

Improving the Classification Ability of Delegating Classifiers Using Different Supervised Machine Learning Algorithms

Basra Farooq Dar¹, Malik Sajjad Ahmed Nadeem¹, Samina Khalid², Farzana Riaz², Yasir Mahmood², & Ghias Hameed²

¹Dept. of Computer Science and Information Technology, University of AJK, Muzaffarabad, Pakistan

²Dept. of CSIT, Mirpur University of Science and Technology, Mirpur AJK, Pakistan

Correspondence: Samina Khalid, Dept. of CSIT, Mirpur University of Science and Technology, Mirpur AJK, Pakistan.

Received: June 6, 2023

Accepted: August 10, 2023

Online Published: August 23, 2023

doi:10.5539/cis.v16n3p22

URL: <https://doi.org/10.5539/cis.v16n3p22>

Abstract

Cancer Classification & Prediction Is Vitaly Important for Cancer Diagnosis. DNA Microarray Technology Has Provided Genetic Data That Has Facilitated Scientists Perform Cancer Research. Traditional Methods of Classification Have Certain Limitations E.G. Traditionally A Proposed DSS Uses A Single Classifier at A Time for Classification or Prediction Purposes. To Increase Classification Accuracy, Reject Option Classifiers Has Been Proposed in Machine Learning Literature. In A Reject Option Classifier, A Rejection Region Is Defined and The Samples Fall in That Region Are Not Classified by The Classifier. The Unclassifiable Samples Are Rejected by Classifier in Order to Improve Classifier's Accuracy. However, These Rejections Affect the Prediction Rate of The Classifier as Well. To Overcome the Problem of Low Prediction Rates by Increased Rejection of Samples by A Single Classifier, the "Delegating Classifiers" Provide Better Accuracy at Less Rejection Rate. Different Classifiers Such as Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), K Nearest Neighbor (KNN) Etc. Have Been Proposed. Moreover, Traditionally, Same Learner Is Used As "Delegated" And "Delegating" Classifier. This Research Has Investigated the Effects of Using Different Machine Learning Classifiers in Both of The Delegated and Delegating Classifiers, And the Results Obtained Showed That Proposed Method Gives High Accuracy and Increases the Prediction Rate.

Keywords: Cancer Claasification, Supervised Classifiers, Prediction Rate

1. Introduction

Cancer is a very severe disease described by tumors or out-of-control cell growth. The cancer facts and figures 2012-2015 showed that it is rapidly becoming the cause of thousands of deaths worldwide every year, A total of 1,658,370 new cancer cases were reported only in the United States in 2015 (Siegel *et al.*, 2015). According to the World Health Organization (WHO) there were 8.2 million deaths from cancer throughout the world in 2012. Cancer classification is one of the major research areas in the medical field. Predicting precisely of different tumor types is a challenge and providing accurate prediction is worthwhile in providing better treatment to the patients (Rabindra kumar singh, 2015) To identity cancer causes and to discover the methods for prevention, identification, treatments and cure, the cancer research has prime importance and it is one of the major research areas in medical field (Mangasarian *et al.*, 2009). A close examination of gene expression is important in many fields of biological research in order to retrieve the required statistics (Rabindra kumar singh, 2015). For a better insight into the problem posed by the cancer disease, real gene expression data of cancer patients is the key (Xue *et al.*, 2004). DNA microarray technology has provided us with a lot of datasets available for various cancer types such as lymphoid cancer, colon cancer, and breast cancer (Lu and Han, 2003). In Machine Learning (ML) the reliability measure of a classifier is its accurate predictions on unseen data (Ferri *et al.*, 2004). Decision Support Systems (DSSs) are knowledge-based systems, which refers to honor the domain knowledge so that it is responsive to automated reasoning, (Druzdzal and Flynn, 1999). Traditionally a DSS classifies or predicts by using a single classifier (Stefanowski, 2007). For complex problems the use of a single classifier results in the reduced accuracy and/or prediction rate of the overall classification task (Stefanowski, 2007). Among available strategies for improving the accuracy of classifiers, "Reject Option" classification using a single classifier is proposed in Machine Learning literature and all the examples within this rejection region are rejected by

