IJARCCE

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

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 11, November 2017

Prediction of Risk Factor for HCV Infection among IDUs - An Approach using C4.5

M.Gomathy[1], Vani Perumal[2]

Assistant Professor, Deptof Computer Application, NBGSM College, NCR Delhi, Haryana, India[1]

Assistant Professor, IT Dept, RCAS, Ministry of Higher Education, Al-Rustaq, Sultanate of Oman[2]

Abstract: Hepatices C viral (HCV) infection is one of the common and hazardous viral infection among injecting drug users (IDU) globally. There are several identified risk factors associated with HCV. In Data Mining, various algorithms are used to mine the necessary information from the large set of data. These are also playing a vital role in numerous disease predictions C4.5 is one of the most popular algorithms for rule based classification and pruning of decision trees. Decision Tree is a supervised classification technique, which is simple, fast and accurate for prediction and decision making. It can be applied to any domain. In this paper, C4.5 algorithm is used to rank the risk factors associated with HCV among IDU's in India. It also identifies the most relevant attributes from a dataset, so that the input space is reduced and simultaneously the performance is improved. In addition to that, this also gives decision tree for effective decision making. From the experimental results, the C4.5 algorithm and the decision tree shows the most important factors associated with HCV infection among IDU's in India.

Keywords: Data Mining, C4.5 Algorithm, IDU, HCV, Entropy, Decision tree

I. INTRODUCTION

From the year of identification, 1989, [1, 2] of Hepatitis C virus and the availability of analyzing the antibody to HCV, the epidemiology of HCV infection has been investigated in many populations. The major group infected and at risk of continuing infection with HCV in India, as in other countries, are people who currently or previously injected or had been injected with illicit drugs. This increased risk is associated with the practice of sharing injecting equipment; especially needles and syringes, similar to Hepatitis B Virus (HBV) and Human Immunodeficiency virus (HIV) [3].

Data mining holds great potential for the healthcare industry to enable health systems to systematically use data and analytics to identify inefficiencies and best practices that improve care and reduce costs. Data mining helps us to extract the hidden predictive information from large databases. The large amounts of data generated by healthcare transactions are too complex and voluminous to be processed and analyzed by conventional methods. Data mining provides the methodology and technology to transform these mounds of data into useful information for decision making. In the present study, decision trees were constructed using the C4.5 classification algorithm to identify the risk factor associated with HCV among IDU's. The C4.5 Algorithm, often referred to as statistical classifier, is based on the concepts of information entropy and information gain. Intuitively, information entropy is the number of bits required to code an event where the higher the probability of the event the lower the number of bits required to code it. Information gain, in turn, is the reduction of entropy when additional information is available. C4.5 uses the fact that each attribute of data can be used to make a decision that splits the data into smaller subsets, which have reduced information entropy. Decision trees are a diagrammatic representation of a decision process, where nodes represent questions about attribute values or ranges of values, and edges represent the possible answers that link question nodes with other nodes down the tree, which represent further questions. In the case of Injecting Drug Users each question node in the tree represents an HCV risk factor and the bottom nodes represent the risk behavior

The data used in this paper is from Integrated Biological and Behavioural Assessment (IBBA). This survey captured Hepatitis C virus associated risk behaviors among injecting drug users (IDUs) in Dimapur of Nagaland in India. There are 440 records collected from this district of Nagaland among IDU's. Each record has the same structure, consisting of a number of attributes or value pairs. One of these attributes represents the category of the record which says the prevalence of high risk of Hepatitis C virus or low risk of Hepatitis C virus from the collected information. Finally a decision tree is constructed from non-categorical attributes. The category attributes are assumed to take the values {high risk, low risk}.The non-category data which has been chosen among various attributes have association on the risk factor of Hepatitis C virus. The domain values which specify the risk behavior among the IDU's are defined for the present investigation are as follows

