2014 IEEE International Conference on Systems, Man, and Cybernetics
October 5-8, 2014, San Diego, CA, USA

One-class Classification for Heart Disease Diagnosis

George Gomes Cabral
Statistics and Informatics Department
Rural Federal University of Pernambuco
Recife, Brazil
george.cabral@gmail.com

Adriano Lorena Inácio de Oliveira
Center of Informatics
Federal University of Pernambuco
Recife, Brazil
alio@cin.ufpe.br

***Abstract*—As has been shown by the recent literature, machine learning techniques are important tools for diagnosing a number of diseases. Hospitals and medical clinics store a large amount of data with respect to the treatment of their patients. However, rarely an analysis of these data is conducted in order to extract intrinsic information for modeling a specific problem. This work presents an analysis of medical data aimed at determining whether or not patients are cardiac. To this end, raw data was collected and preprocessed at a Brazilian local hospital in order to build a new dataset containing only non-invasive information of children with heart murmur symptoms. The gathered data contain information, such as height, weight, gender and birthday date. The collected data was shown to be very imbalanced. Due to this imbalance, we employ the One-class Classification (OCC) paradigm to solve the problem by experimenting five methods; including the FBDOCC, that we proposed in a previous paper. Furthermore, two additional datasets were experimented in order to assess effectiveness of One-Class classifiers on the domain of heart disease detection. The overall results show that the FBDOCC succeeded in this task, yielding, statistically, the best performance for the gathered dataset as well as the other two heart disease datasets.**

I. Introduction

Medicine is a complex domain, not well understood in all its aspects. Therefore it is very fruitful for exploring intelligent systems [16]. Cardiac diseases are highly prevalent and responsible for the highest morbidity and mortality rates, not just in Brazil, but in most modern societies as well [17]. Many different tools and techniques are being continuously developed aimed at analyzing cardiac performance. In many techniques, the diagnosis of a cardiac disease is often based on a combination of the patient's history and his/her physical examination. Its confirmation comes from exams such as the electrocardiogram (ECG) and the echocardiogram (ECHO).

In most cases, particularly in children, these exams are necessary before the establishment of the diagnosis and its severity. However, they are not always available at short notice, especially in developing countries where there is a lack of specialists in many urban centers, such as pediatric cardiologists in pediatric emergency services. This situation may lead to delays in diagnosis with consequent deterioration of the clinical condition of the patient.

A heart murmur is a sound generated from turbulence in the blood flow. It can occur in different areas of the cardiovascular system and in different phases of the cardiac cycle (systole, diastole or both) and have different frequencies and durations.

A cardiac murmur is the most frequent alert sign of a cardiac problem. In a study from Israel [19], 86% of the neonates with a cardiac murmur in the first days of life were shown to have some form of structural heart disease. Another study [18] demonstrated that around 44% of cardiac malformations can be detected in the neonatal period, and that when a heart murmur is heard during physical examination, the chances of a structural heart lesion are around 54%. Thus, the presence of a heart murmur in a neonate does not necessarily imply in a cardiac malformation. In fact, health children can have murmurs, generally referred to as "innocent". The differentiation between a normal, "innocent" murmur from a pathological one is therefore fundamental in a child suspected of a cardiac disease.

Despite not being necessarily associated with a cardiac disease, cardiac murmurs, when detected, particularly in association with other symptoms, are an indication for further investigation to document or rule out such diseases. Experienced cardiologists can often differentiate an innocent from a pathological murmur. However, this recognition ability requires long periods of training and practice and is limited to a small group of professionals.

Computational techniques have been shown effective in providing a better understanding of Medical data [8] [9]. Clinical Decision Support Systems are designed to assist physicians and other health professionals in problems such as image diagnosis or clinical treatment guidelines. To this end, most of the works employ Artificial Intelligence techniques to model each particular problem. However, in the medical data collected to build a model of the problem, the number of examples of healthy patients is often substantially larger than the unhealthy ones. In other words, medical data often presents a substantial class imbalance. Thus, it is important to use specific techniques to cope with imbalanced data.