classifier (Nadeem *et al.*, 2010). By increased rejection rate the prediction rate decreases in a single classifier's classification (Stefanowski, 2007). In order to overcome the low prediction problem, it is proposed that the predictions of a committee are always better than the predictions of a single classifier (Ferri *et al.*, 2004). To address the problems posed by reject-option in classification, the concept of Delegating classifiers has been proposed in the literature (Ferri *et al.*, 2004). The idea behind the delegating classifiers is to send unclassified (Rejected) samples to the next more specialized and efficient classifier, Delegated Classifier (Ferri *et al.*, 2004). The classifier which sends samples to the Delegated Classifier is referred to as Delegating Classifier. Various classifiers have been used in this research work that are proposed in Machine Learning Literature E.g. Support Vector Machines-SVMs (Vapnik, 1999), k -Nearest Neighbor- k NN (Fix and Hodges, 1951), Linear Discriminant Analysis-LDA (Fisher, 1936). Moreover, traditionally the same learner is used in delegation as "Delegating" and "Delegated" classifier e.g. Decision Tree learner (Ferri *et al.*, 2004). The research work has investigated the effects of using different classifiers (SVM, LDA, k NN, etc.) as Delegating and Delegated classifiers for cancer prediction/classification. In this research, it is investigated to improve the classification.

2. Literature Review

Some samples are easier to be classified than others and can be predicted using fewer variables as with a lower prediction cost (Blaise Hanczara, 2021). Machine learning models always make a prediction, even when it is likely to be inaccurate. This behavior should be avoided in many decision support applications, where mistakes can have severe consequences (Kilian Hendrickx, 23 July 2021). In classification with a reject option, the classifier is allowed in uncertain cases to abstain from prediction. (Vojtech Franc, 2023). According to Chow in order to reduce the error rate of a classifier, samples for which posteriori probabilities are insufficiently high should not be classified. Generally, such samples fall in some region in which samples from different classes overlap and an ambiguity arises. Because of inferences of doubts and din in data, the bugs are usually cannot be tolerated. Reject option was presented to reject the samples for extreme misrecognition; it transforms misclassification into rejection (Chow, 1957). It is important to make predictions only when they are reliable. Practically, this results in the addition of a "reject" option to the classifiers. If the confidence to predict the appropriate class of the sample is lowest then the samples are rejected (Nadeem *et al.* 2010). In safety critical applications of decision support systems, prediction errors may lead to a serious loss. In such cases, estimating when the model makes an error can be as important as its average performance. These two objectives are taken into account in classification with a reject option when the classifier is allowed in uncertain cases to abstain from prediction (Vojtech Franc, 2023). A thoughtful classifier would delegate the problematic or inaccurate samples to other more efficient classifiers, or possibly more specialized, classifiers. Delegation is mostly preferable for complex classification because it has several advantages over other collaborative or combining methods: Improved efficiency, because individual classifier learns from smaller examples than the previous one, improved clearness, since each classification derives from a single classifier, an outlook to simplify the overall multi-classifier scenario by introducing delegation (Ferri *et al.*, 2004).

With respect to the idea of delegating classifiers, a step by step procedure is generally designed, in which a classifier decides which of the samples to be categorized and which to reject. The unclassifiable rejected samples are sent to other more efficient classifiers (Ferri *et al.*, 2004). The next more efficient classifier in the chain gets trained on the delegated rejected examples and carries out further classification (Ferri *et al.*, 2004). In the delegation classifier we get trained on training samples, predict the classes of new data with high confidence and delegate the samples with low confidence to the classifier 2 (Temanni *et al.*, 2007). The classifier 2 performs the same as the classifier 1 by predicting the classes of new data samples with high confidence and delegates the samples with low confidence (Temanni *et al.*, 2007). It is an admitted fact that the pragmatic performance of classifiers is dependent on the dataset availed. It is very hard to predict which classifier will deliver its best (Sotoca *et al.*, 2005). Machine learning literature has introduced different types of classifiers and every classifier has its own mode of learning, Support Vector Machines-SVMs (Vapnik, 1996), k -nearest neighbor- k NN (Fix and Hodges, 1951), Linear Discriminant Analysis-LDA (Fisher, 1936). However, the classification accuracy of the classifier can be improved by introducing a reject option with the classifier (Chow, 1957).

