STEM Fellowship Big Data Challenge 2020

Deep learning transcriptomic model for prediction of pan-cancer chemotherapeutic sensitivity

Eddie Guo[1], Mehul Gupta[2], Pouria Torabi[1], and Sunand Kannappan[2]

[1]University of Alberta
[2]University of Calgary

May 31, 2020

Abstract

Emerging precision oncology studies have yet to generate a predictive biomarker that utilizes genes expression profiles to stratify tumours into similar pan-drug sensitivity profiles. This development would allow for identification of candidate drugs for treatment that maximize response and minimize cytotoxic burden. As such, this study utilized cell line sensitivity and molecular profiling data to generate a combinatorial gene expression predictive biomarker, utilizing feature selection and a deep learning architecture. Pan-cancer cohort of cell-line gene expression data from Genomics of Drug Sensitivity in Cancer (GDSC) was clustered into two response groups. Due to the high dimensional nature of the microarray data, biological agnostic feature selection was conducted to highlight genes with the highest predictive value. We developed and tuned a deep learning neural net architecture to predict response groups. Cell line clusters showed a significant difference in the therapeutic response, generated from IC50 values of all available chemotherapeutics. The feature space was reduced to 300 genes, which functional profiling indicated enriched primarily the focal adhesion, ECM-receptor and proteoglycan interaction pathways. Hyperparameter tuning of the deep learning model dictated a 4 hidden layer, diamond shaped architecture with a predictive accuracy of 91.7%.

Keywords
clustering, neural network, transcriptomics, chemotherapeutic response, combinatorial biomarker, molecular profile, therapeutic sensitivity, cancer

1 Introduction

With the advent of high-throughput technology, precision oncology approaches have utilized molecular characteristics of tumours to inform clinical decision-making, including chemotherapeutic regimen. The major focus of these approaches has been the development of targeted therapeutics, which are selective for specific genetic aberrations and expression profiles. Although these approaches may be extremely successful for inducing tumour response, tumours are more likely to gain resistance to therapies with specific targets [1]. Furthermore, not all tumours present with targetable features [1].

Emerging precision oncology approaches have begun to utilize high-throughput technology to potentiate the usage of conventional and less-targeted chemotherapy. Given that many of these less-targeted and consequently more cytotoxic chemotherapies have broad activity, the primary determinants of chemotherapeutic selection include cancer type and certain molecular markers [2]. Nevertheless, it is well established that tumour sensitivity to chemotherapy is heterogeneous between, but also within cancer types – resulting in a subset of patients that fail to respond to cancer type-specific chemotherapy while being exposed to significant side effect burden [3]. Given that evidence suggests that gene expression can mediate drug response, recent advances have utilized individual and combinatorial gene

expression biomarkers to develop predictors of tumour sensitivity to chemotherapeutic compounds [4].

While previous studies have developed predictive biomarkers for specific drugs, the utility of these biomarkers is limited to the determination of whether specific drugs may be effective for a patient [5]. However, a pan-drug predictive biomarker may provide significant clinical utility in the selection of candidate drugs for a certain patient for further investigation. Such a biomarker could be developed if pan-cancer tumours with similar expression have similar drug responses across all chemotherapies, with few exceptions. The availability of pan-cancer cell line databases with in-vitro drug sensitivity analyses along with accompanying gene expression profiling provides an ideal model for such analyses [6]. However, most previous drug sensitivity predictive biomarkers built on cell line databases have utilized classical machine learning combinatorial techniques - which fail to capture the dimensionality of available transcriptomic data. However, advanced deep learning algorithm approaches are capable of handling such dimensionality fail to allow interpretability, and consequently include transcriptomic data volume that is clinically infeasible [7]. As such, deep learning algorithms that utilize feature selection processes may maximize both the accuracy and functionality of such predictive biomarkers.

As such, we set out to generate a combinatorial gene expression predictor of pan-drug chemotherapeutic sensitivity across cell lines of all cancer types – to i) demonstrate that gene expression influences chemotherapeutic response across most drugs, and to ii) motivate future studies into the development of a clinically applicable predictor of candidate chemotherapeutics for tumours of a specific gene expression profile.

Following unsupervised clustering of cell lines into therapeutic response groups with similar pan-drug sensitivity, we show that cancer type does not stratify cell lines by therapeutic response. We utilize a feature selection algorithm to iteratively select and identify a subset of 300 relevant genes that influence chemotherapeutic sensitivity. We then generate a combinatorial predictive model from the feature selected genes utilizing deep neural networks.

2 Methods

Here we develop a deep learning model to accurately classify cancer cell lines into therapeutic response groups using data from the Genomics of Drug Sensitivity in Cancer (GDSC) consortium. Following data collection and curation, we utilized unsupervised learning algorithms to define two groups based on chemotherapeutic response. Next, we used a biologically agnostic feature selection algorithm, Boruta, to select statistically relevant genes for our neural network. We created an optimized neural network that utilizes transcriptomics features to classify patients into therapeutic response groups. See Fig. 1 for an overview of the data analysis pipeline.

Pan-cancer therapeutic response cohorts

To better understand the impact and predictive ability of transcriptomic dysregulation in chemotherapeutic response, a pan-cancer cohort of cell-line and associated therapeutic efficacy data were obtained from the Genomics of Drug Sensitivity in Cancer (GDSC) database. This database includes 1,110 cell lines from various different tumour types, and is thought to represent a relatively comprehensive pan-cancer dataset. In addition, the acquired dataset contained therapeutic efficacy information in the form of half-maximal inhibitory concentration (IC_{50}) values for 251 chemotherapies. These values correspond to the minimal concentration of therapeutic required to induce cell death in 50% of the cells cultured, with lower values being associated with improved drug efficacy. The data was used to generate a matrix with cell-line and accompanying therapeutic information. This dataset was filtered to exclude therapies with less than 80% of data for all cell-lines, followed by the exclusion of cell lines lacking response data for the drugs retained in the first step. This resulted in the inclusion of 548 cell lines (49.4% of the original cell lines) and 117 (46.6%) therapeutics for clustering based analysis.

