International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-2, July 2019

Neural Computation based General Disease Prediction Model

S. Prince Mary, B. Bharathi, Vigneshwari. S, Sathyabama R

Abstract*: To predict the patient disease using soft computing technique is the primary motto of the disease prediction system. Currently, researchers are trying to develop a disease prediction system using pattern mining technique. Here, a technique for disease prediction system using genetic algorithm and artificial neural network is proposed. The genetic algorithm is used for mining the most occurrences of disease sequences rules. To form the disease prediction system, the best rule which is obtained by means of genetic algorithm is used. Artificial neural network is trained to predict the disease. Accuracy of disease prediction is compared with other prediction techniques.

***Keywords*: About four key words or phrases in alphabetical order, separated by commas.**

I. INTRODUCTION

T he process of mining data for finding useful patterns is known as data mining. Data mining techniques are used in all fields. Nowadays usefulness of DMT is executed in Healthcare field. The abstruse medical data are mined for hidden patterns using DMT. The data mining techniques which are applied to find frequent patterns are association rule mining, prediction, classification and clustering. Traditional DMT were used in different domains. DMT is used for various discoveries in medical data such as accurate disease diagnosis, different disease severe level to predict and monitoring health remotely. The special purpose of DMT is to predict cancer, heart disease and so on. [1],[2], Apriori and FPGrowth[3],[4],[5], Unsupervised neural network [6],[7], decision tree algorithms like ID3,C4.5,C5, and CART[8],[9],[10], Navie Bayesian network[8] are used. Soft Computing techniques such as Fuzzy logic, Neural Networks, Genetic Algorithms are used in all computational fields, nowadays these techniques are applied in medical field from mining disease related information. Neural networks are a pivot or group of interconnected artificial neurons. Mining knowledge happens either using a mathematical model or a computational model [11][12]. Fuzzy logic that formulate in an inaccurate reasoning; this system works based on the universe of uncertainty. The core part of the fuzzy logic is linguistic variable [13],[14]. Genetic algorithm works based on the Darwin's "survival of fittest". It is an evolutionary algorithm used for optimization of problems. Genetic algorithm is the best algorithm for large search space and for its optimized solutions.

In existing systems diseases are predicted by various tests or trial and error methods. In the proposed DPS (Disease Prediction System) disease are predicted once the patients get the symptom of the disease. Various patience long term diseases are identified and the data set is formed. Genetic algorithm is applied to find the most occurrence of the disease sequence of various patients. These disease sequences are used to predict the disease using ANN.

II. RELATED WORK

Data mining is the process of mining knowledge from the data. DMT took an important role in information retrieval systems for knowledge discovery from historical data. Currently DMT is used to increase the QoS in medical field by discovering data about diseases, patients, medicine, and diagnosis and so on. DMT is more useful in applications like medical tests, medications, to find the relationship between clinical data and pathological data [15]. FPGrowth and Apriori are the widely used frequent pattern mining algorithms in medical data [16]. Huge volume of by healthcare systems is very complex to be processed and analyzed by traditional techniques. The decision making can be improved with the help of soft computing to discover patterns from the large volume of information. There are several ways to predict the future disease patterns. In [17] an intelligent prediction system for heart diseases using CANIF and Genetic algorithm is introduced, it is a quantitative approach integrated with genetic algorithm. To improve the performance genetic is used, but its computational cost is high. In [18] Genetic algorithm is used for the prediction of heart disease via optimal set of attributes. Optimal set of attributes are generated by genetic algorithm. It needs two steps to predict the disease. [19] Also introduced a heart disease prediction approach using genetic algorithm. Still only the existing patient health is predicted. [20] Used Navie Bayes classifier in medical applications. Two of the most used algorithms used to calculate the probability of objects based on the previous experience. Heart disease is predicted based on multiple techniques. [21] Using associative classification and genetic algorithm used for heart disease to predict.

