AN INTELLIGENT MODEL FOR DETECTION OF POST-OPERATIVE INFECTIONS

Mohamed El-Rashidy [1], Taha Taha [2], Nabil Ayad [3], and Hoda Sroor [4]

[1] Dept. of Computer Science & Eng., Faculty of Electronic Engineering, Menoufiya University, Menouf, Egypt
malrashidy@yahoo.com

[2] Dept. of Electronics& Electrical Communications, Faculty of Electronic Engineering, Menoufiya University, Menouf, Egypt
taha117@hotmail.com

[3] Nuclear Research Center, Atomic Energy Authority, Cairo, Egypt.
n_ayad51@yahoo.com

[4] Dept. of Computer Science & Eng., Faculty of Electronic Engineering, Menoufiya University, Menouf, Egypt
dr.hoda_sroor@yahoo.com

ABSTRACT

An effective intelligent diagnosis model is aiming to provide a comprehensive ANALYSIS to form optimal partitioning REPRESENTATION of patient data, and extracts the most significant features for each partition which raise the accuracy of diagnosis process. Optimal Clustering for support feature machine (OCSFM) is proposed to improve the feature selection in medical data classification comprises clustering, feature selection, and classification concepts which is based on fuzzy C-means, max-min, and support feature machine (SFM) models. Experiments have been conducted on database of surgical patients to detect post-operative infections. The performance of the method is evaluated using classification sensitivity, specificity, overall accuracy, and Matthew's correlation coefficient. The results show that the highest classification performance is obtained for the OCSFM model, and this is very promising compared to NaïveBayes, Linear Support Vector Machine (Linear SVM), Polykernal SVM, artificial neural network (ANN), and SFM models.

KEYWORDS

Clustering, Feature Selection, Classification, SFM Model, Infection.

1. INTRODUCTION

Many patients are admitted to hospital for surgery. During the intervention, they did not show any problem. However, the patients were discomfort from infections that they do not suffer from previously. Infection occurs for various reasons; the most frequent one is that of micro organisms which are found naturally in the skin of the patient when the wound is done during the surgical procedure; however it is minimal to cause inflammation. At present the phenomenon of primary bacteremia occurs when bacteria are presented in blood emerging the leading causes of nosocomial infections related to intravenous therapy. National institute of medical sciences and nutrition, Salvador has recorded between 10 to 15 million nosocomial infections, which are related to about one hundred thousand people of deaths each year. Some people may have infection but they have few or no symptoms, as a result a serious infection has been developed as

contraceptive for woman or death. So, the early detection of infection and finding out its reason is important to increase the chance of successful treatment. Analyzing data for clinical decision

support is a task that has a great importance to help saving time of both patients and doctors and to minimize the risk of making wrong diagnoses.

DOI : 10.5121/ijaia.2011.2404

Machine learning and data mining techniques have been successfully applied to various biomedical domains, for example the detection of tumors, the diagnosis and prognosis of cancers , liver diseases, diabetes, heart disease and other complex diseases [1-4]. One of the core issues in biomedical is data analysis and mining. The goal of predictive data mining in clinical medicine is to derive models that can use patient specific information to predict the outcome of interest and to support clinical decision making. Classification and clustering are an important data mining tasks which are widely used in numerous real world applications. Classification aims at exploring through data objects to find a set of rules which determine the class of each object according to its attributes. These rules are later used to build a classifier to predict the class or missing attribute value of unseen objects whose class might not be known. Clustering is a process aiming at grouping a set of objects into classes according to the characteristics of data so that objects within a class can have high mutual similarity while objects in different classes are dissimilar. SFM gives very good classification results, uses far fewer features to make the decision than support vector machine (SVM) classification and Logical Data Analysis (LAD) [5]. SFM model ignores the variations (pathological types) of diseases, and it uses a union dataset of two disjoint sets, one represents the positive group (patients do not have disease) and the other represents the negative group (patients have disease). Negative and positive groups may have a various reasons to cause them. In other words, diseases may have number of different pathological types.

In this paper we present a new hybrid approach based on SFM that employs advances in classification disease and fuzzy clustering that has less sensitive to class's noisy data. We call this hybrid approach an optimal clustering for Support Feature Machine (OCSFM). The goal of OCSFM is to classify the disease into optimal classes that have a various effected features of disease symptoms. The advantage of OCSFM, it has classes with less sensitive to noise since noise data points will have very low degrees in all classes and selection of features is based on features that have high classifiability. We evaluated the performance of OCSFM on database warehoused from more than one server contained data of surgical patients. This database can be used to determine post-operative infections, and to find out organism names that cause these organisms.

