International Journal of Hospital Research 2014, **3**(1):1-10
www.ijhr.iums.ac.ir
RESEARCH ARTICLE

IJHR

Open Access

A Novel Algorithm for Accurate Diagnosis of Hepatitis B and Its Severity

Khosro Rezaee[1*], Javad Haddadnia [2], Mohammad Rasegh Ghezelbash [3]

[1] Biomedical Engineering Student, Biomedical Engineering Group, Department of Electrical and Computer Engineering, Hakim Sabzevari University of Sabzevar, Sabzevar, Iran [2] Biomedical Engineering Group, Department of Electrical and Computer Engineering, Hakim Sabzevari University, Sabzevar, Iran [3] Department of Neurobiology, Care Sciences and Society (NVS), Karolinska Institute, Stockholm, Sweden

Abstract

Background and Objectives: Accurate detection of type and severity of Hepatitis is crucial for effective treatment of the disease. While several computational algorithms for detection of Hepatitis have been proposed to date, their limited performance leaves room for further improvement. This paper proposes a novel computational method for the diagnosis of Hepatitis B using pattern detection techniques.

Methods: Clinical data of healthy individuals and those suspected of Hepatitis B were collected from the laboratories of Vasei Hospital in Sabzevar (Iran). Using the algorithm, first, data were normalized, then SVM classifier was used for detection of Hepatitis B, and finally, adaptive FCM was applied for measuring the severity of the disease.

Findings: Application of the algorithm to plenary database yielded 98.36%, detection accuracy, 98.44% sensitivity, and 94.06% specificity.

Conclusions: Low PPV and high NPV of the proposed method indicate its high reliability for use in practical diagnosis of Hepatitis B and its severity.

Keywords: Hepatitis B; Chronic Hepatitis; Acute Hepatitis; Diagnosis; Artificial Intelligence; Data Mining; Support Vector Machines (SVM); Fuzzy C-Means (FCM)

Background and Objectives

Hepatitis is the inflammation of the liver, which is commonly known as jaundice [1]. Most people suffering from Hepatitis types B and C do not have any symptoms [2]. When a person becomes infected with Hepatitis, his/her liver becomes inflamed. The virus destroys the normal tissue and, at the end, only the fibrous worn-out tissues will remain [3]. More than 40% of the human population has been infected with the Hepatitis B virus (HBV) worldwide, giving rise to 240 million chronic HBV carriers and ca. 620,000 HBV-associated deaths annually [4].

Since the detection of Hepatitis B virus as the cause of Hepatitis, many high-sensitive measurement methods have been developed [5]. The increasing progress of medical sciences has made it difficult to make clear diagnosis decisions about the disease, especially given the inefficacy of the conventional methods and techniques in medical diagnosis. The emergence of PC-based analyzers and the relevant software can increase the accuracy of medical diagnoses, so that they are now convenient and reliable tools in making an accurate diagnosis on the diseases.

Some of these techniques include statistical methods, machine learning, intelligent systems, data analysis and data mining [6]. To resolve medical diagnosis problems, classifier systems have been proposed and used. Therefore, a variety of classifying methods were applied to USI database to obtain an accurate and high-speed system, a couple of which will be discussed here. Various intelligent methods have been employed for the diagnosis of Hepatitis. The first was proposed by Eason et al. [1]. They used Linear Detection Analysis (LDA) and ANFTS to detect Hepatitis

*Corresponding author: Department of Computer and Biomedical Engineering, Hakim Sabzevari University, Sabzevar, Iran, Tel: +98571-4003587, Fax: +985714002174, Mobile: +989184151281, Email: HYPERLINK "mailto:Kh.rezaee@hsu.aci.ir" Kh.rezaee@hsu.ac.ir

by drawing on the properties and categories of data with 94.16% accuracy. In 2009, Rouhani et al. [7], by using Support Vector Machine (SVM) and neural network (NN), proposed a method for the diagnosis of Hepatitis. In addition to the diagnosis of the disease, this system detected the type and stage of the development of Hepatitis with 97% accuracy. Dogantekin et al. [8] suggested a counseling system called DANFIS for quick diagnosis of Hepatitis; the disease information were organized in a tree structure divided into two branches, namely positive or negative Hepatitis, which could help in diagnosis of the disease. Jilani et al. [9] proposed a system based on NN and PCA for the detection of Hepatitis, which comprised of two stages: first, qualities were extracted using PCA techniques; then the categorization was made by using NN. The accuracy of this system was 99.13% and 100% for training and testing the data, respectively. Uttreshwar et al. [10] employed an automatic system for Hepatitis B diagnosis using Fuzzy deduction and generalized regression NN. It comprised of disease detection using an intelligent system along with the prediction of its development stage by NN. Also various kinds of NNs with different specifications have been used for the diagnosis of diseases such as Hepatitis [11].

A huge burden of research has been done through NNs and fuzzy system for diagnosis of Hepatitis B [12, 13]. Furthermore, the authors [14] present a hybrid method based on SVM combined with Simulated Annealing (SA) that is one of the optimization methods in evolutionary algorithms for Hepatitis diagnosis. The method uses the same datasets, used by the previous studies [15, 16]. By using multi-layer NN with Levenberg marquardt training algorithm [17], the approximate accuracy of 92% was obtained. Modeling Hepatitis C with different types of virus Genome was proposed by Moneim et al. [18] using mathematical patterns, which was a different approach to detect Hepatitis C. Kedziora et al. [19] demonstrated that Phylogenetic trees and Hamming distances best reflect the differences between HCV populations present in the organisms of patients responding positively and negatively to the applied therapy [9].

