Enliven Archive
Intensifying Thoughts

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

ISSN: 2378-5411

Drug Prediction of Cancer Genes Using SVM

Swathi Muppalaneni

Master of Science in Medical Sciences, Long Island University, Brooklyn & Brookville, New York, USA

*Corresponding author: Swathi Muppalaneni,Master of Science in Medical Sciences, Long Island University, Brooklyn & Brookville, New York, USA, E-mail: Smuppalaneni@gmail.com

Received Date: 18th October 2017

Accepted Date: 6th November 2017

Published Date: 12th November 2017

Citation: Swathi M (2017) Drug Prediction of Cancer Genes Using SVM. Enliven: Pharmacovigilance and Drug Safety 4(2): 001.

Abstract

The challenge for customize drug is to choose the right medicine for the individual patient. Drug testing of patients in large clinical trials is a way of assessing their efficacy and toxicity, but testing hundreds of drugs currently under development is impractical. Therefore, preclinical prediction models are highly desirable because of their capability is to know drug acknowledgement to hundreds of cell lines. In this paper, we presents, the classification technique is used to differentiate the drugs. Initially pre-processing is used on the true data that removes the noise, then the features are extracted which are saved into the database along with the information related to drug response. Then the performance parameters such as precision, recall, accuracy and f-measures are determined. The approximate accuracy observed come out to be 0.9315.

Keywords: Drug detection; SVM; Precision; Recall; Accuracy; F-measure

Introduction

In modern years, drug detection efforts have primarily concentrated on recognizing agents that modulate pre-selected singular targets. Although novel drugs have continuously been identified, there is a developing productivity gap, despite more spending on examination and development and advances in technological development [1]. This dilemma arises partly because agents aimed at a particular target frequently show the limited effectiveness and poor safety and resistance profiles that are mostly due to features like network toughness, dismissal, and crosstalk & neutralize actions and anti as well as counter target actions [2]. With such problems in mind, systems-oriented drug scheme has been increasingly highlighted as a potentially more fruitful strategy. This method of drug scheme is sustain by clinical authorities with multi component treatments and multi-targeted agents, and struggles have been directed at the growth of novel multi component therapies [3]. Though with related clinical symptoms, several patients may have separate responses to the corresponding drug or therapy. Therefore, personalized medicine, which makes medical judgments based on patients' genetic content, becomes the foremost directions of the future medical science [4]. In order to produce and access targeted treatments for single patient, one must expedient to the long and costly procedure of drug growth and justification in medical cases, the common way to judge drug effectiveness and toxicity. But the insufficiency of assets has restricted this scheme to useful applications. One potential resolution to this problem is to instantly measure the responsiveness of a patient's tumor cells to a drug of interest in two or three-dimensional (2,3D) in-vitro cultures [5] or in-vivo models like mouse xenograft and hereditarily engineered models. This method has the potential of catching most of the relevant biological characteristics of a patient's tumor, and hence, providing excellent models to test drug sensitivity. Though, such an access is costly, time-consuming and rarely capable of being scaled to screen number of drugs parallely. With the construction of the efficient output methods in the earlier few decades, an alternative system was presented by different research organizations to made genomic analysis of drug acknowledgement from huge panels of cancer cell [6] (Figure 1).

