

Performance Evaluation of Hepatitis Diagnosis using Single and Multi-Classifiers Fusion

Ahmed M. Salah EL-Bohy¹, Atallah I. Hashad², and Hussien Saad Taha³
Arab Academy for Science,
Technology & Maritime Transport, Cairo, Egypt

Abstract- The goal of our paper is to obtain superior accuracy of different classifiers or multi-classifiers fusion in diagnosing Hepatitis using world wide data set from Ljubljana University. We present an implementation among some of the classification methods which are defined as the best algorithms in medical field. Then we apply a fusion between classifiers to get the best multi-classifier fusion approach. By using confusion matrix to get classification accuracy which built in 10-fold cross validation technique. The experimental results show that for all data sets (complete, reduced, and no missing value) using multi-classifiers fusion achieved better accuracy than the single ones.

Keywords- Hepatitis; Classification techniques; Fusion ; WEKA.

I. INTRODUCTION

The diagnosis of some diseases like hepatitis is very difficult task for a doctor, where doctors usually determine decision by comparing the current test results of patients with another one who has the same condition. Hepatitis is one of the most common diseases among Egyptians; as it represents 22% of hepatitis cases around the world. Which motivates us for suggesting new methods to improve the outcomes of existing approaches, as well as to help doctors and specialists to diagnose hepatitis disease survival. [1]

Hepatitis (Greek) means 'liver' and suffix -itis denotes 'inflammation' of the liver and may be due to infectious Or non-infectious causes. The five types of hepatitis viruses are Common infectious causes of liver inflammation and some like Hepatitis A (HAV), B (HBV) and C (HCV) are more frequently seen infectious agents. Inflammation may lead to death of the liver cells (hepatocytes) which severely compromises normal liver function. Acute HBV Infection (less than 6 months) may resemble the fever, flu, muscle aches, joint pains and generally being unwell. Symptoms specifying that states are: dark urine, loss of appetite, nausea, vomiting, jaundice, pain up the liver. Chronic hepatitis B is an infection persisting more than 6 months, the clinical features of that state correspond to liver dysfunction, so signs like these may be noticed: enlarged liver, splenomegaly, hepatosplenomegaly, jaundice, weakness, abdominal pain, confusion and abdominal swelling [2]. The success of treatment depends on an early recognition of the virus, which achieves more exact and less violent treatment options and mortality from Hepatitis falls.

Recently, data-mining has become one of the most treasured tools for operating data in order to produce valuable

information for decision-making [3]. Supervised learning, including classification is one of the most significant brands in data mining, with a recognized output variable in the dataset. Classification methods can achieve high accuracy in classifying mortality cases. The implementations had been applied with a "WEKA" tool which stands for the Waikato Environment for Knowledge Analysis. A lot of papers about applying machine learning procedures for survivability analysis in the field of Hepatitis diagnostic. Here are some examples:

Using Support Vector Machines And Wrapper Method for predicting Hepatitis was introduced achieving maximum accuracy of (74.55%). [4]. but we note that applying SVM classifier only get higher accuracy than the mentioned accuracies even with feature selection.

Improving the accuracy of SVM algorithm using feature selection [5]. Using SVM with Chi-Square achieved accuracy of (83.12 %), But we note that applying another classifier (Logistic, Simple Logistic, SMO, RF, J48) gets higher accuracy than the mentioned one.

Detection of Hepatitis based on SVM and Data Analysis [6] using SVM & wrapper the achieved accuracy is (85%) but we found the results are calculated with 7 attributes out of 25 ones as stated whenever the original number of attributes is 20 only

The rest of this paper is prearranged like this: In sector II, Classification algorithms are discussed. In sector III Evaluation principles are discussed. In sector IV a proposed model is shown. In sector V reports the experimental results. Finally, Sector VI introduces the conclusion.

II. CLASSIFICATION ALGORITHMS

A Bayesian belief network is sometimes named a Bayes net, a belief net, or a causal network, It is a directed, acyclic graph, indicating conditional dependencies. It can be used to guess the probability of events. the Bayesian decision rule assures minimum error if likelihoods and prior probabilities are known [7].

