

Support Center Machine Method for Classification and Help in Medical Diagnosis System

Zefeng Wang^{1*}, Laurent Peyrodie², Hua Cao² and Samuel Boudet²

¹Catholic University of Lille, FMM (HEI-USTB), 41 rue du port, Lille, Nord 59000, France

²Catholic University of Lille, FMM (HEI-USTB), 13 rue Toul, Lille, Nord 59000, France

Abstract

Objectives: A new artificial intelligent method 'Support Center Machines' (SCM) for helping diagnosis and prognosis is applied to a medical system. Methods In data processing, SCM seeks the true centers of each class during machine learning. For application in the medical system, it makes these centers as health-situation models and translates all the health records into a map. All the models, like non disease and diseases, are labeled in this map. Thus, the evolution of patient's health records can be supervised with the map. On the basis of the evolution of the distances from the recent record data to the centers, the system estimates the tendency of the healthy evolution and forecasts the probable situation in the future.

Results: SCM was tested on 'Wisconsin Breast Cancer Data' and compared with LDA and SVM methods. Twenty centers were found to define the healthy map. Based on the test results of four hundred and fifty random data selection for train, SCM has shown a better performance, whose means of correct detection ratios of the breast cancer varied from 91.4% to 95.6%, which were corresponding to 10% of data and 90% of data used to do machine learning. These ratios have increased by 1% to 5%, than SVM and LDA. In addition, the variance of correct detection ratios of SCM results has decreased by 0.8% to 3.0%, compared with SVM and LDA. Even if there were only 10% data for the training, the ratio stayed around 87% with only 3 principal components. When the system used 50% data for the training and tests the others, the mean of ratio was 93% and the best was 95%.

Conclusions: SCM successfully build a disease diagnosis/prognosis system and works out a healthy map. It could display the health record on 2D or 3D map, which allows the clinician to appropriate the interpretation. In addition, if a new situation (symptom / disease) occurs, the practitioner can visualize it and analyze it according to existing maps of SCM.

Keywords: Decision support systems; Computer-assisted diagnosis; Machine learning; Support center machines; Disease prognosis

Introduction

During recent years, with the rapid development of the computer, the clinical decision support systems (CDSSs) play an important role in the medical system [1]. Its goal is to quickly treat the huge massive of biomedical information, so as to get a rapid and accurate help for diagnosis. Those systems may find the undetectable symptom and could deduce the possible disease for people in the future. Several approaches and models have been developed since earliest 60's [2]. Generally, the experts-doctors experience is integrated in a computer-based diagnosis and prognosis system. For example in [3], Wang evaluated drug safety risks by FDA experts, which have been successfully applied to predict the potential safety liabilities of new drugs. In addition, experts-doctors experience could be translated to mathematical models by artificial intelligence or statistics tools in order to realize computer-aided medical diagnosis. For example, [4] used feed forward neural network constructive algorithm (MFNNCA) for medical diagnosis [5]; used decision and probability theory to construct diagnosis systems from a database of typical cases [6]; built a system of computer-aided diagnosis using a Bayesian analysis and displayed diagnostic probabilities in an adaptable format. Typically, recent prognostic methods rely on explicit physiological models [7], which may be combined with traditional models of life expectancy. These models are calculated by the analysis of life tables according to a particular experience of being attacked by a disease.

The method Support Center Machines (SCM) is a novel concept in artificial intelligent field for classification and prediction. It was first applied to faults diagnosis and prognosis of complex system, such as advanced aircraft [8]. It belongs to supervised learning methods that

analyze data and recognize patterns. In theory, SCM seeks the support vectors of true center and sub-centers of each class during the machine learning. The sphere of influence of each center is adjusted by its variance. If the testing data fall in one sphere, they would be attached to the center of this sphere. The sub-centers would deal with some complex condition, such as a sphere within a sphere of other class. Therefore, SCM could resolve the superposition problem of the spheres of influence of different classes. SCM makes each abnormal situation as a dysfunctional model and draws them in the map. Each model has one or more centers and spheres of influence. Additionally, the evolution tendency of the testing data could be monitored in this map. SCM here is different from ESCM, which was proposed by Z Lin in 2008 [9]. ESCM is a complementary algorithm of Support Vectors Machine (SVM), which can tune kernel parameters adaptively and control the number of kernels. SCM in this paper is an independent method to do classification and forecasting of the evolution of data.

In design, SCM owns obvious advantages in comparison with the classical classification methods. In the respect of the practicability, it

***Corresponding author:** Zefeng Wang, Associate professor, Catholic University of Lille, FMM (HEI-USTB), 41 rue du port, Lille, Nord 59000, France, Tel: 0033650618639; Fax:0033650618639; E-mail: zefeng.wang@hotmail.fr, zefeng.wang@icl-lille.fr

Received July 06, 2015; Accepted July 30, 2015; Published August 10, 2015

Citation: Wang Z, Peyrodie L, Cao H, Boudet S (2015) Support Center Machine Method for Classification and Help in Medical Diagnosis System. J Theor Comput Sci 2: 129. doi:10.4172/2376-130X.1000129

Copyright: © 2015 Wang Z, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

can be combined with Principal Component Analysis (PCA) or Linear Discriminant Dimensionality Reduction (LLDR) to reduce high-dimensional data [10] and create a visualization of the data. It can help doctors to intuitively understand the clinical results and to find the nature of disease. In the respect of the system structure, SCM is a supervised and suggestive method. Unlike neural networks [11] and decision tree [12], which distinguishes only the situations known by learning, SCM can detect the strange situation during tests, and add it into database after the verification. It does not need to re-learn all the rules, nor resolving the conflict of new rules and old rules. It is very useful in the medical system, because nobody can assure that all the disease types are already found until today. Even for expert doctors, it is difficult to discover the new disease type without obvious phenotype of patient. In the respect of the algorithm, SCM is also different from the classical forecasting systems, which often focus on researching regression line or hyper-plan, like Support Vector Machines (SVM) [13], based on the different biomedical attributes [14]. SCM could resolve the non-differential problem. In addition, it would monitor the evolution of patient's health records with the map, and deduce the healthy tendency in order to forecast the probable situation in the future. At last, in the respect of the compatibility, the kernel methods (KMs) could be used in this algorithm if the present feature space is not satisfactory to do the classification. KMs approach the problem by mapping the data into a high dimensional feature space, where the data might be more easily classified and each coordinate corresponds to one feature of the data items. SCM would be introduced in the following section.

Methods

Data

The work was tested by the well-known data set Wisconsin Breast Cancer Data [15]. This dataset consists of 569 breasts masses with 357 benign and 212 malignant cases [16].

