



## Comparison of the Effects of Different Dimensional Reduction Algorithms on the Training Performance of Anfis (Adaptive Neuro-Fuzzy Inference System) Model

Ahmet Gürkan YUKSEK<sup>1\*</sup>, Halil ARSLAN<sup>1</sup>, Oğuz KAYNAR<sup>2</sup>, Emre DELIBAS<sup>1</sup>,  
Abdülkadir SEKER<sup>1</sup>

<sup>1</sup>Cumhuriyet University, Computer Engineering Departman, 58140, Sivas / TÜRKİYE

<sup>2</sup>Cumhuriyet University, Management Information Systems, 58140, Sivas / TÜRKİYE

Received: 29.10.2017; Accepted: 21.11.2017

<http://dx.doi.org/10.17776/csj.347653>

**Abstract:** Adaptive Neuro-Fuzzy Inference Systems (ANFIS) is a hybrid artificial neural network (intelligence) approach that utilizes the ability of artificial neural networks to learn, generalize, paralyze and to derive fuzzy logic. The development of models with large numbers of input variables with ANFIS is not very convenient for applications. Dimension reduction methods are proposed as a solution to this problem. Dimensional Reduction is the method used to represent the data in a lower dimensional space. The reduction of the numbers of the input variables using different size reduction methods and the creation of the optimal solution of the probing with the ANFIS model constitute the framework of this work. In this study, we compared the results produced by different dimension reduction methods and investigated which method is more acceptable for ANFIS training.

**Keywords:** ANFIS, Dimensionality Reduction, Deep Learning, Autoencoder.

## Farklı Boyut İndirgeme Yöntemlerinin, Anfis Modelinin Eğitim Performansı Üzerindeki Etkilerinin Karşılaştırılması

**Özet:** Uyarlamalı Sinirsel Bulanık Çıkarım Sistemleri-ANFIS (Adaptive Neuro-Fuzzy Inference Systems-ANFIS), yapay sinir ağları modellerinin öğrenme, genelleme ve paralel hesap yapabilme yetenekleri ile bulanık mantığın çıkarsama yeteneğini kullanan hibrit bir yapay sinir ağı (zekâ) yaklaşımıdır. Giriş değişkenlerinin sayıları fazla olan modellerin ANFIS ile geliştirilmesi uygulamalarda çok elverişli değildir. Boyut indirgeme yöntemleri ise bu soruna bir çözüm yolu olarak önerilmektedir. Boyut indirgeme, verilerin daha düşük boyutlu bir uzayda temsil edilmesi için kullanılan yöntemdir. Farklı boyut indirgeme yöntemleri kullanılarak giriş değişkenlerinin sayılarının indirgenmesi ve ANFIS modeli ile probleme ait en uygun çözümün oluşturulması bu çalışmanın çatısını oluşturmaktadır. Bu çalışmada, farklı boyut indirgeme yöntemlerinin ürettiği sonuçlar karşılaştırılarak ANFIS'in eğitimi için hangi yöntemin kullanılmasının daha kabul edilebilir olduğu araştırılmıştır.

**Anahtar Kelimeler:** ANFIS, Boyut İndirgeme, Derin Öğrenme, Oto Kodlayıcı.

### 1- INTRODUCTION

Dimensionality reduction is to model the dataset in such a way that it can best represent the features of a smaller size space. Dimension reduction of high-dimensional data sets is a significant step in the preparation of preliminary data for applications to

be performed on many real-world data sets [1]. In other words; dimensional reduction is the mapping of data to a lower dimensional space in such a way that there is no non-informative variance of the data or a subspace where the data can be represented, and has a long history as a method for extracting low dimensional features of the data dimension [2].

\* Corresponding author. Email address: [agyuksekk@cumhuriyet.edu.tr](mailto:agyuksekk@cumhuriyet.edu.tr)  
<http://dergipark.gov.tr/csj> ©2016 Faculty of Science, Cumhuriyet University

For real world data, high dimensionality is a very common occurrence and the number of parameters (inputs) to be used in the model must be reduced as much as possible so as to reflect the structure of the problem [3]. As a result, dimensional reduction facilitates the classification, visualization and compaction of high-dimensional data as well as offering many other advantages. Data mining, machine learning, system identification and statistical methods encountered in the literature include a large number of inputs, and their reduction must be applied successfully so as not to impair the representation of the data set [4]. Many different dimension reduction algorithms have been developed for data dimension reduction, pre-processing for supervised learning, and data visualization [5]. In this study, the performances of nonlinear techniques were investigated on artificial and natural tasks. The results of the experiments indicate that nonlinear techniques perform well in selected artificial tasks, but this powerful performance can not cover real world tasks. The paper explains these results by identifying the weaknesses of existing nonlinear techniques and suggests how to improve the performance of nonlinear size reduction techniques. It has contributed to this area because they need methods of dimension reduction in solving many different academic disciplinary problems. Although all of the methods used have similar objectives, approaches to solving the problem are different [6]. The majority of Learning Algorithms have difficulties in the analysis, computation, and processing of data for high-dimensional data processing tasks, which may well lead to unsuccessful results. In short, most of these algorithms are not designed to handle large, complex, and different data such as real-world data sets [1].

Anfis, which is known as an approach that combines parallel computation and learning features of Artificial Neural Networks (ANN) models with fuzzy logic (Fuzzy-Logic FL) inference, is firstly used extensively in the inference operations proposed by Jang [7]. Fuzzy system structures can successfully model the functionalization of different forms with the help of

ANN's learning and generalization ability, which can model linear or non-linear relations with architectural reason, while transferring expert experiences and opinions with the help of verbal expressions. [8-10]. The ANFIS model is a flexible computational technique that is directly and very strongly dependent on the data set used to train the model (all inputs and outputs) and is suitable for modelling the behaviour of complex systems by learning system behaviour from the existing measurement set. For nonlinear functions it is a more appropriate approach than classical statistical tools [11]. Utilizing the current advantages of fuzzy logic and ANN methods, ANFIS has been successfully implemented in recent years, especially in many different disciplines such as control, estimation, classification and time series analysis. It has been used extensively in the prediction of chaotic time series, especially in the modelling of nonlinear functions [7]. Real world problems are represented by nonlinear relationships between many different parameters, and often the number of parameters that model the system is high. The aim of this project is to use the ANFIS for establishing a model for the air quality of Sivas, a middle-east country of Turkey and to employ the developed model as an early warning system. The most revealing definition of air pollution is "the amount of pollutants in the atmosphere that can cause harmful effects on human health, plants, structures and materials" [12]. The comparison of clean and polluted air for different pollutants by the given quantity designation is shown in Table 1.

The model is thought to design the pollution after a certain time by using meteorological and environmental data to be presented to it. The ANFIS model, which will determine the magnitude of 1 output value (Table 2) out of approximately 12 input parameters, is the basis of the study. The data set to be used for estimating and establishing the air quality model is the data obtained from measurements made at 5 different stations in Sivas province between 1990 and 2004 (Figure 1).

