Detection of Hepatitis (A, B, C and E) Viruses Based on Random Forest, K-nearest and Naïve Bayes Classifier

Tahira Islam Trishna
Computer Science &Engineering
North Western University
Khulna, Bangladesh
tahidaislamtrisna@gmail.com

Shimoul Uddin Emon
Computer Science & Engineering
North Western University
Khulna, Bangladesh
shimoul.cse@gmail.com

Romana Rahman Ema
Computer Science & Engineering
North Western University
Khulna, Bangladesh
romanacsejstu@gmail.com

Gazi Imran Hossen Sajal
Computer Science & Engineering
North Western University
Khulna, Bangladesh
imransajal30@gmail.com

Shuvodip Kundu
Computer Science & Engineering
North Western University
Khulna, Bangladesh
sdshuvo65@gmail.com

Tajul Islam
Computer Science & Engineering
North Western University
Khulna, Bangladesh
tajulkuet09@gmail.com

***Abstract*— Recently, the techniques of data mining are broadly used to analyze biomedical data. These techniques have given efficient results in the prediction and classification of diseases severity and diagnosis of sicknesses. Hepatitis is a liver inflammation; it can affect people of all ages. Millions of people globally are thought to be affected by hepatitis. Accurate and early diagnosis of hepatitis can save many patients. Hepatitis is a key challenge for public health care services, due to limited clinical diagnosis of hepatitis disease in its early stages. This paper presents different techniques of data mining techniques used for diagnosing hepatitis and shows the results of different techniques in term of accuracy and training time. We used naive bayes, K-nearest and Random Forest classifier in WEKA software to calculate the result. Naïve bayes is the algorithm that used for solving text classification problems. K-nearest neighbor is a simple, supervised machine learning algorithm that is used for both regression and classification problem. Random Forest is one of the most used algorithms for its simplicity and the reason that it can be used for both regression and classification. In naïve bayes accuracy of our result can be found 93.2%. In Random Forest classifier the accuracy is 98.6% by using 10 fold cross-validation and in K-nearest neghbor it is become 95.8% by using 10 fold cross-validation.**

Keywords—hepatitis A virus(HAV); hepatitis B virus(HBV); hepatitis C virus(HCV); hepatitis E virus(HEV); naïve bayes; Random Forest; K-nearest; classification; accuracy

I. INTRODUCTION

A. Background

Hepatitis is defined as an inflammation of the liver and most commonly it is occured by a viral infection [1] causing overall 1.5 million deaths all over the world every year [2]. This viruses tent to target the cells in the liver.Hepatitis can be occurred due to viruses, bacteria, drugs, etc. This disease can be classified as Acute or Chronic. There are five main Hepatitis viruses, referred to as types A, B, C, D and E.Hepatitis B virus is the most familiar problem and a serious health issue which affects all most 2 billion people all over the world including 350 million chronic carriers from which 75% are in the Asia-Pacific zone [5]. Chronic hepatitis C is a worldwide public health issue estimated to have over 180 million patients that are expected to increase three times by 2020 [6] and worldwide up to 250 million people are thought to be attacked by hepatitis C [7].

Hepatitis E and A are caused by oral infection, contaminated water and unhygienic food [8]. It is specifically occurs in developing countries especially rural communities due to the poor hygienic conditions. Hepatitis B, C and D are considered as global health problem and can easily transfer to individuals through intercourse, blood and birth from infected mothers [8,9]. Hepatitis C more principally spread through blood transfusions and use of un sterilized needles and equipment [9]. For the prevention of Hepatitis B effective vaccine is available globally and locally although very expensive for poor people to afford [keyani1]. For Hepatitis C, globally no vaccines are available and therefore preventive measures and awareness information's disseminated is mandatory.HEV also occurs through transmission by the fecal–oral same as HAV.

B. Paper Organization

Current paper illustrates the following subdivisions as follows: Related works are described in the section II. Section III elaborates the problem formulation and the methodology is formulated in section IV. Section V is presented the result and analysis. Finally in section VI the final conclusion is concluded.

