Enliven Archive
Intensifying Thoughts

Review Article | www.enlivenarchive.org | Enliven: Challenges in Cancer Detection and Therapy

ISSN: 2376-046X

Individual Drug Treatment Prediction in Oncology Based on Machine Learning using Cell Culture Gene Expression Data

Nancy William

Department of Oncology, University of Canberra, Building 1/11 Kirinari St, Bruce ACT 2617, Australia

*Corresponding author: Nancy William, Department of Oncology, University of Canberra, Building 1/11 Kirinari St, Bruce ACT 2617, Australia, E-mail: nancywilliam90@yandex.com

Received Date: 16th September 2018
Accepted Date: 22nd November 2018
Published Date: 23rd November 2018

Citation: William N (2018) Precision Oncology beyond Targeted Therapy: Combining Omics Data with Machine Learning Matches the Majority of Cancer Cells to Effective Therapeutics Enliven: Challenges Cancer Detect Therv 3(1): 003.

In contemporary oncology, there is great effort to develop individual predictors of clinical drug effectiveness. Therefore, for any particular patient with positively diagnosed cancer type as well as a selected drug, the researchers sought to approximate the treatment effect which was caused by the medicine [1]. Machine learning (ML) techniques and gene expression are widely applied to conduct statistical exploration of a set of clinical cases for all patients [2]. Nonetheless, the approach is hindered by a significant challenge that is, the total set of available cases to be analyzed is quite restricted [3]. Alternatively, there are multiple cell cultures in the biotech drug industry which are sustained by gene expression data then assessed to evaluate drug scoring [4]. In this case, the researchers demonstrate how the cell lines data can be integrated into the machine learning analysis to enhance the development of discrete predictors [5]. In personalized medicine, scientists have been able to explore the field of big data collected for a particular patient (for instance, gene expression or mutational data) to predict the effectiveness of a certain treatment regimen or drug for the patient [6]. Personalized medicine is widely applied in oncology and it utilizes machine-learning methods owing to the complex characteristics of the processes that define both progression of cancer as well as the likely methods that facilitate its subdual [7].

Presently, there are many dominant and cutting-edge tools such as the Support vector machines (SVM) that are used to conduct regression analysis and classification [8]. The Support vector machines have proved to be more effective in relations to modifications in input data as compared to other machine-learning algorithms such as the classical multi-layer perceptrons (MLP) [9]. In the training dataset, the SVM need a smaller number of preceding cases while the MLP utilize the least square fitting method [10].The SVM have proved to be essential in predicting the efficiency of data for cancer patient owing to the fact that the current MLPs need many points for the training dataset to sufficiently cover the phase space [11]. In opposite, the separators which are largely SVM-based might tolerably operate with lesser points in the training datasets [12].

Regrettably, it is extremely challenging for majority of the anti-cancer drugs to obtain the multiple gene expressions that were found through the use of the similar investigation platform for the patients received treatment with the similar drug with the identified clinical outcome [13].

With regards to application of the SVM method to effectively perform prediction of anti-cancer drug efficiency, the researchers proposed a new technique that allowed expression-based data to be transferred to the validation (V-) dataset provide description of similar data collected from positively diagnosed cancer patients from training (T-) set that contains expression-based data for cell lines [14,15] The researchers introduced a combination of SVM and kNN methods to stop the SVM process from engaging in pointless adaptation especially while data is being transmitted to the V-set (data for patients) from the T-set (cell line data) [16].

Combining the SVM-with-kNN technique signify that the T-set's K points which are proximal to a particular point in the V-set are relied to construct the SV model [17]. The move means, for each V-set point, the new data transfer method lowers or filters the T-data by use of a floating window surrounding all points in the V-set to disregard the impact of the points which are located far from the T-set [18]. The combination was used to analyze three type of cancer diseases namely lung, renal, and CML cancer [19]. For each condition, the researchers chose the best K value which is relied in maximizing the Area-Under-the-Curve (AUC) to predict the drug response for the group of patients [20]. The highest AUC value surpasses 0.69; the ideal K value seems to be strong and constant consistent with the V-dataset's leave-one-out quality assurance process [21].

