Enliven Archive
Intensifying Thoughts

Review Article www.enlivenarchive.org Enliven: Pharmacovigilance and Drug Safety

ISSN: 2378-5411

How Different Types of Data affects Drug Sensitivity Prediction

Aldin Paul

Lecturer, Department of Pharmaceutical Sciences, Chulalongkorn University, 254 Phayathai Road, Pathumwan, Bangkok 10330, Thailand

*Corresponding author: Aldin Paul, Lecturer, Department of Pharmaceutical Sciences, Chulalongkorn University, 254 Phayathai Road, Pathumwan, Bangkok 10330, Thailand, E-mail: paulaldin@tutanota.com

Received Date: 16th September 2018

Accepted Date: 22nd November 2018

Published Date: 23rd November 2018

Citation: Paul A (2018) How Different Types of Data affects Drug Sensitivity Prediction. Enliven: Pharmacovigilance and Drug Safety 5(1): 002.

According to the article, the NCI-DREAM Drug Sensitivity Prediction problem has been persistent for long; nonetheless, several models and algorithms have developed to tackle its two major sub-challenges [1]. The model includes a weighted Euclidean distance technique that is used to make a prediction as well as ranking the combination of drug especially their ability to neutralize the capability of a diffuse large B-cell lymphoma (DLBCL) cell line [2]. Similarly, the other technique is a bidirectional search algorithm that uses a nonlinear support vector machine (SVM) and a common scheme to predict how the drug compounds can impact the breast cancer cell lines [3]. The bidirectional search algorithm or greedy search algorithm usually makes a combination of the advantages of the support vector machine or kernel techniques as well as the ensemble modeling to correctly make a prediction of the sensitivity of the breast cancer cell lines to drug compounds which were formerly not tested [4]. Current composite models used to predict drug sensitivity use basic algorithms to extract traits from one type of data set [5]. Alternatively, the researcher created an ensemble model (as shown in figure 1 below) that was used in extracting features from different kind of datasets such as DNA copy number disparity, DNA methylation, RNA-seq, gene expression, proteomic data) as opposed to the use of diverse base algorithms on one form of datasets [6]. The researchers used similar base learning algorithms on the five different kinds of datasets [7].

Cancer is regarded as the causative factor of death globally, the development of drug compounds to target cancer cell and provide treatment has highly been prioritized [8]. Nonetheless, the discovery of cancer therapies is regarded as challenging, costly, and time consuming pharmaceutical venture since it entails various clinical evaluations and developmental phases [9]. Currently, simulation and computational models are increasingly preferred to make prediction of the reaction of the cancer cell lines to drug compounds to drug development or discovery process [10]. Such models are significantly helpful in improving the process of designing drugs as well as tackle several problems experienced throughout the process that assists in establishing the drug variations [11].

Furthermore, the researchers argue that the process of determining successful lead drug candidates used in treating cancer can greatly capitalize from such methods used to make prediction of how cancer cell lines are sensitive to drug compounds [12]. Time-series in addition to static conditional gene expression data are widely used to build the silicon models or predictors [13]. Statistical methods, for instance, a combination or integration of random forest and regression analysis have been used to manipulate gene expression data obtained from cancer cells which have treated with diverse drug compounds to predict of the capacity of the drugs to effectively stop cancer cell lines' proliferation [14]. Additionally, researchers have exploited gene expression datasets using the Naïve Bayes classifiers the sequencing data and chromosomal copy number difference obtained from human cancer lines which undergo treatment using twenty-four anticancer drugs to make a prediction of the capacity of the drugs to constrain their propagation [15]. Equally, gene expression data of sixty human cancer cell lines have been examined using a weighted voting classification model to foresee the drug reactions [16].

Moreover, the researchers combined random forest and nearest neighbor methodology to manipulate proteomic data to predict of the drug response of the cell lines [17]. In addition, a weighted voting algorithm which is created through the use of a set of genes which have been expressed differentially was used to accurately group 80 percent of the twenty-six samples to assess the effectiveness of the Docetaxel (an anticancer drug) in treating breast cancer [18, 19]. Majority of the statistical methods are variedly applied to the sequencing, gene expression, and proteomic data to determine how the cancer cell lines usually react to medication [20]. Therefore, review of the article reveal that current practice of making prediction about the cancer cell lines' sensitivity to the drugs can greatly be enhanced by using the ensemble models [21]. The researchers wanted to determine how the diverse forms of data can be relied on more correct prediction of the reaction of cancer cell lines to the drug compounds [22]. The research has greatly added new knowledge especially with how the ensemble methods can be used in capturing and extracting significant features from the diverse kinds of biological data [23, 24].

Furthermore, the simplified weighted Euclidean distance measure can be applied since it has high potential of attaining reasonable outcomes as well as more advanced statistical/similarity evaluations of making predictions about the reaction of cancer cell lines to a mixture of drug compounds [25]. The study can significantly help to introduce significant therapeutic, pharmacological, and clinical changes which can prove beneficial in improving the drug development process since it facilitates the detection of successful lead drug candidates used to treat the diverse types of cancer [26].

