

Classifying Drug Resistance in the NCI60 Cancer Cell Lines Based on the mRNA Expression Levels of the 48 ABC Transporters

Sultanah Alshammari,¹ Vinoo Uurity², Rajeev K. Azad^{2,3}

Departments of 1. Computer Science and Engineering, 2. Biological Sciences, 3. Mathematics
University of North Texas, Denton, Texas 76203

Introduction

One of the main causes of the cancer treatment failure is drug resistance. Cancer cells might be resistant to a single drug or a wider spectrum of chemotherapy named as multi-drug resistance (MDR). Cancerous tumors can become drug resistant via drug pumping channels known as ATP-binding cassette (ABC) transporters. ABC transporters are transmembrane proteins that mainly transport a wide variety of substrates across extra- and intracellular membranes by utilizing ATP hydrolysis. ABC transporters have the ability to pump cancer drugs from inside the cell into the extracellular environment [1].

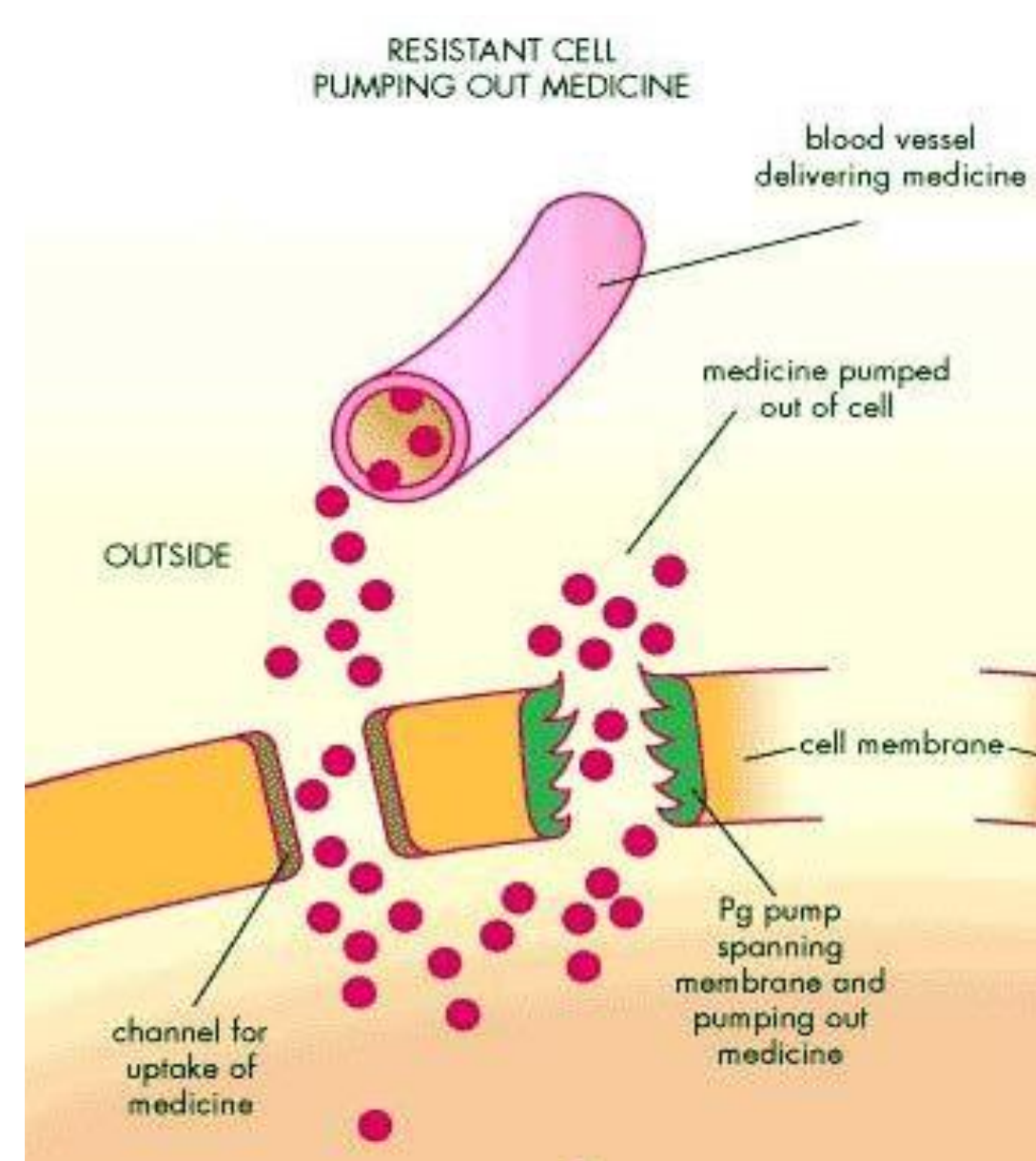


Figure 1: Common drug-resistance pathway [2]

The ability to classify drug resistance of different cancer cell lines using the gene expression data of the ABC transporters is essential for successful treatment of cancer. An accurate classification of drug resistance provides reliable indications to direct the treatment process and could help to decide whether or not to refer a patient for specific drugs.

Classifying cancer resistance to drugs is one of many problems in bioinformatics for which machine learning methods can be used. In this work, we investigate the performance of different classifications methods to distinguish which cancer cell lines show resistance to which drugs based on gene expression profiles of the ABC transporters.

Models and Methods

DATASET: Cancer-drug interaction data were retrieved from the NCI-60's Cell Miner database and gene expression profile from Szakács et al.^[1] The drug interactions are given as a Z score derived from the SRB cytotoxicity assays preformed [3].

