LA-UR-16-27376
Approved for public release; distribution is unlimited.

Title:	A Review of Current Machine Learning Methods Used for Cancer Recurrence Modeling and Prediction
Author(s):	Hemphill, Geralyn M
Intended for:	Report for Pilot 3 of NCI Project
Issued:	2016-09-27

A Review of Current Machine Learning Methods Used for Cancer Recurrence Modeling and Prediction

Cancer has been characterized as a heterogeneous disease consisting of many different subtypes.[17] The early diagnosis and prognosis of a cancer type has become a necessity in cancer research. A major challenge in cancer management is the classification of patients into appropriate risk groups for better treatment and follow-up. Such risk assessment is critically important in order to optimize the patient's health and the use of medical resources, as well as to avoid cancer recurrence.[15,27]

The classification of patients into high or low risk groups has led many biomedical and bioinformatics researchers to study the application of machine learning methods. Machine learning is a suite of techniques within the broad category of predictive analytics. Machine learning methods utilize a variety of statistical, probabilistic, and optimization techniques that allow computers to learn and detect patterns from large, noisy, or complex datasets. This is particularly well-suited to medical applications, particularly when those applications depend on complex proteomic and genomic information.[8] Over the past several decades, machine learning has been frequently used in cancer diagnosis and detection. More recently, it has been applied to cancer prognosis and prediction. The latter is part of a growing trend towards personalized, predictive medicine. This is important not only for the wellbeing of patients but also for the physicians in making the best treatment decisions for their patients.

The fundamental goals of cancer prognosis and prediction are distinct from the goals of cancer diagnosis and detection. In cancer prognosis/prediction there are three predictive objectives: 1) the prediction of cancer susceptibility (risk assessment); 2) the prediction of cancer recurrence; and 3) the prediction of cancer survivability. In the first case, one is trying to predict the likelihood of developing a type of cancer prior to the occurrence of the disease. In the second case, one is trying to predict the likelihood of redeveloping cancer after the apparent resolution of the disease. In the third case, one is trying to predict an outcome (such as life expectancy, survivability, progression) after diagnosis of the disease.[8]

This paper focuses on the application of machine learning methods for predicting the likelihood of a recurrence of cancer. It is not meant to be an extensive review of the literature on the subject of machine learning techniques for cancer recurrence modeling. Other recent papers have performed such a review, and I will rely heavily on the results and outcomes from these papers. The electronic databases that were used for this review include PubMed, Google, and Google Scholar. Query terms used include "cancer recurrence modeling", "cancer recurrence and machine learning", "cancer recurrence modeling and machine learning", and "machine learning for cancer recurrence and prediction". The most recent and most applicable papers to the topic of this review have been included in the references.

Over the past three decades, machine learning methods have been primarily used in cancer diagnosis and detection to identify, classify, detect, or distinguish tumors and other malignancies. It has only been recently that researchers have applied machine learning towards cancer prediction and prognosis. As a result, the body of literature pertaining to the field of machine learning in cancer prediction/prognosis is relatively small.[8]

In addition to the recent increase in the application of machine learning to cancer prediction and prognosis, some of the more obvious trends in recent years include a growing reliance on protein markers and microarray data, a trend towards using a combination of proteomic and clinical data, and a strong bias towards applications in prostate and breast cancer.[8] Among the better designed and validated studies, there appears to be a substantial improvement in the accuracy of cancer susceptibility and outcome prediction using machine learning methods as compared to simple statistical methods. In almost all of these studies, gene expression profiles, clinical variables, and histological information have been incorporated in developing the model.

Machine learning methods appear to have been successfully used in predicting outcomes or risks in nearly a dozen different kinds of cancers. This suggests that machine learning methods in general can be applied to the objective of cancer prediction and prognosis.[8]

Machine learning techniques such as Artificial Neural Networks and Decision Trees have been used for nearly three decades in cancer detection.[17] A growing trend noted in the last decade is the use of other machine learning methods such as Support Vector Machines and Bayesian Networks towards the goal of cancer prediction and prognosis. From a 2007 review of machine learning methods used in cancer prediction and prognosis, it was stated that almost 70% of all reported studies used Artificial Neural Networks as their primary (and sometimes only) method for prediction. Support Vector Machines were a distant second at 9%, and clustering and decision trees each accounted for about 6% of the applications. What I have found is that Artificial Neural Networks are still the most used method, followed by Support Vector Machines.

