Journal of Network and Innovative Computing
ISSN 2160-2174 Volume 3 (2015) pp. 088-104

Computational Intelligence Data Analysis for Decision Support and Health Care Monitoring System

Abdel Hamid Salih Mohamed Salih[1], Ajith Abraham [2]

[1]Sudan University of Science and Technology
Faculty of Computer Science
Khartoum, Sudan
Hamidsalih39@yahoo.com

[2]Machine Intelligence Research Labs (MIR Labs)
Scientific Network for Innovation and Research Excellence, Washington 98071, USA
Email: ajith.abraham@ieee.org

***Abstract*:** Monitoring is a process of continuously gathering data and performing real-time analyses, monitoring can improve the estimation of the current state, identification of the critical situation, and assisted in decision support and planning. The past few years have an increase in the development of Ambient Intelligence health monitoring systems. An important aspect of investigation in such system is how the data is treated and analyzed. The survey of literature in this area presents that data mining analysis is lacking in Ambient Intelligence healthcare monitoring management. Though, there is good understanding of the importance healthcare systems by various authors, their focus was limited to a single aspect of the whole system and without integrated the analysis and decision support using machine learning and data mining methods. Hence, The goal of this paper is devoted to extensive investigation to construct a new novel ensemble health Care decision support for assisting an intelligent health monitoring system and also focus was to reduce the dimensionality of the attributes. Also the paper aims to discusses the findings of machine learning experiments and trend analysis on the simulation wearable sensors patients monitoring data. In the process of addressing the objectives of the paper indicated above, two major phases of experiments were conducted. In the first phase experiment, attempt has been made to investigate the experimental results of the performance of different classification techniques for classifying the data from different simulated wearable sensors used for monitoring different patients with different diseases. So as to construct the Base Classifiers Proposed used in the first experiment are: In the second phase experiment, we investigated various Meta classifiers. Finally new Novel Intelligent Ensemble method was constructed based of Meta classifier voting combining with three base classifiers J48, Random Forest and Random Tree algorithms. Different comparative analysis and evaluation were done using various evaluation methods like Error Metrics, ROC curves, Confusion Matrix, Sensitivity, Specificity and the Cost/Benefit methods. The results obtained show that the Novel Intelligent Ensemble method classifier is very efficient and can achieve high accuracy and, better outcomes that are significantly better compared with the outcomes of the all base classifiers proposed and all meta base classifiers.

***Keywords*:** Monitoring**,** wearable sensors, Base Classifiers, Meta base classifiers, Ensemble methods, Voting.

I. Introduction

Monitoring is a process of continuously gathering data and performing real-time analysis. Monitoring can improve the estimation of the current state, optimization of the business processes, identification of the critical situation and new opportunities, and prediction of the future state and planning. Basic elements of monitoring systems are monitoring processes. Systems monitoring can be decomposed into, gathering data, data preparation, measuring of performances and events, assessment, dissemination of results, interpretation of results. The assessment processes have an important role in the proposed methodology for developing healthcare monitoring and assessment. The assessment processes in this paper, used advanced data mining analysis techniques. In general, such systems are characterized by various data gathering methods such as environment sensors, as well as advanced data analysis methods that assess and predict critical situations. In a hospital healthcare monitoring system, it is necessary to constantly monitor the patient's vital signs, such as blood pressure (BP) and heart rate (HR) to control their health condition. In traditional healthcare monitoring in hospital vital signs monitoring will be done according to personal situation of patient. Often there are three patients situation in hospital. Patients in primary healthcare, patients in intensive care and patients in hospital rooms, all of them need vital signs monitoring to control their health condition. In this paper, we focus on patients in hospital rooms. Systems monitoring can be decomposed into, gathering data, Data preparation, Measuring of performances and events, Assessment, Dissemination of results, Interpretation of results. The assessment processes have an important role in the proposed methodology for developing healthcare monitoring and assessment. The assessment processes in this dissertation, used advanced data mining analysis techniques. Traditional healthcare and services are usually offered within hospitals or medical centers. Chronic diseases are becoming the major causes of the death such as insufficient cardiac heart, asthma, diabetes, and patients with Alzheimer's disease [1]. In traditional healthcare monitoring, experts measure the patient's vital signs and the data was recorded in patients vital signs sheet, so as to be presented to doctors for diagnosis purpose and later will be achieved at statistical office. However, this traditional healthcare monitoring is costly, inefficient and inconvenient for the people with the need of routine checks, since the patients need to frequently visit the hospital, sometimes on a daily basis, or even worse, need a long-stay. In addition, the use of these vital signs sheet shows serious limitations, of which the prime cause lies in the instruments and procedures used. There are huge requirements to move the routine medical check and healthcare services. Ambient Intelligence (AmI) for healthcare monitoring and personalized healthcare is a promising solution to provide efficient medical services, which could significantly lower down the healthcare cost and improve the health care monitoring. Many researchers have defined AmI in different ways. AmI proposes new ways of interaction between people and technology, making it suited to the needs of individuals and the environment that surrounds [2]. Wireless sensor networks (WSNs) are used for gathering the information needed by AmI environments. The principal device in a WSNs is the network node, also called mote. This device, battery powered, has the Radio Frequency Identification (RFID) for the transmission and the reception of the information, an interface between the module and the sensor and a microcontroller. The context is defined as any information used to characterize the situation of an entity, which can be a person, a place or an object [3]. This information is important for defining the interaction between users and the technology that surround them. For these reasons, it is necessary to continuously keep track of information about the users and their environment [1]. The information may consist of many different parameters such as vital signs (e.g. heart rhythm or blood pressure), etc. Thus, distributed sensors throughout the environment and even the users themselves can collect most of the context information. WSNs are used for gathering the information needed by AmI environments. Sensor data is collected from disparate sources and later analyzed to produce information that is more accurate, more complete, or more insightful than the individual pieces. Driving from the concept of AmI, [4] we simulated the environment of Baraha Medical City in Shambat, Khartoum North, in Sudan using the framework reported in [5], [6]. The monitoring system was for thirty patients with CD. For the sake of having a wide controlled analysis system, few works have designed and tested their data mining methods through shapely simulated wearable sensors data. Data simulation would be useful when focus of data processing method is on the efficiency and robustness of information extraction [7]. Another reason to create and use simulated data is the lack of long term and large-scale data sets [7], which helps the proposed data mining systems to deal with huge amount of data. In this paper our experiments are conducted on wearable sensors vital signs data set, which was simulated using a hospital environment. First, we carried out a thorough investigation comparing the performance of various base classifiers. Second, we carried out a thorough investigation comparing the performance of various Meta base classifiers. Third, we investigated Meta classifiers and new Novel Intelligent Ensemble method was constructed based of Meta classifier Voting combining with three base classifiers J48, Random Forest and Random Tree algorithms. The rest of this paper is organized as follows. Section 2 presents the related work and Section 3 describes the computational intelligence methods used and evaluation methods. Section 4 presents the data set and simulation environment of hospital environment. Section 5 pesents the first phase experiments and results. Section 6 presents the second phase experiments and results followed by models comparison, discussion and lessons learned in Section 6.

