Prediction of Complex Heart Disease Using LVQ Algorithm

Subhasis Mitra, Anju kumari, Aparna saha, Antalina Das, Sukanya Roy

Assistant Prof. ,Future Institute Of Engineering & Management , 4/5/1 Bank colony,Dhakuria,Kolkata,WB,India
Btech Student,Future Institute Of Engineering & Management ,B-type Q.no-35/206 jhanjra colony,burdwan,WB,India
Btech Student,Future Institute Of Engineering & Management , Belbagh Bettiah West Champaran,Bihar,India
Btech Student,Future Institute Of Engineering & Management , Nirmala Tea Estate M.A. Sarani Ashrampara,Siliguri,India
Btech Student,Future Institute Of Engineering & Management , A/2/1,no-2 poddernager,Kolkata,India

Abstract:
The diagnosis of heart disease at the early time is important to save the life of people as it is absolutely annoying process which requires extent knowledge and rich experience. By and large the expectation of heart infections in conventional method for inspecting reports . In case of heart disease, its diagnosis is most difficult task. It depends on the careful analysis of different clinical and pathological data of the patient by medical experts, which is complicated process. Due to advancement in machine learning, computer and information technology, the researchers and medical practitioners in large extent are interested in the development of automated system for the prediction of heart disease. In this paper we present a prediction system for heart disease using Learning vector Quantization neural network algorithm. The neural network in this system accepts 13 clinical features as input and predicts that there is a presence or absence of heart disease in the patient, along with different performance measures.

Keywords
heart disease, machine learning, learning vector quantization algorithm

1. INTRODUCTION

Today, the large population in the world is suffering from various types of diseases. The diseases related to heart, kidneys and liver is commonly observed in large number of patients. Among these patients, the count of people suffering from heart disease is increasing day by day. Large number of people in the world die every year due to heart disease. If the heart disease is detected at early stage and the patient is provided with appropriate and adequate treatment at proper stage then the life of large number of people can be saved and also the cost of the treatment can be reduced significantly. So there is a need to develop a prediction system to detect the presence or absence of heart disease in the patient with higher accuracy [1].

In this paper the prediction system is based on artificial neural network (A^N). Artificial neural network are nothing but an information processing system that process the information in the similar manner as the biological nervous system processes. Like human they learn by example and can be configured for specific application such as pattern recognition or data classification through a process of training and learning. In this prediction system we have used artificial neural network because they have ability to derive meaning from complicated or imprecise data which could be used to extract unique patterns or features and detect trends that are more complex to be noticed by either humans or other computer techniques and methods. A neural network after training could be thought of as an expert in particular category of information that has been given to it for analyze. Due to this ability the artificial neural networks are being applied to large number of real world problems. As the diagnosis of heart disease is very complex and difficult process, the medical practitioners and researchers are greatly interested in developing the automated system for diagnosis of heart disease using artificial neural network [2].

2. LITERATURE SURVEY

The number of systems for prediction of different diseases are proposed and implemented by using different techniques and methods. George et al. have proposed decision support system for dementia patients using support vector machines to define and detect agitation transition. In this system two new SVM architectures are presented, which were applied to the detection of agitation and agitation transition. This approach gives the accuracy of 91.4%, which is higher as compared with 90.9% for the traditional SVM [3].

The automated recognition of obstructive sleep apnea syndrome by using SVM classifier is proposed by Haitham and Alan. In this approach, for the detection of the syndrome they computed the features from the phase and magnitude of the signals obtained from thoracic and abdominal respiratory effort and evaluated the classification of whole night normal and apneic epochs. The highest accuracy of this system is 95% [4]. A system for diagnosis of heart disease that is based on support vector machine along with sequential minimal optimization algorithm is presented.

Manuscript received November 5, 2025
Manuscript revised November 20, 2025
https://doi.org/10.22937/IJCSNS.2025.25.11.16

In this system the network structure of Radial Basis Function is also used and it is trained using Orthogonal Least Square algorithm and applied to the dataset based on Indian patients. The result shows that the SVM is equivalently as good as compared to Radial bias function in detection of heart disease with accuracy of 86.42% [5].

A system for classification of myocardial heart disease from ultrasonic images by optimizing the fuzzy membership functions by using genetic algorithm based method is proposed by Tsai and Watanabe. In this method by using the texture features obtained from ultrasonic images, the gaussian distributed membership function is

Genetic algorithm is also used in another approach by Anbarasi et al. where number of tests that are to be conducted by patient is reduced by determining the attributes that involved in the prediction of heart disease. In this approach three classifiers were used and these classifiers were fed with reduced attributes, but the system takes more time for model construction [7].