Support centers machine

The inspiration of SCM was from the astronomy. As in the Galaxy, although there are milliard stars inside, each galaxy has its center and the stars are grouped around their center, such as the solar system. In a philosophical point of view [17], the center theory is ubiquitous and indubitable. Everything has its own center, like atom, cell, and universe and so on. SCM would find the centers by machine learning and use them to distinguish and classify the testing data.

As shown in Figure 1, the algorithm of SCM contains ten steps for the machine training. At first, the data should be normalized for the sake of organizing the fields to minimize redundancy and dependency [18]. The choice of the normalized method is determined by the data condition, Z-Score was used in this paper. But if the variances of different attribute-vectors are too large, Z-Score should not be proposed, because the distribution and the maximum and the minimum of the normalized data are due to their variance [19]. If the boundaries of the vectors of normalized data are quite different, it would be too complex to define a unitary diameter of the sphere of influence for the whole attribute-vectors. Thus, the system should consider other normalized methods in this case, for example 'unity-based normalization' can be generalized to restrict the range of values in the dataset between two any arbitrary points.

Dimensionality reduction of data could be done before or after the normalization, if necessary. The commonly used dimensionality reduction methods include supervised approaches such as linear

discriminant analysis (LDA) [20-21], and unsupervised ones such as principal component analysis (PCA) [22]. When the class label information is available, supervised approaches, such as LDA, are usually more effective than unsupervised ones such as PCA for classification.

According to the results of pre-classification (detailed in next section), the system finds the principal centers and the first group of sub-centers by the confusion matrix. In the field of machine learning, confusion matrix is a specific table layout that allows visualization of the performance of an algorithm. Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class. On the basis of the first confusion matrix of the pre-classification results, the principal centers and first group of sub-centers would be found. If there is still the misclassification by the centers found, the system would start the iteration to find more useful sub-centers to perfectly classify these data. During the learning, the real and useful centers are reserved, and the pseudo or useless centers are deleted. Therefore, the core of the training is to find the principal centers and sub-centers, which constitute the model of SCM. The test data would be used for evaluation. It needs to be emphasized that the rules of normalization must be the same for the training data and the test data. The program flow could be summarized as:

- (1) Normalization of data and treatment of data by PCA or LDA in order to reduce the high dimensionality of data
- (2) Pre-classification
- (3) Calculation of the first confusion matrix from the classification results of pre-classification
- (4) Searching of principal centers and the first group of sub-centers via the first confusion matrix
- (5) Classification by principal centers and the first group of sub-centers
- (6) Re-calculation of confusion matrix
- (7) Calculation of the ratio of misclassification, if the ration is less than or equal to the threshold σ , then go to 8; if the ratio is greater than σ , then go to 9

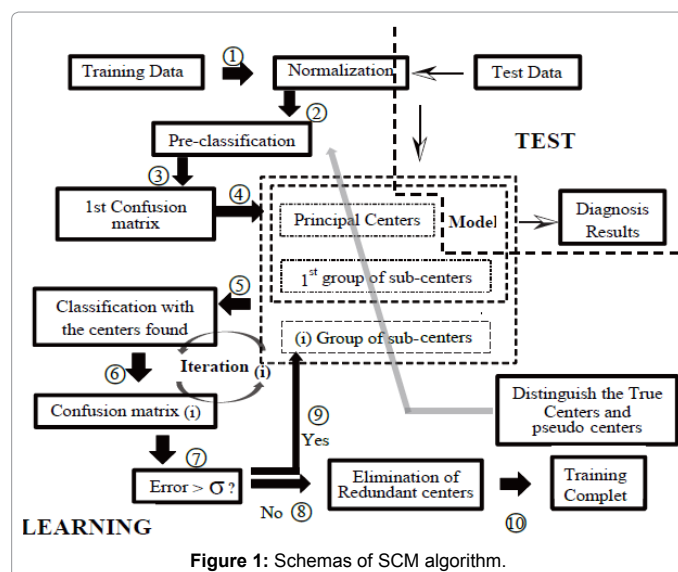


Figure 1: Schemas of SCM algorithm.

- (8) Elimination of Redundant centers by cross-validation
- (9) Searching for new sub-centers from the confusion matrix and adding the new sub-centers into dataset of centers, then go to 5 - iteration
- (10) Training completed.

As presented in the preceding text, all the centers are calculated by the confusion matrix of the results of the previous classification. Whereas, before all of this, the problem of origin should be resolved: Where does the first confusion matrix come from? The simplest is that the system can make some random centers to do the first classification, like K-means method. But here, as a supervised method, there are two ways suggested to complete the pre-classification. The first one is to use the training targets to calculate the principal centers. The system could use them to classify the training data and gain the first confusion matrix of the diagnosis results. The second one is that the system could use other classification methods to accomplish the pre-classification, such as Neural Networks, Gaussian Models Mixture and so on. SCM can either only use their classification results to find the principal centers and the first group of sub-centers, or be used as a complement system to improve the diagnosis performance.

Classification

In comparison with the well-known method SVM, which requires support vectors on the boundary to distinguish different classes? The target of the learning in SCM is to find the centers of each class, and to classify the data by distances between them and all the centers. In general, there is more than one center for each class Figure 2. It means that is necessary to find the sub-centers in order to control the different areas which have distances between themselves but belong to the same class.

As presented in Formula 1, there are k centers C_i^k $k = 1, 2, \dots, n$ of class C_i . Point p belongs to class C_i , if its nearest center belongs to class C_i and it is in the sphere of influence D_i^k of the center. If the point p is out of the sphere, it would be considered as an unknown particular. It means the training doesn't cover this zone. Thus, the system would take it as unknown situation and send a report to users.

$$p \in C_i \Leftrightarrow \exists k, \arg \min_c (|p - C|) = C_i^k \ \& \ |p - C_i^k| \leq D_i^k \quad (1)$$

As an example shown in Figure 3, the data need to be classified into three classes (A, B and C). Each class has 20 points. According to the distribution of the data, only one center for each class is not enough to completely classify all the data. Obviously, if there is only one center defined for each class, the points in the small spheres might be misclassified, because they are likely to be closer to other principal centers than to their own ones. These enclaves might be under the sphere of influence of the different class. It is very clear that the points in sphere A2 are closer to the center of class B than to class A. Therefore, the system would automatically add the sub-centers according to the needs. In order to contain more points, the principal centers normally have a big diameter (like spheres A1, B1 and C1 in Figure 3), though they contain some points which do not belong to them. Then, in order to deal with the misclassified points, sub-centers would be established during the iteration and distinguish these points by their spheres of influence, like spheres A2, B2, B3, B4 and C2 in Figure 3. Finally, a series of centers are established.

