

# Classification Of Emg Signals Using Decision Tree Methods

Selami Keleş, Abdulhamit Subaşı

E-mail:[keles\\_selami@yahoo.com](mailto:keles_selami@yahoo.com), [asubasi@ibu.edu.ba](mailto:asubasi@ibu.edu.ba)

## Abstract

Nowadays, Usage of EMG signals are increasing very fast among the Medical Professionals to determine specific disorders. Recent Computational Intelligence studies show that EMG signals can be processed by machine learning methods. The aim of our study is to implement an accurate system to classify EMG signals using decision tree algorithms. We preprocessed the EMG signals and used autoregressive method (AR) for feature selection. Features are reduced by different filtering methods and applied to decision tree classification algorithms, namely Simple CART, C4.5, Random Forest and Random Tree. EMG signals are classified as Myopathy, Neuropathy and Normal. All the data are compared each other on the table try to find out the best classification and feature reduction methods. While tree algorithms classify the data with the accuracy between %89, 82 and %99, 25, feature reduction slightly affects the accuracy of the classification methods. It has been shown that a successful automatic diagnostic system implemented to classify EMG signals by using decision tree algorithms. Furthermore, future reduction may help to increase the accuracy of the system.

**Keywords:** EMG, Neuropathy, Myopathy, Simple CART, C4.5, Random Tree, Random Forest, Feature reduction.

## 1.INTRODUCTION

Early and accurate diagnosis is important for neuromuscular diseases that help the patient to get full recovery or have better health after therapy. Sometimes, clinical examination is not enough to diagnose and to find the location of disorders [1]. Therefore, it has high importance to find correct location of the disorders to accurate diagnosis and therapy. EMG recordings are more useful than clinical examination to find out the muscle fibers involved in a disorders and abnormal sensory nerve conduction. It allows the clinician to diagnosis without needing a muscle biopsy and raises the clinician response time and helps to treat some disorders.

The analysis of EMG signals can be done only by qualified and professional neurologist. The problem is that, there are few professionals to interpret the EMG waveforms and use the necessary techniques. Therefore, it is important to develop an automated diagnostic system by using EMG signals. The application of Computational Intelligence (CI) techniques can be

used to develop an automated diagnostic system that detects and classifies the neuromuscular diseases by processing EMG signals which helps the neurologists to diagnose the neuromuscular anomalies.

The MUP assessment may not be satisfactory to detect small deviation or miscellaneous patterns of abnormalities [1]. Therefore, to design an accurate automatic EMG signal classification system, different EMG analysis algorithms have been developed [2, 3]

To develop an intelligent diagnostic system, first, EMG signals have to be pre-processed and extracted the characteristic information. Then, extracted features that contain the time and frequency domain information, processes by using wavelet coefficients, Fourier coefficients, autoregressive coefficients or other signal processing techniques. After all, processed information can be used as input to the classifier such as NNs, SVM or Decision Tree to classify the disease.

One of the most popular Machine Learning Method ANN has been widely used to classify the EMG data. In order to increase the classification success, ANN can combine the best of both time and frequency domain measures, but it is not enough for clinical use [4, 5]. Christodoulou and Pattichis used Self Organized Feature Maps and Learning Vector Quantization used to classify MUP's [5]. Genetic algorithms were used by Schizas ve Pattichis to classify the EMG signals [6]. Multilayer Perceptron Neural Networks (MLPNN), Dynamic Fuzzy Neural Network (DFNN) and Adaptive Neuro-Fuzzy Inference System (ANFIS) based classifiers were compared by Subaşı. ANFIS model has reported more successful than others with the accuracy of 95%. [7]. SVM classifier is used by Katsis et al. and the classify the EMG signals with the correct identification rates of 93, 95 and 92% for normal, myopathy and neuropathy, respectively [8]. The result of another comparison research between Combined Neural Network (CNN) and Feedforward Error Backpropagation ANN (FEBANN) classifiers was described by Bozkurt. Even the CNN didn't provide the fast enough classification; it gave slightly higher success than the FEBANN with the accuracy of 92% [9]

There are still challenges to develop an accurate and practical automated system. EMG signals vary patient to patient in a very large range. Signal amplitude and duration changes by patient age. This problem can be solved by designing a signal processing techniques that conserve or capture distinctive information in raw EMG readings. High-quality set of features [10].

