ISSN 2712-0562
Original Research

Sustainable Engineering and Innovation
Vol. 6, No. 2, October 2024, pp.225-238
https://doi.org/10.37868/sei.v6i2.id366

Using artificial intelligence in analyzing principal components

Hasanain Jalil Neamah Alsaedi[1,2]
[1] Statistics, University of Information Technology and Communication, Iraq
[2] Statistics, Acharya Nagarjuna University, India

* Corresponding author E-mail: hasanien.1975@uoitc.edu.iq

Received Feb. 22, 2024
Revised Jun. 29, 2024
Accepted Oct. 09, 2024

Abstract

In this research, the basic concepts of both principal component analysis and artificial neural networks were reviewed, let alone tackling breast cancer, its most important causes and early detection. To add more, obtaining the principal components of a set of explanatory variables using the artificial neural network method was also discussed. One of the most important conclusions reached is that the weights of the artificial neural network based on the (Hebb) rule are close to the values of the characteristic vectors for the correlation matrix. The flexibility of the work of artificial neural networks allows for the expansion of the use of neural networks in other statistical methods. The most important factors causing breast cancer are: marital status, age and breastfeeding, and family history of the disease.

Keywords : Artificial Intelligence, Artificial neutral networks, Principal Components , Breast cancer

1. Introduction

The computer has developed to be an essential aspect in almost all life activities, and the development that has occurred at the level of the machine itself and its components has branched out to include entering and processing knowledge, and making appropriate decisions, until it reached the point of imitating humans in some of their professional activities. Rather, the matter reached more than that, as a new specialty was born that transferred the computer from the state of a large calculator to a thinking machine capable of imitating humans. This specialty was called artificial intelligence. The science of (Artificial intelligence), in cooperation with mathematical and statistical sciences, is the latest innovation created by the human mind in the last five decades of the twentieth century of human life, through which scientists are trying to provide machines with the ability to make inferences Where researchers presented in their paper " BCDR : A breast cancer digital repository " in 2012 the first digital repository that includes breast images and related descriptive data such as clinical history, etc., to diagnose the infection through the computer [1].

Also another study focused on conducted a comparative study between different machine learning algorithms to evaluate the validity of data classification in terms of the efficiency and effectiveness of each algorithm in terms of accuracy, sensitivity and privacy [2].

Another study presented a combination of principal component analysis and support vector machine to develop a model that accurately assesses breast cancer risk and early diagnosis at the early stage of the disease [3]. It is well-known that most statistical methods can easily be solved on computer via using ready-made programs since some of these methods are closely linked to artificial intelligence methods such as artificial neural networks and others. To study a certain phenomenon, the variables of that phenomenon must be determined as the first step to determine the most important of those variables and consider them the basis of the phenomenon under study. Reducing a large number of variables to a smaller one, while retaining the largest amount of information about those variables, can be achieved through the Principal Component Analysis method, which helps direct attention to the significant variable instead of being preoccupied with a number of variables of little interest. Although obtaining the principal components depends on a statistical method based on the discriminant equation |A-λI|, the issue of the large size of the data whose basic components are required has some problems. The covariance matrix for a set of data of size (n), for instance, will contain (p^2) elements, and thus we will need (p^2) of arithmetic operations. Despite the efficiency of computer systems in performing these operations, they are unable to handle this type of data.

According to what has been presented before, the aim of the research was to provide a suitable method for finding the basic components based on artificial neural networks by employing a specific training base. The new method was applied to data from breast cancer patients. Despite the many studies that have focused on breast cancer, none of them took into consideration the case of synchrony (perhaps temporal) in the occurrence between the factors of this disease. In addition, a relatively modern method was presented in representing missing data based on probability.

The research aimed to employ one of the artificial intelligence methods, namely artificial neural networks, in identifying and obtaining the basic components of breast cancer, and building a neural network based on a training base suitable for this purpose. The research also aimed at clarifying how the weights of the neural network approach the characteristic vectors of the correlation matrix for the variables under study.

2. Method

2.1. Eigen values and eigen vectors

Eigen values are also called invariant values or latent values, and eigen vectors are also called invariant vectors or latent vectors. The characteristic roots of a symmetric matrix with real elements are real, and the non-zero characteristic roots of the product of two matrices such as (AB) are equal to the non-zero roots of the product of the two matrices (BA), and here the effect of (AB) is equal to the effect of (BA). For every symmetric matrix (A), there is an orthogonal matrix (P) such that P^' AP=D, where (D) represents a diagonal matrix whose diagonal elements are the characteristic roots of matrix (A), and (P) represents a matrix whose columns are the characteristic vectors of matrix (A).

