International Journal of Pure and Applied Mathematics
Volume 117 No. 20 2017, 853-860
ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
Special Issue

ijpam.eu

Machine Learning Applications in Ovarian Cancer Prediction: A Review

[1]SuthamerthiElavarasu,[2]Viji Vinod,[3]ElavarasanElangovan
[1]Research scholar-Department of Computer Applications,Dr.M.G.R.Educational and Research Institute University Madoravoyal,Chennai,TamilNadu-600095
[2]Head of the department Computer Applications,Dr.M.G.R.Educational and Research Institute University Madoravoyal,Chennai,TamilNadu-600095
[3]DataAnalyst,L&Tinfotech Bangalour
Email: merthi.88@gmail.com[1],Vijivino@gmail.com[2],Elavarasan.Elangovan@lntinfotecch.com[3]

Abstract:

Cancer is the second leading cause of death, alongside heart disease, in both developed and developing countries. Cancer has been characterized as a heterogeneous disease consisting of several different subtypes. Early diagnosis and prognosis of a sort of disease have turned into a need in malignancy explore, as this can encourage the resulting clinical administration of patients. A noteworthy wellspring of familiarity with the battle against tumor as an infection is its shocking assortment, as there are not two appearances of growth, regardless of the possibility that they happen on a similar site. This makes the malignancy perfect for customized prescription. The significance of arranging malignancy patients into high- or low-risk groups has driven many research groups from the bioinformatics and biomedical fields to concentrate the utilization of machine learning techniques. At present, there are top notch open databases including both molecular estimations of tumor, and additionally clinical information on patients. By applying machine learning methods to these databases, it is possible even for non-experimenters to generate plausible assumptions that are supported by the data, which can then be validated on one or more independent data sets. As a result, these techniques have been used to model the progression and treatment of cancerous diseases. Also, the ability of machine learning tools to detect key features from complex data sets reveals their importance. Given the growing trend in the application of machine learning methods in cancer research, in this paper, the most recent publications that use these techniques as an objective to model cancer risk or outcomes are reviewed, and their applications to ovarian cancer are discussed.

Keywords: Machine Learning, Artificial Neural Networks, Bayesian Networks, Support Vector Machines, Ovarian Cancer

INTRODUCTION:

In developing and developed countries, along with heart disease, cancer is the second leading cause of death. Rebecca L. et al. estimated that there would be 600,920 deaths and 1,688,780 new cases of cancer in the United States by 2017 [Siegel,R.L. 2015]. Compared to the prevention of heart disease, the underlying causes and possible mechanisms for cancer are not well understood. As a result, cancer will become the most common cause of death over time, predicted by several experts. Despite numerous times of exertion, upgrades in cancer treatment have not completely satisfied desires. As a regular case from many that could be mentioned, a form of cancer with a very poor prognosis, a multi-country initiative on colorectal cancer, a 0.5% improvement per year resulted in a change in relative survival of five years rate of 59% to 65%, over twelve years [Iversen, L. H., 2016].

Uncontrolled growth of malignant cells causes cancer that contains mutations in a person's normal DNA. Malignant cells undergo normal cells for some time. A remarkable source of difficulty in influencing malignancy as a disease is its confusing assortment. Even when they occur on the same site, there are not two manifestations of cancer alike. There may be multiple mutations of normal DNA within a single cancer tumor. Current research and treatment techniques address what may be referred to as "dominant" mutations within a tumor, as these are the most recognized. Independently of the possibility that the treatment is to kill all dominant mutant cells, other transformations will eventually evolve. This may be the reason why, in most cases, although cancer therapy seems to work, the tumor decreases or may even become undetectable after a time

When the tumor is revealed and shows a new growth explosion. In addition, the recurrent tumor is regularly impermeable to the treatment already applied. Since cancer manifestations vary considerably from one person to another, cancer is an ideal candidate for a "personalized prescription," in which the treatment is tailored for each patient.