Revised Manuscript Received on 30 July 2019.
***** Correspondence Author
S.Prince Mary*, Department of Computer Science Engineering, Sathyabama Institute of Science and technology, Chennai- Tamilnadu India. Email: princemary26@gmail.com,
B.Bharathi, Department of Computer Science Engineering, Sathyabama Institute of Science and technology, Chennai- Tamilnadu India.. Email: bharathivaradhu@gmail.com
S.Vigneshwari*, Department of Computer Science Engineering, Sathyabama Institute of Science and technology, Chennai- Tamilnadu India.. Email: vikiraju@gmail.com
R.Sathyabama, Department of Computer Science Engineering, Sathyabama Institute of Science and technology, Chennai- Tamilnadu India.. Email: rsathyarajeswari@gmail.com

This approach used less number of attributes to predict the disease. Testing of various measures that may relate to the disease. This is used the list of diseases and the symptoms. In TB diagnostic the recent advancements explained.

Soft computing techniques are efficient techniques which can find similar patterns computationally and they can be implemented easily. Among the entire techniques genetic algorithm is the one which is used by many researches for heart disease prediction with the other techniques, which are implemented in two or more steps. In this paper Genetic algorithm alone is used because of its optimal solution generation. ANN is a type of computational design based on the structure and functions of biological neural networks. Used ANN to predict diabetes disease and the comparison with other techniques done and given as ANN gives the best prediction accuracy. Proposed a software based disease prediction system using RBF using sensitivity, time of disease prediction and accuracy measurements.

III. PROBLEM DEFINITION

In order to predict the disease once the patient gets the symptoms in an efficient way, to increase the execution performance of the DPS in term of accuracy, an optimal approach is applied which can find an optimal sequence of diseases. A prediction system is implemented using ANN, because of it produce high accuracy prediction level.

A. *Genetic Algorithm*

Genetic algorithm is an optimal search methodology, based on the Darwin natural selection and genetics biological systems. It works with a group of candidate solutions known as a population. Result of genetic algorithm is an optimal solution one generated by a series of repeated computations. Genetic algorithm is one of the evolutionary algorithms. It provides an optimized result using selection, crossover and mutation steps. Need of genetic algorithm:

- A solution domain which brings the genetic representation.
- To evaluate the solution a fitness function.

An array bits is the representation of the solutions. To measure the quality of the solution a fitness function is used. The fitness function depends upon the problem. Initial population is selected randomly. During the next generation proportion of the population is selected for the new generation. Based on the fitness next solution is selected. Next step is to perform crossover and mutation of the solutions.

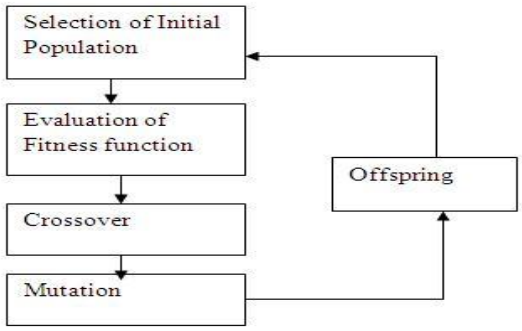

Fig 1. Dataflow Diagram of Genetic Algorithm

Algorithm:

1. Read the patient dataset
2. validate the dataset to make the input format for the Genetic algorithm
3. Optimal solution is obtained from GA and it returns the personal good and global good values.
4. Calculate pGood and gGood on each component of the dataset divided based on the dataset size and block size.
5. Apply fitness function
   F=Sup(X) –Minimum support/ SQRT (Min sup*(1-minsup))/Num
   Sup(X)- pattern support.
6. Based on the value of F find the rules. The rules of highest fitness are saved.
   Apply genetic on these rules.
7. Perform crossover on the selected nodes called as pivot node.
8. Perform mutation based on the pivot node. This process is repeated till the last generation is reached.
9. After performing GA on each index, calculate the gGood for the optimal solution.
10. Using the GA generated rules build the classifiers on the basis of gGood.