II. Related Work

There are several AI and DM methods and techniques used in analyzing sensors data in AmI such like Neural networks, fuzzy Rules, Reasoning, Decision making, and spatial-temporal reasoning and machine learning. These methods and techniques can help accomplish many important tasks in AmI assisted HCM and make the system more efficient [1]. Artificial Neural Network (ANN) is widely used for classification and prediction. [8] proposed an ANN based activity recognition system in order to determine the occurrence of falls. Also [9] presented a multi-layered feed forward neural network (FNNs) as activity classifiers and recognized 8 daily activities with an overall performance of 95%. Hagras, et al. [10] used fuzzy logic-based techniques to learn user preferences in an ordinary living environment. Corchado, et al. [6] developed GerAmI system in conjunction with the Alzheimer Sant ́sima Trinidad Residence of Salamanca, an institute with multiple stories, multiple rooms and upwards of 40 residents. As with all previously mentioned for AmI systems, the GerAmI uses sensors to record patient and user data. However rather than sensors using motion or heat to track users, each resident and staff wears a bracelet containing a unique RFID. As each bracelet's RFID is unique it allows all of the residents and staff to be tracked individually without false data being recorded. This system is unique in that it also tracks the movements of the staff members. This is a major benefit in a system such as this when the medical care providers are on hand as it allows faster reactions to emergencies by alerting staff that are on duty and also located closer to the source of the problem. If intervention or assistance is required a message is sent to the staff members personal digital assistant (PDA). The message contains the name or identifier of the patient in question, the problem that has occurred as well as information from the database about the best way to deal with the situation based on previous events. Doctor et al. [11] developed the iDorm research and focuses on automating a living environment. However, instead of a Markov model, they model resident behavior by learning fuzzy rules that map sensor state to actuator readings representing resident actions. The amount of data created by sensors can create a computational challenge for modeling algorithms. However, the challenge is even greater for researchers who incorporate audio and visual data into the resident model. Activity Prediction and Recognition is widely used in DM and ML field that helps to identify events, which have not yet occurred. Some researchers have used offline data analysis in Activity Prediction and Recognition. Van Laerhoven and Cakmakci [12] proposed a method based on Kohonen Self Organizing Maps (KSOMs) and k-Means clustering, which is able to identify typical motion profiles. This approach relies on active training, used to construct a supervised context transition profile based on a first order Markov process to make the KSOM training procedure converge, the neighborhood radius of the learning neurons must decrease over time [13]. However KSOM have strong dependence of the initialization and is has too unbalanced classes, and also K- Means clustering has problems when clusters are of differing sizes, densities, non-globular shapes, and empty clusters. Duda, et al. [14] presented critical events that can be detected using classification algorithms, for which Bayes classifiers are known to provide good results. However, traditional classifiers do not allow meaningful interruption until the entire model has been evaluated, which is crucial in mobile devices due to limited resources. Limited processing power and high data rates limit the time available for processing one set of sensor values. To overcome this limitation they employed a novel anytime Bayes classifier in a two-phase architecture. On the back-end server a full index structure is stored, which is an extension of previous work was presented by Assent et al., [15] for anytime stream classification. It is trained by sequences of sensor measurements, which correspond to normal situations. Patel, et al. [16] implemented Support Vector Machines (SVM's) to predict clinical scores of the severity of data obtained from wearable sensors in patients with Parkinson's disease. SVM's have the ability to generate nonlinear decision boundaries, by mapping the feature space into a higher dimensional space (using kernels) where classes are linearly separable. Cook and Das [17] presented an AmI application, which is focused on a single environment, which has outfitted with sensors and designed to improve the experience of the resident in the environment., Mozer [18] presented the Neural Network House, and the MavHome was presented by [19]. Helal, et al. [20] developed predicting resident locations, and even resident actions that allow the AmI system to anticipate the resident's needs and assist with (or possibly automate) performing the action. The modeling techniques described so far can be characterized as unsupervised learning approaches. However, if resident activity data is available, then supervised learning approaches can be used to build a model of resident activity and use it to recognize observed activities. Tapia, et al. [21] employed a naive Bayes learner to identify resident activity from among a set of 35 possible classes, based on collected sensor data. However Naïve Bayes is simple probabilistic classifier based on the assumption that the features for a given class are mutually independent, which means that the decisions are made as if all features are equally important. Philipose, et al. [22] enhanced the model with object interaction data. Over the last few years, supporting technologies for AmI have emerged. Automated decision-making and control techniques are available for Building a fully automated AmI application. Simpson, et al. [23] discussed how AI planning systems could be employed to not only remind individuals of their typical next daily activity, but also to complete a task if needed. Augusto and Nugent [24] described the use of temporal reasoning with a rule-based system to identify hazardous situations and return an environment to a safe state while contacting the resident. Few fully implemented applications decision-making technologies have been implemented. Youngblood [25] also have used a reinforcement learner to automate physical environments with volunteer residents, in the MavPad apartment and the MavLab workplace. Amigoni, et al.[26] employed a Hierarchical Task Network (HTN) planner to generate sequences of actions and contingency plans that will achieve the goal of the AmI algorithm. For example, the AmI system may respond to a sensed health need by calling a medical specialist and sending health vitals using any available device (cell phone, email, or fax). If there is no response from the specialist, the AmI system would phone the nearest hospital and request ambulance assistance. [27], has developed novel computer systems enhancing the quality of life of people suffering from Alzheimer's disease and similar disorders, that help an individual perform daily tasks by sensing the individual's location and environment, learning to recognize patterns of behavior, offering audible and physical help, and decision making to alerting caregivers in case of danger. Beck and Pauker [28] described dynamic sequential decision making in medicine using Markov-based approach originally described in terms of medical decision-making. Xiang and Poh [29] utilized dynamic influence diagrams. There are also other's approaches for example. [30] and [31] have utilized decision trees to model temporal decisions. In all cases, the goal is to determine optimal sequences of decisions. Markov decision processes (MDPs) is an efficient technique for determining optimal sequential decisions (termed a "policy") in dynamic and uncertain environments have been used by Schaefer et al., [32] and Alagoz et al., [33]. Several studies have focused on the importance of the ensemble methods in the field of medical health care monitoring. These studies have applied different approaches to the given problem and achieved high classification accuracies. Ensemble methods combined a set of individual methods to obtain a better more accurate and reliable estimates or decisions than can be obtained from using a single model. Classification of sensory data is a major research problem in WSNs. Many researchers have utilized ensemble models in AmI assisted HCM. Fatima, et al. [34] presented Classifier Ensemble Optimization method for activity recognition by optimizing the output of multiple classifiers with evolutionary algorithm. They have combined the measurement level output of different classifiers in terms of weights for each activity class to make up the ensemble. Classifier ensemble learner generates activity rules by optimizing the prediction accuracy of weighted feature vectors to obtain significant improvement over raw classification. Tan and Gilbert [35] presented a comparison of single supervised machine learning and ensemble methods in classifying seven publicly available cancerous data. The authors used C4.5 decision tree, bagged decision tree on seven publicly available cancerous micro array data, and compared the prediction performance of these methods. The experimental results indicate that the ensemble methods consistently perform well over all the datasets in terms of their specificity. A combinational feature selection and ensemble neural network method is introduced by Liu et al., [36] for classification of biomedical data. Many individual algorithms such as self-organizing maps (SOM), learning vector quantization (LVQ), multi-layer perceptron's (MLPs), neural-fuzzy systems, and SVMs were applied to ECG signals. However, these methods have been typically applied to distinguish normal signals from abnormal signals across patients. This is difficult because of the substantial variation in the morphologies of ECG signals across patients. For this reason, Li et al., [37] implemented an ensemble consisting of a standard SVM designed to distinguish normal signals from abnormal signals across patients and a set of one-class SVMs, presented by Scholkopf et al., [38], (one per patient) to distinguish normal signals for a given patient from all other signals [39]. Tu, et al. [40] proposed the use of bagging with C4.5 algorithm, bagging with Naïve bayes algorithm to diagnose the heart disease of a patient. Other study [41] used bagging algorithm to identify the warning signs of heart disease in patients and compared the results of decision tree induction with and without bagging. Chaurasia and Pal [42] used Naive Bayes, J48 Decision Tree and Bagging algorithm to predict the survivability for Heart Diseases patients. Wen [43] conducted experiments on ECG data to identify abnormal high frequency electrocardiograph using decision tree algorithm C4.5 with bagging. Ensemble methods have been used by others researchers in other type of medical data sets. Kaewchinporn, et al. [44] presented a new classification algorithm TBWC combination of decision tree with bagging and clustering.

III. Computational Intelligence

Base Classifiers Used

A brief description of the single classification algorithms, used in this paper is presented in the next subsections.

- **Decision Tree Algorithm J48**

J48 classifier is a simple C4.5 decision tree for classification. Quinlan [45] developed C4.5 algorithm, which is used to generate a Decision Tree. It creates a binary tree. The decision tree approach is most useful in classification problem. With this technique, a tree is constructed to model the classification process. Once the tree is built, it is applied to each tuple in the database and results in classification for that tuple [45]. The C4.5 unlike the IDE3, accepts both continuous and categorical attributes in building the decision tree. It has an enhanced method of tree pruning that reduces misclassification errors, due to noise or too-much detail in the training data set. Decision Trees are produced from the J48 i.e.

- **Partial Decision Trees (PART)**

Frank and Witten describe a rule induction approach without the need for applying a global optimization strategy to generate appropriate rules [46]. PART (Partial Decision Trees) adopts the divide-and-conquer strategy of RIPPER [47] and combines it with the decision tree approach of C4.5 [45]. More precisely, PART generates a set of rules according to the divide-and-conquer strategy, removes all instances from the training collection that are covered by this rule and proceeds recursively until no instance remains. To generate a single rule, PART builds a partial decision tree for the current set of instances and chooses the leaf with the largest coverage as the new rule. Afterwards, the partial decision tree is discarded, which avoids early generalization.

- **Logistic Model Trees (LMT)**

Logistic Model Trees consist of a decision tree structure with logistic regression function at the leaves [48] . As in decision tree, the tested attributes are associated with every inner node. The attributes with *k* values, the node has *k* child nodes for nominal attributes and depending on the value of the attribute the instances are sorted down. For the numeric attributes, the node has two child nodes and comparing the attributes of tested value to a threshold (the instances are sorted down based on threshold [49]. LMT uses pruning of cost complexity. Compared to other algorithm, it is slower to compute.

- **Logit Boost algorithm**

The LogitBoost algorithm was introduced by Friedman et al., [50]. The algorithm is similar to AdaBoost, with the main difference being that LogitBoost performs stage wise minimization of the negative binomial log likelihood, while AdaBoost performs stage wise minimization of the exponential loss. By virtue of using the binomial log likelihood instead of the exponential loss, the LogitBoost algorithm was believed to be more "gentle" and consequently likely to perform better than AdaBoost for classification problems in which the Bayes error is substantially larger than zero.

- **Random Forest**

Random Forest developed by Breiman, [49] is a group of un-pruned classification or regression trees made from the random selection of samples of the training data. Random features are selected in the induction process. Prediction is made by aggregating (majority vote for classification or averaging for regression) the predictions of the ensemble. Each tree is grown as described in [51]**.** By Sampling N randomly, if the number of cases in the training set is N but with replacement, from the original data. This sample will be used as the training set for growing the tree. For M number of input variables, the variable m is selected such that m << M is specified at each node, m variables are selected at random out of the M and the best split on these m is used for splitting the node. During the forest growing, the value of m is held constant. Each tree is grown to the largest possible extent. No pruning is used. Random Forest generally exhibits a significant performance improvement as compared to singletree classifier such as C4.5. The generalization error rate that it yields compares favorably to Adaboost, however it is more robust to noise. The random forest inducing algorithm is derived from random decision a forest that was proposed by Tin Kam Ho of Bell Labs in 1995 [52]. This method combines random selection of features to construct a decision tree with controlled variations. The tree is constructed using algorithm as detailed below.

i) Let N be the number of training classes and M be the number of variables in classifier.

ii) The input variable m is used to determine the node of the tree. Note that m<M.

iii) Choosing n times of training sets with the replacement of all available training cases N by predicting the classes, estimate the error of the tree.

iv) Choose m variable randomly for each node of the tree and calculate the best split.

v) At last the tree is fully grown and it is not pruned.

The tree is pushed down for predicting a new sample. When the terminal node ends up the label is assigned the training sample [53]. This procedure is iterated over all trees and it is reported as random forest prediction.