Identification of pan-cancer therapeutic response cohorts

Cell line therapeutic response matrices obtained from the GDSC consortium were used to evaluate conventional tools used to separate patients into chemotherapeutic response groups as well as bifurcate candidate cell lines into defined response cohorts. Classical cancer classifications, including anatomic location and solid vs. non-solid tumour status, as well as broadly applicable molecular markers – TP53 and KRAS mutation status were used to evaluate whether classical classification systems were able to stratify cell lines into

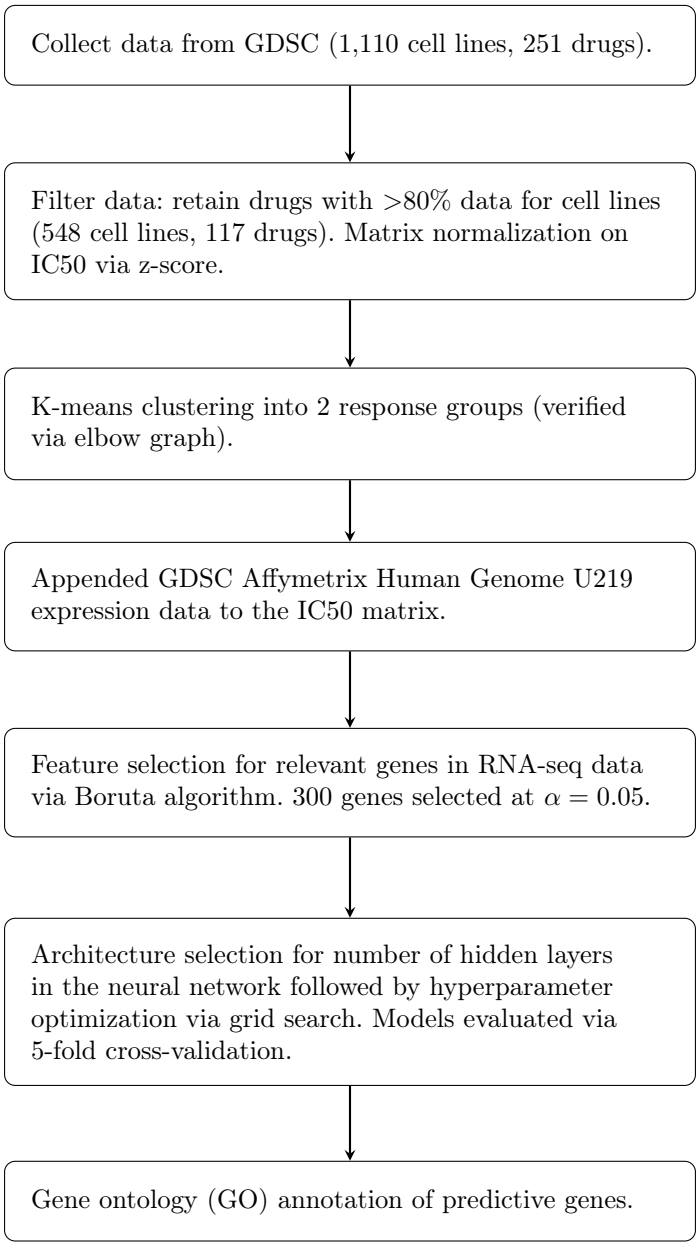


Figure 1: Summary of the data analysis pipeline.

chemotherapeutic response groups ([8], [9], [10], [11]). Cell-lines were separated into subgroups based on anatomic location, solid vs. non-solid tumour status, and mutation/non-mutation groups for TP53 and KRAS respectively. These subgroups were then plotted with the first and second principal components to identify if they clustered together. Observed clustering would be indicative of these classifiers being predictive of chemotherapeutic response.

Following evaluation of existing classifiers, we attempted to create defined cell-line clusters on the basis of the observed chemotherapeutic response of the pan cancer cell-line sample. We developed a Euclidean distance matrix for the retained cell lines based upon their pan-chemotherapy response. This matrix was then used to identify an optimal number of clusters capable of representing the therapeutic heterogeneity identified across the cancer cell lines. K-means clustering was then utilized to assign cell line candidates to appropriate therapeutic response cohorts. Generalized differences in chemotherapeutic efficacy between cohorts were visualized using a heatmap generated by the Pheatmap package in R. Separation between clusters was also visualized using principal component analysis with the factoextra package in R. Following the identification of defined clusters, differences in therapeutic efficacy between the identified cohorts were evaluated. Mann-Whitney U tests

were utilized to compare the half-maximal inhibitory concentration (IC_{50}) values between the groups. We attempted to select the least number of clusters that retained significant differences in IC_{50} values. False discovery rate (FDR) correction was utilized to correct for multiple comparisons.

