer.

1. Introduction

Ovaries are primary female reproductive organs that also play roles in hormone-secreting and egg-production [1]. Ovarian cancer is one of the most frequent gynecologic cancers, also responsible for the high mortality rate for cancer-associated deaths among women [2]. There are many factors for ovarian cancer generation and genetic factors are one of the most important ones [3]. According to the data of the American Cancer Society, the survival rate for ovarian cancer is 44% [4]. According to American Institute for Cancer Research [5], new 295,414 cases were diagnosed globally in 2018. In addition, the economic burden of ovarian cancer was given an average of $140,000 per patient for a year in 2020 [6].

Late diagnosis brings patients to advanced stages of the disease and causes higher mortality rates [3]. Several factors, including menarche and menopause ages, play an important role in the diagnosis of ovarian cancer. Transvaginal ultrasound is the primary diagnostic tool in the diagnosis, while symptoms rarely help to diagnose [7]. In September 2009, the Ovarian Triage Test, which measures the levels of biochemical markers, was approved [7]. Breastfeeding, pregnancy, and the use of contraceptive pills are protective factors [8]. But the use of birth control pills may suppress these markers, which may cause the early diagnosis a challenging issue.

In recent decades, pattern recognition methods have yielded successful results in different application areas [9,10] including accurate diagnosis of many diseases [11–15]. Researchers used various pattern recognition algorithms for feature extraction, feature selection, feature reduction, and classification for this purpose [16]. For example, Talbi and colleagues used Support Vector Machine (SVM) with optimization

* Corresponding author at: Izmir Katip Celebi University, Department of Biomedical Engineering, Balatcik Campus, Cigli, Izmir, Turkey.
E-mail addresses: bartu.yesilkaya@yandex.com (B. Yesilkaya), matjaz.perc@gmail.com (M. Perc), islerya@yahoo.com (Y. Isler).

https://doi.org/10.1016/j.jocs.2022.101775
Received 18 May 2022; Received in revised form 20 June 2022; Accepted 14 July 2022
Available online 20 July 2022

methods of the Genetic Algorithm (GA) and the Particle Swarm Optimization (PSO) to classify ovarian cancer patients from a publicly available dataset at PBSII Data, which includes proteomic patterns in serum. They achieved a maximum classifier accuracy of 99.44% [17]. Rahman and colleagues used five different classifiers (Fine Trees, SVM, K-Nearest Neighbors (KNN), Ensemble Learner, and 15-Neuron multilayer perceptron (MLP)) with the Taguchi method from the same dataset. They achieved classifier accuracies of 88.75% for Fine Trees, 97.20% for Support Vector Machines, 96.80% for K-Nearest Neighbors, 97.60% for Ensemble Learner, and 98.70% for 15-Neuron Multi-Layer Perceptron [18]. Yasodha and Ananthanarayanan used Self-Organizing Maps and Immune Clonal Selection algorithms for feature selection over the same data. They achieved classifier accuracies of 71% for Support Vector Machines, 85% for Multi-Layer Perceptron, 75.6% for Feed Forward Neural Network (FFNN), 87.3% for Radial Basis Function Network (RBFN), 93.21% for General Regression Neural Network (GRNN), and 98.23% for Grammatical Evolution Neural Networks (GENN) [19]. Basegmez and colleagues used three classifiers with two feature selection algorithms over the same data. They reported classifier accuracies of 98.81% for Support Vector Machines, 95.65% for Decision Tree (DT), and 98.88% for Random Forest (RF) [20]. Demircioglu and Bilge investigated the Fisher Correlation Score (FCS) and Welch T-Test (WTS) prior to K-Nearest Neighbors and Support Vector Machines classifiers. They reported a classifier accuracy of 100% [21]. Liu and colleagues examined the effect of high-resolution and low-resolution datasets of ovarian cancer. They reported the classifier accuracy of 100% for the Support Vector Machines classifier with the Dataset Wavelet method. They achieved this accuracy using the Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) together from the low-resolution dataset [22]. Ubaidillah and colleagues also reported classifier accuracies of 78% for Multi-Layer Perceptron and 64% for Support Vector Machines [23]. Li and Ramamohanarao tried to detect ovarian cancer by using the ensemble decision tree and Support Vector Machines in their study in 2004 [24]. Liu and colleagues investigated two feature selection methods of the Fast Correlation-Based Filter (FCBF) and Procrustes Analysis (PA) prior to Support Vector Machines classifier [25]. Gao and colleagues tested some optimization methods combined with the Support Vector Machines classifier [26]. Yesilbas and Guven investigated the effect of the Principal Component Analysis method with the Multi-Layer Perceptron classifier in the detection of ovarian cancer was investigated [27]. Belciug and Gorunescu examined multiple cancer datasets and tested the results using six different classifiers [28]. They reported Adaptive Single-hidden layer Feedforward Neural Network (aSLFN) as the most successful classification method for ovarian cancer diagnosis. Belciug and Ivanescu compared different cancer datasets using Bayesian initialization of extreme learning machine (BiELM) and ELM classifiers [29]. They reported that the BiELM classifier overcomes the ELM classifier. Al-Murad and Hossain investigated two new feature selection methods of Evolutionary Non-Dominated Radial Slot Based Algorithm (ENORA) and Consistency Subset evaluation (CSE), and they tested Multi-Layer Perceptron classifier [30]. In addition to these conventional machine learning algorithms, deep learning methods have been investigated. For example, Kilicarslan and colleagues used a hybrid machine learning model based on the Convolutional Neural Network (CNN) method [31]. They added a dimension reduction layer that uses the Relief method to the model. Wu and Banzhaf used deep learning methods in their study [32]. They compared a few optimization methods and achieved a maximum classifier accuracy of 96.22% using the Optimal Recurrent Neural Networks (ORNN) classifier with the self-organizing-maps-based feature selection. Along with these, there are also articles [33,34] investigating the importance of features using various machine learning algorithms.

As summarized above, there are many machine learning methods were evaluated for the detection of ovarian cancer. Most of these studies relied on complex classifier architecture since they used all features or selected features inefficiently. Principal Component Analysis is one of the commonly used feature reduction methods. We investigated four new feature selection methods based on manifold learning. To show the effect of these methods, we evaluated nine different classifier algorithms using six different feature sets of all features, reduced features via Principal Component Analysis, and reduced features via four manifold learning methods in this study.

