

ORIGINAL RESEARCH

Knowledge-based system for prognosis of specific types of cancer using Elman neural network

Marios Poulos

Department of Archives and Library Science, Ionian University, Ioannou Theotoki, Corfu, Greece

Correspondence: Marios Poulos. Address: Department of Archives and Library Science, Ionian University, Ioannou Theotoki 72-49100, Corfu, Greece. Email: mpoulos@ionio.gr.

Received: November 26, 2012

Accepted: December 20, 2012

Online Published: February 18, 2013

DOI: 10.5430/air.v2n2p62

URL: <http://dx.doi.org/10.5430/air.v2n2p62>

Abstract

This paper focuses on the classification problem of high dimensional patterns and especially of socio-demographic cancer questionnaires. The purpose of this study is to define a predictive indicator of a published clinical study regarding the influence of Hormone Replacement Therapy (HRT) on the growth of cancers, including breast, ovarian, endometrial, and colon cancers. The proposed study, in the preparation stage, combines independent factors of this research using a Bayesian model in order to achieve a normalizing data linked by significant relevant properties of these factors. The specific goal is to determine a standard threshold value in which an independent self-organizing system will decide the correlation between the normalizing data of the preprocessing stage via a well-fitted, recurrent Elman neural network using a threshold factor which is called the distance value. A case study involving a dataset of published clinical research is used and the evaluated procedure is implemented by a well-fitted t-test control.

Key words

Bayesian model, Elman neural network, Endometrial cancer, Prediction system

1 Introduction

1.1 Background and motivation

Artificial neural networks (ANNs) and decision trees (DTs) have been used in cancer detection and diagnosis for nearly 20 years^[1]. This paper focuses on the classification problem of unrelated characteristics, especially those of epidemiologic questionnaires. The baseline questionnaire selected for the present study^[2] is based on socio-demographic features (age, marital status, race, and education), height and body weight, generative and obstetric history, family and personal medical history, cigarette smoking, and use of hormone replacement therapy (HRT). The anonymous questionnaire results are related to findings of Endometrial Thickness (ET) measurements, which were taken using transvaginal ultrasonic examination^[2].

The preparation stage of data processing is the most significant procedure in stochastic and probabilistic decision systems^[3] an important subclass of which is neural networks. Modeling has an enormous impact on the success of a wide range of such complex data analyses as data mining and feature-extraction procedures^[3,4], mainly because the quality of the input data in neural network models strongly influences the results of analyses^[3,5] and the efficiency of their performance, as wrongly prepared data are likely to result in problematic analyses. The appropriate pre-processing of input

data is therefore a significant procedure [3, 6]. A study of the procedure for grouping data collected from a baseline questionnaire, which addresses the selection of the variables with genetic algorithms and those algorithms' linkages to significant relevant properties, would therefore be valuable.

1.2 Problems with the data preparation stage

Grouping non-structural data categories with the specific multi-scale and normalization practices may be overly simplistic [3, 7], as doing so cannot provide solutions for significant multi-factorial unconstructed data that are linked with significant relevant properties. Processes such as demographic data collecting for information management in epidemiology, and educational learning evaluations, often use such data.

Existing solutions have focused on the consolidation phase, during which data management systems consist of a continuous process. The capture procedure stores unconstructed data in a well-constructed database [8], but the basic problem lies in the pre-processing training setting, as the number of possible interaction factors grows exponentially with the inclusion of each additional main effect in the logistic regression model. Logistic regression's ability to deal with interactions involving many factors is therefore limited, as having too many independent variables in relation to the number of observed outcome events is problematic [9, 10].

1.3 Aims and scope

This paper attempts to solve the problems of preparing and normalizing data linked with significant relevant properties and captured from many sources by adapting these factors to a first phase using a Bayesian model that creates normalization conditions with an equitable distribution between the significant relevant properties of these data. It also aims to adapt this normalization to a well-fitted neural network so it can accept pre-processing data for training and testing procedures. This is accomplished using two separate procedures. First, the data variable is implemented in a functional multi-factorial normalization analysis using a normalizing constant. Second, constructed vectors containing normalization values are used in the learning and testing stages of the selected learning vector quantifier neural network. To demonstrate the application of this model, an attempt is made to consolidate and normalize three different groups of factors (socio-demographic characteristics, reproductive and gynecologic history and miscellaneous factors) in order to ensure these data consist of a well-fitted vector of an artificial neural network. The ultimate goal is to achieve a system decision that indicates whether or not there is a significant probability of pathogenic ET (i.e., >3 mm), which would implicate the presence of a type cancer.

In section 2 (Proposed Methods) the integrated data preparation method and the architecture of the neural network are explained. Section three (Experimental Component) consists of the implementation stage of the proposed method. The results are evaluated in section 4. Finally, the conclusion and the possible future uses of this information management method are explained. In brief all these steps are depicted in Figure 1.

