

画像診断と治療支援分野における 生成AIの最前線

森 健策

名古屋大学大学院情報学研究科・教授

名古屋大学情報基盤センター長

国立情報学研究所医療ビッグデータ研究センター長

kensaku@is.nagoya-u.ac.jp



COI

- **共同研究費**

- モリタ製作所
- 朝日インテック

- **知財権実施**

- サイバネットシステム
- モリタ製作所

- **コンサルティング**

- 三友工業株式会社

自己紹介 森 健策

• 学歴

- 1992年 名古屋大学工学部電子工学科卒
- 1994年 名古屋大学大学院工学研究科博士前期課程修了
- 1996年 名古屋大学大学院工学研究科博士後期課程修了

• 職歴

- 1996年 日本学術振興会特別研究員PD
- 1997年 名古屋大学大学院工学研究科助手
- 2000年 名古屋大学大学院工学研究科講師
- 2001年 名古屋大学難処理人工物研究センター助教授
- 2001年 スタンフォード大学脳神経外科客員助教授
- 2003年 名古屋大学大学院情報科学研究科助教授
- 2009年 名古屋大学情報連携推進本部教授
- 2016年 名古屋大学情報基盤センター長
- 2017年 名古屋大学大学院情報学研究科教授
- 2021年 国立情報学研究所医療ビッグデータ研究センター長
- 他にも 名古屋大学情報セキュリティ室長
東海国立大学機構情報戦略室長





CARS

Computer Assisted Radiology and Surgery

CARS 2026



開催期間

2026年7月2日（木）～7月5日（日）



大会長

森健策（名古屋大学）



会場

名古屋大学 豊田講堂・シンポジオン

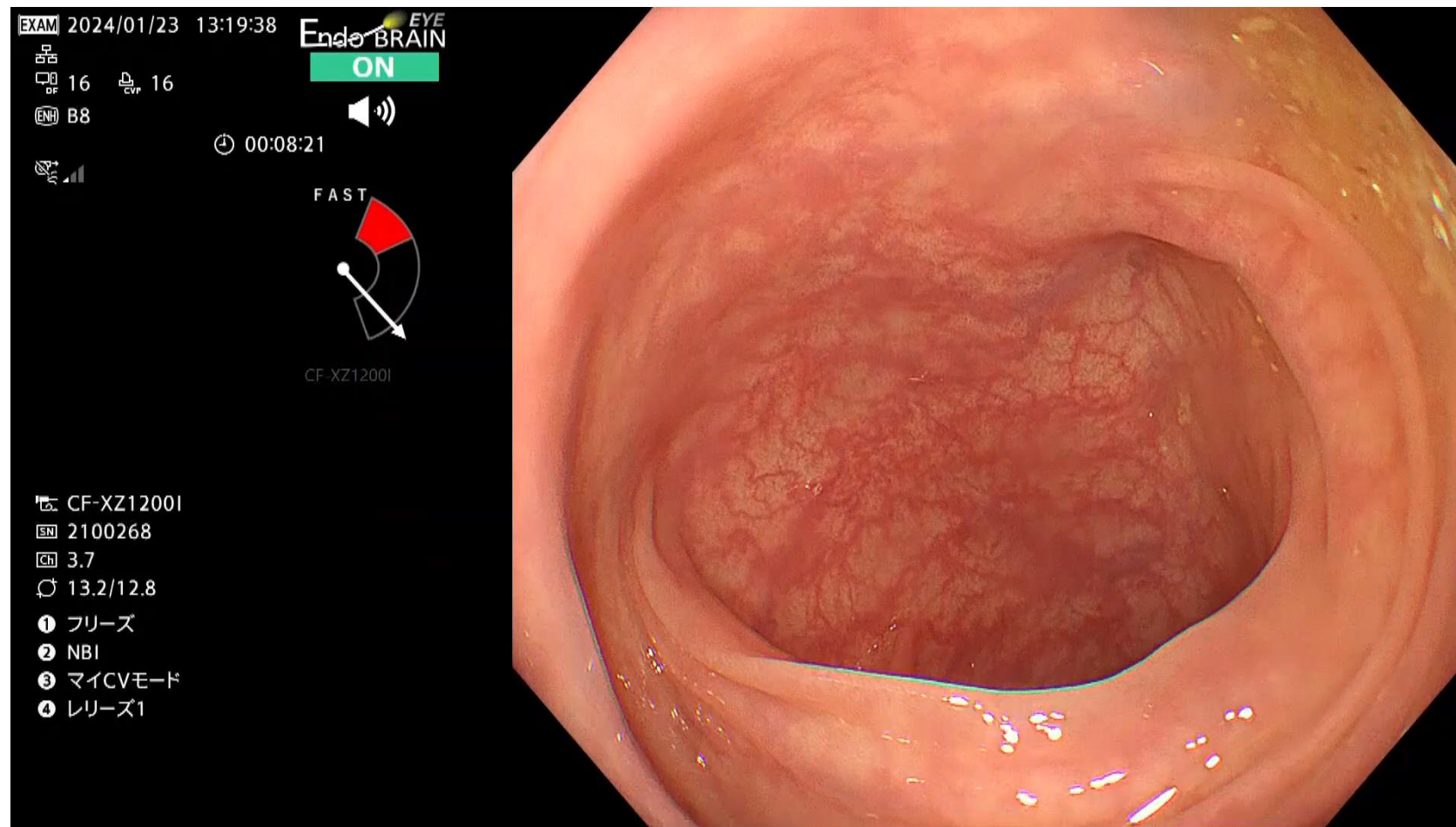


CARS2026内で開催される会議

CAR, ISCAS, CAD, CMI, IPCAI



自己紹介代わりに EndoBRAIN




初AI薬事承認（あらたなカテゴリ） IDATEN適用 診療報酬（0点が60点）

自己紹介代わりに 生成AI



国立研究開発法人 日本医療研究開発機構
Japan Agency for Medical Research and Development



[English Site](#)

[調達情報](#)
[採用情報](#)
[公開情報](#)
[アクセス](#)
[問合せ先](#)
[メルマガ登録](#)

[AMEDについて](#) | [事業紹介](#) | [事業成果](#) | [公募情報](#) | [事務手続](#) | [ニュース](#) | [イベント](#) | [刊行物](#)

[トップ](#) > [ニュース](#) > [プレスリリース](#) > AIを搭載した内視鏡診断支援プログラムが承認—医師の診断補助に活用へ—

プレスリリース

AIを搭載した内視鏡診断支援プログラムが承認—医師の診断補助に活用へ—

[プレスリリース](#)

学校法人 昭和大学
国立大学法人 名古屋大学
国立研究開発法人 日本医療研究開発機構

ポイント

- 大腸の早期がんや前がん病変である腫瘍性ポリープを内視鏡をつかって早期発見・早期治療することは、大腸がん死亡の抑制効果があるとされています。
- ポリープの中には切除する必要がある腫瘍性ポリープと、切除する必要がない腫瘍ではないポリープ（非腫瘍性ポリープ）があり、医師は検査中に両者を的確に判別する必要があります。
- 昭和大学横浜市北部病院消化器センターの工藤進英教授らのグループは、内視鏡画像を人工知

ニュース

- [プレスリリース](#)
- [成果情報](#)
- [トピックス](#)
- [事業に関するお知らせ](#)



NU Research Information
名古屋大学 研究成果発信サイト

[JP](#)

If you don't have time to search for your NIFTIs, go to the bottom of this page, I have prepared for you.