B. *Prediction Algorithm*

ANN is capable of learning by altering the weights. Artificial neural network is used to read the training data set rules generated by the genetic algorithm. There are three layers present, they are

Input Layer: It defines all the input attribute values and their probabilities [26].

Hidden layer: It receives the input from the input layer, at this layer weights are assigned for each input. A weight shows the relevancy of the input to the hidden neuron. Neuron with high weight is assigned to the input. These weights can be negative also. Negative weight shows that input can restrain, rather than support, a particular result.

Output Layer: output layer neurons represent the output as the predictable parameter values for the Data Mining model.

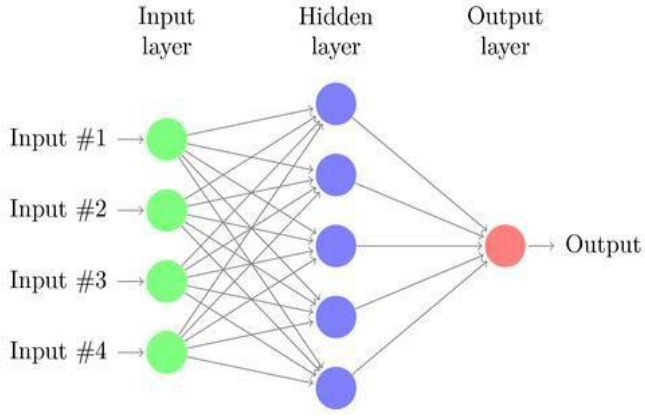

Fig 2. Artificial Neural Network Structure

Supervised ANN is used to read the rules generated by the genetic algorithm, for each neuron input weights are assigned at the hidden layer, the it produce the predicted result according to the learnings.

IV. RESULTS AND DISCUSSIONS

Objective of this work is to detect any disease using soft computing technique a genetic algorithm and feed forward ANN was adopted to develop a disease prediction system. The implementation was developed with a genetic algorithm using Mat Lab. Genetic algorithm produces the prediction rules for the given data set. These rules are applied to train the feed forward ANN. ANN generates the predicted diseases. ANN is trained with the data various results obtained based on the accuracy of disease prediction and prediction time.

Genetic algorithm is used for the purpose of optimization of the healthcare data for monitoring any disease. The proposed model tested using synthetic data set. The data set provide the patience disease with the symptoms. Table 1.Shows the attributes of the dataset used. 200 symptom identified and its respective disease taken to test the proposed system. Performance of the proposed system is evaluated using three parameters such as sensitivity of disease detection, detection time and accuracy. Sensitivity is an input given at runtime. It varies for each patient. To control the sensitivity of the system error value is chosen. When the error value increased for the network its sensitivity reduces. Hence sensitivity and error are inversely proportional attributes.

Table I: Sample Attributes of the dataset

Attribute No.	Attribute
A1	Patient Id
A2	Symptom Id
A3	Disease Id

Table II: Describe the list of symptoms with its id. Symptoms are grouped based on the disease name.

Table II: Sample of Symptoms

Symptom	Symptom name
S1	Abdominal Chest Pain, Fatique, Fever, Joint pain, Stomach Pain
S2	Blood instool, Chest pain, Dark Urine, Dizziness
S3	Frequent watery motion, vomiting, deydration, swollen feet
S4	Blood instool, Chest pain, Dark Urine, Dizziness
S5	Cold sores, wight gain, stomach pain, back pain

Fig.3 shows the sample output for the same case 5 with the sensitivity of values of 30%,50%, and 90% respectively. Table shows that when the sensitivity increases the number of disease decreases.