Principal center and sub-centers with confusion matrix

SCM are primarily based on confusion matrix of the classification results. The principal centers are researched by the diagonal of pre-

classification (or called as first classification) confusion matrix and the sub-centers are found out in the no-diagonal elements of confusion matrix afterwards. It is different from K-means [23] clustering which is an unsupervised algorithm and does not use any information from confusion matrix to find the location centers [24], since SCM could have supervisors in the training. In addition to this, the amount of centers depends on the need, which can be more than the dimension of the matrix. In fact, a new confusion matrix would be studied at every iteration, and the new sub-centers would be found for the no-diagonal cells of the matrix. As shown in Formula 3, a pre-classification is realized to find N principal centers for N classes, which are according to N diagonal cells of the confusion Matrix (M_{nn}). Then, the sub-centers are found in the no-diagonal cells of Matrix. If none of the no-diagonal cells of Matrix is empty, there would be $N-1$ sub-centers for each class via the pre-classification.

\exists Confusion Matrix $M_{n \times n}$ via Pre-classification

$$\text{Principal Centers} \in M_{ij}, \text{ with } i = j, \forall i, j = 1; \quad (2)$$

$$\text{Sub-centers} \in M_{ij}, \text{ with } i \neq j, \forall i, j = 1, 2, 3, \dots, n. \quad (3)$$

Finally in the first iteration, the principal centers and the first group of sub-centers are found by confusion matrix, and the system could do the classification with them. But if the error of the classification cannot be ignored, the system needs to do the second iteration so as to add sub-centers for improving the classification. In this way, sub-centers can be added until the performance goal is achieved. As shown in Figure 4, after the pre-classification, there are 13 points incorrectly classified. In the confusion matrix, 5 points of class A are misclassified into the class B. Thus, a sub-center had to be added to deal with these five points, which is shown as 'sphere of A2' in Figure 3. For the same reason, 'spheres of B2, B3 and C2' are determined by the no-empty items of the no-diagonal of this confusion matrix. With these centers found in the pre-classification, the classification result is improved and presented in the matrix of 'Iteration 1'. But there still remain 4 misclassified points, because some points are influenced by the sub-centers of the other class. For example, there are 3 points of class A close to the sub-center B3 in Figure 3. That is why the item (1,2) of the confusion matrix Iteration 1 displays '3'. Thus, it needs to add a new sub-center of class A for these 3 points, as the sphere of A3. Finally, all the points are correctly classified in the second iteration. All the items of no-diagonal of the confusion matrix 'Iteration 2' are empty. In addition, the sphere of influence of each center could be automatically defined as the distance between the center and the farthest one among the points which are according the

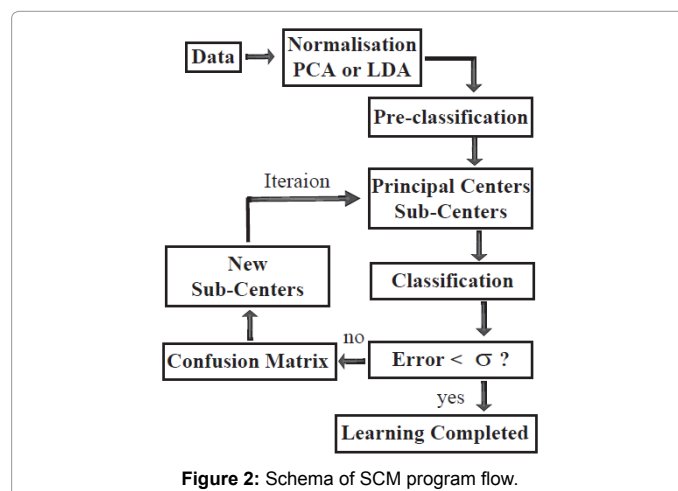


Figure 2: Schema of SCM program flow.

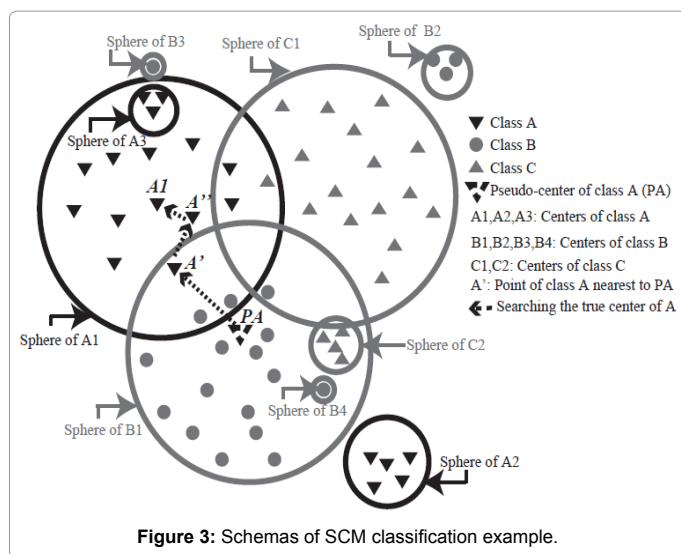


Figure 3: Schemas of SCM classification example.

confusion matrix element. Certainly, the spheres could be also man-made or adjusted by some parameters.

However, it needs to emphasize that the diagonal of confusion matrix is not considered any more after the pre-classification, because the diagonal of the confusion matrix during the iteration represents the corrected-detection results with all the precedent centers. If the system still uses the diagonal to calculate new principal centers, it would cause big problems of classification. For example in Figure 3, the points of class B are distributed in three different zones. The system could not combine the spheres of B1, B2 and B3 into only one sphere, though they are correctly classified at the 'Iteration 1'. The spheres of B2 and B3 are 'enclaves' of class B. If they are forced to be fused, the points of classes A and C, which are distributed between the three zones of class B, would be misclassified, and the sub-centers found in previous steps would be meaningless. So the rules of research of sub-centers after the pre-classification could be concluded as Formula 4.

\exists Confusion Matrix M_k ,

During k st classification, $k = 2, 3, \dots, n$

Sub-centers $\in M_k, i = j, \forall i, j = 1, 2, 3, \dots, n$ (4)

True centers and pseudo-centers

In general case, a center of a class is calculated from the average of the coordinates of the points in the class. In SCM, the centers cannot be considered as the averages. Pseudo-centers might appear during the learning if the system only uses the averages. As illustrated in Figure 3: the real average of all the points of class A is located in the sphere of B1, not in the sphere of class A. This point is called as pseudo-center of class A (PA). If the system considers PA as the center of A, some points of class B would be certainly misclassified into the class A, because they are too close to PA. In fact, this center is only a production of average, not a true point. Thus, after the calculation of the average, the system has to find a true point (A') of class A, which is the nearest one (belonging to class A) to (PA). (A') is a temporal center. The system needs to find (A1) from (A'), and (A1) is located at heartland of this data zone. Normally, data are more densely populated at the heartland, thus the distances between the data here are shorter than the distances between the data around the periphery. Therefore, beginning with the point (A'), the system finds the nearest point (A'') to (A'), then finds the nearest point (A1) to (A''). From the point (A1), the system finds

the nearest point of (A1) is back to (A''), thus the system considers the point (A1) as a true center of class A. At the next iteration, the system would find another true center A2 for the points at left below in Figure 3.