Specifically, ovarian cancer is an exceedingly deadly disease contained epithelial ovarian, essential peritoneal and fallopian tube carcinoma [Hennessy, B. T. 2009] [Schorge, J. O. 2010]. After initial treatment, most patients with ovarian cancer have undetectable diseases and are thought to be in clinical abatement. Most of these women will suffer a recurrence and will eventually succumb to the disease [Siegel, R. 2013,][Prat, J.2012][Chan, J. K. 2006]. As a result, just unobtrusive increases have been acknowledged in 5-year survival, and cure rates have not significantly made strides. Throughout the planned and executed randomized clinical trials using overall survival as the essential endpoint give the most indisputable information after that to build up new models of care, yet are tested by cost and long announcing circumstances. Further, prolonged post-progression survival, significant patient enrolment needs, and routine use of subsequent and cross-operative treatments challenge the overall survival concept as an optimal criterion for all cancer trials, ovary. In addition, the rapid disclosure of potentially imperative and variant

molecular variations in ovarian carcinogenesis has resulted in greater than the research capacity of these targets in extensive and delayed clinical trials [Coleman, R. L. 2013] [Bell-McGuinn, K. 2011].

The goals of this perspective paper are to survey some current advances in the selection of sparse characteristics for regression and classification and to discuss how we could be used in ovarian cancer. Presenting a broad picture of some recent advances in machine learning in the ovarian cancer community and applying some of these techniques to some problems is the main motive of this survey. This paper will encourage the section of the intrigued analysts of the machine learning group in ovarian cancer.

Machine learning techniques:

Machine learning is a part of artificial intelligence, relates the issue of gaining from information tests to the general idea of induction [Bishop,C. M. 2006][Mitchell,T. M. 2006][Azuaje, F. 2006]Each learning procedure comprises of two stages: first, estimation of obscure conditions in a framework from a given dataset and second, utilization of assessed conditions to anticipate new yields of the framework. Machine learning has also been shown to be an intriguing territory in biomedical research with many applications, where satisfactory speculation is obtained through an n-dimensional space for a given arrangement of organic examples, using systems and algorithms [Niknejad, A. 2013]. There are two primary basic sorts of machine learning strategies known as supervised learning and unsupervised learning. In supervised learning, a named set of preparing information is utilized to gauge or guide the information to the coveted yield. On the other hand, under the unsupervised learning strategies, no named illustrations are given, and there is no thought of the yield amid the learning procedure. Subsequently, it is up to the learning plan/model to find designs or finds the gatherings of the info information. This procedure can be considered as a classification problem in supervised learning. The classification task refers to a learning process that classifies the data into a set of finite classes. Clustering and regression are the two other common machine learning tasks. A learning function maps the data into a real-valued variable in the regression problems. Clustering is a typical unsupervised assignment in which one tries to find the classifications or groups keeping in mind the end goal to portray the information things. In light of this procedure, each new example can be appointed to one of the identified clusters concerning the comparative qualities that they share.

Another kind of machine learning strategies that have generally been connected is semi-supervised learning, which is a blend of supervised and unsupervised learning. To build an accurate learning model, it combines marked and unmarked data. This sort of learning is utilized when there are more unlabeled datasets than marked. Data samples are the basic components while applying a machine learning strategy. Each sample is depicted with a few elements, and each component comprises of various sorts of values. Besides, knowing ahead of time the specific kind of information being utilized permits the correct determination of tools and techniques that can be utilized for their investigation. A few information related issues allude to the nature of the information and the preprocessing ventures to make them more reasonable for machine learning. Data quality issues include the presence of aberrant, missing or duplicated sound data and biased, unrepresentative data. While enhancing the data quality, commonly the nature of the subsequent examination is likewise made strides. Likewise, to make the raw data more appropriate for further examination, preprocessing steps ought to be connected that emphasis on the modification of the data. There are a number of different techniques and strategies relevant to data pre-processing that focus on modifying data for better fixation in a specific machine learning method.Among these approaches the absolute most critical methodologies incorporate reducing dimensionality, selecting features, and extracting feature. There are many advantages in reducing dimensionality when data sets have a large number of features. When the dimension is lower, machine learning works better [Pang-Ning.T,2006]. In addition, due to the involvement of fewer features, reducing dimensionality can eliminate irrelevant features, reduce noise and produce more robust learning models. The dimensionality diminishment by choosing new features which are a subset of the old ones is known as feature selection.