Machine learning techniques are popular due to their outstanding performance in handling large-scale datasets with noisy or missing data. The ability of these methods to detect key features from complex datasets amplifies their importance. A variety of these techniques, including Artificial Neural Networks, Bayesian Networks, Support Vector Machines, and Decision Trees have been widely used in cancer research for the development of predictive models, resulting in more accurate and effective decision making.[17]

When applying a machine learning method, it is important to know the specific type of data being used, as this will allow the right tools and techniques to be selected for analysis. Data quality issues such as the presence of noise, outliers, missing or duplicate data, and biased data need to be addressed prior to analysis. In addition, often the data needs to be preprocessed before it can be used in a specific machine learning method. Among preprocessing techniques are dimensionality reduction, feature selection, and feature extraction. The reduction of

dimensionality can eliminate irrelevant features, reduce noise, and produce a more robust model. Some machine learning methods incorporate these preprocessing steps into the model building. The main objective of machine learning techniques is to produce a model that can be used to perform classification, prediction, or estimation.

Success in machine learning is not always guaranteed. A good understanding of the problem and an appreciation of the limitations of the data are critical. One must also understand the assumptions and limitations of the methods being applied. Not all machine learning methods are applicable to all situations. Some are better for certain kinds of problems than others, and it is important to recognize this. Some methods may have assumptions or data requirements that render them inapplicable to the problem being analyzed. Knowing which method is best for any given problem is crucial. For this reason, it is advisable to try several different machine learning methods for any given set of data in order to compare results and see which performs the best and which produces the best predictions or has the smallest classification error rate.

In a more recent review of machine learning methods in cancer prognosis and prediction (2014), it was stated that in many studies, several machine learning techniques were used and compared to find the most optimal one.[17] I have noticed that in my review as well. In the 2014 review, it was also stated that most studies created predictions by integrating either genomic, clinical, histological, imaging, demographic, epidemiological, and proteomic data or different combinations of these.[17] The integration of multidimensional heterogeneous data, combined with the application of different techniques for feature selection and classification, can provide promising tools for inference in the domain of cancer prediction and prognosis.

In a retrospective study, the focus was on the prediction of cancer survival using electronic administrative records and a cancer registry[11]. The approach taken in this study created a model across a wide range of cancers, whereas most other machine learning models are derived for single cancer types or for a limited range of cancers. An advantage stated for this general approach is the ability to predict outcomes in less common cancers where limited data might preclude the development of specific models. An alternative to this generalized approach would be to borrow information across different cancer types as in multitask learning.

Machine learning techniques are classified into three broad categories: Supervised Learning, Semi-supervised Learning, and Unsupervised Learning. Supervised Learning takes a known set of input data and known responses to the data and trains a model to generate reasonable predictions for the response to new data. Examples in this class of methodologies include decision tree techniques (CART), Bayesian methods (Naïve Bayes and variations thereof, Bayesian Model Averaging), Artificial Neural Networks, Instance Based Learning, (K Nearest Neighbors), and Ensemble Methods (Boosting, Bagging, AdaBoost, Gradient Boosting Machines, Gradient Boosted Regression Trees, and Random Forest).

Outcome prediction in cancer research usually relies on traditional supervised learning techniques in which only labeled data can be used for learning, while unlabeled data is

disregarded. Labeled data refers to data from samples with clinical follow-up, and unlabeled data is data from samples without clinical follow-up. Recent studies suggest that unlabeled data, when used in conjunction with the limited amount of labeled data, can produce improvement in learning accuracy.[27]

The category of methods that combines both labeled and unlabeled data is referred to as Semi-supervised Learning. These methods are used for the same applications as Supervised Learning. They are especially useful when the cost associated with labeling is too high to allow for a fully labeled training process. Examples of Semi-supervised Learning methods include regression algorithms such as Ordinary Least Squares Regression, Linear Regression, Logistic Regression, and Stepwise Regression.

Examples of Unsupervised Learning include Clustering Methods (K Means and Hierarchical Clustering) and Principal Components. Unsupervised methods provide a descriptive model. In this case, there is no target to learn. The goal is to explore the data and find some structure within.