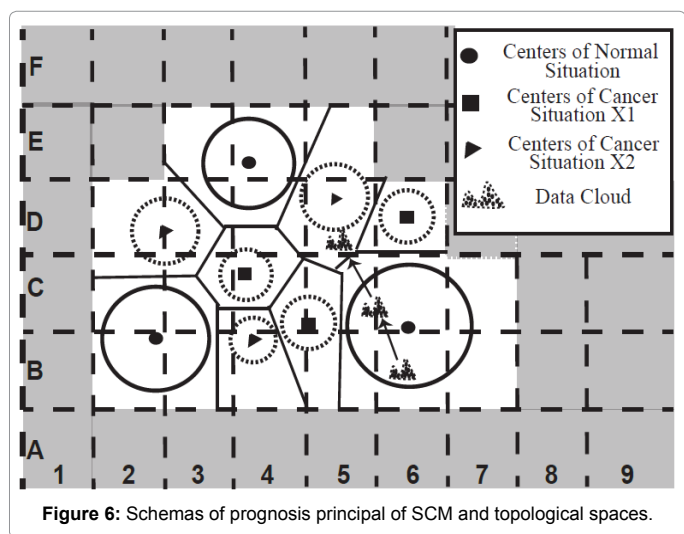
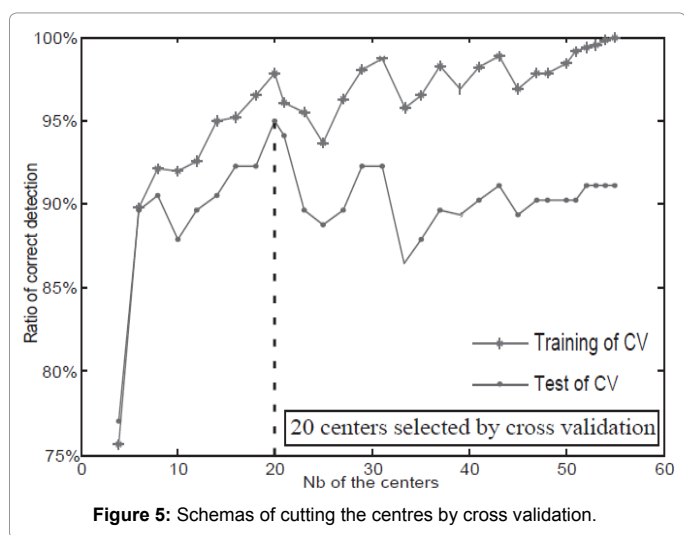
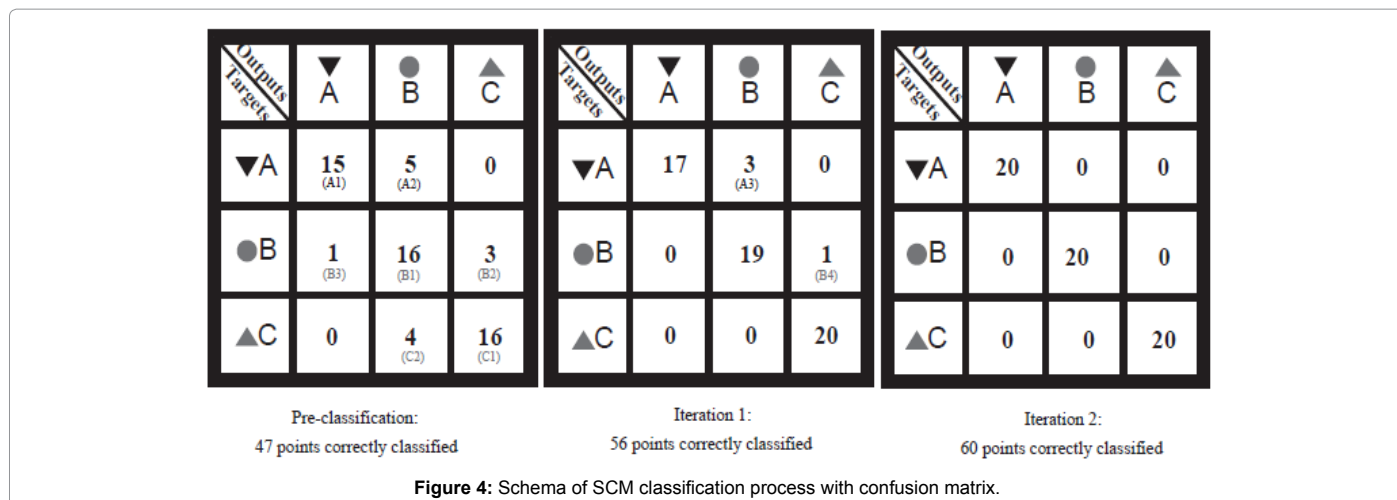
Elimination of redundant centers by cross validation

In fact according to theory, SCM can get 100% corrected detection of the training data. When there is a special point which is isolated and far away from the centers, it could be considered as a center for itself, as the centers B3 and B4 in Figure 3. Thus, redundant centers might appear, and lead to over-fitting: the model fits the training set perfectly, but it can't be generalized at all to new data. It is a phenomenon frequently happening in machine learning, such as neural network and SVM. Whereas it is possible that the system is already over fitting, the method Cross Validation (CV) is needed to avoid this problem [25]. As shown in Figure 5, the corrected detection depends on the amount of centers. According to theory, more centers bring more correct classification. Although 100% corrected detection (CD) could be gained with the training data at the end of learning, it cannot assure a high CD of the test-data. Normally the CD of the test-data would decrease after a threshold because of the over-fitting. Figure 5 shows an example of SCM cross validation on Wisconsin Breast Cancer Dataset. SCM seeks out a part of data as training samples and another part as test data. These training samples are randomly partitioned into k sub-samples of equal size (k is 3 here). Then, a single sub-sample is retained as the validation data for testing the model, and the remaining $k - 1$ sub-sample are used as training data. The cross-validation process is then repeated k times, with each of the k sub-samples used exactly as the validation data. The k resulting from the folds can then be averaged to produce a single estimation. Finally, the system finds 55 centers in total. According to the different number of centers, the system tests the centers with the training samples and the test samples. It is clear that the CD on training data rises to 100% with all the centers. But the CD on test data decreases from the 21st center. It means the over fitting appeared. Thus, the system selects 20 centers rather than 55 centers at last to form the model. The rest 35 centers are considered as redundant centers and eliminated.