Recently, Farquad and Bose [23] have proposed a method that uses a Support Vector Machine (SVM) for preprocessing the data aiming to balance the dataset. Another option to handle imbalanced data is to use machine learning models that deal directly with the imbalance or employ conventional techniques for treating imbalance before the modeling phase. Using the minority class during the modeling phase assumes that the minority amount of data represents well this minority class. In fact, this assumption does not hold for most of the problems. So, in this work, instead of trying to solve the class imbalance, we try to create a closed description of the majority class that does not encompass the data from the minority class. To this end, we experimented five classifiers with different operation forms, namely: Feature Boundaries for One-Class Classification (FBDOCC) [22]; One-Class SVM

978-1-4799-3840-7/14/$31.00 ©2014 IEEE

(OCSVM) [1]; Support Vector Data Description (SVDD) [2]; Gaussian Process for One-class Classification (GP-OCC) [26]; and kernel PCA (kPCA) [3].

This work reports on experiments that compare the aforementioned methods for heart disease detection. To this end, it was collected a dataset with about 1792 children patients having symptoms of heart murmur. This dataset was extracted from an information system of a reputed Brazilian pediatric hospital. Our goal is to create a model to predict heart disease based on the detection of heart murmur on clinical examination as well as other non-invasive information on the patient, such as heart rate, height (in meters), weight. Additionally, two other heart disease datasets were used in the experiments in order to check the capability of one-class classifiers to encompass the normal data in this domain.

This paper is structured into six sections. Section II presents a background on medical diagnosis problem as well as related works. Section III describes the datasets used in the experiments. Section IV briefly presents the classification and parameter optimization techniques used in this paper. Section V details the experiments and analyzes the results. Finally, Section 6 presents conclusions and future works.

II. Related Work

A number of works have reported good results in the use of intelligent methods for medical diagnosis. In [8], Sartakhti et. al. propose the application of a hybrid system comprised of SVMs and Simulated Annealing optimization technique to solve the problem of hepatitis diagnoses. The authors declare that medical diagnostics are quite difficult and only expert doctors should perform visual analyses. Therefore, due to this difficulty, many data mining techniques have been considered for building automatic diagnosis systems for hepatitis in order to support the specialists in their decisions.

Artificial Neural Networks (ANNs) have been used for diagnosing several kinds of diseases. Chest diseases like chronic obstructive pulmonary, pneumonia, asthma, tuberculosis and lung cancer were addressed by Er et. al. [9]. In this study, the authors have used several neural networks based classifiers, such as Multilayer Perceptron neural networks (MLPs), probabilistic neural networks, learning vector quantization and generalized regression. They have also prepared a dataset by using patient's epicrisis reports from a chest diseases hospital's database. There is a number of works in which the Multilayer Perceptron (MLP) topology was successfully employed for disease diagnosis systems, for instance [10].

Medical image diagnosis has became a common application of ANNs [11]. ANNs are adopted for many medical problems due to their capability of optimizing the relationship between the inputs and outputs via distributed computing, training, and processing, leading to reliable solutions. It is important to notice, however, that image segmentation and edge detection processes must be performed beforehand with the aim of extracting meaningful regions of the image [12].

Another highly adopted technique for medical analysis is the Decision Tree classifier (DT) [14]. The main advantage of Decision Trees is their capability to graphically represent the decision/classification. Often, it is quite difficult for researchers of areas other than computing to trust in a decision offered by a model which does not explain how that results were generated (e.g., models such as ANNs and SVMs, known as *black box*). López-Vallverdú et. al. [13] have combined medical criteria to build more robust decision trees for health care decision. In Rahman and Hasan [15], a surveillance system collects on clinical, epidemiological and demographic characteristics of patients in order to classify the critical condition of the patient as low, medium or high. The authors advocate that the decision models created in [15] could be helpful during epidemic when huge number of patients arrive daily.

III. Employed Datasets

This Section describes the data collected for this work and briefly describes the other two UCI datasets, as well.

A. Heart Murmur Dataset

For this dataset, data from 1792 patients with symptoms of heart murmur were extracted from an information system of a reputed Brazilian pediatric hospital. When a child patient arrives at the hospital, standard procedures are used to collect basic information such as heart rate, height (in meters), weight, etc.

The provided information can be summarized as follows (attribute name, data type, statistics): (gender, nominal, 53% male); (weight, real, 19.5 ($\pm$14.6)); (height, real, 1.0 ($\pm$0.32)); (heart rate, integer, 103 ($\pm$25.4)); (birth date, date, -); (attendance date, date, -); and (cardiac (class), nominal, 24.2% yes). It is easy to note that the dataset is unbalanced due to the larger number of healthy patients (non cardiac).