3. Materials and Methods

A huge amount of real microarray gene expression datasets are available nowadays that can be used for cancer classification. For example, lymphoid cancer data set (Shipp *et al.*, 2002), lymphoid cancer data set is the microarray gene expression dataset of cancer patients suffering from DLBCL - Diffuse large B - cell Lymphomas. (Shipp *et al.*, 2002) have measured 7129 gene expression levels from 58 DLBCL patient samples. Among 58 Diffuse large B-cell Lymphomas, 32 are healed patients; however, 26 patients suffer from fatal or stubborn

disease. Colon cancer dataset (Alon *et al.*, 1999), the data set contains gene expression levels of 40 tumor and 22 normal colon tissues for 6500 human genes. Breast cancer dataset (Golub *et al.*, 1999), the data consists of 47 patients suffering from acute lymphoblastic leukemia and 25 patients suffering from acute myeloid leukemia. Individuals in 72 patients got bone marrow sections gained at the time of therapeutic interventions. Acute Lymphoblastic in the two classes (Nina and Lipo, 2007). It is used to rank features for microarray data (Jäger *et al.*, 2002), and for mass spectrometry data (Ilya, 2005). Methods for computing feature ranks are the Correlation coefficient, which evaluate the dependability of variables to the consequences.

$$t = (\bar{X}_1 - \bar{X}_2) / (\bar{s}^2 X_1 - \bar{X}_2)$$

3.1 Classification Techniques

The literal meaning of the classification is the act of distributing things into classes or categories of the same type. The classification technique implemented in the research work is supervised learning classification technique. Supervised learning is concerned with model training with a set of labeled examples, and to predict labels for unlabeled examples ensuring maximum accuracy. Supervised learning uses density estimation according to which there is an arrangement in the input space in a way that certain patterns occur more often than other patterns (Alpaydm, 2014).

3.2 Classifiers

Machine learning literature has proposed various types of classifiers for classification purpose, each classifier has its own mode of operation or classification, and moreover classifier's performance varies from data to data (Sotoca *et al.*, 2005). Support Vector Machines belong to the class of various kernel methods Linear, Polynomial, Radial, and Sigmoid and are progressing in the statistical learning theory. Like all kernel-based learning algorithms they are a general-purpose learning machine, for example Svm Linear and specific kernel functions, for others problems (Hofmann, 2006). Linear discriminant analysis is a statistical approach for supervised dimensionality reduction and classification. It computes an optimal conversion by minimizing the 'within class' distance and maximizing the 'between class' distance simultaneously, thus achieving maximum class discrimination (Ye, 2007). *KNN* is one of the most fundamental and simple classification methods and should be one of the first choices for a classification study when there is little or no prior knowledge about the distribution of the data (Fix and Hodges, 1951).

3.3 Rejection Techniques – Ambiguity Rejection

To reduce the error rate of a classifier, samples for which posteriori probabilities are too high should not be classified. Generally, such samples fall in some regions in which samples from different classes overlap and an ambiguity arises. Because of uncertainties and noise inherent in data, the errors are generally unavoidable (Nadeem *et al.*, 2010).

3.4 Delegation

By increased rejection rate the prediction rate decreases in single classifier classification (Stefanowski, 2007). To address the problems posed by reject-option in classification, the concept of Delegating classifiers has been proposed in the Machine Learning Literature (Ferri *et al.*, 2004). The idea behind the delegating classifiers is to send unclassified rejected samples to the next more specialized and efficient classifier, the Delegated Classifier (Ferri *et al.*, 2004). The classifier which sends samples to Delegated Classifier is referred to as Delegating Classifier. Leukemia (ALL), ALL is a subset of arrays from a large acute lymphoblastic leukemia study. Acute lymphoblastic leukemia ascends from 2 categories of lymphocytes: one is T-cell and the other is B-cell. The details of the datasets used in this work are summarized in Table-1.

