Radiotherapy and Oncology 132 (2019) 171–177

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Original Article

Deep learning provides a new computed tomography-based prognostic biomarker for recurrence prediction in high-grade serous ovarian cancer

Shuo Wang [a,f,1], Zhenyu Liu [a,f,1], Yu Rong [e,1], Bin Zhou [d], Yan Bai [c], Wei Wei [c], Wei Wei [a], Meiyun Wang [c,*], Yingkun Guo [b,*], Jie Tian [a,f,g,*]

[a] CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing; [b] Department of Radiology, Key Laboratory of Birth Defects and Related Diseases of Women and Children of Ministry of Education, West China Second University Hospital, Sichuan University; [c] Department of Radiology, Henan Provincial People's Hospital; [d] Key Laboratory of Birth Defects and Related Diseases of Women and Children of Ministry of Education, West China Second University Hospital, Sichuan University; [e] Department of Radiology, Key Laboratory of Intelligent Medical Image Analysis and Precision Diagnosis in Guizhou Province, Guizhou Provincial People's Hospital; [f] University of Chinese Academy of Sciences, Beijing, China and; [g] Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, Beihang University, Beijing, 100191, china

ARTICLE INFO

Article history:
Received 19 June 2018
Received in revised form 18 October 2018
Accepted 21 October 2018
Available online 1 November 2018

Keywords:
Deep learning
High-grade serous ovarian cancer
Recurrence
Prognosis
Computed tomography
Artificial intelligence
Semi-supervised learning
Auto encoder
Unsupervised learning

ABSTRACT

Background and purpose: Recurrence is the main risk for high-grade serous ovarian cancer (HGSOC) and few prognostic biomarkers were reported. In this study, we proposed a novel deep learning (DL) method to extract prognostic biomarkers from preoperative computed tomography (CT) images, aiming at providing a non-invasive recurrence prediction model in HGSOC.

Materials and methods: We enrolled 245 patients with HGSOC from two hospitals, which included a feature-learning cohort ($n = 102$), a primary cohort ($n = 49$) and two independent validation cohorts from two hospitals ($n = 49$ and $n = 45$). We trained a novel DL network in 8917 CT images from the feature-learning cohort to extract the prognostic biomarkers (DL feature) of HGSOC. Afterward, a DL-CPH model incorporating the DL feature and Cox proportional hazard (Cox-PH) regression was developed to predict the individual recurrence risk and 3-year recurrence probability of patients.

Results: In the two validation cohorts, the concordance-index of the DL-CPH model was 0.713 and 0.694. Kaplan–Meier's analysis clearly identified two patient groups with high and low recurrence risk ($p = 0.0038$ and 0.0164). The 3-year recurrence prediction was also effective (AUC = 0.772 and 0.825), which was validated by the good calibration and decision curve analysis. Moreover, the DL feature demonstrated stronger prognostic value than clinical characteristics.

Conclusions: The DL method extracts effective CT-based prognostic biomarkers for HGSOC, and provides a non-invasive and preoperative model for individualized recurrence prediction in HGSOC. In addition, the DL-CPH model provides a new prognostic analysis method that can utilize CT data without follow-up for prognostic biomarker extraction.

© 2018 The Authors. Published by Elsevier B.V. Radiotherapy and Oncology 132 (2019) 171–177 This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Ovarian cancer (OC) is the leading cause of gynecologic cancer deaths and high-grade serous ovarian cancer (HGSOC) is the most common and most lethal histological type [1,2]. The lethality of HGSOC mainly comes from its high risk of recurrence [3,4]. Consequently, preoperative identification of recurrence in patients with HGSOC is important since it guides the personalized treatment and surveillance planning, such as selecting the agent of chemotherapy [5]. To this end, prognostic biomarkers related to recurrence of HGSOC are needed. Current studies indicated that clinical characteristics such as stage of International Federation of Gynecology and Obstetrics (FIGO), and preoperative serum cancer antigen (CA-125) were associated with recurrence of HGSOC [3,6,7]. However, the clinical biomarkers are invasive and provide only limited information about tumor due to the spatial and temporal pathologic heterogeneity of tumor [8,9].

Benefiting from the ability to noninvasively visualize a cancer's appearance on a macroscopic level, medical imaging demonstrated strong prognostic value [10–12]. Computed tomography (CT), as a

Abbreviations: AUC, area under the receiver operating characteristic curve; C-Index, Harrell's concordance-index; DL, deep learning; HGSOC, high-grade serous ovarian cancer; OC, ovarian cancer; CI, confidence interval.

* Corresponding authors at: Department of Radiology, Henan Provincial People's Hospital, China (M. Wang); Department of Radiology, Key Laboratory of Birth Defects and Related Diseases of Women and Children of Ministry of Education, West China Second University Hospital, Sichuan University, China (Y. Guo); CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing, China (J.Tian).

E-mail addresses: marian9999@163.com (M. Wang), gykpanda@163.com (Y. Guo), jie.tian@ia.ac.cn (J. Tian).

[1] Equal contribution.

https://doi.org/10.1016/j.radonc.2018.10.019
0167-8140/© 2018 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

routinely used medical imaging modality, contains many mineable features associated with prognosis of cancer [13–15] including HGSOC [16]. The existing hand-crafted CT image features can quantify tumor shape and texture information by manually defined feature engineering [8]. However, the intrinsic characteristics of tumor that are difficult to be manually defined require further study.

