Original Research Article

Feature Selection is Critical for 2-Year Prognosis in Advanced Stage High Grade Serous Ovarian Cancer by Using Machine Learning

Alexandros Laios, MD, PhD[1,*] , Angeliki Katsenou, PhD[2,*], Yong Sheng Tan, MBBS[1],
Racheal Johnson, MBChB, MRCOG[1], Mohamed Otify, MBBCh, MRCOG, MSc[1],
Angelika Kaufmann, MBBS, MRCOG, PhD[1], Sarika Munot, MBBS[1],
Amudha Thangavelu, MBChB, MRCOG, MD[1], Richard Hutson, MBBS, MRCOG, MD[1],
Tim Broadhead, MBChB, MRCOG[1], Georgios Theophilou, MBBS, MRCOG, MD[1],
David Nugent, MBChB, MRCOG, PhD[1], and Diederick De Jong, MBBCh, PhD, MSc[1]

Abstract

Introduction: Accurate prediction of patient prognosis can be especially useful for the selection of best treatment protocols. Machine Learning can serve this purpose by making predictions based upon generalizable clinical patterns embedded within learning datasets. We designed a study to support the feature selection for the 2-year prognostic period and compared the performance of several Machine Learning prediction algorithms for accurate 2-year prognosis estimation in advanced-stage high grade serous ovarian cancer (HGSOC) patients.

Methods: The prognosis estimation was formulated as a binary classification problem. Dataset was split into training and test cohorts with repeated random sampling until there was no significant difference (p = 0.20) between the two cohorts. A ten-fold cross-validation was applied. Various state-of-the-art supervised classifiers were used. For feature selection, in addition to the exhaustive search for the best combination of features, we used the-chi square test of independence and the MRMR method.

Results: Two hundred nine patients were identified. The model's mean prediction accuracy reached 73%. We demonstrated that Support-Vector-Machine and Ensemble Subspace Discriminant algorithms outperformed Logistic Regression in accuracy indices. The probability of achieving a cancer-free state was maximised with a combination of primary cytoreduction, good performance status and maximal surgical effort (AUC 0.63). Standard chemotherapy, performance status, tumour load and residual disease were consistently predictive of the mid-term overall survival (AUC 0.63–0.66). The model recall and precision were greater than 80%.

Conclusion: Machine Learning appears to be promising for accurate prognosis estimation. Appropriate feature selection is required when building an HGSOC model for 2-year prognosis prediction. We provide evidence as to what combination of prognosticators leads to the largest impact on the HGSOC 2-year prognosis.

[1]Department of Gynaecologic Oncology, Leeds Teaching Hospitals, St James's University Hospital, Leeds, UK
[2]Department of Electrical and Electronic Engineering, Visual Information Lab, University of Bristol, Bristol, UK

*AL and AK claim equal distribution.

Corresponding Author:
Alexandros Laios, MD, PhD, Postdoc(Ox), Department of Gynaecologic Oncology, St James's University Hospital, Leeds Teaching Hospitals, Beckett Street, Leeds LS9 7TF, UK.
Email: a.laios@nhs.net

Introduction

Cancer of the fallopian tube, ovary or peritoneum ranks as the seventh most common cancer in women and the eighth most common cause of cancer death.[1] It yet remains one of the most difficult cancers to combat with most patients relapsing within 3 years of diagnosis.[2] The majority (90%) of these cancers are epithelial ovarian cancers (EOCs). High-grade serous ovarian cancer (HGSOC) is the most prevalent form among EOCs and is now recognised as a single entity. Indeed, of the women who die of HGSOC, 93% present with advanced-stage (International Federation Obstetrics and Gynaecology FIGO stage-III or IV) disease.[3] Interestingly, HGSOC women who receive surgical treatment have better long-term survival than those who do not, despite being diagnosed at an advanced stage.[4]

The cornerstones of advanced-stage HGSOC treatment are surgical cytoreduction and platinum-based backbone chemotherapy, either as treatment following surgery (adjuvant) or as treatment both before and after surgery (neoadjuvant, NACT).[5] Optimal cytoreduction and initial tumour load are the most significant modifiable markers of survival.[6,7] Following recent publications of landmark randomised studies demonstrating non-inferiority of NACT over primary surgery, it appears that NACT achieves higher complete cytoreduction (R0) rates, but the survival rates are comparable.[6,8] Even when EOC patients undergo complete surgical cytoreduction and systemic chemotherapy, the risk for tumour relapse remains high.

Accurate estimation of EOC patient prognosis can be particularly useful for enhancing diagnostic precision and selection of best treatment protocols. Due to the EOC heterogeneity, a one-size-fits-all FIGO staging system approach is not justified. As the number of clinical and biological parameters under investigation increases daily, it becomes critical to assemble a large and heterogeneous amount of data and construct appropriate models.[9] Prognosis estimation can be difficult with conventional statistics because patient characteristics show multidimensional and non-linear relationship. To develop personalised treatment plans, computational approaches, such as Machine Learning (ML) models can serve the purpose by making predictions using multiple processing layers, including complex structures or multiple non-linear transformations. The evolution of ML technology in the field of gynaecological oncology has been described.[10] We previously demonstrated the feasibility of using a ML approach, the k-NN model, which is very much reflective of 'previous clinical experience' for accurate prediction of complete cytoreduction in advanced-stage HGSOC surgery.[11]

We aimed to develop a data-driven framework by using modern ML to predict the survival outcomes of HGSOC patients from many clinical patient-specific features. We hypothesised that the prognosis prediction of HGSOC patients is multifactorial and could be accurately predicted by using ML algorithms. We performed a comparative analysis to examine the mid-term contribution of selected clinical variables to define their relative survival impact. When developing a cancer prognosis prediction model, model performance is not the sole goal but also extracting the most relevant features to better understand the data and the underlying process. Feature selection is a key step in many classification problems.[12] The study was designed to support the feature selection for different prognostic periods, using the prospectively registered data of HGSOC women, who received surgical treatment. The primary outcome was factor analysis using the Maximum Relevance Maximum Redundancy (MRMR) method for different prognostic periods.[13] The secondary outcome was the performance comparison amongst several ML prediction methods, based on a set of performance metrics,[14] including the accuracy, the sensitivity and specificity of the model, the precision and recall, the f-score and the g-score (or Fowlkes–Mallows index[15]) for different prognostic periods. These results were directly compared to conventional Logistic Regression.

Study Design

The study was structured in two basic workflows, which ultimately integrated into one: the clinical and the engineering workflows. The clinical workflow consisted of the patient input, the patient–clinician interaction and the hospital site part. Most processes in the clinical workflow were related to the data-acquisition, data cleaning, data pre-processing and statistical compilation before feeding them in the engineering workflow. The engineering workflow included all processes related to the data processing feature extraction and ML-based feature selection and prognosis prediction. The workflow, outlined here and described in detail below, is illustrated in the conceptual diagram in Figure 1.