II. RELATED WORKS

Nireekshith.Yarraguntla et al. [9] implemented a system using biosensor that can predict Hepatitis viruses(HAV, HBV, HCV) by immobilizing specific antibodies on the sensing element of the sensor when the sample of the patient bearing hepatitis viruses surface antibody (anti-HbsAg) is settled on the sensing part, Biochemical interactions occur between Hepatitis viruses antigens and the antibodies standing on the sensing element. For this reason, beam structure is changed so that diseases can be detected.

S.L. Znoyko et al. [10] presented real time optical methods for finding the HBV surface antigen. An optical technique has been elaborated for screening of functionalized nanoparticles and reagents for the design of immunochromatographic and capacitive biosensors. The developed immunosensors were tested for rapid quantitative detection of HBV surface antigen, which is the marker for diagnostic of hepatitis B.

E. Alipour et al. [11] designed a capacitive immunosensor to detect the HBV surface antigen. Their result show the presence of HBsAg which affects most of the time to the surface area of the capacitors and for this there is an enhancement in the capacitance. They found the results by developing an immunoassay capacitor using gold nanoparticles as amplifiers.

Chunyan Yang, JinJise Song [13] gave a model on Hepatitis diagnosis which is based on GA-BP neural network algorithm and fuzzy integral. In order to develop the accuracy of hepatitis diagnosis in computer-aided diagnosis system, the medical data from the network public database were optimized by GA-BP neural network algorithm. From the results of calculation, using fuzzy integral can greatly improve the accuracy of hepatitis auxiliary diagnosis.

Huina Wang, Yihui Liu, Wei Huang [16] did their work to detect the risk factors for hepatitis B virus reactivation after the precise radiotherapy in patients with primary liver cancer (PLC). We use sequential forward selection and sequential backward selection to extract features which would be combined into an optimal feature subset, and then establish Bayesian and support vector machine (SVM) classification model. The experimental results showed that the key feature subset has a better classification performance than the initial feature set clearly.

Sara Omer Hussien et al. [7] presented an overview of the recent state-of-the-art data mining techniques used for diagnosing hepatitis and gives the performance of various techniques in term of training time and accuracy. Such review helps in the implementation, development and evaluation of efficient clinical decision support systems; where accurate diagnosis is the most important factor.

Jiaxin Cai, Tingting Chen, Xuan Qiu [3] proposed an automatic diagnosis system of chronic hepatitis C using serum indices data of patients to predict the fibrosis stage and inflammatory activity grade of chronic hepatitis C by training the extreme learning machine. Due to the superiority of extreme learning machine such as simple structure and fast calculation speed, the presented automatic diagnosis system can achieve good diagnosis performance. The proposed automatic diagnosis system is test on real clinical cases of chronic hepatitis C based on serum indices. Experimental results explain that the performance of the given method overcomes that of the state-of-the-art baselines concerning the diagnosis of fibrosis stage and inflammatory activity grade of chronic hepatitis C.

Tasneem A. Gameel et al. [14] predicted the hepatitis C infection progression into cirrhosis or liver cancer. For the prediction of the disease progression, a knowledge discovery framework is proposed consisting of three phases: preprocessing, data mining and prediction. While the preprocessing phase focuses on the discretization of the training data, the data mining phase focuses on mining patients' records using a rule based classifier built by the proposed algorithm to generate a set of unique rules. Eventually, the predictor uses the rules to predict patients' disease progression.

A. Motivation

A large number of the Hepatitis carriers (humans) are completely unaware about their conditions and preventive procedures. The lack of proper health facilities, poor economic status, incompetent health care staff and unawareness about the disease and its prevention lead to a chronic stage of Hepatitis, almost untreatable and so expensive that a poor person could not afford such expense. Although vaccines are there but still no known cure for Hepatitis could be found. Moreover, hepatitis represents a heavy burden on the health care system because of the costs of treatment of liver failure.Timely prediction and accurate diagnosis of disease can save many affected people.

The primary objective of the paper is analysis of data from a hepatitis dataset using different classification techniques to predict the result accurately in each case of data. Major contributions of the paper are:

- To measure useful classification accuracy for the prediction of hepatitis diseases.
- Comparision of various data mining algorithms on the hepatitis dataset.
- Identify the best algorithm performance for the prediction of hepatitis diseases.