Feature Selection

To develop a transcriptomic model predictive of therapeutic response clusters, expression data quantified by the GDSC consortium using the Affymetrix U219 microarray for each candidate cell line was obtained. Minimally processed CEL files were obtained from ArrayExpress (ascension number E-MTAB-3610), and processed using the affy package in R. The resulting normalized expression matrix for candidate cell lines was then merged with the existing dataset. This addition resulted in the loss of 7 cell lines (2 from cluster 1 and 5 from cluster 2), resulting in the inclusion of 541 cell-lines in model generation. The microarray dataset screened the expression levels of 16,382 genes, a model based on that many features is highly likely to overfit, compromising the generalizability of the model on new data. Such a large feature space also adds unnecessary noise and severely limits the accuracy and computational efficiency of the model [12]. Our approach to address these issues was dimensionality reduction through feature selection. The BorutaPy package in Python 3 is a feature selection algorithm based on Random Forest classification which iteratively removes features that are statistically less significant than a shuffled version of the same feature [13].

Classification using an optimized neural network

The neural net was constructed using the Tensorflow Keras sequential deep learning API in Python 3. The model underwent multiple instances of optimization, starting with the manipulation of the overall hidden layer architecture. The classifier's predictive accuracy and misclassification rate were monitored to determine the optimal number of dense hidden layers (Fig. 4a) in addition to iterative manipulation of the number of neurons in each hidden layer. The rectified linear unit (ReLu) was chosen as the neuronal activation function for all the layers except the output layer which used a sigmoid activation as a means of classifying instances into binary classes.

The model was rigorously monitored for and protected against overfitting on the training dataset. To address this, we employed dropout layers with a 0.3 dropout rate and batch normalization layers were employed to improve the generalizability of the model.

The dataset was randomly segregated using the Pareto principle where we reserved 80% of the data for training and the remaining 20% for validation [14]. Model selection was performed by tuning the model's hyperparameters via a grid search and 5-fold cross-validation (Table 2). We performed a grid search with 3-fold cross-validation on the training data (80% of the dataset; 432 training samples, 541 overall) to determine the parameters under which to minimize the binary cross-entropy loss function. GridSearchCV from the scikit-learn library was used as a means of iterating through multiple possibilities of epochs, batch size, optimizer, and kernel initializer to find the optimal model. To prevent class imbalance during training, we used the Synthetic Minority Oversampling Technique (SMOTE) from the imblearn package for Python 3. Each model's performance was evaluated by the area under the receiver operating characteristic (ROC) curve (AUC). Performance evaluation of the final model was performed with the testing set.

3 Results

Clustering of pan-cancer cell lines identifies two distinct therapeutic response cohorts

We assess the ability of common molecular and clinical characteristics used in clinical decisions to stratify cell lines into groups with similar chemotherapeutic performance by subgrouping cell lines based upon these criteria and plotting them against the first and second principal components. Commonly used measures including the system of origin, molecular subtype, as well as TP53, and KRAS mutation status failed to identify defined clusters of cells with similar therapeutic responses (Fig. 2). This likely indicates that these widely used classifiers may be inadequate to stratify patients, and thus improved therapeutic classification system may prove useful for clinical and research settings.

To identify defined cohorts of pan-cancer cell-lines with similar trends in therapeutic efficacy, we employed unsupervised clustering of retained cell-lines. Principal component analysis was used to reduce the dimensionality of the dataset, allowing for visualization of defined therapeutic response cohorts. This process identified two distinct clusters of therapeutic efficacy (Fig. 3a), 362 cell lines identified in

