Contents lists available at *Dergipark*

Journal of Scientific Reports-A

journal homepage: *https://dergipark.org.tr/tr/pub/jsr-a*

E-ISSN: 2687-6167 ***Number 57, June 2024***

RESEARCH ARTICLE

Receive Date: 24.11.2023 *Accepted Date: 02.04.2024*

Examination of the effect of ANN and NLPCA technique on prediction performance in patients with breast tumors

Canan DEMİR[a]*

[a]*Van Yuzuncu Yil University, Vocational School of Health Services, 65100,Van,Türkiye, ORCID: 0000-0002-4204-9756*

Abstract

Breast cancer is among the most prevalent cancer kinds worldwide. The aim of this study is to examine the effect of combining Artificial Neural Networks and Nonlinear Principal Component Analysis techniques on prediction performance in patients with breast tumors. In the application, a network containing 5 layers, including the input, the coding, the bottleneck, the decoding and the output, was used for the 30 variable data set of 569 breast tumor patients. The training algorithm of choice was the Conjugate Gradient Descent (CGD) algorithm. In this study, artificial neural networks (ANN) and nonlinear principal component analysis were coupled. NLPCA was first applied to dimension reduction in artificial neural networks. Using both the original data set and the decreased size, artificial neural networks were used in the second stage to develop prediction models. By contrasting the performance of these two prediction models with one another, the outcomes were understood. 96.37% of the variation was explained by the two fundamental components that were found using NLPCA. The prediction models developed for the original data set and the dimension-reduced data set have R^2 values of 91% and 87%, respectively. The advantages of the NLPCA and ANN combination for breast tumor patients are demonstrated by this study. It is believed that utilizing principal components as inputs can cut down on complexity and extraneous information.

Keywords: Artificial neural networks, Breast tumor, Dimensional reduction, Nonlinear principal component analysis, Variance

1. Introduction

Breast cancer is among the most prevalent cancer kinds worldwide [1], [2]. Breast cancer, which is encountered especially between the ages of 30-59 and is one of the most important causes of death in women, has become a

* Corresponding author
E-mail address: canandemir@yyu.edu.tr

universal problem despite advances in health, advances in diagnosis and treatment methods, and a conscious society [3], [4]. Compared to developed nations, the incidence of breast cancer is rising more quickly in developing low- and middle-income countries, and the death rate is higher [4]. The reasons for this include the adoption of the western lifestyle, fertility, changes in lactation characteristics, exposure to exogenous factors such as malnutrition, fatty diet, alcohol consumption and hormone therapy, lack of health insurance for all individuals, and economic inadequacies [4], [5]. Early detection of breast cancer is important for increasing overall survival and decreasing mortality. On the other hand, whether the cyst is malignant or benign is determined by various biopsy methods. This prevents the cancer from spreading and reduces the ratio of women who need complementary radiotherapy [6].

Although Principal Component Analysis (PCA) emerged in the 1900s, it is still preferred and used by researchers today. Principal component analysis is a linear method and requires assumptions such as that the variables related to the data set are continuous and the relationships between variables are linear. When these assumptions are not fulfilled, (NLPCA) Nonlinear Principal Component Analysis is used. This descriptive dimension reduction technique, which has a linear or nonlinear relationship between them, yields numerical and graphical results for datasets with continuous, categorical, or discrete variables [7], [8].

Artificial neural networks are a type of computational system designed to mimic the functions of the human brain, like the capacity to learn in order to obtain, produce, and find new knowledge [9]. Machine learning is accomplished by ANNs through the use of examples; their algorithms and modes of operation differ from well-known programming techniques. Before the ANN can function successfully, it needs to be trained and its performance evaluated. Artificial Neural Networks are practical instruments utilized in data-driven process modeling, nonlinear process control, pattern recognition, classification, and prediction [10].

The dimensions of the data indicate the amount characteristics that are measured for every observation. Analyzing high-dimensional data is difficult. Dimension reducing strategies are offered to get rid of unnecessary and unnecessary features. Choosing an appropriate dimensionality reduction process can help increase speed up processing and cut down on the time and work needed to retrieve important data. The purpose of this research is to examine the effect of combining ANN and NLPCA techniques on prediction performance on reduced and non-reduced data sets.

