Journal of Network and Innovative Computing
ISSN 2160-2174 Volume 2 (2014) pp. 041-051

Novel Ensemble Decision Support and Health Care Monitoring System

Abdelhmid Salih Mohmed Salih[1] and 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
ajith.abraham@ieee.org

Abstract: In the health care monitoring, data mining is mainly used for classification and predicting the diseases. Various data mining techniques are available for classification and predicting diseases. The aim 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. Extensive investigate of the experimental results of the performance of different meta classifiers techniques for classifying the data from different wearable sensors used for monitoring different diseases was carried. 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. These Meta classifiers used are AdaBoostM1, Bagging, LogitBoost, Random Committee, Stacking, and Voting. 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 results obtained show that the Novel Intelligent Ensemble method classifier achieved better outcomes that are significantly better compared with the outcomes of the all Base Classifiers Proposed and all meta base classifiers used in this paper. 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: **Base Classifiers, Meta base classifiers, Ensemble methods, Voting, wearable sensors.**

1. Introduction

There is growing need to supply constant Health Care Monitoring (HCM) and support to patients with Chronic Diseases (CD) especially the disabled, and elderly. CD is becoming the major causes of the death [1]. In Sudan according to the latest WHO [2], data published in April 2011, Coronary Heart Disease (CHD) deaths reached 10.67% of total deaths. Traditional healthcare and services are usually offered within hospitals or medical centers with traditional monitory of patients with CD, measurements of vital signs are carried with traditional measurements and the corresponding diagnosis are carried. However, this solution is costly, inefficient and inconvenient for the people with the need of routine checks. There are huge requirements to move the routine monitoring medical check and healthcare Services in hospitals. There is an ever-growing need to supply constant care and support to patients with CD, disabled, and elderly. The drive to find more effective ways of providing, such care has become a major challenge for the scientific community [3]. 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. [4]. AmI proposes new ways of interaction between people and technology, making it suited to the needs of individuals and the environment that surrounds those. AmI [5] tries to adapt the technology to the people's needs by means of omnipresent computing elements which communicate amongst them in a ubiquitous way, It also proposes a new way to interact between people and technology, where this last one is adapted to individuals and their context. The context includes both the users and the environment information. The information may consist of many different Parameters such as the building status (e.g. temperature or light), vital signs (e.g. heart rhythm or blood pressure), etc. Wireless sensor networks (WSNs) are used for gathering the information needed by Ami environments. The information may consist of many different sensors such as vital signs (e.g. heart rhythm or blood pressure), etc. Thus, most of the context information can be collected by distributed sensors throughout the environment and even the users themselves.[1]. Sensors data is collected from disparate sources and later need to be classified and analyzed to produce information that is more accurate, more complete, or more insightful than the individual pieces. To deal with the large volume of data produced by these special kinds of wireless networks, one approach is the use of Artificial Intelligence and Data Mining techniques. Data mining plays a vital role in various applications such as business organizations, e-commerce, health care industry, scientific and engineering. In the health care industry, the data mining is mainly used for classification and predicting the diseases from the datasets. Various data mining individual classification methods and ensembles classification methods are available for predicting and classification diseases. Gather information about the context is not enough. However information must be processed, analyzed, reasoning and decision making, since the quality of decision making depends of quality of information by using dynamic mechanisms and methods. 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. These Meta classifiers used are AdaBoostM1, Bagging, LogitBoost, Random Committee, Stacking, and Voting. 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 methods used for evaluation and the base proposed classifiers. Section 4 presents the Experimental Results and is analyzed in Section 5 followed by discussions in Section 6.