The characteristic roots of the matrix (A) of dimensions (n*n) are solutions of the characteristic equation:

$$|A - \lambda I| = 0 \qquad \text{. . . (1)}$$

where the symbol (I) represents a single matrix of order (n*n) and (λ) represents the parameter of the equation (the characteristic roots). The solution to equation (1) above is a polynomial of degree (p) for the parameter (λ), and therefore the matrix (A) has (p) of the characteristic roots, and therefore:

$$|A - \lambda I| = \delta_0(-\lambda)^p + \delta_1(-\lambda)^{p-1} + \delta_2(-\lambda)^{p-2} + \cdots + \delta_{p-1}(-\lambda)^1 + \delta_p \qquad \text{. . . (2)}$$

Since:

$$\delta_0 = 1,$$

$$\delta_1 = (a_{11} + a_{22} + \cdots + a_{nn}) = tr(A)$$

$$\delta_p = |A| = \prod_{i=1}^{p} \lambda_i$$

$$(A - \lambda I)\underline{X} = 0 \qquad ...(3)$$

That is, if the vector (X) is multiplied by the square matrix (A), it will produce the vector itself multiplied by a numerical value (λ), i.e. [4].

2.2. Principal components analysis

The idea of the principal components dates back to 1901 when the scientist Karl Pearson proposed it as a means of solving the problem of multi-collinearity and adopted it as an exploratory method that could be used to arrive at an explanation of the interrelationships between variables. However, the subject did not take its mathematical and statistical form until Kendall began his research on the subject in 1957.

Principal components analysis can be defined as a method for transforming the original explanatory variables into new, unrelated variables. These new variables are called principal components, and each principal component is a linear combination of the original variables. This method is considered as an alternative in the process of estimating the regression model in the event of a multicollinearity problem between the explanatory variables (x_i). According to this method, the data is re-represented according to a certain number of linear combinations into the smallest possible number of explanatory variables derived from the original variables so that they explain most of the total variance of the original values.

The principal component (j) of a sample of (n) different observations on (p) variables can be represented by the following linear combination:

$$PC_j = a_{1j}X_1 + a_{2j}X_2 + \cdots \ddots + a_{pj}X_p \qquad ...(4)$$

where (PC_j) represents the principal component of sequence (j), and the symbol (a_ij) represents the elements of the vector (j) of the matrix (A) associated with the largest characteristic root (λ_j), where (A) can be the sample covariance matrix, and the symbol (X_j) represents the observational variables under study. Since the principal component (PC_j) is a weighted combination of variables that account for the largest amount of the total variance of the data, it can be expressed by the following formula (5):

$$PC_j = w_{1j}X_1 + w_{2j}X_2 + \ldots\ldots + w_{p_j}X_p \qquad ...(5)$$

The weights (w_j) are chosen so as to maximize the variance of the principal component [5].

2.3. Principal Components Inclination

The idea of calculating the principal components depends on the properties of the Eigen Values and their accompanying characteristic vectors. Let us assume the following linear model:

$$Y = \underline{X}\beta + \upsilon \qquad ...(6)$$

where (Y) represents the response variable vector with dimensions (n*1), (β) represents the regression coefficients vector with dimensions (n*1), (υ) represents the random error vector with dimensions (n*1), and (X) represents the matrix of explanatory variables for the variables (p) and n observations with dimensions (n*p) which can be expressed as follows:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdot\cdot & x_{1p} \\ x_{21} & x_{22} & \cdot\cdot & x_{.2p} \\ \vdots & \cdot\cdot & \cdot\cdot & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

The principal components (PC_j) can be written in the following form:

$$\mathrm{PC_j} = \mathrm{a_{1j}}\,\mathrm{X_1} + \mathrm{a_{2j}}\,\mathrm{X_2} + \ldots + \mathrm{a_{pj}}\,\mathrm{X_p}$$

$$\mathrm{PC_j} = \mathrm{X}\,\mathrm{a_j} \qquad \ldots (7)$$

where (PC_j) represents the principal component of the sequence (j), and (a_j) represents the characteristic vector (j) of the matrix (X). In general, the principal component can be written by the following linear combination:

$$pc_j = \sum_{j=1}^{p} a_{k\,j} x_k\ , \qquad j = 1,2,\ldots,p \qquad \ldots (8)$$

To adopt the principal components, some important conditions must be met, which are [6] :

1-The principal components are orthogonal (unrelated)

2-The first principal component (PC_1) explains the largest proportion of the total variance of the original variables, followed by the second principal component (PC_2) and so on.

The variance is considered as the only measure of the amount of information taken from each principal component. For this reason, the principal components are arranged in order of decreasing variance, with the most effective principal component ranked first and the least effective one ranked last. The total variance is the sum of the variances of the explanatory variables, and therefore when using the variance matrix to calculate the principal components, the sum of the variances is equal to the sum of the Eigen Values of the variance matrix, i.e.:

$$\sigma^2 = \sum_{i=1}^{m} S_i^2 = \sum_{i-1}^{m} \lambda_i \qquad \ldots (9)$$

The sum of the variances equals the number of variables when using the correlation matrix to calculate the principal components, because the variables in their standard form will have a variance equal to one and the number of variables, i.e.:

$$\sigma^2 = \sum_{i=1}^{m} S_i^2 = p \qquad \ldots (10)$$

To fulfil the above condition, constraints must be imposed on the coefficients of the basic components (a_ij), these constraints are that the sum of the squares of the coefficients must be equal to one, i.e.: $((aa)^{'}=1)$. [7]

