

A Novel Optimal Configuration of Neural Networks by Multi-Objective Genetic Algorithm and Ensemble-classifier Approach

Maryam Jawad Kadhim¹

Jamal Kh-Madhloom²

Khanapi Abd Ghan³

¹Laser and Optoelectronic Department Kut University College

²Department of Computer Science and Technology Wasit University

³Universiti Teknikal Malaysia Melaka (UTEM)

Maryam.Jawad@alkutcollege.edu.iq

jamalkh@uowasit.edu.iq

khanapi@utem.edu.my

Abstract

Machine learning algorithms have been a hallmark of data mining in image and signal processing. Several studies have proposed various methods for improving classification accuracy. Artificial Neural network (ANN) is one of the most important data mining classification method among predictive algorithms. The performance of ANN is affected by several parameters such as a number of hidden layers neurons, learning function, stop conditions and network architecture. Parameter regulation is a point of critical challenge in this algorithm. The main purpose of this study is to provide a novel approach by using multi-objective genetic algorithm and ensemble classifier to obtain optimal parameters of ANN. To this end, first, a set of neural networks were trained by setting their parameters through the multi-objective genetic algorithm. Next, the best combination of neural networks was selected to make an ensemble classifier. This method was evaluated with five popular and available datasets. Three measurements; accuracy, time and ROC curve were considered to assess the efficiency. The experimental results show that the proposed approach can achieve a trade-off between time and accuracy by the multi-objective genetic algorithm. Moreover, using ensemble-classifiers approach, we increased the reliability of the model. Consequently, the proposed method promotes the detection accuracy in three of selected datasets in comparison of four recent suitable methods.

Keywords: Artificial Neural Network, Multi-Objective Genetic Algorithm, Ensemble-Classifier, Optimal Parameters Regulation.

التكوين الأمثل للشبكات العصبية من خلال الخوارزمية الجينية متعددة الأغراض ونهج مصنف المجموعات

أ.د. خنابي عبد الغني³

جمال خضير مظلوم²

م.م. مريم جواد كاظم¹

¹قسم الهندسة الليزر والالكترونيات البصرية / كلية الكوت الجامعة

²كلية علوم الحاسوب وتكنولوجيا المعلومات / جامعة واسط

³الجامعة التقنية ملاكا ماليزيا

khanapi@utem.edu.my

jamalkh@uowasit.edu.iq

Maryam.Jawad@alkutcollege.edu.iq

المخلص

كانت خوارزميات التعلم الآلي سمة مميزة لاستخراج البيانات في معالجة الصور والإشارات. اقترحت العديد من الدراسات طرقًا مختلفة لتحسين دقة التصنيف. تعد الشبكة العصبية الاصطناعية (ANN) واحدة من أهم طرق تصنيف استخراج البيانات بين الخوارزميات التنبؤية. يتأثر أداء ANN بالعديد من المعلمات مثل عدد من الخلايا العصبية للطبقات المخفية ووظيفة التعلم وظروف التوقف وبنية الشبكة. تنظيم المعلمات هو نقطة التحدي الحاسمة في هذه الخوارزمية. الغرض الرئيسي من هذه الدراسة هو توفير نهج جديد باستخدام خوارزمية جينية متعددة الأهداف ومصنف مجموعة للحصول على المعلمات المثلى لـ ANN. تحقيقاً لهذه الغاية، أولاً، تم تدريب مجموعة من الشبكات العصبية من خلال تحديد معلماتها من خلال الخوارزمية الجينية متعددة الأهداف. بعد ذلك، تم اختيار أفضل مزيج من الشبكات العصبية لعمل مصنف مجموعة. تم تقييم هذه الطريقة بخمس مجموعات بيانات شائعة ومتاحة. ثلاثة قياسات تم النظر في الدقة والوقت ومنحنى ROC لتقييم الكفاءة. تظهر النتائج التجريبية أن النهج المقترح يمكن أن يحقق مفاضلة بين الوقت والدقة بواسطة الخوارزمية الجينية متعددة الأهداف. علاوة على ذلك، باستخدام نهج مصنفات المجموعات، قمنا بزيادة موثوقية النموذج. وبالتالي، فإن الطريقة المقترحة تعزز دقة الكشف في ثلاث من مجموعات البيانات المختارة مقارنة بأربع طرق مناسبة حديثة.

الكلمات المفتاحية: شبكة عصبية اصطناعية، خوارزمية وراثية متعددة الأغراض، مصنف المجموعات، تنظيم المعلمات المثلى

Introduction:

Artificial neural network (ANN) is one of the most popular methods in machine learning. Because of increasing human need to discover knowledge and relationships in the raw data, development of data storage, and retrieving devices, the usage of the automated knowledge discovery methods has been increased. Knowledge discovery is divided into two categories of predicting and describing tasks

[1], [2]. Today, there are several studies about developing and improving algorithms of this area in various scientific and commercial areas [3]–[7].

The term neural network traditionally refers to the network or circuit of biological neuron circuits. However, ANN is used nowadays as a simulation of neurons. In fact, ANNs are inspired from central nervous system. This learning model is used to

estimate the approximation functions and modeling of the relationship between input and output. ANN is consisted of some interconnected networks of neurons that information is exchanged between them. Each of the connections has a numerical weight that is set during the learning process. This algorithm tries to model intelligent neurons as a human nervous system by setting of numerical weight (w) and bias (b) parameters.