The generated training set includes about 3150 days of air pollution and meteorological

measurement values. Table 2 indicates monthly averages of training set parameters between 1990 and 2004. Here, while 12 values form the input parameters for the model is formed, the SO<sub>2</sub> concentration after one day is selected as the output parameter. However, the SO<sub>2</sub> concentration and meteorological data of P.M. (Particulate Matter) concentration has also been added to the data set as input parameters, so that the existing effects of pollution parameters in the atmosphere are also included in the model. Namely, the effect of the air pollution level of a previous time period on the air pollution in the next period of time has been included in these models.

As is known, the numerical surplus of the input parameters in the dataset is an undesirable condition for ANFIS models and is not much preferred because it increases the computation burden, duration and hardware requirements [13]. This disadvantage of the ANFIS models is highlighted in the literature [14, 15], in which studies of multiple ANFIS models are carried out in the literature by segmenting the data set in different studies. The size reduction method can also be presented as an alternative solution method for modelling such large parameter data sets with ANFIS.

**Table 1.** Comparison of dirty atmosphere with clean atmosphere.

Trace Gaz (ppm)	Clean Air	Polluted Air	Polluted/Clean
CO <sub>2</sub>	320	400	1.3
CO	0.1	40 -70	400 -700
CH <sub>4</sub>	1.5	2.5	1.3
N <sub>2</sub> O	0.25	?	?
NO <sub>2</sub> (NO <sub>x</sub> )	0.001	0.2	200
O <sub>3</sub>	0.02	0.5	25
SO <sub>2</sub>	0.0002	0.2	1000
NH <sub>3</sub>	0.001	0.02	2



**Figure 1.** City Layout Plan of Measurement Stations (Satellite View).

**Table 2.** Statistical ratios of the parameters presented to the Air Quality Estimation Model.

Parameter	Unit	Avg.	Max.	Min
Daily Maximum Temperature	°C	12,4	38,2	-14,2
Daily Minimum Temperature	°C	1	18,0	-27,2
Daily Average Precipitation	mm	1,3	55,0	0
Daily Average Wind	m/sec	1,3	6,2	0
Daily Average Humidity	%	68,3	96,0	26,3
Daily Average Pressure	Mb	87,2	88,5	84,8
Daily Average Steam Pressure	Mb	69,4	182	9
Daily Average Cloudiness	-	4,6	10	0
Daily Sunshine Time	Hour	5,9	14,8	0
Daily Sunshine Severity	CAL/CMy	284,6	696,6	0
SO <sub>2</sub>	Mgr/m <sup>3</sup>	144,7	1976	9
P.M (Particulate matter)	Mgr/m <sup>3</sup>	118,9	871	10

While developing an ANFIS model for estimating the intended air quality of this work, it suggests a method that can be employed in similar studies to analyse the effects of these methods on ANFIS by reducing the data set with different dimension reduction methods.

In the second part of the work, ANFIS, which is proposed as a solution method in the problem of air quality estimation, is explained through basic working principles. In the third section, the methods employed to reduce the size of the data set used to train ANFIS are briefly discussed, and then the results of the ANFIS model trained with size reduced data sets using these methods are presented in the related tables and graphs. The results of the methods, comparisons, advantages and disadvantages are discussed in the conclusion section.

## 2- THE STRUCTURE OF ANFIS (ADAPTIVE NEURO-FUZZY INFERENCE SYSTEMS) AND PREDICTION MODEL

ANFIS architecture consists of a fuzzy system of sugeno structure, as a network with artificial neural learning ability. The network consists of the joining of successive nodes arranged in layers in order to perform a specific function, each of which is unique to them [16]. If the fuzzy inference system is derived from the assumption that there are two inputs such as x and y and an output such as z, the set of rules for the first-order Sugeno fuzzy model inference with two fuzzy If-Then rules. The formula is as following:

Rule-1 If x is A<sub>1</sub> is and y is B<sub>1</sub>, then  $f_1 = p_1x + q_1y + r_1$

Rule-2 If x is A<sub>2</sub> is and y is B<sub>2</sub>, then  $f_2 = p_2x + q_2y + r_2$

The reasoning structure for a Sugeno type fuzzy model is shown in Figure 2. The equivalent ANFIS structure, which is equivalent to the indicated structure, is shown in Figure 3. The nodes that are found in the same layer for the ANFIS structure in question have the same node functions as shown in Figure 2. (Here it is stated as the output of the first node in the first layer.)

**Layer I:** Each i node in this layer is an adaptive node the output of which is defined as in Equation 1.

$$O_{i,j} = \mu_{A_i}(x), i = 1,2, \quad \text{Or} \quad 1$$

$$O_{i,j} = \mu_{B_{i-2}}(y), i = 3,4, \quad \text{for}$$

Here x (or y) is the node input A<sub>i</sub> (or B<sub>i-2</sub>), which refers to the fuzzy cluster belonging to the node.

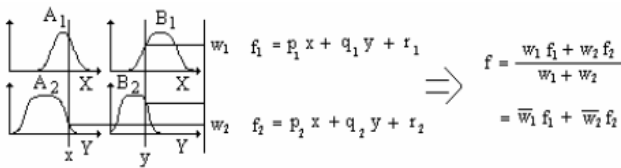


Figure 2. Fuzzy Model with Two Inputs and Two Rules in the First Grade.

In other words, the outputs of the mentioned layer constitute the membership values of the conditions or premise parts of the rules. It can be understood that for A<sub>i</sub> and B<sub>i</sub>, this could be a membership function. For example, the generalized bell curve function expressed in Equation 2 is expressed.

$$\mu_A = \frac{1}{1 + \left[ \left( \frac{x - c_i}{a_i} \right)^2 \right]^{bi}} \quad 2$$

The set {a<sub>i</sub>, b<sub>i</sub>, c<sub>i</sub>} is a set of parameters that is mentioned at Equation 2. The parameters of this layer are also defined as conditional or input parameters.

**Layer II:** Each node in the second layer is a fixed node labelled with, which produces as output the multiplication of the incoming signals. For example;

$$O_{2,i} = w_i = \mu_{A_i}(x) \mu_{B_i}(y), i = 1,2. \quad 3$$

The output of each node forms the degree of realization for each rule. Other T-norm operations that perform fuzzy (and) processing instead of multiplication in Equation 3 can be used as node functions.

**Layer III:** Each node in this layer is a fixed node labelled with an N value. I. node in layer i. calculates the ratio of the rule's realization rate to the sum of all rule's realization ratios.

$$O_{3,i} = \bar{w}_i = \frac{w_i}{w_1 + w_2} = 1,2. \quad 4$$

The outputs of the nodes in the third layer are called as normalized realization ratios in accordance with their computation.

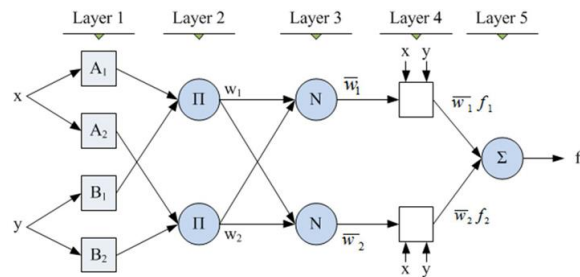


Figure 3. Equivalent ANFIS Construction.

