



An Analysis and Survey on Artificial Neural Network Based Genetic Pattern Recognition for Predicting Hepatitis B and Hepatitis C Virus

M.Venkatesam¹, P.Penchalaiah²

¹Research Scholar, Dept. of Computer Science, Rayalaseema University,
Kurnool, Andhra Pradesh
INDIA

²Research Scholar, Dept of Computer Science, Vikrama Simhapuri University,
Nellore, Andhra Pradesh
INDIA

Abstract: This paper presents an analysis and survey of some of the most valuable work related to the techniques and applications of artificial neural networks for Genetic pattern recognition. Genetic pattern recognition plays a major role in predicting the disease. Automatic analysis of genetic patterns is a challenging problem with many applications. In this paper a neural network and genetic algorithm based pattern recognition system is proposed for predicting Hepatitis B and Hepatitis C virus. The Artificial neural network is used to model the data from the genetic data set and also used to generate casual patterns for the data set. The neural network is trained to learn the casual patterns. The genetic algorithm selects the patterns and finally the selected patterns are analyzed based on genetic knowledge for prediction.

Keywords: Genetic algorithms, Artificial Neural networks, Genetic dataset, Casual patterns, Hepatitis B and C virus.

I. Introduction

During the past several years, Artificial Intelligence has been used extensively in the field of medicine. Artificial intelligence application in the medical field comprises of interpretation and diagnosis of the disease. The diagnosis of a disease which is treated as a classification problem is used to exhibit the presence or absence of the disease. The diagnosis is based on several symptoms and results of medical exams. Many new classification methods are proposed by researchers which are useful in diagnosing the disease. Machine learning techniques such as symbolic and connectionist have been extensively used in medical diagnosis. Artificial neural networks and statistical methods are used for diagnosis the hepatitis disease.

In this paper a disease diagnosis system using different expert methods is proposed. The proposed system includes the discovery of disease causing genes. Identification of genetic patterns which are used for the prediction of the disease, understanding the multiple gene interactions, their relationships and the role each gene plays [1]. The assessment on the hepatitis medical datasets taken from UCI machine learning database [2][3] shown in the Table1.

A. Hepatitis disease

Hepatitis disease occurs due to the inflammation of the liver. A person affected by hepatitis disease may have several disorders like the liver damage. The most commonly seen hepatitis viruses are hepatitis A, hepatitis B and hepatitis C. The other known hepatitis viruses are D, E, F and G. Hepatitis A and E will cause severe infection. Hepatitis B and D will cause ongoing illness. Hepatitis D will be only seen in the people who are already infected by hepatitis B. The main cause for hepatitis virus is due to the glandular fever virus [6].

Table1: The features of Hepatitis disease database.

Feature	Value
Age	10,20,30, 40,50,60,70,80
Sex	Male, Female
Steroid	Yes, No
Antivirals	Yes, No
Fatigue	Yes, No
Malaise	Yes, No
Anorexia	Yes, No
Liver big	Yes, No
Liver firm	Yes, No
Spleen palpabl	Yes, No
Spiders	Yes, No
Ascites	Yes, No

Varices	Yes, No
Bilirubin	0.39,0.80,1.20,2.00,3.00, 4.00
Alk phosphate	33,80,120,160,200,250
Sgot	13,100,200,300,400,500
Albumin	2.1,3.0,3.8,4.5,5.0,6.0
Protime	10,20,30,40,50,60,70,80,90
Histology	Yes, No

The symptoms of hepatitis A, B and C are nausea, diarrhoea, appetite loss, Jaundice, itchy skin, weight loss, vomiting [4].

B. Diagnosis of Hepatitis B

The hepatitis B virus (HBV) will infect the liver [5]. Hepatitis B virus has three commonly used antigens the first one is the surface antigen (HBsAg), the core antigen (HBcAg) and the e antigen (HBeAg).

(i). Antigen HBsAg

The presence of hepatitis B surface antigen (HBsAg) in the blood indicates that the patient is currently affected by the virus. The complete immunity to the Hepatitis B viral infection are provided by the Anti-HBs.

(ii). Antigen HBcAg

Anti-HBc IgM can be used to diagnose an hepatitis B infection.

(iii). Antigen HBeAg

Hepatitis B e antigen is present when the hepatitis B virus is multiplying. The antibody anti-HBe signifies an inactive state of the virus and a lower risk of transmission.

HBV consists of the core particle which is a central portion made up of DNA and the core antigen (HBcAg) and a surrounding envelope which contains the surface antigen (HBsAg). The antigens are present in the blood and the markers are used in diagnosis and evaluation of patients with suspected viral hepatitis. The markers for Hepatitis diagnosis are given in the Table2 below [5].

