A variable structure fuzzy neural network model of squamous dysplasia and esophageal squamous cell carcinoma based on a global chaotic optimization algorithm

Motahareh Moghtadaei, Mohammad Reza Hashemi Golpayegani, Reza Malekzadeh

Complex Systems and Cybernetic Control Lab., Faculty of Biomedical Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran, P.O. Box 1591634311, Iran
Digestive Disease Research Center, Shariati Hospital, Tehran University of Medical Sciences, Tehran, Iran

HIGHLIGHTS
- An algorithm for identification of the risks of squamous dysplasia, and ESCC.
- FNN predicts risk of cancer using the risk factors obtained from case control study.
- Limit the clinical screenings to people with higher risks of cancer.
- Helps in huge cost saving in early detection of esophageal cancer.

ABSTRACT
Identification of squamous dysplasia and esophageal squamous cell carcinoma (ESCC) is of great importance in prevention of cancer incidence. Computer aided algorithms can be very useful for identification of people with higher risks of squamous dysplasia, and ESCC. Such method can limit the clinical screenings to people with higher risks.

Different regression methods have been used to predict ESCC and dysplasia. In this paper, a Fuzzy Neural Network (FNN) model is selected for ESCC and dysplasia prediction. The inputs to the classifier are the risk factors. Since the relation between risk factors in the tumor system has a complex nonlinear behavior, in comparison to most of ordinary data, the cost function of its model can have more local optimums. Thus the need for global optimization methods is more highlighted. The proposed method in this paper is a Chaotic Optimization Algorithm (COA) proceeding by the common Error Back Propagation (EBP) local method. Since the model has many parameters, we use a strategy to reduce the dependency among parameters caused by the chaotic series generator. This dependency was not considered in the previous COA methods. The algorithm is compared with logistic regression model as the latest successful methods of ESCC and dysplasia prediction. The results represent a more precise prediction with less mean and variance of error.

1. Introduction
An estimated 482,300 new esophageal cancer cases and 406,800 deaths occurred in 2008 worldwide (Jemal et al., 2011). Esophageal cancer is among the 5 Leading Cancers in the United States. The estimated numbers of new cases and death occurrences of esophageal cancer expected among men and women is 17,460 and 15,070 respectively, in the United States in 2012 (Siegel et al., 2012).

Based on the global cancer statistics in 2011, 90% of esophageal cancer cases in the highest risk area, stretches from northern Iran through the central Asian republics to North-Central China occurs as squamous cell carcinoma in the middle or upper one-third of the esophagus (Jemal et al., 2011) that is called esophageal squamous cell carcinoma (ESCC). In recent years, incidence rates for ESCC have been steadily declining in several western countries but it has been increasing in certain Asian countries (Jemal et al., 2011).

The five-year survival of ESCC is less than 20% in the United States (Jemal et al., 2010). In Golestan Province, five-year survival of ESCC is only 3.3% with the median survival of only 7 months (Etemadi et al., 2012).

HIGHLIGHTS
- An algorithm for identification of the risks of squamous dysplasia, and ESCC.
- FNN predicts risk of cancer using the risk factors obtained from case control study.
- Limit the clinical screenings to people with higher risks of cancer.
- Helps in huge cost saving in early detection of esophageal cancer.

ARTICLE INFO

Article history:
Received 28 July 2012
Received in revised form 17 October 2012
Accepted 10 November 2012
Available online 19 November 2012

Keywords:
Cancer
Esophagus
Early detection
Classification
Risk factors
great importance. On the other hand, squamous dysplasia is the precursor lesion for ESCC (Wang et al., 2005) and can be detected by endoscopy (Dawsey et al., 1998).

Since endoscopic screening of the whole population is not practical, finding ways to detect people with higher risks of squamous dysplasia, and ESCC can help to limit the clinical screenings. But that is not easy because tumor system is very complicated and is capable of changing its properties. The reason is that the tumor system involves different levels of gene, molecular, cellular, tissue, organ, body and population interacting via various complicated signal transduction pathways (Lee et al., 2003).

Using mathematical modeling to accurately predict cancer initiation can be very helpful. A model can estimate the complicated, unknown dependency between risk factors as inputs and the possibility of occurrence of squamous dysplasia, and ESCC as the output. First an appropriate model is developed using available data and then such a model could be used to specify high risk individuals in other groups minimizing the cost of cancer screening. In other words, the model is used to predict the output, i.e. cancer risk, associated with a new set of input values, i.e. risk factors.