Prospective registered data in the hospital-wide Patient Pathway Manager (PPM) database from 209 HGSOC women undergoing cytoreductive surgery at St James's University Hospital, Leeds from January 2015 to December 2018 were analysed. This database was developed internally for clinical trials and integrated with an electronic patient record system. Our hospital is a tertiary centre, recently accredited by the European Society of Gynaecologic Oncology (ESGO) as a centre of excellence for ovarian cancer surgery. Inclusion criteria included women >18 years of age and FIGO stage III–IV HGSOC. Excluded were women with non-serous and non-epithelial histology, and those undergoing secondary
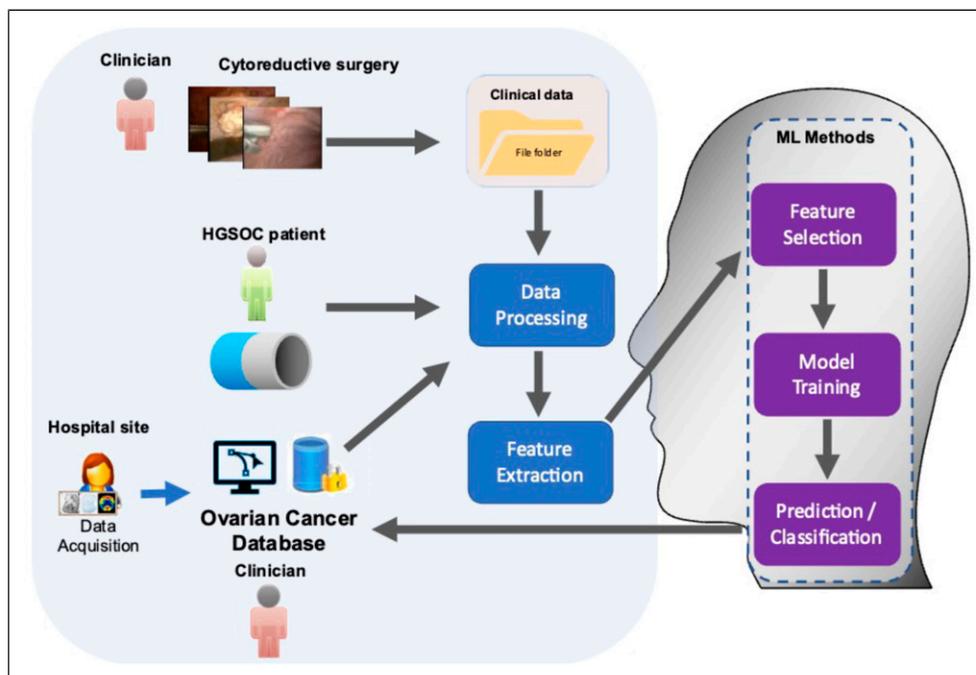
Figure 1. Workflow showing integration of ML algorithms to analyse comprehensive resource of clinical, radiological and surgical data for the development of prognostic ovarian cancer models. The framework for building the predictive ML model comprised 5 steps.

cytoreductive surgeries for recurrent disease. The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Leeds Teaching Hospitals Trust Institutional Review Board (MO20/133163/18.06.20), and informed written consent was obtained. All patients were discussed at the central gynaecological oncology multidisciplinary team (MDT) meeting prior to treatment. Contrast-enhanced Computed Tomography (CT) of the thorax, abdomen and pelvis was performed within a month prior to treatment initiation, interpreted and reported by an MDT radiologist. Three pre-treatment imaging dissemination patterns were identified; intraperitoneal (Group 1), intraperitoneal and lymphatic (Group 2) intraperitoneal and haematogenous patterns (Group 3), respectively. This was confirmed by final histology. Descriptive cohort statistics were summarised by frequency and percentages for binary and categorical variables, and by means and standard deviations (SD) or medians (with lower or upper quartiles) for continuous variables (Table 1). Survival data were summarised using the Kaplan–Meier method. A Cox proportional hazard regression analysis was performed to identify prognostic factors. Statistical tests were two-tailed with a significance level set at $P<.05$. All analyses were performed using SPSS 26® package.

For the prognosis classification, 2 groups were defined using patient survival data; patients who did not relapse or survived beyond 2 years were labeled in the positive class, and patients who relapsed or died before reaching that period were considered in the negative class.

The study was restricted to the most common prognostic variables and focused on predictive model comparisons (Table 1). Blood biomarkers such as preoperative Hb and Ca125 were not included as they appear more reliable to predict surgical outcomes or simply predict malignancy in women with adnexal masses.[16,17] Equally, surveillance modalities are not used to comprehensively evaluate the prognosis of the HGSOC patients provided that the primary objective of follow-up is to detect disease that if treated early can extend survival. It is not to prolong time living with the knowledge that cancer has relapsed without extending survival.[18] Performance variables included age, Eastern Cooperative Oncology Group (ECOG) performance status (PS), radiological intraperitoneal dissemination patterns (IDP), surgical complexity score (SCS),[19] residual disease (RD), chemotherapy regimens, timing of surgery (primary debulking surgery (PDS) or interval debulking surgery (IDS)) and intra-operative disease score (DS), which is a reflection of the tumor burden. Surgical outcomes included: complete cytoreduction (R0), optimal cytoreduction (R1, 1–10 mm) or inadequate cytoreduction (R2>10 mm).[20] The SCS was assigned based on the Aletti classification as low, intermediate and high.[19] The response to chemotherapy and disease progression was defined according to RECIST criteria.[21] The DS was assigned as follows: pelvic disease, lower abdominal, upper abdominal inclusive of miliary disease, as women with miliary disease often have disease in the upper abdomen,[22,23] Progression-free survival (PFS) was defined as the time from the date of diagnosis until

Table 1. Descriptive Statistics of the Advanced-HGSOC Cohort.

Variables (n = 209)	Frequency	Percent (%)
Age, year, mean, SD (range)	64.6±10.6 (41–85)	
Surgical Complexity Score (SCS)		
Low (1–3)	124	59.3
Moderate (4–7)	76	36.4
High (8–12)	9	4.3
Radiological dissemination patterns		
Intraperitoneal	134	64.1
Intraperitoneal and lymphatic	59	28.2
Intraperitoneal and haematogenous	16	7.7
Operation time, mean, SD (min-max)	177±77 (45–485)	
Disease score		
Pelvis (1)	10	4.8
Lower abdomen (2)	187	89.5
Upper abdomen (3)	12	5.7
Timing of surgery		
PDS	46	20
IDS	163	80
Residual disease		
R0	160	76.5
R1	39	18.7
R2	10	4.8
Chemotherapy		
Carboplatin+Taxol	134	64.1
Carboplatin+Taxol+Bevascusimab	22	10.5
Carbo+Taxol+PARP inhibitor	25	12.0
Carboplatin only	22	10.5
No	6	2.9

relapse or death. Overall survival (OS) was defined as the time from the date of diagnosis until death.

The dataset was split into training and test cohorts (80%: 20% ratio) with repeated random sampling, until there was no significant difference ($P = .20$) between the two cohorts, with respect to all variables. Subjects with missing values were omitted. Following the pre-processing stage, all quantitative variables were normalised. Categorical variables were transformed into binary dummy variables. Next, different subsets of data were labelled to solve the prognosis prediction problem. For a given time T, subjects were included or discarded from the subset. To test the HGSOC prognosis, 3 values of the prognosis period T were chosen, namely, one, two and three years. The 5-year prediction was not considered owing to data immaturity. Following preliminary testing, due to unbalanced classes, it was not possible to train a good model for the 1-year and 3-year prognosis prediction. Therefore, we focused on the 2-year prediction analysis. Only subjects with fully curated data were eligible for the 2-year prognosis prediction analysis. The prognosis prediction was then formulated as a binary classification problem. The correction for class imbalance was applied only on our efforts with the 3-year and 5-year prognosis prediction. It was applied before training the models. A repeated random selection of the prevailing class was performed to ensure statistical validity (100 iterations). For the results presented here, we did not apply any such correction on the dataset.

To address data collinearity, feature selection techniques measured the importance of a feature or a set of features according to a given measure. For feature selection, in addition to the exhaustive search for the best combination of features, we used the chi square test of independence[24] and the MRMR method,[13] as typically recommended for categorical data. The outcome of these methods is a feature ranking that shows the weighted importance of the individual features. Both methods were applied for the 2-year prognosis. The resulting rankings were used to select the set of features that led to the highest prediction accuracy. The validity of the feature selection was verified by comparing it to the exhaustive search and regularization methods. Subsequently, the optimal number of important features that would result in the highest prediction accuracy was identified. For this step, a forward selection was followed by starting with the feature of highest importance and subsequently adding features, until we reached the maximum classification accuracy.

