IJARCCE

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

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 11, November 2016

Performance of C4.5 and Naïve Bayes Algorithm to Predict Stomach Cancer - An analysis

Dr. P. Indra Muthu Meena[1], Dr. Vani Perumal[2]

Assistant Commissioner& Oncology Researcher, Chennai, Tamil Nadu, India[1]

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

Abstract: Worldwide cancer survey reports that Stomach cancer is the third leading cause of cancer death in both sexes. The death rate can be minimized in early prediction of the same. Numerous Data Mining algorithms are used to mine the necessary information from the large set of data. It is also playing a significant role in various disease predictions with the aid of different algorithms. This paper concentrates the prediction of the deadly disease Stomach Cancer using two different data mining algorithms named C4.5 algorithm and Naïve Bayes algorithm. It also analyse the performance and the role of both the algorithms in prediction. The experimental results prove the functionalities and the importance of C4.5 and Naïve Bayes algorithm.

Keywords: C4.5 algorithm, Naïve Bayes algorithm, Stomach Cancer prediction, Entropy, Probabilistic Classification.

I. INTRODUCTION

Stomach cancer, also known as gastric cancer, is the accumulation of an abnormal malignant and cancerous group of cells that form a mass in a part of the stomach. According to the World Health Organization, 723,000 cancer-related deaths are caused by stomach cancer each year, globally [1]. It is the fifth most common cancer worldwide, but the third leading cause of cancer-related deaths. Stomach cancers are divided into cardia (the top part of the stomach, near the oesophagus) and non-cardia cancers, depending on where they first appear. It is more common in older adults. The bacterium Helicobacter pylori are an important cause of stomach cancer, particularly non-cardia cancer. Epstein-Barr virus, which is carcinogenic to humans, has also been linked to stomach cancer in some studies. Tobacco use is a cause of stomach cancer. Consumption of alcoholic drinks and foods preserved by salting are probably causes of stomach cancer. Consumption of processed meat is probably a cause of non-cardia stomach cancer [2].

Thus various surveys portrayed the necessity of early prediction of Stomach Cancer. The early prediction plays an imperative role in saving human life from early mortality. Thus a process of dredge up information from the enormous medical dataset is becoming an essential activity now a days. Data mining takes the sole responsibility of this dredging process. Data mining is called as data or knowledge discovery. It is a process of analysing data from different perspectives and summarizing it into useful information. Data mining software is one of a number of analytical tools for analysing data. It allows users to analyse data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

In this paper mining information is done with the help of two algorithms. For the data collection, the data from a questionnaire prepared by an oncology researcher and the database reports of medical patients were utilized. These data are given as an input to the algorithms for the prediction procedure. Also in this paper, the vital attributes which acts as a main cause of stomach cancer are considered for deriving the necessary information. The attributes are:

- Age - Age is classified into four major categories based on the occurrence of cancer. The classification is less than 30, 30-40, 40-50, greater than 50.
- Food habit - It is classified into four major categories like vegetarians (V), non-vegetarians (NV), non-vegetarians combined with frequent goat meat consumers (NV-M), non-vegetarians combined with frequent goat and beef meat consumers (NV-M&B).
- Menopause and Andropause - For female population, menopause is considered and for male population, andropause is considered.
- Profession - Three different broad classifications of profession is taken into account. They are Farmer (F), Sedentary Jobs (S) and Textile Professionals (T).
- Smoking habit - Here Active smokers and Passive smokers are considered as Smokers and the rest are considered as Non-smokers.
- Carbonated drink consumption - This criteria is divided into three categories like Mild (MI), Moderate (MO) and Excessive (EX) drinkers.
- Tobacco Use - This may take two possible values Yes or No.
- Hematology: Blood Group - Based on the occurrence of Stomach cancer, the blood group is divided as as A, B and Other than A&B groups (OT-A&B). A stands for both A+ve, A-ve and B stands for both B+ve, B-ve.

DOI 10.17148/IJARCCE.2016.511103

- Hematology: Oleic Acid measurement - The value 32.93 or above are the two different values considered for prediction.
- Hematology: Palmitic Acid measurement - The value 13.22 or above are measured as the values for the attributes.
- Hematology: Lino Leic Acid measurement - 25.20 or above is the different value for the same.
- Hematology: Amino Acid Profile: Proportion of Lucine and Iso Lucine measurement 2.3 or above is taken for the consideration.