Recently, neural network algorithm is very popular in various areas such as data mining [1], [2] and medical [8], [9] areas. The main advantage of this algorithm is its ability in linear and nonlinear relationship modeling, high accuracy, and stability against noise. However, this method has two major drawbacks: First, this method is not able to display extracted knowledge. In the other word, the extracted knowledge is a black box. Second, setting its parameters is a user-dependent process. These parameters highly affect on the ANN efficiency [1], [2]. Determining the best combination of parameters for neural network leads to enhance exponentially the performance but this problem is a time-consuming process. As a result, the use of brute force searching to find the best combination is unreasonable. The most important parameters, which affect the efficiency, are the number of hidden neurons, number of epochs, learning function, transfer function, error function, learning rate, momentum factor.

With the intention of achieving the optimal configuration of ANN and user dependency reduction in a reasonable time, parameters of ANN can be set by the genetic algorithm. Genetic

algorithm (GA) is an evolutionary approach that is based on random search. This algorithm is often used for optimization problems. GA has the capability of quick searching in the wide area of search space to find the optimal solution. In order to set the ANN parameter configuration with GA, each chromosome in the GA can be encoded as a set of parameters to configure the ANN. In addition, the fitness function can be calculated based on the accuracy in the simplest way. However, another measure such as time, stability and deviation can be added to the fitness function. Therefore, single objective problem is converted to multi-objective. Finally, in the evolutionary process, the last efficiency of ANN is increased, and its parameters will be set automatically.

Ensemble-classifier approach is another class of predictive methods to increase the reliability of predicting model. In this approach, the objective is to enhance the reliability of model by increasing the number of classifiers and diversity. It produces a number of outputs instead of one output and generates final class label by an aggregation such as voting. The best situation of ensemble-classifier is produced in diverse classifiers. By this method, each classifier will cover a part of data and therefore reliability will be increased. If the same classifiers are used, only time will increase and redundancy will occur [10]. For this reason, in this study various criteria have been presented to determine the diversity between classifiers. Finding the best combination of classifiers to shape the ensemble requires very high time cost. As a result, evolutionary process approach is recommended to find the best combination of classifiers to achieve accuracy and diversity,

simultaneously. In this approach, each chromosome includes binary coding, and it determines the membership or lack of membership of each classifier in the final set.

To summarize, the proposed approach is consisted of two stages: generating the best ANNs with the parameter configuration and collecting the best classifiers as an ensemble. As it mentioned, the evaluation of the ANN is a multi-objective problem. Accuracy and time might be determined by two distinct objectives that even in some cases they also act in opposite directions. For instance, the accuracy may be increased by increasing the number of neurons, but time is growing. Therefore, the multi-objective evaluation will be necessary to configure the parameters of ANN. On the other hand, in order to determine the best set of classifiers, accuracy, time, and diversity can be selected as three distinct objectives. The most important characteristic of the genetic algorithm is its ability to support multi-objective mode.

In this study, previous works related to ANN development are reviewed in section 2. For this purpose, a review of previous researches is divided in two categories; ANN development and its application, and development of ensemble classifier. Then, the effect of parameter configuration in the previous research was highlighted. In section 3, the proposed method is presented. Experimental results of the study are provided in section 4. Finally, conclusion and future work are proposed.

Related work: In this section, first considering the previous studies conducted on the configuring of ANN parameters to increase the efficiency; second, reviewing the previous papers about the selection of the optimal combination of classifiers to construct ensemble-classifier and increasing its reliability.

2-1. Configuring ANN parameters

In the medical field, ANN algorithm is regarded as the third top algorithm [8]. Nevertheless, two major drawbacks for this algorithm is seen. First, the lack of an intuitive and standard procedure for evaluation and configuration its parameters. Second, black box function of its extracted knowledge. Several recent studies try to solve these problems to achieve more optimal method from this proper algorithm. In [11], a new method is provided to resolve the first limitation. ANN has configured automatically by using a GA optimization. In this method, ANN parameters, which must be determined by an expert, are configured during an evolutionary process. The purpose of using GA is to achieve the best prediction of advanced cancer compared to diagnosis of expert physicians in the field of medicine. The results show that the proposed method improves the accuracy of prediction without any need for an expert to configure ANN. The main disadvantage of this work is to ignore time and only focus on the accuracy. Furthermore, in a similar study in order to reduce the dependence of ANN to an expert, the evolutionary neural network has been proposed [12]. In the paper [13], an approach was proposed to set the parameters of the ANN using a central composite

design and GA. Three important parameters which affecting the ANN, were considered. ANN parameters were configured with GA and central composite design was used in experiments plan and analyze network behavior based on determined parameters by GA. Generally, it consists of five steps consist of: determining the efficiency ANN parameters, proposing experiments based on a central composite design, defining evaluation function for each evaluation criterion, analyzing variance and determining regression function in order to calculate impact of each factor. The results suggested more performance of ANN with proposed method compared to ANN by randomized parameters and Taguchi method. In the paper [14], a new approach is proposed to update the bias and weight in a combination of MLP and GA. It improves performance using parameter regularization in which, GA sets weights and biases and the ANN is trained by BP and BFGS. The implementation results showed that this combined learning algorithm has proper efficiency of approximation in nonlinear dynamical system.

