IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 12, December 2016

Classification using Convolutional Neural Network for Heart and Diabetics Datasets

Tharani.S[1], Dr. C. Yamini[2]

Research Scholar, Computer Science Department, Sri Ramakrishna College of Arts and Science for Women, Coimbatore, India [1]

Associate Professor, Computer Science, Department, Sri Ramakrishna College of Arts and Science for Women, Coimbatore, India [2]

Abstract: The neural network approach to generate efficient classification rules. Convolution neural network algorithm is a multilayer perceptron that is the special design for identification of two-dimensional data information. Always have more layers: input layer, convolution layer, sample layer and output layer. Deep learning refers to the shining branch of machine learning that is based on learning levels of representations. Convolutional Neural Networks (CNN) is one kind of deep neural network. To perform classification task of heart disease dataset, the neural network is trained using convolutions algorithm. The experiment is conducted with heart disease dataset by considering the single and multilayer neural network modes. The proposed algorithm gives detailed analysis of the process of CNN algorithm both the forward process and back propagation. Then we applied improved convolutional neural network to implement the typical heartdata recognition using weka tool. The experimental result show the best classification accuracy compare with existing classification algorithm.

Keywords: Data Mining, Classification, Convolutional Neural Networks, Heart dataset.

I. INTRODUCTION

Deep learning refers to a subfield of machine learning that is based on learning levels of representations, corresponding to a hierarchy of features, factors or concepts, where higher-lever concepts are defined from lower-lever ones, and the same lower-lever concepts can help to define many higher-lever concepts. Deep learning is learning multiple levels of representation and abstraction, helps to understand the data such as images, audio and text. The concept of Deep Learning comes from the study of Artificial Neural Network; Multilayer Perceptron which contains more hidden layers is a Deep Learning structure.

Feedforward neural network or Multilayer Perceptron with multiple hidden layers in artificial neural networks is usually known as Deep Neural Networks (DNNs). Convolutional Neural Networks (CNN) is one kind of feedforward neural network. In 1960s, when Hubel and Wiesel researched the neurons used for local sensitive orientation-selective in the cat's visual system, they found the special network structure can effectively reduce the complexity of Feedback Neural Networks and then proposed Convolution Neural Network. CNN is an efficient recognition algorithm which is widely used in pattern recognition and image processing. It has many features such as simple structure, less training parameters and adaptability.

An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. Various neural networks are used to diagnosis of disease in medical field such as diabetes, cancer, attacks. Using such type of network diagnosis of disease is very easy task. Using feed forward neural network technique easily prediction of glucose in blood within 75 min. Only 10 patients are assed using NNM model but it not included in training data set. Various input is given to neural network model such as CGM value, insulin dosage metered glucose value, nutritional intake, lifestyle, and emotional factors.

This system gives output as real time prediction of glucose. Using such technique processing time is reduced than time lagged ff. The model predicates 88.6%of normal glucose. The different types of neural network structure such as Multilayer perceptron (MLP), radial basis function (RBF) and general regression neuralnetwork (GRNN). Here PIMA Indians diabetes dataset are used. These structures were applied to PIMA Indians Diabetes (PID). Shows that performance of radial basis function was worse than Multilayer perceptron. General regression neural network (GRNN), Multilayer perceptron (MLP) gives 80.21%, 77.08% classification accuracy respectively. The diagnosis of diabetes using multilayer neural network and probabilistic network of PIMA Indian diabetes database. Diagnosis of diabetes PIMA Indian diabetes dataset is

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 12, December 2016

used. Multilayer neural network is trained by Levenberg (LM) algorithm. Two classes used in PIMA Indian dataset and these classes contain 768 samples. Class1 contain 500 samples and class2 contain 268 samples. And eight attribute are considering. Using such type of network diagnosis of problem is easy. Paper represents not only diagnosis of diabetes but also diagnosis of breast cancer. In this paper diabetes dataset are used that database contain 768 samples taken from patients. Each sample is described by 8 features. Here taken 500 samples from patients who do not have diabetes & 268 samples that have diabetes. From above data, they randomly selected 345 samples for training, 39 samples for cross validation and 384 samples for testing. Using such type of data, they diagnosis of diabetes easily. Network is trained with an augmented cross entropy error function. Main advantages are its reduced risk of data over fitting and reduced cost of future data acquisition.