Deep learning (DL) [17] as an artificial intelligence method has shown promising results in detecting valuable features from medical imaging [18–20]. The computational units in DL are defined as layers that are integrated together to mine the intrinsic characteristics of images [21]. Through a hierarchical neural network structure and convolutional operation, DL extracts the intrinsic characteristics of tumor that showed promising prognostic value [22]. Further study demonstrated that DL feature showed superior performance in comparison to hand-crafted image feature [23].

In this study, we explore an unsupervised DL method to extract the prognostic biomarkers of HGSOC from preoperative CT images, aiming at providing a non-invasive individualized recurrence prediction model in HGSOC. We hypothesize that the features learned by the unsupervised DL method can describe the tumor state thoroughly and reflect the intrinsic characteristics of tumor. Therefore, the DL feature contains much prognostic information on HGSOC. Moreover, the DL method can extract prognostic biomarkers of tumor requiring only tumor image data, which provides a new prognostic analysis method that enables us to utilize a large amount of data without follow-up.

Materials and methods

Patients

The institutional review board approval was granted for this retrospective study, and the requirement for informed consent was waived. We studied 245 patients who were pathologically confirmed to have primary HGSOC from the West China Second University Hospital of Sichuan University (WCSUH-SU, $n = 200$, between February 2010 and September 2015) and Henan Provincial People's Hospital (HPPH, $n = 45$, between May 2012 and October 2016). The contrast-enhanced CT scanning was acquired for all the patients at diagnosis time, and the venous phase image was used for this study (CT scanning parameters in Supplementary Methods 1). Supplementary Methods 2 and Supplementary Fig. S1 provided detailed inclusion and exclusion criteria along with the recruitment pathway.

All the patients underwent primary debulking surgery [3,24], and were followed up every 2–4 months for the first two years, every 3–6 months from the third year, and annually from the 5th year [25]. The date of last follow-up was August 23, 2017 for the WCSUH-SU dataset and March 23, 2018 for the HPPH dataset. The endpoint of this study was recurrence, which was diagnosed combing clinical symptoms, rising CA-125 levels, and radiological findings. The time between clinical remission and first recurrence is defined as recurrence-free survival (RFS) [6,7]. The preoperative clinical characteristics were collected from the Institutional Picture Archiving and Communication System (PACS), including age, preoperative CA-125, FIGO stage, tumor location and maximum tumor diameter (Table 1).

In the WCSUH-SU dataset, patients without follow-up or follow-up time less than 3 years were included into the feature-learning cohort; while patients with 3+ year's complete follow-up information were allocated into a primary cohort and a validation cohort according to the diagnosis time. The first 49 patients who underwent preoperative CT examination between 2010 and 2013 comprised the primary cohort, while the other 49 patients who underwent preoperative CT examination between 2014 and 2015 comprised the validation cohort 1. In the HPPH dataset, all the 45 consecutive patients were used for independent external validation (validation cohort 2). The median follow-up time was 32.83 months for the primary cohort, 31.07 months for the validation cohort 1, and 17.83 months for the validation cohort 2.

Development of the deep learning network

We proposed a novel DL network to extract the intrinsic characteristics of HGSOC from preoperative CT images (Feature learning in Fig. 1). The proposed DL network requires only the CT scanning of HGSOC without follow-up information, which is defined as unsupervised learning.

The main computational processes of DL are convolution, pooling, activation and batch normalization, which are presented in the Supplementary Methods 3. Specifically, the DL network used a convolutional auto-encoder structure [26] that includes an encoder network and a decoder network. The encoder network includes an initial convolutional layer with 24 filters and four subsequent convolutional layers with 16 filters. To accelerate training and avoid covariate shift in the network, a batch normalization layer is inserted between two adjacent convolutional layers. In addition, we used average pooling between convolutional layers to eliminate redundant features. Finally, the encoder network transforms a tumor image into a 16-dimensional mineable feature vector. We refer to this vector as DL feature. Meanwhile, the decoder network uses the DL feature to reconstruct the original tumor image, aiming at evaluating the information capacity included in the DL feature. The decoder network is similar to the encoder network except that it upsamples image by deconvolutional layer. If the DL feature extracts the intrinsic characteristics of tumor, the decoder network should be able to reconstruct the original tumor image from the DL feature. To achieve this goal, we trained the DL network in 102 patients from the feature-learning cohort. A radiologist (5+ years' experience, Y. Rong) located tumor areas in all the CT slices from the 102 patients using a rectangle bounding box (region of interest, ROI), resulting in 8917 tumor images to train the DL network. The network training is an iterative process, which optimizes the network iteratively until it extracts the intrinsic characteristics of HGSOC (details in Supplementary Methods 4).

Deep learning feature extraction

When the DL network is well trained in the feature-learning cohort, we applied it to transform CT image slices of HGSOC into 16-dimensional DL feature. First, we selected the ROI of tumor for each patient according to the following rule: the ROI should cover the primary tumor area in ovaries. If multiple tumor areas are observed in ovaries, multiple ROIs will be selected (Y. Rong and Y. Bai selected the ROIs in the WCSUH-SU dataset and the HPPH dataset respectively, Supplementary Fig. S2 illustrated the ROIs selected by the radiologists). Afterward, the tumor image was standardized by z-score normalization and scaled to 64×64 voxel size, and fed into the DL network. The output of the last convolutional layer in the encoder network was extracted as DL feature [22], which was 16-dimensional. Since the tumor image included multiple 2-dimensional slices, we averaged features from all image slices to acquire the DL feature for the patient.

