

令和 5 年 4 月 24 日現在

機関番号：32645

研究種目：基盤研究(C)（一般）

研究期間：2018～2020

課題番号：18K07027

研究課題名（和文）人工知能による乳癌悪性度予測システムの開発

研究課題名（英文）Breast cancer patients NAC response prediction with artificial intelligence

研究代表者

齋藤 彰（Saito, Akira）

東京医科大学・医学部・教授

研究者番号：10504615

交付決定額（研究期間全体）：（直接経費） 2,500,000円

研究成果の概要（和文）：乳癌では術前化学療法を施行することが多くなってきている。治療開始前の針生検病理標本を使用し、人工知能技術を使用し、その治療効果予測を行うシステムを開発した。人工知能としてはディープラーニング ResNext、機械学習(SVM Random Forest)を使用した。310症例で207症例をトレーニング症例とし、103症例をテスト症例とした。

現時点ではSVMでは症例単位での尤度による判定で、85%の精度で予測できた。Deep Learning Random Forestにおける結果分析を実施しているが、ほぼ同等の結果であると考えられ、総合的に95%程度の予測が可能になると思われる。

研究成果の学術的意義や社会的意義

近年術前化学療法を施行することで、その後の手術切除範囲を縮小し温存できることと、再発時には化学療法の効果を把握することが可能となる。治療開始前の針生検病理標本から癌の細胞核の形態学的特徴量を抽出することで、その効果を予測することが可能となった。また乳癌のサブタイプ別の特徴などを抽出することができた。これらにより患者に追加の負担を掛けることなく、患者が受ける治療の効果の予測が可能となる。また乳癌サブタイプは各種の遺伝子の変異をともなうことで生じており、これらの変異と細胞形態学的関係、化学療法に反応する癌細胞の特徴的形態を示すことが可能となる。

研究成果の概要（英文）：The patients with NAC (Neoadjuvant Chemotherapy) were increasing. we developed NAC response prediction system by using digital pathology technique and AI (artificial intelligence). Regarding AI, we used deep learning and machine learning (SVM and Random Forest). The cases were Tokyo Medical University, Yamaguchi University, and Weill Cornell Medicine, total 310 cases, 207 cases were used as training data set and remaining 103 cases were test cases set. We created 85% accuracy prediction system by SVM with using likelihood score. Now we are analyzing Deep Learning and Random Forest model, there are almost same level accuracy. We are creating these 3 combination model, and that accuracy will become 95%.

研究分野：デジタル病理

キーワード：人工知能 デジタル病理 乳癌 術前化学療法 予後 効果予測

1. 研究開始当初の背景

乳癌の治療は手術により、癌の摘出およびリンパ節の郭清が基本である。しかし乳癌の9割は浸潤癌であり、術後の化学療法を行い、再発・転移を抑えることが必要になる。この手術療法と化学療法の施行順を逆にして化学療法を術前に施行するものを術前化学療法 NAC (Neoadjuvant chemotherapy) と言う。術前あるいは術後に行う化学療法は原則として同じものを実施する。化学療法を術前に実施した場合と術後に実施した場合における生存率あるいは再発・転移などの差はないことが示されている。術前に化学療法を施行することで、化学療法により癌の縮小がみられる場合には、手術摘出範囲を縮小することが可能で温存率が向上する。また NAC によりその後の手術で化学療法の効果を実証することができ、再発・転移した場合の化学療法剤の選択に情報を与えることが可能になるなどの利点がある。温存療法を希望する患者も増加することから、近年 NAC 療法を実施することが多くなってきている。NAC 療法の施行は、手術後に化学療法が必要と考えられる症例 Stage II, III が基本となる。