From the case study papers that I have reviewed, the following modeling and classification methods were used to predict cancer recurrence. Not all of these methods are included under the category of standard machine learning methods. Most notably absent from this list are the Boosting methods such as Gradient Boosting Machines (GBM), AdaBoost, and XGBoost, and Bootstrapped Aggregation methods (Bagging).

1. **Support Vector Machine (SVM)**
   A SVM maps the input vector into a feature space of higher dimensionality and identifies the hyperplane that separates the data points into two classes. The hyperplane can be calculated by one of many different kernels, including nonlinear kernels. The marginal distance between the decision hyperplane and the instances that are closest to the boundary is maximized. The resulting classifier achieves considerable generalizability and can be used for the classification of new samples. Probabilistic outputs can also be obtained. SVMs can be used to perform both linear and nonlinear classification. Applying different kernels to different data sets can dramatically improve the performance of the classifier.[11]
2. **Hierarchical Forward Selection (HFS) wrapper for SVM**
   This method starts from an empty set and then iteratively searches forward in the binary feature subset space. During an iteration, all feature subsets which perform better than their predecessors in the previous iteration are retained using a predefined criterion as the evaluation measure. This process is iterated until a stopping criterion is reached.[22]

3. **Proximal Support Vector Machine**
   The proximal support vector machine is a computationally efficient alternative to the standard SVM. It classifies points depending on proximity to one of two parallel planes that are pushed as far apart as possible, whereas the traditional SVM maximizes the margin between support vector data points.[32]
4. **Artificial Neural Network (ANN)**
   The ANN was originally designed to model the way the brain works with multiple neurons being interconnected to each other through multiple axon junctions. ANNs are composed of an input layer (having neurons for all input variables), a hidden layer (consisting of any number of hidden neurons), and an output layer. Each neuron processes its inputs and transmits its output value to the neurons in the subsequent layer. Each connection between neurons is given a weight during the training phase. The network learns by adjusting the weightings when presented with a combination of inputs and outputs that are known (the training data).[6] Most ANNs are structured using a multi-layered feedforward architecture, meaning there is no feedback. The network's multiple layers of interconnecting neurons have the ability to analyze, learn, and recognize patterns from the input data. Through this process, the network makes predictions about specific outcomes.[16]
5. **Generalized Regression Neural Network (GRNN)**
   This method is based on a radial basis function that can be used for regression, classification, and time series predictions.[5] A radial basis function is a feedforward network consisting of a hidden layer of radial kernels and an output layer of linear neurons. The hidden layer performs a non-linear transformation of input space, and the output layer performs linear regression to predict the desired targets. Although several forms of radial basis functions may be used, Gaussian kernels are the most common.
6. **Probabilistic Neural Network (PNN)**
   This technique is a Bayes-Parzen classifier. It models the Bayesian classifier and minimizes the risk of misclassification. A disadvantage of PNN is the lack of information about the class probability distributions, but an advantage is the better generalization and convergence properties when compared to that of a Bayesian classifier.[5]
7. **Recurrent Neural Network (RNN)**
   This method is a class of Artificial Neural Networks where connections between units form a directed cycle. This creates an internal state of the network that allows it to exhibit dynamic temporal behavior. RNNs can use their internal memory to process arbitrary sequences of inputs, unlike feedforward neural networks. RNNs make use of sequential information. They are called recurrent because they perform the same task for every element of a sequence, with the output depending on the previous computations. You can think of RNNs as having a memory of what has been calculated so far. In contrast, a traditional neural network assumes that all inputs and outputs are independent of each other.

8. **Radial Basis Function Network**

   This is a subtype of Artificial Neural Networks that uses a linear combination of radial basis functions for interpolating the function that maps the variables (features) to the class. (See 5. above for a definition of radial basis function.) It is similar to Logistic Regression in that it outputs a numeric variable that can determine a binary output by selecting a threshold value. It also has the ability to group data using criteria other than just a linear separator.[9]

9. **Random Forest**

   Random Forest is a specific type of ensemble algorithm. With this type of classifier, N decision tree classifiers are made with the intent that the data is run through all of these N classifiers. The final class is based on one of a number of mechanisms such as a weighted average of the individual Decision Trees or a voting majority of the individual Decision Trees. A Random Forest classifier is quick and efficient, and tends to work well with unbalanced or missing data.[9]

10. **Elastic Net**

    The Elastic Net procedure fits a Generalized Linear Model via penalized maximum likelihood. It uses maximum likelihood to select a subset of variables for prediction via a shrinkage method based on a penalized version of least squares. The algorithm uses cyclical coordinate descent that successively optimizes the objective function over each parameter with the others fixed and cycles until convergence.

11. **Least Absolute Shrinkage and Selection Operator (LASSO)**

    This is a special case of the Elastic Net where the elastic net penality α is 0. In this case, when there are multiple correlated predictors, the LASSO procedure will pick one of the correlated coefficents and discard the others. At the opposite extreme, when the elastic net penalty α is 1, the coefficients of correlated predictors are shrunk towards each other and are all included in the predictive model. This is called Ridge Regression.

12. **Bayesian Network**

    This technique is capable of learning the probability density functions of individual pattern classes from a collection of learning samples, designed for pattern classification based on the Bayesian decision rule.[1] A Bayesian Network is a graphical model for probability relationships among a set of features. The network structure is a directed acyclic graph and the nodes of this graph are in one-to-one correspondence with the features. Arcs in the structure represent casual influences among the features. The task of learning a Bayesian Network involves the learning of the directed acyclic graph structure and then the determination of its parameters.[6]

13. **Naïve Bayes**

    Naïve Bayes methods are based on applying Bayes' theorem with the "naïve" assumption of independence between every pair of features, which simplifies the computations. The decoupling of the class conditional feature distributions means that

each distribution can be independently estimated as a one-dimensional distribution, helping to alleviate the problems stemming from the curse of dimensionality.

14. **Iterative Bayesian Model Averaging (BMA)**
    Bayesian Model Averaging combines the effectiveness of multiple models by taking the weighted average of their posterior distributions instead of choosing a single model. BMA can also identify a small number of predictive features, and the posterior probabilities of the selected features and models are available to create an easily interpretable summary of the output. The strength of BMA lies in its ability to account for model uncertainty, an aspect of analysis that is largely ignored by traditional stepwise selection procedures. Traditional methods tend to overestimate the goodness-of-fit between model and data. BMA attempts to solve this problem by selecting a subset of all possible models and making statistical inferences using a weighted average of the posterior distributions for these selected models.[3]
15. **Rule Based Classifier**
    A rule based classifier is a technique for classifying records using a collection of "If . . . Then . . ." rules. They produce descriptive models which are easier to interpret, but give comparable performance to Decision Tree classifiers.[23]
16. **Decision Trees (DT)**
    A Decision Tree is a tree-like structure where each internal node defines the feature or attribute to be classified. The branches define the values of particular attributes and leaf nodes provide the classes that will be output. The feature that best divides the data is the root node of the tree. Decision Trees can be translated into rules where there is a separate rule for each path from root to leaf.[6] Decision trees can handle both numeric and categorical data and generate robust classifiers.
17. **Classification and Regression Tree (CART)**
    With CART, the data is progressively split into subgroups using specific criteria for each split. The splitting process maximizes predictive accuracy. By following a patient's progress down the branches of the Decision Tree, the clinician will arrive at the predicted outcome for the patient.[16]
18. **Chi-squared Automatic Interaction Detection (CHAID)**
    This is a type of Decision Tree technique based upon adjusted significance testing. It can be used for prediction as well as classification, and for detection of interactions among the variables.[1]
19. **Semi-supervised Classification with Low Density Separation (LDS)**
    The LDS algorithm is based on the cluster assumption. It implements two effective procedures to place the decision boundary in low density regions between clusters. First, graph-based distances are derived that emphasize low density regions. Second, a gradient descent approach is used to optimize the objective function in order to find a decision boundary that avoids the high density regions.[27]

20. Logistic Regression

Logistic Regression is a form of parametric regression. It makes no assumptions about the distribution of the independent variables. It is used to predict the probability of occurrence of an event by fitting the data to a logit function which is a linear combination of regression coefficients, input variables, and an intercept constant. The variables may be continuous, discrete, dichotomous, or a combination of these.[6]

21. Generalized Linear Model (GLM)

A GLM consists of three components: a random component specifying the conditional distribution of the response variable given the values of the explanatory variables in the model, a linear function of the explanatory variables, and a smooth and invertible linearizing link function which transforms the expectation of the response variable to the linear predictor.

