

情報理論を用いたガン診断支援システムの機能検証へ向けた プロトタイプの実作

多田 賢一郎[†] 大河原 菜裕[†] 大同 海良[†] 下川 舜[†] 頃安 隆耀[†] 加山 結翔[†] 金道 敏樹[†]
金沢工業大学工学部情報工学科[†]

1. はじめに

近年、がん患者とその検査数の増加により、病理専門医の人手不足が深刻化している。そこで、畳み込みニューラルネットワークなどの人工知能を使った病理画像診断技術の研究を念頭に、病理専門医の負担軽減へ向けた取り組みが注目されている。人工知能による病理画像診断技術は時に、医師と同等かそれ以上の診断精度を示すといった成果を上げている。しかし「がんの判断根拠が不明確である」「教師データには高精度のアノテーションが求められる」という残存課題が存在し、医師が信頼して使用できる段階には至っていない。

本研究室では、判断根拠が明らかな「医師のがん診断」を支援するシステムの開発に CAMELYON16[1]のデータを用いて取り組んでいる。これまでにがんらしさを情報量の密度を多寡で可視化する情報密度法[2, 3, 4]を提案し、がん領域と正常領域を判別可能であるという成果を得ている。一方で、実用化に向けた機能検証の段階に至っていないことやがんと正常を判別するための計算時間が膨大であるという課題が存在する。

そこで我々の研究では今後の実用化に向け、がん診断支援システムのプロトタイプの実作をし、評価を行う。

2. システムの構想と一次試作

2.1 医師のニーズ

先行研究にて、がん領域が小さくても発見可能であり、がん領域であるか怪しい領域を判別可能であることが医師のニーズとして明らかになった。我々はこれらの条件を満たすサーバークライアントシステムのインターフェースを実作する。

2.2 試作したプロトタイプの概要

試作したプロトタイプではバッチ処理とインタラクティブ処理に区別してシステムのプロトタイプを実作する。

図1はプロトタイプの表示画面である。

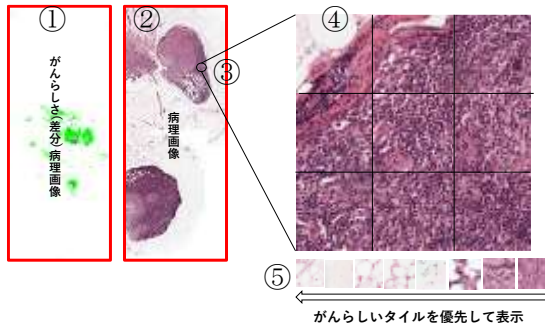


図1 システムのプロトタイプ

がん領域をおおまかに把握するために、①と②にてがんらしさと正常らしさの差分病理画像と診断したい病理画像を表示する。マウスホイールの操作に応じてクローズアップをする。更に詳細を確認するため、③で色付け病理画像にダブルクリック機能と④でピックアップ機能を実装する。選択した箇所を解像度の高いタイルと周辺8近傍タイルを表示することで、がん領域が小さい場合やがん領域が怪しい領域の判別にも対応可能である。⑤にてがんらしいと判断された画像を上位から順に表示するボタンを作成する。

3. 取り組む課題

今回システムのプロトタイプを実作することで計算時間が膨大であるという課題、精度の向上が課題として浮き彫りとなった。医師に提供するがんらしさと正常らしさの差分で色付けを行った病理画像は、支援でも自動診断に準じる精度が求められるが、その精度には至っていない。

3.1 計算時間の短縮

我々が扱う病理画像は、非圧縮状態で50GBあり、扱うには時間が掛かる。表1はプロトタイプ試作時の色付け処理に要した時間と特徴点数である。色付けを行う処理に18時間という膨大な計算時間を要していることが分かる。

表1 処理に要した計算時間と特徴点数

処理概要	プロトタイプ試作時の色付け処理
がんの根拠となる特徴量群	2389 個
正常の根拠となる特徴量群	2571 個
タイルから抽出する特徴点数	3500 個
計算時間	18 時間

コード解析から計算に時間を要していた原因は、「冗長な処理」「タイルから抽出する特徴点数が多い」「がん・正常の根拠となる特徴量の個数が多い」と分かった。非常に大きい病理画像を扱っているため、冗長な処理は実用化に向けて致命的となりえる。そこで処理の効率化と後述する最悪基準タイル選択アルゴリズムを実装することによって計算時間の短縮を図った。その結果が表2である。冗長性除去と最悪基準タイル選択アルゴリズム実装前と比べ、計算時間を1/18へ大幅に短縮できたことが分かる。

表2 処理に要した計算時間と特徴点数

処理概要	最悪基準タイル選択アルゴリズム実施後の色付け処理
がんの根拠となる特徴量群	1324 個
正常の根拠となる特徴量群	1217 個
タイルから抽出する特徴点数	500 個
計算時間	1 時間

Prototype production for functional verification of cancer diagnosis support system using information theory
Kenitirou Tada[†] Mahiro Ookawara[†] Kaira Daidou[†] Syun Simokawa[†] Ryu Koroyasu[†] Yuito Kayama[†] Toshiki Kindo[†]
Department of Information and Computer Science,
College of Engineering, Kanazawa Institute of Technology[†]

3.2 最悪基準タイル選択アルゴリズム

現段階で「がん領域に属するタイルの内、最も正常らしいと認識されているタイル」と「正常領域に属するタイルの内、最もがんらしいと認識されているタイル」を学習させる。これによって、現段階の「がんの根拠となる特徴量群」と「正常の根拠となる特徴量群」のノイズを除去し、新たにがん・正常の根拠となる特徴量を補填する。学習後に情報密度法にしたがってがんらしさと正常らしさを更新し、同じ処理を繰り返す。この学習方法方法を最悪基準タイル選択アルゴリズム[5]という。