Table 1. Summaries of Datasets Used In Presented Research Work

Data Source	Sample Size	Number of Features	Number of Classes
(Shipp <i>et al.</i> ,2002)	58	7129	2
(Alon <i>et al.</i> ,1999)	62	2000	2
(Golub <i>et al.</i> ,1999)	72	7129	2
ALL	128	12625	2

3.5 Data Preprocessing

One of the useful means to recognize relevant features for dimensionality fall/reduction is feature selection.

(Isabelle and Andre, 2003; Huan and Lei, 2005). Data needs preprocessing in order to overcome some problems like overfitting. Overfitting is the situation when the model is heavily complex, which means to say that the model has too many parameters, greater than the number of observations (Refaeilzadeh et al., 2009). K-fold cross-validation is used in this research work to overcome this problem. The dataset is splitted into k approximately alike subgroups. The kth subgroup is used as a test or validation set and the remaining k-1 subgroups are used as training set in k fold cross validation (Refaeilzadeh et al., 2008). Student t-test, estimates whether the means of two classes are statistically unlike by computing a ratio between the difference of two class means and the unevenness of

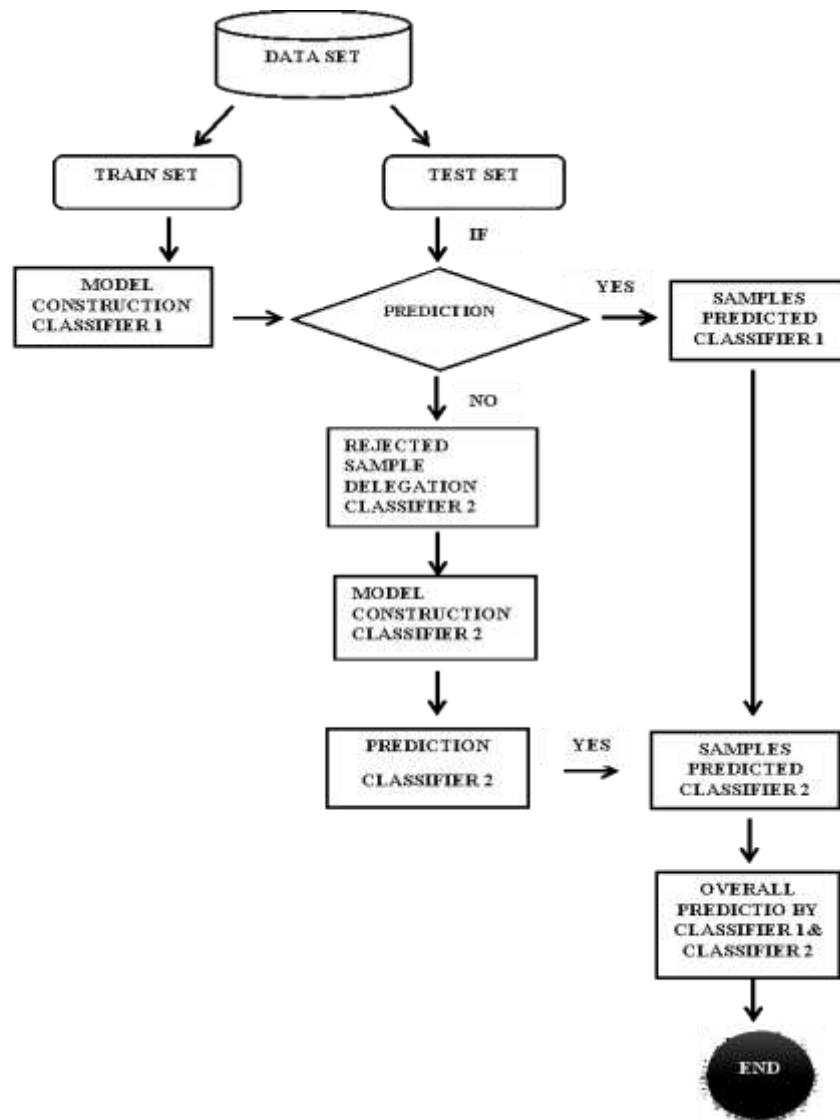


Figure 1. Simulation plan used in current study is present

4. Results and Discussion

Accuracy Rejection Curves- ARC comparison of classifiers “with Delegation” and “without Delegation”:

In this section, obtained results based on different settings are presented. Figure-2 contains the results based on the ALON dataset. At first kNN algorithm is used the classifier (CL1) and then rejected samples and LDA classifier are used to build the classifier (CL2).

