

Hybrid Method HVS-MRMR for Variable Selection in Multilayer Artificial Neural Network Classifier

Ben-Hdech Adil¹, Ghanou Youssef², El Qadi Abderrahim³

^{1,2}TIM Team, High School of Technology, Moulay Ismail University, Meknes, Morocco

³LASTIMI, High School of Technology, Mohammed V University, Rabat, Morocco

Article Info

Article history:

Received Mar 29, 2017

Revised May 5, 2017

Accepted Jul 11, 2017

Keywords:

Heuristic variable selection

Neuronal network

Minimum redundancy

Maximum relevance

Multilayer perceptron

Variable selection

ABSTRACT

The variable selection is an important technique the reducing dimensionality of data frequently used in data preprocessing for performing data mining. This paper presents a new variable selection algorithm uses the heuristic variable selection (HVS) and Minimum Redundancy Maximum Relevance (MRMR). We enhance the HVS method for variable selection by incorporating (MRMR) filter. Our algorithm is based on wrapper approach using multi-layer perceptron. We called this algorithm a HVS-MRMR Wrapper for variables selection. The relevance of a set of variables is measured by a convex combination of the relevance given by HVS criterion and the MRMR criterion. This approach selects new relevant variables; we evaluate the performance of HVS-MRMR on eight benchmark classification problems. The experimental results show that HVS-MRMR selected a less number of variables with high classification accuracy compared to MRMR and HVS and without variables selection on most datasets. HVS-MRMR can be applied to various classification problems that require high classification accuracy.

Copyright © 2017 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Ben-Hdech Adil,

Department of Electrical and Computer Engineering,

National Chung Cheng University,

168 University Road, Minhsiung Township, Chiayi County 62102, Taiwan, ROC.

Email: lsntl@ccu.edu.tw

1. INTRODUCTION

Reducing dimensionality of dataset has become increasingly critical because of the multiplication of data. In many areas, the solution of a system problem is based on a set of database (variables) [1-2]. Increasing the number of these variables that characterizes the problem represents difficulties at many levels such as complexity, computing time, and deterioration of the system problem solution in the presence of noisy data. A method of reducing dimensionality is to find a representation of the original data in a smaller space. Dimensionality reduction can roughly be divided into two categories [3-4]: feature extraction and feature selection. Firstly, Feature extraction generates a small set of novel features by merging the original features. Secondly, Feature selection picks a small set of the original ones.

Variables selection or features selection is a search process used to select a subset of variables for building robust learning models [5] such as neural networks, decision trees and others. Some irrelevant and/or redundant variables exist in the learning data that make learning harder and decrease the performance of learning models. The variables selection methods can be classified into three main categories: filter, wrapper and embedded. Filter methods were the first used for the variables selection. This category allows evaluating the relevance of a variable according to measures that rely on the properties of the learning data. Filter techniques are fast for high-dimensional datasets, but they ignore interaction with the classifier [6]. The wrapper methods use the predictive accuracy of a predetermined learning algorithm to determine the best

subset selected. The accuracy of the learning algorithms is usually high [7]. Wrapper methods tend to find the most suitable feature subset for the learning algorithm, but they are very computationally expensive. Unlike the wrapper and filter methods, embedded methods incorporate the selection of variables during the learning process. The embedded methods combine the advantages of filter and wrapper techniques [8]. The filter approach determines the relevant and redundant variables independent of the classification, such as using only MRMR criterion, so it is not recommended to use it alone [9]. The method filter might improve the selection of variables if it understands how the filtered variables are used by the classifier. The wrapper method evaluates a subset of features by its classification performance using a learning algorithm [10], for example, heuristic variable selection (HVS). In this work, we propose to incorporate MRMR criterion into the ranking scheme of HVS. We are making hybrids by a convex combination of the relevancy given by HVS criterion and the MRMR criterion.

The rest of the paper is organized as follows: Section 2 we present related works. In section 3 we present research method. Section 4 presents the results of our experimental studies including the experimental methodology, experimental results, and the comparison with heuristic variables selection HVS and MRMR Minimum Redundancy Maximum Relevance. The conclusions are drawn in Section 5.

