Journal of Network Communications and Emerging Technologies (JNCET)
Volume 10, Issue 8, August (2020)

Diagnosis of Liver Diseases Using Neural Network Ensemble

Bihter Das
Department of Software Engineering, Firat University, 23119, Elazig, Turkey.
bihterdas@gmail.com

Published Online: 31 August 2020

Abstract – In the last couple of decades, the diagnosis of liver diseases using computational techniques was heavily investigated. This study focuses on the using of a Neural Network ensemble-based method for effective diagnosis of liver diseases. In this study, a Neural Network ensemble-based method was proposed to predict liver diseases. Although ANN and other classification approaches have been heavily investigated in recent years, the using of ensemble-based approaches to predict liver diseases has not been thoroughly investigated. The proposed model relies on using five different ANN nodes. Using the commonly used Indian Liver Patient Dataset, which is provided by the University of California, Irvine, and SAS software suite evaluates the accuracy of the proposed model. The obtained results indicate that in the validation phase the classification accuracy of the proposed model to predict liver diseases is 74.35%. Also, in terms of important classification metrics such as specificity, sensitivity, precision, false-positive rate, false-negative rate and F_1(F-measure) in the training phase the proposed model achieved the rates 36.36%, 89.28%, 78.12%, 63.64 %, 10.72% and 83.33%, respectively. The results are very promising for researchers and practitioners working in related fields.

Index Terms – Liver diseases; classification; artificial neural networks; ensemble-based methods; SAS software suite.

1. INTRODUCTION

The liver is the one of the most important organs and supports almost every other organ in the body. Due to its strategic location and involving several dimensions functions, the liver is also inclined to many diseases. The bare area of the liver is a site that is vulnerable to the passing of infection from the abdominal cavity to the thoracic cavity. The liver has many important functions, including digesting your food and processing and distributing nutrients. There are many kinds of liver diseases and conditions. Some, like hepatitis, are caused by viruses. Others can be the result of drugs or be drinking too much alcohol. Long-lasting injury or scar tissue in the liver can cause cirrhosis. Jaundice, or yellowing of the skin, can be one sign of liver disease [1]. The liver is a vital organ and the Canadian Liver Foundation (CLF) estimates that one out of 10 Canadian citizens suffers because of liver diseases [2]. It was reported that the prevalence of liver diseases is in increase both in the European Union (EU) zone and the United States of America [3]. Common types of liver diseases are Cirrhosis, Liver Cancer, Hepatitis (Hepatitis A, Hepatitis B, Hepatitis C, Hepatitis D, and Hepatitis E, Liver transplant, Alcoholic Liver Disease (ALD), Non-Alcoholic Fatty Liver Disease (NAFLD) and Non-Alcoholic Steato Hepatitis (NASH). Since the number of people suffering from liver diseases has been increasing around the world, using the computational method for early diagnosis is of the utmost importance [4,5].

In this paper, classification is performed for the effective diagnosis of liver disease by using ensemble-based methods. The Indian Liver Patient Dataset in the UCI repository was used for classification [6]. All 11 different feature sets in the data set were used for classification measurements using ensemble-based methods. In the literature, classification has been made using this data set, but less than 11 feature sets have been classified. The importance of this study is that an effective classification is made using the popular neural network ensemble method for 11 feature set.

The remainder of the paper is organized as follows. The related works of this study is presented in Section 2. Materials and methodology of the proposed system is presented in Section 3. The implemented of the proposed system is presented in Section 4. The components of the SAS base software are introduced briefly. The implementation constraints are also given in this section. The experimental results and discussion are reported in Section 5. The SAS base software represents several statistical evaluation tests and different graphics for the users. Brief explanation can also be found in this section. Finally, section 6 concludes the paper.

2. RELATED WORKS

In the last couple of decades, the diagnosis of liver diseases using computational techniques was heavily investigated. There are many different studies and publications related to the classification of liver diseases [7,8] and its diagnosis by using the neural network ensemble method in the academic literature [9,10,11,12]. Furthermore, it is seen that neural network ensemble method is used for the detection of various diseases [10,11] In this section, several of them are presented.

Netzer et al. in [9] proposed the using of an ensemble-based algorithm to identify breath gas marker of candidates suffering from alcoholic fatty liver disease (AFLD). The authors used Ion molecule reaction mass spectrometry (IMR-MS), and studied 126 human breath gas samples comprising 91 cases suffering because of AFLD, NAFLD and cirrhosis and 35 healthy ones for control. Their results showed that the using of Stacked Feature Ranking (SFR) as a new feature selection modality had better performance than single feature selection methods and proved that ensemble methods could be very

efficient in the prediction of liver diseases, particularly AFLD. In [13] Acharya et al. studied fatty liver disease and using Computer Aided Diagnostic (CAD) techniques on the ultrasound images of 20 abnormal and 15 normal livers they obtained the accuracy of 93.3% based on the Steatosis Classification Index (SCI). In [14] pulse waves of fatty liver patients, cirrhosis patients, and healthy individuals were collected and analyzed using unsupervised learning Principal Component Analysis (PCA) and supervised learning Least Squares regression (LS) and Least Absolute Shrinkage and Selection Operator (LASSO) with cross-validation. The results indicated that when 7 parameters were used the highest accuracy, 93%, was obtained using supervised learning's (LS and LASSO). However, when 2 parameters were used, the highest accuracy was 84%. In [13] Abdar used Rapid Miner and IBM SPSS Modeler and the Indian Liver Patient Dataset (ILPD) [6] to evaluate the performance of various classification algorithms. While Support Vector Machine (SVM), C4.5, Random Forest, Chi-square Automatic Interaction Detector (CHAID), Artificial Neural Network (ANN), Linear Regression, Naive Bayes and k-Nearest Neighbor (k-NN) were used in Rapid Miner, C5.0, KNN, ANN, CHAID, Logistic Regression, SVM, Bayesian networks and Quick, Unbiased, Efficient Statistical Tree (QUEST) were used in IBM SPSS Modeler. Although in Rapid Miner ANN-based approach achieved the accuracy of 70.81%, 9.30%, in IBM SPSS Modeler C5.0-based approach achieved the accuracy of 87.91%. Pérez-Ortiz et al. in [15] showed that combining threshold models by independently separating each class the proposed natural and general ensemble method can compete with other state-of-the-art classifiers such as AdaBoost, EBC (SVM) or KDLOR.