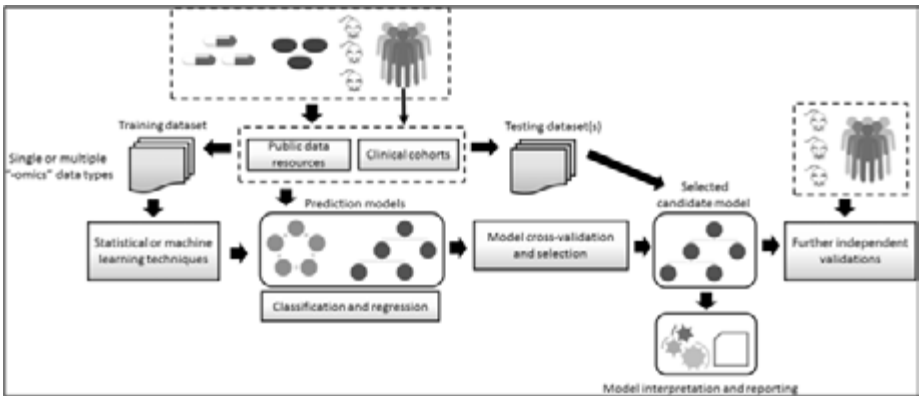

Figure1: Steps to analyze drug response

Utmost of these techniques are depends on gene expression profile. The above figure shows the different steps used for predicting the drug response. There are mostly four steps that are used for knowing the drug response [7]. Firstly, the data set are selected and then pre-processed. For pre-processing one can use computer system and filtering system to remove noise. Secondly the features are take out from the relevant data. Then the extracted data from human being is saved into the metadata along with the drug response information. The prediction issues might be defined as a classification and regression problem. In the testing phase the data is inputted that is not used during the training phase [8-11].

Related Work

In this section, the related work in the field of drugs classification performed by a number of researchers is discussed along with the dataset and outcomes. (Table 1).

References	Proposed work	Database	Outcomes
[12]	Proposed the domain-tuned-hybrid technique that infers the network of drug-target interactions.	(www.clinicaltrials.gov)	A sequence of kinase comprises of DLG_4 (PSD-95) & BDKRB2, a G protein-integrated Kin's receptor in family have been identified.
[13]	Proposed a new computational algorithm to know the drug response of every patient by means of the personal genomic profiles, as well as pharmacogenomic and drug sensitivity data.	The data has been taken from GDSC data base.	It is concluded that the proposed drug prediction algorithm has been utilized enhance the reliability to determine optimal drugs for individual patients and also form a significant factor in the precision medicine infrastructure for oncology care.
[14]	Proposed three various methods to diagnose cancer in human being.	Affymatrix best-match data set available at www.affymetrix.com	Similarity metrics have been examined for different cancer types by suing Cosine and Semi-correlation function. It is determined that ANN perform better than other two techniques. But it is expensive than other two techniques.
[15]	Used Fuzzy algorithm and SVM algorithm for cancer prediction.	The datasets of Cancer gene and the visualizations http://www.biolab.si/supp/bi-cancer/projections	It is concluded that the accuracy of SVM- k-NN is high as compared to fuzzy logic.
[16]	Presented a technique by utilizing GA (Genetic algorithm) for enhancing the prediction of accuracy for Breast cancer.	A dataset of 133 patients that are suffered from breast cancers have been taken for the 1st, 2nd and 3rd stage.	The prediction accuracy up to 92 % has been achieved by using Genetic Algorithm (GA).
[17]	Proposed a machine learning model system that has been used to categorize cell line chemosensitivity exclusively based on proteomic profiling.	http://discover.nci.nih.gov/datasets.jsp	It is concluded that Chemosensitivity has been predict accurately by using proteomic method.

Table 1: Comparative analysis of existing work

Proposed Methodology

In this article, the work used SVM classification algorithms for detecting drugs as per patient. Initially pre-processing has been used on the test data for removing the noise, then the features are extracted which are saved into the database along with the information related to drug response. SVM is used to classify the drugs as per patient requirement.

Dataset

The dataset of the research work has been taken from (cancer X-gene org.) universal genomics of drug sensitivity repository is listed in table 2. The values are taken for different drugs named as RNF14, CCNE1, MKL1, HHAT, LIFR, FDXR, DUS P21, SEMA4F, ORI 2D3, PIK3CA, PIK3CG, MAPK8, ARHGEF12.