- MStatus - This attribute specifies the Marital Status among the IDU's which hold values as either married or unmarried
- Type of drug - It is classified into two categories based on usage of heroin drug and other drugs
- LOC - It specify the location of the IDU which can take two values whether it is used inside home or other places

IJARCCE

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

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 11, November 2017

- Shared Needles/Syringe - This attribute can have two possible values likes sharing, not sharing the needles/Syringes before usage
- No.of Shared - This attributes take two values to show the sharing of drugs is between more than two IDU persons.
- Freq.Shar N/S -Frequency of Shared Syringe/Needle. This attributes specifies the frequency of sharing the Syringe/needle.
- Inj Pre Syn- Injecting from prefilled syringe. It gives information about the IDU's injecting equipment's, new/utilized.
- Shar CC -Shared Common Container -This attribute indicates the containers used by the IDU's for sharing of drugs.
- Period-It indicates the period of abstaining from injecting drugs by IDU's.

II. REVIEW OF RELATED RESEARCH

Many researchers have been developed factors associated with HCV viral infection patients among injecting drug users globally including India from clinical information using different data mining techniques. Apart from that, numerous prediction algorithms are designed with the help of implementing Data mining algorithms. Some of them are discussed here.

AbdElrazek et.al. utilized data mining intelligent computational analysis to quantify the findings among HCV transmission in Egypt among the Neonates born to hepatitis C virus positive mothers. They aimed to investigate the factors contributing for vertical HCV transmission in Egypt and identified that optimizing the HCV viral load in HCV-positive women might prevent vertical HCV transmission to neonates.[4]

Shungo Imai BS et.al. used decision tree analysis to find the combinations of multiple risk factors that would increase the risk of nephrotoxicity associated with vancomycin. They compared the predictive performance with that of multiple logistic regression analysis and showed the usefulness of Decision Tree analysis and identified that predictive accuracy of the Decision Tree and logistic regression models were similar and accurate. They also suggested the decision tree models for the evaluation of adverse drug reactions among the patients (intravenous administrations of VCM). [5]

Dr. P. Indra Muthu Meena and Dr. Vani Perumal analyzed the performance of C4.5 algorithm and Navie Bayes algorithm to predict the cancer accuracy. They utilized C4.5 algorithm to rank the most important attribute which causes stomach and Naïve Bayes algorithm is used to predict stomach cancer for any given report, based on the set of classifiers. [6]

T.Miranda Lakshmi et.al. used decision tree to the analyse the qualitative data in student's performance in the examinations and other activities. They used ID3, C4.5 and CART algorithms to compare the results which also showed the performance of the algorithms in data mining. [7]

DebasishBasu et.al. showed a high seroprevalence of anti-HCV antibody in IDUs. In the substance users, HCV positivity was significantly and independently associated with several clinical, behavioral and personality risk factors. Logistic regression analysis was used to determine the risk factors of HCV infection. They have proved the four risk factors strongly associated with HCV positivity in multivariate analysis were sharing syringe, reuse of injection accessories, blood transfusion and IDU status. [8]

Lopamudra Ray Saraswati et.al. analyzed PWID (persons who inject drugs) are at high risk for HIV, Hepatitis B and Hepatitis C infections. They also reported the prevalence of HIV, HBV and HCV infections and correlates of HIV-HCV co-infection among male PWIDs in Delhi, India. They identified the Higher frequency of injection days/month was associated with a higher likelihood of HCV mono-infection and HIV-HCV co-infections. Multinomial logistic regression was employed to identify predictors of HIV, HCV and HIV-HCV co-infection. [9]

Santosh Kumar Sharma and Shri Kant Singh examined co-infection with Human Immunodeficiency Virus and Hepatitis C Virus among injecting drug users in North-eastern states of India. Univariate with Chi-square test and Binary logistic regression were used for analyzing the factors associated with co-infection with HIV and HCV. They performed a comparative study on the prevalence of HIV and HCV among these populations and indicated that most of the IDUs who have HIV were already infected with Hepatitis C Virus, which leads to many HCV-related liver disease, and increased risk for cirrhosis and liver cancer. [10]