2. Materials and methods

2.1. Data acquisition

NCI PBSII data is one of the ovarian cancer datasets which contains a list of women's data that have the maximum risk factor. The dataset includes 91 normal subjects and 162 patients with ovarian cancer. Each sample in the dataset covers chemical information, intensity, and magnitude, which equals to the total of 15154 features [19]. Detailed information can be obtained from https://data.mendeley.com/datasets/jbjd5fmggh/2.

2.2. Data transformation using Principal Components Analysis (PCA)

Principal Component Analysis is a multivariate statistical transformation technique to eliminate similarity between variables. It creates new linearly-independent perpendicular variables. The number of these new variables is the system parameter, covering the percentage ratio of the variance of the initial variables. Each new variable is called the principal component. Detailed information and mathematical expressions can be found in a well-known book [35] and a recent article [36].

2.3. Data transformation using manifold learning methods

2.3.1. t-Distributed Stochastic Neighbor Embedding (t-SNE)

t-Distributed Stochastic Neighbor Embedding is a tool for visualizing high-dimensional data. While t-Distributed Stochastic Neighbor Embedding converts similarities between data points into joint probabilities, It tries to minimize the Kullback–Leibler distinctions between these probabilities of the low-dimensional embedding and high-dimensional data. t-Distributed Stochastic Neighbor Embedding has a non-convex cost function. In other words, different results can be achieved with each initiation. This method has several advantages over existing techniques: Describe the structure on a single map at many scales, uncover data in multiple and different manifolds or clusters, reducing the tendency in the center to crowded spots.

The t-Distributed Stochastic Neighbor Embedding method is a variation of the Stochastic Neighbor Embedding method. This variation preserves the local structures of the data in lower dimensional space and reveals an important global structure [37]. Using the same formulation as SNE for high-dimensional space, t-Distributed Stochastic Neighbor Embedding uses the formula in Eq. (1) for the output space.

$$q_{ij} = \frac{(1 + \|y_i - y_j\|)^2)^{-1}}{\sum_{k \neq l}(1 + \|y_i - y_j\|)^2)^{-1}} \tag{1}$$

Unlike the Gaussian distribution of the SNE method, the t-distribution is used in this study. A reduced version of the cost function with the gradient descent algorithm is given in Eq. (2).

$$\frac{\delta C}{\delta y_i} = 4 \sum_j (p_{ij} - q_{ij})(y_i - y_j)(1 + \|y_i - y_j\|^2)^{-1} \tag{2}$$

2.3.2. Multidimensional Scaling (MDS)

Multidimensional Scaling is a statistical analysis that aims to graphically display the multidimensional data. The data is formed as a result of measuring the 'n' unit in terms of the 'p' variable, in a way that preserves the binary similarities between the units [38]. In Multidimensional Scaling, a graphical representation of the units can be provided based on the similarities between the units, as well as the graphical representation of the variables based on the similarity between the variables.

Multidimensional Scaling can also be defined as a dimension reduction technique. It provides a graphical representation of multidimensional data by placing similar units close to each other and dissimilar units far from each other with a non-technical point of view. To provide a graphical representation of multidimensional data, size reduction is performed in a way that preserves similarity in lower dimensions. Moreover, Multidimensional Scaling analysis is widely used as statistical tool in the visualization of multidimensional categorical data, as it is based on the similarities between units or variables. Multidimensional Scaling is a technique that creates a map showing the relative positions of many objects and tabulates the distances between them. The map can be one-dimensional or multidimensional. A table of distances is known as the proximity matrix. There are two methods to solve Multidimensional Scaling, metric (classical) and non-metric. Classical multidimensional scaling (CMDS) attempts to generate original metrics or distances. However, non-metric multidimensional scaling (NMMDS) tries to construct sequences of distances, as it only knows that degrees of distances exist [39].

To explain the basics of Classical multidimensional scaling, it will be useful to understand the method of Torgerson (1952), one of the pioneers of this technique. In Torgerson's algorithm, a distance matrix D assumes that a configuration of X dimensions approximates the interpoint distances n (n = 1, 2, or 3) in a low-dimensional space. Briefly, the elements dij make up the distance matrix D can be calculated from X using the formula:

$$d_{ij} = \sqrt{\sum_{k=1}^{n} (x_{ik} - y_{ik})^2} \tag{3}$$

2.3.3. Isometric Feature Mapping (ISOMAP)

Unlike Principal Component Analysis, isometric feature mapping (ISOMAP) is a nonlinear dimension (feature) reduction method. Isometric Feature Mapping attempts to map points on a high-dimensional nonlinear manifold to a lower-dimensional set of coordinates. In addition, Isometric Feature Mapping is powerful when dealing with high signal-to-noise ratio (SNR) systems. Isometric Feature Mapping can be used in two ways: visualization and classification. The steps for processing high-dimensional data using Isometric Feature Mapping can be summarized as follows [40]:

- Step 1: Neighborhood graph
  Construct a Neighborhood plot for all data points or the adjacent matrix from the dataset. In this step, it is necessary to identify the neighbors in the 'M' manifold. There are two methods for this. The first one is the maximum Euclidean search distance. In this method, any point connects to all points around it with a small radius of ϵ. The second one is to use the nearest neighbor 'k' number. Neighborhood relations are created as a weighted graph on data points. The weight edges between adjacent points are defined as $d_x(i, j)$. Fig. 1 shows how to create a local Neighborhood graph.
- Step 2: Calculation of geodetic distances
  The G graph consists of small jumps. All geodetic distances $d_G(i, j)$ between all points in the 'M' manifold are approximated by combining the small Euclidean distances. Then, the shortest path distance $d_G(i, j)$ calculated on graph G. Fig. 2 shows the estimate of $d_G(i, j)$ based on Isometric Feature Mapping.
- Step 3: Downsize
  For minimization, the Multidimensional Scaling technique is applied to graph distance matrix $d_G(i, j)$.