The vector of model parameters shown in Equation (6) can be estimated by the least squares method, as in the following formula (11):

$$\hat{\beta} = (X'X)^{-1}X'Y \qquad \ldots (11)$$

Assuming an orthogonal matrix such that:

$$A'A = AA' = I \qquad \rightarrow \qquad A'(X'X)A = D$$

Where (D) is a diagonal matrix that can be expressed as follows:

$$= \begin{bmatrix} \lambda_1 & 0 & .. & 0 \\ 0 & \lambda_2 & .. & 0 \\ : & .. & .. & : \\ 0 & 0 & 0 & \lambda_p \end{bmatrix} \qquad , \quad \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p > 0$$

Thus, the columns of matrix (A) will represent the characteristic vectors associated with the characteristic roots (λ_i), and a new set of explanatory variables is obtained in the form of linear combinations that can be represented by the following formula (12):

$$PC = XA \qquad ...(12) \qquad [8]$$

2.4. Principal components capabilities

Using the least squares method, the parameters of the principal components can be estimated as follows:

$$y = p\alpha + v$$

$$\begin{aligned} v'v &= (y - p\alpha)'(y - p\alpha) \\ &= (y' - \alpha' p')(y - p\alpha) \\ &= y'y - y'p\alpha - \alpha' p' y + \alpha' p' p\alpha \end{aligned}$$

$$\therefore v'v = y'y - 2\alpha' p' y + \alpha' p' p\alpha$$

Taking the partial derivative of the above relation with respect to α^' yields

$$\frac{\partial v'v}{\partial \alpha'} = -2p'y + 2p'p\alpha$$

By setting the derivative equal to zero and then pre-multiplying both sides by (P^' P)^(-1), we get:

$$(P'P)^{-1}P'Y = (P'P)^{-1}(P'P)\,\hat{\alpha} \quad \rightarrow \quad \therefore \hat{\alpha} = (P'P)^{-1}P'Y \qquad ...(13)$$

By substituting (P^') with (A´X), formula (13) can be expressed as follows:

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\alpha}_{p-r} \end{bmatrix} = \begin{bmatrix} P_r' P_r & \cdots & 0 \\ 0 & \cdots & P_{p-r}' P_{p-r} \end{bmatrix}^{-1} \begin{bmatrix} A_r' \\ A_{p-r}' \end{bmatrix} X'Y$$

where (A_r) represents a matrix of rank (p*r) and (A_(p-r)) represents a matrix of rank (p*(p-r)), and since the elements (k-r) of the matrix (D) are equal to zero, the least squares estimator of α_r is equal to:

$$\alpha_r^{\wedge} = \left(P_r^{`} P_r\right)^{-1} A_r' X'Y \qquad ...(14)$$

Thus, the capabilities of the principal components can be expressed in terms of the values and characteristic vectors of the data matrix in the following formula (15):

$$b_{PC} = A_r' \alpha_r^{\wedge} \qquad ...(15)$$

The covariance matrix of the principal component estimators can be expressed by the following formula (16):

$$Var\,(b_{PC}) = A_r\,(P_r' P_r)^{-1}\,A_r'\,\sigma_Y^2\,I \qquad ...(16)$$

Principal component estimates are efficient, but biased. [9]

2.5. Artificial intelligence

It can be said that the aim of artificial intelligence (AI) becomes clearer if we try to get a better understanding of the nature of human intelligence, which is the only competitor to artificial intelligence and its source. This is done in order to know the role that human intelligence played in the development of artificial intelligence. Opinions differed about the definition of human intelligence by many scientists. Via the dependence on many studies, the knowledge that a person finds within himself without understanding its reason and which helps him understand things the first time is called human intelligence. [10]

Artificial intelligence is used in many fields, all of which are aimed at serving computer work, the most important of which are:

- Expert systems, which are intelligent programs that contain a lot of information possessed by a human expert in a specific field of knowledge. These programs use the laws of thinking, such as logic, common sense, and others, to reach results related to the stored information. This is due to the presence of an inference engine whose function is to detect important rules and use them.[11]
- Natural Language Processing, which seeks to understand natural languages with the aim of directly instructing the computer in this language, and then enabling the computer to converse with people by answering certain questions.
- Speech Analysis: Providing the computer with the ability to understand human speech by receiving sounds from the outside, reassembling them, recognizing them, and then responding to them.
- Vision: Providing the computer with optical sensors that enable it to recognize people or existing shapes.
- Robotics is an electromechanical machine that receives commands from a computer and performs certain tasks. Artificial intelligence in this field includes giving the robot the ability to move, understand its surroundings, and respond to a number of external factors.
- Education: The most important of which is automated augmented learning, which is an attempt to benefit from computer capabilities in the fields of education. [12]

2.6. Artificial neutral networks

It is an information processing system. It has certain performance features in a manner that mimics biological neural networks. Neural networks are used in many fields and applications. The idea of the artificial neural network is based on the biological nerve cell (the basic unit of the nervous system), whose task is to receive, interpret, and then transmit electromechanical signals. The human nervous system is built from a huge and complex number of nerves that are connected to other cells in the body.