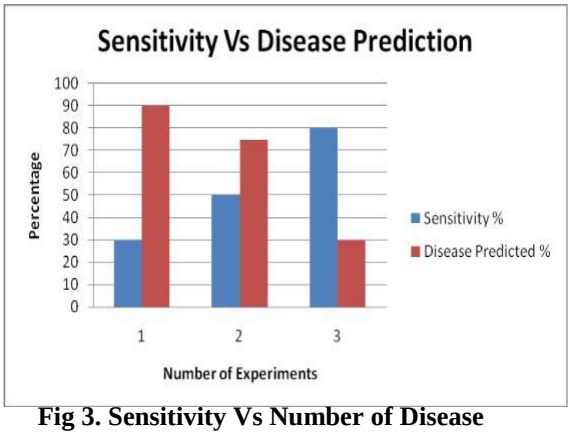

Fig 3. Sensitivity Vs Number of Disease Predicted

Comparison of disease prediction using different algorithm such as GA trained ANN compared with other soft computing techniques. Table III. Shows that the proposed GA trained ANN accuracy in prediction is high than the other techniques

Table III: Comparisons of proposed technique with other prediction techniques

Techniques Applied	Prediction Accuracy
GA trained Feed forward ANN	98.45
Feed Forward ANN	97.46
SVM	75
KNN	78.32

Fig 4. Show the performance comparison graph of proposed technique with the various disease prediction techniques. It shows that the proposed techniques prediction accuracy is more than the existing prediction techniques

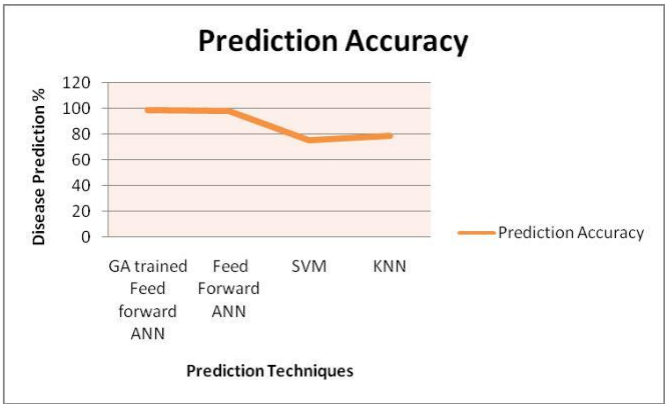

Fig 4. Comparison of Prediction Accuracy with various Techniques

Accuracy of the disease prediction is evaluated and compared with other soft computing techniques. Genetic based feed forward ANN disease prediction system generates best accurate prediction of diseases than the other techniques

V. CONCLUSION

In this paper we have proposed a technique for disease prediction using genetic algorithm based feed forward neural network. The proposed technique contains two phases as mining optimum rules for diseases using genetic algorithm and then feed forward artificial neural network is trained using the result of genetic algorithm. Feed forward ANN predicts the disease of any patient by giving their symptoms to the network. Accuracy of prediction is compared with the existing disease prediction techniques such as simple feed forward ANN, SVM, KNN algorithms. The proposed system is giving high accuracy in prediction of any disease than the existing systems. The proposed genetic based feed forward can be applied to real dataset in healthcare systems.