- **Random Tree**

A random tree is a tree constructed randomly from a set of possible trees having *K* random features at each node. Random trees can be generated efficiently and the combination of large sets of random trees generally leads to accurate models. A random tree is a tree formed by stochastic process. Types of random trees include Uniform spanning tree, Random minimal spanning tree, Random binary tree, Random recursive tree, Treap, Rapidly exploring random tree, Brownian tree, Random forest and branching process [49].

- **K- nearest neighbor IBK**

IBK (commonly known as K- nearest neighbor). instance-based learning is lazy [53], deferring the real work as long as possible, whereas other methods are eager, producing a generalization as soon as the data has been seen. In instance-based learning, each new instance is compared with existing ones using a distance metric, and the closest existing instance is used to assign the class to the new one. This is called the nearest-neighbor classification method. Sometimes more than one nearest neighbor is used, and the majority class of the closest k neighbors (or the distance- weighted average, if the class is numeric) is assigned to the new instance. This is termed the k-nearest-neighbor method. Computing the distance between two examples is trivial when examples have just one numeric attribute: it is just the difference between the two attribute values. It is almost as straightforward when there are several numeric attributes: generally, the standard Euclidean distance is used.

A. **Meta-learning**

Meta-learning aims to compute a number of independent classifiers by applying learning programs to a collection of independent and inherently distributed databases in parallel. The “base classifiers” computed are then collected and combined by another learning process. The most popular meta-learning algorithms are bagging and boosting. Bagging [54], is a method for generating multiple classifiers (learners) from the same training set. The final class is chosen by, e.g., voting. Combining multiple models approach is to making decisions more reliable is to combine the output of different models. They can all, more often than not, increase predictive performance over a single model. And they are general techniques that can be applied to numeric prediction problems and to classification tasks.

- **AdaBoostM1**

AdaBoost.M1is a well-known algorithm for boosting weak classifiers [55]. AdaBoostM1 is a member of a broader family of iterative machine learning algorithms that build the final classifier through a finite series of improvements to the classifier. AdaBoost.M1 is the most straightforward generalization of boosting algorithm. It is adequate when the weak learner is strong enough to achieve high accuracy.

- **LogitBoost**

One of the boosting algorithms developed recently, is introduced for predicting protein structural classes. Boosting was originally proposed to combine several weak classifiers to improve the classification performance. Later on, Freund and Schapir proposed a more capable and practical boosting algorithm, the so-called AdaBoost. [56]. Ada- Boost, an abbreviation for Adaptive Boosting, is a metal earning algorithm. It tries to build a weak classifier iteratively on others according to the performance of the previous weak classifiers.

- **Bagging**

The term bagging (for “bootstrap aggregating”) was coined by Breiman [57], who investigated the properties of bagging theoretically and empirically for both classification and numeric prediction. [57]. Combining the decisions of different models means amalgamating the various outputs into a single prediction. The simplest way to do this in the case of classification is to take a vote (perhaps a weighted vote); in the case of numeric prediction, it is to calculate the average (perhaps a weighted average). Bagging and boosting both adopt this approach, but they derive the individual models in different ways. In bagging, the models receive equal weight, whereas in boosting, weighting is used to give more influence to the more successful ones—just as an executive might place different values on the advice of different experts depending on how experienced they are. Bagging (bootstrap aggregating), generates a collection of new sets by resampling the given training set at random and with replacement. These sets are called bootstrap samples. New classifiers are then trained, one for each of these new training sets. They are amalgamated via a majority vote. [55], [57].

- **Stacking**

Stacked generalization [53], originated with[58], who presented the idea in the neural network literature, and was applied to numeric prediction by [57]. Stacked generalization, is a different way of combining multiple models. Although developed some years ago, it is less widely used than bagging and boosting, partly because it is difficult to analyze theoretically and partly because there is no generally accepted best way of doing it—the basic idea can be applied in many different variations. Unlike bagging and boosting, stacking is not normally used to combine models of the same type. The usual procedure would be to estimate the expected error of each algorithm by cross validation and to choose the best one to form a model for prediction on future data.

- **Random Committee**

Classifier that ensembles randomizable base classifiers, it builds an ensemble of base classifiers and averages their predictions. Each one is based on the same data but uses a different random number seed. This only makes sense if the base classifier is randomized; otherwise, all classifiers would be the same. The random committee algorithm is a diverse ensemble of random tree classifiers. In the case of classification, the random committee algorithm generates predictions by averaging probability estimates over these classification trees. [53].

B. Ensemble methodology

The main purpose of an ensemble methodology is to combine a set of models, each of which solves the same original problem, in order to obtain a better composite global model with more accurate and reliable estimates or decisions than can be obtained from using a single model. The main discovery is that the ensemble classifier is constructed by ensemble machine learning algorithms, such as bagging and boosting approaches, often performs much better than the single classifiers that make them up. The idea of ensemble methodology is to build a predictive model by integrating multiple models. It is well known that ensemble methods can be used for improving prediction performance. The learning procedure for ensemble algorithms can be divided into the following two parts. [59].

1. Constructing base classifiers/base models: the main tasks of this division are:
   (a) Data processing: Prepare the input training data for building base classifiers and attributes selection to reduce the dimensionality of the attributes.
   (b) Base classifier constructions: build base classifiers on the data set with a learning algorithm.
2. Voting: the second stage of ensemble methods is to combine the base classifiers models built in the previous step into the final ensemble model.

- **Voting**

There are various kinds of voting systems. Two main voting systems are generally utilized, namely weighted voting and un-weighted voting. In the weighted voting system, each base classifier holds different voting power. On the other hand, in the unweight system, individual base classifier has equal weight, and the winner is the one with most number of votes. The simplest kind of ensemble is the way of aggregating a collection of prediction values base level giving different voting power for its prediction. The final prediction obtains the highest number of votes. Voting includes the weighted average (of each base classifier holds) when using regression problem and majority voting when doing classification and the weighted-majority output is given by:

$$\arg\max\left(\sum_{i=1}^{k} P_i(x), w_i\right) \quad (1)$$

Pi(x) is the results of the prediction of ith prediction model and Pi(x, w) is indicator function defined as:

$$p_i(x,w)=\begin{cases}1 & x=w\\ 0 & otherwise\end{cases} \quad (2)$$

Majority voting has some benefits that it does not require any additional complex computation and any previous knowledge. However, this approach leads to the result that it is difficult to analyze and interpret. The second strategy is un-weighted, which gives some predictor higher weight if they achieve more accuracy than others (the winner is the one with the most number of votes) [60], [61]. Combining rules are the simplest combination approach and it is probably the most commonly used in the multiple classifier system [62]. This combination approach is called non-trainable combiner, because combiners are ready to operate as soon as the classifiers are trained and they do not require any further training of the ensemble as a whole [63]. A theoretical framework for fixed rules combination was proposed by [64]. It includes the sum, product, max, min, average and median rules. In this Thesis we have used the Maximum rule. Maximum rule is based on information provided by the maximum value of:

$$P\left(x^i \mid w_k\right) \quad (3)$$

Across all class labels, it finds the maximum score of each class between the classifiers and assigns the input pattern to class with the maximum score among the maximum scores as following [63].

$$f(x)=w_j\ ,\ j=\arg\max\left(\max p\left(x^i \mid w_k\right)\right) \quad (4)$$

As shown in Figure 1, the dataset (which are simulation sensors data in our case) are used to train and test the system, each classifier in the system is trained using the training data set, and then give an output. The outputs of all classifiers are combined using one of fixed rules that mentioned previously to give the final decision. In this Thesis the author investigated Meta classifiers and a new Novel Intelligent Ensemble method was constructed based of Meta classifier Voting combining with three base classifiers J48 [45], Random Forest [51] and Random Tree [49] algorithms.

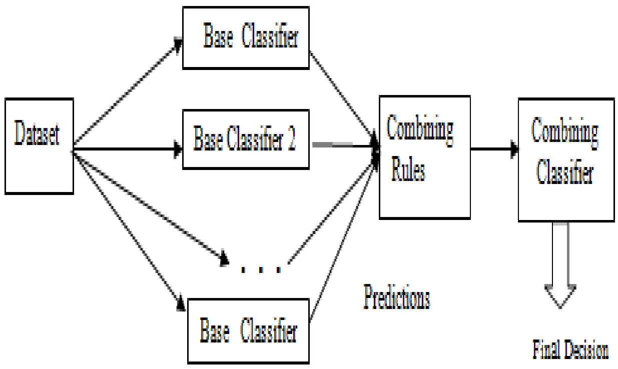

Figure 1: Ensemble using combination rule with voting

C. Cross-validation

In this paper the author applied a 10-fold cross validation test option. Cross-Validation (CV) is a statistical method of evaluating and comparing learning algorithms by dividing data into two segments: one used to learn or train a model and the other used to validate the model. The basic form of CV is k-fold CV. In k-fold CV the data is first partitioned into k equally (or nearly equally) sized segments or folds. Subsequently k iterations of training and validation are performed such that, within each iteration a different fold of the data is held-out for validation while the remaining k -1 folds are used for learning. The advantage of K-Fold Cross validation is that all the examples in the dataset are eventually used for both training and testing [4].