The inputs are received by the cell body, which works to excite them until those excitations accumulate and exceed the required level. Then the cell is stimulated and sends a response appropriate to that excitation. Through the properties of the biological neuron, the artificial neuron was formed, as each artificial neuron receives a set of values entered into the network, each of which is multiplied by a weight which represents the strength of the connection point between artificial cells and corresponds to the nodes that connect biological neurons to each other, and all weighted inputs are collected to determine the level of influence in the artificial neuron. Figure 1 illustrates the structure of the neuron: [13]:

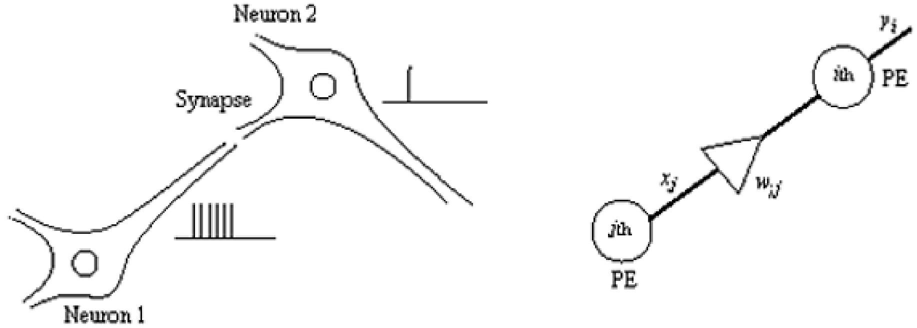

Figure 1. Structure of a biological nerve cell and its counterpart in an artificial nerve cell

The processing elements that make up the neural network are arranged in layers, which arrange the nerve nodes (Neurons) in the biological cell. What is more is that the nodes in each layer follow one method in processing information. The arrangement of the nodes in the layers and the form of connection within or between the layers is called the geometric structure or structure of the neural network, and thus we have two types of networks according to the number of their layers, which are:

- Single-layer networks.
- Multi-layer networks.

The layers consist of the input layer, the output layer, and the hidden layer(s) between them. Note that calculating the number of layers of the neural network does not include the input layer because it does not perform any computational operation.

The most distinctive feature of neural networks is their ability to overcome the limitations of traditional solutions. One of their properties is that they are characterized by parallelism, as all cells in the network work at the same time, which makes them capable of processing data at high speed despite the huge amount of data. In addition, because each cell of the neural network has its own memory unit that stores data, the loss of one of the elements does not cause a loss of data, but rather a marginal change in the effectiveness of the cells [14].

2.7. Artificial neural networks in statistics

Statistics, with its many methods and means, is useful in training artificial neural networks to display the outputs of pre-trained networks. Neural networks are trained by performing some operations that will formulate the weights of those networks. If the training is successful, applying it to a set of inputs will produce a good and desired set of outputs. There are two approaches to training methods: Deterministic and statistical. The first follows a step-by-step approach to modify the network weights based on their current values as well as the input values. The second one is based on pseudo-random changes in the weight values and remains the one that has the most advanced results. Despite the limitations of single-layer networks in the problems they can solve, for many years no method was known to train multilayer networks until statistical training provided a way out of this dilemma. [15].

The most important types of neural networks that rely on statistical concepts in their training method are: Probabilistic Neural Network:

- The concept of Probability Theory has entered neural networks to reveal what is known as the Probabilistic Neural Network (PNN), which is used in classifying samples and relies on the probability density function to determine the category to which the sample belongs. Among the most important features of the Probabilistic Neural Network are the speed of training and ease of implementation. [16]
- Back-propagation Neural Network is an algorithm used to reduce error in neural networks by adjusting weights and biases. The network training algorithm relies on the concept of stepwise regression to find the minimum square error of the output value produced by the network, allowing the network to learn from its mistakes [17].

The following Table 1 presents some statistical terms and their equivalents in neural networks:

Table 1. Statistical terms and their equivalents in neural networks

Statistic	Neural Network
Variables	
Independent Variable	Input
Predicted Values	Output
Dependent Variable	Target or Training Values
Residuals	Errors
Estimation	Training, Learning, Adaptation Self organization
Estimation Criterion	Error Function, Cost Function
Observations	Patterns or Training Pairs
Parameters Estimates	(Synaptic) Weights

2.8. Artificial neural networks in principal components analysis

According to the traditional method, the principal components are calculated, assuming that the symmetric variance-covariance matrix (or correlation matrix) has a dimension that can be dealt with while performing calculations in known computer programs such as (SAS, SPSS, MINITAB, ...) so that the inverse, variance-covariance matrix, and other algebraic operations can be obtained, in addition to the necessity of completeness of the data and obtaining at least a positive specific matrix.

Since most of the research is directed towards data compression and feature extraction, especially for radiological images or images taken from satellites and others, it has become necessary to resort to a new method and style that depends on artificial intelligence such as neural networks to facilitate dealing with the large volume of data [18].