NAC はアンスラサイクリン系の薬剤をまず投与し、その後タキサン系の薬剤を使用して約6-8カ月の治療を実施する。アンスラサイクリン系としては Epirubicin, Doxorubicin に Cyclophosphamide を加えるもの 5-Fluorouracil を用いる場合など多くの組み合わせからその薬剤の頭文字から EC, FEC などと表現される。タキサン系としては Paclitaxel, Docetaxel などが使用される。また Her2 陽性患者には Trastuzumab, Pertuzumab などが追加される。

NAC の治療効果評価は終了後に施行される手術摘出標本で、0: 治療前と変化なし 1: 1/3 が縮小 2: 2/3 が縮小 3: 浸潤癌の消失 の4段階で、1, 2 については残存癌の変異で a, b に分けて計6段階の評価がなされている。

一方、WSI (Whole Slide Image) スキャナーの普及で、病理標本のデジタル画像化が可能となり、デジタル化された病理画像をコンピュータで解析を行うデジタル病理の普及が進展している。人工知能を使用し、多くは癌部位の検出が目的で実施されているが、再発の可能性あるいは治療の効果予測などの試みも実施されている。

2. 研究の目的

NAC 療法施行には必ず浸潤癌であることの確定診断および乳癌のサブタイプ (ホルモンレセプタと Her2 の発現有無) を確定するため、針生検により NAC 前の組織診断を実施する。これらの針生検組織画像をデジタル画像化の上、NAC 治療の効果予測を人工知能でどこまで予測できるかを研究目的とする。WSI によって得られる画像は1Gバイト近くになるため、そのままの画像ではコンピュータの制約から解析はできないため ROI (Region of Interest) と呼ぶ小領域に分割して解析を実施する。解析ではこの ROI 単位を独立に解析を進めていく。人工知能のモデル構築にあたり、癌か否かは病理学的に判断可能であり人工知能のモデル構築の教育ができる。これらは Supervised モデルと呼ばれるが、治療効果については組織画像を見ても病理学的には判断ができない。このため、この症例は NAC の効果判定では 0 とか 3 であった症例と教育できるが、当然 NAC 評価で 0 であっても NAC に反応する ROI としない ROI が混在することになる。これらのモデルは weakly supervised model と呼ばれる。

針生検画像を用いて人工知能 weakly supervised model を構築し、NAC 治療効果予測を実施する。また各 ROI での結果を症例単位の結果にどのように集約して症例単位の予測結果につなげるか、また各種ある人工知能手法での結果の違いをどのように評価するか、また人工知能が判断に使用した画像の特徴量の評価基準分析を二次的な研究目的とした。

3. 研究の方法

東京医科大学 136 症例 山口大学 26 症例 Weill Cornell Medicine 148 症例 計 310 症例を解析対象とし、207 症例を人工知能モデルの構築用トレーニング症例に使用して残りの 103 症例をテスト症例とした。複数の施設での症例を使用することで、施設による差を排除することが可能である。サブタイプとしては TN (Triple Negative: ホルモンレセプタ Her2 共に陰性) H+ (ホルモンレセプタのみ陽性) Her2+ (Her2 レセプタのみ陽性) H+Her2+ (ホルモンレセプタと Her2 レセプタ陽性) 症例はそれぞれ 56 (18%), 143 (46%), 40 (13%), 71 (23%) であった。これらの比率は一般的な患者の分布と同じで、日米での差はなかった。

人工知能としては現在広く使用されている Deep Learning で画像を扱う CNN (Convolutional Neural Network)の中から ResNext、機械学習法 ML (Machine Learning)手法としては SVM (Support Vector Machine) RF(Random Forest)の手法を独立に実施することとした。MLの場合は、人工知能に与える画像の特徴量抽出作業が必要であり、癌細胞の核を抽出し、それらを計測することが必要となる。

解析手法は次のような手順で実施した。

- ・全 310 症例の治療前の針生検 HE 染色スライドを NanoZoomer (浜松ホトニクス社) で単層 20 倍で撮影して、デジタル画像を取得する。
- ・CNN (ResNext) 解析用としては図 1 に示すように元のデジタルスライドイメージ a において