2-2. Second approach: Ensemble ANN

One of the most powerful tools in construction of high accurate classifier is Ensemble classifier. In this approach, several classifiers instead of one make decision. Therefore, accuracy and reliability may be increased. In order to produce final decision among several decisions, voting is suggested [15]–[17]. The main goal of this approach is to increase accuracy through collection

of the diverse classifiers. This performance enhances along with increasing of executing time.

In the paper [15], a new efficient approach was proposed based on ensemble ANN for medicine application. In this approach, after reading and preprocessing of data, wide range of ANN classifiers are constructed with different number of neurons. The most important parameter in this approach is the number of neurons. For this purpose, from 1 to 1928 neurons were considered for construction of ANN. Then, ANNs are sorted based on the accuracy and 100 ANNs are selected from pooling. In the next step, to construct a final set, Kohavi-Wolpert (KW) was considered. The KW is the measurement for diversity. Results of this proposed method indicated that diversity concentration could improve accuracy of the ensemble set. Besides, inexplicit process of selecting the optimal classifiers combination for maximum diversity, and very high complexity of ANNs construction from pooling of 1928 neurons are considered as this model disadvantages. In a same paper, a novel metric was used to calculate diversity between classifiers. Paper [18] tried to present an approach based on GA optimization techniques to select a subset of classifiers based on accuracy and variance inflation factor. In this approach, each gene considered as one classifier and binary encoding was proposed. In addition, the fitness function defended as a mean of accuracy. This approach focuses only on accuracy, and it is not appropriate to deal with multi-linearity problem. For this reason, the variance inflation factor was used to solve this problem and measuring linearity between classifiers. In the other words, VIF criterion determines the impact

of a classifier on the final predicting model. Experimental results of this study are suggested that in an ensemble approach, the selection of classifiers with lower linearity and higher diversity is more effective. In the paper [19], online ensemble approach has been proposed for ANN. This approach had two steps: in the first, usual training was carried out and in the second, a new method is used to optimal selection by giving a weight to each classifier. Results of evaluation of this online method indicated considerable improvement of classifier performance compared to other online general methods.

Materials and Methods

The main objective of the proposed method can be categorized into two targets. The first one is to regulate the parameters of the ANN independent of the expert in a multi-objective manner. The second one is to use the ensemble-classifiers based on GA to increase the reliability and diversity among the classifiers. In both steps, the procedures carry out based on GA. In the first step, evolutionary process should set the parameters of the ANN and in the second one; the best combination of classifiers is selected. Figure 1 shows the architecture of the proposed method in this paper.

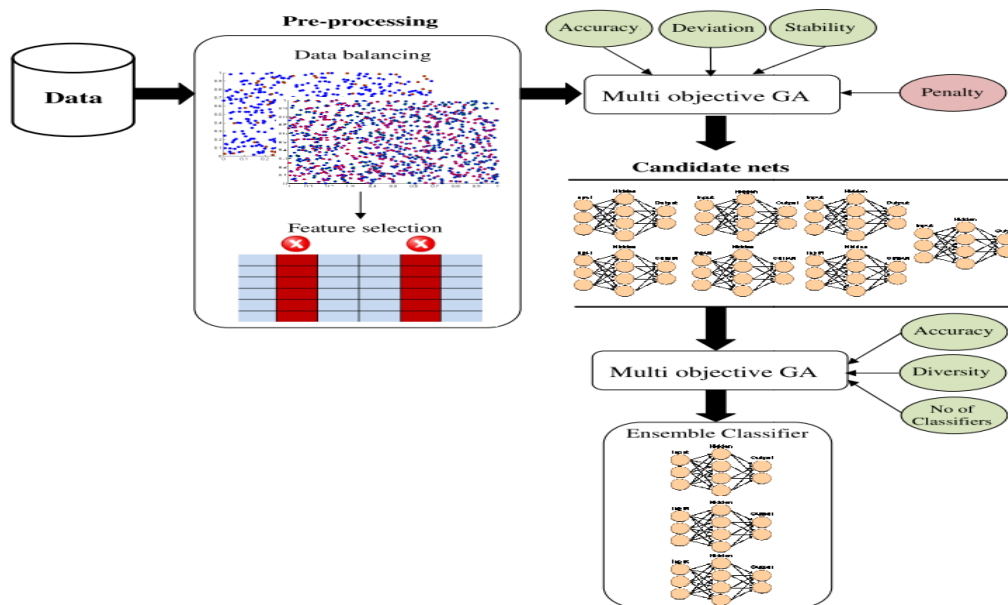


Figure 1: the proposed approach framework

The proposed method consists of four steps, as follows:

Step 1) Data pre-processing

At the first step, the data should be balanced so that the efficiency of classifier in predicting

minority and the majority class label to be preserved. Data may be imbalanced for various reasons such as data collection policy, nature of study area, and incorrect division of data into train and test. Some reasons can be solved without any process. For instance, incorrect division of data may be occurred by holdout sampling method and

can be solved by cross-validation. On the other side, some of reasons should be processed. Today, various methods have been proposed for balancing the data. In the one category, majority class is reduced to be equal with the minority and on the opposite side; minority data are replicated to be equal with the majority. In the proposed method, in order to solve the problem of unbalanced data, two ways are used. The data is divided by cross-validation and the SMOTE method [20] has been used to remove skew in data. After data balancing, feature selection was applied to reduce time in the GA process. As mentioned, both GA and ANN are time-consuming. Consequently, the best way to reduce the executing time is feature selection. In

this step, a combination of GA and incremental search is used [21] to select the best features.