Methods

A typical pattern recognition system includes four parts: feature extraction, feature selection, design and training of classifier, and ultimately, the results. To convert or modify data, one method or technique out of several available options (depending on the type of data) should be selected. In this study, normalization procedure has been chosen. SVM classifier and FCM clustering were, respectively, used to detect Hepatitis B, and to measure its severity. The characteristics of the database compiled from Vasei Hospital (Sabzevar, Iran) and the Medical Sciences Clinics of Sabzavar were used.

The data used in this study included the laboratory test results, which were collected from Sabzevar University Clinic. The data consisted of 350 samples, of which 114 cases had Hepatitis B, including 93 acute and 21 chronic cases, and 236 were healthy subjects. Ten features (inputs) were used to determine the type of Hepatitis and four to detect the stage of Hepatitis B. Table 1 displays 14 features of the database for five subjects.

Proposed Algorithm

The proposed system (algorithm) was organized as shown in Fig. 1. It is an acknowledged fact that database normalization generally yields better results. After normalizing the data in the range [0-1], classification and clustering methods are applied.

SVM Classifier Training

Support Vector Machine (SVM) is a classification and regression device, which helps maximizing the accuracy of prediction using learning theory machine. This technique is a supervised learning algorithm that has a variety of functions, including segmentation and classification [20]. In two-dimensional mode, this system can be used with a separator line. SVM is a discriminative classifier, which is formally defined by a separating hyper plane [21]. A SVM algorithm is one of the pattern recognition algorithms, which is used when the pattern recognition and object classification are needed in particular classes. Matrix model and kernel function are required for implantation of this method. By selecting kernel function parameters and C value, the α_i values of learning algorithm were used by nonlinear Programming Solver (PS). The new data can be classified using α_i values and support vectors. Fig. 2 displays an accurate picture of the formation of super surface by SVM. In this figure, the convex layer has been drawn around the points of class -1 and class +1. Line P shows the nearest distance between two convex layers. First, a convex layer around the points of each class is assumed. Moreover, h is both super surface separator and the perpendicular line that cuts P in half. B is width from the origin of super surface with the maximum separation boundary. If b is ignored, the answers will only include super surfaces that pass through the origin. The vertical distance of super surface from the origin is achieved

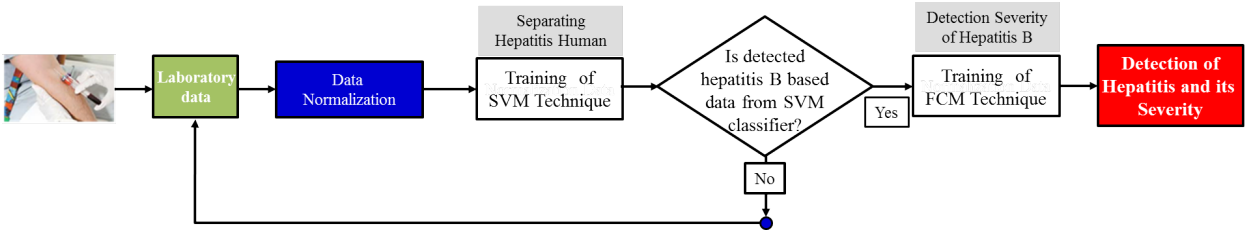

Figure 1 Implementation stages of the proposed algorithm for detection of Hepatitis

by dividing the absolute value of the parameter b by w. The main idea is to select a suitable separator, which has the maximum distance from the neighboring points in both classes. This answer shares the largest boundary with the points related to two different classes, and can be bounded with two parallel super surfaces, passing through at least one of the two points.

These vectors are called “support vectors”. Mathematical formulas of these two parallel super surfaces, which constitute the separating boundary, have been shown in Eqs. (1) and (2) indicating the first super surface and second super surface, respectively:

$$wx - b = 1 \quad (1)$$

$$wx - b = -1 \quad (2)$$

Another important point is that if the training data is in form of linearly resolution, the two border super surfaces can be chosen with no data between them. Then it maximizes the distance between these two parallel super surfaces. By applying geometry theorems, the distance between these two super surfaces gets equal to lwl/2. Thus, the lwl should be minimized. Moreover, data points should not be situated on the inner boundary area. For this, a mathematical limit was added to the formal definition. As can be seen, there are multiple lines providing a solution for the problem in Fig. 3.