2. Related Work

Sensor data is collected from disparate sources and later analyzed to produce information that can help in assisted health care monitoring that is more accurate, more complete. There are several machine learning techniques and data mining methods and Techniques used in analyzing sensors data in AmI. These methods and techniques can help accomplish many important tasks in AmI assisted healthcare monitoring and make the system more efficient. Jafari et al. [6] proposed an ANN based activity recognition system in order to determine the occurrence of falls. Their system works with single sensor placed on to the chest of the subjects. However ANN Require more tuning parameters than support vector machines, and also ANN is sensitive to noise (a validation set may help here) and missing values in the training samples need to be replaced or removed. Augusto et al. [7-11, 12-16] used Event-Condition-Action (ECA) Rules and various extensions of them for applications in Smart Homes and supported living for the elderly. Bager et al. [17] developed another Smart Home application and area of testing, for the Medical Automation Research Center. This system uses probabilistic methods to determine patterns in behavior. Based on a series of sensors, one in each room, the system monitors the duration of time that the user spends in each room. Although these systems have shown improvements over other systems of its type. However it is still lacking in one major area. Both of the above systems deal with the elderly living alone. This is due to the fact that there is no identifier on the person using the system. Corchado et al. [12] developed GerAmI system that has got around this problem. The GerAmI system was developed in conjunction with the Alzheimer Santísima 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 radio frequency identification chip (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. Shyamal et al. [18] have 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. Vapnik [19] has selected SVM's due to their success in many classification problems; SVM's optimize an objective function that is convex, hence guaranteed to find an optimal solution. However many other classification algorithms only guarantee that local optima be reached. Reasoning algorithms offer is the ability to predict and recognize activities that occur in AmI environments. Tapia et al. [20] 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. 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 Augusto and Nugent [21] have 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. University of Washington [22] 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 [23] described dynamic sequential decision making in medicine using Markov-based approach originally described in terms of medical decision-making. There are also others approaches for example. Leong [24] and Stahl [25] have utilized decision trees to model temporal decisions. In all cases, the goal is to determine optimal sequences of decisions. Other several studies have been reported that they 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 Ambient Intelligence (AmI) assisted healthcare monitoring. Fatima et al. [26] presented Classifier Ensemble Optimization method for activity recognition they have been proposed 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 [27] have presented a comparison of single supervised machine learning and ensemble methods in classifying seven publicly available cancerous data. The experimental results indicate that 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. [28] 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 [29] 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. [30] (one per patient) to distinguish normal signals for a given patient from all other signals [31]. Other several studies have been reported that they have focused on the importance of the Ensemble methods in the field of medical health care. These studies have applied different approaches to the given problem and achieved high classification accuracies. Tu et al. [32] proposed the use of bagging with C4.5 algorithm, bagging with Naïve bayes algorithm to diagnose the heart disease of a patient. Tu et al. [33] 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 et al. [34] used Naive Bayes, J48 Decision Tree and Bagging algorithm to predict the survivability for Heart Diseases patients. Pan wen et al. [35] conducted experiments on ECG data to identify abnormal high frequency electrocardiograph using decision tree algorithm C4.5 with bagging. Kaewchinporn et al [36] presented a new classification algorithm TBWC combination of decision tree with bagging and clustering. This algorithm is experimented on two medical datasets: cardiocography1, cardiocography2 and other datasets not related to medical domain. Li et al. [37] experimented on ovarian tumor data to diagnose cancer-using C4.5 with and without bagging. Cao et al. [38] proposed a new decision tree based ensemble method combined with feature selection method backward elimination strategy with bagging to find the structure activity relationships in the area of chemo metrics related to pharmaceutical industry. Liu et al. [39] experimented on breast cancer data using C5 algorithm with bagging to predict breast cancer survivability. Tan et al. [40] used C4.5 decision tree, bagged decision tree on seven publicly available cancerous micro array data, and compared the prediction performance of these methods.

3. Intelligent Data Analysis

3.1. Base Classifiers Used

We used the following base classifiers available in Weka.[41] for a series of complete tests with outcomes presented in this paper.

A) Decision tree algorithm J48

J48 classifier is a simple C4.5 decision tree for classification. 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 [42- 43].

B) Logistic Model Trees (LMT)

A logistic model tree (LMT) [44] is an algorithm for supervised learning tasks, which is combined with linear logistic regression and tree induction. LMT creates a model tree with a standard decision tree structure with logistic regression functions at leaf nodes. In LMT, leaves have an associated logic regression functions instead of just class labels.

