

A Hybrid Machine Learning Classification Algorithm for Medical Science

Swarnendu Kundu, Deblina Banerjee

P.G. Student, SCOPE, VIT UNIVERSITY, VELLORE,India

P.G. Student, Department of Information Technology,GCECT,MAKAUT,KOLKATA,India

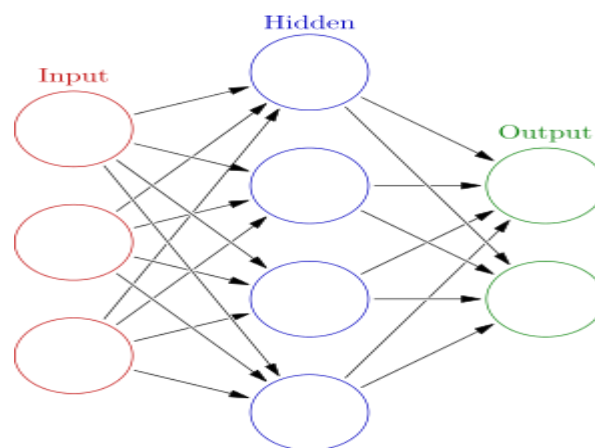
ABSTRACT: Machine learning plays a vital role in the digital world. Its works efficiently on the medical science. There are many classification algorithms to classify the data or predict the data, it may be in medical image, medical dataset etc. But in the classification algorithm, features selection is play a key role to predict or classify the data. In real time, Medical Dataset are very huge and also in high dimension. So, it works slow in learning rate and also higher cost in computational. Feature selection is expected to deal with the high dimensionality of datasets in terms of reduced feature set. In this paper we are merge, artificial neural network (ANN) for prediction or classification and Genetic Algorithm is used for the feature selection. At-last we are compare with the other classification algorithms like Random forest, KNN, Support Vector Machine(SVM).

KEY WORDS: Machine Learning, Genetic Algorithm (GA), Artificial Neural Network (ANN).

I.INTRODUCTION

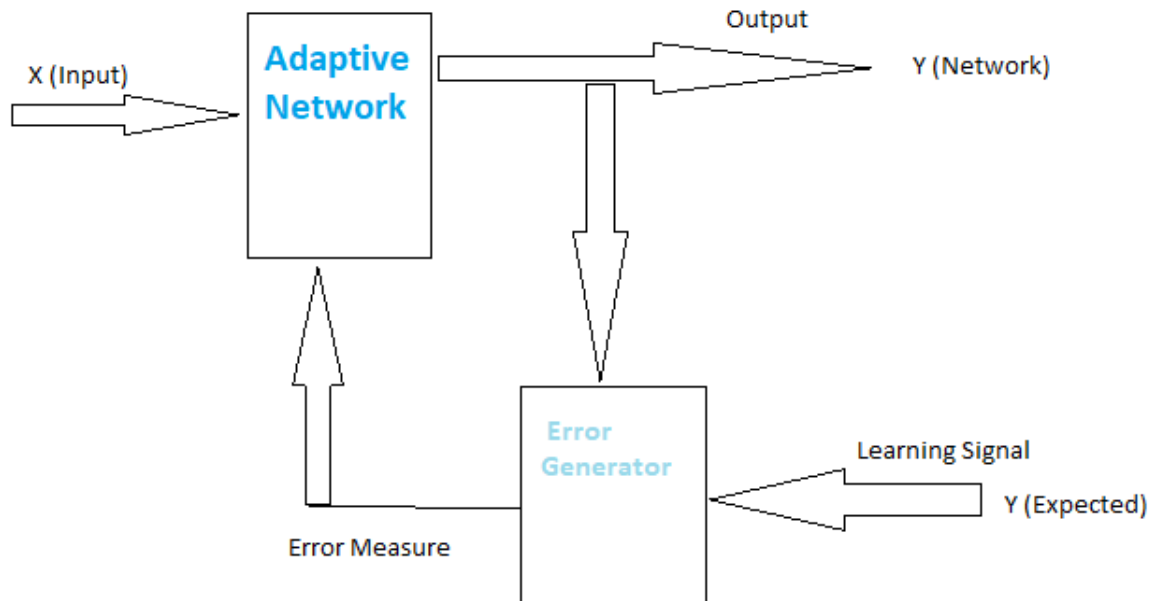
Now a days extract the information from the medical dataset with the help of Machine learning. Machine learning tasks are typically classified into three categories i) “Supervised learning” ii) “Unsupervised learning” iii) “Reinforcement learning”. Supervised learning is inferring a function with labeled training data. Unsupervised learning is inferring a function with no labeled training data. In this Paper, we are focused on supervised learning for labeled data. We are merge Artificial neural network (ANN) for classification and Genetic algorithm for features selection.