In recent years, another focus of the researchers working in the related fields was to show the benefits of the using of data mining and text mining algorithms for disease prediction. Zhang et al. in [16] investigated the use of text mining for NAFLD to understand Traditional Chinese Medicine (TCM) pathogenesis. Using SinoMed and PubMed datasets and relying on the theory of 'formulae-pattern-disease' correlation, it was shown that the probable TCM pathogenesis of NAFLD is related to the biological process of lipid metabolism disorder, inflammation, and metabolic regulation confusion. In [17] Aneeshkumar and Venkateswaran proposed Reverse Sequential Covering Algorithm to predict AFLD, NAFLD of male (NAFLD-M) and female (NAFLD-F). In [18] random forest algorithm was used to diagnose chronic hepatitis B and argued that due to the costly process of diagnosis and treatment of this disease, as well as complications of drugs, the use of data mining to predict chronic hepatitis B is quite useful. In [19] mining frequent pattern tree was used for pain-related decision making for chronic disease patients.

Literature survey shows that one of the most popular classifiers for liver disease prediction is based on ANNs. In [20] based on the Wisconsin Diagnostic Breast Cancer (WDBC) dataset, ILPD dataset, the Vertebral Column Data Set (VCDS) and Heart Disease Data Set (HDDS), Weng et al. implemented different types of ANN classifiers for the detection of four diseases, and validated the results using cross validation tests. When the ILPD was used, the highest accuracy rate was 0.7938. As it was used in [20] although different datasets are available, one of the most popular datasets is the ILPD. It was collected by Ramana et al. in 2012 from the North East of Andhra Pradesh region in India [8] and is available in the data repository of the University of California, Irvine (UCI). It includes information about 583 individuals located in two classes, as well as except column for class of patients, has 10 other columns related to 441 male patient records and 142 female patient records. One of the restrictions of this dataset is that patients older than 89 years old are considered as the age of 90.

3. MATERIALS AND METHODS

Artificial neural networks (ANN) are one of the most popular and even the most successful approach used in data mining and machine learning applications, originally developed to imitate the neurophysiology of the human brain [8]. ANN consists of three layers: input, hidden and output, and Multi-layer Perceptrons (MLP) is one of the most common types. Being an appropriate solution for non-linear problems, ANN can be used for various purposes [21] like disease prediction [22], classification [23], object and image recognition [24] seismic events and earthquake prediction [25,26] and temperature and weather forecasting [27]. Most feedforward ANNs rely on the backpropagation algorithm, a method for computing the error gradient for a feedforward network [10,11]. An artificial neuron model is in Figure 1 and its net output is calculated using Equation 1.

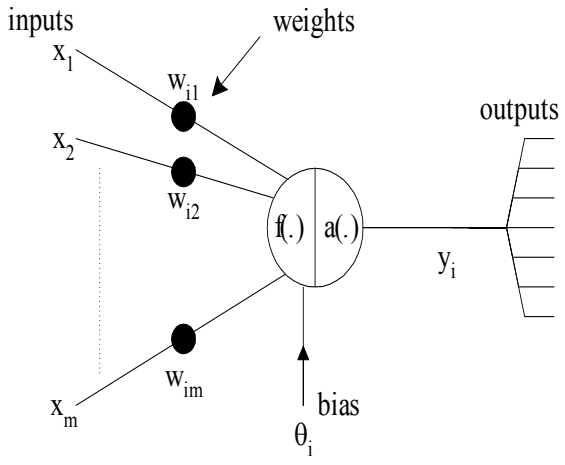

Fig. 1 Artificial neuron model [28].

$y\,(t+1) = a\,(\sum_{j=1}^{m} w_{ij} x_j(t) - \theta_i)$ and

$$f_i \triangleq net_i = \sum_{j=1}^{m} w_{ij} x_j - \theta_i \quad (1)$$

where, $X = (X_1, X_2, \ldots, X_m)$ represents *m* input of the neuron, W_i represents the weight for input X_i,
θ_i represents a bias value, and F (y) is an activation function.