The first law for training artificial neural networks, which are considered as one of the training methods for solving problems with linear functions, was developed by the psychologist (Donald Hebb). The idea of training is summarized by assuming the existence of a single linear neuron with d inputs, as shown in Figure 2 below:

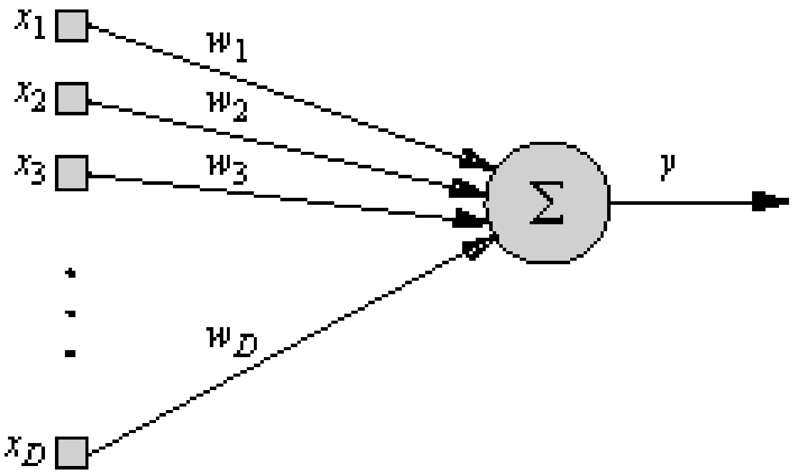

Figure 2. Hebb's processing unit

Since:

$$y = \sum_{i=1}^{d} w_i x_i \qquad ...(17)$$

and (w_i) represents the weight and is calculated according to the following formula:

$$\therefore \mathrm{w_i}(\mathrm{n}+1) = \mathrm{w_i}(\mathrm{n}) + \eta . \mathrm{y}(\mathrm{n}) . \mathrm{x_i^{\oplus}}(\mathrm{n}) \qquad ...(18)$$

where (n) represents the number of training times (iteration number), and (η) represents the learning rate. [19].

The idea of using a neural network to obtain the principal components is based on using a simple unsupervised feed-forward neural network. In fact, this network includes input units (xi) with a number of (n) accompanied by a matrix of weights with forward association with the input units, such that the column vector (wi) represents the associations between the inputs (xi) and the outputs (yi). Assuming a two-layer feed-forward neural network with (N) inputs and M outputs, such that (M<N) and all neurons are linear, the network output will be in the following form:

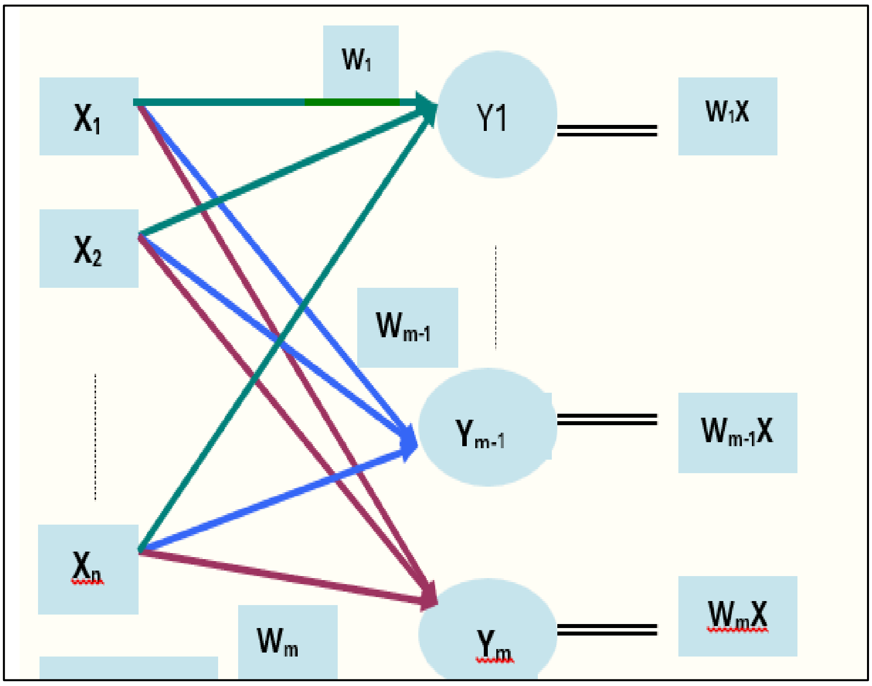

Figure 3. Architecture of the principal components network

$$\left.\begin{array}{l} y_i = \sum_{j=1}^{N} w_{ij}(n).x_j \\ y_i = w_i^T x \end{array}\right], i = 1,2,\ldots,m \qquad ...(19)$$

Figure 3 shows the architecture of the principal components network, which consists of a single feed-forward layer in which the weights are fully connected to the neurons to produce the output.

When (i=1) then (x^⊕=x) which is the case of a single neuron, the weight of the neuron will approach the first principal component. When (i=2) and assuming that all the weights of the first neuron have approached their final value then (x^⊕=x-y_1.w_1 (n)), i.e. the neuron has perceived the input from which the first principal component was removed, so (w_1 (n)) will approach the coefficient of the second principal component from the training set, and so on.

