Journal of Infection and Public Health 12 (2019) 13–20

Contents lists available at ScienceDirect

Journal of Infection and Public Health

journal homepage: http://www.elsevier.com/locate/jiph

A predictive method for hepatitis disease diagnosis using ensembles of neuro-fuzzy technique

Mehrbakhsh Nilashi[a,*], Hossein Ahmadi[b], Leila Shahmoradi[c,d], Othman Ibrahim[a], Elnaz Akbari[e,f,**]

[a] School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia
[b] Department of Health Information Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran
[c] Halal Research Center of IRI, FDA, Tehran, Iran
[d] Health Information Management Department, School of Allied Medical Sciences, Tehran University of Medical Sciences (TUMS), Iran
[e] Department for Management of Science and Technology Development, Ton Duc Thang University, Ho Chi Minh City, Vietnam
[f] Faculty of Electrical & Electronics Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam

ARTICLE INFO

Article history:
Received 15 August 2017
Received in revised form 14 August 2018
Accepted 20 September 2018

Keywords:
Viral hepatitis
Public health
NIPALS
SOM
ANFIS ensemble
Decision trees

ABSTRACT

Background: Hepatitis is an inflammation of the liver, most commonly caused by a viral infection. Supervised data mining techniques have been successful in hepatitis disease diagnosis through a set of datasets. Many methods have been developed by the aids of data mining techniques for hepatitis disease diagnosis. The majority of these methods are developed by single learning techniques. In addition, these methods do not support the ensemble learning of the data. Combining the outputs of several predictors can result in improved accuracy in classification problems. This study aims to propose an accurate method for the hepatitis disease diagnosis by taking the advantages of ensemble learning.
Methods: We use Non-linear Iterative Partial Least Squares to perform the data dimensionality reduction, Self-Organizing Map technique for clustering task and ensembles of Neuro-Fuzzy Inference System for predicting the hepatitis disease. We also use decision trees for the selection of most important features in the experimental dataset. We test our method on a real-world dataset and present our results in comparison with the latest results of previous studies.
Results: The results of our analyses on the dataset demonstrated that our method performance is superior to the Neural Network, ANFIS, K-Nearest Neighbors and Support Vector Machine.
Conclusions: The method has potential to be used as an intelligent learning system for hepatitis disease diagnosis in the healthcare.

Introduction

Viral hepatitis prevails all over the world and has been on of is a key global public health issue [1,2]. Hepatitis is an inflammation of the liver, most commonly caused by a viral infection which has caused yearly an estimated 1.5 million deaths worldwide [3]. Viral liver diseases are among the most important communicable disease worldwide different species of viruses (e.g. Epstein–Barr, Cytomegalovirus, Herpes simplex, Coxsackie virus, Adenovirus, Mumps, Yellow fever). However, the term viral hepatitis generally implies the five hepatotropic viruses: Hepatitis A [4], B [5,6], C [7], D [8] and E virus [9]. Hepatitis B virus (HBV) and hepatitis C virus (HCV) have been the most common viral causes of hepatic diseases universally [10] which their infections initially begin from acute infection and progress to chronic infection. World Health Organization (WHO) [13] showed that about 130–150 million people globally are chronically infected with hepatitis C infection [11]. Tatoos and piercing, drug abuse, sexual contact with hepatitis carrier, hemodyalisis, blood transfusions and health workers have been the main risk factors of the hepatitis [12]. The hepatitis disease diagnosis is mainly done by a routine blood testing. Medical diagnostics of hepatitis is quite difficult or a physician as many factors should be considered in the disease diagnosis procedure. Hence, the development of the automatic and accurate diagnosis systems can be helpful for the hepatitis detection and accordingly for decision-making by a physician.

* Corresponding author.
** Corresponding author at: Department for Management of Science and Technology Development, Ton Duc Thang University, Ho Chi Minh City, Vietnam.
E-mail addresses: nilashidotnet@hotmail.com, nilashidotnet@yahoo.com (M. Nilashi), elnaz.akbari@tdtu.edu.vn (E. Akbari).

https://doi.org/10.1016/j.jiph.2018.09.009

Table 1
Related work on hepatitis disease diagnosis.

Disease	Author	Techniques														
		SVM	SA	PCA	RS	C 4.5	FL	ARIS	GA	RF	ELM	EI	FLD	ABC	EA	SC
Hepatitis	Sartakhti et al. [16]	*	*													
	Çalişir and Dogantekin [17]	*		*												
	Chen et al. [18]	*		*												
	Kaya and Uyar [19]				*						*					
	Polat and Güneş [20]					*	*	*								
	Tan et al. [21]								*							
	Zhang and Suganthan [22]									*						
	Nahato et al. [23]						*				*					
	Gorunescu and Belciug [24]											*				
	Ben-Israel and Levin [25]												*			
	Zorarpacı and Özel [26]													*	*	
	Yang et al. [27]															*
	Shah et al. [28]	*		*												

SVM: support vector machine, SA: simulated annealing, RS: rough set, ELM: extreme learning machine, FL: fuzzy logic, ARIS: artificial immune system, GA: genetic algorithm, RF: random forests, EI: evolutionary inspired, FLD: fisher linear discriminant, ABC: artificial bee colony, EA: evolution algorithm, SC: spectral clustering.