In section 2, related works will be explained. In section 3, classification criteria's will be described. In section 4, each step of our proposed OCSFM will be detailed. In section 5, the results and the performance characteristics of the proposed approach will be discussed. The concluding remarks will be offered in section 6.

2. RELATED WORKS

OCSFM model is an integration of both characteristics of supervised and unsupervised models and is based on clustering, feature selection, and classification concepts, which uses fuzzy clustering, max-min, and SFM models that employ advances in classification of medical data.

2.1. Fuzzy Clustering

Clustering is a process that brings a set of objects together into classes according to data characteristics, so that objects within a class can have high mutual similarity, while objects in different classes are dissimilar. Existing clustering models could be classified into three subcategories hierarchical, density based, and partition-based approaches. Hierarchical algorithms organize objects into a hierarchy of nested clusters; hierarchical clustering can be further divided into agglomerative (bottom-up) and divisive (top-down) methods [6-9]. Density based algorithms describe the density of a data which are set by the density of its objects; the clustering involves the search for dense areas in the object space [10-12].The core idea of Partition based algorithms is to partition data directly into disjoint classes. This subcategory includes several algorithms as k-means, fuzzy c-means, P3M, SOM, graph theoretical approaches, and model based approaches [9] and [13-18]. These approaches assume a predefined number of clusters. In addition, these approaches (except the fuzzy/possibilistic ones) always make brute force decisions on the class borders and thus, it may be easily biased by noisy data. This fact makes these fuzzy/possibilistic approaches less sensitive to noisy data. Fuzzy c-means algorithm (FCM) is an iterative partitioning method [19]. It partitions data samples into c fuzzy classes, where each sample x_j belongs to a class k with a degree of believe which is specified by a membership value u_{kj} between zero and one such that the generalized least squared error function J is minimized.

$$J = \sum_{j=1}^{n} \sum_{k=1}^{c} (u_{kj})^m d(x_j, y_k) \tag{1}$$

Where *m* is a parameter of fuzziness, c is the number of classes, y_k is the center of class *k*, and $d(x_j, y_k)$ expresses the similarity between the sample x_j and the center y_k. The summation of the membership values for each sample is equal to one, and

$$0 < \sum_{k=1}^{c} u_{kj} \quad \text{and} \quad \sum_{k=1}^{c} u_{kj} = 1 \qquad \forall j = 1,....,n \tag{2}$$

this guarantees that no class is empty. This approach is called probabilistic clustering, since that the membership degrees for a given data point formally resemble the probabilities of being a member of the corresponding class. This makes the possibilistic clustering less sensitive to noise since noise data points will have very low degrees in all classes. The minimizations of J are resulted in the following membership function and class center.

$$u_{kj} = \frac{1}{\sum_{i=1}^{c} \left(\frac{d(x_j, y_k)}{d(x_j, y_i)} \right)^{\frac{2}{m-1}}} \tag{3}$$

u_{kj} is a possibility degree that measures how much typical is data point x_j to class *k*. The membership degree of x_j to a cluster not only depends on the distance between x_j and that class, but also the distances between x_j and the other classes. The partitioning property of a probabilistic clustering algorithm, which distributes the weight of x_j on the different classes, is due to this equation. Although it is often desirable, the relative character of the membership degrees in a probabilistic clustering approach can lead to counterintuitive results.

$$y_k = \frac{\sum_{j=1}^{n} (u_{kj})^m x_j}{\sum_{j=1}^{n} (u_{kj})^m} \tag{4}$$

This choice makes y_k proportional to the average intra class distance of *k*, and is related to the overall size and shape of the class.

2.2. Max-Min Median Initialized Fuzzy C-means

Fuzzy c-means algorithms (FCM) are sensitive to the initial center choices, especially for noisy data, since the classes are separated groups in a feature space. The basic max-min approach was first proposed by Tou and Gonzalez in Ref. [20]. It is desirable to select the initial centers which are well separated. The next sample is chosen such that the minimum distance between it and all previously selected samples is a maximum. Specifically, to select the third sample, the first two elements of rows r and w are compared, the minimum is chosen; the minimum of the second elements in rows r and w is chosen, etc. This gives the minimum distance between sample 1 and samples r and w, between sample 2 and samples r and w, etc. A new n element vector (for all n samples) of minima results; the maximum of this vector determines the third sample to retain (max min distance). This procedure is repeated until c samples are selected. Each selected sample is then replaced by the median point of its p nearest neighbors.

2.3. Support Feature Machine Algorithm

The selection feature of voting scheme based on one matrix A = (aij) is an n × m, i = 1,…, n, j = 1,…, m, where n is the number of samples and m is the number of features. The classification is correct when the average distances from sample i to all other samples in the same class at feature j (intra class distance) is smaller than the average distances to all samples in different class at the same feature (inter class distance). Therefore, the entry aij = 1 indicates that the nearest neighbor rule is correctly classified sample i at feature j, 0 otherwise. The best subset of features is selected, which gives the majority correct votes (value 1's) that have the maximum number of correct classified samples [5].