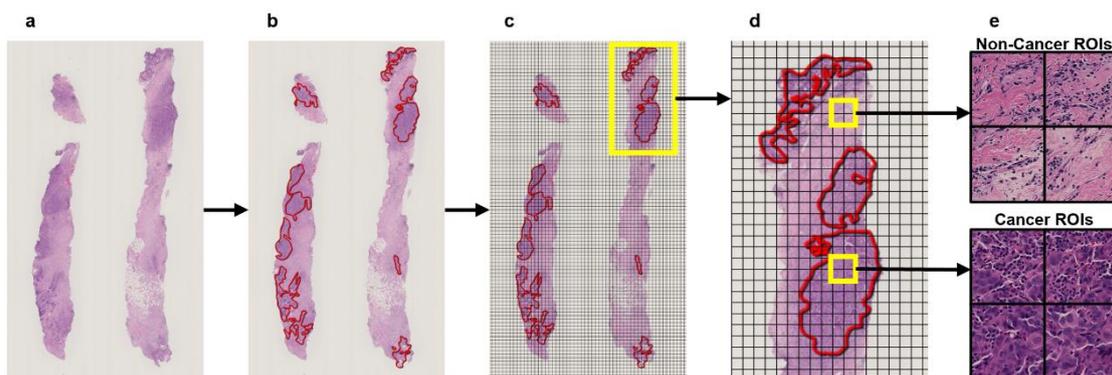


図 1 CNN 用解析作業手順

癌領域をアノテーションする (b)。全領域を 256x256 pixels で分割し (d) 癌領域と非癌領域に分割した ROI(Region of Interest)に分割する。各 pixel で半分以上が癌領域ならば癌領域 pixel とした。全体で 858,863 ROI が抽出された。

- ・SVM と Random Forest 用の作業は図 2 に示すように 2048x2048 pixels で癌領域から人手で ROI 画像を選択(a)し ROI 画像(b)を取得した。

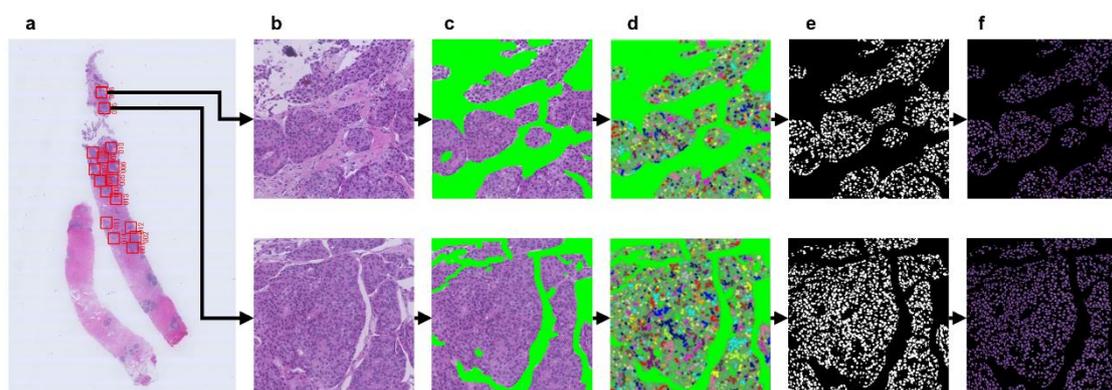


図 2 SVM と Random Forest 用解析手順

各 ROI には間質、リンパ球凝集領域、壊死あるいは生検での挫滅組織もあり、これらの領域は緑でマスクした画像 (c) を作成し、Icastik を使用して核抽出して(d)核のみのマスク画像(e)を作成、c の上に e をオーバーレイすることで細胞核のみの画像(f)を生成する。