III. Material And Methods

A. Clicinal Data

All the data were collected from different hospitals. We went to the hospitals and discussed with a few number of specialists. We also discussed with the hepatitis patients of the hospitals collected their symptoms and also the tests and result from their test report. The dataset is checked by a gastroenterologist. There are 25 attributes for the automatic diagnosis system. The list of these 25 attributes is given below within a Table I. The attributes has two parts: symptoms and tests. These parts are briefly described in the section IV.

The entire dataset is divided into two sets one is used for training and the other to test them. These samples include five classes (HEP A, HEP B, HEP C, HEP E and NONE). The attribute list of data is given below in TABLE I.

TABLE I. Attributes List Of Data

Number	***Attributes***	***Parts***
1	AGE	SYMPTOMS
2	GENDER	
3	ANOREXIA	
4	JAUNDICE	
5	FEVER	
6	RASH	
7	FATIGUE	
8	NAUSEA	
9	WEAKNESS	
10	LIVER PAIN	
11	MASALE PAIN	
12	ITCHING SKIN	
13	PREGNENCY	
14	Anti HAV IgM	TESTS
15	Anti HAV IgG	
16	Anti HBe IgM	
17	HbsAg	
18	HbeAg	
19	Anti HBsAg IgG	
20	Anti HBc IgM	
21	Anti HCV	
22	RIBA	
23	PCR	
24	Anti HEV IgM	
25	Anti HEV IgG	

B. Classifiers:

We used Naïve Bayesian, K Nearest-Neighbor (KNN) and J48 classifier in Weka 3.8 version software.

Naive Bayesian is a probabilistic statistical classifier. The term "naive" refer to a conditional independence among features or attributes as it assumes that the presence of features in a data set is not associated to the presence other feature. The "naive" assumption lessens computation complexity to a general multiplication of probabilities. High speed is the main advantage of the Naive Bayesian classifier as it is the simplest algorithm among classification algorithms [new]. The simplicity of this algorithm helped in handling datasets with high dimensional feature space easily. In addition, naive Bayesian classifier develops accurate parameter estimations with small training data sets only as it calculates the pairs and attributes frequencies from the training datasets [7]. The algorithm calculate posterior probability P(c|x) from probability of x, y (p(x), p(c) and p(x|c)) as illustrated in the equation .

$$P(C \mid X) = \frac{\mathrm{P}(X \mid C)\, P(C)}{\mathrm{P}(\mathrm{X})} \qquad \text{(I)}$$

Random Forest classifier is an ensemble of decision tree classifiers [25]. Each decision tree on the forest is trained on a bootstrap samples or subsamples from the original training data, which adds more diversity to the decision trees. In each tree, samples are pushed from the root node to the leaf node by performing a binary test at each internal node. In those binary test a certain feature is compared with a threshold value [24]. In the training step, for each node a binary test on a particular feature is identified by optimizing the information gain in the training dataset. Depending on the outcome of the binary test, each training sample is sent to the corresponding child node and this process is recursively repeated until the number of samples in a node falls below a certain threshold.

K Nearest-Neighbour (KNN) is one of the easiest classification techniques. If there is no prior knowledge about the distribution of country data KNN is one of the first choices for these identification problems. KNN classifier has been both a benchmark classifier algorithm. KNN

classifier performance determined by choice of K as well as the distance metric applied. In the predetermining of the K value difficult when the points are uniformly distributed [22].

The classification itself is performed in two steps: (1) finding K nearest neighbors in the training dataset, and: (2) assigning a label based on majority voting amongst the K neighbors [23]. The neighbors are determined by calculating the Euclidian distance between the new observation and the instances in the training dataset.

IV. METHODOLOGY

A. Data Preprocessing Phase

The clinical appearances of hepatitis are virtually the same, regardless of which hepatitis virus is the cause (Table II). Fever, anorexia, nausea, vomiting, and jaundice are typical. Dark urine, pales faces, and elevated transaminase levels are seen. Most cases resolve effortlessly within 1 month. Hepatitis A has a short incubation period (3-4 weeks) in contrast to that of hepatitis B, which is 10 to 12 weeks.