A line would be unsuitable if it passes too close to the points, because it will be noise sensitive and will not be generalized correctly. Therefore, our aim should be to discover the line passing as far possible as from all the points. Hence, the function of SVM algorithm is based on discovering the hyper plane that gives the largest minimum distance to the training of Hepatitic and non-Hepatitic cases. The SVM theory is formed based on this distance. We can maximize the margin of the training data of Hepatitic and non-Hepatitic cases by optimal

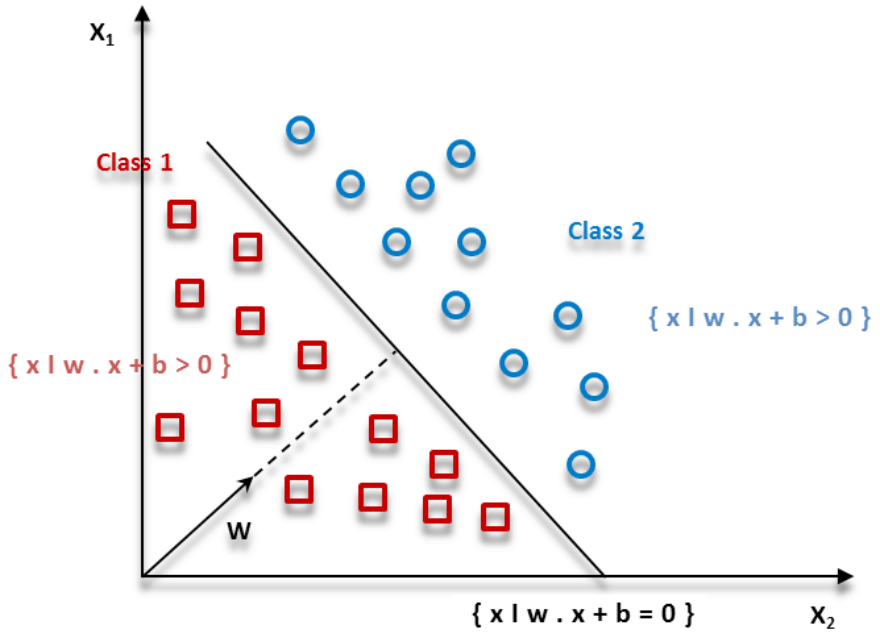

Figure 2 The formation of super surface separating two data classes in two-dimensional space

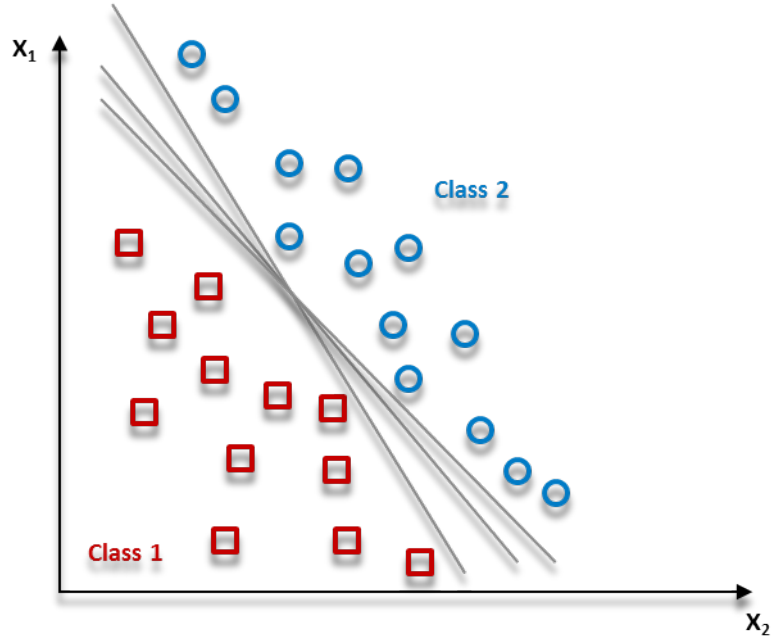

Figure 3 Multiple lines offering a solution for the problem

separating of the hyper plane. For two classes of data such as Hepatitis, which are shown in Fig. 3, we use SVM algorithm, and consequently, the optimal hyper plane will be detected as in Fig. 4.

The optimal hyper plane is introduced based on Eq. (3), which is used to define formally a hyper plane [22]:

$$f(x) = \beta_0 + \beta^T x \quad (3)$$

In this equation, β is s the weight vector, β0 is the bias and x predicates the training cases closest to the hyper plane. The optimal hyper plane can be represented in an infinite number of different procedures

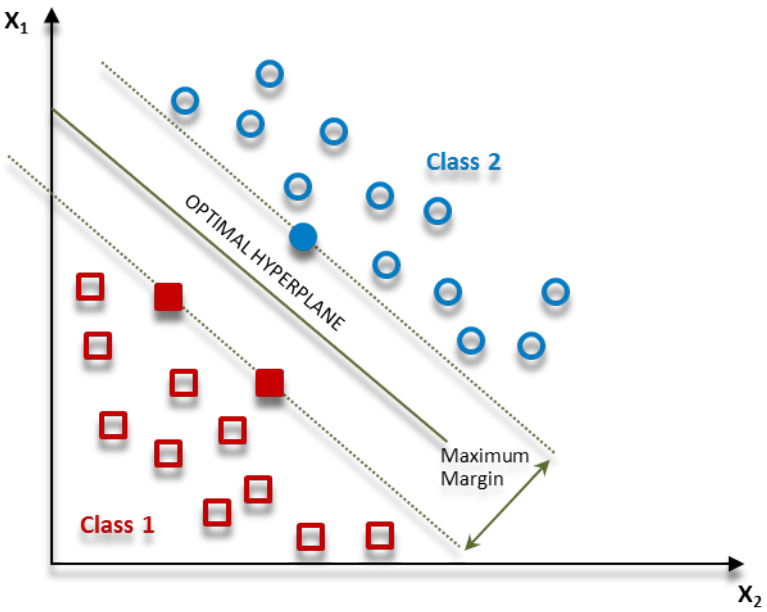

Figure 4 Applying the SVM technique into two classes of data such as Hepatitis and non-Hepatitis, shown in Figure 3, to find optimal hyper plane