C) Random Forest

Random forest [45] is an ensemble classifier that consists of many decision tree and outputs the class that is the mode of the class's output by individual trees. The algorithm for inducing a random forest was developed by Breiman and Cutler. Random Forests grows many classification trees without pruning. Then each decision tree classifies a test sample and random forest assigns a class, which have maximum occurrence among these classifications.

D) Random Tree

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 [46].

E) PART

Rule-based learning, especially decision trees (also called classification trees or hierarchical classifiers) is a rule generator that uses J48 to generate pruned decision trees from which rules are extracted [47].

F) IBK

The lazy IBk (commonly known as K- nearest neighbor) is one of classification algorithms that uses distance weighting measures with capability of various attributes like Date attributes, Numeric attributes, Unary attributes, Nominal attributes, Missing values, Binary attributes and Empty nominal attributes. K-nearest neighbours classifier. Can select appropriate value of K based on cross-validation. Can also do distance weighting [48].

3.2. Base Meta Classifiers Used

A) AdaBoostM*1*

AdaBoost.M1is a well-known algorithm for boosting weak classifiers. [49]. 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.

B) Bagging

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. [47-50].

C) LogitBoost

One of the boosting algorithms developed recently, is introduced for predicting protein structural classes. Logit Boost is one of the boosting algorithms developed in recent years. 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. [51]. 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.

D) Random Committee

Classifier that ensembles randomizable base classifiers. Each base classifier is built using a different random number seed based one the same data. The final prediction is a straight average of the predictions generated by the individual base classifiers.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.

E) Stacking

Stacking can be regarded as a generalization of voting, where meta-learner aggregates the out puts of several base classifiers [52]. Stacking often performance better than any single one of the trained models. It has been successfully used on supervised learning tasks (regression) [53].

3.3. 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. [54]:

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.

A) 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:

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

Pi(x) is the results of the prediction of i^{th} 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) [55- 56]. There is different combination approaches used. In this paper we employed combining rules approach.

B) **Combining Rules**

Combining rules are the simplest combination approach and it is probably the most commonly used in the multiple classifier system [57]. 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 [58]. A theoretical framework for fixed rules combination was proposed by Kittler [59]. It includes the sum, product, max, min, average and median rules. In this paper we have used the Maximum rule. Maximum rule is based on information provided by the maximum value of

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

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 [58].

$$f(x) = w_j \ , \ \ j = \arg\max\{\max \ p\left(x^i \mid w_k\right)\} \quad (3)$$

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.

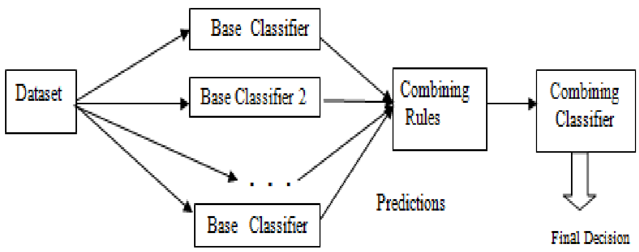

Figure 1: Ensemble using Combination rule with Voting

C) Attribute selection

It is often an essential data processing step prior to applying a learning algorithm. Reduction of the attribute space that leads to a better understandable model and simplifies the usage of different visualization technique. Attribute selection reduces dataset size by removing irrelevant and redundant attributes. It finds a minimum set of attributes such that the resulting probability distribution of data classes is as close as possible of original distribution.

D) Cross-Validation Method

In this paper, we 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.