The data set used for this algorithm were collected from the various districts from Tamil Nadu, India. The size of the training data set is one hundred with different values for all the above mentioned twelve attributes.

II. REVIEW OF RELATED RESEARCH WORK

A handful of researchers have utilized C4.5 algorithm and Naïve Bayes algorithm for numerous applications. A brief review of some recent and significant researches is presented here.

Miranda Lakshmi et. al. used student qualitative data taken from educational data mining and they analyse the performance of the decision tree algorithm using ID3, C4.5 and CART [3].

Seema Sharma et. al. proposed a C4.5 classifier based on the various entropies (Shannon Entropy, Havrda and Charvt entropy, Quadratic entropy) instance of Shannon entropy for classification [4].

Davinder Kaur et. al. provide a review of the decision tree algorithms. At first they present concept of Data Mining, Classification and Decision Tree. Then they present ID3 and C4.5 algorithms and then they made comparison of these two algorithms [5].

Badr Hssina et. al. present the classical algorithm, ID3. They also discuss C4.5 in more detail. Then they made a comparison between the two algorithms and others algorithms such as C5.0 and CART [6].

Harvinder Chauhan and Anu Chauhan implemented C4.5 algorithm using weka data mining tool using publicly available datasets of different size. They also gave insights into the rate of accuracy it provides when a dataset contains noisy data, missing data and large amount of data [7].

Dr. S. Vijayarani et.al used Naïve Bayes Algorithm and Support Vector Machine for the prediction of Liver disease. They predicted normal liver diseases, CBCL, Acute Hepatitis and Outliers using six attributes [8].

Ankita et.al utilized Naïve Bayes Classifier for the prediction of Swine Flu disease. They have used the values of eight different attributes for the prediction [9].

Sukhmeet Kaur et.al used Naïve Bayes algorithm for the prediction of future manufacturing of number of cars which is useful for the car manufacturing industry. They produced the prediction results. Then they compared the prediction results with the actual and real world values in order to validate the results. The utilized Naïve Bayes algorithm for predicting the result [10].

Dhamodharan et.al predicted three major liver diseases such as Liver cancer, Cirrhosis and Hepatitis with the help of distinctive symptoms. The authors used Naïve Bayes algorithm and FT Tree algorithm for the prediction of those diseases. Comparison of the performance of these two algorithms has been done based on the classification of accuracy measure. Based on the experimental results they concluded that the Naïve Bayes algorithm as a better algorithm for the prediction of the diseases with maximum classification accuracy than the other different algorithms [11].

Dhanashree et.al implemented a classifier approach for the detection of heart disease. Also they have shown how Naïve Bayes algorithm can be used for classification. They have used thirteen parameters as classifiers for prediction of heart disease [12].

Rosalina et.al predicted a hepatitis prognosis disease using SVM and Wrapper Method. First they have used wrapper methods to remove the noise features then they have used the classification process. Features selection was implemented to minimize the noisy or irrelevant data. Using the experimental results they observed the increased accuracy rate in the clinical laboratory test cost with minimum execution time. They achieved the target by combining Wrappers Method and Support Vector Machine techniques [13].

Omar S. Soliman et.al has projected a hybrid classification system for HCV diagnosis, using Modified Particle Swarm Optimization algorithm and Least Squares Support Vector Machine. In their paper, they used Feature vectors which are extracted using Principle Component Analysis algorithm. LS-SVM algorithm is sensitive to the changes of values of its parameters, so they used Modified-PSO Algorithm to search for the optimal values of LS-SVM parameters. They obtained in a less number of iterations. Their proposed system was implemented and evaluated in the benchmark HCV data set from UCI repository of machine learning databases. Then the database was compared with another classification system. That particular system utilized PCA and LS-SVM. From their experimental results, they proposed a new system, which obtained maximum classification accuracy than the other systems [14].