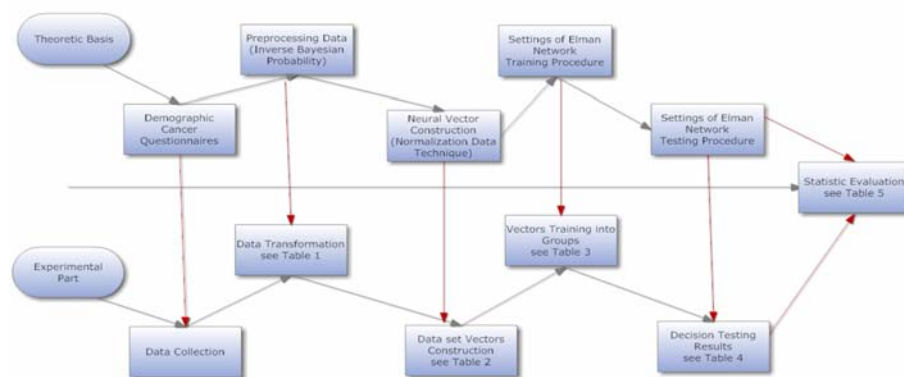


Figure 1. A block-diagram representing the entire path of cancer prognosis (starting from data-extraction to cancer detection)

2 Proposed methods

2.1 Integrated data preparation method

According to previous work ^[7], the proposed technique to formulate related features of non-structural data is based on extracting related features from text data or documents by semantic analysis, and formulating an event-specific summary. However, this technique does not support the case of the proposed method because it is based on the % of related results. In this case, the solution of this conversion could be found in the Bayes' theorem, according to which the posterior probability measure is proportional to the product of the prior probability measure and the likelihood function. Proportional to implies that one must multiply or divide by a normalizing constant to assign a measure of 1 to the whole space, i.e., to get a probability measure. In a simple discrete case:

$$P(H_0 | D) = \frac{P(D | H_0)P(H_0)}{P(D)} \quad (1)$$

where P(H0) is the prior probability that the hypothesis is true; P(D|H0) is the conditional probability of the data given that the hypothesis is true, but given that the data are known it is the likelihood of the hypothesis (or its parameters) given the data; P(H0|D) is the posterior probability that the hypothesis is true given the data. P(D) should be the probability of producing the data, but on its own is difficult to calculate, so an alternative way to describe this relationship is as one of proportionality:

$$P(H_0 | D) \sim P(D | H_0)P(H_0) \quad (2)$$

Since P(H|D) is a probability, the overall possible (mutually exclusive) hypothesis should be 1, leading to the conclusion that

$$P(H_0 | D) = \frac{P(D | H_0)P(H_0)}{\sum_i P(D | H_i)P(H_i)} \quad (3)$$

In this case, the reciprocal of the value

$$P(D) = \sum_i P(D | H_i)P(H_i) \quad (4)$$

where,

$$k(j)=1/P(D) \quad (5)$$

is the normalizing constant and where $j=1$, and n is the number of parameters. Thus, vector $k(j)$ of size $1 \times j$ is constructed.

Thus, using the advantage that Bayes' theorem is a solution to the problem of inverse probability, the inverse probability that $ET < 3$ mm ^[11] is calculated. Then Tables 1, 2 and 3 from ^[3] are transformed into a two column probability that $ET \geq 3$ mm and $ET < 3$ mm. More details of this transformation are in section 3.

2.2 Neural network architecture

In this study, a well-defined neural network (NN) was selected to provide an appropriate weight in the process of learning for input vectors which present variability according to previous study ^[11, 12]. Furthermore, this NN could be trained to

distinguish and produce both spatial and temporal patterns that resolve the problem of the factors' threshold value variability in order to support a rule-based decision procedure. Thus, the recurrent Elman neural network was selected over BP and SOM, because according to previous work on similar problems, a comparison of their architectures (i.e., a multilayer perceptron (MLP) trained with the resilient backpropagation (RPROP)) showed that the best prediction accuracy was obtained with the extended Elman neural network^[13]. Furthermore, one of the problems of SOM is that it is not a clustering approach but rather is a way of mapping a complex, multidimensional space onto a simpler two-dimensional space, as characterized by the network^[14].

The Elman network is based on a two-layer network with feedback in the first (hidden recurrent) layer and a second output layer. This recurrent connection permits the Elman network to both detect and generate time-varying patterns^[15]. The hidden recurrent layer consists of neurons with a hyperbolic tangent activation function (tansig) as described in the application^[16] and is represented by the following equation:

$$\text{tansig}(X) = \frac{1 - e^{-2x}}{1 + e^{-2x}} \quad (6)$$

The output layer is characterized by a linear activation function^[17] (Casella, 2006). In this case the two different types of input vectors (high risk, i.e., $ET \geq 3$; and low risk, i.e., $ET \leq 3$ for developing a type of cancer) respond to the function.

Summing and activation functions

A modified Elman, with non-linear neurons in the hidden layer and linear neurons in the lasting layers, is employed, and the hyperbolic tangent function (see Eq. (6)) has been adopted as the activation function of the non-linear neurons^[18]. In this case, the significant problem is the discrimination thresholds between the input classes of vectors^[18, 19], where the threshold is an adaptive value obtained as a function of the maximum of the difference between the log-likelihood of the maximum category and the log-likelihood of the medium and minimum categories, respectively^[20].

It is known that in a biological context a neuron is activated when it detects electrical signals from the neurons to which it is connected^[21]. If these signals are sufficient, the neuron will send further electrical signals to the neurons connected to it. An activation function is similar: the artificial neuron will output a value (α) based on inputs received. It is almost always the case that a neuron will output an (α) value between [0, 1] or [-1, 1]. This normalization of data occurs by using the summed inputs as a parameter to a normalizing function called the "activation function"^[22]. In the Elman network, the activation function is accomplished by the network training procedure, which receives x_j input training vectors, and the following polynomial y term in sigma-pi units is extracted:

$$y = f\left(\sum w_{ij}x_j\right) \quad (7)$$

where j is the number of training vectors in the input; and i is the number of training epochs.