3.3 精度の向上

図2は最悪基準タイル選択アルゴリズム導入前の色付けを行った病理画像である。左から「病理画像(一部)」「正常らしさで色付けを行った病理画像」「がんらしさで色付けを行った病理画像」「がんらしさと正常らしさの差分で色付けを行った差分病理画像」である。差分病理画像内のアノテーションが施されている領域ががん領域である。がん領域ではない領域の色付けが濃いことが見て取れる。

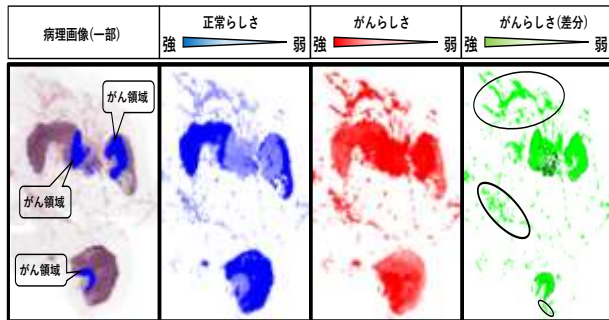


図2 プロトタイプ試作時に色付けを行った病理画像

がん領域に属するタイル90枚,正常領域に属するタイル90枚を用意してがんらしさを算出した(図3)。正常らしいタイルほどマイナスの値となり,がんらしいタイルほどプラスの値となる。プロトタイプ試作時のがんらしさ分布をみると,がん領域の組織のみが描画されたタイル(赤)はプラスの値を取る。一方で正常領域の組織のみが描画されたタイル(青)は,がん領域のタイルより値は小さいものの,プラスの値を取るタイルが大半であることが確認できる。これは正常領域の組織のみが描画されたタイルの中にがんらしいと判別されてしまうタイルが存在することを意味する。精度向上を目指して,がんらしさ分布を図3左図から図3右図の分布に近づける必要がある。

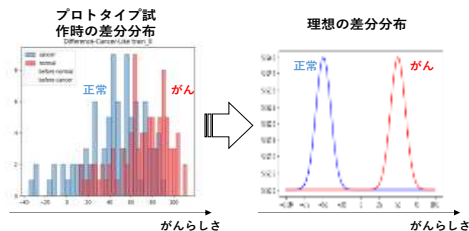


図3 理想とプロトタイプ試作時のがんらしさ分布

図4は最悪基準タイル選択アルゴリズムの実施前後のがんらしさの分布(上図)とがんらしさと正常らしさの差分病理画像(下図)である。最悪基準タイル選択アルゴリズム実施前よりもがん領域領域に属するタイル分布

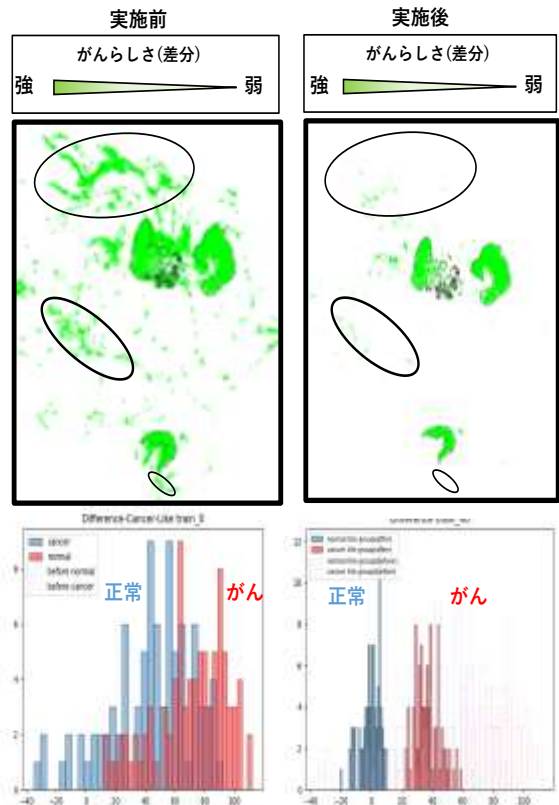


図4 最悪基準タイル選択アルゴリズム実施後のがんらしさ分布と差分病理画像

と正常領域に属するタイル分布が分離されており,精度の向上が確認できる。しかしがんらしさの値については未だに差分値が0以上の正常領域で,誤認識を引き起こす正常タイルが存在している。更に精度の向上を目指すには,現段階で誤認識されているタイルを学習させていく必要がある。

4. 総括

本研究では,機能検証に向けたプロトタイプの試作を行い,実用化に向けた評価を行った。同時に,最悪基準タイル選択アルゴリズムによる計算時間の高速化,精度の向上の可能性を確認した。実用化に向けた課題として,さらなる計算時間の短縮が挙げられる。医師の負担軽減に貢献するには現在の全体の計算時間では不十分であると考える。今後は実用化を念頭に機能検証と改善を行っていく。

参考文献

- [1] Ginneken van B. et al., "CAMELYON16", 2016
- [2] 金道敏樹, 田村一希, 田島和征, 「情報量を用いた画像特徴量解析とそのガン画像診断への応用」, 2020
- [3] 陸田駿弥, 鎌田千晶, 佐藤大樹, 他: 「情報量を用いたガン画像診断—特徴量の特性解析による性能向上の検討—」 情報処理学会第84回全国大会
- [4] 金道敏樹, 陸田駿弥, 山田壮亮: 「病理画像のガンらしさと正常らしさを情報量で評価する情報密度法」, MEDICAL IMAGING TECHNOLOGY (in press)
- [5] 金道敏樹, 「病理画像のがんらしさと正常らしさを情報量で評価する情報密度法」, 第4回医学と数理研究会