P.K. Chelleng et.al. discussed about hepatitis C infection among IDUs and their unsafe injecting practices, practice of tattooing in remote tribal areas forms the basis of prevalence of parentally transmitted viral diseases. They also measure the risk behaviors and seroprevalence of hepatitis C virus antibodies amongst IDUs of Mizoram, a State of the northeast India. They proved unsafe injecting practices were found to be associated with a higher risk of acquiring hepatitis C infection and also showed that syringe and needle exchange programme alone was not sufficient as a preventive strategy for control of hepatitis C infection among IDUs.[11]

Sunil S Solomon el.at. used Cross-sectional analysis to estimate the prevalence of HIV , HCV and HBV co-infection as well as current risk behaviors among HIV positive and negative injection drug users in Chennai, India. They also used multivariate analysis to show the attribute injecting at a dealer's place and duration of injection drug use were positively associated with prevalent HIV infection. They also showed the attribute Alcohol consumption was negatively associated with HIV. HIV positive IDUs were as or more likely compared to HIV negative IDUs to report recent high-risk injection-related behaviors. [12]

Mohammed A.Farahatel.at. designed a framework to predict the response of Chronic HCV genotype four patients to Direct-Acting Anti virals (DAAs) by applying Data Mining Techniques (DMT) on clinical information, and then extract the result of DMT to be a Knowledge Base for the application to perform the prediction process. Data from 420 patients infected with hepatitis C virus genotype 4 from different centers in Europe and Egypt are analyzed. Patients were treated with four different regimens of DAAs with or without pegylated interferon. Initially they applied Data preprocessing phase to prepare the data before applying the DMT, Then in DM phase to apply DMT. Then in the evaluation phase they evaluated the performance and accuracy of the built model using a data mining evaluation technique. Feature selection algorithm has been applied on each group. Decision Tree has been applied for the prediction, after that extraction of the result of DTs was performed. This helped to construct a Knowledge Base application to perform the prediction operation. The method gave 90.9% acceptable and best accuracy results. [13]

V.Kirubha, S.ManjuPriya analyzed the application and importance of data mining in healthcare domain They analyzed some of the techniques used in datamining for various disease prediction. Decision Tree, NaïveBayes, Neural Networks, Fuzzy Logic, SVM, Multilayer Perceptron , Simple Logistic were used to predict Heart Disease, Kidney Disease, Liver Disease, Diabetes, Cancer. They compared various algorithms used for disease prediction from previous study made by other researchers and showed that Datamining provides good results in disease diagnosis when appropriate tools and techniques are applied. The author also showed mining is the promising field for healthcare predictions. [14]

Mohammed M.Eissa et al. used Rough Granular Neural Network model and Artificial Neural Network (ANN) for Making Treatment Decisions of Hepatitis C. The rough set technique had been used to discover the dependency among the attributes and to reduce the attributes, and their values before the original information, remove redundant information, reduce the dimension of the information space, provide a simpler neural network training sets, and then construct and train the neural network. The experimental results show that the proposed hybrid model can acquire the advantages from the two data-mining methods Rune Space (RS) and Artificial Neural Network (ANN). In addition, the integration of the Rough Sets and ANN together can produce a positive effect, enhancing model performance. [15]

E. M. F. El Houby, et.al. applied knowledge discovery technique to predict HCV patients' response to treatment, which is a combined therapy Peg-IFN and RBV, according to a set of features. The proposed framework consists of two phases which are pre-processing and data mining. Associative Classification (AC) has been used to predict response to treatment in patients. AC technique has been used to generate a set of Class Association Rules (CARs). The most suitable CARs are selected to build a classifier which predicts patient's response to treatment from the selected features. The accuracy of the algorithm is high reach up to 90%. [16]