2. Material and methods

In the study, a dataset consisting of 30 variables belonging to 569 breast tumour patients (357 benign, 212 malignant) provided from the free data access site was used as the application material. On July 15, 1992, Olvi Mangasarian provided the dataset, which was generated by Dr. William H. Wolberg, a doctor at the University of Wisconsin Hospital in Madison, Wisconsin, USA [11]. The variables and characteristics of 569 patients with breast tumors are given in Table 1. Matlab (MathWorks. Matlab version R2013b. Natick: 2013) and SPSS (Version 16.0 of IBM SPSS. Spss for Windows. IBM, Chicago, 2010) package program were used for data analysis.

Table 1. Study variables and their characteristics.

No.	Variable name
1.	mean radius
2.	radius Standart Deviation
3.	worst Radius
4.	mean texture
5.	texture SD

6. worst texture
7. mean perimeter
8. perimeter SD
9. worst perimeter
10. mean area
11. area SD
12. worst area
13. mean smoothness
14. smoothness SD
15. worst smoothness
16. mean compactness
17. compactness SD
18. worst compactness
19. mean concavity
20. concavity SD
21. worst concavity
22. mean concave points
23. concave points SD
24. worst concave points
25. mean symmetry
26. symmetry SD
27. worst symmetry
28. mean fractal dimension
29. fractal dimension SD
30. worst fractal dimension

2.1 Theory of nonlinear principal component analysis in neural networks

NLPCA has been designed to identify nonlinear interactions in cases when the relationship between the variables is not linear [12]. The NLPCA is implemented using the neural network (or "neuron") shown in Figure 1.

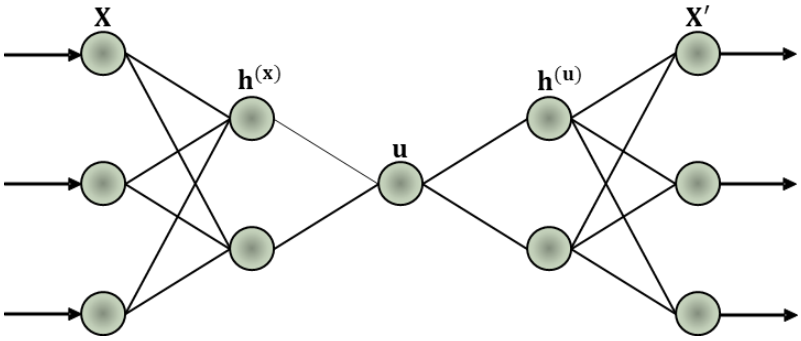

Fig. 1. Nonlinear principal components analysis using a neural network model.

The transfer function is mapped to the input column vector x of length l, or f_1, by the first hidden layer, denoted by $h^{(x)}$, which is the column vector of length m. Consequently, the expression for $h_k^{(x)}$ is as follows:

$$h_k^{(x)} = f_1\left[\left(W^{(x)}x + b^{(x)}\right)_k\right] \quad (1)$$

A weight matrix measuring "m x 1" is represented by $w^{(x)}$ in equation (1), where $b^{(x)}$ is the column vector in m dimensions containing the bias parameters (k = 1,..., m). Similar to the first transfer function, f_2, this one maps the nonlinear basic component (u) from the encoder layer to the hidden layer. It is expressed as follows for this nonlinear basic component:

$$u = f_2\left(w^{(x)}h^{(x)} + \overline{b}^{(x)}\right) \quad (2)$$

The transformation function f_1 is nonlinear, whereas f_2 is typically assumed to be a linear function (its exact function is unknown, but it can typically be a sigmoid function or hyperbolic tangent). The final hidden layer, $h^{(u)}$, is mapped from u by the transformation function f_3. Hence, $h_k^{(u)}$ is expressed as follows (k = 1,.., m);

$$h_k^{(u)} = f_3\left[\left(w^{(u)}u + b^{(u)}\right)_k\right] \quad (3)$$

Equation comes after this function (4).

$$x_i' = f_4\left[\left(W^{(u)}h^{(u)} + \bar{b}^{(u)}\right)_i\right] \quad (4)$$

The error function j = <‖x-x^'‖2> is used here. By identifying the ideal values, W(x), b(x), w(x), ¯b (x), w(u), b(u), W(u) and ¯b (u) are minimized [13]. This reduces the mean squared error between the neural network output x' and the original data x.