2. RELATED WORKS

The variables selection is generally defined as a search process to find a subset of "relevant" characteristics from those of the original set [5], [11-13]. The concept of relevance of a subset of variables always depends on the objectives and system requirements. The problem of variables selection for classification task can be described as follows: given the original set G , of N features, find a subset F consisting of N' relevant features where $N' < N$. The selection of a subset F allows maximizing the performance of the classification by constructing learning models.

2.1 The Heuristic Variable Selection

Let's consider that a multilayer perceptron (MLP) [14] is neural networks architecture defined by $A(I, H, O)$ where I input layer, H hidden layers and O output layer, and W weight matrix. The value of w_{ij} connection between two neurons j and i reflects the importance of their relationship. This value can be positive or negative depending on if the connection is excitatory (+) or inhibitory (-). Yacoub et al proposed a method for variable selection named heuristic variable selection HVS [15]. The HVS criterion is interested in the strength of these connections. This strength is quantified by $|w_{ij}|$. The partial contribution $\pi(i, j)$ of the hidden neuron j on the output i is given by the proportion of all the connection strength arriving to neuron i see Figure 1.

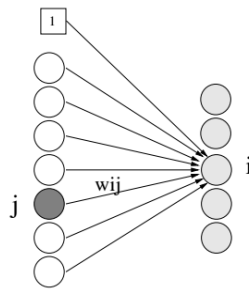


Figure 1. The partial contribution of the neuron j connect with i

$$\pi_{i,j} = \frac{|w_{i,j}|}{\sum_j^N |w_{i,j}|} \quad (1)$$

For estimate the relative contribution of unit j is final decision of the system. The unit j sends connections to, a set of units (i) with partial contributions $\pi_{i,j}$ Figure 2.

$$C_j = \sum_i^N \pi_{i,j} \delta_i \quad (2)$$

Where

$$\delta_i = \begin{cases} 1 & \text{if unit } i \in O \\ C_i & \text{if unit } i \in H \end{cases} \quad (3)$$

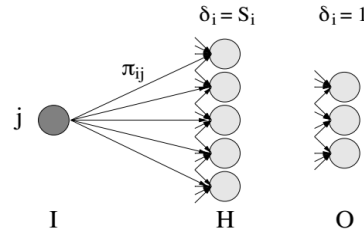


Figure 2. The relative contribution of the neuron j

The selection algorithm based on HVS criterion as follows:

- Learning the MLP until we reach a local minimum.
- Calculate the relevance of each variable according to 2.
- Sort variables in ascending order of relevance.
- Remove the lower variable importance
- Back to 1) to the last variable

2.2. Minimum Redundancy Maximum Relevance

The MRMR (Minimum Redundancy Maximum Relevance) method [16] selects variables that have the maximally relevance with the target class and which are also minimally redundant. In this work, to find a maximally relevant and minimally redundant set of variables, we use mutual information based MRMR criterion. The calculation of redundancy and relevance of a variable is given by equations (4) and (5). The $I(i, Y)$ is the mutual information between class labels y and variable i . This allows us to quantify the relevance of variable i to the classification. The relevance of variable is given by:

$$Rl_i = \frac{1}{|S|^2} \sum_y I(i, Y) \quad (4)$$

Where

$$I(i, Y) = \sum_{i \in S} \sum_{y \in Y} p(i, y) \log \left(\frac{p(i, y)}{p(i)p(y)} \right) \quad (5)$$

The redundancy of a variable subset is determined by the mutual information among the variables. The redundancy of variable i with the other variables is given by:

$$Rd_i = \frac{1}{|S|^2} \sum_{i \in S} I(i, j) \quad (6)$$

Where S and $|S|$ respectively denote the set of variables and its size and $I(i, j)$ is the mutual information between i and j . The score of a variable is the combination of these two factors:

$$SC_i = \frac{Rl_i}{Rd_i} \quad (7)$$

The measures of relevance and redundancy of variables can be formed in several ways, but the quotient of the relevance by redundancy select highly relevant variables with less redundancy [17]. After this individual variable evaluation, a sequential search technique is used with a classifier to select the final subset of variables. A classifier is used to evaluate the subsets starting with the variable that has the best score, the best two, until we find the subset that minimizes the classification error.