D. Attribute selection

Attribute selection (AS), also called feature selection. It is often an essential data processing step prior to applying a learning algorithm. Reduction of the attribute dimensionality leads to a better understandable model and simplifies the usage of different visualization technique and is the process of identifying and removing as much irrelevant and redundant information as possible. Reduces the dimensionality of the data, may allow learning algorithms to operate faster and more effectively and, accuracy can be improved later on future classification. It finds minimum set of attributes such that resulting probability distribution of data classes is as close as possible of original distribution. Methods used for AS can be classified into two types. The filter approach and Wrapper approach The filter approach actually precedes the actual classification process. The filter approach is independent of the learning algorithm, computationally simple fast and scalable. Using filter method, AS is done once and then can be provided as input to different algorithms. [65]. Wrapper approach uses the method of classification itself to measure the importance of attribute set, hence the AS depends on the algorithm model used. Wrapper methods are too expensive for large dimensional database in terms of computational complexity and time since each attribute set considered must be evaluated with the classifier algorithm used. Filter methods are much faster than wrapper methods and therefore are better suited to high dimensional data sets. Some of these filter methods do not perform attribute selection but only attribute ranking hence they are combined with search method when one needs to find out the appropriate number of attributes. Such filters are often used with forward, backward elimination, bi-directional search, best-first search, and other methods [65], [66]. Various AS techniques have been proposed in the ML/DM literature. In this Thesis, we used WEKA tool [67] for pre-processing experiments, to reduce the attributes dimensionality and formulated a new dataset, which was derived from the original dataset after implementing several AS algorithms, Such as:

- Correlation-based Feature Selection (CFS)

CFS is a filter algorithm that ranks feature subsets according to a correlation based heuristic evaluation function.CFS assumes that useful feature subsets contain features that are predictive of the class but uncorrelated with one another. CFS computes a heuristic measure of the “merit” of a feature subset from pair-wise feature correlations and a formula adapted from test theory. Heuristic search is used to traverse the space of feature subsets in reasonable time; the subset with the highest merit found during the search is reported.[65].

- Principal Component Analysis (PCA)

PCA technique reduces the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set. This is achieved by transforming to a new set of variables, the principal components (PCs), which are uncorrelated, and which are ordered so that the first few retain most of the variation present in all of the original variables. [68].

- Gain Ratio (GR) attribute evaluation

GR is a modification of the information gain that reduces its bias. GR takes number and size of branches into account when choosing an attribute. It corrects the information gain by taking the intrinsic information of a split into account. Intrinsic information is entropy of distribution of instances into branches (i.e. how much info do we need to tell which branch an instance belongs to). Value of attribute decreases as intrinsic information gets larger. [69].

- Relief Attribute Evaluation

The main idea of Relief algorithm [70], is to evaluate and estimate the quality of attributes according to distinguishing values between the instances that are near to each other. Both Relief and its extension ReliefF [71], are aware of the content information and can correctly estimate the quality of attributes in classification tasks with strong dependencies between attributes. [72].

- Wrapper attributes Selection

It depends on an induction algorithm to estimate the merit of feature subsets [65].

E. Evaluation Approaches and Techniques

We evaluate our classifiers by measuring their performance by various methods and performance matrices. The following methods are used in our experiments.

- Evaluation of time to build a model for each classifier.
- Mean Absolute Error (MAE):
- Root Mean Squared Error (RMSE)
- Kappa Statistics (KS)
- ROC curves.
- AUC (Area Under ROC Curve) is also taken under consideration.
- Confusion Matrix
- Cost/Benefit methods

IV. Data Set and Simulation Environment of Hospital Environment

A. Software simulation platform

In this research, the researcher used the GerAmi framework reported in [5],[6], for simulating the Hospital environment of Elbaraha Medical City in Shambat, Khartoum North, Sudan with 30 patients. Platform is built in Java and C# with GUI compatibly with mobile phones and personal computers. Depending on the patient’s ailment, different number of sensors was attached and the data is transmitted to a central server using tiny low power sensor devices. The sensor uses an 8MHz CPU with 10 KB RAM and provided 100 meters range. Various other sensors are attached depending on the patient’s requirement. *Example:*

1. Oxymeter: Heart rate and blood oxygen saturation levels with current location
2. ECG: Samples signal 12 bits @ 120 Hz
3. Motion Capture and EMG Sensors: Special-purpose sensors to capture limb motion and muscle activity.

How the sensors transmit data?

-- Sensors locally filter, compress, or analyze data to reduce radio congestion

-- Data from critical patients given higher priority Scalable platform to handle 100's of real patients.

B. Data Set and Simulation of Hospital Environment

We simulated the environment of Baraha Medical City in Shambat, Khartoum North, Sudan using the framework reported in [5], [6]. The hospital is situated in a 600 Square meter lot with a garden within the compound. The hospital has five floors with a 75-bed capacity and provides complete medical services for patients. The Hospital receives patients who suffer from chronic diseases such as heart diseases, asthma, diabetes and abnormal blood pressure etc. Also people in post-surgery state needs continuous monitoring of their health condition, especially the vital signs, until their health status becomes stable. Each patient when arrived to the hospital with chronic diseases, the patient first received primary healthcare treatment. Then the patient will be assigned to individual room. In our simulation, we allocated 6 chronic ill patients in each floor (total 30 patients) as we focused only on the monitoring and providing medical service for patients with chronic or terminally ill diseases. Depending on the critical condition of the patent, each patient was attached with several sensors. For thirty patients, there were a total of 300 readings at any measuring instant. As the results of the simulation, we obtained the wearable sensors monitoring. In this project, our main task is to develop a decision support system that could assist the hospital management to assess the situation of the hospital as Normal or Abnormal (too many medical emergencies) so that more medical help could be sorted. The simulated dataset consist of 750 instances and 300 attributes.

C. Pre-Processing

Data preparation or preprocessing is always important in machine learning and pattern recognition process. Though, there are various types of preprocessing tasks like handling missing values, minimizing noises, Transformation: changes the forms of the data into the ones appropriate for the data mining task by using different operations, dimensionality reductions etc. Transformation of dataset was done to meet our data-mining task in this research. In the data set obtained from simulation wearable sensors patient’s monitoring data, there is no missing data and also no noises.

D. Attribute Selection (AS)

Evaluator	Search Method	Final No of Attributes
Correlation-based Feature Selection (CFS)	Best first	6
Correlation-based Feature Selection (CFS)	Greedy Step wise (forwards)	6
Gain Ratio (GR) attribute evaluation	Attribute Ranker	300
Principal Component Analysis (PCA)	Attribute ranking	10
Relief Attribute Evaluation	Attribute ranking	300
Wrapper attributes Selection	Greedy Stepwise (forwards)	-
Wrapper attributes Selection	Best first	-

Table 1: Performance of the evaluator and search method used

Attribute Selection (AS) plays an important role in classification. This is one of the Preprocessing techniques in data mining. In this paper we investigated available attribute selection methods (Evaluators) and Search Methods in Weka tool [67], to the Dataset using full training set (300 attribute . We found that Correlation-based Feature Selection (CFS) Evaluator with Best first, search methods reduces the dimensionality of the attributes to six attributes. Table 1 summarizes the results.

V. First Phase Experiments and Results

The Base algorithms Proposed

The aim of this Section is to investigate the experimental results of the performance of different classification techniques for the simulation wearable sensors dataset to select the base classifiers with highest performance accuracy to be used in the second phase in the next section. The performance factors used for analysis are accuracy and error measures. The accuracy measures are TP rate, F Measure, ROC area, Sensitivity and Specificity. The error measures are Mean Absolute Error, Root Mean Squared Error and Kappa Statistics. We investigated algorithms available in WEKA [67] with our dataset using cross-validation with 5 fold and 10 fold with test options available. Finally we found that cross-validation give the best classification using 10 Fold cross-validation with the reduced the dimensionality obtained in the previous step to six attributes selection . Then we selected algorithms with classification accuracy between 90% to 100% as the proposed Base Classifiers. The Base algorithms Proposed in our investigation in this research are: K- nearest neighbor (IBk), Attribute Selected algorithm, Bagging, Random Committee, Rule-based learning (PART), Decision tree algorithm J48, Logistic Model Trees(LMT), Random Forest, Random Tree, as illustrated in Table 2 the correctly classified for each base classifier in term of percentile using 5 and 10 fold cross-validation. Table 3 depicts time required to build the model for each algorithm. From table 3 it is inferred that Random Tree model and IBK classifies are very quickly in comparison to other models.

6 Selection Attribute		
Test Options: Cross-Validation		
	5 Fold	**10 Folds**
Classifiers	**Correctly Classified**	**Correctly Classified**
IBk	88.7248 %	90.3356 %
Attribute Selected Classifier	89.1275 %	91.9463 %
Bagging	88.4564 %	90.4698 %
Random Committee	93.557 %	95.0336 %
PART	89.9329 %	91.8121 %
J48	89.6644 %	92.8859 %
LMT	92.6174 %	92.2148 %
Random Forest	92.8859 %	94.2282 %
Random Tree	92.8859 %	94.8993 %

Table 2: Base classifiers proposed

Algorithm	Time taken to build model (Sec)
IBk	0.001
Attribute Selected Classifier	0.14
Bagging	0.12
Random Committee	0.05
PART	0.04
J48	0.02
LMT	2.46
Random Forest	0.04
Random Tree	0.001

Table 3: Time required building the model

Table 4 depicts the various error metrics analyzed in the data set. It is inferred from Table 4, that Random Committee has the highest Kappa Statistic value and also has better accuracy compared with the others classifiers.

Algorithm	MAE	RMSE	KS	Correctly Classified
IBk	0.0978	0.3104	0.8062	673 90.33 %
Attribute Selected Classifier	0.1008	0.2631	0.8384	685 91.94 %
Bagging	0.1527	0.2609	0.8089	674 90.46 %
Random Committee	0.0643	0.1931	0.9004	708 95.03 %
PART	0.101	0.264	0.8355	684 91.81 %
J48	0.0865	0.2518	0.8574	692 92.88 %
LMT	0.0854	0.2454	0.844	687 92.21 %
Random Forest	0.0961	0.219	0.8843	702 94.22 %
Random Tree	0.051	0.2258	0.8977	707 94.89 %

Table 4: Performance measures comparison

In the next step we investigated and analyzed the performance of the Base Classifiers Proposed obtained in the previous step using the same wearable sensors dataset with the reduced number of attributes. The investigation and analysis of the performance is based on: Classifier performance in term of recall precision, f measure and false alarm rate; Classification performance for normal class; Classification performance for abnormal class and Classification performance of each classifier in term of Sensitivity and Specificity. Table 5 depicts the classifier performance of each classifier in term of recall precision, *f* measure and false alarm rate. It is inferred from Table 5 that Random Committee model has the highest precision and lowest false alarm rate, and the same recall as Radom Tree.

