Advances in Smart Systems Research ISSN 2050-8662
Vol. 6. No. 2 : pp.41-54 : k17is-223

International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2017, 6-8 September 2017, Marseille, France

Statistical Analysis and Optimization of Classification Methods of Big Data in Medicine

George K. Fourfouris[1], George - Peter K. Economou[2], Spiridon Likothanassis[3]

[1]Department of Computer Engineering and Informatics, University of Patras, 26504
fourfouris@ceid.upatras.gr

[2]Department of Computer Engineering and Informatics, University of Patras, 26504
gpoikonomou@ceid.upatras.gr

[3]Department of Computer Engineering and Informatics, University of Patras, 26504
likothan@ceid.upatras.gr

Abstract

The process of big data classification has been extensively explored for the past decades, being it essential for Machine Learning. Likewise, performing statistical analysis and investigating the optimization factors of applied algorithms, while evaluating datasets based on recent experience, are important as well in order to specify the best case scenarios parameters' that enhance their accuracy. In this paper search on and implementation of classification algorithms, such as the state-of-the-art k-Nearest Neighbors (k-NN), the alternative k-Nearest Neighbors on Feature Projections (k-NNFP) algorithm and Vote Feature Intervals (VFI), over big medical datasets from UCI Machine Learning Repository, is presented. These algorithms, moreover, are statistically analyzed on a standardized Arrhythmia Dataset [1] in order to extract the scenarios that enhance the precision execution for multiple k-fold cross validations and the number of nearest neighbors. Additionally, feature selection and extraction methods are implemented in order to exploit and establish best-case scenarios.

1. Introduction

In recent decades, the problem of categorization and learning how to classify sets of objects, has been extensively studied in the field of Machine Learning; hence, many state-of-the-art algorithms, approaches and derivatives have been proposed in various fields of human expertise. Classification algorithms can greatly vary to include distance-based between training and classification examples ones, feature-based algorithms that depend on calculations, which are evaluated over each feature separately, as well as vote feature-based processes that hold the classification information after voting procedures per classification example feature [15], [16].

K-Nearest Neighbors [2], [3] (k-NN) is a state-of-the-art classification method that calculates the k training examples that are closer to the classification one in its feature space. In addition, k-NN extracts the most frequent class as the dominant class of the classification example. The distance function that is commonly used in the k nearest training example's procedure is the Euclidean distance. *K-Nearest Neighbors on Feature Projections* [4] (k-NNFP) is an alternative approach of k-NN algorithm that represents the classification knowledge as sets of projections of each training data feature. Classification examples take into account feature votes that produce the outcome class. The *Vote Feature Intervals* [5] (VFI) classification algorithm, on the other hand, represents the classification knowledge as a set of feature intervals of each training data feature. Classification examples are classified by feature votes such as in k-NNFP case, equivalently.

Among these algorithms there are a lot of factors and metrics that are very important, such as the best number of Nearest Neighbors, or the accuracy and CPU time spent during the evaluation of each algorithm. The research, especially, tries to illuminate the various scenarios, which actively affect the accuracy of the results that is a dominant factor in Medical Science over diagnosis process, as a matter of fact, and more. In this paper, *Section 2*, gives the definitions of the classification algorithms that are used for the statistical analysis and optimization, as introduced above. *Section 3* and *Section 4* describe in more detail the experimental and evaluation frameworks for the classification algorithms' statistical analysis. In addition, we reference in *Section 5* the group outcomes and the best-case scenarios of the algorithms. In *Section 6* we determine the main references and key points of the statistical analysis and evaluation, which focus on advantages and disadvantage of the classification algorithms and the results' outcomes. Apart from the statistical analysis, further Decision Tree and Feature Selection research outcomes that less feature space is actively used for the dataset classification. Finally, a discussion over future work and further experimentation concludes the paper.

2. The Classification Algorithms

We already mentioned that the process of classification is fundamental in Machine Learning. In medical problems the application of a classification process usually represents the prognosis or diagnosis of an illness. Prognostics of breast cancer, thyroid disease, or rheumatology diagnosis are different cases of those classification applications [6]. This paper evaluates three (3) classification algorithms, presented in the following, on a standardized Arrhythmia Dataset. The dataset is available on UCI Machine Learning Repository web pages [1].