Three principle approaches exist for selection of features, namely integrated, filters and envelopes approach. A new set of functions can be made from the underlying set that catches all the significant data in a dataset in the case of feature extraction. The creation of new sets of functionalities makes it possible to collect the described advantages of the reduction of the dimensionality. However, application of feature selection techniques may lead to specific flaws in the creation of lists of predictive features. Several studies in the literature deal with the lack of agreement between the lists of predictive genes discovered by different groups, the need for thousands of samples to achieve the desired results, the lack of biological interpretation of predictive signatures, and the dangers of information leakage recorded in published studies [Drier, Y. 2011][Dupuy, A. 2007][Ein-Dor, L. 2005][Ein-Dor,L.2006]

At the point when the data are preprocessed, and we have defined the sort of learning assignment, a rundown of machine learning techniques including Artificial Neural Networks (ANNs), Bayesian Networks (BNs), Support Vector Machines (SVMs) and Decision Trees (DTs) is available. Based on the intensity of this survey, we will refer only these machine learning techniques that have been widely used in the literature for a case study of cancer prediction and prognosis. We distinguish the patterns in regards to the sorts of machine learning strategies that are utilized, the sorts of information that are incorporated and in addition the assessment techniques utilized for surveying the general execution of the techniques utilized for growth expectation or illness results. A variety of pattern recognition or classification problems deals by ANNs. With combination of the input variables, they are formed to generate an output. Multiple hidden layers that mathematically represent neural connections are usually used for this process. Although in several classification tasks, ANNs serve as gold standard methods, they suffer from some disadvantages [Ayer, T.2010]. Their non-specific layered structure eventually becomes tedious, while it can cause an exceptionally poor performance. In addition, this specific system is represented as a "black

box" technology. Trying to find out how he plays the ranking procedure or why an ANN did not work is virtually hard to identify. Fig. 1 depicts the structure of an ANN with its interconnected group of nodes.

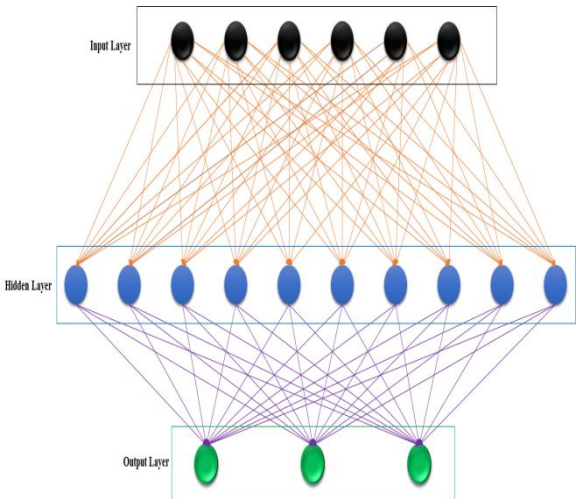
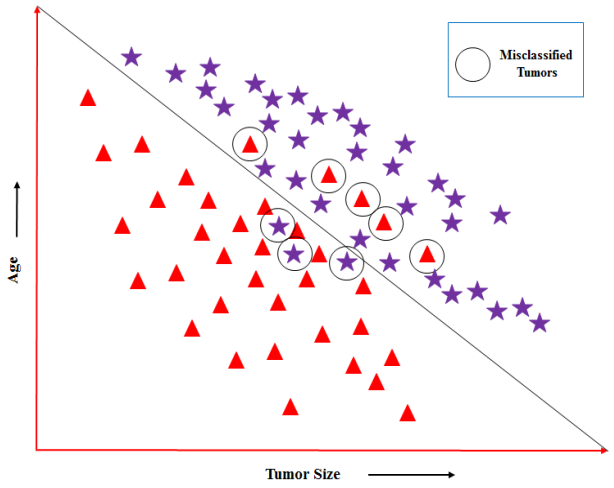