Step 2) Configure the ANN parameters and make ANNs pool.

The final shape in the proposed model is ensemble ANNs which it should achieve a high accuracy and diversity. Before this, we should construct optimal ANNs and then select the best combination among them. In order to construct optimal ANN, multi-objective GA was used. In GA, each chromosome represents a setting for ANN and has nine genes. Figure 2 provides a view of the chromosome encoding.

No. of neurons	Epoch	Training Func	Error Func	Transfer Fun L1	Transfer Func L2	Maximum Validation	Momentum factor	Learning rate
----------------	-------	---------------	------------	-----------------	------------------	--------------------	-----------------	---------------

Figure 2: chromosome encoding in the genetic algorithm

The first gene shows the number of neurons in the hidden layer. According to the [11], the number of genes can be defined as below:

$$\frac{1}{2} \text{features} \leq \text{Neurons} \leq 3 * \text{Featur} \quad (\text{Equation 1})$$

Obviously, as the number of neurons increase, learning accuracy and training time also will increase. The second gene determines the number of epochs in ANN for training. Like before, more epoch causes more training time. Furthermore, if the number of epochs be less the training cannot be completed. According to experience, this gene can have a range number between 50 to 500. It is important that, if the number of epochs be

considered less or more than optimal point, the under or over fitting may be occurred. The third gene chooses the learning function. Given the choice of MATLAB, the values of Table 1 were used for this gene. The fourth gene deals with calculation of error in ANN learning process. The values of these genes are also determined by the values shown in Table 1. Fifth and sixth genes, determine the transfer function for first and second layers. These genes will also select an arbitrary among the values shown in Table 1. The seventh gene is the maximum number of validation checks. By this parameter, the number of consecutive iterations that the validation performance fails to decrease is determined. This criterion can be

known as an early stop mechanism. For this gene, a number between three and eight is selected. The eighth gene is momentum factor that is adds a fraction of the previous movement to the current. This technique helps ANN to avoid local minima. The high value of this parameter speeds up the

convergence but may cause a random and instable behavior of the model. For this gene, values changed in range from 0.6 to 0.95. Ninth gene is learning rate that is controls the weight and bias changes during training. For this gene, the values from 0.1 to 0.9 are selected.

Table 1: Learning, transfer and error functions in Matlab that was used in simulation

Learning functions	trainb , trainbfg, traingb, traingf, traingp, traingd, traingda, traingdm, traingdx, trainlm, trainoss, traingp, traingru, traingcg
Transfer functions	Compet, hardlim, hardlims, logsig, netinv, poslin, radbas, satlin, satlins, tansig
Error functions	mae, mse, sae, sse

After encoding, other GA operators should be set. For this purpose, GA parameters set as: Generation = 20, population = 40, Cross over = Two point, mutation = Adaptive feasible and Selection = roulette wheel was selected. In order to calculate fitness, each chromosome is evaluated by accuracy, stability, and standard deviation. The first fitness score is accuracy. Because of using k-fold cross validation method for dividing data, test accuracy can be calculated at any step. In equation 2, test error is seen.

$$Acc = \frac{\sum_{i=1}^k AccTest(M_i)}{k} \quad (\text{Equation 2})$$

Where, k represents the number of data division in k-fold cross validation and AccTest (Mi) is accuracy in ith step of cross validation. In order to increase impact of accuracy in fitness function, cubic of accuracy is used. The second fitness score

is stability. Equation 3 shows stability calculation. This criterion indicates difference between training and test accuracy. When this value is lower, it represents high reliability of model when faced with unseen data.

$$STA = Max(|AccTest(M_i) - AccTrain(M_i)|) \longrightarrow 1 \leq i \leq k \quad (\text{Equation 3})$$

To calculate this, the difference between training and test accuracy of each k-fold in cross validation is calculated and the highest difference is reported. The third score is similarity of the results of each model in cross validation. If the results of ANN for different part of data are similar, it will have high reliability. For this purpose, the standard deviation can be used as follow:

$$STD = \sqrt{\frac{1}{k} \sum (AccTest(M_i) - \mu)^2} \quad (\text{Equation 4})$$

Where μ is mean of AccTest across k folds. Finally, time can be determined as a penalty in fitness score. For this purpose, in the following equation, the considered values are seen.

$$\begin{aligned}
 P=0 & \quad \text{time} < 3\text{min} \\
 P=0.1 & \quad 3 \leq \text{time} \leq 5\text{min} \\
 P=0.2 & \quad 5 \leq \text{time} \leq 10\text{min} \\
 P=0.3 & \quad \text{time} > 10\text{min}
 \end{aligned}
 \tag{Equation 5}$$

According to the above descriptions, GA starts to search and find the best ANN parameters. Lastly, as the output of this step, 19 chromosomes are selected from the elite population, and they are used as input of the next step to make an ANNs pool. The most important characteristic of the produced ANNs in this step is the ability of achieving to a balance between different criteria.