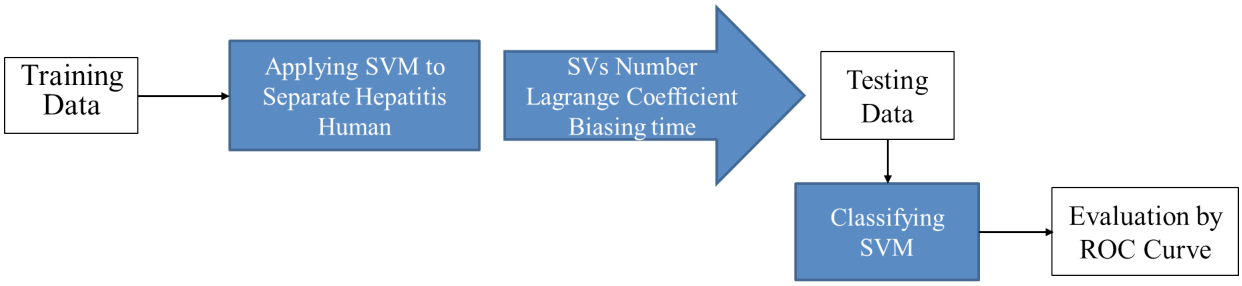

Figure 5 Implementation of SVM algorithm to separate patients from healthy subjects

by scaling of β and β0. One best way faced with the constraint is defined as (4):

$$\left|\beta_0 + \beta^T x\right| = 1 \quad (4)$$

In general, the training cases closest to the hyper plane are called "support vectors". This is known as the canonical hyper plane. We use the result of geometry that gives the distance between a point x and a hyper plane (β, β0):

$$Dis = \frac{\left|\beta_0 + \beta^T x\right|}{\|\beta\|} \quad (5)$$

For the canonical hyper plane form, the numerator is equal to one, and the distance to the support vectors is as

$$Dis_{SupportVector} = \frac{\left|\beta_0 + \beta^T x\right|}{\|\beta\|} = \frac{1}{\|\beta\|}$$

The margin, denoted as M, is twice the distance to the closest cases equal to

$$M = \frac{2}{\|\beta\|}$$

Finally, the problem of maximizing M is equivalent to the problem of minimizing a function L (β) subject to some constraints.

The constraints model is a requirement for the hyper plane to classify correctly all of the training cases xi. Formally:

$$\min_{\beta,\beta_0} L(\beta) = \frac{1}{2}\|\beta\|^2 \ \ subject\ to\ y_i(\beta_0 + \beta^T x_i) \geq 1 \forall i \quad (6)$$

Where, yi indicates each of the labels of the training cases. As a special feature, SVMs simultaneously minimize the empirical classification error and maximize the geometric margin. The effectiveness of SVM