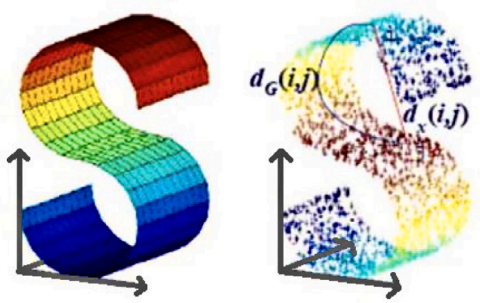


Fig. 1. Generating the local neighborhood graph.

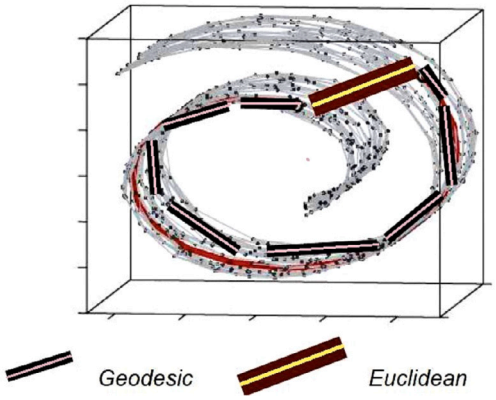


Fig. 2. Estimation of $d_G(i, j)$ based on ISOMAP according to Geodesic and Euclidean distances.

2.3.4. Locally Linear Embedding (LLE)

Local linear mapping or the Locally Linear Embedding (LLE) method is a dimension reduction technique based on manifold learning. Manifold learning is a newly developed technique for nonlinear dimension reduction. The investigated data is assumed to be in an embedded nonlinear manifold within the higher dimensional space [40]. Manifold learning algorithms operate to uncover these parameters to find a low-dimensional representation of the data.

The Locally Linear Embedding algorithm was first described by Roweis and Saul (2000) as a way to project complex high-dimensional data into a much lower-dimensional space for analysis. The main idea is to create local and linear models of the data from high-dimensional space and maintain local distances during reduction to low-dimensional space. The Locally Linear Embedding algorithm can be thought of as a nonlinear manifold learning algorithm. A primary approach for providing relationships between high- and low-dimensional representations of data points with the same locally linear relationships was obtained by the nonlinear manifold learning algorithm. The Locally Linear Embedding algorithm has been used for extremely high dimensional data, such as face image data [41].

Locally Linear Embedding is an eigenvalue–eigenvector based method. The Locally Linear Embedding method is based on simple geometric concepts [38]. This algorithm reconstructs each data point using only its nearest neighbors according to the Euclidean distance. It also characterizes the low-dimensional local geometry of the data points by finding linear coefficients.

2.4. Classification

There are many classification techniques used in supervised machine learning in the literature [42]. Among them, we used nine well-established algorithms in this study. The following subsections give brief explanations about these techniques. A detailed information about these methods can be found in popular books [35].

2.4.1. Stochastic Gradient Descent

Gradient descent (GD) is an extensively popular method in machine learning and deep learning as an optimization tool. The gradient is the slope, which is the derivative, of a given function. GD is an iterative method used to find the value of parameters that minimize the cost function. In the Stochastic Gradient Descent (SGD) algorithm, only a few randomly selected samples from the dataset are used to minimize the cost function instead of using all samples together [43].

2.4.2. AdaBoost

AdaBoost is a powerful ensemble classifier method that gives decisions by utilizing several weak classifiers together. The final decision is the weighted decisions of these weak classifiers. Weak classifiers are slightly better predictors than random guessing, which provides more flexibility in the design of the poor classifier set [44].

2.4.3. Random Forests

The Random Forests (RF) creates and trains a model consisting of multiple decision trees [45]. In the classifier model, each inner node represents the feature in the relevant sample, each branch represents the test result, and the leaf node represents the class label. Each decision tree is structured using randomly selected values from the input data. If the original feature vector has m features, each tree uses n randomly selected features. Decision trees (a forest) are allowed to grow before exceeding their capacity. After the forest is trained, the forest is evaluated using each test sample.

2.4.4. Naive Bayes

Naive Bayes algorithm is an easy implementation of the Bayes' Theorem [35]. It is a probability-based classification method. It calculates the prediction for test samples based on joint probability of the train data [46].

2.4.5. Decision trees

Decision trees have a structure of roots and leaves. There are numerous different implementations of it depending on the tree structure [35]. Decision trees are often preferred classification techniques due to their faster train and test processes. In addition, the results obtained can be easily interpreted [47]. The training phase consists of constructing a tree structure and obtaining classification rules [48].

2.4.6. Logistic Regression

The Logistic Regression estimates possible classes from categorically distributed variables [35]. It is a regression (curve-fitting) method, in fact, but the desired values are discrete numbers (i.e., categories) instead of real-valued numbers [49].

2.4.7. K-Nearest Neighbors (K-NN)

K-Nearest Neighbors (KNN) is a simple classifier. It assigns a class to the test sample as the majority class of its k neighbors [35]. Although various distance measuring functions are available, the Euclidean distance is probably the most used measure among them [16]:

$$distance(x, y) = \sqrt{\sum_{i=1}^{k} (x_i - y_i)^2} \tag{4}$$

where x_i and y_i are the ith features of x and y, respectively.

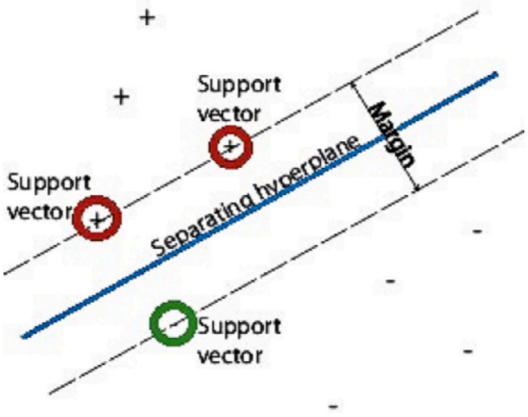


Fig. 3. Separation of two different labeled data with using geometric drawing of SVM.

