

# 卵巣明細胞癌の遺伝子変異を病理スライドから予測可能なmachine learning classifierの開発によるprecision medicineの探求

Development of machine learning classifiers for predicting gene mutations in ovarian clear cell carcinoma using pathological glass slides

東京慈恵会医科大学 産婦人科学講座 助教 横溝 陵

## 要約

卵巣癌の病理組織型の一つである卵巣明細胞癌について、病理標本から得られる病理学的所見と遺伝学的検査の結果を学習データとしてdeep learningを行い、病理標本を画像データとして入力することによりactionableな遺伝子変異を同定できる分類器の作成を行い、precision medicineのさらなる発展に資することを最終的な目的として本研究を実施した。Whole slide imaging処理を行った卵巣明細胞癌110例のHE染色標本を用いて作成した分類器を用いて再発、progression free survival及びoverall survivalの予測を試みた。結果として、卵巣明細胞癌の再発を93%の精度で予測することができ、receiver operating characteristic曲線の曲線下面積は0.98、感度及び特異度は0.92以上であった。また、progression free survival及びoverall survivalについても約90%の精度で予測可能であった。本研究の成果として、卵巣明細胞癌に対する機械学習を用いた予後予測ツールを作成することができた。現在も引き続き、ゲノム情報及び臨床情報との関連について、作成したパイプラインを活用し解析を進めている。本研究の実現により、将来的なprecision medicineの全国的な幅広い普及に寄与することが期待される。

## 緒言

卵巣癌は全世界で313,000人以上が発症し、207,000人近くが死亡する婦人科悪性腫瘍である。なかでも卵巣明細胞癌は日本人女性に多く見られる組織型であり、上皮性卵巣癌の25~30%を占めている<sup>1</sup>。卵巣明細胞癌は約8割の症例が早期癌として発見され根治的治療により治癒する事が多いが、早期発見にも関わらず予後不良な転帰を辿る症例が存在するため、既知の手法を上回る正確な予後予測法の確立が望まれている。

これまでの研究では、年齢や進行期等の臨床情報、画像データ、CRPや腫瘍マーカー等の血液検査データを用いた統計学的解析により予後予測や種々のバイオマーカーの探索が行われてきた。しかし、大規模なデータセットを用いた多変量解析のみでの予後予測には限界があり、近年技術革新が目覚ましい機械学習の応用が期待されている。実際に、腫瘍学分野における機械学習の有用性に関する研究は近年活

発に行われており、卵巣癌に関しては、予後予測における機械学習の有用性について我々の研究室から報告した研究が注目されている<sup>2</sup>。

近年の卵巣癌診療では、病理学的所見に基づくリスク層別化に加え、遺伝学的検査により治療標的や予後予測因子となりうる遺伝子変異を同定し、患者ごとに個別化された治療 (precision medicine) が導入され始めているが、遺伝学的検査により得られるビッグデータを最大限活用することで、今後さらなる発展が期待されている。そのような背景の中、我々は、病理標本から得られる病理学的所見と遺伝学的検査の結果を学習データとしてdeep learningを行い、病理標本を画像データとして入力することによりactionableな遺伝子変異を同定できる分類器の作成を行い、precision medicineのさらなる発展に資することを目的として、本研究を実施した。

表1 対象症例の臨床病理学的因子

PARAMETER	PATIENTS (%), MEAN ± SD
Patients	110
Age (years old)	51.6 ± 10.5
Stage	
I	74 (67.3)
II	9 (8.2)
III	22 (20.0)
IV	5 (4.5)
Ascites volume	
<500mL	88 (80.0)
≥500mL	22 (20.0)
Ascites cytology	
Negative	15 (13.6)
Positive	92 (83.7)
NA	3 (2.7)
Surgery	
Optimal	76 (69.0)
Suboptimal	28 (25.5)
Fertility sparing	6 (5.5)
Chemotherapy (more than 3 cycles)	
With	97 (88.2)
Without	13 (11.8)
Follow-up period (days)	2794.1 ± 1777.4

NA, not applicable; SD, standard deviation.