E) Methods Used for Evaluation of Algorithms

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

4. Experimental Results

A) Data set and simulation of Hospital Environment

We simulated the environment of Baraha Medical City in Shambat, Khartoum North, in Sudan using the framework reported in [60-61]. It is situated in a 600 Sq. 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. In our simulation, we allocated chronic 6 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. Depending on the criticality of the patient's condition, when a sensor finds values that fall in the danger zone an automated alarm is triggered notifying the nurses and doctors through mobile network or Wifi systems [60-61]. In this project, our main task is to develop Novel Intelligent Ensemble Health Care Decision Support and Monitoring 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. We apply attribute selection method to reduce the number of the attributes. All the 300 attributes were labeled as A, B, C, Z, …KN. We investigated the Decision tree algorithm J48, Logistic Model Trees (LMT), Random Tree, Random Forest, PART and the lazy IBk classifiers using WEKA [41] and finally managed to reduce to only 6 attributes: AK, CM, CP, CW, FJ and KN. We found that cross-validation give the best classification with 10 Fold. Then the overall accuracy for all classifiers was done. Second we tested various Meta Classifiers and have chosen the following classifiers for a series of complete tests with outcomes presented in this paper. AdaBoostM1, Bagging, Logit Boost, Random Committee, Stacking and Voting. In the third stage we constructed Novel Ensemble Methods using Voting Meta Classifier that combine the base models built in the previous step into the final ensemble model.

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 1: Performance Measures comparison of individual Meta Classifiers

Combined Classifiers	Correctly Classified	MAE	RMSE	Kappa statistic	Time to build a model
Voting + 5 classifiers	710 95.302 %	0.123 9	0.2206	0.906	2.51 seconds
Voting + 3 classifiers	711 95.436 %	0.102 5	0.2077	0.9086	0.07 seconds
Voting + 2 classifiers	707 94.899 %	0.086 6	0.204	0.8977	0.05 seconds

Table 2: Performance Measures comparison for ensemble models

Experiments are conducted on wearable sensors vital signs data set, which was simulated using a hospital environment. The aim of this paper is to build Novel Ensemble Methods and investigate the experimental results of the performance of different ensemble methods for the simulation wearable sensors dataset. Comparative analysis and evaluation have been done using various evaluation methods and 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. In the preprocessing step we have changed the class attribute to Abnormal or Normal where an ‘Abnormal’ specifies 1 class and a ‘Normal’ Specifies 0 class. Table 1 depicts the detailed results of the execution required to build the model for each base meta classifier. From Table 1, it is inferred that Random Committee model outperform the others meta classifiers with MAE = 0.06 and 95.0336 % Correctly Classified.

Table 1 depicts the various error metrics analyzed in the data set. It is inferred from Table 1 that Random Committee has the lowest MAE and highest Kappa Statistic value. Random Committee is an appropriate model for classifying the hospital situation in a minimal span of time with higher accuracy. Tables 2 depicts the classifier performance using Ensemble Model of Meta Voting Classifiers combining with various single Meta 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).

Tables 2 depict 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 2 that the ensemble (Voting + 3 classifiers) has the least MAE and RMES than ensemble (Voting + 5) 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 others ensembles. It is inferred from Tables 1 and 2 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.

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	0.023809
Voting + 2 classifiers	0.9383	0.9544	0.94627	0.041237

Table 3: The classification performance of each Ensemble model in term of recall precision, f- measure and false alarm 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 4: The classification performance of each Ensemble model in term of recall, precision, f- measure and Roc Area for Normal and Abnormal class.

Combined Classifiers	Parameter	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 5: The classification performance of each Ensemble model in term of Sensitivity and Specificity

Table 3 depicts the performance of each classifier in term of recall precision and f-measure and false alarm rate. It is inferred from table 2 that Ensemble (Voting + 3 classifiers) model has the highest precision and lowest false alarm rate and highest F-measure than the others Ensemble and with the same recall value as (Voting + 2 classifiers), and highest recall value than (Voting + 5 classifiers). Table 4 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 4 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. Figure 2 depicts the Area under ROC of Ensemble (Voting + 5 classifiers).

Table 5 depicts the classification performance of each classifier in term of Sensitivity, Specificity. It is inferred from Table 5 that 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. Table 6 depicts the overall Ensemble performance ranked by accuracy. It is inferred from Table 6 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 6: Overall Ensembles performance ranked by accuracy

Table 7 depicts the overall Ensembles and Meta classifiers performance ranked by accuracy. It is inferred from Table 7 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
Logit Boost	0.82281