II. RELATED WORK

Naïve bayes is machine learning technique for constructing classifier. It is simple probabilistic classifier based on bayes theorem with strong independent assumption. It is highly scalable, requiring a number of parameters linear in number variable. It's called also independence bayes. Naive bayes classifier is also used to diagnosis of diabetes in early as possible. The two methodologies decision tree and naïve Bays. It gives simpler solution to the problem for diagnosis of diabetes especially in women. Here used J48 decision tree. Input is PIMA Indian diabetes dataset in csv format. Classification type of data mining has been applied to PIMA Indian diabetes dataset and pre-processing are done using weka tool. Decision tree is a tree structure, which is form of flow chart. Using nodes and internodes classification and prediction are done. Roots and internodes are used as test cases that separate the instances with different features. Internal nodes are result of attribute cases. Leaf nodes denote the class variable. Class variable determine if person has diabetes or not. Output of decision tree gives either tested positive or tested negative.

Naïve bays are sequential in natures. Bays algorithm is applied for overcome limitation of existing system. It is applied on larger dataset in real time. Using Naïve bays and decision tree diagnosis of diabetes is efficient way. To predict chances ofdiabetic patient getting various diseases like heart disease. The existing system applying Naïve Bayes data mining classifier technique which produces an optimal prediction model using minimum training set. Using such attribute such as age, sex, blood pressure and blood sugar and find the diabetic disease like heart disease. The optimally adjusted morphological operators to be used for exudates detection on diabetic retinopathy patients no dilated pupil and low-contrast images. Using naïve bayes technique they diagnosis of diabetic retinopathy patient. Naïve bayes classifier requires small amount of training data for classification. It can be used for both binary and multi class classification problems.

Support Vector Machine (SVM), that machine learning method as the classification technique for diagnosis of diabetes with high level of performance. SVM focuses on classification of diabetes disease from high dimensional medical dataset. Data are trained by using SVM supervised learning. Advantage of support vector machine they give flexibility to diagnosis of disease. They also give unique solution. Disadvantage of using support vector machine is lack of transparency of result.The ensemble based classification technique for diabetes datasets rather than single method. Paper present three types of decision tree such as ID3, C4.5and CART are used as base classifier. Proposed work shows better performance as compared too single as well as other ensemble techniques.

The decision tree approaches with imbalance data is presented by many of the researchers, one of the contribution is done by Ali Mirza Mahmood as a comprehensive review of current methods for constructing models for learning from class imbalanced data. He also presented a critical review of the nature of the problem.

Data mining is process of extracting data from huge database. In medical system, large amount of data are present. There are many properties of data mining as Automatic discovery of patterns, Prediction of likely outcomes, Creation of actionable information, Focus on large data sets and databases. Various data mining technique such as clustering, decision tree, and association rule all these are used in data base for extracting medical data. Using such techniques diagnosis of disease is very easy task. In existing system, many classifications technique are used to diagnosis of diseases such as neural network, naive Bays, support vector machine, decision tree.

The process of discovering or extracting new patterns from large data sets involving methods from statistics and artificial intelligence. Classification and prediction are the techniques used to make out important data classes and predict probable trend. The Decision Tree is an important classification method in data mining classification. It is commonly used in marketing, surveillance, fraud detection, scientific discovery. As the classical algorithm of the decision tree ID3, C4.5, C5.0 algorithms have the merits of high classifying speed, strong learning ability and simple construction. However, these algorithms are also unsatisfactory in practical application. When using it to classify, there does exists the problem of inclining to choose attribute which have more values, and overlooking attributes which have less values.