M. ElHefnawi, et.al. made a prediction of response to Interferon-based therapy in Egyptian patients with Chronic Hepatitis C using machine-learning approaches. They used Artificial Neural Network (ANN) and Decision Tree(DT) techniques in data mining and showed the maximum accuracy for ANN and DT. [17]

M. M. Eissa, et.al. introduced a Hybrid Rough Genetic Model to classify the effects of a new medication for HCV treatment through Hybrid Rough Genetic Model which has been used to predict response to new medications for HCV treatment in patients with hepatitis C virus. The hybrid model used 4 phases like data preprocessing, data reduction, rule generation and classifications of HCV data .The author showed the advantages from the two data-mining (Rough Sets and Genetic Algorithms) methods used in this model and therefore, produce superlative results. The Integrating

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 11, November 2017

Rough Sets and Genetic Algorithms together can produce a positive effect as well as it enhanced the model performance. [18]
Enas M. F. El Houby compared the performance of different data mining techniques' in predicting patients' response to treatment of HCV from clinical information. Three data mining techniques Artificial Neural Network, Analogues Complexing and Decision Tree were used to show the accuracy of the outcome. In evaluation phase, all the models built for various candidate features subsets have been evaluated using test dataset .The author concluded by showing the highest performance and accuracy of each models as for the Analogues Complexing is 92% , Artificial Neural Network is 78% and 80% for Decision Tree. [19]

Lin E, et.al. have used two classification algorithms, including Multilayer Feed Forward Neural Network (MFNN) and logistic regression are used for comparisons. An MFNN is one type of Artificial Neural Network models where connections between the units do not form a directed cycle. These classifiers were performed using the datamining algorithm Waikato Environment for Knowledge Analysis (WEKA) software. They used numerical forms1 for "SVR (Sustained virological response)" and 0 for "NR (Non-viral response)", respectively. To measure the performance of prediction the author has used the Receiver Operating Characteristic (ROC) methodology and calculated the area under the ROC curve. [20]

Masayuki Kurosaki, et. al. used classification and regression tree and Statistical analysis to build a predictive model of response to the treatment in HCV. The software automatically explore the data to search for optimal split variables, builds a decision tree structure and finally classifies all subjects into particular subgroups that are homogeneous with respect to the outcome of interest. [21]

Kazuaki Chayama et al. made a statistical analysis using R software package. CART analysis was used to generate a decision tree by classifying patients by SVR, based on a recursive partitioning algorithm with minimal cost-complexity pruning to identify optimal classification factors. The association between SVR and individual clinical factors was assessed using logistic regression. Data was collected from 840 genotype 1b chronic hepatitis C patients. [22]

NaglaaZayed et al. made a study on retrospective data belonging to 3719 adult patients with chronic HCV infection of both sexes who were diagnosed by anti-HCV antibodies. Weka implementation of C4.5 (WEKA J48) decision-tree learning algorithm was applied using 19 clinical, bio-chemical, virologic and histologic pre-treatment attributes form the data of 3719 Egyptian patients with chronic HCV. The universality of the decision-tree model was validated using both internal and external validation to confirm the reproducibility of the results. They applied Statistical and Multivariable logistic regression analysis on the data. [23]

III. APPLICATION OF C4.5 IN INJECTING DRUG USERS' DATA SET

This session specifies the algorithm, approach and investigational results by applying C4.5 in the IDU's dataset.

A. C4.5 Algorithm

C4.5 algorithm was introduced by Quinlan for inducing Classification Models also referred as Decision Trees, from the observed data. In the observed data set, each record contains the same structure of data. The data can have any number of attributes or value pairs. One of these attributes represents the category of the record. The problem is to build a decision tree on the basis of the observation about the non-category attributes predicts correctly the value of the category attribute. The category attribute can take values like {true, false}, or {Predicted, not predicted}, or {success, failure}, or something equivalent. In any case, one of its values will mean failure. If there are n equally probable possible messages, then the probability p of each is 1/n and the information conveyed by a message is -log(p) = log(n). In general, if we are given a probability distribution P = (p1, p2,.., pn) then the Information conveyed by this distribution, also called the Entropy of P, is: I(P) = -(p1*log(p1) + p2*log(p2) + .. + pn*log(pn)). If a set T of records are partitioned into disjoint exhaustive classes C1, C2, .., Ck on the basis of the value of the categorical attribute, then the information needed to identify the class of an element of T is Info(T) = I(P), where P is the probability distribution of the partition (C1, C2, .., Ck): P = (|C1|/|T|, |C2|/|T|,.., |Ck|/|T|)