The prognosis estimation problem was formulated as a binary classification problem. Various state-of-the-art supervised classifiers, suitable for the type and size of the dataset, were trained and tested, including Support-Vector-Machines (SVMs),[25] K-Nearest Neighbors (K-NNs),[26] Ensemble Classifiers,[27] Naïve Bayes,[28] and Logistic Regression.[29] The SVMs are highly accurate even for non-linear problems. Different kernels SVMs are flexible to identify the optimum hyperplane, to best separate the data into their categories, albeit slow for large datasets. The K-Nearest Neighbors are robust classifiers for low-dimensionality classification problems. Ensemble methods are frequently used for categorical data due to their inherent properties. They combine several different decision trees to produce better predictive performance compared to single decision trees. Bagging is a combination of decision trees to optimise the variance. We also experimented with probabilistic techniques for classification, such as Naïve Bayes and Logistic Regression. Naïve Bayes algorithms are built on the concept of conditional probability; these classifiers are computationally efficient, thus scalable to the size of the dataset and the feature set cardinality. Similarly, Logistic Regression, conventionally used in the clinical setting, gives off fast results, but has the difficulty of capturing non-linear relationships in the dataset. Due to the limitations of the dataset with respect to its size and the classes' cardinality, 'data-hungry' deep-learning based classification methods were not included in this comparison. We considered Logistic Regression as our benchmark method. To promote reproducibility, the code and the model parameters

promotes were made publicly available: https://github.com/AngKats/OCPrognosis

Results

A total of 209 HGSOC patients were identified from the hospital-wide database PPM. The cohort characteristics are summarised in Table 1. The median age and median SCS were 66 (41–85) years and 3 + 1 (1–8), respectively. Of these patients, 46/209 (20%) underwent PDS and 163/209 (80%) underwent IDS, respectively. Complete (R0) and optimal (R1) cytoreduction was achieved in 160/209 (76.5%) and 39/209 (18.7%) patients, while 10/209 (4.8%) had RD>1 cm (R2). Cox regression analysis for PFS and OS identified significant prognostic variables (Table 2). The median PFS and OS for the entire cohort were 19 months (95% CI 16.4–21.6) and 38 months (95% CI 34.4–41.6), respectively. In the complete cytoreduction group, the median PFS and OS were 20 months (95% CI 16.8–23.3) and 41 months (95% CI 30.5–51.5), respectively. In the incomplete cytoreduction group, the median PFS and OS were 18 months (95% CI 14.3–21.7) and 28 months (95% CI 18.3–37.6), respectively, (Figure 2A and B). Women with intraperitoneal-only pattern of their disease

distribution had the highest rate of complete cytoreduction (77.9%), resulting in markedly improved OS compared to the other subgroups (*P:* .05) (Figure 2C and D). 172/209 patients with fully curated data were eligible for the 2-year prognosis prediction analysis. 104/172 (60%) and 55/172 (32%) patients had disease recurrence or died of disease within 2 years, respectively.

We estimated the relative importance of the features using the chi-square test and the MRMR approaches. The results are shown in Figure 3. For the 2-year survival prediction, the mean predictive accuracy of the ML models reached 73%. As expected, the feature importance between PFS and OS outcomes was not identical. For the 2-year OS prognosis prediction, the two best performance results were achieved with the SVM – Quadratic Kernel classifier using the top-3 features (standard chemotherapy, low DS and increased SCS) selected by the MRMR algorithm Area-Under-Curve (AUC = .66) and the k-NN (5 Neighbors) with the top-4 features (standard chemotherapy, no RD, PS and low DS selected by the chi-square test [AUC = .63]) (Figure 4). The combination of good PS, PDS and increased SCS best predicted 2-year PFS with the accuracy reaching 63.5% (AUC = .62) by the SVM – Quadratic Kernel – classifier.

Table 2. Cox-Regression with Progression-Free Survival and Overall Survival as Outcomes.

Variables	Progression-free survival (PFS)						Overall survival (OS)					
	Univariate analysis			Multivariate analysis			Univariate analysis			Multivariate analysis		
	HR	P	95% CI	HR	P	95% CI	HR	P	95% CI	HR	P	95% CI
Age	.997	.742	.983–1.01	.995	.661	.971–1.019	1.004	.672	.98–1.03	.983	.284	.951–1.015
ECOG performance status (PS) (0)	1.000	.133		1.000	.19		1.000	.002		1.000	.131	
ECOG performance status (PS) (1)	0.5	.085	.23–1.1	.46	.08	.2–1.1	.289	.006	.12–.69	.39	.061	.14–1.04
ECOG performance status (PS) (2)	.52	.115	.24–1.15	.55	.17	.23–1.3	.367	.027	.15–.91	.53	.23	.2–1.49
ECOG performance status (PS) (3)	.75	0.5	.32–1.73	.71	.44	.3–1.7	.716	.47	.29–1.71	.71	0.5	.26–1.94
IP dissemination (1)	1.000	.158		1.000	.188		1.000	.009		1.000	.007	
IP dissemination (2)	0.1	.630	.357–1.1	.734	.336	.392–1.378	0.5	.048	.25–.99	.556	.127	.262–1.182
IP dissemination (3)	.49	.811	.44–1.48	1.035	.919	.534–2.01	.957	.904	.46–1.95	1.226	.603	.570–2.637
PDS	1.43	.084	.95–2.14	1.610	.039	1.026–2.529	1.648	.087	.93–2.92	2.008	.039	1.035–3.894
Residual disease (RD)	.671	.034	.464–.97	.656	.046	.433–.992	.422	<.001	.27–.66	.437	.001	.264–.724
Surgical complexity score (SCS)-low	1.000	.494		1.000	.852		1.000	.763		1.000	.825	
Surgical complexity score (SCS)-intermediate	.98	.958	.452–2.12	1.102	.865	.359–3.387	1.173	1.173	.36–3.75	.596	.535	.116–3.061
Surgical complexity score (SCS)-high	.795	.574	.358–1.76	.926	.957	.377–2.43	.99	.99	.29–3.27	1.226	.572	1.67–2.685
Operation time	1.000	.921	.998–1.02	1.001	.686	.997–1.004	.708	.999	.997–1.01	.999	.522	.994–1.02
Carboplatin and Taxol	25,04	.03	2.93–213.7	34.56	.002	3.83–311.65	18.09	.008	2.11–155.1	43.77	.001	4.45–430.19
Disease score (DS) (1)	1.000	.947		1.000	.810		1.000	.592		1.000	.516	
Disease score (DS) (2)	.941	.914	.31–2.83	1.292	.676	.389–4.28	.483	.308	.12–1.95	.717	.671	.154–3.332
Disease score (DS) (3)	.883	.884	.38–2.01	.984	.972	.41–2.364	.671	.442	.24–1.85	.552	.287	.185–1.649
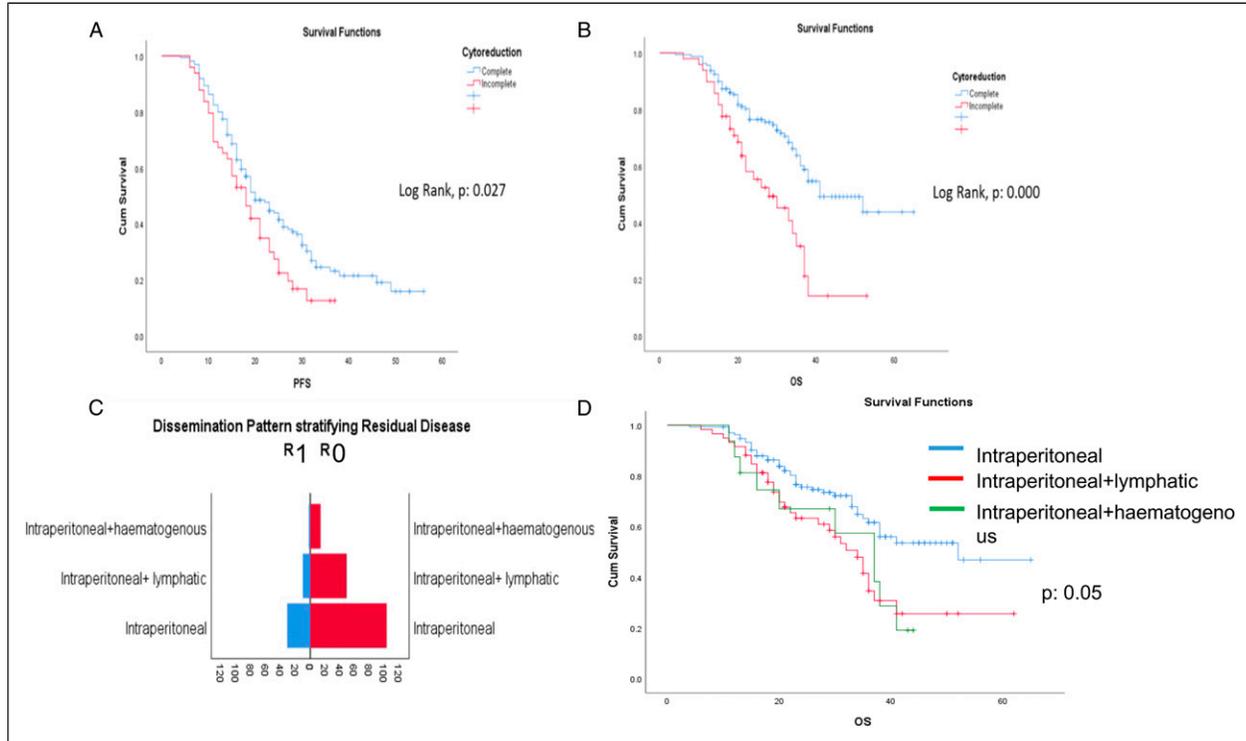
Figure 2. Cohort survival outcomes. Kaplan–Meier curves demonstrating (A) PFS and (B) OS analysed by complete and incomplete cytoreductive outcomes. (C) Stratification of residual disease according to intraperitoneal dissemination pattern. (D) Kaplan–Meier curves demonstrating OS according to IDP. Haematogenous metastases negatively affect OS, potentially highlighting difficulty to achieve complete cytoreduction (p:0.000).