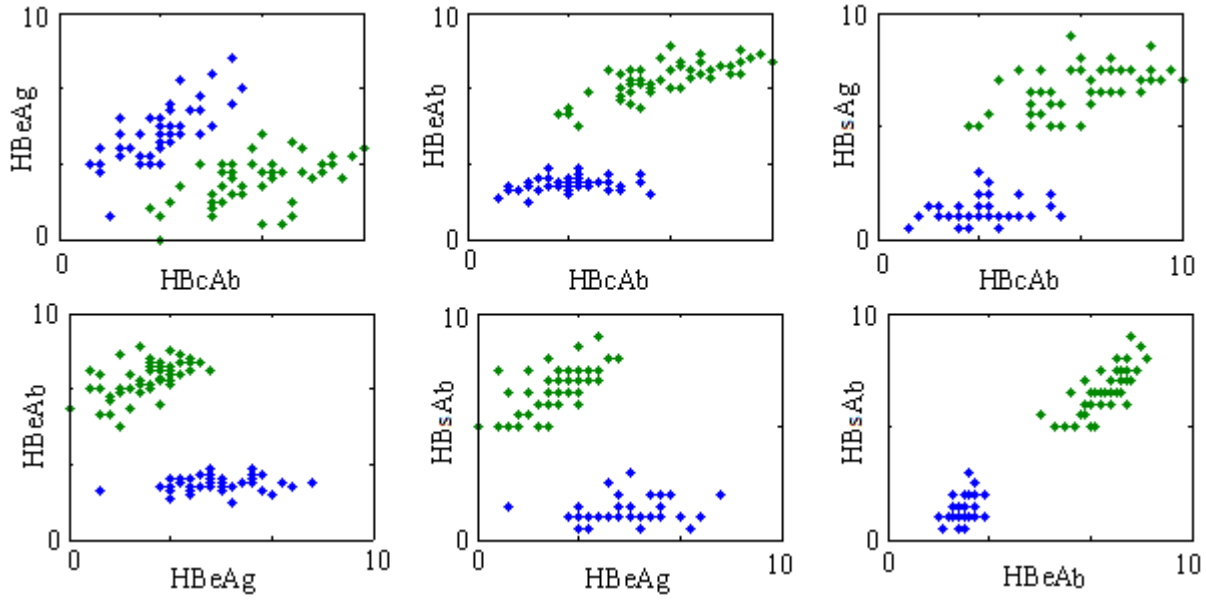

Figure 6 The ratio of features to each other for proper clustering

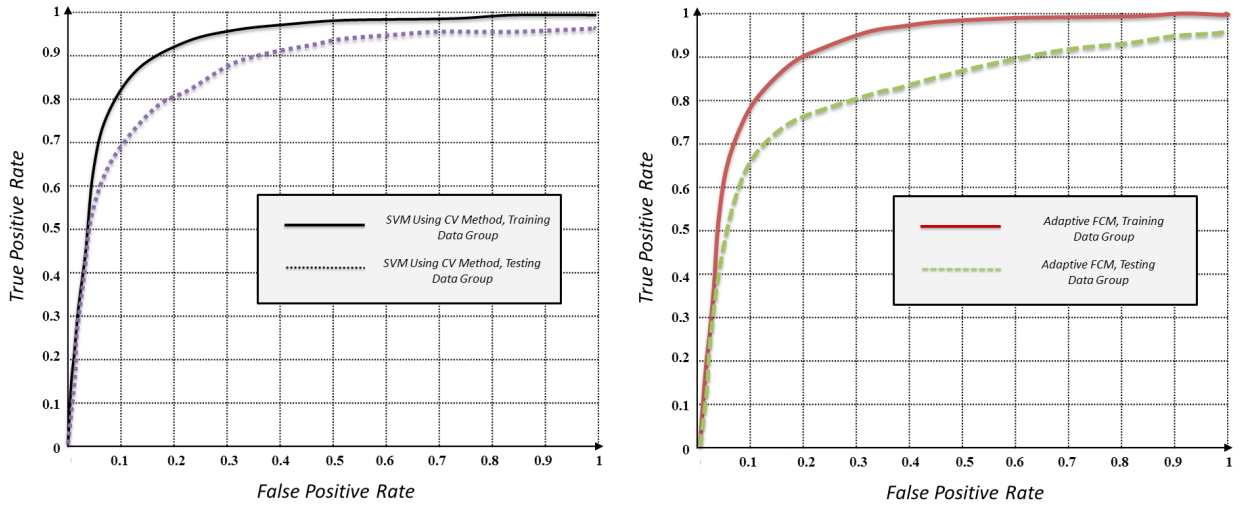

Figure 7 ROC curve for the SVM and FCM techniques in the test and training data

depends on the selection of kernel, the kernel parameters, and soft margin parameter C. The Gaussian kernel is a common choice, which has a single parameter γ. The best combination of C and γ is often selected by a grid search with exponentially growingsequences of C and γ, for example, $C \in \{2^{-5}, 2^{-3}, ..., 2^{13}, 2^{15}\}$ and $\gamma \in \{2^{-15}, 2^{-13}, ..., 2^{1}, 2^{3}\}$

The final model used for testing and classifying new data is then trained on the whole training set using the selected parameters to find Hepatitic and non-Hepatitic cases [23]. To simulate SVM in this study, MATLAB software (Version 7.14) was used. Given the above point and the fact that the classification is done for two classes, i.e. Hepatitis B and healthy subjects, all data were put in a two-dimensional array. Then, as shown in Fig. 5, the data separates the Hepatitis B patients and healthy subjects using SVM. After determining the data related to the individual patients, these data are again positioned in a two-dimensional array.

Adaptive FCM Clustering

Clustering is one of the most important training methods without supervision. A cluster is a series of corresponding data. In clustering, the aim is to divide the data into clusters with data in each cluster having the maximum resemblance with each other, and minimum similarity with the data in other clusters. FCM clustering method is very useful in various clustering-related issues. Considering X as the sample, the aim of FCM is to minimize the Cost function (7):

$$J = \sum_{j=1}^{N} \sum_{i=1}^{c} U_{ij}^{m} \left\| X_j - V_i \right\|^2 \quad (7)$$

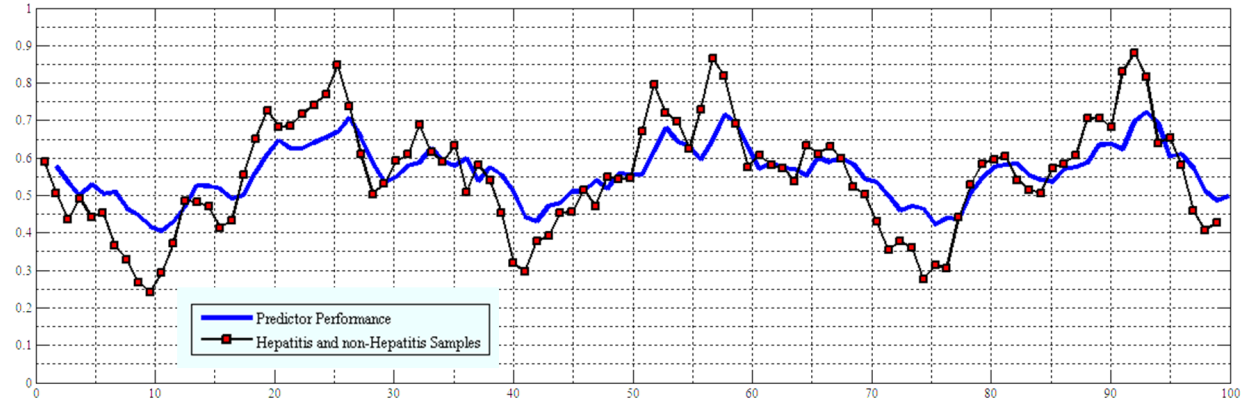

Figure 8 There is little difference between the predicted outputs of the algorithm and the physician's diagnosis