In a previous study by Etemadi et al., 2012, statistical models have been used for modeling and prediction of squamous dysplasia, and ESCC using their risk factors. But in fact, for most practical applications relating complex natural and biological phenomena, like occurrence of squamous dysplasia, and ESCC, statistics is imprecise and cannot be helpful for several reasons. First of all, although statistical analysis is used for modeling unexplained variations in the data, in biological applications, events often consist of a sequence of possible but not probable occurrences. Second, in the statistical approaches, it is necessary to calculate the probability distributions. While in complex biological applications the a priori probabilities and the state conditional densities of classes are unknown, and usually the data cannot represent the probability distributions. That is why statistical methods are not successful in predicting the risk of squamous dysplasia and ESCC.

For other types of cancers, like lung cancer (Blatt et al., 2007; Di Natale et al., 2003), breast cancer (Das and Battacharya, 2008), and liver cancer (Huiyan et al., 2010) non statistical models for cancer identification using risk factors has been studied. High performance of non-statistical methods in cancer diagnosis and prediction has been proved by many recent articles (Feng et al.; Hsieh et al., 2012; Naghibi et al., 2012; Saritas et al., 2012).

Several chaotic prediction methods have been developed to predict complex phenomena, such as the local-region method, adaptive rational function filter, and the artificial neural network method (Yang et al., 2008). Among all, taking advantages of nonlinear learning and noise tolerance, neural networks are the most appropriate tools (Rabuňál and Dorrado, 2006).

Particularly, the Fuzzy Neural Network (FNN) is a better approximator with advantages such as simplicity of use and less computational demand. Since fuzzy logic is based on approximate reasoning and computational intelligence, the FNN has an advantage in dealing with the nonlinear time series (Senju et al.). Usefulness and success of neuro fuzzy techniques has been studied for breast cancer detection (Das and Battacharya, 2008).

After constructing the FNN structure, the next step is to find a proper training method. The prevalent method for training FNN is Error Back Propagation (EBP) (Alpaydın, 2004). EBP algorithm (Russell and Norvig, 2010) is based on gradient descent method, which depends on the initial weights and thus it may converge to a suboptimal point. A number of different evolutionary algorithms have been proposed to improve the capability of the neural network. Being less sensitive to local minima, the commonly used global optimization methods, Genetic Algorithm (GA) and Particle Swarm Optimization (PSO), are much more effective than traditional gradient-based methods such as EBP (Wu, 2011; Guo and Dong, 2011; Tong and Mintram, 2010). A GA based neuro fuzzy method (Das and Battacharya, 2008) and a PSO based neural network method (Zhang et al., 2012) have been used for breast cancer classification recently. But in fact GA and PSO methods are not suitable in problems with large variations and complexities. In such problems, using GA method requires large population size that causes the rate of convergence to be too slow. On the other hand, because of the problem’s complexity, the conventional PSO has longer computation time and may easily get trapped in a local minimum in dealing with complex problems, which implies it cannot be used efficiently in such complicated optimization problem. In order to enhance the performance of the conventional PSO, Liang Zhao and Yupu Yang created a Cooperative Random learning Particle Swarm Optimization (CRPSO) by increasing the number of sub-swarms and get the trade-off between local and global search (Zhao et al., 2008). The undeniable disadvantages of CRPSO encountering with complex problem is a relatively long computation time and a discontinuous search space. For complex problems, it has to increase the number of sub-swarms and so it makes the optimization algorithm useless for online applications.

Beyond all discussed problems, discontinuity of the search region restricts the accuracy in all mentioned methods. Recently a more efficient strategy called Chaotic Optimization Algorithm (COA) has been proposed and successfully applied to some typical nonlinear global optimization problems (Yang et al., 2007). Unrepeatability of COA makes it optimum and continuity of the search region makes it precise but too slow. To solve the latter problem, we propose using EBP in combination with chaotic optimization in this article. In this way, EBP helps in initializing of the search bounds for COA.

In Section 2, we introduce the FNN model based on hybrid method of COA in cooperation with EBP. In Section 3, the application of the method for prediction of the risk of squamous dysplasia and ESCC is discussed. The results are represented and its performance is compared with logistic regression method, in Section 4. In Sections 5 and 6, we have discussion and conclusions respectively.