2.4.8. Support Vector Machine (SVM)

The main advantage of Support Vector Machines is to solve the classification problem by transforming it into an optimization problem. In this way, the number of calculation processes will be reduced and a faster solution can be obtained compared to other techniques [50]. Support Vector Machines is a member of linear binary class classifiers. Classification of objects in datasets is mainly based on tagging objects as −1 (first class) or +1 (other class). The labeling process varies depending on the specification of the research [51]. Another important function of Support Vector Machines is to create an optimal hyperplane (linear decision boundary) that can distinguish differently labeled data points and maximize the distance between the support vectors [52,53]. The mathematical explanation of Support Vector Machines can be summarized as follows. According to Eq. (5); each entry point can be displayed as "xi" and the labels can be expressed as "f(x)", "w" represents the normal and weight vector of the hyperplane, and "b" represents the trend and constant value.

$$f(x) = wx + b \tag{5}$$

The geometric drawing of the linear Support Vector Machines model is shown in Fig. 3 for the classification of two classes and two dimensions.

Two parallel lines are called the boundary plane. The dark-colored plane passing through the middle of the boundary planes and separating both planes equally is expressed as a hyperplane [54,55]. Support Vector Machines uses core functions such as polynomial and Gauss to create nonlinear adaptive data at higher dimensions where a linear decision boundary can be found.

2.4.9. Multi-Layer Perceptron (MLP)

Multi-Layer Perceptron is one of the important approaches used in machine learning. Multi-Layer Perceptron consists of three layers (input layer, hidden layer, and output layer); each layer consists of adaptive processing units that are interconnected and called neurons. A neuron [56] is a general calculation unit that receives m inputs and produces a single output. The parameter that distinguishes the output of neurons is their connection weights. Each neuron in a layer is connected to all neurons in a top layer with different weights. The input layer multiplies the incoming data with weights and transmits it to the hidden layer [57]. A transfer function is used to output from these multiplication results collected in the hidden layer. An example of the (m = 2) Multi-Layer Perceptron architecture is shown in Fig. 4.

Each neuron in the Multi-Layer Perceptron in the entry and latent layer is connected to the neurons of the next layer which is the classical neural network structure. In the Multi-Layer Perceptron architecture, the number of hidden layers can be increased if necessary [35]. Calculating the number of parameters required to train a network containing

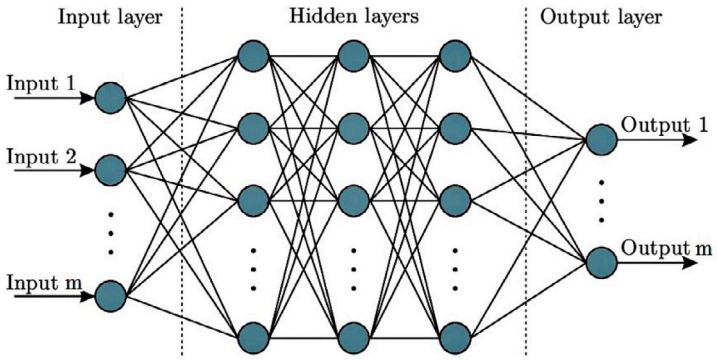


Fig. 4. General structure of Multi-Layer Perceptron architecture with using artificial neurons.

L layers and N neurons in each layer can be a difficult problem. Working with a large stack of parameters can result in an untrained network, although it is not practical. Deep learning networks have been proposed to tackle such problems.

2.5. Performance evaluation

The commonly used performance evaluation metrics are the accuracy (ACC), the area under the curve (AUC), sensitivity (SEN), specificity (SPE), and F1-score [16,35]. The dataset is divided into two groups, train and test data, to prevent the classifiers from memorizing the dataset instead of learning. The classifier parameters are determined with the train data where classifier performances are estimated over the test data. In the literature, there are many validation methods to split the dataset: 70% for training and 30% for testing [18,23,30,31], 40% for training and 60% for testing [21], 90% for training and 10% for testing [27,32], leave-one-out method [16], three-fold method [25], five-fold method [20], and ten-fold method [17,22,24,28,29]. On the other hand, some studies ignored this rule and used all data for training and testing [19,26]. In this study, we preferred splitting dataset as 70% for training and 30% for testing. In addition, we iterated each classifier algorithm ten times and averaged the classifier success measures to reduce the effect of randomness in classifier performances.

3. Results

In this study, a classification study was carried out for the diagnosis of ovarian cancer using the Orange data mining tool program.

Firstly, we applied nine classifier methods to dataset. These classifiers are AdaBoost, Multi-Layer Perceptron, Logistic Regression, Naive Bayes, Decision Tree, Random Forest, k-Nearest Neighbors and Support Vector Machines are examined. We used the Classification widgets in the program for this. All classifiers were used in default mode. The entire dataset was used as training data.

Secondly, we applied Principal Component Analysis using the Preprocess widget. Then, same classifiers were tested and classifier performances were saved, Similarly, we applied four distinct manifold learning methods instead of Principal Component Analysis using Manifold Leaning widget. Thus, same classifiers were tested and classifier performances were saved.

Parameters are determined by the user through the program interface while using manifold methods. In this study, The Principal Component Analysis (torgerson) method was chosen as the initialization instead of the random initialization when using the Multidimensional Scaling technique. The maximum number of iterations is set to 300 in Multidimensional Scaling technique. In the Locally Linear Embedding algorithm, ‘standard’ was chosen as the method. In this method, the number of neighbors and maximum iterations were determined as 100. In the Isometric Feature Mapping method, the number of neighbors was set to 100. In the t-Distributed Stochastic Neighbor Embedding method, outputs were obtained after the metric was set as Euclidean, perplexity as 30, early exaggeration as 12, learning rate as 200, maximum iterations as 1000 and initialization as Principal Component Analysis. The highest accuracy were achieved using these parameters.

As mentioned in the article titled ‘A Python toolkit for dimensionality reduction quality assessment’ [1], a similar accuracy was obtained using less data, without using the entire dataset, by using the Principal Component Analysis method. As stated in the article, ‘The first 5 PCs hold 78.7% of the total original information’ [58]. We observed this situation in orange data mining tool.

According to the Principal Component Analysis, 86.9% of the original information was hold in the first 10 components. The first 70 components represent 98% of the information. The program provides this information for the first 100 principal components.