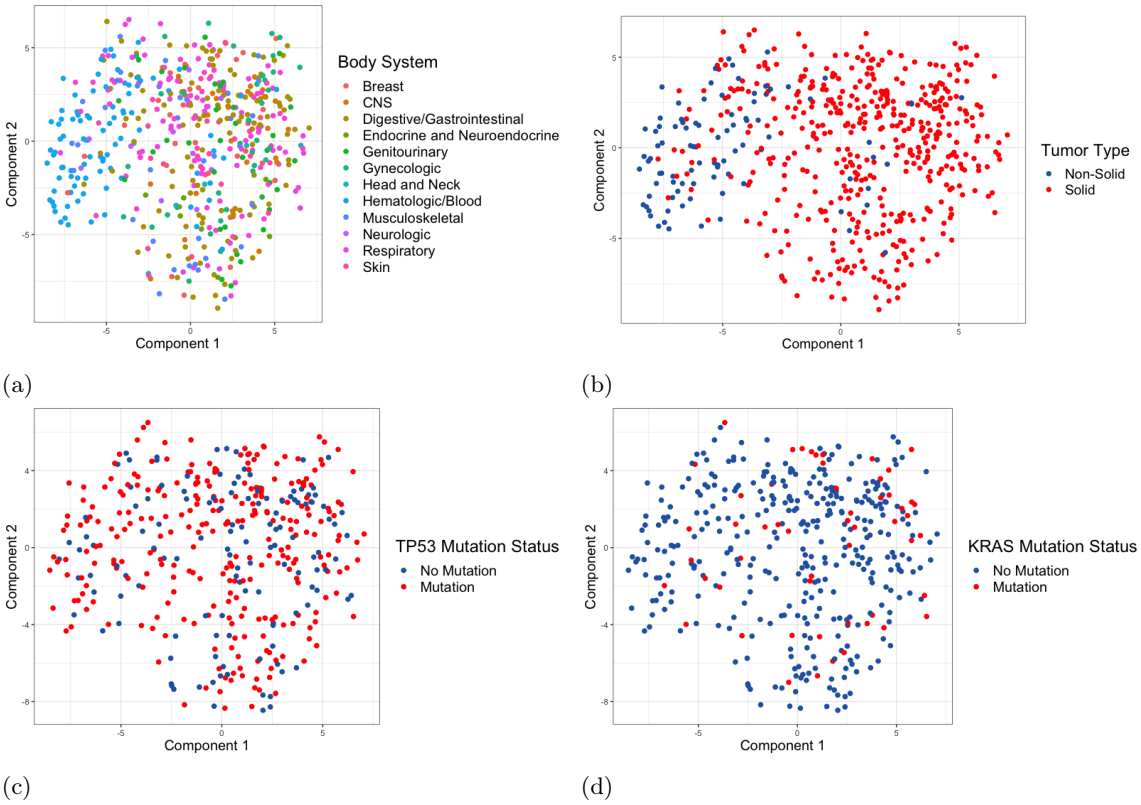


Figure 2: Principal component analysis of pan-cancer cell line therapeutic efficacy, generated from IC_{50} values of all available chemotherapeutics.The horizontal axis shows the first principal component, the vertical axis the second component. Cell lines are visualized based on major cancer type classifications, including (2a) body system of tumour and (2b) solid vs. non-solid tumour status. Cell lines were also visualized on major molecular markers, including (2c) TP53 mutation status, and (2d) KRAS mutation status. Legends demonstrate visualized colour.

response group 1, and 186 cell lines identified in response group 2. Further analysis of these clusters demonstrate that a subset of therapeutics performs substantially differently between the different cohorts (Figure x - heatmap). To quantify differences in therapeutic response between clusters, IC50 values were compared between candidate cell lines (Fig. 3b). Of the 117 therapies included, 95 had significant differences in efficacy between the two cohorts identified. This suggests that these cohorts represent groups of cell lines with vastly different therapeutic responses. Therefore the ability to accurately stratify into these cohorts may be a valuable tool for stratification prior to chemotherapeutic treatment.

Boruta selects 300 genes

The Boruta feature selection algorithm identified 300 relevant genes from the original set of 16,382 genes at $\alpha = 0.05$ with a maximum tree depth of 5. Of importance, the algorithm selects genes that are estimated to have highest predictive value rather than biological significance. This biologically agnostic feature selection was key to limiting preliminary bias in the model. Functional profiling indicated that the top three pathways enriched by the selected genes were pathways in focal adhesion, extracellular matrix-receptor, and proteoglycan interactions in cancer.

A neural network with four hiddeen layers accurately classifies patients into responder and non-responder cohorts

Unsupervised learning in the form of the K-means clustering of the cancer cell line transcriptomes indicated substantially different responses to chemotherapies. Using these distinct therapy response cohorts, we developed a deep learning binary classification to predict drug response based on transcriptome data. We initially analyzed five neural network architectures, each corresponding to 1-5 hidden layers (Fig. 6). Hyperparameter optimization via grid search returned similar results for each model: 50 epochs, batch size of 32, Adagrad as the optimizer, and a normal kernel initializer.

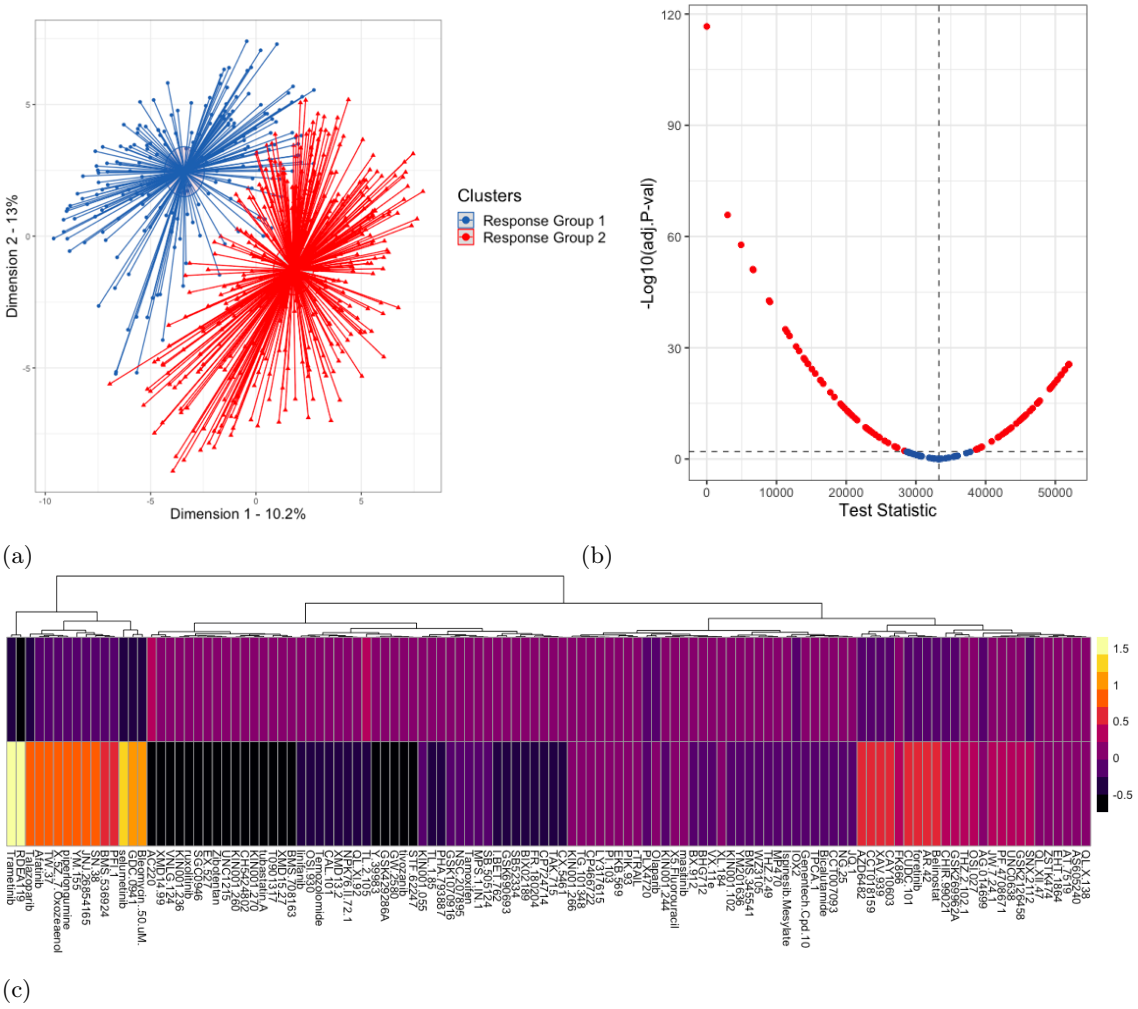
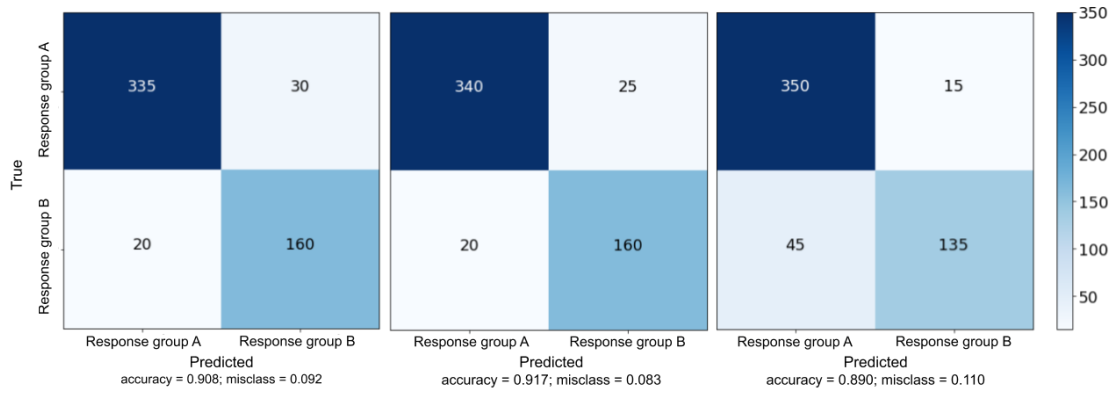


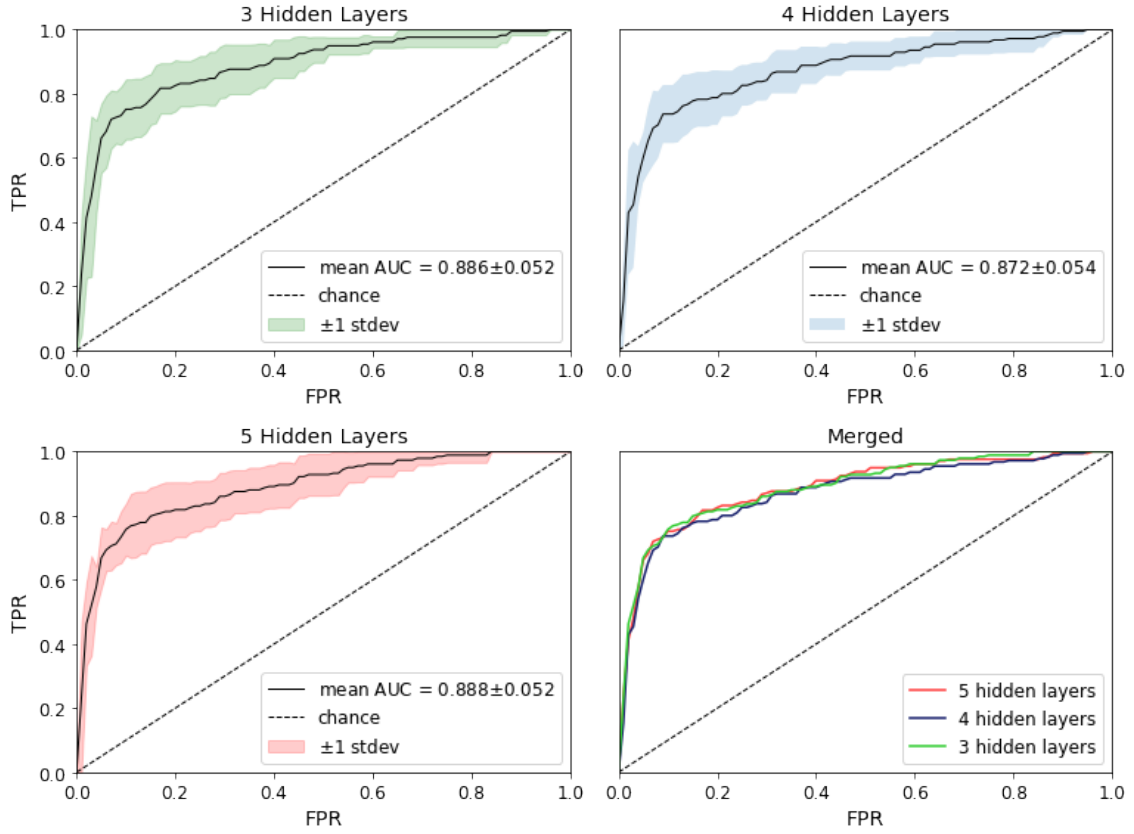
Figure 3: (3a) Principal component component analysis of pan-cancer cell line therapeutic efficacy, generated from IC_{50} values of all available chemotherapeutics. The horizontal axis shows the first principal component, the vertical axis the second component. The two identified therapeutic response clusters are indicated in red and blue respectively. (3b) Volcano plot identifying chemotherapeutics with significantly different IC50 values between therapeutic response clusters. Drugs identified in red meet the criteria for significance (FDR adjusted $p < 0.05$). (3c) Heatmap of therapeutic IC50 for the two identified therapeutic response clusters. Columns represent individual chemotherapies, and are clustered according to euclidean distance. Colours range from yellow to purple, with a shift toward the latter indicating increased efficacy of the corresponding chemotherapeutic.