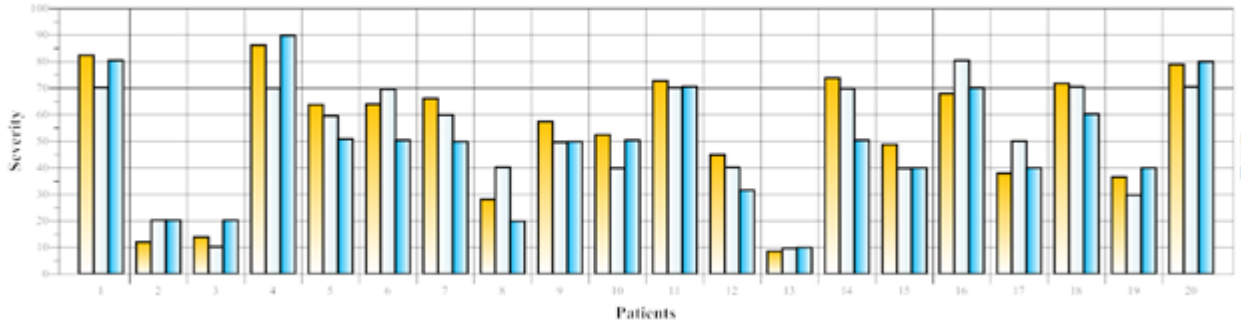

Figure 9 Statistical comparison of physical opinions about the severity of the hepatitis disease for 20 patients; yellow: detection of hepatitis severity by physician 1, detection of hepatitis severity by physician 2

Where, c is the number of clusters, m is the degree of fuzziness, and V is the initial center for each cluster. Also, the degree of allocation of each data to the desired cluster according to (8) is calculated through the membership function as follows:

$$U_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{\|X_j - V_i\|}{\|X_j - V_k\|}\right)^{2/(m-1)}} \quad (8)$$

The value of new centers for each cluster according to (9) is calculated as below:

$$V_i = \frac{\sum_{j=1}^{N} U_{ij}^m X_j}{\sum_{j=1}^{N} U_{ij}^m} \quad (9)$$

These steps are repeated until the difference between Us (the degree of data allocation to each cluster) in the new stage and Us in the previous step level is less than a threshold value. In traditional FCM, the results are not close to the reality, when different vectors not the same contribution to the cluster.

To improve the performance of FCM, initial cluster centers are selected in another way. The cluster sample of FCM algorithm [24] [25] is very sensitive to the number of cluster centers; the cluster centers' initialization often artificially get significant errors, and even may get the actual opposite results. The detailed procedure of this algorithm is shown in Table 2.

Results

To test the diagnosis of Hepatitis B in 350 samples collected from the database of Sabzevar, cross validation was used in which 10% of the samples were test data, and the rest were training data. This procedure was recycled until all data were tested. SVM or RBF kernel (with radius equal to 1, error penalty for class 1 equal to C = 50 and for class -1 equal to C = 10) was used. Overall, the proposed algorithm has the following properties:

- *Increased accuracy of the performance*
- *Negligible average error*
- *Reduced user intervention*
- *Suitable stimulator for Hepatitis disease*

To test the severity of Hepatitis B, 114 samples from the database of Hepatitis B patients were selected. To determine the stage of disease in Hepatitis B patients, adaptive FCM clustering was used. This sample included four features, including HBsAb, HBeAg, HBeAb, and HBsAb. Fig. 6 represents the ratio of features in true clustering. In this figure, the ratio of features has been calculated to determine which features are in the clustering with higher precision. Fig. 7 presents ROC curves for the proposed system in SVM technique in training and testing conditions. The ratio of the first and second features in true clustering is not very significant, and the data sets are relatively similar. In other features, however, including first to third, first to fourth, second to third, second to fourth, and third to fourth, the data have been classified into two clusters with high accuracy. As shown in the above figure, the ratio of the second and fourth features in the true clustering of data is not so effective, because the data sets are very similar. However, in other cases, including first to third, first to fourth, second to third, second to fourth and third to fourth, data have been classified into two clusters with high accuracy.

Sensitivity and specificity of the SVM technique according to CV were at an accepted level, because the system was oriented upward in the bisector of the first quadrant, and thus, proper output of the system could be trusted. The performance of the algorithm in identifying the patients from others for 100 random samples of the data is shown in Fig. 8. Characteristics of the diseases were normalized in the interval [0-1], and finally, the av-

erage of these random samples was calculated. The difference between the actual outputs of the algorithm is minimal in comparison with a state of the disease state. There is little difference between the predicted outputs of the algorithm and the physician's diagnosis in the diagnosis of Hepatitis B and its severity.