Fig.1.An illustration of the structure of ANNs

Fig.2. An illustration of a linear Support Vector Machines (SVMs) classification



SVM is a more recent approach of machine learning techniques applied in the field of cancer prognosis/prediction. At first, SVMs delineate info vector into a component space of higher dimensionality and recognize the hyperplane that isolates the information focuses on two classes. The negligible separation between the choice hyperplane and the occurrences that are nearest to limit is expanded. The subsequent classifier accomplishes impressive generalizability and can in this way be utilized for the dependable classification of new specimens. It is significant that probabilistic yields can likewise be gotten for SVMs [Platt, J. C., 1999]. Fig. 2 outlines how an SVM may function with a specific end goal to group tumors among amiable and malignant in light of their size and patients' age [Adams, S2012]. The identified hyperplane can be thought as a choice limit between the two groups. Clearly, the presence of a choice limit takes into account the discovery of any misclassification created by the strategy.

Machine learning applications in cancer prediction:

In cancer susceptibility, recurrence, and survival prediction; extensive research has been conducted on the use of machine learning methods. Due to a large number of articles published in this domain, a thorough review has been done to maintain the most relevant articles. Most of the research work used different types of input data: epidemiological data, demographic, genomics, imaging, histological, clinical, or a combination of these. In last decade, we observed that the survey based on machine learning applications in cancer prediction is rapidly increased. In this review, a significant number of relevant papers have been extracted and are presented. We have analyzed the works that use recognizable machine learning techniques and integrated data from heterogeneous sources to predict desirable outcomes. We have concentrated mainly on studies that have been published over the last five years in order to present the latest state of the art in the field and their progress compared to the old publications. Tables 1 represent some of the publications presented in this review that used machine learning techniques for cancer susceptibility, recurrence, and survival predictions. The type of cancer, the machine learning technique, the type of data and the overall accuracy obtained by each proposed method are introduced. In particular, we studied prediction of cancer susceptibility, prediction of cancer recurrence and prediction of cancer survival according to machine learning techniques. Here, we presented the most accurate predictive model in the case where articles where multiple machine learning methods are applied.

APPLICATION OF MACHINE LEARNING TO OVARIAN CANCER :

OVARIAN CANCER I:

As we discussed before, ovarian cancer is the fifth most fatal form of cancer in women after pancreatic, colon, breast, and lung cancer. Standard cutting edge treatment for ovarian tumor

Table 1: Summary of machine learning techniques used for cancer prediction

Types Prediction	Approach	Cancer Type	Type of Data	Accuracy	Reference
	ANN	Breast Cancer	Demographic	AUC= 0.965	[Ayer, T.2010]
	SNM	Multiple Myeloma	SNPs	Yields = 71%	[Waddell, M.2005]
Susceptibility	SVM	Breast Cancer	SNPs	PredictivePower = 69%	[Listgarten, J.2004]
	BN	Colon Carcinomatosis	Pathologic	AUC= 0.71	[Stojadinovic, A.2011]
	BN	Oral Cancer	Imaging tissue genomic	Accuracy 100%	[Exarchos,K. P2012]
	SVM	Breast Cancer	Pathologic	Accuracy = 89%	[Kim, W.2012]
Recurrence	SVM	Cervical Cancer	Pathologic	Accuracy = 68%	[Tseng,C. J.2014]
	SVM	Breast Cancer	Clinical, Population	Accuracy = 95%	[Ahmad,L. G.2013]
	ANN	Lung Cancer	Gene Expression	Accuracy= 83.5%	[Chen,Y. C.2014]
	SVM	Oral Cancer	Genomic	Accuracy = 75%	[Chang,S. W2013]
	SVM	Breast Cancer	Genomic	Accuracy = 97%	[Xu,X.,2012]
Survival	BN	Breast Cancer	Clinical	AUC = 0.851	[Gevaert, O.2006]
	SVM	Oral Cancer	Molecular	Accuracy = 98%	[Rosado, P.2013]

