

敵対的生成ネットワークによる間質性肺疾患患者の予後予測
Use of generative adversarial network for
survival prediction in patients with interstitial lung diseases

植村 知規^{†,‡} 渡利 千夏[†]
Tomoki Uemura Chinatsu Watari

Näppi Janne[†] 弘中 亨[†] 金 亨燮[‡] 吉田 広行[†]
Janne Näppi Toru Hironaka Hyougseop Kim Hiroyuki Yoshida

1. はじめに

関節リウマチ(RA)は、関節滑膜を病変の主座とする慢性炎症性疾患であり、日本における罹患率は人口の0.6-10%と推定されている[1]。間質性肺疾患(ILD)は関節リウマチによって引き起こされる代表的な呼吸器病変の1つであり、関節リウマチ関連間質性肺疾患(RA-ILD: Rheumatoid Arthritis-associated Interstitial Lung Diseases)と呼ばれる。関節リウマチ患者の5-10%が間質性肺疾患を患っているとされる。RA-ILDは、平均生存期間3-8年と予後不良な疾患であり、その臨床経過は多様であるため、個々の患者の予後予測は困難である。

RA-ILDの予後予測に対し、臨床バイオマーカーとして性別(Gender)、年齢(Age)、呼吸機能(Physiology)から算出されるGAP-index[2]が広く用いられている。一方、胸部CT画像の画像解析による、画像主体の予後予測手法は未だ確立されていない。

また、近年、敵対的生成ネットワーク(GANs: Generative Adversarial Networks)は画像生成タスクにおいて目覚ましい成果を上げており、医用画像における諸問題に対してGANを用いた研究が多く報告されている。GANを用いた予後予測手法は既に報告されているが、既存の手法は肺機能検査等の数値データからGANによって生存時間の予測を行うのみであるため、画像解析による予測は不可能である。

本研究では、GANによる胸部CT画像から直接的に生存時間の推定を行うことを可能とする、pix2survと呼ばれる新たな予後予測モデルを提案し、既存のバイオマーカーによる予後予測との比較、評価を行う。

2. 関連研究

Ryersonら[2]は、性別(Gender)、年齢(Age)、呼吸機能検査(Physiology)から算出されるGAP-indexを用いた予後予測モデルを提案し、評価指標となるC-index(Concordance index)において74.6%を達成した。Watariら[3]は畳み込みネットワークから得られた特徴量を用いたモデルを提案し、C-index: 85.5%を達成した。これらのモデルはCox比例ハザードモデルによる予後予測モデルであるため、ある時間での生存確率を予測するのみであり、患者に対する生存時間を推定することは不可能である。

Paidamoyoら[4]はGANに基づく予後予測モデルを提案し、真の観測時間と予測時間との相対絶対誤差(RAE: Relative Absolute Error)において23.6%を達成した。このモデルは、患者の生存時間を直接的に推定することが可能である一方、

呼吸機能検査等の検査量からなるベクトルデータをGANへの入力として用いており、画像ベースの予測モデルではない。

3. 手法

本研究では、GANによる予後予測モデルを提案する。提案モデルはCT画像を入力とし、生存時間を直接的に予測することが可能な、画像解析をベースとした新たな予後予測モデルである。本研究では、提案モデルをpix2survと呼ぶ。以下に詳細を述べる。

3.1 pix2surv

pix2survは時間の推定を行うTime Generator(G)、入力された画像対が真の時間とCT画像のペアか、予測された時間とCT画像とのペアかを識別するDiscriminator(D)から成る、conditional GAN(cGAN)[5]である。pix2survの概略図を図1に示す。

Gは画像のエンコード部と生存時間の予測を行う予測部から成る。エンコード部は4層の畳み込み層から成り、それぞれバッチ正規化層、活性化関数としてLeaky-ReLUを持つ。予測部は3層の全結合層から成り、1, 2層への入力の20%はDropoutにより無視される。Dは5層の畳み込み層から成り、最終層を除きバッチ正規化層、活性化関数としてLeaky-ReLUを持つ。最終層の活性化関数はシグモイド関数であり、入力された画像対が真の時間とCT画像のペアである確率を出力する。