References

1. Hejase HA, Chan C (2015) Improving Drug Sensitivity Prediction Using Different Types of Data. CPT Pharmacometrics Syst Pharmacol 4: 98-105.
2. Lee JK, Havaleshko DM, Cho H, Weinstein JN, Kaldjian EP, et al. (2007) A Strategy for Predicting the Chemosensitivity of Human Cancers and Its Application to Drug Discovery. Proc Natl Acad Sci U S A 104: 13086-13091.
3. Mori S, Chang JT, Andrechek ER, Potti A, Nevins JR (2009) Utilization of Genomic Signatures to Identify Phenotype-Specific Drugs. Plos One 4: e6772.
4. Staunton JE, Slonim DK, Coller HA, Tamayo P, Angelo MJ, et al. (2001) Chemosensitivity Prediction by Transcriptional Profiling. Proc Natl Acad Sci U S A 98: 10787-10792.
5. Berlow N, Davis LE, Cantor EL, Seguin B, Keller C, et al. (2013) A New Approach for Prediction of Tumor Sensitivity to Targeted Drugs Based On Functional Data. BMC Bioinformatics 14: 239.
6. Chang JC, Wooten EC, Tsimelzon A, Hilsenbeck SG, Gutierrez MC, et al. (2003) Gene Expression Profiling for The Prediction of Therapeutic Response to Docetaxel in Patients with Breast Cancer. Lancet 362: 362-369.
7. Riddick G, Song H, Ahn S, Walling J, Borges-Rivera D, et al. (2010) Predicting in Vitro Drug Sensitivity Using Random Forests. Bioinformatics 27: 220-224.
8. Jordi B, Caponigro G, Stransky N, Venkatesan K, Margolin AA, et al. (2012) The Cancer Cell Line Encyclopedia Enables Predictive Modelling of Anticancer Drug Sensitivity. Nature 483: 603-607.
9. Ma Y, Ding Z, Qian Y, Shi X, Castranova V, et al. (2006) Predicting Cancer Drug Response by Proteomic Profiling. Clinical Cancer Research 12: 4583-4589.
10. Iwao-Koizumi K, Matoba R, Ueno N, Kim SJ, Ando A, et al. (2005) Prediction of Docetaxel Response in Human Breast Cancer By Gene Expression Profiling. J Clin Oncol 23: 422-431.
11. Harrell, Frank (2003) Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. Bioinformatics 1: 257-258.
12. Saeys Y, Inza I, Larranaga P (2007) A Review of Feature Selection Techniques in Bioinformatics. Bioinformatics 23: 2507-2517.
13. Borisy AA, Elliott PJ, Hurst NW, Lee MS, Lehar J, et al. (2003) Systematic Discovery of Multicomponent Therapeutics. Proc Natl Acad Sci 100: 7977-7982.
14. Burger M, Engl HW (2000) Training Neural Networks with Noisy Data as Ill-Posed Problem. Advances in Computational Mathematics 13: 335-354.
15. Hijazi H, Wu M, Nath A, Chan C (2012) Ensemble Classification of Cancer Types and Biomarker Identification. Drug Dev Res 73: 414-419.
16. Hussein H, Chan C (2013) A Classification Framework Applied to Cancer Gene Expression Profiles. J Healthc Eng 4: 255-283.
17. Malik Y, Ketany M, Manevitz L, Showe LC, Showe MK (2009) Classification and Biomarker Identification Using Gene Network Modules and Support Vector Machines. BMC Bioinformatics 10: 337.
18. Swathi M (2018) Enhancement of K-Mean Clustering for Genomics of Drugs. Enliven: J Genet Mol Cell Biol 5: 001.
19. Wu J, Gan M, Jiang R (2011) A Genetic Algorithm for Optimizing Subnetwork Markers for The Study of Breast Cancer Metastasis. In Natural Computation (ICNC), 2011 Seventh International Conference On 3: 1578-1582.
20. Jahid J, Ruan J (2011) Identification of Biomarkers in Breast Cancer Metastasis by Integrating Protein-Protein Interaction Network and Gene Expression Data. In Genomic Signal Processing and Statistics (GENSIPS) 2011 IEEE International Workshop on 60-63.
21. Heiser LM, Sadanandam A, Kuo WL, Benz SC, Goldstein TC, et al. (2012) Subtype and Pathway Specific Responses To Anticancer Compounds In Breast Cancer. Proceedings of The National Academy of Sciences 109: 2724-2729.
22. Jordi B, Caponigro G, Stransky N, Venkatesan K, Margolin AA, et al. (2012) The Cancer Cell Line Encyclopedia Enables Predictive Modelling of Anticancer Drug Sensitivity. Nature 483: 603-607.
23. Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, et al. (2012) Systematic Identification of Genomic Markers of Drug Sensitivity in Cancer Cells. Nature 483: 570-575.
24. Swathi M (2017) Clustering Enhancement Using Similarity Indexing to Reduce Entropy. Enliven: Bioinform 4: 001.
25. Dietterich TG (2000) Ensemble Methods in Machine Learning. In International Workshop On Multiple Classifier Systems 1-15.
26. Chang CC, Lin CJ (2011) LIBSVM: A Library for Support Vector Machines. ACM Transactions on Intelligent Systems and Technology (TIST) 2: 27.
27. Ying L, Han J (2003) Cancer Classification Using Gene Expression Data. Information Systems 28: 243-268.
28. Swathi M (2017) Drug Prediction of Cancer Genes Using SVM. Enliven: Pharmacovigilance and Drug Safety 4: 001.