REFERENCES

1. Prather JC, Lobach D F, Goodwin L K, et.al. MedicalData Mining: Knowledge Discovery in a Clinical DataWarehouse",1997.
2. Hai-Bing Ma, Jin Zhang, Ylng-Jie Fan, Yun-Fa W. Mining Requent Patterns Based On Is+-Tree. IEEE(2004), P1208-1213.
3. Latha Parthiban, " Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm", International Journal of Biological and Life Sciences.(2007); 3:3.
4. Shruti Ratnakar, Rajeswari K, Rose Jacob, Prediction Of Heart Disease Using Genetic Algorithm For Selection Of Optimal Reduced Set Of Attributes, International Journal Of Advanced Computational Engineering And Networking,(2013) , Volume-1, Issue-2, (P): 2320-2106 .
5. Prince Mary, S., Tom, N. " A framework to facilitate choice of cloud carrier providers", Journal of Advanced Research in Dynamical and Control Systems, 2017, 1943-023X.
6. Nilakshi P. Waghulde, Nilima P,Patil. Genetic Neural Approach for Heart Disease Prediction. International Journal of Advanced Computer Research (ISSN (print): 2249-7277 Volume-4 Number-3 Issue-16 September-2014.
7. Harish, P., Vigneshwari, S., Ravi Teja, K.B.S.,Enhancing the security of cloud storage for medical data retrieval using double encryption with data anonymization,Pakistan Journal of Biotechnology,Vol. 14 (Special Issue II) Pp. 75- 78 (2017)
8. Eswar V.O.S,Vinil.B,Ankayarkanni B and Albert, "Integrated Collective Node Behavior Analysis with Onion Protocol for Best and Secured Data Transmission", 2018 International Conference on Communication and Signal Processing (ICCSP), Chennai, 2018, pp. 0918-0921.
9. Shiyamala, R , Sathyabama, R and B. Bharathi. (2015). , "Hybrid method for moving object exploration in dynamic scene: a study", Global Journal of Pure and Applied Mathematics, vol.11,No.8, pp. 4969 – 4964.
10. Kamalesh M. D., Albert Mayan J., Felix Y., Sumanth B. S., & Sai Tej B. (2018), " MAGRISYS: A Smart And Ubiquitous Controlled – Environment Agriculture System " , 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) , pp: 1-5 , 2018
11. Thamizhselvi. S, Prince Mary S. " A Survey about Data Prediction in Wireless Sensor Networks with Improved Energy Efficiency", Research Journal of Pharmaceutical, Biological and Chemical Sciences, March – April 2016 ,7(2) Page No. 2118-2120.
12. Ananthi, S.,Periwal, A.,Prince Mary, S."Data security based on big data storage" 2016, Global Journal of Pure and Applied Mathematics, Volume 12, Number 2 (2016), Pp. 1491-1500.
13. Revathy, R., and K. Priyadharshini. "Review on Big Data Analytics in Distribution of Power Systems." Journal of Computational and Theoretical Nanoscience 16, no. 5-6 (2019): 1829-1831.
14. Joshila Grace, L.K., Maheswari, V. , Mathematical frame work for efficient mining of multi level patterns , Journal of Computational and Theoretical Nanoscience , Volume 13, Number 11, November 2016, pp. 8298-8305(8).
15. Vijai Chandra Prasad.R, Yashwanth Sai M,Niveditha P.R, Dr.Sasipraba .T, Vigneswari S and S.Gowri ,"low cost automated facial recognition system",2017 second ieee international conference on electrical, computer and communication technologies , 22-24, february, 2017, SVS Engineering College,Coimbatore.DOI 10.1109/ICECCT.2017.8117829,
16. Vigneshwari.S, Mary Posonia. A,Gowri,S, "An Efficient framework for Document Retrieval in Relationship Strengthened Personalized Ontologies",Soft Computing inData Analytics, part of AISC volume 78,pp, 735-742, 2018
17. Bhuvaneswari R, Kalaiselvi K, Naïve Bayesian Classification Approach in Healthcare Applications, International Journal of computer Science and Telecommunication,(2012); vol. 3, no. 1, pp. 106-112.
18. Akhil Jabbar M , Priti Chandra, Deekshatulu B L. Heart Disease Prediction System using Associative Classification and Genetic Algorithm. International Conference on Emerging Trends in Electrical, Electronics and Communication Technologies-ICECIT,(2012).
19. Paulsen JS, Hayden M, Stout JC, Langbehn DR, Aylward E, Ross CA, Guttman M, Nance M, Kieburtz K, Oakes D, Shoulson I, Kayson E, Johnson S, Penziner E. Preparing for preventive clinical trials: the Predict-HD study. Archives of neurology 2006; 63: 883-890.
20. Santha Sheela A.C and Jayakumar C."Duplicate Web Pages Detection With The Support Of 2D Table Approach", Journal of Theoretical and Applied Information Technology,Vol. 67, No.1, SEP 2014.