As a subset of artificial intelligence, machine learning applies statistical techniques to optimize a performance criterion using example data or past experience. There are mainly two types of these techniques, supervised and un-supervised. Clustering techniques such as Self-Organizing Map (SOM) and Expectation Maximization (EM) have been in the first group and prediction machine learning techniques such as adaptive Neuro-Fuzzy Inference System (ANFIS) and Support Vector Machine (SVM) in the second group. The use machine learning techniques in developing the methods and decision support systems have been important. Due to diseases diagnosis importance to mankind, several studies have been conducted on developing methods for their classification [14,15]. Some of the studies on hepatitis disease diagnosis which have used data mining techniques for the development of the methods are presented in Table 1. From this table, we can see that the majority of the methods have been developed by the no-ensemble data mining techniques. In addition, the most methods developed by supervised classification techniques in the previous researches do not use ensembles of prediction data mining techniques. As the accuracy prediction of standard supervised methods can be improved by ensemble learning techniques, also called committees of predictors, and the use of the ensemble learning for the disease diagnosis task, in this study accordingly a new method is proposed using ensembles of ANFIS to improve the predictive accuracy of the hepatitis disease diagnosis systems. In addition, we use a clustering technique, SOM, for clustering the data and Non-linear Iterative Partial Least Squares (NIPALS), as a dimensionality reduction technique, to reduce the dimensions of the data. Furthermore, we use decision trees (DT) for the selection of the most important features in the dataset. The combination of these techniques for the first time is proposed in this study for hepatitis disease diagnosis. We evaluate the proposed method on a real-world dataset which can be accessed from Data Mining Repository of UCI [15]. Overall, the contributions of this research are as follows:

- A hybrid machine learning approach is proposed for hepatitis disease diagnosis using SOM, NIPALS and ANFIS ensemble.
- SOM is used for the clustering of data in the experimental dataset.
- NIPALS is used for dimensionality reduction and improving the accuracy of clustering.
- CART is used for the selection of the most important features.
- ANFIS ensemble is used for hepatitis disease diagnosis.

Our study at hand is orgnized as follows: we present the method in Section “Methodology”. Section “ANFIS ensemble evaluation” presents ANFIS ensemble results and methods comparision. Finally, conclusions and future work is provided in Section “Conclusion and future work”.

Methodology

This paper proposes a new machine learning approach for hepatitis disease diagnosis using a set of real data. The method is developed using NIPALS to reduce the dimensions of the data, SOM to cluster the data, CART for feature selection and ANFIS ensemble for hepatitis disease diagnosis. In Fig. 1, we present the proposed method. As can be seen form this figure, several machine learning techniques are used for hepatitis disease diagnosis. In the first step, we try to cluster the data using SOM. The aim of the clustering is to improve the readability of the data for classification task. The NIPALS technique is used for data dimensionality reduction to improve the clustering quality. In the next step, decision trees are used for selecting the most important features of the dataset in each cluster. In the last step, we use ANFIS ensemble for hepatitis disease diagnosis from a set of medical data.

The proposed method is evaluated on a real-world dataset. The dataset is obtained from the machine learning repository at UCI. The dataset includes 155 records in two different classes which are die in 32 cases (20.6%) and live in 123 cases (79.4%). The dataset includes 19 attributes (13 binary and 6 attributes with 6–8 discrete values) with many missing values. For method evaluation, we first pre-processed the data. From this analysis, we found that there is a high correlation among the input variables. It has been shown that this correlation can affect on the disease prediction accuracy [29]. Accordingly, in this study we used NIPALS to estimate the missing values in the dataset and overcome the multi-collinearity problem.

NIPALS

Principal component analysis (PCA) has been an effective dimensionality reduction technique applied in many application fields. It is mainly used as a statistical technique for multivariate analysis. In addition, it is used for data compression task to retain the essential information from the data [29]. NIPALS algorithm is a nonlinear iterative PLS for finding the eigenvectors [30], which can effectivley handle the missing values in the dataset [31]. Using NIPALS, the pricipal components for the data are generated. The best pricipal components are then selected to be used in the clustering task.