To fully evaluate the effectiveness of the model, we considered other performance metrics that could also capture the balance of the data classes, irrespective of the prediction accuracy. Therefore, we calculated and reported in Table 3, an extended set of metrics such as precision (positive predictive value), recall (sensitivity), f-score, g-score and the AUC classes' values.[30]

Discussion

Women with HGSOC have a heterogeneous response to treatment and prognosis. Establishing the prognosis of HGSOC women remains a critical part of their evaluation. Machine Learning appears a promising approach for accurate prognosis estimation.[31] We demonstrated the feasibility and validity of using feature selection algorithms to ensure the highest performance of the 2-year prognosis ML prediction model. We employed the chi-square test of independence[24] and the MRMR method[13] for categorical data in a stepwise fashion, and verified the validity of the feature selection by comparing it to the exhaustive search method. After applying feature ranking with the described methods, we followed a feed-forward selection approach,[32] considering the ranking of the features for each different ML model. Forward selection is an iterative method in which, at each iteration, we continue to add the feature which best improves our model, until an addition of a new variable does not improve the performance of the model. The feed-forward selection helped define the set of lower number of features that provided the highest accuracy of prediction.

Classification problems typically involve a high time complexity and low performance when many features are used but will have a low time complexity and high performance for a minimum size and the most effective features.[33] HGSOC prognosis is a complex matter and failure to address this, can lead to a less meaningful interpretation of outcome data. Nevertheless, our effort allowed us to minimise redundancy and identify those discriminant features with the maximal relevance to the 2-year prediction estimation.

We adopted a binary classification approach to exploit the use of predictive ML models. Several different ML models were explored and tested. The SVM and k-NN algorithms outperformed the Logistic Regression model with respect to prediction accuracy indices. The maximum accuracy reached 73%. The predictive accuracy of the 2-year PFS was lower than the 2-year OS for all models due to cardinality of the classes. Firstly, the data classes were imbalanced, as indicated for the 2-year prognostic periods. Unbalanced classes lead to insufficient training for the less populated class, thus biasing the prediction towards the more populated class. This was reflected in the difference between the AUC values for the 2 classes, but also in the wide variation amongst other classification performance metrics, against the accuracy, as reported in Table 3. This justified the use of AUC as a performance
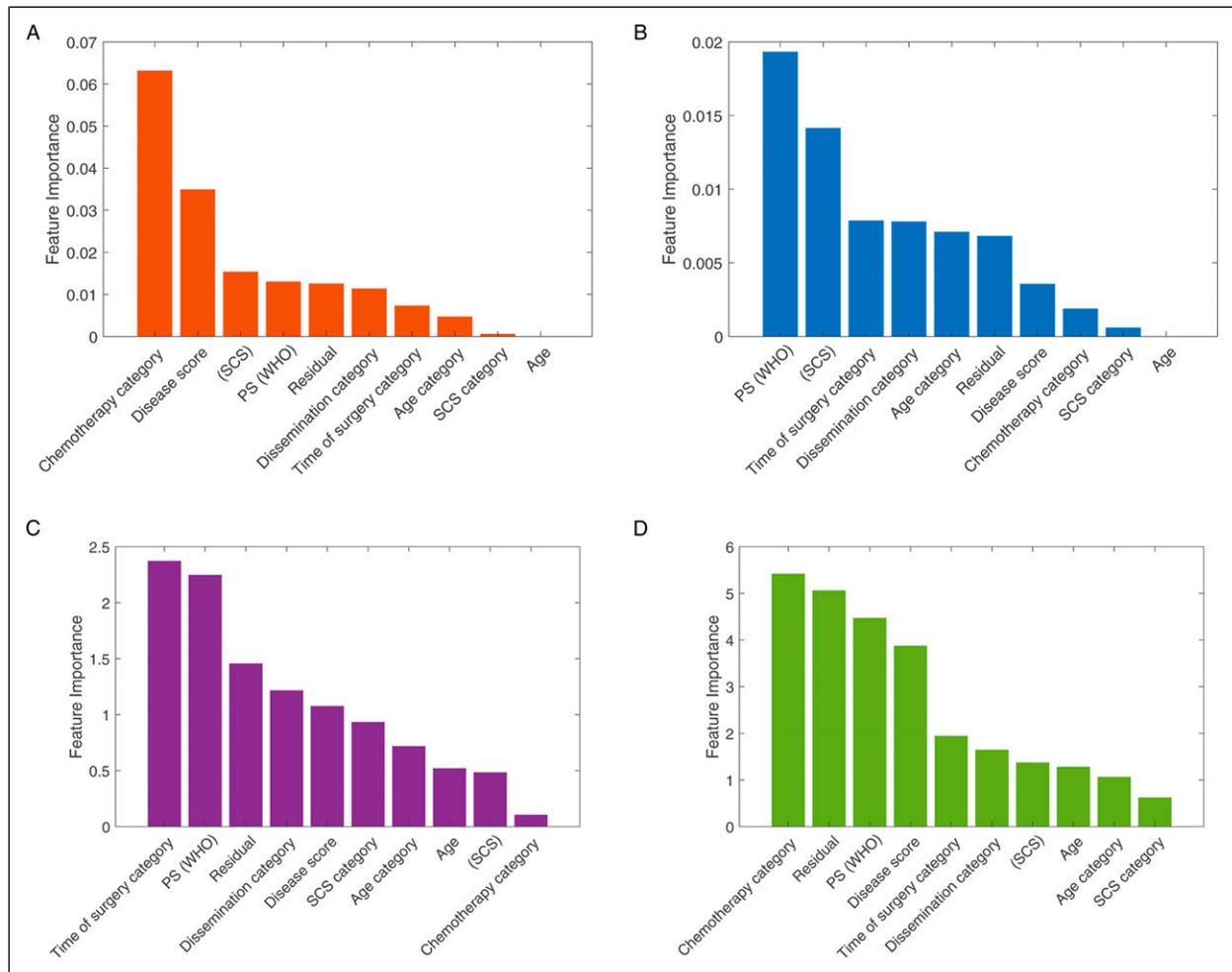
Figure 3. Feature ranking graphs for 2-year PFS: (A) Univariate feature ranking for classification using chi-square tests. (B) Multivariate feature ranking using MRMR algorithm; feature ranking graphs for 2-year OS: (C) Univariate feature ranking for classification using chi-square tests. (D) Multivariate feature ranking using MRMR algorithm.