**Layer IV:** Each node in IV layer is an adaptive node whose node function is as in Equation 5.

$$O_{4,i} = \bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i) \quad 5$$

Here  $\bar{w}_i$ , is output from III. Layer and {p<sub>i</sub>, q<sub>i</sub>, r<sub>i</sub>} values are set consisting of parameters in this layer. The parameters in this layer will be accepted as output or result parameters.

**Layer V:** In this layer, which is the last layer, there is a fixed node, denoted by Σ, which collects the whole of the signals it receives to calculate the total output.

$$O_{5,i} = \text{OverallOutput} \quad \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad 6$$

Thus, an adaptive network structure with the function of the Sugeno fuzzy model is established [7].

The primary problem that should be addressed in this study is that the size of the data set to be used for training the model (12 inputs, 1 output) is larger than the tolerances for training the ANFIS and therefore the model is being trained unstable or untrained state. Decreasing the size of the data set by selecting parameters for the solution is a commonly used method. By choosing the most important disadvantage (according to various parameter selection methods or individual expertise and experience), the behaviour of the parameters derived from the training set of the model can never be represented. This is an undesirable situation for adaptive methods such as ANFIS. The more different behaviours and effects modelled during the system training, the better the success is. For this reason, any parameter in the data set which is formed from the measurable values and which has an influence on the solution of the problem should be done without leaving the model. The most important step during the creation of ANFIS-like flexible system models and the most effective factor on the success of the model is the organization of the data set [17]. The generated dataset contains direct factors that affect the outcome of the model to be developed [18]. Four basic steps to be followed when using models such as ANFIS in complex data analysis operations are shown in Figure 4. In the first step, the problem description and the expected results that guide in the following steps are determined. The second part is the preparation of the data set which can answer the expected results and model the problem in the best way. It is trained by the ANFIS iterations that were decided to be built in the third stage. In the last stage, data mining is performed on the results produced by the ANFIS model and actual results, and the success of the model is tested [19].

The ANFIS basic algorithm combines the idea of artificial neural networks learning, combining

features such as finding the most optimal and parallel-connected architectures, and combining the advantages of fuzzy logic such as decision-making and expert knowledge, experience and visibility. Thus, fuzzy control systems are given the ability to provide decision-making and expert knowledge, such as fuzzy control, while neural networks are given the power to learn, generalize and calculate [20]. The main purpose of the ANFIS system is to use artificial neural learning methods to determine its own structure, variables. In fuzzy logic systems, two important adjustments are made: variable setting and structure setting. Structural adjustment; the number of variables (parameters) to be used in the model, the number of rules, and the partitioning of the definition spaces of each input / output variable. Once the probabilistic rule structure is established, controller variables need to be set. In the setting of the variables, the appropriate centres, slopes, widths of the membership functions and weights of the fuzzy logic rules are calculated [21]. The ANFIS model was used to solve the problem of estimating SO<sub>2</sub> value for the next day with the "Neural Fuzzy Logic" method. The MATLAB program was used for the training of the ANFIS model and for the size reduction calculations of the data set. The size of the data set is reduced by different dimension reduction methods which are examined in detail in the next section. All reduced datasets to be used in ANFIS training are divided into four parts as "Training Set", "Verification Set", "Test Set" and "Control Set". The Training, Validation and Test sets in this section are presented to the modelling models for training ANFIS models and used to generate model parameters. The control data set was used for the purpose of checking the accuracy of the values produced by the models in which the trainings ended. The actual outputs were presented to the ANFIS model to compare and to determine the success of the model.

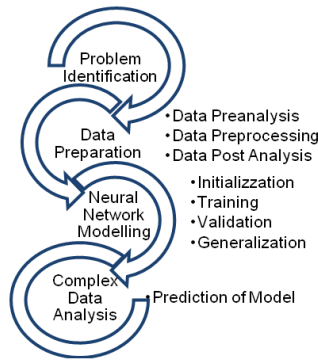


Figure 4. Flexible Programming Data Analysis Process.

### 3- REDUCTION OF DIMENSIONS OF METHOD DATA SET AND TRAINING OF ANFIS MODEL

A number of different methods are used for dimension reduction in the literature, and a broad summary of these methods is presented by Maaten [5] as an academic review. In this study, the methods of dimensional reduction are summarized under two main groups as methods that optimize an objective function that does not include local optima (Group I) and methods that optimize objective functions that include local optima (Group II). Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA), IZOMAP and LLE (Local Linear Embedding) methods employed in this study are investigated in the first group methods. Similarly, Sammon Mapping and Auto Encoder LLC (Locally Linear Coordination) methods are investigated as group II methods. There are no basic criteria to consider when deciding on the choice of methods. Experiments on the data set have been examined by making (default) experiments on the most general structures of the methods.

#### 3.1 Principle Component Analysis (PCA)

PCA, also known as the Karhunen-Loeve method, is a multivariate analysis that reduces the size by describing the covariance / variance structure of a set of variables, which is used in disciplines of recognition, classification, image compression, face recognition, seismic sequence analysis, and data mining. It is a statistical method. The variance structure of an original  $p$  variable is fewer than the present and can be summarized as a method of

explaining these variables over new variables with linear components [22]. More specifically,  $(p)$  is the linear components of the original variables with no correlation between and fewer than the original number of variables ( $p > k$ ) [23]. The variable  $(p)$  of the number of measurements  $(n)$  expressing the dependency structure in the mentioned method;  $(k)$  new variables that is linear, orthogonal and independent of each other. It finds the most general properties in a high-dimensional data, reducing the number of dimensions and compressing the data. It is certain that some features will be lost with size reduction; however, it is intended that these lost properties have little knowledge of the dataset [23]. It is often preferred because it is an approach to remove the dependency structure in the related works and reduce the size of the data set. It aims to find the strongest image within the data set and to express the diversity of the data set with a small set of dimensions selected from the whole dimension. Another significant advantage is that they are removed from the dataset as a result of the size reduction, since the noisy loudness is weaker than the pattern. Values and vectors of the covariance matrix or correlation matrix are used to find the linear components of the  $p$  variable in the data matrix [24].

In the PCA, up to  $p$  number of basic components (Principle Components) can be obtained from a data matrix containing  $p$  number of variables.

$$\begin{aligned}
 Y_1 &= a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \\
 Y_2 &= a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p, \dots \dots \dots \\
 Y_p &= a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p
 \end{aligned}
 \tag{7}$$

At the equation 7  $Y_1, Y_2, \dots, Y_p$  for PC and  $a_{ij}$ ,  $i$ . PC for  $j$ . shows the weight of the variable. PC weights  $(a_{ij})$ , is calculated to form the determined conditions:

The first principal component contributes most to the total variance, while the other components contribute to the total variance in decreasing amounts.