A preprocessing was performed in order to make the data suitable for a proper use by the classifiers. The *height* attribute, for instance, was collected in two different units: meters and cm. In this case all values were converted to meters. Another pre-processing was performed for the attribute *weight* which was converted from grams to kilograms. The attribute *gender* did not need any pre-processing. The attribute *age* was created according to the birth date.

Furthermore, the attributes *Correctly Filled Gender*, *Body Mass Index* (BMI) and *Body Surface Area* (BSA) were derived from the raw data. The attribute *Correctly Filled Gender* is a flag that was added due to the high percentage of missing data for the gender attribute.

The attribute BMI is a worldwide known international measure used to classify an individual based on his weight and height. The BSA attribute represents the area of skin in square meters. This attribute is an important factor for calculating the correct dosage of some medicines. For instance, chemotherapeutic medicines are more related to the BSA than to the weight. Equation 1 computes the BSA attribute. In Eq. 1, the units are square meters, cm and kilograms for the attributes BSA, height and weight, respectively.

$$BSA = 0,007184 \times Height^2 \times Weight^{0,425} \tag{1}$$

After this phase, the final dataset contains eight input features and the target, as following: (i) Gender; (ii) BMI; (iii)

Heart Rate; (iv) Correctly Filled Gender; (v) Age; (vi) height; (vii) BSA; (viii) weight; and (ix) Cardiac.

B. Heart Disease Dataset

This data set is composed of 270 instances (150 of the normal class and 120 of the abnormal class) and 76 features; however, most of published experiments refer to using a subset of 14 features. This data set contains four different values for the presence of heart disease. Thus, instances belonging to these values were relabeled to compose only one class; the abnormal class. This data set is available at the UCI repository [27].

C. SPECTF Heart Data Set

This dataset is also provided by the UCI repository and is composed of 267 instances (55 of the normal class and 212 of the abnormal class) and 44 continuous feature patterns. The data describes diagnosing of cardiac Single Proton Emission Computed Tomography (SPECT) images. Each of the patients is classified into two categories: normal and abnormal. The database was processed to extract features that summarize the original SPECT images. As a result, 44 continuous feature patterns were created for each patient.

IV. Computational Intelligent Techniques

This Section briefly reviews the methods experimented to tackle the problem of the early detection of cardiac disease. Furthermore, the optimization method (Particle Swarm Optimization - PSO), used to search for the best parameters of the techniques FBDOCC, One-class SVM, SVDD and kernel PCA is also detailed in this Section.

A. Features Boundaries Detection for One-class Classification - FBDOCC

The underlying idea behind our FBDOCC (Feature Boundaries Detector for One-Class Classification) [22] method is to explore all feature dimensions of the problem for each instance in a training set containing only instances of the normal class in order to find the boundaries which tightly encompass this training distribution. Let $T = \{t_1^l, t_2^l, .., t_n^l\}$ be a training set containing n vectors of size l, and let X be a Hilbert space with an induced metric ρ which measures the similarity between two vectors such that each $t_i^l \in X$. Let ρ the Euclidean distance, the FBDOCC generates $2l$ new artificial prototypes for each training instance t_i in a relatively small distance (defined by the parameter r) to it. Each artificial prototype p_j aims at representing a piece of the border between the normal and the abnormal (unknown) classes. The aim is to generate one hyper sphere with radius th for each prototype p_j and check whether or not there is any training instance (except the generating instance) inside this hyper sphere.

If all the training instances t_i are located outside the hyper sphere defined by p_j (i.e., $\rho(t_i, p_l) > th \ \forall \ t_i \in T$), then the following occurs: (1) information about p_j is stored in order to reproduce it as a positive prototype in the test phase, and (2) the training instance which has created this prototype is added to the set of negative instances. Once information about one of the $2l$ artificial prototypes is stored, the remaining $2l - 1$ prototypes generated by the current training instance are then discarded. The positive prototypes define the abnormal class whereas the set of negative prototypes represents the normal class. Once the training phase is finished, the model consists of a set formed by negative instances (*NegSet*) and a data structure containing minimal information in order to recreate the positive instances set (*PosSet*).