2.9. Data description

The data were collected from Hazem Al-Hafez Hospital for Cancer and Nuclear Therapy, a health institution that receives patients with various types of cancer, in addition to thyroid tumors and others. It has a specialized medical scientific staff, including surgeons, physicists and chemists, and is a center for the treatment of various cancers. A sample size of (93) cases was selected, and the study included (36) variables. Table 2 shows the incidence rates of breast cancer.

Table 2. Breast cancer incidence rates

A-According to Age		**B-According to weight**	
age	**Incidence rate**	**weight**	**Incidence rate**
-20	6%	-50	22%
-30	15%	-60	26%
-40	27%	-70	23%
-50	34%	-80	18%
-60	12%	-90	7%
-70	6%	-100	6%

C-According to marital status		**D-According to Breast-feeding**	
Marital status	**Incidence rate**	**Breast-feeding rate**	**Incidence rate**
Unmarried	16%	Non-existent	20%
Married before 35 years	21%	Acceptable (less or equal a year)	46%
Married after 35 years	63%	excellent	34%

E-Start of Menstruation		**F-Period between the start of menstruation and the onset of menopause**	
period	Incidence rate	period	Incidence rate
Early (before 11 years	16%	Not yet	6%
Normal (between 11-15 years)	62%	Short (less than 45 years)	3%
Late (after 15 years	22%	Normal (45-47 years)	40%
		Long (more than 47 years)	51%

3. Results and discussion

An artificial neural network was designed to obtain the principal components of the sample, operating according to the learning algorithm and applied in several cases according to a selected value for the learning rate (η) and the number of rotations (Iteration) to reach the most accurate results. The network is a feed-forward consisting of an input layer and an output layer, the input layer in turn consists of (36) nodes representing the number of variables under study, taking their real values without converting them to binary numbers, and the output layer represents the basic components to be obtained, and the number of nodes here is (35) nodes, i.e. (35) components to be obtained, which were later reduced to (18) basic components according to the stopping criterion (COM), and Figure 4 illustrates this:

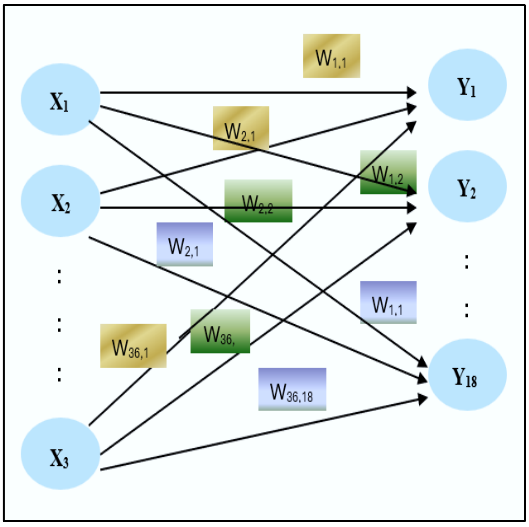

Figure 4. Network architecture of the principal components of breast cancer

Where (Xi) represents the explanatory variable with sequence (i) where i= 1, 2,….,36. And (w) represents the network weight matrix that corresponds to the characteristic vector matrix of the original data, where (wij) represents the weight with sequence (i) in column (j).

After applying the algorithm to the data under study, the following results were reached:

Table 3. Distinctive roots (Eigen values) and their explanation of sample variance

Comp.	Eigen values	% of Variance	Cumulative %	Comp.	Eigen values	% of Variance	Cumulative %
1	3.873	10.759	10.759	19	.716	1.990	86.307
2	2.735	7.598	18.357	20	.621	1.724	88.031
3	2.433	6.759	25.117	21	.594	1.650	89.681
4	2.337	6.491	31.607	22	.572	1.588	91.269
5	2.178	6.049	37.656	23	.498	1.384	92.654
6	1.943	5.397	43.054	24	.431	1.198	93.851
7	1.872	5.201	48.255	25	.385	1.070	94.921
8	1.701	4.726	52.981	26	.335	.931	95.852
9	1.628	4.524	57.504	27	.297	.824	96.676
10	1.431	3.976	61.480	28	.278	.773	97.449
11	1.305	3.626	65.106	29	.251	.698	98.147
12	1.246	3.460	68.566	30	.205	.570	98.717
13	1.163	3.231	71.798	31	.149	.415	99.131
14	1.100	3.054	74.852	32	.128	.354	99.486
15	1.007	2.799	77.650	33	.107	.298	99.784
16	.870	2.416	80.067	34	0.00572	.159	99.943
17	.783	2.176	82.242	35	0.00206	0.00573	100.000
18	.747	2.075	84.317	36			100.000

Table 3 shows that the first distinct root explains (10.759%) of the total variance, then decreases to (7.598%) for the second root, and so on. By taking the cumulative amount of what the basic components explain, it is found that (78%) of the sample variance has been explained by the first fifteen distinct roots out of the thirty-six distinct roots, while the last twenty-one distinct roots do not explain more than (22%) of the total variance. Accordingly, this result can be relied upon by taking the first fifteen basic components.