2.3.1. Voting Scheme (V-SFM)

The selection feature of voting scheme based on one matrix A = (aij) is an n × m, i = 1,…, n, j = 1,…, m, where n is the number of samples and m is the number of features. The classification is correct when the average distances from sample i to all other samples in the same class at feature j (intra class distance) is smaller than the average distances to all samples in different class at the same feature (inter class distance). Therefore, the entry aij = 1 indicates that the nearest neighbor rule is correctly classified sample i at feature j, 0 otherwise. The best subset of features is selected, which gives the majority correct votes (value 1's) that have the maximum number of correct classified samples [5].

2.3.2. Averaging Scheme (A-SFM)

The selection feature of averaging scheme is based on two matrices. The first is an n×m intra class distance matrix D = (dij), and the other is an n×m inter class distance matrix $\overline{D} = (\overline{d}ij)$. The entry of the intra class matrix dij is the intra class distance, the entry of the inter class matrix $\overline{d}ij$ is the inter class distance. After the two matrix are constructed, the selection of features is derived from the sum of intra class average distances (dij) are smaller than the sum of inter class average distances ($\overline{d}ij$) in the selected features [5].

3. Classification Criterion

The performance of data classification is commonly presented in terms of sensitivity and specificity. Sensitivity measures the fraction of positive test samples that are correctly classified as positive, then we define

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{5}$$

where TP and FN denote the number of true positives and false negatives, respectively. Specificity measures the fraction of negative test samples that are correctly classified as negative. Let FP and TN denotes the number of false positives and true negatives, respectively, then we define

$$\text{Specificity} = \frac{TN}{TN + FP} \qquad (6)$$

An overall accuracy is defined as

$$\text{Accuracy} = \frac{TN + TP}{TP + FP + TN + FN} \qquad (7)$$

The Matthew's correlation coefficient (MCC) is a powerful accuracy evaluation criterion of machine learning methods. Especially, when the number of negative samples and positive samples are obviously unbalanced; MCC gives a better evaluation than overall accuracy with a lot of machine learning methods, such as SVM, ANN and BNN [1]. MCC should be used as an additional evaluation criterion.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \qquad (8)$$

4. PROPOSED MODEL

We propose a new hybrid algorithm based on fuzzy C-means, max-min, and support feature machine approaches is called OCSFM. The goal of OCSFM is to classify the data points into optimal number of representative classes that have smallest average distance (intra class distance), and greatest average distance to all different class (inter class distance). This makes OCSFM has classes less sensitive to noise since lowest noise degree data point's in all classes, and maximize classification accuracy. The flowchart of our proposed algorithm is shown in Figure 1.

Our model is composed of six main steps. In the first step (Clustering), cluster data points in order to form optimal partitioning representative of classes with smallest intra class distance and greatest inter class distance, considering in the clustering process the maximization of classification accuracy. In the second step (Selected Features), find the optimal subset of features in order to have the maximum number of samples correctly classified into those partitioning classes. In the third step (Classification) training samples are classified according to those selected features, and compute the performance of data classification which is presented in terms of TP, TN, FP, and FN to obtain MCC.

In the fourth step (Classes representatives points), we use max-min method to select classes representatives, it chooses a median of one class from those partitioning classes as a start point to select another classes representatives points as separate as possible from start point. In the fifth step (Multi step max-min algorithm), find an optimal representative partitioning for a fixed number of classes, each iteration of the optimization process is based on clustering, selected features, and classification is obtained by the max-min method but it changes start point with another class median. Iteration is stopped when each of classes medians is selected as a start point, therefore number of iteration for multi step max-min algorithm are equal to classes medians (number of classes). In the sixth step (Optimal classes number), compute an optimal classes number of partitioning classes by highest classification accuracy. For this, multi step max-min algorithm is repeated with increasing the number of partitioning classes from $c_{\min}$ to $c_{\max}$, using MCC as a validity measure in Equation (8), the optimal number that is produced the highest MCC.