Algorithm	**Recall**	**Precision**	**F-measure**	**False alarm rate**
IBk	0.916	0.905	0.911	0.085
Attribute Selected classifier	0.914	0.914	0.913	0.076
Bagging	0.893	0.905	0.898	0.084
Random Committee	0.938	0.957	0.947	0.038
PART	0.924	0.908	0.914	0.080
J48	0.9185	0.9316	0.924	0.061
LMT	0.905	0.9316	0.917	0.062
Random Forest	0.927	0.951	0.938	0.044
Random Tree	0.9383	0.9544	0.945	0.041

Table 5: Performance in term of recall precision, *f* measure and false rate

As an example of classifier error illustration, Figure 2 depicts the Classifier error of Random Committee. The blue crosses indicate the Normal class and red crosses indicate the Normal class and squares indicate not classified.

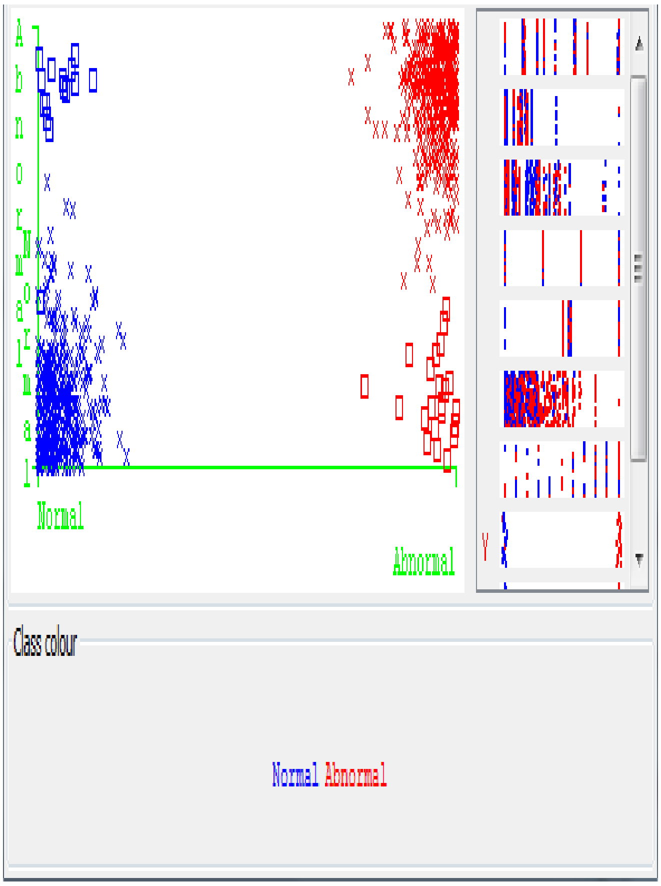

Figure 2: Classifier error of random committee

Table 6 depicts the algorithm performance of each classifier in term of recall precision and *f* measure for Normal class is summarized. Figure 4 depicts the Area under ROC of Random Committee classifier with highest area under ROC. Table 7 depicts the classifier performance of each classifier in term of recall precision, and f measure for abnormal class.

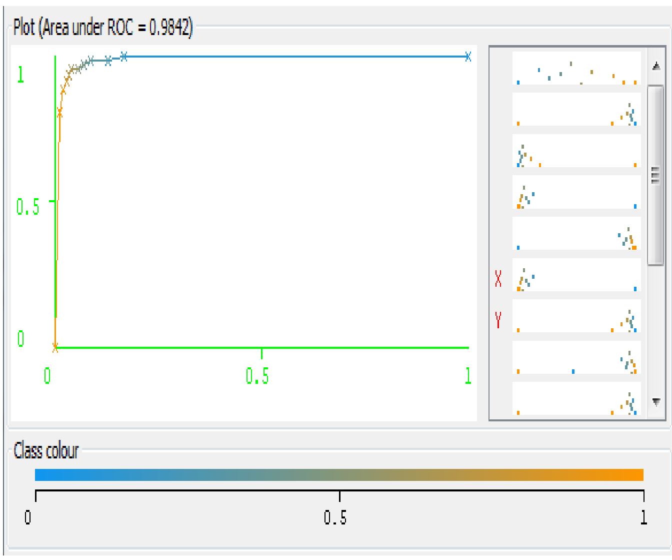

Figure 3: AUR of random committee classifier

Classifiers	TP Rate	FP Rat	Precision	Recall	F-Measure	ROC Area
IBk	0.883	0.109	0.878	0.883	0.881	0.891
Attribute selected classifier	0.906	0.122	0.869	0.906	0.887	0.926
Bagging	0.880	0.112	0.875	0.880	0.878	0.953
Random Committee	0.943	0.071	0.922	0.943	0.932	0.984
PART	0.926	0.124	0.869	0.926	0.897	0.955
J48	0.903	0.109	0.881	0.903	0.892	0.934
LMT	0.886	0.094	0.894	0.886	0.890	0.948
Random Forest	0.937	0.084	0.909	0.937	0.923	0.973
Random Tree	0.932	0.074	0.919	0.932	0.925	0.929

Table 6: Classification performance for normal class

Classifiers	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
IBk	0.901	0.094	0.915	0.901	0.908	0.910
Attribute selected classifier	0.878	0.094	0.913	0.878	0.895	0.926
Bagging	0.888	0.120	0.893	0.888	0.891	0.953
Random Committee	0.929	0.057	0.948	0.929	0.938	0.984
PART	0.876	0.074	0.930	0.876	0.902	0.955
J48	0.891	0.097	0.912	0.891	0.901	0.934
LMT	0.906	0.114	0.899	0.906	0.903	0.948
Random Forest	0.916	0.063	0.943	0.916	0.929	0.973
Random Tree	0.926	0.068	0.938	0.926	0.932	0.929

Table 7: Classification performance of each classifier for abnormal class

Table 8 depicts the classification performance of each classifier in term of Sensitivity and Specificity with the Random Committee model having the highest Specificity and also high Sensitivity. Random Committee model also has the highest accuracy and the IBK model has the lowest accuracy.

Classifiers	Sensitivity	Specificity	Accuracy
IBk	0.8907	0.914	0.9033
Attribute Selected Classifier	0.941	0.923	0.9194
Bagging	0.8932	0.915	0.9046
Random Committee	0.938	0.961	0.9503
PART	0.924	0.92	0.9221
J48	0.918	0.938	0.9288
LMT	0.905	0.937	0.9221
Random Forest	0.927	0.955	0.9422
Random Tree	0.938	0.958	0.9489

Table 8: Performance model in term of sensitivity and specificity.

Table 9 depicts overall algorithm performance ranked by accuracy. It inferred from Table 9 that Random Committee has the highest accuracy and the IBK model has the lowest accuracy comparing with the rest of the classifiers in this research.

Algorithm	Accuracy
Random Committee	0.9503
Random Tree	0.9489
Random Forest	0.9422
J48	0.9288
PART	0.9221
LMT	0.9221
Attribute Selected Classifier	0.9194
Bagging	0.9046
IBk	0.9033

Table 9. Overall algorithm performance ranked by accuracy

VI. Second Phase Experiments and Results

Novel Ensemble Decision Support and Health Care Monitoring System

As stated in methodology section and according to literature, in this thesis the main purpose of an ensemble methodology is to combine a set of models, each of which solves the same original problem, in order to obtain a better composite global model with more accurate and reliable estimates or decisions than can be obtained from using a single model. The aim of this Section is devoted to extensive investigation to construct a new novel ensemble healthcare decision support for assisting an intelligent health monitoring system. Experiments are conducted using the same dataset with the reduced dimensionality of the attributes obtained in the phase one experiments. This second phase consist of two steps. First step, extensive investigation of the experimental results of the performance of different Meta classifiers techniques for classifying the dataset. Second step, in this second phase experiments the researcher selected five of the classifiers with highest performance accuracy obtained in the first phase experiments (see Table 9) as base classifiers in the second phase experiment. These base classifiers used in this phase: Random Tree, Random Forest, J48, PART, LMT to construct a novel ensemble model. The accuracy measures are TP rate, F Measure, ROC area, Sensitivity and Specificity. The error measures are Mean Absolute Error, Root Mean Squared Error and Kappa Statistics. First step, we tested various Meta classifiers and have chosen the following classifiers for a series of complete tests with outcomes presented in this research. We found that cross-validation give the best classification with 10 Fold. These Meta classifiers used are AdaBoostM1, Bagging, LogitBoost, Random Committee, Stacking, and Voting as depicted in Table 10. Table 10 depicts the various error metrics analyzed in the data set. It is inferred from Table 10 that Random Committee has the lowest MAE and highest Kappa Statistic value, Random Committee is an appropriate model for classifying the hospital situation with MAE = 0.06 and 95.0336 % were correctly classified.

Meta - Classifier	MAE	RMSE	KS	Correctly classified
AdaBoostM1	0.2957	0.3794	0.6051	599 80.4027 %
Bagging	0.1527	0.2609	0.8089	674 90.4698 %
Logit Boost	0.2725	0.3593	0.644	613 82.2819 %
Random Committee	0.0643	0.1931	0.9004	708 95.0336 %
Stacking	0.4983	0.4992	0	394 52.8859 %
Vote	0.4983	0.4992	0	394 52.8859 %

Table 10: Performance measures comparison of individual meta classifiers

Second step, the researcher investigated and constructed various Voting ensembles by combining methods based on meta classifier Voting and combined with previous selected base classifiers obtained in phase one in previous subsection. Table 11 depicts various ensemble models of Meta Voting Classifiers combining with various single classifiers. Voting combining: J48, LMT, Random Forest, Random Tree, PART (Voting + 5 classifiers), Voting combining: J48, Random Forest, Random Tree (Voting + 3 classifiers), and Voting combining: Random Forest, Random Tree (Voting + 2 classifiers).