Haung and Wang have proposed a system using Support Vector Machine where feature selection and optimization of Support Vector Machine parameter is done using genetic algorithm. This system uses less number of input parameters for support vector machine and it is evaluated on 11 real world datasets with improved accuracy of 89.6% [8].

Rajkumar and Sophia have proposed the use of Data mining algorithm in diagnosis of heart disease with an accuracy of 52.33%. They have combined the ECG attributes and clinical symptoms to detect the heart disease. The algorithms used by this system are Naive Bayes algorithm, Decision list algorithm and KNN algorithm [9].

Due to the higher accuracy and learning rate the artificial neural network k(ANN) algorithms can also be used in the prediction of heart disease [10]. Kumaravel et al. have proposed automatic diagnosis system for heart diseases using neural network. In this system ECG data of the patients is used to extract features and 38 input parameters are used to classify 5 major types of heart diseases with accuracy of 63.6 - 82.9% [11].

Chen et al. have proposed the prediction system for heart disease using learning vector quantization algorithm. This heart disease prediction system can help the medical practitioners in evaluating a patient's heart disease based on the clinical data of the patient. The approach includes three steps. Firstly they have selected 13 important clinical features, and then developed an artificial neural network algorithm for classifying heart disease based on these clinical features. The accuracy of prediction is near 80%

constructed and the genetic algorithm based fuzzy classifier is used in classification. In this technique 96% of classification accuracy is achieved [6].

In the system for heart disease prediction, Learning vector quantization (LVQ) neural network algorithm is used. The main purpose of using LVQ is that it creates prototypes that are easy to interpret for experts in the respective application domain. LVQ network is a nearest neighbor pattern classifier based on competitive learning. It is generally applicable in problems of M

A. Learning vector quantization (L VQ) Algorithm

Learning vector Quantization is a special case of competitive network, which uses supervised learning methodology. It is the well known algorithm that deals with the process of classifying the patterns or selecting prototypes. LVQ network is a nearest neighbor pattern classifier based on competitive learning. It is generally applicable in problems of non-linear separation and used for data classification in large extent.

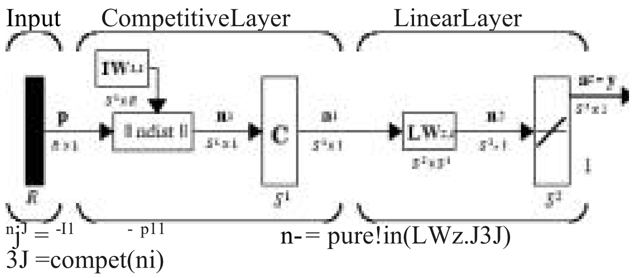

Fig. 1. LVQ Network Architecture [13]

3. METHODOLOGY

In the system for heart disease prediction, Learning vector quantization (LVQ) neural network algorithm is used. The main purpose of using LVQ is that it creates prototypes that are easy to interpret for experts in the respective application domain. LVQ network is a nearest neighbor pattern classifier based on competitive learning. It is generally applicable in problems of nonlinear separation and used for data classification in large extent. In this paper the system consists of two steps, in the first step 13 clinical attributes are accepted as input and then the training of the network is done with training data by LVQ algorithm [12].

The architecture of LVQ is similar to the Kohonen self organizing map, but the main difference is that in LVQ, the topological structure at the output unit is not considered. The LVQ network architecture is shown in figure 1. Where R is the number of elements in input vector, SI is the number of competitive neurons and S2 is the number of linear neurons. As shown in figure 1, LVQ network has a competitive layer and a linear layer. When the input vector is given to the competitive layer it firstly learns to classify that input vectors in to classes in the same way as the competitive layers of cluster with self-organizing map neural network. The linear layer then transforms classes of competitive layer into target classifications that is defined by the user. The classes learned by the competitive layer are referred to as subclasses and the classes of the linear layer as target classes. The layers in LVQ have one neuron per class that is competitive and linear layers have one neuron for each class. Thus, there can be total S1 subclasses learn by competitive layer and S2 target classes that are formed by linear layer by combining the S1 subclasses from competitive layer, such that SI is always larger than S2.