3.1 Ensemble-based methods

Ensemble methods combine several decision trees classifiers to produce better predictive performance than a single decision tree classifier. The main principle behind the ensemble model is that a group of weak learners come together to form a strong learner, thus increasing the accuracy of the model. When we try to predict the target variable using any machine learning technique, the main causes of difference in actual and predicted values are noise, variance, and bias. Ensemble helps to reduce these factors except noise, which is irreducible error. Recently, ensemble methods which rely on a combination of several models have become extremely popular [28, 29]. The assumption behind the ensemble methods is that the combination can increase the accuracy compared to each of the models. The ensemble-based methods create new models by combining the posterior probabilities for class targets or the predicted values for interval targets from multiple predecessor models [10,11] As proven in many studies, the use of ensemble methods improves the performance of algorithms and results. Seven common types of ensembles are:

1. Bayes optimal classifier [30]
2. Bootstrap aggregating (Bagging) [32]
3. Boosting [33, 34]
4. Bayesian parameter averaging [35]
5. Bayesian model combination [36]
6. Bucket of models [37]
7. Stacking [38]

In the literature, researchers working on ensemble methods have focused on improving the performance of algorithms for the prediction of diseases including heart disease [10] valvular heart disease [10,11] liver disease [9], and Erythemato - Squamous Diseases [39]. Accordingly, researchers have compared the performance of various ensemble methods. In [40] using 23 datasets, the accuracy of Bagging and Boosting methods were compared and it was shown that almost always Bagging was more accurate than a single classifier, but sometimes was less accurate than Boosting. However, when ANN was used, Boosting was less accurate than the single classifier. Dietterich in [7] used Bagging and Boosting on C4.5 decision trees. He proved that without classification noise or little Boosting and randomization had better performance than Bagging while with substantial classification noise Bagging had much better performance than Boosting and randomization. The ensemble-based methods are different in according to the terms of implementation and in terms of the procedure by which single classifiers have been produced, and/or the procedure by which the classifiers have been combined. A schematic view of ensemble model has been shown in Figure 2.

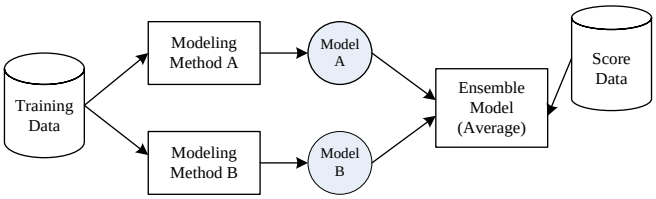

Fig. 2 A schematic view of ensemble models [10,11].

As given in [41], ensemble methods include for classification tasks mainly include 4 main building blocks as follows:

1. *Training set:* In this block, a labeled dataset is used for ensemble training. *A* is the set of input attributes which include *n* attributes that $A = \{a_1, ..., a_i, ..., a_n\}$ and *y* to show the class variable or the target attribute.
2. *Base Inducer:* It is an appropriate approach as an induction algorithm and shows relationship between the input attributes and the target attribute.
3. *Diversity Generator:* Different classifiers are generated by this component.
4. *Combiner:* Finally, in this block, different classifiers are combined.

3.2 Bagging and Boosting

In this part, briefly presents the two ensemble classification methods which bagging and boosting from the aforementioned base classifiers.

Bagging also called Bootstrap aggregation, can improve unstable estimation or classification schemes and decreases the variance of prediction through generating additional data and, as result, more data becomes available for training [42] Although increasing the number of data for the training phase does not guarantee the accuracy improvement but narrowly tunes the prediction to the expected result through decreasing the variance. On the other hand, boosting includes two main steps. While in the first step, it utilizes subsets of original dataset in order to create a series of averagely performing models, it boosts their performance through combining them together using a specific cost function or based on majority vote. In bagging, each classifier is adjusted on a randomly drawn training set with the probability of drawing any given example being equal. Samples are drawn with replacement, so that some examples may be selected multiple times while others may not be selected at all [43].

Boosting is an ensemble technique that attempts to create a strong classifier from a number of weak classifiers. This is done by building a model from the training data, then creating a second model that attempts to correct the errors from the first model. Using techniques like Bagging and Boosting helps to decrease the variance and increased the robustness of the model. Combinations of multiple classifiers decrease variance,

especially in the case of unstable classifiers, and may produce a more reliable classification than a single classifier [43].

4. PROPOSED APPRROACH AND IMPLEMENTATION

In this study, the proposed methodology, which is illustrated in Figure 3, is implemented with the SAS Base Software 9.1.3 platform. The obtained liver disease dataset from UCI were used to implement the proposed approach. There are 11 different features in the data set. In the literature applications, only 3, 4, 5 or 7 of these features were classified. The data obtained from the UCI repository by the researchers and used in the applications were measured by different methods. There have been methods with low or higher classification performances. However, in this study, 11 features were used simultaneously with ANN based ensemble method and a high-performance classification was made in the whole data set. That is, the difficult path was chosen and the classification performance of the ensemble method was tested.

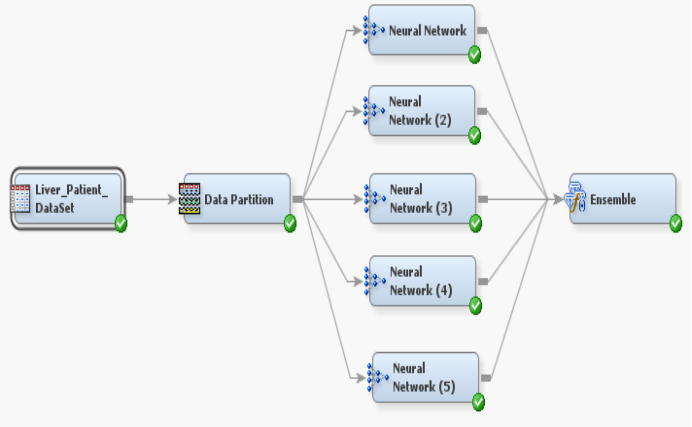

Fig. 3 Performance analysis of the ensemble method for liver disease diagnosis.