## **2.EMG**

EMG can be defined as a method of analyzing neuromuscular conditions depends on cell action potentials for the duration of muscle action. The specification of the EMG signal is 0.01-10mV and 10-2000Hz on average. This signal has information about location, reason of disorder and type of illness. For example, while the EMG pulse duration shows the location and metabolic condition of the muscle [11], odd spikes may point to the myopathy.

Electromyograph records the Motor Unit Action Potential (MUAP). EMG can be categorized into needle or fine wire EMG and surface (sEMG). While EMG signals are recording, some instruments are required including, electrodes, a signal acquisition system and signal filters. Generally, EMG instruments are produced with typical settings for signal characteristics such as filter bandwidth, gain and input impedance [12].

The needle electrode or wire electrode can reach the individual motor unit and get the action potential more accurately than the surface EMG. Surface EMG electrode is more useful than needle or wire electrodes, because it is used by attaching the body instead of inserting anything in it. EMG signals are recorded at hospital lab by Electromyographers [10].

### 3. Myopathy

Myopathy is a muscle disorder especially skeletal muscle, which is caused by several reasons such as injury of muscle group or some genetic mutation. It obstructs the proper tasks of muscle fibers. The patient suffering with myopathy has weak muscle and has difficulties to perform regular tasks. Depending on the severity of disease, sometimes it is impossible to make any movement by using affected muscle. There are a number of types of myopathy including; Muscular dystrophy, Congenital muscular dystrophy, Duchenne muscular dystrophy, Becker muscular dystrophy, Emery–Dreifuss muscular dystrophy, Myotonic muscular dystrophy, Distal muscular dystrophy, limb–girdle muscular dystrophy, facioscapulohumeral muscular dystrophy and oculopharyngeal muscular dystrophy [10]

#### Neuropathy

Simply, Neuropathy is the term for describing damage to nerves of nerves system. It causes pain and some disability. Neuropathy can be caused by variety of precipitating factors including infection, diabetes; alcohol abuse, cancer chemotherapy and injury. When a single nerve is affected, it is called Mono-neuropathy. When a group of nerves or all nerves of peripheral nerve are affected, it is called Polyneuropathy. Poly neuropathies are similar because of inadequate manner in which sensory nerves react to malfunction. EMG diagnosis is not considerably useful for Polyneuropathy, because the patients with polyneuropathy have normal electrophysiological characteristics [10]

#### Decision Tree Classifiers;

The Decision Tree is a classification algorithm that classifies a pattern by asking questions, in which the next question asked depends on the answer to the present question [13]. It uses a “divide-and-conquer” approach to solve the learning problems [14]. Decision Tree learning methods are one of the most popular inductive inference algorithms and

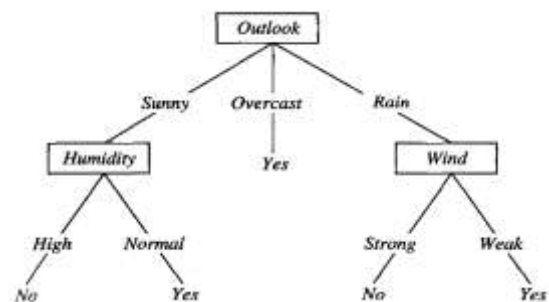


Figure-1

have been used a wide range of task about medical diagnosis [15].