Decision Tree (DT) Tree that the root and each interior node is marked with a question. The arcs represent each possible answer to the concomitant question. Each leaf node represents a forecast of a problem solution. A prevalent technique for classification; Leaf node leads the class to which the corresponding tuple belongs. It's Model is a computational

model comprising of the three parts: Decision Tree Algorithm to create the tree Algorithm that applies the tree to the data, creation of the tree is the most exciting part. Processing is mostly a search similar to that in a binary search tree (although DT may not be binary). Advantages: Easy to understand. Easy to generate rules[8]

Support Vector Machine (SVM) SVMs are amongst the best (and many believe are definitely the best)“on-the-shelf ” supervised learning algorithm. It’s derived from statistics in 1992.SVM is widely used in multiple applications pattern recognition, classification and regression. The SVMs work on an underlying principle, which is to insert a hyper-plane between the classes and orient it in such a way so as to keep it at the maximum distance from the nearest data points These data points, which appear closest to the hyper-plane, are known as Support Vectors. [9]

Sequential Minimal Optimization (SMO) humble, easy to implement, is generally quicker, and has better scaling properties for difficult SVM problems than the usual SVM training algorithm SMO quickly solve the SVM QP problem without extra matrix storage (the memory used is linear with the training data set size) and without using numerical QP optimization steps at all. SMO can be used for online learning. While SMO has been shown to be operative on sparse data sets and especially fast for linear SVMs, the algorithm can be extremely slow on non-sparse data sets and on problems that have many support vectors. Regression problems are especially prone to these matters because the inputs are usually non-sparse real numbers (as opposed to binary inputs) with solutions that have many support vectors. Because of these restrictions, there have been limited reports of SMO being successfully used on regression problems. [10]

Logistic Regression (LR) is a famous well known classifier, it could be used to extend classification results into a deeper analysis. It’s not widely used due to its slow response in data mining especially when compared with SVM in large data sets (not our case) It models the relationship between a dependent and one or more independent variables, and consents us to look at the fit of the model as well as at the significance of the relationships [11]

Simple Logistic: We use simple logistic regression when we have one nominal variable with two values (dead/alive, male/female) and one measurement variable. The nominal variable is the dependent variable, but the measurement variable is not it is an independent one.

Simple logistic regression is analogous to linear regression, except the dependent variable is nominal, not a measurement. [12]

Random Forest: Random forests change how the classification or regression trees are constructed. In standard trees, it uses the best split among all variables to split each node In a random forest, each node is split using the best among a subset of predictors arbitrarily chosen at that node. it works by one of two methods, boosting and bagging.it has the advantages of: handling thousands of input variables without deleting any variable, giving estimation of variables importance in the classification, and It also has an active method for estimating

missing data and keeps accuracy when a large amount of the data are missing [13].

SGD is an abbreviation for Stochastic Gradient Descent: the gradient of the loss is estimated each sample at a time and the model is updated along the way with a decreasing strength schedule. SGD has been excellently applied to large-scale and sparse machine learning problems often come across the text classification and natural language processing [14].

K-star or K* is an instance-based classifier. The class of a test instance is based on the training instances similar to it, as determined by some similarity function. The difference between it and the other instance-based learners in that K* uses an entropy-based distance function. Instance-based learners classifying instance by matching it to a database of pre-classified cases. The fundamental assumption is that similar instances will have similar classifications [15].

III. EVALUATION PRINCIPLES

Evaluation method is based on the confusion matrix. The confusion matrix is an imagining implement usually used to show presentations of classifiers. It is used to display the relationships between real class attributes and predicted classes. The grade of efficiency of the classification task is calculated with the number of exact and unseemly classifications in each conceivable value of the variables being classified in the confusion matrix [16]

Table 1. Confusion matrix

		Predicted Class	
		Negative	Positive
Outcomes	Negative	TN	FN
	Positive	FP	TP

For instance, in a 2-class classification problem with two predefined classes (e.g., Positive, negative) the classified test cases are divided into four categories:

- True positives (TP) correctly classified as positive instances.
- True negatives (TN) correctly classified negative instances.
- False positives (FP) incorrectly classified negative instances
- False negatives (FN) incorrectly classified positive instances.

To evaluate classifier performance. We define accuracy term that is defined as the entire number of truly classified instances divided by the entire number of available instances for an assumed operational point of a classifier.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

IV. PROPOSED METHODOLOGY

We propose a reliable method for diagnosing Hepatitis and better classifying mortality cases that may be caused by Hepatitis infection based on data mining using WEKA as follows:

A. Data preprocessing

Pre-processing steps are applied to the data before classification:

First step is Data Cleansing: eliminating or decreasing noise and the treatment of missing values. Hepatitis data set has a lot of missing values, especially in attributes 18th, 15th, and 17th as they are 67, 29, 16 values respectively. After removing the instances has these missing values the (155) instance data set converted to (83) instances with minor missing values in other attributes (Only 9 missing values in the reduced data set). We also convert the data set into a new (third) one by removing these (9) missing values to be (80) instance instead of (83) instance.