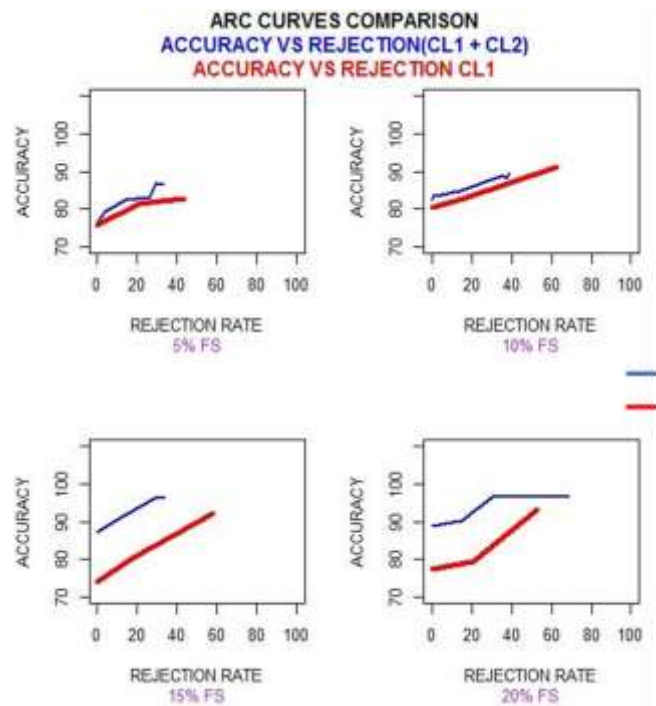


Figure 2. Results obtained by using kNN (CL1) and LDA (CL2)

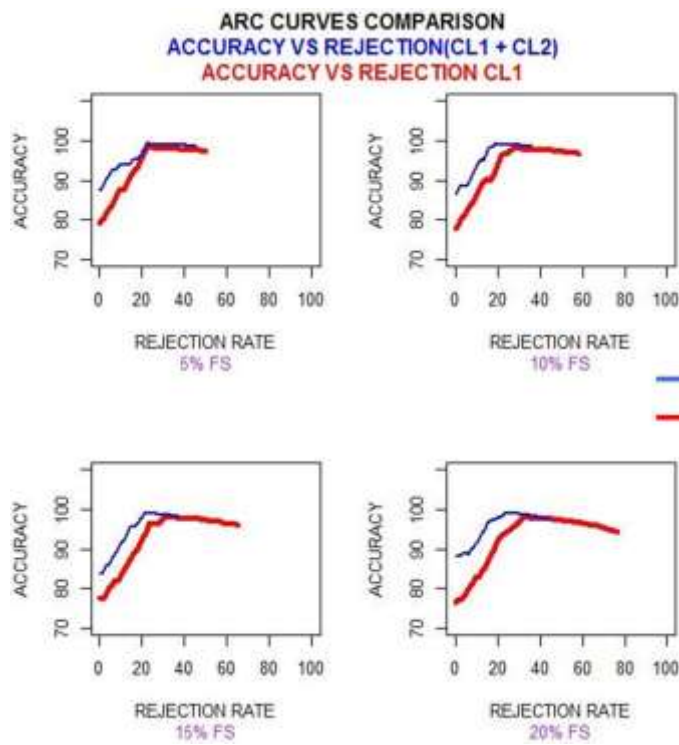


Figure 3. Results obtained by using SVM Polynomial (CL1) and LDA (CL2)