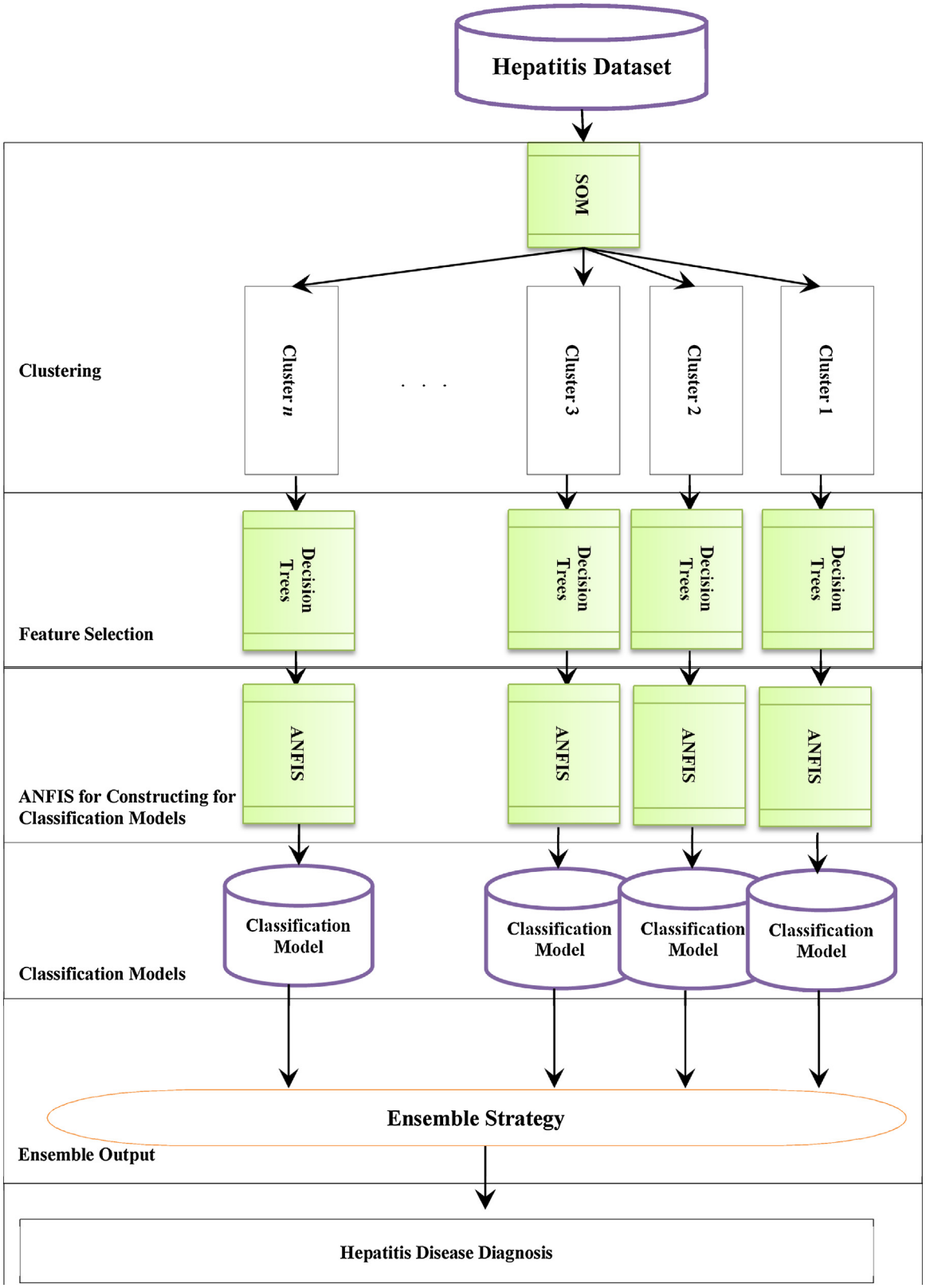

Fig. 1. Research methodology.

SOM

SOM has been one of the most commonly used networks for clustering and classification purposes. It has an effective training procedure to handle large sets of data which can visualize the *n*-dimensional data to lower dimensions. To do so, it uses a grid of neurons in the lower data space. SOM is an unsupervised machine learning techniques for data clustring aim [32]. This clustering technique is based on artificial neural network. By the use of SOM, the data is visualized on a low-dimensional (typically

two-dimensional) space. This technique does not need a class for training and accordingly clustering the data. In this technique to perform the clustering, the selection of maximum epochs, number of neurons and learning rate are important. In addition, the main tasks in this clustering are: selection of random inputs, compute the winner neurons and update neurons. In SOM, for each point on the grid, which is denoted by its coordinate position (x, y), a neuron (M) and a weight vector $w_{x,y}$ are considered. Basically, in SOM clustering, Euclidean distance measure is used to find the Best Matching Unit (BMU) by calculating the distance between each node's weight vector in the data space and the current input vector. Hence, for N dimensions in the data vector, the distance of V (current input vector) and W (the node weight vector) are calculated using Euclidean distance.

In this study, SOM technique for clustering is applied on hepatitis dataset. We selected different clustering size for SOM. The learning rate in SOM was set to be 0.05. For SOM, SOM 2×2 (4 clusters), SOM 2×3 (6 clusters), SOM 3×3 (9 clusters) and SOM 3×4 (12 clusters) were considered. We found that SOM 2×3 generated the best quality for data clustering.