The results we obtained before using Principal Component Analysis are shown in Table 1. Then, to see the effect of Principal Component Analysis, we increased the number of Principal Components one-by-one and obtained the accuracy. Classification results from the first principal component to the thirtieth principal component. Among the accuracy we achieved without Principal Component Analysis, we achieved the highest result came from Logistic Regression model. It is observed that the accuracy increases when the number of principal components increases.

4. Discussion

Many studies compared different classifier performances in the classification and diagnosing problems. Our accuracy is higher than many studies. We achieved very high accuracy without applying any Feature selection or feature extraction methods. The accuracy of other studies are shown in Table 2. This dataset, which includes 15154 columns and 253 rows, has been classified more quickly by using Principal Component Analysis, which is a dimension reduction method. By using only, the first 15 principal components, the success obtained from the entire dataset has been achieved.

According to Table 1, 29th Principal Component has 99% accuracy with using Multi-Layer Perceptron classifier. This accuracy was passed over with using Stochastic Gradient Descent on the 30th Principal Component and with using the Logistic Regression on the 17th Principal Component. By looking at these table, we can comment that Support Vector Machines, Multi-Layer Perceptron, Logistic Regression and Stochastic Gradient Descent outperform other classifiers when looking at the first 30 components. These classifiers achieved a accuracy of close to 100% and kept their success stable from low component to high components. In general, accuracy for all classifiers increased rapidly after the first 5 components and exceeded 80%. This leads us to the idea that the first 5 components for this dataset carry meaningful information for classification.

When we examine the classifier successes according to the Multidimensional Scaling method, we see that the Logistic Regression model

Table 1
Maximum Classifier performances tested in this study using all methods for variable components. The number of features (or components) is given in parentheses to achieve the maximum accuracies for the corresponding classifier.

Classifier	Manifold learning methods					
	ALL	PCA	MDS	LLE	ISOMAP	T-SNE
k-Nearest Neighbors	0.914 (15154)	0.933 (23)	0.934 (20)	**0.962** (12)	0.909 (14)	0.932 (10)
Decision Tree	**0.973** (15154)	0.896 (28)	0.892 (21)	0.910 (6)	0.884 (17)	0.909 (12)
Support Vector Machines	0.978 (15154)	0.994 (27)	0.992 (28)	**0.996** (19)	0.965 (19)	0.941 (10)
Stochastic Gradient Descent	0.964 (15154)	0.997 (30)	0.993 (25)	**0.998** (30)	0.984 (19)	0.930 (10)
Random Forest	**0.974** (15154)	0.928 (16)	0.928 (6)	0.925 (12)	0.908 (9)	0.926 (10)
Multi-Layer Perceptron	0.956 (15154)	0.990 (29)	0.986 (12)	**0.996** (9)	0.969 (13)	0.945 (10)
Naive Bayes	0.844 (15154)	0.916 (23)	0.906 (6)	**0.937** (17)	0.902 (15)	0.880 (10)
Logistic Regression	0.673 (15154)	0.994 (17)	**0.998** (28)	0.992 (28)	0.982 (19)	0.917 (5)
AdaBoost	**0.916** (15154)	0.885 (6)	0.877 (5)	0.906 (6)	0.880 (6)	0.914 (5)
Average	0.951 (15154)	0.948 (22.1)	0.945 (16.7)	**0.958** (15.4)	0.931 (14.5)	0.921 (**9.1**)

brings us to 99.8% success in the fastest way. It was observed that 99.8% success was achieved with Logistic Regression using the 28th component and 99.3% success was achieved with Stochastic Gradient Descent using the 25th component as seen in Table 1. Considering the Logistic Regression classifier, it can be concluded that Multidimensional Scaling is more successful than Principal Component Analysis. In addition, four classifiers are more successful than other classifiers in Multidimensional Scaling as in Principal Component Analysis.

When we examine the classifier accuracy according to the Locally Linear Embedding method,we see that the Stochastic Gradient Descent model brings us to 99.8% accuracy in the fastest way. It was observed that 99.8% accuracy was achieved with Stochastic Gradient Descent using the 30th component and 99.6% accuracy was achieved with MLP using the 9th component as seen in Table 1. Considering the Stochastic Gradient Descent classifier, it can be concluded that Locally Linear Embedding is more successful than Principal Component Analysis. In this method, unlike Principal Component Analysis, a high accuracy could not be obtained with the Logistic Regression classifier. Isometric Feature Mapping and t-Distributed Stochastic Neighbor Embedding methods also were examined but these methods are more slowly and less successfully than Principal Component Analysis.

As can be seen in Table 1, ovarian cancer was tried to be detected with 9 classifiers and 5 methods. The accuracy of these methods and classifiers in detecting ovarian cancer were calculated by the Orange Data Mining program. All these calculations were found by taking the average of the accuracy obtained by running the program 10 times. Using the K-NN classifier, a 96.2% accuracy was achieved with only 12 components in the Locally Linear Embedding method. It is seen that the Locally Linear Embedding method is quite successful for this classifier compared to other manifold methods. By using the Decision Tree classifier, the components obtained with the Manifold methods could not find better results than the existing features. The highest accuracy obtained with this classifier is 97.3% and it was obtained using all features. The Support Vector Machines classifier achieved a accuracy of 99.6% using only 19 components obtained by the Locally Linear Embedding method. In order for the Principal Component Analysis and Multidimensional Scaling methods to achieve the same accuracy with this classifier, it is necessary to use much more features. Although the Stochastic Gradient Descent classifier achieved high results with Principal Component Analysis and Multidimensional Scaling methods, it achieved the highest accuracy with the Locally Linear Embedding method. Stochastic Gradient Descent achieved the highest accuracy in this study with 99.8% by using all 30 components that we obtained with the Locally Linear Embedding method. Unlike other classifiers, with the Logistic Regression classifier, it achieved the most successful result with the Multidimensional Scaling method. Logistic Regression achieved the highest accuracy in the study with 99.8% by using 28 components that we obtained with the Multidimensional Scaling method. As with the Decision Tree and Adaboost classifiers, the Random Forest classifier achieves the highest success when all features are used. The Multi-Layer Perceptron classifier achieved the highest result with the Locally Linear Embedding method by using the least number of components. A 99.6% accuracy was achieved using only 9 features. With this result, the obvious superiority of the Locally Linear Embedding method came to the fore. Naive Bayes classifier found the highest accuracy as 93.7%. It obtained this accuracy using the Locally Linear Embedding method. When the first 30 components were examined, it was observed that the first 17 components reached this result. As a result, it is seen that the Locally Linear Embedding method is more successful in 5 out of 9 classifiers. As a result of all these comments, it can be said that the Locally Linear Embedding method outperforms the other methods which used in this study. We recommend using the Locally Linear Embedding method instead of the Principal Component Analysis method, which is used for easy examination of large datasets in the medical field.