The network described here uses backpropagation as the learning algorithm. This algorithm permits us to modify the weights in the network in response to the errors produced for the training data. In the most general terms, the algorithm can be understood as a way to accomplish credit/blame assignment. More specifically, the algorithm involves the following weight adjustment equation^[16, 23]:

$$w_{ij} = n\delta_i a_j \quad (8)$$

This equation defines the weight change between any two units, i and j , as the product of three terms. The first term, n , is a scaling constant and is referred to as the learning rate. It is typically a small value, so learning occurs in small increments.

The last term, a_i , is the activation of the sender and implements the credit/blame aspect of the algorithm. The middle term, δ_i , is calculated as follows:

$$\delta_i = f'(net) \quad (9)$$

where the error (in the case of an output unit) simply represents the discrepancy between the target output of the unit and the actual output, while $f'(net)$ represents the derivative of the receiver unit's activation function, given its current net input. Eq. (2) is transformed by the following equation:

$$y = f(g - \theta_h) \quad (10)$$

where

$$g = \prod_{i=1, j=1}^N x_i^{w_{ij}} \quad (11)$$

N is the number of experimental training epochs; $f(x)$ is the activation function, which normally takes the form of a sigmoidal or threshold function; g is the total input stimulus; and θ_h is the threshold.

In this section, the coordinated trajectory dynamics clearly represent a solution that can process the maximum, medium and minimum performance cases, is selected. The argument is based on the dynamic properties of the linearized systems derived in the analysis (see section 2.1). This proposed system is an application of the counting solutions of analog computation theory, except that this system can produce output predictions that are linearly separable [24]. The Elman network is set to respond within the following thresholds, which are determined by the following piecewise-linear function [24].

$$f(x) = \begin{cases} +1 \rightarrow \text{if} & +0.5 < x \leq +1.5 \\ +2 \rightarrow \text{if} & +1.5 < x \leq +2.5 \end{cases} \quad (12)$$

In this study, function $f(1)$ corresponds to the maximum risk case with a low threshold value of 0.5, and function $f(2)$ corresponds to the low risk case with a low threshold value of 1.5. The minimum threshold of 0.5 is considered to be the value of a low risk threshold, which is compared to an extracted value denoting a candidate set. This configuration is established by default in the activation learning procedure, where a neuron that is activated by a class of input vectors, u^1 (maximum case), returns threshold values between 0.5 and 1.5. The same is true for u^2 . So the purpose of a specific minimum threshold is to compare each extracted value of the unknown vector being tested using the purellin function with a value of 0.5.

Thus, the inputs to the net are coded in binary using 1 or 2 values (see Eq. (12)). Each input is weighted with its appropriate weight, W_{ji} , and the sum of the weighted inputs and the bias form the input to the transfer function. The outputs of the transfer function are then fed into the hidden layer as inputs. The output prediction value is derived using the same procedure from the hidden layer to the output layer. The log-sigmoid transfer function used and the relationship between inputs and outputs are as follows:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (13)$$

In the training procedure, the extracted weights (w_1 and w_2) create the rule by organizing the normalized input vectors in relation to the thresholds. In this way, the constructed Elman neural network creates a global “cited” coefficient. The extracted distance values depend exclusively on a well-fitted neural network. This adaptation depends upon the correct ranking of the trained vectors in the established categories, a suitable setting for the trained vector values for each category, and finally, the selected number of neurons.

2.3 Extraction of the cited distance

In the network testing procedure, each candidate testing vector is related to the extracted weights of the training procedure via Eq. (12), and this relationship is depicted by the following equation:

$$a = -\frac{\ln |(f(x) + 0.5)|}{x} \quad (14)$$

Finally, the risk (i.e., the probability of developing an $ET \geq 3$ mm), is calculated by the following:

$$Risk = \frac{\theta_h}{a} \quad (15)$$

where, in this case:

$$\theta_h = 0.5 \quad (16)$$

The idea from previous work ^[25] was selected as a benchmark for describing the ideal vector’s high risk case. In this work, as a benchmark consideration, a set of a vector’s values were used that has a null probability (high risk) of developing $ET \geq 3$ mm. Thus, as the cited distance is greater, there is a greater probability that the ET will increase to become $ET \leq 3$ mm and thus, the probability of cancer is decreased.

3 Experimental component

3.1 Integrated data preparation method

In this stage, the Integrated Data Preparation Method (see section 2.1) is applied to a distribution of factors (e.g., demographic, reproductive and others) for women stratified according to whether they were receiving Hormone Replacement Therapy (HRT). These data were originally obtained by ^[2]. The present study was implemented in two steps.

In the first step, we prepared the normalizing data in accordance with the proposed method (see section 2.2.1). More specifically, the normalized constants (k_j) according to Eq. (5) were constructed, where j is the number of characteristics included in the work ^[2]. To illustrate this normalization, an example is given using Table 1 ^[2] and, specifically, data pertaining to women aged 55-59:

1) $ET \geq 3$ mm

- The probability that one of the women has $ET \geq 3$ mm is $\alpha = 71.7/100$
- If a woman has $ET \geq 3$ mm, then the probability that she will currently be receiving HRT is $b = 80.5/100$
- If a woman does not have $ET \geq 3$ mm, then the probability is that she is not currently receiving HRT is $c = 61.7$.