First T is partitioned on the basis of the value of a non-categorical attribute X into sets T1,T2,..,Tn then the information needed to identify the class of an element of T becomes the weighted average of the information needed to identify the class of an element of Ti, i.e. the weighted average of Info(Ti): Info(X,T) = Sum for i from 1 to n of $|Ti|/|T|$ * Info(Ti) and the quantity Gain(X,T) defined as Gain(X,T) = Info(T) - Info(X,T) This represents the difference between the information needed to identify an element of T and the information needed to identify an element of T after the value of attribute X has been obtained, that is, this is the gain in information due to attribute X. Thus we can predict which information offers a greater informational gain than all the other information. This notion of gain is used to rank

IJARCCE

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

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 11, November 2017

attributes and to construct decision trees where at each node is located the attribute with greatest gain among the attributes not yet considered in the path from the root.

The intent of this ordering is twofold:
(i) To create small decision trees so that records can be identified after only a few questions.
(ii) To match a hoped for minimality of the process represented by the records being considered.
Hence the C4.5 algorithm can be used to identify the most relevant attribute which is the key risk factors associated with HCV infection among IDU's in India.

B. The Proposed Methodology

As mentioned above the observed dataset related to Injecting Drug User's contains Four hundred and forty records of different values for nine different non-categorical attributes. The attributes and the possible values are given in Table 1.

TABLE I: Description about the non categoryattributes and its possible values

Attribute	Description	Possible Value
MStatus	Marital Status	0-Married , 1- Unmarried
TDrug	Type of drug	1-Heroin , 2 –Other Drug
Loc	Location	1- Own house, >1 Other place
Shar N/S	Shared Needles /Syringe	0-no, 1-yes,
No. Shared	Number of person shared	>2, <2
Freq.Shar N/S	Frequency of Shared Syringe/Needle	1-Every time , >1 not always
Inj Pre Syn	Injecting from prefilled syringe	1-Every time , >1 not always
Shar CC	Shared Common Container	1-Every time , >1 not always
Period	Period of stopped common container	1-less than 6months 2-greater than 6months

Using the data table that contains attributes and class of the attributes, the homogeneity (or heterogeneity) is measured based on the classes. If a table is pure or homogenous, it contains only a single class. If a data table contains several classes, then it says that the table is impure or heterogeneous. Entropy value is calculated to identify the degree of impurity.The simplified pseudo-code for the algorithm is as follows:

(1) Find the most informative attribute (i.e. the one with the lowest entropy or the largest information gain) in the given set of samples.
(2) Create a decision node (Level 0) that splits on the selected attribute; this node will have a decision question on an attribute's value and will partition the samples in relation to such a value (high risk/low risk) .If all the samples belong to the same partition, the corresponding node becomes a class node.
(3) Create a Child node for each remaining case, and proceed from step 1 for all elements that remain in the corresponding partition.
From the specified attributes, identify the class of element of A (Non Category Data) relative to S(category data) which is used to calculate the entropy of S i.e. Info(A)=I(S) Therefore from the dataset S of 440 instances , 50 are "High Risk ", 390 are "Low Risk". From the Information conveyed by this distribution (high risk, low risk), the Entropy of S is specified as Entropy (Risk) =I (HIGH RISK,LOWRISK).