Combined Classifiers	Base Classifiers				
	J48	LMT	Random Forest	Random Tree	PART
Voting + 5 classifiers	J48	LMT	Random Forest	Random Tree	PART
Voting + 3 classifiers	J48	-----	Random Forest	Random Tree	----
Voting + 2 classifiers	------	------	Random Forest	Random Tree	-----

Table 11. Combined classifiers

Tables 12 depicts the classifier performance of each classifier in term of MAE, RMSE, Kappa statistic, Time to build a model and % Correctly Classified. It is inferred from Table 12 that the ensemble (Voting + 3 classifiers) has the least MAE and RMES than ensemble (Voting + 5 classifiers) and the same Kappa Statistic value as (Voting + 5 classifiers). Ensemble (Voting + 3 classifiers) has the highest MAE and RMES than ensemble (Voting + 2 classifiers) and the highest Kappa Statistic value than Voting + 2 classifiers. But in terms of Correctly Classified instances, ensemble (Voting + 3 classifiers) has the highest Correctly Classified instances than the other considered ensembles.

Combined Classifiers	Correctly Classified	MAE	RMSE	Kappa statistic	Time to build a model
Voting + 5 classifiers	710 95.30 %	0.1239	0.2206	0.906	2.51 seconds
Voting + 3 classifiers	711 95.43 %	0.1025	0.2077	0.9086	0.07 seconds
Voting + 2 classifiers	707 94.89 %	0.0866	0.204	0.8977	0.05 seconds

Table 12: Performance measure comparison for ensemble models

Ensemble	Recall	Precision	F-measure	False Alarm rate
Voting + 5 classifiers	0.9270	0.97720	0.951433	0.02133
Voting + 3 classifiers	0.9318	0.9743	0.95257	0.0238
Voting + 2 classifiers	0.9383	0.9544	0.94627	0.0412

Table 13: Performance in term of recall precision, and false rate

Ensemble	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	PRC Area	Class
Voting + 5 classifiers	0.977	0.069	0.927	0.977	0.951	0.987	0.985	Normal
	0.931	0.023	0.979	0.931	0.954	0.987	0.986	Abnormal
Voting + 3 classifiers	0.974	0.063	0.932	0.974	0.953	0.982	0.972	Normal
	0.937	0.026	0.976	0.937	0.956	0.982	0.979	Abnormal
Voting + 2 classifiers	0.954	0.056	0.938	0.954	0.946	0.981	0.969	Normal
	0.944	0.046	0.959	0.944	0.951	0.981	0.981	Abnormal

Table 14: Performance classification of combining models in term of ROC

It is inferred from Table 12 that ensemble (Voting + 3 classifiers) has the best Correctly Classified than all individual Meta Classifiers, but individual Meta Classifiers Random Committee has the lowest MAE and RMSE than the Ensembles combined model. Table 13 depicts the performance of each classifier in term of recall precision and f-measure and false alarm rate. It is inferred from table 13 that Ensemble (Voting + 3 classifiers) model has the highest F-measure and the highest precision as (Voting + 5 classifier) and lowest false alarm rate. In the term of recall with the same recall value as (Voting + 2 classifiers), and highest recall value than (Voting + 5 classifiers). Table 14 depicts the algorithm performance of each classifier in term of recall precision and f-measure for Normal and Abnormal classes is summarized. It is inferred from Table 14 that (Voting + 5 classifiers) model has the highest ROC Area and also highest PRC Area than the others Ensemble in classification the class Normal and Abnormal classes but in term of F- Measure the ensemble (Voting + 3 classifiers) has highest F- Measure the others. In terms of Sensitivity, Specificity. Table 15 depicts the classification performance of each classifier in term of Sensitivity, Specificity. It is inferred from Table 15 that the Ensemble (Voting + 3 classifiers) model has the highest Accuracy than the others ensemble. But in terms of Specificity and Sensitivity the Ensemble (Voting + 2 classifiers) is highest.

Combined Classifiers	Sensitivity	Specificity	Accuracy
Voting + 5 classifiers	0.92702	0.93147	0.95302
Voting + 3 classifiers	0.93188	0.93654	0.95436
Voting + 2 classifiers	0.93837	0.94416	0.94899

Table 15: Performance combining models in term of sensitivity and specificity

Table 16 depicts the overall Ensemble Voting performance ranked by accuracy. It is inferred from Table 16 that Ensemble (Voting + 3 classifiers) model has the highest accuracy and the Ensemble (Voting + 2 classifiers) model has the lowest accuracy.

Algorithm	Accuracy
Voting + 3 classifiers	0.95436
Voting + 5 classifiers	0.95302
Voting + 2 classifiers	0.94899

Table 16: Overall ensembles performance ranked by accuracy

Table 17 depicts the overall Ensembles and Meta classifiers performance ranked by accuracy. It is inferred from Table 17 that Ensemble (Voting + 3 classifiers) model has the highest accuracy and the Meta classifiers Stacking and Voting models have the lowest accuracy.

Models	Accuracy
Ensemble (Voting+ 3 classifiers)	0.95436
Ensemble (Voting + 5 classifiers)	0.95302
Random Committee	0.95033
Ensemble (Voting + 2 classifiers)	0.94899
Bagging	0.90469
Logit Boost	0.82281
AdaBoostM1	0.80402
Stacking	0.52885
Vote	0.52885

Table 17: Ensemble and meta classifiers performance ranked by accuracy

Figure 5 depicts the classification error of Ensemble (Voting+ 3 classifiers) performance, the blue crosses indicated Normal class classification, the red crosses indicate the Abnormal class classified, the red squares indicated Abnormal class unclassified and the blue squares indicated Normal class unclassified.

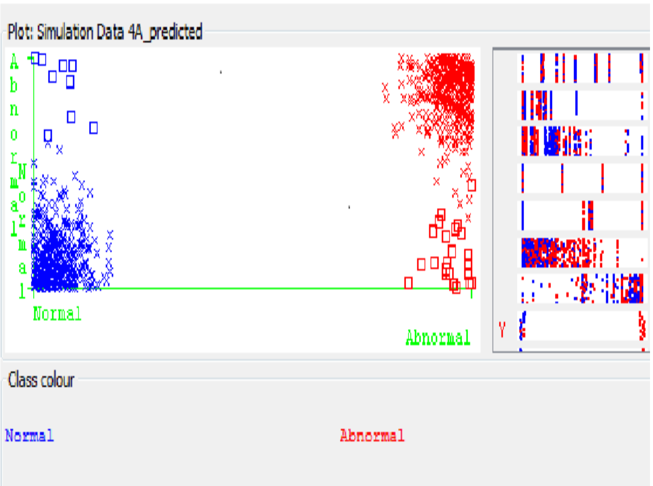

Figure 4: Error of ensemble (voting+ 3 classifiers) performance

Figure 6 depicts the abnormal Class, Area under Roc of Ensemble (Voting + 3 classifiers) with highest area under ROC. (0.9816).

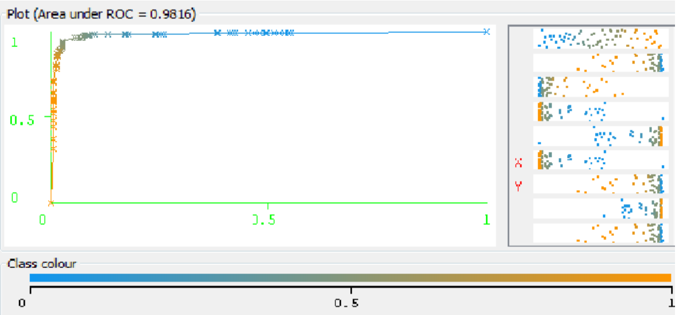

Figure 5: Abnormal class, area under roc of ensemble (voting+ 3 classifiers)

Figure 7 depicts the results when the cost is 0, Random is 394 and the difference between the values of the cost function between the random selection and the current value of the cost is called Gain, as indicated in the right side of the frame. In the context of abnormal situation, the Gain can be interpreted as the benefit obtained by using the classification model instead of random selection of the same number of patients. In our experiments, the gain (Benefit) obtained is 0. Threshold curve depicts the dependence of the part of class "Abnormal" patients retrieved in the course of predicting selected from the whole dataset (i.e. only those selected for which the estimated probability of having abnormal disease exceeds the chosen threshold). The confusion matrix for the current value of the threshold is shown in the Confusion Matrix frame at the left bottom corner of the window.

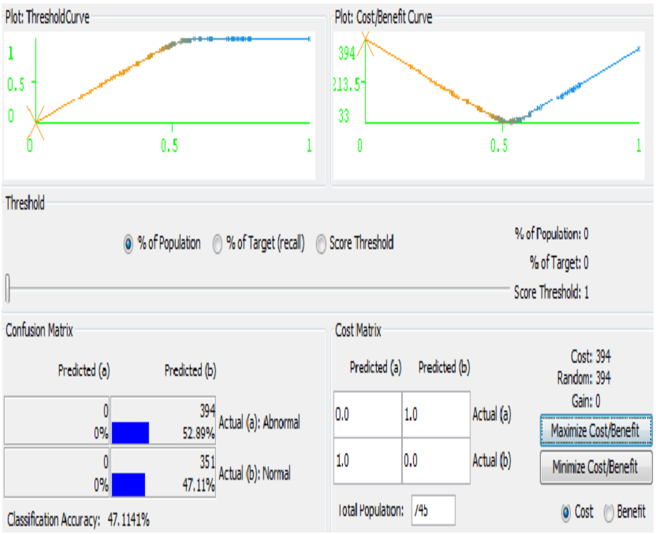

Figure 6: Maximize cost/benefit of class abnormal

Figure 7 depicts the results when the cost is 33, Random is 370.97 and the Gain is 337.97. In the context of abnormal disease, the Gain can be interpreted as the benefit obtained by using the classification model instead of random selection of the same number of patients. In our experiments, the gain (Benefit) obtained is 337.97 and the classification Accuracy is 95.5705, which means that using Cost/Benefit, we can obtain more classification accuracy than ROC Curve.