Sneha et.al used Apriori algorithm to classify the web pages based on the results extracted after submitting the query to the search engine. Then they applied Naïve Bayes algorithm to calculate the probability of each feature to classify the page into the respective class based on its probability [15].

Karthik et.al applied a soft computing technique for intelligent diagnosis of liver disease. They implemented classification detection and its type detection in three phases. In the first phase, they classified the disease using Artificial Neural Network classification algorithm. In the second phase, they generated the classification rules using Learn by Example algorithm. In the third phase fuzzy

DOI 10.17148/IJARCCE.2016.511103

rules were applied to identify the types of the liver disease. Thus they used ANN for classification [16].
Chaitrali S. Dangare et.al has structured prediction systems for Heart disease using more number of input attributes. They used the data mining classification techniques like Decision Trees, Naive Bayes and Neural Networks. The performances of these techniques are compared, based on accuracy. Their analysis shows that out of these three classification models Neural Networks has predicted the heart disease with highest accuracy [17].
Dipali Bhosale et.al used Naïve Bayes algorithm for feature selection. First they noted classification results without doing any kind of feature selection techniques. They also used Co-relation based Feature Selection, Wrapper, and Information Gain on the data sets. Then, by using these three feature selection techniques they separate feature subsets which are chosen for each technique then they passed the selected features to the classifiers and they derived the final results [18].

III. APPLICATION OF C4.5 IN STOMACH CANCER DATASET

This section deals with the complete algorithm, prediction methodology and experimental results by applying C4.5 in the Cancer 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 is 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 $\frac{|T_i|}{|T|}$ * Info(Ti) and the quantity Gain(X,T) defined as

$$\text{Gain}(X,T) = \text{Info}(T) - \text{Info}(X,T)$$

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. We can use this notion of gain to rank attributes and to build 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 are 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(Occam's Razor).
Thus the C4.5 algorithm can be implemented to predict the value of category attribute based on a set of observed values of non-categorical attributes.

B. Experimental Results of Stomach Cancer Prediction by C4.5

As mentioned above the observed dataset related to Stomach Cancer contains one hundred records of different values for twelve non-categorical attributes. The attributes and the possible values are given in Table 1.

TABLE I THE NON-CATEGORICAL ATTRIBUTES AND THEIR POSSIBLE VALUES

S. No	Attribute	Possible values
1.	Age	<30, 30-40, 40-50, >50
2.	Food Habit	V, NV, NV-M, NV-M&B
3.	Menopause/ Andropause	Yes, No
4.	Profession	F, S, T
5.	Smoking Habit	Yes, No
6.	Carbonated Drink Consumption	MI, MO, EX
7.	Blood Group	A, B, OT-A&B
8.	Oleic Acid ≥ 32.93	Yes, No
9.	Palmitic Acid ≥ 13.22	Yes, No
10.	Linoleic Acid ≥ 25.20	Yes, No
11.	Amino Acid: Prop. of L&IL ≥ 2.3	Yes, No
12.	Tobacco Consumption	Yes, No

DOI 10.17148/IJARCCE.2016.511103

For the above mentioned twelve different attributes, based on the observation of one hundred record, the Table. 2 shows the total number of entries which supports Yes and No in Stomach Cancer prediction. For the given values, the Information conveyed by this distribution, also called the Entropy of P is computed in I(P). First the information needed to identify the class of an element of T, where T set of records partitioned into disjoint exhaustive classes is computed (i.e) Info(T) = I(P) for the categorical attribute, where P is the probability distribution of the partition (C1, C2, .., Ck).

Then based on the values of the different attributes, Info(X, T) is computed for all the twelve different non-categorical attributes for building the decision tree. Using this Info(X, T) and Info(T), Gain(X, T) is calculated to rank the attributes and to determine which attribute offers a greater informational gain. Table 3 gives Info(X, T) and Gain(X, T) for all the twelve attributes.

