Enliven Archive
Intensifying Thoughts

Review Article | www.enlivenarchive.org | Enliven: Journal of Anesthesiology and Critical Care Medicine

ISSN:2374-4448

Using Support Vector Machines (Svm) to Facilitate Classification of Cancer

Nick Ray

Independent Researcher in Centre for Pain, University of Bath, 1 West Level 3, BA2 7AY, United Kingdom

***Corresponding author**: Nick Ray, Independent Researcher in Centre for Pain, University of Bath, 1 West Level 3, BA2 7AY, United Kingdom, Email: rayn82@yandex.com

Received Date: 15th July 2018
Accepted Date: 15th September 2018
Published Date: 16th September 2018

Citation: Ray N (2018) Using Support Vector Machines (Svm) to Facilitate Classification of Cancer. Enliven: J Anesthesiol Crit Care Med 5(1): 005.

Scientists are now able to concurrently perform screening or expression of many genes using DNA micro-arrays to establish their status that is whether they are silent, hyperactive, or active in cancerous or normal tissue [1]. Given the fact that such new micro-array devices produce massive volumes of critical data, there is a need to develop new analytical methods that are capable of sorting out whether the cancer cellshaveunique signatures of gene tissue over the several types of cancer tissue or the normal cells [2]. The current multiple data sets which are publicly available on open source platforms such as Internet are characterized by many problems, for instance,a rather smaller number of experiments and a huge amount of gene expression values per experiment [3]. Additionally, it is evident that data analysis can be conducted from diverse perspectives. The majority of the empirical studies in the available relate to gene clusters which have been revealed through unsupervised learning methodologies [4]. Clustering is usually conducted together with other data dimensions; for instance, each experiment might link to one patient who might carry or not have a certain disease [5]. Therefore, clustering is mostly done with the intention of grouping the patients with comparable clinical history or record [6]. Presently, it is evident that the application of supervised learning to classify cancer and proteins has increasingly gained prominence [7].

In this case, the researchers seek to prove that the application of advanced classification algorithms such the Support Vector Machines (SVM) can greatly help in extracting a minor subsets of highly discriminant genes to create highly dependable cancer classifiers [8]. In addition, the researchers wanted to demonstrate how Support Vector Machines (SVMs) can be applied in data mining and knowledge discovery [9]. The Support Vector Machines play a great role with regards to discovering informative patterns as well as attributes or features (for instance, significant genes) [10]. Furthermore, the researchers seek to illustrate that SVMs have both quantitative and qualitative advantages as compared to other gene section techniques related to Colon cancer data. In addition, their method outdoes other schemes as it demonstrates high classification performance for minor gene subsets while making the selection of genes that have reasonable significance to cancer diagnosis and treatment [11].

The researchers seek to explore massive patterns of gene expression data which have been recorded on DNA micro-arrays to tackle the problem of selection of a minor gene subset [12]. Furthermore, they developed appropriate classifier used during drug discovery and genetic diagnosis using accessible training examples from healthy and cancer patients [13]. The previous proposed solutions to tackle the problem used correlation methods to select the genes [14]. The researchers use the Support Vector Machine techniques which are founded on the Recursive Feature Elimination (RFE) to develop a new technique of gene selection [15]. Additionally, they conducted experiments to illustrate that the selection of genes using their method will produce superior classification performance with high biological relevance to cancer [16]. In comparison to the baseline method, the researchers' techniques spontaneously eradicate gene redundancy as well as produce high quality subsets of gene [17]. The technique discovered two genes in patients with leukemia that produce zero leave-one-out error whereas sixty four genes are essential for the baseline method to achieve the best outcomes (that is one leave-one-out error) [18]. Furthermore, in the colon cancer database, the baseline attained 86% accuracy while the researchers' method achieved 98% accuracy using the four genes [19].

The Support Vector Machines usually integrate feature and pattern selections in a single consistent framework. They performed experiments involving two diverse cancer databases that demonstrated that classification performance can be impacted when one considers mutual information between the genes in gene selection process [20]. In comparison to the baseline technique that makes implied orthogonality presumptions, the researchers' method obtained considerable improvements in the gene selection process [21]. Additionally, they validated the biological significance of the genes which were acquired through the Support Vector Machines. The high rated genes acquired through the process all had a probable link to cancer [22]. Alternatively, the other technique was irrelevant to cancer diagnosis despite selecting genes that are associated with the separation [23]. The SVM allowed the researchers to get nested subsets of genes that are compatible to a model selection method that achieves an optimal number of genes [24]. Furthermore, the investigation illustrates that the Recursive Feature Elimination is superior in relations to data over fitting as compared to other techniques inclusive of combinatorial search [25].

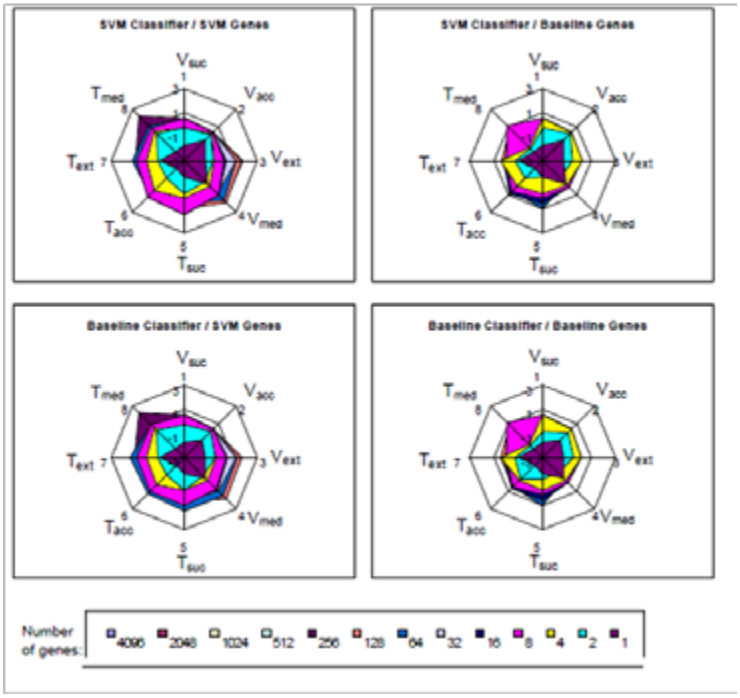

Figure 1: Performance comparison between SVMs and the baseline method [23].