最終的に 4,911 ROI 画像から 2,516,356 個の乳癌細胞を抽出した。これらの細胞核ごとに

82個の細胞核特徴量を CellProfiler で計測する。この個々の細胞核単位の計測データは CFLCM に入力して ROI 単位に Heterogeneity を含めた 960 個の特徴を算出する。

4. 研究成果

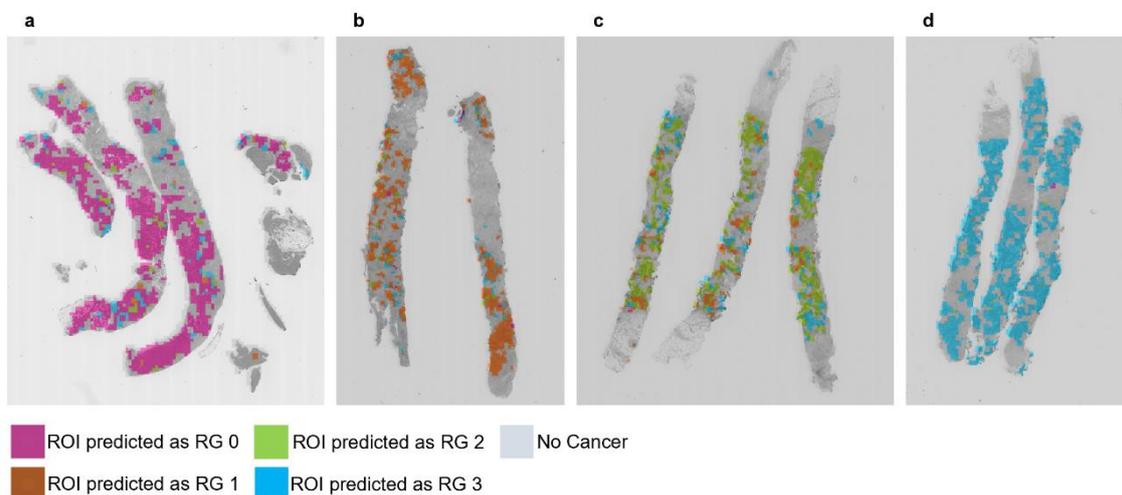


図3 CNNの解析結果

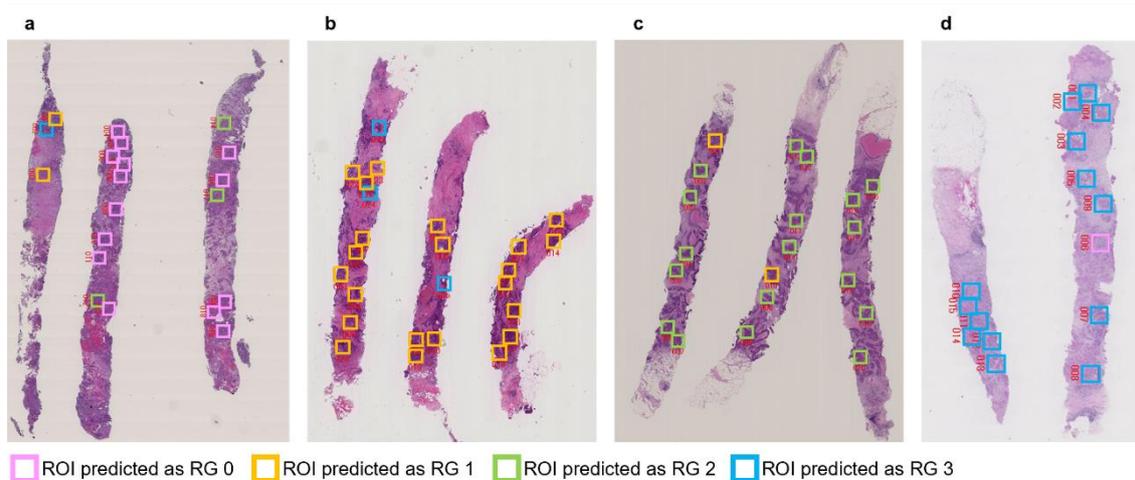


図4 SVMでの解析結果

CNN、SVM および Random Forest 各解析結果は ROI 単位に判断されるが、症例ごとに各 ROI の平均をもとめ、最終的に症例単位の結果とした。