## 方法

2000年から2012年の期間において、東京慈恵会医科大学附属病院でprimary surgeryが行われた卵巢明細胞癌患者を対象とした。臨床情報が不十分な症例及び適切なHE染色標本が入手不可能であった症例は解析対象から除外した。HE染色標本については婦人科病理医によりWHO分類（2014年）に基づく組織型のレビューが行われ、進行期についてはFIGO2014に基づいて決定した。2019年の段階で診療記録を後方視的にレビューし、年齢、進行期、腹水量、腹水または腹腔洗浄細胞診、術式、化学療法、再発の有無、PFS及びOS等の臨床病理学的因子について調査した（表1）。

続いて、病理標本の電子データ化とデータ処理を行うパイプラインの作成を行った。臨床情報と紐付

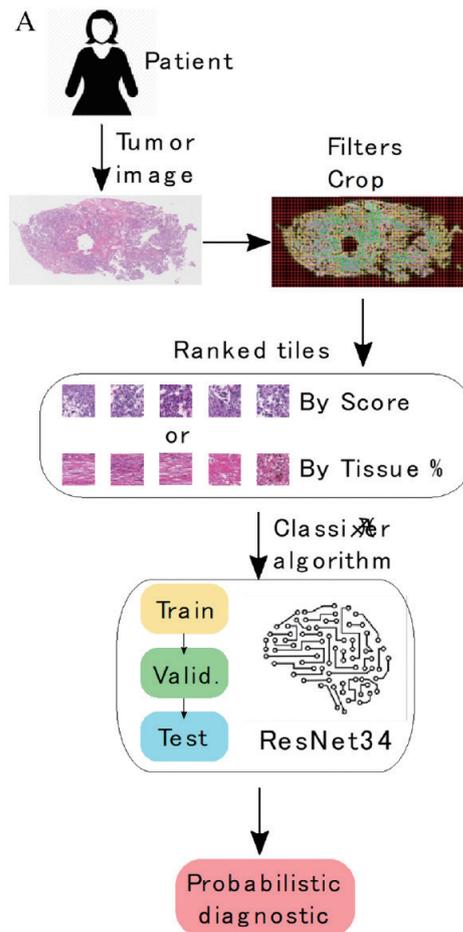


図1. 本研究におけるパイプライン

患者から得られた卵巢明細胞癌組織のHE染色標本についてデジタル画像を取得し、ノイズ等を除去するためにフィルタリングを行い、画像の品質向上を図った。その後各画像を小さなタイルに切り出し、“Tissue(タイル中の組織の占める割合)”と“Score(色の要素を付加)”によりランク付けし、各タイルの情報について定量化した。対象症例の70%の画像をTraining setに、15%をValidation setに、15%をTest setに割振り、深層学習ニューラルネットワークResNet34に適用することで予測精度を検証した。

く卵巢明細胞癌のHE染色標本110例について、バーチャルスライド化（Whole slide imaging処理）を行った。得られた画像データから、背景やノイズの影響を除外するフィルタリング処理を行い、さらにRGBスケールを用いて定量化データとした。パイプラインの作成については、フレームワークとしてPyTorchを、neural networkのモデルとしてResNet34を用いてDeep learningを行った（図1）。

並行した解析として、同一コホートについてFFPE標本からゲノム抽出し、actionableな84遺伝子を対象にターゲットシーケンスを行った上で、ゲノム情報及び臨床情報との関連について、作成したパイプラインを活用して画像データと統合して解析を行っている（継続中）。本研究は、研究計画について倫理委員会における承認を得た上で実施した。

