Local Dimensionality Reduction and Supervised Learning within Natural Clusters for Biomedical Data Analysis

Mykola Pechenizkiy, Alexey Tsymbal, and Seppo Puuronen

Abstract—Inductive learning systems have been successfully applied in a number of medical domains. Nevertheless, the effective use of these systems requires data preprocessing before applying a learning algorithm. Especially it is important for multidimensional heterogeneous data presented by a large number of features of different types. Dimensionality reduction is one commonly applied approach. The goal of this paper is to study the impact of human expert produced natural clustering on dimensionality reduction for classification in the nosocomial infections domain. We compare several data mining strategies that apply dimensionality reduction by means of feature extraction or feature selection for subsequent classification. We show experimentally on this microbiological data that local dimensionality reduction within natural clusters results in a better feature space for classification in comparison with the global search in terms of generalization accuracy.

I. INTRODUCTION

CURRENT electronic data repositories, especially in medical domains, contain huge amount of data including also currently unknown and potentially interesting patterns and relations that can be found using knowledge discovery and data mining (DM) methods [5]. Inductive (supervised) learning systems have been successfully applied in a number of medical domains, for example, in localization of a primary tumor, prognostics of recurrence of breast cancer, diagnosis of thyroid diseases, and rheumatology [10].

However, researchers and practitioners realize that the effective use of these learning systems requires preceding data preprocessing. This is especially important for multidimensional heterogeneous data, presented by a large number of features of different types. The so-called “curse of dimensionality” [2] pertinent to many learning algorithms, denotes the drastic raise of computational complexity and classification error on data having large number of features. Hence, the dimensionality of feature space is often reduced before classification is undertaken. Generally, dimensionality reduction (DR) is only one effective approach to data reduction among others like instance selection or data selection [12]. We see the goal of DR to be in: (1) reducing the quantity of data with a focus on relevant data, and (2) improving the quality of data and/or its representation for a supervised learning method. Consequently, achievement of these goals results in a reduced amount of data, relevance of this reduced data to the domain and supervised learning method applied, and finally, improvement of the performance of the learning method.

There are a number of DR techniques, and according to the adopted reduction strategy they are usually divided into feature selection (FS) and feature extraction (FE) (also called feature discovery) approaches [12]. The key difference between FS and FE is that in the former a subset of original features only is selected while the latter is based on generation of a completely new feature space through a functional mapping, keeping in fewer dimensions as much information about the data as possible [12]. Many FS techniques are usually sensitive to interacting or correlated features. That is why transformation of the given representation might be preferable.

For some problem domains a feature subset may be useful in one part of the instance space, and at the same time it may be useless or even misleading in another part of it. Therefore, it may be difficult or even impossible in some datasets to remove irrelevant and/or redundant features and leave only useful ones by means of global FS. However, if it is possible to find local homogeneous regions of heterogeneous data, then there are more chances to apply FS successfully (individually to each region). For FE the decision whether to proceed globally over the entire instance space or locally on different parts of the instance space is also one of the key issues. It can be seen that despite being globally high dimensional and sparse, data distributions in some domain areas are locally low dimensional and dense, for example in physical movement systems [19].

One possible approach for local FS or local FE would be...
clustering (grouping) of the whole data set into smaller regions. Generally, different clustering techniques can be used for this purpose, for example the k-means or EM techniques [20]. However, in this paper we emphasize the possibility to apply so-called natural clustering aimed at using contextual features for splitting whole heterogeneous data space into more homogeneous clusters. Usually, features that are not useful for classification alone but are useful in combination with other (context-sensitive) features are called contextual (or environmental) features [18].

In this paper we apply our natural clustering approach to real clinical data trying to construct local models that would help in the better prediction of antibiotic resistance and in understanding its development.

In our experimental study we apply k-nearest neighbor classification (kNN) to build antibiotic sensitivity prediction models. We apply the principle of natural clustering, grouping the instances into partitions related to certain pathogen types. We apply three different wrapper-based sequential FS techniques and three eigenvector-based FE techniques globally and locally and analyze their impact on the performance of kNN classifier.