Prognosis system and visualization of the dysfunctional models map

In a general way, the majority of prognosis systems are based on regression methods, like stepwise logistic regression [26], Kalman filter [27], Regression Linear [28], Neural Network [29], Support Vector Regression (SVR) [30]. But sometimes these regression tools are not very suitable for the forecasting case. At first, in the prognosis case, the symptoms are not easily perceived. Normally, more attributes of bio-information collected and analyzed, the prognosis result would be more affirmed and convinced. It is not easy to find the correct linear relation among a mass of attributes. Second, the linear relation does not exist sometimes. The same disease might have different phenotypes, some of them are undetectable. As an example shown in Figure 6, each situation of cancer might have three phenotypes which are presented as the centers and their influence sphere. It is difficult to find any regression relation between the three phenotypes of the same cancer. On contrary, SCM could resolve these problems. On the basis of the result of

SCM training, the system could draw a map of all the dysfunctional models, whether the relation of the data is linear or not. In addition to this, with the help of PCA or LDA, the system can reduce the high dimensionality. With the two or three principal components via PCA or LDA, the system could easily visualize the map in 2D or 3D. It can visually explain the progress of the disease and we can intuitively



understand the prognosis results. As the example shown in Figure 6, there are three situations to study: Normal Situation, Cancer Situation 1 and Cancer Situation 2. After the machine training, each class finds three centers. The data collected from the periodic physical examination

can be drawn at this map, as the cloud of data. Based on the physical examinations of different period, the system can find the evolution of data. As the example shown, if the cloud of data moves to some center of cancer situation, it means this risk situation occurs with higher possibility later. In this case, the prognosis system would send an alarm to the doctor, so as to prevent it. In fact, the system would calculate the mean distance or the minimum distance between the cloud of data and each center, and infer the prognosis conclusion.

In order to obtain results in form of probability, the system could calculate the probability of each situation occurrence from the distances measured from the sample data to each situation center. As shown in Equation 5, there are n situations considered. P_i is the probability of situation i occurrence, the distance from sample data to the center of situation i is d_i . The sum of all the probabilities of all situations occurrences is 1. It should be noted that the probability of occurrence could be calculated by mathematic model, for example Gaussian model. SCM offers the centers and the variances which are indispensable in Gaussian model. But not all the data distribute as the form of Gaussian. It is simple and clear to use distances to calculate the probability of occurrence which is applied in this paper.

$$P_i = 1 - \left(\frac{d_i}{\sum_{i=1}^n (d_i)} \right) \forall i = 1, 2, 3, \dots, n \quad (5)$$

Detection of unknown situation with topological space

The performance of the system largely depends on the reliability of the training data. Nevertheless, there are not all-embracing data. From time to time, new virus or cancer situation appears, new phenotype appears and new disease appears. It is difficult to find these new risks in time, let alone forecast them. Thus, an intelligence system is expected to have the ability to detect the unknown situations. In order to achieve this, topological s -pace is introduced [31]. As the example shown in Figure 6, the whole space is regularly divided into 54 sub-spaces. The highlight areas mean that the training data appeared in these sub-spaces. The gray areas are strange. Certainly, the points which appear in gray areas can be classified by comparison of distance. Nevertheless, if they appear in some area in high frequency, it might be an important signal, which contains some hidden information. It may be a new situation or phenotype of cancer. However, it is meaningful and noteworthy.

$$\text{If } \exists K : d_k > \max(d_i) * \lambda, \text{ then } K \in \{\text{unknown}\} \quad (6)$$

As shown in Equation 6, if the distance (d_k) from sample (K) data to centers of situation known is farther than the maximum of distances (d_i) which were saved during the machine learning, the sample (K) should be considered in an unknown zone. λ is a multiple factor, which could adjust the boundary of zone known. In general, λ is considered as 1. If the machine learning is based on small samples, it might lead to the deficiency of learning. Thus λ could be adjusted as a parameter >1 , which can amplify the boundary of zone known. On contrary, if it might occur over-fitting, then λ could be adjusted as a parameter between (0, 1), which can give up some 'enclaves' of centers, so as to reduce the over-fitting. After the judge of Equation 6, the decision system would make a report to the user, so as to analyze its nature. All the process can be done during the test, therefore the system is semi-supervised. It can update itself during the test without any supervisor.

Material and Application

Wisconsin Breast Cancer Dataset consists of 569 samples in total. For each sample, there are thirty characteristics derived and described as radius, perimeter, area, compactness, smoothness, concavity, concave points, symmetry, fractal dimension and texture [32]. In order to verify the robustness of SCM, the dataset were divided into 9 groups randomly according to different proportion of training data and test data. As shown in Figure 7, from 10% to 90%, different proportion of data were selected for the training in the different groups, the system tests the rest of data. In addition, for each group, the test would be repeated for 50 times. The selection of training data for each test was random. Therefore, there was (50×9) 450 times of test in total. At last, the system would calculate the mean and the variance of each group test. Thus, the performance of SCM would be shown from the small samples of training to relative large samples of training.

In order to confirm the classification capacity of SCM, two different series comparison tests were made as shown in Figure 7. One was to compare SCM with SVM, where SCM uses targets of training data to find the principle centers in the pre-classification; the other was to compare SCM with LDA, where SCM use discriminant analysis to find the principle centers in the pre-classification. As a well-known method, SVM was already used to diagnosis this database and gave a good accuracy [33]. It is necessary to confirm that if SCM could work as well as SVM. Another reason to do this comparison test is that SVM generally combine PCA for the sake of reduction of the high dimension of data, so SCM does. It would be interesting to compare them by using the same pretreatment of data. Thanks to PCA, three principal components were instead of 30 original vectors to do the test. The sum of variance explained about the three principal components was about 72.63%, which is shown in Figure 8a. In addition, it should be noted that the effectiveness of SVM depends on the selection of kernel, the kernel's parameters, and soft margin parameter C. A common choice Gaussian kernel was used in SVM here, which has a single parameter γ in $\kappa(x_i, x_j) = \exp(-\|x_i - x_j\|^2)$. The best combination of C and γ were selected by a grid search with exponentially growing sequences $C \in 2^{-3}, 2^{-2}, \dots, 2^9$ and $\gamma \in 2^{-10}, 2^{-9}, \dots, 2^1, 2^3$.

At last, the best combination $C = 1$ and $\gamma = 0.1$ were taken to be used in the comparison test with SCM. The second series of tests were to compare SCM with LDA. SCM would use LDA as a pretreatment tool. Except for PCA, SCM can also use discriminant analysis to accomplish the data pretreatment. As shown in Figure 9a, 5 principal vectors via LDA were selected for the sake of reducing the dimension of data. It is clearly that the ratio of correct detection arrived at gentle zones from the 5st vector. This is a small test to find how many vectors via LDA are appropriate, neither more nor less. Indeed, the correct detection ratio

could be increased with more vectors, but it would spend more time on calculation. After the step of reduction of data dimension, SCM would aid the LDA to analyze and classify the data. Normally, LDA uses linear regression (LR) methods to classify the data. But not all the data have obvious regression relation; they might be distributed in some isolated zones as the example in Figure 3. Thus, SCM would be used here to find the enclaves. At last, the diagnosis results of the cooperation 'LR+SCM' would be compared with the results of the system 'LR' independent.