Fig. 3 shows how the performance evaluation of the ensemble method is realized using SAS software suite [43]. Before realizing the performance evaluation study, we used different number of independent neural network nodes and decided that ideally, we needed five independent neural network nodes. The confusion matrix given in Table 1 was used in this study. Basically, the confusion matrix is used to describe the performance of a classification model on a set of test data for which the true values are known [10]. The confusion matrixes for training and validation datasets are listed in Tables 2 and 3.

Table1 Confusion matrix used in this study.

Actual	Predicted	
	Disease (positive)	No-disease (negative)
Positive	TP	FP
Negative	FN	TN

Where TP, FN, FP and TN are as follows:
TP = *True Positive:* the number of positive examples correctly classified.
FN = *False Negative:* the number of negative examples misclassified as positive.
FP = *False Positive:* the number of positive examples misclassified as negative.
TN = *True Negative:* the number of negative examples correctly classified.

Table 2 Confusion matrix for training dataset.

Actual	Predicted	
	Disease (positive)	No-disease (negative)
Positive	298	66
Negative	34	68

Table 3 Confusion matrix for validation dataset.

Actual	Predicted	
	Disease (positive)	No-disease (negative)
Positive	75	21
Negative	9	12

One of the most popular data repositories is the University of California, Irvine (UCI) [6]. For this reason, the liver disease data set was taken from this repository. The Indian Liver Patient Dataset (ILPD) have been collected from north east of Andhra Pradesh, India by Dr. Ramana et al. in 2012 [8]. The data set include information about 583 individuals which are located in two classes, as well as except column for class of patients, ILPD has other 10 columns which are related to 441 male patient records and 142 female patient records. One of the important points about this database is that patients whose are older than 89 years old, are considered as the age of 90. More details about this data set is presented as follows:

1. ***Age:*** Age of the patient [4-90]
2. ***Gender:*** Gender of the patient [Male - Female]
3. ***TB:*** Total Bilirubin [0.4-75]
4. ***DB:*** Direct Bilirubin [0.1-19.7]
5. ***Alkphos:*** Alkaline Phosphotase [63-2110]
6. ***Sgpt Alamine:*** Aminotransferase [10-2000]
7. ***Sgot Aspartate:*** Aminotransferase [10-4929]
8. ***TP:*** Total Proteins [2.7-9.6]
9. ***ALB:*** Albumin [0.9-5.5]
10. ***A/G Ratio:*** Albumin and Globulin Ratio [0.3-2.8]
11. ***Selector field:*** (class 1- 416 liver patient records and class 2- 167 non-liver patient records)

5. EXPERIMENTAL RESULTS AND DISCUSSIONS

In order to visually evaluate the performance of the proposed ensemble model, SAS Enterprise Miner 5.2 provides tools. The liver disease dataset was provided in two classes. While class 1 was for liver disease patients, class 2 was for non-liver

disease patients. The dataset was divided in two groups that 70% of the liver disease dataset was used in the training phase of the proposed ensemble model and the rest of the liver disease dataset (30%) was used in the validation phase of the ANN ensemble system. Based on Table 2 and Table 3, the confusion matrix was applied for both the training (TRAIN) and validation (VALIDATE) phases. The measured metrics of the phases are given in Table 4. As given in Table 4, the accuracy rate of the proposed ensemble model was 78.44% and 74.35% for the training and validation phases, respectively. For both the training and validation phases, the confusion matrix displaying the classification results of the proposed ensemble model are presented in Table 5 and Table 6. The assessment Score Rankings (ASR) and Assessment Score Distribution (ASD) of the proposed ensemble model for both the training (TRAIN) and validation (VALIDATE) phase are given in Tables 7,8,9 and 10. The CL curve basically reflects several statistics on the vertical axis for both groups of observations. Whole observations in the scored dataset are arranged through the posterior possibilities of the event level in descending order for a binary target whereas whole observations are arranged from the highest to the lowest expected profit for a nominal or ordinal target. Depth (deciles or groups) of the observations are shown on the horizontal axis of the CL curve.

In order to visually evaluate the performance of the proposed ensemble model, SAS Enterprise Miner 5.2 provides tools. Fig. 4 shows the Cumulative Lift (CL) curve of the model. The CL curve basically reflects several statistics on the vertical axis for both groups of observations. Whole observations in the scored dataset are arranged through the posterior possibilities of the event level in descending order for a binary target whereas whole observations are arranged from the highest to the lowest expected profit for a nominal or ordinal target. Depth (deciles or groups) of the observations are shown in the horizontal axis of the CL curve. Fig. 5 shows the score distribution graph for the training and validation phases. Some information about the proportions of events, nonevents and other values are plotted on the vertical axis of the score distribution graph. The model score of a bin is also presented on the horizontal axis. Based on each of the prediction of the target and the number of buckets used, the model score is different. Whole observations are categorized into the bin through the posterior possibilities of the event level and the number of buckets while for an interval targets, whole observations are categorized through the actual predicted values of the target. The score distribution graph displays information for higher model score by a higher percentage of events and as well as information for lower model scores by a higher percentage of nonevents. The classification table chart in Fig. 6 gives information through a stacked bar chart of the classification results for a categorical target variable. It provides information for both the training and validation phases in two groups: correct group and incorrect group. Total observations were also plotted on the vertical axis. The horizontal axis displays the target levels that observations actually belong to. The color of the stacked bars identifies the target levels that observations are classified into. The height of the stacked bars represents the percentage of total observations. The graph in Figure 5 is derived using the data in Table 7 and Table 8. Moreover, the graph in Figure 6 is derived using the data in Table 9 and Table 10.