3. RESEARCH METHOD

Using the filter methods alone for example MRMR, may not give the best performance because it operates independently the classifier and is not involved in the selection of variables. On the other hand, HVS does not take into account the redundancy among variables. Our objective is to improve the variables selection HVS by introducing an MRMR filter to minimize the redundancy among relevant variables. As seen later, this improves the performance of classifier by compromising relevancy and redundancy of variables.

In our approach of HVS-MRMR variables selection, the variables are selected by a convex combination of the relevancy given by HVS contributions and the MRMR criterion. For i the variable, the ranking Measure R_i is given by

$$R_i = \alpha|C_i| + (1 - \alpha)Sc_i \quad (8)$$

Where the parameter $\alpha \in [0,1]$ determines the compromise between HVS and MRMR criterion,

The search strategy is one of the properties of the variable selection algorithms. There are three strategies, forward selection, backward elimination and stepwise selection. In forward selection, variables are progressively incorporated into larger and larger subsets. In backward elimination one starts with the set of all variables and progressively eliminates the least promising ones [3]. In our algorithm we use the strategy backwards elimination. To better compromise with redundancy and Relevancy of variables, we use S(i)MRMR criterion for ranking. Also, we use $|C_i|$ the criterion of HVS as the measure of relevance variables.

Algorithm 1 illustrates HVS-MRMR variables selection method. In each iteration, we identified the least important variable after ranking the variables in the set G . The variable least significant are retired and the remaining subset G will go through process iterative, each time we remove a variable, relearning is required. The algorithm removes the variables one by one until the last variable.

Let's consider that S_p is a subset of variables and p is the number of these variables.

Algorithm 1 : HVS-MRMR for variables selection

```

Begin
  Set  $\alpha$ 
  Given set of variable,  $S \subset G$ 
  Repeat :
    Train the MLP with test dataset;
    For each  $i \in S_p$  do
      Compute the  $C_i$  by equation (2)
      Compute the  $Sc_i$  by equation (7)
      Compute the  $R_i$  by equation (8)
    End for
    Select the variable  $i^* = \arg \min\{R_i\}$ 
    Update  $S_{p-1} = G \setminus \{i^*\}$ 
  Until  $G = \{\emptyset\}$ 
End

```

For selecting a subset S_p of variables, algorithm.1 generates a set of N neuronal networks (N is the number of variables) having less and less variables. The choice of subset includes using a statistical test (Fisher test) and searches among all networks $MLP(p)$ those that are statistically near $MLP(p^*)$. This principle provides a set of neuronal networks such as $E(p^i) \approx E(p^*)$. The subset of selected variables is the smallest subset of p^0 statistically near $MLP(p^*)$ variables. Where $E(p^i)$ is the error of MLP for subset S_p .

Where,

$E(p)$ is the error of MLP for subset S_p ,

$$MLP(p^*) = \arg \min_{MLP(p)} E(p) \quad (9)$$

$$p^0 = \min_i \{i\} \quad (10)$$

4. RESULTS AND ANALYSIS

To evaluate the performance of the HVS-MRMR, we use some data sets for real world classification. Table 1 shows the detailed information of each dataset. They were partitioned into three sets: a training set, a validation set and a testing set. The training and testing sets were used to train MLP and to evaluate the classification accuracy of trained MLP, respectively. The validation set is used to estimate prediction error for MLP.

Table 1 Descriptions and Partitions for Different Classifications Datasets

Dataset	Variables	Number of			Classes
		Training examples	Validation examples	Testing examples	
<i>Diabetes</i>	8	384	192	192	2
<i>Cancer</i>	9	349	175	175	2
<i>Glass</i>	9	108	53	53	6
<i>Vehicle</i>	18	242	211	211	4
<i>Hepatitis</i>	19	77	39	39	2
<i>Waveform</i>	21	2500	1250	1250	3
<i>Horse</i>	21	172	86	86	2
<i>Ionosphere</i>	34	175	88	88	2

4.1 Preprocessing

We preprocessed the datasets by rescaling input variables values between 0 and 1 the linear normalization function: After normalization, all features of these examples are between zero and one. The standardization formula is as follows:

$$x_{i,new} = \frac{x_{i,old} - x_{i,min}}{x_{i,max} - x_{i,min}} \quad (11)$$

Where $x_{i,new}$ and $x_{i,old}$ are the new and old value of attribute, respectively. The $x_{i,max}$ and $x_{i,min}$ are the maximum and minimum value, respectively of variable i .