The entropy of the set S:
Entropy (Risk) = I (HIGH RISK, LOWRISK)
= 0.51079

Gain(S, A)= Entropy(S) $- \sum \frac{|Sv|}{|S|} Entropy(Sv)$

To determine the best attribute for a particular node in the tree, information gain is applied. The information gain, Gain (S, A) of an attribute A, relative to the collection of attributes S is calculated as first partition S(risk)on the basis of the value of a non-categorical attribute A into sets S1, S2, .., Sn . To find Info(S, A) and Gain(S,A) for all the attributes, the above calculated Entropy of S is used.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 11, November 2017

Information gain is calculated to all the attributes. Table II describes the information gain of all the attributes of qualitative parameters

TABLE II: The non-category attributes and their gain

Attributes	Info(S,A)	Gain((S,A)
MStatus	0.44465	0.066137601
TDrug	0.20794	0.302841003
Shar N/S	0.359045	0.151744242
No.of Shared	0.430134	0.08065507
Freq.Shar N/S	0	0.51079
Inj Pre Syn	0.069614	0.441175371
Shar CC	0.156527	0.354262301
Loc	0.29625	0.214531857

The Gain value is used to rank attributes and to build decision trees. Each node portrays the attribute with greatest gain among the attributes to build a decision tree so that records can be identified easily and to match a minimality of the process from the considered records. By ranking the Gain value, we have identified that the attribute with greatest gain value will be the root node. From the Table 2, the attribute "Frequency of sharing the needles, syringe" is the strongest attribute to show the prevalence of hepatitis C virus among the injecting Drug User.

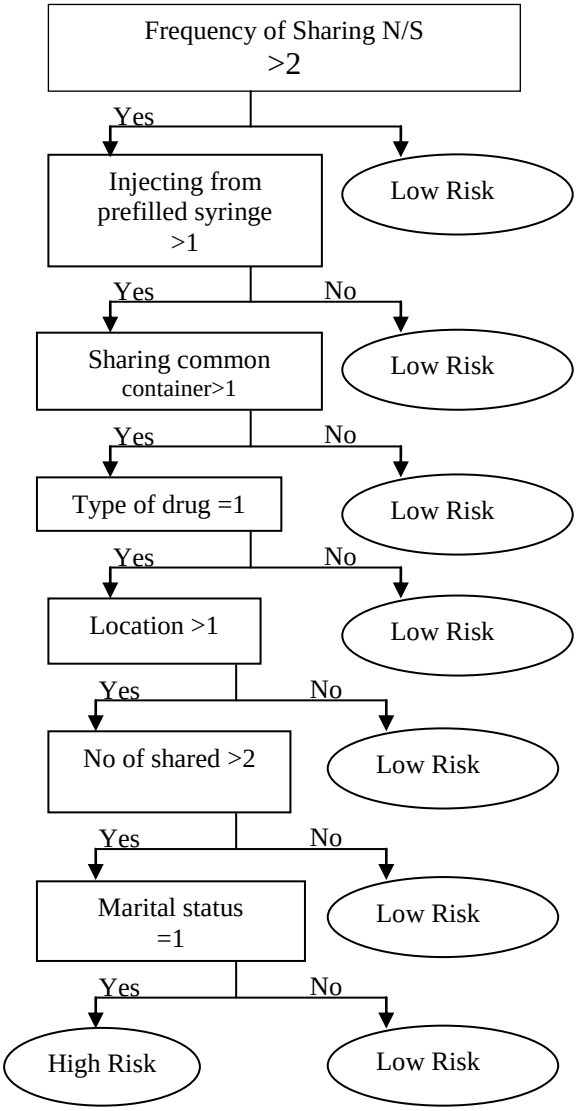

Fig.1.Decision Tree to show the prevalence of hepatitis C virus (HCV) among the injecting Drug User

Therefore the decision tree can be drawn with this attribute as root node which can be effectively used to show the prevalence of high risk of hepatitis C virus or low risk of hepatitis C virus among IDU's. The next attribute to show the prevalence of hepatitis C virus among the injecting Drug User is "injecting from prefilled syringe". Other attributes to be considered are sharing the Common Container, type of the drug, location of usage, Number of persons shared and Marital Status. Fig .1. shows the tree structure which specifies strongest attribute to identify the prevalence of hepatitis C virus among the injecting Drug User.