Decision tress

CART is mainly used to model a response variable (y) based on some input variables (x). As a supervised learning and nonparametric statistical technique, CART uses recursive partitioning for regression and classification tasks. The CART is mainly used to find the patterns from large sets of data by inducing decision trees. In fact, the main objective of the CART is to discover the relationship between the explanatory variables (inputs) and response variable (outputs). This technique constructs the prediction models by recursive binary splitting which Gini index, a generalization of the binomial variance, is used as the impurity function. CART tries to initially grow overly large trees from the observations and then prune them to smaller sizes by minimizing the classification errors. In addition, CART technique use a 10-fold cross validation approach in which 9 folds are used for model training and one fold is used for model testing. In CART the regression model is fitted to each node to provide the predicted values of the output variable. In Table 2, a sample of induced decision tree using CART for the experimental dataset is presented.

Ensembles of ANFIS

Fuzzy set theory, introduced by Zadeh [33], is mainly used when the information is incomplete or imprecise. Fuzzy logic provides the degree of membership to the features of the objects. In fuzzy logic, membership functions are used to characterize fuzziness. There are many types of membership functions which some of them are: Triangular, Generalized Bell-Shaped, Gaussian and Π-Shaped. One of the main issues of fuzzy logic is that it is not able to automatically learn the models from the data [34]. In contrast, neural network has solved this issue. However, neural network is not suitable for modeling when the information is incomplete or imprecise. Accordingly, the hybrid of neural network and fuzzy logic, neuro-fuzzy, has solved these issues. One of these techniques is ANFIS [35] (see Fig. 2) which has the ability to automatically learn the model from the imprecise data.

Nowadays, the use of ensemble learning has attracted the interest of the researchers for developing the methods for classification and prediction tasks. In ensemble learning paradigm, a collection of different predictors is constructed whose individual responses are then combined to label test instances. Accordingly, combining the outputs of several predictors results in improved accuracy in many prediction and classification problems. Specifically, the generalization performance of the ensemble learning has been proved to be much better than a single individual ensemble member. In fact, these improvements in performance and accuracy are because of the combination of accurate predictors which their errors are complementary. The final ensemble decision can be reached by combining the individual predictions of the ensemble members. To do so, different approaches, for example, average prediction,

Table 2
A part of induced decision trees from the hepatitis dataset.

Decision Tree A

- ASCITES < 1.5000
  - SPLEEN PALPABLE < 0.5000
    - AGE < 68.5000 then Class = Live
    - AGE >= 68.5000 then Class = Die
  - SPLEEN PALPABLE >= 0.5000
    - PROTIME < 53.5000
      - HISTOLOGY < 1.5000 then Class = Live
      - HISTOLOGY >= 1.5000
        - PROTIME < 26.5000
          - PROTIME < 11.5000
            - BILIRUBIN < 1.1000 then Class = Live
            - BILIRUBIN >= 1.1000 then Class = Die
          - PROTIME >= 11.5000 then Class = Live
        - PROTIME >= 26.5000 then Class = Die
    - PROTIME >= 53.5000
      - SGOT < 50.5000 then Class = Die
      - SGOT >= 50.5000 then Class = Live

Decision Trees B

- SPIDERS >= 1.5000
  - BILIRUBIN < 5.6000
    - STEROID < 1.5000
      - SGOT < 47.5000 then Class = Live
      - SGOT >= 47.5000
        - LIVER BIG < 0.5000 then Class = Die
        - LIVER BIG >= 0.5000
          - LIVER BIG < 1.5000 then Class = Live
          - LIVER BIG >= 1.5000
            - VARICES < 1.5000 then Class = Die
            - VARICES >= 1.5000
              - ALK PHOSPHATE < 205.5000 then Class = Live
              - ALK PHOSPHATE >= 205.5000 then Class = Die
    - STEROID >= 1.5000 then Class = Live
  - BILIRUBIN >= 5.6000 then Class = Die