NLPCA uses the following total parameters: (m+f+1)(M1+M2)+m+f. The number of neurons hidden in the decoding and coding is generally regulated by a rule, which also determines which M. Although a bigger M improves the network's capacity for modeling that is nonlinear, it may additionally cause excessively fitting. The accuracy could be low for a lower M because of the network's constrained representation capacity. The complexity of nonlinear functions that the network is capable of producing is indicated by the value of M. Only when f_4 is a linear function and M = 1 can there be a linear relationship between the x_i' variables. All of x_i' is linearly related to a single hidden neuron, as shown by Equation (4). Nonlinear solutions have a M≥2 [14].

Three hidden layers are shown in Figure 1, positioned between input layer x on the left and output layer x' on the right (shown by a circle). The coding layer sits next to the input layer, and is followed by the analysis and "bottleneck" layers (which contain a single neuron, u). The neuron in the bottleneck, u, provides compression or dimension reduction of the data; this is known as the nonlinear principal component [15].

2.2 Backpropagation

A typical backpropagation network consists of an input layer, an output layer, and one or more hidden layers. Even so, the quantity of hidden layers is theoretically unlimited, one or two are typically used. Every layer is completely linked to the layers below it [16].

Supervised learning is the underlying principle of multi-layered networks. During training, the network must be shown both input and output values. For every input, the network's job is to generate the corresponding output. The multilayer network's learning rule is a generalized Delta learning rule that uses the least squares approach as its foundation. The rule of learning is also known as the "Generalized Delta Rule" because of this. There are 2 phases to this rule. The network output is calculated in the first phase, known as forward calculation, and the weights are adjusted in the second phase, known as reverse calculation [17], [18].

Calculation in advance: This stage is where the network output is calculated. The training set's data is now displayed on the network's input layer, marking the start of information processing. As previously stated, the input layer does not process any information. Without any modifications, the newly received entries are routed up to the intermediate stratum. This is what the formula $y_i = x_i$ indicates.

Through the use of connecting weights, every unit in the intermediate layer is given the data from the input unit. The middle layer units' net input is computed as follows:

$$n^1 = \left(\sum_{i=1}^{j} w_{ir} x_i + b_i\right) \tag{5}$$

In Equality (5)

w_{ir}: i. the input layer element,
r: the connection's weight value to the middle layer unit,
x_i: The input layer's output from the i. processor unit,
n_1: It displays the middle layer unit's output value,
The following is an expression of the output layer's unit's net input;

$$n^2 = \sum_{k=1}^{s} w_k^2 a_k^1 + b^2 \tag{6}$$

Network output;

$$\hat{y} = g\left\{\sum_{k=1}^{s} w_k f\left(\sum_{i=1}^{j} w_i x_i + b^1\right) + b^2\right\} \tag{7}$$

and the following equation is used to calculate the total error [19];

$$E_D = \frac{1}{2}\sum_{i=1}^{j}(y - \hat{y})^2 \tag{8}$$

The transfer function can be either the hyperbolic tangent function or the sigmoid function, though this is not necessary. It should be mentioned the chosen function can be derived because back propagation requires consideration of the function's derivative [20].

Backward calculation: The network is run entirely backwards during the back propagation phase. The derivative of the transfer function typically modifies the discrepancy between the final layer's output and the intended output, which then propagates back to the previous layer (s). As a result, the error is decreased in the following cycle. Up until the input layer is reached, this process continues for the preceding layer or layers.

2.3 Artificial neural networks (ANN)

Artificial neural networks, which are modeled after the architecture of the brain's nerve cells in humans, are composed of interconnected processing elements of varying weights [21]. The input layer, variable weight factors, summation function, activation function, and output layer are the five primary components that make up an artificial neural network cell in this system (Figure 2).