Most hepatitis A virus infections are the mostly unseen symptom and are detected only by the presence of IgM antibody. It is important to detect IgM antibody for HAV test. The increment of 4-fold in anti-HAV IgG antibody can also be used as a current infection. Isolation of the virus in cell culture is possible but not available in the clinical laboratory. The two most important serological tests for the diagnosis of early hepatitis B are the tests for HBsAg and for IgM antibody. Both occur in the serum earlier on the disease. The availability of HBsAg, Anti-HBV IgM, and HBeAg is indicating the early stage (acute) of hepatitis B. If only Anti-HBc IgG is presented it indicates Window period. The persistent HBsAg, HBeAg, and attendance of Anti-HBc IgG refer to the chronic stage with viral activity and infectivity. The emergence of Anti-HCV indicates the current infection by hepatitis C virus. There is another test namely RIBA (Recombinant immunoblot assay) for confirmation of HCV. HEV antibody test is not readily obtainable, the diagnosis is therefore typically made by detecting IgM antibody to HEV. High probability of infected by HEV in pregnancy period of a woman. Mode of transmission for HBV and HCV are almost the same but the fact which matters is the percentage. According to the percentage we categorize mode of transmission into Horizontal and Vertical.

Horizontal Transmission:

- Spread of the infected unexposed blood products, sexual intercourse, injection drug use and tattoos acupuncture needles.

Vertical Transmission:

- Transmission from hepatitis infected mother to child.

TABLE II. TEST AND TRANSMISSION

Virus	***Mode of Transmission***	***Laboratory Test Used for Diagnosis***
1.HAV	Fecal-oral	IgM HAV
2.HBV	Blood, sexual, at birth	HBsAg, HBsAb, IgM HBcAb
3.HCV	Blood, sexual	HCV Ab
5.HEV	Fecal-oral	None

The visualization diagram overall methodology is given below:

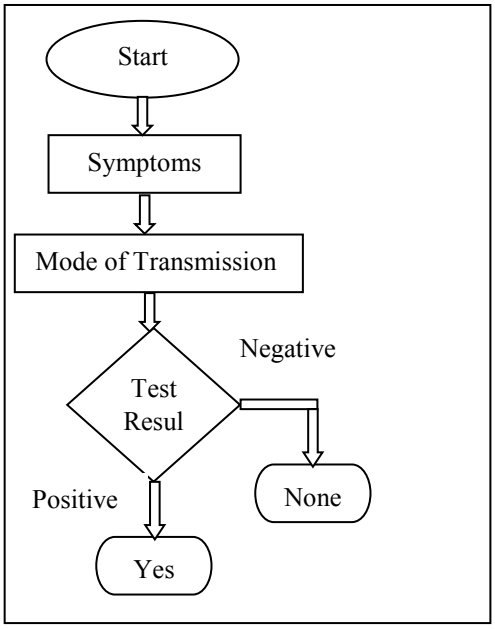

Fig. 1: Visualization diagram of methodology

B. Data Mining Phase

Different data mining techniques are provided in this section which are used for diagnosing hepatitis in the literature. The data mining

techniques are applied to the training set to do clustering, classification or prediction. There are two phases in data mining techniques:

- Training/learning phase in which the model is trained by pairing the input with expected output. There are three learning approaches namely, supervised learning, semi-supervised and unsupervised learning.
- Testing/validation phase: used to evaluate the quality of the model. Fig. 1 shows the overall data mining process.

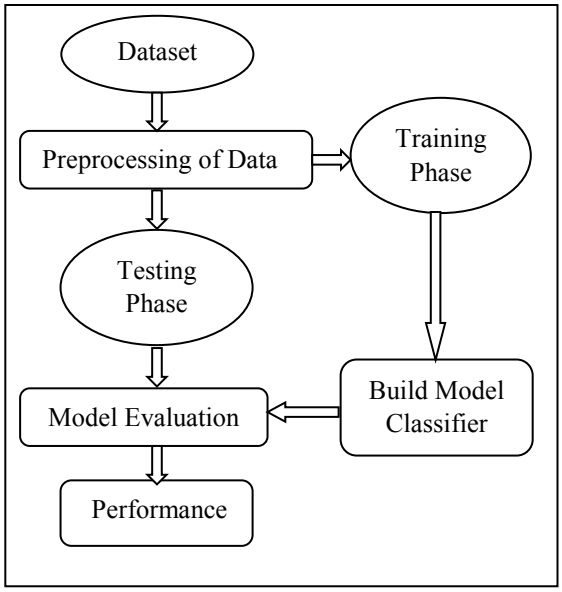

Fig. 2: Data mining process.