Diabetes mellitus is one of the most serious health challenges in both developing and developed countries. According to the International Diabetes Federation, there are 285 million diabetic people worldwide. This total is expected to rise to 380 million within 20 years. Due to its importance, a design of classifier for the detection of Diabetes disease with optimal cost and better performance is the need of the age. The Pima Indian diabetic database at the UCI machine learning laboratory has become a standard for testing data mining algorithms to see their prediction accuracy in diabetes data classification. The

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 12, December 2016

proposed method uses Support Vector Machine (SVM), a machine learning method as the classifier for diagnosis of diabetes. The machine learning method focus on classifying diabetes disease from high dimensional medical dataset. The experimental results obtained show that support vector machine can be successfully used for diagnosing diabetes disease.
Gene expression profiles, which represent the state of a cell at a molecular level, have great potential as a medical diagnosis tool. Compared to the number of genes involved, available training data sets generally have a fairly small sample size in cancer type classification. These training data limitations constitute a challenge to certain classification methodologies. A reliable selection method for genes relevant for sample classification is needed in order to speed up the processing rate, decrease the predictive error rate, and to avoid incomprehensibility due to the large number of genes investigated. Improved binary particle swarm optimization (IBPSO) is used in this study to implement feature selection, and the K-nearest neighbour (K-NN) method serves as an evaluator of the IBPSO for gene expression data classification problems. Experimental results show that this method effectively simplifies feature selection and reduces the total number of features needed. The classification accuracy obtained by the proposed method has the highest classification accuracy in nine of the 11 gene expression data test problems, and is comparative to the classification accuracy of the two other test problems, as compared to the best results previously published.

III.PROPOSED APPROACH

Deep learning refers to the shining branch of machine learning that is based on learning levels of representations. Convolutional Neural Networks (CNN) is one kind of deep neural network. It can study concurrently. The proposed detailed analysis of the process of CNN algorithm both the forward process and back propagation. Then we applied the particular convolutional neural network to implement the typical diabetic dataset problem by java with weak. In addition, by measuring the actual time of forward and backward computing, analysed the maximal speed up and parallel efficiency theoretically.

A. Role of Convolutional Neural Networks

Generally, the structure of CNN includes two layers one is feature extraction layer, the input of each neuron is connected to the local receptive fields of the previous layer, and extracts the local feature. Once the local features are extracted, the positional relationship between it and other features also will be determined. The other is feature map layer; each computing layer of the network is composed of a plurality of feature map. Every feature map is a plane, the weight of the neurons in the plane are equal. The structure of feature map uses the sigmoid function as activation function of the convolution network, which makes the feature map have shift invariance. Besides, since the neurons in the same mapping plane share weight, the number of free parameters of the network is reduced. Each convolution layer in the convolution neural network is followed by a computing layer which is used to calculate the local average and the second extract, this unique two feature extraction structure reduces the resolution.
CNN is mainly used to identify displacement, zoom and other forms of distorting invariance of two-dimensional graphics. Since the feature detection layer of CNN learns by training data, it avoids explicit feature extraction and implicitly learns from the training data when we use CNN. Furthermore, the neurons in the same feature map plane have the identical weight, so the network can study concurrently. This is a major advantage of the convolution network with respect to the neuronal network connected to each other. Because of the special structure of the CNN's local shared weights makes it have a unique advantage in speech recognition and image processing. Its layout is closer to the actual biological neural network. Shared weights reduce the complexity of the network. In particular multi-dimensional input vector image can directly enter the network, which avoids the complexity of data reconstruction in feature extraction and classification process.

B. Feature Selection Algorithm

Diverse feature ranking and feature selection techniques have been proposed in the machine learning literature. The purpose of these techniques is to discard irrelevant or redundant features from a given feature vector. For the purpose of this experiment, we used feature ranking and selection methods with two basic steps of general architecture: subset generation and subset evaluation for the ranking of each feature in every dataset. Filter method was used to evaluate each subset.

- **Information Gain**

The proposed feature selection both class membership and the presence/absence of a particular term are seen as random variables and one computes how much information about the class membership is gained by knowing the presence/absence statistics as is used in decision tree induction. Indeed, if the class membership is interpreted as a random variable C with two values, positive and negative, and a word is likewise seen as a random variable T with two values, present and absent, then using the information-theoretic definition of mutual information we may define Information Gain as:

$$IG(t)=H(C)H(C|T)= \Sigma\tau,cP(C=c,T=\tau)\ln[P(C=c,T=\tau)/P(C=c)P(T=\tau)] \quad (1)$$

Here, τ ranges over {present, absent} and c ranges over {c+, c–}. As pointed out above, this is the amount of information about C (the class label) gained by knowing T (the presence or absence of a given word).