結果として CNN 86.4% SVM 85.4% Random Forest 82.24%の精度で NAC に対する効果を予測できた。

3つのモデルを総合化する手法として、1つは複数のモデルの結果が一致をしていることを確認していく Ensemble learning での方法を使用した場合 94.2%の精度となる。

しかし各モデルでは結果に対して尤度値を出力しており、モデルとしては結果に対する一種の保証と考え、CNN, SVM, Random Forest の順で尤度が低ければ次のモデルの結果を採

用するという流れパイプラインを作成(図5)して結果を求め最終的に 95.15%の精度となった(図5)。

また SVM 解析では NAC の反応 (0-3 の 4 群) のみでなく、各症例のサブタイプ (ホルモンレセプター有無と Her2 の陽性陰性) の区分を加えた 16 群 (実際には NAC の反応が 0 で Her2+症例はなかった) の方が分離状態は良いことが分かった。

(図6) このことは、乳癌サブタイプにはそれぞれの細胞核における形態学的特徴が含まれている可能性を示唆しており、今後の研究としたい。

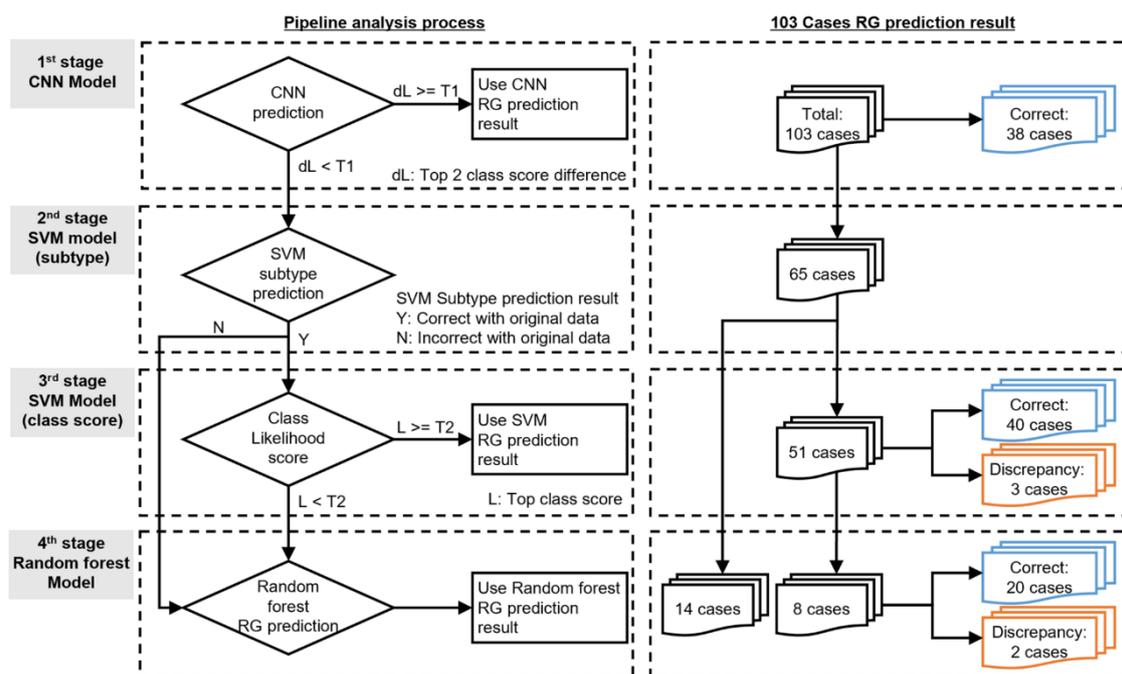


図5 NAC 反応予測における AI モデルパイプライン

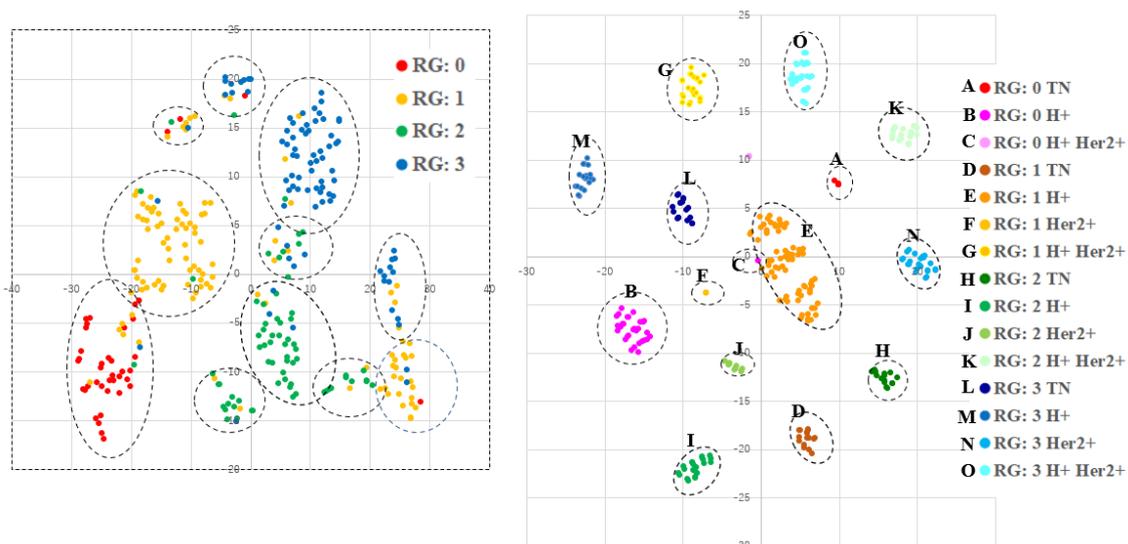


図6 SVM 結果 4 群と 15 群での分離結果(tSNE)

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件/うち国際共著 1件/うちオープンアクセス 1件）

1. 著者名 Bin Shen, Akira Saito, Ai Ueda et al.	4. 巻 0