22. Generalized Additive Model (GAM)

A GAM is a Generalized Linear Model in which the linear predictor depends linearly on unknown smooth functions of some predictor variables, and the interest is focused on inference about these smooth functions. GAMs were developed to blend properties of Generalized Linear Models with Additive Models.

23. Partial Least Squares (PLS) Regression

PLS is a statistical method that is similar to Principal Components Regression. But instead of finding hyperplanes of maximum variance between the response and independent variables as does Principal Components, it finds a Linear Regression model by projecting the predicted variables and the observable variables onto a new space.

24. Kernelized Partial Least Squares (K-PLS)

K-PLS is a method for obtaining a Nonlinear Regression model in the space of the original input variables. This is basically a nonlinear version of Partial Least Squares.

25. K Nearest Neighbor (KNN)

K Nearest Neighbor is based on the concept that similar data will generally exist in close proximity to each other. If the cases are tagged with a classification label, then the label of an unclassified case can be determined by observing the class of its K Nearest Neighbors.[6]

26. Nomogram

A Nomogram is a graphical representation of an algorithm that incorporates multiple variables to predict a specific outcome. It typically utilizes Logistic Regression as the statistical method for computing a prediction. Nomograms are widely used in predicting outcomes for cancer patients, including survival, risk of occurrence, and the extent of the disease.[16]

27. Semi-supervised Methods (Self-learning and Co-training)

Self-learning is a technique in which the labeled set of observations L is used to build the initial classifier, and the unlabeled set U is utilized to enhance its accuracy by adding the unlabeled samples with the highest classification confidence to the training data, thus

allowing the classifier to learn based on its own decision. This classifier is then used to perform several iterations over the unlabeled data.[14]

Co-training is a technique that requires two views of the data. A separate classifier is constructed for each view. Each classifier is then used to train the other by classifying unlabeled samples and training the other view with the samples that have the highest classification confidence.[14] In many real applications, a large supply of unlabeled data is available and can be collected with little effort. However, it can be challenging to obtain labeled data, since it is often costly, difficult, or time-consuming to generate the labels for the data. Semi-supervised Learning exploits the knowledge of the input structure from unlabeled data and at the same time utilizes the information provided by the labeled data.[28]

28. Coupling Approach of two Sub-modules (a predictor and a descriptor)

A Coupling Approach involves two sub-modules (a predictor and a descriptor). The predictor algorithm generates the predicted output for cancer survivability, and any relevant machine learning technique that produces a predictive model can be used. The model generates pseudo labels that are assigned to unlabeled data before treating them as if they were labeled. As the labeled data increases, the predictive performance increases. The descriptor algorithm post-processes the results from the predictor model and provides variable importance and patient segmentation.[28]

29. Kaplan-Meier Survival Analysis Method

Kaplan-Meier curves and estimates of survival data are a familiar way of dealing with differing survival times (times-to-event), especially when not all the subjects continue in the study. They can be used to estimate disease free survival where the event of interest is a relapse of the cancer rather than death.

30. Cox Proportional Hazard Model

Whereas the Kaplan-Meier method with log-rank test is useful for comparing survival curves in two or more groups, Cox Regression (Proportional Hazards Model) allows analyzing the effect of several risk factors on survival. A Cox model is a statistical technique for exploring the relationship between the survival of a patient and several explanatory variables. It provides an estimate of the treatment effect on survival after adjustment for other explanatory variables. It also allows the estimation (the hazard or risk) of death for an individual, given their prognostic variables.

References

1. Afshar HL, Ahmadi M, Roudbari M, Sadoughi F. Prediction of Breast Cancer Survival Through Knowledge Discovery in Databases. Global Journal of Health Science; Vol 7, No 4; 2015. ISSN 1916-9736. E-ISSN 1916-9744.

2. Ahmad LG, Eshlaghy AT, Poorebrahimi A, Ebrahimi M, Razavi AR. Using Three Machine Learning Techniques for Predicting Breast Cancer Recurrence. J Health Med Inform 2013, 4:2. http://dx.doi.org/10.4172/2157-7420.1000124.