The researchers did a permutation test to determine whether the support vector machine-with- kNN process is not overtrained. The test whereby the real non-responder/responder flags for the V-dataset samples were substituted with arbitrary values [22]. With regards to the three types of cancer the unsystematic permutation tremendously

reduced the aggregate number of cases when the Area-Under-the-Curve (AUC) was less than 0.69 [23]. Furthermore, the p-value was 4•10-4 for the null hypothesis meaning the reduction was triggered by an arbitrary chance concurrently for the dataset of the three conditions [24]. The researchers were confident that their method had numerous benefits of both international (such as support vector machine) and national (for instance the kNN method) machine learning methods [25].

Drug GEO reference Disease type	Lung cancer, sorafenib [14] GSE31428	Renal cancer, sorafenib (current study)	CML, imatinib [15] GSE2535
Samples	37 (23 responders, 14 non-responders)	28 (13 responders, 15 non-responders)	28 (16 responders, 12 non-responders)
Optimal *K* value	29	162	104
AUC for optimal *K* value	0.72	0.81	0.78
p-value for Gaussian test	0.16	0.04	0.07

The outcome of performance test for kNN technique (Source: Borisov, Nikolay, Victor Tkachev, Ilya Muchnik, and Anton Buzdin (2017) Individual Drug Treatment Prediction in Oncology Based on Machine Learning Using Cell Culture Gene Expression Data. In Proceedings of the 2017 International Conference on Computational Biology and Bioinformatics 1: 1-6).