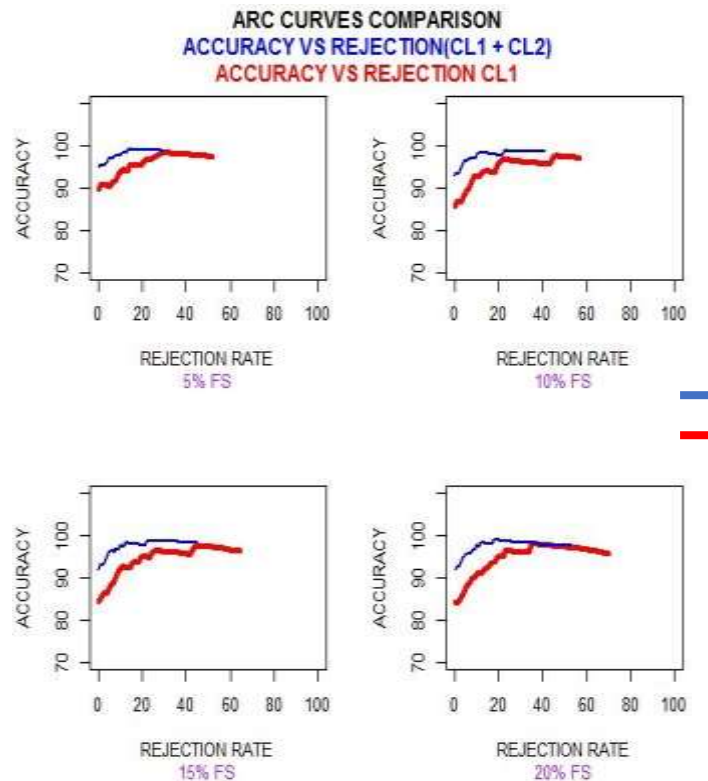


Figure 4. Results obtained by using SVM Polynomial (CL1) and LDA (CL2)

The choice of a classification method has its certain importance. The choice of the classification method depends upon how reliable and efficient it is. A single classifier however classifies the samples efficiently and accurately. A classifier with a reject option has a reliable and improved accuracy rate (details discussed earlier). However, by increased sample rejection the prediction rate is decreased. The overall prediction rate is improved, if the rejected samples are delegated to the next more efficient classifier. So that prediction rate as well as accuracy rate of overall classification task is improved. Results showed that in delegation the overall classification workload is divided among the classifiers, and it is proved to be very efficient. A classifier classifies the example for which it is specialized and delegates the rejected examples to the next more efficient classifier which is more efficient to classify them. However, it is the requirement of successful delegation that the classifier that gets the rejected examples should be more efficient and more reliable to successfully classify the rejected example. The research work has investigated the use of different learning algorithms in both delegating and delegated classifiers which further increased the accuracy and decreased the error rate. This research investigated the fact that each learning algorithm has its own specific details for learning and complexity of a problem is based on data too.

5. Conclusion and Main Outcomes

In the present study we investigated the effects of using different machine learning algorithms in the scenario of delegating classifiers and we have improved the classification ability of a decision support system. The improved classification ability is of prime importance in almost every field of life. For example, in tumor classification and therapy, a decision support system with improved classification ability helped improve the quality of care and facilitated by providing more accurate, effective, and reliable diagnosis and treatments, and by avoiding errors due to misclassification.

References

- Alon, U. B., Naama, A. N., Daniel, W. G., Kenneth, S., Ybarra, M., & Danieland, J. L. A. (1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proceedings of The National Academy of Sciences*, *96*, 6745-6750. <https://doi.org/10.1073/pnas.96.12.6745>
- Alpaydm, E. (2014). *Introduction to machine learning*. MIT press.

