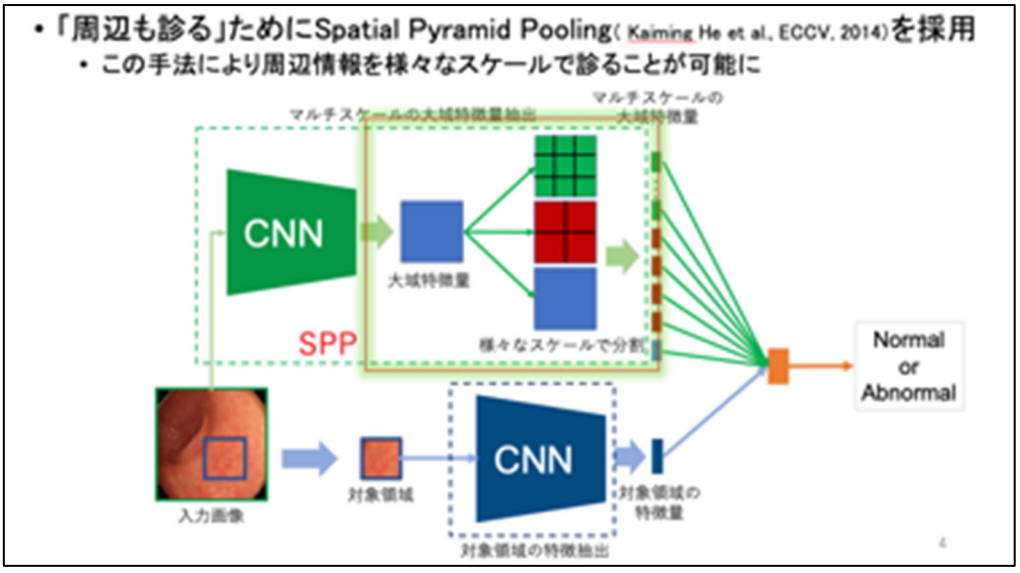
2023 年 3 月 31 日

「胃癌 AI 診断の精度向上」のための研究
【研究成果報告】

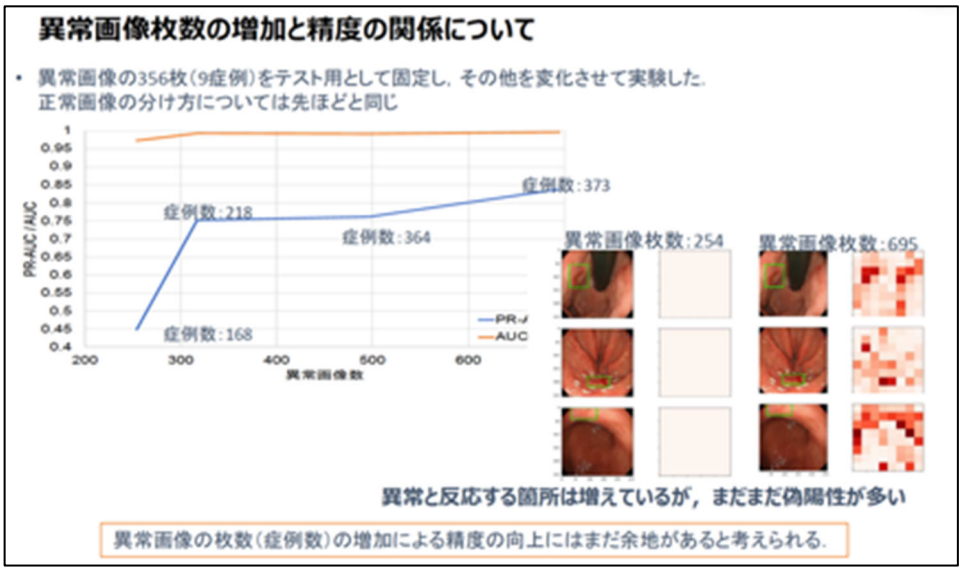
研究の目的:
胃癌には病理学的に多くの種類が存在し、細分類によって対応する治療法が異なることから、AI により胃癌で日常的に遭遇する確率の高い、乳頭腺癌、高分化管状腺癌、中分化管状腺癌、低分化腺癌、印環細胞癌、粘液癌の 6 種類について細分類する試みを行う。