The paper is organized as follows. In Section 2 we briefly consider dimensionality reduction techniques used in the study. In Section 3 data used in our experiments and its nature are described. In Section 4 we describe how natural clustering approach was applied to our data. In Section 5 we present the results of experiments with the FE and FS techniques applied globally for the whole data set and locally in clusters for further classification. Finally, in Section 6 we briefly conclude with a summary and present the directions of further research.

II. DIMENSIONALITY REDUCTION TECHNIQUES USED IN THE STU

A. Feature Extraction Techniques

Principal Component Analysis (PCA) is one of the most commonly used FE techniques. It is based on extracting the axes on which data shows the highest variability [9]. Although PCA “spreads out” the data in the new basis (new extracted axes), and can be of great help in unsupervised learning, there is no guarantee that the new axes are consistent with the discriminatory features in a classification problem.

Another approach is to account for class information during the FE process. One technique is to use some class separability criterion (for example, from Fisher’s linear discriminant analysis), based on a family of functions of scatter matrices: the within-class covariance, the between-class covariance, and the total covariance matrices. Parametric and nonparametric eigenvector-based approaches that use the within- and between-class covariance matrices thus taking into account class information have been analyzed and compared [17]. Both the parametric and nonparametric approaches use the simultaneous diagonalization algorithm to optimize the relation between the within- and between-class covariance matrices. The difference between the approaches is in calculation of the between-class covariance matrix. The parametric approach accounts for one mean per class and one total mean, and therefore may extract at most number_of_classes-1 features. The nonparametric method tries to increase the number of degrees of freedom in the between-class covariance matrix, measuring the between-class covariances on a local basis. Our previous experiments with parametric and nonparametric FE approaches show that nonparametric FE is often more robust to different dataset characteristics and often results in higher classification accuracy of such basic supervised learning techniques as Naïve Bayes, C4.5 and kNN comparing to parametric FE [13].

B. Feature Selection Techniques

Greedy hill climbing is one of the simplest search strategies that consider sequential changes to the current feature subset [Caruana]. Often, it is just the addition or deletion of a single feature from the subset at a time. We selected the most commonly used sequential strategies for FS [1]: forward feature selection (FFS), backward feature elimination (BFE), and bidirectional search (BS). The first strategy starts with no features and successively adds new ones. On the contrary, the second one begins with all the features and step-wisely deletes features one-by-one. Bidirectional search proceeds in both forward and backward directions in turn.

Search algorithms that implement these strategies may consider all possible changes to the current subset and then select the best, or may simply choose the first change that improves the merit of the current feature subset. In either case, once a change is accepted, it is never reconsidered (that is why the name “greedy”). The FS process can stop adding/deleting features when none of the evaluated subsets improves the previous result or, alternatively, the search can continue to produce and evaluate new feature subsets while the result does not start to degrade. The evaluation of selected feature subsets in our study was based on the wrapper paradigm that assumes interaction between the FS process and the classification model [8].

III. DOMAIN: NOSOCOMIAL INFECTIONS

A. Nosocomial Infections

Nosocomial infections and antibiotic resistance (AR) are highly important problems that impact the morbidity and mortality of hospitalized patients as well as their cost of care. It is known that 3 to 40 percent of patients admitted to hospital acquire an infection during their stay, and that the risk for hospital-acquired infection, or nosocomial infection, has risen steadily in recent decades. Formally, nosocomial infections are defined as infections arising after 48 hours of hospital admission. Infections arising earlier are assumed to be arisen prior to admission, though this is not always true [7]. The frequency of nosocomial infections depends mostly on the type of conducted operation being greater for “dirty” operations (10-40%), and smaller for “pure” operations (3-7%). For example, such a serious infectious disease as
meningitis is often the result of nosocomial infection. Analysis of microbiological data included in antibiograms collected in different institutions over different periods of time is considered as one of the most important activities to restrain the spreading of AR and to avoid the negative consequences of this phenomenon. Traditional hospital infection control surveillance and localization of hospital infection often relies on the manual review of suspected cases of nosocomial infections and the tabulation of basic summary statistics which requires considerable time and resources, and produced measures and patterns are often not up-to-date. Advanced computer-based analysis methods might help to discover more potentially useful patterns faster. [3].

It has been widely recognized lately that sophisticated, active, and timely intra-hospital surveillance is needed. Computer-assisted infection control surveillance research has focused on identifying high-risk patients, the use of expert systems to identify possible cases of nosocomial infection, and the detection of deviations in the occurrence of predefined events [6].