Second step is Relevance Analysis: Statistical correlation analysis is used to discard the redundant features from further analysis. The data set has one irrelevant attribute named 'sex' which has no effect in the classification process;

Final, step is Data Transformation: The dataset is transformed by normalization, which is one of the greatest public tools used by inventors of automatic recognition classifications to get superior results. Data normalization hurries up training time by initialing the training procedure to reach feature within the same scale. The aim of normalization is to transform the attribute values to a small-scale range.

B. Single classification task

Data classification is the process of organizing data into categories for its most effective and efficient use. A well-structured data classification system makes it easy to get, handle, and retrieve necessary data.

Classification consists of conveying a class label to a set of unclassified cases. Supervised classification: The set of possible classes is known in advance. Unsupervised classification: Set of possible classes is unknown. After classification we can try to assign a name to that class. Unsupervised classification is known as clustering. The presumed model depends on analyzing the training dataset. The derivative model characterized in several procedures, such as simple classification rules, decision trees and another. Basically data classification is a two-stage process, in the initial stage; a classifier is built signifying a predefined set of notions or data classes.

This is the training stage, where a classification technique builds the classifier by learning from a training dataset and their related class label columns or attributes. In next stage this model is used for measurement. In order to deduce the predictive accuracy of the classifier an independent set of the training instances is used. We evaluate the most frequently classification techniques that mentioned in recently published researches in the medical field to accomplish the highest accuracy classifier's result with each data set.

C. Multi-classifiers fusion classification task

Fusion is a combination between 2 or more classifiers to enhance classifier performance (Accuracy). Our procedure will be electing the highest 2 classifiers in accuracy calculating accuracy, then the 3rd, then 4th and so on until accuracy decreases then stop & take the maximum accuracy obtained as shown in Figure 1 We name the number of classifiers used in the fusion process "fusion level" best accuracy with other single classifiers predicting to improve accuracy. Repeating the same process till the latest level of fusion, according to the number of single classifiers to pick the highest accuracy through all processes.

We propose 2 methods (1st Directly applying a classifier to the raw data & 2nd preprocess the data before applying classifier).

Import the Dataset.

- Convert nominal values to binary ones (in case of preprocessing).
- Normalize each variable of the data set, so that the values range from 0 to 1 (in case of preprocessing).
- Create a separate training (learn) and testing sets by indiscriminately drawing out the data for training and for testing.
- Picking and adapt the learning procedure
- Perform the learning procedure
- Calculate the performance of the model on the test set