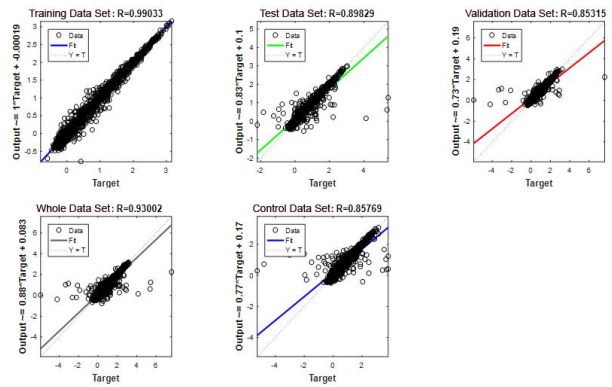
$$\begin{aligned}
 a_{i1}^2 + a_{i2}^2 + \dots + a_{ip}^2 &= 1 && (i=1,2,\dots, p) \\
 a_{i1}^2 a_{j1}^2 + a_{i2}^2 a_{j2}^2 + \dots + a_{ip}^2 a_{jp}^2 &= 0 && (\text{for each } i \neq j)
 \end{aligned}$$

In PCA, each component is calculated as a linear function of all the components in the analysis. Derived PCs have a maximum contribution to the total variance, respectively. According to the second condition, PC weights must be a sum of squares. In order to satisfy this condition, all variables in the analysis need to be purified from variant. In other words, it is necessary to standardize the original variables before PCA. The third condition, which is mandatory, is used to determine the scale of the new variables. Because it is possible to increase the variance by changing the scale of PC weights [25].

The data set was reduced from 12 input parameters to 4 input parameters using coding developed by MATLAB using the PCA method. The reconstructed data set as 4 inputs 1 output is used to set up, train and verify the ANFIS model by dividing randomly into 4 different parts (training, verification, test and control) as mentioned in the previous section. In the succeeding sub-sections, the same methods were used for different size reduction methods and reduced size data sets. The method of separating the used data set was done by using the same algorithm in all processes. As a result of the studies performed with PCA, the values shown in Figure 5 and Table 3 were obtained.

**Table 3.** Statistical Values Produced by ANFIS Model Established by PCA.

	Model Reduced by ANFIS - PCA				
	Training	Test	Validation	All	Control
Mse	0.12532	0.4158	1.17649	0.63036	0.5031
Rmse	0.01570	0.1729	1.38414	0.39736	0.2531
Mape	0.67831	1.0068	0.92783	0.82282	0.9414
R	0.99033	0.8982	0.85315	0.93002	0.8776
R <sup>2</sup>	0.9807	0.806	0.72786	0.86493	0.7356



**Figure 5.** Results of the PCA-ANFIS Model with Data Sets.

### 3.2. Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is especially used in machine learning problems such as feature pattern recognition [26], face recognition [27], feature extraction [28], and data size reduction. The goal in this method is to reflect the properties of a dataset to a lower dimensional space that can represent the properties of a good class to avoid over-incompatibility and reduce computational costs. It is a classification method developed by RA Fisher [29] in 1936 and is simple, mathematically consistent and often more complex produce results that are as accurate as the methods. The LDA is based on the concept of a linear combination search (predictors) that best separates two classes (objectives). In a high dimensional space,  $R^n$  is a method of determining the class of an  $x$  value of  $x \in R^n$ . The method obtains a separation vector by maximizing the ratio of the interclass distribution measure to the in-class distribution measure, and can be formulated as:

$$\max_w \frac{w^T S_b w}{w^T S_w w} \quad 8$$

Here  $S_b$  and  $S_w$  distribution matrixes, respectively  $w \in R^m$ , and

$$S_b = \frac{1}{n} \sum_{i=1}^c n_i (m_i - m)(m_i - m)^T \quad 9$$

$$S_w = \frac{1}{n} \sum_{i=1}^c \sum_{j=1}^{n_j} (x_{ij} - m_i)(x_{ij} - m_i)^T \quad 10$$

Here  $m_i$  is the average (or centre) of class  $i$  and  $m$  is the total average (or centre) of all classes. It can be estimated in general using the  $m_i$  and  $m$  training set,

for instance  $m_i = \frac{1}{n} \sum_{j=1}^{n_j} x_{ij}$  and

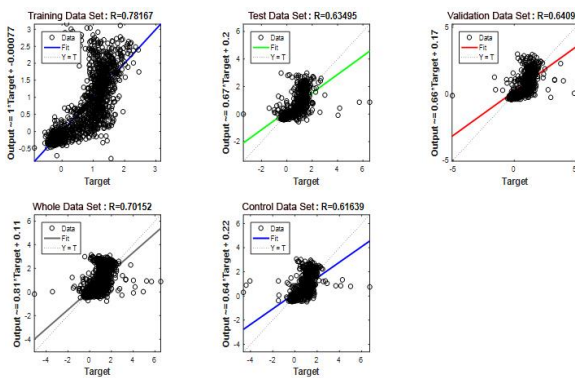
$m = \frac{1}{n} \sum_{i=1}^{c-1} \sum_{j=i+1}^c x_{ij}$ . In addition,  $S_t = S_b + S_w$  and  $S_t$  is the sum distribution matrix. Here the criterion can also be formulated as a traceability problem:

$$\max_w \frac{tr(W^T S_b W)}{tr(W^T S_w W)} \quad (11)$$

A series of separator vectors (Equation 8) can be obtained by the value decomposition of  $S_w - 1S_b$  according to the class order distribution matrix in exact order. The projection matrix can be constructed with vectors of the largest values; this is the best optimal solution of Equation 9. The data set was reduced using the LDA method as described in the previous section and used to establish, train and validate the ANFIS model. As a result of the studies with LDA, the results shown in Figure 6 and Table 4 were obtained.

**Table 4.** Statistical Values Produced by ANFIS Models Established with LDA.

Model Reduced by ANFIS -LDA					
	Training	Test	Validation	All	Control
Mse	0.56335	0.73083	0.713947	0.64767	0.94511
Rmse	0.31737	0.53411	0.509721	0.41947	0.89323
Mape	1.23844	1.71614	2.367759	1.64020	1.36759
R	0.78167	0.63495	0.640091	0.70152	0.61639
R <sup>2</sup>	0.61100	0.40316	0.40971	0.49213	0.37993