2. Data description

Two datasets are used in this study. The first dataset is relating to ESCC. This dataset is from the Golestan Case-Control Study that was conducted from 2003 to 2007 (Nasrollahzadeh et al., 2008). The ESCC dataset includes 300 biopsy-proven ESCC cases and 571 age and sex-matched neighborhood controls.

The second dataset is relating to dysplasia. This dataset is collected from individuals visiting Atrak Clinic, a gastroenterology research clinic in Gonbad City, Golestan Province, between 2002 and 2007 (Nasrollahzadeh et al., 2008). The dysplasia dataset includes 26 individuals with dysplastic lesions and 698 controls. Video endoscopy with Lugol’s iodine staining, questionnaire and biopsies helped to develop this dataset.

In each of these, the data set includes all risk factors known in this region to be associated with dysplasia and ESCC. Based on the previous studies (Nasrollahzadeh et al., 2008; Abnet et al., 2008; Islami et al., 2009; Islami et al., 2009; Akbari et al., 2006), these risk factors included: age, ethnicity, place of residence, tobacco smoking, opium use, socio-economic status, oral health, family history, tea temperature, and water source.

The conduct of studies performed to obtain the dataset including risk factors were reviewed and approved by the Institutional Review Boards of the Tehran University Digestive
Disease Research Center (DDRC), the US National Cancer Institute (NCI), and International Agency for Research on Cancer (IARC) (Nasrollahzadeh et al., 2008).

Place of residence of patients include the counties of Gonbad, Minoodasht, Kalaleh, Azadshahr, and Ramian in eastern Golestan province. Approximately half of the residents of the study area are of Turkmen ethnicity, and the rest Persians, Kurds, Turks, and others (Nasrollahzadeh et al., 2008).

For opium use, and tobacco smoking, average intensity, total duration, and cumulative use (average intensity multiplied by duration of use) were categorized into three groups: no use, low use (< median use in controls), and high use (> median use). Age of first use was similarly categorized (Nasrollahzadeh et al., 2008).

The socio-economic status indicators that were included in the dataset include education level, head of the household’s education level, marital status, number of first degree relatives, relatives and family structure. In order to create a wealth score for wealth of each individual, based on appliances and other variables, a multiple correspondence analysis (MCA) is utilized on house ownership, house structure, house size in square meters (m²), number of people living together in the current house, ownership of household appliances, including bath in the residence, personal car, motorbike, black and white TV (B/W TV), color TV, refrigerator, freezer, vacuum and washing machine, and the duration of owning these appliances, and current job or the most recent occupation for retired or disabled subjects (Islami et al., 2009).

To describe the oral health in terms of numbers, each patient’s teeth, the number of decayed, missing, or filled teeth were counted, and the age of first adult tooth loss, oral hygiene habits (tooth brushing or rinsing the teeth with salt water), and denture use were recorded. All oral hygiene practices other than at least daily brushing were lumped into a single category due to small numbers (Abnet et al., 2008).

The data relating the family history contains information on all of the first- and second-degree relatives and first cousins. Information was obtained regarding the vital status of these family members and all occurrences of esophageal cancer and other cancers. Current age, age at diagnosis of cancer, site of cancer, age of death, clinical and pathological diagnosis of cancer were recorded for all first-degree relatives. The presence of parental consanguinity was recorded for cases and controls (Akbari et al., 2006).

The amount of consumed tea, tea steeping time, and tea temperature degree that was estimated by the time interval between tea being poured and drunk were also recorded for each case (Islami et al., 2009). Patients’ water source, access to piped water, and years having access to piped water were also recorded (Nasrollahzadeh et al., 2008).

In general in order to express the seven mentioned risk factors in terms of numbers, there is a total of 90 and 77 numbers recorded for each subject in the ESCC and dysplasia datasets respectively.

To better describe the datasets, three important histograms for describing these risk factors in the dataset. These 90 numbers are set according to the membership function described in Eq. (1).

Where $A_{ji}(x)$ are the available fuzzy sets of the model and $A_{ji}(x_i)$ gives the membership value of the input $x_i$ to the $j$th fuzzy set according to the membership function described in Eq. (1).

The parameters $c_j$ and $\sigma_j$ are mean and standard deviation for the $j$th fuzzy membership function, respectively. The third and fourth layers are used to implement the fuzzy rules of the TSK fuzzy model.