2.1 K-Nearest Neighbors algorithm (k-NN)

The *K-Nearest Neighbor* (k-NN) algorithm is a state-of-the-art classification algorithm in machine learning with a great variety of applications and alternative approaches. The Nearest Neighbor (NN) approach, tries to find the closest neighbor of a given unclassified example. The class label of the neighbor with the closest distance in feature space is assigned to the unclassified example.
Respectively, the *k-Nearest Neighbors* process finds the k closest neighbors to a given unclassified example. The class label that is more frequently appeared to the k neighbors, is assigned to the unclassified example, equivalently.

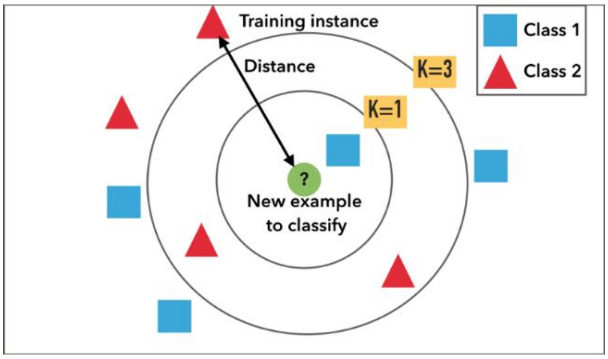


Figure 1. Examples of *k-NN* algorithm for K = 3 and *NN* algorithm for K = 1

2.2 K-Nearest Neighbors on Feature Projections algorithm (k-NNFP)

The *K-Nearest Neighbors on Feature Projections* (k-NNFP) algorithm is an alternative approach of the state-of-the-art k-NN classification process. This approach is based on extracted votes applied on a set of classes, which derive after k-NN performance in each feature separately of the unclassified example. The most important characteristic of this classification algorithm is that the training set is stored as an equivalent projection set on feature space. In the dataset that is delivered below in *Table 1* three training examples with two features and a class feature per example is presented.

Training Example	f_0	f_1	Class
Example 1	1	6	**B**
Example 2	4	3	**A**
Example 3	10	5	**A**

Table 1. Training Examples

The *k-NNFP* algorithm classifies a given unclassified examples as follows. Firstly, the k-closest training examples are calculated for each feature, and their classes are stored, separately. In the next step, the sum of each class is calculated in

feature storage space. Finally, the class with the max sum is performed to be the class of the unclassified example.
In example, given an unclassified example *(<3, 2>)*, where f_0 = 3, f_1 = 2 and the label class is unknown, the k-NNFP approach classifies the example to the 'A' class for
k = 1 and *k = 2*, as it is presented in *Table 2*.

k	*f_0*	*f_1*	*Features' Bag*	*Class*
1	[A]	[A]	[A, A]	**A**
2	[A, B]	[A, A]	[A, B, A, A]	**A**

Table 2. *K-NNFP* classification for unclassified example (<3, 2>)

Considering their differences and the similarities, the *k-NN* approach focuses on the distances between training dataset and unclassified examples that provide the outcome classification on feature space, in opposite with the *k-NNFP* approach, which is mainly focused on each feature contribution to the classification process via the majority of feature votes.

2.3 Vote Feature Intervals

The *Vote Feature Intervals* (VFI) algorithm extends the predefined *k-NNFP* approach in terms of the training and the classification process. The main concept of the algorithm includes the storage of the feature *End_Points* and the definition of the feature *Intervals*, in the *Training Process*, as well as, the calculation of the feature *Votes* for the unclassified examples and the summary of those votes over all features for each class in the *Classification Process*.
End_Points are the minimum and the maximum values for each class and *Intervals* are the sets of spaces between the sorted *End_Points* for each feature. Given an unclassified example, *Votes* for each feature and class are calculated. Moreover, the class with the highest features' *Vote* sum outcomes is class of the unclassified example.