TABLE III TWELVE ATTRIBUTES & TOTAL NUMBER OF ENTRIES SUPPORT YES/NO IN STOMACH CANCER PREDICTION

Attributes	**Possible Values**	**Number of Stomach Cancer Prediction**	
		Yes	**No**
Age	< 30	7	9
	30 – 40	17	11
	30 – 40	23	5
	> 50	23	5
Food Habit	V	8	12
	NV	18	9
	NV-M	19	7
	NV-M&B	25	2
Menopause/ Andropause	Yes	46	6
	No	24	24
Profession	F	26	8
	S	12	15
	T	32	7
Smoking Habit	Yes	54	11
	No	16	19
Carbonated Drink Consumption	MI	18	11
	MO	20	11
	EX	32	8
Blood Group	A	32	6
	B	26	12
	OT-A&B	12	12
Oleic Acid ≥ 32.93	Yes	48	18
	No	22	12
Palmitic Acid ≥ 13.22	Yes	47	18
	No	23	12
Linoleic Acid ≥ 25.20	Yes	48	13
	No	22	17
Amino Acid: Prop.of L&IL ≥ 2.3	Yes	50	15
	No	20	15
Tobacco Consumption	Yes	46	6
	No	24	24

From the above mentioned observed values, Entropy of P (i.e) I(P) is calculated using the total hundred records.

I(P) = 0.88129

To find Info(X, T) and Gain(X, T) for all the attributes, the above calculated Entropy of P is used. From the computed values of Info(X,T) and Gain(X,T), it is observed that Tobacco Consumption and Menopause/Andropause offers a greater informational gain. Thus the notion of gain is used to rank attributes and to build 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.

Then the immediate greater level of informational gain is obtained by Smoking Habit and there are very less differences among all the attributes since all are playing a vital role in stomach cancer prediction. The order of informational gain for the remaining attributes are Blood Group, Food Habit, content of Linoleic Acid and Amino Acid, Carbonated drink consumption, Age, content of Oleic Acid and Palmitic Acid consecutively.

TABLE IIII NON-CATEGORICAL ATTRIBUTES AND THEIR INFO(X, T) & GAIN(X, T)

Attribute - X	**Info(X, T)**	**Gain(X, T)**
Age	0.807933	0.073357
Food Habit	0.763479	0.117811
Menopause/ Andropause	0.748292	0.191707596
Profession	0.800005596	0.139994404
Smoking Habit	0.774505178	0.165494822
Carbonated Drink Consumption	0.857339507	0.082660493
Blood Group	0.821017333	0.118982667
Oleic Acid ≥ 32.93	0.876398528	0.063601472
Palmitic Acid ≥ 13.22	0.877930124	0.062069876
Linoleic Acid ≥ 25.20	0.841273838	0.098726162
Amino Acid: Prop.of L&IL ≥ 2.3	0.851407242	0.088592758
Tobacco Consumption	0.748292404	0.191707596

IV. APPLICATION OF NAÏVE BAYES IN STOMACH CANCER DATASET

In this section the theory behind the Naïve Bayes classification algorithm, the prediction methodology and experimental results by applying the same in Cancer dataset is discussed in detail.

A. Naïve Bayes Algorithm

DOI 10.17148/IJARCCE.2016.511103

The entire Naïve Bayes classification algorithm consists of three steps. They are given below:

Step. 1: Each data sample is represented as an n dimensional feature vector, X = (x1, x2….. xn). This depicts n measurements made on the sample from n different attributes like A1, A2…..An respectively.
Step. 2: Assume that there are n classes, C1, C2……Cn in a data sample X, the classifier will predict that X belongs to the class having the highest posterior probability, which is conditioned as:
if and only if: P(Ci/X)>P(Cj/X) for all 1< = j< = n and j!= i Thus P(Ci|X) is maximized. As mentioned above, thus the class Ci for which P(Ci|X) is maximized and referred as maximum posteriori hypothesis.
Step. 3: As described above, P(X) is constant for all classes, only P(X|Ci) P(Ci) need be maximized. But for the class, prior probabilities are unknown. Hence it is assumed that the classes are equally likely, and so that P(X|Ci) P(Ci) is maximized. If they are not equally likely, P(X|Ci) is maximized. Note that the class prior probabilities can be estimated by P(Ci) = Si/S On X, the Naïve Bayes probability assigns an unknown sample X to the class Ci. The idea based on this algorithm is now applied for the cancer data set for the prediction of stomach cancer.