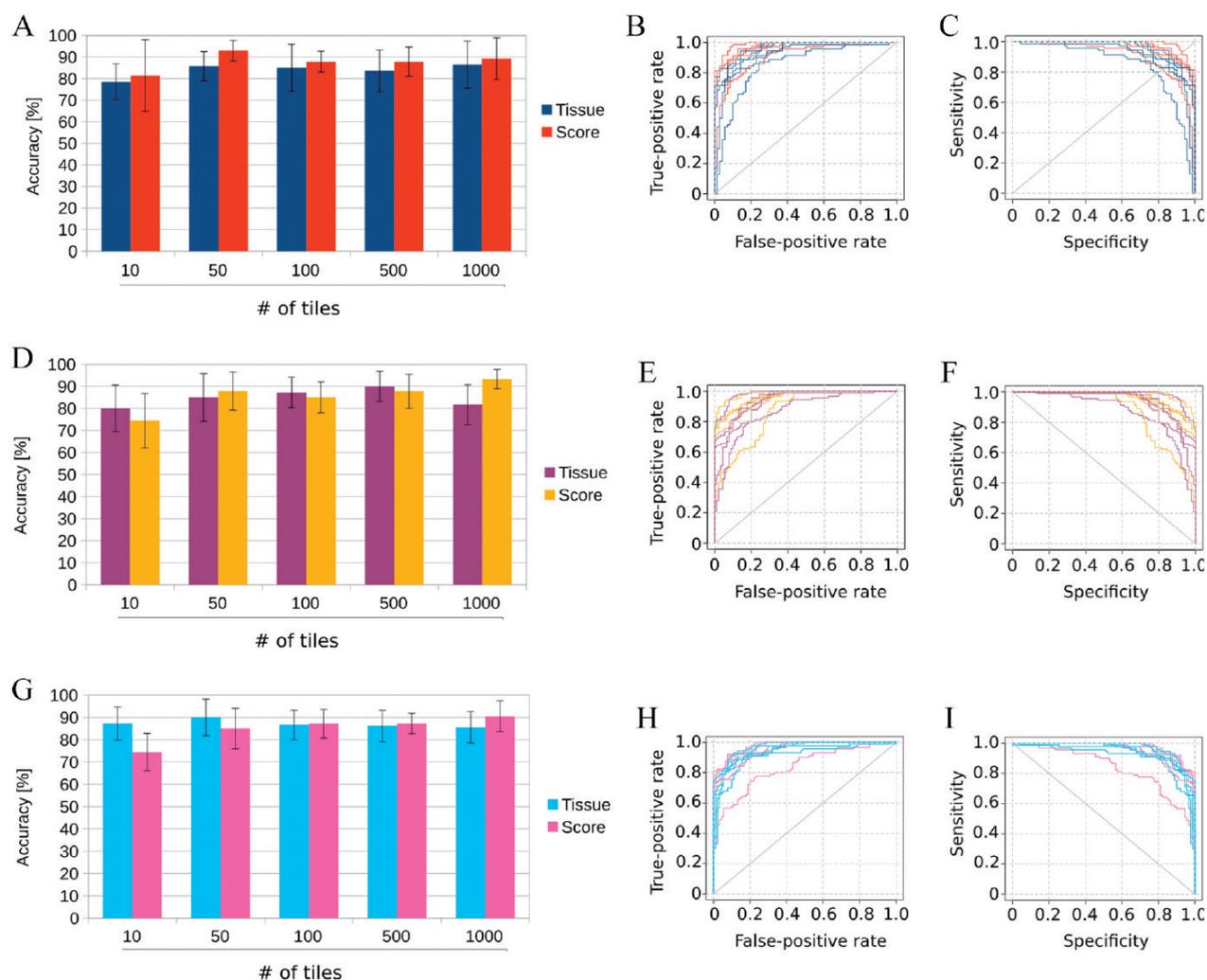


図2 作成した分類器の性能

(A) 再発/無再発の予測精度. 異なる数のタイルを使用し, "Tissue"または"Score" によってランク付けした場合の精度についての比較. (B) ROC 曲線. (C)感度・特異度曲線. (D~ F) PFSに関する予測についての予測精度, ROC 曲線, 感度・特異度曲線. (G~ I) PFSに関する予測についての予測精度, ROC 曲線, 感度・特異度曲線.