2) $ET < 3$ mm

- The probability that one of these women has $ET < 3$ mm is $a = (100 - 71.7 = 28.3) / 100$
- If a woman has $ET < 3$ mm, then the probability is $(100 - 61.7 = 38.3) / 100$ that she will not have be currently receiving HRT
- If a woman does not have $ET < 3$ mm, then the probability that she is currently receiving HRT is $(100 - 80.5 = 19.5) / 100$

These data are calculated in terms of cases $ET \geq 3$ mm and $ET < 3$ mm via the Bayes function of MATLAB ^[26]:

```

%% ET >= 3mm
a=71.7
b=61.7
c=80.5
k1=bayes(a/100,b/100,c/100,1)
%% ET < 3mm
a1=100-a;
b1=100-b;
c1=100-c;
k2=bayes(a1/100,b1/100,c1/100,1)

```

where

a is the a priori probability (prior, marginal or actual)

b is the first conditional probability, $p(Y|X)$ [option=1]; first interaction

probability, $p(XY)$ (true positive) [option=2] or conditional

probability, $p(Y|X)$ [option=3]

c is the second conditional probability, $p(Y|\sim X)$ [option=1]; second interaction

probability, $p(X\sim Y)$ (false positive) [option=2] or interaction

probability, $p(\sim XY)$ [option=3]

and the values $k1=0.66$ and $k2=0.44$ are calculated (see Table 1).

Thus, in Table 1 the integrated data preparation is constructed; it contains the 45 characteristics of 15 groups which belonged in three categories (socio-demographic characteristics, reproductive and gynecologic history and miscellaneous factors).

Table 1. Integrated data preparation from Tables 1, 2 and 3 ^[2]

Groups	Characteristics	Normalized Constant (k) ET \geq 3mm	Normalized Constant (k) ET \leq 3mm	Reference
Age, y	55-59	0.66	0.44	Table 1 ^[2]
	60-64	0.56	0.65	ibid.
	65-69	0.53	0.62	ibid.
	70-74	0.53	0.51	ibid.
Marital Status	Married	0.62	0.52	ibid.
	Unmarried	0.52	0.60	ibid.
Education	Less than high school	0.47	0.56	ibid.
	High school graduate	0.56	0.59	ibid.
	Beyond high school	0.58	0.56	ibid.
	College graduate	0.62	0.50	ibid.
Ethnicity	White	0.58	0.55	ibid.
	Black	0.72	0.39	ibid.
	Other	0.32	0.71	ibid.
Parity	Nulliparous	0.62	0.53	Table 2 ^[2]
	Parous	0.58	0.55	ibid.
	Unknown	0.32	0.62	ibid.
Age at menarche, y	<12	0.67	0.39	ibid.
	12-13	0.56	0.58	ibid.
	\geq 14	0.59	0.58	ibid.
Age at menopause, y	<40	0.68	0.32	ibid.
	40-49	0.50	0.60	ibid.
	\geq 50	0.63	0.51	ibid.
Uterine fibroid	Ever	0.72	0.41	ibid.
	Never	0.57	0.56	ibid.
Oral contraceptive use	Ever	0.57	0.55	ibid.
	Never	0.61	0.53	ibid.
HRT use	Never	1	0.24	ibid.
	Former	1	0.23	ibid.
	Current	0	0.55	ibid.
Years since menopause	<5	0.76	0.33	ibid.
	5-9	0.59	0.54	ibid.
	10-14	0.54	0.55	ibid.
	\geq 15	0	0.41	ibid.
Smoking	Never	0.61	0.50	Table 3 ^[2]
	Former	0.57	0.58	ibid.
	Current	0.54	0.60	ibid.
Diabetes	Ever	0.69	0.33	ibid.
	Never	0.58	0.56	ibid.
Hypertension	Ever	0.64	0.47	ibid.
	Never	0.56	0.57	ibid.
BMI(kg/m ²)	<23.25	0.42	0.63	ibid.
	23.26-26.03	0.56	0.58	ibid.
	26.04-29.88	0.61	0.61	ibid.
	>29.89	0.73	0.43	ibid.

In the second step, $n=2$ classes (i.e., case 1 and case 2) of input $k(j)$ vectors according to classification problems. The size of each vector is determined by 44 elements which included all the characteristics in a particular order. An example of two vectors (case 1 and case 2) is presented in Table 2. For each case, a total of $4*2*4*3*3*3*3*2*2*3*4*3*2*2*4=5971968$ vectors are formed; these consist of the raw values from the original clinical study ^[2].

Table 2. Vector examples for case1 and case2

Groups	Characteristics	Vector Case1 ET \geq 3mm	Vector Case2 ET \leq 3mm	Reference
Age, y	55-59	0.6600	0	Table 1 ^[2]
	60-64	0	0.65	ibid.
	65-69	0	0	ibid.
	70-74	0	0	ibid.
Marital Status	Married	0	0	ibid.
	Unmarried	0.5200	0.60	ibid.
Education	Less than high school	0	0.56	ibid.
	High school graduate	0	0	ibid.
	Beyond high school	0	0	ibid.
	College graduate	0.6200	0	ibid.
Ethnicity	White	0	0	ibid.
	Black	0	0	ibid.
	Other	0.3200	0.71	ibid.
Parity	Nulliparous	0	0.53	Table 2 ^[2]
	Parous	0	0	ibid.
	Unknown	0.3200	0	ibid.
Age at menarche, y	<12	0	0.39	ibid.
	12-13	0.5600	0	ibid.
	\geq 14	0	0	ibid.
Age at menopause, y	<40	0	0.32	ibid.
	40-49	0	0	ibid.
	\geq 50	0.3200	0	ibid.
Uterine fibroid	Ever	0.7200	0.41	ibid.
	Never	0	0	ibid.
Oral contraceptive use <<	Ever	0	0.55	ibid.
	Never	0	0	ibid.
HRT use	Never	1.0000	0.24	ibid.
	Former	0	0	ibid.
	Current	0	0	ibid.
Years since menopause	<5	0.7600	0.33	ibid.
	5-9	0	0	ibid.
	10-14	0	0	ibid.
	\geq 15	0	0	ibid.
Smoking	Never	0.6100	0	Table 3 ^[2]
	Former	0	0.58	ibid.
	Current	0	0	ibid.
Diabetes	Ever	0.6900	0	ibid.
	Never	0	0.56	ibid.
Hypertension	Ever	0.6400	0	ibid.
	Never	0	0.57	ibid.
BMI(kg/m2)	<23.25	0.4200	0	ibid.
	23.26-26.03	0	0.58	ibid.
	26.04-29.88	0	0	ibid.
	>29.89	0	0	ibid.