Based on the diagnosis model, which was formed during the machine learning, the system would forecast the risk of future for the patients with benign tumors. The probability of risk would be evaluated by the distances from their data to the malignant centers and shown in a table. If the data is far away from the zones known where are determined by the spheres of centers, the prognosis system would consider it as a case unknown and make a report. In addition, the time training would be compared among the different systems.

Results

The comparison between SCM and SVM is shown in Figure 8b. On basis of different proportions of the training data, the mean and variance of the correct detection ratio of the two methods are presented. Firstly, the well-known method - SVM confirmed the availability of the new method - SCM. The means of correct detection ratios by SCM varied from 91.4% to 95.6%, it showed a little better than SVM (86.4% - 92.5%). From the view of the means, SCM had 2% to 3% advantage; from the view of the variance, SCM variance interval (1.9% - 3.1%) was a little smaller than SVM (2.1% - 4.86%), so the performance of SCM was more stable. Secondly, the behavior of the correct detection (CD) ratios reveals the robustness of SCM. Even if there were only 10% data for the training and 90% data for the test, the CD ratios of the tests stayed around 87% in SVM and 91% in SCM. The comparison between 'LDA + SCM' and LDA is shown in Figure 9b. Obviously, the regression method was not quite sensitive to the proportion of training data. The CD ratio stayed around at 90% - 92.5% for every proportion of training data. On contrary, SCM was sensitive to the proportion of training data. The mean of CD ratios was increased with more training data, from 92.4% (with the data set which contained 10% training data) to 96.93% (with the data set which contained 90% training data). Consequently, SCM made the CD ratios to increase 1% to 5%. It demonstrated that SCM worked well as a complement system for LDA to improve the diagnosis results. In addition, the variances of 'SCM+LDA' system (1.9% - 3.2%) were smaller than LDA system (2.7% - 4.9%). In another word, the LDA system became more stable with SCM.

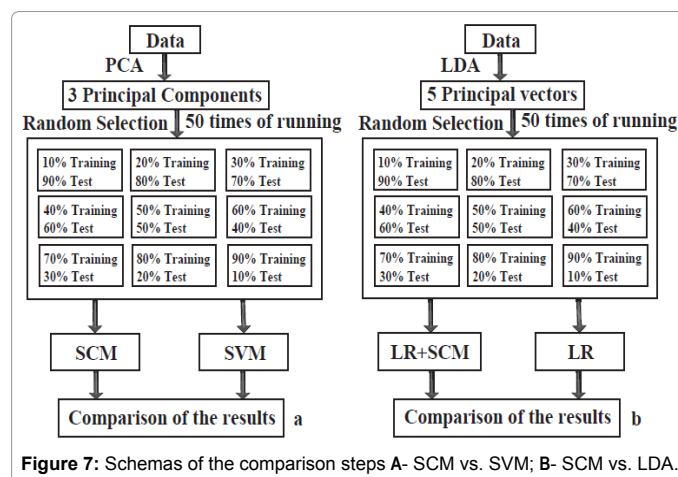


Figure 7: Schemas of the comparison steps A- SCM vs. SVM; B- SCM vs. LDA.

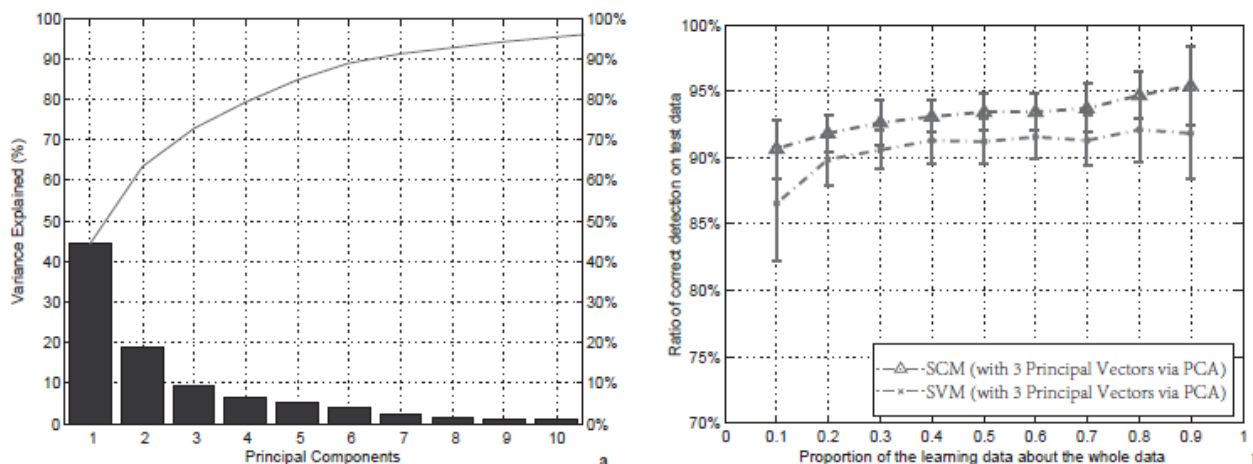


Figure 8: A: Principal component analysis on database. B: SCM vs SVM on correction detection ratio according to the proportion of training data with 3 principal vectors via PCA.

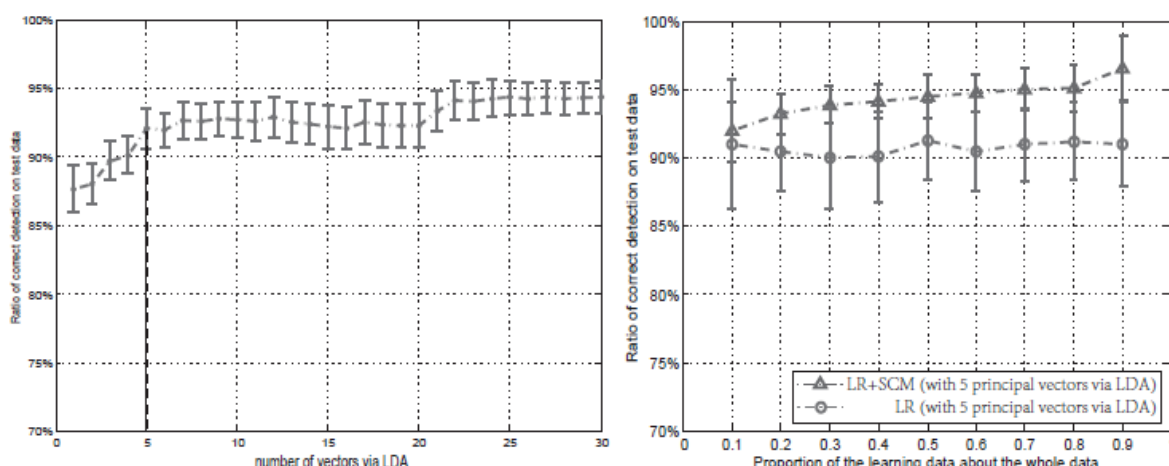


Figure 9: A: Ratio of correct detection according to number of vectors via LDA. B: SCM vs. LDA on correction detection ratio according to the proportion of training data (with 5 first vectors via LDA).