References

1. Borisov N, Tkachev V, Muchnik I, Buzdin A (2017) Individual Drug Treatment Prediction in Oncology Based on Machine Learning Using Cell Culture Gene Expression Data. In Proceedings of the 2017 International Conference on Computational Biology and Bioinformatics 1: 1-6
2. Min-Wei H, Chih-Wen C, Wei-Chao L, Shih-Wen K, Chih-Fong T (2017) SVM and SVM Ensembles in Breast Cancer Prediction. PloS one 12: e0161501.
3. El Naqa, Issam, Ruijiang Li, Martin J. Murphy (2015) Machine Learning in Radiation Oncology. Theory Appl 2: 57-70.
4. Osuna, Edgar, Robert Freund, Federico Girosi (1997) An Improved Training Algorithm for Support Vector Machines. In Neural Networks for Signal Processing [1997] VII. Proceedings of the 1997 IEEE Workshop 1: 276-285.
5. Peter B, Shawe-Taylor J (1999) Generalization Performance of Support Vector Machines and Other Pattern Classifiers. Advances in Kernel Methods—Support Vector Learning 43-54.
6. Vapnik V, Chapelle O (2000) Bounds on Error Expectation for Support Vector Machines. Neural computation 12: 2013-2036.
7. Robin X, Turck N, Hainard S, Lisacek F, Sanchez JC, et al. (2009) Bioinformatics for Protein Biomarker Panel Classification: What is needed to Bring Biomarker Panels into in Vitro Diagnostics?. Expert Review of Proteomics 6: 675-689.
8. Swathi M (2017) Clustering Enhancement Using Similarity Indexing to Reduce Entropy. Enliven: Bioinform 4: 001.
9. Minsky Marvin, Seymour A. Papert. Perceptrons: An Introduction to Computational Geometry. MIT press, 2017.
10. Buzdin AA, Zhavoronkov AA, Korzinkin MB, Roumiantsev SA, Aliper AM et al. (2014) The OncoFinder Algorithm for Minimizing the Errors Introduced by the High-Throughput Methods of Transcriptome Analysis. Frontiers in Molecular Biosciences 1: 8.
11. Borisov NM, Terekhanova NV, Aliper AM, Venkova LS, Smirnov PY, et al. (2014) Signaling Pathways Activation Profiles Make Better Markers of Cancer Than Expression of Individual Genes. Oncotarget 5: 10198-10205.
12. Chih-Chung C, Chih-Jen L (2011) LIBSVM: A Library for Support Vector Machines. ACM Transactions on Intelligent Systems and Technology (TIST) 2: 27.
13. Buzdin AA, Zhavoronkov AA, Korzinkin MB, Venkova LS, Zenin AA, et al. (2014) Oncofinder, a New Method for the Analysis of Intracellular Signaling Pathway Activation Using Transcriptomic Data. Frontiers In Genetics 5: 55.
14. Kuzmina NB, Borisov NM (2011) Handling Complex Rule-Based Models of Mitogenic Cell Signaling (On The Example Of ERK Activation Upon EGF Stimulation). Int Proc Chem Biol Environ Eng 5: 67-82.
15. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H et al. (2013) Genomics of Drug Sensitivity in Cancer (GDSC): A Resource for Therapeutic Biomarker Discovery in Cancer Cells. Nucleic acids Res 41: D955-D961.
16. Blumenschein GR, Saintigny P, Liu S, Kim ES, Tsao AS, et al. (2013) Comprehensive Biomarker Analysis and Final Efficacy Results of Sorafenib in the BATTLE Trial. Clinical Cancer Research 19: 6967-6975.
17. Crossman LC, Mori M, Hsieh YC, Lange T, Paschka P, et al. (2005) In Chronic Myeloid Leukemia White Cells from Cytogenetic Responders and Non-Responders to Imatinib Have Very Similar Gene Expression Signatures. Haematologica 90: 459-464.
18. Mulligan G, Mitsiades C, Bryant B, Zhan F, Chng WJ, et al. (2007) Gene Expression Profiling and Correlation with Outcome in Clinical Trials of the Proteasome Inhibitor Bortezomib. Blood 109: 3177-3188.
19. Swathi M (2018) Enhancement of K-Mean Clustering for Genomics of Drugs. Enliven: J Genet Mol Cell Biol 5: 001.
20. Arimoto R, Prasad MA, Gifford EM (2005) Development of CYP3A4 Inhibition Models: Comparisons of Machine-Learning Techniques and Molecular Descriptors. J Biomol Screen10: 197-205.
21. Balabin RM, Smirnov SV (2012) Interpolation and Extrapolation Problems of Multivariate Regression in Analytical Chemistry: Benchmarking the Robustness on Near-Infrared (NIR) Spectroscopy Data. Analyst 137: 1604-1610.

22. Betrie GD, Tesfamariam S, Morin KA, Sadiq R (2013) Predicting Copper Concentrations in Acid Mine Drainage: A Comparative Analysis of Five Machine Learning Techniques. Environ Monit Assess 185: 4171-4182.
23. Shabalin AA, Tjelmeland H, Fan C, Perou CM, Nobel AB (2008) Merging Two Gene-Expression Studies Via Cross-Platform Normalization. Bioinformatics 24: 1154-1160.
24. Rudy J, Valafar F (2011) Empirical Comparison of Cross-Platform Normalization Methods for Gene Expression Data. BMC Bioinformatics 12: 467.
25. Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003) A Comparison of Normalization Methods for High Density Oligonucleotide Array Data Based on Variance and Bias. Bioinformatics 19: 185-193.
26. Swathi M (2017) Drug Prediction of Cancer Genes Using SVM. Enliven: Pharmacovigilance and Drug Safety 4: 001.
27. Wang Q, Liu X (2015) Screening of Feature Genes in Distinguishing Different Types of Breast Cancer Using Support Vector Machine. OncoTargets and Therapy 8: 2311-2317.
28. Chih-Chung C, Chih-Jen L (2011) LIBSVM: A Library for Support Vector Machines. ACM Transactions on Intelligent Systems and Technology (TIST) 2: 27.