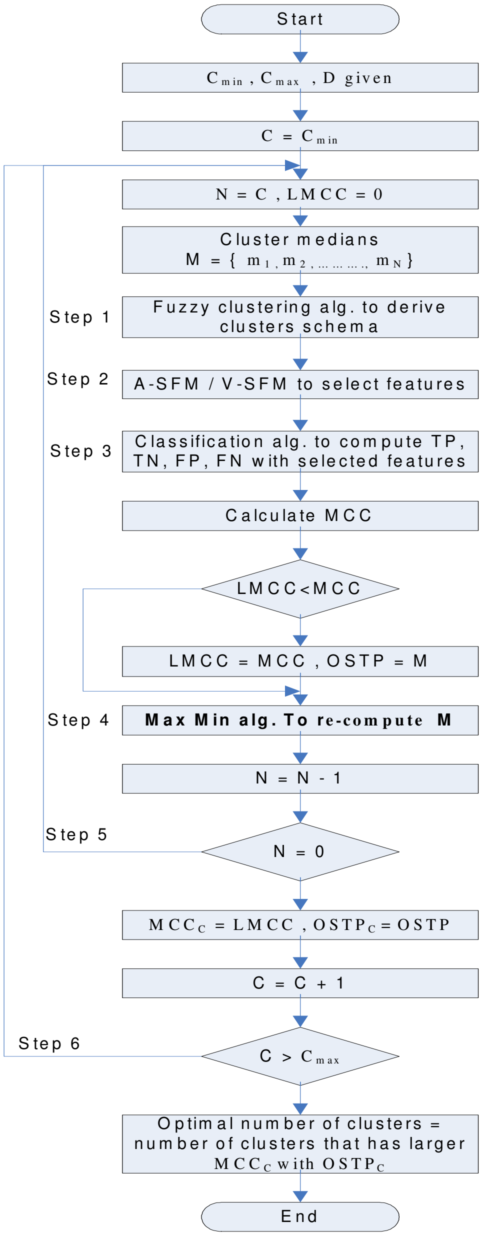
Start
C_{min} , C_{max} , D given
C = C_{min}
N = C , LMCC = 0
Cluster medians
M = { m_1 , m_2 , ……… , m_N }
Step 1
Fuzzy clustering alg. to derive clusters schema
Step 2
A-SFM / V-SFM to select features
Step 3
Classification alg. to compute TP, TN, FP, FN with selected features
Calculate MCC
LMCC<MCC
LMCC = MCC , OSTP = M
Step 4
Max Min alg. To re-compute M
N = N - 1
Step 5
N = 0
MCC_C = LMCC , $OSTP_C$ = OSTP
C = C + 1
Step 6
C > C_{max}
Optimal number of clusters = number of clusters that has larger MCC_C with $OSTP_C$
End

Figure 1. Flowchart of our proposed algorithm

4.1. OCSFM Algorithm

In our proposed approach (OCSFM) in algorithm 1 for a given number of classes $c\left(c_{\min} \le c \le c_{\max}\right)$, $c_{\min}$ and $c_{\max}$ are user specified parameters, and represents respectively, the lower and upper bounds of the expected optimal number of classes. We find the optimal partitioning c_o using the modified multi step max-min algorithm, and use MCC as a validity measure in Equation(8), the optimal number of classes is producing the highest MCC.

```
Algorithm 1: OCSFM algorithm
Input: Data Set D = { $d_0, d_1, \ldots, d_n$ }, $c_{\min}$ and $c_{\max}$ the minimal and the maximal
numbers of expected clusters, respectively.
Output: $c_o$ An optimal cluster scheme.
begin
Let C ← $c_{\min}$.
while (c <= $c_{\max}$) do
    1. Randomly choose a data point $m$ from D as the starting point.
    2. Perform Max Min algorithm (Algorithm 3) to compute the clusters medians
       M={ $m_1, \ldots, m_c$ }.
    2. Perform Multi step Max Min algorithm (Algorithm 2) to find the optimal
       partitioning for C clusters with  highest MCC.
    3. Let C = C + 1.
end while
Let $c_o$ ← the optimal partitioning with the highest MCC.
return ($c_o$)
end
```