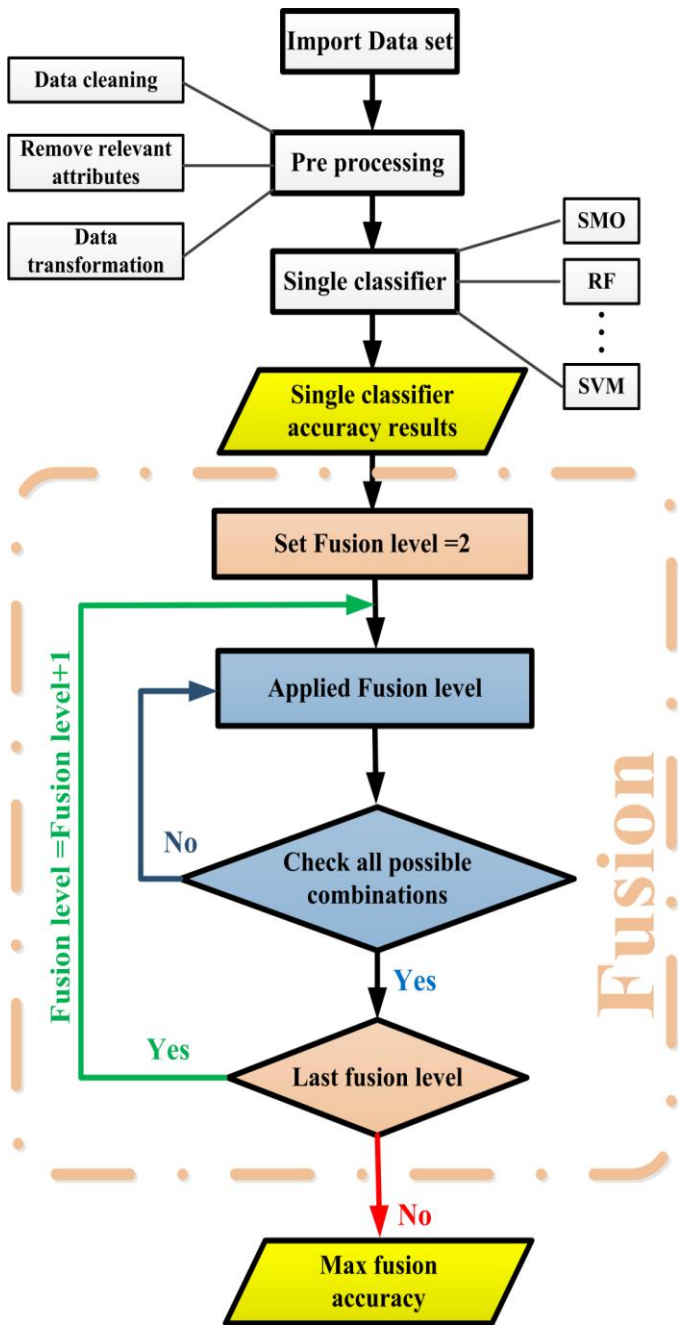


Figure 1 .Proposed Hepatitis diagnosis algorithm

V. EXPERIMENTAL RESULTS

According to the mentioned work in sector IV proposed methodology-A- data preprocessing we now have 3 data sets with 155, 83, 80 instances respectively & last two data sets has been preprocessed as mentioned in sector IV proposed methodology-A- data transformation To calculate the proposed model, first we calculate accuracy using 9 single classifiers then we compute accuracy in fusion process up to 9 classifiers together using the highest classifiers combination in each fusion level (All results are illustrated graphically in Figure 2).

A. Single classification task

Table 2 shows the comparison between accuracies over nine single classifiers (Bayes Net, SVM, Logistic, SGD, Simple Logistic, SMO, K*, J48, and RF). Highest accuracy in each data set is highlighted in yellow color in Table 2. Using 155 instances (complete) data set the highest accuracy by SMO classifier is 85.17%. Using 83 instances (reduced) data set the highest accuracy by K* classifier is 91.67%. Using 80 instances (no missing value) data set the highest accuracy by Bayes Net classifier is 88.75%.

SMO classifier achieves better accuracy than SVM classifier illustrated in previous work for the three data sets without any preprocessing or feature selection [4-6]

Furthermore, we note that RF classifier achieves the second rank among single classifiers for the whole three data sets and it's highlighted in cyan color in Table 2.

Finally K* classifier bare as a good classifier as it achieves highest accuracy in one data set and it got the second place in another one equal with RF classifier in 80 instances (no missing value) data set

Table 2 Single classifier results

# instance \ Classifier	155 complete	83 Reduced	80 no missing value
BayesNet	83.21	86.81	88.75
SVM	79.38	82.08	83.75
Logistic	82.58	77.92	81.25
SGD	84.54	84.17	81.25
Simple Logistic	83.88	80.42	85.00
SMO	85.17	84.03	85.00
K*	81.96	91.67	87.50
J48	83.79	82.22	86.25
RF	85.13	90.56	87.50

Table 3 shows the comparison between accuracies over multi classifiers fusion between classifiers up to nine classifiers all possible combination has been executed as illustrated in proposed algorithm Figure 1 and the table contains the highest accuracy at each fusion level.

Highest accuracy in each data set for all fusion levels is highlighted in yellow color in Table 3. Using 155 instances (complete) data set the highest fusion accuracy obtained by using K*, J48, RF, Simple Logistic, and Logistic classifiers is 87.04%. Using 83 instances (reduced) data set the highest fusion accuracy obtained by using K*, RF, SVM, and BayesNet classifiers is 92.92%. Using 80 instances (no missing value) data set the highest fusion accuracy obtained by using K*, J48, RF, and BayesNet classifiers is 91.25%.

The results in Table 3 indicates

First data preprocessing minimized the fusion level / number of classifiers by one classifier as we get the best result / highest accuracy at fusion level 4th level instead of 5th level in case of 155 instances (complete) data set

Second K* and RF are performed as outstanding classifiers in fusion for Hepatitis data set in all cases (complete, Reduced, and no missing value) with / without preprocessing

Finally the importance of BayesNet classifier is appeared for less or no missing values data as it increases accuracy by a noticeable amount in both 83 instances (reduced) and 80 instances (no missing value) data sets

Table 3 Fusion classifiers results

# instance \ Fusion Level	155 Complete (%)	83 Reduced (%)	80 no missing value (%)
2 nd	86.42	91.67	88.75
3 rd	86.46	89.31	88.75
4 th	86.42	92.92	91.25
5 th	87.04	91.67	90.00
6 th	87.04	90.42	90.00
7 th	86.42	90.42	91.25
8 th	85.75	86.67	88.75
9 th	85.03	85.42	88.85