3. Annest A, Bumgarner RE, Raftery A, Yeung KY. Iterative Bayesian Model Averaging: A Method for the Application of Survival Analysis to High-dimensional Microarray Data. BMC Bioinformatics 2009, 10:72. doi:10.1186/1471-2015-10-72.

4. Bastani M, Vos L, Asgarian N, Deschenes J, Graham K, Mackey J, Greiner R. A Machine Learned Classifier That Uses Gene Expression Data to Accurately Predict Estrogen Receptor Status. PLoS ONE 8 (12): e82144. doi: 10.1371/journal.pone.0082144. https://journals.plos.org/plosone/article?id-10.1371/journal.pone.0082144.

5. Bilurkar N, Prakash A. GRNN and PNN Models for Cancer Prognosis and Prediction. International Journal of Innovative Technology and Research. ISSN 2320-5547.

6. Chaplot N, Dhyani P, Rishi OP. A Review on Machine Learning Concepts for Prediction Based Applications. International Journal of Computational Science, Engineering & Technology (ISSN 2320-4648), Vol I, Issue II, March 2013.

7. Chen Y, Ke W, Chiu H. Risk Classification of Cancer Survival using ANN with Gene Expression Data from Multiple Laboratories. Computers in Biology and Medicine 48 (2014) 1-7.

8. Cruz JA, Wishart DS. Applications of Machine Learning in Cancer Prediction and Prognosis. Cancer Inform. 2007 Feb 11; 2:59-77. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2675494/.

9. Dellsperger DJ. Outcome Prediction in Head and Neck Cancer Patients using Machine Learning Methods. Iowa Research Online. http://ir.uiowa.edu/etd/4606.

10. Ellis HP, Greenslade M, Powell B, Spiteri I, Sottoriva A, Kurian KM. Current Challenges in Glioblastima: Intratumour Heterogeneity, Residual Disease, and Models to Predict Disease Recurrence. Front Oncol. 2015; 5: 251. doi: 10.3389/fonc.2015.00251.

11. Gupta S, Tran T, Luo W, Phung D, Kennedy RL, Broad A, Campbell D, Kipp D, Singh M, Khasraw M, Matheson L, Ashley DM, Venkatesh S. Machine-learning Prediction of Cancer Survival: a Retrospective Study using Electronic Administrative Records and a Cancer Registry. BMJ Open 2014; **4**: e004007 doi: 10.1136/bmjopen-2013-004007.

12. Hasan R, Srivastava G, Alyass A, Sharma R, Saraya A, Chattopadhyay TK, DattaGupta S, Walfish PG, Chauhan SS, Ralhan R. Prediction of Recurrence Free Survival for Esophageal Cancer Patients Using a Protein Signature Based Risk Model. Oncotarget, Advance Publications 2016. www.impactjournals.com/oncotarget/

13. Huang TM, Kecman V, Kopriva I. Kernel Based Algorithms for Mining Huge Data Sets: Supervised, Semi-supervised, and Unsupervised Learning. Series Studies in Computational Intelligence, Vol 17. Springer Verlag, Berlin, Heidelberg, 2006.

14. Ibrahim R, Yousri NA, Ismail MA, El-Makky NM. miRNA and Gene Expression based Cancer Classification using Self-Learning and Co-Training Approaches. Bioinformatics and Biomedicine (BIBM), 2013 IEEE International Conference. doi: 10.1109/BIBM.2013.6732544.

15. Kim Q, Kim KS, Noh DY, Kim AW, Jung YS, Park MY, Park RW. Development of Novel Breast Cancer Recurrence Prediction Model Using Support Vector Machine. J Breast Cancer 2012 June; 15 (2): 230-238. http://dx.doi.org/10.4048/jbc.2012.15.2.230.

16. Kazem MA. Predictive Models in Cancer Management: A Guide for Clinicians. http://dx.doi.org/10.1016/j.surge.2016.06.002.

17. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine Learning Applications in Cancer Prognosis and Prediction. Computational and Structural Biotechnology Journal 13 (2015) 8-17. http://dx.doi.org/10.1016/j.csbj.2014.11.005.

18. Kulkarni S, Bhagwat M. Predicting Breast Cancer Recurrence using Data Mining Techniques. International Journal of Computer Applications (0975-8887), Vol 122, No 23, July 2015.