所見レポートを作成してください。

了解しました。CT画像を確認し、所見レポートを作成いたします。

右肺上葉のすりガラス影は前回より改善しています。右肺上葉のすりガラス影は前回同様です。左下葉の濃厚影は前回よりやや縮小。左下葉の濃厚影は前回同様です。左下葉の濃厚影は前回同様です。左下葉の濃厚影は前回同様です。炎症後変化を疑います。背景肺に気腫性変化あり。胸水なし。有意なサイズのリンパ節腫大なし。肝臓癌あり。

再発はありますか

再発の兆候は見られません。ただし、左下葉の濃厚影は前回と比較してやや縮小していますが、持続しているため、定期的なフォローアップが必要です。炎症後の変化である可能性が高いですが、再発の可能性も完全には否定できません。主治医の先生と相談の上、適切なフォローアップ計画を立てることをお勧めします。

会話履歴を削除して新しい会話を開始

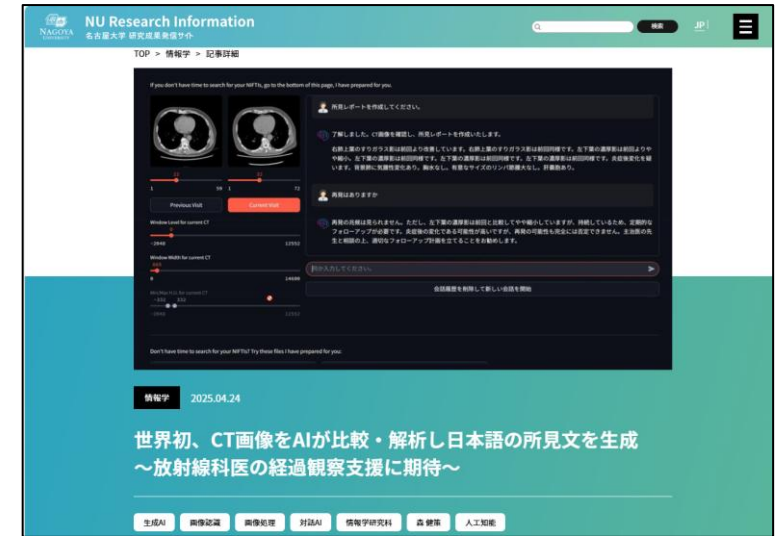
Don't have time to search for your NIFTIs? Try these files I have prepared for you:

情報学 2025.04.24

世界初、CT画像をAIが比較・解析し日本語の所見文を生成 ~放射線科医の経過観察支援に期待~

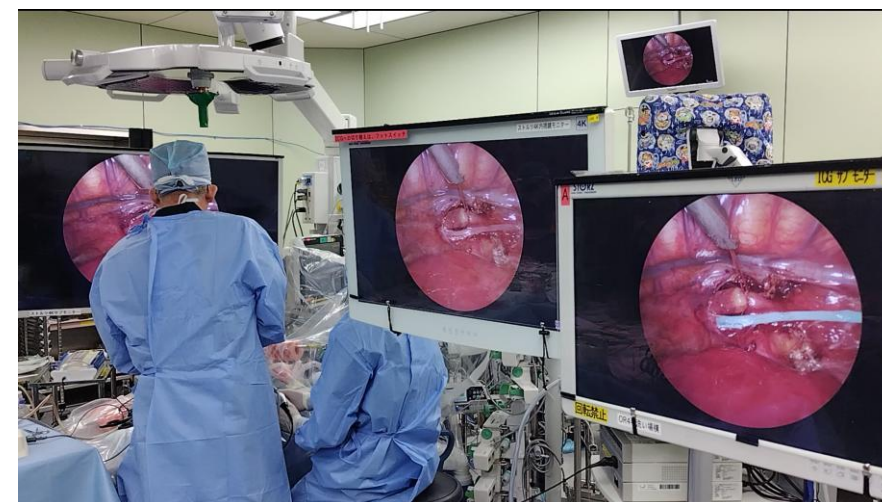
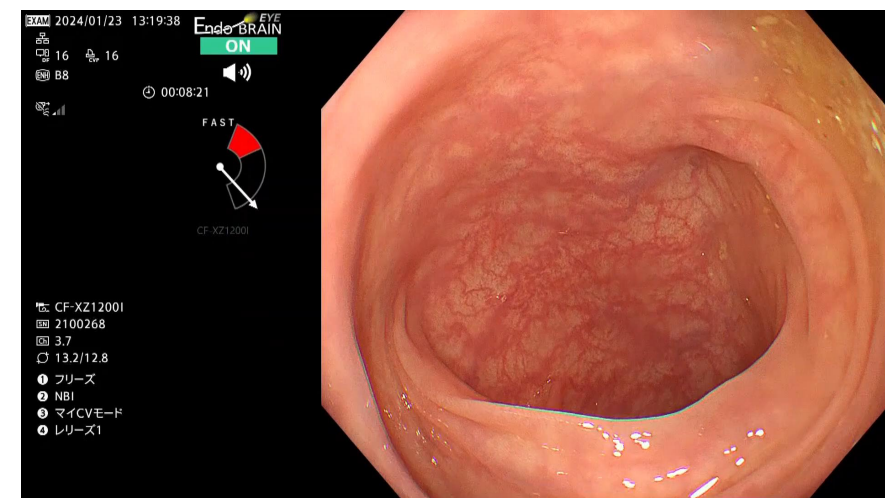
医療分野における生成AIによる自動化

- 医療分野の自動化
 - 診断
 - 治療
 - 医療情報・医事
 - 投薬
 - カウンセリング
 - オーダリング
 - 病院案内・診察受付
- 様々な分野で自動化が試みられている
 - 特に医療情報・医事関係はSIPにおける重要課題
 - 第2期 SIP AIホスピタル (PD 中村先生)
 - 第2期 統合ヘルケアシステム (PD永井先生)
(LLMによる医療情報処理)