研究期間:2018 年 10 月から 2023 年 3 月 31 日まで

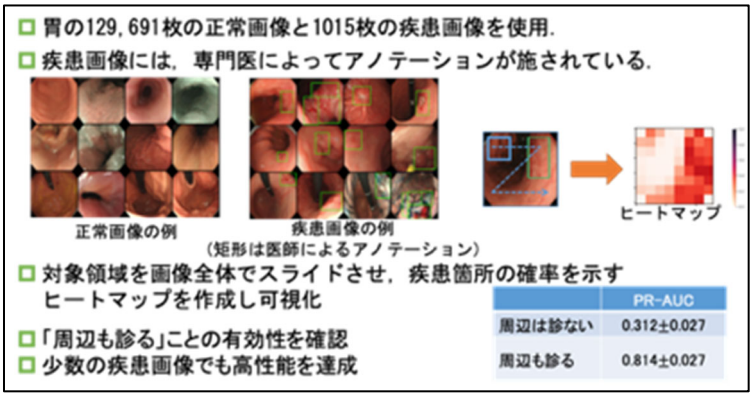
研究経過:
2018 年度は、倫理承認済み 2 施設から、胃の多様性画像 130,742 枚(正常画像 3,346 症例分 129,691 画像と異常画像 382 症例分 1,051 枚)を NII(東京大学 AI 研究チーム)に提供した。
東京大学 AI 研究チームとの協議により、まずは AI による正常、異常の識別を試みることとし、ネットワークの開発(下図)を開始した。



2019 年度は東京大学 AI 研究チームにより、上記ネットワークを利用した胃癌の検出と AI による深層学習による再分類正解率を評価した。正常画像・異常画像の 80%の症例を学習用に、20%の症例 をテスト用に分割し、医師と同じく周辺領域も観る Spatial Pyramid Pooling(Kaiming He et.al, ECCV, 2014)を採用し(下図)、AI による異常検出分類を 3 回実施した。初期は識別率 PR-AUC で 0.596±0.029 であったが、正常画像より異常画像は得られる量が圧倒的に少ないため、このクラス間バランスが AI 識別精度に大きな影響を与えることが解った。
そのため正常画像と異常画像の枚数が同程度となるように、生成モデル(GAN)を用いて異常画像を大量生成し、昨年度開発したネットワークを使用し、生成画像を含む場合と含まない場合で精度評価を実施し比較した。その結果、生成画像を含むことにより識別精度が向上することが確認された。(下図)



さらに、周辺領域を観る大域・局所特徴量の層を深くした結果、0.814±0.027 まで精度が大きく向上した。



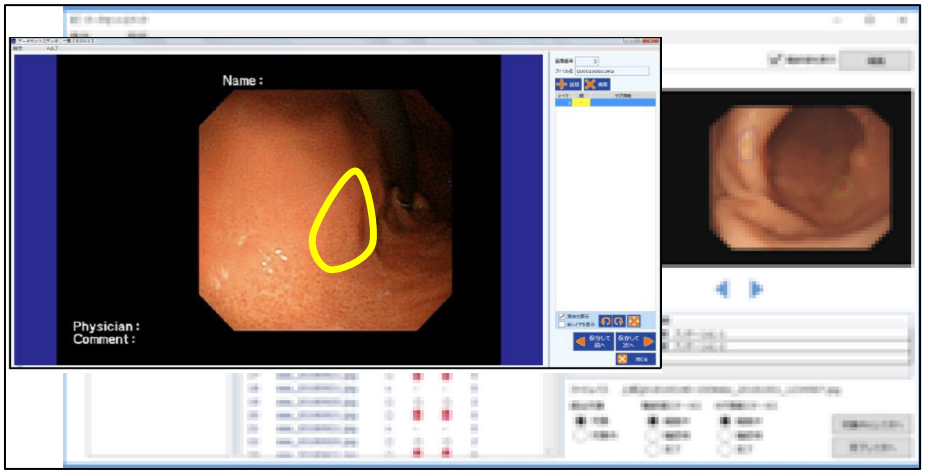
また、AI による胃がんの TYPE 識別の準備段階として、研究参加 39 施設で病変部にアノテーションを加えた胃がん画像と肉眼型所見、病理診断 TXT のデータセット作成を行った。最終的に 1,752 画像のデータセット（下図）を生成し、東京大学 AI 研究チームへ提供した。