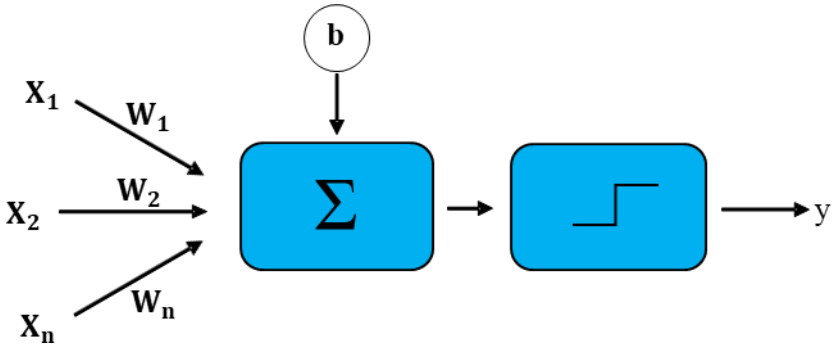

Fig. 2. Artificial neuron.

The data that the artificial nerve cell receives from the outside world is known as an input. The values that establish the significance of information reaching an artificial neuron and its impact on the cell are called weights. When the calculation starts, the weights are given randomly and the weights continue to be changed until the error is minimum according to the process. Learning in artificial neural networks occurs by changing these weights. According to the addition function, each weight in the artificial neuron is multiplied by the input it belongs to. After multiplying these inputs by the weights affecting the inputs, the results are added arithmetically. The result of the addition operation is summed with the threshold value and sent to the activity function. In the calculation, "w_1, w_2,…,w_n" gives the weights and "x_1, x_2,…,x_n" gives the values of the input parameters.

$$w_1x_1 + w_2x_2 + \cdots + w_nx_n = \sum_{i=1}^{n} w_ix_i \quad (9)$$

An output value is created by processing the weighted input value with the help of the activation function [8].

2.3.1 Levenberg-marquardt (LM) algorithm

The Hessian matrix does not need to be computed in order to approximate the quadratic training rate using the LM algorithm. The Hessian matrix can be predicted as follows when fed network training frequently uses a sum of squares as the performance function:

$$H = J^T J \tag{10}$$

and can be calculated as gradient.

$$g = J^T e \tag{11}$$

With respect to the weights and deviations, the initial derivatives of the network errors are represented by the Jacobian matrix J, and the network error vector e. It is much simpler to calculate the Jacobian matrix than the Hess matrix, and can be done so with a standard backpropagation technique. The following Newton-like modify is produced by the Levenberg-Marquardt algorithm using this approximation of the Hessian matrix:

$$x_{k+1} = x_k - [J^T J + \mu I]^{-1} J^T e \tag{12}$$

When the scalar μ in equation (11), the Hessian matrix and Newton's method are used. When μ exceeds zero, the function transforms into a gradient with a small step size. The goal is to switch to Newton's method as soon as possible because it is faster, more accurate, and has almost no error. As a result, μ is only raised in the unlikely event that a temporary step would result in an increase in performance function; otherwise, it is lowered after each successful step (decrease in performance function). The outcome is that the performance function continuously decreases with each algorithm iteration [22].

2.3.2 Hyperbolic tangent function

With a marginally larger output space, similar to the sigmoid activation function is the hyperbolic tangent activation function. Although the results of the sigmoid activation function fall within the range of [0, 1], the tangent hyperbolic activation function gives results in the range of [-1, 1] [8].

$$y = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{13}$$

In this study, R^2 and MSE were used as accuracy measures. One of these performance measures, R^2 is the accuracy rate decision coefficient of the model. The high value of this coefficient indicates that the predictive relationship is good. MSE, on the other hand, is a measure of high performance because it is an error measure. The error is expressed by the formula et, where t is the time series observed in the time interval rt to show a certain time and pt is the estimated time series [23].

$$MSE = \frac{1}{n}\sum_{i=1}^{n} e_t^2 \tag{14}$$

3. Results and discussions

NLPCA works by training a feedforward neural network whose inputs are reproduced. Figure 3 depicts the architecture of the neural network that was used to implement NLPCA.