V. Results And Analysis

A. Performance Measurement

The accuracy rate is used as the evaluation exponent of the model in the classification problem, and the accuracy rate represents the ability of the classifier to judge the whole data. In short, the accuracy rate means the number of right classified dataset partitioned by the number of all datasets. In general we know, the higher the accuracy, the better the classifier. But for unbalanced data, the correct recognition of the minority is often more meaningful than the overall accuracy, so accuracy is not appropriate for all cases. Taking a classification problem as an example, its confusion matrix is given in Table III:

TABLE III. Confusion Matrix

Class	***Predicted as positive***	***Predicted as negative***
Positive in fact	TP	TN
Negative in fact	FP	FN
Total number	P	N

$$\text{Accuracy} = \frac{TP+TN}{P+N} \quad \text{(II)}$$

$$\text{TPR/Sensitivity} = \frac{TP}{P} \quad \text{(III)}$$

$$\text{TNR/Specificity} = \frac{TN}{N} \quad \text{(IV)}$$

TPR represents the correct proportion of the true positive cases; TNR represents the correct proportion of the true negative cases. The most ideal classifier means all the samples will be classified correctly, so that means: FP=0, FN=0. Therefore, for the ideal classifier: TPR=1, TNR=1.

The experiment took the k-fold cross validation. Suppose the sample number is S and then divided into k piece, each piece has S/k samples. The k-fold cross validation defines as follows:

$$\overline{x} = \frac{1}{k} \sum_{i=1}^{k} x_i \quad \text{(V)}$$

The different result of key class of selected by sequential forward selection with Naive Bayes classifier, Random Forest and K-nearest classifier are given in Table IV, Table V, Table VI respectively and average results are in Table VII.

TABLE IV. Naïve Bayes Using 10 Fold

Class	***Accuracy***	***Error***	***TPR***	***TNR***	***Precision***
HEP A	94%	6%	95.8%	93.7%	61.6%
HEP B	98%	2%	100%	97%	89%
HEP C	97%	3%	98.6%	96.8%	76.6%
HEP E	94%	6%	98.6%	93.9%	63.2%
NONE	83%	17%	67.9%	99.2%	98.9%

The results of the accuracy, error rate, TPR, TNR and precision are given in the TABLE IV for the Naïve bayes classifier using 10 fold cross validation.

TABLE V. RANDOM FOREST USING 10 FOLD

Class	*Accuracy*	*Error*	*TPR*	*TNR*	*Precision*
HEP A	99%	1%	92.4%	99.7%	97%
HEP B	99%	1%	99.7%	1%	1%
HEP C	99%	1%	94.5%	99.1%	92%
HEP E	99%	1%	93.8%	99.5%	95.1%
NONE	97%	3%	97.5%	96.5%	96.8%

The results of the accuracy, error rate, TPR, TNR and precision are given in the TABLE V for the Random Forest.

TABLE VI. K-NEAREST USING 10 FOLD

Class	*Accuracy*	*Error*	*TPR*	*TNR*	*Precision*
HEP A	97%	3%	85.4%	98.5%	85.4%
HEP B	98%	2%	96.6%	98.8%	95.3%
HEP C	98%	2%	89%	99.3%	93.5%
HEP E	96%	4%	78.6%	97.5%	77%
NON E	90%	10%	90.8%	90.2%	90.8%

The results of the accuracy, error rate, TPR, TNR and precision are given in the TABLE VI for the K-nearest classifier.

TABLE VII. AVERAGE RESULTS OF CLASSIFIER

Classifier	*Accuracy*	*Error*	*TPR*	*TNR*	*Precision*
Naïve Bayes	93.2%	6.8%	92.2 %	96.1 %	77.8%
Random Forest	98.6%	1.4%	95.6 %	98.9 %	96.2%
K-nearest	95.8%	4.2%	88.1 %	96.9 %	88.4%

The average results of the accuracy, error rate, TPR, TNR and precision are given for the all three classifiers in the table VII.