	RNF14	CCNE1	FHIT	MKL1	HHAT	LIFR	FDXR	DUSP21	SEMA4F	OR12D3	PIK3CA	PIK3CG	MAPK8	ARH-GEF12
23132-87	0.121773	-0.11566	-0.65141	-0.20408	-0.28739	-0.31293	3.768469	-1.00507	-0.29991	0.932495	-1.00629	0.061634	-1.13654	-0.21235
5637	0.387102	0.165806	1.2908	-0.59954	0.302017	-0.56414	0.13507	0.55189	-0.3745	-0.77753	-0.17119	-0.57571	1.709916	0.011803
639-V	0.195854	0.190151	-0.898	0.158226	-0.60775	0.024976	-0.6937	-0.40014	0.217455	-1.06417	1.261287	-0.36173	-0.63186	2.618048
647-V	0.291048	-0.00734	-1.09848	1.978738	0.440794	-0.40334	-0.33061	-0.79399	0.91943	-0.08569	2.333559	0.034483	0.064457	0.054486
697	-0.70217	-0.37302	0.22472	-1.17345	-0.39963	-0.59252	-0.59464	0.099633	-0.68706	1.013707	1.197107	1.076214	0.53159	-0.7202
786-0	1.227811	-0.36199	0.214307	0.669712	0.34973	-0.3965	-0.55762	5.152416	-0.3099	-0.69748	0.238684	-0.52096	2.486837	1.248128
8-MG-BA	-0.41364	1.404851	0.30255	0.018807	0.926164	-0.47816	0.235431	1.478572	0.328081	-1.05283	-0.61115	-0.45432	-0.88701	-0.81737
8505C	-0.39743	-0.37748	-1.16099	-1.12578	-0.68144	-0.11148	-0.64511	1.282889	-0.9419	-1.02415	-0.44013	-0.56143	-0.38579	0.833196
A101D	-0.21889	0.138503	0.776327	0.781485	0.126079	0.043527	1.356456	1.143151	0.661404	1.034824	-1.18288	-0.52697	-1.28766	-0.78915
A172	1.365816	0.116666	0.67243	0.082871	1.01644	-0.37352	1.735213	0.572314	0.64084	1.92375	-0.84549	0.105821	-0.10344	-0.53547
A2058	1.109898	0.285524	0.214927	-0.35279	0.421385	0.052534	-0.54717	1.656064	-0.57773	-0.8492	0.599685	1.253219	0.286329	0.576426
A253	-0.51117	-0.41784	-0.5311	-0.14399	-0.45761	0.342237	0.031753	0.344865	0.213101	0.184154	0.857091	-0.59303	-0.45698	0.105821
A2780	0.16888	-0.18677	1.358228	-0.1573	-0.92135	0.348209	0.225513	-0.71992	-0.2539	-0.90801	0.824231	-0.52367	0.326265	0.309628
A3-KAW	-0.97779	-0.18628	1.700061	-0.49525	-0.34379	-0.61359	0.104446	-0.41223	-0.40205	-1.05972	-0.79451	4.28677	-0.42871	-1.13658
A375	0.999935	-0.30101	-1.11785	-0.62168	0.600566	0.817344	-0.23241	1.89238	-0.73087	2.167392	1.184382	0.040307	-1.86898	1.163266
A4-Fuk	-1.08639	-0.20736	0.084383	-0.80444	-0.45342	-0.58945	-0.16271	-0.45125	-0.53783	1.187272	-0.97623	4.741809	-0.49011	-1.19241
A427	-0.65391	0.544717	-0.51561	-1.01138	-0.22358	-0.46744	3.145288	0.800438	2.095959	1.097242	-0.56594	-0.47735	0.446775	-0.80129
A431	-1.04264	-0.28991	-1.1408	-0.53303	-0.45577	-0.5973	-0.51071	-0.91367	-0.46109	-0.48646	0.336214	0.146732	0.701286	-0.40186
A498	0.528933	0.113951	-0.34161	-0.14573	-0.09543	-0.57563	0.038716	-1.12317	1.272929	1.263448	-0.54538	-0.54811	-0.2116	-0.41346
A549	0.968953	-0.14762	-0.41139	0.077694	-0.304	0.122767	-0.02898	-0.94752	0.055467	-0.35162	-0.43379	0.163398	0.685902	0.319673
A704	2.427497	-0.38385	0.271403	-0.46646	0.299353	-0.58434	0.230743	0.959545	1.199781	-0.74363	-0.25135	-0.56187	-0.07743	1.329561
ABC-1	-0.36187	-0.34273	0.309576	0.507807	-0.00098	0.60606	-0.40288	-1.01459	-0.31694	-0.45557	3.682774	0.180777	1.41016	1.14122
ACHN	-0.10372	-0.36356	3.120561	-0.60284	1.261212	1.076124	1.421226	0.146082	0.14064	1.048009	0.470658	-0.5267	1.685573	2.099579
ACN	1.753898	-0.30297	0.301769	-0.66623	0.185998	-0.42682	-0.63572	2.305204	0.888672	-0.42574	0.390736	-0.54676	0.329424	-0.37015

Table 2: Original dataset

Support Vector Machine (SVM)

SVM is known as a binary classifier which is utilized for differentiating only two categories. SVM used a hyper plane for classification purpose. In the below figure, we have taken two categories of data along with straight lines. Here, blue circle represents 1st class data whereas red blocks represents 2nd class data. Thus the operation of SVM is for finding the hyper plane that offers the minimum distance to the training data. This distance is known as margin in SVM. Here in this figure, from class one only 1 dataset comes under the hyper plane whereas from class two 2 datasets comes in the range f hyper plane. Thus SVM train these data that is comes in the hyper plane (Figure2).