**Figure 6.** Results of the LDA-ANFIS Model with Data Sets.

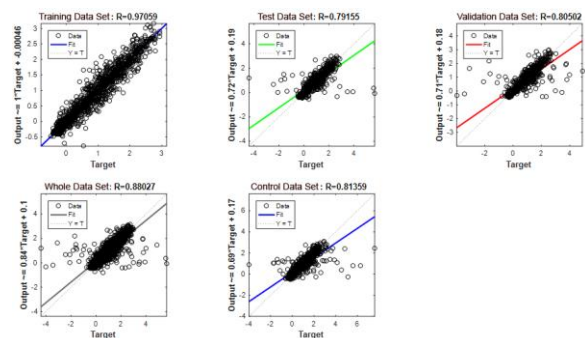
### 3.3. Isomap

PCA has proved successful in many applications, but it has caused problems because it targets pairwise Euclidean distance preservation and does not

add to the distribution of neighbouring data points. In the case of a high-dimensional image where the data is distributed on / off a curvilinear (manifold) structure, the PCA may consider the two data points to be close together, while the points may have much more distances on the distributor. In this context, izomap [30] is a technique that solves this problem by trying to protect double geodesic / curvilinear distances between data points. The geodetic distance is the distance between two points measured on the distributor. In Izomap, the geodetic distances between the data points  $x_i$  ( $i = 1, 2, \dots, n$ ) in an X data set are the nearest neighbours of each data point  $x_{ij}$  ( $j = 1, 2, \dots, k$ ). The shortest path between the two points on the graph is an estimate of the geodetic distance between these two points and can be easily calculated using Dijkstra's or Floyd's shortest path algorithm [31], [32]. A bi-directional geodesic distance (MP) matrix is constructed and the geodetic distances between all data points in X are calculated.  $Y_i$ , which is the low dimensional representation of  $x_i$  data points in the low dimensional Y space, is calculated by applying classical scaling (PCA) to the MP matrix. Izomap has been successfully applied to tasks such as wood examination [33], visualization of biomedical data [34], and head exposure estimation [35].

**Table 5.** Statistical Values Produced by ANFIS Models Established with IZOMAP.

Model Reduced with ANFIS - ISOMAP					
	Training	Test	Validation	All	Control
Mse	0.21745	1.03928	1.05007	0.75332	1.75015
Rmse	0.04728	1.08012	1.10265	0.56749	3.06303
Mape	0.85655	0.92641	1.02677	0.91657	1.81246
R	0.97059	0.79155	0.80502	0.88027	0.81359
R <sup>2</sup>	0.94204	0.62655	0.64805	0.77487	0.66192



**Figure 7.** Results of ISOMAP-ANFIS Model with Data Sets.



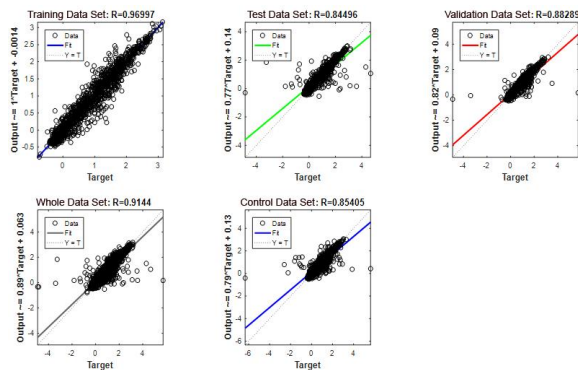
As a result of the studies done with ISOMAP, the results shown in Figure 7 and Table 5 were obtained.

### 3.4. Local Linear Embedding

Local Linear Embedding (LLE) [36] is a technique similar to Izomap in that it constructs graphical representations of data points. Unlike izomap, LLE tries to preserve the local properties of the data. Preservation of local properties allows successful placement of non-convex distributors. In LLE, the local properties of the data distributor are formed by writing the high-dimensional data points as linear components of their nearest neighbours. In the low dimensional representation of the data, LLE attempts to preserve the reconstruction weights as well as possible in linear combinations. LLE suggested the use of linear variants of the algorithm [37]. Super resolution [38] has been successfully applied to the sound source identification [39]. Despite these successes, there are also experimental results reporting poor performance in visualizing even basic synthetic biomedical data sets [34].

**Table 6.** Statistical Values Produced by ANFIS Model Established by LLE.

Model Reduced with ANFIS - LLE					
	Training	Test	Validation	All	Control
Mse	0.21969	0.57182	0.65714	0.4623	0.5403
Rmse	0.04826	0.32698	0.43184	0.2137	0.2919
Mape	0.93614	1.07356	0.77678	0.9306	2.8713
R	0.96997	0.84496	0.88289	0.9144	0.85405
R <sup>2</sup>	0.94084	0.71395	0.77949	0.8361	0.72940



**Figure 8.** Results of LLE-ANFIS Model with Data Sets.

As a result of the studies with LLE, the results shown in Figure 8 and Table 6 were obtained.

### 3.5. Sammon's Mapping (SM)

SM [40], a class of MDS (Multidimensional Scaling), weights and rearranges the contribution of the classical scaling cost function to the cost function of each pair (i, j) by inverting the binary distances of this pair in the high dimensional space  $d_{ij}$ . In this way, the cost function determines approximately equal weights to obtain each of the binary distances, thus obtaining the local structure of the data better than the classical scale. Mathematically, Sammon Mapping is expressed in Equation 12.

$$(Y) = \frac{1}{\sum_{i,j} d_{ij}} \sum_{i \neq j} \frac{(d_{ij} - ||y_i - y_j||)^2}{d_{ij}} \quad 12$$

Here,  $d_{ij}$  denotes the binary Euclidean distance between the high dimensional  $x_i$  and  $x_j$  data points. The front constant is added to reduce the slope of the cost function. The minimization of the Sammon cost function is usually done by the Pseudo-Newton method [38]. The weakness of the SM is that the weight value determined in order to maintain the distance value is larger.

**Table 7.** Statistical Values Produced by ANFIS Model Established by Sammon Mapping.

Model Reduced with ANFIS - Sammon					
	Training	Test	Validation	All	Control
Mse	0.13751	0.48647	1.49566	0.79011	0.76050
Rmse	0.0189	0.23665	2.23701	0.62427	0.57836
Mape	0.77835	0.83794	0.76028	0.78873	0.95873
R	0.98834	0.86567	0.82977	0.91507	0.88145
R <sup>2</sup>	0.97681	0.74938	0.68851	0.83735	0.77695

As a result of the work done by Sammon Mapping method, the results shown in Figure 9 and Table 7 were obtained.

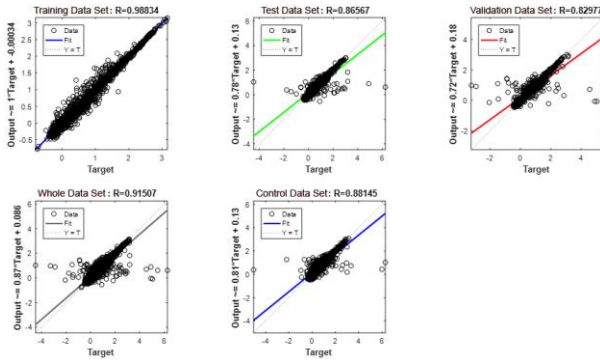


Figure 9. Results of Sammon-ANFIS Model with Data Sets.