The output of the third layer is:

$$\mu_j = \prod_{i=1}^{I} A_{ji}(x_i)$$

In the fourth layer, $\pi_j$ which is the normalized value of $\mu_j$ is multiplied by the function $f_j$ of the consequent part of the rule. The output layer, i.e. fifth layer, has one node and the output of this layer is the output value of the TSK fuzzy model. The output value $y$ is the class assigned to the control and case subjects, i.e. 0 and 1 respectively.

$$y = \sum_{j=1}^{M} \pi_j f_j$$

So the main part of the procedure is to find the optimum values for these parameters $(f,c,\omega)$ considering the proper computational time for online application.

### 3.2. Hybrid chaotic optimization algorithm

Chaos optimization algorithm as a novel method of global optimization has attracted much attention (Yang et al., 2007). In principle, chaos optimization always reaches to the global minimum and more easily escapes from local minima than other stochastic optimization algorithms. Searching the global optimum by the chaos optimization algorithm, utilizes the nature of chaos sequences such as pseudo randomness, ergodicity and irregularity (Yang et al., 2007).

The proposed algorithm in this paper uses a hybrid method to enhance the effective computation. This approach includes two major steps. Firstly, the conventional EBP is employed to initialize search boundaries for COA search.

Secondly, COA is used for global search. In this part a chaotic sequence is defined based on the Logistic map in its chaotic mode as described in Eq. (4), in which $v_k$ is the value of the Logistic sequence in the $k$th iteration.

$$v_{k+1} = 4v_k(1-v_k)$$

The generated sequence of chaotic points is mapped to a sequence of parameter values of the FNN model. Then, the objective function is calculated with respect to the generated parameter values, and the values with the minimum objective function are chosen as the current optimum.

### 3.3. The algorithm description:

1. Constructing the $I$-input single-output FNN as described above.
2. Training the FNN by EBP method as described in Eq. (5) until reaching an RMS error value that cannot be reduced more.
than a predetermined threshold, ε, in t iterations.

\[
\begin{align*}
    f_{ij}^{n+1} & = f_{ij}^{n} \cdot \frac{x^T(y^n - T)}{b^n} \mu_j^n \\
    o_{ij}^{n+1} & = o_{ij}^{n} \cdot \frac{x^T(y^n - T)}{b^n} \mu_j^n \frac{(2(x - c_{ij}^n))^2}{(c_{ij}^n)^2} \\
    c_{ij}^{n+1} & = c_{ij}^{n} \cdot \frac{x^T(y^n - T)}{b^n} \mu_j^n \frac{(2(x - c_{ij}^n))^2}{(c_{ij}^n)^2}
\end{align*}
\]  

(5)

In which \( f_{ij}, o_{ij}, c_{ij}, \mu_j, \) and \( b^n \) are the values of \( f_{ij}, o_{ij}, c_{ij}, \mu_j, \) and \( b \) in the \( n \)th iteration, \( x \) is the step size, \( T \) is the target output, \( x \) is the input, and \( b \) is defined as:

\[
b = \sum_{j=1}^{M} \mu_j \]

(6)

3. Continuing the training stage by switching to the chaotic optimization search as the steps below:

I. Considering the latest EBP weights as the initial conditions:

\[
\begin{align*}
    F^* & = F_{\text{EBP}} \\
    W^* & = W_{\text{EBP}} \\
    C^* & = C_{\text{EBP}}
\end{align*}
\]

(7)

II. Producing chaotic search space for renewing the weights around the previous most optimum weights, i.e. \( F^*, W^*, C^* \) described as

\[
\begin{align*}
    F_{n+1} & = F_n + z_1(V_1 - 0.5) \\
    W_{n+1} & = W_n + z_2(V_2 - 0.5) \\
    C_{n+1} & = C_n + z_3(V_3 - 0.5)
\end{align*}
\]

(8)

where \( F_{n+1}, W_{n+1}, \) and \( C_{n+1} \) are the values of \( f, \omega \) and \( c \) respectively in the \( n \)th iteration. In Eq. (8), \( V_1, V_2, V_3 \) are the vectors that contain the sequential numbers generated by Logistic function (Eq. 4) in its chaotic state and the adaptive parameters \( z_1, z_2, z_3 \) are the real numbers associated to the required bound for the search space; \( z_1, z_2, z_3 \) determine the length of steps for regenerating the new weights toward the optimum value. In other words, they play the same role of the step size in EBP algorithm. On the other hand, it can be easily verified that the logistic