3.2 Neural network architecture

This stage is divided into two sub-stages, which are: a) the appropriate pre-processing of the vectors which train the Elman network, and b) the experimental setting.

3.2.1 Pre-processing for training

The final step of the proposed method is to use two classes (case 1 and case 2) of 20 vectors each after bootstrapping experimentation [27]. The vectors used were selected from those formed during the procedures outline in section 3.1.

3.2.2 Experimental setting

Finally, in this stage the Elman network was optimally trained using 40 vectors (1×44 each; 20 for each category). During the experimental training the neural network conditions imposed were satisfied after 382 training rounds (see Figure 2). The best convergence (284 epochs) using 12 neurons with a sum squared error of 0.02 was obtained (see Figure 2). This procedure was achieved via the following function of the nnet Toolbox of MATLAB:

```
[w1,b1,w2,b2]=initelm(P,12,C)
```

```
[w1,b1,w2,b2]=trainelm(w1,b1,w2,b2,P,C)
```

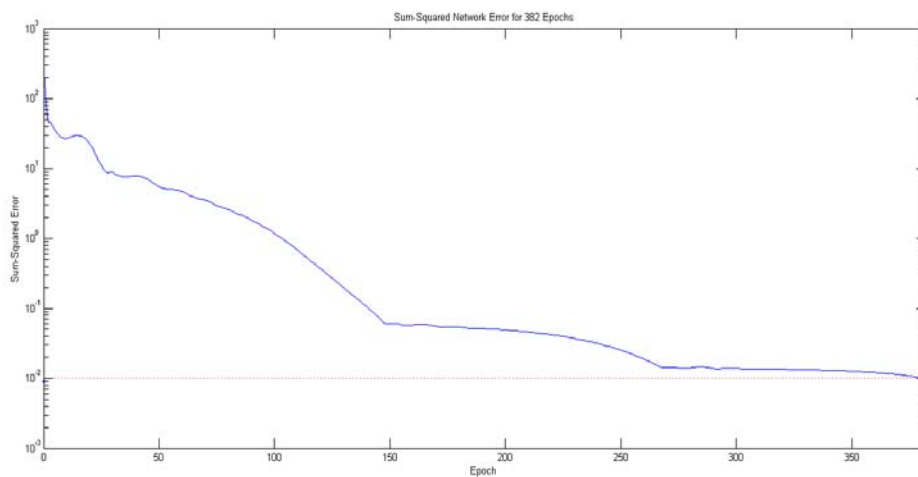


Figure 2. Sum-squared network error for 382 epochs

4 Results

4.1 Testing procedure

One hundred vectors were included in the testing procedure, in which the raw values of each vector were converted to membership values in order use to a real health case as a test control. This conversion is often practiced in engineering applications; a continuous function is used to convert raw values into membership values for the complete range $[0, 1]$ [28, 29]. In this study, the log-sigmoid transfer function (see section 2.2.1) was used as a continuous function. In this way, 50 vectors (with true values) representing each case were selected.

As an example, two characteristic vectors of these cases are presented in Table 3. According to Eq. (5) and using the sigmoid function `simuelm` of MATLAB, the values (a) of the ranking in relation to the threshold values $0.5 < x < 2.5$ (see section 2) were obtained. In this procedure, two kinds of vectors participated: the high risk ($ET \geq 3$ mm; u) and low risk

($ET < 3$ mm; m). Furthermore, the cited distance was calculated according to Eq. (6), where $\theta_h = 0.5$. For example, in Table 3, the first case vector $ET \geq 3$ is compared with the weight matrixes w_1 and w_2 , and return a value of 0.5194, while in the second case vector $ET \leq 3$ is compared with the weight matrixes w_1 and w_2 and return a value of 2.2831.