Step 3) GA ensemble ANN

In the previous section, ANNs was configured in a multi-objective way. In this step, we should select an optimal combination to shape the ensemble set. In order to build the ensemble, GA was applied to select the best combination of ANNs from pooling. Besides, the voting scheme was used to aggregate the answers and produce a single predict. Each chromosome at this step has binary encoding, which is composed of 19 genes. Every gene associated with an ANN that configured in the previous. If gene is set, it means that corresponding ANN is considered in the final combination, and otherwise, it will not in the final combination. Binary encoding of chromosome can be seen in Figure 3.

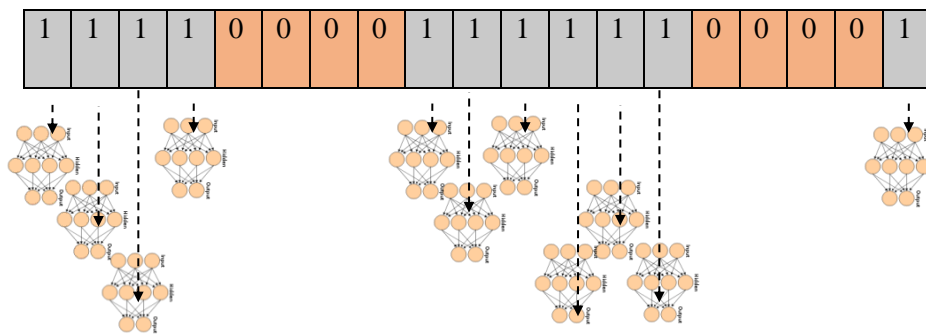


Figure 3: Binary encoding of each chromosome in step 3

For simplification, the parameters of GA in this step set as the GA parameters in the previous step. To write fitness function and evaluate each chromosome, three objectives are considered: accuracy of the final model, diversity of classifiers, and number of classifiers. Accuracy of the model is determined based on Equation 2

through cross validation. For importance of accuracy, its square is used in the evaluation. To calculate diversity of classifiers, Kohavi-Wolpert [22] criterion is used. In addition, the number of selected classifiers is evaluated as a penalty. In figure 4, pseudo code of the proposed approach can be seen.

```

Void main()
{
    data=Load(address);
    [feature]=Ga_FS(data, size(data,2)*0.7);    //feature selection
                                                by GA. 0.7 Of the main features were selected

    d1=data[:,feature];
    d2=SMOTE(d1); //data balancing
    input=d2[:,1:end-1];
    target=d2[:,end];

    Ann []nets=new Ann[19];
    nets=FindANNParameters(input,target); //GA step 1

    Ensemble model=new Ensemble();
    model=FindCombination(nets,input,target); // GA step 2
}

```

```

Ann[] FindANNParameters(input,target)
{
    Genetic ga=new Genetic();
    Ga.population=40; Ga.Generation=20; Ga.CrossOver=TwoPoint;
    Ga.Mutation= AdaptiveFeasible; Ga.Selection= RouletteWheel;

    Chromosome ch=new Chromosome(9);    //Each chromosome contains          9 genes

    ch[1].bound=[0.5*#feature ... 3*#feature];
    ch2.bound=[5 ...500];          ch3.bound=[1 ... 14];
    ch[4].bound=[1 ... 10];          ch[5].bound=[1 ... 10];
    ch[6].bound=[1 ... 4];          ch[7].bound=[3 ...8];
    ch[8].bound=[0.6 ... 0.95]; ch[9].bound=[0.1 ... 0.9];

    [result,score]=ga.run(ch,@fitness1)
    Ann[] pool=result[1:19];
    Return pool;
}

result fitness1(Chromosome ch)
{
    Ann a=petternNet(ch[1]);
    a.epoch=ch[2];          a.learning=ch[3];          a.transfer=(ch[4],ch[5]);
        a.error=ch[6];
    a.validation=ch[7];          a.momentum=ch[8];
    a.learning=ch[9];

    a.train(data);
    retrun Evaluate(a,data); //Calculate accuracy, STDdev, stability
                                                and time
}

```

```

model FindCombination (nets,input,target)
{
    Genetic ga=new Genetic();
    Ga.population=40; Ga.Generation=20; Ga.CrossOver=TwoPoint;
    Ga.Mutation= AdaptiveFeasible; Ga.Selection= RouletteWheel;

    Chromosome ch=new Chromosome(19);    //Each chromosome contains
                                                19 genes with the binary encoding

    [result,score]=ga.run(ch,@fitness2)
    Model m=Ensemble(find(result==1),voting);
    Return m;
}

result fitness2(Chromosome ch)
{

```

Figure 4: pseudo code of the proposed approach

Experiment Results

In this section, a framework of designed experiment is expressed and then the results of proposed model are validated by several exist datasets. Different kind of datasets are selected to cover a wide variety of properties. In addition, the results are compared with the several algorithms.

Five available datasets from UCI Machine Learning Repository were selected. In data collection, it has been tried to investigate the effect of number of records, variables, and imbalance class variable on the algorithm. All of the data sets are without missing value. The characteristics of data sets are presented in table 2.

Table 2: Characteristics of data sets

Data set	Records	Columns
IRIS	150	4
Wisconsin Breast Cancer	699	10
Liver	345	7
Tic Tac Toe	958	9
Heart	270	13

One of the advantages of proposed approach is that it can be mounted on any traditional algorithms for ANN. In this study, a usual feed forward neural network was used. Evolutionary ensemble algorithm proposed in [23] and evolutionary assembled algorithm proposed in [11] are chosen to make a comparative study because of similarity with proposed algorithms in this paper. The experiments were implemented on Matlab in PC with 2.4 GHz processor and 2GB memory and win7. In order to evaluation of the results and final classifier, four metrics time, accuracy, Roc curves and area under ROC (AUC) were taken into account [16].

In this section, first the results of simulations are separately indicated. Next, the results of three metric time, accuracy and AUC for each dataset are presented in a tabular format to reach an overall view. Subsequently, the time of the proposed approach will be investigated in a detail. For this purpose, four step of the proposed approach is taken into account. In the following, stability and ROC analysis is expressed. Finally, advantage and disadvantage of the proposed method is discussed.

Table 3: Time, accuracy and AUC of the proposed method and comparatives methods

Data set	Method	Time (Ms)	Accuracy (%)		AUC
			Train	Test	
IRIS	Traditional	51	76.64	73.67	0.74
	Evolutionary ensemble	504	82.14	74.61	0.79
	Evolutionary assembled	688	78.67	74.98	0.77
	Proposed	1294.3	75.54	74.67	0.81
Breast Cancer	Traditional	150	83.06	79.51	0.72
	Evolutionary ensemble	3100	82.31	81.16	0.77
	Evolutionary assembled	3547	80.76	80.62	0.71
	Proposed	3228.1	84.23	84.05	0.82
Liver	Traditional	226	76.33	73.33	0.67
	Evolutionary ensemble	1000	75	74.91	0.7
	Evolutionary assembled	545	76.36	75	0.73
	Proposed	4502.5	75.26	75.85	0.76
Tic Tac Toe	Traditional	51	76.64	73.48	0.74
	Evolutionary ensemble	504	82.14	74.61	0.79
	Evolutionary assembled	688	78.67	74.98	0.77
	Proposed	1293.6	75.54	74.67	0.81
Heart	Traditional	25	73.54	67.9	0.65
	Evolutionary ensemble	200	86.24	79.01	0.74
	Evolutionary assembled	622	95.38	90.47	0.84
	Proposed	811.1	86.77	85.18	0.79

In table 3, proposed method was compared through three metrics. It can see clearly that, running time of the proposed approach is longer than other ones. On the other side, traditional feed forward ANN has a lowest time. However, keep in mind, this time is for training that can be classified as an off-line. On-line time can be introduced as a response time for predict new data. Consequently, we can conclude that, time of the proposed approach is the challenge. Two reasons can be cited for this challenge. First, in the proposed

approach GA used twice. As mentioned before, evolutionary algorithms such as GA known as a time-consuming process. Second, in order to balance data, SMOTE was used that over-sampling approach is applied. Therefore, the number of records will rise up. In figure 5, time of the proposed approach divided into four sub-part and average time for each part was plotted. The idea for this chart is to identify time-consuming parts of the proposed approach.

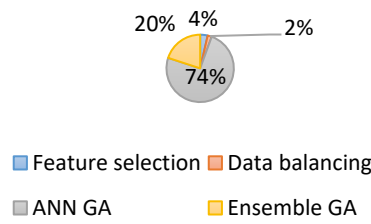


Figure 5: Time detail of the proposed method

It is clear that, ANN configuration by GA is the most time-consuming part. Therefore, this part can be considered as improvements in the future. Accuracy is one of the most common and widely used methods in evaluation. This method is the simplest numerical evaluation method so that it can easily be interpreted. We can do two analyses through this metric: first, test accuracy of the model, shows the validity of the produced model to predict unseen data. Second, stability analysis can be done in order to determine reliability of the

model. Test accuracy of the proposed approach is better than others except in Tic tac toe data set. In general, improve in test accuracy by the proposed approach is obviously visible. Stability can be defined as the difference between training and test accuracy. In the other words, stability reveals the dependence of the model to the data. As this measure is decrease, reliability will be increase. In figure 6, training accuracy and test accuracy was plotted as a bar chart. The aim of this chart is to show stability.

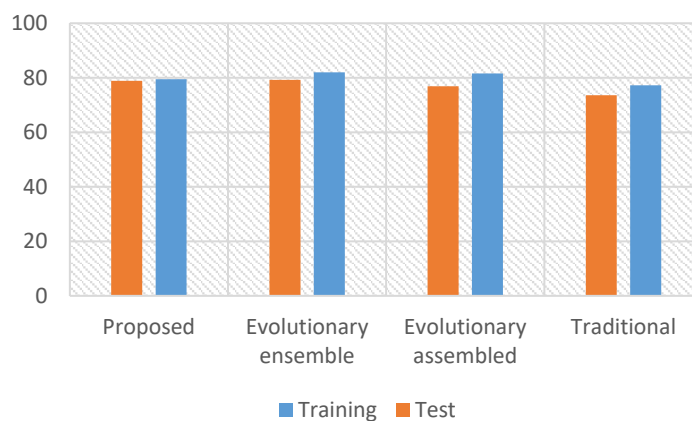


Figure 6: Average accuracy for training and test

Figure 6 shows that stability of the proposed approach is more than other ones. It is clear by the difference between training and test accuracy. Therefore, it shows the proposed approach is more robust against changes in data.

One of the important visual metrics for evaluation is ROC curve. ROC curves were plotted using true positive rate (TPR) and false positive rate (FPR). In this chart, each model displays by a set of points as a curved line. The model is more efficient whatever the distance of the graph curve to the point (0, 1) is less. The best model can be chosen

by measuring this distance for each model. ROC curve can be used to select optimal classifier and identify sub-optimal ones. In this curve, class distribution or cost does not consider. In figure 7, ROC curve for five datasets and four methods was plotted. In order to achieve numerical evaluation through ROC curve, AUC was suggested. AUC represents the area under the ROC curve and the numerical efficiency of different classifiers can be measuring and comparing. According to the results of table 3 and figure 7, the proposed method indicates an optimal manner in comparison of Evolutionary.

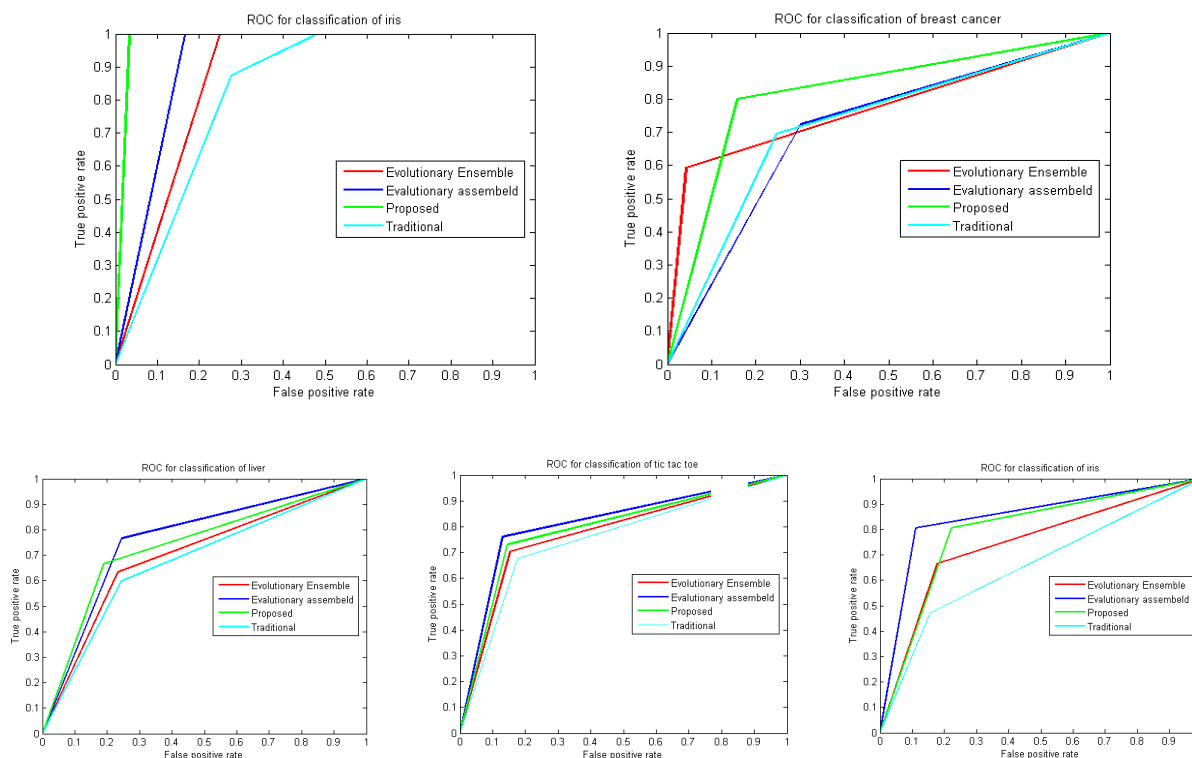


Figure 7: Roc curves of four methods on the five datasets

Conclusions

Neural network is one of the most and widely used classification algorithm in data mining. This method is more important due to the high power in modeling of relationships among data. In order to achieve the best results of neural network, it is necessary to set the parameters optimally. To this end, need to skill and experience that, an expert person does it but in the most cases, parameters are set randomly. Therefore, an automatic way is essential to parameters regularization. In addition, the efficiency of classification algorithms is assessed by various parameters.

The main goals of the proposed method are summarized to two phases. phase1, setting the parameters of the neural network independent of an expert, in order to minimize time and maximize accuracy. Phase2, increasing the reliability and diversity within the classifier model using the multiple classifiers. In both phases Genetic algorithms is applied. In the first phase, in an evolution process, the more affected parameters of the neural network are determined. In the second phase, from the selected classifiers the best combination is selected.

Five available datasets with varied features were considered to evaluate the efficiency of the proposed approaches. Several algorithms were used to compare with the proposed approach. Finally, accuracy, runtime, AUC and ROC curve criteria were used to evaluate the results. AUC and ROC curve indicate that the proposed method is more efficient compared to other methods. Furthermore, it has reasonable results according to accuracy. Moreover, the proposed method shows

the highest stability. Despite of all mentioned advantages for the proposed method, because of using genetic algorithm twice and balancing algorithm, this approach has high time complexity.

References

- [1] D. J. Hand, "Data Mining Based in part on the article 'Data mining' by David Hand, which appeared in the Encyclopedia of Environmetrics.," in Encyclopedia of Environmetrics, Chichester, UK: John Wiley & Sons, Ltd, 2013.
- [2] Xindong Wu, Xingquan Zhu, Gong-Qing Wu, and Wei Ding, "Data mining with big data," IEEE Trans. Knowl. Data Eng., vol. 26, no. 1, pp. 97–107, Jan. 2014.
- [3] H.-C. Lee, R. Kosoy, C. E. Becker, J. T. Dudley, and B. A. Kidd, "Automated cell type discovery and classification through knowledge transfer," Bioinformatics, vol. 33, no. 11, pp. 1689–1695, Jun. 2017.
- [4] N. Capuano and D. Toti, "Experimentation of a smart learning system for law based on knowledge discovery and cognitive computing," Comput. Human Behav., vol. 92, pp. 459–467, Mar. 2019.
- [5] F. Liu, X. Zhou, J. Cao, Z. Wang, H. Wang, and Y. Zhang, "Arrhythmias Classification by Integrating Stacked Bidirectional LSTM and Two-Dimensional CNN," Springer, Cham, 2019, pp. 136–149.

- [6] F. Kargarfard, A. Sami, and E. Ebrahimie, "Knowledge discovery and sequence-based prediction of pandemic influenza using an integrated classification and association rule mining (CBA) algorithm," *J. Biomed. Inform.*, vol. 57, pp. 181–188, Oct. 2015.
- [7] M. Parto Dezfouli and M. R. Daliri, "Single-Trial Decoding from Local Field Potential Using Bag of Word Representation," *Brain Topogr.*, vol. 33, no. 1, pp. 10–21, Jan. 2020.
- [8] N. Esfandiari, M. R. Babavalian, A.-M. E. Moghadam, and V. K. Tabar, "Knowledge discovery in medicine: Current issue and future trend," *Expert Syst. Appl.*, vol. 41, no. 9, pp. 4434–4463, Jul. 2014.
- [9] D. Delen, A. Oztekin, and L. Tomak, "An analytic approach to better understanding and management of coronary surgeries," *Decis. Support Syst.*, vol. 52, no. 3, pp. 698–705, Feb. 2012.
- [10] G. Giacinto and F. Roli, "An approach to the automatic design of multiple classifier systems," *Pattern Recognit. Lett.*, vol. 22, no. 1, pp. 25–33, Jan. 2001.
- [11] A. M. Vukicevic, G. R. Jovicic, M. M. Stojadinovic, R. I. Prelevic, and N. D. Filipovic, "Evolutionary assembled neural networks for making medical decisions with minimal regret: Application for predicting advanced bladder cancer outcome," *Expert Syst. Appl.*, vol. 41, no. 18, pp. 8092–8100, Dec. 2014.
- [12] M. Castellani, "Evolutionary generation of neural network classifiers—An empirical comparison," *Neurocomputing*, vol. 99, pp. 214–229, Jan. 2013.
- [13] M. Bashiri and A. Farshbaf Geranmayeh, "Tuning the parameters of an artificial neural network using central composite design and genetic algorithm," *Sci. Iran.*, vol. 18, no. 6, pp. 1600–1608, Dec. 2011.
- [14] M. J. Er and F. Liu, "Parameter Tuning of MLP Neural Network Using Genetic Algorithms," Springer, Berlin, Heidelberg, 2009, pp. 121–130.
- [15] S.-T. Luo and B.-W. Cheng, "Diagnosing Breast Masses in Digital Mammography Using Feature Selection and Ensemble Methods," *J. Med. Syst.*, vol. 36, no. 2, pp. 569–577, Apr. 2012.
- [16] R. C. Prati, G. E. A. P. A. Batista, and M. C. Monard, "A Survey on Graphical Methods for Classification Predictive Performance Evaluation," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 11, pp. 1601–1618, Nov. 2011.
- [17] M. A. P. Dezfouli, M. P. Dezfouli, and H. S. Rad, "A novel approach for baseline correction in 1H-MRS signals based on ensemble empirical mode decomposition," in 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2014, 2014, pp. 3196–3199.
- [18] M.-J. Kim and D.-K. Kang, "Classifiers selection in ensembles using genetic

- algorithms for bankruptcy prediction,” *Expert Syst. Appl.*, vol. 39, no. 10, pp. 9308–9314, Aug. 2012.
- [19] M. Eastwood and B. Gabrys, “Generalised bottom-up pruning: A model level combination of decision trees,” *Expert Syst. Appl.*, vol. 39, no. 10, pp. 9150–9158, Aug. 2012.
- [20] C. Bunkhumpornpat, K. Sinapiromsaran, and C. Lursinsap, “DBSMOTE: Density-Based Synthetic Minority Over-sampling TEchnique,” *Appl. Intell.*, vol. 36, no. 3, pp. 664–684, Apr. 2012.
- [21] O. Ludwig and U. Nunes, “Novel Maximum-Margin Training Algorithms for Supervised Neural Networks,” *IEEE Trans. Neural Networks*, vol. 21, no. 6, pp. 972–984, Jun. 2010.
- [22] V. L. Berardi and G. P. Zhang, “An empirical investigation of bias and variance in time series forecasting: modeling considerations and error evaluation,” *IEEE Trans. Neural Networks*, vol. 14, no. 3, pp. 668–679, May 2003.
- [23] P. M. Leod, B. Verma, and M. Zhang, “Optimizing configuration of neural ensemble network for breast cancer diagnosis,” in *2014 International Joint Conference on Neural Networks (IJCNN)*, 2014, pp. 1087–1092.