Some examples of prognosis results are shown in Table 1. Although the tumors in 357 patients are benign, the tumors have the risk to become malignant. Thus the patients need to be supervised and taken care of carefully. SCM could take out results of distances to Benign and to Malignant, which could be used to forecast the probability of malignant risk. In Table 1, the first column is the id of patients which are benign in the targets and in the diagnosis results. There are 11 examples of patients presented in the table. The second column is the distance between the samples data and the center of benign, and the third column is the distance between the samples data and the center of malignant. Obviously, the samples data are close to Benign than to Malignant, but there are some signals of dangerous. The fourth column presents the probabilities of cancer occurrence, which was calculated by Equation 5 explicated in Section 2. For example, the patients ID 8510426 and ID 861853. They had more than 40% probability to be transformed to malignant according to the prognosis results. It is important to note that the probability here is relative. For example, although the distance to malignant of patient ID 8510653 (distance = 2.6155) was smaller than ID 8510426 (distance = 3.1994), patient (ID 8510653) had no risk of malignant. Because his data was much more close to benign, even at the center of benign class.

Obviously, when the patient data are were from the centers, the uncertainty becomes more important. Thus the risk was raised. During the machine learning, the maximum of distances saved was 5.6934 to benign center and 6.1816 to malignant. As shown in Table 2, there were three patients whose data was too far to the centers. As a result, the situations of the three patients were considered as unknown, the system SCM sent out a report to warn the users (clinician). These calculated results are valuable to aid the clinicians to monitor the patient's condition and discover unknown condition. The probabilities of risk are quantitative indicators. They are easy to be understood by clinicians who are not data miner. The final decision or diagnosis report is made by clinicians. With a normal computer (Inter(R) Core(TM) i3 CPU 2.27GHz and RAM 4.0G), SVM needed 2160 seconds to complete the whole training (540 times running - 50 times random selections of data \times 9 data sets of different proportions of training data) and LDA independent system cost 966 seconds. But SCM needed 6640 seconds to accomplish the training.

Discussion

SCM is a flexible method. It can be either combined with PCA or LDA, which can reduce the dimension of data and make a visualization

ID	D to Benign	D to Malignant	Risk
8510426	2.5428	3.1994	44.28%
8510653	0.0000	2.6155	0.00%
857373	1.6665	2.6661	38.46%
8610908	0.9878	2.3273	29.80%
8612080	1.9527	3.0430	39.09%
861597	0.0000	2.1852	0.00%
861598	2.3861	3.6812	39.33%
861648	0.0000	2.2334	0.00%
861853	2.0836	2.9106	41.72%
862009	1.2492	2.2688	35.51%

D: Distance.

Table 1: Risk prognosis of transformation to Malignant.

ID	D to Benign	D to Malignant	Report
8710441	13.1465	17.8973	unknown
89143602	7.9523	9.1428	unknown
894047	8.0938	8.5146	unknown

D: Distance

Table 2: Detection of unknown situation.

of data with the principal vectors. In fact, LDA is not really a tool of data compression. It includes label information of the data and maximally separates the distribution of different classes in the new feature space. Its principle is to maximize the ratio of between class variance to the within class variance in any particular data set thereby guaranteeing maximal separability [34]. It means that SCM could work in the new feature space of LDA to accomplish the classification, where is already maximally separated the data of different classes.

On contrary PCA does not include label information of the data, it makes a good dimensionality reduction. But it cannot consider that it is less useful for classification [35]. It remains the main characteristics in a low-dimensions feature space. SVM and SCM accomplished the classification on this feature space and the best precision given was about 95%, in which case we used 50% of data for machine learning and test the others.

According to the diagnosis results with the comparison with SVM and LDA, SCM was confirmed by the test of Wisconsin breast cancer data. As presented in [36], SCM was also confirmed in the test of fault diagnosis and prognosis for aircraft, which is a huge dataset included more than millions flight records. Thus, SCM is a practical method to diagnosis and prognosis different types and different volume of data.

In the prognosis results, although the probabilities of malignant risk in future for the patients, who actually have benign tumor, are taken out clearly, there isn't any record of these patients in the database to confirm the prognosis results. But at least, the prognosis results can propose the doctors to pay attention to the patients who have a great probability of malignant risk. The report of distances to malignant and to benign clearly shows the breast healthy map of each patient. Doctors can monitor the distances to evaluate the situation of patient.

At last, SCM costs much more time on machine learning. As a novel method, it could not roundly surpass the classic method which is used for several decades. However, it needs time to develop and improve. On contrary, one advantage of SCM must be pointed here. If there are some new data, for example new phenomenon of the disease appears, SCM does not need to relearn the total database. It could be trained only by the new data and add the new centers in the map. Thus, although it costs more time on the first machine learning, it saves much

more time if the system needs to add new rules in order to deal with new cases. As in the tests, the system used only 10% of data for machine learning and built the model at the first step. If the system needs to learn 10% data to improve itself, it doesn't need to relearn the previous 10% data. It can find the new centers from new data, and fuse them into old centers if they are too close.

Conclusion and Perspectives

The aim of this paper is to introduce a novel diagnosis and prognosis method SCM, which can be used for both complex system, and medical system. In the last section, SCM has shown the robust and advantages from the comparisons with SVM and LDA. SCM can work independently as well as a complement system for other classification systems. The diagnosis results of the tests on the data set "Wisconsin Breast Cancer Data" have shown a good performance of SCM in the comparisons with SVM and LDA. In fact, SCM can not only make a diagnosis, but also a prediction of disease with a healthy map, which was made during the machine learning. Based on this map, the risk probability of malignant in future is presented in the prognosis report. In addition, this map is possible to be visualized on 2D or 3D, which is intuitive and useful to study the disease progresses.

As a novel artificial intelligent method, SCM has many rooms for improvement and it is worthy of further investigation. At first, Kernel Methods (KMs) can be combined in the system. KMs approach the problem by mapping the data into a high dimensional feature space, where each coordinate corresponds to one feature of the data items [37], transforming the data into a set of points in a Euclidean space [38]. In this new feature, the data might be classified more easily. In SCM, KMs can be used in two places: one is in LDA, which can combine KMs and manipulate LDA in a new feature space; the other is about calculation of the distances between the centers and data in algorithm of SCM. Secondly, the pre-classification can be done with other methods, like classification tree [39] or neural networks. Finally, there are many topological spaces forms to choose. On basis of the different database, a suited topological form could be used to improve the utilization of space.