胃がん TYPE 診断 AI 学習用データセット

病理診断 TXT（例）

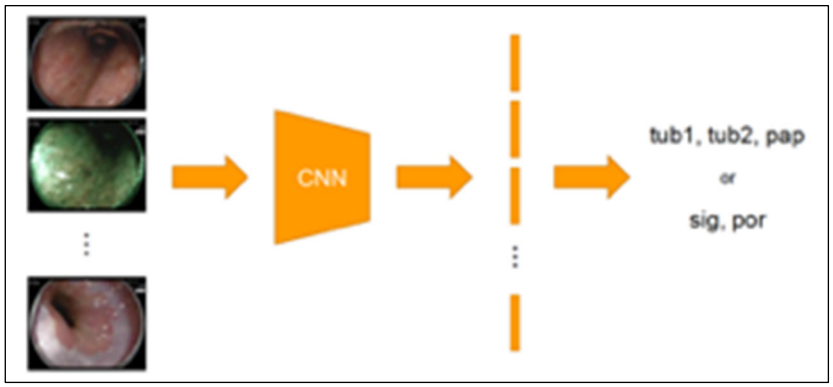
性別	年齢 歳 【手入力】	主部位 1	主部位 2	内視鏡診断（精査時）				病理組織診断（内視鏡治療もしくは外科手術検体）							
				術前主肉眼型	術前深達度	術前UL	術前大きさ(長径) mm 【手入力】	病理主肉眼型	病理深達度	病理UL	病理大きさ(長径) mm 【手入力】	主組織型	副組織型	リンパ管侵襲	静脈侵襲
男	83	M	小弯	0-IIc	SM1	なし	15	0-IIc	M	なし	15	por	sig	なし	なし

点描画アノテーションツール「データセットエディタ」（日本消化器内視鏡学会）



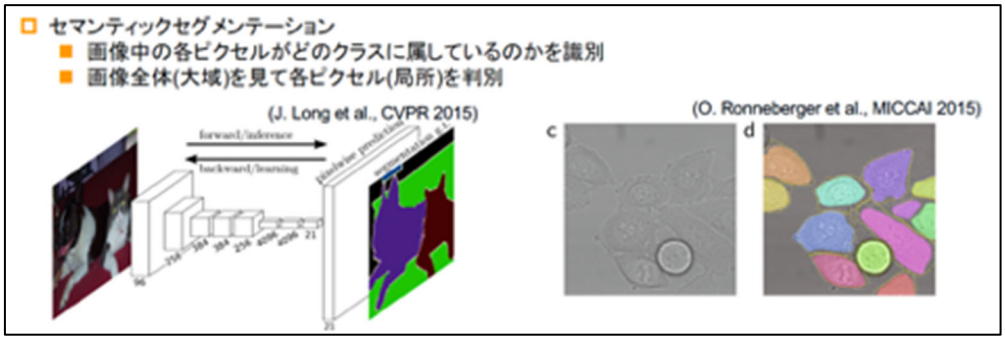
2020 年度は、東大 AI チームで 1,752 画像のデータセットをもとに、トライアルとして腺管を形成している癌（sig, por）と形成していない癌（tub1, tub2, pap）の二値分類を試みた。（下図）しかしながら正答率は低調であった。

主組織系判定の初期検討モデル （東京大学　黒瀬先生資料より）



初期モデルは、病変検出に物体検出アルゴリズムを適用してきたが、内視鏡画像の病変部は多様な姿かたちをしており、そのような不定形の物体を扱うことを苦手としている点から、周辺領域を観る大域・局所特徴量の層を深く見る方式から、局所識別手法に切り替えると同時に病変検出をセグメンテーションアルゴリズムに変更を試みた。