For example, suppose that in competitive layer there are three neurons 1, 2, and 3 and all of them learn subclasses of the input space that belongs to the linear layer target class 3. Then competitive neurons 1, 2, and 3 will have linear weights of 1.0 to neuron 3 in the linear layer, and weights of 0 to all other linear neurons. Thus, the linear neuron produces a 1 if any of the three competitive neurons (1, 2, or 3) wins the competition and outputs a 1. This is how the subclasses of the competitive layer are combined into target classes in the linear layer.

The LVQ network is made up of two layers that are a competitive layer and a linear layer. In the competitive layer, each neuron is assigned to a particular class. The number of classes in the competitive layer is determined by the number of neurons in hidden layer. There may be number of different neurons in the competitive layer that can be assigned to the same class. Each of those classes is then assigned to one of the neuron in the linear layer. The number of neurons in the competitive layer are always greater or equal to the number of neurons in the linear layer. In the competitive layer, the neuron learns a prototype vector which allows it to classify a region of the input space. Euclidean distance is used to measure the similarity between the input vector and any of the weight vectors. Since some of these classes may be identical, they are really subclasses. The second layer (the linear layer) of the LVQ network is then used to combine subclasses that are formed by competitive layer into a single class. This is done by using the weight matrix. This process is known as ’’training” of neural network. The algorithmic steps are as follows [2]:

The algorithm is as follows:

Step 1: Initialize weights (reference) vectors.

Initialize learning rate.

Step 2: While stopping is false, do step 3-7

Step 3: For each training input vector x, do step 4-5

Step 4: Compute j using squared Euclidean distance $D(j) = I(Xi - Wij)^2$

Find j when D (j) is minimum

Step 5: Update Wj as follows:

If t= Cj then

$^{w}j(new) = {}^{W}j(old) + {}^{a[x-w}j(old)^{]}$

If t 4 CJ , then

$^{w}j(new) = {}^{w}j(old)^{-a[x-w}j(old)^{]}$

Step 6: Reduce the learning rate.

Step 7: Test for stopping condition

The condition may be fixed number of iteration or the learning rate reaching a sufficiently small value. In this prediction system the Cleveland heart disease database is used to fed the input to neural network. The network is having three layers. The learning vector quantization algorithm is trained with random order incremental training. In the input layer of the network there are 13 neurons that accept the 13 values of clinical information from the heart disease database. The hidden layer neurons can be varied in order to reduce error and increase accuracy. The output layer consists of single neuron that indicates whether the heart disease is present or absent.

4. EXPERIMENTAL RESULTS

A. *Data Source*

The performance of the system is evaluated on Cleveland heart disease database that was taken from dataset repository of UCI. This database consists of 303 records with each having 13 clinical attributes that include age, sex, type of chest pain, resting blood pressure, cholesterol, fasting blood sugar, resting ECG, maximum heart rate, exercise induced angina, old peak, slope, number of vessels coloured and thal respectively. In this database out of 303 records 164 belong to healthy category and 139 belong to heart disease [14].

B. *Performance Evaluation*

The system for prediction for heart disease using multilayer perceptron neural network is implemented in MATLAB R2012b. In this system the database is divided in to two sets randomly that is training set and testing set. Out of total records 80% records are used for training and testing is done by using remaining 20% records. The evaluation of performance of the system is done by computing the percentage value of different parameters like Accuracy, Specificity and Sensitivity by using following equations [15]:

Sensitivity=TP/(TP+FN)

Specificity=TN/(TN+FP)

Accuracy=(TN+TP)/(TN+TP+FN+FP)

Where, TP = number of samples classifies as true while they were true.

TN = number of samples classifies as false while they were actually false

FN = number of samples classifies as false while they were actually true.

FP = number of samples classifies as true while they were actually false.

C. *Performance with Different Number of Neurons*

The performance with varying number of neurons of the system is shown in table I. Out of 303 instances the network is trained with 242 instances and remaining 61 instances are used for testing. The system gives the highest accuracy of 85.55% for 18 neurons in hidden layer with 100 iterations. The different performance measures like TP, FP, TN and FN for different number of neurons is shown in table II.

Table 1: PERFORMANCE OF THE SYSTEM WITH DIFFERENT NUMBER OF NEURONS

No. of Neurons	Acc	Sens.	Spec.	Error
6	74.74	68.11	76.41	25.93
8	78.51	72.98	79.86	21.49
10	79.62	74.60	84.02	20.38
12	81.85	80.95	82.62	18.15
14	82.96	82.53	83.33	17.04
16	84.84	80.95	87.5	15.56
18	85.55	83.33	87.5	14.45

of total samples 242 samples are used for training and remaining 61 samples are used for testing.