Table 2
Comparison Classifier Performances to Similar Studies from The Literature. Abbreviations of the algorithms are RNN (Optimal Recurrent Neural Networks), SOM (Self-Organizing Map), MLP (Multi-Layer Perceptron), aSLFN (Adaptive Single-Hidden Layer Feedforward Neural Network), BiELM (Bayesian Initialization of Extreme Learning Machine), CNN (Convolutional Neural Network), LR (Logistic Regression), SGD (Stochastic Gradient Descent), IFS (Integrated Feature Selection), CBFS (Correlation Based Feature Selection), PCA (Principal Component Analysis), MDS (Multi-Dimensional Scaling), and LLE (Locally Linear Embedding).

Study	Classifier	Optimization	Validation	ACC
[23]	SVM	–	70% + 30%	0.64
[28]	aSLFN	–	10-fold	0.72
[23]	MLP	–	70% + 30%	0.78
[29]	BiELM	–	10-fold	0.80
[22]	SVM	Procrustes	10-fold	0.95
[32]	RNN + SOM	–	90% + 10%	0.96
[31]	CNN	–	70% + 30%	0.98
[18]	MLP	Taguchi	70% + 30%	0.98
[17]	SVM	PSO	10-fold	0.99
[30]	MLP	IFS	70% + 30%	0.99
This study	**LR**	**MDS**	70% + 30%	**0.99**
This study	**SGD**	**LLE**	70% + 30%	**0.99**
[20]	SVM	CBFS	5-fold	1.00
[21]	SVM	Fisher Score	40% + 60%	1.00
[24]	SVM	CS4	10-fold	1.00
[25]	SVM	PCA + LDA	3-fold	1.00
[27]	MLP	PCA	90% + 10%	1.00

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Matjaž Perc was supported by the Slovenian Research Agency (Grant Nos. P1-0403 and J1-2457).

References

[1] J.S. Richards, S.A. Pangas, et al., The ovary: basic biology and clinical implications, J. Clin. Invest. 120 (4) (2010) 963–972, http://dx.doi.org/10.1172/JCI41350.

[2] T. Arora, S. Mullangi, M.R. Lekkala, Ovarian cancer, StatPearls [Internet] (2021) URL https://www.ncbi.nlm.nih.gov/books/NBK567760/.

[3] I.N.G. Budiana, M. Angelina, T.G.A. Pemayun, Ovarian cancer: Pathogenesis and current recommendations for prophylactic surgery, J. Turkish German Gynecol. Assoc. 20 (1) (2019) 47, http://dx.doi.org/10.4274/jtgga.galenos.2018.2018.0119.

[4] A. Pietrangelo, R. Chen, The outlook for ovarian cancer: prognosis, life expectancy, and survival rates by stage, 2020, URL https://www.healthline.com/health/cancer/ovarian-cancer-outlook.

[5] M. Gentry, World cancer research fund international (WCRF), Impact 2017 (4) (2017) 32–33, http://dx.doi.org/10.21820/23987073.2017.4.32.

[6] L. Chin, R.N. Hansen, J.J. Carlson, Economic burden of metastatic ovarian cancer in a commercially insured population: A retrospective cohort analysis, J. Managed Care Specialty Pharm. 26 (8) (2020) 962–970, http://dx.doi.org/10.18553/jmcp.2020.26.8.962.

[7] J.S. Carter, L.S. Downs Jr., Ovarian cancer tests and treatment, The Female Patient 36 (4) (2011) 30, URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3154728/.

[8] L.J. Havrilesky, P.G. Moorman, W.J. Lowery, J.M. Gierisch, R.R. Coeytaux, R.P. Urrutia, M. Dinan, A.J. McBroom, V. Hasselblad, G.D. Sanders, et al., Oral contraceptive pills as primary prevention for ovarian cancer: a systematic review and meta-analysis, Obstet. Gynecol. 122 (1) (2013) 139–147, http://dx.doi.org/10.1097/AOG.0b013e318291c235.

[9] A. Agarwal, R. Maheswaran, V. Sehgal, R. Khosa, B. Sivakumar, C. Bernhofer, Hydrologic regionalization using wavelet-based multiscale entropy method, J. Hydrol. 538 (2016) 22–32, http://dx.doi.org/10.1016/j.jhydrol.2016.03.023.

[10] R. Shukla, A. Agarwal, C. Gronott, K. Sachdeva, P.K. Joshi, Farmer typology to understand differentiated climate change adaptation in the Himalayas, Sci. Rep. 9 (2019) 20375, http://dx.doi.org/10.1038/s41598-019-56931-9.

[11] T. Sarac, Estimation for survival of lung cancer patients depending on different attributes, J. Intell. Syst. Appl. 1 (2) (2018) 145–148, http://dx.doi.org/10.54856/jiswa.201812044.

[12] E. Sayilgan, O. Karabiber Cura, Y. Isler, Use of clustering algorithms and extreme learning machine in determining arrhythmia types, in: 2017 25th Signal Processing and Communications Applications Conference (SIU), IEEE, 2017, pp. 1–4, http://dx.doi.org/10.1109/SIU.2017.7960236.

[13] A. Narin, M. Ozer, Y. Isler, Effect of linear and non-linear measurements of heart rate variability in prediction of PAF attack, in: 2017 25th Signal Processing and Communications Applications Conference (SIU), IEEE, 2017, pp. 1–4, http://dx.doi.org/10.1109/SIU.2017.7960358.

[14] E. Sayilgan, Y.K. Yuce, Y. Isler, Evaluation of mother wavelets on steady-state visually-evoked potentials for triple-command brain-computer interfaces, Turk. J. Electr. Eng. Comput. Sci. 29 (5) (2021) 2263–2279, http://dx.doi.org/10.1038/s41598-019-56931-9.