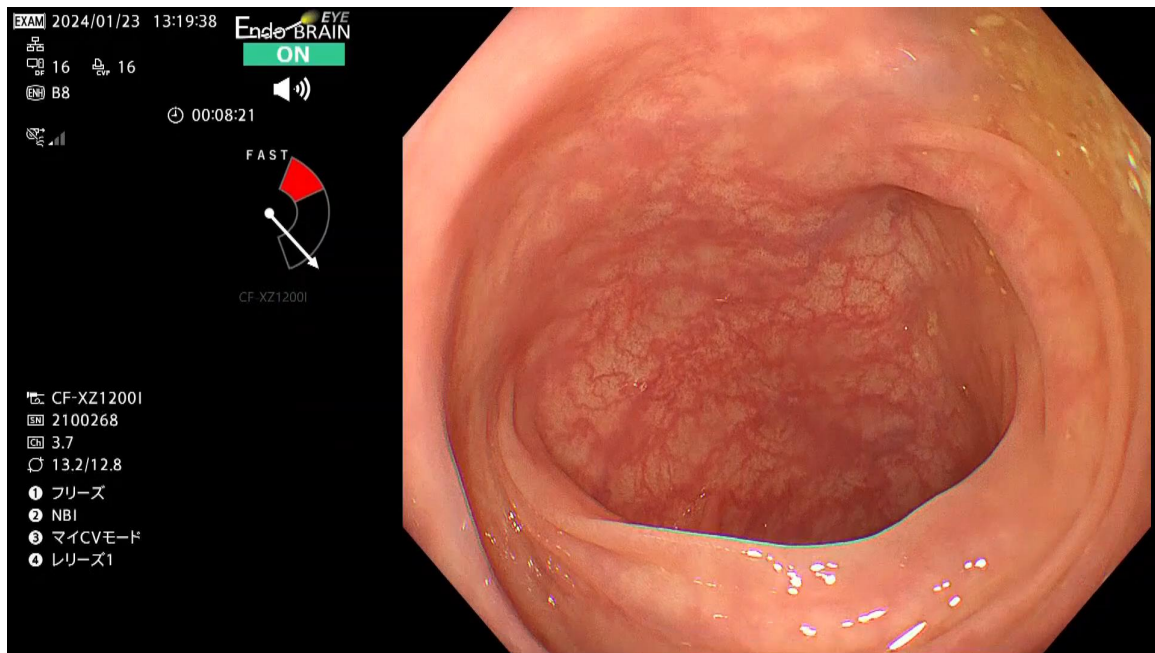


生成AI系の急激な進化

- **AIの進化**
 - 認識AI → 生成AI → 対話AI → 身体性AI
- **古典的なAI医療機器**
 - 画像から異常部位を見つける
(さかのぼれば1960年代)
 - ロボットがターゲットを追いかける
- **これからのAI医療機器**
 - 情報を生み出す
 - コンテキストを理解する
 - 身体性を持つ
 - メディカルスタッフと対話する
 - ジョークも言う？
 - 単一目的から複数目的/包括目的