indicator. The accuracy may not be often adequate for assessing model performance, as it tends to give advantage to models that always output the class with the highest frequency. Secondly, AUC is independent of cut-off point choices, and hence keeps the choice of clinical applications open beyond the analysis. Another explanation for the results comes from the inherent nature of the predictive parameters. Progression-free-survival is by nature heavily quantised, as time to relapse is potentially associated with the pre-scheduled screening. On the other hand, by definition, OS has a higher temporal resolution. For those cases where the data classes were unbalanced, the tested methods performed similarly to Logistic Regression.

The mean prediction accuracy figures indicate the potential in building eventually a combinational classifier that could potentially outperform conventional Logistic Regression, which is commonly used in the clinical setting. A maximum accuracy at 73% is satisfactory, but closer to 80% would have been preferable. The size of the dataset and the inherent characteristics of the categorical data are the main reasons for

these results. Another reason may be the high correlation amongst the variables that may render the model partly unstable due to collinearity (which further exists when the variables are increased. To address this, we examined the correlation amongst the variables and produced a correlation heatmap of the features included in the models. A rather weak correlation amongst features was demonstrated (Figure 5). Only in the 2 cases where we chose to include both categorical and the continuous variable, for example, age and age category, did we observe high correlation values. The low correlation indicates that we do not need to apply feature selection to alleviate features for their collinearity, but rather to identify the combination of features that can provide a reliable prognosis prediction.

We acknowledge the complexity of the predicting variables; some were not ready-made and converted into categorical classifiers. Starting with simple classifiers and then gradually proceeding with more complex classifiers, remains one of the ML principles, which could potentially affect the prediction accuracy of the model.[33] Nonetheless, the ML
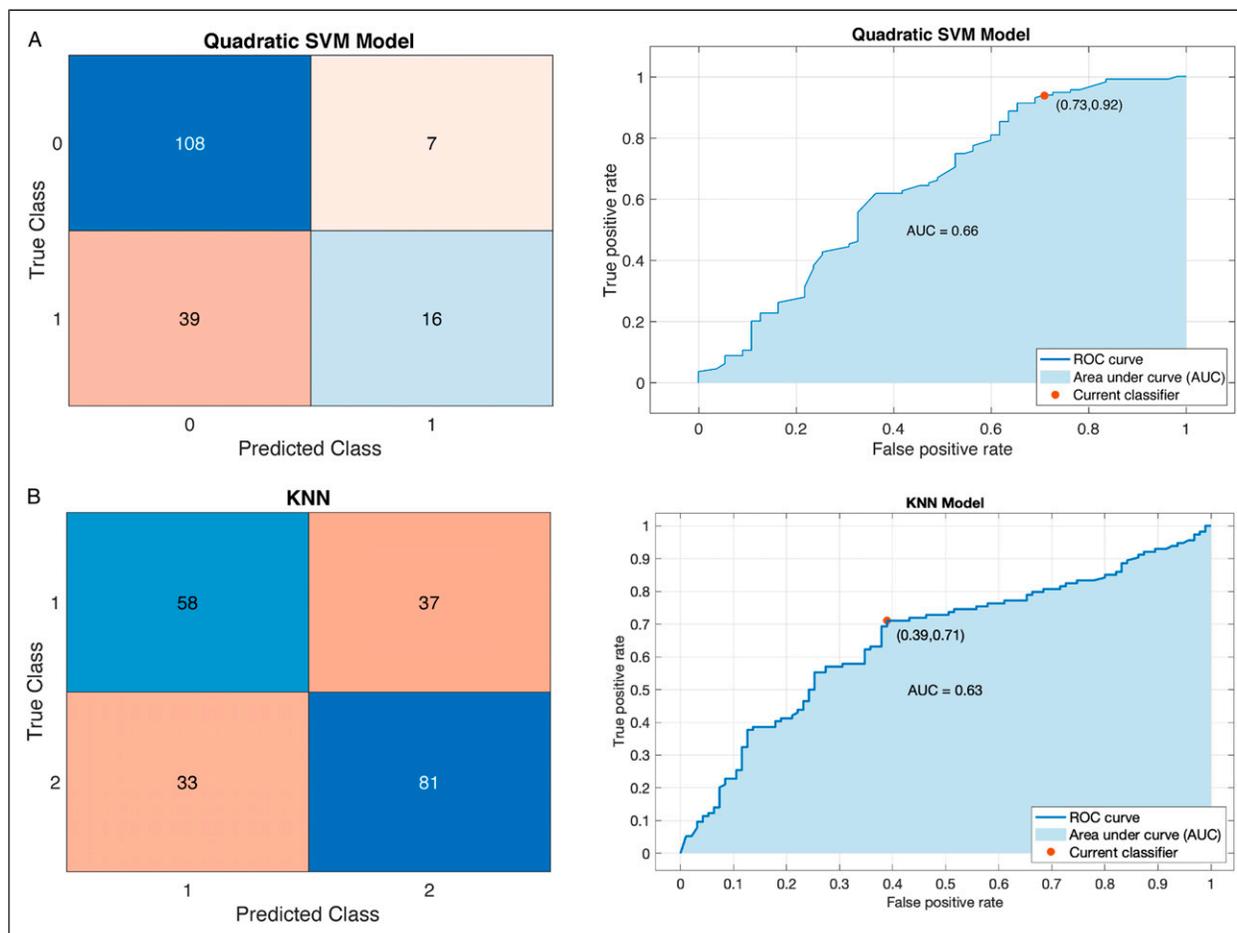
Figure 4. Example of a confusion matrix showing a) prediction accuracy for 2-year OS by use of (A) the SVM classifier with Quadratic Kernel (AUC: .66) (B) the k-NN (AUC: .63). The example shows that the prediction is more accurate for the negative class compared to the positive class.

approach is proving versatile. Both recall and precision, often inversely related, were greater than 80%. In this way, many potential clinical applications could be captured by this model, should this be used in a cancer diagnostic system, where sensitivity and positive predictive value are greatly appreciated.

Enshaei et al. compared a variety of algorithms and classifiers with conventional Logistic Regression statistical approaches to demonstrate the role of ML in providing prognostic and predictive data for ovarian cancer patients.[34] In a cohort of 668 patients, he demonstrated that an artificial neural network algorithm could predict OS with high accuracy (93%) and an AUC of .74, which outperformed Cox regression. Novel 'radiomic' descriptors of ovarian tumour phenotype and prognosis have been recently validated in a reliable and reproducible fashion.[35,36] The value of ML and conventional systems to provide critical diagnostic and prognostic prediction for patients with EOC before initial intervention based on blood biomarkers has been also demonstrated.[37] Cohort expansion to a larger sample size is expected to improve predictability.