\[ \text{Fig. 1. Histograms of: (a) age, (b) frequency of brushing teeth (1: Once a day, 2: Twice a day, 3: Three time a day, 4: Other (specify), 5: Never), and (c) tea temperature (estimated by the time interval between tea being poured and drunk, and categorized in minutes as 1: 4 min or more, 2: 2–3 min, and 3: 0–1 min), for ESCC dataset (total number of ESCC cases = 300, total number of control subjects = 571).} \]
Fig. 2. Histograms of (a) age, (b) frequency of brushing teeth (1: Once a day, 2: Twice a day, 3: Three times a day, 4: Other (specify), 5: Never), and (c) tea temperature (estimated by the time interval between tea being poured and drunk, and categorized in minutes as 1: 4 min or more, 2: 2–3 min, and 3: 0–1 min) for dysplasia dataset (total number of dysplasia cases = 26, total number of control subjects = 698).

Fig. 3. Structure of the fuzzy neural network.
probability density function (PDF) is symmetric (Yang et al., 2007) as shown in Fig. 4. So, each value of the sequence generated by the Logistic map belongs to any of the intervals [0,0.5] or [0.5,1] with equal possibility. For this reason, the additional terms to the old weights in the sequence generated by the Logistic map belongs to any of the intervals [0,0.5] or [0.5,1] with equal possibility. For this reason, the additional terms to the old weights in the right sides of Eq. (8), create both the positive and negative numbers with the chaotic turbulence for exiting from the probably local minimum of the error, in every possible direction.

There is one important issue that should be considered here. As shown in Fig. 4, although the PDF values of the Logistic chaotic sequence are nearly uniform around the middle of the interval [0,1], they are much larger near two ends. That makes the search not to be beneficial because numerous searches are performed close to the two ends. It is more critical if the global optimum does not locate near the two ends. For this reason using the idea described in (Yang et al., 2007), we have thrown away the intervals near the two ends properly.

III. Calculating the new error value, $E^{n+1}$:

$$E^{n+1} = \frac{\left( \sum_{p=1}^{N} (y_{n+1}^p - T^p)^2 \right)}{N}$$

where, $N$ is the number of subjects in the train set, $y_{n+1}^p$ is the model output for the $p$th subject in the $n+1$th iteration, and $T^p$ is the target output for the $p$th subject which is 1 for patient subjects and 0 for control subjects.

a) If $E^{n+1} \leq E^n$, then $E^n = E^{n+1}$, $P^n = P^{n+1}$, $W^n = W^{n+1}$, $C^n = C^{n+1}$.

b) If $E^{n+1} > E^n$, stop the Algorithm; else repeating the algorithm from II ($\varepsilon_0$ is a predetermined small acceptable value for error).

To verify how the classification results will generalize to an independent data set, testing datasets are used by applying a cross validation method, the 10-fold algorithm. We can estimate how accurately a predictive model will perform in practice. In the 10-fold algorithm, the original sample is randomly partitioned into 10 subsamples. The training and testing procedures are repeated 10 times (folds). In each fold, 90% of the data is used for training, and the rest 10% of the data is used for testing the performance of the algorithm. Each of the 10 subsamples are used exactly once as the validation data (Geisser, 1993). This way one can make sure that the classifier can generalize classification for different data sets of the same type. On the other hand the test dataset helps to avoid overfitting during training procedure. In principle, while training the model, first the prediction error for the test data set decreases, but at the point that overfitting begins, this error suddenly increases. The training procedure should be stopped at this point.

Recall that in this study, the unknown dependency between risk factors, as inputs, and the possibility of occurrence of squamous dysplasia and ESCC, as the output, is approximated using a fuzzy model with $I$ inputs. The second to fourth layers model the fuzzy rules and the fuzzy sets are regarded as the connection weights. The membership functions of the FNN are modeled in the second layer where respectively, $c^j$ and $\omega_0$ are the mean (center) and standard deviation for the $j$th Gaussian membership function that is connected to the $i$th input as described in Eq. (1).

Since the result is highly influenced by the structure of the model, i.e. number of hidden neurons, choosing the best structure is very important. For a complex system, it is much better to use a variable structure model. This way, during the training procedure the model is adapted to the most efficient structure. We follow this strategy in finding the most efficient structure for the model in this study.