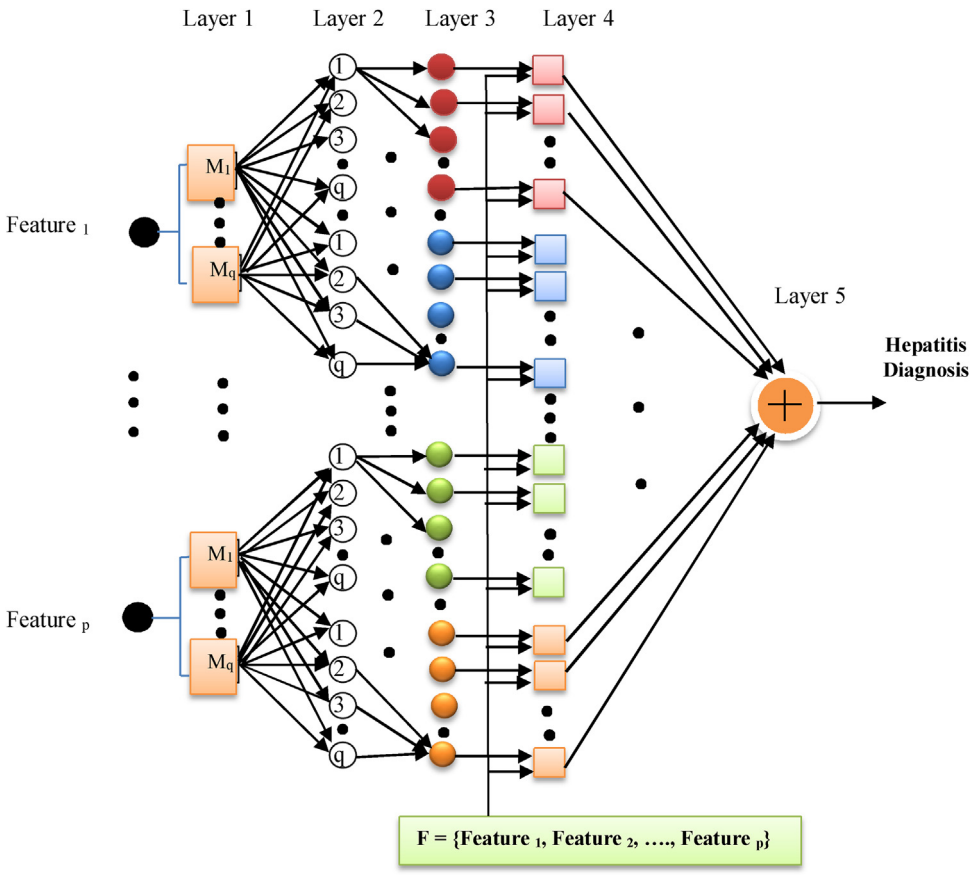

Fig. 2. Hepatitis diagnosis by ANFIS model.

majority voting and non-linear decision functions can be used. It should be noted that in general there is no strong evidence supporting that more complex combination approaches can perform better performance.

In this work, we have considered four types of MF which are Triangular, Generalized Bell-Shaped, Gaussian and Π-Shaped. In this study we use Centroid of Area (COA) for defuzzification task. The method centroid, COA, has been one of the most widely used defuzzification strategy. This method calculates the center of gravity of the fuzzy set A, which correspond to the center of the area under the membership function $\mu_A(z)$. The crisp value of the fuzzy set A using this technique (z_{COA}) is calculated as:

$$z_{COA} = \frac{\int_z \mu_A(Z)\, z dz}{\int_z \mu_A(Z)\, dz}. \quad (1)$$

ANFIS ensemble

It has been shown that solely using a single inductive learning method cannot achieve a generalization performance in all possible classification tasks [36]. In additions, the accuracy of using supervised learning methods more depends on the nature of the problem investigated and observed data. Despite this consideration, ensemble learning has proved to have an excellent performance in numerous learning tasks of practical applications [37]. In this research, multiple components of ANFIS learner are trained for hepatitis disease diagnosis. Accordingly, for each ANFIS model we calculated the classification accuracy, for output and then integration by average approach was applied for the final ensemble decision [38]. As a straightforward approach, the final decision by integration by average approach is obtained from the sum of results generated by each ANFIS model divided by the number of ANFIS models used for prediction task.

ANFIS ensemble evaluation

This study aimed to provide a method for hepatitis diagnosis using a set of input parameters. Accordingly, we tried to discover the relationships between the input parameters and the output to classify the hepatitis disease. All classification models were constructed by the ANFIS technique. The discovered fuzzy rules from the experimental dataset assisted the models to generalize the following relationship: $Y=f(X_1, X_2, \ldots, X_n)$) to accuratly classify the disease. For hepatitis dataset, input parameters are (X): "Age", "Sex", "Steroid, 'Antivirals', "Fatigue", "Malaise", "Anorexia", "Liver Big", "Liver Firm", "Spleen Palpable", "Spiders Ascites", "Arices", "Bilirubin", "Alk 'Phosphate', "Sgot", "Albumin", "Protime and Histology", and output parameter (Y) is the class of disease.

To construct the classification models and evaluate them, two sets of the data were used, training, test and checking sets. The training set was used to construct the models, test and checking sets were used for models verification. Specifically, we considered 20% of data in each cluster for test set, 20% of data for checking set and 60% for training set. To evaluate the prediction models of ANFIS technique, we used Receiver Operating Characteristic (ROC) chart.

In this research totally six ANFIS ensemble prediction models were developed as 6 clusters have been generated by SOM. In each cluster, different types of MFs were designed and considered for fuzzification task. The types of MFs were considered as: Triangular, Generalized Bell-Shaped, Gaussian and Π-Shaped. In addition,

there linguistic variables (Low, Moderate and High) are used as MFs degree for each feature. Hence, we totally developed 6 ANFIS ensemble models using four types of MFs. For each ensemble ANFIS, we then applied integration by average approach for the final decision.

In Fig. 3, we present the results of ANFIS in three-dimensional plots. These plots are generated form the discovered fuzzy rules from the data and with the aid of implemented membership functions. These figures simply show the interdependencies between each two input variables and the output. To experimentally show the effectiveness of ANFIS ensemble, we conduct the experiments on the hepatitis dataset and compare with the other methods on the classification accuracy of the disease diagnosis. The classification accuracy of ANFIS ensemble was measured by ROC in each cluster. From the results, we found that the ANFIS ensemble has provided good classification accuracies for all clusters. In addition, the results showed that the average classification accuracy of all clusters measured by ROC is AUC (Area Under Curve) = 0.9306. In Fig. 4, AUC values are presented for cluster 3 and cluster 5. It should be noted that after 100 epochs and using 10-cross validation the AUC is calculated for ANFIS classification models.