3.2 損失関数

pix2survの損失関数LはAdversarial loss: L_{cGAN} と打ち切り(censored)データおよび非打ち切り(non-censored)データに関

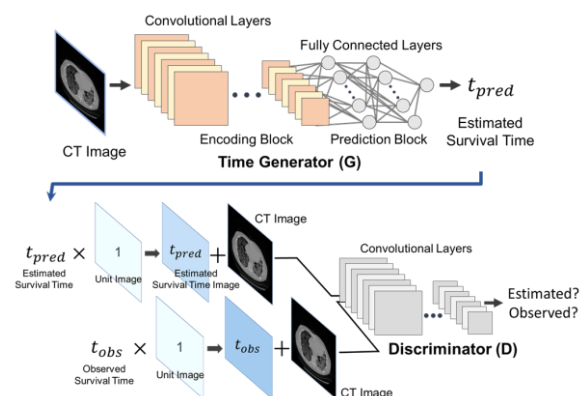


図1 pix2survの概略図

[†] Massachusetts General Hospital and Harvard Medical School

[‡] 九州工業大学 Kyushu Institute of Technology

する損失 $L_{censored}$, $L_{noncensored}$ から成り、次式で与えられる。

$$L = L_{cGAN} + L_{censored} + L_{noncensored} \quad (1)$$

L_{cGAN} は D と G に関する損失関数であり、 D の予測に対して次式のように定められる。

$$L_{cGAN} = E_{x,t \sim p_{data}(x,t)} [D(x,t)] + E_{x \sim p_{data}(x), z \sim p_z(z)} [1 - D(G(x,z))] \quad (2)$$

ここで、 x は入力画像、 t は真の観測時間、 z はガウス分布から得られる潜在ベクトルである。 L_{cGAN} に関して、 D は最大化、 G は最小化を目的に、それぞれのパラメータの最適化が行われる。

$L_{censored}$ および、 $L_{noncensored}$ は G に関する損失関数であり、予測と真値(観測時間)との絶対誤差によって次式で定められる。

$$L_{censored} = E_{x \sim p_{data}(x), z \sim p_z(z)} [\max(0, t - G(x,z))] \quad (3)$$

$$L_{noncensored} = E_{x \sim p_{data}(x), z \sim p_z(z)} [|t - G(x,z)|]$$

予後予測において、観測された時間には打ち切りデータが存在し、打ち切りデータはイベント(患者の死亡)が観測されていない。従って、打ち切りデータの予測時間が観測時間を下回る場合のみ損失を与える。

3.3 生存時間の予測

1患者のスキャンあたり複数枚のCT画像が得られる。従って、1患者あたりに複数の予測時間が得られる。本研究では、1患者から得られる複数の予測時間の中央値を、最終的な患者あたりの予測時間として得る。

4. データセットおよび評価方法

実験は RA-ILD 71 症例を用いた。すべての症例において、患者は胸部 CT 撮影および呼吸機能検査を受けており、医師による臨床的知見、並びに生検によって RA-ILD であると診断されている。

モデルの評価指標として C-index および RAE を用いた。RAE は相対絶対誤差であり、次式で与えられる。

$$RAE = \frac{1}{N} \sum_i \frac{|t_i^{pred} - t_i^{obs}|}{t_i^{obs}} \quad (4)$$

ここで、 N は患者数、 t_i^{pred} , t_i^{obs} はそれぞれ、ある患者 i の pix2surv による予測時間および、真の観測時間を表す。

また、値のバイアスを考慮するため、反復回数を 500 回としたブートストラップ法を用いて C-index および RAE を算出した。また、一般的に用いられるバイオマーカーである GAP-index および、年齢と性別(Age + Gender)を用いた予後予測モデルとの比較を行い、優位水準を $p = 0.05$ と定め