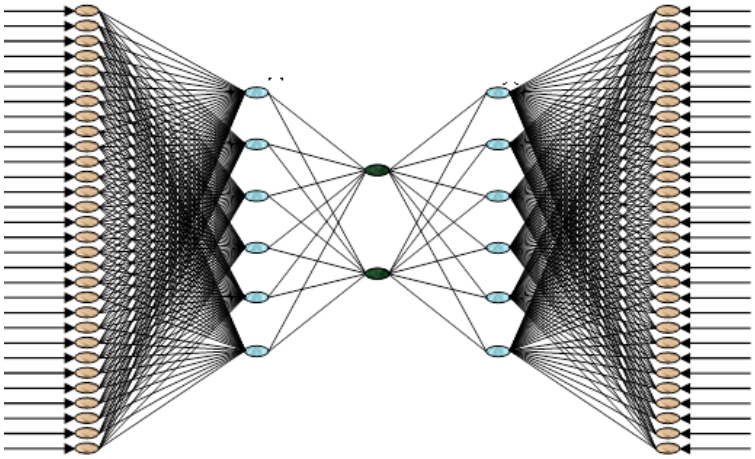

Fig. 3. 30-6-2-6-30 NLPCA model.

In the application, a network containing 5 layers, including the input, the coding, the bottleneck, the decoding and the output, was used for the 30 variable data set of 569 breast tumor patients. The number of iterations in the network architecture, which was selected as the number of neurons 30-6-2-6-30, was determined as 3000. This architecture was decided as a result of trial and error. The coding and decoding layer uses a non-linear hyperbolic tangent function, while the other layers use linear functions. The training algorithm of choice was the Conjugate Gradient Descent (CGD) algorithm.

The original data set consists of a 30x569 matrix. Two main components obtained by NLPCA method in artificial neural networks are a 2x569 dimensional matrix and this matrix explained 96.37% of the total variance (Figure 4).

Fig. 4. Two-dimensional Nonlinear Principal Component.

In order to see the effectiveness of NLPCA, prediction models were created with both the original data set and the obtained principal components. (Table 2).

Table 2. Comparison results for two data sets.

	R^2	**MSE**
Original data set (30x569)	0.91	0.0382
Two principal components (2x569)	0.87	0.0533

The data set was divided into 3 parts as training, verification and testing. Accordingly, the data lines were divided into three as; 70% training, 15% verification and 15% test data. A single hidden layer with 10 neurons in networks and the hyperbolic tangent function in the layers were used. Networks were trained with the Levenberg-Marquardt algorithm. 500 iterations were foreseen for each training. For both data sets, all properties in the networks were kept equal. R^2 shows what percentage of the change in the dependent variable is explained by the independent variables. The high value of the R^2 determination coefficient indicates that our prediction is successful and the prediction is accurate. In feedforward networks, mean square error (MSE) is generally used as a performance function. The MSE function is the average of the sum of squares of the difference between the desired output values and the values produced by the network. The network with the best performance, that is, the network with the smallest test error was evaluated. Accordingly, were found as $R^2 = 0.91$ and MSE = 0.0382 for the original data set, and $R^2 = 0.87$ and MSE = 0.0533 for the dimension reduced data set.

Many studies have been conducted on the use of PCA as a pre-processor to multi-layer perceptron neural networks in various domains. In order to categorize the quality of food products, O'Farrella et al. employed PCA as a pre-treatment, followed by neural networks with a hidden layer and back propagation algorithm [24]. Bucinski et al. integrated ANN and PCA. They used PCA to reveal some important features in patients with uterine cancer and a back propagation ANN model to classify these patients into two categories [6]. Ayesha et al. stated that most dimensionality reduction applications have a need for classification and prediction accuracy. They also stated that as the size of the data increases, dimensionality reduction is necessary to reduce uncertainty in decision-making [25]. Zebari et al. reported that to improve the accuracy of learning features and reduce training time, dimensionality reduction is necessary as a pre-processing step that can remove irrelevant data, noise, and redundant features [26].