Table 3. Vector examples for case 1 ($ET \geq 3$ mm) and case 2 ($ET < 3$ mm)

Groups	Characteristics	Truth Data $ET \geq 3$ mm	Truth Data $ET < 3$ mm	Reference
Age, y	55-59	1	0	Table 1 ^[2]
	60-64	0	1	ibid.
	65-69	0	0	ibid.
	70-74	0	0	ibid.
Marital Status	Married	1	0	ibid.
	Unmarried	0	1	ibid.
Education	Less than high school	0	0	ibid.
	High school graduate	0	1	ibid.
	Beyond high school	0	0	ibid.
	College graduate	1	0	ibid.
Ethnicity	White	0	0	ibid.
	Black	1	0	ibid.
	Other	0	1	ibid.
Parity	Nulliparous	1	0	Table 2 ^[2]
	Parous	0	0	ibid.
	Unknown	0	1	ibid.
Age at menarche, y	<12	1	0	ibid.
	12-13	0	0	ibid.
	≥ 14	0	1	ibid.
Age at menopause, y	<40	1	0	ibid.
	40-49	0	1	ibid.
	≥ 50	0	0	ibid.
Uterine fibroid	Ever	1	0	ibid.
	Never	0	1	ibid.
Oral contraceptive use	Ever	0	1	ibid.
	Never	1	0	ibid.
HRT use	Never	0	0	ibid.
	Former	0	0	ibid.
	Current	0	1	ibid.
Years since menopause	<5	1	0	ibid.
	5-9	0	0	ibid.
	10-14	0	1	ibid.
	≥ 15	0	0	ibid.
Smoking	Never	1	0	Table 3 ^[2]
	Former	0	0	ibid.
	Current	0	1	ibid.
Diabetes	Ever	1	0	ibid.
	Never	0	1	ibid.
Hypertension	Ever	1	0	ibid.
	Never	0	1	ibid.
BMI(kg/m ²)	<23.25	0	1	ibid.
	23.26-26.03	0	0	ibid.
	26.04-29.88	0	0	ibid.
	>29.89	1	0	ibid.

Table 4. Calculated α and Risk values (r) after testing procedure. Risk evaluation cut-offs: $r > 0.5243$ (see equation 20), with (SR) is denoted the significance of the r -value on the results

Case 1 ET \geq 3 mm		Case2 ET<3 mm	
a value	Risk (r)	a value	Risk (r)
1.0026	0.4987	1.6168	0.3093
0.5879	0.8505 (SR)	1.4302	0.3496
0.5576	0.8967(SR)	1.6838	0.2969
0.6940	0.7205(SR)	1.6016	0.3122
0.7235	0.6911(SR)	1.7070	0.2929
0.8110	0.6165(SR)	1.4666	0.3409
0.5785	0.8643(SR)	1.7212	0.2905
0.8863	0.5641(SR)	1.4508	0.3446
0.7988	0.6259(SR)	1.4529	0.3441
0.8989	0.5562(SR)	1.5335	0.3261
0.9775	0.5115	1.5757	0.3173
1.2320	0.4058	1.5848	0.3155
0.9070	0.5513(SR)	1.9438	0.2572
0.9305	0.5373(SR)	1.8062	0.2768
1.0237	0.4884	2.0298	0.2463
1.0499	0.4762	2.2013	0.2271
1.1010	0.4541	1.6479	0.3034
1.2859	0.3888	1.4229	0.3514
0.9522	0.5251(SR)	1.8480	0.2706
1.3873	0.3604	1.5611	0.3203
1.4638	0.3416	2.2064	0.2266
1.5450	0.3236	1.7212	0.2905
1.2567	0.3979	1.4550	0.3436
1.1536	0.4334	1.9642	0.2546
0.6449	0.7753(SR)	1.9565	0.2556
0.7969	0.6274(SR)	1.3756	0.3635
0.5959	0.8391(SR)	1.4763	0.3387
0.5081	0.9841(SR)	1.6913	0.2956
0.6277	0.7966(SR)	1.8933	0.2641
0.7914	0.6318(SR)	2.3924	0.2090
0.5186	0.9641(SR)	2.0323	0.2460
0.5206	0.8596(SR)	1.5340	0.3259
0.8504	0.5880(SR)	1.9163	0.2609
0.7749	0.6452(SR)	1.8769	0.2664
0.7553	0.6620(SR)	1.9476	0.2567
0.8352	0.5987(SR)	2.2190	0.2253
0.8817	0.5671(SR)	1.6714	0.2992
0.8478	0.5898(SR)	1.8322	0.2729
0.7128	0.7015(SR)	1.8367	0.2722
0.7786	0.6422(SR)	1.7356	0.2881
1.0023	0.4989	1.5956	0.3134
0.6789	0.7365(SR)	1.9980	0.2503
1.0003	0.4999	2.2309	0.2241
1.1235	0.4450	1.6112	0.3103
0.6323	0.7908(SR)	2.1231	0.2355
0.9320	0.5365(SR)	1.7820	0.2806
0.7329	0.6822(SR)	1.9674	0.2541
1.6902	0.2958	2.0323	0.2460
0.7372	0.6782(SR)	2.4323	0.2056
0.8329	0.6003(SR)	1.8361	0.2723

All the results of testing the Elman neural network are presented in detail in Table 4 and specifically in the column entitled, “a value”. The value specifically represents the reaction of the purelin function in an input testing vector. The purelin function, which was implemented via Eq. (14), compared the values of each vector with the weight matrixes w_1 and w_2 of the learning system, returning the value. Furthermore, the risk of each a value is calculated according to Eq. (15) (see Table 4).