In addition to performance comparison, we identified the features with the highest discriminant power (top-4) for the 2-year HGSOC prognosis prediction. Although the list for features was slightly different between chi-square test and the MRMR algorithm, some features were common for both methods. Equally, we compared our feature selection methods with regularization methods, such as Lasso,[38] and Elastic Net,[39] as shown in Figure 6. As expected, these methods resulted in a different ranking of the features, as they are usually applied on higher dimensional feature space. Nevertheless, the result confirmed a common subset of features including RD, ECOG PS and DS that appeared on the top-5 from all tested methods, thus confirming the validity of the employed feature selection methods (Figure 6).

The probability of achieving a cancer-free state (PFS) was maximised through a combination of primary surgery, good ECOG status, IDP and maximal surgical effort. In the era of precision medicine, the use of either NACT or PDS with no definite mechanisms to predict outcomes can lead to significant variations in practice. Previously, patient stratification was proposed according to patterns of tumour

Table 3. Predictive Accuracy of the ML Models and Comparisons with Conventional Logistic Regression for the 2-Year PFS and OS.

OS 2-years							
Model	Accuracy	AUC_P	AUC_N	Precision	Recall	F-score	G-score
SVM – Quadratic Kernel	72.9%	.66	.418	.7182	.9076	.8018	.8074
SVM – Cubic Kernel	68.2%	.58	.41	.7252	.8719	.7917	.7951
Logistic Regression	66.5%	.59	.413	.7209	.9169	.8071	.8130
Gaussian Naïve Bayes	66.0%	.63	.463	.6934	.9879	.8148	.8276
KNN – 5 neighbors	71.8%	.63	.443	.7009	.8656	.7742	.7787
KNN – 10 neighbors	69.4%	.62	.433	.7081	.8350	.7661	.7688
Ensemble – Bagged Trees	68.8%	.60	.432	.7086	.8425	.7695	.7725
Ensemble – Subspace Discriminant	71.8%	.61	.411	.7154	.9270	.8071	.8141
PFS 2-years							
Model	Accuracy		AUC	Precision	Recall	F-score	G-score
SVM – Quadratic Kernel	65.50%	.62	.469	.5160	.8893	.6530	.6774
SVM – Cubic Kernel	58.20%	.52	.485	.4309	.7286	.5415	.5603
Logistic Regression	56.50%	.58	.468	.5049	.8478	.6384	.6619
Gaussian Naïve Bayes	58.80%	.55	.49	.4356	.8373	.5731	.6039
KNN – 5 neighbors	57.60%	.54	.452	.4574	.5834	.5127	.5165
KNN – 10 neighbors	56.18%	.58	.446	.4643	.5947	.5214	.5254
Ensemble – Bagged Trees	55.30%	.52	.494	.4180	.7497	.5367	.5598
Ensemble – Subspace Discriminant	59.40%	.58	.475	.5112	.9096	.6546	.6819
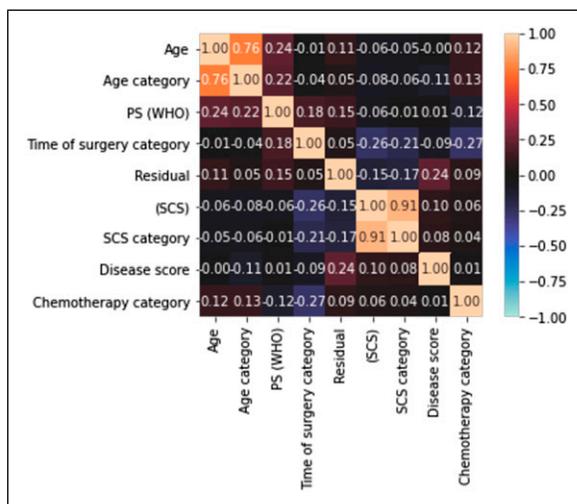
Figure 5. Correlation heatmap of the features included in the ML models demonstrating the correlation amongst the features using a variation of Pearson's R correlation coefficient. The colours in the heatmap represent the correlation coefficients. A weak correlation amongst features was demonstrated.

spread (reflecting the biologic behaviour of HGSOC), response to chemotherapy and prognosis to make a more rational decision between PDS and NACT-IDS.[40] Our data may provide the potential for more tailored approaches. The value of RD following PDS remains less diluted than following IDS and does carry the anticipated survival effect.[41] Both NACT and PDS have the same efficacy when used at their maximal possibilities, but their toxicity profile is different.[42] Nevertheless, most patients with advanced-stage HGSOC should benefit from primary surgery.

For the 2-year OS period, only PS retained its survival benefit, in addition to standard chemotherapy, status of complete cytoreduction and the tumor burden. Good performance status remains pivotal and, efforts to optimise baseline functional status and minimizing surgical complications may improve discharge rates and post-operative functional status.[42] The extent of disease at surgery (DS), in line with current literature, was more prognostic of OS than PFS. Indeed, the finding of bulky and diffuse disease spread may reflect high biological aggressiveness or long disease existence, allowing for advanced growth.[43] At a second glance, this is all interesting, as the factors predicting recurrence and death would not be separable, under the proportional hazard's assumption. We surmise that surgery and good medical health confer a transient survival benefit, but for overall prognosis, factors suggestive of the tumour biological behaviour including response to standard chemotherapy may be equally influential.

In our study, complete surgical cytoreduction remained an independent determinant of survival, potentially on the presumption of increased surgical effort.[44] Where surgery results in residual disease, the survival advantage from surgery is lost (Figure 6). Whilst we acknowledge that such results may be influenced by patient selection and chemotherapy exposure, they are comparable to international peers. In our cohort, the prolonged median overall survival of up to 38 months was comparable with that reported in the SCORPION trial[45] and substantially better than the 27 months from the individual patient meta-analysis of the EORTC and CHORUS trials.[46] Complete surgical resection, to 'reset the clock', may partly overcome the negative effect of tumour load, in line with a recent study.[47] Standard
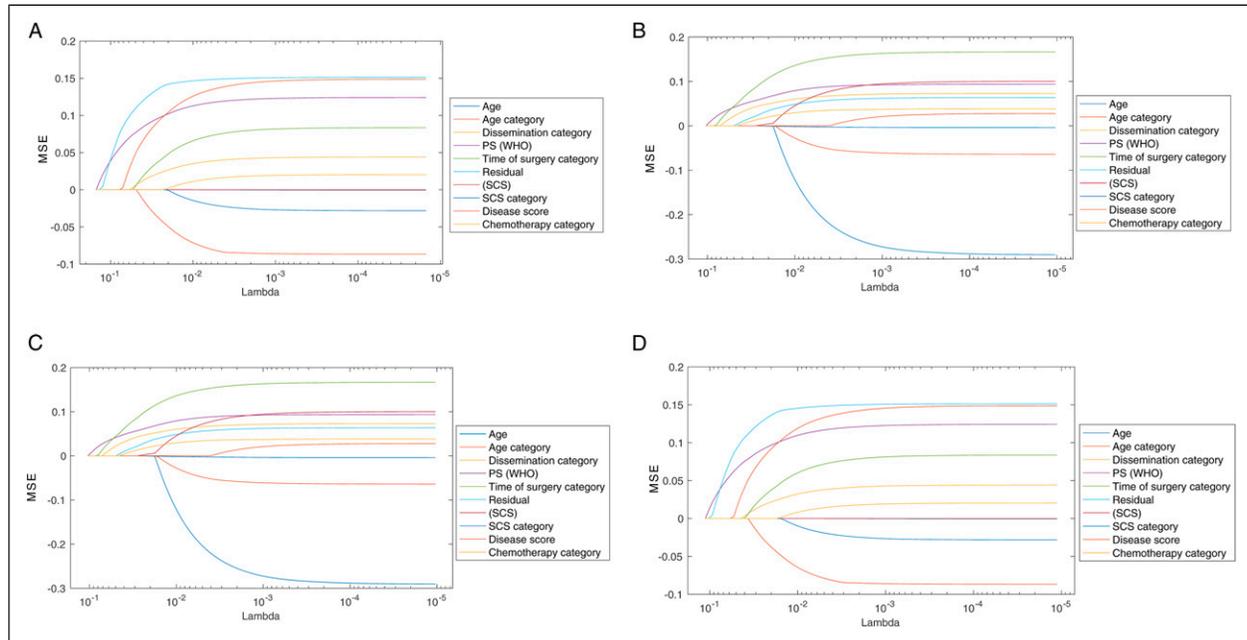
Figure 6. (A) Feature ranking for PFS based on the Lasso method. (B) Feature ranking for OS based on the Lasso method. (C) Feature ranking for PFS based on the Elastic Nets method. (D) Feature ranking for OS based on the Elastic Nets method.