Table2: The markers for Hepatitis diagnosis.

Marker	Value
AgHBs	Positive
AgHBs	Negative
AgHBe	Positive
AgHBe	Negative
anti – VHD	Positive
anti – VHD	Negative
anti – VHC	Positive

The basic rule is: IF (AgHBs=Positive) AND (AgHBe=Positive) AND (anti-VHD=Negative) THEN Hepatitis B. The interpretation of sets of results from hepatitis B blood serological tests is given in the Table3 below [5].

Table3: The markers for Hepatitis diagnosis.

Status	Tests	Results
Susceptible, not infected, not immune	HBsAg anti-HBc anti-HBs	Negative Negative Negative
Immune due to natural infection	HBsAg anti-HBc anti-HBs	Negative Positive Positive
Immune due to hepatitis B vaccination	HBsAg anti-HBc anti-HBs	Negative Negative Positive
Acutely infected	HBsAg anti-HBc IgM anti-HBc anti-HBs	Positive Positive Positive Negative
Chronically infected	HBsAg anti-HBc IgM anti-HBc anti-HBs	Positive Positive Negative Negative

C. Diagnosis of Hepatitis C

Hepatitis C is a liver disease that causes inflammation and infection in the liver. Hepatitis C leads to chronic infections like cirrhosis (scarring), liver failure, liver cancer. Hepatitis C can be infected by blood. The cells are damaged when HCV virus enters the body, penetrates the liver cells and begins to damage the cells. Hepatitis C exists in two forms: acute and chronic. Acute hepatitis C causes symptoms to set in quickly, where as the chronic hepatitis C symptoms will not be apparent at once since hepatitis C develops over a period of months.

The HCV patients features include: Blood test features, Patient characteristics features. The blood test features includes 9 features [14] which are Albumin, Alfa-Feto protein, viral load, cirrhosis, Alanine Amino Transferase (ALT), Aspartate Amino Transferase (AST), fibrosis stage and Histology Activity index (HAI). The patient characteristics features include 3 features which are gender, age and Body Mass Index (BMI). The Table4 below describes the HCV patient's features [14].

Table4: List of HCV patient's features.

Feature name	Values
Fibrosis stage	0-6
Cirrhosis	0=No(189) ; 1= Yes(31)
Age	Years
Genotype	Genotype 4 = 173; Non-Genotype 1 = 47
Histology Activity Index(HAI)	1 - 15
Gender	1= M(176); 2 = F(44)
Viral load	0.006 – 5.05 copies/ml
Body Mass Index (BMI)	16.6 – 43.2 Kg/cm ²
Alanine Amino Transferase(ALT)	0.8 – 7.05
Albumin	2.5 – 5.4 g/L
Alfa-Feto Protein	0.02 – 3.25
Aspartate Amino Transferase(AST)	0.01 – 0.3
Response	0= non-responder (128) , 1 = responder (92)

The fundamental goals of Hepatitis B&C virus prediction is concerned with two predictive foci:

- 1) The prediction of disease causing genes
- 2) Identification of genetic patterns which are used for the prediction of HBV and HCV disease.

The prediction of HBV and HCV disease involves the generation of patterns for a given genetic data set. The patterns are selected based on their strength and the selected patterns are analyzed based on the genetic knowledge.

D. Artificial Neural Networks

An Artificial neural network set of simple processing units called neurons which communicate by sending signals to each other over a large number of weighted connections [7]. The basic unit of an artificial neural network which is a neuron receives inputs, processes them by simple combination and threshold operations and outputs a final result. The Artificial neural network is used to model the data from the genetic data set and also used to generate casual patterns for the data set. The neural network is trained to learn the casual patterns. Artificial neural networks are used for pattern recognition and prediction.

E. Genetic Algorithms

A genetic algorithm is a searching technique used in computing to find exact or optimal solutions to optimization and search problems [8]. Genetic algorithms perform the searching process by combining the selection, crossover and mutation operators with a goal of finding the best solution to a problem. Genetic algorithms search for the optimal solution until a specified termination criterion is met. The genetic algorithm selects the patterns and finally the selected patterns are analyzed based on genetic knowledge.

F. Genetic Pattern Recognition

Genetic pattern recognition is the process of identifying the gene combinations which relates to a given trait and is determined by multiple genes called polygenic trait [9]. The exact gene combinations that will lead to the trait are to be identified. The genetic pattern discovery program (GPDP) is used to identify the genes and the patterns of their interactions. The conventional methods to the genetic pattern recognition are used only to recognize individual genes without analyzing their interaction patterns. Due to the heterogeneity of polygenic traits, the analysis of genetic data sets is extremely complex. The results of complex gene interactions which involves the dominant or recessive relationships between alleles of a single gene as well as the non-linear interactions between alleles at separate loci. In such complicated situations, more than one genotype should produce the phenotype involved in the expression of a complex trait.