画像診断支援システム出力の劇的変化



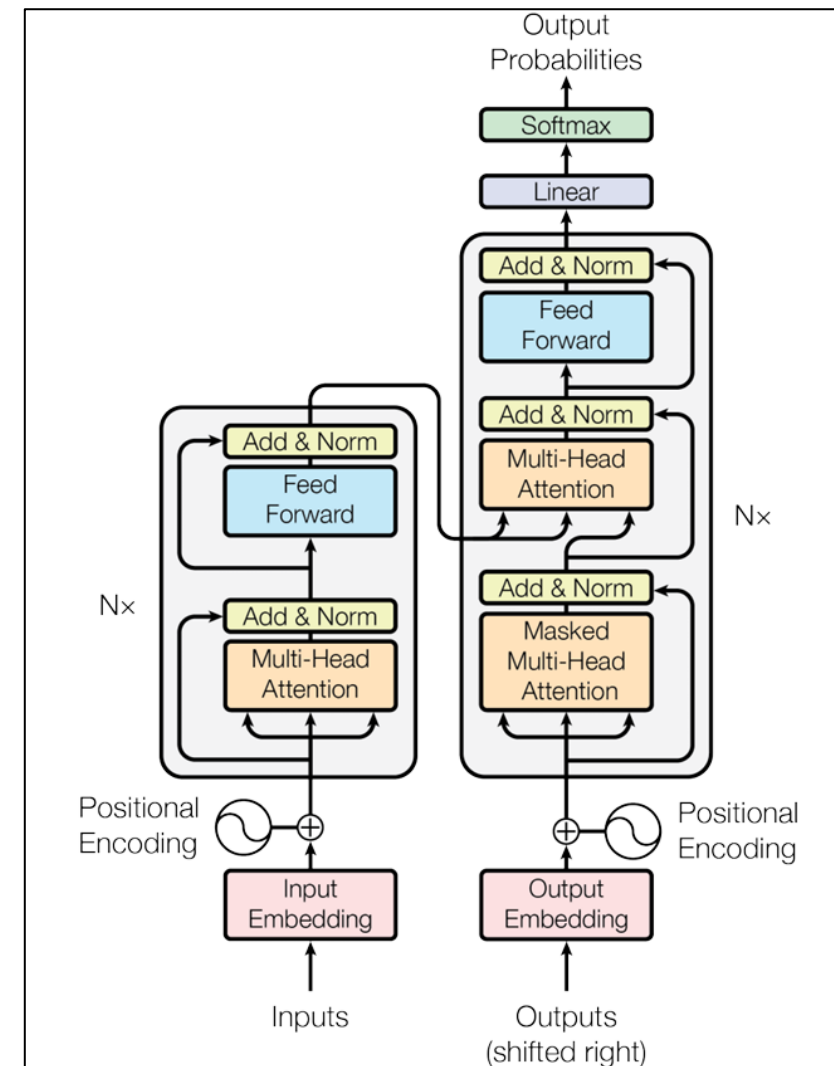
従来SaMD例
病変部位を提示



生成AI SaMD例
総合的な診断

生成AI

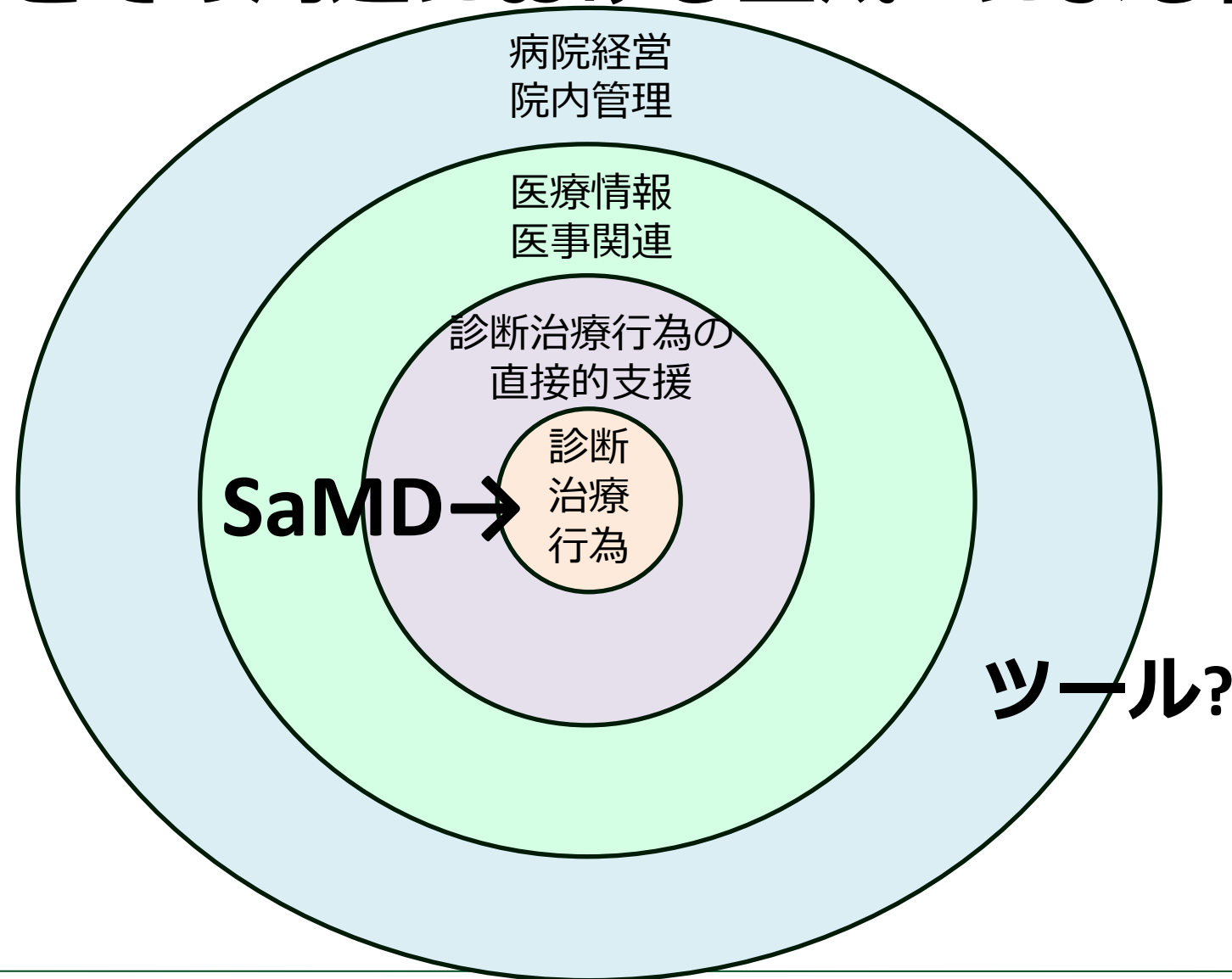
- Generative AI
 - ChatGPTなどの対話型生成AIによって注目
 - 世界が大きく変わるきっかけ (開発、製品、生活、投資、政策など)
 - 言語による対話、マルチモーダルデータ処理
 - 事前学習されたTransformerが基本
- Transformerが世界に与えたインパクトは超大
 - 超々大規模データと超々大規模データ計算
- 生成AIは医療機器あるいはその開発においてどのような役割を果たすのか？



生成AIによるSaMD

- **何が生成AIを用いたSaMDか？ (生成AI SaMD定義)**
 - 診断結果を文章として出力するSaMD
 - 対話機能を有するSaMD
 - 対話機能を有する手術ナビゲーション・ロボット (ロボットが理解できるコマンドを作るLMM)
 - 治療方法を自動的に見出すSaMD
- **適用範囲はどこまでか？ (適用範囲)**
 - 法律で定義される医療行為の直接的支援から間接的支援のどこまでか？
- **生成AIの評価基準をどうするか？ (生成AI SaMD評価)**
 - 生成されたテキストの評価
 - 総合的診断への評価
 - ▶ 例 ポリープ発見支援とは明らかに異なる
- **生成AI開発で用いるデータは？ (学習データ)**
 - どのデータを用いて訓練したのか？ そのデータに対する同意は？
 - ▶ 特に事前学習に用いたデータ
- **生成AI特有の自信満々の誤りを生み出す可能性があるSaMDへの対応は？ (ハルシネーション)**
 - 誤りをどこまで許容するのか？ (見落とし・拾いすぎの議論を超えた対応)

診断治療とその周辺における生成AIによる自動化



診断分野における自動化

- 古くは1950年代から研究
 - Nash FA. Differential diagnosis: an apparatus to assist the logical faculties. Lancet 1954;263:874–875, p. 874
 - DJ Croft, Is computerized diagnosis possible?, 1972
 - G LODWICK, Computer Diagnosis of Primary Bone Tumors - A Preliminary Report, Radiology, 1963 PMID: PMC2464549
 - EH. Shortliffe, MYCIN: A KNOWLEDGE-BASED COMPUTER PROGRAM APPLIED TO INFECTIOUS DISEASES, Proc Annu Symp Comput Appl Med Care. 1977,
 - 鳥脇 胸部X線自動診断, 1960s-
 - 土井 マンモグラフィ, late 1980s-
 - IBM Syste360のころから自動診断の試み (名大はFACOM 230-60)
- 自動診断、診断支援、そして、また自動診断へと
 - 生成AIによる自動診断手法の高度化
 - 文章を理解し、文章を生成できるようになった

電子計算機による胸部X線写真の病巣陰影識別に関する基礎的実験

鳥脇 純一郎, 福村 晃夫, 小池 和夫, 高木 良雄
 医用電子と生体工学, 1968年6巻3号 p. 207-214

何をもってAIとするかの議論は別として コンピュータによる画像診断支援はこのころから

研究

(207)

電子計算機による胸部X線写真の病巣陰影識別に関する基礎的実験*

鳥脇純一郎**・福村 晃夫**・小池 和夫***・高木 良雄***

Automatic Recognition of Abnormal Shadows in Chest Roentgenograms*

Jun-ichiro TORIWAKI**, Teruo FUKUMURA**, Kazuo KOIKE, M.D.***, Yoshio TAKAGI, M.D.***

Computer diagnosis of the roentgenogram is regarded as a typical two-dimensional pattern recognition problem. As the first step to approach the problem, we investigated the characteristics of the density distribution of the chest roentgenograms. Then we studied the method of automatic recognition of rib boundaries.

As the next step, we studied in this paper the method of automatic recognition of abnormal shadows in chest roentgenogram.

First, simple models of abnormal shadows in chest roentgenogram consisting of (1) some irregular figures whose uniform densities are known and whose shapes, sizes and positions are unknown, and (2) additive Gaussian noise with 0 mean and known variance, are assumed and recognition method is derived. Only outline of the method is stated here.

Next, this method is applied to the recognition of abnormal shadows in the real chest roentgenogram. In this experiment, input patterns (roentgenograms) are processed according to the following procedure: (i) measurement and sampling of film density, (ii) recognition of boundaries of dorsal portion of ribs, (iii) subtraction of density corresponding to dorsal portion of ribs, (iv) equalization of average density level in the horizontal direction, (v) recognition of boundaries of ventral portion of ribs, (vi) subtraction of density corresponding to ventral portion of ribs, (vii) recognition of abnormal shadows (tuberculous lesions).