4.2. Multi Step Max-Min Algorithm

The multi step max-min algorithm is used to find an optimal representation of partitioning classes for a fixed number of classes. In the multi step max-min algorithm, each iteration of the process is based on the partition obtained by the max-min method.

```
Algorithm 2: Multi step Max-Min approach with MCC as an optimization
criterion
Input: Data set D = { $d_0, d_1, \ldots, d_n$ }, number of clusters C, starting point $m$.
Output: Optimal partitioning $c_o$ = {C1,…, Cc} for a fixed number of classes.
begin
1. Compute the cluster representative for $c_o$ using Algorithm 4
2. for $i$ ← 2 to c do
    3.1. Recompute Cluster medians with $m_i$ as a start point using Algorithm 3.
    3.2. Recompute the cluster representative for $\overline{c}_o$ using Algorithm 4
    3.3. Perform selected features algorithm (Algorithm 5 for V-SFM or Algorithm 6 for A-
         SFM) to find optimal separability features F ={ $f_1, \ldots, f_n$ }for $c_o$ and $\overline{c}_o$.
    3.4. Perform nearest neighbor classification algorithm (Algorithm 7) to find TP, TN, FP,
         and FN for $c_o$ and $\overline{c}_o$.
    3.5. Compute the MCC of both cluster schemes $c_o$ and $\overline{c}_o$ using Eq. (9).
    3.6. if (MCC($\overline{c}_o$) > MCC($c_o$)) then Let $c_o$ ← $\overline{c}_o$
end for
end
```

```
Algorithm 3: Max-Min approach
Input: Data set D = { $d_0, d_1, \ldots, d_n$ }, number of clusters C, starting point $m$.
Output: Cluster medians M = { $m_1, \ldots, m_c$ }.
begin
1. Let $m_1 \leftarrow m$, and M = { $m_1$ }.
2. for $i \leftarrow 2$ to c do
2.1. for all $\forall d_h \in D \setminus M$ do
    Compute all distances $dis(d_h, m_j) \forall m_j \in M$ Save only the minimum distance in a set DIS
end for
2.2. Compute $m_i$ the data point with the maximum distance value in DIS
2.3. Let $M \leftarrow M \cup \{m_i\}$
end for
return (M)
end
```

The max-min method tries to select class representatives by making classes as separate as possible. The basic max-min approach was first proposed in [20] and is summarized in Algorithm 3. Our approach is summarized in Algorithm 2. Initially, we compute a class scheme using the max-min method starting from the given initial point *m*. Thereafter, a refinement of this scheme is performed using the same max-min method but with the computed object medians as new starting points.

4.3. Classes Representatives

After obtaining the new set of representatives using algorithm 2, in algorithm 4 each data point is assigned to the cluster that has the nearest representative (median). The whole process is repeated until the set of representatives becomes stable, or a maximal number of iterations are reached.

```
Algorithm 4: Classes representatives approach
Input: Data set D = { $d_0, d_1, \ldots, d_n$ }, Clusters medians M = { $m_1, \ldots, m_c$ }.
Output: New Clusters medians $\overline{\mathrm{M}}$ = { $\overline{\mathrm{m}}_1, \ldots, \overline{\mathrm{m}}_\mathrm{c}$ }.
 begin
 flag $\leftarrow$ off.
while flag = off do
1. for each $\mathrm{d}_\mathrm{j} \in D$ do
choose the nearest representative, say $m_i$ group $d_j$ into cluster $c_i$
(whose representative is $m_i$).
2. for i $\leftarrow$ 1 to c do
compute $\overline{m_i}$, the object median for $\overline{c_i}$ as its new representative using Equation (9).
3. Let $\overline{\mathrm{M}}$ = { $\overline{\mathrm{m}}_1, \ldots, \overline{\mathrm{m}}_\mathrm{c}$ }.
4. if (M = $\overline{\mathrm{M}}$) OR (maximal iterations are reached) then flag $\leftarrow$ on.
    else flag $\leftarrow$ off.    end if
end while
end
```

4.4. Selected features

The classification accuracy of every feature can be evaluated by applying A-SFM or V-SFM that gives the accuracy information for all features, those formulated to incorporate all the classification decisions made by each feature and select the best subset of features which maximizes the classification accuracy. In Algorithm 5, the voting scheme with the SFM. In Algorithm 6, the averaging scheme with the SFM.

```
Algorithm 5: Selected features V-SFM approach
Input: Optimal partitioning $c_o$ = {C1,…, Cc} that have j = 1,…, m features, positive
and negative point declaration.
Output: optimal separability features F = { $f_1$,…, $f_n$ }
begin
1. Let POS set is empty.
2. for i $\leftarrow$ 1 to c do
    2.1. if( number of positive data points for $c_i$ > number of negative data points for $c_i$ )
    then POS = POS $\cup$ $c_i$ .
    end if
end for
3. for i $\leftarrow$ 1 to c do
    3.1. if( number of negative data points for $c_i$ > number of positive data points for $c_i$ )
    then begin
        Compute intra class matrix A = (aij) (the entry is 1 when the average distance from data
        point I to all other data points in the same class is smaller than the average distance to all
        data points in POS set for each feature j). end if
end for
4. for i $\leftarrow$ 1 to c do
    4.1. if( number of negative data points for $c_i$ >number of positive data points for $c_i$ )
    then begin
        4.1.1. $f_i$ $\leftarrow$ empty.
        4.1.2. for j $\leftarrow$ 1 to m do
            4.1.2.1 if ( $\sum a_{ij}$ is high e.g. (greater than a half number of $c_i$ data points)) then $f_i$ =
            $f_i$ $\cup$ j. end if
        end for
        4.1.3. F=F$\cup$ $f_i$ .
    end if
end for
return (F)
end
```

```
Algorithm 6: Selected features A-SFM approach
Input: Optimal partitioning $c_o$ = {C1,…, Cc} that have j = 1,…, m features, positive
and negative point declaration.
Output: optimal separability features F = { $f_1$,…, $f_n$ }
begin
1. Let POS set is empty.
2. for i $\leftarrow$ 1 to c do
    2.1. if( number of positive data points for $c_i$ > number of negative data points for $c_i$ )
```

```
then POS = POS $\cup c_i$ . end if
end for
3. for i $\leftarrow$ 1 to c do
3.1. if( number of negative data points for $c_i$ > number of positive data points for $c_i$ )
then begin
3.1.1. Compute intra class matrix D = (dij) (average distance between each data point i
in $c_i$ and all other data points in the same class for each feature j).
3.1.2 .Compute inter class matrix $\overline{D} = (\overline{d}ij)$ (average distance between each data point i
in $c_i$ and all data points in POS set for each feature j). end if
end for
4. for i $\leftarrow$ 1 to c do
4.1. if( number of negative data points for $c_i$ >number of positive data points for $c_i$ )
then begin
4.1.1. $f_i \leftarrow$ empty.
4.1.2. for j $\leftarrow$ 1 to m do
4.1.2.1 if ( $\sum d_{ij} < \sum \overline{d}_{ij}$ ) then $f_i = f_i \cup$ j. end if
end for
4.1.3 F=F $\cup f_i$ .
end if
end for
return(F)
end
```