その結果、以前より高い精度（0.5633→0.6378）を得ることができた。（下図）



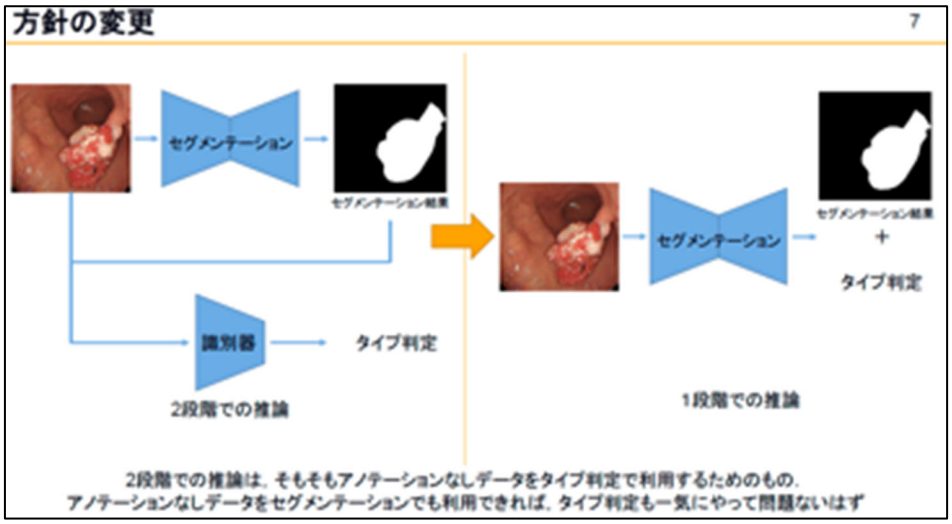
Grid Level AP (PR-AUC)						
Fold	1	2	3	4	5	平均
Deeplab v3+	0.4538	0.6685	0.6572	0.7179	0.6918	**0.6378**
[A. Hayakawa, CARS, 2019]	0.4319	0.5524	0.5566	0.6530	0.6229	**0.5633**

2021 年度は、さらに病変検出と腺管ありなしの2値分類精度を高めるために、今までに判明した点から、これまでタイプ識別に画像をそのまま利用している。しかし画像上は、がんの部分よりも正常部が支配的であるので、ネットワークががんの部位か正常部かの判別ができておらず、タイプ識別が難しくなっている可能性がある。

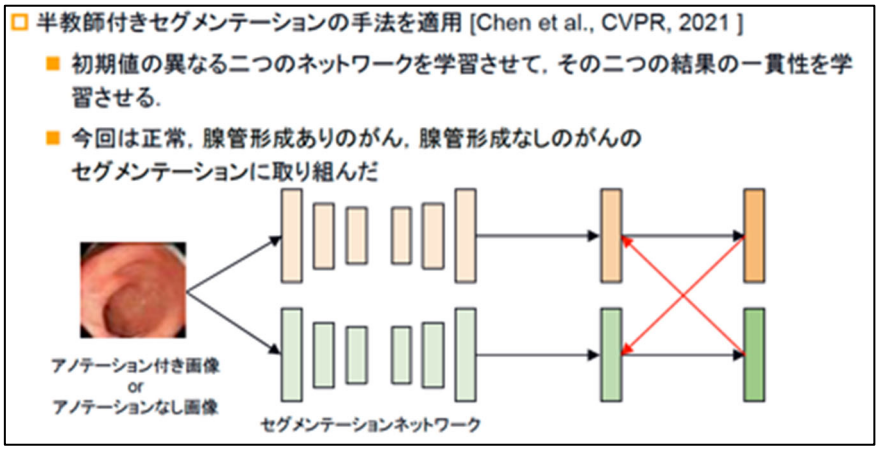
よって、今まで通りアノテーション付き画像のみを利用したとしても、精度向上は難しいのではないか。

アノテーション付き同一病変フォルダ内には、アノテーションなし画像も大量にあり、それを利用できないか。

以上の仮説を立て、これまでの教師付き学習から半教師付き学習を行う点、これまで病変部位判定と腺管ありなしのタイプ判定の 2 段階判定から、部位とタイプを一気に行うように方針変更を行った。(下図)