4. Conclusion

In this study, NLPCA was used in artificial neural networks. NLPCA reduced the original data set to two dimensions, with a very high variance explanation rate (96.37%). In dimension reduction methods, original variables contribute to the dimensions to a certain extent. The variables that contribute the most to the dimensions are; radius, perimeter and area variables. Here, variance can be thought as the equivalent of the information in the original data set. In other words, a 30x569 dimensional large matrix was transformed into a much smaller 2x569 dimensional matrix and 96.37% of the information contained was preserved.

In order to see the effectiveness of dimension reduction in patients with breast tumor, prediction models were created in artificial neural networks with both the dimension reduced and non-dimensionally-reduced data set. In practice, multilayer feedforward artificial neural network model was used to see the prediction success of artificial neural networks and malignant tumor was predicted. 30 variable inputs given in the data set variable, and the breast tumor (357 benign, 212 malignant) variable was taken as the target variable. Thus, the probability of predicting malignant tumor for both the original data set and the reduced size data set was determined. Accordingly, when the

prediction models are evaluated, the performance of dimension reduced data seems very close to the performance of non-dimensionally-reduced data.

For the purpose of training machine learning algorithms, not every characteristic in the generated datasets is significant. Certain features might not matter, while others might not have an impact on the prediction's result. Machine learning algorithms are not burdened as much when these unimportant or irrelevant features are ignored or eliminated [27], [28]. In conclusion, this research reveals the benefits of the combination of NLPCA and ANN in dimensionality reduction. It is thought that using basic components as input can reduce unnecessary information and complexity.

Acknowledgement

There is no conflict of interest with any person/institution in the article.

References

[1] A. Haydaroğlu, S. Dubova, S. Özsaran, et al., "Breast Cancer in Ege University "Evaluation Of 3897 Cases"", *The Journal of Breast Health*, vol. 1, pp. 6-11, 2005.
[2] C. Hocaoglu, G. Kandemir and F. Civil, "The Influence of Breast Cancer to Family Relationships: a Case Report", *The Journal of Breast Health*, vol. 3, pp. 163-165, 2007.
[3] C. Eroğlu, M. A. Eryılmaz, S. Civcik, et al., "Breast Cancer Risk Assessment: 5000 Cases", *International Journal of Hematology and Oncology*, vol. 20, pp. 27-33, 2010.
[4] N. Akyolcu and G. A. Ugras, "Breast Self-Examination: how Important is it in Early Diagnosis?", *The Journal of Breast Health*, vol. 7, pp. 10-14, 2011.
[5] A. Tumer, and H. Baybek, "The Risk Level of Breast Cancer at The Working Women", *The Journal of Breast Health*, vol. 6, pp. 17-21, 2010, https://hdl.handle.net/20.500.12809/4656.
[6] A. Buciński, T. Bączek, T. Waśniewski and M. Stefanowicz, "Clinical data analysis with the use of artificial neural networks (ANN) and principal component analysis (PCA) of patients with endometrial carcinoma", *Rep. Pract. Oncol. Radiother*, vol. 10, pp. 239-248, 2005, doi:10.1016/S1507-1367(05)71096-8.
[7] S. I. V. Sousa, F. G. Martins, M. C. M. Alvim-Ferraz and M.C. Pereira, "Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations", *Environmental Modelling & Software*, vol. 22, pp. 97-103, 2007, doi:10.1016/j.envsoft.2005.12.002.
[8] C. Demir and S. Keskin, "Artificial neural network approach for nonlinear principal components analysis", *International Journal of Current Research*, vol. 13, no. 1, pp. 15987-15992, 2021, doi:10.24941/ijcr.40671.01.2021.
[9] M. O. Kaya, C. Colak and E. Ozdemir, "The Prediction of Prostate Cancer Using Different Artificial Neural Network Models with The Help of Prostate Specific Antigen", *İnönü University Journal of Health Sciences*, vol. 1, pp. 19-22, 2013.
[10] S. Samarasinghe, "Neural networks for applied sciences and engineering: from fundamentals to complex pattern recognition", *Crc Press*, pp. 245-281. 2006.
[11] W. Wolberg, O. Mangasarian, N. Street and W. Street, "Breast Cancer Wisconsin (Diagnostic)", *UCI Machine Learning Repository*, 1995, doi:10.24432/C5DW2B.
[12] M. Scholz, F. Kaplan, C. L. Guy, J. Kopka and J. Selbig, "Non-linear PCA: a missing data approach", *Bioinformatics*, vol. 21, pp. 3887-3895, 2005, doi:10.1093/bioinformatics/bti634.
[13] W. W. Hsieh, "Nonlinear principal component analysis by neural networks", *Tellus*, vol. 53, pp. 599-615, 2001, doi:10.3402/tellusa.v53i5.12230.
[14] M.A. Kramer, "Nonlinear principal component analysis using auto-associative neural networks", *AIChE Journal*, vol. 37, pp. 233-43, 1991, doi:10.1002/aic.690370209.
[15] W. W. Hsieh, "Nonlinear principal component analysis of noisy data", *Neural Networks*, vol. 20, no. 4, pp. 434-443, 2006. doi:10.1016/j.neunet.2007.04.018.
[16] D. Anderson and G. McNeill, "Artificial Neural Networks Technology", *New York: Rome Laboratory RL/C3C Griffiss AFB.* 83. A011, 1992.
[17] A. Gülbağ, "Quantitative Determination Of Volatile Organic Compounds By Using Artificial Neural Network And Fuzzy Logic Based Algorithm", [dissertation], Sakarya: Sakarya University, 2006.
[18] H. Güler, "Prediction of Elements On Corrosion Behavior Of Zinc-Aluminum Alloys Using Artificial Neural Networks", [dissertation]. Sakarya, Sakarya University, 2007.
[19] H. Okut, "Bayesian Regularized Neural Networks for Small n Big p Data. Rosa JLG", Artificial Neural Networks-Models and Applications. London: InTechOpen, 2016.
[20] V.S.A. Kargı, "Artificial Neural Network Models and an Application at a Textile Firm", [dissertation]. Bursa, Uludağ University, 2013.
[21] B. Taşar, F. Üneş, M. Demirci and Y. Z. Kaya, "Forecasting of Daily Evaporation Amounts Using Artificial Neural Networks Technique",