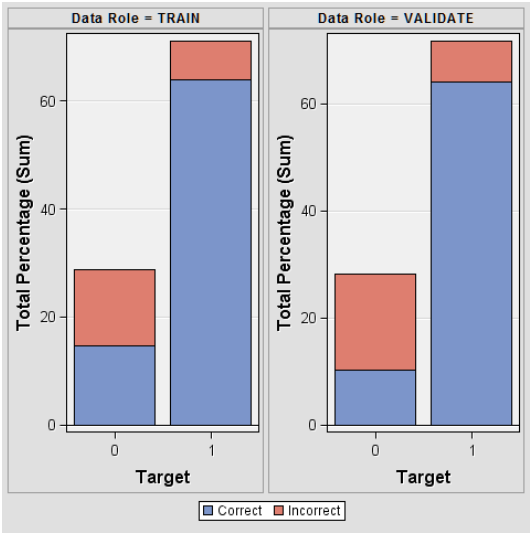

Fig. 6 Classification table chart of the ensemble mod

Table 4 The measured metrics for the proposed ensemble model (%).

Status	Specificity	Sensitivity	Precision	FPR	FNR	F_1	Accuracy
TRAIN	50.74	89.75	81.86	49.26	10.25	85.63	78.44
VALIDATE	36.36	89.28	78.12	63.64	10.72	83.33	74.35

Table 5 The classification table in the training phase.

Target	Outcome	Target Percentage (%)	Outcome Percentage (%)	Frequency Count	Total Percentage (%)
0	0	66.6667	50.7463	68	14.5923
1	0	33.3333	10.2410	34	7.2961
0	1	18.1319	49.2537	66	14.1631
1	1	81.8681	89.7590	298	63.9485

Table 6 The classification table in the validation phase.

Target	Outcome	Target Percentage (%)	Outcome Percentage (%)	Frequency Count	Total Percentage (%)
0	0	57.1429	36.3636	12	10.2564
1	0	42.8571	10.7143	9	7.6923
0	1	21.8750	63.6364	21	17.9487
1	1	78.1250	89.2857	75	64.1026

Table 7 Assessment Score Rankings: Data Role = TRAIN.

Depth	Gain	Lift	Cumulative Lift	Response %	Cumulative Response (%)	Number of Observations	Posterior Probability
5	**40.3614**	**1.40361**	**1.40361**	**100.000**	**100.000**	**24**	**0.99902**
10	**40.3614**	**1.40361**	**1.40361**	**100.000**	**100.000**	**23**	0.99658
15	**40.3614**	**1.40361**	**1.40361**	**100.000**	**100.000**	**23**	0.98944
20	**40.3614**	**1.40361**	**1.40361**	**100.000**	**100.000**	**24**	0.97062
25	36.7624	1.22053	1.36762	86.957	97.436	23	0.94812
30	35.3485	1.28156	1.35349	91.304	96.429	23	0.91373
35	32.6587	1.16968	1.32659	83.333	94.512	24	0.85215
40	30.6037	1.15951	1.30604	82.609	93.048	23	0.81745
45	29.6672	1.22053	1.29667	86.957	92.381	23	0.79243
50	27.1084	1.03745	1.27108	73.913	90.558	23	0.75878
55	24.5230	0.99423	1.24523	70.833	88.716	24	0.72203
60	22.3150	0.97643	1.22315	69.565	87.143	23	0.67685
65	20.4422	0.97643	1.20442	69.565	85.809	23	0.62996
70	17.1825	0.76029	1.17182	54.167	83.486	24	0.59097
75	16.2995	1.03745	1.16299	73.913	82.857	23	0.54676
80	13.6439	0.73232	1.13644	52.174	80.965	23	0.50566
85	9.6021	0.46787	1.09602	33.333	78.086	24	0.45883
90	7.2762	0.67129	1.07276	47.826	76.429	23	0.41280
95	4.5582	0.54924	1.04558	39.130	74.492	23	0.34290
100	0.0000	0.12205	1.00000	8.696	71.245	23	0.20603

Table 8 Assessment Score Rankings: Data Role = VALIDATE.

Depth	Gain	Lift	Cumulative Lift	Response %	Cumulative Response (%)	Number of Observations	Posterior Probability
5	**39.2857**	**1.39286**	**1.39286**	**100.000**	**100.000**	**6**	**0.99883**
10	**39.2857**	**1.39286**	**1.39286**	**100.000**	**100.000**	**6**	0.99630
15	**39.2857**	**1.39286**	**1.39286**	**100.000**	**100.000**	**6**	0.98984
20	33.4821	1.16071	1.33482	83.333	95.833	6	0.95365
25	30.0000	1.16071	1.30000	83.333	93.333	6	0.91130
30	31.5476	1.39286	1.31548	100.000	94.444	6	0.86628
35	32.4913	1.39286	1.32491	100.000	95.122	5	0.83226
40	30.3951	1.16071	1.30395	83.333	93.617	6	0.80943
45	26.1456	0.92857	1.26146	66.667	90.566	6	0.79469
50	22.7603	0.92857	1.22760	66.667	88.136	6	0.76931
55	17.8571	0.69643	1.17857	50.000	84.615	6	0.74305
60	15.7445	0.92857	1.15744	66.667	83.099	6	0.70863
65	15.7699	1.16071	1.15770	83.333	83.117	6	0.65853
70	15.5052	1.11429	1.15505	80.000	82.927	5	0.60820
75	12.3782	0.69643	1.12378	50.000	80.682	6	0.56876
80	8.1687	0.46429	1.08169	33.333	77.660	6	0.53291
85	7.2500	0.92857	1.07250	66.667	77.000	6	0.46990
90	3.8073	0.46429	1.03807	33.333	74.528	6	0.42485
95	0.7334	0.46429	1.00733	33.333	72.321	6	0.34218
100	0.0000	0.83571	1.00000	60.000	71.795	5	0.19658

Table 9 Assessment Score Distribution: Data Role = TRAIN.