The instances are classified by sorting them down the tree from root to some leaf node which the classification is provided in decision tree algorithms. The attributes of the instance are tested at each node and sent to the sub node or leaf node from one of the branch which correspond the possible values of that attribute [15]. The numeric attributes are tested by comparing a pre-defined constant value at the node and it gives two or three-way split depends on the several different possibilities. [14]. Trained trees can be shown by a set of if-then rules to increase human readability [15]. An example of three is shown in the figure-1 which is adopted from Quinlan research [16].

#### **4.C4.5**

C4.5 is develop by Ross Quinlan [17] to make complex decision trees more understandable by using a list of rules of the form “If X and Y and Z and ....then class A” where rules are grouped together for each class. When the first rule is found which satisfies the condition of case, the instance is classified. If there is no rule which is satisfied by the case, it is sent to default class. The basic disadvantage of the C4.5 algorithm is requirement of high amount of CPU time and system memory[18].

#### **5.Random Tree**

Random Decision Tree is a randomly trained ensemble of decision trees which is proposed by Fan et al. [19]. The features are randomly selected at each node, while training trees phase is proceeding. A selected discrete feature never selected again till it is vain to use the same discrete feature more than once. Conversely, it is possible to choose continues features several times as long as every time, using randomly selected splitting value. Each tree gives raw posterior probabilities at the classification phase and outputs of each tree in the ensemble are averaged for last posterior profanities estimation. It is proofed that the Random Decision Tree is highly accurate classification method for both 0-1 loss and cost-sensitive loss function. [20].

#### **6.Random Forest**

Random forest is a tree algorithm which composed of a number of tree predictors. In this algorithm, each tree is shown by a random vector which is independently taken from the same distribution in the forest. As the number of the tree increase in the forest, the generalization error converges to a limit. The strength of the individual tree and relationship between the trees affects the generalization error. Once all trees in the forest produce a result, they are voted for the most passible class [21]. It is one of the most successfulclassification

methods among the available algorithms for many data type [22], but opposite to other decision tree methods, it makes classification which is difficult to deduce by human [23].

## **7. Classification and Regression Trees(CART)**

Classification and Regression Trees (CART), proposed by Breiman et al., [24], was a revolutionary improvement of Machine Learning and Data Mining fields which can be used almost any domain such as electrical engineering, biology and medical researches. It is a binary repeated division process which can work with the nominal and continuous data.

The raw form of data is processed without requiring binning. The growing trees aren't halted by using stopping rules till it reaches maximum size and then clipped back to the root by cost-complexity pruning method. The pruned next split contributes the overall performance of the tree. The CART algorithm is projected to grow a sequence of nested pruned trees that all of them are nominee of the optimal trees. To find out the "right sized" tree, the predictive success of every tree is evaluated at the pruning process. The performance of the tree is measured by test data or cross validation method and tree is selected after evolution, because CART doesn't have any internal performance measurement method depending on training data. [18]

## **8. AR model**

An Autoregressive (AR) model is used to estimate the different kinds of natural fact in signal processing and statistic fields which were originally proposed by Yule. It contains a set of linear estimation formulas which is used to predict the output of a system depends on the previous output. [25, 26]

There are a number of methods to estimate the AR model parameters. Some of them are the Yule-Walker, Burg(1968), covariance and modified covariance methods. It is easy to access and use these methods in many software packages such as MATLAB (<http://www.mathworks.com/products/matlab/>) and Signal Processing Toolbox.