## 結果

Whole slide imaging処理を行った卵巢明細胞癌のHE染色標本110例から作成した分類器を用いて、再発/非再発、progression free survival (PFS), overall survival (OS) の予測を試みた。結果として、卵巢明細胞癌の再発を93%の精度で予測することができ、receiver operating characteristic (ROC) 曲線の曲線下面積は0.98、感度/特異度が0.92以上であった。また、PFS/OSについても約90%の精度で予測可能であった (図2)。

## 考察

本研究の成果として、卵巢明細胞癌に対する機械学習を用いた予後予測ツールを作成することができた<sup>3</sup>。これまでに、年齢、超音波所見、血液検査デー

タ等を組み合わせて、付随器腫瘍における卵巢癌の診断予測に関する研究が行われてきている。機械学習を用いたアプローチは重要なツールであると考えられる一方、卵巢明細胞癌の予後予測に関する機械学習アルゴリズムの開発は、卵巢明細胞癌の発生率が世界的に低いため困難であった。本邦では卵巢明細胞癌の発生率が高いため比較的多くのサンプルを入手可能であり、本研究ではそのリソースを活用して分類器を開発することができた。卵巢明細胞癌はStage Iであっても、一般的な画像検査等で検出されない微小転移が存在しうるため、病理画像を用いて再発を予測出来る本研究成果は、診療における意思決定に寄与する可能性がある。

近年、他のがん種において病理画像から遺伝子異常を予測する研究がされており、本研究計画におい

でも卵巣明細胞癌において同様のコンセプトで研究を展開する予定である。つまり、病理画像を入力することで、actionableな遺伝子の変異に関する情報を入手でき、治療標的・予後予測因子の同定が可能となる分類器の開発を目指している。また、本研究開発の成果が実際に臨床の場で使用可能か、外的妥当性について検証を進める計画である。本研究の実現は、どのような医療機関でも普遍的に利用可能な病理標本であるHE染色標本を出発点とした個別化医療の実現へと繋がり、将来的にはprecision medicineの全国的な幅広い普及に寄与することが期待される。

## 謝 辞

本研究の実施に際しまして公益財団法人神澤医学

研究振興財団からのご支援を頂きましたことに深謝申し上げます。

## 参考文献

1. Okamoto A, Glasspool RM, Mabuchi S, et al. Gynecologic Cancer InterGroup (GFIG) consensus review for clear cell carcinoma of the ovary. *Int J Gynecol Cancer*. 2014;24:S20-S25.
2. 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 Cancer Res*. 2019;25:3006-3015.
3. Yokomizo R, Lopes TJS, Takashima N, et al. O3C Glass-Class: A Machine-Learning Framework for Prognostic Prediction of Ovarian Clear-Cell Carcinoma. *Bioinform. Biol. Insights*. 2022;16:1-9

## Abstract

Ovarian clear cell carcinoma (OCCC), one of the histopathological types of ovarian cancer, has a poor prognosis when it recurs; however, it is difficult to precisely predict the risk of recurrence. Here, we analyzed pathological images of OCCC to elucidate the relationship between pathological findings and recurrence, and using machine learning, we established a classifier to predict the recurrence and several other prognosis indicators of this disease. In total, 110 patients with OCCC treated with primary surgery at a single institution were enrolled in this study. We used the deep-learning neural networks to process the whole slide images of OCCC obtained by digitally scanning the original hematoxylin and eosin-stained glass slides. The images were preprocessed and used as input to the machine learning pipeline. We fine-tuned its parameters to predict the recurrence, progression-free survival, and the overall survival days of all patients. We predicted the recurrence of OCCC with an overall accuracy of 93%, area under the receiver operating characteristic curve of 0.98, and sensitivity/specificity above 0.92 using Resnet 34. Furthermore, we predicted progression-free survival/overall survival of the patients with ~90% accuracy. In conclusion, our study demonstrates the feasibility of using a machine learning system to predict different features of OCCC samples using histopathological images as input. We are currently developing a machine learning system to predict gene mutation from histopathological images. Our novel application provides accurate prognosis information and aids in the development of personalized treatment strategies.