II. Related Work

There are several theories to explain about polygenic traits. One theory explains about the existence of a threshold called the threshold trait. The gene contributes to some extent. If the combined effect of the multiple genes exceeds the threshold, then only the trait will appear. The expression for the trait is all or none. In the other theory there is no threshold for the trait. The expression of the trait is continuous, which is based on the combined effect of the genes.

In both the theories, a quantitative model is used to simulate the genetic mechanism. A linear system which states that the gene effects are linearly additive is simple but will be of little use. A non-linear solution is mandatory.

Another alternative solution is to use the probabilistic approach. In the case of the threshold traits the probabilistic approach will be misleading since it fails to recognize that the probability of a trait is zero when the trait is under the threshold.

Researchers have applied data mining using different machine learning techniques for analyzing and identifying hidden patterns inside HCV datasets. Artificial neural networks are used for HCV & HBV proteases cleavage and achieved high prediction accuracy [10][11].

One of the models proposed by Soheila Molaei [12] uses artificial neural networks to predict HCV phylogeny from the visual restriction map data. The goal of this model is to predict various HCV types and subtypes from a complex set of restriction patterns by the use of expert system. The gene classification is done using three methods based on artificial neural networks, principal component analysis and support vector machine (SVM) were proposed. The model is proved to be accurate predictor of HCV genotype.

An intelligent system for the diagnosis of Hepatitis B disease was developed by Dakshata Panchal [5]. The intelligent system uses a generalized regression network which gives the result whether the patient is tested positive for Hepatitis B or not and the severity of the patient. The Generalized regression network is a type of neural network which uses kernel-based approximation to perform regression.

A Rough-granular neural networks model which uses rough sets and artificial neural networks to make efficient data analysis and suggestive predictions was proposed by Mohammed M.Eissa [13]. The rough sets are used to reduce and choose the most relevant attributes which reduces the number of input vectors to artificial neural networks which increases the classification accuracy of the proposed model. The resulting optimal dataset is given as input to the constructed neural network with supervised learning algorithm for classification. The overall performance of the proposed model is superlative.

An associative classification approach has been used to predict the Hepatitis C virus from patient's blood test [14]. For prediction a knowledge discovery framework was developed. The knowledge frame work involves two phases: Pre-processing phase and the data mining phase. The associative classification approach generates a set of class association rules (CARs) in the data mining phase. A classifier is build from the selected suitable CARs which are used to predict the response to treatment. More data is needed to learn and to extract more patterns.

A comparison between Backpropagation and Naïve Bayes classifier was proposed by Bekir Karlik [15] to predict and diagnose hepatitis disease. Both these classifiers require small amount of training data to estimate the parameters which are needed for classification. Naïve Bayes classifier is highly efficient since its classification decision may be correct even if its probability estimates are inaccurate. These two are to be compared with the other classifiers methods such as Fuzzy-classifier, Support-Vector Machines (SVM), Radial-basis functions (RBF), Conic section function (CSF) neural networks, Learning vector quantization (LVQ), k-nearest and others.

An automatic diagnosis system based on artificial neural networks (ANN) which deals with the feature extraction and classification for predicting hepatitis C virus was introduced by Tahseen A. Jilani [16]. The classification accuracy obtained using this automatic diagnosis system for training data is 99.1%. The viral markers such as hepatitis enzymes, blood picture, urine analysis such as bilirubinuria, mild proteinuria, urobilinogen, serology for viral markers are not considered for predicting hepatitis disease. Moreover Fuzzy learning models are to be applied on a large number of patitents.

In the published literature a few efforts for genetic pattern discovery are available. One is the work that describes the pattern discovery-based method for detecting multi-locus genetic association by Zhong Li [17]. In this literature, the method for dissecting complex genetic traits in humans for potential genetic heterogeneity is not addressed properly. The method will be useful only for dissecting complex genetic traits in humans, particularly when oligogenic/polygenic inheritance and gene-gene interactions are present.

Ghumbre Shashikant Uttreshwar [18] implemented an intelligent system for classifying Hepatitis B virus using logical inference and Self-Organizing maps. The intelligent system uses logical inference to decide on the type of hepatitis the patient is infected, whether it is hepatitis B virus or not. Then the intelligent system uses kohen's self-organizing maps applied to hepatitis data for predicting the hepatitis B and the severity level of the patient. The proposed intelligent system will be a promising tool for faster and more accurate prediction of hepatitis B.