4.5. Classification

After obtaining the optimally selected features, in algorithm 7 dataset is classified according to those selected features F. V-SFM classifies an unlabeled class sample to the class with majority voting from all selected whose baseline training samples are more similar to the sample based on the dimension of selected features. After each data point is labeled by SFM schemes, accuracies of SFM schemes can be calculated by comparing the labeled class with the actual class of each sample [5].

```
Algorithm 7: Nearest neighbor classification approach
Input: Data set D = { $d_0, d_1, \ldots, d_n$ }, Optimal partitioning $c_o$ = {C1,..., Cc}, optimal
separability features F = { $f_1, \ldots, f_n$ }.
Output: TP, TN, FP, and FN the true positive, true negative, false positive, and false negative,
respectively.
begin
1. for each $\mathrm{d}_j \in D$ do
begin
1.1. choose the nearest neighbor according to features F for each class, say $c_i$ .
1.2. if( number of negative data points for $c_i$ > number of positive data points for $c_i$ ) then if( $d_j$
is negative data point) then TN=TN+1 else FN=FN+1 end if
    else if( $d_j$ is positive data point) then TP=TP+1 else FP=FP+1 end if
end if
end for
return (TP,TN.FP.FN)
```

5. Experiments and Discussion

In this study, database of surgical patients was used and analyzed. They have been collected from more than one server of Egyptian hospitals. There are 446 records in this database. Each record in the database has 15 features which are believed to be a good indicator for the infections. These features include age, gender, clinical department name, operation name, operation risk index, health degree of patient (from 1 to 5), actual duration for operation, duration ideal for operation, wound class (none, mild, moderate, severe) of inflammation, length of stay sick before and after the operation, the period between first dose of anti biotic and starting operation, patient temperature during the operation, infection index (non-infected, infected), and name of organism that cause infection. In this database, 101 records (patients) have infection and 345 records (patients) have no infection.

All the experiments were implemented and performed on AMD Phenom 9550 Quad Core 2.2 GHz workstation with 4 gigabytes of memory running on Windows Server 2003. All calculations and algorithms were implemented and run on ORACLE 10G. All programs were written by Java language. We divided the data into training and testing phases, in test stage, 5-fold cross validation method was applied and the top performances are tabulated. The classification performance in the training phase was used to identify the best parameter setting and test it into the testing phase. We evaluated the classification performance of OCSFM with NaïveBayes, Linear SVM, Polykernal SVM, ANN, and SFM approaches; we supported those approaches with two different selected features models, A-SFM that has a good performance than V-SFM [5] and genetic model. We used these models to know the best selection feature model and the classification approach that can be used to hit the highest performance. We propose two distinct experiments.

First experiment, NaïveBayes, Linear SVM, Polykernal SVM, ANN, and SFM approaches are used a union dataset of two disjoint sets; one represents the positive group (patients do not have infections) and the other the negative group (patients have infections). In other words training data is divided into two classes one represent the positive data points and the other the negative data points, ANN is built with two hidden layer, and twelve features are selected by A-SFM model. Sensitivity, specificity, overall accuracy and MCC for each of training data and testing data of those approaches with OCSFM approach that are based on the subset of selected features from A-SFM model in the classification data are summarized in table 1.