B. One-class SVM

Based on SVMs, Scholkopf et. al. [1] have proposed the one-class SVM (OCSVM). In OCSVM, The kernel function maps the training objects to a feature space. In the feature space, OCSVM then recognizes the origin as the only member of the second class (the novelty class). In contrast to the SVDD (Support Vector Data Description) [2] which tries to find the less bulky hyper sphere which contains almost all of the training objects, the One-class SVM tries to find the hyper plane which separates the training data with maximal distance from the origin in the feature space. The goal is to maximize the margin of separation to the origin. As in the multi-class SVM, slack variables denoted by $\vec{\xi_i}$ enable some training objects to fall outside the side of the hyper plane which represents the normal class (i.e., misclassifies some training objects). A training sample is a support vector when it is misclassified or falls inside the description of the hyper plane.

C. SVDD - Support Vector Data Description

The Support Vector Data Description (SVDD) classifier [2] works by finding the smallest sphere enclosing the data. As in multi-class SVMs, slack variables, denoted by $\vec{\xi_i}$, are associated to each data sample. This enables some of the training data samples to fall outside the description (i.e., are misclassified as outliers) when the minimum radius is found.

In order to separate the data from the origin with maximum margin, the following quadratic problem must be solved:

$$\vec{min}\frac{1}{2}\|\omega\|^2 - \rho + \frac{1}{\nu\ell}\sum_{i=1}^{l}\vec{\xi_i} \tag{2}$$

where ω is the normal vector to the separating hyper plane, ℓ is the number of training samples and ρ is the offset, subject to $(\omega \bullet \Phi(x_i)) \geq \rho - \xi_i \quad i = 1, 2, ..., \ell \quad \xi_i \geq 0$.

If ω and ρ solve this problem, then we have found a function $f(x) = sign\left((\omega \bullet \Phi(x))\right) - \rho)$ such that if $f(x) > 0$, the object x is classified as normal. Otherwise, x is classified as novelty.

D. GP-OCC - Gaussian Procce for One-Class Classification

In [26], Gaussian Process Priors were successfully applied for OCC problem and a kernel based learning is formulated in a Bayesian framework. The authors propose the idea of using an adequate Gaussian process in order to derive useful membership scores. Achieving the smaller mean of the prior, for instance, from the normal data and comparing it to the mean of a test object z region indicates whether or not z belongs to the normal class. This idea restricts the space of probable latent functions (used as kernels) to the ones where the values gradually decrease as the object gets far from the training set distribution.

Kemmler et. al. [26] experimented four different measures to achieve the scores that are derived from the predictive distribution, namely, the mean (μ_*), the negative variance ($-\sigma_*^2$), the predictive probability (P) and a heuristic (H). Thus, for a training set X with objects in the form (x, y), where x is the features vector and y the label of the class, and being (x_*, y_*) a test object, the following Equations show each measure. Considering these four measures, the authors motivated the use of the mean and the negative variance. Additionally, the experiments conducted in [26] show that the negative variance outperforms the mean measure.

E. kernel PCA

In [3], Hoffmann proposes a method aimed at giving lower classification errors by generating a decision boundary, in general, according to the author, tighter than that of one-class SVM, yielding thus a better description of the data. The author introduces the application of a kernel version of the standard PCA (Principal Component Analysis), namely Kernel PCA, to address the problem of novelty detection. In contrast to the standard PCA, where the eigenvectors and eigenvalues are computed from the covariance matrix, the Kernel PCA computes them from a kernel matrix. Instead of using the dot product, the kernel matrix is constructed by adopting a kernel function which receives two data points as arguments. So, each entry of the matrix is in the form $k_{i,j} = \kappa(x_i, xj)$, where κ is the kernel function and x_i and x_j are two data instances.

Considering a dataset consisting of n data points, the operation of subtracting the mean of each entry in the standard PCA is then replaced by the use of Equation 3. Thus, the features in high dimensional space of the kernel matrix become zero mean.

$$k_{i,j} = k_{i,j} - \frac{1}{n}\sum_{r=1}^{n} k_{i,r} - \frac{1}{n}\sum_{r=1}^{n} k_{j,r} + \frac{1}{n^2}\sum_{r,s=1}^{n} k_{r,s} \tag{3}$$

After computing the eigenvector matrix α in the kernel space, the measure of novelty is calculated according to the reconstruction error [25].

F. Optimization Method

Particle Swarm Optimization (PSO) was proposed by Kennedy and Eberhart [5]. This technique is usually aimed at finding optimal solutions for non-linear problems. Inspired by the social behavior of a group of birds, the main intuition behind PSO is to build a set (population) of particles to simulate the movements performed by the birds while searching for food in a specific region. This technique explores the social behavior of an intelligent set of individuals along with their ability to communicate to find a global solution. Each PSO particle represents a solution (model) in an n-dimensional space, where n stands for the number of parameters to be optimized.