表1 各モデルの C-index および RAE

Model	C-index [95% CI]	RAE [95% CI]	P-value
pix2surv	88.0 [82.1, 93.9]	16.9 [11.3, 22.6]	-
GAP	77.1 [68.7, 85.5]	22.3 [16.4, 28.3]	<0.001
Age+Gender	64.5 [59.3, 75.0]	27.3 [19.8, 34.7]	<0.001

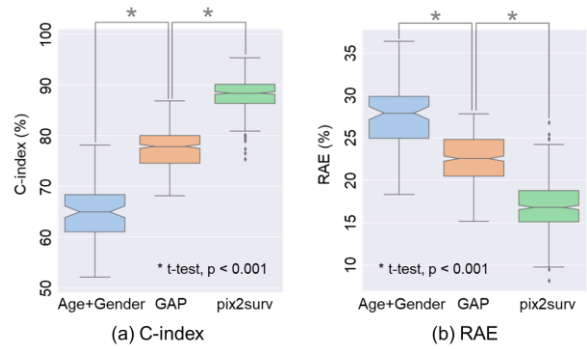


図2 各モデルの(a)C-index, (b)RAE の箱ひげ図

た両側 t 検定により、モデル間の C-index 及び RAE の差の統計的検定を行った。

5. 実験結果

pix2surv は C-index: 88.0%, RAE: 16.9%, GAP は C-index: 77.1%, RAE: 22.3%, Age + Gender は C-index: 64.5%, RAE: 27.3% となった。C-index および RAE の箱ひげ図を図2に示す。また、各モデルの C-index および、RAE を 95% Confidence Interval(CI)を含め、表1に示す。

pix2surv は C-index および RAE に関し、最も良い結果を示し、C-index において GAP よりも 10.9%、Age + Gender よりも 23.5% 高く、RAE において GAP よりも 5.4% 低く、Age + Gender よりも 10.4% 低い。これらの差は統計的に有意であり ($p < 0.001$)、従って、提案モデルである pix2surv が予後予測において優れていることが示された。

6. 結論

本研究では RA-ILD の GAN を用いた予後予測モデルである pix2surv を提案し、性能の評価を行った。pix2surv は CT 画像から直接的に生存時間の予測を可能とする新たな予後予測モデルであり、既存のバイオマーカーと比較して優れた性能を示した。従って、提案モデルである pix2surv は RA-ILD 症例の予後予測に有効なバイオマーカーに成り得ることが示された。

参考文献

- [1] Yamanaka H., Sugiyama N., Inoue E., Taniguchi A., Momohara S., "Estimates of the prevalence of and current treatment practices for rheumatoid arthritis in Japan using reimbursement data from health insurance societies and the IORRA cohort (I)", *Modern Rheumatology* Vol 24, No. 1 (2014).
- [2] Ryerson C. J., Vittinghoff E., Ley B., Lee J. S., Mooney J. J., Jones K. D., Elicker B. M., Wolters P. J., Koth L. L., King T. E., Jr, Collard, H. R., "Predicting survival across chronic interstitial lung disease: the ILD-GAP model", *Chest* Vol. 145, No. 4 (2014).
- [3] Watari C., Matsuhiro M., Näppi J. J., Nasirudin R. A., Hironaka, T., Kawata, Y., Niki N, Yoshida, H., "Radiomic texture-curvature (RTC) features for precision medicine of patients with rheumatoid arthritis-associated interstitial lung disease", *Proc. SPIE Medical Imaging 2018* (2018).
- [4] Chapfuwa P., Tao C., Li C., Page C., Goldstein B., Carin L, Henao R., "Adversarial Time-to-Event Modeling", arXiv 1804.03184 (2018).
- [5] Mirza M, Osindero S, "Conditional Generative Adversarial Nets", arXiv 1411.1784 (2014).