References

- Musen MA, Shahar Y, MD EHS (2006) Clinical Decision-Support Systems. Biomedical Informatics: 698-736.
- Lodwick GS (1965) A probabilistic approach to the diagnosis of bone tumors. Radiol Clin North Am 3: 487-497.
- Wang K, Weng Z, Sun L, Sun J, Zhou SF, et al. (2015) Systematic drug safety evaluation based on public genomic expression (connectivity map) data: Myocardial and infectious adverse reactions as application cases. Biochem Biophys Res Commun 457: 249-255.
- Kamruzzaman SM, Hasan AR, Siddiquee AB, Mazumder MEH (2004) Mediacal diagnosis using neural network. 3rd International Conference on Electrical and Computer Engineering 28-30.
- Stensmo M, Sejnowski TJ (1996) Automated medical diagnosis based on decision theory and learning from cases. Wnrlid Canwss on Neural Networks: 1227-1231.
- Horrocks JC, McCann AP, Staniland JR, Leaper DJ, De Dombal FT (1972) Computer-aided diagnosis: description of an adaptable system, and operational experience with 2,034 cases. Br Med J 2: 5-9.
- de Keizer NF, Bonsel GJ, Goldfad C, Rowan KM (2000) The added value that increasing levels of diagnostic information provide in prognostic models to estimate hospital mortality for adult intensive care patients. Intensive Care Med 26: 577-584.
- Wang Z, Zarader J, Argentieirie S (2012) A novel aircraft fault diagnosis and prognosis system based on support centre machines. Condition Monitoring and Diagnosis (CMD): 723-728.

9. Lin Z, Hao Z, Yang X (2008) Evolutionary support centre machine. *IEEE World Congress on Computational Intelligence*. *IEEE Congress on:* 1901-1906.
10. Agresti A (1996) *An Introduction to Categorical Data Analysis*, John Wiley and Sons.
11. Kumar K, Abhishek (2012) Artificial neural networks for diagnosis of kidney stones disease. *IJ Information Technology and Computer Science* 7: 20-25.
12. Abdelhalim A, Traore I (2009) A new method for learning decision trees from rules. *Machine Learning and Applications, ICMLA '09. International Conference on:* 693-698.
13. Orrù G, Petterson-Yeo W, Marquand AF, Sartori G, Mechelli A (2012) Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review. *Neurosci Biobehav Rev* 36: 1140-1152.
14. Zurac S, Negroiu G, Petrescu S, Andrei R, Tebeica T, et al. (2012) Spectrum of morphologic alterations of regression in cutaneous melanoma - potential for improving disease prognosis. *Rom J Intern Med* 50: 145-153.
15. K. Bache, M Lichman, Uci machine learning repository, Report, University of California, Irvine, School of Information and Computer Sciences (2013). URL: <http://archive.ics.uci.edu/ml>
16. WH Wolberg, WN Street, OL Mangasarian, Wisconsin diagnostic breast cancer (wdbc) (1995). URL: <http://www.cs.wisc.edu/olvi/uwmp/cancer.html>
17. B Giordano, De l'infinito universo e mondi (on the infinite universe and worlds).
18. Lawrence G, Boyd I, McIntyre P, Isaacs D (2004) Surveillance of adverse events following immunisation: Australia 2002 to 2003. *Commun Dis Intell Q Rep* 28: 324-338.
19. Larsen RJ, Marx ML (2000) *An Introduction to Mathematical Statistics and Its Applications*.
20. Fisher RA (1936) The use of multiple measurements in taxonomic problems. *Annals of Eugenics* 7: 179-188.
21. Fukunaga K (1990) *Introduction to statistical pattern recognition*, Academic Press Professional, Inc., San Diego, CA, USA.
22. Jolliffe IT (2002) *Principal Component Analysis*, Pringer-Verlag, New York, 2nd edition.
23. Queen JBM (1967) Some methods for classification and analysis of multivariate observations, In: *Proceeding of 5th Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press: 281-297.
24. Honarkhah M, Caers J (2010) Stochastic simulation of patterns using distance-based pattern modeling. *Mathematical Geosciences* 42: 487-517.
25. Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *International joint Conference on artificial intelligence* 2: 1137-1145.
26. Wolberg WH, Street WN, Heisey DM, Mangasarian OL (1995) Computer-derived nuclear features distinguish malignant from benign breast cytology. *Hum Pathol* 26: 792-796.
27. Kalman RE (1960) A new approach to linear filtering and prediction problems. *Transactions of the ASME Journal of Basic Engineering* 82: 33-45.
28. Tofallis C (2009) Least squares percentage regression. *Journal of Modern Applied Statistical Methods*.
29. Dunis CL, Jalilov J (2001) Neural network regression and alternative forecasting techniques for predicting financial variables. *Neural network world* 12: 113-140.
30. Drucker H, Burges CJC, Kaufman L, Smola A, Vapnik V (1997) Support vector regression machines, In: *Advances in Neural Information Processing Systems* 9, MIT Press: 155-161.
31. Rudolph L, Han L, Charles E (2009) Modelling emotional development via finite topological spaces and stratified manifolds, Report.
32. Mu T, Nandi AK (2008) Breast cancer diagnosis from fine-needle aspiration using supervised compact hyper spheres and establishment of confidence of malignancy. *Signal Processing Conference:* 1-5.
33. W Duch, Datasets used for classification: Comparison of results, Report, Computational Intelligence Laboratory, Department of Informatics, Nicolaus Copernicus University (2010).
34. Martinez MA, KAC Pca versus Ida (2001) PCA versus LDA. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 23: 228-233.
35. Lopez M, Ramirez J, Gorriiz JM, Salas-Gonzalez D, Alvarez I, et al. (2009) Automatic tool for Alzheimer's disease diagnosis using PCA and bayesian classification rules. *Electronics Letters* 45: 389-391.
36. Wang Z (2013) Fault diagnosis and prognosis system for aircraft, Thesis, Soutenue en.
37. Afif MH, Hedar AR, Hamid THA, Mahdy YB (2013) Ss-svm (3svm): A new classification method for hepatitis disease diagnosis. *IJACSA* 4: 53-58.
38. Song K, Zhang Z, Tong TP, Wu F (2012) Classifier assessment and feature selection for recognizing short coding sequences of human genes. *J Comput Biol* 19: 251-260.
39. Wang Z, Zarader JL, Argentieri S, Youssef K (2013) A decision system for aircraft faults diagnosis based on classification trees and PCA. *Advances in Intelligent Systems and Computing* 193: 411-422.