Neural network architectures containing 3-5 hidden layers performed similarly with approximately 90% accuracy. The architectures with 1 and 2 hidden layers performed less optimally with 80% accuracy. We proceeded to validate the architectures with 3-5 hidden layers using 5-fold cross-validation. Of note, the model with 4 hidden layers had the lowest false positive rate (FPR = 3.67%) and a false negative rate (FNR = 4.59%). The model with five hidden layers had the highest FPR of the models evaluated (8.26%).

A receiver operating characteristic (ROC) curve was plotted for each of the neural network variants as an alternative evaluative method under uneven class sizes (Fig. 4b). The relative ratio between the model's false positive classification rate and its true positive rate was averaged between 5 K-Folds. The mean Area Under Curve (AUC) for the 5 trials was used to compare the three network architectures. However, the differences between the various architectures for AUC was not significatly different. Consequently, confusion matrices (Fig. 4a) and associated misclassification rates were used to pick the optimal model. The neural net with 4 hidden layers boasted the best performance overall with a 91.7% accuracy and



(a)



(b)

Figure 4: (4a) From left to right: confusion matrices for the 3, 4, and 5 hidden layer neural network models evaluating the true positive, false positive, true negative, and false negative rate. The models classify patient RNA-seq datasets into chemotherapy response cohorts. (4b) ROC curves for 3, 4, and 5 hidden layers neural network models with confidence bands of ±1 standard deviation. The models classify patient RNA-seq datasets into chemotherapy response cohorts. Each model was subject to 5-fold cross-validation and the mean across all trials was plotted.

an 8.3% misclassification rate.

These results suggest that the developed transcriptomic deep learning model is able to accurately classify patients into therapeutic response cohorts. To better understand the underlying transcriptomic heterogeneity underlying the therapeutic response cohorts, a KEGG pathway enrichment analysis was performed on predictive genes used in the deep learning model. This analysis identified enrichment of gene sets associated with focal adhesion and ECM interaction. In addition, the set was enriched with genes associated with PI3K signalling and leukocyte invasion among

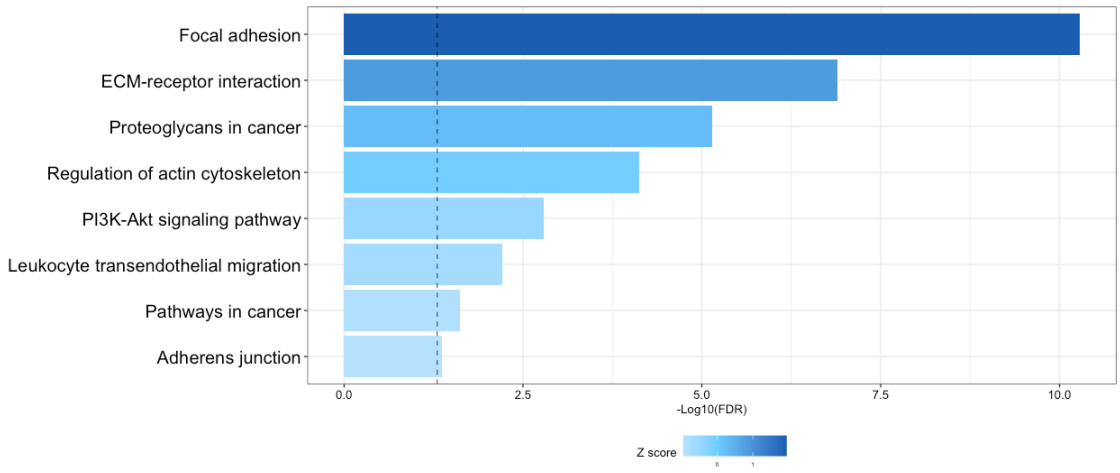


Figure 5: KEGG pathway functional enrichment for predictive genes included in the deep learning model conducted using g:Profiler.

predictive transcripts (Fig. 5).

4 Discussion

In this study we developed an accurate deep learning-based model for classifying cancer patients based on RNA expression data into therapeutic response cohorts on a relatively small cell line therapeutic response dataset. Using unsupervised clustering techniques, we segregated cancer patients into two defined therapeutic response groups with significantly different responses to a multitude of standard chemotherapies. [EXPAND HERE]

The relatively low accuracy of the neural networks with 1 and 2 hidden layers (82.6% and 70.8% respectively) suggests that the therapeutic response cohorts cannot be separated by a linear classifier. Furthermore, the high FNR of the network with 5 hidden layers as compared to the 3 and 4 hidden layer networks indicates overfitting. To this end, either a 3 or 4 hidden layer network is the ideal architecture for analyzing our data. Interestingly, there is no significant difference between AUC for these models (Fig. 4b). This indicates that the major factor for model evaluation above two hidden layers for this dataset are the FPR and FNR of the classifications.