3.6. AutoEncoder

AutoEncoder (AE) is a non-consultant machine learning method that is based on artificial neural networks [41]. Although studies with artificial neural network models yielded very good results in non-complex data sets, it is observed that the same successful results cannot be produced for more complex structured data sets. The AE model has used a similar network structure to improve the success rate in more complex data sets and to eliminate this important deficiency of artificial neural networks [42]. In the architecture of the AE model, the number of neurons in the input layer is generally greater than the number of neurons in the hidden layer, and is equal to the number of neurons in the input layer and the number of neurons in the output layer, since the input data set and the output data set are identical [43]. AE is to transform an n-dimensional feature vector into a smaller-sized vector with the lowest loss. An AE consists of an encoder and a decoder section. Encoder: first reads all the properties in the input layer. This information is transferred to the next layer (hidden) as shown in Equation 13. Equation 13 represents the value of the jth neuron in the input layer,  $y_i$  denotes the number of neurons in the hidden layer,  $n$  denotes the number of neurons in the input layer,  $n$  denotes the weight of the  $i$  th neuron in the hidden layer of the  $j$ th neuron in the  $w_{ij}$  input layer,  $f$  denotes the activation function [43].

$$y_i = f(\sum x_j * w_{ij}; n_j = 1) \tag{13}$$

The values obtained are transferred to the output layer as shown in Equation 14, and the calculation of the final values is ensured. Equation 14 shows

the  $j$ th neuron in the output layer, the  $i$ th neuron in the hidden layer, the  $j$ th neuron in the output layer from the  $i$ th neuron in the hidden layer,  $m$  the number of neurons in the hidden layer, and  $f$  the activation function.

$$x_j' = f(\sum y_i * w_{ij}; m_i = 1) \tag{14}$$

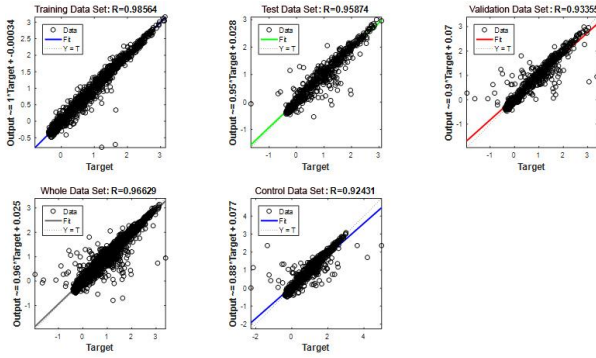
The main goal in the AE model is that the value of aspects derived from the continuation of the previous two steps is similar to the value of those in the input layer. In order to approximate these values, weight values are updated using the backpropagation algorithm. The backpropagation algorithm minimizes the difference between the two values as given in Equation 15.

$$\min \sum (x_j' - x) \tag{15}$$

They gradually perform size reduction operations on the AE data set that is connected to the next one. In this structure, the output of each AE is taken as the input of the next AE. The output of the last AE is the input data of the artificial neural network that will carry out supervised learning. To this network in the last layer, AEs only perform size reduction and no supervised learning takes place. The classification process is performed by this artificial neural network, which is placed at the end of the architecture [42, 43].

Table 8. Statistical Values Produced by ANFIS Models Established with AE.

	Model Reduced with ANFIS – AE				
	Training	Test	Validation	All	Control
Mse	0.15254	0.25188	0.31774	0.2296	0.38945
Rmse	0.0232	0.06344	0.10096	0.0527	0.15167
Mape	0.44291	0.66103	0.575824	0.5306	0.54707
R	0.98564	0.95874	0.93355	0.96629	0.92431
R <sup>2</sup>	0.97148	0.91918	0.87151	0.93371	0.85434



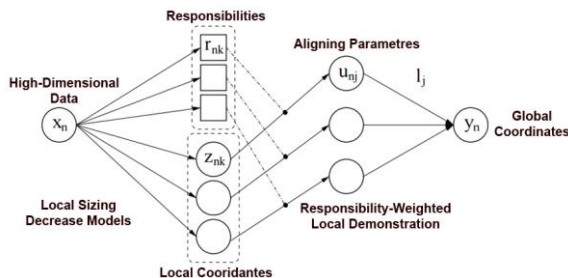
**Figure 10.** Results of AutoEncoder-ANFIS Model with Data Sets.

As a result of the studies performed with AE, the results shown in Figure 10 and Table 8 were obtained.

**3.7. LLC (Locally Linear Coordination)**

The LLC calculates some local linear models and then performs a global alignment of the linear models. This process consists of two steps: Step 1: calculating a combination of local linear models on the data by the average of the EM algorithm; Step 2: aligning the local linear models to obtain low-dimensional data representation using a fraction of the LLE. The LLC transaction is shown in the figure.

Two main ideas have encouraged this method. The first is to use a convex cost function with a unique minimum at the desired global coordinate. The other is to restrict the global coordinates ( $y_n$ ) to the dataset ( $x_n$ ), solely based on the representations of the nodes and responsibilities of  $z_{nk}$  and  $r_{nk}$ , so that the power of the structure of the mixture model is reduced and balanced to mitigate the size of the optimization problem.



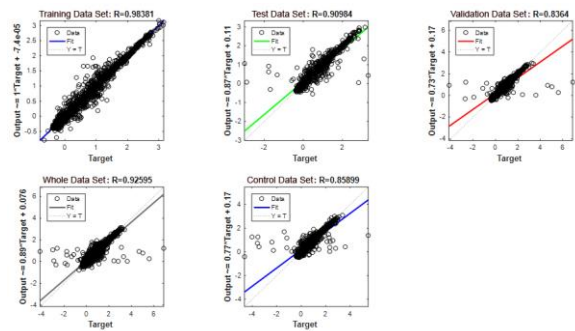
**Figure 11.** Responsibilities-deriving global coordinates through weighted local coordinates.

The basic weakness of LLC is that the fitting of the combination of the factor analysers is sensitive to the presence of local maximums in the log probability function (Teh, Yee Whye, 2003).

As a result of the studies carried out with the LLC, the results shown in Figure 12 and Table 9 were obtained.

**Table 9.** Statistical Values Produced by ANFIS Models Established by LLC.

	Model Reduced with ANFIS – LLC				
	Training	Test	Validation	All	Control
Mse	0.16188	0.49158	0.53563	0.38081	0.58901
Rmse	0.02620	0.24165	0.28690	0.14502	0.34694
Mape	0.57388	0.90867	0.53402	0.64761	2.28090
R	0.98381	0.90984	0.83644	0.92595	0.85899
R <sup>2</sup>	0.96788	0.82780	0.69956	0.85738	0.73786



**Figure 12.** Results of the LLC-ANFIS Model with Data Sets.

**4. CONCLUSION**

Training of Fuzzy Neural Networks model; membership function, numbers of membership functions belonging to the input units and parameters of the learning algorithm of the model are determined by the experts and the most appropriate ones of the fuzzy logic rules of the model are found by the iterative method. Table 10 summarizes the R2 and MSE (Mean Square Error) values produced by the Fuzzy Neural Network model for "Training, All and Control Data Sets" by different dimension reduction methods used in this study. The comparative graphs of these values are as shown in Figure 13 and Figure 14. All values in

Tables 3-9 are the statistical values of the outputs produced by the data sets used during the ANFIS training and control phase. Similarly, the graphs in Figures 5-12 show the relationship between the results produced by the data sets used for ANFIS training and control and the actual results for each data set separately.