3. Experiments Framework

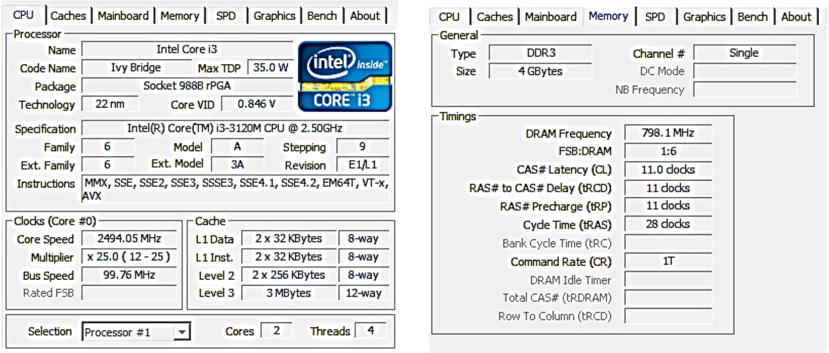
As it mentioned in previous sections, this paper performs statistical analysis and optimization scenarios on classification algorithms. This analysis uses the Arrhythmia Dataset from UCI Machine Learning Repository as the input of the experiments. The dataset consists of *279 features*, *1 class* feature and *452 examples*. Each example is corresponded to one class projection out of the 16 that are showed in *Table 3*.

Type of Arrhythmia	Class Number	Number of examples
Normal	1	245
Ischemic changes (Coronary Artery Disease)	2	44
Old Anterior Myocardial Infarction	3	15
Old Inferior Myocardial Infarction	4	15
Sinus tachycardia	5	13
Sinus bradycardia	6	25
Ventricular Premature Contraction (PVC)	7	3
Supraventricular Premature Contraction	8	2
Left bundle branch block	9	9
Right bundle branch block	10	50
1. degree AtrioVentricular block	11	0
2. degree AV block	12	0
3. degree AV block	13	0
Left ventricular hypertrophy	14	4
Atrial Fibrillation or Flutter	15	5
Others	16	22

Table 3. The Arrhythmia Dataset Classes and number of the initial examples per class.

The classification algorithms, which are used for our experiments, are the *k-NN* algorithm, the alternative *k-NNFP* algorithm and *VFI* algorithm. Experiments are performed *for k = 5, 10, 20, 50, 100, 200* nearest neighbors. Additionally, the *k-fold cross validation* technique is used, for *10, 20, 50* folds, in order to endure the analysis outcomes. For each case of nearest neighbor number and number of folds, we calculate the average out of ten executions, in order to extract our results. The results are referred to classification accuracy metric of the algorithm evaluation, which corresponds to the number of correctly classified examples out of the overall examples population. In addition, CPU time is taken into account as a metric for the statistical analysis.

Last but not least, there are two scenarios of algorithms' evaluation. Firstly, dataset is purely used on classification evaluation with no preprocessing. The second scenario prunes the features that contain missing values and uses the pruned dataset for classification.



Processor
Name: Intel Core i3
Code Name: Ivy Bridge
Max TDP: 35.0 W
Package: Socket 988B rPGA
Technology: 22 nm
Core VID: 0.846 V
Specification: Intel(R) Core(TM) i3-3120M CPU @ 2.50GHz
Family: 6
Model: A
Stepping: 9
Ext. Family: 6
Ext. Model: 3A
Revision: E1/L1
Instructions: MMX, SSE, SSE2, SSE3, SSSE3, SSE4.1, SSE4.2, EM64T, VT-x, AVX

Clocks (Core #0)
Core Speed: 2494.05 MHz
Multiplier: x 25.0 (12 - 25)
Bus Speed: 99.76 MHz
Rated FSB:

Cache
L1 Data: 2 x 32 KBytes, 8-way
L1 Inst.: 2 x 32 KBytes, 8-way
Level 2: 2 x 256 KBytes, 8-way
Level 3: 3 MBytes, 12-way

Selection: Processor #1
Cores: 2
Threads: 4

General
Type: DDR3
Size: 4 GBytes
Channel #: Single
DC Mode:
NB Frequency:

Timings
DRAM Frequency: 798.1 MHz
FSB:DRAM: 1:6
CAS# Latency (CL): 11.0 clocks
RAS# to CAS# Delay (tRCD): 11 clocks
RAS# Precharge (tRP): 11 clocks
Cycle Time (tRAS): 28 clocks
Bank Cycle Time (tRC):
Command Rate (CR): 1T
DRAM Idle Timer:
Total CAS# (tRDRAM):
Row To Column (tRCD):

Figure 2. Computer System for the statistical analysis evaluation
(CPU frequency and Memory space).