In the proposed work, SVM is used to classify the drugs categories that which drug is suitable for the patient. On the basis of rule set created by SVM the drugs are predicted for the target.

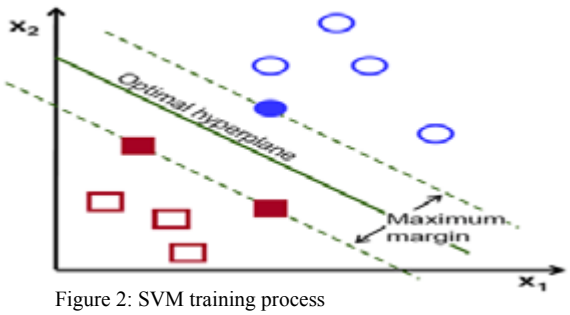

Figure 2: SVM training process

Result Analysis

In this section, the results are shown that are obtained after the simulation of the model. The parameters namely Precision, Recall, Accuracy and f-measure are used to obtain the results (Table 3).

Above table comprises of the values of precision, recall, accuracy and f-measures obtained by numerous samples taken for the classification. The graphical analysis is shown below (Figure 3).

No. of samples	Precision	Recall	Accuracy	f-measure
5	0.9245	0.9175	0.9275	9278
10	0.9541	0.9298	0.9378	94.58
15	0.9615	0.9204	0.9532	91.68
20	0.9308	0.9469	0.9136	94.86
25	0.9127	0.9586	0.9257	95.96

Table3: Simulation parameters

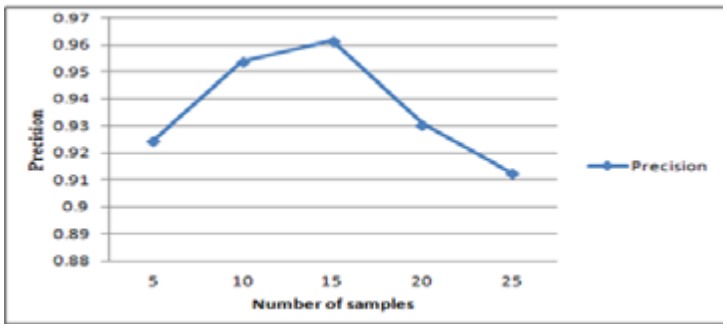

Figure 3: Precision

Above figure shows the precision value for 5, 10, 15, 20 and 25 number of samples. X-axis represents number of samples, y-axis represents precision values. It is being concluded from the above graph that the average value for precision is 0.93672 (Figure 4).

The above figure represents the recall values for the 5, 10,15,20,25 number of samples. Recall has a average value of 0.9346. The recall rate is increases with the increase in the number of samples (Figure 5).

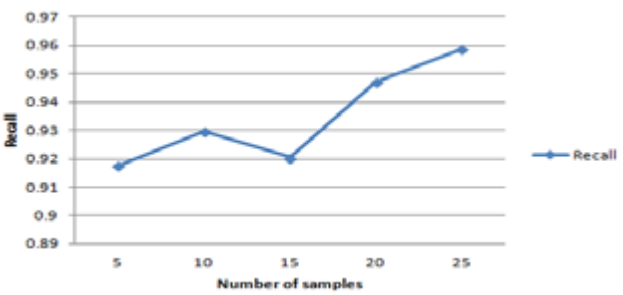

Figure 4: Recall

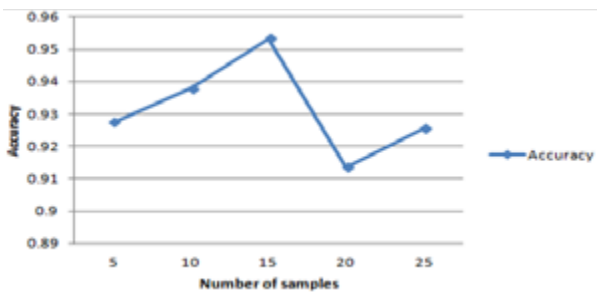

Figure 5: Accuracy

The above figure shows the accuracy for number of samples such as 5, 10,15,20,25. The average value of accuracy obtained for the proposed work is approximately equal to 0.9315 (Figure 6).

