Journal of Biomedical Informatics 59 (2016) 185–200

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

IntelliHealth: A medical decision support application using a novel weighted multi-layer classifier ensemble framework

Saba Bashir, Usman Qamar *, Farhan Hassan Khan

Computer Engineering Department, College of Electrical and Mechanical Engineering, National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan

ARTICLE INFO

Article history:
Received 5 August 2015
Revised 1 November 2015
Accepted 6 December 2015
Available online 15 December 2015

Keywords:
Machine learning
Classification
Bagging
Disease prediction
Multi-layer
Ensemble technique

ABSTRACT

Accuracy plays a vital role in the medical field as it concerns with the life of an individual. Extensive research has been conducted on disease classification and prediction using machine learning techniques. However, there is no agreement on which classifier produces the best results. A specific classifier may be better than others for a specific dataset, but another classifier could perform better for some other dataset. Ensemble of classifiers has been proved to be an effective way to improve classification accuracy. In this research we present an ensemble framework with multi-layer classification using enhanced bagging and optimized weighting. The proposed model called "HM-BagMoov" overcomes the limitations of conventional performance bottlenecks by utilizing an ensemble of seven heterogeneous classifiers. The framework is evaluated on five different heart disease datasets, four breast cancer datasets, two diabetes datasets, two liver disease datasets and one hepatitis dataset obtained from public repositories. The analysis of the results show that ensemble framework achieved the highest accuracy, sensitivity and *F*-Measure when compared with individual classifiers for all the diseases. In addition to this, the ensemble framework also achieved the highest accuracy when compared with the state of the art techniques. An application named "IntelliHealth" is also developed based on proposed model that may be used by hospitals/doctors for diagnostic advice.

1. Introduction

Data mining in the medical domain collects data from the past experiences and analyzes them to identify trends and solutions to the present situations [1]. It is an efficient analytical methodology for detecting unknown and valuable information, from large volume of medical data [2]. Data mining techniques can be used for development of predicative models that enable classification and prediction. Learning is the process of building a scientific model after discovering knowledge from data. Which brings us to the concept of machine learning, which can be formally defined as "the complex computation process of automatic pattern recognition and intelligent decision making based on training sample data" [3]. Depending on the availability of the data, machine learning classifiers are categorized into supervised learning and unsupervised learning [3]. In supervised learning, training data is available and a learning model is trained. Popular methods include Artificial Neural Network (ANN), Support Vector Machine (SVM), and Decision Trees. In unsupervised learning, no label is given in sample data. Examples include *k*-means clustering and self-organization map. An ensemble approach combines the results of single classification techniques and results in better performance as compared to a single classifier [4,5]. Multiple techniques are utilized for constructing the ensemble and each results in different diagnosis accuracy. Bagging [6], boosting [7] and stacking [8] are the most common ensemble approaches.

In the last few decades, multiple machine learning techniques have been widely used for classification and prediction of heart disease [9], breast cancer [10], diabetes [11], liver disease [12] and hepatitis [13]. The comparison of these techniques is shown in Tables 1–5.

It can be seen from the above tables that extensive research has been conducted in this field. However, there is no agreement on which methodology is better than others: one classifier could perform better than others in respect of a given dataset or a specific disease, while a further approach could outperform the others when dealing with a different dataset or a disease.

This uncertainty about which model represents an optimal solution has been overcome in this research by introducing a novel ensemble approach named "HM-BagMoov" which is able to exploit the potentials of several classifiers.

* Corresponding author. Tel.: +92 3235577710.
E-mail addresses: saba.bashir@ceme.nust.edu.pk (S. Bashir), usmanq@ceme.nust.edu.pk (U. Qamar), farhan.hassan@ceme.nust.edu.pk (F.H. Khan).

http://dx.doi.org/10.1016/j.jbi.2015.12.001
1532-0464/

Table 1
Machine learning techniques used on heart disease datasets.

Author/reference	Year	Technique	Accuracy (%)
Thenmozhi et al. [14]	2014	*K*-Mean based on MAFIA	74
Chitra et al. [15]	2013	Cascaded Neural Network	84
		Support Vector Machine	82
Shouman et al. [16]	2013	Gain ratio Decision Tree	79.1
		Naïve Bayes	83.5
		K Nearest Neighbor	83.2
Shouman et al. [17]	2012	*K* mean clustering with Naïve Bayes algorithm	78.62
Ghumbre et al. [18]	2011	Support Vector Machine	84.05
		Radial Basis Function	82.24

Table 2
Machine learning techniques used on breast cancer datasets.

Author/reference	Year	Technique	Accuracy (%)
Chaurasia et al. [19]	2014	Sequential Minimal Optimization	94.2
Ashfaq Ahmed et. al. [20]	2013	Support Vector Machine	75.00
		Random forest	75.00
Salama et al. [21]	2012	SMO + J48 + MLP + IBK	77.31
Lavanya et al. [22]	2012	Hybrid approach	93.96
Lavanya et al. [23]	2011	CART with feature selection	94.56

Table 3
Machine learning techniques used on diabetes datasets.

Author/ reference	Year	Technique	Accuracy (%)
Gandhi et al. [24]	2014	*F*-score feature selection + SVM	75
Stahl et al. [25]	2014	Sliding Window Bayesian Model Averaging	61
Aslam et al. [26]	2013	Genetic programming with comparative partner selection	79
Nirmala Devi et al. [27]	2013	*K* nearest neighbor	87.4
Christobel et al. [28]	2012	*K* nearest neighbor with imputation and scaling	73.38
Zolfaghari [29]	2012	BP Neural Network and SVM	88
Lee [30]	2011	Fuzzy diabetes ontology	75

Table 4
Machine learning techniques used on liver disease datasets.

Author/reference	Year	Technique	Accuracy (%)
Vijayarani et al. [31]	2015	Support vector machine	61.2
Jin et al. [32]	2014	Naïve Bayes	53.9
		Decision tree	69.4
		Multi layer perception	67.9
		K nearest neighbor	65.3
Sugawara et al. [33]	2013	Self-Organizing Map	66
Kumar et al. [34]	2013	Rule based classification model	72
Ramana et al. [35]	2012	NN + CFS	73
		MLP	74
		J48	73
		Simple CART	69
		K star	68

Table 5
Machine learning techniques used on hepatitis datasets.

Author/reference	Year	Technique	Accuracy (%)
Pushpalatha et al. [36]	2014	Model framework	80
El Houby [37]	2013	Associative classification	80
Karthikeyan et al. [38]	2013	Naïve Bayes	84
		J48	83
		Random forest	83
Kaya et al. [39]	2013	Rough set-ELM	86.4
Kumar et al. [40]	2012	Support vector machine	79.33

The main contributions of this paper are summarized as follows:

- An optimal combination of classifiers is presented which is comprised of Naïve Bayes, Linear Regression, Quadratic Discriminant Analysis, Instance Based Learner, Support Vector Machine, Artificial Neural Network and Random Forest.
- The ensemble approach uses bagging with multi-objective optimized weighted vote based technique and multi-layer classification.
- We compare our approach with existing state of the art techniques to prove the superiority of our proposed ensemble.
- An application named "IntelliHealth" has been developed based on the proposed ensemble model that may be used by hospitals/doctors for diagnostic advice.

The proposed HM-BagMoov ensemble approach is an enhanced version of our BagMoov ensemble approach [55]. However the major difference between the two techniques is that BagMoov ensemble approach follows the taxonomy of flat structure where all base classifiers are arranged in single layer with bagging. HM-BagMoov is a multi-layer classification scheme to further enhance the prediction. In HM-BagMoov, the number of classifiers has been increased from 5 to 7 as well as utilization of enhanced bagging and optimized weighting. We have also introduced feature selection, missing value imputation, noise removal and outlier detection to further improve the accuracy.

The rest of the paper is organized as follows: Section 2 presents the proposed ensemble framework. Section 3 provides the results and discussion from the experiments carried out. Section 4 provides details of the IntelliHealth application. Finally, conclusions and recommendations for future work are summarized in Section 5.

2. HM-BagMoov ensemble framework

The proposed framework involves data acquisition, pre-processing, classifier training and HM-BagMoov (Hierarchical Multi-level classifiers Bagging with Multi-objective optimized voting) ensemble model for disease prediction based on three layered approach.

2.1. Data acquisition and pre-processing module

Data acquisition and pre-processing module includes feature selection, missing value imputation, noise removal and outlier detection.

• *F-Score feature selection:* The proposed framework uses *F*-Score feature selection method to select the most important and related features from medical datasets. *F*-Score is a simple method that distinguishes the two classes with real values. Following equation is used to calculate the *F*-Score of a feature in given dataset:

$$F(i) = \frac{\left(X_i^{\prime(+)} - X_i^{\prime}\right)^2 + \left(X_i^{\prime(-)} - X_i^{\prime}\right)^2}{\frac{1}{n_+ - 1}\sum_{k=1}^{n_+}\left(X_{k,i}^{(+)} - X_i^{\prime(+)}\right)^2 + \frac{1}{n_- - 1}\sum_{k=1}^{n_-}\left(X_{k,i}^{(-)} - X_i^{\prime(-)}\right)^2} \quad (1)$$

where the average of the *i*th feature of the whole, positive and negative datasets are denoted by X_i', $X_i'(+)$ and $X_i'(-)$ respectively. $X_{k,i}(+)$ is the *i*th feature of the *k*th positive instance whereas $X_{k,i}(-)$ is the *i*th feature of the *k*th negative instance. The numerator shows the

discrimination between the positive and negative sets, and the denominator defines the one within each of the two sets. The *F*-score value for each feature is calculated and then for selecting the appropriate features, a threshold value is used. It is obtained by calculating the average *F*-Score of all features in the dataset. The feature is added to the feature space which has *F*-Score value greater than threshold. For a given vector x_k, where $k = 1, \ldots, n$, and number of positive and negative instances are denoted by n_+ and n_-. The *F*-Score of *i*th feature is then calculated using Eq. (1). A large value of *F*-Score indicates that the feature is more discriminative.

• *Missing data imputation using KNN approach:* The KNN approach is used for missing data imputation. The proposed procedure is named as "KNNimpute". It is defined as, given a set of instances with incomplete pattern, the *K* closest cases with known attribute values are identified from the training cases for which the attribute values need to be determined. Once *K*-nearest neighbors are identified, the missing attribute values are then identified. An appropriate distance measure is used to determine the *K*-nearest neighbors and then to impute the missing value. The proposed missing data imputation technique uses heterogeneous Euclidean-overlap metric (HEOM) for distance measure between two variables. HEOM has a benefit that it can calculate different distance measures for different types of attributes. For instance, x_a and x_b are two variables and distance between them is denoted by $d(x_a, x_b)$. Then following formula is used to calculate the distance between them:

$$d(x_a, x_b) = \sqrt{\sum_{j=1}^{n} d_j(x_{aj}, x_{bj})^2} \tag{2}$$

where distance is determined for *j*th attribute. d_0 is an overlap distance function which assigns a value of 0 if qualitative features are same other $d = 1$. For example:

$$d_0(x_{aj}, x_{bj}) = \begin{cases} 0 & x_{aj} = x_{bj} \\ 1 & x_{aj} \neq x_{bj} \end{cases} \tag{3}$$

Consider *j*th attribute is missing for an instance *x* i.e. $m_j = 1$. The distance of *x* from all training instances is calculated and *K*-nearest neighbors are identified using the given notation:

$$V_x = \{V_k\}_{k=1}^{K} \tag{4}$$

where *k*-nearest neighbors of *x* are arranged in increasing distance order. For instance, v_1 is the closest neighbor of *x*. Once *K*-nearest neighbors are identified, the unknown value is determined by an estimate from *j*th feature values of V_x.