The *k-NN, k-NNFP* and *VFI* classification algorithms have been developed, both from built-in Mathworks functions and from scratch. The simulations have been established in *MATLAB r2017* environment on a computer with specifications that are referenced above in *Figure 2*.

4. Evaluation Framework

In this section an extended presentation of the *k-NN, k-NNFP* and *VFI* classification algorithms' evaluation is performed on Arrhythmia Dataset. Comparison on average *accuracy* outcomes among different *nearest neighbors* and *k-fold cross validation* scenarios are referenced, as well.

The extracted accuracy from the classification algorithms' evaluation increases in accordance with the increase of nearest neighbor and folds number. Moreover, the analysis of the classification shows that the extracted accuracy seems to reach its peak in multiple scenarios. Experiments do not include pruning preprocessing of the missing value features.

Nearest Neighbors	10-fold		20-fold		50-fold	
5	65.11	0.20	60.87	0.25	71.11	0.68
10	64.22	0.20	80.00	0.62	84.44	0.64
20	69.11	0.21	68.26	0.54	87.78	0.64
50	70.89	0.17	69.57	0.31	87.78	0.60
100	71.11	0.20	65.22	0.54	77.78	0.60
200	71.11	0.21	65.22	0.32	88.89	0.59

Table 4. *K-NN* evaluation on Arrhythmia Dataset without missing value features pruning. Accuracy metric measures (%) the correctly classified examples out of all dataset population. CPU time (sec) measures the overall execution time.

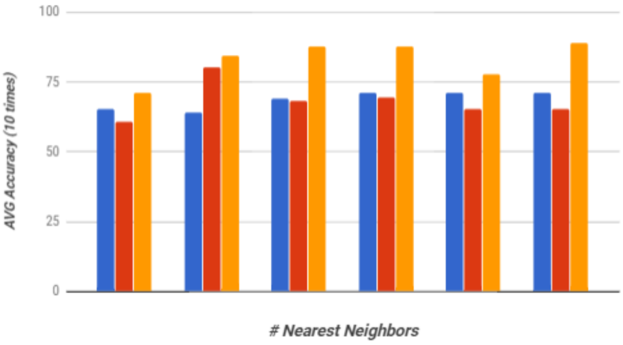


Figure 3. Graphical representation of *Table 4*. The *x-axis* represent the *NN number* (5/10/20/50/100/200). The first bars (blue color) reference the ***10-fold*** series, the second bars (red color) reference the ***20-fold*** series and the third bars (yellow color) reference the ***50-fold*** series.

Nearest Neighbors	10-fold		20-fold		50-fold	
5	71.11	17.21	69.57	16.93	80	17.21
10	57.78	16.51	73.33	17.75	100	17.12
20	71.11	16.71	66.67	18.40	88.89	18.89
50	71.11	17.54	82.61	18.45	88.89	18.75
100	71.11	18.43	78.26	18.89	100	19.54
200	71.11	19.20	65.22	20.25	88.89	20.48

Table 5. *K-NNFP* evaluation on Arrhythmia Dataset without missing value features pruning. Accuracy metric measures (%) the correctly classified examples out of all dataset population. CPU time (sec) measures the overall execution time.

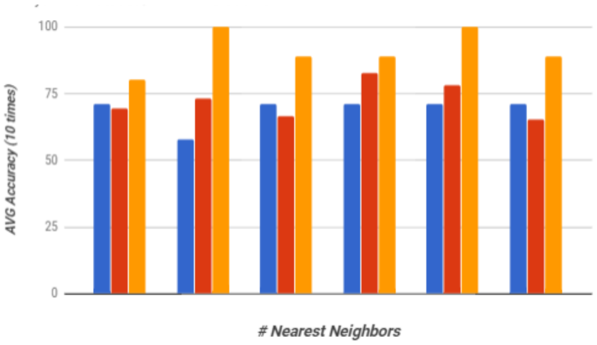


Figure 4. Graphical representation of *Table 5*. The *x-axis* represent the *NN number* (5/10/20/50/100/200).The first bars (blue color) reference the ***10-fold*** series, the second bars (red color) reference the ***20-fold*** series and the third bars (yellow color) reference the ***50-fold*** series.