CLASSIFICATION MODELS:

Support Vector Machine (SVM): SVM is a supervised learning method that requires training data to find an appropriate hyper-plane to separate different classes of objects in high dimensional space. New examples are mapped onto the same space for prediction.

Decision Trees (DT): DT are a type of classification method utilizing flow-chart type structures to model outcomes. The structure is composed of three types of nodes: decision, chance and end nodes. This type of classification method allows the user to make probabilistic conclusions.

Models and Methods Cont.

Artificial neural network (ANN): ANN is a multilayer perceptron with a feed-forward model that maps sets of input data onto a set of appropriate outputs. It uses three or more layers of nodes to distinguish data that is not linearly separable, or separable by a hyper-plane.

EXPERIMENT

1. The input dataset was divided into two subsets– a training dataset to train the classifier, and a test dataset to assess the methods.
2. For WEKA Implementation, 10-fold cross-validation was used and for JustNN, SPSS was used to randomly choose 10 drugs to use for our training and verification datasets.
3. Implementation of the selected classifiers is based on WEKA (Waikato Environment for Knowledge Analysis)^[4] environment and JustNN for ANN.
4. Multiple ANNs are constructed with variable number of hidden layers, ranging from 1 to 4, using 1 to 50 nodes per layer with randomized initial weights.
5. Linear, polynomial and radius-based (RBF) kernels were used in SVM training.
6. The C4.5 algorithm was used for the decision tree classifier.

PERFORMANCE METRICS: The three commonly used performance metrics including accuracy, sensitivity and specificity were used to evaluate the classification technique.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