4.2 Statistical evaluation

4.2.1 The cross-correlation procedure

In this stage, the extracted set case 1 ($\hat{C}x$) risk-coefficients (see Table 3) are submitted to the cross correlation r ^[30] along with another set ($\hat{C}y$) of risk-coefficients in order to control the degree of homogeneity of each set case in relation to other set case. The cross-correlation equation is described as follows:

$$r = \frac{\sum_{i=1}^{k-1} (\hat{C}x_i - \bar{\hat{C}x}_i)(\hat{C}y_i - \bar{\hat{C}y}_i)}{\sqrt{\sum_{i=1}^{k-1} (\hat{C}x_i - \bar{\hat{C}x}_i)^2 \sum_{i=1}^{k-1} (\hat{C}y_i - \bar{\hat{C}y}_i)^2}} \quad (17)$$

The extracted cross-correlation coefficient is a number between -1 and 1, which measures the degree to which two variable sets are linearly related. In our study, it is considered that the $\hat{C}y$ set has a perfect positive linear relationship with the set $\hat{C}x$ when the cross-correlation coefficient is approximately 1.

Taking into account Eq (1) the value $r = 0.0139$.

4.2.2 Tests of (least-squares) correlation coefficients

In this case, the t -value can also be converted to a Pearson r -value to measure the effect. Thus, the significance of r is determined in relation to the homogeneity per set. Then, the basic principle is to test the null hypothesis that the means of the two sets are equal and are randomly drawn from normal populations with equal variances. A t -statistic can be calculated as:

$$\sigma = \sqrt{\frac{(1-r^2)}{k-2}} \quad (18)$$

$$t = \frac{r - \mu}{\sigma} = \frac{r}{\sqrt{\frac{(1-r^2)}{k-2}}} = \frac{r\sqrt{k-2}}{\sqrt{1-r^2}} \quad (19)$$

The next step was to determine the appropriate value of the r coefficient in order to characterize it as a significant linear relationship between the correlated sets in our experiment. Thus, having $k = 44$, and the degree of freedom $v = k - 2 = 42$, $\alpha = 0.01$ and thus the critical $t_{\alpha/2} = 3.538$ was chosen. Then the significant value of r was calculated as follows:

$$r = \frac{t_{a/2}}{\sqrt{t_{a/2}^2 + k - 2}} = 0.5243 \quad (20)$$

In conclusion, in this case, r may be characterized as significant when the null hypothesis is rejected ($1 \leq r \leq 0.5243$), where the value 1 corresponds in absolute probability related with case 1 ($ET \geq 3$ mm).

Thus, using the risk values and taking into account $r=0.523$, Table 3 is modified in Table 4 as follows:

Table 5. Cross-correlation Risk coefficients (r)

Case 1 ($ET \geq 3$ mm)		Case 2 ($ET < 3$ mm)	
Significantly Related	Non-significantly Related	Significantly Related	Non-significantly Related
34	16	0	50

The results of the evaluation in relation to the probability (risk) of the method showed agreement with the literature research [2] and particularly in Case 2 of the non-association (non-significantly relevant) with abnormal incidents. In addition, and in case 1 the evaluation results showed an agreement with the research in relation to the association between endometrial thickness and risk factors [2].

5 Conclusions

This study focused on the classification problem of high-dimensional patterns and especially of socio-demographic cancer questionnaires. The purpose of this study was to define a predictive indicator of a published clinical research study regarding the influence of Hormone Replacement Therapy (HRT) on the growth of different cancers (e.g., breast, ovarian, endometrial, and colon). This study combined independent factors from the clinical research using a Bayesian model in order to achieve normalizing data linked by significant relevant properties of these factors. The goal was to determine a standard threshold value at which an independent self-organizing system would decide the correlation between the normalization of the data in the preprocessing stage via a well-fitted, recurrent Elman neural network using a threshold factor (distance value). For a case study of this implementation, a dataset of published clinical research was used and the evaluated procedure was implemented by a well-fitted t -test control. The statistical results showed that case data with a significant probability $ET \geq 3$ mm could be trusted to complete a questionnaire at first grade preventive medicine by women. For this purpose, an online questionnaire based on the proposed methodology was created in the Laboratory Information Technologies of the Ionian University. [http://lit.ionio.gr/index.php?option=com_content&view=article&id=69&Itemid=101&lang=el].

Future research would entail the application of the proposed methodology to several other epidemiological questionnaires.

References

- [1] Cruz, J.A., Wishart, D.S. Applications of machine learning in cancer prediction and prognosis. *Cancer Inform.* 2006; 2: 59.
- [2] Sit, A.S.Y., Modugno, F., Hill, L.M., Martin, J., Weissfeld, J.L. Transvaginal ultrasound measurement of endometrial thickness as a biomarker for estrogen exposure. *Cancer Epidemiol Biomarkers Prev.* 2004; 13: 1459-1465. PMID:15342446
- [3] Yu, L., Wang, S., Lai, K.K. An integrated data preparation scheme for neural network data analysis. *IEEE Knowl Data Eng.* 2006; 18: 217-230. <http://dx.doi.org/10.1109/TKDE.2006.22>
- [4] Hu, X. DB-H reduction: a data preprocessing algorithm for data mining applications. *App Math Lett.* 2003; 16: 889-895. [http://dx.doi.org/10.1016/S0893-9659\(03\)90013-9](http://dx.doi.org/10.1016/S0893-9659(03)90013-9)
- [5] Sattler, K.U., Schallehn, E. A data preparation framework based on a multidatabase language, in: *Proceedings of the International Symposium on Database Engineering and Applications.* 2001: 219-228.