10-fold		20-fold		50-fold	
23.33	6.53	30.43	7.01	38.89	14.43

Table 6. *VFI* evaluation on Arrhythmia Dataset without missing value features pruning. Accuracy metric measures (%) the correctly classified examples out of all dataset population. CPU time (sec) measures the overall execution time.

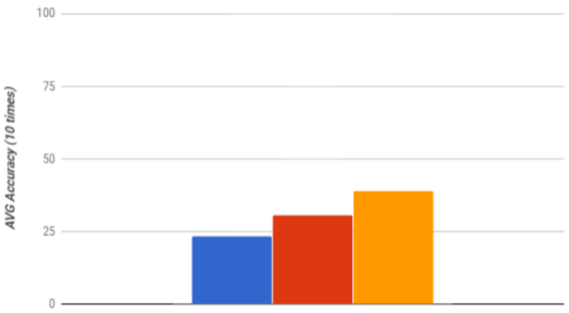


Figure 5. Graphical representation of *Table 6*. The first bar (blue color) reference the ***10-fold*** outcome, the second bar (red color) reference the ***20-fold*** outcome and the third bar (yellow color) reference the ***50-fold*** outcome.

In the next step, we perform classification evaluation of the algorithms with the preprocessing phase. This phase prunes the features of the initial dataset, which include missing values. The missing value features pruning phase is very important, because in real life if a feature is unknown, then it will be ignored and will not participate in the prediction process. Below we present the statistical analysis only for the *k-NN* and *k-NNFP* classification algorithms, because the *VFI* approach ignores the missing values features on prediction process, by default.

Nearest Neighbors	10-fold		20-fold		50-fold	
5	62.44	0.17	71.74	0.28	91.11	0.62
10	66.00	0.18	78.70	0.23	92.22	0.67
20	69.78	0.18	79.57	0.28	98.89	0.67
50	71.11	0.19	82.17	0.25	100	0.65
100	71.11	0.17	82.61	0.27	100	0.66
200	71.11	0.18	82.61	0.29	100	0.64

Table 7. *K-NN* evaluation on Arrhythmia Dataset with missing value features pruning. Accuracy metric measures (%) the correctly classified examples out of all dataset population. CPU time (sec) measures the overall execution time.

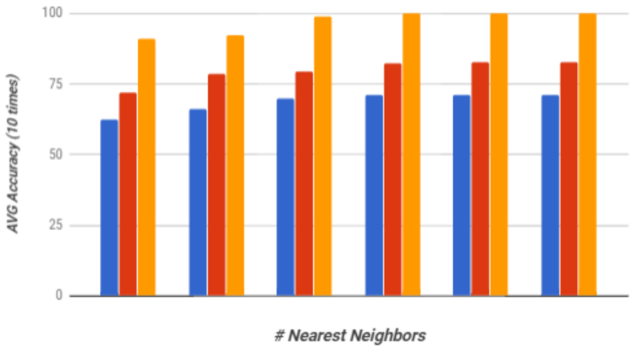


Figure 6. Graphical representation of *Table 7*. The *x-axis* represent the *NN number* (5/10/20/50/100/200).The first bars (blue color) reference the ***10-fold*** series, the second bars (red color) reference the ***20-fold*** series and the third bars (yellow color) reference the ***50-fold*** series.

Nearest Neighbors	10-fold		20-fold		50-fold	
5	71.11	18.17	82.61	19.26	100	19.95
10	71.11	17.56	82.61	19.42	100	19.21
20	71.11	19.25	82.61	17.85	100	17.20
50	71.11	18.48	82.61	17.48	100	17.70
100	71.11	17.82	82.61	18.06	100	18.43
200	71.11	18.78	82.61	19.14	100	20.03

Table 8. *K-NNFP* evaluation on Arrhythmia Dataset with missing value features pruning. Accuracy metric measures (%) the correctly classified examples out of all dataset population. CPU time (sec) measures the overall execution time.