The $I \times 1$ vectors $x_i$ and $x_0$ denote the lower and upper bounds of the input data, $x$, respectively. The lower bound is defined as the minimum values of $x$, and the upper bound is defined as the maximum values of $x$ as described in Eq. (10), where $N$ is the total number of cases in the train set.

$$\begin{align*}
    x_{0i} &= \min\{x_i \text{ for all } N\} \\
    x_{ui} &= \max\{x_i \text{ for all } N\} \quad i = 1,2,\ldots,I
\end{align*}$$

These vectors show the approximation region of the second layer of the FNN, because all of the data values lay in the region formed between the lower and upper bound of data, i.e. $x_0$ and $x_u$.

As defined in Eq. (1), each Gaussian membership function in the second layer of the FNN is described by its mean (center) and standard deviation, i.e. $c$ and $\omega$. To locate the centers of fuzzy membership functions inside the approximation region of $x$, a two dimensional variable structure grid is used, i.e. the number of nodes is variable and can be changed during the constructive and pruning phases in the train algorithm as will be explained later. Each grid node corresponds to the center of one fuzzy membership function. For this reason, the grid is called the center grid. The center grid is initialized inside the approximation region of $x$ with $2I$ boundary grid nodes as $c_{ij} = (c_{i1}, c_{i2}) = (x_{ij}, x_{ui})$, $i = 1,\ldots,I$, $j = 1,2$. These boundary grid nodes cannot be removed because they form the boundary of the approximation region which does not change during the training procedure. In each coordinate, additional grid nodes, i.e. fuzzy membership functions, will be added into and then can be removed during the training procedure as described below. The initial grid is shown in Fig. 5. Note that the operations of adding and removing the fuzzy membership functions are applied to each coordinate independently.

a) Constructive phase: If $E^* > \varepsilon_0$, $E^* > \varepsilon_0$, and $E^{n+1} - E^n < 0$, we find that the number of fuzzy membership functions is insufficient. At this time, a fuzzy membership function should be added to the network by adding a new grid node. To do that, first, the nearest grid node in the center grid to the current input, $X$, is determined, that is $c_{ij}$.

![Fig. 4. PDF of the logistic chaotic sequence.](image)
Then, the grid node \( c_{i,j} \) is found, in a way that \( \mathbf{X} \) is located between \( c_{i,j} \) and \( c_{i,j} \). Next, the adding operation is performed for each coordinate independently, provided that:

1. \( |x_i - c_{i,j}| > |c_{i,j} - c_{i,j}|/4; \)
2. \( |x_i - c_{i,j}| > d_{\text{threshold}}, \) where \( d_{\text{threshold}} \) is a parameter that determines the minimum grid distance in the \( i \)th coordinate.

After the new structure is defined, the training phase, i.e. stages 2 and 3 of the algorithm are repeated.

b) Pruning phase: When \( E \leq c_0 \), to prevent network redundancy, the least significant fuzzy membership functions of the existing network are removed by removing the related grid node. The grid node located at \( c_{i,j} \) is removed if:

1. \( c_{i,j} \notin \{x_i,x_n\}; \)
2. The grid node in the \( i \)th coordinate with its coordinate equal to \( c_{i,j} \) is in the higher than or in the same layer as the highest layer of the two neighboring grid nodes in the same coordinate;
3. \( |x_i - c_{i,j}| < \tau |c_{i,j} - c_{i,j}|, \) for \( \tau \in (0.0.5). \)

After the new structure is defined, the training phase, i.e. stages 2 and 3 of the algorithm are repeated.

c) When \( E^* \) increases to \( E^* > c_0 \), the previous structure with \( E^* < c_0 \) is restored and the training is terminated.

The schematic of the grid after performing a constructive and a subsequent pruning phase is demonstrated in Fig. 6.

4. Results

The training error of the FNN by EBP method, at the first step of our methodology was reached to some value and remained without more reduction after considerable iterations. Although the prediction result of the FNN on the test set after training with the EBP method is considerable, but FNN’s weights can be assumed as an estimation of the desired weights. These weights are used as the initial values for the next step of training, which is COA. As mentioned earlier, the COA convergence is not dependent on its initial condition but searching around the target would lead to more efficient and faster search. That’s why the EBP algorithm was used and the final weights of the network due to the EBP are the initial condition of the COA.

The accuracy, specificity, precision and sensitivity of the model for ESCC and dysplasia risk prediction for the train and test datasets are represented in Tables 1–4.