AdaBoostM1	0.80402
Stacking	0.52885
Vote	0.52885

Table 7: Overall Ensembles and Meta classifiers performance ranked by accuracy

Figure 1 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

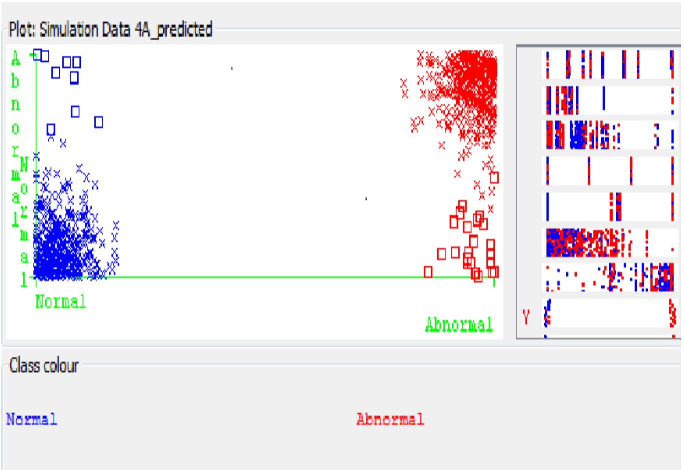

Figure 2: The classification error of Ensemble (Voting+ 3 classifiers) performance

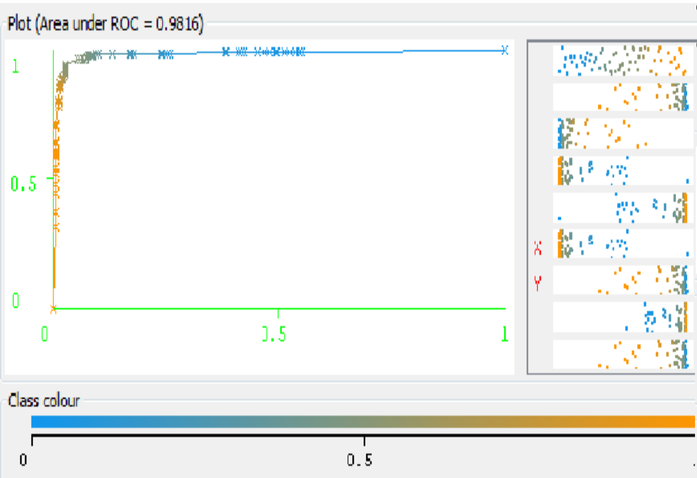

Figure 3: Class Abnormal, Area under ROC of Ensemble (Voting+ 3 classifiers) performance.

Figure 2 depicts Class Abnormal, Area under Roc of Ensemble (Voting + 3 classifiers) with highest area under ROC. (0.9816). Figure 4 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, indicated at 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 experiment 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.

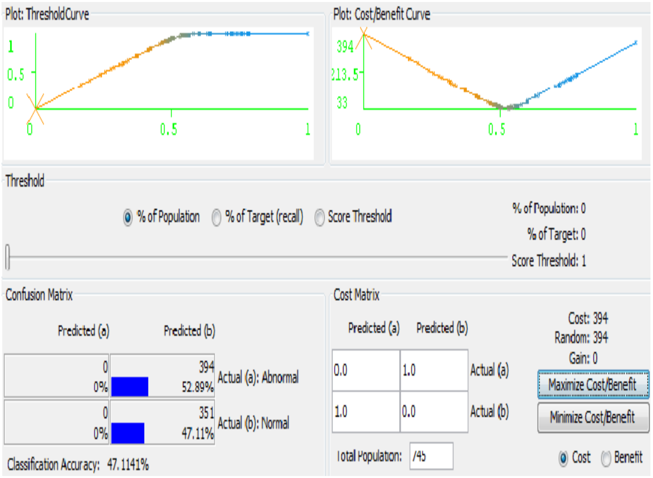

Figure 4: Maximize Cost/Benefit of class Abnormal

Figure 5 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 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.

Figure 5: Minimize Cost/Benefit of class Abnormal

5. Discussions

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.

6. Conclusions

The main goal of this paper is to evaluate Ensembles design and Combining different algorithms to develop Novel Intelligent Ensemble Health Care Decision Support and Monitoring System to classify the situation of an emergency hospital based on the Vital Signs from Wearable Sensors. We reduced the number of attributes from 300 attributes to 6 attributes. We explored various ensembles combining model and evaluated the models with various methods of evaluation based on Error Metrics, ROC curves, Confusion Matrix, Sensitivity, Specificity and the Cost/Benefit methods. We compared the performance of the entire classifiers and 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.