In the standard implementation of PSO, particles move inside the multi-dimensional search space using a combination of attraction to the best solution found by this individual particle and an attraction to the best particle belonging to its neighborhood [6] [7]. A neighborhood is defined as a subset of the swarm whose particles are able to establish communication. The swarm moves in the search space by updating the velocity and position of each particle [4].

G. Performance Measures

The one-class classifiers will be assessed considering four metrics: (i) the Area Under the Curve (AUC) of the Receiver Operating Characteristic curves (ROC) of each model; (ii) the Matthews Correlation Coefficient (MCC); (iii) the Sensitivity; (iv) the Specificity.

The AUC (Area Under the Curve) is frequently used to evaluate one-class classifiers and methods for novelty detection [2] [24]. The Receiver Operating Characteristic (ROC) curve is composed by the Sensitivity (y-axis) and False Positive rate, or 1 - Specificity (x-axis). Each point on the ROC curve represents a sensitivity/specificity pair corresponding to a particular decision threshold; or model. The Sensitivity (Equation 4) is the success rate for the positive class and the Specificity (Equation 5) the success rate for the negative class.

$$Sensitivity = \frac{TP}{TP + FN} \tag{4}$$

$$Specificity = \frac{TN}{TN + FP} \tag{5}$$

In Equations 4 and 5, for a given operation point of the ROC curve, TP is the number of true positives, FN the number of false negatives, TN the number of true negatives and FP the number of false positives.

The MCC provides a balanced evaluation of the prediction. In other words, it can be used even if the classes are of very different sizes [21] [20]. It returns a value between -1 (total disagreement between prediction and observation) and $+1$ (perfect prediction). Equation 6 depicts its computation.

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FN)(TP + FP)(TN + FP)(TN + FN)}} \tag{6}$$

V. Experiments and Discussion

This Section reports on the experiments carried out for evaluating the performance of the aforementioned methods. Each entire dataset was shuffled 25 times and for each shuffle, fifty percent (50%) of the normal examples of the original dataset were used for training the classifiers. The rest of the examples, including the examples from the minority class, were equally divided between the validation and test sets respecting each class proportion. Therefore, each result value shown from now on is the average of these 25 randomizations.

For each experiment, the Wilcoxon Rank Sum statistical test with significance of 95% was performed in order to identify the best result. The Tables I, II, and III present the best results in boldface. This test is encouraged by Demsar [28] and was also performed by Wu and Ye in [29].

A. Parameter Selection

All the classifiers, except the GP-OCC (where no parameter selection is needed), employed the PSO method aimed at optimizing the specific parameters of each classifier for each one of the 25 randomizations of each data set. A preliminary optimization was carried out for the methods OCSVM, SVDD and FBDOCC in order to find the best problem features set of the Heart Murmur data set. For the three classifiers, the optimizations converged for the following six features of the problem: Weight, Height, Heart Rate, Age, BSA and BMI. Thus, for all the experiments with all classifiers, only these six features were adopted. The fitness function was designed to maximize the validation MCC constrained to the sensitivity lower bound of 81%. This value was a consensual choice of local specialists; the idea is to prioritize the detection of positive cases (cardiac) over the recognition rate of normal (negative) cases.

B. Results for SPECTF Heart Data Set

Table I comprises all the aforementioned measures considering all the methods for the data set SPECTF Heart Data Set. The models of Table I were all chosen based on the validation error and then assessed on the test set. Each cell value consists of the average obtained from twenty-five executions to find out the best model. This average is followed by its standard deviation. According to this table, the FBDOCC obtained better performances for all the metrics and It is clear the better performance in terms of MCC and Specificity.

TABLE I. RESULTS FOR THE SPECTF HEART DATA SET.

Metric	FBDOCC	OCSVM	SVDD	GP-OCC	kPCA
AUC	**0.848** $\pm(0.028)$	0.831 $\pm(0.033)$	0.824 $\pm(0.040)$	0.818 $\pm(0.031)$	0.757 $\pm(0.230)$
MCC	**0.500** $\pm(0.057)$	0.408 $\pm(0.104)$	0.386 $\pm(0.097)$	0.383 $\pm(0.083)$	0.393 $\pm(0.068)$
Sensitivity	**0.857** $\pm(0.028)$	**0.842** $\pm(0.025)$	0.834 $\pm(0.025)$	0.840 $\pm(0.023)$	0.829 $\pm(0.022)$
Specificity	**0.765** $\pm(0.090)$	0.675 $\pm(0.147)$	0.656 $\pm(0.127)$	0.604 $\pm(0.111)$	0.675 $\pm(0.101)$

C. Results for Heart Disease Dataset

Table II presents the results obtained in the simulations for the Heart Disease Dataset. This Table shows that the FBDOCC detects almost 100% of the abnormal (cardiac) patients without a significant decrease in the recognition rate of normal patients.