Recurrence analysis

To evaluate the prognostic value of the DL feature, we used a multivariate Cox-PH regression to build the association between the DL feature and recurrence of HGSOC (recurrence analysis in Fig. 1). We trained the Cox-PH model using data (DL feature, recurrence time and status) from the primary cohort, and then validated

Table 1
Clinical characteristics of patients from the feature-learning, primary, and validation cohorts.

Characteristics	Feature-learning cohort (n = 102)	Primary cohort (n = 49)	Validation cohort 1 (n = 49)	Validation cohort 2 (n = 45)
Age (years)[a]	52.85 ± 10.22	52.92 ± 9.25	50.98 ± 8.33	59.93 ± 10.47
CA-125 (U/ml)[b]	1536.55 (39.3–9385.2)	1648.12 (23.9–6596.5)	2166.05 (102.0–8000.0)	1560.50 (237–7210)
Tumor diameter (mm)[c]	88.54 ± 42.60	87.365 ± 34.10	85.79 ± 35.67	77.56 ± 30.13
Tumor location (%)				
Left	31 (30.39%)	11 (22.45%)	7 (14.28%)	10 (22.22%)
Right	21 (20.59)	12 (24.49%)	9 (18.37%)	16 (35.56%)
Bilateral	50 (49.02%)	26 (53.06%)	33 (67.35%)	19 (42.22%)
FIGO stage (%)				
I	6 (5.88%)	2 (4.08%)	1 (2.04%)	0 (0.00%)
II	11 (10.78%)	3 (6.12%)	2 (4.08%)	1 (2.22%)
III	72 (70.59%)	35 (71.43%)	41 (83.67%)	30 (66.67%)
IV	13 (12.75%)	9(18.37%)	5 (10.21%)	14 (31.11%)

Note: [a]Data are the mean ± std. [b]Data are the mean value with the range in the parentheses. [c]Tumor diameter is the largest diameter of the tumor in axial view CT image. The clinical characteristics of validation cohort 1 and feature-learning cohort do not have significant difference with the primary cohort (p = 0.179–0.824 between the validation cohort 1 and primary cohort, p = 0.580–0.970 between the feature-learning cohort and primary cohort). No significant difference was found between the validation cohort 2 and the primary cohort in terms of CA-125, tumor diameter, tumor location and FIGO stage (p = 0.145–0.875).

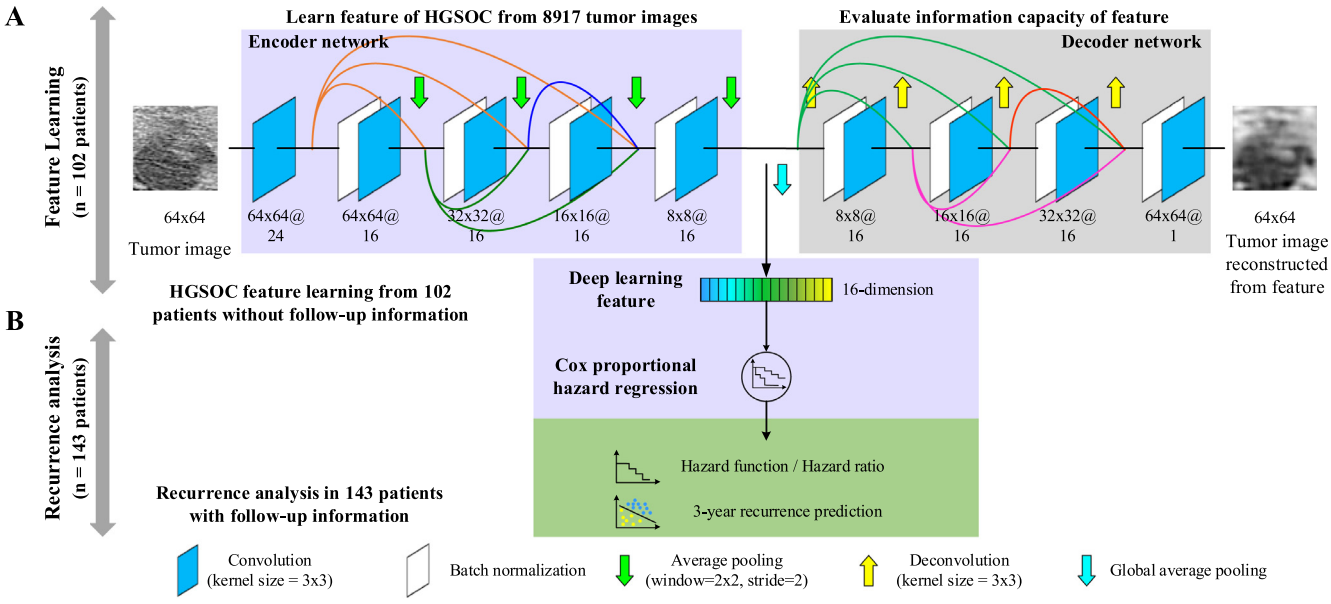

Fig. 1. The framework of the proposed DL-CPH model for recurrence prediction. This framework includes two parts: (a) HGSOC feature learning and (b) recurrence analysis. The feature learning part is a convolutional auto encoder structure that encodes ovarian cancer into DL feature (16-dimensional). The numbers under the layers of the DL network represent the *feature map width* × *feature map height @ feature map channel*. Inside the encoder and the decoder network, skip connections are used, which are similar to the DenseNet architecture [33]. The recurrence analysis part includes a multivariate Cox proportional hazard regression that uses the DL feature to predict recurrence.

its performance in the two independent validation cohorts. The Cox-PH model using the DL feature was defined as the DL-CPH model in this study.