A major limitation of our study was the availability of large datasets to train our model. Here we faced a $p \gg n$ problem as machine learning models expect that the number of features p is much larger than the number of observations n. To minimize this bias, we applied the Boruta algorithm to reduce our 16,382 genes by 541 cell lines dataset to a 301 by 541 matrix. The algorithm has been shown in various journals to be an effective feature selector method in high dimensional omics datasets [15]. To prevent overfitting of the reduced matrix, we applied batch normalization and dropout layers immediately preceding each hidden layer. Of note, a neural network architecture where the initial hidden layers diverge and the latter hidden layers converge provides the most accurate classifications of the RNA-seq data.

Future investigations will look to validate the efficacy of the model in prediction of chemotherapy response in various forms of cancer using the current transcriptomic signature. It is likely that the model accuracy will vary between therapy targets, as such, further studies can make use of our project pipeline to create a stratified model whereby the drug class and target are additional inputs. Furthermore, it may be interesting to select relevant genes using a different feature selector. The selected algorithm in this paper, Boruta, operates on patterns of statistical relationships rather than biological relationships. Our use of Boruta was inspired by its efficacy demonstrated by prior studies of the algorithm as compared to other feature selectors [15, 16]. Computational efficiency and the resulting feature set quality were also motivators for choosing BorutaPy over other selection algorithms such as univariate selection or principle component analysis. It is possible that a feature selection method informed by gene function and linkage disequilibrium could yield a different set of relevant genes.

5 Conclusions

Using transcriptomics data from the cancer cell lines, two chemotherapeutic response clusters were identified via unsupervised learning in the form of K-means Clustering. A feature selection algorithm was used to select a 300 gene signature which served as inputs to multiple neural networks. We determined that the network with 4 hidden layers was the most accurate model, producing a binary classifier to predict patient therapy response with 91.7% accuracy. Future studies will investigate the efficacy of our model to predict chemotherapy response in various forms of cancer.

Acknowledgements

We wish to acknowledge the STEM Fellowship for organizing the 2020 Big Data Challenge, as well as Roche, SAS, Canadian Science Publishing, Digital Science, Altmetric, and Overleaf for their contributions that enabled this competition. We would like to thank our mentor, Dr. Daiva Nielsen for her feedback on our paper.

References

[1] Akshat Pathak, Sanskriti Tanwar, Vivek Kumar, and Basu Dev Banarjeec. Present and future prospect of small molecule & related targeted therapy against human cancer. *Vivechan Int J Resp*, 9(1):36–49, March 2018.

[2] Brian A Baldo and Nghia H Pham. Adverse reactions to targeted and non-targeted chemotherapeutic drugs with emphasis on hypersensitivity responses and the invasive metastatic switch. *Cancer Metastasis Rev*, 32(3-4):723–761, December 2013.

[3] Joseph A. Sparano, Robert J. Gray, Della F. Makower, Kathleen I. Pritchard, Kathy S. Albain, Daniel F. Hayes, Charles E. Geyer Jr, Elizabeth C. Dees, Matthew P. Goetz, Jr. John A. Olson, Tracy Lively, Sunil S. Badve, Thomas J. Saphner, Lynne I. Wagner, Timothy J. Whelan, Matthew J. Ellis, Soonmyung Paik, William C. Wood, Peter M. Ravdin, Maccon M. Keane, Henry L. Gomez Moreno, Pavan S. Reddy, Timothy F. Goggins, Ingrid A. Mayer, Adam M. Brufsky, Deborah L. Toppmeyer, Virginia G. Kaklamani, Jeffrey L. Berenberg, Jeffrey Abrams, and George W. Sledge. Adjuvant chemotherapy guided by a 21-gene expression assay in breast cancer. *N Engl J Med*, 379(2):111–121, July 2018.

[4] Kevin Shee, Jason D. Wells, Amanda Jiang, and Todd W. Miller. Integrated pan-cancer gene expression and drug sensitivity analysis reveals slfn11 mrna as a solid tumor biomarker predictive of sensitivity to dna-damaging chemotherapy. *PLoS One*, 14(11):e0224267, November 2019.

[5] Xuewei Wang, Zhifu Sun, Michael T Zimmermann, Andrej Bugrim, and Jean-Pierre Kocher. Predict drug sensitivity of cancer cells with pathway activity inference. *BMC Med Genomics*, 12(15), January 2019.

[6] Wanjuan Yang, Jorge Soares, Patricia Greninger, Elena J Edelman, Howard Lightfoot, Simon Forbes, Nidhi Bindal, Dave Beare, James A Smith, I Richard Thompson, Sridhar Ramaswamy, P Andrew Futreal, Daniel A Haber, Michael R Stratton, Cyril Benes, Ultan McDermott, and Mathew J Garnett. Genomics of drug sensitivity in cancer (gdsc): A resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res*, 41:D955–61, November 2012.

[7] Mehreen Ali and Tero Aittokallio. Machine learning and feature selection for drug response prediction in precision oncology applications. *Biophys Rev*, 11(1):31–39, February 2019.

[8] Stacey Shiovitz and William M Grady. Molecular markers predictive of chemotherapy response in colorectal cancer. *Curr Gastroenterol Rep*, 17(2):431, February 2015.

[9] Vincenzo Catalano, Anna Maria Baldelli, Paolo Giordani, and Stefano Cascinu. Molecular markers predictive of response to chemotherapy in gastrointestinal tumors. *Crit Rev Oncol Hemat*, 38(2):93–104, May 2001.

[10] Ikuo Sekine, John D Minna, Kazuto Nishio, Tomohide Tamura, and Nagahiro Saijo. A literature review of molecular markers predictive of clinical response to cytotoxic chemotherapy in patients with lung cancer. *J Thorac Oncol*, 1(1):31–37, January 2006.