TABLE II. RESULTS FOR THE DATASET HEART DISEASE.

Metric	FBDOCC	OCSVM	SVDD	GP-OCC	kPCA
AUC	**0.779** $\pm(0.035)$	0.593 $\pm(0.039)$	0.617 $\pm(0.049)$	0.638 $\pm(0.046)$	0.632 $\pm(0.043)$
MCC	**0.850** $\pm(0.046)$	0.120 $\pm(0.098)$	0.189 $\pm(0.115)$	0.200 $\pm(0.083)$	0.204 $\pm(0.090)$
Sensitivity	**0.990** $\pm(0.037)$	0.838 $\pm(0.021)$	0.839 $\pm(0.028)$	0.831 $\pm(0.022)$	0.832 $\pm(0.020)$
Specificity	**0.832** $\pm(0.057)$	0.265 $\pm(0.090)$	0.327 $\pm(0.112)$	0.345 $\pm(0.072)$	0.348 $\pm(0.090)$

D. Results for Heart Murmur Dataset

In Table III, the results for the children heart disease detection based on non invasive data are presented. The results show that the FBDOCC performed similar to the SVM based classifiers. Given a lower bound for the error rate in the positive cases (i.e., 81%), the methods face the trade-off of detect the positive cases and correct detect the negative cases. The FBDOCC performed slightly better than the SVM classifiers.

Figure 1 depicts the results of all the experiments grouping the datasets regarding each performance measure. For the AUC, the FBDOCC and the SVM based classifiers perform similar, except for the Heart Disease Dataset, where the FBDOCC significantly outperforms all the methods. The MCC varies from -1 to 1, however, all the classifiers reached positive rates, so, only the positive part of the chart is depicted. For this measure, the FBDOCC outperforms all the classifiers, in particular for the Heart Disease Dataset. The rate of correct detection of the positive cases had the value lower bounded in 81%, so, all the classifiers performed well for this measure. However, the FBDOCC had a remarkable better performance than the other classifiers for the Heart Disease Dataset. The Specificity analysis shows that, except for the Heart Disease Dataset, all the methods performed well. The FBDOCC achieved good rates for all the datasets.

TABLE III. RESULTS FOR THE DATASET HEART MURMUR.

Metric	FBDOCC	OCSVM	SVDD	GP-OCC	kPCA
AUC	**0.833** $\pm(0.010)$	0.828 $\pm(0.014)$	**0.832** $\pm(0.013)$	0.723 $\pm(0.017)$	0.723 $\pm(0.020)$
MCC	**0.491** $\pm(0.044)$	0.464 $\pm(0.034)$	0.466 $\pm(0.040)$	0.262 $\pm(0.036)$	0.309 $\pm(0.031)$
Sensitivity	**0.807** $\pm(0.033)$	**0.812** $\pm(0.003)$	**0.813** $\pm(0.002)$	**0.812** $\pm(0.003)$	**0.813** $\pm(0.003)$
Specificity	**0.700** $\pm(0.056)$	0.668 $\pm(0.036)$	0.669 $\pm(0.041)$	0.448 $\pm(0.040)$	0.499 $\pm(0.035)$