Performance Measurement Criteria

Four criteria are used for determining the performance of Hepatitis diseases detector:

Balanced mean of accuracy and sensitivity (F-Measure)*: According to Equation (7), this means the measurement of system performance using accuracy and sensitivity:*

$$FMeasure = \frac{2N_{TP}}{2N_{TP} + N_{FP} + N_{FN}} \quad (7)$$

Accuracy (AC): *According to equation (8), this means summing up of true positive and negative rates of the received samples during the system test:*

$$Ac = \left(\frac{N_{TP} + N_{TN}}{N_{TP} + N_{TN} + N_{FP} + N_{FN}}\right) \quad (8)$$

Confidence coefficient of performance (Kappa): *Kappa coefficient indicates confidence in the performance of a system):*

$$Kappa = \frac{2(N_{TP}N_{TN} + N_{FN}N_{FP})}{(N_{TP} + N_{FN})(N_{TN} + N_{FN}) + (N_{TN} + N_{FP})(N_{TP} + N_{FP})} \quad (9)$$

Area under the ROC curve (Az): *This factor also indicates the performance of system, and its high rate shows the proportion of sensitivity and specificity, and consequently high relative accuracy. This area can be calculated for the patients using point sensitivity and specificity:*

$$A_z = \frac{1}{N_{TP}.N_{TN}} \sum_{v=1}^{N_{TP}} \sum_{nv=1}^{N_{TN}} I(N_{TP_v}, N_{TN_{nv}}) \quad (10)$$

In these equations, NTP is the number of retina images in which there is no pathological tissue, and the software has correctly diagnosed the retina tissue as healthy. NFN is the number of retina images in which there is no pathological tissue but the software has not correctly diagnosed the pathological tissue of retina. NTN is the number of retina images in which there is pathological tissue, and the software has correctly diagnosed the pathological tissue of retina. NFP is the number of retina images in which there is no pathological tissue; however, the software has correctly diagnosed the tissue of retina as unhealthy. Finally, TP and TN are the ordered lists of unique vessel and non-vessel pixels, respectively, and I is the function, in which:

$$I(x, y) = \begin{cases} 1 & x > y \\ 0.5 & x = y \\ 0 & otherwise \end{cases} \quad (11)$$

This holds true whether the pairs represent absolute pixel counts (TPv, TNnv) or their relative fractions (TPF and TNF) as they have been plotted. In Table 3, the accuracy and confusion matrix for training and testing the SVM and FCM techniques have been calculated. To test the detection of Hepatitis severity in 114 samples, cross validation method was used in which 10% of samples were test data, and the rest were training data; this procedure was recycled until all data were tested. The final factors of accuracy, sensitivity and specificity were calculated after classifying and clustering as 98.36%, 98%.44 and 94.09%, respectively.

Discussion

To evaluate the proposed algorithm with the results of previous studies, a comparison of system performance in terms of accuracy has been shown in Table 4. It is to be noted that the new database is different from that used in previous studies in terms of data number. Moreover, in the new database, some clinical symptoms have been removed due to their insignificant effect. The comparison of the proposed system with the previous ones involves only the detection of Hepatitis, and there is no precedent for detecting the severity of Hepatitis. In comparison with valid techniques, the performance of our proposed algorithm is acceptable. However, Kappa coefficient is not high, but low PPV and high NPV of the system are a warranty of the system, and both the clinical specialist and the patients can trust the software and its output. Altogether, the ROC of SVM classifier and FCM clustering method show the capability of system to detect Hepatitis B and measure its severity.

Two physicians commented about the severity of Hepatitis illness of the patients based on four clinical main factors, and assessed the stage of disease progression. They were unable to determine the severity of the disease with high accuracy, because there were four factors indicating almost continuous changes. So, we asked them to indicate the severity of Hepatitis using the numbers 1 to 10. Furthermore, the output of the algorithm is much more accurate in comparison with the physician's opinions. The output of the algorithm determined the severity of Hepatitis using a numerical distance in [0-100]. In Fig. 9, a statistical comparison has been made with regard to the physician's opinions about the severity of the Hepatitis disease for 20 patients and healthy individuals.

Conclusions

An algorithm for detecting the type and severity of Hepatitis was proposed. In this algorithm, first, the samples collected from database were normalized at [0-1] range, where classifier and clustering were trained based on the data subsets. In SVM technique, the accuracy was in the range of 90%-99% for the trained data, and 89%-97% for tested data. After identifying Hepatitis B, FCM algorithm was implemented on the Hepatitis B sample to determine the severity of the disease. By assigning the degree of allocation to each sample, this algorithm created two clusters; one related to acute cases and the other to chronic ones. The current system increases speed and accuracy of diagnosis, and offers reliable results based on software analysis.

Abbreviations

Competing Interests

The authors declare no competing interests.