A . Artificial Neural Network(ANN):ANN collect the information by identify the patterns and relationships in data and trained through the experience. It contains some weights of every connection from one node to another node. There are 3 components in Artificial neutral network. (i) Input (ii) Hidden Layer (iii) Output layer. It is classified into two: (i) Forward Propagation (ii) Backward Propagation.

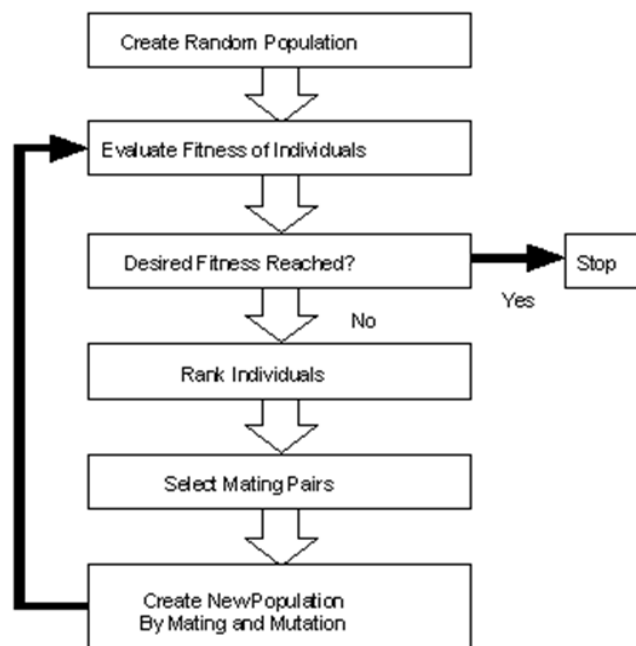


Fig(i): ANN Diagram

In our Paper , we are used Supervised network with back-propagation learning rule model. Because, it is a well organized algorithm for computation of gradients. It fixed the error of the Output error of the Neural network and the actual output. Fixing the weights or by finding better activation function with a good stable derivate.



B. Genetic Algorithms(GA):It is used for generating good quality solution for optimization and search problem. Here we are used as feature selection of the dataset. The operators of GA are mutation,crossover and selection.

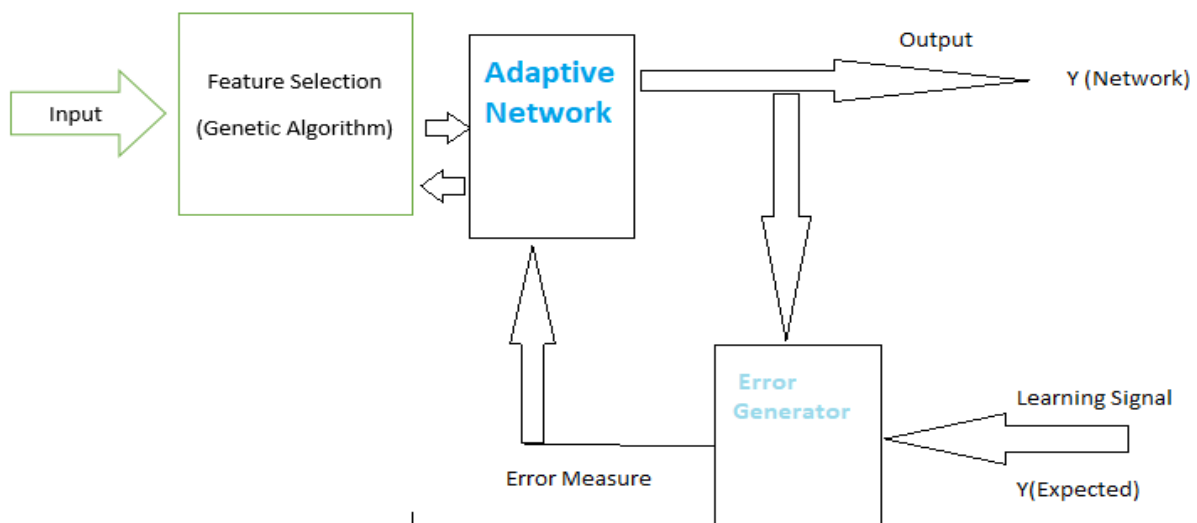


Fig(iii):Typical Genetic Algorithm Flowchart

II. LITERATURE SURVEY

Class decomposition ^[1] , Neural are adjust by each incoming link and classification of non-linear. Here in this paper ,genetic algorithm is used for optimizing Random forest. M. Bader-El-Den^[8] says that each chromosome has a RF(Random_forest) solution with different trees. Here number of feature is not addressed for optimization. The number of trees are not optimized. But a variable length chromosome are used, for allowing navigation in this solution space. However, result is good. Azer^[5] tested with the medical dataset with the support vector machine(SVM) . He decided that LPSVM is good in “diagnosis aid”. Azer^[5] proposed a hybrid model of random forest(RF) and Genetic algorithm(GA). He used genetic algorithm(GA) as a feature selection before applying the random forest for optimization. He used lymph data set. However result is comparatively good. Burton^[13] presented compare with the ANN and SVM for predicting and classification. He use the breast cancer dataset. Result is comparatively good.