Our method is also compared with some other classifiers, ANFIS, K-Nearest Neighbors (K-NN), Neural Network (NN) and SVM. We apply these techniques on the same dataset without incorporating SOM and NIPALS for the classification task. In addition to these methods, we compare our method with the previous methods [12,17,18,19] developed for hepatitis diagnosis.

For SVM, the Radial Basis Function (RBF) kernel which is more accurate than other types of SVM kernels was used [39,40]. Accordingly, two parameters of SVM, (C) and γ were found using the exhaustive search approach. In addition, exponentially growing sequences of γ from 2^{-10} to 2^{9} and C form 2^{-15} to 2^{10} were tested to find the optimal values for these parameters of SVM. For NN, we used three-layers feedforward back-propagation. In addition, the resilient back-propagation training algorithm was used to train the three-layers feedforward back-propagation model.

Table 3
Classification accuracy for different classifiers.

Method	Accuracy
K-NN	71.41%
SVM	81.17%
NN	78.31%
ANFIS	79.67%
NIPALS–SOM–ANFIS ensemble	93.06%
PCA–LSSVM	95.00%
PCA–AIRS	94.12%
LFDA–SVM	96.77%
RES–ELM	100%

The result is presented in Table 3. It can be seen from the result that the method which uses dimensionality reduction (NIPALS), clustering (SOM) and ensembles of ANFIS (*NIPALS–SOM–ANFIS Ensemble*) (93.06%) is more accurate compared to the NN (78.31%), ANFIS (79.67%), SVM (81.17%) and K-NN (71.41%) methods. These results show that use of clustering and dimensionality reduction with the aid of ensembles of ANFIS has been effective in hepatitis disease classification. The results further reveal that SVM has outperformed ANFIS, K-NN and NN in hepatitis disease classification. Compared to the previous methods in the literature, the method which uses Rough Set (RS) and Extreme Learning Machine (ELM) [19] provides the best results (100%). In addition, we can see that the methods PCA–LSSVM (95.00%) and LFDA–SVM (96.77%) which combine dimensionality reduction with the SVM have obtained good accuracies on the experimental dataset. Overall, from the comparison of ANFIS and NIPALS–SOM–ANFIS Ensemble, we can conclude that the use of clustering, dimensionality reduction and ensemble learning has improved the accuracy of ANFIS. However, as the use of ensemble learning was effective in improving the classification accuracy, an opportunity for future work is therefore to develop our method based on the techniques presented in the literature which have provided better results compared to the ANFIS. Accordingly, we view other techniques such as Support

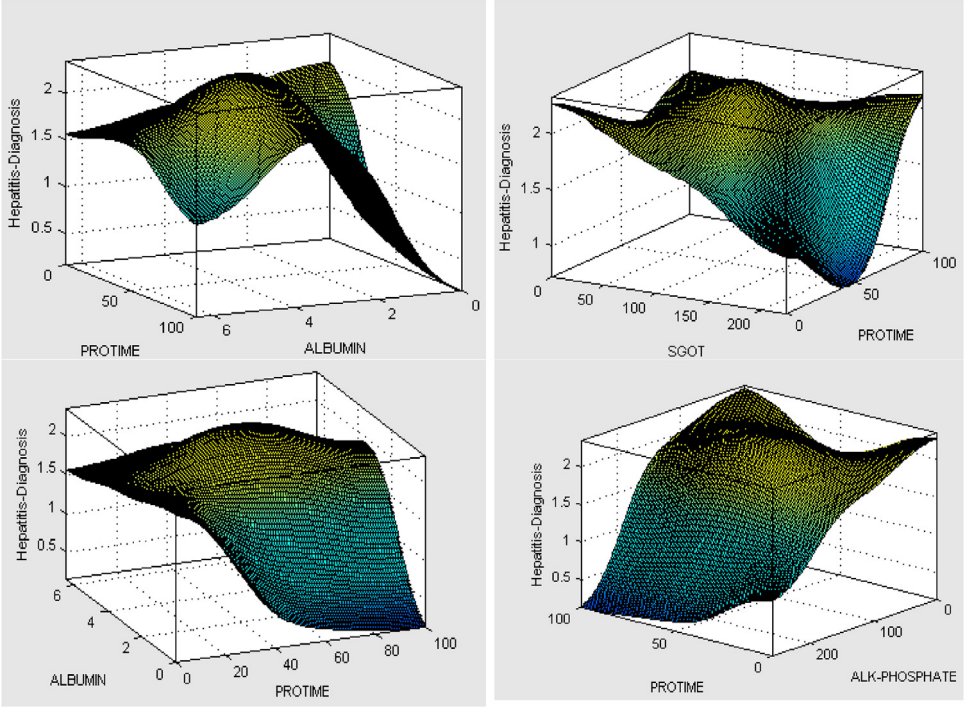

Fig. 3. Hepatitis diagnosis using fuzzy rules.

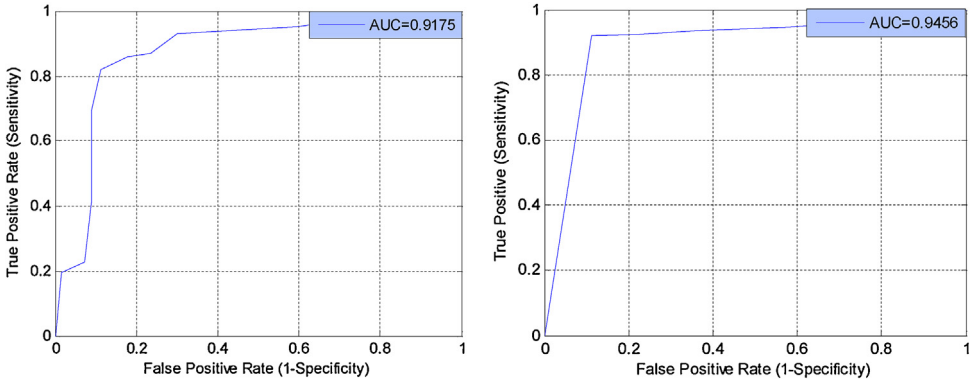

Fig. 4. AUC for cluster 3 and cluster 5.

Vector Machine, Rough Set and Extreme Learning Machine to be complementary to our method.

Conclusion and future work

Accuracy has been one of the important concerns of researchers in developing the methods for diseases diagnosis. There are many methods in the literature for diseases classification. However, the most methods are developed using single learning techniques. This study has investigated the effectiveness of ensemble learning techniques for the prediction of hepatitis disease using several parameters, Age", "Sex", "Steroid, 'Antivirals', "Fatigue", "Malaise", "Anorexia", "Liver Big", "Liver Firm", "Spleen Palpable", "Spiders Ascites", "Arices", "Bilirubin", "Alk 'Phosphate', "Sgot", "Albumin", and "Protime and Histology". Accordingly, we proposed to rely on ANFIS as the supervised and SOM as the un-supervised machine learning techniques. The accuracy of SOM is improved using NIPALS. In addition, we develop our method for ensemble learning by using several types of membership function in ANFIS technique. Our method is evaluated on a real-world dataset obtained from UCI. The accuracy of our method on this dataset measured by ROC was 93.06%.

In this study, non-incremental ANFIS has been implemented for learning the classification models. In addition, the method developed by ANFIS does not support the incremental learning and it requires to recompute all the training data in constructing the prediction models. Accordingly, in order to improve the computation time of hepatitis diagnosis, it is suggested to develop this method to incrementally update the trained models when new information is available, which can be more efficient in memory requirement.