For each patient, the Cox-PH model predicts a hazard score indicating the individual recurrence risk. This hazard score was used for recurrence-free survival prediction and to stratify patients into high- and low-risk groups concerning recurrence. In addition, The Cox-PH can predict the recurrence probability of patient in a specific time point. In this study, we used the Cox-PH model to predict the 3-year recurrence probability.

Comparing deep learning feature with clinical characteristics

Since clinical characteristics were used as prognostic biomarkers in HGSOC [6,7], we compared the prognostic value of the DL feature to the clinical information. We built a clinical model involving age, FIGO stage, preoperative CA-125, tumor location and tumor diameter as features, and Cox-PH regression for recurrence prediction.

Since clinical characteristics and CT imaging reflect HGSOC from different perspectives, we therefore explore the combination of these two information. Among the five clinical characteristics, we used backward step-wise selection with the likelihood ratio test to select clinical predictors, which employed Akaike information criterion as the stopping rule [27]. Afterward, we built a Cox-PH model combining the DL-predicted hazard score and clinical predictor, and defined it as combined model.

Statistical analysis

We used Harrell's concordance-index (C-Index) to measure the concordance between the DL-predicted recurrence risk and the actual recurrence time. A C-index score around 0.70 indicates a good model, whereas a score around 0.50 means random results without predictive performance [22]. When assessing the 3-year recurrence prediction, we used area under the receiver operating characteristic curve (AUC) and accuracy to evaluate the discriminatory performance of the DL-CPH model. Moreover, calibration

curves accompanied by the Hosmer–Lemeshow test were plotted to assess the DL-CPH model, where a non-significant statistic implied that the DL-CPH model was perfectly calibrated and close to the perfect model [28].

To assess whether the DL-CPH model would improve patient outcomes, we used decision curve analysis to examine clinical consequences based on threshold probability, from which the net benefit could be derived [15,29].

When assessing the clinical characteristics between the primary and validation cohorts, the independent samples *t* test was adopted to evaluate the significance of the mean value on age and preoperative CA-125. The chi-squared test was used to assess the difference of categorical variables. In all statistical tests, *p*-values smaller than 0.05 were considered significant. All the statistical analyses were conducted with *R* software (version 3.0.1). The Cox-PH model was implemented by *lifelines* package in Python 2.7. The DL network was implemented by *Keras 2.1.5*. We made the DL network and the DL-CPH model of this study available at http://www.radiomics.net.cn/post/111.

Results

In the primary cohort, the DL-CPH model achieved good performance on RFS prediction (Table 2, C-Index = 0.717, [95% confidence interval (CI): 0.683–0.755], Hazard Ratio (HR) = 2.711, [95% CI: 2.503–2.919]). In the two independent validation cohorts, the predictive performance of the DL-CPH model was further confirmed (C-Index = 0.713 [95% CI: 0.681–0.750] in the validation cohort 1; C-Index = 0.694, [95% CI: 0.658–0.730] in the validation cohort 2). Importantly, the C-Index of the DL-CPH model was higher than the clinical model (C-Index = 0.448, [95%CI: 0.402–0.492] in the validation cohort 1; C-Index = 0.631, [95%CI: 0.588–0.674] in the validation cohort 2).

In addition, the strong association between the DL-predicted hazard score and the RFS was further demonstrated by the Kaplan–Meier analysis in Fig. 2. We used the median hazard score of the primary cohort as cut-off value to split patients into high- and low-risk groups [10,13]. Significant discrimination between the RFS of the two groups was observed in the three cohorts ($p < 0.0001$ in the primary cohort; $p = 0.0038$ in the validation cohort 1; $p = 0.0164$ in the validation cohort 2, log-rank test).

To further characterize the association between the DL feature and recurrence, we used the DL-CPH model to predict 3-year recurrence probability for patients. Table 2 and Fig. 3a indicated that the DL-CPH model achieved an AUC of 0.833 (95% CI: 0.792–0.874) in the primary cohort and AUC = 0.772 (95% CI: 0.721–0.820) in the validation cohort 1. The DL-predicted probability also showed significant difference between patients who relapsed less than three years and longer than three years ($p < 0.0001$ in the primary cohort; $p = 0.0010$ in the validation cohort 1, Fig. 3b). Good calibration in Supplementary Fig. S3a indicated that the DL-CPH model did not systematically under-predict or over-predict the 3-year recurrence probability because the Hosmer–Lemeshow test yielded a non-significant statistic to the perfect model ($p = 0.475$ and $p = 0.404$ in the primary cohort and validation cohort 1). The decision curve in Supplementary Fig. S3b showed that if the threshold probability of a patient or doctor is bigger than 30%, using the DL-CPH model to predict 3-year recurrence added more

Table 2
Model performance on predicting RFS and 3-year recurrence.