Posterior Probability Range	Number of Events	Number of Nonevents	Mean Posterior Probability	Percentage (%)
0.95-1.00	**102**	**2**	**0.98554**	**22.3176**
0.90-0.95	28	3	0.93045	6.6524
0.85-0.90	15	1	0.87278	3.4335
0.80-0.85	35	7	0.82240	9.0129
0.75-0.80	27	6	0.77725	7.0815
0.70-0.75	21	11	0.72619	6.8670
0.65-0.70	15	6	0.67704	4.5064
0.60-0.65	20	12	0.62570	6.8670
0.55-0.60	16	9	0.57347	5.3648
0.50-0.55	19	9	0.52672	6.0086
0.45-0.50	10	17	0.47316	5.7940
0.40-0.45	11	14	0.42342	5.3648
0.35-0.40	5	11	0.37311	3.4335
0.30-0.35	4	3	0.33302	1.5021
0.25-0.30	2	7	0.27430	1.9313
0.20-0.25	0	7	0.22343	1.5021
0.15-0.20	1	7	0.17850	1.7167
0.10-0.15	1	2	0.14307	0.6438

Table 10 Assessment Score Distribution: Data Role = VALIDATE.

Posterior Probability Range	Number of Events	Number of Nonevents	Mean Posterior Probability	Percentage (%)
0.95-1.00	**21**	**1**	**0.98904**	**18.8034**
0.90-0.95	7	0	0.92098	5.9829
0.85-0.90	6	1	0.87022	5.9829
0.80-0.85	10	1	0.81981	9.4017
0.75-0.80	8	4	0.78200	10.2564
0.70-0.75	5	4	0.73548	7.6923
0.65-0.70	5	1	0.69020	5.1282
0.60-0.65	5	2	0.62046	5.9829
0.55-0.60	4	2	0.57760	5.1282
0.50-0.55	4	5	0.53048	7.6923
0.45-0.50	0	1	0.45009	0.8547
0.40-0.45	4	6	0.42875	8.5470
0.35-0.40	1	1	0.36841	1.7094
0.30-0.35	1	1	0.31141	1.7094
0.25-0.30	1	1	0.29091	1.7094
0.20-0.25	1	0	0.24651	0.8547
0.15-0.20	0	2	0.16140	1.7094
0.10-0.15	1	0	0.12468	0.8547

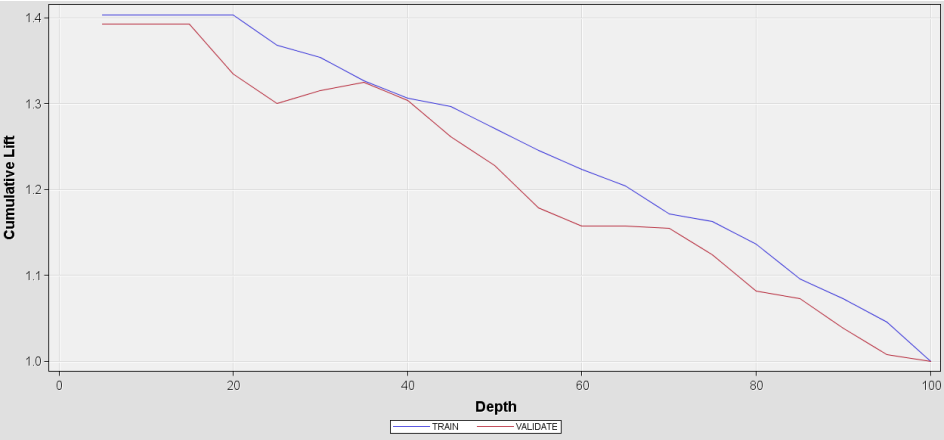

Fig. 4 Cumulative lift curve of the ensemble model.

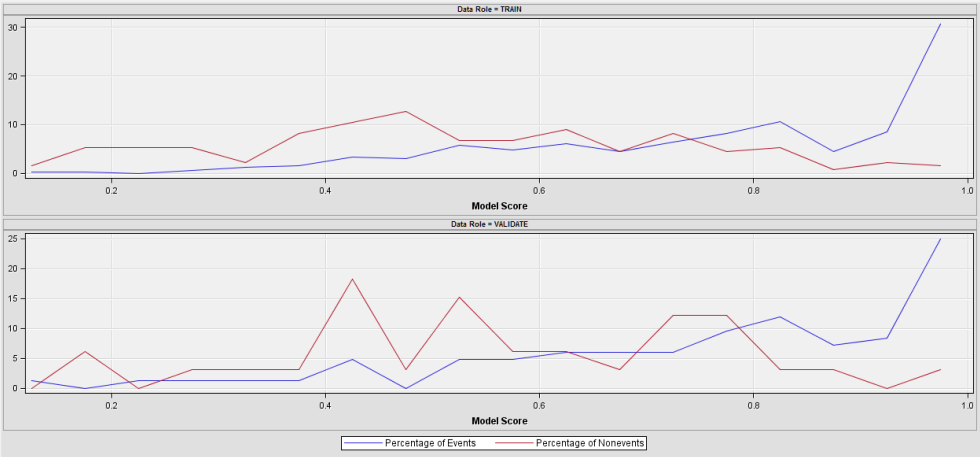

Fig. 5 Score distribution graph of the ensemble model.