Nosocomial infections are the inevitable consequence of long treatment, especially in Intensive Care Units (ICUs). The first step of nosocomial infection arising is the colonization of skin and mucous tunic by hospital microorganism cultures. The peculiarity of these cultures is the acquisition of unpredictable AR according to the policy of the use of antimicrobial medications in the present department or institution. To treat nosocomial infections, at first a microbiological investigation is normally conducted. In this investigation pathogens are isolated and for each isolated bacterium, an antibiogram is built (represents bacterium’s resistance to a series of antibiotics). The user of the test system can define the set of antibiotics used to test bacterial resistance. The result of the test is presented as an antibiogram that is a vector of couples (antibiotic/resistance). The information included in this antibiogram is used to prescribe an antibiotic with a desired level of resistance for the isolated pathogen. The antibiogram is not uniquely identified given a bacterium species, but it can vary significantly for bacteria of the same species. This is due to the fact that the same bacteria of the same species may have evolved differently and have developed different resistances to antibiotics. However, very often groups of antibiotics have similar resistance when tested on a given bacterium species, despite its strains [6].

AR is an especially difficult problem for nosocomial infections in hospitals because they attack critically ill patients who are more vulnerable to infections than the general population and therefore require more antibiotics. Heavy use of antibiotics in these patients hastens the mutations in bacteria that bring about drug resistance [16]. According to the Center for Disease Control and Prevention (CDC) statistics, more than 70 percent of bacteria that cause hospital-acquired infections are resistant to at least one of antibiotics most commonly used to treat infections. Persons infected with drug-resistant organisms are more likely to have longer hospital stays and require treatment with second or third choice drugs that may be less effective, more toxic, and more expensive [16]. In short, antimicrobial resistance drives up health care costs, increases the severity of disease, and increases the death rates of some infections.

B. Source and Nature of Data

The data used in our analysis were collected in the Hospital of N.N Burdenko Institute of Neurosurgery using “Vitek-60” analyzer (developed by bioMérieux) over the years 1997-2004 and information systems ”Microbiologist" (developed by the Medical Informatics Laboratory of the institute) and "Microbe" (developed by Russian company "MedProject-3").

In our previous pilot yet many-sided analysis of this data we applied a number of different DM techniques trying to build an accurate predictive model, to explore and understand our data and find valuable association rules [14].

C. Data Organization

Each instance of the data used in the analysis represents one sensitivity test and contains the following features (Table 1): pathogen that is isolated during the microbe identification analysis, antibiotic that is used in the sensitivity test and the result of the sensitivity test (sensitive $S$, resistant $R$, or intermediate $I$), obtained from “Vitek” according to the guidelines of National Committee for Clinical Laboratory Standards (NCCLS) [6].

The information about sensitivity analysis is connected with a patient, his/her demographical data (sex, age) and hospitalization in the Institute (main department, whether the test was taken while the patient was in ICU, days spent in the hospital before, etc.). Each instance of microbiological test in the database corresponds to a single specimen that may be blood, liquor, urine, etc. In this study we focus on the analysis of meningitis cases only, and the specimen is liquor.

For the purposes of this analysis we picked up all 4430 instances of sensitivity tests related to the meningitis cases of the period of January 2002 – July 2004. Grouping binary features for pathogens and antibiotics were introduced so that 17 pathogens and 39 antibiotics were combined into 6 and 15 groups respectively.

Thus, each instance in our data has 30 features (beside the ID-like attributes for records, patients, antibiotics, pathogens and so on) that included information corresponding to a single sensitivity test augmented with data concerning the type of the antibiotic used and the isolated pathogen, and clinical features of the patient and his/her demographics, and the microbiology test result as the class attribute (Table 1).

IV. NATURAL CLUSTERING APPROACH

We apply so-called natural clustering, i.e. clustering based on the knowledge of domain area experts. The main reason to do so was our belief (supported by expert opinion and pilot exploratory many-sided data analysis) that the patterns of AR have different nature and, therefore, their behavior might be
enterococcus clustering is depicted in Figure 1.

In our database, the whole data set can be divided into two (the hierarchy was introduced by the grouping of the features).