Results

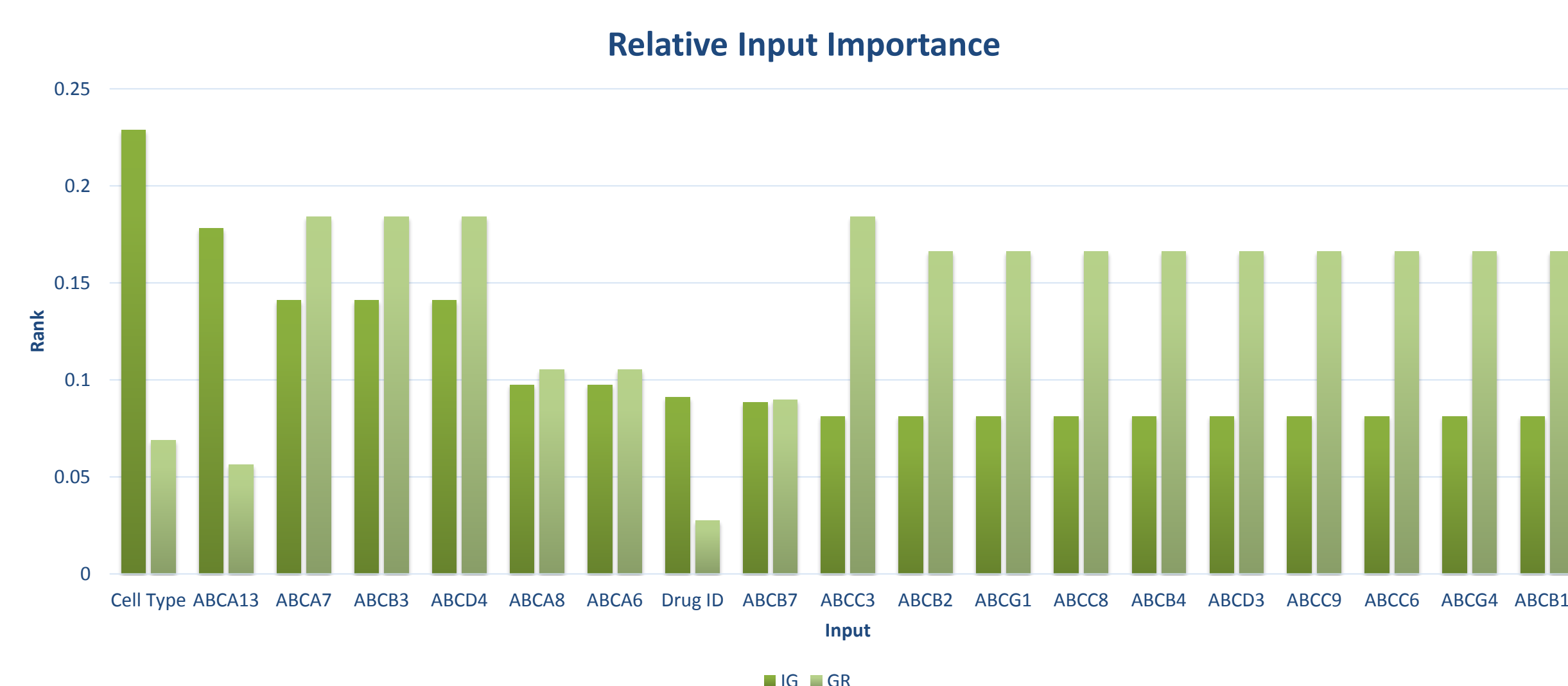


Figure 2: The ranking of the input attributes based on information gain (IG) and grain ratio (GR)

Results Cont.

In comparison with the other models, SVM has best overall performance. Comparative assessment of all the classification algorithms are presented in the table below.

Model	Accuracy	Sensitivity	Specificity
ANN	0.7786	0.5304	0.8534
SVM	0.8191	0.5217	0.9154
SVM*	0.8279	0.5217	0.9295
C4.5	0.7872	0.4782	0.8873