The above figure represents the f-measure value obtained for the proposed work for different number of samples such as 5, 10 15 20 and 25 respectively. 93.97 is the f-measure average value.

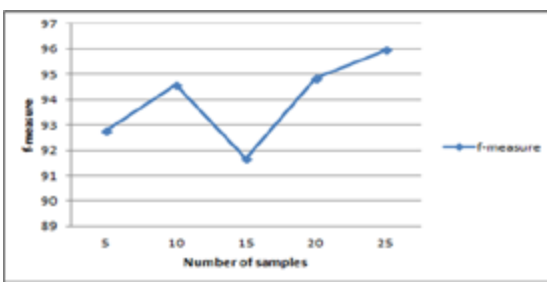

Figure 6: f-measure

Conclusion

The resolution of precision medicine couples on the ability to efficiently translate genomic data into actionable, customized diagnosis and treatment regimens for different patients. This demands to identify a genomic disease indication from a patient, then coordinating it with the most effective therapeutic intrusion. The typical data set is to build auspicious models linking genomic context to treatment would be systematically characterized drug sensibilities across a huge cohort of patients, but this information is time-intensive to produce, prohibitively pricey, and restricted in the field of drugs that can be examined. So, the proposed work is used to utilize the concept of support vector machine algorithm and hence the results are analyzed. It is concluded that the average values of precision, recall, accuracy and f-measures are 0.93672, 0.9346, 0.9315 and 0.9315 respectively.

References

1. Ma Y, Ding Z, Qian Y, Shi X, Castranova V, et al. (2006) Predicting cancer drug response by proteomic profiling. Clin Cancer Res 12: 4583-4589.
2. Ma Y, Ding Z, Qian Y, Wan YW, Tosun K, et al. (2009) An integrative genomic and proteomic approach to chemo sensitivity prediction. Int J Oncol 34: 107-115.
3. Jia J, Zhu F, Ma X, Cao Z, Cao ZW, et al. (2009) Mechanisms of drug combinations: interaction and network perspectives. Nat Rev Drug Discov 8: 111-128.
4. Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, et al. (2013) The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet 45: 1113-1120.
5. Huang L, Li F, Sheng J, Xia X, Ma J, et al. (2014) DrugComboRanker: drug combination discovery based on target network analysis. Bioinformatics 30: 228-236.
6. Hofree M, John P Shen, Hannah Carter, Andrew Gross, Trey Ideker, et al. (2014) Network-based stratification of tumor mutations. Nat Methods 10: 1108-1115.
7. Kandoth C, Schults N, Cherniack AD, Akbani R, Liu Y, et al (2013) Integrated genomic characterization of endometrial carcinoma. Nature 497: 67-73.
8. Barretina J, 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. 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.
10. 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: 955-961.
11. Li F, Huang L, Sheng J, Wong S (2016) Network-based Computational Drug Combination Prediction, IEEE.
12. Nguyen TP, Priami C, Caberlotto L (2015) Novel drug target identification for the treatment of dementia using multi-relational association mining. Scientific reports 5: 11104.
13. Sheng J, Li F, Wong ST (2015) Optimal drug prediction from personal genomics profiles. IEEE journal of biomedical and health informatics 19: 1264-1270.
14. Shamsaei B, Gao C (2016) Comparison of some machine learning and statistical algorithms for classification and prediction of human cancer type. In Biomedical and Health Informatics (BHI), 2016 IEEE-EMBS International Conference 296-299.
15. Maulik U, Chakraborty D (2014) Fuzzy preference based feature selection and semisupervised SVM for cancer classification. IEEE transactions on nanobioscience 13: 152-160.
16. Hu W (2015) High accuracy gene signature for chemosensitivity prediction in breast cancer. Tsinghua Science and Technology 20: 530-536.
17. Ma Y, Ding Z, Qian Y, Shi X, Castranova V, et al. (2006) Predicting cancer drug response by proteomic profiling. Clin Cancer Res 12: 4583-4589.