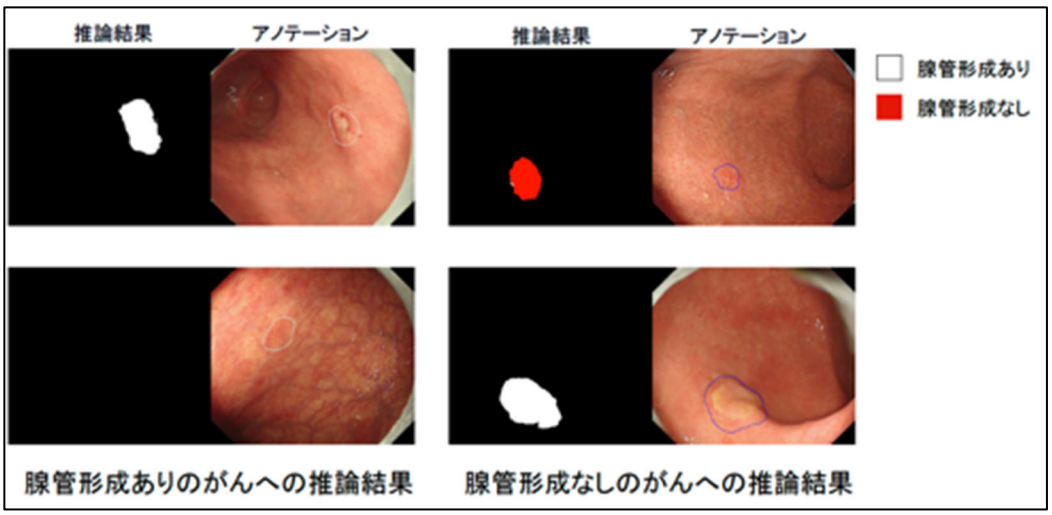
AIネットワークはセグメンテーションの手法[Chen et.al, CVPR, 2021]を適用した。（下図）



評価の結果（下図）、アノテーションを半分しか利用していない結果の方がよく見えるが、試行ごとに精度が全く安定しておらず、評価が難しいという判断となった。

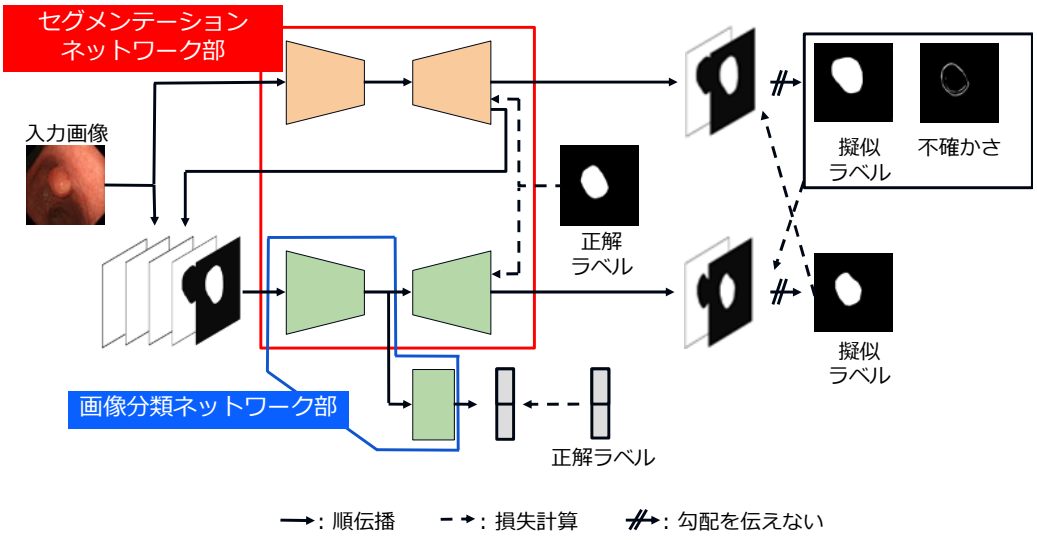
mIoU / mean pixel accuracy / classification accuracy
いずれも%表記

	mIoU			classification acc (参考)	
	正常	癌（腺管形成あり）	癌（腺管形成なし）	癌（腺管形成あり）	癌（腺管形成なし）
all	94.014	19.146	6.914	64.623	27.272
half	94.598	33.332	4.613	77.359	68.182



2022 年度は、病変部位判定と腺管ありなしのタイプ判定を同時に行った結果を踏まえ、考えられる方策を検討した。教師データの中身を見直したところ、IIa,IIc の判定が難しいセットも混在しており、これが AI の精度に影響を与えているのではと推論し、データセットをチェックし判定が難しいものを手作業により除外することとした。

また、AI 側も病変のセグメンテーションと同時に正常、非正常（IIa,IIc）を判定する改良を加えた。除外後のデータセットを学習、評価、試験用に9:0.5:0.5 の割合で分割し、改良を加えた AI ネットワーク（下図）で評価した。