Figure 7: Minimize cost/benefit of class abnormal

VII. Models Comparison, Discussion and Lessons learned

As mentioned in previous Sections, two major phases of experiments were conducted. In the first phase experiment, attempt has been made to investigate the experimental results of the performance of different classification techniques for classifying the data from different simulated wearable sensors used for monitoring different patients with different diseases. We explored and evaluated the models with various methods of evaluation based on Error Metrics, ROC curves, Confusion Matrix, Sensitivity and Specificity. Empirical results indicate that the execution time of Random Committee algorithm is lowest for classification in comparison with the rest of classification algorithms, and the LMT algorithm has the higher execution time. The MSE error of the classification values for Random Committee is lower in comparison with the rest of the based proposed classifiers, and the Meta bagging classifier has higher MSE error in comparison with the rest of the base proposed classifiers. In terms of recall precision, f measure and false alarm rate the Random Committee model has the highest precision and lowest false alarm rate, and the same recall as Random Tree. In term of recall precision and f-measure for Normal class, it is inferred that Random Committee model has the highest precision and also high recall. With higher true positive rate and minimum false rate also with higher ROC Area when the classification is Normal class in comparison of the rest of the classifiers. Attribute Selected Classifier has the lower precision in comparison with the rest. Also from the performance of each classifier in term of recall precision and f measure for abnormal class, Random Committee model has the highest precision and also high recall (with higher true positive rate and minimum false rate), also has highest ROC Area in comparison with other classifiers. While PART classifier has the lowest precision the same as Attribute Selected Classifier but with highest in ROC Area compare with Attribute Selected Classifier. From Sensitivity, Specificity and Accuracy perspective, the Random Committee model has the highest Specificity and also high Sensitivity the same as Random Tree but with highest accuracy of all the classifiers. While IBK classifier has the lowest Sensitivity, Specificity and Accuracy compare with the rest of the classifiers. To sum up, from the execution and accuracy point of view, Random committee model can be identified as the best choice for analysis and decision model among all the other classifier algorithms. Random committee provides an advantage that with a reduced attribute set a better classification performance. Empirical results illustrate that Random committee classifier with selection attribute method gives better accuracy, error rate and reduced false alarm rate and with the highest Sensitivity and Specificity. The aim of the second phase, is to explore various ensembles combining models and evaluate the models with various methods of evaluation based on Error Metrics, ROC curves, Confusion Matrix, Sensitivity, Specificity and the Cost/Benefit methods. To construct a new novel ensemble healthcare decision support for assisting an intelligent health monitoring system. Experiments are conducted using the same dataset with the reduced dimensionality of the attributes obtained in the phase one experiments. We summarize the obtained results from the evaluation conducted in the previous Sections. The results indicate that the execution time of Ensemble (Voting + 2) classifiers algorithm is lowest for classification in comparison with the rest of ensemble classification algorithms, and the Ensemble (Voting + 5 classifiers) classification algorithm has the higher execution time. The MSE error of the classification values for Ensemble (Voting + 2 classifiers) is lower in comparison with the rest of the based proposed classifiers, and the Ensemble (Voting + 5 classifiers) classifier has higher MSE error in comparison with the rest of the base proposed classifiers. In terms of recall precision, *f* measure and false alarm rate the Ensemble (Voting + 5 classifiers) model has the highest precision and lowest false alarm rate, and the has the highest recall lower in comparison with the rest of the ensembles models. In term of recall precision and f- measure for Normal class it is inferred that Ensemble Voting + 3 classifiers model has the highest precision than Ensemble Voting + 5 classifiers model, but with lowest recall than Ensemble (Voting + 5 classifiers), has highest recall and highest TP Rate than Ensemble (Voting + 2 classifiers), and with minimum false rate than Ensemble (Voting + 5 classifiers) also with higher Roc Area and higher PRC when the classification is Normal class in comparison of the Ensemble (Voting + 2 classifiers). The Ensemble (Voting + 5 classifiers) has higher ROC Area and higher PRC when the classification is Normal class in comparison with the rest. In the case of class Abnormal we found that Ensemble Voting + 3 classifiers has highest True Positive Rate, minimum false rate and highest recall in comparison with the rest. We found the Ensemble (Voting + 5 classifiers), has highest ROC Area and higher PRC when the classification is abnormal class in comparison with the rest. From Sensitivity, Specificity and Accuracy perspective, the Ensemble Voting + 2 classifiers model has the highest Specificity and also high Sensitivity followed by Ensemble (Voting + 3 classifiers) model. From Accuracy perspective, the Ensemble (Voting + 3 classifiers) model has the highest Accuracy in comparison with the rest. To sum up, from the execution and accuracy point of view, Ensemble (Voting + 3 classifiers) model can be identified as the best choice for analysis and detection model among all the other classifier ensembles modes algorithms for our data set. Ensemble (Voting + 3 classifiers) provides an advantage that with a reduced feature set a better classification performance and is able to offer a better decision support system. The last evaluation method used in our experiments is Cost/Benefit method. As indicate in the result section using Cost/Benefit method minimizes the cost and increases the classification accuracy. In our experiment the gain (Benefit) obtained is 337.97 and the classification Accuracy is 95.5705 this mean that using Cost/Benefit we can obtain more classification accuracy than ROC Curve. The main Goal of this Section is to evaluate ensemble design and combining different algorithms to develop a novel intelligent ensemble healthcare decision support and a monitoring system to classify the situation of an emergency hospital based on the Vital Signs from Wearable Sensors. We compared the performance of the entire Individual base classifiers; Meta classifiers and Ensemble combine models. Empirical results illustrate that Voting combining with J48, Random Forest, Random Tree (Voting + 3 classifiers) model, with selection attribute method gives better accuracy, with high recall and high f- measure. Our Novel Intelligent Ensemble Health Care Decision Support and Monitoring can optimize the results and improve assisted health care monitoring.

References

[1] A. Salih and A. Abraham, "A Review of Ambient Intelligence Assisted Healthcare Monitoring," *International Journal of Computer Information Systems and Industrial Management (IJCISIM)*, vol. 5, pp. 741-750, 2013.

[2] D. I. Tapia, S. Rodrǵuez, J. Bajo, and J. M. Corchado, "FUSION@, a SOA-based multi-agent architecture," in *International Symposium on Distributed Computing and Artificial Intelligence 2008 (DCAI 2008)*, 2009, pp. 99-107.

[3] G. D. Abowd, A. K. Dey, P. J. Brown, N. Davies, M. Smith, and P. Steggles, "Towards a better understanding of context and context-awareness," in *Handheld and ubiquitous computing*, 1999, pp. 304-307.

[4] A. S. M. Salih and A. Abraham, "Novel Ensemble Decision Support and Health Care Monitoring System," *Journal of Network and Innovative Computing*, vol. 2, pp. 041-051, 2014.

[5] J. A. Fraile, J. Bajo, J. M. Corchado, and A. Abraham, "Applying wearable solutions in dependent environments," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 14, pp. 1459-1467, 2010.

[6] J. M. Corchado, J. Bajo, and A. Abraham, "GerAmi: Improving healthcare delivery in geriatric residences," *Intelligent Systems, IEEE*, vol. 23, pp. 19-25, 2008.

[7] T. H. N. Vu, N. Park, Y. K. Lee, Y. Lee, J. Y. Lee, and K. H. Ryu, "Online discovery of Heart Rate Variability patterns in mobile healthcare services," *Journal of Systems and Software*, vol. 83, pp. 1930-1940, 2010.

[8] R. Jafari, W. Li, R. Bajcsy, S. Glaser, and S. Sastry, "Physical activity monitoring for assisted living at home," in *4th International Workshop on Wearable and Implantable Body Sensor Networks (BSN 2007)*, 2007, pp. 213-219.

[9] J.-Y. Yang, J.-S. Wang, and Y.-P. Chen, "Using acceleration measurements for activity recognition: An effective learning algorithm for constructing neural classifiers," *Pattern recognition letters*, vol. 29, pp. 2213-2220, 2008.

[10] H. Hagras, V. Callaghan, M. Colley, G. Clarke, A. Pounds-Cornish, and H. Duman, "Creating an ambient-intelligence environment using embedded agents," *Intelligent Systems, IEEE*, vol. 19, pp. 12-20, 2004.

[11] F. Doctor, H. Hagras, and V. Callaghan, "A fuzzy embedded agent-based approach for realizing ambient intelligence in intelligent inhabited environments," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol. 35, pp. 55-65, 2005.

[12] K. Van Laerhoven and O. Cakmakci, "What shall we teach our pants?," in *Wearable Computers, The Fourth International Symposium on*, 2000, pp. 77-83.

[13] T. Kohonen, "The self-organizing map," *Proceedings of the IEEE*, vol. 78, pp. 1464-1480, 1990.

[14] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*: John Wiley & Sons, 2012.

[15] I. Assent, R. Krieger, F. Afschari, and T. Seidl, "The TS-tree: efficient time series search and retrieval," in *Proceedings of the 11th international conference on Extending database technology: Advances in database technology*, 2008, pp. 252-263.