Dicle University Engineering Faculty Journal of Engineering, vol. 9, no. 1, pp. 543-551, 2018.
[22] Trainlm [Internet]. [access date 29 november 2021]. Access address:
https://uk.mathworks.com/help/deeplearning/ref/trainlm.html
[23] Y. Gültepe, "A Comparative Assessment on Air Pollution Estimation by Machine Learning Algorithms", *European Journal of Science and Technology*, vol. 16, pp. 8-15, 2019, doi:10.31590/ejosat.530347.
[24] M. O'Farrella, E. Lewisa, C. Flanagana, W. B. Lyonsa and N. Jackman, "Combining principal component analysis with an artificial neural network to perform online quality assessment of food as it cooks in a large-scale industrial oven", *Sensors and Actuators B: Chemical*, vol. 107, pp. 104–112, 2005, doi:10.1016/j.snb.2004.09.050
[25] S. Ayesha, M. K. Hanif, and R. Talib, "Overview and comparative study of dimensionality reduction techniques for high dimensional data", *Information Fusion*, vol. 59, pp. 44-58, 2020, https://doi.org/10.1016/j.inffus.2020.01.005.
[26] R. Zebari, A. Abdulazeez, D. Zeebaree, D. Zebari and J. Saeed, "A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction", *Journal of Applied Science and Technology Trends*, vol. 1, no. 1, pp. 56-70, 2020, doi:10.38094/jastt1224.
[27] G. T. Reddy, M. P. K. Reddy, K. Lakshmanna, R. Kaluri, D. S. Rajput, G. Srivastava and T. Baker, "Analysis of dimensionality reduction techniques on big data", *Ieee Access*. 8, 54776-54788. 2020, doi: 10.1109/ACCESS.2020.2980942
[28] I. K. Omurlu, F. Cantas, M. Ture and H. Ozturk, "An empirical study on performances of multilayer perceptron, logistic regression, ANFIS, KNN and bagging CART", *Journal of Statistics and Management Systems*, vol. 23, no. 4, pp. 827-841, 2020, doi:10.1080/09720510.2019.1696924.