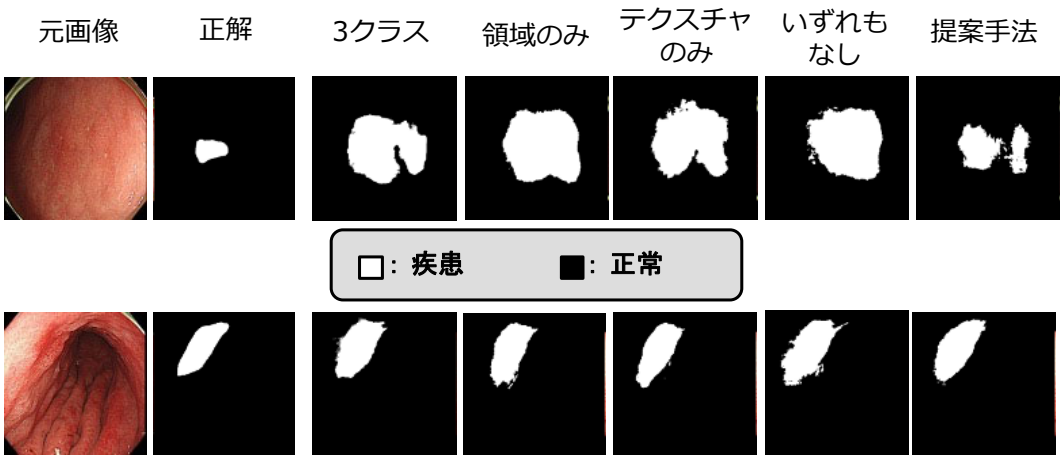
このネットワークは、アノテーションの付与されたものとされていないものを同時に扱って学習する半教師付き学習の枠組みであり、胃がんの有無を検知するセグメンテーションを行うネットワークとIIa,IIc の画像分類を行うネットワークから構築されている。

実験は、結果としては、従来手法をベースにしたIIa,IIc 及び正常箇所を直接セグメンテーションで分類するものと、提案手法でセグメンテーションネットワークの出力のみを利用したもの、画像分類の際に細かい特徴のみを使ったもの、いずれも使わなかったもの、提案手法の5種類で比較実験を行った。分類の精度は依然高くはない（下図）状態であった。

胃がんの有無を判定するセグメンテーションの結果

	0-IIa	0-IIc	疾患部計
3クラスで直接セグメンテーション	0.1228±0.0515	0.3102±0.0462	0.6086±0.0037
領域のみ	-	-	0.6158±0.0069
テクスチャのみ	-	-	0.5973±0.0436
いずれもなし	-	-	0.6050±0.0153
領域 + テクスチャ (提案手法)	-	-	**0.6261± 0.0102**

セグメンテーション結果の例



IIa,IIc の判定結果

	F1 値	AUROC
領域のみ	0.7230±0.0342	0.7245±0.0103
テクスチャのみ	0.7210±0.1028	0.7033±0.0538
いずれもなし	0.6968±0.0972	0.7205±0.0397
領域 + テクスチャ (提案手法)	**0.7596±0.0296**	**0.7563±0.0295**

以上が、これまでに得られた結果であり依然として精度は高くはないが、今回の AI 識別のネットワークが半教師付き学習であり、アノテーション付きのデータを増加させ教師付き学習として解くことができればまだ精度は上がる可能性も残されている。

最後に期待通りの高い精度が得られなかったことは残念ではありますが、これまでデータの取得およびアノテーションにご尽力いただいた先生方に感謝いたします。

注）本報告書中の図表は、東京大学　情報理工学研究科　原田研究室　黒瀬 優介 先生の資料より抜粋しております。また、別紙の最新打ち合わせ資料につきましても、本研究の研究者が研究遂行を目的として使用許可を頂いておりますので、取扱いにご注意をお願い申し上げます。