B. Performance Analysis

The original feature set has different accuracy, sensitivity, specificity, precision, F-score by using different classifier. The experimental result of Random Forest and K-nearest classifier is higher than Naive Bayes classifier. The accuracy of Naive Bayes classifier reached 93.2% by using 10 fold cross-validation. This may occur because of a large number of training data in the experiment. When uses 10 fold cross-validation, the accuracy of Random Forest classifier reached 98.6% and the accuracy of K-nearest classifier reached 95.8%. The result shows that Random Forest classifier is better than Naive Bayes classifier and K-nearest classifier.

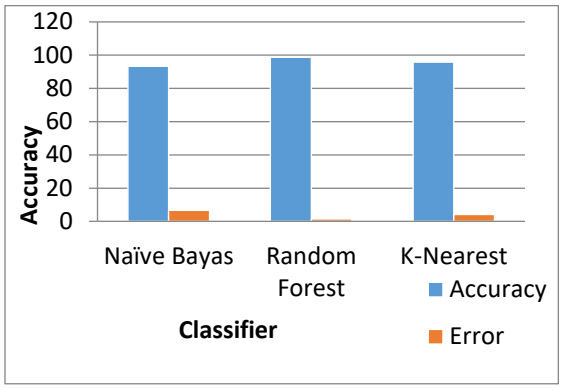

Fig. 3: Comparison graph of accuracy of classifiers

VI. CONCLUSION

In, rural Bangladesh, Hepatitis also known as 'silent killer' is spreading at alarming rate claiming lives of millions of people every year yet it is difficult for health authorities to evaluate its burden and impact. This work introduces an automatic diagnosis system using different data mining processes on Weka to predict the hepatitis viruses. The proposed system has been applied on real data of hepatitis patients. Experimental result prove that the proposed system is effective for diagnosing the hepatitis viruses. In this paper we didn't work with hepatitis D because there is no specific test for hepatitis D. Moreover hepatitis D is hardly seen in Bangladesh.

In this paper Weka data mining tool is used. It has four interfaces: Explorer, Experiment, Knowledge flow and Simple CLI. Out of this we used only the Explorer interface. Each interfaces has its own classifier algorithms. We have used only 3 classifier.

In the future, larger data set will be considered to build the model which will produce more unique rules so giving higher accuracy. To enhance the accuracy more, different rule weighting techniques are suggested. It is possible to extend the research by using different classification techniques.

REFERENCES

[1] Keyani S. S. , Mumtaz A. A. , Ahmad A. "Hepatitis Surveillance System for rural Pakistan through web and mobile based technologies", 11th Annual High Capacity Optical Networks and Emerging/Enabling Technologies (Photonics for Energy), 2014, pg. no. 1-5

[2] S. Pushpalatha, Dr. Jagdish G. Pandya, "Designing A Framework for Diagnosing Hepatitis Disease Using Data Mining Techniques", IEEE 2017, pg. no. 1-6

[3] Jiaxin Cai, Tingting Chen, Xuan Qui, "Fibrosis and Inflammatory Activity Analysis of Chronic Hepatitis C