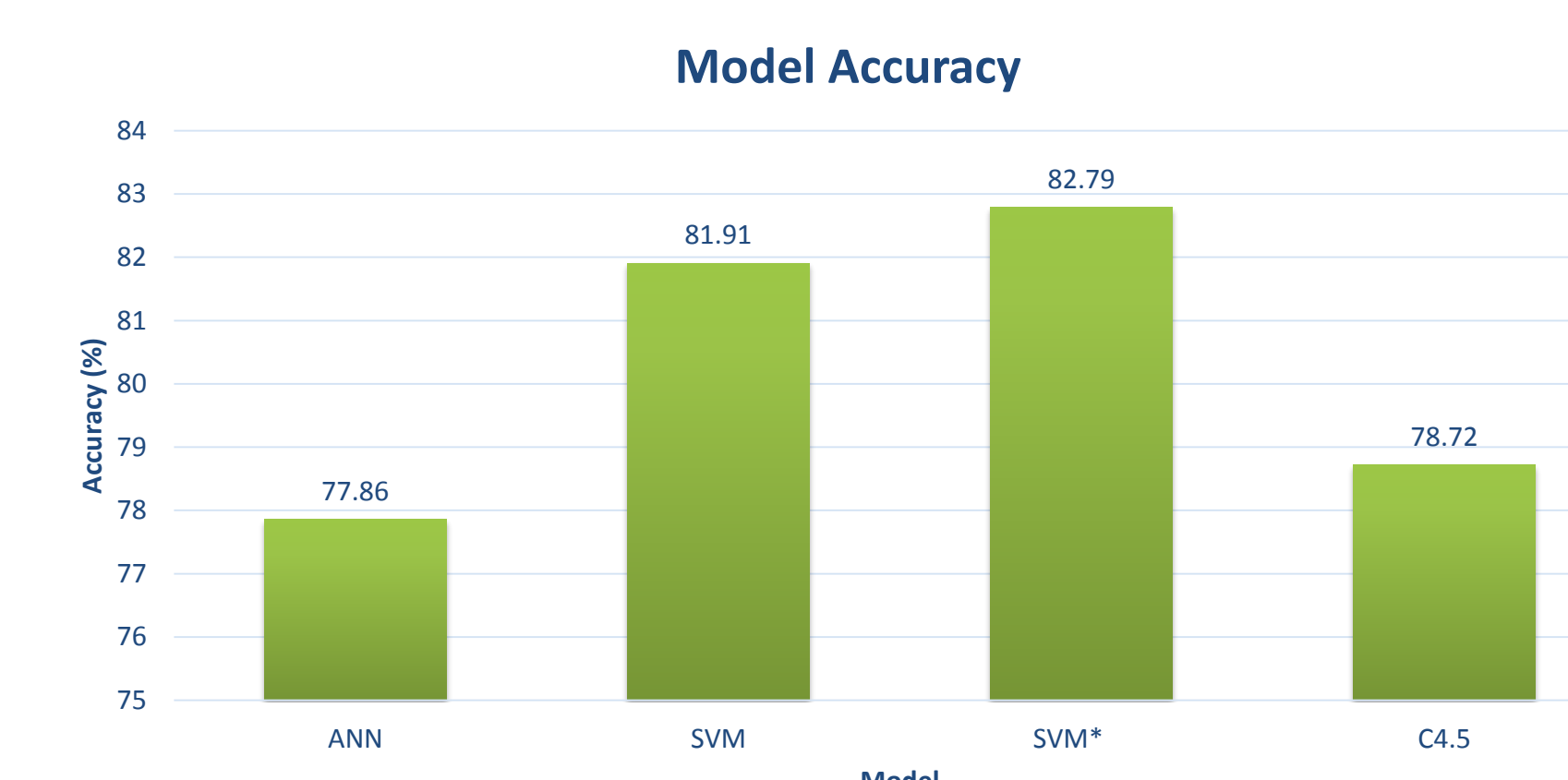


Figure 3: Accuracy (in percent) of the classification methods. SVM* represents the SVM modeled using only the top 17 inputs as defined by WEKA's attribute rankings

Conclusions

- Using supervised machine learning methods to predict drug sensitivity in cell lines is a useful approach.
- ABCA7, ABCD4 and ABCB3 appear to have the largest influence on drug sensitivity in the NCI-60's melanoma cell lines.
- ANN with two hidden layers and 30 neurons resulted in the highest accuracy. By running different experiments, one is able to decide the best ANN structure.
- For experiments involving SVM, the linear kernel resulted in the best classification accuracy among all the trials.
- Although decision tree is one of the most popular machine learning approaches in addressing classification problems, the decision tree algorithm used (C4.5) produced poor results.

Future Work

This study focused on NCI-60 based melanoma cell lines. The above models need to be further tested and refined through the use of other cancer cell types. We also hope to try and classify this data using unsupervised learning methods.

References

- [1] Szakács et al. (2004). Predicting drug sensitivity and resistance. *Cancer Cell*. 2004 Aug 1; 6(2):129-137.
- [2] Chakraborty A. (2013). Sanguine BioScience. "Multidrug-Resistance in Cancer: ABC-Transporters". Web. 6 Mar 2014.
- [3] Reinhold WC et al. (2012). CellMiner: A Web-Based Suite of Genomic and Pharmacologic Tools to Explore Transcript and Drug Patterns in the NCI-60 Cell Line Set. *Cancer Res*. 2012 Jul 15;72(14):3499-511.
- [4] Tan PN, Steinbach M, Kumar V (2005). *Introduction to Data Mining*. Addison-Wesley. 2005.
- [5] Hall M et al. (2009). *The WEKA Data Mining Software: An Update*; SIGKDD Explorations, Volume 11, Issue 1. 2009.