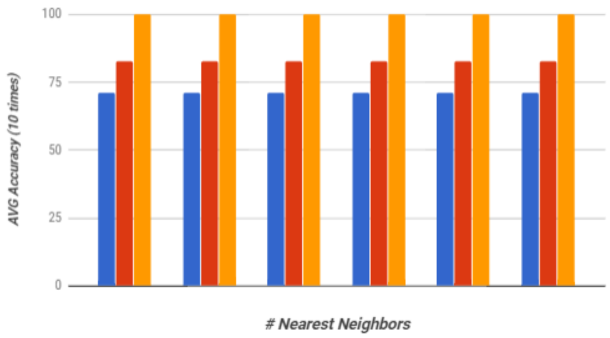


Figure 7. Graphical representation of *Table 8*. The *x-axis* represent the *NN number* (5/10/20/50/100/200).The first bars (blue color) reference the ***10-fold*** series, the second bars (red color) reference the ***20-fold*** series and the third bars (yellow color) reference the ***50-fold*** series.

After performing several experiments we have observed that the missing value features pruning claim is truly verified. Classification algorithms reach their peak accuracy in the most cases. This claim makes us consider that the given dataset calculates its label class with features, which are not needed.. In other words, the initial dataset probably contains extra features with noise that should be pruned or discard. Extra analysis on the Arrhythmia dataset is approached in order to find those features that play active role on the prediction process. This analysis process contains the extraction of the Arrhythmia dataset classification Decision Tree [7], in order to extract the rules and features that actively participate in the classification process.

From the Classification Tree bellow we have observed that only *41* out of *279 features* actively are taken into account on the classification process. This outcome poses another consideration about the existence of dominant features on the initial Arrhythmia Dataset.

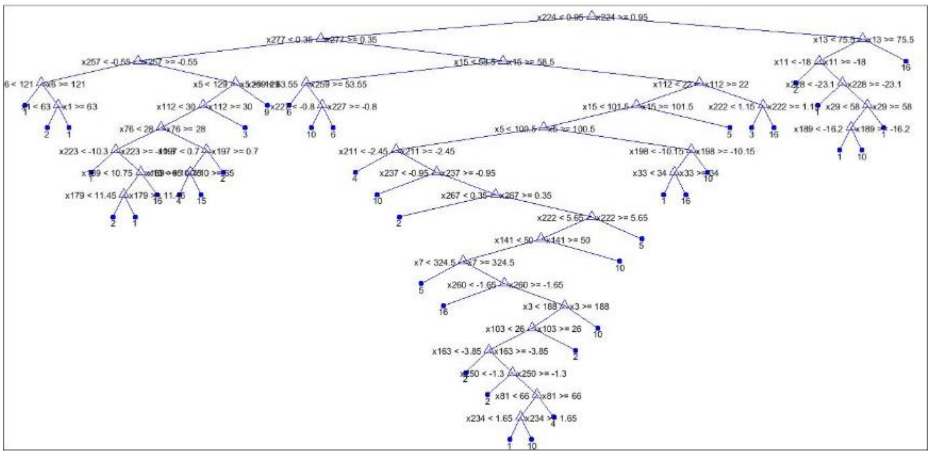


Figure 8. Classification Decision Tree for the Arrhythmia Dataset.

Feature Selection [8] process with Nearest Component Analysis [9] (*NCA*) has showed that there is a feature weight barrier that includes lots of features. Over and below that barrier feature weights are mostly scattered. In the *Figure 9* is represented an example of the referred barrier and in *Table 9* are presented 10 examples of Feature Weight barriers.

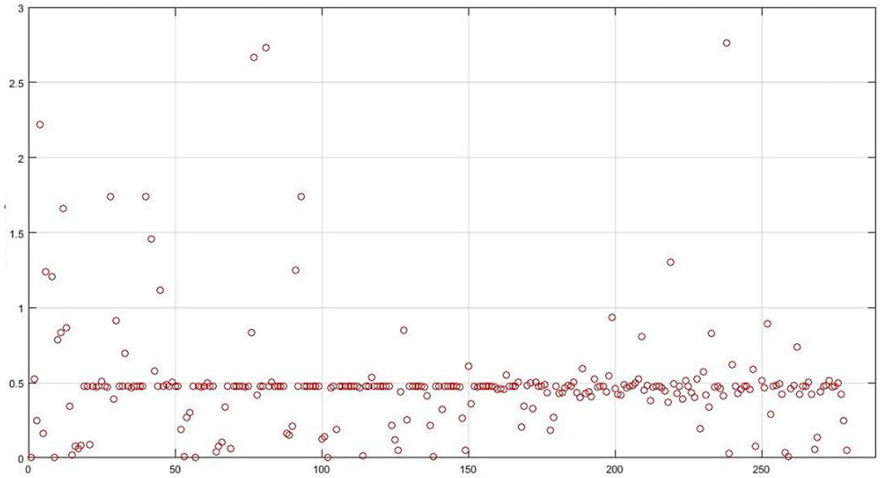


Figure 9. Evaluation of Feature Selection process with *NCA* algorithm example (*x-axis Feature weight/y-axis Feature index*).