C. Back Propagation Algorithm

Convolution neural network algorithm is a multilayer perceptron that is the special design for identification of

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 12, December 2016

two-dimensional image information. Always has more layers: input layer,convolution layer, sample layer and output layer. CNN algorithm has two main processes: convolution and sampling. Convolution process: use a trainable filter Fx, deconvolution of the input image (the first stage is the input image, the input of the after convolution is the feature image of each layer, namely Feature Map), then add a bias bx, we can get convolution layer Cx. A sampling process: n pixels of each neighbourhood through pooling steps, become a pixel, and then by scalar weighting Wx + 1 weighted, add bias bx + 1, and then by an activation function, produce a narrow n times feature map Sx + 1. The key technology of CNN is the local receptive field, sharing of weights, sub sampling by time or space, so as to extract feature and reduce the size of the training parameters. The advantage of CNN algorithm is that to avoid the explicit feature extraction, and implicitly to learn from the training data; The same neuron weights on the surface of the feature mapping, thus network can learn parallels , reduce the complexity of the network; Adopting sub sampling structure by time or space, can achieve some degree of robustness, scale and deformation displacement; Input information and network topology can be a very good match, It has unique advantages in speech recognition and image processing.

$$O^{(l,k)}_{x,y} = \tanh \sum_{t=0}^{f-1} \sum_{r=0}^{K_h} \sum_{c=0}^{K_w} W^{(k,t)}_{(r,c)} O^{(l-1,t)}_{(x+r,x+c)} + Bias^{(l,k)} \qquad (2)$$

among them, f is the number of convolution cores in a feature pattern. output of neuron of row x , column y in the l th sub sample layer and k th feature pattern：

$$O^{(l,k)}_{x,y} = \tanh(W^k \sum_{r=0}^{S_h} \sum_{c=0}^{S_w} O^{(l-1,t)}_{(x\times S_h+r,y\times S_w+c)} + Bias^{(l,k)} \qquad (3)$$

the output of the j th neuron in l th hide layer H ：

$$O_{(l,j)} = \tanh(W^k \sum_{k=0}^{S-1} \sum_{x=0}^{S_h} \sum_{y=0}^{S_w} W^{(j,k)}_{(x,y)} O^{(l-1,k)}_{(x,y)} + Bias^{(l,k)} \qquad (4)$$

among them, s is the number of feature patterns in sample layer.output of the i th neuron l th output layer F

$$O_{(l,i)} = tanh(\sum_{j=0}^{H} O_{(l-1,j)} W^{l}_{(i,j)} + Bias^{(l,i)} \qquad (5)$$

D. Modified Back propagation

Backpropagation, let's warm up with a fast matrix-based algorithm to compute the output from a neural network. We actually already briefly saw this algorithm near the end of the last chapter, but I described it quickly, so it's worth revisiting in detail. In particular, this is a good way of getting comfortable with the notation used in backpropagation, in a familiar context. Backpropagation is a neural network learning algorithm. The neural networks field was originally kindled by psychologists and neurobiologists who sought to develop and test computational analogy of neurons. Roughly speaking, a neural network is a set of connected input/output units in which each connection has a weight associated with it. During the learning phase, the network learns by adjusting the weights so as to be able to predict the correct class label of the input tuples.

Output deviation of the k th neuron in output layer O:

$$d(O^0_k) = y_k - t_k \qquad (6)$$

Input deviation of the k th neuron in output layer:

$$d(I^0_k) = (y_k - t_k)\varphi(v_k) = \varphi(v_k) d(O^0_k) \quad (7)$$

Weight and bias variation of k th neuron in output O:

$$\Delta W^0_{k,x} = d(I^0_k) y_{k,x} \quad (8)$$

$$\Delta Bias^0_k = d(I^0_k) \qquad (9)$$

Output bias of k th neuron in hide layer H:

$$d(O^H_k) = \sum_{i=0}^{i++} d(I^0_k) W_{i,k} \quad (10)$$

Input bias of k th neuron in hide layer H:

$$d(I^H_k) = \varphi(v_k) d(O^H_k) \quad (11)$$

Weight and bias variation in row x , column y in the m th feature pattern ,a former layer infront of k neurons in hide layer H.

$$\Delta W^{H,k}_{m,x,y} = d(I^H_k) y^m_{x,y} \quad (12)$$

$$\Delta Bias^H_k = d(I^H_k) \qquad (13)$$

Output bias of row x , column y in m th feature pattern ,sub sample layer S

$$d(O^{S,m}_{x,y}) = \sum_{k}^{i++} d(I^H_{m,x,y}) W^{H,k}_{m,x,y} \qquad (14)$$

Input bias of row x , column y in m th feature pattern ,sub sample layer S

$$d(I^{S,m}_{x,y}) = \varphi(v_k) d(O^{S,m}_{x,y}) \qquad (15)$$

Weight and bias variation of row x , column y in m th feature pattern ,sub sample layer S

$$\Delta W^{S,m} = \sum_{x=0}^{fh} \sum_{y=0}^{fw} d(I^{S,m}_{[\frac{x}{2}],[\frac{y}{2}]}) O^{C,m}_{x,y} \qquad (16)$$

Among them, C represents convolution layer.

$$\Delta Bias^{S,m} = \sum_{x=0}^{fh} \sum_{y=0}^{fw} d(O^{S,m}_{x,y}) \qquad (17)$$

Output bias of row x, column y in k th feature patter, convolution layer C

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
IJARCCE
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 12, December 2016

$$d\left(O_{x,y}^{C,k}\right) = d\left(I_{\left[\frac{x}{2}\right],\left[\frac{y}{3}\right]}^{S,k}\right) W^{k} \quad (18)$$

Input bias of row x, column y in k th feature patter, convolution layer C

$$d\left(I_{x,y}^{c,k}\right) = \varphi(v_k) d\left(O_{x,y}^{C,k}\right) \quad (19)$$

Weight variation of row r, column c in m th convolution core, corresponding to k th feature pattern in l th layer, convolution C.

$$\Delta W_{r,c}^{k,m} = \sum_{x=0}^{fh} \sum_{y=0}^{fw} d\left(I_{x,y}^{C,k}\right) O_{x+r,y+c}^{l-1,m} \quad (20)$$

Total bias variation of the convolution core

$$\Delta Bias^{C,k} = \sum_{x=0}^{fh} \sum_{y=0}^{fw} d\left(I_{x,y}^{C,k}\right) \quad (21)$$

E. Algorithm Process flow

Step1: Select the arff dataset.
Step2: Feature selection using information gain and ranking
Step3: Classification algorithm
Step4: Each Feature calculate fx value of input layer
Step5: bias class of each feature calculate
Step6: Next produce the feature map it go to forward pass input layer
Step7: Calculate the convolution cores in a feature pattern
Step8: Produce sub sample layer and feature value.
Step9: Back propagation input deviation of the k th neuron in output layer.
Step10: Finally give the selected feature and classification results.

IV. EXPERIMENTAL RESULTS

The performance of the proposed CNN algorithm based on information gain feature selection techniques applied and compares the existing algorithm.

TABLE I COMPARISON GRAPH FOR ID3, J48, FAST ROTATIONAL FOREST, CNN FOR HEART DATASET

Algorithm	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC
ID3	0.661	0.159	0.665	0.666	0.669	0.696
J48	0.769	0.242	0.769	0.769	0.768	0.798
FRF	0.818	0.189	0.818	0.818	0.818	0.897
CNN	0.825	0.179	0.825	0.825	0.825	0.910

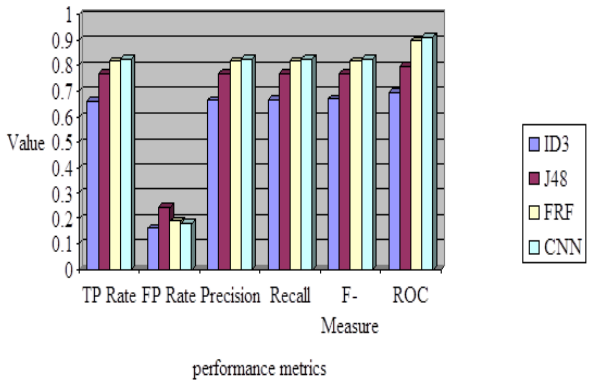

Fig. 1.Comparison Graph for ID3, J48, Fast rotational forest, CNN

TABLE III COMPARISON GRAPH FOR ID3, J48, FAST ROTATIONAL FOREST, CNN FOR PIMA INDIAN DIABETES

Algorithm	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC
ID3	0.618	0.257	0.671	0.621	0.635	0.710
J48	0.738	0.327	0.735	0.738	0.736	0.752
FRF	0.764	0.341	0.759	0.764	0.755	0.819
CNN	0.948	0.485	0.745	0.948	0.813	0.820

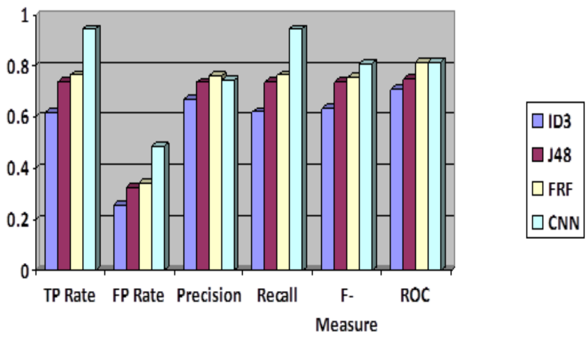

Fig. 2.Comparison Graph for ID3, J48, Fast rotational forest, CNN for pimaIndian diabetes

V. CONCLUSION

The neural network approach to generate efficient classification rules is proposed. To perform classification task of medical data, the neural network is trained using Convolutions algorithm. The experiment is conducted with heart disease dataset by considering the single and multilayer neural network modes. Convolution neural network algorithm is a multilayer perceptron that is the special design for identification of two-dimensional image information. Always has more layers: input layer, convolution layer, sample layer and output layer. In addition, in a deep network architecture the convolution layer and sample layer can have multiple.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 12, December 2016