III. METHODOLOGY



Fig(iv): Hybrid Model of ANN and GA

The above hybrid model GA is used a Feature Selection of the input X and then send it to ANN . ANN is proceed for prediction or classification. If there is any error of the Output and the expected output ,the back-propagation neural networks held the error by fixing the weights or by finding better activation function with a good stable derivate.

Algorithm:

- Step 1:** Input X (x1,x2,...,x_n) attributes
- Step 2:** Ga= GA(fitness=X) // Genetic algorithm(GA) Function for feature selection.
- Step 3:** hidden_layer= No. of hidden layer of this model;
- Step 4:** Nn= neuralnet(Label~Ga, traindata, hidden_layer);
- Step 5:** Rmse= Rmse(Nn, expect_output);
- Step 6:** Adjust the weight or activation function; // Backpropagation Neutral network
- Step 7:** Loop from step3 to step4 untill its Convergence.
- Step 8:** Plot(Nn);

Dataset Description

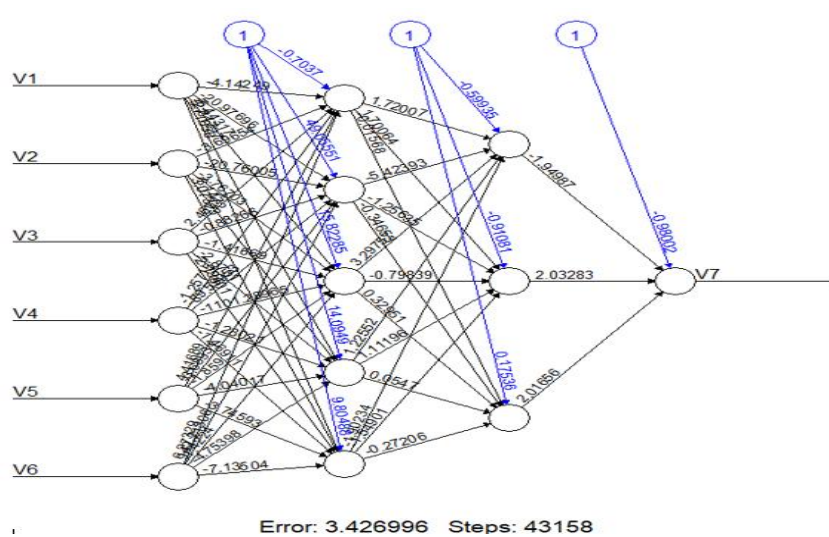
We are using Heart disease data set, which contains 14 Columns and 281 Rows. It is available in the UCI repository.

IV. EXPERIMENTAL RESULTS

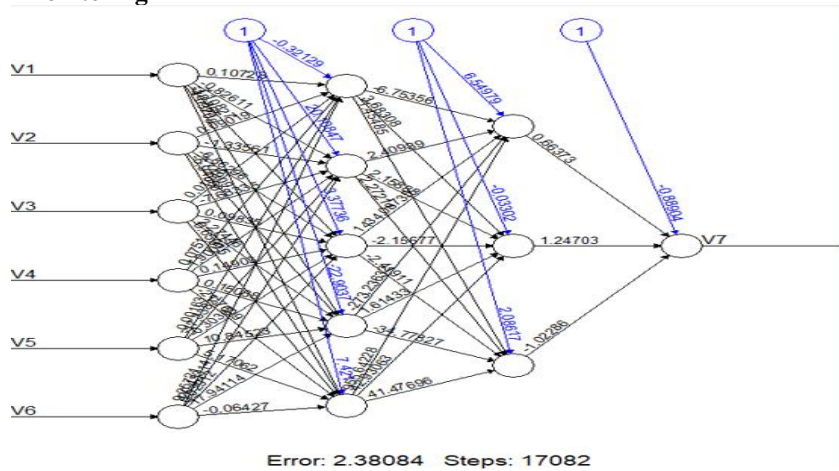
Table 1: Result of Hybrid Model

ParaMeter of Genetic Algorithm	Result
Type	Real-Valued
Population Size	50
No. of Generation	100
Elitism	2
Crossover Probability	0.8
Mutation Probability	0.1
Iteration	100
Fitness function	47.7
No. of Hidden Layer	3