chemotherapy does not reduce the eventual likelihood of death from ovarian cancer per se. Despite the generally accepted use of chemotherapy, delayed initiation of chemotherapy is associated with adverse clinical outcomes. It is advocated to start adjuvant chemotherapy within five to six weeks following debulking surgery.[48]

Strength of this study was the feature selection, aka the selection of the prediction variables, prior to building the classifiers. Except for our exhaustive search for the best combination of features, the literature is rich in various methodologies, including forward selection and recursive feature elimination.[49] In that sense, we focused solely on clinical pre-operative and intra-operative features, which was perhaps more practical and easier to obtain than molecular, genomic or radiomic features, thus the developed models are expected to have more clinical applicability. We did not address the value of surveillance modalities to detect recurrence during follow-up as we religiously follow the international guidelines. Another strength was the inclusion of initial disease distribution imaging data that proved more simplistic but useful than potential integration of 'radiomics' data. In our prognostic model, we included IDPs, which were pathologically verified, to demonstrate the anatomical extent of disease. Such preoperative imaging information is essential for prognostication and can be used to predict surgical resectability. Baseline IDP can be a prognostic factor, potentially addressing the aggressiveness of the disease and the difficulty to achieve complete cytoreduction (Figure 2C and D). Classification of such patterns can help counsel patients initially on their prognosis and identify those who might benefit from intraperitoneal chemotherapy to complement their treatment.[50]

This analysis comprised a homogenous fully curated cohort, which enabled a close collaboration with computer engineers toward prognosis improvements using multifactor analysis.[51] The stimulating debate whether ML-based algorithms are 'smarter' than human brains is largely irrelevant. The algorithms are reproducible because ML retains the strength of the structural model used for the prognosis prediction, even when applied in other populations and reveal different prediction features. Our effort represented a single institution experience, albeit we acknowledge the different practices worldwide, deriving from varying interpretations of evidence. Standardisation of surgical practice and identification of centres of excellence will potentially benefit patients from a maximal effort approach at all possible levels.[52]

Conclusions

We investigated the prediction of survival in advanced-stage HGSOC using clinical variables. We focused our analysis on the comparison of several classification models, including conventional regression analysis, under the same resampling conditions. Appropriate feature selection is required when building an HGSOC model for 2-year prognosis prediction by ML. For HGSOC prognosis, one should consider not only the patient's disease burden but also their overall medical status and ability to undergo extensive surgery, resulting in survival benefits alongside with standard chemotherapy.

Declaration of Conflicting Interests

Funding

Ethical Approval Statement

This study was approved by the Leeds Teaching Hospitals (LTHT) Institutional Review Board (MO20/133163/18.06.20) and informed written consent was obtained.

Abbreviations

The following abbreviations are used in this manuscript: AUC, Area under Curve; CT, Computer Tomography; DS, Disease Score; ECOG, Eastern Cooperative Oncology Group; EOC, Epithelial Ovarian Cancer; FIGO; Federation International of Obstetrics and Gynaecology; IDP, Intraperitoneal Dissemination Pattern; IDS, Interval Debulking Surgery; K-NN, K-Nearest Neighbor; ML, Machine Learning; MRMR, Minimum Redundancy Maximum Relevance; NACT, Neoadjuvant Chemotherapy; OS, Overall Survival; PFS, Progression Free Survival; PS, Performance Status; RD, Residual Disease; R0, No Residual-Complete Cytoreduction; SCS, Surgical Complexity Score; SD, Standard Deviation; SJUH, Saint James's University Hospital; SVM, Support-Vector-Machine.

ORCID iDs

Alexandros Laios ⓘ https://orcid.org/0000-0002-4870-7393
Diederick De Jong ⓘ https://orcid.org/0000-0003-0081-674X

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68:394-424.

2. Buechel M, Herzog TJ, Westin SN, Coleman RL, Monk BJ, Moore KN. Treatment of patients with recurrent epithelial ovarian cancer for whom platinum is still an option. *Ann Oncol*. 2019;30:721-732.

3. National Cancer Institute. SEER stat fact sheets: ovary cancer [online], ovary.html. 2015.

4. van der Burg MEL, van Lent M, Buyse M, et al. The effect of debulking surgery after induction chemotherapy on the prognosis in advanced epithelial ovarian cancer. *N Engl J Med*. 1995; 332:629-634.

5. Querleu D, Planchamp F, Chiva L, et al. European society of gynaecological oncology (ESGO) guidelines for ovarian cancer surgery. *Int J Gynecol Canc*. 2017;27:1534-1542.

6. Wright AA, Bohlke K, Armstrong DK, et al. Neoadjuvant chemotherapy for newly diagnosed, advanced ovarian cancer: society of gynecologic oncology and American society of clinical oncology clinical practice guideline. *J Clin Oncol*. 2016; 34:3460-3473.

7. Elattar A, Bryant A, Winter-Roach BA, Hatem M, Naik R. Optimal primary surgical treatment for advanced epithelial ovarian cancer. *Cochrane Database Syst Rev*. 2011;2011:Cd007565.

8. Kehoe S, Hook J, Nankivell M, et al. Primary chemotherapy versus primary surgery for newly diagnosed advanced ovarian cancer (CHORUS): an open-label, randomised, controlled, noninferiority trial. *Lancet*. 2015;386:249-257.

9. Chen C, He M, Zhu Y, Shi L, Wang X. Five critical elements to ensure the precision medicine. *Canc Metastasis Rev*. 2015;34: 313-318.

10. Zhou J, Zeng ZY, Li L. Progress of artificial intelligence in gynecological malignant tumors. *Canc Manag Res*. 2020;12: 12823-12840.

11. Laios A, Gryparis A, DeJong D, Hutson R, Theophilou G, Leach C. Predicting complete cytoreduction for advanced ovarian cancer patients using nearest-neighbor models. *J Ovarian Res*. 2020;13:117.

12. Al-Rajab M, Lu J, Xu Q. Examining applying high performance genetic data feature selection and classification algorithms for colon cancer diagnosis. *Comput Methods Progr Biomed*. 2017; 146:11-24.

13. Hanchuan Peng H, Fuhui Long F, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Anal Mach Int*. 2005;27:1226-1238.

14. Powers D. Evaluation: from precision, recall and F-Measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*. 2020. ArXiv: abs/2010:16061. doi:10.9735/2229-3981.

15. Halkidi M, Batistakis Y, Vazirgiannis M. On clustering validation techniques. *J Intell Inf Syst*. 2001;17:107-145.

16. Bachmann R, Brucker S, Stäbler A, et al. Prognostic relevance of high pretreatment CA125 levels in primary serous ovarian cancer. *Mol Clin Oncol*. 2021;14(1):8.

