IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 7, Issue 2, February 2019

Detection of Cancer using Machine Learning

Omprakash B[1], Raksha K R[2], Sheetal M[3], Soujanya B[4], Yogitha K R[5]

Assistant Professor, Dept. of Information Science & Engg., Atria Institute of Technology, Bangalore, India[1]

Student, Dept. of Information Science & Engg., Atria Institute of Technology, Bangalore, India[2,3,4]

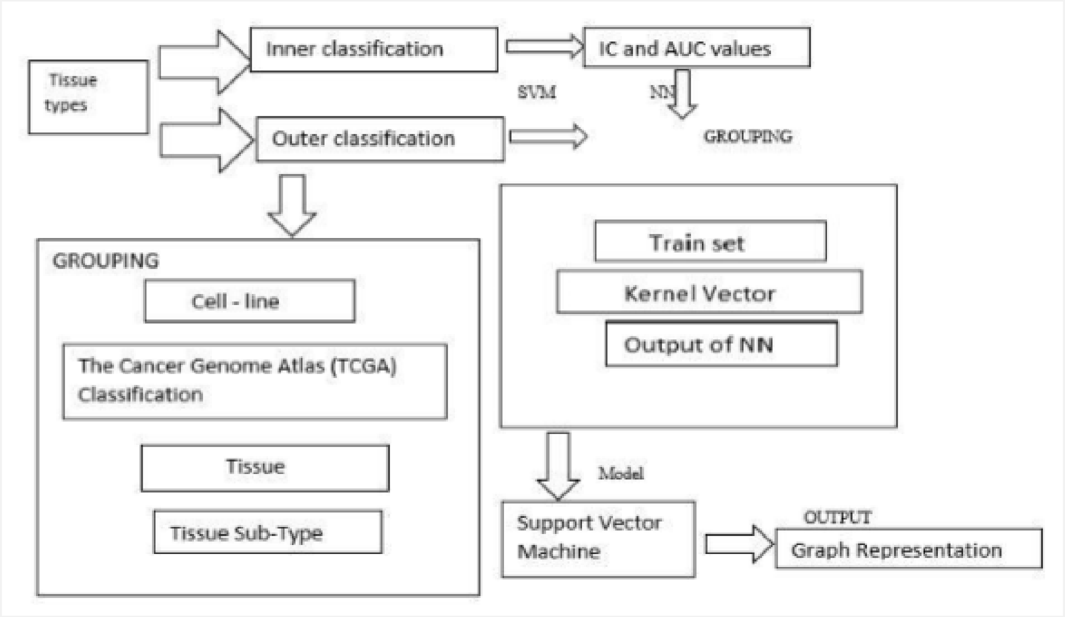
Abstract: Cancer is a heterogeneous disease. There are various options available for cancer treatment. Type of cancer treatment is influenced by various factors such as cancer type, the severity of cancer and the most important is genetic heterogeneity. So in such complex environment, the targeted drug treatments are likely to be irresponsive or respond differently. To study the anticancer drug response we need the classification of cancerous profile. Hence, there is a need to analyze cancer data for predicting optimal treatment option. Analysis of cancerous profile predicts and discovers potential drugs and drug targets. Main aim is to provide machine learning based classification technique for cancerous profile.

Keywords: Cancer, Machine learning techniques, Neural Network (NN) and Support Vector Machine (SVM)

I. INTRODUCTION

All living organisms are made up of basic unit of life, called Cells. Individual cells describe a completely complex functionality. All living organisms are made up of basic unit of life, called Cells. Individual cells describe a completely complex functionality. Genes are the carrier of genetic information within the Cell. Genetics is a branch of science that has evolved ever since study of genes started. Genetics also studies about the expression level of the genes, to determine the up and down state of the gene. Gene's classification and clustering methods are the integral part of any analysis in the micro array data. Current classification methods rely primarily on the cancer's tissue of origin (for example, whether a tumor first developed in the lung or the brain) and on the microscopic appearance and location of cancerous cells. TCGA's principal aims are to generate quality control, merge, analyze and interpret molecular profiles at the DNA, RNA, protein and epigenetic levels for hundreds of clinical tumors representing various tumor types and their subtypes. Microarrays have become important tools for profiling global gene expression patters of cells/tissues. Currently, such studies involve many thousands of genes but only a few hundred of fewer samples.

II. ARCHITECTURE DIAGRAM



Initially we will fetch the input data which are the tissue types. Fetched input is subjected to pre – processing where we find IC and AUC values which becomes tissue values and it is used for inner clustering in proposed algorithm. In outer

DOI 10.17148/IJIREEICE.2019.7210

classification, we use Neural Network (NN) and Support Vector Machine (SVM) model to group the data. Data is grouped with respect to tissue, each is tissue is grouped separately. Finally, we compare the efficiency by calculating accuracy of SVM And NN clustering by plotting graph.

III. METHODOLOGY

The proposed methodology encompasses of hybrid algorithm which contains inner and outer classification. The proposed algorithm is divided into three sections:
a) Dataset Pre-processing
b) Clustering using Neural Network
c) Classification using Support Vector Machine.

The steps or methods followed to develop and implement the project is as follows, fetching the tissue types based on four compound target values. Preprocess the data and find out the IC50 and AUC values as inner clustering. Classify the data based on tissue, this is done using feed forward neural network, using tissue groups are formed and count of it is made. Later using SVM which is based on statistical learning, this model takes two inputs itself and output of NN clustering (inner classification) We calculate the Precision, Recall, F-measure and Accuracy to plot the graph and compare the efficiency .The graph we get will show how efficient the drug works as sample size increases.

SVM (Support Vector Machine) :
Machine learning and. data mining. In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis.

Neural Network:
An artificial neural network is a network of simple elements called artificial neurons, which receive input, change their internal state (activation) according to that input, and produce output depending on the input and activation.

IV. CONCLUSION

In this survey, how to detect cancer has been shown. From the knowledge we gained by referring multiple project reports and survey papers belonging to the domain of Machine learning, we conclude oncogenomic research domain aims at identifying and analyzing cancer related genes and thus helps in diagnosis at genotype level. In order to address aforementioned challenges the proposed technique is an attempt to solve classification problem for cancerous genomic profiles. Our technique is based on concept of utilizing SVM and NN machine learning algorithm. Result provides comparative analysis of model performance when the sample size is varied. As the sample size increase model performance also increases, which shows positive aspect towards the robustness and adaptively of the model.

REFERENCES

[1]. Alexandre R Zlotta, “Genome sequencing identifies a basis for everolimus sensitivity,” European urology, Vol. 64, No. 3, pp. 29-33, 2013.
[2]. P Ganesh Kumar, T Aruldoss Albert Victoire, P Renukadevi and Durairaj Devaraj, “Design of fuzzy expert system for microarray data classification using a novel genetic swarm algorithm,” Expert Systems with Applications, Vol. 39, No. 2, pp. 1811–1821, 2012.
[3]. Todd R Golub, Donna K Slonim, Pablo Tamayo, Christine Huard, Michelle Gaasen beek, Jill P Mesirov, Hilary Coller, Mignon L Loh, James R Downing and Mark A Caligiuri “Molecular classification of cancer: class discovery and class prediction by gene expression monitoring,” Science, Vol. 286, No. 5439, pp. 531–537, 1999.
[4]. John N Weinstein, Eric A Collisson, Gordon B Mills, Kenna R Mills Shaw, Brad A Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander and Joshua M Stuart, “The cancer genome atlas pan-cancer analysis project,” Nature Genetics, Vol. 45, No. 10, pp. 1113–1120, 2013.

DOI 10.17148/IJIREEICE.2019.7210