Table 2. Performance of HVS-MRMR for different classifications datasets, St. dev. given after the \pm sign

Dataset	Measure	Without VS	HVS	MRMR	HVS-MRMR
Diabetes	Mean No. of variable	8.00±0.00	5.90±0.92	6.50±0.63	5.55±1.02
	Acc %	75.68±1.01	76.32±2.06	76.25±2.01	76.63±3.23
Cancer	Mean No. of variable	9.00±0.00	6.83±1.01	5.60±0.92	6.13±0.98
	Acc %	97.96±0.88	98.43±1.06	98.17±1.52	98.68±1.88
Glass	Mean No. of variable	9.00±0.00	5.13±0.55	4.90±0.45	4.33±0.65
	Acc %	74.21±5.62	76.61±4.66	76.73±5.20	77.03±4.95
Vehicle	Mean No. of variable	18.00±0.00	4.51±0.67	5.33±0.64	4.90±0.76
	Acc %	73.37±4.87	74.35±3.57	74.21±4.03	75.17±3.47
Hepatitis	Mean No. of variable	19.00±0.00	3.73±0.87	5.10±0.76	3.55±0.92
	Acc %	70.63±3.60	77.65±2.79	76.32±3.43	78.67±3.15
Waveform	Mean No. of variable	21.00±0.00	4.85±0.96	5.50±0.81	4.85±1.04
	Acc %	85.30±1.23	85.91±2.54	85.27±2.98	84.91±3.13
Horse	Mean No. of variable	21.00±0.00	7.94±1.87	6.93±1.65	6.55±1.45
	Acc %	84.51±2.52	86.03±2.10	85.27±2.05	86.21±1.95
Ionosphere	Mean No. of variable	34.00±0.00	6.52±2.03	7.11±1.87	6.87±1.97
	Acc %	94.66±2.03	95.81±1.95	96.27±2.25	96.31±2.98

Where

Mean No. of variable : the average numbers of variables selected

Acc: The classification accuracy

Without VS: Without variable selection

4.2 Parameter Estimation

In this paragraph, we have specified some parameters of HVS-MRMR. These are described as follows: The initial weights for an MLP were randomly chosen in the value between -1 and 1. The training error threshold value for cancer, diabetes, glass, hepatitis, horse, ionosphere, vehicle, and waveform datasets was set to 0.002, 0.003, 0.04, 0.02, 0.003, 0.02, 0.01 and 0.03 respectively. Also, the validation error threshold value was set to 0.001, 0.002, 0.025, 0.018, 0.001, 0.014, 0.007 and 0.025 for cancer, diabetes, glass, hepatitis, horse, ionosphere, vehicle, and waveform datasets, respectively. The learning rate and initial weight values are the parameters of the well known back-propagation algorithm [18-19]. From the suggestions of many previous works [20-21] and after some preliminary tests these values were set. The α value for the HVS-MRMR variable selection was then determined empirically from the set {0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8} based on the best tenfold cross-validation performance.

4.3 Results

Table 2 shows the results of HVS-MRMR, MRMR and HVS over 20 independent runs on eight classification datasets. The classification accuracy (Acc) in Table 2 refers to the percentage of classifications accuracy produced by trained MLP on the testing set of a classification dataset and (Mean No. of variable) presents the average numbers of variables selected. It can be observed from table 2 that HVS-MRMR was selected a smaller number of variables for different benchmark datasets. For example, HVS-MRMR selected on average 6.13 variables from a set of 9 variables for solve the cancer dataset. It also selected on average 6.87 variables from a set of 34 variables for solve the Ionosphere dataset. In fact, HVS-MRMR selected a small number of variables for each datasets with more variables.