The Yule-Walker technique is based on a partial form of the autocorrelation approximate to guarantee a positive semi defined autocorrelation matrix. Alternatively, the Burg method uses a form of order-recursive least square method which approximates the parameters by minimizing errors of the linear system. [10]

## **9. Feature selection algorithms**

An important issue is handling irrelevant features in pattern recognition field. Feature Selection (FS) method is necessary to find out the important features to classify the data accurately, because it was not considered how to overcome a large amount of irrelevant feature in many pattern recognition methods, while they were designing. [27,28,29] Mostly, the feature selection methods are used to increase the model performance, to abstain the

overfitting, to get faster and more cost effective models and to understand the processes which produce the data. Beside the advantages, FS methods add new complexity layer to the models. [30], searching the optimal subset of relevant features. FS methods can be grouped in to three categories by the way of allaying relevant features search with building classification model; filter methods, wrapper methods and embedded methods. [31]

## **10. Materials and Methods**

### **10.1. Subjects and Data Acquisition**

The patients which samples are taken from and the control group were chosen at Neurology Department of University of Gaziantep. Measurements are taken by an EMG system (Keypoint; Medtronic Functional Diagnostics, Skovlunde, Denmark) with standard settings. The signal was obtained from biceps brachii muscle by using a concentric needle electrode (0.45 mm diameter with a recording surface area 0.07 mm<sup>2</sup>; impedance at 20 Hz below 200 K $\Omega$ ). 5 Hz to 10 KHz band-pass filter was applied to the raw signal and sampled at 20 KHz for 5 s with 12-bit resolution. Then 8 KHz low-passed filter was applied.

The signals are recorded from three to five different points in muscle for standardization. And also the needles are inserted in to muscle until it reaches the medial or posterior border of the muscle (at least 3-5 mm deep). The needles are moved 3-5 mm to ensure to record different MUPs at every recoding season.

The signals were taken from the biceps brachii muscle of the patients under isometric condition at just about 30% of Maximum Voluntary Contraction (MVC). Before the patient diagnosis, general examination and clinical history of the patient were considered and EMG and nerve conduction tests were regarded. Unless, the EMG diagnosis results were uncertain and some other clinical reason; the muscle biopsies were not done.

The data which was used for this study were collected from 27 different subjects and analyzed. Details about the subject are given below as in [3]

- 7 healthy subjects, (3 males, 4 females,) ages between 10 to 43 years (mean age $\pm$ standard deviation (S.D.): 30.2 $\pm$ 10.8 years)
- 7 myopathic subjects (4 males, 3 females) ages between 7 to 46 years, (mean age $\pm$ standard deviation (S.D.): 21.5 $\pm$ 13.3 years)
- 13 neuropathic subjects (8 males, 5 females) ages between 7 to 55 years, (mean age $\pm$ standard deviation (S.D.): 25.1 $\pm$ 17.2 years)

We used the dataset which is recorded, preprocessed and features are extracted by Subaşı (2006) for his research namely “Classification of EMG Signals Using Combined Features and Soft computing” in this study.

## 10.2.Data set

The dataset has 129 features which were extracted by AR model from recorded EMG signals and contains three classes which are “Normal”, “Neuropathy” and “Myopathy”. As shown in the table-2

Class	Number of instance
Normal	400
Neuropathy	399
Myopathy	400
Total	1199

Table-2

## 11.WEKA

WEKA is open source software issued under the GNU General Public License which contains machine learning algorithms for data mining tasks. It is developed for contributing to a theoretical framework for the field by Machine Learning Group at University of Waikato, New Zealand. It composed of easy to use tools which can be applied directly to the dataset. Data pre-processing, classification, regression, clustering, association rules, and visualization are tools in WEKA. And also, well known classification algorithms such as Neural Network, Bayesian, SVM and Decision Tree are available in this tool. It can either get the data from a database or a file. The file format “.arff” and “.cvs” are supported by WEKA. [32]

### 11.1.Experiments

The data mining tool WEKA was used for both feature selection and classification tasks.10-fold-Cross validation method was used to train and test the classifiers. In 10-fold cross-validation, the original sample is randomly partitioned into 10 subsamples. One of the subsample is reserved as the validation data for testing the model, and the residual 9 subsamples are used as training data. The cross-validation process is then repeated 10 times, with each of the 10 subsamples used exactly once as the validation and training data. [33]

The data set weretested by four Decision Tree algorithms which are C4.5, Random Tree, Random Forest and Simple CART and the results wererecorded on a table. Then, the Feature Selection methods wereapplied to the data set to determine non effective or comparably less effective features and ineffectual featureswereremoved from data set. The new data set

wastested by four Decision Tree algorithms and the results wererecorded on a table again. This process wasrepeated whit elevendifferent feature selection methods which are listed below. Totally, 48 different testswere done for this study and the total accuracy of each test wasrecorded on a table (Table-3).