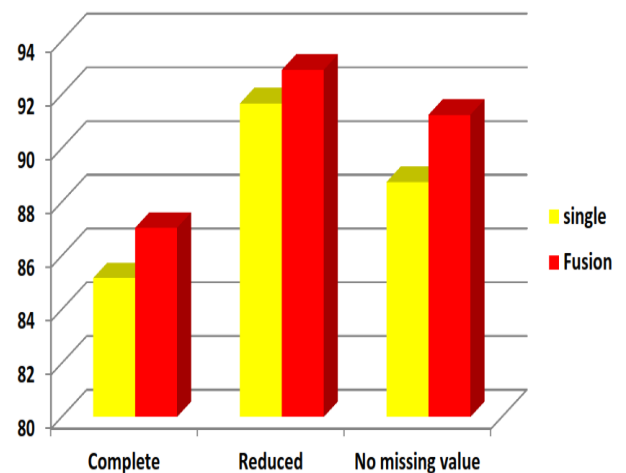


Figure 2 Single / Fusion classifier accuracy comparison

VI. CONCLUSION

The experimental results clarified that multi-classifiers fusion reaches better accuracy than single classifier. For the three data sets.

Although no unique classifier achieves the highest accuracy as a single classifier for all data sets but RF classifier playing a great role as a single classifier for the 3 mentioned data sets (complete, reduced, and no missing value). Also the K* classifier is a common master classifier between all fusion combination, adding RF to it in fusion process gives an outstanding results. BayesNet classifier consider as a break through dealing with Hepatitis reduced and no missing value data sets in fusion as it improves accuracy by 1.25% and 2.5% from best single classifier respectively and increase it by 3.61% and 2.5% from pervious fusion level

REFERENCES

- [1] World Health Organization at <http://www.who.int>
- [2] Tomasz KANIK. "Hepatitis disease diagnosis using Rough Set modification of the pre-processing algorithm." Information and communication technologies international conference (2012).
- [3] Fayyad, Usama, Gregory Piatetsky-Shapiro, and Padhraic Smyth. "From data mining to knowledge discovery in databases." *AI magazine* 17.3 (1996).
- [4] A.H.Roslina and A.Noraziah. "Prediction Of Hepatitis Prognosis Using Support Vector Machines And Wrapper Method." Seventh International Conference on Fuzzy Systems and Knowledge Discovery (2010).
- [5] Varun Kumar.M , Vijay Sharathi.V and Gayathri Devi.B.R. " Hepatitis Prediction Model based on Data Mining Algorithm and Optimal Feature Selection to Improve Predictive Accuracy." *International Journal of Computer Applications (0975 – 8887) Volume 51– No.19, August 2012.*
- [6] C. Barath Kumar , M. Varun Kumar , T. Gayathri, and S. Rajesh Kumar " Analysis and Prediction of Hepatitis Using Support Vector Machine." *International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 2235-2237.*
- [7] Jyotirmay Gadewadikar, Ognjen Kuljaca1, Kwabena Agyepong, Erol Sarigul, Yufeng Zheng and Ping Zhang, Exploring Bayesian networks for medical decision support in breast cancer detection, *African Journal of Mathematics and Computer Science Research Vol. 3(10), pp. 225-231, October 2010.*
- [8] Ross Quinlan, (1993) C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, San Mateo, CA.
- [9] Chen Y., Wang G., and Dong S. "Learning with progressive transductive support vector machine", *Pattern Recognition Letters, Vol. 24, Pages: 1845-1855*
- [10] John C. Platt. "Sequential Minimal Optimization:A Fast Algorithm for Training Support Vector Machines." *Technical Report MSR-TR-98-14April.*
- [11] Mitchell T. "Machine Learning, mcgraw Hill." (1997).
- [12] John H. Mcdonald. "Handbook of Biological Statistics." At <http://www.strath.ac.uk>
- [13] Breiman, Leo. "Random forests." *Machine learning* 45.1 (2001): 5-32.
- [14] Y. Tsuruoka, J. Tsujii, and S. Ananiadou. " Stochastic gradient descent training for l1-regularized log-linear models with cumulative penalty " *AFNLP/ACL 2009.*
- [15] Mahmood, Deeman Y., and Mohammed A. Hussein. "Intrusion Detection System Based on K-Star Classifier and Feature Set Reduction." *International Organization of Scientific Research Journal of Computer Engineering (IOSR-JCE) Vol 15 (2013): 107-112.*
- [16] P. Cabena, P. Hadjinian, R. Stadler, J. Verhees and A.Zanasi, *Discovering data mining from concept to implementation.* Uppersaddle River, N.J.: Prentice Hall, 1998.