![Fig. 5. Schematic of initial center grid of the FNN with 21 nodes.](image)

**Table 1**

<table>
<thead>
<tr>
<th>Accuracy (%)</th>
<th>Specificity (%)</th>
<th>Precision (%)</th>
<th>Sensitivity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

The accuracy, specificity, precision and sensitivity are defined as:

\[
\text{accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN + TN)}
\]

(11)

\[
\text{precision} = \frac{TP}{(TP + FP)}
\]

(12)

\[
\text{specificity} = \frac{TN}{(TN + FP)}
\]

(13)

\[
\text{sensitivity} = \frac{TP}{(TP + FN)}
\]

(14)

In which, \( TP \) stands for the number of true positives, i.e. sick people correctly diagnosed as sick, \( FP \) stands for the number of false positives, i.e. healthy people incorrectly identified as sick, \( TN \) stands for the number of true negatives, i.e. healthy people correctly identified as healthy, and finally \( FN \) stands for the number of false negatives, i.e. sick people incorrectly identified as healthy.

The Receiver Operating Characteristic (ROC) diagrams for ESCC and dysplasia risk prediction are presented in Figs. 7 and 8 respectively.

An ROC space is defined by false positive rate (FPR) and true positive rate (TPR) as \( x \) and \( y \) axes respectively, which depicts relative trade-offs between true positive (benefits) and false positive (costs). To draw an ROC curve, after the training phase is finished, the TPR and FPR of the prediction of the model for train and test datasets are calculated, and the TPR is plotted versus FPR.

A comparison of the results of the proposed method with the results of multivariate logistic regression models to predict ESCC is managed in Table 5.

![Fig. 6. Schematic of the center grid after a constructive and a following pruning phase. The black nodes are added to the grid in the constructive phase and the node marked with a square is removed from the grid in the pruning phase.](image)
5. Discussion

In this paper, a variable structure FNN model based on a hybrid training method consists of EBP and COA for ESCC and squamous dysplasia risk prediction was proposed.

ANN is commonly used for classification due to advantages such as nonlinear learning and noise tolerance. Among all ANNs, a FNN as a classifier can approximate the relation between inputs and outputs with any degree of accuracy, given enough basis functions with shorter computation time. On the other hand finding the optimum structure of the model is of great importance. So a variable structure FNN model was chosen in this paper.

Although this approach uses a black box model and is not related to pathophysiology of the disease, but as a machine learning technique, it can estimate the unknown dependency between the set of given input variables, i.e. the risk factors and its response variable, i.e. 0 for patients and 1 for healthy subjects. This dependency forms the model that is used to predict the future output associated with a different set of input values, e.g. test datasets.

The reproducibility and generalization of the classifier is verified by checking its performance on an independent dataset that can well represent the population, and yet sufficiently distinct from the train dataset, i.e. the test datasets by applying a cross validation strategy, i.e. the 10-fold algorithm.

Over-fitting to the training dataset is a major challenge in machine learning methods. In this study, we have used early stopping technique based on the test dataset classification error in each fold of the 10-fold algorithm.

While the number of cases in the first dataset relating ESCC is 300, in the total number of 871, in the second dataset that is relating to dysplasia, the number of cases with dysplasia is only 26 in the total number of 724, which is a very small fraction. It is clear that using this data in the training phase, the model cannot identify cases from controls properly. For this reason, the data relating cases in the dysplasia dataset were repeated twelve times to have a dataset with 312 cases and 698 controls. Since in the second dataset the number of cases is very small, the probability distribution of cases cannot be determined with good accuracy, and that is why regression models have problem in modeling and predicting this data, and repeating the cases in the dataset does not help to find a better approximation of the probability distribution of the data.

On the other hand, although statistical analysis is used to model unexplained variations in the data, in biological applications, events often consist of a sequence of possible but not probable occurrences. For these reason, using a fuzzy model is more realistic than a statistical model.

In order to fit the model with the data, applying a machine learning technique is necessary. In this paper, a hybrid training method was proposed for this purpose. Since both global optimization and faster convergence are two key issues of training algorithms, the hybrid method used in this work can save much computing time and enhances the computational efficiency of algorithm. This is carried out by initializing the search region for COA by the conventional EBP.