III. Conclusions and Future Work

The genetic patterns which are identified by the different classification methods will have several advantages like prevention, diagnosis and treatment of the disease. These classification methods will be useful in qualitative understanding of the genes, study of the gene patterns, understanding the evolutionary relationships among the genes and the understanding of the evolutionary relationships among the different species.

In this paper we have investigated several classification methods for identification of genetic patterns which are used for the prediction of HBV and HCV disease. The accuracy obtained by applying these classification methods for the prediction of the disease is nearly 98%. We will apply a new and a novel approach for genetic pattern dissection “Genetic pattern discovery program (GPDP)” which will be useful in the discovery of disease causing genes and understanding the multiple gene interactions, their relationships and the role each gene plays.

References

- [1] Mahdich Adeli, Hassan Zarabadipour, “Automatic disease diagnosis systems using pattern recognition based genetic algorithm and neural networks”, Recent Advances in Applied & Biomedical Informatics and Computational Engineering in Systems Applications.
- [2] P.N. Suganthan, “Structural pattern recognition using genetic algorithms, Pattern Recognition”, Vol.35, No.X, 2002,pp. 1883--1893.
- [3] Blake,C.L.,&Merz,C.J.(1996).UJI repository of machine learning databases. <http://www.ics.uci.edu/~mlearn/MLRepository.html>
- [4] http://kidshealth.org/parent/infections/bacterial_viral/hepatitis.html
- [5] Dakshata Panchal and Seema Shah, “Artificial Intelligence Based Expert System for Hepatitis B Diagnosis”, International Journal of Modeling and Optimization, Vol. 1, No. 4, October 2011.
- [6] <http://www.natcol.co.uk/symptomproducts.php?condition=414>
- [7] www.learnartificialneuralnetworks.com
- [8] <http://www.genetic-programming.com>
- [9] Limin Fu, Neural networks in computer intelligence, Tata McGraw-Hill, Genetic Pattern Recognition, pp. 397-410.
- [10] Wang, D., Larder BA, Revell AD, Montaner J, Harrigan R., "A comparison of three computational modelling methods for the prediction of virological response to combination HIV therapy," Artificial Intelligence in Medicine, 47, 63-74, 2009.
- [11] Lau-Corona, D., M. Gutiérrez-Ruiz, G. Robles-Díaz, and D. Kershenobich, "Effective use of fibro test to generate decision trees in hepatitis C," Journal of Gastroenterology, 15, pp. 2617-2622, 2009.
- [12] Soheila Molaei, Somaye Naser Eslami, “Development and performance appraisal of an expert system for predicting HCV genotype using a virtual restriction map and artificial neural network”, World of Sciences Journal; 2013[04].
- [13] Mohammed M.Eissa Mohammed Elmogy Mohammed Hashem, “Rough - Granular Neural Network Model for Making Treatment Decisions of Hepatitis C”, The 9th International Conference on INFormatics and Systems (INFOS2014) – 15-17 December Data Engineering and Knowledge Management Track.
- [14] Enas M.F. El Houby, “Analysis of Associative Classification for Prediction of HCV Response to Treatment”, International Journal of Computer Applications (0975 – 8887) Volume 63– No.15, February 2013.
- [15] Bekir KARLIK, “Hepatitis Disease Diagnosis Using Backpropagation and the Naive Bayes Classifiers”, Journal of Science and Technology, Volume : 1 / Number : 1 / Year :2011.
- [16] Tahseen A. Jilani, Huda Yasin and Madiha Mohammad Yasin, ”PCA-ANN for Classification of Hepatitis-C Patients” International Journal of Computer Applications (0975 – 8887) Volume 14– No.7, February 2011.
- [17] Zhong Li, Aris Floratos, David Wang, and Andrea Califano, “A Pattern Discovery-Based Method for Detecting Multi-Locus Genetic Association” Department of Computational Genetics, High Throughput Biology Inc., Livingston, NJ.
- [18] Ghumbre Shashikant Uttreshwar and A.A. Ghatol, “Hepatitis B Diagnosis Using Logical Inference and Self-Organizing Map” Journal of Computer Science 4 (12): 1042-1050, 2008 ISSN 1549-3636© 2008 Science Publications.

Acknowledgments

The authors are deeply indebted to Dr. U. Mallikarjunaiah for his support and valuable suggestions on this manuscript. We thank Madhu Babu, Research Scholar for his observations and for careful reviews of this manuscript.