6. CONCLUSION

In the last couple of decades, even though classification techniques like ANN widely were investigated, ensemble-based approaches started to become new popular for various predictions. But there are limited studies, which investigates the using of ensemble-based approaches for disease prediction. In this respect, this paper focused on this topic and proposed a new neural network ensemble model to predict liver diseases. The proposed approach was implemented in SAS software analytics platform, and its accuracy was investigated using the Indian Liver Patient Dataset, which is a well-known dataset for liver diseases. The results obtained in the classification applied with the proposed approach are given in detail in tables. And the graphs derived from these values in the tables are presented as figures. It is seen that the performance of the classification results made from the data with high number of features continues to increase. The obtained results are very promising since in the validation phase the proposed model's classification accuracy was 74.35% and the model achieved the rates 36.36%, 89.28%, 78.12%, 63.64%, 10.72% and 83.33% in terms of specificity, sensitivity, precision, false-positive rate, false-negative rate and F1(F-measure). Moreover, the proposed model's suitability is shown with the cumulative lift chart, the score distribution graph and the classification bar chart.

REFERENCES

[1] Liver Disease, National Institute of Diabetes and Digestive and Kidney Disease (NIDDK), https://www.niddk.nih.gov/, Accessed, 2019.
[2] LDCR: Liver Disease in Canada Report, http://www.liver.ca/liver-health/. Accessed: 20 April 2019.
[3] Zatoński, W.A., Sulkowska, U., Mańczuk, M., et al. Liver cirrhosis mortality in Europe, with special attention to Central and Eastern Europe. Eur. Addict. Res. 16:193-201, 2010.
[4] HEPAMAP. A roadmap for hepatology research in Europe: An overview for policy makers. Available from: http://epha.org/IMG/pdf/EASL_policy_makers_summary.pdf. Accessed: April 2018.
[5] American Liver Foundation. The Liver Lowdown – Liver Disease: the big picture. Available from: http://www.liverfoundation.org/ Accessed: April 2018.
[6] UCI, ILPD (Indian Liver Patient Dataset), https://archive.ics.uci.edu/ml/datasets/ILPD [Accessed April 26, 2016].
[7] Abdar, M., A Survey and Compare the Performance of IBM SPSS Modeler and Rapid Miner Software for Predicting Liver Disease by Using Various Data Mining Algorithms. Cumhuriyet Science Journal, 36(3): 3230-3241, 2015.
[8] Bendi Venkata Ramana, M. S. Prasad Babu and N. B. Venkateswarlu, A Critical Comparative Study of Liver Patients from USA and INDIA: An Exploratory Analysis, International Journal of Computer Science Issues, ISSN: 1694-0784, May 2012.
[9] Netzer, M., Millonig, G., Osl, M., Pfeifer, B., Praun, S., Villinger, J., Vogel, W. and Baumgartner, C., A new ensemble-based algorithm for identifying breath gas marker candidates in liver disease using ion molecule reaction mass spectrometry. Bioin. 25(7): 941-947, 2009.
[10] Das, R., Turkoglu, I. and Sengur, A., Effective diagnosis of heart disease through neural networks ensembles. Exp. Syst. Appl. 36(4): 7675-7680, 2009. doi: 10.1016/j.eswa.2008.09.013.
[11] Das, R., Turkoglu, I. and Sengur, A., Diagnosis of valvular heart disease through neural networks ensembles. Com. Meth. Pro. Bio. 93(2): 185-191, 2009. doi: 10.1016/j.cmpb.2008.09.005.
[12] I. Valavanis, S. G. Mougiakakou, K. S. Nikita, and A. Nikita, "Computer aided diagnosis of CT focal liver lesions by an ensemble of neural network and statistical classifiers," in *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541)*, Jul. 2004, vol. 3, pp. 1929–1934 vol.3, doi: 10.1109/IJCNN.2004.1380907
[13] Acharya, U.R., Sree, S.V., Ribeiro, R., Krishnamurthi, G., Marinho, R.T., Sanches, J. and Suri, J.S., Data mining framework for fatty liver disease classification in ultrasound: a hybrid feature extraction paradigm. Med. Phys. 39(7): 4255-4264, 2012. doi: 10.1118/1.4725759.
[14] Nanyue, W., Youhua, Y., Dawei, H., Bin, X., Jia, L., Tongda, L., Liyuan, X., Zengyu, S., Yanping, C. and Jia, W., Pulse Diagnosis Signals Analysis of Fatty Liver Disease and Cirrhosis Patients by Using Machine Learning. Scie. Wor. J. 2015. doi: 10.1155/2015/859192.
[15] Pérez-Ortiz, M., Gutiérrez, P.A., Hervás-Martínez, C., Briceño, J. and de la Mata, M.: June. An ensemble approach for ordinal threshold models applied to liver transplantation. In Neural Networks (IJCNN), 2012.
[16] Zhang, H.Q., Li, J., Zheng, G., Jiang, M., Li, L., Tan, Y. and Lu, A.P.: December. Using text mining to understand traditional Chinese medicine pathogenesis of nonalcoholic fatty liver disease. In Bioinformatics and Biomedicine (BIBM), 2013 IEEE International Conference on, 2013.