Based on Extreme Learning Machine", IEEE 2018, pg. no. 177-181
[4] Sana AnsariI, Imran ShafiI, Aiza Ansari2, Jamil AhmadI, Syed Ismail Shah, "Diagnosis Of Liver Disease Induced By Hepatitis Virus Using Artificial Neural Networks", IEEE 2011, pg. no. 8-12
[5] Faisal Nouroz, Sidra Shaheen, Ghulam Mujtaba, Shumaila Noreen, "An overview on hepatitis C virus genotypes and its control",Volume 16, Issue 4, October 2015, pg. no. 291-298
[6] Iram Zubair, Bilal Wajid, "Comparison of APRI, FIB-4 and Fibro Test in Prediction of Fibrosis and Cirrhosis in Patients with Hepatitis C", IEEE 2018, pg. no. 222-227
[7] Sara Omer Hussien, Sara Sir Elkhatem, Nisreen Osman, Ashraf Osman Ibrahim, "A Review of Data Mining Techniques for Diagnosing Hepatitis", IEEE 2017, pg. no. 1 - 6
[8] Barbara G. Wells,PharmD, FASHP, FCCP, BCPP Dean and Professor Executive Director, Research Institute of Pharmaceutical Sciences School of Pharmacy, The University of Mississippi Oxford, Mississippi, Pharmacotherapy Handbook Seventh Edition, pg. no. 273-281
[9] Nireekshith.Yarraguntla, Naveena Tirumala, Syed.Shameem, K.srinivasa rao, "Detection of Hepatitis viruses (HBV, HAV, HCV) in serum using MEMS based Bio-Sensor", IEEE 2018, pg. no. 405-409
[10] S.L. Znoyko1, V.A. Bragina1,E. Alipour,H. Ghourchian,P.I. Nikitin1, "Real-Time Optical Methods for Development of Nanoparticle-Based Biosensors for Detection of Hepatitis B Surface Antigen", IEEE 2018, pg. no. 578-578
[11] E. Alipour1, H. Ghourchian1,2, S.L. Znoyko3,P.I. Nikitin3,4, "Designing a Capacitive Immunosensor for Detection of Hepatitis B Surface Antigen", IEEE 2018, pg. no. 562-562
[12] S Pushpalatha, Dr. Jagdesh Pandya, "Comparison of Data Models for Hepatitis Diagnosis – Data Mining Technique", Volume 4, Issue 1, January 2014, pg. no. 1092-1095
[13] Chunyan Yang, JinJie Song, "Research on Hepatitis Auxiliary Diagnosis Model Based on Fuzzy Integral and GA - BP Neural Network", IEEE 2017, pg. no. 664 - 667
[14] Tasneem A. Gameel, Sherine Rady, Khaled A. El-Bahnasy, Sanaa M. Kamal, "Prediction of Liver Cancer Development Risk in Genotype 4 Hepatitis C Patients using Knowledge Discovery Modeling", IEEE 2017, pg. no. 410 - 415
[15] Waqar Hassan, Nadia Jamil, M. Israr, Maliha Wajeeh, "Study of gender preference of hepatitis virus in suspected patients from Islamabad and Karachi", IEEE 2018, pg. no. 231-234
[16] Huina Wang, Yihui Liu, Wei Huang, "Random Forest and Bayesian Prediction for Hepatitis B Virus Reactivation", IEEE 2017, pg. no. 2060 - 2064
[17] J Hepatol, 'Journal of Hepatology', 2016. [online]. Available:https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4914770/ [Accessed: 05.02.2019]
[18] Current Trending News, 'Viral Hepatitis (A, B, C, D, E) – Causes, Symptoms, Diagnosis & Treatment',2018. [online].Available: https://www.currenttrending.com/viral-hepatitis/ [12.02.2019]
[19] Dayana C. Tejera Hernandez, "An Experimental Study of K* Algorithm", 2015. [online]. Available: http://www.mecs-press.org/ijieeb/ijieeb-v7-n2/IJIEEB-V7-N2-3.pdf. [Accessed: 12- 02- 2019]
[20] Bamshad Mobasher, "Classification Via Decision Trees in WEKA",Available:http://facweb.cs.depaul.edu/mobasher/classes/ect584/weka/classify.html.[Accessed: 15.02.2019]
[21] W. Levinson, REVIEW OF Medical Microbiology and Immunology. 14th Edition, p. 342-352
[22] Ahmed S EI Rawas, Hamhi A Mahmoud, "Quality Assurance for Economy Classification based on Data Mining Techniques", Bus Eco J 2017, Vol 8(4): 333
[23] Stéphanie Allogba, Christine Tremblay, "K-Nearest Neighbors Classifier for Field Bit Error Rate Data", 2018 Asia Communications and Photonics Conference (ACP).
[24] Suman Sedai, Pallab Kanti Roy, Rahil Garnavi, "Right Ventricle Landmark Detection Using Multiscale Hog and Random Forest Classifier", IEEE 2015, pg. no. 814-818
[25] R. Gayathri Devi, Sumanjani .P, "Improved classification techniques by combining KNN and Random Forest with Naive Bayesian Classifier", IEEE 2015, pg. no. 1-4