The tested Feature Selection methods:

- Information Gain,
- Chi Squared Attribute Evaluator
- Filtered Attribute Evaluator
- Consistency Subset Evaluator
- Filtered Subset Evaluator
- Gain Ratio Attribute Evaluator
- One-R Attribute Evaluator
- Principal Components
- Relief Attribute Evaluator
- SVM Attribute Evaluator
- Symmetrical uncertainty Attribute Evaluator

**11.2.Results**

	All Features (No Reduction)	Information Gain	Chi Squared Attribute Evaluator	Filtered Attribute Evaluator	Consistency Subset Evaluator	Filtered Subset Evaluator	Gain Ratio Attribute Evaluator	One R Attribute Evaluator	Principal Components	Relief F Attribute Evaluator	SVM Attribute Evaluator	Symmetrical uncertainty Attribute Evaluator	Average	Max	Min
j48,( C4.5):	96,33	96,25	96,58	96,16	96,25	97,08	96,41	96,25	91,99	96,50	96,41	96,33	96,05	97,08	91,99
Random Forest	98,50	98,67	98,83	99,17	99,25	98,67	98,92	98,92	93,58	98,83	98,50	99,00	98,40	99,25	93,58
Random Tree	96,66	97,16	95,50	97,33	96,00	97,50	97,08	96,91	89,82	97,25	96,75	96,25	96,18	97,50	89,82
Simple CART	96,50	96,41	96,50	96,41	96,58	96,66	96,50	96,41	91,41	96,41	96,58	96,50	96,07	96,66	91,41
Average	97,00	97,12	96,85	97,27	97,02	97,48	97,23	97,12	91,70	97,25	97,06	97,02	96,68	97,62	91,70
Max	98,50	98,67	98,83	99,17	99,25	98,67	98,92	98,92	93,58	98,83	98,50	99,00	98,40	99,25	93,58
Min	96,33	96,25	95,50	96,16	96,00	96,66	96,41	96,25	89,82	96,41	96,41	96,25	96,05	96,66	89,82



Table-3

The accuracy of the classifier varies from %89.82 to %99.25. The most successful algorithm is Random Forest which can classify the data with %99.25 accuracy by using feature selection method “Consistency Subset Evaluator”. The Classification algorithms C4.5, Random Tree and Simple CART classify the data with the similar accuracy, between %96.33 and %96.50.

Reducing the features by using feature selection methods does not considerably affect the accuracy of the classification algorithms except the “Principal Component”. Principal Component decreases the classification success of all the algorithms which we test. Using feature selection “Filtered Subset Evaluator” increases the success of C4.5, Random Tree and Simple CART classification algorithms, but not considerable, less than %1.

Statistics information for Random Forest with the feature selection method Consistency Subset Evaluator

Correctly Classified Instances: 1190- 99.2494 %

Incorrectly Classified Instances:9 - 0.7506 %

Kappa statistic :0.9887

Mean absolute error : 0.023

Root means squared error : 0.087

Relative absolute error : 5.1793 %

Root relative squared error : 18.4468 %

Total Number of Instances : 1199

### Detailed Accuracy by Class

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.990	0.005	0.990	0.990	0.990	0.998	Normal
	0.995	0.003	0.995	0.995	0.995	0.998	Myopathy
	0.992	0.004	0.992	0.992	0.992	0.999	Neuropathy
Weighted Average	0.992	0.004	0.992	0.992	0.992	0.999	

### Confusion Matrix

Classified as →	Normal	Myopathy	Neuropathy	Accuracy
Normal	396	2	2	%99.00
Myopathy	1	398	1	%99.50
Neuropathy	3	0	396	%99.25

Confusion Matrix shows that none of Neuropathy classified as Myopathy and the Myopathy is classified with the maximum accuracy (%99.50) among the 3 classes.