2. 論文標題 Development of multiple AI pipelines that predict neoadjuvant chemotherapy response of breast cancer using H&E-stained tissue	5. 発行年 2023年
3. 雑誌名 The Journal of Pathology; Clinical Research	6. 最初と最後の頁 182-194
掲載論文のDOI（デジタルオブジェクト識別子） 10.1002/cjp2.314	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

〔学会発表〕 計3件（うち招待講演 0件/うち国際学会 0件）

1. 発表者名 上田亜衣 齋藤 彰 黒田雅彦 石川 孝
2. 発表標題 乳癌病理画像の人工知能解析による術前化学療法の効果・予後予測
3. 学会等名 第28回日本乳癌学会総会
4. 発表年 2020年

1. 発表者名 永松由衣 齋藤 彰 梅津知宏 黒田雅彦
2. 発表標題 新たな人工知能の手法による乳癌のNAC療法の効果予測システムの開発
3. 学会等名 第111回 日本病理学会総会
4. 発表年 2022年

1. 発表者名 沈 淋 齋藤 彰 黒田雅彦
2. 発表標題 人工知能パイプラインによる乳癌のNAC療法の効果予測モデルの開発
3. 学会等名 第112回 日本病理学会総会
4. 発表年 2023年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分担 者	黒田 雅彦 (Kuroda Masahiko) (80251304)	東京医科大学・医学部・主任教授 (32645)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------