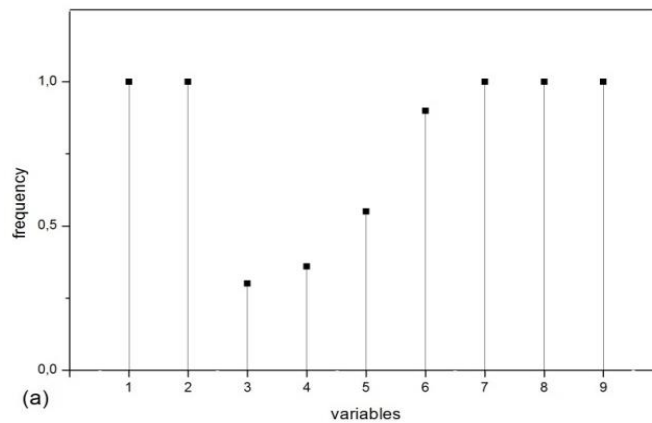


Figure 3 shows the frequency of selected by HVS-MRMR for Cancer dataset

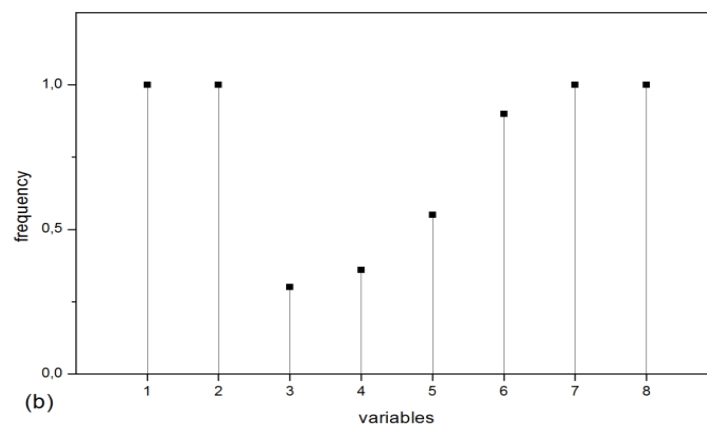


Figure 4 shows the frequency of selected by HVS-MRMR for Diabetes dataset

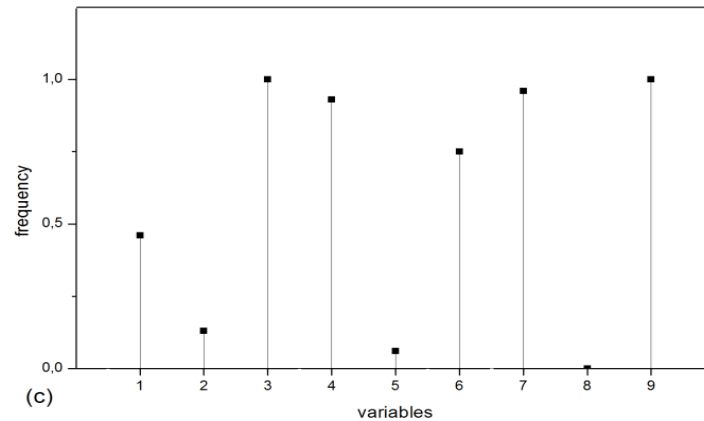


Figure 5 shows the frequency of selected by HVS-MRMR for Glass dataset

In order to determine the essence of selected variables, we measured the frequency of variables. The frequency of a variable i can be defined as

$$f_i = \frac{H_i}{T} \quad (12)$$

Where H_i is the number of times a particular variable is selected in all test and T is the total number of test.

Figure 3, Figure 4 and Figure 5 show the frequency of variables selected for diabetes, cancer, and glass datasets, respectively. It can be seen in Figure 3 that HVS-MRMR selected features 1, 2, 6, 7, 8 and 9 of the Cancer dataset vary frequently. The frequency of selection for these variables is one or nearly one.