	Ex1	Ex 2	Ex 3	Ex 4	Ex 5	Ex 6	Ex 7	Ex 8	Ex 9	Ex 10
Ft. Weight Barrier	0.4746	0.0497	0.9112	0.6892	0.8303	0.6892	0.6892	0.4746	0.8303	0.0023
Below Barrier	115	67	80	112	94	100	108	134	90	42
Above Barrier	82	103	109	76	105	97	96	58	93	120
On Barrier	82	109	90	91	80	82	75	87	96	117

Table 9. Feature Weight outcome examples from *Feature Selection* with *NCA* algorithm process.

Specifically, *Table 9* refers that *~1/3* of the feature space is important, essentially. Several examples, which are referenced above verify the claim of less feature space dominance. Features, which are below the barrier are less important for the classification process than the features which are above, likewise. In this way, dataset classification is truly affected by noisy data that are directly connected with the outcome accuracy and execution. Even if this outcome seems minor to the Informatic world, it is very important for the Medical world and so to the prediction or in-time treatment process, moreover.

5. The Results

Classification algorithms that are used in medical fields have to perform accurately [10], [11], [12], [13], [14]. Nevertheless, extended statistical analysis is needed in order to extract the best case scenarios or moreover to search for the factors that probably will enhance the classification accuracy. These scenarios are highly important in order not only to find the best accuracy scenarios but to directly target those irrelevant factors and noisy data that affect the classification process and accuracy.

First of all, *k-NN* and *k-NNFP* are similar algorithms that handle very well datasets with or without missing values on feature space. The *k-NNFP* algorithm is observed to perform more accurately than the initial *k-NN*. Additionally, *k-NNFP* reaches its peak accuracy in most of the nearest neighbors and folds number cases, when the Arrhythmia Dataset contains missing value features and in all of the cases when the Arrhythmia Dataset does not, equivalently. Nevertheless, *k-NN* algorithm handles both missing value featured dataset and pruned dataset with the same accuracy in all cases and scenarios. *VFI* is a discard missing value features algorithm, by default, and it seems to handle not as much accurate the Arrhythmia Dataset for different number of folds scenarios.

In extension of our experimental scenarios, we have discovered that actually a small subset of the feature space of the Arrhythmia Dataset does actively participate in the prediction process. This outcome refers to the fact of feature space reduction and the existence of noisy data. Moreover, classification decision tree showed that only *41* out of *279* features (~15% of the features) are important for the classification process. This outcome proves that feature selection is needed, in order to extract the prediction of the most important features number. The examples have showed that there is a barrier of feature weight. In most of the cases, *~1/3* of the features seems to play more important role in the prediction process. The rest *2/3* of the features seem to play less important role on classification process. Likewise, the claim of noisy feature space and data, which are actively used in classification process, in verified once more. Moreover, diagnosis process and probable treatment could accurately take place with less dataset feature space and calculations.

6. Conclusion

Similar classification algorithms such as *k-NN*, *k-NNFP* and *VFI* handle very accurately big datasets with medical entries. Nevertheless, statistical analysis on those datasets seems to be needed in order to determine the different scenarios that will optimize the algorithms' metrics and the overall extracted accuracy. This optimization is very important because probable noisy scenarios may actively affect the outcome accuracy and execution results of the classification algorithms. Those results, in addition, especially accuracy, are dominant for the diagnosis process as long as for the in-time treatment.

The *k-NN* algorithm is observed to be the algorithm that handles, accurately, both scenarios with datasets that include missing value features or not. Respectively, *k-NNFP* algorithm even it performs more accurately than *k-NN*, in general, in all of the cases it reaches the peak performance, only on missing value pruned features scenarios. *VFI* algorithm seems to be the less accurate algorithm in all of our experimental scenarios.