Models	Cohorts	C-Index (95% CI)	AUC (95% CI)	ACC (95% CI)
Clinical Model	Primary	0.680 (0.642, 0.717)	0.774 (0.727, 0.826)	0.735 (0.689, 0.784)
	Validation 1	0.448 (0.402, 0.492)	0.443 (0.381, 0.506)	0.449 (0.396, 0.503)
	Validation 2	0.631 (0.588, 0.674)	0.400 (0.268, 0.536)	0.541 (0.480, 0.598)
DL-CPH Model	Primary	0.717 (0.683, 0.755)	0.833 (0.792, 0.874)	0.776 (0.733, 0.820)
	Validation 1	0.713 (0.681, 0.750)	0.772 (0.721, 0.820)	0.714 (0.665, 0.760)
	Validation 2	0.694 (0.658, 0.730)	0.825 (0.765, 0.893)	0.730 (0.678, 0.786)
Combined Model	Primary	0.738 (0.705, 0.773)	0.865 (0.830, 0.903)	0.796 (0.757, 0.840)
	Validation 1	0.700 (0.659, 0.742)	0.760 (0.709, 0.809)	0.694 (0.647, 0.741)
	Validation 2	0.729 (0.696, 0.761)	0.762 (0.693, 0.831)	0.703 (0.646, 0.759)

Note: CI represents confidence interval. C-Index represents Harrell's concordance index, which measures the performance of the RFS prediction. AUC represents area under the receiver operating characteristic curve, and ACC is accuracy. AUC and ACC evaluate the performance of the 3-year recurrence prediction.

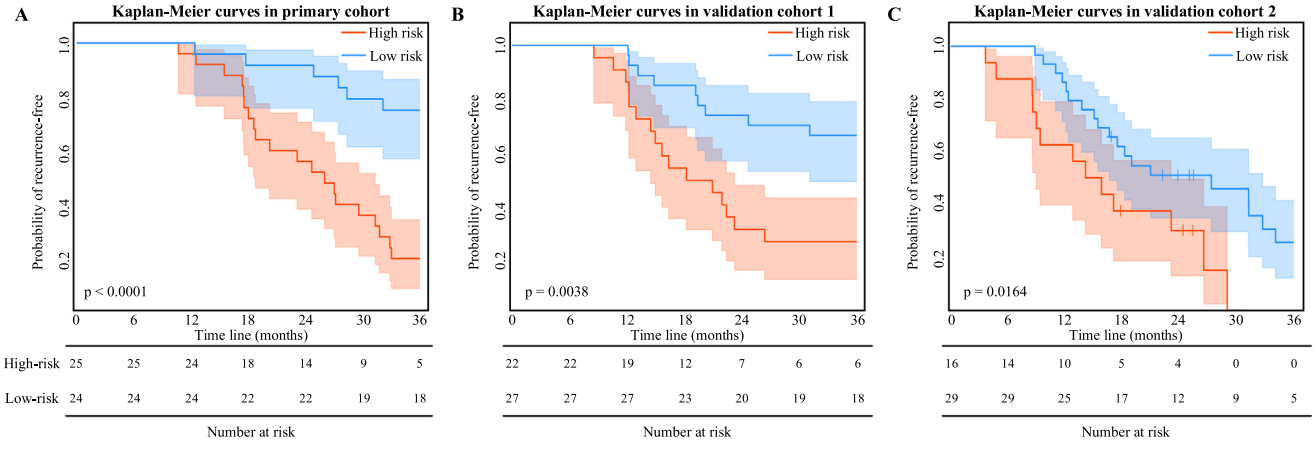

Fig. 2. Kaplan–Meier's analysis of the DL-CPH model. (a) Kaplan–Meier's analysis in patients from the primary cohort. The vertical lines indicate censored data, and the shadow indicates the 95% confidence interval. (b) Kaplan–Meier's analysis in patients from the validation cohort 1. (c) Kaplan–Meier's analysis in patients from the validation cohort 2.

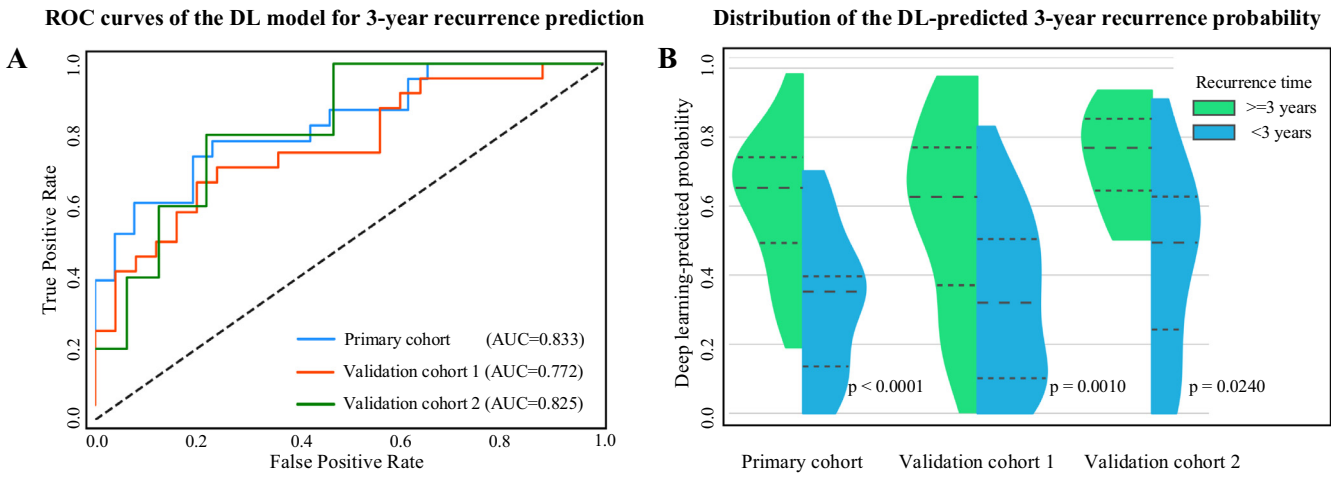

Fig. 3. The 3-year recurrence prediction of the DL-CPH model. (a) ROC curves of the DL-CPH model for 3-year recurrence prediction. (b) Distribution of the DL-predicted 3-year recurrence probability in the three cohorts.

benefit than the treat-all or the treat-none schemes. In the independent validation cohort 2, similar performance was observed (AUC = 0.825, 95% CI: 0.765–0.893). The DL-predicted probability illustrated significant difference between patients who relapsed less than three years and longer than three years ($p = 0.0240$).