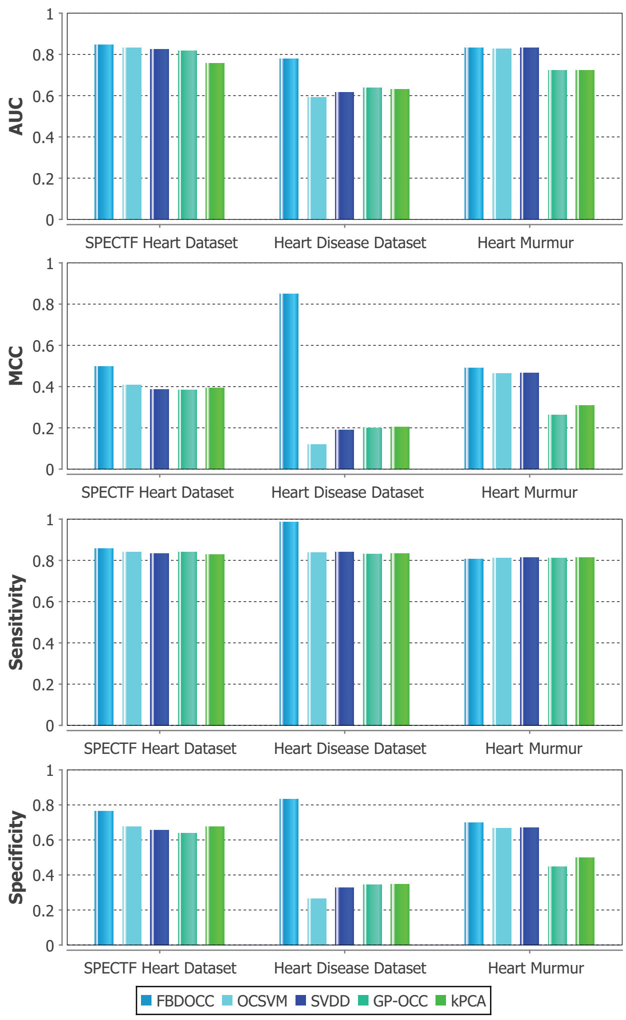

Fig. 1. Summary of all the experiments.

VI. CONCLUSION

Heart diseases are one of the major problems of public health. The early prediction of a heart disease is very important. In this context, the damage of classifying a healthy person as cardiac is substantially smaller than the damage of classifying a cardiac person as healthy. This may cause a delay on the treatment, thereby increasing the severity of the disease. Non-invasive data are cheap, fast to be acquired and may be relevant when applied to develop models able to predict if a child is or not cardiac prone. Five machine learning methods based on the OCC paradigm were investigated for this problem. This paradigm was chosen due to problem of the class imbalance. The underlying intuition is that the low number of positive examples may not adequately represent the unhealthy cases, but may support the validation of the description involving the healthy examples.

As a two-class problem treated by One-class Classification methods, this work has prioritized the case of the disease detection due to its higher importance against the recognition of healthy patients. Therefore, a sensitivity rate of 81% was fixed aimed at searching for models of classifiers with the lowest error rates constrained to this sensitivity level. Since this work was the very first one to use the collected data, a feature selection was preliminarily performed. The four metrics used in the experiments show that the methods FBDOCC, OCSVM and SVDD performed well, however, the FBDOCC significantly outperformed the other methods for the Heart Disease Dataset and, according the Wilcoxon Rank Sum test, obtained the best results for all the datasets. The results also show that the use of the OCC paradigm may support the screening phase of the treatment. In particular, the capability of using a predefined error rate for the positive class makes the use of the OCC very appropriate for the medical diagnosis problem domain.

As future work, the inclusion of other noninvasive data, such as family health history, may be considered to improve the correct decision rate. Furthermore, this methodology may be extended to other typical childhood diseases.

ACKNOWLEDGMENT

This work was supported by the National Institute of Science and Technology for Software Engineering (INES[1]), funded by CNPq and FACEPE, grants 573964/2008-4 and APQ-1037-1.03/08. Additionally, the authors would like to thank the professionals from the Pediatric Cardiology Network of the Heart Circle in the provinces of Pernambuco and Paraiba - Brazil (RCP-CirCor), for their unconditional support during the preparation of this manuscript.

REFERENCES

[1] B. Schölkopf, J. C. Platt, J. S. Taylor, A. J. Smola and R. C. Williamson. Estimating the support of a high-dimensional distribution. Neural Computation. Vol. 13(7), pp. 1443–1471, 2001.

[2] D. M. J. Tax.One-class classification concept-learning in the absence of counter-examples. Ph.D. thesis, Technische Universiteit Delft, 2001.

[3] H. Hoffmann. Kernel pca for novelty detection. Pattern Recognition. Vol. 40, pp. 863–874, 2007.

[4] J. Kennedy. Why does it need velocity? In: Swarm Intelligence Symposium. pp. 38–44, 2005.

[5] J. Kennedy and R. Eberhart. Particle swarm optimization. In: IEEE International Conference on Neural Networks (ICNN'95). Vol. 4, pp. 1942–1947, 1995.

[6] A. P. Engelbrecht. Fundamentals of Computational Swarm Intelligence. Wiley & Sons, 2005.