Table 2: PERFORMANCE OF THE SYSTEM WITH DIFFERENT NUMBER OF EPOCHS

No. of Epochs	Acc	Sens.	Spec.	Error
100	81.85	79.36	84.02	18.15
125	82.59	80.15	84.72	17.41
150	84.81	83.33	86.11	15.19
200	85.18	84.15	89.58	14.82

Table 2 shows the system performance with varying number of epochs with highest accuracy of 85.18%.In order to compute the accuracy and performance of the prediction ayatem,the number on neurons in hidden layeris set to 6. The HDPS:Heart Disease Prediction System [12] developed An artificial neural network using LVQ algorithm for classifying heart disease based on the 13 clinical features with accuracy of near about 80%.

5. CONCLUSIONS

In this paper an efficient system for prediction of heart disease is developed. The system accepts the 13 clinical features as input and the training of network is done by using Learning Vector Quantization algorithm. It gives the presence or absence of heart disease. The system also computes different parameters like accuracy, sensitivity, specificity, training time, testing time, error and ROC curve. To improve the performance of the prediction system, it is trained with different number of neurons and also the number of training epochs are varied. The results shows that our system when compared with others gives the highest accuracy of 85.55%.

The prediction system for heart disease can be enhanced to incorporate other neural network algorithms so that the accuracy can again be increased and it can be deployed in large number of health care centers.

REFERENCES

[1] "Medline plus: Heart diseases," http://www.nlm.nih.gov/medlineplus/heartdiseases.htm l
[2] S.N.Sivanandam and S.N.Deepa, "Principles of soft computing" ,Wiley India Edition, 2007.
[3] G. E. Sakr, I. Elhajj, and H. Huijer, "Support vector machines to define and detect agitation transition", IEEE Transactions On Affective Computing, vol. I, pp. 98-108, December 2010.
[4] M. Haitham, A. Angari, and A. Sahakian, "Automated recognition of obstructive sleep apnea syndrome using support vector machine classifier", IEEE Transactions On Information Technology In Biomedicine, vol. 16, pp. 463-468, May 2012
[5] Jayshril S. Sonawanel, D. R. PatiI, "Prediction of Heart Disease Using Learning Vector Quantization Algorithm", 978-1-4799-3064-7/14/$31.2014 IEEE.
[6] D. Y. Tsai and S. Watanabe, "Method for optimization of fuzzy reasoning by genetic algorithms and is application to discrimination of myocardial heart disease", In proceedings of IEEE Nuclear Science Symposium and Medical Imaging Conference (NSSMIC),pp. 2239-2246, December 1966.
[7] M. Anbarasi, E. Anupriya, and N. Iyengar, "Enhanced prediction of heart disease with feature subset selection using genetic algorithm", International Journal of Engineering Science and Technology, vol. 2, pp. 5370-5376, November 2010.
[8] C. L. Huang and C. J. Wang, "A ga-based feature selection and parameters optimization for support vector machines", Expert Systems with Applications, vol. 31, pp. 231-240, October 2006.
[9] A. Rajkumar and G. S. Reena, "Diagnosis of heart disease using datamining algorithm", Global Journal of Computer Science and Technology, vol. 10, pp. 38- 43, December 2010.
[10] W. G. Baxt, "Application of artificial neural networks to clinical medicine", Lancet, vol. 346, pp. 1135-1138, October 1995.
Fig. 4. ROC Curve
[11] N. Kumaravel, K. Sridhar, and N. Nithiyanandam, "Automatic diagnoses of heart diseases using neural network", In Proceedings of the Fifteenth Biomedical Engineering Conference, pp. 319-322, March 1996.
[12] A. Chen, S.Huang, and E.Lin, "Hdps: Heart disease prediction system", In proceedings of IEEE Conference on Computing in Cardiology, pp. 557-560, September 2011.
[13] "Dtreg", http://www.dtreg.com/mlfu.htm.
[14] UCI, "Uci machine learning repository: Heart disease data set", http://archive.ics.uci.edu/ml/datasets/Heart+Disease.ht ml.
[15] I. Dessai, "Intelligent heart disease prediction system using probablistic neural network", International Journal of Advanced Computer Theory and Engineering, vol. 2, pp. 3844, August 2013.
[16] Sunila, P. Panday, and N. Godara, "Decision support system for cardivascular heart disease diagnosis using improved multilayer perceptron", International Journal of Computer Application, vol. 45, pp. 12-20, May 2012.