17. Håkansson F, Høgdall EVS, Nedergaard L, et al. Risk of malignancy index used as a diagnostic tool in a tertiary centre for patients with a pelvic mass. *Acta Obstet Gynecol Scand*. 2012; 91:496-502.

18. Rustin GJS. What surveillance plan should be advised for patients in remission after completion of first-line therapy for advanced ovarian cancer? *Int J Gynecol Canc*. 2010;20: S27-S28.

19. Aletti GD, Eisenhauer EL, Santillan A, et al. Identification of patient groups at highest risk from traditional approach to ovarian cancer treatment. *Gynecol Oncol*. 2011;120:23-28.

20. du Bois A, Reuss A, Pujade-Lauraine E, Harter P, Ray-Coquard I, Pfisterer J. Role of surgical outcome as prognostic factor in advanced epithelial ovarian cancer: a combined exploratory analysis of 3 prospectively randomized phase 3 multicenter trials. *Cancer*. 2009;115:1234-1244.

21. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Canc*. 2009;45:228-247. Oxford, England: European journal of cancer; 1990.

22. Torres D, Kumar A, Wallace SK, et al. Intraperitoneal disease dissemination patterns are associated with residual disease, extent of surgery, and molecular subtypes in advanced ovarian cancer. *Gynecol Oncol*. 2017;147:503-508.

23. Horowitz NS, Miller A, Rungruang B, et al. Does aggressive surgery improve outcomes? Interaction between preoperative disease burden and complex surgery in patients with advanced-stage ovarian cancer: an analysis of GOG 182. *J Clin Oncol.* 2015;33:937-943.

24. Manning CD, Raghavan P, Schütze H. *Introduction to Information Retrieval.* Cambridge: Cambridge University Press; 2008.

25. Lee Y. Support vector machines for classification: a statistical portrait. *Methods Mol Biol.* 2010;620:347-368.

26. Malley JD, Kruppa J, Dasgupta A, Malley KG, Ziegler A. Probability machines: consistent probability estimation using nonparametric learning machines. *Methods Inf Med.* 2012;51:74-81.

27. Haque MN, Noman N, Berretta R, Moscato P. Heterogeneous ensemble combination search using genetic algorithm for class imbalanced data classification. *PLoS One.* 2016;11:e0146116.

28. Kim W, Kim KS, Park RW. Nomogram of Naive Bayesian model for recurrence prediction of breast cancer. *Healthcare Inf Res.* 2016;22:89-94.

29. Harrell FE. *Regression Modelling Strategies: with applications to Linear Models, Logistic Regression and Survival Analysis.* New York: Springer; 2010.

30. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One.* 2015;10:e0118432.

31. Huang S, Yang J, Fong S, Zhao Q. Artificial intelligence in cancer diagnosis and prognosis: opportunities and challenges. *Canc Lett.* 2020;471:61-71.

32. Dy JG. Unsupervised feature selection. In: Motoda H, Liu H, eds. *Computational Methods of Feature Selection.* New York: Chapman and Hall/CRC; 2007:35-56.

33. Guo H, Li Y, Mensah GK, et al. Resting-state functional network scale effects and statistical significance-based feature selection in machine learning classification. *Comput Math Methods Med.* 2019;2019:9108108.

34. Enshaei A, Robson CN, Edmondson RJ. Artificial intelligence systems as prognostic and predictive tools in ovarian cancer. *Ann Surg Oncol.* 2015;22:3970-3975.

35. Lu H, Arshad M, Thornton A, et al. A mathematical-descriptor of tumor-mesoscopic-structure from computed-tomography images annotates prognostic- and molecular-phenotypes of epithelial ovarian cancer. *Nat Commun.* 2019;10:764.

36. Gerestein C, Eijkemans M, de Jong D, et al. The prediction of progression-free and overall survival in women with an advanced stage of epithelial ovarian carcinoma. *BJOG.* 2009;116:372-380.

37. Kawakami E, Tabata J, Yanaihara N, et al. Application of artificial intelligence for preoperative diagnostic and prognostic prediction in epithelial ovarian cancer based on blood biomarkers. *Clin Canc Res.* 2019;25:3006-3015.

38. Zhao P, Yu B. On model selection consistency of Lasso. *J Mach Learn Res.* 2006;7:2541-2563.

39. De Mol C, De Vito E, Rosasco L. Elastic-net regularization in learning theory. *J Complex.* 2009;25:201-230.

40. Makar AP, Tropé CG, Tummers P, Denys H, Vandecasteele K. Advanced ovarian cancer: primary or interval debulking? five categories of patients in view of the results of randomized trials and tumor biology: primary debulking surgery and interval debulking surgery for advanced ovarian cancer. *Oncol.* 2016;21:745-754.

41. Fotopoulou C, Sehouli J, Aletti G, et al. Value of neoadjuvant chemotherapy for newly diagnosed advanced ovarian cancer: a European perspective. *J Clin Oncol.* 2017;35:587-590.

42. Roy AG, Brensinger CM, Latif N, et al. Assessment of poor functional status and post-acute care needs following primary ovarian cancer debulking surgery. *Int J Gynecol Canc.* 2020;30:227-232.

43. Zivanovic O, Sima CS, Iasonos A, et al. The effect of primary cytoreduction on outcomes of patients with FIGO stage IIIC ovarian cancer stratified by the initial tumor burden in the upper abdomen cephalad to the greater omentum. *Gynecol Oncol.* 2010;116:351-357.

44. Eisenkop SM, Spirtos NM, Friedman RL, Lin WCM, Pisani AL, Perticucci S. Relative influences of tumor volume before surgery and the cytoreductive outcome on survival for patients with advanced ovarian cancer: a prospective study. *Gynecol Oncol.* 2003;90:390-396.

45. Fagotti A, Ferrandina MG, Vizzielli G, et al. Randomized trial of primary debulking surgery versus neoadjuvant chemotherapy for advanced epithelial ovarian cancer (SCORPION-NCT01461850). *Int J Gynecol Canc.* 2020;30:1657-1664.

46. Vergote I, Coens C, Nankivell M, et al. Neoadjuvant chemotherapy versus debulking surgery in advanced tubo-ovarian cancers: pooled analysis of individual patient data from the EORTC 55971 and CHORUS trials. *Lancet Oncol.* 2018;19:1680-1687.

47. Angeles MA, Rychlik A, Cabarrou B, et al. A multivariate analysis of the prognostic impact of tumor burden, surgical timing and complexity after complete cytoreduction for advanced ovarian cancer. *Gynecol Oncol.* 2020;158:614-621.

48. Timmermans M, van der Aa MA, Lalisang RI, et al. Interval between debulking surgery and adjuvant chemotherapy is associated with overall survival in patients with advanced ovarian cancer. *Gynecol Oncol.* 2018;150:446-450.

49. Efroymson MA. Multiple regression analysis. In: Ralston A, Wilf HS, eds. *Mathematical Methods for Digital Computers.* New York: John Wiley; 1960:191-203.

50. Tanner EJ, Black DR, Zivanovic O, et al. Patterns of first recurrence following adjuvant intraperitoneal chemotherapy for stage IIIC ovarian cancer. *Gynecol Oncol.* 2012;124:59-62.

51. Chen JH, Asch SM. Machine learning and prediction in medicine - beyond the peak of inflated expectations. *N Engl J Med.* 2017;376:2507-2509.

52. Fotopoulou C, Concin N, Planchamp F, et al. Quality indicators for advanced ovarian cancer surgery from the European Society of Gynaecological Oncology (ESGO): 2020 update. *Int J Gynecol Canc.* 2020;30:436-440.