Table 1. Training and Testing Performance in % sensitivity, specificity, overall accuracy and MCC of NaïveBayes, Linear SVM, Polykernal SVM, ANN, SFM, OCSFM approaches using A-SFM approach to select features.

Classification algorithm	**Training Data**				**Testing Data**			
	Sens.	**Spec.**	**Accu.**	**MCC**	**Sens.**	**Spec.**	**Accu.**	**MCC**
NaïveBayes	92.72	72.13	89.91	60.57	92.11	55.55	85.11	49.90
Linear SVM	99.74	26.22	89.68	46.60	100	11.11	82.97	30.29
Polykernal SVM	99.74	24.59	89.46	44.95	100	11.11	82.97	30.29
ANN	100	39.34	91.70	59.91	100	22.22	85.11	43.32
SFM	90.75	39.62	83.95	30.37	94.74	11.11	78.72	9.41
OCSFM	**95.08**	**71.69**	**91.97**	**65.74**	**94.74**	**66.66**	**89.36**	**64.29**

Second experiment, the classification approaches classified data based on the subset of selected features from genetic model, five features are selected. The performance criteria's of training data and testing data classification are summarized in table 2. In this experiment, Linear SVM, Polykernal SVM, and SFM models can not calculated MCC as TN and FN are equaled zero.

Table 2. Training and Testing Performance in % sensitivity, specificity, overall accuracy and MCC of NaïveBayes, Linear SVM, Polykernal SVM, ANN, SFM, OCSFM approaches using Genetic approach to select features.

Classification algorithm	**Training Data**				**Testing Data**			
	Sens.	**Spec.**	**Accu.**	**MCC**	**Sens.**	**Spec.**	**Accu.**	**MCC**
NaïveBayes	94.81	19.69	83.45	28.69	94.74	22.22	80.85	23.91
Linear SVM	100	00.00	86.96	--	100	00.00	80.85	--
Polykernal SVM	100	00.00	86.96	--	100	00.00	80.85	--
ANN	99.14	17.31	88.47	32.41	100	00.00	80.85	--
SFM	100	00.00	86.96	--	100	00.00	80.85	--
OCSFM	**97.40**	**33.96**	**88.97**	**42.37**	**94.74**	**22.22**	**80.85**	**23.91**

The results are shown in table 1 that our proposed model OCSFM enhances the behavior of classification models by clustering a dataset into optimal partitioning which improved overall accuracy and MCC in both training and testing data by sensitive rates. A-SFM model selected the best effective features in the classification data rather than genetic model, which appeared on the results of Tables 1 and 2. OCSFM hits the highest rate of classification accuracy with A-SFM model. The improvement of classification accuracy is very crucial in the infection identification process; the improvement of sensitivity can reduce the rate of false detection of non-infection. Similarly, the improvement of specificity can reduce the rate of false detection of infection; avoidance of infection complication can increase the chance of successful treatment. The improvement can be appeared clearly on using OCSFM model, and it is considered as the important purpose that can help physicians to better detect the infection, and the name of organism that cause it.

6. CONCLUSIONS

New hybrid model has been proposed and applied to detect post-operative infections. This approach is a combination of clustering, selected features and classification approaches. Results have indicated that our proposed approach can minimize noise data points, smallest intra class distance, greatest inter class distance in all classes, optimal selection of features in order to have the maximum number of samples correctly classified, and highest classification accuracy. Here, after applying our intelligent model, the overall accuracy, and the Matthew's correlation coefficient have been improved comparing with NaïveBayes, Linear SVM, Polykernal SVM, ANN, and SFM approaches using A-SFM and Genetic approaches to select features. Results showed that A-SFM selected features provided good classification performance than genetic approach which showed the significant improvement in both accuracy and robustness, and the optimal selected features can be used as the reference for making decision in hospital and provide the reference for the researchers.

REFERENCES

[1] Cheng H., Juan Sh., Wen J., Yanhui G., and Ling Z., "Automated breast cancer detection and classification using ultrasound images: A survey", Pattern Recognition, 43, 299-317, 2010.

[2] Riccardo B., and Blaz Z., "Predictive data mining in clinical medicine: Current issues and guidelines", international journal of medical informatics, 77, 81-97, 2008.

[3] Rong-Ho Lin, "An intelligent model for liver disease diagnosis", Artificial Intelligence in Medicine, 47, 53-62, 2009.

[4] Yue H., Paul M., Norman B., and Roy H., "Feature selection and classification model construction on type 2 diabetic patient's data", Artificial Intelligence in Medicine, 41, 251-262, 2007.