### 11.3. Discussion

Our study shows that it is possible to implement an accurate automatic diagnostic system to classify the EMG signals as Myopathy, Neuropathy and Normal by using Decision Tree algorithms. All the Decision Tree based classification algorithms which we analysed in this study can be used as classifier for creating such a system, but we recommended using Random Forest, as classifier and “Consistency Subset Evaluator” among feature selection

methods for reducing the features. The performance of this system gives the maximum accuracy (%99.25) among the others. The other Decision Tree based Classifiers C4.5, Random Tree and Simple CART may be used without feature reduction. When the results are compared at the Table-3, feature selection methods enhance the performance less than %1.

Among the feature reduction methods, we don't suggest to use "Principle Component" for selecting effective features, because it noticeably decreases the achievement of all classification methods. It decreases the performance of Random Tree from % 96.66 to %89.82.

## 12. CONCLUSION

This study shows that it is possible to design a high performance automatic diagnostic system by using EMG signals which are taken from 27 different subjects. It is necessary to test our system by using data set which is taken more than 200 subjects.

## REFERENCES

- [1] Preston, D.C., Shapiro B. (2005). *Electromyography and Neuromuscular Disorders: Clinical-electrophysiologic Correlations*. Philadelphia, PA:Elsevier. ISBN-13: 978-0-7506-7492-8
- [2] Richfield E. K., Cohen B. A., Albers J. W., Review of quantitative and automated needle electromyographic analyses, *IEEE Trans Biomed Eng.* 28(7) (1981) 506-514.
- [3] Subasi A., Yilmaz M., Ozcalik H. R., Classification of EMG signals using wavelet neural network, *J. Neuroscience Meth.* 156 (2006) 360–367.
- [4] Hassoun M. H., Wang C., Spitzer A. R., NNERVE: Neural network extraction of repetitive vectors for electromyography—Part I: Algorithm, *IEEE Trans. Biomed. Eng.* 41 (1994) 1039–1052.
- [5] CHRISTODOULOU C.I., PATTICHIS C.S., "Unsupervised pattern recognition for the classification of EMG signals", *IEEE Trans Biomed Eng* 1999, 46(2):169–78
- [6] SCHIZAS C.N., PATTICHIS C.S., "Learning systems in biosignal analysis", *BioSystems* 1997, 41:105–25.
- [7] Subaşı. A, Classification of EMG Signals Using Combined Features and Soft Computing Techniques, *Applied Soft Computing*, Elsevier 2012, G Model ASOC-1517; No. of Pages 11
- [8] Katsis C.D., Goletsis Y., Likas A., Fotiadis D.I., Sarmas I. A novel method for automated EMG decomposition and MUAP classification, 2005 Elsevier B.V *Artificial Intelligence in Medicine* (2006) 37, 55—64