【研究参加施設】

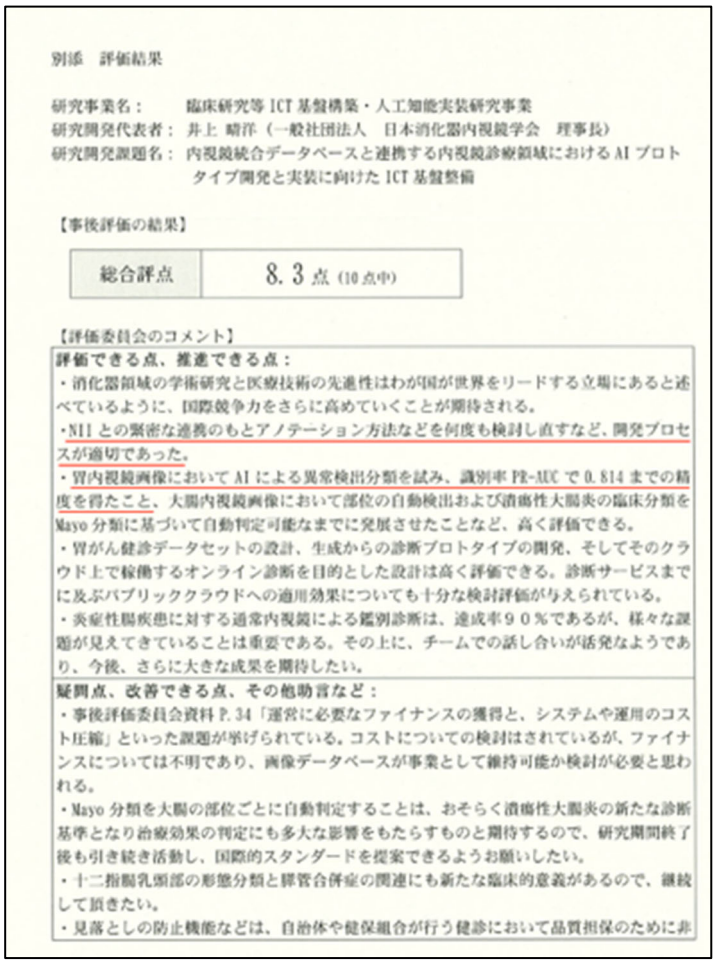
京都第二赤十字病院、福井県立病院、昭和大学藤が丘病院、県立静岡がんセンター、長崎みなとメディカルセンター、国立国際医療研究センター国府台病院、国立大学法人広島大学病院、大阪府立病院機構大阪急性期・総合医療センター、愛媛大学医学部附属病院、金沢大学附属病院、春回会井上病院、大阪市立大学医学部附属病院、弘前大学医学部附属病院、東京慈恵会医科大学葛飾医療センター、国立大学法人東北大学病院、近畿大学医学部附属病院、慶應義塾大学病院、大阪市立総合医療センター、石川県立中央病院、市立豊中病院、和歌山県立医科大学附属病院、順天堂大学医学部附属順天堂医院、山口県厚生農業協同組合連合会周東総合病院、独立行政法人国立病院機構函館病院、国立大学法人群馬大学医学部附属病院、名古屋大学医学部附属病院、愛知県がんセンター、公立大学法人福島県立医科大学附属病院、国家公務員共済組合連合会斗南病院、国家公務員共済組連合会虎の門病院、国立研究開発法人国立がん研究センター東病院、国立がん研究センター中央病院、筑波大学附属病院、旭川医科大学病院、独立行政法人労働者健康安全機構関西ろうさい病院、東京都立墨東病院、東京大学医学部付属病院、国立大学法人神戸大学医学部附属病院、山口大学医学部附属病院

計39施設

【AMED 事後評価結果】

先生方のご尽力のおかげで、NII を含めた参加 6 学会の中で最高クラスの評価をいただきました。

（本研究に関するコメントは下記赤線の通りです）



別添　評価結果

研究事業名：　臨床研究等 ICT 基盤構築・人工知能実装研究事業
研究開発代表者：井上 晴洋（一般社団法人　日本消化器内視鏡学会　理事長）
研究開発課題名：内視鏡統合データベースと連携する内視鏡診療領域における AI プロトタイプ開発と実装に向けた ICT 基盤整備

【事後評価の結果】

総合評点	8.3 点（10 点中）

【評価委員会のコメント】

評価できる点、推進できる点：

・消化器領域の学術研究と医療技術の先進性はわが国が世界をリードする立場にあると述べているように、国際競争力をさらに高めていくことが期待される。

・NII との緊密な連携のもとアノテーション方法などを何度も検討し直すなど、開発プロセスが適切であった。

・胃内視鏡画像において AI による異常検出分類を試み、識別率 PR-AUC で 0.814 までの精度を得たこと、大腸内視鏡画像において部位の自動検出および潰瘍性大腸炎の臨床分類を Mayo 分類に基づいて自動判定可能なまでに発展させたことなど、高く評価できる。