Comprises of some type of taxane (paclitaxel) combined with some type of platinum (cisplatin or carboplatin), alluded to from now on as platinum-based chemotherapy. To first-line treatment, the response of the patient is not uniform. Since it is unrealistic to monitor a patient constantly to evaluate reaction to treatment, one can utilize progression-free survival or overall survival as to some degree blemished intermediaries for patient reaction. Based on the TCGA Agilent data set, four different classifiers are introduced by Misganaw et al. [MISGANAW, B.2015]. A recent article states that 139 studies have reported an association between biomarker expression and overall survival (OS) with univariate analysis, whereas with multivariate analysis an associationbetween the number of studies that evaluated an association between biomarker expression and progression-free survival with univariate analysis and multivariate analysis were respectively 66 and 20 studies [XU, L.,2013]. A phenomenalsurvey of a few reviews can be found in Sabatier's research [SABATIER, R.2009]. In any case, none of these papers contains a molecular signature, that is, a methodology for changing over measured estimations of the biomarkers more often than not quality expression levels into a numerical score. The improvement of such a signature and not only a biomarker board is one of the inspirations for the review by.[MISGANAW, B.2015]

To develop an appropriate set of molecular signatures, the TCGA ovarian cancer database of gene expression values, measured on an Agilent stage is utilized as the training dataset [CANCER GENOME ATLAS RESEARCH NETWORK. 2011]. Two different sets of test data are utilized, in particular: the TCGA ovarian malignancy database of gene expression values measured on an

Affymetrix and the Tothill dataset, which is additionally measured on an Affymetrix stage and is from Australia [TOTHILL, R. W.2008]. Signature testing using gene expression values of the same tumors, but measured on a different platform, is used to demonstrate the validity of the signature on the platforms while testing the signature on a completely unique dataset from another mainland serves to set up the legitimacy of the signature crosswise over differing populaces.

As mentioned above, the goal is to find molecular signatures that can discriminate against extreme players, i.e. those who respond very well to platinum-based chemotherapy and those who respond very poorly. After some trial and error, details of which are reported in [MISGANAW, B.2015], the patient population in the training data was divided into three equal percentiles, with 33% identified as "non-responders (NR)", with the highest 33% being referred to as "super-responders (SR) ", and the 33% means being referred to as" average responder "(MR)." The "patient response" can be measured in two different ways: progression-free survival and global survival. Although these two are largely correlated, the correlation is by no means perfect. For example, overall survival is determined not only by the effectiveness of therapy, but by other factors such as age, general health, and so on. Progression-free survival is also subjective, as the date on which a tumor is recorded as having progressed is the day on which it was observed to have progressed when in fact, the progression would have occurred on an unknown date between that observed date And the date of the previous audit. Thus, the disparity between the recorded date of progression and the actual date of recidivism could be several months. It is not a priori clear which clinical parameter would lead to better predictions. Subsequently, in[MISGANAW, B.2015], indicators are created in light of every parameter, and their execution is analyzed.