Results of the experiment are stated here in detail. They show that the method used in this paper was proved useful for detection of some kinds of abnormal shadows in chest roentgenograms.

1. ま え が き

胸部X線写真の電子計算機による処理は典型的なパターン認識の問題である。そして、この場合に対象とするパターンがきわめて複雑であり、しかも性質が互いに異なる成分から構成されていることから、各成分に適合した基本的処理方法を開発することがまずはじめに必要とされる。これら基本的処理方法の第一として肋骨境界線の識別法の研究がすでに報告されている

1) 本論文はそれにつづくものとして病巣陰影識別の方法について述べたものである。ここで用いられる方法は塊状図形の処理方法であり、将来完成されるべきシステム全体のなかでは、先に肋骨境界線の識別に用いられた線状図形の処理方法²⁾とらんで中心的役割を果たすものである。ただし、本文ではこの部分の処理実験のみを目的としているため、必要な処理の過程がすべて自動化されているわけではない。また、処理方法の導出、能力の解析などの基礎的検討の詳細は別の報告³⁾にゆずり、本文では実際の胸部X線写真に対する識別実験の結果について詳しく述べている。なお、先の報告³⁾にも述べたように病巣陰影識別実験のみが目的とされる場合でも肋骨は識別され除去されなくてはならない。本文の実験ではすでに報告された方法を用いてこの過程を実行しているため、肋骨境界線

* 昭和42年11月27日原稿受付 (Received on Nov. 27th 1967)
 ** 名古屋大学工学部電気工学教室 (Department of Electrical Engineering, Faculty of Engineering, Nagoya University)
 *** 国立療養所梅ヶ丘光風園 (National Sanatorium Ume-no-koikefuen)

電子計算機による胸部X線写真の病巣陰影識別に関する基礎的実験

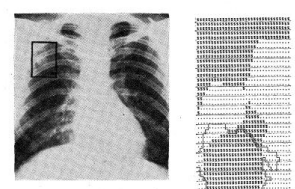
(211)

とおき、 K' には肋骨の濃度の実測値、 r には適当な値を入れて得られる値を用いた。
 また、定数 σ/K (σ は雑音の標準偏差)は適当な一つの試料写真から得られた値をそのまま共通に用いている (これは、すべての写真について S/N 比が大体一定であるとみなしたことになる)。
 この実験では種々の基礎的データをうることを目的としているため手順 II 以下は実行していない。
 さて、各処理の結果を一括して、第4図、第5図、第7図に示す。それぞれについて以下順に述べていこう。

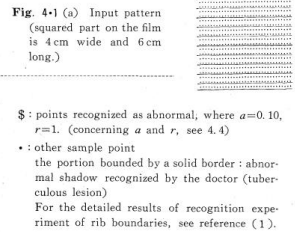
4-5 識別実験の結果(1)——比較的良好な識別が行なわれる場合

まず、第4図に比較的良好な識別が行なわれる場合を示す。両者とも肋骨と重複している部分も含めて大体医師の診断結果と一致した領域を抽出している。以下注意されるべき点をあげておく。

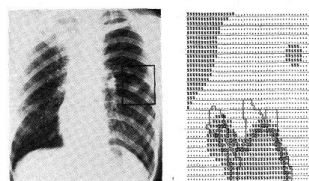
1) 第4-1図 (b) は医師による識別結果がやや左上りの傾向をもつものに対して計算機による結果は右上りの傾向を示す。これは計算機の処理が肋骨境界付近



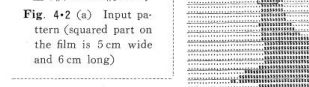
第4-1図 (a) 入力パターン (四角枠はフィルム上で横4cm×縦6cm)
 Fig. 4-1 (a) Input pattern (squared part on the film is 4cm wide and 6cm long.)



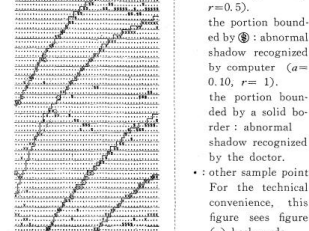
第4-1図 (b) 病巣陰影識別結果
 Fig. 4-1 (b) Result of recognition experiment of abnormal shadows



第4-2図 (a) 入力パターン (四角枠はフィルム上で横5cm×縦6cm)
 Fig. 4-2 (a) Input pattern (squared part on the film is 5cm wide and 6cm long)



第4-2図 (b) 肋骨境界線識別結果 (背部肋骨)
 Fig. 4-2 (b) Result of recognition experiment of rib boundaries (dorsal portion)



第4-2図 (c) 病巣陰影識別結果
 Fig. 4-2 (c) Result of recognition experiment of abnormal shadows

\$: points recognized as abnormal, where $a=0.10$, $r=1$. (concerning a and r , see 4.4)
 * other sample point
 the portion bounded by a solid border: abnormal shadow recognized by the doctor.
 (All curves running through more than five points are shown here.)

第4-2図 (b) 肋骨境界線識別結果 (背部肋骨)
 Fig. 4-2 (b) Result of recognition experiment of rib boundaries (dorsal portion)

の著しく濃度の大きい部分 (これは肋骨消去後もその近くの点より大きい目の値として残ってくる) の影響を受けているためである。

電子計算機による胸部X線写真の病巣陰影識別に関する基礎的実験

(213)

1	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0
2	3	2	0	0	0	0	0	0	0
2	3	4	4	0	0	0	0	0	0
1	1	2	4	8	4	2	1	1	1
0	0	0	0	0	0	4	4	3	2
0	0	0	0	0	0	2	3	2	2
0	0	0	0	0	0	0	1	2	2
0	0	0	0	0	0	0	0	1	1
0	0	0	0	0	0	0	0	0	1

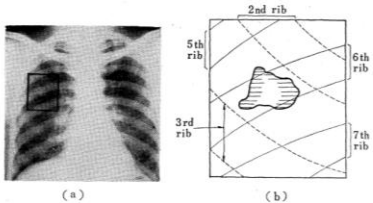
第6図 前胸部肋骨境界線識別に用いられた重み行列
 Fig. 6 Weight matrix employed in the recognition experiment of rib boundaries (ventral portion)

4-7 識別実験の結果(3)——本文の実験では識別されない場合

本実験において計算機では識別されなかった例を第7図に示す。この場合、病巣陰影の濃度は意外に小さく、肋骨消去および水平方向レベル変動補正の誤差と同程度になっている。ただし、この場合には濃度計のわずかな非線形性も影響しているようである。もちろんここに病巣陰影があることは明らかである (これは被験者の位置を変えてこの部分が肋骨と重ならないようにした別の写真で確かめられている)。

4-8 問題点の検討

上述の結果からみて今後解決されるべき問題は非常に多い。ここに述べた方法ではかなり限られた種類の病巣陰影しか扱えないのは別として、一応処理できるパターンのみを考えると、処理方法の点でも