.

Acknowledgment

The authors are grateful to the Ministry of health in the Khartoum state for letting the researchers get access to the healthcare monitoring data at Khartoum state hospitals. Our Special thanks to the management and staff at Al Baraha medical city in Shambat, Khartoum North Hospital for their cooperation with us in this research.

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), ISSN 2150-7988 Volume 5 (2013), pp. 741-750.

[2] http://www.worldlifeexpectancy.com/sudan-coronary-heart-disease

[3] Nealson,J.L, and Moreno, "Application of software Agent technology in Health care domain", vol..212.

[4] Dante I, Tapia, Ajith Abraham, Juan M. Corchado and Ricardo S. Alonso "Agents and ambient intelligence: case studies", J Ambient Intell Human Comput DOI 10.1007/s12652-009-0006 -2

[5] Aarts E, Roovers R. "Embedded system design issues in ambient intelligence". In: Basten T, Geilen M, Groot HD (eds) Ambient intelligence: impact on embedded system design. Kluwer, Norwell, pp. 11–29 2003.

[6] R. Jafari, W. Li, R. Bajcsy, S. Glaser, and S. Sastry, "Physical activity monitoring for assisted living at home," in BSN, 7, pp. 213–219, 2007.

[7] Augusto, J. C., and Nugent , C. D. "Designing Smart Homes The Role of Artificial Intelligence", Lecture Notes in Artificial Intelligence, Springer, Berlin, vol. 4008, 2006b.

[8] Augusto, J. C., Nugent, C. D.,artin, S., and OlphertL, C. "Towards personalization of services and an integrated service model for Smart Homes applied to elderly care". In Proceedings of the International Conference on Smart Homes and Health Telematics. Montreal, Canada, From Smart Homes to Smart Care, S. Giroux and H Pigot, Eds. IOS Press, pp. 151–158, 2005.

[9] Augusto, J. C., Nugent, C. D., AND Black, N. D. "Management and analysis of time-related data in Smart Homes environments". In Proceedings of the Mediterranean Conference on Medical and Biological Engineering (Medico), pp. 4–7, 2004.

[10] Augusto, J. C., and Nugent, C. D. "The use of temporal reasoning and management of complex events in Smart Homes". In Proceedings of the European Conference on Artificial Intelligence (ECAI'04). R. L. de M´antaras and L. Saitta, Eds. IOS Press, pp. 778–782, 2004.

[11] Augusto, J. C., and Nugent, C. D. "Smart homes can be smarter. In Designing Smart Homes: The Role of

Artificial Intelligence", J. C Augusto and C. D. Nugent, Eds., Lecture Notes in Artificial Intelligence, vol. 4008, Springer, Berlin, pp. 1–15, 2006a.

[12] Corchado, J. M., Bajo, J., A. Abraham,. "GerAmi: Improving Healthcare Delivery in Geriatric Residences". In IEEE Intelligent Systems, pp. 19-25, 2008.

[13] G. M. Youngblood. "Automating inhabitant interactions in home and workplace environments through data-driven generation of hierarchical partially-observable Markov decision processes". PhD thesis, The University of Texas at Arlington, 2005.

[14] S. W. Loke. "Representing and reasoning with situations for context-aware pervasive computing: a logic programming perspective". The Knowledge Engineering Review, 19(3), pp. 213-233, 2005.

[15] F. Doctor, H. Hagras, and V. Callaghan. "A fuzzy embedded agent-based approach for realizing ambient intelligence in intelligent", Phd thesis. The University of Texas at Arlington, 2005.

[16] Luhr, S., West, G., & Venkatesh, S. " Recognition of emergent human behaviour in a smart home: A data mining approach". Pervasive and Mobile Computing, 3(2), pp. 95-116, 2007.

[17] Bager, T. S., Brown, D. E., Alwan, M. "Health-Status Monitoring Through Analysis of Behavioral Patterns". In IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans, 35(1) ,pp. 22-27, 2005.