- [9] BOZKURT M. R. Emg İşaretlerinin Modern Öntemlerle Önişlenmesi Ve Sınıflandırılması. T.C. Sakarya Üniversitesi Fen Bilimleri Enstitüsü Doktora Tezi 2007
- [10] Begg R., Lai D. T. H., Palaniswami M., Computational Intelligence in Biomedical, EngineeringCRC Press, Boca Raton Florida, 2008
- [11] Emly, M., L. Gilmore, and S. Roy (1992). Electromyography. IEEE Potentials 11(2), 25–28.
- [12] Preedy, V. and T. Peters (2002). Skeletal Muscle: Pathology, Diagnosis and Management of Disease. U.K.: Cambridge University Press.
- [13] Duda R.O. “Pattern Classification” Richard O., Peter E. Hart and David G. -2nd ed. ISBN: 0-471-05669-3,394, 2001, Stork. John Wiley & Sons Inc.
- [14] Witten I. H., Frank E., Hall M. A., “Data Mining Practical Machine Learning Tools and Techniques”, Third Edition, 64, 2011 Elsevier Inc.
- [15] Mitchell T. M. “Machine Learning” 52, 53: McGraw-Hill Science/Engineering/Math; (March 1, 1997) ISBN: 0070428077
- [16] Quinlan J.R. Induction of Decision Trees, Machine Learning, Morgan Kaufmann (1), 81-106, 1986
- [17]. Quinlan, J. R. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993
- [18] Wu X., · Kumar V., · Quinlan J. R., · Ghosh J., · Yang Q., · Motoda H. · McLachlan G. J., · Ng A., · Liu B., · Yu P. S., · Zhou Z., · Steinbach M., · Hand D. J., · Steinberg D., Top 10 algorithms in data mining SURVEY PAPER 27-28Springer-Verlag London Limited 2007
- [19]. Fan,W.; Wang, H.; Yu, P. S., Ma, S., Is random model better? On its accuracy and efficiency. In Proceedings of Third IEEE International Conference on Data Mining (ICDM-2003).
- [20]. Fan W. On the Optimality of Probability Estimation by Random Decision Trees IBM T.J. Watson Research, Hawthorne, New York, 2004 American Association for Artificial Intelligence, weifan@us.ibm.com, (www.aaai.org).
- [21] BreimanLeo Random Forests inMachine Learning, 45, 5–32, 2001, Statistics Department, University of California, Berkeley, CA 94720 Kluwer Academic Publishers. 2001
- [22] Caruana, Rich; Karampatziakis, Nikos; Yessenalina, Ainur (2008). "An empirical evaluation of supervised learning in high dimensions". Proceedings of the 25th International Conference on Machine Learning (ICML).
- [23] Berthold, Michael R. (2010). Guide to Intelligent Data Analysis. Springer London.

- [24]. Breiman L, Friedman JH, Olshen RA, Stone CJ Classification and regression trees. Boca Rota, London New York, Washington DC., Chapman&Hall/CRC, 1984.
- [25] [http://en.wikipedia.org/wiki/AR\\_model](http://en.wikipedia.org/wiki/AR_model), 15.04.2012
- [26] NIST/SEMATECH e-Handbook of Statistical Methods, <http://www.itl.nist.gov/div898/handbook/>, 12.04.2012.
- [27] Guyon I., Elisseeff A. An introduction to variable and feature selection. J. Mach Learn Res., 3, 1157–1182. 2003
- [28] Liu,H. and Motoda,H. Feature Selection for Knowledge Discovery and Data Mining. Kluwer Academic Publishers, Norwell, MA. 1998
- [29] Yu,L. and Liu,H. Efficient feature selection via analysis of relevance and redundancy. J. Mach. Learn. Res., 5, 1205–1224. 2004
- [30] Daelemans, Walter, Véronique Hoste, Fien De Meulder and Bart Naudts, 'Combined Optimization of Feature Selection and Algorithm Parameter Interaction in Machine Learning of Language.' Proceedings of the 14th European Conference on Machine Learning (ECML-2003), Lecture Notes in Computer Science 2837, Springer-Verlag, Cavtat-Dubrovnik, Croatia, 84-95, 2003.
- [31] Saeys Y., Inza I. aki and Larranaga P. Areview of feature selection techniques in bioinformatics V. 23 P. 2507–2517 BIOINFORMATICS, Advance Access publication 2007 doi:10.1093/bioinformatics/btm344
- [32] <http://www.cs.waikato.ac.nz/ml/weka/>, 12.04.2012
- [33] McLachlan G.J., Do K.A. Ambrose, C. Analyzing microarray gene expression data. John Wiley & Sons. 2004.