A) Result of GA Monitoring=1



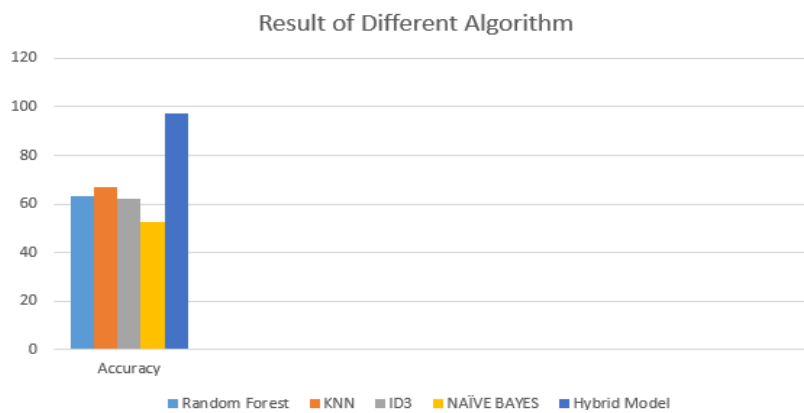
B) Result of GA Monitoring=2



Fig(A) and Fig(B): Result of the Hybrid Model

Table 2: Compare with the Other Classification Algorithm

Classification Algorithm	Accuracy Result
Random-Forest	63.3%
<i>K-Nearest-Neighbors (KNN)</i>	66.9%
ID3	62.3%
Naïve Bayes	52.3%
Hybrid_Model_of_Artificial_Neuron_Network and Genetic Algorithm	97.3%



Fig(v): Graphical Representation

Our Proposed model are compare with the different types of classification algorithm like ID3,Random Forest,KNN,Naïve Bayes. KNN are not gives the good result. But Random forest and KNN are good then the Native Bayes. However, our proposed model gives the most significant result.

V.CONCLUSION AND FUTURE WORK

This hybrid model is working efficiently better compare to other classification algorithm. Random forest is very good classification algorithm for large dataset, but it is not good for the small dataset. Random forest is also taken huge time for classification. Here we, select the features through genetic algorithms and a make classification or prediction through Artificial Back-propagation neutral network. In future, we are trying to merge deep learning with Genetic algorithm for better results.

REFERENCES

[1] E. Alfaro , M. Gámez , N. García , adabag: An R package for classification with boosting and bagging, J. Statistical Software 54 (2) (2013) 1–35 .
 [2] R. Analytics, S. Weston, doParallel: Foreach parallel adaptor for the parallel package, 2014 . R package version 1.0.8, URL <http://CRAN.R-project.org/package=doParallel> .
 [3] B. Antal, A. Hajdu, An ensemble-based system for automatic screening of diabetic retinopathy, Knowl. Based Syst. 60 (2014) 20–27, doi: 10.1016/j.knosys.2013.12.023 .
 [4] A.T. Azar , S.M. El-Metwally , Decision tree classifiers for automated medical diagnosis, Neural Comput. Appl. 23 (7–8) (2013) 2387–2403 .
 [5] A .T. Azar , S.A . El-Said , Performance analysis of support vector machines classifiers in breast cancer mammography recognition, Neural Comput. Appl. 24 (5) (2014) 1163–1177 .
 [6] A.T. Azar , H.I. Elshazly , A.E. Hassaniien , A.M. Elkorany , A random forest classifier for lymph diseases, Comput. Methods Prog. Biomed. 113 (2) (2014) 465–473 . [7] K. Bache, M. Lichman, UCI machine learning repository, 2013 . <http://archive.ics.uci.edu/ml> . [8] M. Bader-El-Den , M. Gaber , Garf: towards self-optimised random forests, in: Neural Information Processing, Springer, 2012, pp. 506–515 .
 [9] I. BoussaïD , J. Lepagnot , P. Siarry , A survey on optimization metaheuristics, Inf. Sci. 237 (2013) 82–117 .
 [10] L. Breiman , Bagging predictors, Mach. Learn. 24 (2) (1996) 123–140 .
 [11] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32, doi: 10.1023/A:1010933404324 .
 [12] M. Charytanowicz, J. Niewczas, P. Kulczycki, P.A. Kowalski, S. Łukasik, S. Zak, Information technologies in biomedicine: Volume 2, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 15–24. doi: 10.1007/978-3-642-13105-9_2 .
 [13] M. Burton, M. Thomassen, Q. Tan, and T. A. Kruse, “Gene expression profiles for predicting metastasis in breast cancer: a cross-study comparison of classification methods,” The Scientific World Journal, vol. 2012, Article ID 380495, 11 pages, 2012