・胃がん健診データセットの設計、生成からの診断プロトタイプの開発、そしてそのクラウド上で稼働するオンライン診断を目的とした設計は高く評価できる。診断サービスまでに及ぶパブリッククラウドへの適用効果についても十分な検討評価が与えられている。

・炎症性腸疾患に対する通常内視鏡による鑑別診断は、達成率９０％であるが、様々な課題が見えてきていることは重要である。その上に、チームでの話し合いが活発なようであり、今後、さらに大きな成果を期待したい。

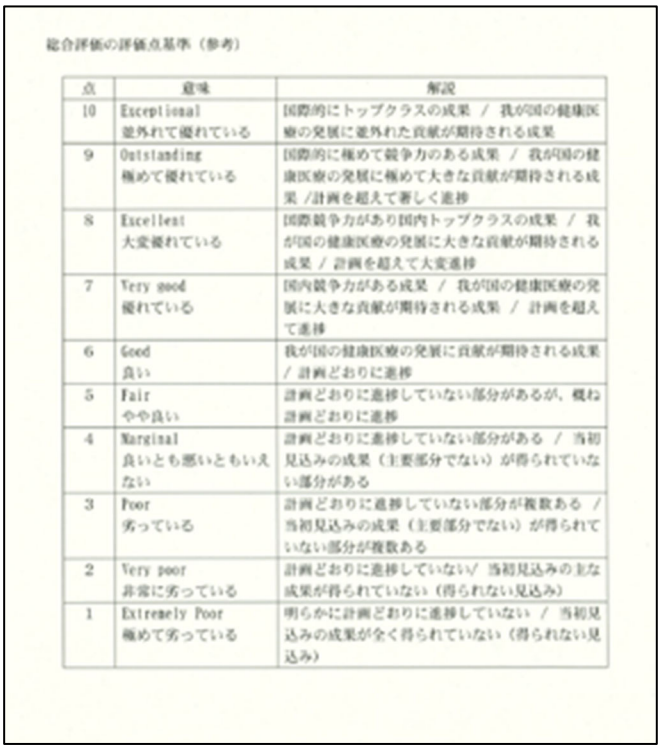
疑問点、改善できる点、その他助言など：

・事後評価委員会資料 P. 34「運営に必要なファイナンスの獲得と、システムや運用のコスト圧縮」といった課題が挙げられている。コストについての検討はされているが、ファイナンスについては不明であり、画像データベースが事業として維持可能か検討が必要と思われる。

・Mayo 分類を大腸の部位ごとに自動判定することは、おそらく潰瘍性大腸炎の新たな診断基準となり治療効果の判定にも多大な影響をもたらすものと期待するので、研究期間終了後も引き続き活動し、国際的スタンダードを提案できるようお願いしたい。

・十二指腸乳頭部の形態分類と膵管合併症の関連にも新たな臨床的意義があるので、継続して頂きたい。

・見落としの防止機能などは、自治体や健保組合が行う健診において品質担保のために非



総合評価の評価点基準（参考）

点	意味	解説
10	Exceptional 並外れて優れている	国際的にトップクラスの成果 / 我が国の健康医療の発展に並外れた貢献が期待される成果
9	Outstanding 極めて優れている	国際的に極めて競争力のある成果 / 我が国の健康医療の発展に極めて大きな貢献が期待される成果 /計画を超えて著しく進捗
8	Excellent 大変優れている	国際競争力があり国内トップクラスの成果 / 我が国の健康医療の発展に大きな貢献が期待される成果 / 計画を超えて大変進捗
7	Very good 優れている	国内競争力がある成果 / 我が国の健康医療の発展に大きな貢献が期待される成果 / 計画を超えて進捗
6	Good 良い	我が国の健康医療の発展に貢献が期待される成果 / 計画どおりに進捗
5	Fair やや良い	計画どおりに進捗していない部分があるが、概ね計画どおりに進捗
4	Marginal 良いとも悪いともいえない	計画どおりに進捗していない部分がある / 当初見込みの成果（主要部分でない）が得られていない部分がある
3	Poor 劣っている	計画どおりに進捗していない部分が複数ある / 当初見込みの成果（主要部分でない）が得られていない部分が複数ある
2	Very poor 非常に劣っている	計画どおりに進捗していない/ 当初見込みの主な成果が得られていない（得られない見込み）
1	Extremely Poor 極めて劣っている	明らかに計画どおりに進捗していない / 当初見込みの成果が全く得られていない（得られない見込み）

以上