[18] S. Patel, R. Hughes. "Using Wearable Sensors to Predict the Severity of Symptoms and Motor Complications in Late Stage Parkinson's Disease". 30th Annual International IEEE EMBS Conference, British Columbia, Canada, pp. 20-24, 2008

[19] V. Vapnik, "The Nature of Statistical Learning Theory". New York, NY: Springer-Verlag, 1995.

[20] E. M. Tapia, S. S. Intille, and K. Larson. "Activity recognition in the home using simple and ubiquitous sensors". In Proceedings of Pervasive, pp. 158-175, 2004.

[21] J. Augusto and C. Nugent. "The use of temporal reasoning and management of complex events in smart homes". In R. L. de M¶antaras and L. Saitta, editors. In Proceedings of European Conference on Artificial Intelligence (ECAI), pp. 778- 782, 2004. IOS Press (Amsterdam, The Netherlands), 2004.pp. 22-27, 2004.

[22] University of Washington. The assistive cognition project. URL: http://www.cs.washington.edu/ assistcog.

[23] J.R. Beck and S.G. Pauker, "The Markov process in medical prognosis", Med Decision Making, 3(4), pp. 419–458, 1983.

[24] T.Y. Leong, "Dynamic decision modeling in medicine: a critique of existing formalisms". In Proceeding Annu Symp Comput Appl Med Care, (AMIA, Washington, DC) , pp. 478–484, 1993.

[25] J.E. Stahl, "Modelling methods for pharmacoeconomics and health technology assessment: an overview and guide", Pharmacoeconomics, 26(2), pp. 131–148, 2008.

[26] I. Fatima , M. Fahim , Y.K. Lee , S.Y. Lee, "Classifier Ensemble Optimization for Human Activity Recognition in Smart Homes". http:/uclab.khu.ac.kr/resources/publication/c_268.pd f

[27] A. Tan, D. Gilbert, Ensemble machine learning on gene expression data for cancer classification, Applied Bioinformatics 2 (3), pp. 75–83, 2003.

[28] B. Liu, Q. Cui, T. Jiang, S. Ma, "A combinational feature selection and ensemble neural network method for classification of gene expression data", BMC Bioinformatics 5, (136), pp. 51- 131, 2004.

[29] P. Li, K. Chan, S. Fu, S.M. Krishnan," An abnormal ecg beat detection approach for long-term monitoring of heart patients based on hybrid kernel machine ensemble", in: Nikunj C. Oza, Robi Polikar, Josef Kittler, Fabio Roli (Eds.), Proceedings of the Sixth International Workshop on Multiple Classifier Systems, Springer, Berlin, pp. 346–355, 2005.

[30] B. Scholkopf, J.C. Platt, J. Shawe-Taylor, A.J. Smola, R.C. Williamson, Estimating the support of a high-dimensional distribution, Neural Computation 13 pp. 1443–1471, 2001.

[31] http//www.fipa.org (Foundation for Intelligent Physical Agents).

[32] My Chau Tu, Dongil Shin, Dongkyoo Shin ,"Effective Diagnosis of Heart Disease through Bagging Approach", 2nd International Conference on Biomedical Engineering and Informatics,2009.

[33] My Chau Tu, Dongil Shin, Dongkyoo Shin, "A Comparative Study of Medical Data Classification Methods Based on Decision Tree and Bagging Algorithms" Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing, 2009.

[34] V. Chauraisa and S. Pal, ―Data Mining Approach to Detect Heart Diseases‖, International Journal of Advanced Computer Science and Information Technology (IJACSIT),Vol. 2, No. 4,2013, pp 56-66.

[35] Pan Wen, "Application of decision tree to identify a abnormal high frequency electrocardiograph", China National Knowledge Infrastructure Journal, 2000.

[36] Kaewchinporn .C, Vongsuchoto. N, Srisawat. A " A Combination of Decision Tree Learning and Clustering for Data Classification", 2011 Eighth International Joint Conference on Computer Science and Software Engineering (JCSSE).