Authors' Contributions

Acknowledgements

References

1. Eason G, Noble B, Sneddon IN. On certain integrals of Lipschitz-Hankel type involving products of bessel functions. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 1955, 247(935):529-51.
2. Shariati S, Haghighi MM. Comparison of anfis neural network with several other ANNs and Support Vector Machine for diagnosing hepatitis and thyroid diseases. In: *Computer Information Systems and Industrial Management Applications (CISIM), 2010 International Conference on: 8-10 Oct. 2010 2010*; 2010:596-99.
3. Guan P, Huang DS, Zhou BS. Forecasting model for the incidence of hepatitis A based on artificial neural network. *World J Gastroenterol* 2004, 10(24):3579-82.
4. Ott JJ, Stevens GA, Groeger J, Wiersma ST. Global epidemiology of hepatitis B virus infection: New estimates of age-specific HBsAg seroprevalence and endemicity. *Vaccine* 2012, 30(12):2212-19.
5. Shibata H. The history of hepatitis B virus-related determination tests and inspection and the measurements of problems in Japan. *Rinsho Byori* 2013, 61(9):846-51.
6. Polat K, Güneş S. Hepatitis disease diagnosis using a new hybrid system based on feature selection (FS) and artificial immune recognition system with fuzzy resource allocation. *Dig Sig Process* 2006, 16(6):889-901.
7. Rouhani M, Haghighi MM. The Diagnosis of hepatitis diseases by support vector machines and Artificial neural networks. In: *Computer Science and Information Technology - Spring Conference, 2009 IACSITSC '09 International Association of: 17-20 April 2009 2009*; 2009:456-58.
8. Dogantekin E, Dogantekin A, Avci D. Automatic hepatitis diagnosis system based on Linear Discriminant Analysis and Adaptive Network based on Fuzzy Inference System. *Expert Sys Appl* 2009, 36(8):11282-86.
9. Jilani TA, Yasin H, Yasin MM. PCA-ANN for classification of Hepatitis-C patients. *Int J Comput Appli (0975–8887)* 2011, 14(7).
10. Uttreshwar GS, Ghatol AA. Hepatitis B diagnosis using logical inference and generalized regression neural networks. *Advance Computing Conference, 2009 IACC 2009 IEEE International: 6-7 March 2009 2009*; 2009:1587-95.
11. Sheikhan M, Ghoreishi SA. Application of covariance matrix adaptation–evolution strategy to optimal control of hepatitis B infection. *Neural Comput Appli* 2013, 23(3-4):881-94.
12. Madhusudhanan S, Karnan M, Rajivgandhi K. Fuzzy based ant miner algorithm in datamining for Hepatitis. In: *Signal Acquisition and Processing, 2010 ICSAP '10 International Conference on: 9-10 Feb. 2010 2010*; 2010:229-32.
13. Neshat M, Yaghobi M. Designing a fuzzy expert system of diagnosing the hepatitis B intensity rate and comparing it with adaptive neural network fuzzy system. *Proceedings of the World Congress on Engineering and Computer Science* 2009, 2:797-802.
14. Sartakhti JS, Zangooei MH, Mozafari K. Hepatitis disease diagnosis using a novel hybrid method based on support vector machine and simulated annealing (SVM-SA). *Comput Meth Prog Biomed* 2012, 108(2):570-9.
15. Çalişir D, Dogantekin E. A new intelligent hepatitis diagnosis system: PCA–LSSVM. *Expert Sys Appli* 2011, 38(8):10705-08.
16. Chen H-L, Liu D-Y, Yang B, Liu J, Wang G. A new hybrid method based on local fisher discriminant analysis and support vector machines for hepatitis disease diagnosis. *Expert Sys Appli* 2011, 38(9):11796-803.
17. Bascil MS, Temurtas F. A study on hepatitis disease diagnosis using multilayer neural network with levenberg marquardt training algorithm. *J Med Syst* 2011, 35(3):433-6.
18. Moneim IA, Mosa GA. Modelling the hepatitis C with different types of virus genome. *Comput Math Methods Med* 2006, 7(1):3-13.
19. Kedziora P, Figlerowicz M, Formanowicz P, Alejska M, Jackowiak P, Malinowska N, Fratczak A, Blazewicz J, Figlerowicz M. Computational methods in diagnostics of chronic hepatitis C. *Bulletin of the Polish Academy of Sciences, Technical Sciences* 2005, 53(3).

20. Cristianini N, Shawe-Taylor J. *Support Vector Machines*: Cambridge: Cambridge University Press; 2000.
21. Shin K-S, Lee TS, Kim H-j. An application of support vector machines in bankruptcy prediction model. *Expert Sys Appli* 2005, 28(1):127-35.
22. Hastie TJ, Tibshirani RJ, Friedman JH. The elements of statistical learning: data mining, inference, and prediction: Springer; 2009.
23. Hsu C-W, Chang C-C, Lin C. A practical guide to support vector classification technical report department of computer science and information engineering. *National Taiwan University, Taipei* 2003.
24. Suresh K, Madana Mohana R, Rama Mohan Reddy A, Subramanyam A. Improved FCM algorithm for Clustering on Web Usage Mining. In: *Computer and Management (CAMAN), 2011 International Conference on: 2011*. IEEE; 2011:1-4.
25. Esmaili Torshabi A, Riboldi M, Imani Fooladi AA, Modarres Mosalla SM, Baroni G. An adaptive fuzzy prediction model for real time tumor tracking in radiotherapy via external surrogates. *J Appl Clin Med Phys* 2013, 14(1):4008.
26. Polat K, Güneş S. Prediction of hepatitis disease based on principal component analysis and artificial immune recognition system. *Appli Math Comput* 2007, 189(2):1282-91.
27. Neshat M, Sargolzaei M, Nadjaran Toosi A, Masoumi A. Hepatitis disease diagnosis using hybrid case based reasoning and particle Swarm optimization. *ISRN Artificial Intelligence* 2012, 2012:6.

Please cite this article as:
Khosro Rezaee, Javad Haddadnia, Mohammad Rasegh Ghezelbash. A Novel Algorithm for Accurate Diagnosis of Hepatitis B and Its Severity. *International Journal of Hospital Research* 2014, **3**(1):1-10.