The missing value features prune process extends our consideration for the existence of extra noisy features as well as for the feature selection and weight dominance. The Arrhythmia dataset classification decision tree verifies this consideration and results that only *~15%* of the feature space actively participates in the prediction process. Moreover, feature selection with Nearest Component Analysis experimental process outcomes the fact that *~1/3* of the feature space is more important for classification.

Future work would contain either the effect on the accuracy of the features, which Classification Decision Tree process has extracted or the evaluation of the statistical analysis on different medical datasets.

7. References

[1] H. Altay Guvenir, Burak Acar, Gulsen Demiroz, Ayhan Cekin. *Arrhythmia Data Set*, UCI Machine Learning Repository, 1998.

[2] O. Sutton, *Introduction to kNearet Neighbour Classification and Condensed Nearest Neighbour Data Reduction*, 2012.

[3] Leif E. Peterson, Scholarpedia, 4(2):1883, 2009.

[4] Aynur Akkus, and H. Altay Güvenir, *K Nearest Neighbor Classification on Feature Projections*, in Proceedings of ICML'96, Lorenza Saitta (Ed.), Morgan Kaufmann, Bari, Italy, 1996

[5] G. Demiröz, and H. A. Güvenir, *Classification by Voting Feature Intervals*, in Proceedings of 9th European Conference on Machine Learning (ECML-97), Maarten van Someren and Gerhard Widmer (Eds.) Springer-Verlag, LNAI 1224, Prague, Czech Republic, 1997, p. 85-92

[6] Kononenko, I*, Inductive and Bayesian learning in medical diagnosis*, Applied Artificial Intelligence an International Journal, 7(4), 1993, p. 317-337.

[7] Kamiński, B., Jakubczyk, M., Szufel, P.. "A framework for sensitivity analysis of decision trees". Central European Journal of Operations Research, 2017

[8] Chandrashekar, G., & Sahin, F. A survey on feature selection methods. Computers & Electrical Engineering, 40(1), p. 16-28, 2014.

[9] Bro, R., & Smilde, A. K., Principal component analysis. Analytical Methods, 6(9), 2014, p. 2812-2831.

[10] Bennasar, M., Setchi, R., Bayer, A., & Hicks, Y. *Feature selection based on information theory in the Clock Drawing Test*. Procedia Computer Science, 22, 2013, p. 902-911.

[11] Azzopardi, C., Hicks, Y. A., & Camilleri, K. P. *Exploiting gastrointestinal anatomy for organ classification in capsule endoscopy using locality preserving projections*. In Engineering in Medicine and Biology Society (EMBC), 35th Annual International Conference of the IEEE, 2013, p. 3654-3657.

[12] Ghodsi, M., Sanei, S., Hicks, Y., Lee, T., & Dunne, S. (2007, September). *A facial pattern recognition approach for detection of temporomandibular disorder*. In Signal Processing Conference, 15th European, 2007, p. 1950-1954.

[13] T. Shigemori, H. Kawanaka, Y. Hicks, R. Setchi, H. Takase and S. Tsuruoka, "*Dementia Detection Using Weighted Direction Index Histograms and SVM for Clock Drawing Test*", 20th International Conference on KnowledgeBased and Intelligent Information & Engineering Systems (KES-2016), 2016, p. 1240-1248.

[14] K. Fukuma, V. B. S. Prasath, H. Kawanaka, B. J. Aronow and H. Takase, "*A study on nuclei segmentation, feature extraction and disease stage classification for human brain histopathological images*", 20th International Conference on KnowledgeBased and Intelligent Information & Engineering Systems (KES-2016), 2016, p. 1202-1210.

[15] Kezirian, E. J., Hohenhorst, W., & de Vries, N. Drug-induced sleep endoscopy: the VOTE classification. European Archives of Oto-Rhino-Laryngology, 268(8), 2011, p. 1233-1236.

[16] Broekaert, S., Roy, R., Okamoto, I., Van Den Oord, J., Bauer, J., Garbe, C. & Elder, D. E. Genetic and morphologic features for melanoma classification. Pigment cell & melanoma research, 23(6), 2010, p 763-770.