We can see from the figure that clusters gram− and gram+ clusters. The results of our experiments demonstrated that the average accuracy of classifiers (except with C4.5) over subgroups (i.e. when they are applied locally within each cluster) is higher in comparison with the global classifiers accuracy.

These results motivated us to continue the study of our natural clustering approach with regard to applying different DR techniques globally and locally for further construction of local classification models.

V. EXPERIMENTAL STUDY

A. Experiment Design

In Figure 2 we present the main idea behind our experimental setup.

In our experiments we were aimed at checking three major hypotheses:

1. whether natural clustering is an efficient approach for the construction of local models;
2. whether dimensionality reduction within natural clusters produces better representation spaces for further supervised learning in comparison with global dimensionality reduction;
3. whether feature extraction techniques behave differently globally and within natural clusters in comparison with feature selection approaches with regard to their effect on generalization accuracy.

For these purposes we have collected the accuracies for classification with and without DR on the whole data set, and on each natural cluster produced also with and without DR.

In our experimental studies we used an instance-based classifier (kNN), the FFS, BFE, and BS feature selection techniques available in the machine learning library with Java implementation “WEKA 3.4.2” [20]. We used the conventional PCA and the class-conditional parametric (Par) and nonparametric (NPar) FE techniques [13], which we implemented within the same library. We used k=7 and the inverse distance for kNN since these parameters were found to be the best combination for our data in our pilot studies.

different for different natural clusters. Thus, we were interested in how the accuracy of classification models varies from one cluster to another and whether it is possible to achieve better accuracy applying a classifier locally in each cluster instead of the use of global classification.

The data for our analysis is relatively high dimensional and heterogeneous; heterogeneity is presented by a number of contextual (environmental) features. Semantically, the sensitivity concept is related first of all to the pathogen and antibiotic concepts. For our study binary features that describe the pathogen grouping were selected as prior environmental features, and they were used for hierarchical natural clustering (the hierarchy was introduced by the grouping of the features). In our database, the whole data set can be divided into two nearly equal-sized natural clusters: gram+ and gram−. Then, the gram+ cluster consists of the staphylococcus and enterococcus clusters, and gram− cluster consists of the enterobacteria and nonfermentes clusters. This natural clustering is depicted in Figure 1.

We can see from the figure that clusters gram− and gram+ are approximately of the same size. The further division of gram+ cluster results in less balanced clusters. The further division of gram− cluster results in highly unbalanced clusters by the number of instances in each.

In our many-sided exploratory analysis of this data, we applied several classifiers to the whole dataset and individually to the gram− and gram+ clusters. The results of our experiments demonstrated that for example for the Naïve Bayes and C4.5 classifiers the differences of accuracies between the clusters were quite big and at the same time the average accuracy over the clusters is always higher than the corresponding global accuracy achieved with the whole dataset. However, with instance-based classifiers (like kNN-classifier) the difference between average of local accuracies and global accuracy was insignificant for the division of dataset into gram− and gram+ clusters.

We also applied the same classification algorithms for one antibiotic group consisting of three subgroups. And the experimental results demonstrated that the average accuracy of classifiers (except with C4.5) over subgroups (i.e. when they are applied locally within each cluster) is higher in comparison with the global classifiers accuracy.

Major dataset characteristics: 4430 instances, 3 classes, 30 features: 4 numerical, 2 categorical and 24 binary. After binarization of categorical features we have got 48 features, of which 4 are numerical and 44 – binary.