C. Experimental Results of Stomach Cancer Prediction by Naïve Bayes
Now for the same data set, Naïve Bayes algorithm is applied for the prediction of stomach cancer. Table 4 is the resultant of the observation and application of the algorithm of the data set. This also gives the probability of the various attributes.

TABLE IVV PROBABILITY OF THE TWELVE ATTRIBUTES FOR STOMACH CANCER PREDICTION

Attributes	**Possible Values**	**Number of Stomach Cancer Prediction**	
		Yes	**No**
Age	< 30 30 – 40 30 – 40 > 50	7/70 17/70 23/70 23/70	9/30 11/30 5/30 5/30
Food Habit	V NV NV-M NV-M&B	8/70 18/70 19/70 25/70	12/30 9/30 7/30 2/30
Menopause/ Andropause	Yes No	46/70 24/70	6/30 24/30
Profession	F S T	26/70 12/70 32/70	8/30 15/30 7/30
Smoking Habit	Yes No	54/70 16/70	11/30 19/30
Carbonated Drink Consumption	MI MO EX	18/70 20/70 32/70	11/30 11/30 8/30
Blood Group	A B OT-A&B	32/70 26/70 12/70	6/30 12/30 12/30
Oleic Acid ≥ 32.93	Yes No	48/70 22/70	18/30 12/30
Palmitic Acid ≥ 13.22	Yes No	47/70 23/70	18/30 12/30
Linoleic Acid ≥ 25.20	Yes No	48/70 22/70	13/30 17/30
Amino Acid: Prop.of L&IL ≥ 2.3	Yes No	50/70 20/70	15/30 15/30
Tobacco Consumption	Yes No	46/70 24/70	6/30 24/30

From the table, first all possible probabilities conditioned on the target attribute of Stomach Cancer is computed by using Naïve Bayesian method.
P(St. Cancer = Y) = 70/100 = 0.7
P(Blood Group=A | St. Cancer = Y) = 0.457
P(Blood Group=B | St. Cancer = Y) = 0.371
P(Blood Group=OT-A&B | St. Cancer = Y) = 0.171
P(St. Cancer = N) = 30/100 = 0.3
P(Blood Group=A | St. Cancer = N) = 0.2
P(Blood Group=B | St. Cancer = N) = 0.4
P(Blood Group=OT-A&B | St. Cancer = N) = 0.4
Like the above mention method, all the attribute's possible probabilities were calculated using the probability table of given twelve attributes. Then by implementing the Naïve Bayes classifier and the information derived by using the Classification algorithm for the above data set, the possibility of stomach cancer can be predicted for any given required information.

V. CONCLUSION

Prediction and Classification are the principal data mining techniques which are largely used in healthcare sectors for medical diagnosis and predicting diseases. This work C4.5 algorithm and Naïve Bayes algorithm for the prediction of stomach cancer. Comparison of these two algorithms are made based on the performance and experimental results of the same. From the results it is observed that, C4.5 algorithm is used to rank the most important attribute which causes stomach cancer, based on the attribute the prediction can be done and Naïve Bayes algorithm is used to predict stomach cancer for any given report, based on the set of classifiers. Both algorithms using classifiers and training data set for prediction.
To predict the occurrence of cancer in any test data, both algorithms can be implemented one by one to obtain the accurate result. This proposed methodology attained promising results, which may infer to utilize Information Technology for the prediction of Stomach Cancer. The execution time of both the algorithms are also minimum. These algorithms run in linear time complexity for any

size of data set. From the experimental results and the real time data set, this methodology concludes that the combination of C4.5 algorithm and Naïve Bayes algorithm is the most suitable technique for the prediction of Stomach Cancer.

ACKNOWLEDGMENT

My heart full thanks to the Almighty for bestowing on me his everlasting love which has made me what I am today and also given me the courage to carry me through all the tomorrows. I would like to thank and render my heartfelt gratitude and abundant love to my Parents and family members for being always with me and for their vital encouragement and support.