第7図 (a) 入力パターン (四角枠はフィルム上で横5cm×縦6cm)
 (b) (a)のスケッチ
 Fig. 7 (a) Input pattern (squared part on the film is 5cm wide and 6cm long)
 (b) Sketch of the material in (a)

問題となるのは曲線の変形とかレベル変動の補正といったことでは人が行なっている部分の処理法である。前者は一応別の機会に論ずるとして、後者についてはつぎの三つの方向が考えられる。

問題となるのは曲線の変形とかレベル変動の補正といったことでは人が行なっている部分の処理法である。前者は一応別の機会に論ずるとして、後者についてはつぎの三つの方向が考えられる。

- (1) ここで人が行なったことをアルゴリズムを確立して計算機で行なう。
- (2) このような変動成分の影響を受けないような処理方法、あるいはこの成分をより簡単に除去する方法の開発。
- (3) 適当な入力装置を介して計算機と人間の情報交換を容易にし、計算機でやりにくい部分は人間が代行する。

現在(2)の方向の検討を進めているが、一般に、より高度の能力をもつ情報処理システムの開発という方向としては(2)、早い時点で何らかの実用的装置を構成するならば(3)がそれぞれ有用で、かつ興味深い方向ではないかと思われる。

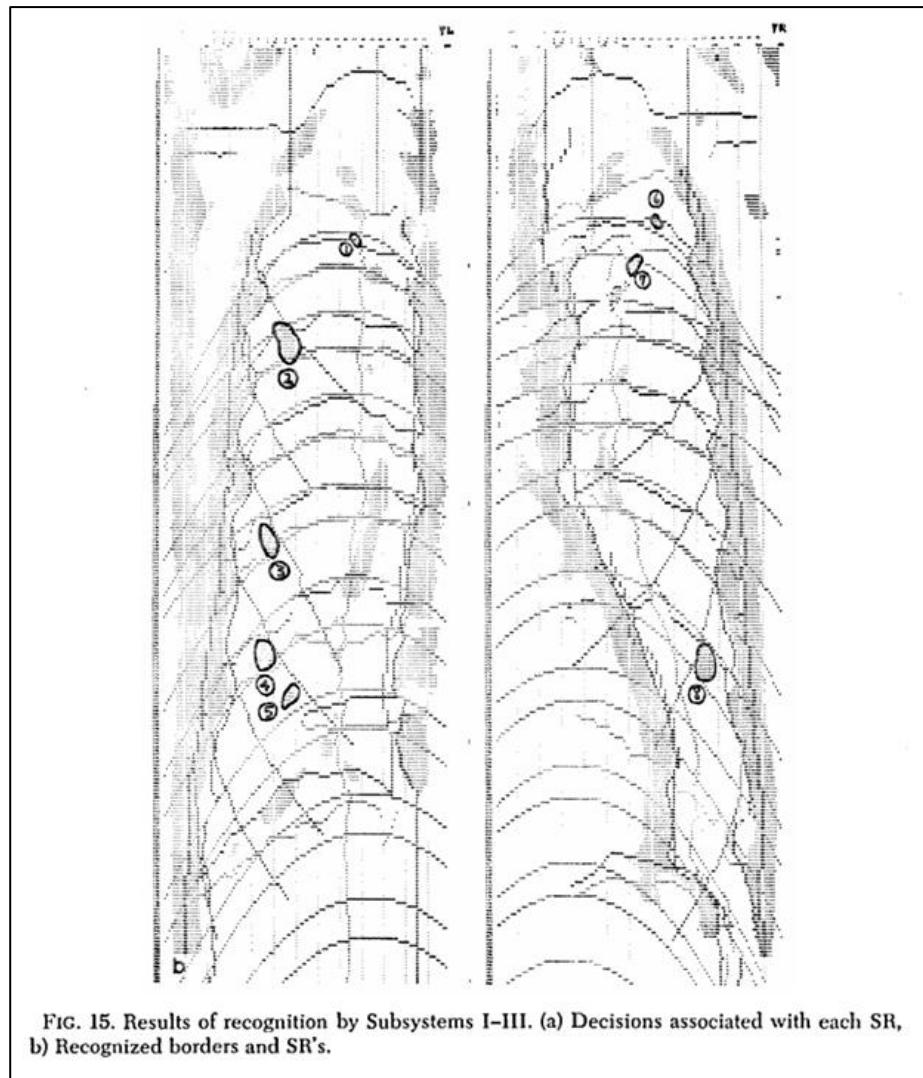
処理方法とらんでもう一つの重要な点はパターン認識の性質に関する定量的な資料の積み重ねである。実際、パターンの性質が明確であるほど処理方法の開発は容易であり、しかもよい結果が得られることは肋骨境界線識別の場合をみれば明らかである。

5. む す び

本文では胸部X線写真における病巣陰影識別のための基礎的実験の結果を述べた。ここではX線写真処理システムのもつ基本的処理の一つである random mass pattern の識別方法の基礎的検討が主要な目的であるが、その前段階としていくつかの段階が必要であるため、それらの処理 (肋骨境界線の識別も含む) に関しても、多くの結果が得られている。

さて、本実験の結果、ある種の病巣陰影については一応この方法で満足すべき結果が得られることが知られた。しかしながら、ここに用いられたパターンはいずれもきわめて識別しやすいものであることから、その能力はまだかなり限定されたものといわざるをえない。しかも両面の処理全体を自動化するにはまだまだ多くの別の機能が実現されなければならない。もちろんその場合でも中心になるのは線状図形の処理機能と塊状図形の処理機能であり、それらのあいだを適当な処理を経て接続するという構成になるであろう。その意味で、本論文は前の報告¹⁾²⁾とあわせて胸部X線写真診断

1970年初頭での自動診断 (名大 鳥脇・末永)



当時名大大型計算機センターで動作していた
FACOM 230-60 (写真は富士通ミュージアム)

J. Toriwaki, Y. Suenaga, T. Negoro, et al., Pattern
recognition of chest x-ray images, Computer Graphics
and Image Processing, 2 (1973), pp. 252-271