In the two independent validation cohorts, the DL-CPH model yielded a higher AUC than the clinical model (AUC = 0.443, [95% CI: 0.381–0.506] in the validation cohort 1; AUC = 0.400, [95% CI: 0.268–0.536] in the validation cohort 2) with significant difference ($p = 0.0045$ and 0.0361 in the validation cohort 1 and validation cohort 2, DeLong's test).

Among the clinical characteristics, FIGO stage was identified as an independent predictor for recurrence. Consequently, we constructed a combined model integrating both clinical predictor (FIGO stage) and the DL-predicted hazard score. The combined model showed an improvement over the clinical model or DL-CPH model alone (C-Index = 0.738, 95% CI: 0.705–0.773) in the primary cohort in terms of the C-Index. This result was further confirmed in the validation cohort 2 (C-Index = 0.729, 95% CI: 0.696–0.761).

Since most OC were diagnosed with advanced stage (FIGO stage III, IV), we performed a stratified analysis to evaluate the performance of the DL-CPH model in advanced HGSOC. Similar results were observed in Supplementary Table S1, Supplementary Fig. S4 and Supplementary Fig. S5. In the primary cohort, the DL-CPH model achieved C-Index of 0.706 (95% CI: 0.669–0.746). This performance was confirmed in the validation cohort 1 (AUC = 0.712, 95% CI: 0.675–0.749) and validation cohort 2 (AUC = 0.687, 95% CI: 0.649–0.723). When predicting the 3-year recurrence probability, the DL-CPH model yielded an AUC of 0.857 (95% CI: 0.815–0.897) in the primary cohort; and an AUC of 0.763 (95% CI: 0.713–0.818) in the validation cohort 1; and an AUC of 0.825 (95% CI: 0.712–0.814) in the validation cohort 2. In addition, good calibration curves were also observed.

Discussion

The DL network includes thousands of neuron paths to extract the intrinsic characteristics of HGSOC. This self-learning structure enables us to quantify the prognostic features of HGSOC that are difficult to be manually defined. Through a stacked neural network structure, the DL network manages to encode tumor into multi-level features reflecting various characteristics of HGSOC. In Fig. 4a, we visualized the proposed DL network [30,31]. Each layer in the network extracted different characteristics of HGSOC, from simple low-level features to complex high-level features. Filters from the first layer extracted CT intensities of tumor. Afterward, the second convolutional layer extracted tumor edge information. When the network went deeper, the convolutional layers extracted abstract and complex features such as more complicated edges in *Conv. 3* layer, shapes in *Conv. 4* layer and a combination of multiple features in the *Conv.5* layer.

Tumors with different recurrence times can activate different signal pathways of the DL network and finally be encoded into features with different values. We fed two tumor images from two patients (recurrence time = 8.53 and 31.07 months) into the DL network, and observed different responses as shown in Fig. 4b. The filters from the DL network had weak response on the patient with short recurrence time and high response on the patient with long recurrence time. This was further demonstrated in Supplementary Fig. S6 that depicted the distribution of patients in the DL feature space (reduced to 2-dimensional by principal component analysis algorithm [32] for display convenience). This figure indicated that patients with longer recurrence time were separated from patients with shorter recurrence time. Furthermore, we depicted the patients with recurrence time longer than 3 years and shorter than 3 years, and these two classes of patients could be divided easily in the DL feature space.

The good prognostic value of the DL feature probably comes from the DL network design, which includes an encoder network compressing tumor images into a lower dimensional representation (the DL feature) and a decoder network evaluating the information capacity of the DL feature. Each layer in the encoder network aims at using few features to include most information of the tumor. Therefore, redundant information is eliminated, and only features that can reflect the intrinsic characteristics of HGSOC are preserved. We used the DL feature to reconstruct the original tumor image in Supplementary Fig. S7. This figure indicated that the tumor image reconstructed from the DL feature was similar to the original tumor image, which demonstrated that the DL feature included the intrinsic characteristics that are essential to represent a tumor.