- Blaise, H. A. H. (2021). CASCARO: Cascade of Classifiers for Minimizing the Cost of Prediction. *pattern recognition*.
- Chow, C. K. (1957). *An Optimum Character Recognition System Using Decision Functions' Transactions on Electronic Computers EC-6*, 247-254. <https://doi.org/10.1109/TEC.1957.5222035>
- Druzdzal, M. J., & R. R. Flynn (1999). Decision support systems. *Encyclopedia of library and information science*. A. Kent. Marcel Dekker, Inc. Last Login 1003: 2010.
- Ferri, C., P. Flach & J. Hernández-Orallo (2004). Delegating classifiers. *Proceedings of the twenty-first international conference on Machine learning, ACM*. <https://doi.org/10.1145/1015330.1015395>
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 72, 179-188. <https://doi.org/10.1111/j.1469-1809.1936.tb02137.x>
- Fix, E., & J. L. Hodges (1951). *Discriminatory analysis-nonparametric discrimination: consistency properties*. DTIC Document. <https://doi.org/10.1037/e471672008-001>
- Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., ... Lander, E. S. (1999). *Molecular classification of cancer: class discovery and class prediction by gene expression monitoring*. *Science*. <https://doi.org/10.1126/science.286.5439.531>
- Hofmann, M. (2006). Support Vector Machines - *Kernels and the Kernel Trick*. *Hauptseminar report*.
- Huan, L. Y., & Lei (2005). Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on Knowledge and Data Engineering*, 17, 491-502. <https://doi.org/10.1109/TKDE.2005.66>
- Isabelle, G., & E. Andre (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, 3, 1157-1182.
- Jäger, J., R. Sengupta & W. L. Ruzzo (2002). Improved gene selection for classification of microarrays. *Proceedings of the eighth Pacific Symposium on Biocomputing: 3-7 January 2003; Lihue, Hawaii*. https://doi.org/10.1142/9789812776303_0006
- Kilian Hendrickx, L. P. (23 JULY 2021). *Machine Learning with a Reject Option*.
- Lu, Y., & J. Han (2003). Cancer classification using gene expression data. *Information Systems*, 284, 243-268. [https://doi.org/10.1016/S0306-4379\(02\)00072-8](https://doi.org/10.1016/S0306-4379(02)00072-8)
- Mangasarian, O. L., W. N. Street & W. H. Walberg (2009). *Breast Cancer Diagnosis and Prognosis via Linear Programming*, *Cite seer*.
- Nadeem, M. S. A., J-D Zucker & Hanczar, B. (2010). Accuracy-Rejection Curves (ARCs) for Comparing Classification Methods with a Reject Option *Journal of Machine Learning Research*, 8, 65-81.
- Nina, Z., & W. Lipo (2007). A Modified T-test Feature Selection Method and Its Application on the HapMap Genotype Data. *Genomics, Proteomics & Bioinformatics*, 5, 242-249. [https://doi.org/10.1016/S1672-0229\(08\)60011-X](https://doi.org/10.1016/S1672-0229(08)60011-X)
- Rabindra kumar singh, M. D. (2015). Feature Selection of Gene Expression Data for Cancer Classification: A Review. *Procedia computer sciences*.
- Refaeilzadeh, P., L. Tang & H. Liu (2009). *Cross-Validation*. *Encyclopedia of Database Systems*. Liu, L. and M. T. Özsu, Springer US: 532-538. https://doi.org/10.1007/978-0-387-39940-9_565
- Shipp, M. A., K. N. Ross, P. Tamayo, A. P., Weng, J. L., Kutok, R. C. T., Aguiar, M., ... T. R. Golub (2002). Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat Med*, 81, 68-74. <https://doi.org/10.1038/nm0102-68>
- Siegel, R. L., K. D. Miller & A. Jemal (2015). Cancer statistics, 2015. *CA: a cancer journal for clinicians*, 651, 5-29. <https://doi.org/10.3322/caac.21254>
- Sotoca, J., J. Sanchez & R. Mollineda (2005). A review of data complexity measures and their applicability to pattern classification problems. *Actas del III Taller Nacional de Minería de Datos y Aprendizaje.-TAMIDA: 77-83*.
- Stefanowski, J. (2007). On combined classifiers, rule induction and rough sets. *Transactions on rough sets VI*, Springer: 329-350. https://doi.org/10.1007/978-3-540-71200-8_18
- Temanni, M., Ramzi, S. A., Nadeem, D. P. B., & J. D. Zuker (2007). *Camda 07, Spain*.

- Vapnik, V. N. (1999). *An overview of statistical learning theory*. Neural Networks, IEEE Transactions on 105: 988-999. <https://doi.org/10.1109/72.788640>
- Vojtech Franc, D. P. (2023). Optimal Strategies for Reject Option Classifiers. *Journal of Machine Learning Research*.
- Xue, R., J. Li & D. J. Streveler (2004). Microarray gene expression profile data mining model for clinical cancer research. System Sciences, 2004. *Proceedings of the 37th Annual Hawaii International Conference on, IEEE*.
- Ye, J. (2007). *Least squares linear discriminant analysis*. Proceedings of the 24th international conference on Machine learning, ACM. <https://doi.org/10.1145/1273496.1273633>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).