References

[1] Balkhy HH, El-Saed A, Sanai FM, Alqahtani M, Alonaizi M, Niazy N, et al. Magnitude and causes of loss to follow-up among patients with viral hepatitis at a tertiary care hospital in Saudi Arabia. J Infect Public Health 2016;10(4):379–87.
[2] Lavanchy D. Hepatitis B virus epidemiology, disease burden, treatment, and current and emerging prevention and control measures. J Viral Hepat 2004;11(2):97–107.
[3] Lee WM. Hepatitis B virus infection. New Engl J Med 1997;337(24):1733–45.
[4] Almuneef MA, Memish ZA, Balkhy HH, Qahtani M, Alotaibi B, Hajeer A, et al. Epidemiologic shift in the prevalence of Hepatitis A virus in Saudi Arabia: a case for routine Hepatitis A vaccination. Vaccine 2006;24(27):5599–603.
[5] Al-Thaqafy MS, Balkhy HH, Memish Z, Makhdom YM, Ibrahim A, Al-Amri A, et al. Hepatitis B virus among Saudi National guard personnel: seroprevalence and risk of exposure. J Infect Public Health 2013;6(4):237–45.
[6] Al-Thaqafy MS, Balkhy HH, Memish Z, Makhdom YM, Ibrahim A, Al-Amri A, Al-Thaqafi A. Improvement of the low knowledge, attitude and practice of hepatitis B virus infection among Saudi National Guard personnel after educational intervention. BMC Res Notes 2012;5(1):597.
[7] Shepard CW, Finelli L, Alter MJ. Global epidemiology of hepatitis C virus infection. Lancet Infect Dis 2005;5(9):558–67.
[8] Wu JC, Chen TZ, Huang YS, Yen FS, Ting LT, Sheng WY, et al. Natural history of hepatitis D viral superinfection: significance of viremia detected by polymerase chain reaction. Gastroenterology 1995;108(3):796–802.
[9] Haagsma EB, van den Berg AP, Porte RJ, Benne CA, Vennema H, Reimerink JH, et al. Chronic hepatitis E virus infection in liver transplant recipients. Liver Transplant 2008;14(4):547–53.
[10] Perz JF, Armstrong GL, Farrington LA, Hutin YJ, Bell BP. The contributions of hepatitis B virus and hepatitis C virus infections to cirrhosis and primary liver cancer worldwide. J Hepatol 2006;45(4):529–38.
[11] Issur M, Götte M. Resistance patterns associated with HCV NS5A inhibitors provide limited insight into drug binding. Viruses 2014;6(11):4227–41.
[12] 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. Digit Signal Process 2006;16(6):889–901.
[13] The coordination of telematics in healthcare, 1997, World Health Organisation; Geneva, Switzerland http://www.hon.ch/libraray/papers/gulbinat html.
[14] Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. Artif Intell Med 2001;23(1):89–109.
[15] Newman DJ, Hettich S, Blake CL, Merz CJ. UCI repository of machine learning databases; 1998.
[16] 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 Methods Programs Biomed 2012;108(2):570–9.
[17] Çalişir D, Dogantekin E. A new intelligent hepatitis diagnosis system: PCA–LSSVM. Expert Syst Appl 2011;38(8):10705–8.
[18] Chen HL, Liu DY, 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 Syst Appl 2011;38(9):11796–803.
[19] Kaya Y, Uyar M. A hybrid decision support system based on rough set and extreme learning machine for diagnosis of hepatitis disease. Appl Soft Comput 2013;13(8):3429–38.
[20] Polat K, Güneş S. Medical decision support system based on artificial immune recognition immune system (AIRS), fuzzy weighted pre-processing and feature selection. Expert Syst Appl 2007;33(2):484–90.
[21] Tan KC, Yu Q, Heng CM, Lee TH. Evolutionary computing for knowledge discovery in medical diagnosis. Artif Intell Med 2003;27(2):129–54.
[22] Zhang L, Suganthan PN. Random forests with ensemble of feature spaces. Pattern Recognit 2014;47(10):3429–37.
[23] Nahato KB, Nehemiah KH, Kannan A. Hybrid approach using fuzzy sets and extreme learning machine for classifying clinical datasets. Inf Med Unlocked 2016;2:1–11.
[24] Gorunescu F, Belciug S. Evolutionary strategy to develop learning-based decision systems. Application to breast cancer and liver fibrosis stadialization. J Biomed Inf 2014;49:112–8.
[25] Ben-Israel A, Levin Y. The geometry of linear separability in data sets. Linear Algebra Appl 2006;416(1):75–87.
[26] Zorarpacı E, Özel SA. A hybrid approach of differential evolution and artificial bee colony for feature selection. Expert Syst Appl 2016;62:91–103.
[27] Yang P, Zhu Q, Huang B. Spectral clustering with density sensitive similarity function. Knowl Based Syst 2011;24(5):621–8.
[28] Shah SMS, Batool S, Khan I, Ashraf MU, Abbas SH, Hussain SA. Feature extraction through parallel probabilistic principal component analysis for heart disease diagnosis. Physica A 2017;482:796–807.
[29] Nilashi M, Ibrahim O, Ahani A. Accuracy improvement for predicting Parkinson's disease progression. Sci Rep 2016;6.
[30] Wold H. Causal flows with latent variables: partings of the ways in the light of NIPALS modelling. Eur Eco Rev 1974;5(1):67–86.
[31] Preda C, Duhamel A, Picavet M, Kechadi T. Tools for statistical analysis with missing data: application to a large medical database. Stud Health Technol Inform 2005;116:181.
[32] Kohonen T, Oja E, Simula O, Visa A, Kangas J. Engineering applications of the self-organizing map. Proc IEEE 1996;84(10):1358–84.
[33] Zadeh LA. Fuzzy sets. Inf Control 1965;8(3):338–53.

[34] Nilashi M, bin Ibrahim O, Ithnin N. Multi-criteria collaborative filtering with high accuracy using higher order singular value decomposition and Neuro-Fuzzy system. Knowl Based Syst 2014;60:82–101.
[35] Jang JSR. ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans Syst Man Cybern 1993;23(3):665–85.
[36] Woods K, Kegelmeyer WP, Bowyer K. Combination of multiple classifiers using local accuracy estimates. IEEE Trans Pattern Anal Mach Intell 1997;19(4):405–10.
[37] Hansen LK, Salamon P. Neural network ensembles. IEEE Trans Pattern Anal Mach Intell 1990;12(10):993–1001.
[38] Soto J, Melin P, Castillo O. (2013, April). A new approach for time series prediction using ensembles of ANFIS models with interval type-2 and type-1 fuzzy integrators. In Computational intelligence for financial engineering & economics (CIFEr), 2013 IEEE Conference on (pp. 68–73). IEEE.
[39] Nilashi M, Ibrahim O, Ahani A. Accuracy improvement for predicting Parkinson's disease progression. Sci Rep 2016;6:34181.
[40] Nilashi M, Ibrahim O, Ahmadi H, Shahmoradi L, Farahmand M. A hybrid intelligent system for the prediction of Parkinson's disease progression using machine learning techniques. Biocybern Biomed Eng 2018;38(1):1–15.