Table 3. Comparison among HVS-MRMR, HVS, MRMR ADHOC, GPSFSCD, EIR-MLPFS and ANNIGMA-WRAPPER for the cancer, diabetes, glass, hepatitis, horse, ionosphere, vehicle, and waveform datasets

Dataset	HVS	MRMR	ADHOC	GPSFSCD	EIR-MLPFS	ANNIGMA-WRAPPER	HVS-MRMR
Diabetes	76.32	76.25	71.2	–	–	77.8	76.63
Cancer	98.43	98.17	–	96.84	89.40	96.5	98.68
Glass	76.61	76.73	70.5	–	44.10	–	77.03
Vehicle	74.35	74.21	69.6	78.45	74.60	–	75.17
Hepatitis	77.65	76.32	–	–	–	–	78.6
Waveform	85.91	85.27	–	–	–	–	84.91
Horse	86.03	85.27	–	–	–	–	86.21
Ionosphere	95.81	96.27	–	–	90.60	90.2	96.31

“–” means not available

We can be observed that our method achieved the best classification accuracy among all other algorithms for five out (Cancer, Glass ,out, Hepatitis, Horse and Ionosphere) of eight datasets. For the remaining three datasets, HVS-MRMR achieved as a second best. while HVS (Waveform), ANNIGMA-WRAPPER(Diabetes) and GPSFSCD(Vehicle) achieved the best classification accuracy for one dataset each.

We can be said that the variables selection increases the classification accuracy by ignoring the irrelevant variables from the original feature set. The variables selection is an important task in such a process is to select necessary information (irrelevant variables). Otherwise, the performance of classifiers might be decreased.

The efficacy of embedding of MRMR filter in HVS was evidenced by improved classification performance on benchmark datasets. In this paper, the proposed algorithm outperformed other methods in the classification on the most database tested, it was able to select the relevance variables among datasets. We can be choose from these data with the analysis of performance of the subset of variables that have strong relationship with the classification. However in terms of execution time, our proposed approach consumes more than HVS and MRMR.

5. CONCLUSION

It is very important to remove the redundant and irrelevant variables in data before applying some data mining techniques to analyze the data sets. In our research, we suggest a new method of variables selection based on HVS criterion and MRMR criterion, called the HVS-MRMR, to integrate the procedures of variable selection filter and variable selection wrapper to improve the performance of classification.

We applied HVS-MRMR for four classification problems. The experiment results show that HVS-MRMR variables selection selected a less number of variables with high classification accuracy compared to MRMR, HVS.

In a forthcoming research work, we intend to improve this approach to find better subset of variables selected and to improve classification accuracy [26-27].

ACKNOWLEDGEMENTS

We express our thanks to the associate editor and anonymous referees whose valuable suggestions helped to improve this work significantly and also to their colleague.