[17] Aneeshkumar, A.S. and Venkateswaran, C.J., Reverse Sequential Covering Algorithm for Medical Data Mining. Proc. Comp. Scie. 47: 109-117, 2015. doi: 10.1016/j.procs.2015.03.189.
[18] Wang, H., Ma, J., Wu, Z. and Xiao, C.: A medical support treatment model based on data mining. In Computer Science & Education (ICCSE), 2015.
[19] Jung, H., Chung, K.Y. and Lee, Y.H., Decision supporting method for chronic disease patients based on mining frequent pattern tree. Mul. Tool. Appl. 74(20): 8979-8991, 2015. doi: 10.1007/s11042-013-1730-3.
[20] Weng, C.H., Huang, T.C.K. and Han, R.P., Disease prediction with different types of neural network classifiers. Telematics and Informatics. 33(2): 277-292, 2016. doi: 10.1016/j.tele.2015.08.006.
[21] Nawi, N.M., Khan, A. and Rehman, M.Z.: A new back-propagation neural network optimized with cuckoo search algorithm. In Computational Science and Its Applications–ICCSA, 2013.
[22] Corley CD, Pullum LL, Hartley DM, Benedum C, Noonan C, Rabinowitz PM, et al. (2014) Disease Prediction Models and Operational Readiness. PLoS ONE 9(3): e91989. doi: 10.1371/journal.pone.0091989
[23] Karimpouli, S., Fathianpour, N. and Roohi, J., A new approach to improve neural networks' algorithm in permeability prediction of petroleum reservoirs using supervised committee machine neural network (SCMNN). J. Pet. Sci. Eng. 73(3): 227-232, 2010. doi: 10.1016/j.petrol.2010.07.003.
[24] Qiakai, N.I., Chao, G. and Jing, Y.: Research of face image recognition based on probabilistic neural networks. In Control and Decision Conference (CCDC), 2012.
[25] Shah, H., Ghazali, R. and Nawi, N.M., Using artificial bee colony algorithm for MLP training on earthquake time series data prediction. arXiv preprint arXiv:1112.4628, 2011.
[26] De Souza, F.T., A data-based model to locate mass movements triggered by seismic events in Sichuan, China. Environ. Monit. Assess. 186(1): 575-587, 2014. doi: 10.1007/s10661-013-3400-3.
[27] Maqsood, I., Khan, M.R. and Abraham, A., An ensemble of neural networks for weather forecasting. Neural. Comput. Applic. 13(2): 112-122, 2004. doi:10.1007/s00521-004-0413-4.
[28] Hanbay, D., Turkoglu, I. and Demir, Y., An expert system based on wavelet decomposition and neural network for modeling Chua's circuit. Exp. Syst. Appl. 34(4): 2278-2283, 2008. doi:10.1016/j.eswa.2007.03.002.
[29] Sołtys, M., Jaroszewicz, S. and Rzepakowski, P., Ensemble methods for uplift modeling. Data Min Knowl. Disc. 29(6): 1531-1559, 2015.
[30] Mitchell, T.M., Machine learning. 1997. Burr Ridge, IL: McGraw Hill, 45: 995, 1997.
[31] Breiman, L., Bagging predictors. Mach. Lear. 24(2): 123-140, 1996.
[32] Hinde, A., Sahu, A., Apley, D. and Runger, G., Preimages for variation patterns from kernel PCA and bagging. IIE Tran. 46(5): 429-456, 2014.
[33] Opitz, D. and Maclin, R., Popular ensemble methods: An empirical study. J. Arti. Intel. Research, 169-198, 1999.
[34] Dietterich, T.G., An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. Mach. Lear. 40(2): 139-157, 2000.
[35] Hoeting, J.A., Madigan, D., Raftery, A.E. and Volinsky, C.T., Bayesian model averaging: a tutorial. Statistical science, 382-401, 1999. doi:10.2307/2676803.
[36] Monteith, K., Carroll, J.L., Seppi, K. and Martinez, T.: Turning Bayesian model averaging into Bayesian model combination. In Neural Networks (IJCNN), 2011.
[37] Bensusan, H. and Giraud-Carrier, C.: Discovering task neighbourhoods through landmark learning performances. In Principles of Data Mining and Knowledge Discovery, 2000.
[38] Ozay, M. and Vural, F.T.Y., A new fuzzy stacked generalization technique and analysis of its performance. arXiv preprint arXiv:1204.0171, 2012.
[39] Ozcift, A. and Gulten, A., A robust multi-class feature selection strategy based on rotation forest ensemble algorithm for diagnosis of erythemato-squamous diseases. J. Med. Syst. 36(2): 941-949. doi: 10.1007/s10916-010-9558-0.
[40] Abdar, M., Zomorodi-Moghadam, M., Das, R., IH Ting, "Performance analysis of classification algorithms on early detection of Liver disease", Expert Systems with Applications 67 (1), 239-251, 2017.
[41] Rokach, L., Ensemble-based classifiers. Artif. Intell. Rev. 33(1-2): 1-39, 2010. doi: 0.1007/s10462-009-9124-7.
[42] Bühlmann, Peter. "Bagging, boosting and ensemble methods." In Handbook of Computational Statistics, pp. 985-1022. Springer Berlin Heidelberg, 2012.
[43] Das, R. and Sengur, A., Evaluation of ensemble methods for diagnosing of valvular heart disease. Exp. Syst. Appl .37(7): 5110-5115, 2010. doi:10.1016/j.eswa.2009.12.085.
[44] http://support.sas.com/documentation , [last accessed 01 August, 2019].

Author

Bihter Das graduated B.S. and M.S. degrees from the Department of Computer Science at the Fırat University in 2004 and 2007 respectively. Then she completed his Ph.D. degree at the Department of Software Engineering at the same university in 2018. She also worked between September 2017 and June 2018 as a visiting scholar at the Department of Computing Science at the University of Alberta, Edmonton, Canada. Her current research areas include bioinformatic, Digital Signal Processing, Genome Analysis, Classification.