If the *j*th feature value is continuous value, then missing value is replaced by mean value of its *K* nearest neighbors. If the *j*th input feature is qualitative value then the missing value is imputed by determining the category of value using V_x. The mode of $\{V_k\}_{k=1}^{K}$ is imputed where same importance is given to all neighbors. A weighted method assigns a weight α_k to each V_k and closer neighbors are assigned greater weight. The grouping is performed using discrete value in the *j*th input feature. The distance weighted method for KNN classifier where α_k is calculated by:

$$\alpha_k(x) = \frac{d(V_k, x) - d(V_k, x)}{d(V_k, x) - d(V1, x)} \tag{5}$$

where $\alpha_k(x) = 1$ when all distances are equal. For imputation, consider the *j*th feature is qualitative having *S* possible discrete values. The imputed value is given by

$$S = \text{argmax}\ \{\alpha_x^s\} \tag{6}$$

where *s* is the possible category that is determined as imputed value. The missing value imputation method can be defined by flowchart given in Fig. 1.

• *Outlier detection and elimination:* We have used Grubb's test for outlier values that may exist in medical datasets. There are several reasons for outliers in disease data such as abnormal condition of patient or equipment malfunction or recording error. Grubb's test is also called ESD (Extreme Studentized Deviate) method. This method is beneficial as it considers complete dataset values for outlier detection and elimination. This test is performed as long as all outliers will be detected. It is a two-step process. First step is to calculate how far the outliers are from others. The ratio *G* is calculated in second step that is the difference between dataset value and mean divided by standard deviation. Following formula is used to determine outliers using Grubb's test.

$$G = \frac{\max_{i=1\ldots n}|x_i - \bar{x}|}{\sigma} \tag{7}$$

where x_i is the element of dataset, $\bar{x}$ is mean of dataset and σ is standard deviation of dataset. A statistical significance (α) level is used as threshold value. The Grubb's test calculated value is compared with α level threshold value. If the calculated value is higher or lower than significant level, then this value is considered as an outlier.

• *Noise removal:* The proposed pre-processing involves noise removal using clustering approach. It can also result in data reduction as irrelevant and noisy data is removed from medical dataset. The data is clustered into groups having similar characteristics. The centroid is calculated for each cluster and then maximum distance value is measured for each cluster. If this distance value is larger than specified threshold than it is considered as noise. Threshold value is a parameter provided by user and is measured using mean of all values within a cluster. We have used pair-wise distance in order to calculate Euclidean distance between pair of objects for *n* by *p* data matrix. The rows of *X* show instances or observations whereas columns indicate the variables or attributes. Following formula is used to measure Euclidean distance between pair of attributes for two instances:

$$d_{rs}^2 = (x_r - x_s)(x_r - x_s)' \tag{8}$$

where

$$\bar{x}_r = \frac{1}{n}\sum_j x_{rj} \tag{9}$$

and

$$\bar{x}_s = \frac{1}{n}\sum_j x_{rs} \tag{10}$$

Following steps are used to remove noise from disease dataset: Calculate pair-wise distance using Euclidean distance between pair of objects; Take square distance. Calculate maximum values from square distance values; Calculate threshold value for each cluster; If distance > threshold then the value is noise and removed from dataset.

2.2. Classifier training module

The classifier training module performs further computations on pre-processed data. Each individual classifier is trained using the training data in order to make them useful for disease prediction.

2.3. HM-BagMoov ensemble

In this section we explain the construction of the HM-BagMoov ensemble. A crucial step of most ensemble methods is the selection of the optimal set of models to be included in the ensemble. Two

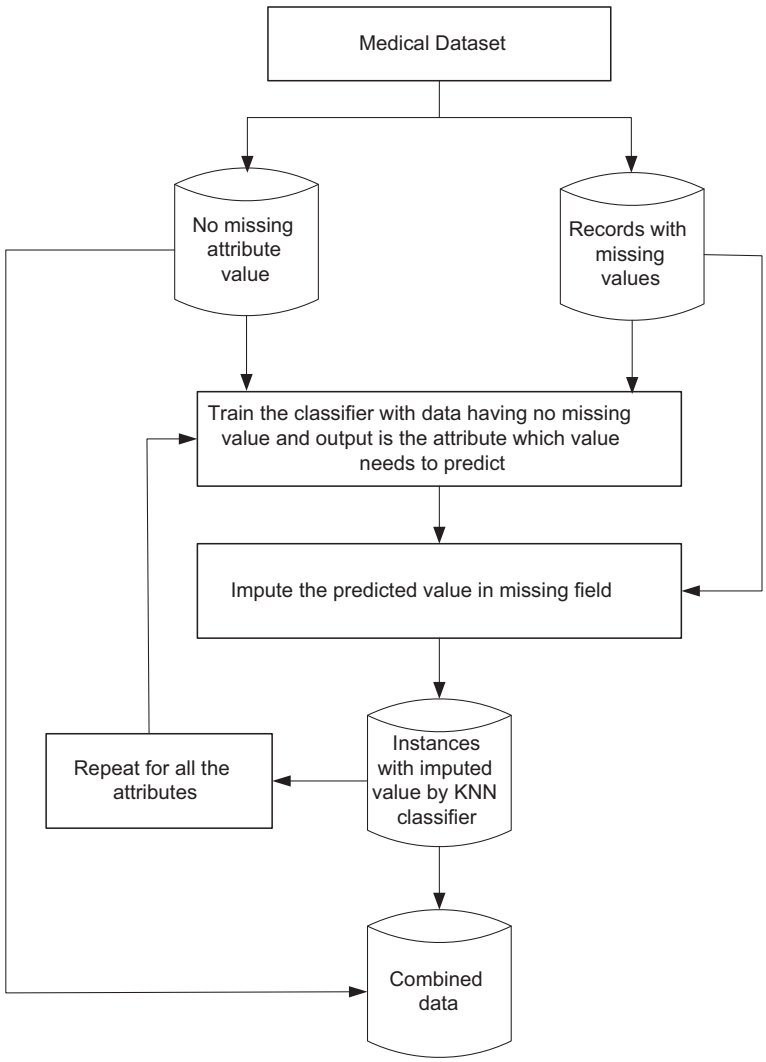


Fig. 1. Proposed missing data imputation method.

conditions that should be satisfied to achieve a good quality ensemble are accuracy and prediction diversity.

2.3.1. Multi-objective optimization weighted voting ensemble scheme

There are numerous combing models for ensembles such as Genetic Programming (GP), which is an evolutionary based optimization technique [76,77]. For example, authors in [76] proposed a classifier stacking based evolutionary ensemble method for predicting amino acid sequences associated with breast cancer. The base classifiers are stacked and GP is employed to generate the meta-classifier which combines the prediction of the classifiers. However, Unweighted [41] and Weighted Voting [42] are two of the most common methods for combining models. In unweighted voting also called majority voting, each model outputs a class value and the class with the most votes is the one proposed by the ensemble. For example, simple majority based ensemble approaches are developed for cancer prediction using amino acid sequences [74,75]. Ali et al. [74] proposed an 'IDM-PhyChm-Ens' classification method by utilizing homogeneous ensemble method where majority voting is used to combine the results.

Weighted voting scheme has an advantage over majority voting that in case of different predictions by each base classifier for an instance, the final prediction will be the classifier which has highest weight associated with it. Weighted voting technique determines the final prediction where weights can be assigned on the basis of classification accuracy. The classifier which has high accuracy will attain high weight and vice versa. The final prediction will be done based on highest weighted vote. The biasness of accuracy results can be generated if we have biased dataset in case of unbalanced classes. An unbiased metric is needed in order to assign the weights to base classifiers instead of accuracy measure. Thus we propose using an enhanced version of weighted vote based ensemble framework based on *F*-Measure (Precision + Recall).

The precision is defined as follows:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \tag{11}$$

where true positives are correctly classified actual positives and false positives are actual negatives that are incorrectly labeled as positives.

Recall is defined as follows:

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{12}$$

where false negatives are incorrectly labeled data as negatives that are actual positives.

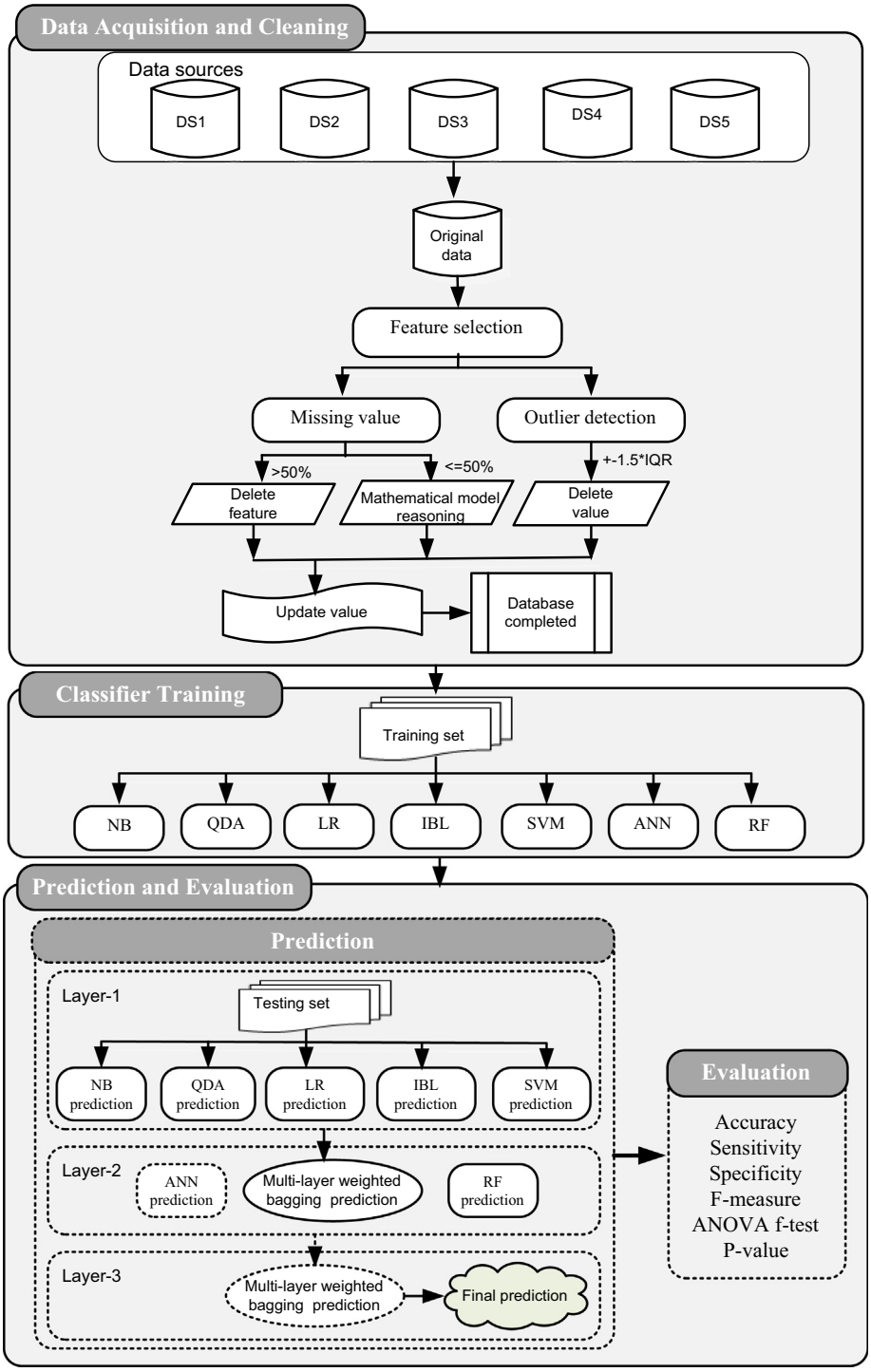


Fig. 2. Detailed architecture of proposed ensemble framework.

The *F*-Measure is the weighted average of Recall and Precision and is represented by:

$$F\text{-Measure} = 2 * \frac{\text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}} \quad (13)$$

HM-BagMoov therefore, will use multi-objective optimized weighting scheme where the objective functions will be precision and recall. Formally, it can be stated as: Find the number of vectors V_k where $V_k = \{v_1, v_2, v_3, \ldots, v_n\}$ for each classifier C_k and $C_k = \{C_1, C_2, C_3, \ldots, C_n\}$ such that simultaneously optimize the N objective criteria, while satisfying the constraints, if any.