IV. CONCLUSION AND FUTURE WORK

Various data mining techniques can be collaborated with C4.5 classifier algorithm and can be used to check the prevalence of hepatitis C virus among IDU's. This proposed approach attained capable results, which may lead to further attempts to utilize Information Technology for the prediction other diseases among IDU's. Classifiers are playing a vital role in Data mining techniques, which are used in the prediction of diseases. From the experimental results and the real time data set, this work concludes that C4.5 algorithm is the most suitable algorithm to check the prevalence of hepatitis C virus among IDU's. The execution time of the same algorithm is also minimum. In future additional research is also needed to examine behaviors that might cause HCV infection in the absence of needle borrowing, such as sharing injection equipment used to dissolve and divide drugs.This technique using C4.5 can also be combined with other techniques and applied to the Data set to obtain high accuracy of results.

ACKNOWLEDGMENT

First and foremost, I would like to thank almighty for his abounded blessings showered on me in all my endeavour of life. My sincere thanks all my well-wishers who supported me and helped me to learn both in the scientific arena, and also on a personal level which brought a big impact on me in completing the paper.

REFERENCES

[1] Zanetti, Alessandro Remo, Global surveillance and control of hepatitis C Report of a WHO Consultation organized in collaboration with the Viral Hepatitis Prevention Board, Antwerp, Belgium. J Viral Hepat. 1999;6:35–47. [PubMed]
[2] Centre for Diseases Control and Prevention. 2015. Hepatitis C: 25 years Since Discovery. https://www.cdc.gov/knowmorehepatitis/media/pdfs/hepc-timeline.pdf
[3] Samiran Panda et.al "Alarming epidemics of human immunodeficiency virus and hepatitis C virus among injection drug users in the northwestern bordering state of Punjab, India: prevalence and correlates" International Journal of STD & AIDS,2013, DOI: 10.1177/0956462413515659.
[4] AbdElrazek et.al ,"Prediction of HCV vertical transmission: what factors should be optimized using data mining computational analysis " LIVER International Journal ,16 June 2016 Volume 37, Issue 4 April 2017 .
[5] Shungo Imai BS et.al "Usefulness of a decision tree model for the analysis of adverse drug reactions: Evaluation of a risk prediction model of vancomycin-associated nephrotoxicity constructed using a data mining procedure" "Journal of Evaluation in Clinical Practice-International Journal of Public Health Policy and Health Services Research" 23 May 2017 10.1111/jep.12767.
[6] Dr. P. Indra Muthu Meena, Dr. Vani Perumal "Performanceof C4.5 and Naïve Bayes Algorithm to Predict Stomach Cancer - An analysis" International Journal of Advanced Research in Computer and Communication Engineering ISO 3297:2007 Certified Vol. 5, Issue 11, November 2016.
[7] T. Miranda Lakshmi et.al.," An Analysis on Performance of Decision Tree Algorithms using Student's Qualitative Data", I.J. Modern Education and Computer Science, 2013, 5, 18-27.
[8] DebasishBasu et.al.," Hepatitis C virus (HCV) infection & risk factors for HCV positivity in injecting & non-injecting drug users attending a de-addiction centre in northern India"Indian J Med Res 142, September 2015, pp 311-316 DOI:10.4103/0971-5916.166596.
[9] Lopamudra Ray Saraswati et.al., "HIV, Hepatitis B and C among people who inject drugs: high prevalence of HIV and Hepatitis C RNA positive infections observed in Delhi, India" . BMC Public Health (2015) 15:726 .
[10] Santosh Kumar Sharma and Shri Kant Singh, "Co-infection with Human Immunodeficiency Virus and Hepatitis C Virus among Injecting Drug Users in North-Eastern States of India" Journal of Prevention & Infection Control 2016 Vol.2 No.2:9 ISSN 2471-9668.
[11] P.K. Chelleng et.al., "Risk of hepatitis C infection among injection drug users in Mizoram, India" Indian J Med Res 128, November 2008, pp 640-646 Regional Medical Research Centre, Northeast Region (Indian Council of Medical Research), Dibrugarh, India.
[12] Sunil S Solomon el.at.,"High prevalence of HIV, HIV/hepatitis C virus co-infection and risk behaviors among IDUs in Chennai, India: A Cause for Concern"NIH Public Access-J Acquir Immune DeficSyndr. 2008 November 1; 49(3): 327–332. doi:10.1097/QAI.0b013e3181831e85.
[13] Mohammed A.Farahatel.at.," Response Prediction for Chronic HCV Genotype 4 Patients to DAAs" (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 12, 2016.
[14] V.Kirubha1, S.ManjuPriya "Survey on Data Mining Algorithms in Disease Prediction" International Journal of Computer Trends and Technology (IJCTT) – Volume 38 Number 3 - August 2016.
[15] Mohammed M.Eissa et al., "Rough - Granular Neural Network Model for Making Treatment Decisions of Hepatitis C," the 9th International Conference on INFOrmatics and Systems (INFOS2014), December, 2014.
[16] E. M. F. El Houby and M. S. Hassan, "Using associative classification for treatment response prediction," Journal of Applied Sciences Research, vol. 8, no. 10, pp. 5089–5095, 2012.
[17] M. ElHefnawi, M. Abdalla, S. Ahmed et al., "Accurate prediction of response to Interferon-based therapy in Egyptian patients with Chronic Hepatitis C using machine-learning approaches, "in Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 803–810,August, 2012 .
[18] M. M. Eissa, M. Elmogy, M. Hashem, and F. A. Badria,"Hybrid rough genetic algorithm model for making treatment decisions of hepatitis C," in Proceedings of the 2nd Conference of Engineering and Technology and International (ICET '14), German University in Egypt, Cairo, Egypt, 2014.
[19] [15] Enas M. F. El Houby, "A Framework for Prediction of Response to HCV Therapy Using Different Data Mining Techniques," Hindawi Publishing Corporation, 2014.