[15] E. Sayilgan, Y.K. Yuce, Y. Isler, Evaluation of wavelet features selected via statistical evidence from steady-state visually-evoked potentials to predict the stimulating frequency, J. Faculty Eng. Archit. Gazi Univ. 36 (2) (2021) 593–605, http://dx.doi.org/10.1038/s41598-019-56931-9.

[16] Y. Isler, A detailed analysis of the effects of various combinations of heart rate variability indices in congestive heart failure, (Ph.D. thesis), Dokuz Eylul University, 2009.

[17] E.-G. Talbi, L. Jourdan, J. Garcia-Nieto, E. Alba, Comparison of population based metaheuristics for feature selection: Application to microarray data classification, in: 2008 IEEE/ACS International Conference on Computer Systems and Applications, IEEE, 2008, pp. 45–52, http://dx.doi.org/10.1109/AICCSA.2008.4493515.

[18] M.A. Rahman, R.C. Muniyandi, K.T. Islam, M.M. Rahman, Ovarian cancer classification accuracy analysis using 15-neuron artificial neural networks model, in: 2019 IEEE Student Conference on Research and Development (SCOReD), IEEE, 2019, pp. 33–38, http://dx.doi.org/10.1109/SCORED.2019.8896332.

[19] P. Yasodha, N. Ananthanarayanan, Detecting the ovarian cancer using big data analysis with effective model, Biomed. Res. (0970-938X) (2018) http://dx.doi.org/10.4066/biomedicalresearch.29-17-155.

[20] H. Basegmez, E. Sezer, C.S. Erol, Optimization for gene selection and cancer classification, in: Multidisciplinary Digital Publishing Institute Proceedings, Vol. 74, 2021, p. 21, http://dx.doi.org/10.3390/proceedings2021074021.

[21] H. Demircioglu, H. Bilge, Analysis of gene expressions in ovarian cancer data set by using data mining, Marmara J. Pure Appl. Sci. 27 (4) (2015) 125–134, http://dx.doi.org/10.7240/mufbed.89154.

[22] Q. Liu, Q. Gu, Z. Wu, Feature selection method based on support vector machine and shape analysis for high-throughput medical data, Comput. Biol. Med. 91 (2017) 103–111, http://dx.doi.org/10.1016/j.compbiomed.2017.10.008.

[23] S.H.S.A. Ubaidillah, R. Sallehuddin, N.A. Ali, Cancer detection using aritifical neural network and support vector machine: A comparative study, Jurnal Teknologi 65 (1) (2013) http://dx.doi.org/10.11113/jt.v65.1788.

[24] J. Li, K. Ramamohanarao, A tree-based approach to the discovery of diagnostic biomarkers for ovarian cancer, in: Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer, 2004, pp. 682–691, http://dx.doi.org/10.1007/978-3-540-24775-3_80.

[25] Y. Liu, Dimensionality reduction and main component extraction of mass spectrometry cancer data, Knowl.-Based Syst. 26 (2012) 207–215, http://dx.doi.org/10.1016/j.knosys.2011.08.006.

[26] L. Gao, M. Ye, C. Wu, Cancer classification based on support vector machine optimized by particle swarm optimization and artificial bee colony, Molecules 22 (12) (2017) 2086, http://dx.doi.org/10.3390/molecules22122086.

[27] D. Yesilbas, A. Guven, Classification of ovarian cancer with neural networks using mass spectrometry data, Cukurova Univ. J. Faculty Eng. 36 (3) (2021) 781–790, http://dx.doi.org/10.21605/cukurovaumfd.1005791.

[28] S. Belciug, F. Gorunescu, Learning a single-hidden layer feedforward neural network using a rank correlation-based strategy with application to high dimensional gene expression and proteomic spectra datasets in cancer detection, J. Biomed. Inform. 83 (2018) 159–166, http://dx.doi.org/10.1016/j.jbi.2018.06.003.

[29] S. Belciug, R.C. Ivanescu, A Bayesian framework for extreme learning machine with application for automated cancer detection, An. Univ. Craiova-Math. Comput. Sci. Ser. 46 (1) (2019) 189–202, URL http://inf.ucv.ro/~ami/index.php/ami/article/view/1039/675.

[30] A. Al-Murad, M.F. Hossain, An integrated feature selection method for neural network to classify ovarian cancer, in: 2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI), 2021, pp. 1–6, http://dx.doi.org/10.1109/ACMI53878.2021.9528106.

[31] S. Kilicarslan, K. Adem, M. Celik, Diagnosis and classification of cancer using hybrid model based on relieff and convolutional neural network, Med. Hypotheses 137 (2020) 109577, http://dx.doi.org/10.1016/j.mehy.2020.109577.

[32] M. Elhoseny, G.-B. Bian, S. Lakshmanaprabu, K. Shankar, A.K. Singh, W. Wu, Effective features to classify ovarian cancer data in internet of medical things, Comput. Netw. 159 (2019) 147–156, http://dx.doi.org/10.1016/j.comnet.2019.04.016.

[33] K. Li, H. Xu, X. Liu, Analysis and visualization of accidents severity based on LightGBM-TPE, Chaos Solitons Fractals 157 (2022) 111987, http://dx.doi.org/10.1016/j.chaos.2022.111987.

[34] L. Chen, J. Sun, K. Li, Q. Li, Research on the effectiveness of monitoring mechanism for “yield to pedestrian” based on system dynamics, Physica A 591 (2022) 126804, http://dx.doi.org/10.1016/j.physa.2021.126804.

[35] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, second ed., Wiley-Interscience, 2000.

[36] A. Sahin, N. Ozcan, G. Nur, Ovarian cancer prediction using PCA, K-PCA, ICA and random forest, J. Intell. Syst. Appl. 4 (2) (2021) 103–108, http://dx.doi.org/10.54856/jiswa.202112168.

[37] L. Van der Maaten, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (11) (2008) URL https://www.jmlr.org/papers/v9/vandermaaten08a.html.

[38] C.J. Burges, Dimension Reduction: A Guided Tour, Now Publishers Inc, 2010, http://dx.doi.org/10.1561/2200000002.