References

1. Isabelle Guyon, Jason Weston, Stephen Barnhill, Vladimir Vapnik (2002) Gene Selection for Cancer Classification Using Support Vector Machines. Machine Learning 46: 389-422.
2. Vellodi A, Foo Y, Cole TJ (2005) Evaluation of three biochemical markers in the monitoring of Gaucher disease. J Inherit Metab Dis 28: 585-592.
3. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, et al. (2000) Distinct Types of Diffuse Large B-Cell Lymphoma Identified by Gene Expression Profiling. Nature 403: 503.
4. Swathi M (2018) Enhancement of K-Mean Clustering for Genomics of Drugs. Enliven: J Genet Mol Cell Biol 5: 001.
5. Alon U, Barkai N, Notterman DA, Gish K, Ybarra S, et al. (1999) Broad Patterns of Gene Expression Revealed by Clustering Analysis of Tumor and Normal Colon Tissues Probed by Oligonucleotide Arrays. Proc Natl Acad Sciences 96: 6745-6750.
6. Ben-Hur A, David Horn, Hava TS, Vladimir V (2000) A Support Vector Method for Clustering. In International Conference on Pattern Recognition 3: 12-20.
7. Guyon I, Boser B, Vapnik V (1993) Automatic Capacity Tuning of Very Large VC-Dimension Classifiers. Advances in Neural Information Processing Systems 5:147-155.
8. Bernhard E. Boser, Isabelle M. Guyon, Vladimir N. Vapnik (1992) A Training Algorithm for Optimal Margin Classifiers. In Proceedings of the Fifth Annual Workshop on Computational Learning Theory, 3:144-152.
9. Avrim L. Blum, Pat Langley (1997) Selection of Relevant Features and Examples in Machine Learning. Artificial Intelligence 97: 245-271.
10. Bradley PS, Olvi LM, Nick WS (1998) Feature Selection via Mathematical Programming. INFORMS Journal on Computing 10: 209-217.
11. Swathi M (2017) Drug Prediction of Cancer Genes Using SVM. Enliven: Pharmacovigilance and Drug Safety 4: 001.
12. Bredensteiner, Erin J, Kristin PB (1999) Multicategory Classification by Support Vector Machines. In Computational Optimization 12: 53-79.
13. Ron Kohavi, George H. John (1997) Wrappers for Feature Subset Selection. Artificial Intelligence 97: 273-324.
14. Brown MP, Grundy WN, Lin D, Cristianini N, Sugnet CW, et al. (2000) Knowledge-Based Analysis of Microarray Gene Expression Data by Using Support Vector Machines. Proc Natl Acad Sci 97: 262-267.
15. Chapelle Olivier, Vladimir Vapnik, Olivier Bousquet, Sayan Mukherjee (2002) Choosing Multiple Parameters for Support Vector Machines. Machine Learning 46: 131-159.
16. Cortes C, Vladimir V (1995) Support Vector Machine. Machine Learning 20: 273-297.
17. Mukherjee S, Pablo ASDT, Donna S, Alessandro Verri, et al. (1999) Support Vector Machine Classification of Microarray Data. Molecular and Cellular Biology 1: 2-32.
18. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster Analysis and Display of Genome-Wide Expression Patterns. Proc Natl Acad Sci 95: 14863-14868.
19. Swathi M (2017) Clustering Enhancement Using Similarity Indexing to Reduce Entropy. Enliven: Bioinform 4: 001.
20. Ibid.
21. Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M, et al. (2000) Support Vector Machine Classification and Validation of Cancer Tissue Samples Using Microarray Expression Data. Bioinformatics 16: 906-914.

22. Tony Jebara, Tommi Jaakkola (2000) Feature Selection and Dualities in Maximum Entropy Discrimination. In Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence: 291-300.
23. Hastie T, Tibshirani R, Michael E, Brown P, Doug R, et al. (2000) Gene Shaving: A New Class of Clustering Methods for Expression Arrays. Genome Biology 2: 1-3.
24. Michael Kearns, Yishay Mansour, Andrew Y. Ng, Dana Ron (1997) An Experimental and Theoretical Comparison of Model Selection Methods. Machine Learning 27: 7-50.
25. Isabelle Guyon, Jason Weston, Stephen Barnhill, Vladimir Vapnik (2002) Gene Selection for Cancer Classification Using Support Vector Machines. Machine Learning 46: 395.
26. Thorsteinsdottir U, Krosl J, Kroon E, Haman A, Hoang T, et al. (1999) The OncoproteinE2A-Pbx1a Collaborates with Hoxa9 to Acutely Transform Primary Bone Marrow Cells. Mol Cell Biol 19: 6355-6366.
27. Perou CM, Jeffrey SS, Van De Rijn M, Rees CA, Eisen MB, et al. (1999) Distinctive Gene Expression Patterns in Human Mammary Epithelial Cells and Breast Cancers. Proc Natl Acad Sci 96: 9212-9217.
28. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster Analysis and Display of Genome-Wide Expression Patterns. Proc Natl Acad Sci 95: 14867.