In addition, we evaluated the robustness of the deep learning features concerning the radiologist bias in selecting the ROIs of tumor. Two radiologists annotated ROIs independently on 40 patients that were randomly selected from the whole dataset. Afterward, we calculated the intra-class correlation coefficient (ICC) of the deep learning features using the ROIs selected by the two radiologists. The ICC values show that the deep learning

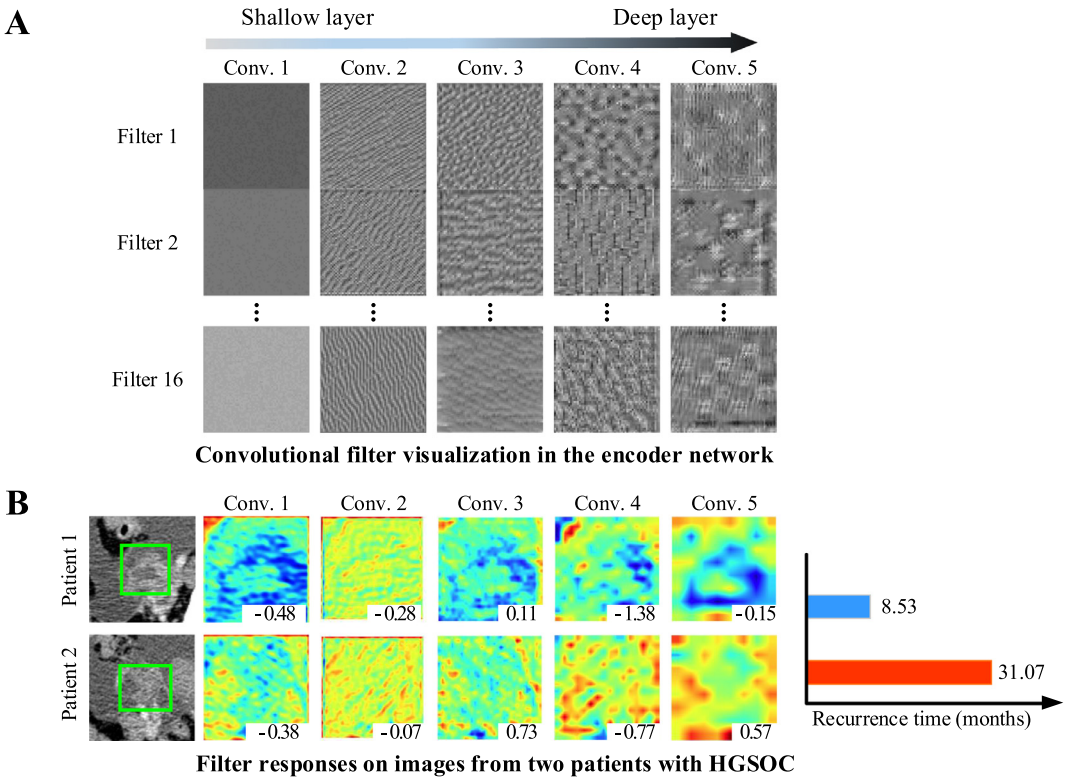

Fig. 4. DL feature visualization. (a) Convolutional filters from the five convolutional layers of the encoder network. Each convolutional layer includes 24 or 16 filters (24 filters for the first convolutional layer, and 16 filters for other convolutional layers), and we randomly selected three filters for visualization. (b) Filter responses on two patients with different recurrence time. The green box in the first column is the tumor location. Red and blue colors represent intensive and weak response. For display convenience, the response maps from the same convolutional layer are normalized. The number on the bottom right corner is the average value of the response map. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

features are stable between two radiologists (ICC range between 0.83 to 0.98 for all the deep learning features). Since we do not require precise tumor boundary segmentation, the ROIs selected by different radiologists do not have large effects to the deep learning features.

Despite the encouraging results, our study has several limitations. First, the DL network extracted the intrinsic characteristics of HGSOC. However, the specific connection between the DL feature and genetic changes were not explored. In the future work, we can explain the DL feature in genetic level by combining the genetic profile of these patients. Second, the DL-CPH model is separate from the Cox-PH. Therefore, the integration of Cox analysis and DL architecture needs further exploration.

To conclude, this study shows that deep learning can provide new CT-based prognostic biomarkers related to the recurrence of HGSOC, which demonstrated stronger prognostic value than clinical characteristics. We also developed a non-invasive DL-CPH model to predict the recurrence of HGSOC by preoperative CT imaging, aiming at assisting individualized treatment and surveillance planning in HGSOC. Moreover, we proposed a novel method to mine the intrinsic characteristics of HGSOC by unsupervised learning. This method can take advantage of the large amount of data without the need for follow-up information.

Acknowledgements

This paper was supported by the National Key Research and Development Plan of China (2017YFA0205200, 2016YFC0103001, YS2017YFGH000397, 2017YFC1308700, 2017YFC1309100), National Natural Science Foundation of China (81227901, 81527805, 81772012, 81501549, 81720108021, 81641168), the Beijing Natural Science Foundation (7182109), Beijing Municipal Science & Technology Commission (Z171100000117023, Z161100002616022), Chinese Academy of Sciences (GJJSTD20170004, QYZDJ-SSW-JSC005), Henan Province Scientific and Technological Cooperation Project (152106000014).

The authors would like to acknowledge the instrumental and technical support of Multi-modal biomedical imaging experimental platform, Institute of Automation, Chinese Academy of Sciences.

Conflicts of interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.radonc.2018.10.019.

References

[1] Richardson DL, Sill MW, Coleman RL, Sood AK, Pearl ML, Kehoe SM, et al. Paclitaxel with and without pazopanib for persistent or recurrent ovarian cancer: a randomized clinical trial. JAMA Oncol 2018;4:196–202.
[2] Ledermann J, Raja F, Fotopoulou C, Gonzalez-Martin A, Colombo N, Sessa C, et al. Newly diagnosed and relapsed epithelial ovarian carcinoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2013;24:vi24–32.
[3] Rizzuto I, Stavraka C, Chatterjee J, Borley J, Hopkins TG, Gabra H, et al. Risk of ovarian cancer relapse score: a prognostic algorithm to predict relapse following treatment for advanced ovarian cancer. Int J Gynecol Cancer 2015;25:416.
[4] Goode EL, Block MS, Kalli KR, Vierkant RA, Chen W, Fogarty ZC, et al. Dose-response association of CD8+ tumor-infiltrating lymphocytes and survival time in high-grade serous ovarian cancer. JAMA Oncol 2017;3. e173290-e.