生成AIでFORTRANコードをPythonに

mori@taka:~/slip/slip2df\$ more disp6i.f

```

SUBROUTINE DISP6I(IS,JS,D1,IE,JE,IX1,IX2,JX1,JX2,N1,MT,KP)      02908000
C ***** 7-LEVELS PRINT DATA NO SAIDAI-CHI TO SAISYO-CHI O TSUKAU *02909000
C ***** CODED BY J.TORIWAKI 1974.11.15 *****02910000
C ***** D1(IE,JE)=INSATSU SURU NYURYOKU DATA (INTEGER) *****02911000
C ***** I=IX1--IX2, J=JX1--JX2 NO HANI NO D1(I,J) NO MAX. TO MIN. 02912000
C ***** O MOTOMETE IWK1, IWK2 TO SURU. IWK1 TO IWK2 NO AIDA O 02913000
C ***** N1-TOBUN SHITE D1 O RYOSHI-KA 02914000
C ***** N1=RYOSHI-KA NO LEVEL-SU N1.LE.7 *****02915000
C ***** MT=HYODAI, KP=KAI-PAGE SHITEI (1=KAI-PAGE, 0=KAI-PAGE SHINAI02916000
C ***** IS,JS=GYO,RETSU NO BANGO INSATSU NO SYOKICHI *****--02917000
C INTEGER D1(IE,JE),III(130),ICH(7),MT(10) 02918000
C integer D1(IE,JE)
C character*1 III(130), ICH(7)
C character III1*130, ICH1*7, MT(10)*4
C equivalence (III,III1), (ICH,ICH1)
C DATA ICH/28H. + = * 0 IO *M $ / 02919000
C data ICH1 / 7H.-+*=*OM /
C *****-----*****02920000
C M1=MAX0(1,MIN0(7,N1)) 02921000
C IWK1=D1(IX1,JX1) 02922000
C IWK2=D1(IX1,JX1) 02923000
C DO 30 J=JX1 ,JX2 02924000
C DO 30 I=IX1 ,IX2 02925000
C IWK3=D1(I,J) 02926000
C IF(IWK3.GT.IWK1) IWK1=IWK3 02927000
C IF(IWK3.LT.IWK2) IWK2=IWK3 02928000
30 CONTINUE 02929000
C U1=IWK1 02930000
C B1=IWK2 02931000
C DD1=(U1-B1)/M1 02932000
  
```

```

02932000
C *****-----*****02933000
C DO 20 JK=1,JE,130 02934000
C JK1=JK-1 02935000
C JC=MIN0(JE,JK+129)-JK1 02936000
C IF(KP.NE.0) WRITE(6,200) 02937000
C WRITE(6,201) MT,B1,U1,M1,(J1,J1=JS+JK1+10,JS+JK1+130,10) 02938000
C *****-----*****02939000
C DO 20 I=1,IE 02940000
C IL=I+IS 02941000
C DO 10 J=JK,MIN0(JE,JK+129) 02942000
C 10 III(J-JK1)=ICH(MAX0(1,MIN0(M1 ,IFIX(1.0+(D1(I,J)-B1)/DD1)))) 02943000
C WRITE(6,202) IL,(III(J),J=1,JC) 02944000
C 20 WRITE(6,203) (III(J),J=1,JC) 02945000
C 20 continue
C 200 FORMAT(1H1) 02946000
C 201 FORMAT(11X,10A4/1H , ' LOWER T=',E11.4,5X,'UPPER T=',E11.4,I7, 02947000
C 1' LEVELS ' /6X,13I10) 02948000
C 202 FORMAT(1H ,I4,1X,130A1) 02949000
C 203 FORMAT(1H+,5X,130R1) 02950000
C *****-----*****02951000
C RETURN 02952000
C END 02953000
  
```

1974.11のコード

ChatGPTによるFORTRANコードのPython変換

了解です。これは2次元整数配列 D1(IE,JE) を、値の範囲（最小～最大）を最大7段階に量子化して、各画素を1文字の記号に置き換え、ASCII画像のようにプリント出力するサブルーチンです（古いラインプリンタ／端末向けの可視化）。

1) 目的（このルーチンがやっていること）

指定した領域

- I = IX1..IX2
- J = JX1..JX2

の D1(I,J) の最大値 IWK1 と最小値 IWK2 を求め、その範囲 [IWK2, IWK1] を

- N1 段階（ただし 1～7 にクリップ）

で等間隔に分割して量子化し、量子化レベルを以下の記号にマッピングして、行ごとに出力します。

Plain text

```

ICH1 = ".-+*=0M"   (7段階)
レベル1   : .
レベル2   : -
レベル3   : +
レベル4   : =
レベル5   : *
レベル6   : 0
レベル7   : M
  