19. Land WH, Qiao X, Margolis DE, Ford WS, Paquette CT, Perez-Rogers JF, Borgia JA, Yang JY, Deng Y. Kernelized Partial Least Squares for Feature Reduction and Classification of Gene Microarray Data. BMC Syst Biol 2011: 5 (Suppl 3): 513. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3287568/.

20. Liu L, Messer K, Baron JA, Lieberman DA, Jacobs ET, Cross AJ, Murphy G, Martinez ME, Gupta S. A Prognostic Model for Advanced Colorectal Neoplasia Recurrence. Cancer Causes Control. doi 10.1007/s10552-016-0795-5.

21. McQuisten KA, Peek AS. Comparing Artificial Neural Networks, Generalized Linear Models and Support Vector Machines in Building Predictive Models for Small Interfering RNAs. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2760777/.

22. Mi H, Petitjean C, Dubray B, Vera P, Ruan S. Robust Feature Selection to Predict Tumor Treatment Outcome. Artificial Intelligence in Medicine 64 (2015) 195-204. http://dx.doi.org/10.1016/j.artmed.2015.07.002.

23. Murti S, Mahantappa. Using Rule Based Classifiers for the Predictive Analysis of Breast Cancer Recurrence. Journal of Information Engineering and Applications. ISSN 2224-5782 (print) ISSN 2225-0506 (online) Vol 2, No 2, 2012.

24. Nithya, B. An Analysis on Applications of Machine Learning Tools, Techniques and Practices in Health Care System. International Journal of Advanced Research in Computer Science and Software Engineering. Vol 6, Issue 6, June 2016. www.ijarcsse.com.

25. Podolsky MD, Barchuk AA, Kuznetcov VI, Gusarova NF, Gaidukov VS, Tarakanov SA. Evaluation of Machine Learning Algorithm Utilization for Lung Cancer Classification Based on Gene Expression Levels. Asian Pac J Cancer Prev, 17 (2), 835-838.

26. Ruan J, Jahid MJ, Gu F, Lei C, Huang YW, Hsu YT, Mutch D, Chen CL, Kirma NB, Huang THM. A novel Algorithm for Network-based Prediction of Cancer Recurrence. http://dx.doi.org/10.1016/j.ygeno.2016.07.005.

27. Shi M, Zhang B. Semi-supervised Learning Improves Gene Expression-based Prediction of Cancer Recurrence. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3198572/.

28. Shin H, Nam Y. A Coupling Approach of a Predictor and a Descriptor for Breast Cancer Prognosis. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4101306/.

29. Stoean R, Stoean C. Modeling Medical Decision Making by Support Vector Machines, Explaining by Rules of Evolutionary Algorithms with Feature Selection. Expert Systems with Applications 40 (2013) 2677-2686.

30. Thottakkara P, Ozrazgat-Baslanti T, Hupf BB, Rashidi P, Pardalos P, Momcilovic P, Bihorac A. Application of Machine Learning Techniques to High Dimensional Clinical Data to Forecast Postoperative Complications. PLoS ONE 11 (5): e0155705. doi:10.1371/journal.pone.0155705.

31. Umesh DR. Predicting Breast Cancer Recurrence Using Machine Learning Techniques. International Journal of Latest Trends in Engineering and Technology. Vol 5, Issue 1, January 2015.

32. Van Stiphout RGPM, Postma EO, Valentini V, Lambin P. The Contribution of Machine Learning to Predicting Cancer Outcome. Proceedings of the BNAIC 2010, University of Luxembourg, pp 1-7. http://bnaic2010.uni.lu/proceedings.html.

33. Vidyasagar M. Machine Learning Methods in the Computational Biology of Cancer. Proc. R. Soc. A 470: 20140081. http://dx.doi.org/10.1098/rspa.2014.0081.

34. Wei Q. Application of Machine Learning Techniques to Acute Myeloid Leukemia. University of Washington Thesis Paper.

35. Zhang YD, Wang J, Wu CJ, Bao ML, Li H, Wang XN, Tao J, Shi HB. An Imaging-based Approach Predicts Clinical Outcomes in Prostate Cancer Through a Novel Support Vector Machine Classification. Oncotarget, Advance Publications 2016. www.impactjournals.com/oncotarget/.