[16] S. Patel, R. Hughes, N. Huggins, D. Standaert, J. Growdon, J. Dy, *et al.*, "Using wearable sensors to predict the severity of symptoms and motor complications in late stage Parkinson's Disease," in *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, 2008, pp. 3686-3689.

[17] D. J. Cook and S. K. Das, "How smart are our environments? An updated look at the state of the art," *Pervasive and mobile computing*, vol. 3, pp. 53-73, 2007.

[18] M. Mozer, "Lessons from an adaptive house," Architectural Engineering, 2004.

[19] S. K. Das, D. J. Cook, A. Battacharya, E. O. Heierman III, and T.-Y. Lin, "The role of prediction algorithms in the MavHome smart home architecture," *Wireless Communications, IEEE*, vol. 9, pp. 77-84, 2002.

[20] S. Helal, B. Winkler, C. Lee, Y. Kaddoura, L. Ran, C. Giraldo, *et al.*, "Enabling location-aware pervasive computing applications for the elderly," in *Pervasive Computing and Communications, 2003.(PerCom 2003). Proceedings of the First IEEE International Conference on*, 2003, pp. 531-536.

[21] E. M. Tapia, S. S. Intille, and K. Larson, *Activity recognition in the home using simple and ubiquitous sensors*: Springer, 2004.

[22] M. Philipose, K. P. Fishkin, M. Perkowitz, D. J. Patterson, D. Fox, H. Kautz, *et al.*, "Inferring activities from interactions with objects," *Pervasive Computing, IEEE*, vol. 3, pp. 50-57, 2004.

[23] R. Simpson, D. Schreckenghost, E. F. LoPresti, and N. Kirsch, "Plans and planning in smart homes," in *Designing Smart Homes*, ed: Springer, 2006, pp. 71-84.

[24] J. C. Augusto and C. D. Nugent, "The use of temporal reasoning and management of complex events in smart homes," in *ECAI*, 2004, p. 778.

[25] G. M. Youngblood, "Automating inhabitant interactions in home and workplace environments through data-driven generation of hierarchical partially-observable Markov decision processes," 2005.

[26] F. Amigoni, N. Gatti, C. Pinciroli, and M. Roveri, "What planner for ambient intelligence applications?," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on,* vol. 35, pp. 7-21, 2005.

[27] UniversityofWashington. (2014). *The assistive cognition project*
Available:
http://www.cs.washington.edu/assistcog.
. (Access Date 15 December 2014)

[28] J. R. Beck and S. G. Pauker, "The Markov process in medical prognosis," *Medical Decision Making,* vol. 3, pp. 419-458, 1983.

[29] Y. Xiang and K.-L. Poh, "Time-critical dynamic decision modeling in medicine," *Computers in biology and medicine,* vol. 32, pp. 85-97, 2002.

[30] T.-Y. Leong, "Dynamic decision modeling in medicine: a critique of existing formalisms," in *Proceedings of the Annual Symposium on Computer Application in Medical Care*, 1993, p. 478.

[31] J. E. Stahl, "Modelling methods for pharmacoeconomics and health technology assessment," *Pharmacoeconomics,* vol. 26, pp. 131-148, 2008.

[32] A. J. Schaefer, M. D. Bailey, S. M. Shechter, and M. S. Roberts, "Modeling medical treatment using Markov decision processes," in *Operations research and health care*, ed: Springer, 2004, pp. 593-612.

[33] O. Alagoz, H. Hsu, A. J. Schaefer, and M. S. Roberts, "Markov decision processes: a tool for sequential decision making under uncertainty," *Medical Decision Making,* 2009.

[34] I. Fatima, M. Fahim, Y.-K. Lee, and S. Lee, "Classifier ensemble optimization for human activity recognition in smart homes," in *Proceedings of the 7th International Conference on Ubiquitous Information Management and Communication*, 2013, p. 83.

[35] A. C. Tan and D. Gilbert, "Ensemble machine learning on gene expression data for cancer classification," 2003.

[36] B. Liu, Q. Cui, T. Jiang, and S. Ma, "A combinational feature selection and ensemble neural network method for classification of gene expression data," *BMC bioinformatics,* vol. 5, p. 136, 2004.

[37] P. Li, K. L. Chan, S. Fu, and S. M. Krishnan, "An abnormal ecg beat detection approach for long-term monitoring of heart patients based on hybrid kernel machine ensemble," in *Multiple Classifier Systems*, ed: Springer, 2005, pp. 346-355.

[38] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural computation,* vol. 13, pp. 1443-1471, 2001.

[39] fipa.org. *Foundation for Intelligent Physical Agents*. . Available: http//www.fipa.org (Access Date 20 September 2013)

[40] M. C. Tu, D. Shin, and D. Shin, "Effective diagnosis of heart disease through bagging approach," in *Biomedical Engineering and Informatics, 2009. BMEI'09. 2nd International Conference on*, 2009, pp. 1-4.

[41] M. C. Tu, D. Shin, and D. Shin, "A comparative study of medical data classification methods based on decision tree and bagging algorithms," in *Dependable, Autonomic and Secure Computing, 2009. DASC'09. Eighth IEEE International Conference on*, 2009, pp. 183-187.

[42] V. Chaurasia and S. Pal, "Data Mining Approach to Detect Heart Diseases," *International Journal of Advanced Computer Science and Information Technology (IJACSIT) Vol,* vol. 2, pp. 56-66, 2014.

[43] P. Wen, "Application of decision tree to identify abnormal high frequency electro-cardiograph," *Physics Experimentation,* vol. 11, p. 011, 2009.

[44] C. Kaewchinporn, N. Vongsuchoto, and A. Srisawat, "A combination of decision tree learning and clustering for data classification," in *Computer Science and Software Engineering (JCSSE), 2011 Eighth International Joint Conference on*, 2011, pp. 363-367.

[45] J. R. Quinlan, *C4. 5: Programs for Machine Learning*: Morgan Kaufmann, 1993.

[46] E. Frank and I. H. Witten, "Generating accurate rule sets without global optimization," 1998.

[47] W. W. Cohen, "Fast effective rule induction," in *Proceedings of the twelfth international conference on machine learning*, 1995, pp. 115-123.

[48] N. Landwehr, M. Hall, and E. Frank, "Logistic model trees," in *Machine Learning: ECML 2003*, ed: Springer, 2003, pp. 241-252.

[49] L. Breiman, "Random forests," *Machine learning,* vol. 45, pp. 5-32, 2001.

[50] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors)," *The annals of statistics,* vol. 28, pp. 337-407, 2000.

[51] Symposium. (2005). *Symposium. volume 1, July, 2005*.Available:
http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm#proxSymposium. (Access Date 29 September 2014)

[52] L. B. a. A. Cutler. (2001). *"Random Forest (2001)"*. . Available:http://stat
www.berkeley.edu/users/breiman/RandomForests/cc_home.htm. (Access Date 15 May 2014)

[53] I. H. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*: Morgan Kaufmann, 2005.

[54] P. K. Chan and S. J. Stolfo, "Experiments on multistrategy learning by meta-learning," in *Proceedings of the second international conference on information and knowledge management*, 1993, pp. 314-323.

[55] Y. Freund and R. E. Schapire, "A desicion-theoretic generalization of on-line learning and an application

to boosting," in *Computational learning theory*, 1995, pp. 23-37.
[56] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of computer and system sciences,* vol. 55, pp. 119-139, 1997.
[57] L. Breiman, "Stacked regressions," *Machine learning,* vol. 24, pp. 49-64, 1996.
[58] D. H. Wolpert, "Stacked generalization," *Neural networks,* vol. 5, pp. 241-259, 1992.
[59] J. R. Quinlan, "Bagging, boosting, and C4. 5," in *AAAI/IAAI, Vol. 1*, 1996, pp. 725-730.
[60] C. W. Wang, "New ensemble machine learning method for classification and prediction on gene expression data," in *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*, 2006, pp. 3478-3481.
[61] M. Panda and M. R. Patra, "Ensemble voting system for anomaly based network intrusion detection," *Int. J. of Recent Trends in Engineering and Technology,* vol. 2, 2009.
[62] L. Felföldi, "Classifier Combination Systems and their Application in Human Language Technology," *Research Group on Artificial Intelligence, Szeged, Hungary,* 2008.
[63] L. I. Kuncheva, "" Fuzzy" versus" nonfuzzy" in combining classifiers designed by Boosting," *Fuzzy Systems, IEEE Transactions on,* vol. 11, pp. 729-741, 2003.
[64] J. Kittler, M. Hatef, R. P. Duin, and J. Matas, "On combining classifiers," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 20, pp. 226-239, 1998.
[65] M. A. Hall, "Correlation-based feature selection for machine learning," The University of Waikato, 1999.
[66] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *bioinformatics,* vol. 23, pp. 2507-2517, 2007.
[67] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: An update, software available at http://www. cs. waikato. ac. nz/ml/weka," *SIGKDD Explorations,* vol. 11, 2009.
[68] I. Jolliffe, *Principal component analysis*: Wiley Online Library, 2002.
[69] J. Han, M. Kamber, and J. Pei, "Data mining: concepts and techniques (the Morgan Kaufmann Series in data management systems)," *Morgan Kaufmann,* 2000.
[70] K. Kira and L. A. Rendell, "A practical approach to feature selection," in *Proceedings of the ninth international workshop on Machine learning*, 1992, pp. 249-256.
[71] I. Kononenko, "Estimating attributes: analysis and extensions of RELIEF," in *Machine Learning: ECML-94*, 1994, pp. 171-182.
[72] M. Robnik-Šikonja and I. Kononenko, "An adaptation of Relief for attribute estimation in regression," in *Machine Learning: Proceedings of the Fourteenth International Conference (ICML'97)*, 1997, pp. 296-304.