OVARIAN CANCER:

The purpose of this section is to explore whether taking advantage of the ordered nature of the label space offers an advantage over label processing as being nominal, i.e. not having an order. Along these lines, patients with ovarian cancer in the AGGA aggregate data set of TCGA data are divided into four consisting of 25 percentiles in terms of progression-free survival. The choice of four groups as opposed to three is done to give more "structure" to the problem. Then, the different multi-class ranking methods are applied, and the results are compared. Gene expression profiles, measured on Agilent platforms as well as Affymetrix, and clinical data from 566 ovarian cancer patients were downloaded from the Atlas Genome website [The Cancer Genome Atlas,2017]. Additional datasets were obtained from the gene expression omnibus database: the Tothill dataset [Tothill, R. W.,2008] consisted of 285 samples measured on a flat plate -form Affymetrix and the Yoshihara dataset [Yoshihara, K.,2010] consists of 110 samples measured on an Agilent platform. Progression-free survival times in the TCGA data range from 0 to 5480 days. The 25, 50 and 75 percentiles of the progression-free survival days, corresponding to 233, 404 and 723 days, were used to divide the samples into four risk groups of 142, 141, 142 and 140. Next, 70% of samples from each of four groups were selected at random for use as training data. There were four separate sets of test data: the remaining 30% of the samples from the Agilent TCGA data, the Affymetrix TCGA data set, the Tothill data, and the data of Yoshihara. In the training data and each of the four test data sets, the gene expression values of each gene were normalized to zero and to the unit variance. In [Vidyasagar, M., 2014] selected set of genes, a linear classifier was constructed using a l_1-norm SVM algorithm.

Future direction:

The training and test data must follow the same probability distribution is one of the essential premises of machine learning. The purpose behind this suspicion is plainly obvious: if an algorithm is formed on a data set and tested on an entirely different dataset, it generally cannot have good performance on the test data. In the early years, levels of gene expression were measured using microarrays. This approach has resulted in relatively inaccurate and fairly no repeatable estimations. The present pattern is to supplant microarrays by something many refer to as RNA-seq. Not at all like microarray estimations, are RNA-seq estimations to a great degree precise. Along these lines it is important to fit both the legacy microarray information with more up to date RNA-seq information. So, with new RNA-seq data, it is essential to harmonize both the existing microarray data. There is a generally new region in machine learning known as "transfer learning" in which the preparation information and test information are permitted to have different probability distributions. Be that as it may, the two sets of data must come from a common domain. The measurements are read counts and are thus evaluated in integer numbers, while the microarray measurements are actually evaluated I the case of RNA-seq. As a result, the transfer learning methodology as it currently exists is not specifically relevant. Therefore, it is possible and extended the transfer learning methodology to circumstances where training and test data come from distinct domains.

Conclusion:

In this survey, we discussed the concepts of machine learning while describing their application in cancer prediction. Most studies proposed in recent years focus on the development of predictive models using supervised machine learning techniques and classification algorithms to predict valid outcomes of the disease. On the basis of the analysis of their results, it is clear that the integration of multidimensional heterogeneous data, combined with the application of different techniques of selection and classification of functionalities, can provide promising tools for inference in cancer.

References:

[1] Siegel, R. L., Miller, K. D., & Jemal, A. (2015). Cancer statistics, 2015. *CA: a cancer journal for clinicians*, *65*(1), 5-29

[2] Iversen, L. H., Green, A., Ingeholm, P., Østerlind, K., & Gögenur, I. (2016). Improved survival of colorectal cancer in Denmark during 2001–2012–The efforts of several national initiatives. *Acta Oncologica*, *55*(sup2), 10-23

[3] Hennessy, B. T., Coleman, R. L., & Markman, M. (2009). Ovarian cancer. *The Lancet*, *374*(9698), 1371-1382

[4] Schorge, J. O., Modesitt, S. C., Coleman, R. L., Cohn, D. E., Kauff, N. D., Duska, L. R., & Herzog, T. J. (2010). SGO White Paper on ovarian cancer: etiology, screening and surveillance. *Gynecologic oncology*, *119*(1), 7-17

[5] Siegel, R., Naishadham, D., Jemal, A., Cancer statistics, (2013). CA Cancer J. Clin., 63 (1), 11–30

[6] Prat, J. (2012). New insights into ovarian cancer pathology. *Annals of oncology*, *23*(suppl 10), x111-x117

[7] Chan, J. K., Cheung, M. K., Husain, A., Teng, N. N., West, D., Whittemore, A. S., ... & Osann, K. (2006). Patterns and progress in ovarian cancer over 14 years. *Obstetrics & Gynecology*, *108*(3, Part 1), 521-528

[8] Coleman, R. L., Monk, B. J., Sood, A. K., & Herzog, T. J. (2013). Latest research and treatment of advanced-stage epithelial ovarian cancer. *Nature reviews Clinical oncology*, *10*(4), 211-224

[9] Bell-McGuinn, K., Konner, J., Tew, W., et al., (2011). New drugs for ovarian cancer, Ann. Oncol. 22 (Suppl. 8), 77–82

[10] Bishop, C. M. (2006). Pattern recognition. *Machine Learning*, *128*, 1-58

[11] Mitchell, T. M. (2006). *The discipline of machine learning* (Vol. 9). Carnegie Mellon University, School of Computer Science, Machine Learning Department

[12] Azuaje, F. (2006). Witten IH, Frank E: Data Mining: Practical Machine Learning Tools and Techniques 2nd edition

[13] Niknejad, A., & Petrovic, D. (2013). Introduction to computational intelligence techniques and areas of their applications in medicine. *Med Appl Artif Intell*, *51*

[14] Pang-Ning T, Steinbach M, Kumar V. (2006). Introduction to data mining

[15] Drier, Y., & Domany, E. (2011). Do two machine-learning based prognostic signatures for breast cancer capture the same biological processes?. *PloS one*, *6*(3), e17795

[16] Dupuy, A., & Simon, R. M. (2007). Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. *Journal of the National Cancer Institute*, *99*(2), 147-157

[17] Ein-Dor, L., Kela, I., Getz, G., Givol, D., & Domany, E. (2005). Outcome signature genes in breast cancer: is there a unique set?. *Bioinformatics*, *21*(2), 171-178

[18] Ein-Dor, L., Zuk, O., & Domany, E. (2006). Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. *Proceedings of the National Academy of Sciences*, *103*(15), 5923-5928

[19] Ayer, T., Alagoz, O., Chhatwal, J., Shavlik, J. W., Kahn, C. E., & Burnside, E. S. (2010). Breast cancer risk revisited. *Cancer*, *116*(14), 3310-3321

[20] Platt, J. C., Cristianini, N., & Shawe-Taylor, J. (1999, November). Large margin DAGs for multiclass classification. In *Proceedings of the 12th International Conference on Neural Information Processing Systems* (pp. 547-553). MIT press

[21] Adams, S. (2012). Is coursera the beginning of the end for traditional higher education?. *Leadership*, *4*, 31PM

[22] Waddell, M., Page, D., & Shaughnessy Jr, J. (2005, August). Predicting cancer susceptibility from single-nucleotide polymorphism data: a case study in multiple myeloma. In *Proceedings of the 5th international workshop on Bioinformatics* (pp. 21-28). ACM

[23] Listgarten, J., Damaraju, S., Poulin, B., Cook, L., Dufour, J., Driga, A., ... & Zanke, B. (2004). Predictive models for breast cancer susceptibility from multiple single nucleotide polymorphisms. *Clinical cancer research*, *10*(8), 2725-2737

[24] Stojadinovic, A., Nissan, A., Eberhardt, J., Chua, T. C., Pelz, J. O., & Esquivel, J. (2011). Development of a Bayesian Belief Network Model for personalized prognostic risk assessment in colon carcinomatosis. *The American Surgeon*, *77*(2), 221-230

[25] Exarchos, K. P., Goletsis, Y., & Fotiadis, D. I. (2012). Multiparametric decision support system for the prediction of oral cancer reoccurrence. *IEEE Transactions on Information Technology in Biomedicine*, *16*(6), 1127-1134

[26] Kim, W., Kim, K. S., Lee, J. E., Noh, D. Y., Kim, S. W., Jung, Y. S., ... & Park, R. W. (2012). Development of novel breast cancer recurrence prediction model using support vector machine. *Journal of breast cancer*, *15*(2), 230-238

[27] Tseng, C. J., Lu, C. J., Chang, C. C., & Chen, G. D. (2014). Application of machine learning to predict the recurrence-proneness for cervical cancer. *Neural Computing and Applications*, *24*(6), 1311-1316

[28] Ahmad, L. G., Eshlaghy, A. T., Poorebrahimi, A., Ebrahimi, M., & Razavi, A. R. (2013). Using three machine learning techniques for predicting breast cancer recurrence. *J Health Med Inform*, *4*(124), 3

[29] Chen, Y. C., Ke, W. C., & Chiu, H. W. (2014). Risk classification of cancer survival using ANN with gene expression data from multiple laboratories. *Computers in biology and medicine*, *48*, 1-7

[30] Chang, S. W., Abdul-Kareem, S., Merican, A. F., & Zain, R. B. (2013). Oral cancer prognosis based on clinicopathologic and genomic markers using a hybrid of feature selection and machine learning methods. *BMC bioinformatics*, *14*(1), 170

[31] Xu, Y., Zhang, Y., Zou, L., Wang, M., & Li, A. (2012, October). A gene signature for breast cancer prognosis using support vector machine. In *Biomedical Engineering and Informatics (BMEI), 2012 5th International Conference on* (pp. 928-931). IEEE

[32] Gevaert, O., De Smet, F., Timmerman, D., Moreau, Y., & De Moor, B. (2006). Predicting the prognosis of breast cancer by integrating clinical and microarray data with Bayesian networks. *Bioinformatics*, *22*(14), e184-e190

[33] Rosado, P., Lequerica-Fernández, P., Villallaín, L., Peña, I., Sanchez-Lasheras, F., & De Vicente, J. C. (2013). Survival model in oral squamous cell carcinoma based on clinicopathological parameters, molecular markers and support vector machines. *Expert Systems with Applications*, *40*(12), 4770-4776

[34] Misganaw, B., Ahsen, E., Singh, N., Baggerly, K. A., Unruh, A., White, M. A., & Vidyasagar, M. (2015, December). Optimized prediction of extreme treatment outcomes in ovarian cancer. In *Decision and Control (CDC), 2015 IEEE 54th Annual Conference on* (pp. 1254-1258). IEEE

[35] Xu, L., Cai, J., Yang, Q., Ding, H., Wu, L., Li, T., & Wang, Z. (2013). Prognostic significance of several biomarkers in epithelial ovarian cancer: a meta-analysis of published studies. *Journal of cancer research and clinical oncology*, *139*(8), 1257-1277

[36] Sabatier, R., Finetti, P., Cervera, N., Birnbaum, D., & Bertucci, F. (2009). Gene expression profiling and prediction of clinical outcome in ovarian cancer. *Critical reviews in oncology/hematology*, *72*(2), 98-109

[37] Cancer Genome Atlas Research Network. (2011). Integrated genomic analyses of ovarian carcinoma. *Nature*, *474*(7353), 609-615

[38] Tothill, R. W., Tinker, A. V., George, J., Brown, R., Fox, S. B., Lade, S., ... & Traficante, N. (2008). Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clinical cancer research*, *14*(16), 5198-5208

[39] The Cancer Genome Atlas. (TCGA). http://cancergenome.nih.gov, (accessed on 25 April 2017)

[40] Yoshihara, K., Tajima, A., Yahata, T., Kodama, S., Fujiwara, H., Suzuki, M., ... & Kudo, Y. (2010). Gene expression profile for predicting survival in advanced-stage serous ovarian cancer across two independent datasets. *PloS one*, *5*(3), e9615

[41] Vidyasagar, M. (2014, July). Machine learning methods in the computational biology of cancer. In *Proc. R. Soc. A* (Vol. 470, No. 2167, p. 20140081). The Royal Society