[5] Luvero D, Milani A, Ledermann JA. Treatment options in recurrent ovarian cancer: latest evidence and clinical potential. Ther Adv Med Oncol 2014;6:229–39.
[6] Kurta ML, Edwards RP, Moysich KB, McDonough K, Bertolet M, Weissfeld JL, et al. Prognosis and conditional disease-free survival among patients with ovarian cancer. J Clin Oncol 2014;32:4102.
[7] Lee C, Simes R, Brown C, Lord S, Wagner U, Plante M, et al. Prognostic nomogram to predict progression-free survival in patients with platinum-sensitive recurrent ovarian cancer. Br J Cancer 2011;105:1144.
[8] Aerts HJ. The potential of radiomic-based phenotyping in precision medicine: a review. JAMA Oncol 2016;2:1636–42.
[9] Lambin P, Leijenaar RT, Deist TM, Peerlings J, de Jong EE, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017;14:749.
[10] Zhang B, Tian J, Dong D, Gu D, Dong Y, Zhang L, et al. Radiomics features of multiparametric MRI as novel prognostic factors in advanced nasopharyngeal carcinoma. Clin Cancer Res 2017;23:4259–69.
[11] van Timmeren JE, Leijenaar RTH, van Elmpt W, Reymen B, Oberije C, Monshouwer R, et al. Survival prediction of non-small cell lung cancer patients using radiomics analyses of cone-beam CT images. Radiother Oncol 2017;123:363–9.
[12] Jensen GL, Yost CM, Mackin DS, Fried DV, Zhou S, Court LE, et al. Prognostic value of combining a quantitative image feature from positron emission tomography with clinical factors in oligometastatic non-small cell lung cancer. Radiother Oncol 2018;126:362–7.
[13] Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014;5:4006.
[14] Huang Y, Liu Z, He L, Chen X, Pan D, Ma Z, et al. Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (i or ii) non—small cell lung cancer. Radiology 2016;281:947–57.
[15] Y-q Huang, C-h Liang, He L, Tian J, C-s Liang, Chen X, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. J Clin Oncol 2016;34:2157–64.
[16] Nougaret S, Lakhman Y, Gönen M, Goldman DA, Miccò M, D'Anastasi M, et al. High-grade serous ovarian cancer: associations between BRCA mutation status, CT imaging phenotypes, and clinical outcomes. Radiology 2017;285:472–81.
[17] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436.
[18] Kermany DS, Goldbaum M, Cai W, Valentim CC, Liang H, Baxter SL, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell 2018;172. 1122–31. e9.
[19] Wang S, Zhou M, Liu Z, Liu Z, Gu D, Zang Y, et al. Central focused convolutional neural networks: developing a data-driven model for lung nodule segmentation. Med Image Anal 2017;40:172–83.
[20] Lustberg T, van Soest J, Gooding M, Peressutti D, Aljabar P, van der Stoep J, et al. Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer. Radiother Oncol 2018;126:312–7.
[21] Thompson RF, Valdes G, Fuller CD, Carpenter CM, Morin O, Aneja S, et al. Artificial intelligence in radiation oncology: a specialty-wide disruptive transformation? Radiother Oncol.
[22] Chaudhary K, Poirion OB, Lu L, Garmire LX. Deep Learning based multi-omics integration robustly predicts survival in liver cancer. Clin Cancer Res 2017. clincanres. 0853.2017.
[23] Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nat Biomed Eng 2018;2:158.
[24] Hillman RT, Chisholm GB, Lu KH, Futreal PA. Genomic rearrangement signatures and clinical outcomes in high-grade serous ovarian cancer. JNCI-J Natl Cancer Inst 2018;110.
[25] Morgan RJ, Armstrong DK, Alvarez RD, Bakkum-Gamez JN, Behbakht K, L-m Chen, et al. Ovarian cancer, version 1.2016, NCCN clinical practice guidelines in oncology. J Natl Compr Cancer Netw 2016;14:1134–63.
[26] Masci J, Meier U, Cireşan D, Schmidhuber J. Stacked convolutional auto-encoders for hierarchical feature extraction. International Conference on Artificial Neural Networks. Springer; 2011. p. 52–9.
[27] Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1–W73.
[28] Liu Z, Zhang X-Y, Shi Y-J, Wang L, Zhu H-T, Tang Z-C, et al. Radiomics analysis for evaluation of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. Clin Cancer Res 2017. clincanres. 1038.2017.
[29] Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. Lancet Oncol 2015;16:e173–80.
[30] Kotikalapudi Rac. keras-vis. GitHub. https://github.com/raghakot/keras-vis2017.
[31] Yosinski J, Clune J, Nguyen A, Fuchs T, Lipson H. Understanding neural networks through deep visualization. arXiv preprint arXiv:150606579. 2015.
[32] Abdi H, Williams LJ. Principal component analysis. Wiley Interdiscip Rev Comput Stat 2010;2:433–59.
[33] Huang G, Liu Z, Weinberger KQ, van der Maaten L. Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 3.