Table 4 shows the results reached using the (Hebb) algorithm, where the amount of commonality represents the extent of the correlation between the variables and the extracted factors, and the larger and closer it is to one, the more this indicates that the variable is associated with the extracted factors, and if it falls between zero and one, this indicates that the overlap is partial between the variables and factors.

Through observing the amount of prevalence, we find that the variable (X1) age has the strongest association with the fifteen factors (the basic components that were previously identified), then (x11) number of pregnancies, and so on until it reaches the variable that has the least association with the group of factors, which is (x26) the need for blood exchange. As previously concluded, the first factor is the most important of the other factors in explaining (10.749%) of the total variance.

Table 4. Distinctive roots and their explanation of sample variance

variables	Amount of prevalence	variables	Amount of prevalence	variables	Amount of prevalence
X_1	0.930	X_{13}	0.729	X_{25}	0.765
X_2	0.753	X_{14}	0.820	X_{26}	0.613
X_3	0.673	X_{15}	0.870	X_{27}	0.810
X_4	0.787	X_{16}	0.796	X_{28}	0.745
X_5	0.903	X_{17}	0.792	X_{29}	0.794
X_6	0.849	X_{18}	0.744	X_{30}	0.728
X_7	0.816	X_{19}	0.740	X_{31}	0.851

X_8	0.684	X_{20}	0.693	X_{32}	0.683
X_9	0.771	X_{21}	0.638	X_{33}	0.790
X_{10}	0.69	X_{22}	0.781	X_{34}	0.815
X_{11}	0.922	X_{23}	0.890	X_{35}	0.820
X_{12}	0.875	X_{24}	0.633	X_{36}	0.736

Table 5. The basic factors that explain the highest possible amount of the total variance were distinguished, in addition to distinguishing the variables influencing each basic factor through the matrix of rotated factor loadings

factors	Factors variance	The amount of variance explained by a factor	Factors	Factors variance	The amount of variance explained by a factor
1	3.0256	8.4	9	1.6619	4.6
2	2.4689	6.9	10	1.5724	4.4
3	2.177	6.0	11	1.5220	4.2
4	2.1124	5.9	12	1.5127	4.2
5	1.8857	5.2	13	1.5003	4.2
6	1.8836	5.2	14	1.4883	4.1
7	1.8722	5.2	15	1.4220	4.0
8	1.8309	5.1	total		78

Table 5 shows the variance of each of the basic factors and the amount it explains of the total variance, which is the sum of what the factors together explain (78%) of the total variance. The first factor sorted the variables (X9: marital status), (X18: breast-feeding) to a good degree, and it also distinguished the variables (X12: number of children) and (X10: age at marriage), and thus this factor can be called marital status. The second factor sorted the variables (X1: age), (X5: age at diagnosis) and (X19: breast-feeding period), in addition to sorting the variables (X3: height) and (X11: number of pregnancies), and thus this factor can be called age and breastfeeding. The third factor classified the variables (X27: presence of a family history of the disease), (X29: presence of a second-degree disease) and (X17: use of different medications) and distinguished (X8: onset of menstruation) and (X16: pregnancy stabilizer) as the second degree, and thus this factor can be called a familial disease. The fourth factor classified the variables (X14: use of therapeutic hormones) and (X16: use of pregnancy stabilizer) and (X13: number of miscarriages) as the first degree and the variables (X15: use of contraceptives), (X17: use of other medications) as the second degree, and thus this factor can be called a therapeutic hormone. The fifth factor sorted the variables (X34: Eating pickled foods) and (X35: Eating spicy foods) and sorted the factors (X32: Eating foods rich in animal and vegetable fats), and thus this factor can be called the type of nutrition. The sixth factor sorted the variables (X6: Age at the start of menstruation) and (X7: Regularity of menstruation) to a high degree as well, and to a second degree this factor sorted each of (X3: Height) and (X15: Use of contraceptives), and thus this factor can be called menstrual problems. The seventh factor sorted the variables (X22: breast surgery), (X23: site of breast surgery) to a large degree and the variables (X25: lymph node surgery), as well as (X16: use of pregnancy stabilizer) and thus this factor can be called breast and lymph node surgery. The eighth factor sorted the variables (X30: exposure to radiation) and its negative effect, i.e. increased exposure to radiation affects the incidence of the disease (X26: the need for blood exchange or replacement), (X10: age at marriage) and this factor can be called exposure to radiation. The next factor sorted the variables (X31: taking vitamins) as reducing it negatively affects the incidence of the disease and vice versa, (X12: number of children), (X3: height). The tenth factor sorted the variables (X20: presence of breast pathological problems), (X24: ovarian surgery), (X28: presence of disease in first-degree relatives) This factor can be called breast and uterine pathological problems. The eleventh factor sorted the variables (X4: level of education) and (X36: smoking). This factor can be called general culture. The twelfth factor sorted the variables (X11: number of pregnancies), (X12: number of children), (X15: use of contraceptives) and (X32: intake of foods rich in animal and vegetable fats) and in the second degree sorts (X16: pregnancy stabilizer) and (X19: breastfeeding) and accordingly this factor can be called pregnancy and number of children. The thirteenth factor sorted the variables (X33: Eating sweets) in the first degree and the variables (X32: Eating foods rich in animal and vegetable fats) and (X31: Eating vitamins) in the second degree, which gives an idea that this factor is concerned with nutrition. The penultimate factor sorted the variables (X8: Regularity of the menstrual cycle), (X21: Location of the breast disease problem). The last factor sorted the variables (X2: weight) and (X25: lymph node surgery) and thus this factor can be called the change in lymph node hormones.