[7] D. Bratton and J. Kennedy. Defining a standard for particle swarm optimization. In: Swarm Intelligence Symposium. pp. 120–127, 2007.

[8] J. S. Sartakhti, M. H. Zangooei and K. Mozafari. Hepatitis disease diagnosis using a novel hybrid method based on support vector machine and simulated annealing (SVM-SA), Computer Methods and Programs in Biomedicine. Vol. 108(2), pp. 570–579, 2012.

[9] O. Er, N. Yumusak and F. Temurtas. Chest diseases diagnosis using artificial neural networks. Expert Systems with Applications. Vol. 37(12), pp. 7648–7655, 2010

[10] H. Temurtas, N. Yumusak and F. Temurtas. A comparative study on diabetes disease diagnosis using neural networks. Expert Systems with Applications, Vol. 36, pp. 8610-8615, 2009.

[11] W. Qian, T. Zhukov, D. S. Song and M.S. Tockman. Computerized analysis of cellular features and biomarkers for cytologic diagnosis of early lung cancer. Analytical and Quantitative Cytology and Histology. Vol. 29(2), pp. 103–111, 2007.

[12] J. Jiang, P. Trundle and J. Ren. Medical image analysis with artificial neural networks. Computerized Medical Imaging and Graphics. Vol. 34(2) pp. 617-631, 2010.

[13] J. A. L. Vallverdu, D. Riano and J. A. Bohada. Improving medical decision trees by combining relevant health-care criteria. Expert Systems with Applications. Vol. 39(14), pp. 11782-11791, 2012.

[14] J. R. Quinlan. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.

[15] R. M. Rahman and F. R. Md. Hasan. Using and comparing different decision tree classification techniques for mining ICDDR,B Hospital Surveillance data, Expert Systems with Applications, Vol. 38(9), pp. 11421-11436, 2011.

[16] W. Horn. AI in medicine on its way from knowledge-intensive to data-intensive systems. Artificial intelligence in medicine. Vol. 23(1), pp. 5-12, 2001.

[17] World Health Organization. The world health report 2010 - Health systems financing: the path to universal coverage, 2010.

[18] S. Ainsworth, J. P. Wyllie and C. Wren. Prevalence and clinical significance of cardiac murmurs in neonates. Archives of disease in childhood. Fetal and neonatal edition. Vol. 80(1), pp. 43-45, 1999.

[19] A. Rein and S. OmokhodionJ. Significance of a cardiac murmur as the sole clinical sign in the newborn. CLIN PEDIATR. n. 39, pp. 511-520, 2000.

[20] R. Gupta, A. Mittal and K. Singh. A Novel and Efficient Technique for Identification and Classification of GPCRs. IEEE Transactions on Information Technology in Biomedicine. Vol. 12(4), pp. 541-548, 2008.

[21] P. Baldi, S. Brunak, Y. Chauvin, C. A. F. Andersen and H. Nielsen. Assessing the accuracy of prediction algorithms for classification: an overview. Bioinformatics, 16(5), pp. 412-424, 2000.

[22] G. G. Cabral and A. L. I. Oliveira. One-class classification through optimized feature boundaries detection and prototype reduction. International Conference on Artificial Neural Networks. pp. 693–700, 2012.

[23] M. Farquad and I. Bose. Preprocessing unbalanced data using support vector machine. Decision Support Systems, Vol. 53(1), pp. 226-233, 2012.

[24] L. Cao, H. P. Lee and W. K. Chong. Modified support vector novelty detector using training data with outliers. Pattern Recognition Letters. Vol. 24 (14), 2479–2487, 2003.

[25] K. I. Diamantaras and S. Y. Kung. Principal Component Neural Networks. Wiley, New York, 1996.

[26] M. Kemmler, E. Rodner and J. Denzler. One-class classification with gaussian processes. In 10th Asian Conference on Computer Vision, Vol. 2, pp. 1025-1036, 2010.

[27] A. Frank and A. Asuncion. UCI machine learning repository, 2010 (available at http://archive.ics.uci.edu/ml).

[28] J. Demsar. Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research. Vol. 7, pp. 1-30, 2006.

[29] M. R. Wu and J. P. Ye. A small sphere and large margin approach for novelty detection using training data with outliers. IEEE Trans. Pattern Analysis and Machine Intelligence. Vol. 31(11), pp. 2088-2092, 2009.

[1] www.ines.org.br