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patient and hospitalization related</strong></td>
<td></td>
</tr>
<tr>
<td>sex</td>
<td>Male, Female</td>
</tr>
<tr>
<td>age</td>
<td>[0;72], mean 29.8</td>
</tr>
<tr>
<td>recurring stay</td>
<td>[True, False]</td>
</tr>
<tr>
<td>days of stay in NSI before test</td>
<td>[0;317], mean 87.5</td>
</tr>
<tr>
<td>days of stay in ICU</td>
<td>[0;237], mean 34</td>
</tr>
<tr>
<td>days of stay in ICU before test</td>
<td>[0;169], mean 31.6</td>
</tr>
<tr>
<td>specimen was received bacterium was isolated when patient is in ICU</td>
<td>[True, False]</td>
</tr>
<tr>
<td>main department</td>
<td>[0,...,9]</td>
</tr>
<tr>
<td>department of stay</td>
<td>[0,...,11]</td>
</tr>
<tr>
<td><strong>Pathogen and pathogen groups</strong></td>
<td></td>
</tr>
<tr>
<td>pathogen name</td>
<td>{Pat_name1,..., Pat_name17}</td>
</tr>
<tr>
<td>group1</td>
<td>{True, False}</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
<tr>
<td>group6</td>
<td>{True, False}</td>
</tr>
<tr>
<td><strong>Antibiotic and antibiotic groups</strong></td>
<td></td>
</tr>
<tr>
<td>antibiotic name</td>
<td>{Ant_name1,..., Ant_name39}</td>
</tr>
<tr>
<td>group1</td>
<td>{True, False}</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
<tr>
<td>group15</td>
<td>{True, False}</td>
</tr>
<tr>
<td>sensitivity</td>
<td>{Sensitive, Intermediate, Resistant}</td>
</tr>
</tbody>
</table>

**TABLE I**

**DATASET FEATURES**

**FIGURE I**

**NATURAL CLUSTERING OF DATA WITH REGARD TO PATHOGENS**
B. Experimental Results

The main results of experiments are given in Table 2. Each row of the table contains the name of the dataset (cluster), number of instances in it, accuracy of 7-NN classifier and the number of features used for each FS (FFS, BFE, and BS) and FE (PCA, Par, NPar) technique averaged over 30 test runs. The column noFS is related to the case when no FS was applied and noFE to the case when no FE was applied but all the categorical features were binarized.

The first row corresponds to the (global) results on the whole data set. The last row corresponds to overall accuracy achieved with the most appropriate selection of sub data sets (clusters): *staphylococcus*, *enterococcus*, and *gram–* (best). We have not analyzed the difference between selected feature subsets. However, we can see that in many cases the number of features (original or transformed) selected is different in different clusters and this number depends on whether DR was applied globally or locally. This also supports our hypothesis about the heterogeneity of data.

In Figure 3 comparison of local and global results of 7-NN classifier for 7 different clusters (including the whole data set) are shown. Results show similar behavior of FS and FE across the 7 different clusters. Analyzing the histograms one by one we can see that the DR techniques for our data result in the best classification accuracy when applied locally to *staphylococcus*, *enterococcus*, and *gram–* clusters.

Applying 7-NN with DR locally in the *gram+* and *gram–* clusters (see Figure 3a) does not outperform global accuracy. However, we can see that accuracy results for cluster *gram+* are much higher than for the *gram–* cluster. The FE methods were almost equally good for the *gram+* cluster. For the *gram–* cluster, Par was the worst and NPar was the best. But still 7-NN without any FE performed slightly better for *gram–*. The FS methods had no effect on 7-NN accuracy for the *gram–* cluster. And BS was the only FS method that increased 7-NN accuracy for *gram+.*

In Figure 3b we can see that applying 7-NN and DR techniques individually to the *staphylococcus* and *enterococcus* clusters significantly changes the overall accuracy. Local Par outperforms local 7-NN by 1.3% (avg Par vs. avg noFE), NPar decreases the accuracy of 7-NN by 2%, and PCA has no effect. Local FS decreases the performance of 7-NN by 3.5 - 4.5%. Relatively low accuracy for the *enterococcus* cluster does not decrease much the average accuracy since this cluster is rather small and contains only 5.7% of instances from *gram+* while *staphylococcus* contains 94.3%. However, the analysis of how good or bad the performance of certain local model is helps to understand which subsets of instances are harder to classify or which subsets of instances are very noisy.

In Figure 3c we can see that dividing the cluster *gram–* further into *enterobacterias* and *nonfermentes* does not increase the accuracy of 7-NN both with and without local FE or FS.

Now, if we compare the FE and FS horizontal triples of histograms we can see that for our data the sequential strategies for FS have no success. Exceptionally, BS was successful when applied individually to *gram+* and *gram–* clusters. The FE methods have more diverse behaviour. So, PCA is the best for the *enterococcus* cluster while Par is the best for *staphylococcus* (Figure 3b top). NPar demonstrated the best accuracy for global FE on the whole data set. This leads to an idea of adaptive selection of FE method for each cluster, that is the use of PCA in one cluster, and Par or NPar in some other cluster may result in significantly higher overall accuracy. Figure 4 shows the classification accuracies for our data with regards to the selection of the best DR techniques within corresponding clusters: locally in *gram+* and globally in *gram–*.