2.3.2. Enhanced bagging scheme

Bagging is a "Bootstrap" ensemble method and also termed as Bootstrap Aggregation. Bootstrap Aggregating (bagging) aims to sample data sets for an ensemble of classifiers.

Bagging methods form a class of algorithms which build several instances of a black-box estimator on random subsets of the original training set and then aggregate their individual predictions to

form a final prediction. These methods are used as a way to reduce the variance of a base estimator, by introducing randomization into its construction procedure and then making an ensemble out of it [73]. As Bagging ensemble provides a way to reduce overfitting, bagging methods work best with strong and complex models (e.g., fully developed decision trees, Naïve Bayes, SVM, etc.), in contrast with boosting methods and other ensemble techniques [73].

We have applied an enhanced bagging scheme to the HM-BagMoov ensemble. At first stage, original training set for each dataset is divided into multiple bootstrap sets with replacement. In order to create bootstrap samples from training set of size m, t multinomial trails are performed, where one of the m examples is drawn for each trial. The probability of each example to be drawn in each trial is $1/m$. The proposed ensemble chooses a sample r from 1 to m and the rth training example is added to bootstrap training set S. Moreover, it is possible that some of the training examples will not be selected in bootstrap training sets whereas others may be chosen one time or more. At second stage, classifiers' training is performed using bootstrap training sets generated during the first stage. We also state that each classifier in the bagging approach should not have the same weight as each classifier has a different individual performance level. The HM-BagMoov ensemble will return a function $h(d)$ that classifies new samples into class y having the highest weight from the base models h_1, h_2, h_3, ..., h_t.

2.3.3. Selection of classifiers

The ensemble frameworks are classified into two categories [43,44]: homogeneous ensemble frameworks and heterogeneous ensemble frameworks. The homogenous ensemble framework uses base classifiers of same type whereas heterogeneous ensemble framework uses base classifiers of different types. The basic idea behind ensemble classifiers is that they perform better than their components when the base classifiers are not identical. A necessary condition for the ensemble approach to be useful is that member classifiers should have a substantial level of disagreement, i.e., they make error independently with respect to one another. The limitations of homogeneous ensemble frameworks can be removed by using heterogeneous ensemble frameworks. For heterogeneous ensembles, ensemble creation is usually a 2-phase process (sometimes called overproduce & select): many different base models are generated by running different learning algorithms on the training data, then the generated models are combined to form the ensemble. Multiple studies show that the strength of heterogeneous ensemble is related to the performance of the base classifiers and the lack of correlation between them (model diversity). We have also introduced multi-layer classification to further enhance the prediction. In order to keep the ensemble computationally less expensive we have used a three layer approach. The detailed architecture of proposed ensemble model is shown in Fig. 2.

The choice of base classifiers is based on diversity where those classifiers are selected from the literature recently used for disease classification and prediction in medical domain. We have performed literature survey of almost 200 research papers and selected classifiers which tend to produce the high classification and prediction accuracy. Table 6 shows the literature review of machine learning techniques that have consistently produced high accuracy. These classifiers become our base classifiers for building the proposed ensemble.

2.3.3.1. Layer-1 classifiers. The diversity among classifiers is achieved by selecting entirely different type of classifiers. Naïve Bayes (NB) is probability based classifier which has high classification and prediction accuracy [45]. Linear Regression (LR) is linear classifier which has the advantages of simplicity, interpretability, scientific acceptance and widespread availability [46]. Quadratic Discriminant Analysis (QDA) is a common tool for classification which has quadratic decision surface and requires no parameters to tune the algorithm [47]. It is inherently multiclass and proven to work well in practice.

The individual Naïve Bayes classifier considers each attribute independently without taking into account the relation between them whereas the proposed ensemble model can handle dependency and relation between given attribute set. The linear regression classifier and quadratic discriminant analysis are used in proposed model in order to perform correlation analysis between attribute sets. Thus ensemble model resolves the limitation of individual Naïve Bayes classifier by handling correlation.

Instance Based Learner (IBL) classifier is distance based classifier which does not build model explicitly [48]. It may simply store a new instance or throw an old instance away. Instance based learner avoids the risk of overfitting to noise in the training set.

Table 6
Literature review of selected machine learning techniques along with accuracy range.

Machine learning techniques	Papers referenced	Accuracy (%)
Naïve Bayes	Shouman et al. [16]	83.5
	Pattekari and Parveen [78]	80
	Shouman et al. [17]	78.62
	Shouman et al. [16]	83.5
	Peter and Somasundaram [53]	80
	Ramana et al. [66]	56
	Karthik et al. [67]	55
Support Vector Machine	Ghumbre et al. [18]	85.05
	Chitra et al. [15]	82
	Ashfaq Ahmed et. al. [20]	75
	Aruna et al. [79]	96.84
	Christobel et al. [28]	96.84
	Zolfaghari et al. [29]	88
	Ramana et al. [35]	58
	Ramana et al. [66]	82
	Kousarrizi et al. [80]	98
K Nearest Neighbor	Shouman et al. [16]	83.2
	Christobel et al.[28]	71.9
	Ramana et al. [35]	63
	Chen et al. [2]	72
	Rajkumar et al. [81]	45.67
	Polat et al. [82]	87
Linear Regression	Kurt et al. [83]	75.3
	Liao et al. [84]	97.5
	Ma et al. [85]	75
	Dubberke et al. [86]	75
	William et al. [87]	80
	Wilson et al. [88]	85
Quadratic Discriminant Analysis	Ster et al. [89]	96.80
	Georgiou-Karistianis et al. [90]	80
	Zhang et al. [91]	78
	Maroco et al. [92]	80
	Drent et al. [93]	90
	Srivastava et al. [94]	85.7
Artificial neural network	Das et al. [95]	89.01
	Chitra et al. [15]	85
	Chen et al. [96]	80
	Chunekar et al. [97]	70.72
	Kahramanli et al. [98]	84.2
	Ramana et al. [35]	73
	Chen et al. [96]	80
	Das et al. [95]	89.01
Random Forest	Tu et al. [99]	78.9
	Shouman et al. [16]	79.1
	Ashfaq Ahmed et al. [20]	75
	Tu et al. [99]	81.41
	Shouman et al. [17]	84.1

Table 7
Comparison of HM-BagMoov ensemble framework with individual classifiers for heart disease.

Classifiers	Cleveland Dataset				Eric Dataset			
	Acc (%)	Sen (%)	Spec (%)	*F*-M (%)	Acc (%)	Sen (%)	Spec (%)	*F*-M (%)
NB	77.23	81.71	71.94	76.51	68.90	77.78	57.61	66.19
SVM	80.86	93.90	65.47	77.15	78.47	89.74	64.13	74.81
LR	83.50	88.41	77.70	82.71	77.99	88.89	64.13	74.51
QDA	65.68	68.29	62.59	65.32	46.41	10.26	92.39	18.46
kNN	64.36	68.90	58.99	63.56	65.55	68.38	61.96	65.01
RF	69.64	84.76	51.80	64.30	69.86	90.60	43.48	58.76
ANN	79.21	79.88	78.42	79.14	76.08	80.34	70.65	75.19
HM-BagMoov	**86.20**	**94.12**	**76.32**	**84.29**	**83.82**	**91.74**	**73.74**	**81.76**
	SPECT Dataset				SPECTF Dataset			
NB	80.52	76.36	81.60	78.90	78.28	23.64	92.45	37.65
SVM	67.04	85.45	62.26	72.04	79.40	0.00	100	0.00
LR	83.15	38.18	94.81	54.44	78.28	9.09	96.23	16.61
QDA	83.52	36.36	95.75	52.71	20.60	100	0.00	0.00
kNN	79.40	7.27	98.11	13.54	71.91	36.36	81.13	50.22
RF	79.40	0.00	100	0.00	79.40	0.00	100	0.00
ANN	79.78	50.91	87.26	64.30	78.65	50.91	85.85	63.92
HM-BagMoov	**84.77**	**34.73**	**97.75**	**51.25**	**83.03**	**7.45**	**99.11**	**13.85**
	Statlog Dataset							
NB	78.52	82.00	74.17	77.89				
SVM	81.85	94.67	65.83	77.66				
LR	82.59	87.33	76.67	81.65				
QDA	68.15	64.00	73.33	68.35				
kNN	65.56	68.67	61.67	64.98				
RF	71.11	81.33	58.33	67.94				
ANN	78.15	79.33	76.67	77.98				
HM-BagMoov	**87.93**	**94.67**	**79.50**	**86.42**				

Acc = Accuracy.
Spec = Specificity.
Sen = Sensitivity.
F-M = *F*-Measure.
NB = Naïve Bayes.
SVM = Support Vector Machine.
LR = Linear Regression.
QDA = Quadratic Discriminant Analysis.
kNN = *k* Nearest Neighbor.
RF = Random Forest.
ANN = Artificial Neural Network.
Bold values indicate the results of the proposed algorithm.

Table 8
Comparison of HM-BagMoov with state of art techniques for Cleveland Dataset.

Reference	Year	Accuracy (%)	Sensitivity (%)	Specificity (%)
Yang et al. [52]	2015	73.74	–	–
Peter et al. [53]	2015	78.2	–	–
Kiruthika et al. [54]	2014	78	–	–
Shouman et al. [16]	2013	80	–	–
Bashir et al. [55]	2015	84.16	92.68	74.10
HM-BagMoov	**2015**	**86**	**94**	**76**

Support Vector Machines (SVM) are used for classification, regression and outlier detection [49]. It is effective in high dimensional space and use different kernel functions for making decisions. The individual IBL algorithm has limitations such as it is computationally intensive and requires lot of storage space. The SVM algorithm performs the feature selection by using only subset of data chosen based on information gain. The ensemble model has resolved the storage problem by selecting only necessary and useful features for disease analysis and prediction. Moreover SVM algorithm decreases the overfitting issue and increase the prediction and classification accuracy. In any scenario where one classifier has some limitation, the other classifier performs well, consequently giving better performance.

The HM-BagMoov ensemble framework utilizes diverse set of base classifiers that differ in their decisions. Hence, each classifier in HM-BagMoov ensemble framework has diverse set of qualities that complement each other to make an accurate ensemble framework for disease prediction. Moreover the diversity parameter can be determined by the extent to which each individual classifier disagree about the probability distributions for the test datasets. Thus, all of the five selected classifiers complement each other very well. One of the main problems with ensemble models is that what they deliver in increased accuracy may get lost in translation. As a result, we have limited the number of different models in an ensemble to five. For a given instance, each trained classifier will predict one class either 0 or 1. For each classifier, the weight is calculated based on *F*-Measure of the training dataset. Therefore, five outputs will be obtained from these five classifiers in form of either 0 or 1.

2.3.3.2. Layer-2 classifiers. At layer-2 the output of layer-1 is combined using weighted bagging ensemble approach with further two different classifiers. In order to achieve diversity among the classifiers, we have used Artificial Neural Networks (ANN) and Random Forest (RF) classifier. In Layer-1 no decision trees based classifier is used. So instead of using a single decision tree based classifier we have opted for RF. Random Forest operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random forest classifiers almost always have lower classification error and deal really well with uneven data sets that have missing variables

Table 9
Comparison of HM-BagMoov ensemble framework with individual classifiers for breast cancer datasets.

Classifiers	UMC Dataset				WPBC Dataset			
	Acc (%)	Sen (%)	Spec (%)	*F*-M (%)	Acc (%)	Sen (%)	Spec (%)	*F*-M (%)
NB	68	73.9	75.04	73.92	72.76	36.36	16.5	22.70
SVM	70.31	73.76	89.62	80.92	76.29	0	0	0.00
LR	71.33	74.16	91.12	81.77	78.87	63.02	46.5	53.51
QDA	31.85	66.67	6	11.01	23.71	23.71	100	38.33
kNN	62.22	69.76	81.6	75.22	66.13	28.67	15	19.70
RF	70.28	11.76	95.02	20.94	76.77	100.00	2.13	4.17
ANN	65.73	30.59	80.60	44.35	78.28	87.42	48.94	62.75
HM-BagMoov	**73.45**	**74.19**	**95.57**	**83.53**	**80.34**	**90**	**19.5**	**32.05**
	WDBC Dataset				WBC Dataset			
NB	93.45	95.03	94.67	94.85	96.7	91.95	99.58	95.61
SVM	94.84	94.27	97.16	95.69	95.86	94.02	95.28	94.65
LR	95.26	93.78	99.44	96.53	95.85	96.26	91.68	93.91
QDA	79.63	78.97	93.78	85.74	70.53	95.24	15.81	27.12
kNN	78.04	77.67	91.31	83.94	63.09	45.73	36.63	40.68
RF	88.05	69.81	98.88	81.84	92.85	87.14	95.85	91.29
ANN	96.13	92.92	98.04	95.41	96.42	96.27	96.51	91.29
HM-BagMoov	**96.56**	**95**	**99.16**	**97.04**	**97.11**	**95.78**	**95.85**	**95.81**

Bold values indicate the results of the proposed algorithm.

[50]. It has better *f*-scores than a single decision tree and it is fast to build. ANN has iterative network learning where many problems are solved without finding and describing method of such problem solving, without building algorithms, without developing programs, even without any case of knowledge about the nature of solved problems [51]. Neural network classifier with back propagation learning capability has very low prediction error. The most important advantage of using ANN is the ability to implicitly detect complex nonlinear relationships among variables. Such a characteristic makes the ANNs suitable for prediction and pattern recognition applications. The final prediction is obtained at layer-3. Again, for each classifier, the weight is calculated based on *F*-Measure of the training dataset.

2.3.3.3. Layer-3 classification. Layer-3 is used to generate the final prediction result for a given instance. It accepts the result of previous layer i.e. three class predictions and combine them using novel multi-layer weighted bagging ensemble approach. The output of multi-layer weighted bagging ensemble classifier will be final prediction for a given instance. For example, either class 0 or class 1 will be the predicted class using proposed HM-BagMoov ensemble approach.

3. Results

3.1. Dataset description

The proposed framework is executed on five different heart disease datasets, four breast cancer datasets, two diabetes datasets, two liver disease datasets and one hepatitis dataset. Each dataset contains diverse set of attributes that ultimate determine the class of disease (healthy/diseased). Four heart disease datasets (SPECT, SPECTF, Heart disease and Statlog) are taken from UCI[1] data repository, and the fifth dataset (Eric) is taken from ricco[2] data repository. Three breast cancer datasets (WDBC, WBC, WPBC) are obtained from UCI data repository whereas one dataset (UMC Breast cancer dataset) is taken from Wisconsin clinical sciences center[3] repository.

Table 10
Comparison of HM-BagMoov ensemble framework with state of art techniques Wisconsin Breast Cancer dataset.

Reference	Year	Accuracy (%)	Sensitivity (%)	Specificity (%)
Sørensen et al. [56]	2015	92	90	65
Zand [57]	2015	84.5	70	57
Chaurasia et al. [58]	2014	74.47	76.2	92.5
Chaurasia et al. [19]	2014	–	95.8	98
Ashfaq Ahmed et. al. [20]	2013	75	–	–
HM-BagMoov	**2015**	**97**	**95.78**	**95.85**

Diabetes datasets are named as Pima Indian Diabetes Dataset (PIDD) and BioStat Diabetes Dataset (BDD). PIDD is taken from UCI repository whereas BDD is taken from Biostat data repository.[4] Liver disease datasets are named as BUPA liver disease dataset and ILDP (Indian liver patient dataset) liver disease dataset and both are taken from UCI data repository. Hepatitis dataset is named as hepatitis disease dataset which is taken from UCI data repository. In order to maintain consistency, the class labels of each dataset are replaced with 0 and 1 where 0 represents the absence of disease and 1 indicates the presence of heart disease. Further details of the datasets along with the sample set of each dataset can be found in the Supplementary material, Section A.

The multi-layer ensemble model is applied to each test set. Ten confusion matrices obtained from each fold of cross validation are then summed up into final confusion matrix. The average prediction results for all subsets are calculated and then analyzed.

3.2. Disease classification and prediction

3.2.1. Heart disease prediction

Table 7 shows the comparison of accuracy, sensitivity, specificity and *F*-measure results of the HM-BagMoov for all datasets with individual classifiers. It can be seen from Table 6 that, HM-BagMoov ensemble framework produces the highest accuracy level for all the datasets when compared to other individual classifiers. HM-BagMoov ensemble framework shows a consistent accuracy level of around 84%, whereas other classifiers are not stable as

[1] http://archive.ics.uci.edu/ml/datasets.html [last accessed 25 September 2014].
[2] http://www.eric.univ-lyon2.fr/~ricco/dataset/heart_disease_male.xls [last accessed 25 September 2014].
[3] https://www.lri.fr/~antoine/Courses/Master-ISI/TD-TP/breast-cancer.arff [last accessed 2 April 2015].
[4] http://biostat.mc.vanderbilt.edu/wiki/pub/Main/DataSets/diabetes.html [last accessed 25 September 2015].

Table 11
Comparison of HM-BagMoov ensemble framework with individual classifiers for diabetes datasets.

Classifiers	Pima Indian Diabetes Dataset				Biostat Diabetes Dataset			
	Acc (%)	Sen (%)	Spec (%)	*F*-M (%)	Acc (%)	Sen (%)	Spec (%)	*F*-M (%)
NB	71.09	81.15	72.6	76.64	85.61	51.45	49.76	50.59
SVM	75.95	78.22	87	82.38	90.57	87.83	50.43	63.34
LR	77.08	79.15	88.4	83.52	91.07	86.67	54.19	66.74
QDA	57.94	82.98	46.8	59.85	14.88	14.88	100	25.91
kNN	69.94	75.96	79	77.45	85.37	54.55	10.1	17.04
RF	65.10	3.73	98.00	7.19	85.11	100.00	0.00	0.00
ANN	74.74	57.84	83.80	68.44	89.83	95.34	58.33	72.38
HM-BagMoov	**78.21**	**78.65**	**92.6**	**85.05**	**93.07**	**86.31**	**65.19**	**74.28**

Bold values indicate the results of the proposed algorithm.

Table 12
Comparison of HM-BagMoov ensemble framework with state of art techniques for PIDD dataset.

Reference	Year	Accuracy (%)	Sensitivity (%)	Specificity (%)
Kandhasamy et al. [59]	2015	71.74	53.81	80.4
Bashir et al. [60]	2014	74.48	81.4	61.5
Gandhi et al. [24]	2014	75	–	–
Tapak et al. [61]	2013	75.3	13.3	99.9
Karthikeyani et al. [62]	2012	74.8	–	–
HM-BagMoov	**2015**	**77.21**	**77.65**	**91.6**

observed in the results. HM-BagMoov ensemble framework achieved best accuracy of 87.93% with 94.67% sensitivity, 79.50% specificity, and 86.41% *F*-Measure. Table 8 shows a comparison of accuracy, sensitivity, specificity and *F*-measure for HM-BagMoov ensemble framework with other state of art classification techniques for the Cleveland heart disease dataset. The comparison of the results shows that HM-BagMoov ensemble framework performed much better than state of the art classification techniques.

3.2.2. Breast cancer prediction

The evaluation of HM-BagMoov ensemble framework is also performed for breast cancer datasets. The comparison of HM-BagMoov ensemble framework is performed with other classifiers for breast cancer datasets as shown in Table 9. The highest accuracy of 97% and 95.78% sensitivity is achieved for WBC dataset whereas high specificity of 99.16% and 97.04% *F*-Measure is achieved for WDBC datasets respectively by HM-BagMoov ensemble framework. It has produced the highest accuracy across all the datasets, no single classifier has achieved that.

Table 10 shows accuracy, sensitivity, specificity and *f*-Measure comparison of HM-BagMoov ensemble framework with state of the art techniques for breast cancer datasets. HM-BagMoov ensemble framework produces the highest accuracy level for all the datasets when compared to other state of the art techniques. It has achieved an accuracy of 97% for WBC dataset whereas 95.78% sensitivity and 95.85% specificity are achieved when compared with the state of art techniques for the Wisconsin Breast Cancer dataset.

3.2.3. Diabetes prediction

The comparison of HM-BagMoov ensemble framework is also performed with other classifiers for diabetes datasets as shown in Table 11. We have used two diabetes datasets i.e. Pima Indian Diabetes Dataset (PIDD) and Biostat Diabetes Dataset (BDD) for comparison of results. Accuracy, sensitivity, specificity and *F*-Measure comparison is performed for HM-BagMoov ensemble framework with other classifiers. The analysis of the results indicates that HM-BagMoov has achieved the highest prediction accuracy for both the datasets. It has achieved 77.21% accuracy for PIDD and 93.07% for BDD datasets respectively.

Table 12 shows accuracy, sensitivity, specificity and *F*-measure comparison of HM-BagMoov ensemble framework with other state of art classification techniques. The analysis of the results indicates that HM-BagMoov ensemble framework has achieved the highest accuracy of 77.21%, when compared with the state of the art techniques for PIDD dataset.

3.2.4. Liver disease prediction

The comparison of HM-BagMoov ensemble framework is also performed for liver disease datasets with other classifiers. Table 13 shows accuracy, sensitivity, specificity and *F*-Measure comparison of different classifiers for BUPA liver disease dataset and Indian liver patient dataset (ILPD). The analysis of the results indicates that HM-BagMoov ensemble framework has achieved the highest accuracy in both liver disease datasets. It has achieved 72.7% accuracy for ILPD dataset and 70.16% accuracy for Bupa liver disease dataset when compared with other classifiers.

Table 14 shows accuracy, sensitivity, specificity and *F*-Measure comparison of HM-BagMoov ensemble framework with the state of the art classification techniques for liver disease datasets. The

Table 13
Comparison of proposed HM-BagMoov ensemble framework with individual classifiers for liver disease datasets.

Classifiers	ILPD Dataset				BUPA liver disease dataset			
	Acc (%)	Sen (%)	Spec (%)	*F*-M (%)	Acc (%)	Sen (%)	Spec (%)	*F*-M (%)
NB	68.39	46.84	75.35	57.77	60.02	63.8	72.59	67.91
SVM	71.36	0	0	0.00	67.51	67.18	86.02	75.44
LR	70.66	50	5.44	9.81	67.52	69.41	79.06	73.92
QDA	70.10	0	0	0.00	42.03	0	0	0.00
kNN	65.2	38.66	34.15	36.27	63.78	67.21	73.92	70.41
RF	71.70	99.76	1.80	3.53	59.42	7.59	97.00	14.07
ANN	68.44	84.13	29.34	43.51	72.46	60.69	81.00	69.39
HM-BagMoov	**72.7**	**76**	**2.76**	**4.44**	**70.16**	**68.81**	**89.52**	**77.81**

Bold values indicate the results of the proposed algorithm.

Table 14
Comparison of proposed HM-BagMoov ensemble framework with state of art techniques for ILPD dataset.

Reference	Year	Accuracy (%)	Sensitivity (%)	Specificity (%)
Gulia et al. [63]	2014	67.2	–	–
Jin et al. [64]	2014	65.3	72.7	46.7
Sug [65]	2012	67.82	–	–
Ramana et al. [66]	2011	62.6	55.86	67.5
Karthik et al. [67]	2011	55	–	–
HM-BagMoov	**2015**	**72.7**	**75**	**1.76**

Table 15
Comparison of proposed HM-BagMoov ensemble framework with individual classifiers for hepatitis dataset.

Classifiers	Acc (%)	Sen (%)	Spec (%)	*F*-M (%)
NB	83	66.67	36.67	47.32
SVM	85	66.67	54.17	59.77
LR	84.38	70.17	55	61.67
QDA	83.75	68.42	40	50.49
kNN	74.04	21.43	9.17	12.84
RF	83.12	100.00	18.75	31.58
ANN	83.77	91.80	53.13	67.30
HM-BagMoov	**87.04**	**77.27**	**51.67**	**61.93**

Bold values indicate the results of the proposed algorithm.

Table 16
Comparison of HM-BagMoov ensemble framework with state of art techniques.

Reference	Year	Accuracy (%)	Sensitivity (%)	Specificity (%)
Houby [68]	2014	69	87.3	50.6
Karthikeyan et al. [38]	2013	84	–	–
Neshat et al. [69]	2012	70.29	–	–
Kumar et al. [13]	2011	83.12	–	–
Khan et al. [70]	2008	72	83	66
HM-BagMoov	**2015**	**87.04**	**77.27**	**51.67**

analysis of the results indicates that HM-BagMoov has achieved the highest accuracy of 72.7% for ILPD dataset when compared with other classifiers.

3.2.5. Hepatitis prediction

Table 15 shows accuracy, sensitivity, specificity and *F*-Measure comparison of HM-BagMoov ensemble framework with other classifiers. The analysis of the results indicates that HM-BagMoov ensemble framework has achieved the highest accuracy, sensitivity and *F*-Measure of 87.04%, 77.27% and 61.93% respectively when compared with individual classifiers.

Table 16 shows accuracy, sensitivity, specificity and *F*-Measure comparison of HM-BagMoov ensemble framework with the state of the art techniques for hepatitis disease dataset. The analysis of the results indicates that HM-BagMoov ensemble framework has achieved the highest accuracy when compared with the state of the art techniques.

Looking at the performance of HM-BagMoov ensemble framework, it can be seen that our approach yields the highest accuracy, sensitivity, specificity and *F*-Measure for all the diseases and datasets. While some classifiers have produced good results for a single dataset, that same classifier when applied to another dataset has produced poor results. Whereas HM-BagMoov ensemble has consistently produced the highest accuracy for all datasets. This is contributed to our ensemble framework, which utilizes an optimal model of class diversity. Each classifier in HM-BagMoov ensemble framework has diverse set of qualities that complement and overcome the limitations of each other to make an accurate ensemble framework for disease prediction.

3.3. Discussion

Looking at the performance of our approach and existing approaches, it can be seen that our approach yields the highest accuracy for all the diseases. This increased in accuracy is contributed to our ensemble approach which uses bagging with multi-objective optimized weighted vote based technique and multi-layer classification.

3.3.1. Single layer vs. multi-layer classification

The multi-layer classifiers allows for flexibility in many ways as compared to single layer classifiers. The main advantage is that we can use different models for respective granularity of problems. If we use different classifiers for different layers, we can use different features for each layer; and the classification tasks can be more refined [71]. Another advantage of multiple layers is that we can split up the imbalanced classification problems in two more or relatively more balanced classification problems [71]. The multi-layer technique is theoretically modest and scalable for hierarchical training and classification and can be applicable for large medical datasets [72].

In layered architecture, the negative sample is smaller at top level layer, because it includes the item only from same top level. As a result of that, training is faster at top level layer [72]. Also the hierarchical structure always provide better classification quality than the flat structure, because of computing errors at each level. Hence, it reduces the risk of making an error in the top down classification process [71].

The empirical evaluation of HM-BagMoov approach in layered structure and putting all seven classifiers (NB, QDA, LR, IBM, SVM, RF, ANN) in single layer is also performed. Table 17 shows the comparison of single layer ensemble classifier and layered HM-BagMoov ensemble approach for heart disease datasets. It is

Table 17
Comparison of HM-BagMoov with single layered ensemble classifier for heart disease datasets.

Dataset	Technique	Accuracy (%)	Sensitivity (%)	Specificity (%)	*F*-Measure (%)
Cleveland	Single layer Ensemble	84.16	92.07	74.82	82.55
	HM-BagMoov	86.20	94.12	76.32	84.29
Eric	Single layer Ensemble	78.95	87.18	68.48	76.71
	HM-BagMoov	83.82	91.74	73.74	81.76
SPECT	Single layer Ensemble	83.15	30.91	96.70	46.84
	HM-BagMoov	84.77	34.73	97.75	51.25
SPECTF	Single layer Ensemble	78.28	14.55	94.81	25.22
	HM-BagMoov	83.03	7.45	99.11	13.85
Statlog	Single layer Ensemble	84.81	92.00	75.83	83.14
	HM-BagMoov	87.93	94.67	79.50	86.42

Table 18
Comparison of HM-BagMoov with other ensemble and voting approaches for heart disease datasets.

Classifiers	Cleveland Dataset				Eric Dataset			
	Acc (%)	Sen (%)	Spec (%)	*F*-M (%)	Acc (%)	Sen (%)	Spec (%)	*F*-M (%)
Bagging	84.16	92.68	74.10	82.36	81.82	89.74	71.74	79.74
AdaBoost	79.21	83.54	74.10	78.54	77.03	80.34	72.83	76.40
Majority voting	82.84	87.80	76.98	82.04	77.51	87.18	65.22	74.62
Accuracy-Weighting	84.16	92.95	74.21	82.53	81.50	85.55	70.99	77.59
HM-BagMoov	**86.20**	**94.12**	**76.32**	**84.29**	**83.82**	**91.74**	**73.74**	**81.76**
	SPECT Dataset				SPECTF Dataset			
Bagging	82.77	32.73	95.75	48.78	79.03	5.45	98.11	10.33
AdaBoost	83.90	36.36	96.23	52.78	78.28	45.45	86.79	59.66
Majority voting	83.52	36.36	95.75	52.71	79.78	9.09	98.11	16.64
Accuracy-Weighting	83.99	32.50	93.99	48.29	80.03	5.45	95.50	10.31
HM-BagMoov	**84.77**	**34.73**	**97.75**	**51.25**	**83.03**	**7.45**	**99.11**	**13.85**
	Statlog Dataset							
Bagging	85.93	92.67	77.50	84.41				
AdaBoost	78.89	83.33	73.33	78.01				
Majority voting	82.59	90.67	72.50	80.57				
Accuracy-Weighting	85.55	92.67	77.50	84.40				
HM-BagMoov	**87.93**	**94.67**	**79.50**	**86.42**				

Acc = Accuracy.
Spec = Specificity.
Sen = Sensitivity.
F-M = *F*-Measure.
Bold values indicate the results of the proposed algorithm.

Table 19
Comparison of HM-BagMoov with other ensemble and voting approaches for breast cancer datasets.

Classifiers	UMC Dataset				WPBC Dataset			
	Acc (%)	Sen (%)	Spec (%)	*F*-M (%)	Acc (%)	Sen (%)	Spec (%)	*F*-M (%)
Bagging	71.05	73.75	91.60	81.71	77.24	100.00	20.00	33.33
AdaBoost	64.34	28.24	79.60	41.68	79.80	92.72	38.30	54.21
Majority voting	72.38	20.00	94.53	33.01	82.32	98.01	31.91	48.15
Accuracy-Weighting	72.95	45.46	85.55	59.37	82.85	97.95	32.91	49.26
HM-BagMoov	**73.45**	**74.19**	**95.57**	**83.53**	**80.34**	**90**	**19.5**	**32.05**
	WDBC Dataset				WBC Dataset			
Bagging	94.19	94.60	95.04	94.81	97.57	94.13	95.08	94.60
AdaBoost	96.13	94.34	97.20	95.75	95.85	92.95	97.38	95.11
Majority voting	95.61	89.15	99.44	94.01	96.71	97.38	95.44	96.40
Accuracy-Weighting	96.01	94.23	98.50	96.31	96.98	93.28	95.01	94.13
HM-BagMoov	**96.56**	**95**	**99.16**	**97.04**	**97.11**	**95.78**	**95.85**	**95.81**

Bold values indicate the results of the proposed algorithm.

clear from empirical evaluation that the arrangement of classifiers in layered structure performs better than the classifiers arranged in a single layer.

3.3.2. HM-BagMoov comparison with other ensemble and voting techniques

The experimental comparison of HM-BagMoov is performed with other ensemble approaches such as Bagging, Majority voting, Adaboost and voting approaches such as Majority voting and Accuracy based weighting (Accuracy-Weighting) techniques in order to show the superiority of enhanced bagging and optimized weighting. Accuracy, sensitivity, specificity and *f*-measure comparison is performed on heart disease, breast cancer, liver, diabetes and hepatitis datasets; and results are shown in Tables 18–22. It is clear from the analysis that HM-BagMoov approach is better than simple Bagging, Majority voting, Adaboost and Accuracy-weighting when compared to other ensemble and voting techniques).

3.3.3. Time complexity comparison of HM-BagMoov with other techniques

The training time of HM-BagMoov is also compared wrt time with individual as well as other ensemble approaches as shown in Table 23. HM-BagMoov does has high time complexity as compared to single classifiers. This is expected as HM-BagMoov is using a combination of classifiers with an enhanced bagging. However we have overcome this limitation by using off-line learning model

Table 20
Comparison of HM-BagMoov with other ensemble and voting approaches for diabetes datasets.

Classifiers	Pima Indian Diabetes Dataset				Biostat Diabetes Dataset			
	Acc (%)	Sen (%)	Spec (%)	*F*-M (%)	Acc (%)	Sen (%)	Spec (%)	*F*-M (%)
Bagging	77.99	75.96	85	80.22	90.29	84.83	62.48	71.95
AdaBoost	76.43	52.99	89.00	66.42	88.83	96.79	43.33	59.87
Majority voting	76.30	50.00	90.40	64.39	91.07	98.54	48.33	64.86
Accuracy-Weighting	77.00	65.54	85.55	74.21	92.24	95.55	45.76	61.88
HM-BagMoov	**78.21**	**78.65**	**92.6**	**85.05**	**93.07**	**86.31**	**65.19**	**74.28**

Bold values indicate the results of the proposed algorithm.

Table 21
Comparison of HM-BagMoov with other ensemble and voting approaches for liver disease datasets.

Classifiers	ILPD Dataset				BUPA liver disease dataset			
	Acc (%)	Sen (%)	Spec (%)	*F*-M (%)	Acc (%)	Sen (%)	Spec (%)	*F*-M (%)
Bagging	71.87	46.59	16.92	24.82	69.02	66.72	92.99	77.69
AdaBoost	66.55	81.49	29.34	43.15	68.41	53.79	79.00	64.00
Majority voting	71.53	99.04	2.99	5.81	71.88	45.52	91.00	60.68
Accuracy-Weighting	71.90	75	1.00	1.97	67.01	59.99	75.98	67.04
Enhanced Bagging	**72.7**	**76**	**2.76**	**4.44**	**70.16**	**68.81**	**89.52**	**77.81**

Bold values indicate the results of the proposed algorithm.

Table 22
Comparison of HM-BagMoov with other ensemble and voting approaches for hepatitis dataset.

Classifiers	Acc (%)	Sen (%)	Spec (%)	*F*-M (%)
Bagging	85.50	75.63	48.99	59.46
AdaBoost	83.12	95.08	37.50	53.79
Majority voting	83.12	94.26	40.63	56.78
Accuracy-Weighting	85.55	75.60	50.99	60.90
Enhanced Bagging	**87.04**	**77.27**	**51.67**	**61.93**

Bold values indicate the results of the proposed algorithm.

which is commonly used in complex expert systems [100]. System has high time complexity only when it is being trained. At runtime, only the learned model is used to make a prediction. The model can be updated/re-trained when the system is not being used.

4. IntelliHealth: an intelligent medical decision support application

Using the HM-BagMoov ensemble framework we have developed an application named "IntelliHealth". The IntelliHealth med-

Table 23
Time comparison of HM-BagMoov with individual and ensemble classifiers.

Classifiers	Time (ms) per instance						
	Heart disease datasets					Diabetes datasets	
	Cleveland	Eric	Statlog	SPECT	SPECTF	PIMA	Diabetes
NB	0.1	.05	.08	.06	.51	.09	.13
SVM	.03	.01	.01	.02	.03	.01	.02
LR	.01	.02	.01	.01	.01	.01	.01
QDA	.03	.05	.03	.04	.11	.02	.03
kNN	.11	.08	.11	.13	.26	0.2	.19
Bagging	35.42	22.24	38.02	26.88	118.9	74.47	63.95
Majority Voting	0.47	0.27	0.47	0.36	1.29	.75	.85
AdaBoost	16.17	14.41	21.15	1.87	86.24	16.13	8.28
HM-BagMoov	**35.10**	**22.1**	**36.9**	**25.9**	**118.2**	**73.5**	**62.07**
	Breast cancer datasets				Liver disease datasets		Hepatitis dataset
	UMC	WPBC	WDBC	WBC	ILPD	BUPA	Hepatitis
NB	.06	.43	.32	.09	.1	.1	0.14
SVM	.02	.04	.02	.01	.02	.02	.03
LR	.02	.04	.01	.01	.01	.01	.03
QDA	.02	0.1	.06	.03	.02	.01	.06
kNN	.1	.17	.36	.24	0.2	0.1	.09
Bagging	18.11	187.6	490.9	42.07	55.95	31.79	38.19
Majority Voting	.24	2.13	6.03	0.58	0.7	0.39	0.49
AdaBoost	9.6	138.9	186.3	16.96	13.46	14.33	2.33
HM-BagMoov	**17.98**	**150.9**	**467.3**	**40.2**	**54.82**	**30.09**	**36.7**

Bold values indicate the results of the proposed algorithm.

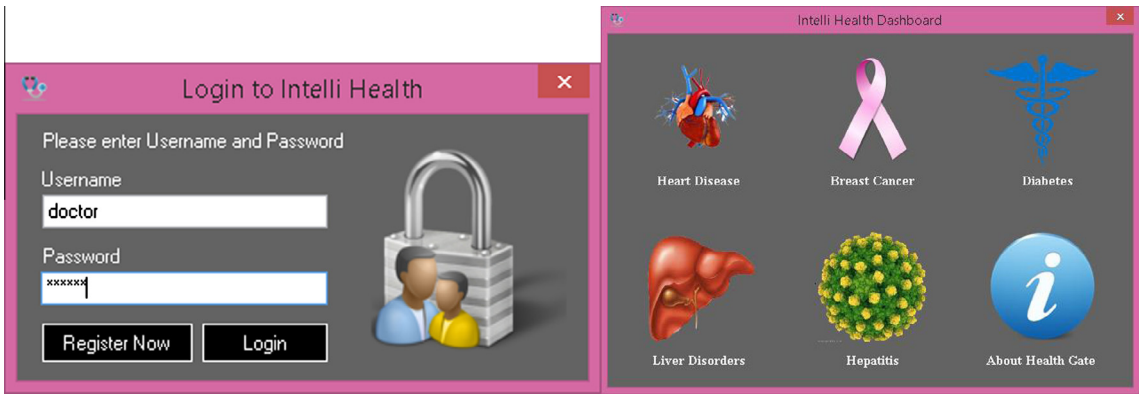


Fig. 3. The login and dashboard of the application.

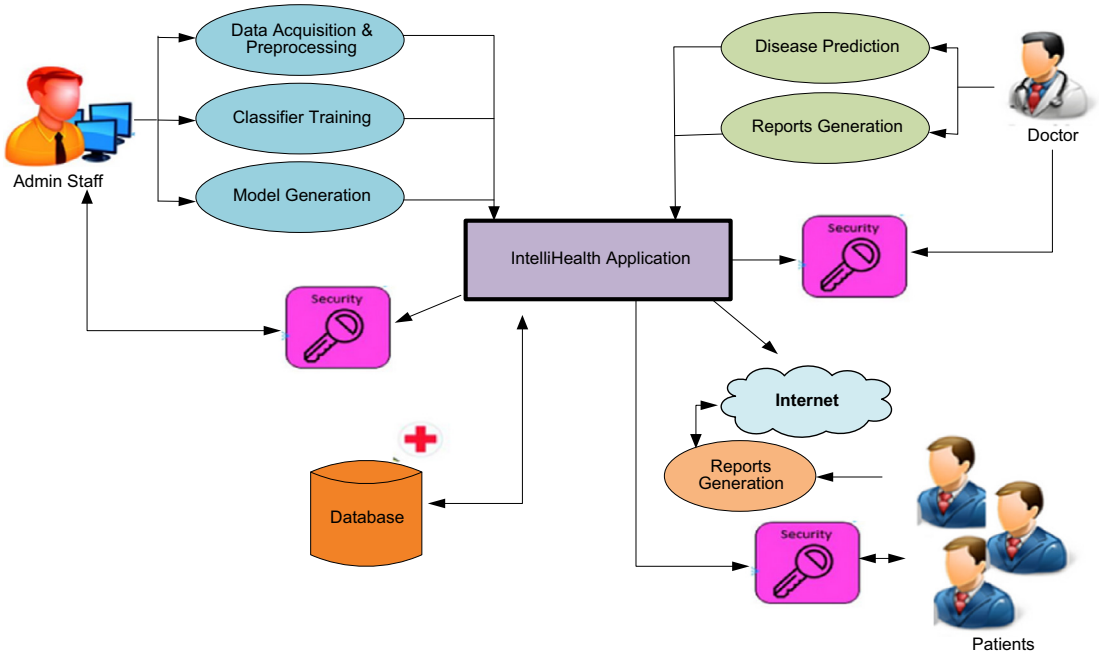


Fig. 4. Architecture of proposed IntelliHealth application.

ical application is developed with the purpose of assisting the physician in diagnosing diseases. The login and dashboard of the application is show in Fig. 3, whereas the architecture of Intellihealth application is shown in Fig. 4.

4.1. Modules of IntelliHealth

The application is divided into four main modules:

1. Data acquisition and preprocessing module.
2. Classifier training and model generation.
3. Disease prediction module.
4. Report generation.

1. Data acquisition and preprocessing module

This module involves entering the patient data. Preprocessing is then applied to the data at backend. The refined data is saved in the database and a unique patient ID is generated for each patient. Each patient is identified by a unique patient's ID which is generated when its information is entered into the system. Fig. 5 shows the form that is used to enter a new patient's data and preprocess it. The form that we have generated is dynamic in nature and attributes can be edited for each disease, thus providing complete customization.

After entering the patient information, the admin staff member will click on "Add New Patient" button. The information will be

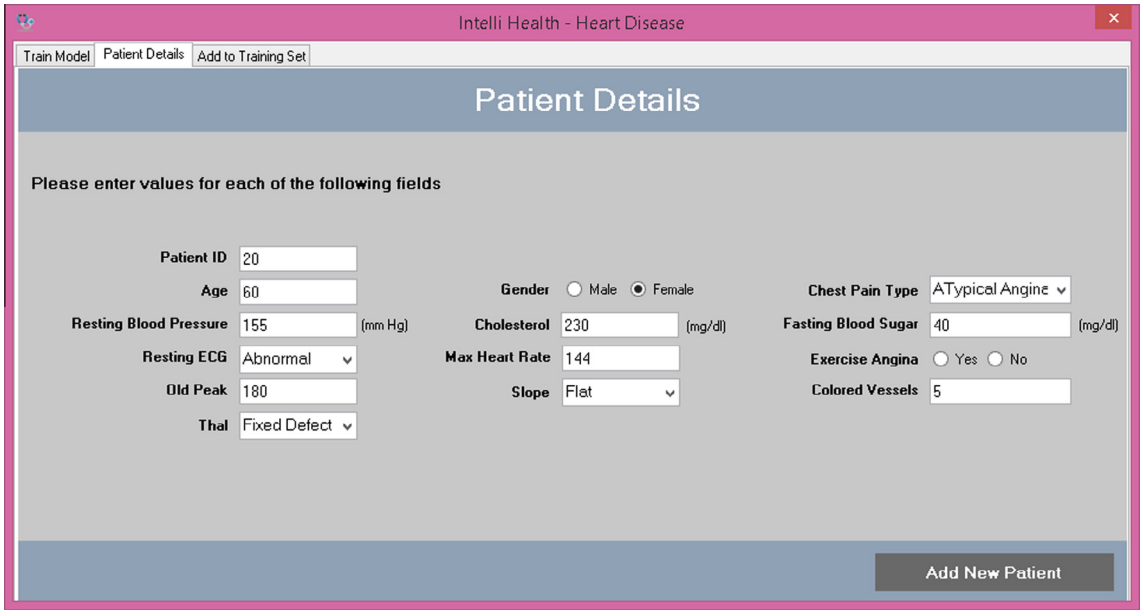
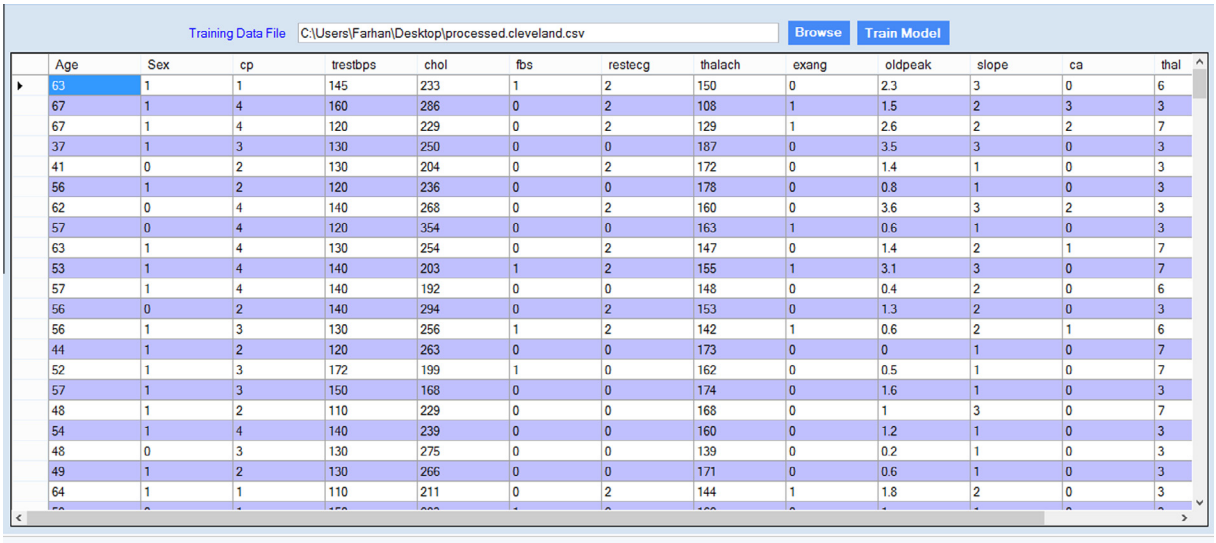


Fig. 5. Add patient's detail screen.



Age	Sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal
63	1	1	145	233	1	2	150	0	2.3	3	0	6
67	1	4	160	286	0	2	108	1	1.5	2	3	3
67	1	4	120	229	0	2	129	1	2.6	2	2	7
37	1	3	130	250	0	0	187	0	3.5	3	0	3
41	0	2	130	204	0	2	172	0	1.4	1	0	3
56	1	2	120	236	0	0	178	0	0.8	1	0	3
62	0	4	140	268	0	2	160	0	3.6	3	2	3
57	0	4	120	354	0	0	163	1	0.6	1	0	3
63	1	4	130	254	0	2	147	0	1.4	2	1	7
53	1	4	140	203	1	2	155	1	3.1	3	0	7
57	1	4	140	192	0	0	148	0	0.4	2	0	6
56	0	2	140	294	0	2	153	0	1.3	2	0	3
56	1	3	130	256	1	2	142	1	0.6	2	1	6
44	1	2	120	263	0	0	173	0	0	1	0	7
52	1	3	172	199	1	0	162	0	0.5	1	0	7
57	1	3	150	168	0	0	174	0	1.6	1	0	3
48	1	2	110	229	0	0	168	0	1	3	0	7
54	1	4	140	239	0	0	160	0	1.2	1	0	3
48	0	3	130	275	0	0	139	0	0.2	1	0	3
49	1	2	130	266	0	0	171	0	0.6	1	0	3
64	1	1	110	211	0	2	144	1	1.8	2	0	3

Fig. 6. Load training set and generate model screen.

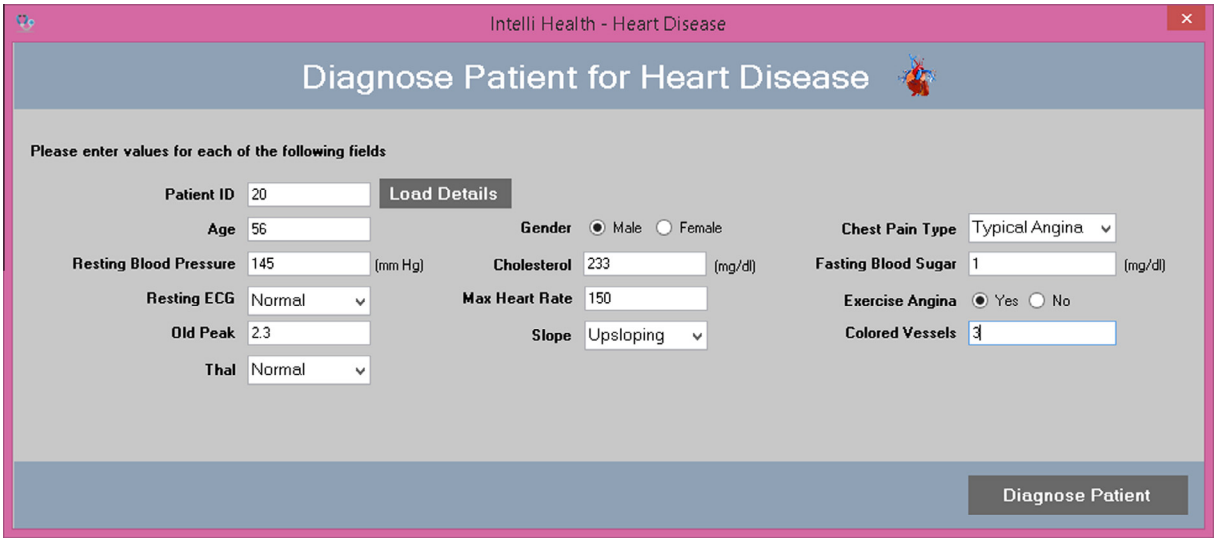


Fig. 7. Diagnose patient screen displayed to doctor.

added into the database. The patient detail along with disease diagnosis can be added into the training set on doctor's alert.

2. Classifier training and model generation

The second module of IntelliHealth application is classifier training and model generation. After entering patient's data, classifier training is performed based on historical/previous data. After classifier training, a model is generated at the backend using the HM-BagMoov ensemble. The model training is performed using "Train model" tab. Fig. 6 shows the screen that will be displayed when model training is performed.

3. Disease prediction module

The third module of proposed application is disease prediction. The disease prediction module screens will be accessible to only the doctor. When the doctor enters the patient's ID then its relevant information will be generated. The doctor can then make disease prediction based on training model and personal knowledge. Fig. 7 shows the screen that will be displayed for disease prediction to the doctor. The doctor will enter the unique patient's ID and click on "Load Details" button. The respective information entered by the admin staff member will be displayed in each field. The doctor will then click on "Diagnose Patient" button. A message will be generated if the patient has a possibility of a disease and further tests should be carried out or patient is healthy.

4. Report generation

The report generation module will generate medical report for each patient. Both the doctor and the patient will have access to report generation module. The doctor can handover the report to patient in printed form or can email the patient. The patient can also generate the reports itself by login to the system. Currently the report generation is limited to displaying the patient information and wherever the patient has been diagnosed with a specific disease or not.

4.2. Users of proposed IntelliHealth application

There are three main users of IntelliHealth application. Each user has its login ID and password that will use for interaction with the system.

1. **Admin/I.T. Staff:** The Admin staff member will have access to following two modules:
   - Data acquisition and preprocessing module.
   - Classifier training and model generation module.
2. **Doctor:** Doctor will have access to the following modules:
   - Disease prediction module.
   - Report generation module.
3. **Patient:** The patient will have access to following module:
   - Report generation module.

5. Conclusion

In this paper we have proposed an ensemble framework with multi-layer classification using enhanced bagging and optimized weighting. The framework was evaluated on five different heart disease datasets, four breast cancer datasets, two diabetes datasets, two liver disease datasets and one hepatitis dataset. The analysis of the results indicates that HM-BagMoov ensemble framework achieved the highest accuracy, sensitivity and *F*-Measure when compared with individual classifiers for all the diseases. In addition to this ensemble framework also achieved the highest accuracy when compared with the state of the art techniques. Moreover, an application named IntelliHealth application is also developed based on proposed ensemble model that can help practitioners and patients for disease classification and prediction based on disease symptoms. In future, a publically accessible web server can be developed where people can have access to the application through internet and can perform any query effectively and efficiently.

Conflict of interest

The authors declare that there are no conflict of interest.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jbi.2015.12.001.

References

[1] C. Fernández-Llatas, J.M. García-Gómez (Eds.), Data Mining in Clinical Medicine, Humana Press, 2015.
[2] J.H. Chen, T. Podchiyska, R.B. Altman, OrderRex: clinical order decision support and outcome predictions by data-mining electronic medical records, J. Am. Med. Inform. Assoc. (2015).
[3] S. Dua, X. Du, Data Mining and Machine Learning in Cyber Security, CRC Press, 2011.
[4] A. Ahmad, G. Brown, Random ordinality ensembles: ensemble methods for multi-valued categorical data, Inf. Sci. 296 (2015) 75–94.
[5] B. Sluban, N. Lavrač, Relating ensemble diversity and performance: a study in class noise detection, Neurocomputing 160 (2015) 120–131.
[6] F. Moretti, S. Pizzuti, S. Panzieri, M. Annunziato, Urban traffic flow forecasting through statistical and neural network bagging ensemble hybrid modeling, Neurocomputing (2015).
[7] M.J. Kim, D.K. Kang, H.B. Kim, Geometric mean based boosting algorithm with over-sampling to resolve data imbalance problem for bankruptcy prediction, Expert Syst. Appl. 42 (3) (2015) 1074–1082.
[8] S. Kang, S. Cho, P. Kang, Multi-class classification via heterogeneous ensemble of one-class classifiers, Eng. Appl. Artif. Intell. 43 (2015) 35–43.
[9] M. Kitakaze, M. Asakura, A. Nakano, S. Takashima, T. Washio, Data mining as a powerful tool for creating novel drugs in cardiovascular medicine: the importance of a "back-and-forth loop" between clinical data and basic research, Cardiovasc. Drugs Ther. (2015) 1–7.
[10] F. Zhu, P. Patumcharoenpol, C. Zhang, Y. Yang, J. Chan, A. Meechai, B. Shen, Biomedical text mining and its applications in cancer research, J. Biomed. Inform. 46 (2) (2013) 200–211.
[11] B.R. Shah, L.L. Lipscombe, Clinical diabetes research using data mining: a Canadian perspective, Can. J. Diabetes 39 (3) (2015) 235–238.
[12] N. Lavrač, Selected techniques for data mining in medicine, Artif. Intell. Med. 16 (1) (1999) 3–23.
[13] M.V. Kumar, V.V. Sharathi, B.R.G. Devi, Hepatitis prediction model based on data mining algorithm and optimal feature selection to improve predictive accuracy, Int. J. Comput. Appl. 51 (19) (2012) 13–16.
[14] K. Thenmozhi, P. Deepika, Heart disease prediction using classification with different decision tree techniques, Int. J. Eng. Res. Gen. Sci. 2 (6) (2014).
[15] R. Chitra, D.V. Seenivasagam, Heart disease prediction system using supervised learning classifier, Int. J. Softw. Eng. Soft Comput. 3 (1) (2013).
[16] M. Shouman, T. Turner, R. Stocker, Integrating clustering with different data mining techniques in the diagnosis of heart disease, J. Comput. Sci. Eng. 20 (1) (2013).
[17] M. Shouman, T. Turner, R. Stocker, Integrating Naive Bayes and K-means clustering with different initial centroid selection methods in the diagnosis of heart disease patients, Comput. Sci. Inform. Technol. (2012) 125–137.
[18] S. Ghumbre, C. Patil, A. Ghatol, Heart disease diagnosis using support vector machine, in: International Conference on Computer Science and Information Technology (ICCSIT'), Pattaya, 2011.
[19] V. Chaurasia, S. Pal, A novel approach for breast cancer detection using data mining techniques, Int. J. Innov. Res. Comput. Commun. Eng. 2 (1) (2014).
[20] K. Ashfaq Ahmed, S. Aljahdali, S.N. Hussain, Comparative prediction performance with support vector machine and random forest classification techniques, Int. J. Comput. Appl. 69 (11) (2013).
[21] G.I. Salama, M.B. Abdelhalim, M.A. Zeid, Breast cancer diagnosis on three different datasets using multi-classifiers, Int. J. Comput. Inform. Technol. 01 (01) (2012).
[22] D. Lavanya, K.U. Rani, Ensemble decision tree classifier for breast cancer data, Int. J. Inform. Technol. Converg. Serv. (IJITCS) 2 (1) (2012).
[23] D. Lavanya, K.U. Rani, Analysis of feature selection with classification: breast cancer datasets, Ind. J. Comput. Sci. Eng. (IJCSE) 2 (5) (2012).
[24] K.K. Gandhi, N.B. Prajapati, Diabetes prediction using feature selection and classification, Int. J. Adv. Eng. Res. Develop. (2014).
[25] F. Stahl, R. Johansson, E. Renard, Ensemble Glucose Prediction in Insulin-Dependent Diabetes. Data Driven Modeling for Diabetes, Springer, 2014.
[26] M.W. Aslam, Z. Zhu, A.K. Nandi, Feature generation using genetic programming with comparative partner selection for diabetes classification, Expert Syst. Appl. (2013).
[27] M. Nirmala Devi, S. Appavu, U.V. Swathi, An amalgam KNN to predict diabetes mellitus, in: International Conference on Emerging Trends in Computing, Communication and Nanotechnology (ICE-CCN), 2013.
[28] Y.A. Christobel, P. SivaPrakasam, The negative impact of missing value imputation in classification of diabetes dataset and solution for improvement, IOSR J. Comput. Eng. (IOSRJCE) 7 (4) (2012) 5.
[29] R. Zolfaghari, Diagnosis of diabetes in female population of Pima Indian heritage with ensemble of BP neural network and SVM, IJCEM Int. J. Comput. Eng. Manage. 15 (2012).
[30] C. Lee, A fuzzy expert system for diabetes decision support application, IEEE Trans. Syst. Man Cybern. B Cybern. 41 (1) (2011).
[31] S. Vijayarani, S. Dhayanand, Liver disease prediction using SVM and Naïve Bayes algorithms, Int. J. Sci. Eng. Technol. Res. (IJSETR) 4 (4) (2015).
[32] H. Jin, S. Kim, J. Kim, Decision factors on effective liver patient data prediction, Int. J. BioSci. BioTechnol. 6 (4) (2014).
[33] K. Sugawara, N. Nakayama, S. Mochida, Acute liver failure in Japan: definition, classification, and prediction of the outcome, J. Gastroenterol. (2012).
[34] Y. Kumar, G. Sahoo, Prediction of different types of liver diseases using rule based classification model, Technol. Healthcare 21 (5) (2013) 417–432.
[35] B.V. Ramana, M.P. Babu, Liver classification using modified rotation forest, Int. J. Eng. Res. Develop. 1 (6) (2012) 17–24.
[36] S. Pushpalatha, J. Pandya, Data model comparison for hepatitis diagnosis, Int. J. Emerg. Res. Manage. Technol. 3 (7) (2014) 138–141.
[37] E.M.F. El Houby, Analysis of associative classification for prediction of HCV response to treatment, Int. J. Comput. Appl. 63 (15) (2013).
[38] T. Karthikeyan, P. Thangaraju, Analysis of classification algorithms applied to hepatitis patients, Int. J. Comput. Appl. 62 (15) (2013).
[39] Yılmaz Kaya, Murat Uyar, A hybrid decision support system based on rough set and extreme learning machine for diagnosis of hepatitis disease, 2013.
[40] M.V. Kumar, V.V. Sharathi, B.R.G. Devi, Hepatitis prediction model based on data mining algorithm and optimal feature selection to improve predictive accuracy, Int. J. Comput. Appl. 51 (19) (2012).
[41] J.F. Díez-Pastor, J.J. Rodríguez, C. García-Osorio, L.I. Kuncheva, Random balance: ensembles of variable priors classifiers for imbalanced data, Knowl.-Based Syst. (2015).
[42] M.A. King, A.S. Abrahams, C.T. Ragsdale, Ensemble learning methods for pay-per-click campaign management, Expert Syst. Appl. 42 (10) (2015) 4818–4829.
[43] H. Parvin, M. MirnabiBaboli, H. Alinejad-Rokny, Proposing a classifier ensemble framework based on classifier selection and decision tree, Eng. Appl. Artif. Intell. 37 (2015) 34–42.
[44] J. Mendes-Moreira, A.M. Jorge, J.F. de Sousa, C. Soares, Improving the accuracy of long-term travel time prediction using heterogeneous ensembles, Neurocomputing 150 (2015) 428–439.
[45] S.H. Park, J. Fürnkranz, Efficient implementation of class-based decomposition schemes for Naïve Bayes, Mach. Learn. 96 (3) (2014) 295–309.
[46] H. Hino, K. Koshijima, N. Murata, Non-parametric entropy estimators based on simple linear regression, Comput. Stat. Data Anal. 89 (2014) 72–84.
[47] S. Bose, A. Pal, R. SahaRay, J. Nayak, Generalized quadratic discriminant analysis, Pattern Recogn. 48 (8) (2015) 2676–2684.

[48] D. Lin, X. An, J. Zhang, Double-bootstrapping source data selection for instance-based transfer learning, Pattern Recogn. Lett. 34 (11) (2013) 1279–1285.
[49] S. Datta, S. Das, Near-Bayesian support vector machines for imbalanced data classification with equal or unequal misclassification costs, Neural Networks 70 (2015) 39–52.
[50] I. Ibarguren, J.M. Pérez, J. Muguerza, I. Gurrutxaga, O. Arbelaitz, Coverage-based resampling: building robust consolidated decision trees, Knowl.-Based Syst. 79 (2015) 51–67.
[51] S. Senan, Robustness analysis of uncertain dynamical neural networks with multiple time delays, Neural Networks (2015).
[52] J. Yang, Y. Lee, U. Kang, Comparison of prediction models for coronary heart diseases in depression patients, Int. J. Multimedia Ubiq. Eng. 10 (3) (2015) 257–268.
[53] J. Peter, Somasundaram, Probabilistic classification for prediction of heart disease, Aust. J. Basic Appl. Sci. 9 (7) (2015) 639–643.
[54] C. Kiruthika, S.N.S. Rajini, An ill-identified classification to predict cardiac disease using data clustering, Int. J. Data Min. Tech. Appl. 03 (2014) 321–325.
[55] S. Bashir, U. Qamar, F.H. Khan, BagMOOV: a novel ensemble for heart disease prediction bootstrap aggregation with multi-objective optimized voting, Australas. Phys. Eng. Sci. Med. (2015) 1–19.
[56] K.P. Sørensen, M. Thomassen, Q. Tan, M. Bak, S. Cold, M. Burton, M.J. Larsen, T. A. Kruse, Long non-coding RNA expression profiles predict metastasis in lymph node-negative breast cancer independently of traditional prognostic markers, Breast Cancer Res. (2015).
[57] H.K.K. Zand, A comparative survey on data mining techniques for breast cancer diagnosis and prediction, Ind. J. Fundam. Appl. Life Sci. 5 (S1) (2015) 4330–4339.
[58] V. Chaurasia, S. Pal, Data mining techniques: to predict and resolve breast cancer survivability, Int. J. Comput. Sci. Mob. Comput. 3 (1) (2014) 10–22.
[59] J.P. Kandhasamy, S. Balamurali, Performance analysis of classifier models to predict diabetes mellitus, Proc. Comput. Sci. 47 (2015).
[60] S. Bashir, U. Qamar, F.H. Khan, An efficient rule based classification of diabetes using ID3, C4.5 and CART ensemble, in: Frontier Information Technology, IEEE, Islamabad, Pakistan, 2015.
[61] L. Tapak, H. Mahjub, O. Hamidi, J. Poorolajal, Real-data comparison of data mining methods in prediction of diabetes in Iran, Healthcare Inform. Res. 19 (3) (2013) 177–185.
[62] V. Karthikeyani, I.P. Begum, K. Tajudin, I.S. Begam, Comparative of data mining classification algorithm (CDMCA) in diabetes disease prediction, Int. J. Comput. Appl. 60 (12) (2012).
[63] A. Julia, R. Vohra, P. Rani, Liver patient classification using intelligent techniques, Int. J. Comput. Sci. Inform. Technol. 5 (4) (2014) 5110–5511.
[64] H. Jin, S. Kim, J. Kim, Decision factors on effective liver patient data prediction, Int. J. BioSci. BioTechnol. 6 (4) (2014) 167–178.
[65] H. Sug, Improving the prediction accuracy of liver disorder disease with oversampling, Appl. Math. Electric. Comput. Eng. (2012) 331–335.
[66] B.V. Ramana, M.P. Babu, N.B. Venkateswarlu, A critical study of selected classification algorithms for liver disease diagnosis, Int. J. Datab. Manage. Syst. (IJDMS) 3 (2) (2011).
[67] S. Karthik, A. Priyadarishini, J. Anuradha, B.K. Tripathy, Classification and rule extraction using rough set for diagnosis of liver disease and its types, Adv. Appl. Sci. Res. 2 (3) (2011) 334–345.
[68] E.M.F. Houby, A framework for prediction of response to HCV therapy using different data mining techniques, Adv. Bioinform. 2014 (2014).
[69] M. Neshat, M. Sargolzaei, A.N. Toosi, A. Masoumi, Hepatitis disease diagnosis using hybrid case based reasoning and particle swarm optimization, Artif. Intell. (2012).
[70] D.A. Khan, F.T. Zuhra, F.A. Khan, A. Mubarak, Evaluation of diagnostic accuracy of apri for prediction of fibros is in hepatitis c patients, J. Ayub. Med. Coll. Abbottabad. (2008).
[71] K. Zolfaghar, N. Verbiest, J. Agarwal, N. Meadem, S.C. Chin, S.B. Roy, et al., Predicting risk-of-readmission for congestive heart failure patients: a multi-layer approach, 2013. Available from: <arXiv:1306.2094>.
[72] M.A.A. Burungale, D.A. Zende, Survey of Large-scale Hierarchical Classification.
[73] Ensemble Methods. <http://scikit-learn.org/stable/modules/ensemble.html> (accessed 05.10.15).
[74] S. Ali, A. Majid, A. Khan, IDM-PhyChm-Ens: intelligent decision making ensemble methodology for classification of human breast cancer using physicochemical properties of amino acids, Amino Acids 46 (4) (2014) 977–993.
[75] A. Majid, S. Ali, M. Iqbal, N. Kausar, Prediction of human breast and colon cancers from imbalanced data using nearest neighbor and support vector machines, Comput. Methods Programs Biomed. 113 (3) (2014) 792–808.
[76] S. Ali, A. Majid, Can-Evo-Ens: classifier stacking based evolutionary ensemble system for prediction of human breast cancer using amino acid sequences, J. Biomed. Inform. (2015) 256–269.
[77] A. Majid, S. Ali, HBC-Evo: predicting human breast cancer by exploiting amino acid sequence based feature spaces and evolutionary ensemble, Amino Acids 47 (1) (2014) 217–221.
[78] S.A. Pattekari, A. Parveen, Prediction system for heart disease using Naïve Bayes, Int. J. Adv. Comp. Math. Sci. 3 (3) (2012) 290–294.
[79] S. Aruna, S.P. Rajagopalan, L.V. Nandakishore, Knowledge based analysis of various statistical tools in detecting breast cancer, CCSEA, CS IT 02 (2011) 37–45.
[80] M.R.N. Kousarrizi, F. Seiti, M. Teshnehlab, An experimental comparative study on thyroid disease diagnosis based on feature subset selection and classification, Int. J. Electric. Comput. Sci. IJECS-IJENS 12 (01) (2012).
[81] A. Rajkumar, G.S. Reena, Diagnosis of heart disease using data mining algorithm, Glob. J. Comput. Sci. Technol. 10 (10) (2010).
[82] K. Polat, S. Sahan, S. Gunes, Automatic detection of heart disease using an artificial immune recognition system (AIRS) with fuzzy resource allocation mechanism and k-nn (nearest neighbour) based weighting preprocessing, Expert Syst. Appl. 32 (2007) 625–631.
[83] I. Kurt, M. Ture, A.T. Kurum, Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease, Expert Syst. Appl. 34 (1) (2008) 366–374.
[84] J.G. Liao, K. Chin, Logistic regression for disease classification using microarray data: model selection in a large pand small n case, Bioinformatics 23 (15) (2007) 1945–1951.
[85] Y. Ma, L. Zuo, J. Chen, Q. Luo, X. Yu, Y. Li, J. Xu, S. Huang, L. Wang, W. Huang, M. Wang, G. Xu, H. Wang, Modified glomerular filtration rate estimating equation for Chinese patients with chronic kidney disease, J. Am. Soc. Nephrol. (2006).
[86] E.R. Dubberke, K. Reske, M. Olsen, L. McDonald, V. Fraser, Short- and long-term attributable costs of *Clostridium difficile*-associated disease in nonsurgical inpatients, Clin. Infect. Dis. 46 (2008) 497–504.
[87] William A. Knaus, Elizabeth A. Draper, Douglas P. Wagner, Jack E. Zimmerman, APACHE II: a severity of disease classification system, Crit. Care Med. 3 (10) (1985).
[88] P.W.F. Wilson, R.B. D'Agostino, D. Levy, A.M. Belanger, H. Silbershatz, W.B. Kannel, Prediction of coronary heart disease using risk factor categories, Am. Heart Assoc. (1998) 1837–1847.
[89] B. Ster, A. Dobnikar, Neural networks in medical diagnosis: comparison with other methods, in: Proceedings of the International Conference on Engineering Applications of Neural Networks, 1996, pp. 427–430.
[90] N. Georgiou-Karistianis, M.A. Gray, D.J.F. Domínguez, A.R. Dymowski, I. Bohanna, L.A. Johnston, A. Churchyard, P. Chua, J.C. Stout, G.F. Egan, Automated differentiation of pre-diagnosis Huntington's disease from healthy control individuals based on quadratic discriminant analysis of the basal ganglia: the IMAGE-HD study, Neurobiol. Dis. 51 (2013) 82–92.
[91] M.Q. Zhang, Identification of protein coding regions in the human genome by quadratic discriminant analysis, Proc. Natl. Acad. Sci. U.S.A. 94 (2) (2015).
[92] J. Maroco, D. Silva, A. Rodrigues, M. Guerreiro, I. Santana, A. Mendonç, Data mining methods in the prediction of dementia: a real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests, BMC Res. Notes 4 (2011) 299.
[93] M. Drent, P.G. Mulder, S.S. Wagenaar, H.C. Hoogsteden, H. Velzen-Blad, J.M. Bosch, Differences in BAL fluid variables in interstitial lung diseases evaluated by discriminant analysis, Eur. Respir. J. (1993).
[94] S. Srivastava, M. Gupta, B. Frigyik, Bayesian quadratic discriminant analysis, J. Mach. Learn. Res. 8 (2007) 1277–1305.
[95] R. Das, I. Turkoglu, A. Sengur, Effective diagnosis of heart disease through neural networks ensembles, Expert Syst. Appl. (2009) 7675–7680.
[96] H. Chen, B. Yang, J. Liu, D. Liu, A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis, Expert Syst. Appl. 38 (2011) 9014–9022.
[97] V.N. Chunekar, H.P. Ambulgekar, Approach of neural network to diagnose breast cancer on three different data set, in: International Conference on Advances in Recent Technologies in Communication and Computing, 2009, pp. 893–895.
[98] H. Kahramanli, N. Allahverdi, Design of a hybrid system for the diabetes and heart diseases, Expert Syst. Appl. 35 (1–2) (2008) 82–89.
[99] M.C. Tu, D. Shin, D. Shin, Effective diagnosis of heart disease through bagging approach, in: 2nd International Conference on Biomedical Engineering and Informatics, 2009, pp. 1–4.
[100] F. Ricci, L. Rokach, B. Shapira, Introduction to Recommender Systems Handbook, Springer, US, 2011. pp. 1–35.