The results obtained from the control dataset determine which method makes a more favourable reduction for the problem under study. When an examination is made on the graphs in Figure 13 and Figure 14, it is observed that "AE" produces better values for the Control Set than all other methods. The statistical values in Table 10 show this more clearly. Relations and distributions between these results, which are produced by ANFIS and the statistical values of the results produced by the ANFIS model trained by the reduced data set given in [Table 3- Table 9], and the graphics given in the range [Figure 5 - Figure 12] are shown.

When these tables and graphics are examined, it is concluded that the data set reduced by using AE produces the best acceptable statistical values for almost all reserved sets, in particular the lowest MSE value for the "Control Data Set" Table 8, Table 10 and Figure 13 and the highest R2 Table 8, Table 10 and Figure 14.

Likewise, when the distribution graphs are examined, it can be seen that not only the size of R2 is the largest, but also the distribution that shows the relation between the values produced by ANFIS and the real values are of importance. This acceptance can be seen from the values produced by the FIT line equations expressing the relation between the line  $Y = T$  and the line FIT in the graphs (Figure 15). The closer this value is to 1, the stronger the relationship is. These values produced by AE can be easily seen from the graphs produced by most acceptable methods in almost all methods.

The predicted air quality estimate for this study is nonlinear. As can be seen from the results obtained the performance of the ANFIS model on the non-linear model can be produced much more effectively and represents the nature of the actual data set on the data set obtained by the Auto

Encoder method, which has been accepted in many studies. In fact, size reduction methods are preferred in applications where the number of input parameters in the data set is too large. This study shows that size reduction methods can be successfully applied for training of models falling into the unstable state by increasing the number of input parameters such as ANFIS. The input set consisting of 12 parameters was used in the training of the ANFIS model by reducing 4 parameters and the control results derived from the generated ANFIS model shows that the success of the model is acceptable.

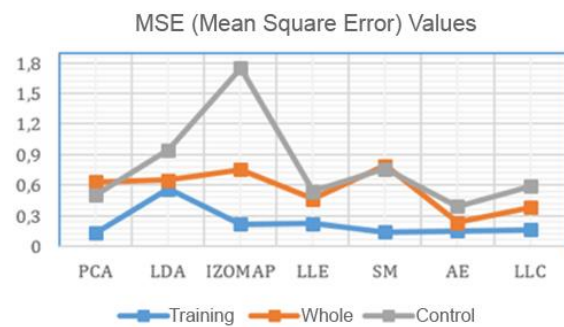


Figure 13. MSE Comparison Graph of ANFIS Results According to Different Reduction Methods.



Figure 14. R² Comparison Graph of Results of ANFIS Results According to Different Reduction Methods.

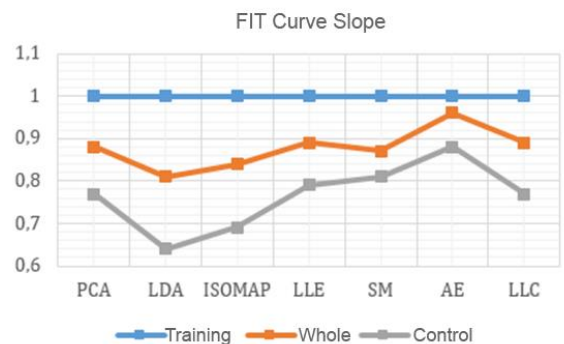


Figure 15. FIT Curve Comparison Scheme of ANFIS Results According to Different Reduction Methods.

**Table 10.** MSE, R<sup>2</sup> and FIT Values of ANFIS Model Results According to Different Reduction Methods.

	PCA		
	Training	All	Control
Mse	0.12532	0.63036	0.5031
R <sup>2</sup>	0.9807	0.86493	0.7356
FIT	1	0.88	0.77
	LDA		
Mse	0.56335	0.64767	0.94511
R <sup>2</sup>	0.61100	0.49213	0.37993
FIT	1	0.81	0.64
	ISOMAP		
Mse	0.21745	0.75332	1.75015
R <sup>2</sup>	0.94204	0.77487	0.66192
FIT	1	0.84	0.69
	LLE		
Mse	0.21969	0.4623	0.5403
R <sup>2</sup>	0.94084	0.8361	0.72940
FIT	1	0.89	0.79
	Sammon Mapping		
Mse	0.13751	0.79011	0.76050
R <sup>2</sup>	0.97681	0.83735	0.77695
FIT	1	0.87	0.81
	AE		
Mse	0.15254	0.2296	0.38945
R <sup>2</sup>	0.97148	0.93371	0.85434
FIT	1	0.96	0.88
	LLC		
Mse	0.16188	0.38081	0.58901
R <sup>2</sup>	0.96788	0.85738	0.73786
FIT	1	0.89	0.77

## REFERENCES

- [1]. Houari R., Bounceur A., Kechadi M-T., A-Tari K., ve Euler R., "Dimensionality reduction in data mining: A Copula approach", *Expert Syst. Appl.*, v. 64, pp. 247–260, 2016.
- [2]. Burges C. J. C., "Dimension Reduction: A Guided Tour", *Found. Trends® Mach. Learn.*, v. 2, 4, pp. 275–365, Ağu. 2010.
- [3]. Tzeng D.-Y. ve Berns R. S., "A review of principal component analysis and its applications to color technology", *Color Res. Appl.*, v. 30, 2, pp. 84-98, Nis. 2005.
- [4]. Sorzano C. O. S., Vargas J., ve Montano A. P., "A survey of dimensionality reduction techniques", *ArXiv14032877 Cs Q-Bio Stat*, 2014.
- [5]. Van Der Maaten L., Postma E., ve Van den Herik J., "Dimensionality reduction: a comparative", *J Mach Learn Res*, v. 10, pp. 66-71, 2009.
- [6]. Ali Godshi, "Dimensionality Reduction A Short Tutorial", 2006. [Online]. Available at: [http://www.stat.washington.edu/courses/stat539/spring14/Resources/tutorial\\_nonlin-dim-red.pdf](http://www.stat.washington.edu/courses/stat539/spring14/Resources/tutorial_nonlin-dim-red.pdf). [Retrieved: 12-Feb-2017].
- [7]. Jang J. S. R., "ANFIS: adaptive-network-based fuzzy inference system", *IEEE Trans. Syst. Man Cybern.*, v. 23, 3, pp. 665-685, 1993.
- [8]. Cybenko G., "Approximation by superpositions of a sigmoidal function", *Math. Control Signals Syst.*, v. 2, 4, pp. 303-314, 1989.
- [9]. Hornik K., Stinchcombe M., ve White H., "Multilayer feedforward networks are universal approximators", *Neural Netw.*, v. 2, 5, pp. 359-366, 1989.
- [10]. Hornik K., "Approximation capabilities of multilayer feedforward networks", *Neural Netw.*, v. 4, 2, pp. 251–257, 1991.
- [11]. Jang J. S. R., Sun C. T., ve Mizutani E., "Neuro-fuzzy and soft computing, a computational approach to learning and machine intelligence", *MyScienceWork*, 1997.
- [12]. Wark K., ve Warner C. F., "Air pollution: its origin and control", 1981.
- [13]. Prasad K., Gorai A. K., ve Goyal P., "Development of ANFIS models for air quality forecasting and input optimization for reducing the computational cost and time", *Atmos. Environ.*, v. 128, pp. 246–262, 2016.
- [14]. Khoshnevisan B., Rafiee S., Omid M., ve Mousazadeh H., "Development of an intelligent system based on ANFIS for predicting wheat grain yield on the basis of energy inputs", *Inf. Procepp. Agric.*, v. 1, sayı 1, pp. 14-22, 2014.
- [15]. Yuksek A. G., Karaboga D., "Hava kirliliği tahmininde çoklu regresyon analizi ve yapay sinir ağları yönteminin karşılaştırılması - Sayfa 264 - Ulusal Tez ve Araştırma Merkezi - Akademik Tezler ve Araştırmalar". [Online]. Available at: <http://ulusaltezmerkezi.com/hava-kirliligi-tahmininde-coklu-regresyon-analizi-ve-yapay-sinir-aglari-yonteminin-karsilastirilmasi/264/>. [Retrieved: 04-Apr-2017].
- [16]. Tsoukalas L. H., ve Uhrig R. E., *Fuzzy and Neural Approaches in Engineering*, 1st baskı. New York, NY, USA: John Wiley & Sons, Inc., 1996.
- [17]. Hu X., "DB-HReduction: A data preprocessing algorithm for data mining applications", *Appl. Math. Lett.*, v. 16, sayı 6, pp. 889-895, 2003.
- [18]. Sattler K. U., ve Schallehn E., "A data preparation framework based on a