We compare the impact of FS and FE on classification either globally or locally with the most appropriate selection of clusters: *staphylococcus*, *enterococcus* (joined into averaged results of *gram+*), and *gram–*. Due to space limitations we do not present a separate figure but list here the main conclusions of this comparison: (1) Natural clustering is useful for our data (in terms of increasing generalization accuracy) only by means of FE with any (global or local) DR; (2) FE was useful (it improved generalization accuracy) both when applied globally and locally, while FS increases the accuracy of 7-NN only when applied locally to the *gram+* and *gram–* clusters. This fact supports the hypothesis about heterogeneity of our data; (3) FS applied locally on this data results in higher accuracy produced by 7-NN comparing to local FE. However we need to point out that this was due to the binarization of categorical features (that is required for FE) that leads to increase in redundant binary features. 7-NN produces almost 3% higher accuracy results for data presented by original categorical (not binarized) features. Perhaps, by analyzing possible reasons of why the accuracy of 7-NN on this data decreases after binarization, we can improve the overall situation in the FS-FE competition; (4) Par produced very poor results when applied globally, but performed surprisingly well in some of the clusters. NPar was quite stable across
different clusters, and it was the best FE technique for the situation when DR was applied globally.

VI. CONCLUSION AND FUTURE DIRECTIONS

DR is an effective approach to data reduction aimed at focusing on relevant features and improving the quality of data representation for classification. We experimentally compared and showed the benefits of local and global DR by means of FS and FE. In this study we applied the natural clustering approach aimed at using contextual features for splitting a real-world clinical data set into more homogeneous clusters in order to construct local models that would help in the better prediction of antibiotic resistance.

The results of our experiments demonstrate that the proper selection of a local DR technique can lead to a significant increase of predictive accuracy in comparison with the global 7NN classification with or without DR. The amount of features extracted or selected locally is always smaller than that in the global space that also shows the usefulness of natural clustering in coping with data heterogeneity (and higher dimensionality).

Our future research efforts are going to be directed towards the comparison of a mixture of FE models for classification built on natural clusters and on clusters produced by traditional clustering techniques. We analyzed spatial contextual features related to categorization of different pathogens. We believe that natural clustering according to features that contain implicit or explicit information about timestamp of certain instance may give interesting results in different time contexts.

Another challenging goal is the adaptive selection of FE method for each cluster according to certain characteristics of the cluster. So, the appropriate use of PCA in one cluster, and Par or NPar in some other cluster may result in significantly higher overall accuracy.

ACKNOWLEDGMENT

We would like to thank Dr. Michael Shifrin and Dr. Irina Alexandrova from N. N. Burdenko Institute of Neurosurgery, Russian Academy of Medical Sciences, Moscow, Russia for the provided dataset and helpful discussions. We also thank WEKA developers for the source code used in our work.

REFERENCES

FIGURE 2. EXPERIMENTAL APPROACH

TABLE 2: THE BASIC EXPERIMENTAL RESULTS: ACCURACY OF 7NN CLASSIFIER FOR EACH APPROACH AMD NUMBER OF FEATURES USED BY CORRESPONDING APPROACH FOR CLASSIFICATION ON AVERAGE

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Inst</th>
<th>Accuracy of 7-NN classifier and number of features used</th>
<th>Feature Selection</th>
<th>Feature Extraction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>FFS</td>
<td>BFE</td>
</tr>
<tr>
<td>gram+ (g+)</td>
<td>2134</td>
<td>.787 5 .784 5 .798 5 .788 24</td>
<td>.745 19</td>
<td>.752 1</td>
</tr>
<tr>
<td>staphiloc.</td>
<td>2013</td>
<td>.799 5 .757 5 .756 5 .799 23</td>
<td>.766 20</td>
<td>.785 1</td>
</tr>
</tbody>
</table>

Figure 3. Accuracy results of 7NN classifier with different approaches: