

Artificial Intelligence Systems as Prognostic and Predictive Tools in Ovarian Cancer

A. Enshaei¹, C. N. Robson¹, and R. J. Edmondson²

¹Medical School, Northern Institute for Cancer Research, University of Newcastle Upon Tyne, Newcastle upon Tyne, UK;

²Chair of Gynaecological Oncology, Faculty Institute for Cancer Sciences, University of Manchester, Manchester Academic Health Science Centre, St. Mary's Hospital, Manchester, UK

ABSTRACT

Background. The ability to provide accurate prognostic and predictive information to patients is becoming increasingly important as clinicians enter an era of personalized medicine. For a disease as heterogeneous as epithelial ovarian cancer, conventional algorithms become too complex for routine clinical use. This study therefore investigated the potential for an artificial intelligence model to provide this information and compared it with conventional statistical approaches.

Methods. The authors created a database comprising 668 cases of epithelial ovarian cancer during a 10-year period and collected data routinely available in a clinical environment. They also collected survival data for all the patients, then constructed an artificial intelligence model capable of comparing a variety of algorithms and classifiers alongside conventional statistical approaches such as logistic regression.

Results. The model was used to predict overall survival and demonstrated that an artificial neural network (ANN) algorithm was capable of predicting survival with high accuracy (93 %) and an area under the curve (AUC) of 0.74 and that this outperformed logistic regression. The model also was used to predict the outcome of surgery and again showed that ANN could predict outcome (complete/

optimal cytoreduction vs. suboptimal cytoreduction) with 77 % accuracy and an AUC of 0.73.

Conclusions. These data are encouraging and demonstrate that artificial intelligence systems may have a role in providing prognostic and predictive data for patients. The performance of these systems likely will improve with increasing data set size, and this needs further investigation.

Prognostic information has always been important to patients and clinicians, but with the advent of a greater range of treatment options for patients, prediction has become increasingly important as well. This is particularly true with ovarian cancer, which is an extremely heterogeneous disease. Patients present at various stages of the disease, and the tumors not only are of varying grades but also display a range of histologic subtypes.

Currently, outside of clinical trials, patients are treated with surgical cytoreduction followed by platinum-based chemotherapy.¹ Whereas age,² stage,³ grade,⁴ histologic type,⁵ preoperative CA125 levels,⁶ and result of surgery⁷ all have been associated with overall survival, multivariate analysis consistently shows only stage and result of surgery to be independent prognostic factors.⁸

However, studies have shown new treatment options including antiangiogenic treatment,⁹ poly(ADP-ribose) polymerase (PARP) inhibitor therapy,¹⁰ and hormonal therapy¹¹ to be effective for at least some subgroups of patients with ovarian cancer. Although biomarkers are being developed to identify patients likely to respond to each of these treatments,^{12–14} it is unclear how these treatments should be scheduled and prioritized. Furthermore, although the role of surgery in ovarian cancer is well established, it is clear that not all patients benefit from surgery, and for some patients, the morbidity of the procedure outweighs any potential benefit. It is therefore

Electronic supplementary material The online version of this article (doi:10.1245/s10434-015-4475-6) contains supplementary material, which is available to authorized users.

© Society of Surgical Oncology 2015

First Received: 25 September 2014;
Published Online: 10 March 2015

R. J. Edmondson
e-mail: richard.edmondson@manchester.ac.uk

important to develop more powerful prognostic and predictive tools that allow patients and clinicians to plan treatment both optimally and realistically.

Artificial intelligence (AI) systems are in common everyday use and provide many benefits including ability to handle enormous amounts of data, to cope with missing data items, and to evolve in the presence of new data. In general, they attempt to find better solutions rather than the absolutely best or correct solution.¹⁵

Simplistically, AI systems comprise an algorithm and a classifier. Several different algorithms and classifiers have been developed, all of which have relative merits. Further information regarding these is given in the supplementary materials.

This study therefore investigated the role of AI systems in creating a prognostic and predictive tool for ovarian cancer. We generated a historical data set from patients treated at our institution and extracted data that would be readily available after initial surgery and staging, mimicking the clinical situation, as well as overall survival data. We then used a systems-based approach to investigate the optimal method for determining a range of outcomes.

METHODS

The Data Set

The Newcastle ovarian cancer tissue and data set has been described previously.¹⁶ Briefly, the data set comprises 668 cases of epithelial ovarian cancer. All patients underwent maximal-effort primary cytoreductive surgery followed by platinum-based chemotherapy either with or without paclitaxel between 1995 and 2005 at the Northern Gynaecological Oncology Centre, Gateshead, UK. No patient received intraperitoneal chemotherapy during the study period.

For each tumor sample, all histology reports and patient records were retrieved wherever possible, and data were extracted after ethical approval. Data were extracted from the patient record, including age, International Federation of Gynecology and Obstetrics (FIGO) stage, grade, histologic subtype, preoperative CA125, and outcome of surgery (Table 1, supplementary materials). All these data items are available to a clinician immediately after staging surgery. In fact, given the increased use of pretreatment biopsy and the routine use of cross-sectional imaging, these data items all can be provided to a clinician before any treatment is mandated. These data items were termed “features” for the purpose of this study. Although other features could have been considered for inclusion in the study, the aforementioned list was considered the minimum data set that would be available to all clinicians.

An important trend during the study period was the increase in complete cytoreduction rates from 30 % in 1995 to 55 % in 2005. Overall survival (all-cause mortality) also was recorded as an end point. Median survival increased from 25 months in 1995 to 35 months in 2005.

Generation of the AI System

A system then was created to generate overall patient survival as the output. The system was developed on a Java-based platform capable of analyzing multiple parameters using any of three algorithms [decision tree (DT), artificial neural network (ANN), and Bayesian network (BN)] and any of three classifiers [support vector machines (SVM), naïve Bayes (NB), and K-nearest neighbor (KNN)] (for further information, see the supplementary materials).

The system was thus capable of making multiple comparisons between different combinations of algorithms and classifiers using the data set as described earlier. Each analysis was stored and compared with the percentage performance of each technique on the appropriate data sets and sub-data sets.

Experimental Methodology

To compare different machine learning techniques and data sets, we first generated a process definition (Fig. 1). Briefly, the data set was first randomized. The initial training and evaluation of the model started with 190 cases randomly selected from the data set. Sequentially, 18 new cases then were added to the data set until all 668 available cases were included. This process allowed investigation of the model’s performance by increasing the number of data items and allowed discovery of the optimal stop-training point. Ultimately, this means that for each outcome investigated, a total of 7668 analyses were performed.

Each interval of the process contained both a training process and an evaluation process. The tenfold cross-validation was used to estimate the performance prediction.¹⁷

Survival was used as the end point in all analyses. The survival period of the patients in the data set ranged from 1 to 116 months. In each analysis, the median survival period for that particular cohort was calculated, and the accuracy of the system was calculated by the ability of the system to predict survival more or less than the median for that cohort. Patients for whom follow-up data were not available to at least the median time point were excluded from the analysis. Other performance indicators used were accuracy, recall, mean absolute error (MAE), root mean squared error (RMSE), and area under the curve (AUC), reviewed in the study of Cort and Kenji.¹⁸

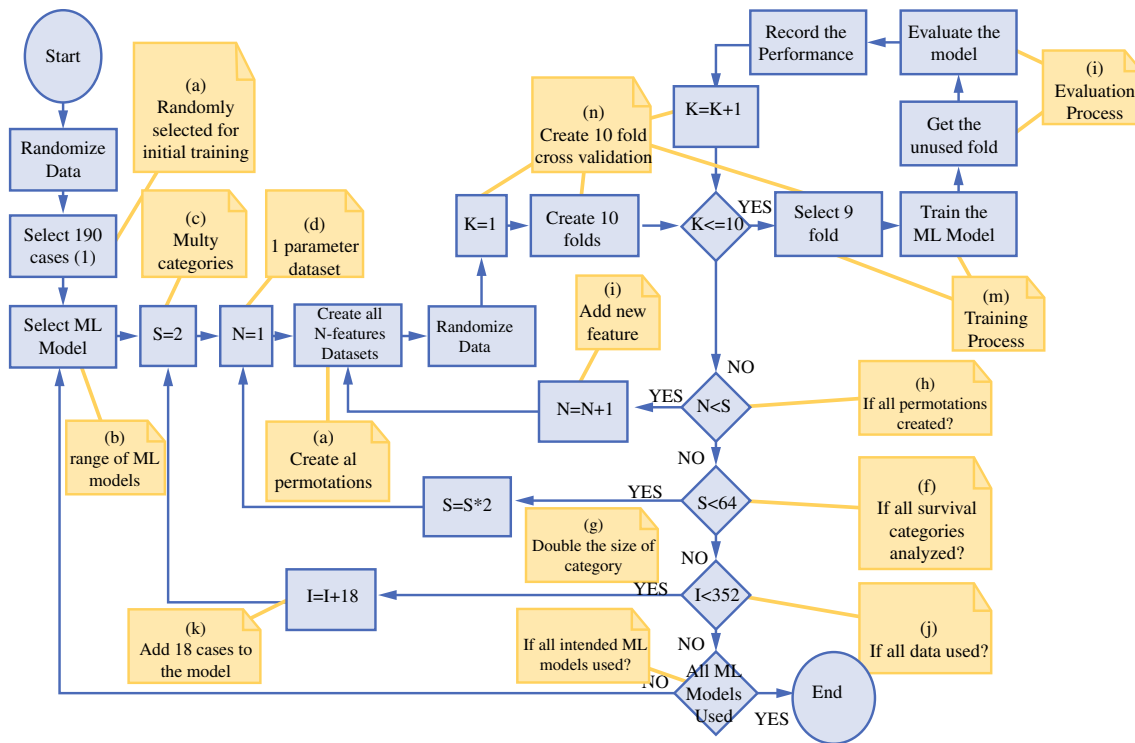


FIG. 1 Process definition algorithm. From the total cohort, 190 cases were selected at random and used as the test set. The remaining cases then were used as a validation set

RESULTS

Artificial Neural Networks Have the Highest Accuracy for Prognosis

Ability to Predict More or Less Than Median Survival. The model was used to predict survival, dichotomized around the median survival of the cohort. As outlined earlier, the model was tested initially with a random set of 190 cases and validated with the remainder. In this test, ANN outperformed the Bayesian networks and the decision tree, which had only limited predictive capacity (Fig. 2), and gave an overall predictive accuracy of 89 %. The AUC for the receiver operating characteristic curve (ROC) was 0.74, suggesting good prediction, and the MAE and RMSE had low scores of 0.16 and 0.30, respectively. Thereafter, all analyses were performed using ANNs.

When the number of cases in the validation set was increased in steps of 18 cases, the accuracy of the system also improved slightly to 92 % when the maximum 658 cases were used for test cases and the remaining 10 were used as validation cases.

Ability to Determine Prognosis for Smaller and Smaller Groups. The ability to determine survival to more or less than a single time point is of limited clinical value. We therefore wanted to investigate the performance of the system to determine prognosis in increasing numbers of

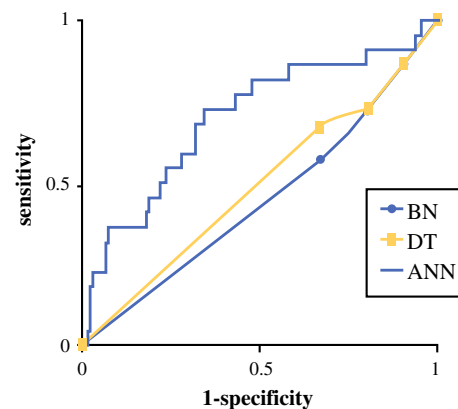


FIG. 2 Receiver operating characteristic (ROC) curve showing performance of model in predicting survival above the median using three different algorithms: Bayesian networks (BN), decision-tree analysis (DT), and artificial neural network (ANN). The latter clearly outperforms both BN and DT. Area under the curve (AUC) is 0.74

categories. We therefore divided the cohort into quartiles and then into 8, 16, 32, and ultimately, 64 categories.

As expected, the accuracy of the model decreased, whereas the MAE and RMSE increased in relation to prediction of survival with increasing numbers of categories from 89 % for 2 categories to just 6 % for 64 categories (Fig. 3).

Ability to Predict 5-Year Survival. Overall survival can be defined in terms of median survival or 5-year survival

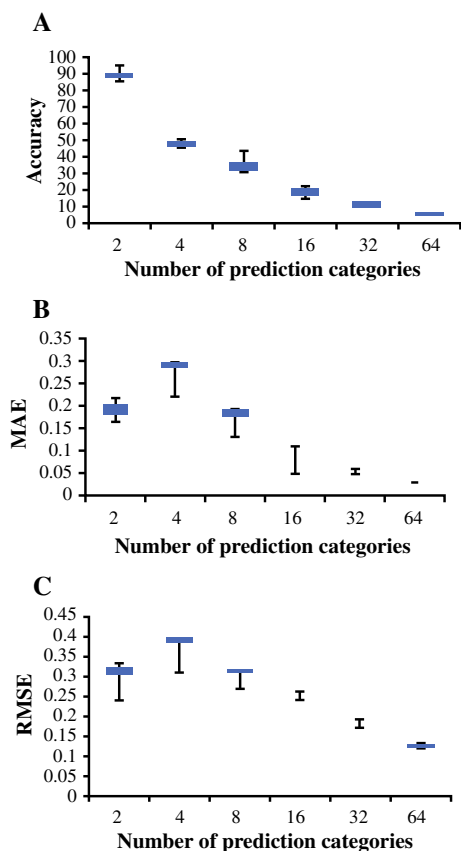


FIG. 3 a As the number of prediction categories was increased from 2 to 64, accuracy of the model decreased, although error as measured by b mean absolute error (MAE) and c root mean squared error (RMSE) was not adversely affected

rates. To allow comparison of our model against other published data, we therefore calculated the accuracy of our system in predicting 5-year (60-month) survival. The accuracy of the model in predicting 5-year survival was 93 %. The error rate values were relatively small, with an MAE of 0.16 and RMSE of 0.30. The area under the ROC curve (AUC) of 0.74 indicated that the model performed well.

AI Can be Used to Provide Predictive Information Regarding Surgical Resection

In addition to use of the model as a prognostic tool, we tested its ability to predict outcome. We chose to evaluate the ability of the system to test outcome of surgery because this is arguably one of the most important treatment decisions needing to be made with a patient.¹⁹ The features used were age, stage, grade, histologic type, and preoperative Ca125. All these features would be available to a surgeon after a biopsy but before performance of a laparotomy.

The accuracy of the model to predict complete, optimal, or suboptimal cytoreduction was 59.75 %, with an MAE value of 0.34 and an RMSE value of 0.42. The AUC for complete cytoreduction curve was 0.75. The AUC for optimal cytoreduction was 0.66, and the AUC for suboptimal cytoreduction was 0.71 (Fig. 4).

The question then was further refined to give a prediction of “complete and optimal” as the one category and “suboptimal” as the other category. In this situation, the performance of the model improved to give an accuracy of 77.7 % associated with an MAE of 0.29 and an RMSE of 0.38.

Comparison of the AI System with Conventional Statistics

To have a complete comparison between models, logistic regression models as standard conventional statistical prediction models were developed. The results of the logistic regression models then were compared against the ANN results.

Although Cox proportional hazards models are more commonly used for survival analysis and can be applied to predict survival probabilities, such models are better suited for making comparisons between two different treatment arms (comparing the hazards between each). Given that this study aimed to develop prediction for an individual patient, logistic regression models were more appropriate.

The model, using ANN as the algorithm, consistently outperformed logistic regression in its ability to predict survival more or less than the median for each cohort. In 5998 (75 %) of 7998 analyses, ANN had a greater accuracy than logistic regression (Fig. 5).

Using AUC to make a direct comparison, the AUC for logistic regression was 0.62 compared with the AUC of 0.72 for ANN. Compared with logistic regression, ANN also was associated with lower MAE and RMSE scores.

DISCUSSION

This report describes the development of an AI model capable of providing prognostic and predictive information in ovarian cancer. The model uses readily available data items as input variables. We have measured success of the model in predicting overall survival, which is the most clinically important outcome measure. Thus, this tool may have direct clinical relevance for patients and their treating physicians. It also may be of use in the research setting in which reliable outcome data must be obtained without long follow-up periods.

With the system optimized, we are able to predict survival for patients in our cohort with an accuracy of 93 %

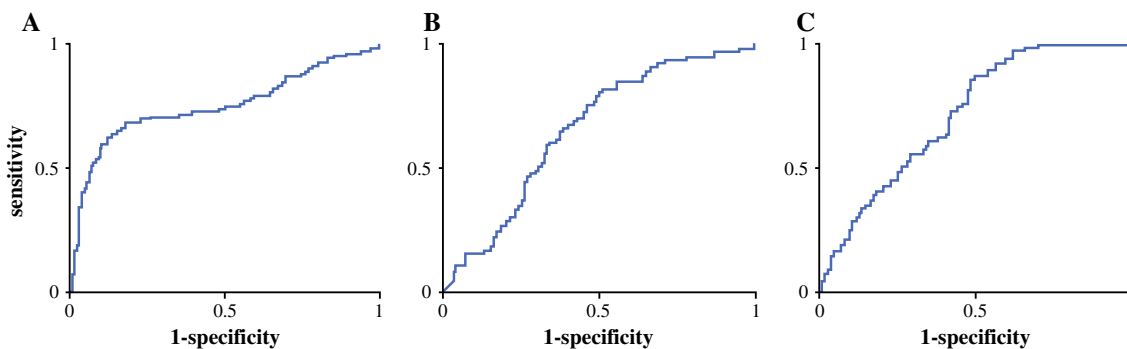


FIG. 4 Receiver operating characteristic (ROC) curve demonstrating ability of the model to predict the outcome of surgery using an artificial neural network (ANN) algorithm. The model was used to

predict **a** complete [area under the curve (AUC), 0.75], **b** optimal cytoreduction (AUC, 0.66), and **c** suboptimal cytoreduction (AUC, 0.71)

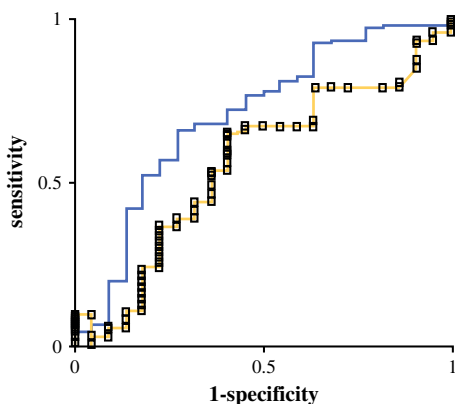


FIG. 5 Receiver operating characteristic (ROC) curve demonstrating prediction of survival less than the median survival of the cohort comparing artificial neural network (ANN) with logistic regression (LR). The ANN is associated with better prediction [area under the curve (AUC), 0.72] than LR (AUC, 0.62)

compared with the median survival of the group (AUC, 0.72). Clearly, the ability to predict whether a patient will survive to more or less than a single time point has limited clinical use, but given the size of this data set, we believe that these are encouraging results. The ability of the system to predict survival with more outcome categories decreased, but even when 64 outcome categories were used (which is almost equivalent to predicting survival by month), the system could accurately predict survival in 6 % of cases.

We also used the model to provide predictive information and chose outcome of surgery (the ability to identify patients likely to end up with suboptimal cytoreduction) as an important clinical question.¹⁹ To date, conventional predictive systems have had limited accuracy.²⁰ Although our model also had only limited accuracy (77 %), this was based on data easily available to all clinicians and did not take into account any findings from radiologic investigations or comorbidities of the patient. Such features can easily be incorporated into an AI system, and future

investigations will examine whether incorporating such features improves the accuracy of the model further. These results are consistent with those of Jefferson et al.,²¹ who showed that neural networks can outperform logistic regression in predicting surgical outcome in lung cancer.

Our results suggest that the model described in this report is a better predictor of outcome than a system based on conventional statistics. Two systems based on conventional statistics have been described for use in ovarian cancer: Gerestein et al.²² developed a nomogram using three or five parameters that has a c statistic of 0.67, whereas Teramukai et al.²³ developed a similar nomogram, which has subsequently been validated in a UK-based population of patients.²⁴ Both of these models are finite (i.e., their performance is fixed and will not improve with time). In contrast, a systems-based approach, such as the one described in this report, can be iterative, and performance is likely to improve as the data set enlarges. In this report, we demonstrated modest increases in performance as the number of cases was increased, which supports this concept. It also is likely that as the number of cases in the data set increases, the relative weightings of some of the input parameters will change. Thus, we will continue to input as many data items as are available and monitor the performance of the system to test this further.

Ovarian cancer is an extremely heterogeneous disease with a variety of different histologic subtypes and a wide range of responses to treatment. By developing a model in this disease, we are confident that similar models also can be developed in more homogeneous cancers. Artificial intelligence systems also have other significant advantages over conventional models. In particular, they can cope with missing data items and will incorporate algorithms to overcome this (see supplementary data). Their greatest advantage, however, is that they are iterative, and performance improves as the data sets increase in size and relevance. Therefore, although this model has been tested in only one center, it could be adapted for use in multiple

centers but would need to “learn” how to predict in each center given that the outcomes (e.g., for surgery) are likely to be different.

Current research in ovarian cancer is focused on developing new treatments, many of which will be biomarker driven (i.e., a test will be available that will dictate whether the patient is likely to respond to the treatment in question). Such a model is already standard care in breast cancer, with a test that identifies the presence of the HER2 receptor and dictates whether a patient receives herceptin treatment.

Therefore, within a relatively short time, we likely will have a multitude of new treatments available, all associated with a particular biomarker test. The challenge for the clinician will be to prioritize these treatments and determine the optimal treatment strategy. Because conventional algorithms probably will not be capable of handling such complex interactions as are likely to exist, the development of AI systems will become paramount. We hope that the system described in this report can be the basis for the development of these more complex models.

Future work will include increasing the size of the data set to increase the performance of the system and validating it further on other external data sets of patients.

ACKNOWLEDGMENTS Support for A. Enshaei was provided by a Grant from the EPSRC and Catherine Cookson Foundation. The data set was created using a Grant from the Northern Cancer Care and Research Society.

CONFLICT OF INTEREST There are no conflicts of interest.

REFERENCES

1. ESMO Minimum clinical recommendations for diagnosis, treatment, and follow-up of ovarian cancer. *Ann Oncol.* 2001;12:1205–7.
2. Thigpen T, et al. Age as a prognostic factor in ovarian carcinoma. *Cancer.* 1993;71:606–14.
3. Benedet J, et al. FIGO staging classifications and clinical practice guidelines in the management of gynecologic cancers. *Int J Gynecol Obstet.* 2000;70:209–62.
4. van Houwelingen J, et al. Predictability of the survival of patients with advanced ovarian cancer. *J Clin Oncol.* 1989;7:769–73.
5. Makar AP, et al. The prognostic significance of residual disease, FIGO substage, tumor histology, and grade in patients with FIGO stage III ovarian cancer. *Gynecol Oncol.* 1995;56:175–80.
6. Hogdall E. Cancer antigen 125 and prognosis. *Curr Opin Obstet Gynecol.* 2008;20:4–8.
7. Bristow RE, et al. Survival effect of maximal cytoreductive surgery for advanced ovarian carcinoma during the platinum era: a meta-analysis. *J Clin Oncol.* 2002;20:1248–59. DOI:10.1200/JCO.20.5.1248.
8. du Bois, A, Harter P. The role of surgery in advanced and recurrent ovarian cancer. *Ann Oncol.* 2006;17(Suppl 10):x235–40.
9. Perren TJ, Swart AM, Pfisterer J, Ledermann JA, Pujade-Lauraine E, Kristensen G, Carey MS, Beale P, Cervantes A, Kurzeder C, du Bois A, Sehouli J, Kimmig R, Stähle A, Collinson F, Essapen S, Gourley C, Lortholary A, Selle F, Mirza MR, Leminen A, Plante M, Stark D, Qian W, Parmar MK, Oza AM, ICON7 Investigators. A phase 3 trial of bevacizumab in ovarian cancer. *N Engl J Med.* 2011;365(26):2484–96.
10. Ledermann JA, Harter P, Gourley C, Friedlander M, Vergote IB, Rustin GJS, et al. Phase II randomized placebo-controlled study of olaparib (AZD2281) in patients with platinum-sensitive relapsed serous ovarian cancer (PSR SOC). *J Clin Oncol.* 2011;29(Suppl 15):5003.
11. Sjoquist K, et al. The Role of Hormonal therapy in Gynaecological Cancers – Current Status and Future Directions. *Int J Gynecol Cancer.* 2011;21:1328–33.
12. Collinson F, et al. Predicting response to bevacizumab in ovarian cancer: a panel of potential biomarkers informing treatment selection. *Clin Cancer Res.* 2013;19:5227–39.
13. Mukhopadhyay A. et al. Development of a functional assay for homologous recombination status in primary cultures of epithelial ovarian tumor and correlation with sensitivity to poly(ADP-ribose) polymerase inhibitors. *Clin Cancer Res.* 2010;16:2344–51.
14. Elattar A, et al. Androgen receptor expression is a biological marker for androgen sensitivity in high-grade serous epithelial ovarian cancer. *Gynecol Oncol.* 2012;124:142–47.
15. Narayanan A, Keedwell E, Olsson B. Artificial intelligence techniques for bioinformatics. *Appl. Bioinform.* 2002;1:191–222.
16. Wilkinson SJ, et al. Expression of gonadotrophin releasing hormone receptor I is a favorable prognostic factor in epithelial ovarian cancer. *Hum Pathol.* 2008;39:1197–204.
17. Dziuda D. Data Mining for Genomics and Proteomics: Analysis of Gene and Protein Expression Data. Wiley, New Jersey, 2010.
18. Cort JW, Kenji M. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Res.* 2005;30:79–82.
19. Bristow RE. Predicting surgical outcome for advanced ovarian cancer, surgical standards of care, and the concept of kaizen. *Gynecol Oncol.* 2009;112:1–3.
20. Salani R, et al. Limited utility of conventional criteria for predicting unresectable disease in patients with advanced stage epithelial ovarian cancer. *Gynecol Oncol.* 2008;108:271–5.
21. Jefferson MF, et al. Comparison of a genetic algorithm neural network with logistic regression for predicting outcome after surgery for patients with nonsmall cell lung carcinoma. *Cancer.* 1997;79:1338–42.
22. CG Gerestein ME, de Jong D, van der Burg MEL, Dykgraaf RHM, Kooi GS, Baalbergen A, Burger CW, Ansink AC. The prediction of progression-free and overall survival in women with an advanced stage of epithelial ovarian carcinoma. *BJOG Int J Obstet Gynaecol.* 2009;116:372–80.
23. Teramukai S, et al. PIEPOC: A new prognostic index for Advanced Epithelial Ovarian Cancer Japan Multinational Trial Organization OC01-01. *J Clin Oncol.* 2007;25:3302–6. DOI:10.1200/JCO.2007.11.0114.
24. Clark TG, et al. Validation of a new prognostic index for advanced epithelial ovarian cancer: results from its application to a UK-based cohort. *J Clin Oncol.* 2007;25:5669–70. DOI:10.1200/JCO.2007.14.2976.