COA has two main advantages: First, un-repeatability of COA in a continuous search region gives it an important priority. COA is un-repeatable because in this algorithm, the optimization is realized through chaos. The algorithm uses unrepeatable chaotic sequences generated through logistic map to search for the optimum values. Since chaos can go through all states without repetition, the search would be more efficient than other continuous search algorithms like random search. Second, the continuity of search region makes it precise but too slow. To solve the latter problem, we use EBP to restrict the search bounds for COA. This way, EBP finds an optimum in a short time, which is probably just a local minimum. In the next step, COA starts a precise search around it until reaching a better optimum. To find a more reliable optimum, the adaptive parameters, \(a_1, a_2, a_3\), are considered that are tuned in order to satisfy the required bounds of the search.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Classifier performance for ESCC risk prediction for the test dataset.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy (%)</td>
<td>Specificity (%)</td>
</tr>
<tr>
<td>87.3563</td>
<td>88.2765</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Classifier performance for dysplasia risk prediction for the train dataset.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy (%)</td>
<td>Specificity (%)</td>
</tr>
<tr>
<td>96.5333</td>
<td>62.8571</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Classifier performance for dysplasia risk prediction for the test dataset.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy (%)</td>
<td>Specificity (%)</td>
</tr>
<tr>
<td>70.53</td>
<td>80.9637</td>
</tr>
</tbody>
</table>

Fig. 7. ROC diagram comparing the sensitivity and specificity of the model to predict ESCC in Golestan province.

Fig. 8. ROC diagram comparing the sensitivity and specificity of the model to predict squamous dysplasia in Golestan province.
area, as described in Eq. (8). The mentioned parameters can enlarge the search region or limit it by getting larger or smaller respectively.

With two benefits, simple calculating and global power of chaos optimization, the algorithm finds the global optimum much faster than other evolutionary methods.

The results obtained in this study, as described in Table 5, prove that the variable structure FNN model that uses hybrid COA algorithm leads to better discrimination between the two classes and eventually better prediction results in comparison to multivariate logistic regression model for ESCC and squamous dysplasia risk prediction.

The perfect classification method leads to the point (0,1) in the ROC space, representing 100% sensitivity and 100% specificity. The ROC space is defined by FPR and TPR as x and y axes respectively. The proposed classification method yields points above the line of no-discrimination close to the perfect classification in the ROC diagram as shown in Figs. 7 and 8 which means that the model is a good classifier and a good predictor.

6. Conclusion

The proposed mathematical method was proved to be helpful in predicting the risk of ESCC and squamous dysplasia using the risk factors obtained from case control epidemiologic study. In order to make sure how successfully this model could be used in clinical practice we need to test this model in our ongoing prospective cohort study of esophageal cancer in Northern Iran. If the model is proved to be successful, then it could be used in clinical practice and lead to huge cost saving for screening and early detection of esophageal cancer.

References


area, as described in Eq. (8). The mentioned parameters can enlarge the search region or limit it by getting larger or smaller respectively.

With two benefits, simple calculating and global power of chaos optimization, the algorithm finds the global optimum much faster than other evolutionary methods.

The results obtained in this study, as described in Table 5, prove that the variable structure FNN model that uses hybrid COA algorithm leads to better discrimination between the two classes and eventually better prediction results in comparison to multivariate logistic regression model for ESCC and squamous dysplasia risk prediction.

The perfect classification method leads to the point (0,1) in the ROC space, representing 100% sensitivity and 100% specificity. The ROC space is defined by FPR and TPR as x and y axes respectively. The proposed classification method yields points above the line of no-discrimination close to the perfect classification in the ROC diagram as shown in Figs. 7 and 8 which means that the model is a good classifier and a good predictor.

6. Conclusion

The proposed mathematical method was proved to be helpful in predicting the risk of ESCC and squamous dysplasia using the risk factors obtained from case control epidemiologic study. In order to make sure how successfully this model could be used in clinical practice we need to test this model in our ongoing prospective cohort study of esophageal cancer in Northern Iran. If the model is proved to be successful, then it could be used in clinical practice and lead to huge cost saving for screening and early detection of esophageal cancer.

References


Table 5

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Method</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESCC</td>
<td>Variable Structure FNN using Hybrid COA algorithm</td>
<td>85.76</td>
<td>88.27</td>
</tr>
<tr>
<td></td>
<td>Multivariate logistic regression</td>
<td>80.6</td>
<td>82.4</td>
</tr>
<tr>
<td>Squamous dysplasia</td>
<td>Variable Structure FNN using Hybrid COA algorithm</td>
<td>75.83</td>
<td>80.9637</td>
</tr>
<tr>
<td></td>
<td>Multivariate logistic regression</td>
<td>61.5</td>
<td>69.5</td>
</tr>
</tbody>
</table>