[5] Ya-Ju F., and Wanpracha A. Ch., "Optimizing feature selection to improve medical diagnosis", Ann Oper-Res, 174, 169-183, 2010.

[6] Eisen M., Spellman P., Brown P., and Botstein D., "Cluster analysis and display of genome wide expression patterns", Natl Acad Sci USA, 95(25), 14863-14868, 1998.

[7] Blatt M., Wiseman S., and Domany E., "Super-paramagnetic clustering of data", Phys Rev Lett, 76, 3251-3254, 1996.

[8] Rose K., "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems", IEEE, 86(11), 2210-2239, 1998.

[9] Herrero J., Valencia A., and Dopazo J., "A hierarchical unsupervised growing neural network for clustering gene expression patterns", Bioinformatics, 17(2), 126-136, 2001.

[10] Jiang D., Tang C., and Zhang A., "Cluster analysis for gene expression data: a survey", IEEE Trans Knowl Data Eng, 16(11), 1370-1386, 2004.

[11] Jiang D., Pei J., and Zhang A., "DHC: a density-based hierarchical clustering method for time series gene expression data", the 3rd IEEE symp on bioinformatics and bioengineering, Maryland, USA, 393-400, 10-12 March 2003.

[12] Hinneburg A., and Keim D., "An efficient approach to clustering in large multimedia database with noise", the 4th int conf on knowledge discovery and data mining, NY, USA, 58–65, 27–31 August 1998.

[13] Au W., Chan K., Wong A., and Wang Y., "Attribute clustering for grouping, selection, and classification of gene expression data", IEEE/ACMTrans Comput Biol Bioinform, 2(2), 83–101, 2005.

[14] Bickel D., "Robust cluster analysis of microarray gene expression data with the number of clusters determined biologically", Bioinformatics, 19(7), 818–824, 2003.

[15] Guthke R., Schmidt-Heck W., Hann D., and Pfaff M., "Gene expression data mining for functional genomics", the European symp on intel techn, Aachen, Germany, 170–177, 2000.

[16] Romdhane L.B., Shili H., and Ayeb B.," Mining microarray gene expression data with unsupervised possibilistic clustering and proximity graphs", Appl Intell, 10.1007, 10489-009, 2009.

[17] Shamir R., and Sharan R., "CLICK: A clustering algorithm for gene expression analysis", the int conf on intelligent systems for molecular biology, CA, USA, 307–316, 19–23 August 2000.

[18] Yeung K., Fraley C., Murua A., Raftery A., and Ruzz W., "Model-based clustering and data transformations for gene expression data", Bioinformatics, 17(10), 977–987, 2001.

[19] Bezdek J., "Pattern Recognition with Fuzzy Objective Function Algorithms", New York: Plenum, 1981.

[20] Tou J., and Gonzalez R., "Pattern recognition principles", Addison-Wesley, Reading, 1974.

Authors

Mohamed A. El-Rashidy obtained his Master degree in computer science and engineering, 2008. Currently, he is working as a Lecturer Assistant in the Dept. of Computer Science and Engineering, Faculty of Electronic Engineering, 32952, Menouf, Menoufiya University -Egypt. Areas of interest of the author include datamining and bioinformatics.

Taha E. Taha was born in Tanta, Egypt, on October 11, 1946. He received the B.Sc. degree (with distinction) in communication engineering from Menoufia University, Egypt, in June 1969, the M.Sc. degree in communication engineering from Helwan University, Egypt, in April 1978, and the Ph.D. degree (very honorable) in electronic engineering from the National Polytechnic Institute, Toulouse, France, in June 1985**.** From September 1969 to July 1978, he was a Demonstrator, in July 1978, he was an Assistant Lecturer, in November 1985, he was a Lecturer, in February 1990, he was an Assistant Professor, and in September 1995, he was named Professor, all in the Faculty of Electronic Engineering, Menoufia University, Communication Department,. He was appointed Vice Dean from February 2002 to October 2005, and Head of the Communication Department, from November 2005 to July 2007. At present, he is an Emeritus Professor at the same department. His main research interests are surface acoustic wave devices, optical devices, superconductor devices, medical applications of ultrasound, and bioinformatics.

Nabil M. A. Ayad received Ph.D degree in CSE from Cairo University, in 1984. He is working as vice chairman for reactors division, Nuclear Research Center, Atomic Energy Authority- Egypt. He is a member of IEEE. His main research interests database and networks.

Hoda S. Sroor received Ph.D degree in CSE from Menoufiya University, in 1991. She is working as Professor in Dept. of Computer Science and Engineering, Faculty of Electronic Engineering, 32952, Menouf, Menoufiya University- Egypt, her main research interests parallel processing and database.