REFERENCES

- [1] Hinton Geoffrey E, Salakhutdinov Ruslan R. "Reducing the Dimensionality of Data with Neural Networks." *Science*, 2006; 313(5786); p. 504-507.
- [2] Demers David, Cottrell G W. "n-linear dimensionality reduction." In: *Advances in neural information processing systems*, 1993; 5; 580-587.
- [3] Fodor, Imola K. "A Survey of Dimension Reduction Techniques." Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, 2002; 9; 1-18.
- [4] Wei, Hua-Liang Et Billings, Stephen A. "Feature Subset Selection and Ranking for Data Dimensionality Reduction." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007; 29(1); 162-166.
- [5] Guyon Isabelle, Elisseeff André. "An Introduction to Variable and Feature Selection." *Journal of Machine Learning Research*. 2003; 3(3); 1157-1182.
- [6] Liuhan, Motoda Hiroshi, YU Lei. "Feature Selection with Selective Sampling." In: Proceedings of the Nineteenth International Conference on Machine Learning, *ICML*. 2002; 395-402.
- [7] Dy Jennifer G, BRODLEY Carla E. "Feature Selection for Unsupervised Learning." *Journal of machine learning research*. 2004; 5; 845-889.
- [8] Kannan S, Senthamarai et Ramaraj N A. "Novel hybrid Feature Selection via Symmetrical Uncertainty Ranking Based Local Memetic Search Algorithm." *Knowledge-Based Systems*, 2010; 23(6); 580-585 .
- [9] Peng Hanchuan, Lon Fuhui, Ding Chris. "Feature Selection Based on Mutual Information Criteria of Max-dependency, Max-relevance, and Min-redundancy." *IEEE Transactions on pattern analysis and machine intelligence*. 2005; 27(8); 1226-1238.
- [10] Kohavi Ron, John George H. "Wrappers for Feature Subset Selection." *Artificial intelligence*, 1997; 97(1-2), 273-324.
- [11] Chormunge Smita, Jena Sudarson. "Efficient Feature Subset Selection Algorithm for High Dimensional Data." *International Journal of Electrical and Computer Engineering*. 2016; 6 (4); 1880.
- [12] Liu Chuan, Wang, Wenyong, Zhao Qiang, et al. "A New Feature Selection Method Based on a Validity Index of Feature Subset." *Pattern Recognition Letters*, 2017.
- [13] Sela Enny Itje, Hartati Sri, Harjoko Agus, et al. "Feature Selection of the Combination of Porous Trabecular with Anthropometric Features for Osteoporosis Screening." *International Journal of Electrical and Computer Engineering*. 2015; 5(1) 78.
- [14] Rosenblatt Frank. "Principles of Neurodynamics. Perceptrons and the Theory of Brain Mechanisms." Cornell Aeronautical Laboratory Inc Buffalony, 1961.
- [15] Yacoub Meziane, Bennani Y. "HVS: A Heuristic for Variable Selection in Multilayer Artificial Neural Network Classifier." *Intelligent Engineering Systems through Artificial Neural Networks*, St. Louis, Missouri. 1997; 7; 527-532.
- [16] Peng Hanchuan, Long Fuhui, Ding Chris. "Feature Selection Based on Mutual Information Criteria of Max-dependency, Max-relevance, and Min-redundancy." *IEEE Transactions on pattern analysis and machine intelligence*. 2005; 27(8); 1226-1238.

- [17] Ding Chris, Peng Hanchuan. "Minimum Redundancy Feature Selection from Microarray Gene Expression Data." *Journal of bioinformatics and computational biology*, 2005; 3(02); 185-205.
- [18] Rumelhart D E, McClelland J L, PDP Research Group. "Parallel Distributed Processing." *IEEE*. 988; 1; 443-453.
- [19] Yang, Yang, Hu, Jun, et Zhang, Mu. "Predictions on the Development Dimensions of Provincial Tourism Discipline Based on the Artificial Neural Network BP Model." *Bulletin of Electrical Engineering and Informatics*, 2014; 3(2); 69-76.
- [20] Islam, Md M., Yao, Xin, et Murase, Kazuyuki. "A Constructive Algorithm for Training Cooperative Neural Network Ensembles." *IEEE Transactions on neural networks*, 2003; 14(4); 820-834.
- [21] Prechelt Lutz, et al. "Proben1: A Set of Neural Network Benchmark Problems and Benchmarking Rules." 1994.
- [22] Richeldi Marco, Lanzi Pier Luca. "ADHOC: A tool for performing effective feature selection. In : *Tools with Artificial Intelligence*." Proceedings Eighth IEEE International Conference on. *IEEE*. 1996; 102-105.
- [23] Muni Durga Prasad, PAL Nikhil R, DAS Jyotirmay. "Genetic Programming for Simultaneous Feature Selection and Classifier Design." *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*. 2006; 36(1); 106-117.
- [24] Gasca E, Sánchez J S, Alonso R. "Eliminating Redundancy and Irrelevance Using a New MLP-based Feature Selection Method." *Pattern Recognition*. 2006; 39(2), 313-315.
- [25] Hsu Chun-Nan, Huang Hung-Ju, Dietrich S. "The Annigma-wrapper Approach to Fast Feature Selection for Neural Nets." *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*. 2002 ; 32(2) ; 207-212.
- [26] Ben-Hdech A, Ghanou Y, Abderrahim E. L. "*Robust Multi-combination Feature Selection for Microarray Data*." *Advances in Information Technology: Theory and Application*, 2016.
- [27] Ghanou, Y., & Bencheikh, G. "Architecture Optimization and Training for the Multilayer Perceptron Using Ant System." *IAENG International Journal of Computer Science*. 2016; 28; 10.