[20] Lin E, Hwang Y, Wang SC, "Pharmacogenomics of drug efficacy in the interferon treatment of chronic hepatitis C using classification algorithms," Advances and Applications in Bioinformatics and Chemistry journal, 2010.
[21] Masayuki Kurosaki, et al, "A predictive model of response to peginterferon ribavirin in chronic hepatitis C using classification and regression tree analysis," Hepatology Research, 2010.
[22] Kazuaki Chayama et al., "Factors predictive of sustained virological response following 72 weeks of combination therapy for genotype 1b hepatitis C," J Gastroenterol (2011) 46:545–555, 2011.
[23] NaglaaZayed et al., "The assessment of data mining for the prediction of therapeutic outcome in 3719 Egyptian patients with chronic hepatitis C," clinics and Research in Hepatology and Gastroenterology, 2012.

BIOGRAPHIES

Mrs M.Gomathy received her M.C.A degree from the Presidency College, Madras University and M.Phil from Bharathidasan University. She is currently working as Assistant Professor in the Department of Computer Application, NBGSM College, NCR Delhi, Haryana, India.

Dr. Vani Perumal received M.C.A degree from the Department of Computer Applications, Bharathidasan University, M.Phil and Ph.D degree from the Department of Computer Science, Mother Teresa Women's University. She is currently working as an Assistant Professor in Rustaq College, Ministry of Higher Education, Sultanate of Oman. Her research interest includes Data mining, Machine learning, Pattern recognition, Biometric Image processing, Data Compression and Nano Technology.

.

DOI 10.17148/IJARCCE.2017.61112