- multidatabase language”, içinde *Proceedings 2001 International Database Engineering and Applications Symposium*, 2001, pp. 219-228.
- [19]. Yuksek A. G., Tüzemen E. Ş., Arslan H., ve Elagöz S., “Anfis Kullanılarak ZnO İnce Filmin Yansıtma Özelliklerinin Modellenmesi Ve Regresyon Analizi İle Sonuçlarının Karşılaştırılması”, 2. *Ulusal Yönetim Bilişim Sistemleri Kongresi*. [Online]. Available At: <http://Docplayer.Biz.Tr/755252-2-Ulusal-Yonetim-bilisim-sistemleri-kongresi.html>. [Retrieved: 20-Feb-2017].
- [20]. “Çimento Hammadde ve Üretim Sürecinde Esnek Hesaplama - Documents”, *Documents.tips*. [Online]. Available at: <http://documents.tips/documents/cimento-hammadde-ve-ueretim-suerecinde-esnek-hesaplama.html>. [Retrieved: 06-Apr-2017].
- [21]. Sadık Önal, “Mn/n-GaAs schottky diyotunun hidrostatik basınç altında elektriksel karakterizasyonu”, *Gazi Osman Paşa Üniversitesi Yüksek Lisans Tez*, 2007.
- [22]. Jolliffe I., “Principal Component Analysis”, in *Wiley StatsRef: Statistics Reference Online*, John Wiley & Sons, Ltd, 2014.
- [23]. Yıldız K., Çamurcu Y., ve Doğan B., “Veri Madenciliğinde Temel Bileşenler Analizi ve Negatif Matris Çarpanlarına Ayırma Tekniklerinin Karşılaştırmalı Analizi”, *Akad. Bilişim*, 2010.
- [24]. Maćkiewicz A., ve Ratajczak W., “Principal components analysis (PCA)”, *Comput. Geosci.*, v. 19, pp. 303-342, 1993.
- [25]. Ali Sait Albayrak, *Uygulamalı Çok Değişkenli İstatistik Teknikleri*. Asil Yayın Dağıtım, 2006.
- [26]. McLachlan G., *Discriminant Analysis and Statistical Pattern Recognition*. John Wiley & Sons, 2004.
- [27]. Li Y., Kittler J., ve Matas J., “Effective Implementation of Linear Discriminant Analysis for Face Recognition and Verification”, in *Computer Analysis of Images and Patterns*, 1999, pp. 234-242.
- [28]. Yang J., Ye H., ve Zhang D., “A new LDA-KL combined method for feature extraction and its generalisation”, *Pattern Anal. Appl.*, v. 7,1, pp. 40-50, 2004.
- [29]. Fisher R. A., “The Use of Multiple Measurements in Taxonomic Problems”, *Ann. Eugen.*, v. 7, 2, pp. 179-188, 1936.
- [30]. Tenenbaum J. B., de Silva V., ve Langford J. C., “A Global Geometric Framework for Nonlinear Dimensionality Reduction”, *Science*, v. 290, 5500, pp. 2319–2323, 2000.
- [31]. Dijkstra E. W., “A note on two problems in connexion with graphs”, *Numer. Math.*, v. 1, 1, pp. 269-271, 1959.
- [32]. Floyd R. W., “Algorithm 97: Shortest Path”, *Commun ACM*, v. 5, 6, s. 345-, 1962.
- [33]. Niskanen M., ve Silven O., “Comparison of dimensionality reduction methods for wood surface inspection”, 2003, v. 5132, pp. 178-188.
- [34]. Lim I. S., Ciechomski P. de H., S. Sarni S., ve D. Thalmann, “Planar arrangement of high-dimensional biomedical data sets by isomap coordinates”, içinde *16th IEEE Symposium Computer-Based Medical Systems, 2003. Proceedings.*, 2003, pp. 50-55.
- [35]. Raytchev B., Yoda I., ve Sakaue K., “Head pose estimation by nonlinear manifold learning”, içinde *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, 2004, v. 4, s. 462-466 Vol.4.
- [36]. Roweis S. T. ve Saul L. K., “Nonlinear Dimensionality Reduction by Locally Linear Embedding”, *Science*, v. 290, 5500, pp. 2323-2326, 2000.
- [37]. Kokiopoulou E. ve Saad Y., “Orthogonal Neighborhood Preserving Projections: A Projection-Based Dimensionality Reduction Technique”, *IEEE Trans. Pattern Anal. Mach. Intell.*, v. 29, 12, pp. 2143–2156, 2007.
- [38]. Chang H., Yeung D.-Y., ve Xiong Y., “Super-resolution through neighbor embedding”, içinde *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, 2004, v. 1, pp. I-I.
- [39]. Duraiswami R., ve Raykar V. C., “The manifolds of spatial hearing”, içinde *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, 2005, v. 3, s. iii/285-iii/288 Vol. 3.
- [40]. Sammon J. W., “A Nonlinear Mapping for Data Structure Analysis”, *IEEE Trans. Comput.*, v. C-18, 5, pp. 401-409, 1969.
- [41]. Badino L., Canevari C., Fadiga L., ve Metta G., “An auto-encoder based approach to unsupervised learning of subword units”, içinde *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 7634-7638.

- [42]. Kaynar O., Görmez Y., Ve Işık Y. E., “Oto Kodlayıcı Tabanlı Derin Öğrenme Makinaları İle Spam Tespiti” 3. International Management Information Systems
- [43]. A. Ng, “Sparse autoencoder”, *CS294A Lect. Notes*, v. 72, 2011, pp. 1-19, 2011.