REFERENCES

[1] Guyon I, Weston J, Barhill S, Vapnik V. Gene selection for cancer classification using support vector machines. Mach Learn 2002;46:389–422.
[2] Ding C, Peng H. Minimum redundancy feature selection from microarray gene expression data. J BioinformComputBiol 2005;3:185–205.
[3] Chuang L-Y, Chang H-W, Tu C-J, Yang C-H. Improved binary PSO for feature selection using gene expression data. ComputBiolChem 2008;32:29–38.
[4] Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell 2005;27.
[5] T A, T H, de Peer Y V, P D, Y S. Robust biomarker identification for cancer diagnosis with ensemble feature selection methods. Bioinformatics 2010;26:392–8.
[6] Y. Angeline Christobel, P.Sivaprakasam, " A New Classwise k Nearest Neighbor (CKNN) Method for the Classification of Diabetes Dataset", IJEAT, Volume-2, Issue-3, February 2013, pp. 396-400, ISSN: 2249 – 8958.
[7] Kumari V. Anuja, Chitra R. (2013). Classification of Diabetes Disease Using Support Vector Machine. International Journal of Engineering Research and Applications. Vol. 3, pp. 1797-1801, ISSN: 2248-9622.
[8] Hung, et al. "Feature selection and classification model construction on type 2 diabetic patients" data", Journal of Artificial Intelligence in Medicine, pp 251-262, Elsevier, 2008.
[9] B. R. Patel and K. K. Rana, "A Survey on Decision Tree Algorithm for Classification", International Journal of Engineering Development and Research (IJEDR), vol. 2, no. 1, pp. 1-5.
[10] S. Manohar, A. Mittal, S. Naik and A. Ambre, "A Dynamic Classifier using Decision Tree Algorithm", International Journal of Advanced Research in Computer Science and Software Engineering, vol. 5, no. 1, (2015), pp. 628-631.
[11] G. S. Babu and S. Suresh, "Meta-cognitive RBF network and its projection based learning algorithm for classification problems," Applied Soft Computing Journal, vol. 13, no. 1, pp. 654–666, 2013.
[12] G.-B. Huang and L. Chen, "Enhanced random search based incremental extreme learning machine," Neurocomputing, vol. 71, no. 16–18, pp. 3460–3468, 2008.
[13] T.Matias, F. Souza, R. Ara´ujo, and C.H.Antunes, "Learning of a single-hidden layer feedforward neural network using an optimized extreme learningmachine,"Neurocomputing, vol. 129, pp. 428–436, 2014.
[14] D. Chyzhyk, A. Savio, and M. Gra˜na, "Evolutionary ELM wrapper feature selection for Alzheimer's disease CAD on anatomical brain MRI," Neurocomputing, vol. 128, pp. 73–80, 2014.
[15] F. Han, H.-F. Yao, and Q.-H. Ling, "An improved evolutionary extreme learning machine based on particle swarm optimization," Neurocomputing, vol. 116, pp. 87–93, 2013.
[16] D. E.Goodman Jr., L. C. Boggess, andA. B.Watkins, "Artificial immune system classification of multiple-class problems," in Proceedings of the Artificial Neutral Networks in Engineering Conference (ANNIE '02), pp. 179–184, November 2002.
[17] M. Karabatak and M. C. Ince, "An expert system for detection of breast cancer based on association rules and neural network," Expert Systems with Applications, vol. 36, no. 2, pp. 3465–3469, 2009.
[18] P. Jaganathan and R. Kuppuchamy, "A threshold fuzzy entropy based feature selection for medical database classification," Computers in Biology and Medicine, vol. 43, no. 12, pp. 2222–2229, 2013.
[19] F. J. Mart´ınez-Estudillo, C. Herv´as-Mart´ınez, P. A. Guti´errez, and A. C.Mart´ınez-Estudillo, "Evolutionary product-unit neural networks classifiers," Neurocomputing, vol. 72, no. 1–3, pp. 548–561, 2008.
[20] C.Herv´as-Mart´ınez, F. J.Mart´ınez-Estudillo, andM.Carbonero-Ruz, "Multilogistic regression by means of evolutionary product-unit neural networks," Neural Networks, vol. 21, no. 7, pp. 951–961, 2008.
[21] G. Reibnegger, G.Weiss, G.Werner-Felmayer, G. Judmaier, and H.Wachter, "Neural networks as a tool for utilizing laboratory information: comparison with linear discriminant analysis and with classification and regression trees," Proceedings of the National Academy of Sciences of the United States of America, vol. 88, no. 24, pp. 11426–11430, 1991.
[22] K. Polat, S. Sahan, H. Kodaz, and S. G¨unes, "A new classification method to diagnosis heart disease: supervised artificial immune system (AIRS)," in Proceedings of the Turkish Symposium on Artificial Intelligence and Neural Networks (TAINN '05), 2005.
[23] R. Das, I. Turkoglu, and A. Sengur, "Effective diagnosis of heart disease through neural networks ensembles," Expert Systems with Applications, vol. 36, no. 4, pp. 7675–7680, 2009.
[24] H. Kahramanli and N. Allahverdi, "Design of a hybrid system for the diabetes and heart diseases," Expert Systems with Applications, vol. 35, no. 1-2, pp. 82–89, 2008.
[25] A. SlowikandM. Bialko, "Training of artificial neural networks using differential evolution algorithm," in Proceedings of the Conference on Human System Interaction (HSI '08), pp. 60–65, IEEE, May 2008.
[26] S.-W. Lin, T.-Y. Tseng, S.-Y.Chou, and S.-C.Chen, "Asimulatedannealing- based approach for simultaneous parameter optimization and feature selection of back-propagation networks," Expert Systems with Applications, vol. 34, no. 2, pp. 1491–1499, 2008.
[27] J. Yu, S. Wang, and L. Xi, "Evolving artificial neural networks using an improved PSO and DPSO," Neurocomputing, vol. 71, no. 4–6, pp. 1054–1060, 2008.
[28] S. Salcedo-Sanz, A. Pastor-S´anchez, L. Prieto, A. Blanco-Aguilera, and R. Garc´ıa-Herrera, "Feature selection in wind speed prediction systems based on a hybrid coral reefs optimization—extreme learning machine approach," Energy Conversion and Management, vol. 87, pp. 10–18, 2014.
[29] J. Kennedy and R. C. Eberhart, "Particle swarm optimization,"in Proceedings of the IEEE International Conference on Neural Networks, vol. 4, pp. 1942–1948, IEEE, Perth, Australia, December 1995.
[30] A. Frank and A. Asuncion, "UCI Machine Learning Repository," University of California, School of Information and Computer Science, Irvine, Calif, USA, 2010, http://archive.ics.uci .edu/ml..