[11] I F Faneyte, J G Schrama, J L Peterse, P L Remijnse, S Rodenhuis, and M J van de Vijver. Breast cancer response to neoadjuvant chemotherapy: predictive

markers and relation with outcome. *Br J Cancer*, 86:406–412, February 2003.

[12] Huan Liu. *Feature Selection*, pages 402–406. Springer US, Boston, MA, 2010.

[13] Rudnicki W. Kursa, M. Feature selection with the boruta package. *Journal of Statistical Software*, 36(11), 2010.

[14] Shie-Yui Liong, Soon-Thiam Khu, and Weng-Tat Chan. Derivation of pareto front with genetic algorithm and neural network. *Journal of Hydraulic Engineering*, 6(1), January 1998.

[15] Frauke Degenhardt, Stephan Seifert, and Silke Szymczak. Evaluation of variable selection methods for random forests and omics data sets. *Brief Bioinform*, 20(2):492–503, March 2019.

[16] Theodore Sakellaropoulos, Konstantinos Vougas, Sonali Narang, Filippos Koinis, Athanassios Kotsinas, AlexanderPolyzos, Tyler J Moss, Sarina Piha-Paul, Hua Zhou, Eleni Kardala, Eleni Damianidou, Leonidas G Alexopoulos, Iannis Aifantis, Paul A Townsend, Mihalis I Panayiotidis, Petros Sfikakis, Jiri Bartek, Rebecca C Fitzgerald, Dimitris Thanos, Kenna R Mills Shaw, Russell Petty, Aristotelis Tsirigos, and Vassilis G Gorgoulis. A deep learning framework for predicting response to therapy in cancer. *Cell Reports*, 29(11):3367–3373, December 2019.

Supplementary Data

Figures

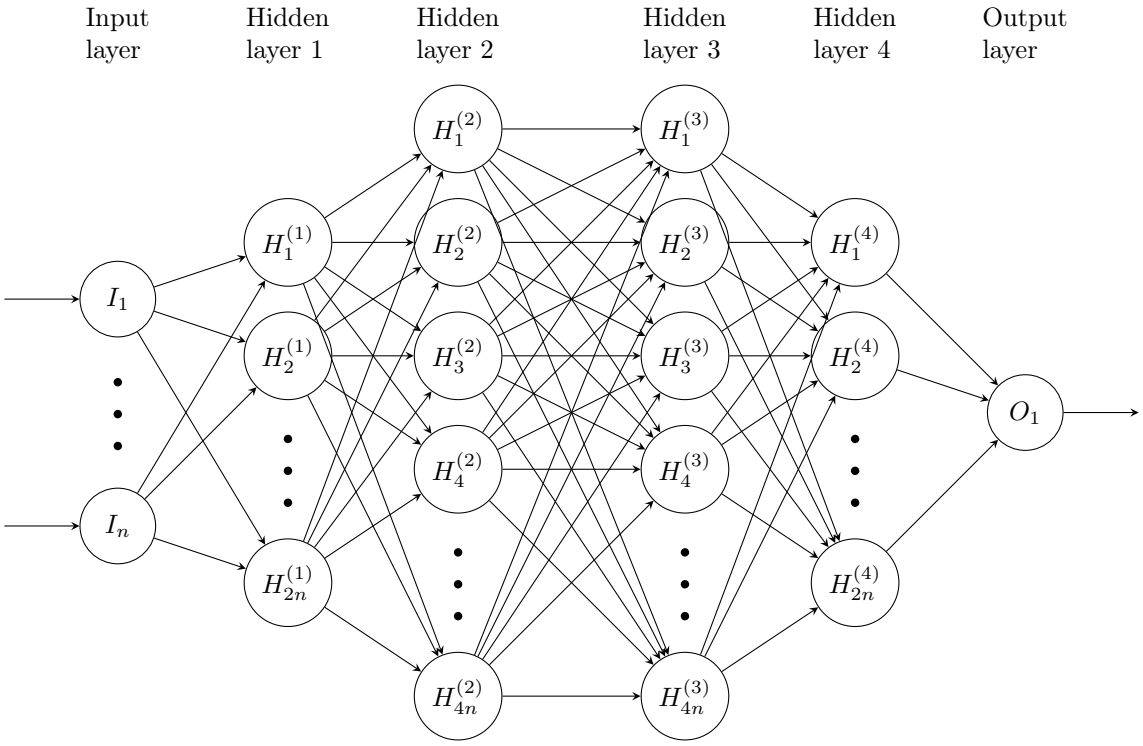


Figure 6: Neural network architecture representation with four hidden layers ($n = 300$). Inputs include the feature-selected genes from the RNA-seq dataset. Each hidden layer has a dropout rate of 0.3 and is subject to batch normalization.

Tables

Table 1: Grid search parameters to optimize all neural network architectures presented in this paper (3, 4, and 5 hidden layers). Each grid search underwent 3-fold cross-validation on the training data.

Epochs	Batches	Optimizer	Kernel initializer
25	15	Stochastic gradient descent	Normal
50	32	Adagrad	Uniform
75	64	Adam	Glorot uniform

Table 2: Neural network architectures and the number of neurons per layer ($n = 560$).

Architecture	Number of neurons
5 hidden layers	300 inputs, 2n, 4n, 4n, 2n, n, binary output
4 hidden layers	300 inputs, 2n, 4n, 4n, 2n, binary output
3 hidden layers	300 inputs, 2n, 4n, 2n, binary output