[39] I.K. Fodor, A survey of dimension reduction techniques, Tech. rep., Lawrence Livermore National Lab., CA (US), 2002, http://dx.doi.org/10.2172/15002155.

[40] A. Ghodsi, Dimensionality Reduction a Short Tutorial, Vol. 37, (38) Department of Statistics and Actuarial Science, Univ. of Waterloo, Ontario, Canada, 2006, p. 2006, URL https://www.math.uwaterloo.ca/~aghodsib/courses/f06stat890/readings/tutorial_stat890.pdf.

[41] A.B. Graf, F.A. Wichmann, Gender classification of human faces, in: International Workshop on Biologically Motivated Computer Vision, Springer, 2002, pp. 491–500, http://dx.doi.org/10.1007/3-540-36181-2_49.

[42] H.S. Bıcakcı, M. Santopietro, M. Boakes, R. Guest, Evaluation of electrocardiogram biometric verification models based on short enrollment time on medical and wearable recorders, in: 2021 International Carnahan Conference on Security Technology (ICCST), 2021, pp. 1–6, http://dx.doi.org/10.1109/ICCST49569.2021.9717372.

[43] L. Bottou, Stochastic gradient learning in neural networks, Proc. Neuro-Nimes 91 (8) (1991) 12.

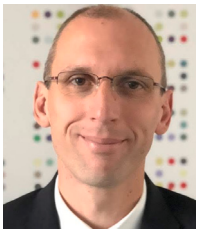
[44] B. Kegl, Introduction to AdaBoost, 2009, URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.679.8866&rep=rep1&type=pdf.

[45] G. Biau, E. Scornet, A random forest guided tour, TEST 25 (2016) 197–227, http://dx.doi.org/10.1007/s11749-016-0481-7.

[46] B. Chandra, M. Gupta, M. Gupta, Robust approach for estimating probabilities in naive-Bayes classifier, in: International Conference on Pattern Recognition and Machine Intelligence, Springer, 2007, pp. 11–16, http://dx.doi.org/10.1007/978-3-540-77046-6_2.

[47] S.X. Wu, W. Banzhaf, The use of computational intelligence in intrusion detection systems: A review, Appl. Soft Comput. 10 (1) (2010) 1–35, http://dx.doi.org/10.1016/j.asoc.2009.06.019.
[48] C. Kaya, O. Yıldız, Intrusion detection with machine learning techniques: Comparative analysis, Marmara J. Pure Appl. Sci. 3 (2014) 89–104, http://dx.doi.org/10.7240/mufbed.24684, URL http://hdl.handle.net/11424/3976.
[49] E.A. Sagbas, S. Ballı, Transportation mode detection by using smartphone sensors and machine learning, Pamukkale Univ. J. Eng. Sci. 22 (5) (2016) 376–383, http://dx.doi.org/10.5505/pajes.2015.63308.
[50] S. Osowski, K. Siwek, T. Markiewicz, MLP and SVM networks-a comparative study, in: Proceedings of the 6th Nordic Signal Processing Symposium, 2004. NORSIG 2004, IEEE, 2004, pp. 37–40, URL https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1344517.
[51] T. Hastie, R. Tibshirani, J. Friedman, The elements of statistical learning, Cited on (2009) 33, http://dx.doi.org/10.1007/978-0-387-84858-7.
[52] C.M. Bishop, Pattern recognition, Mach. Learn. 128 (9) (2006) URL https://www.springer.com/gp/book/9780387310732.
[53] A. Ben-Hur, J. Weston, A user's guide to support vector machines, in: Data Mining Techniques for the Life Sciences, Springer, 2010, pp. 223–239, http://dx.doi.org/10.1007/978-1-60327-241-4_13.
[54] C.J. Burges, A tutorial on support vector machines for pattern recognition, Data Min. Knowl. Discov. 2 (2) (1998) 121–167, http://dx.doi.org/10.1023/A:1009715923555.
[55] K. Soman, R. Loganathan, V. Ajay, Machine Learning with SVM and Other Kernel Methods, PHI Learning Pvt. Ltd., 2009.
[56] N. Ozer, Y. Isler, H. Ozer, A computer software for simulating single-compartmental model of neurons, Comput. Methods Programs Biomed. 75 (1) (2004) 51–57, http://dx.doi.org/10.1016/j.cmpb.2003.08.002.
[57] C. Baxter, Q. Zhang, S. Stanley, R. Shariff, R.-R. Tupas, H. Stark, Drinking water quality and treatment: the use of artificial neural networks, Can. J. Civil Eng. 28 (S1) (2001) 26–35, http://dx.doi.org/10.1139/l00-053.
[58] Y. Zhang, Q. Shang, G. Zhang, pyDRMetrics-A Python toolkit for dimensionality reduction quality assessment, Heliyon 7 (2) (2021) e06199, http://dx.doi.org/10.1016/j.heliyon.2021.e06199.

Bartu Yesilkaya graduated from the Dokuz Eylül University with B.Sc. Electrical and Electronics Engineering degree in 2016. In 2020, he achieved M.Sc. degree from department of Biomedical Technologies in Izmir Katip Celebi University. Since 2018, he has been working in Biomedical Engineering as a research assistant at the same university. He is currently studying to achieve Phd degree at the department of biomedical technologies and his research interests are biomedical signals and machine learning algorithms.



Matjaž Perc is professor of physics at the University of Maribor. He is a member of Academia Europaea and the European Academy of Sciences and Arts, and among top 1% most cited physicists according to 2020 and 2021 Clarivate Analytics data. He is also the 2015 recipient of the Young Scientist Award for Socio and Econophysics from the German Physical Society, and the 2017 USERN Laureate. In 2018 he received the Zois Award, which is the highest national research award in Slovenia. In 2019 he became Fellow of the American Physical Society. Since 2021 he is Vice Dean of Natural Sciences at the European Academy of Sciences and Arts.

Yalcin Isler is associate professor of Biomedical Engineering at Izmir Katip Celebi University. His main research interests cover biomedical signal processing, pattern recognition, medical device design, programming, and embedded systems. In addition to academics, he founded Islerya Medical and Information Technologies Company, Izmir, Turkey, in 2017. He was an IEEE Student Member (2009–2011) and IEEE Member (2011–2014) and has been IEEE Senior Member since 2014.