- [6] Pastor-Satorras, R., Vespignani, A. Epidemics and immunization in scale-free networks, in: Bornholdt, S. and Schuster, H.G. (eds.), Handbook. 2003.
- [7] Poulos, M., V. Belesiotis, & N. Alexandris. A Classroom Observation Model Fitted to Stochastic and Probabilistic Decision Systems. *Artificial Intelligence Applications and Innovations*. 2010: 30-36.
- [8] Tsumoto, S. Mining diagnostic rules from clinical databases using rough sets and medical diagnostic model. *Inf Sci*. 2004; 162: 65-80. <http://dx.doi.org/10.1016/j.ins.2004.03.002>
- [9] Moore, J.H., Williams, S.M. New strategies for identifying gene-gene interactions in hypertension. *Ann Med*. 2002; 34: 88-95. PMID:12108579 <http://dx.doi.org/10.1080/07853890252953473>
- [10] Ritchie, M.D., Motsinger, A.A., Bush, W.S., Coffey, C., S., Moore, J.H. Genetic programming neural networks: a powerful bioinformatics tool for human genetics. *Appl Soft Comp*. 2007; 7: 471-479. PMID:20948988 <http://dx.doi.org/10.1016/j.asoc.2006.01.013>
- [11] Eddy, S. R. 'What is Bayesian statistics? *Nature biotechnology*. 2004; 22(9): 1177-1178. PMID:15340486 <http://dx.doi.org/10.1038/nbt0904-1177>
- [12] Poulos, M., S. Papavaslopoulos, N., Alexandris, E., Vlachos. Comparison between auto-cross-correlation coefficients and coherence methods applied to the EEG for diagnostic purposes. *Medical Science Monitor*. 2004; 10(10): 99-108.
- [13] Boukadoum, M., Lounis, H., Houari, G.M.,Sahraoui, A. Siveton, V. A comparison between using a neural network and a fuzzy regression system to predict the values of hydro power system variables. In: *Proceedings of the International Conference on Artificial Intelligence and Applications (AIA)*. 2002.
- [14] Iyer, V. R. Response of human fibroblasts to serum. *Science*. 1999; 283: 83-87. PMID:9872747 <http://dx.doi.org/10.1126/science.283.5398.83>
- [15] Nolfi, S., Parisi, D., Elman, J.L. Learning and evolution in neural networks. *Adapt Behav*. 1994; 3: 5-28. <http://dx.doi.org/10.1177/105971239400300102>
- [16] Elman, J.L. Learning and development in neural networks: the importance of starting small. *Cognition*. 1993; 48: 71-99. [http://dx.doi.org/10.1016/0010-0277\(93\)90058-4](http://dx.doi.org/10.1016/0010-0277(93)90058-4)
- [17] Casella, F. Neural network library in Modelica. In: *Proceedings of the 5th International Modelica Conference*. Vienna, Austria. 2006: 549-557.
- [18] Engelbrecht, A P., Cloete, I. Selective learning using sensitivity analysis, in: *Neural Networks Proceedings, 1998. IEEE World Congress on Computational Intelligence. The 1998 IEEE International Joint Conference on*. 1998; 2: 1150-1155.
- [19] Pinker, S. How the mind works. *Ann N Y Acad Sci*. 1999; 882: 119. PMID:10415890 <http://dx.doi.org/10.1111/j.1749-6632.1999.tb08538.x>
- [20] Wong, F., Park, K.H., Bien, Z. A skin color model parameter prediction by using Elman network, in: *Proceedings of the 6th Symposium on Advanced Intelligent Systems*. Korea. 2005: 704-709.
- [21] O'Reilly, R.C., Munakata, Y. *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain*. MIT Press. 2000.
- [22] Specht, D.F. Probabilistic neural networks. *Neural Netw*. 1990; 3: 109-118. [http://dx.doi.org/10.1016/0893-6080\(90\)90049-Q](http://dx.doi.org/10.1016/0893-6080(90)90049-Q)
- [23] Valle-Lisboa, J.C., Reali, F., Anastasia, H., Mizraji, E. Elman topology with sigma-pi units: an application to the modeling of verbal hallucinations in schizophrenia. *Neural Netw*. 2005; 18: 863-877. PMID:15935616 <http://dx.doi.org/10.1016/j.neunet.2005.03.009>
- [24] Rodriguez, P., Wiles, J., Elman, J.L. A recurrent neural network that learns to count. *Connect Sci*. 1999; 11: 5-40. <http://dx.doi.org/10.1080/095400999116340>
- [25] Papavaslopoulos, S., Poulos, M., Korfiatis, N., Bokos, G. A non-linear index to evaluate a journal's scientific impact. *Inf Sci*. 2010; 180: 2156-2175. <http://dx.doi.org/10.1016/j.ins.2010.01.018>
- [26] Trujillo-Ortiz, A., Hernandez-Walls, R., Trujillo-Perez, F.A., Castro-Castro, N., 2009. Bayes' theorem: the discrete case. Available from: <http://www.mathworks.com/matlabcentral/fileexchange/25203>. Accessed on DATE.
- [27] Davison, A.C., Hinkley, D.V. *Bootstrap methods and their application [software]*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press. 1997.
- [28] Cheng, Q., Agterberg, F.P. Fuzzy weights of evidence method and its application in mineral potential mapping. *Nat Resour Res*. 1999; 8: 27-35. <http://dx.doi.org/10.1023/A:1021677510649>
- [29] Brown, W., Groves, D., Gedeon, T. Use of fuzzy membership input layers to combine subjective geological knowledge and empirical data in a neural network method for mineral-potential mapping. *Nat Resour Res*. 2003; 12: 183-200. <http://dx.doi.org/10.1023/A:1025175904545>
- [30] Morrison, N., Donald, F., 1976. *Multivariate Statistical Methods*. McGraw-Hill Book Company, New York.