[37] Jinyan Li, Huiqing Liu, See-Kiong Ng and Limsoon Wong," Discovery of significant rules for classifying cancer diagnosis data", Bioinformatics 19(Suppl. 2) Oxford University Press 2003.

[38] Dong-Sheng Cao, Qing-Song Xu ,Yi-Zeng Liang, Xian Chen, "Automatic feature subset selection for decision tree-based ensemble methods in the prediction of bioactivity", Chemometrics and Intelligent Laboratory Systems..

[39] Liu Ya-Qin, Wang Cheng, Zhang Lu," Decision Tree Based Predictive Models for Breast Cancer Survivability on Imbalanced Data" , 3rd International Conference on Bioinformatics and Biomedical Engineering , 2009.

[40] Tan AC, Gilbert D. “Ensemble machine learning on gene expression data for cancer classification”, Appl Bioinformatics. 2003;2(3 Suppl):S,pp.75-83,2003.
[41] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009);The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.
[42] R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufman, San Mateo, CA, 1993.
[43] Margaret H. Danham,S. Sridhar, ” Data mining, Introductory and Advanced Topics”, Person education , 1st ed., 2006
[44] Aman Kumar Sharma, Suruchi Sahni, “A Comparative Study of Classification Algorithms for Spam Email Data Analysis”, IJCSE, Vol. 3, No. 5, pp. 1890-1895,2011
[45] Niels Landwehr, Mark Hall, Eibe Frank. Logistic Model Trees. Machine Learning. 95(1-2), pp.161-205, 2005.
[46] Breiman Leo. Random Forests. Machine Learning. 45(1): 5-32, 2001.
[47] Breiman, L.: Bagging predictors. Machine Learning 24(2), pp. 123–140, 1996
[48] Witten, I.H. & Frank, E. 2000. Weka machine learning algorithms in java, in Data Mining: Practical Machine Learning Tools and Techniques with Java.
[49] Freund, Y., Schapire, R.E.: A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting. In EuroCOLT '95, pp. 23–37. Springer-Verlag: London (1995).
[50] Liang, G. & Zhang, C. (2011), Empirical study of bagging predictors on medical data, in P. Vam- plew, A. Stranieri, K.-L. Ong, P. Christen & P. J. Kennedy, eds, `Australasian Data Mining Conference, AusDM 2011', Vol. 121 of CRPIT, ACS, Ballarat, Australia, pp. 31-40, 2011.
[51] Freund, Y., Schapire, R., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. 55, pp. 119–139, 1997.
[52] Wolpert, D. (1992), `Stacked generalization', Neural Networks 5(2), pp. 241-259, 1992.
[53] Breiman, L., Stacked Regression, Machine Learning, 24, pp. 49-64, 1996.
[54] Quinlan, J. R., 1996. Bagging, boosting, and C4.5. Proceedings of the Thirteenth National Conference on Artificial Intelligence: pp. 725-730 1996.
[55] 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, 2006.
[56] M. Panda and M. R. Patra, "Ensemble voting system for anomaly based network intrusion detection," International journal of recent trends in engineering, vol. 2, pp. 8-13, 2009.
[57] L. Felföldi, "Classifier Combination Systems and their Application in Human Language Technology," Research Group on Artificial Intelligence, Szeged,Hungary, 2008.
[58] L. Kuncheva, “Fuzzy vs Non-fuzzy in combining classifiers designed by boosting “ IEEE Transaction on Fuzzy Systems, vol. 11 , no. 6, pp. 729- 741 , Dec 2003.
[59] J. Kittler, M. Hatef, R. Duin and J. Matas, "On combining classifiers," IEEE Trans Pattern Analysis and Machine Intelligence, vol. 20, no. 3, p. 226–239, 1998.
[60] Juan Fraile, Javier Bajo, Juan Corchado, Ajith Abraham, Applying wearable solutions in dependent environments, IEEE Transactions on Information Technology in Biomedicine, IEEE Press, 14(6):1459-67, 2010.
[61] Juan Corchado, Javier Bajo and Ajith Abraham, GerAmi: Improving the delivery of health care in geriatric residences, IEEE Intelligent Systems, IEEE Computer Society Press, 23(2):19-25, 2008.