REFERENCES

[1] Christian Nordqvist (2016) Stomach Cancer: Causes, Symptoms, and Treatments [Online]. Available: http://www.medicalnewstoday.com/articles/257341.php
[2] World Health Organizationv(2012) GLOBOCAN 2012: Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012 [Online]. Available: http://globocan.iarc.fr
[3] 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, vol. 5, pp. 18-27, May 2013.
[4] Seema Sharma et. al., "Classification Through Machine Learning Technique: C4.5 Algorithm based on Various Entropies", International Journal of Computer Applications, vol. 82, pp. 20-27, Nov. 2013.
[5] Davinder Kaur et. al., "Review of Decision Tree Data Mining Algorithms: ID3 and C4.5", in Proc. International Conference on Information Technology and Computer Science, 2015, pp. 5-8.
[6] Badr Hssina et. al., "A comparative study of decision tree ID3 and C4.5", International Journal of Advanced Computer Science and Applications, Special Issue on Advances in Vehicular Ad Hoc Networking and Applications, pp. 13-19, 2013.
[7] Harvinder Chauhan, Anu Chauhan, "Implementation of decision tree algorithm c4.5", International Journal of Scientific and Research Publications, vol. 3, pp. 1-3, Oct. 2013.
[8] Dr. S. Vijayarani, Mr.S.Dhayanand, "Liver Disease Prediction using SVM and Naïve Bayes Algorithms", International Journal of Science, Engineering and Technology Research, vol.4, pp. 816-820, Apr. 2015.
[9] Ms. Ankita et. al., "Naïve Bayes Classifier for Prediction of Swine Flu Disease", International Journal of Advanced Research in Computer Science and Software Engineering, vol. 5, pp. 120-123, Apr. 2015.
[10] Sukhmeet Kaur, Kiran Jyoti, "Predicting the future of car manufacturing industry using Naïve Bayse Classifier", International Journal for Science and Emerging Technologies with Latest Trends, vol. 4, pp. 25-34, 2012.
[11] Dhamodharan. S, "Liver Disease Prediction Using Bayesian Classification", in Proc. 4th National Conference on Advanced Computing, Applications & Technologies, 2014, pp. 1-3.
[12] Dhanashree S et. al., "Heart Disease Prediction System using Naïve Bayes", International Journal of Enhanced Research in Science Technology & Engineering, vol. 2, pp. 1-5, Mar. 2013.
[13] Rosalina. A. H, Noraziah. A, "Prediction of Hepatitis Prognosis Using Support Vector Machine and Wrapper Method" , IEEE, pp. 2209-2222, 2010"
[14] Omar S.Soliman, Eman Abo Elhamd, "Classification of Hepatitis C Virus using Modified Particle Swarm Optimization and Least Squares Support Vector Machine", International Journal of Scientific & Engineering Research, vol. 5, pp. 122-129, Mar. 2014.
[15] Sneha K. Dehankar et. al., "Web Page Classification using Apriori Algorithm and Naïve Base Classifier", International Journal of Advanced Research in Computer Science and Management Studies, vol. 3, pp. 527-533, Apr. 2015.
[16] Karthik. S et. al., "Classification and Rule Extraction using Rough Set for Diagnosis of Liver Disease and its Types", Advances in Applied Science Research, vol.2, pp. 334-345, Jun. 2011.
[17] Chaitrali S. Dangare and Sulabha S. Apte, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques", International Journal of Computer Applications, vol. 47, pp. 44-48, Jun.2012.
[18] Dipali Bhosale and Roshani Ade, "Feature Selection based Classification using Naive Bayes, J48 and Support Vector Machine", International Journal of Computer Applications, vol.99, pp. 14-18, Aug. 2014.

BIOGRAPHIES

Dr. P. Indra Muthu Meena received the M.Sc degree from the Department of Zoology, University of Madras and Ph.D degree from the Department of Biotechnology, Mother Teresa Women's University. She is currently working as an Assistant Commissioner in the Commercial Tax Department, Tamil Nadu, India. She is also doing research in Oncology. Her research interest includes Monoclonal Antibodies, Cancer Biology and Genome analysis.

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 of Applied Sciences, Ministry of Higher Education, Sultanate of Oman. Her research interest includes Data mining, Machine learning, Pattern recognition, Biometric Image processing and Data compression.

DOI 10.17148/IJARCCE.2016.511103