4. Conclusion

- It can be concluded that the weights of the artificial neural network based on the (Hebb) rule are close to the characteristic vectors of the correlation matrix. The flexibility of the work of artificial neural networks allows expanding the use of neural networks in other statistical methods.
- The highest incidence of breast cancer is in the age group of 50 years and the weight group of 60 kilograms, and in married women after 35 years, and in breast-feeding women with an acceptable rate, and when the menstrual cycle rate is early.
- It can also be concluded that the most important factors causing breast cancer in the environment surrounding the city of Mosul are social status, age and breastfeeding, and family infection with the disease.

References

[1] M. G. Lopez *et al.*, “BCDR: a breast cancer digital repository,” *15th International conference on experimental mechanics*, vol. 1215, pp. 113–120, 2012.

[2] H. Asri, H. Mousannif, H. A. Moatassime, and T. Noel, “Using machine learning algorithms for breast cancer risk prediction and diagnosis,” *Procedia Computer Science*, vol. 83, pp. 1064–1069, 2016.

[3] B. A. Akinnuwesi, O. Babafemi, and B. S. Macaulay, “Breast cancer risk assessment and early diagnosis using Principal Component Analysis and support vector machine techniques,” *Informatics in medicine unlocked*, vol. 21, 2020.

[4] H. Abdi, “The eigen-decomposition: Eigenvalues and eigenvectors,” *Encyclopedia of measurement and statistics*, pp. 304–308, 2007.

[5] B. M. S. Hasan and A. M. Abdulazeez, “A review of principal component analysis algorithm for dimensionality reduction,” *Journal of Soft Computing and Data Mining*, vol. 2, no. 1, pp. 20–30, 2021.

[6] A. A. Afifi, S. May, and V. A. Clark, *Computer-aided multivariate analysis, fourth edition*, 4th ed. Philadelphia, PA: Chapman & Hall/CRC, 2003.

[7] N. M. Mohankumar, M. Jain, H. Wan, S. Ganguli, K. D. Wilson, and D. M. Anderson, “Principal component regression to study the impact of economic factors on disadvantaged communities,” *arXiv [stat.AP]*, 2024.

[8] D. G. Kleinbaum, L. L. Kupper, and K. E. Muller, *A , Applied regression analysis and other multivariable methods*, vol. 601. Belmont, CA: Duxbury press, 1988.

[9] F. Richard and L. Gunst Rober, *Regression Analysis and its Application”*. New York, U.S.A: Decker Inc, 1980.

[10] S. J. Russell and P. Norvig, *Artificial intelligence: a modern approach*. Pearson, 2016.

[11] B. A. Frederick, D. Matthias, and O. E. Taylor, “Expert System for Control and Maintenance of Steam Package Boiler Drum and Feed Water Using Rule-Based Fuzzy Logic Techniques,” *American Journal of Computer Science and Technology*, vol. 6, no. 2, pp. 80–95, 2023.

[12] S. I. Yahya *et al.*, "A New Design Method for Class-E Power Amplifiers Using Artificial Intelligence Modeling for Wireless Power Transfer Applications," *Electronics,* vol. 11, no. 21, p. 3608, 2022.

[13] P. C. L. Hui, Ed., *Artificial neural networks: Recent advances, new perspectives and applications*. London, England: IntechOpen, 2023.

[14] K. O’Shea and R. Nash, “An Introduction to Convolutional Neural Networks,” *arXiv [cs.NE]*, 2015.

[15] P. Wasserman, “Neural computing: Theory and practice,” *Van Nostrand Reinhold google schola*, vol. 4, pp. 1121–1126, 1989.

[16] B. Mohebali, A. Tahmassebi, A. Meyer-Baese, and A. H. Gandomi, “Probabilistic neural networks: a brief overview of theory, implementation, and application. Handbook of probabilistic models,” pp. 347–367, 2020.

[17] T. P. Lillicrap, A. Santoro, L. Marris, C. J. Akerman, and G. Hinton, “Backpropagation and the brain,” *Nat. Rev. Neurosci.*, vol. 21, no. 6, pp. 335–346, 2020.

[18] S. G. Hoggar, “Mathematics of digital images: Creation, compression, restoration, recognition (hoggar, S.g.; 2006) [book review],” *IEEE Signal Process. Mag.*, vol. 25, no. 4, pp. 128–129, 2008.

[19] C. Chatterjee, V. P. Roychowdhury, and E. P. Chong, “On relative convergence properties of principal component analysis algorithms,” *IEEE Trans. Neural Netw.*, vol. 9, no. 2, pp. 319–329, 1998.