```

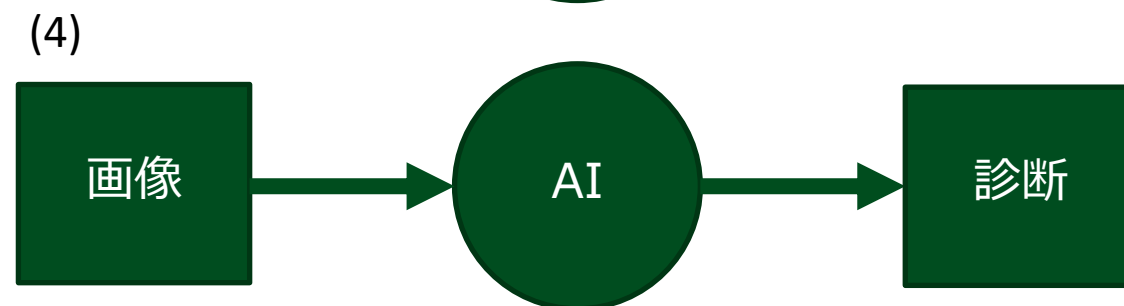
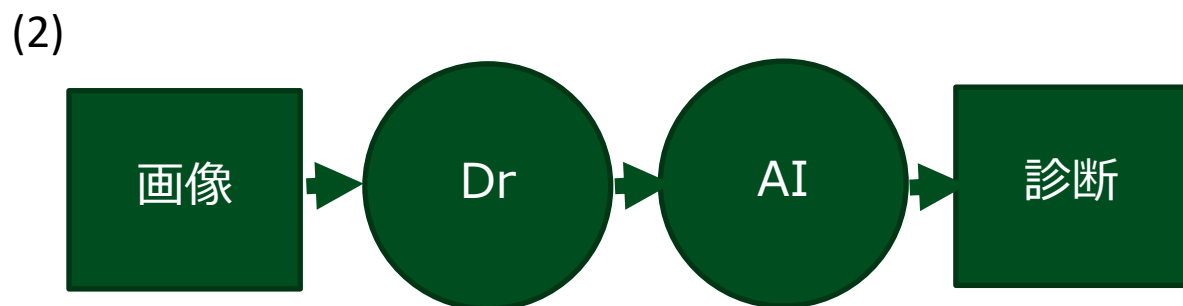
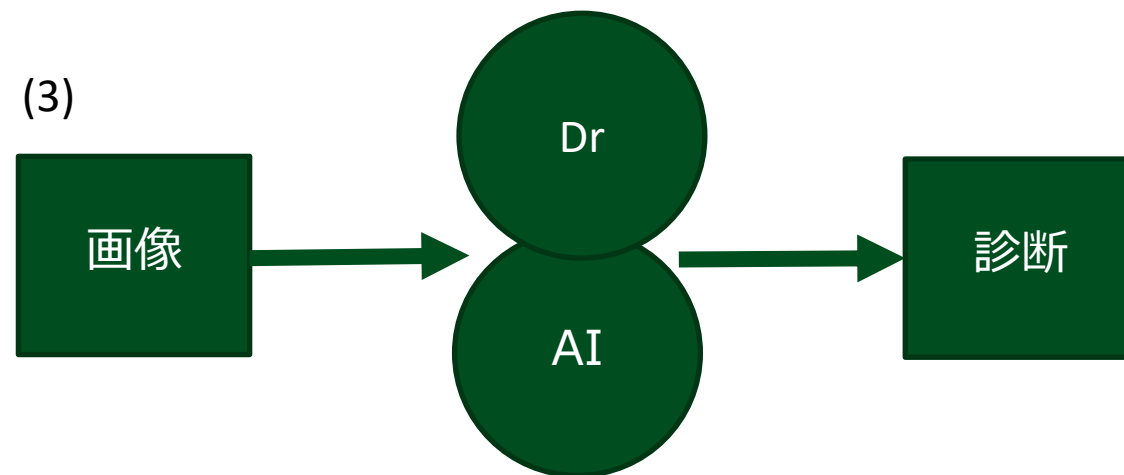
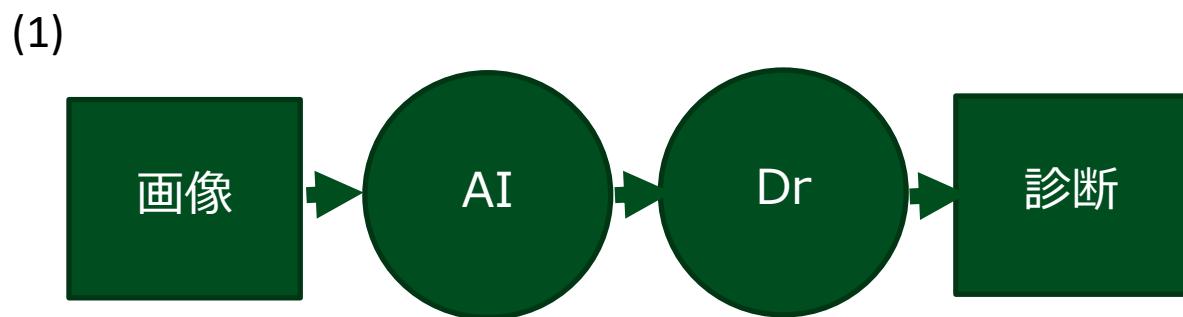


```

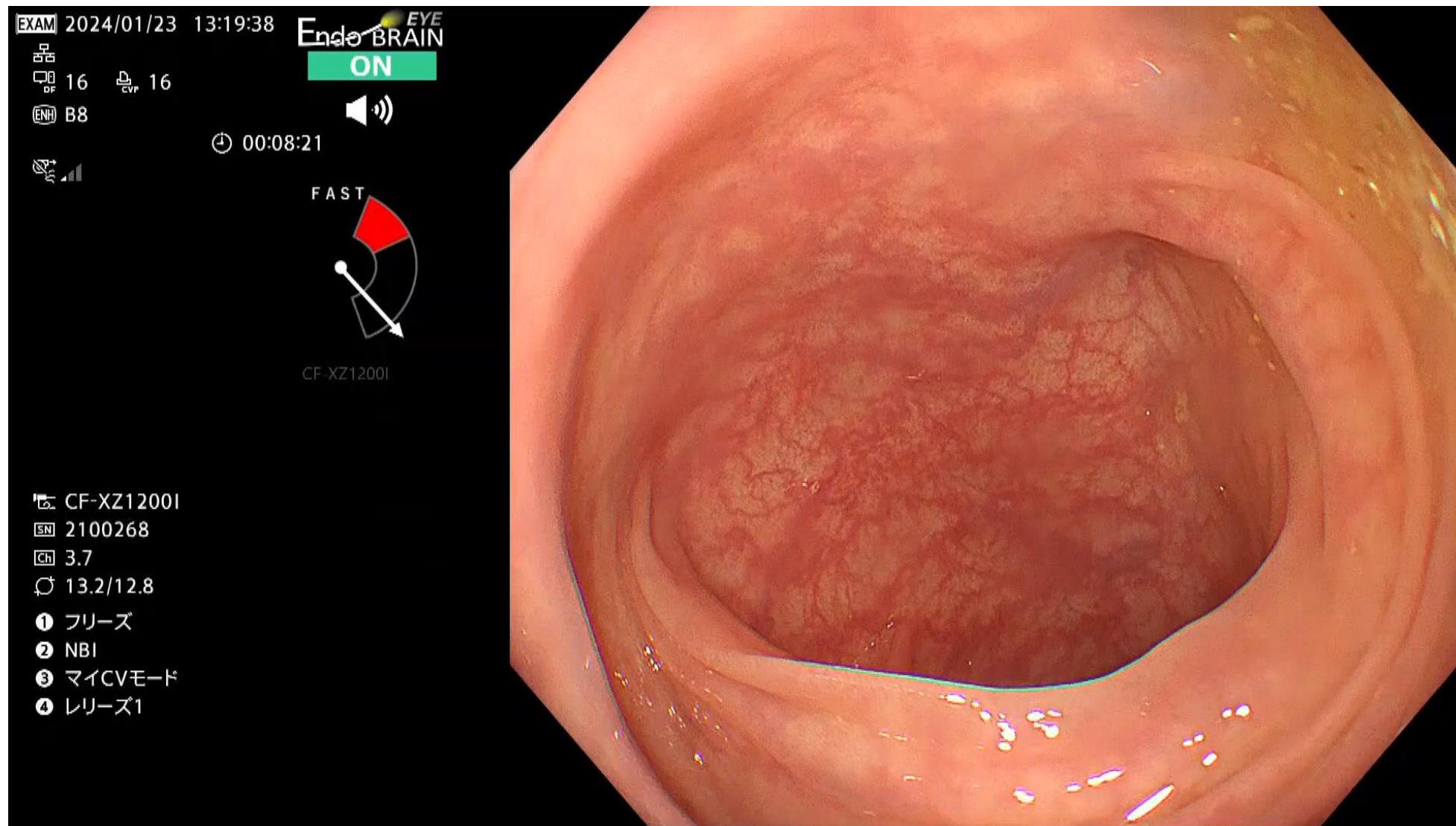
ASCII_7 = ".-+*=0M"
def image_to_ascii(
    img: Image.Image,
    width: int = 130,
    levels: int = 7,
    charset: str = ASCII_7,
    char_aspect: float = 0.5,
) -> list[str]:
    if mx == mn:
        # constant image -> middle-ish tone
        idx = np.full_like(a, fill_value=(levels - 1)
            // 2, dtype=np.int32)
    else:
        # FORTRAN-like quantization:
        # level = 1 + (val - mn) / ((mx-mn)/levels)
        # then clamp to 1..levels
        dd = (mx - mn) / levels
        lvl = 1.0 + (a - mn) / dd
        lvl = np.clip(lvl, 1.0, float(levels))
        idx = (lvl.astype(np.int32) - 1) # 0..levels-1
    lines = ["".join(charset[i] for i in row) for row in idx]
    return lines
  
```


診断の自動化

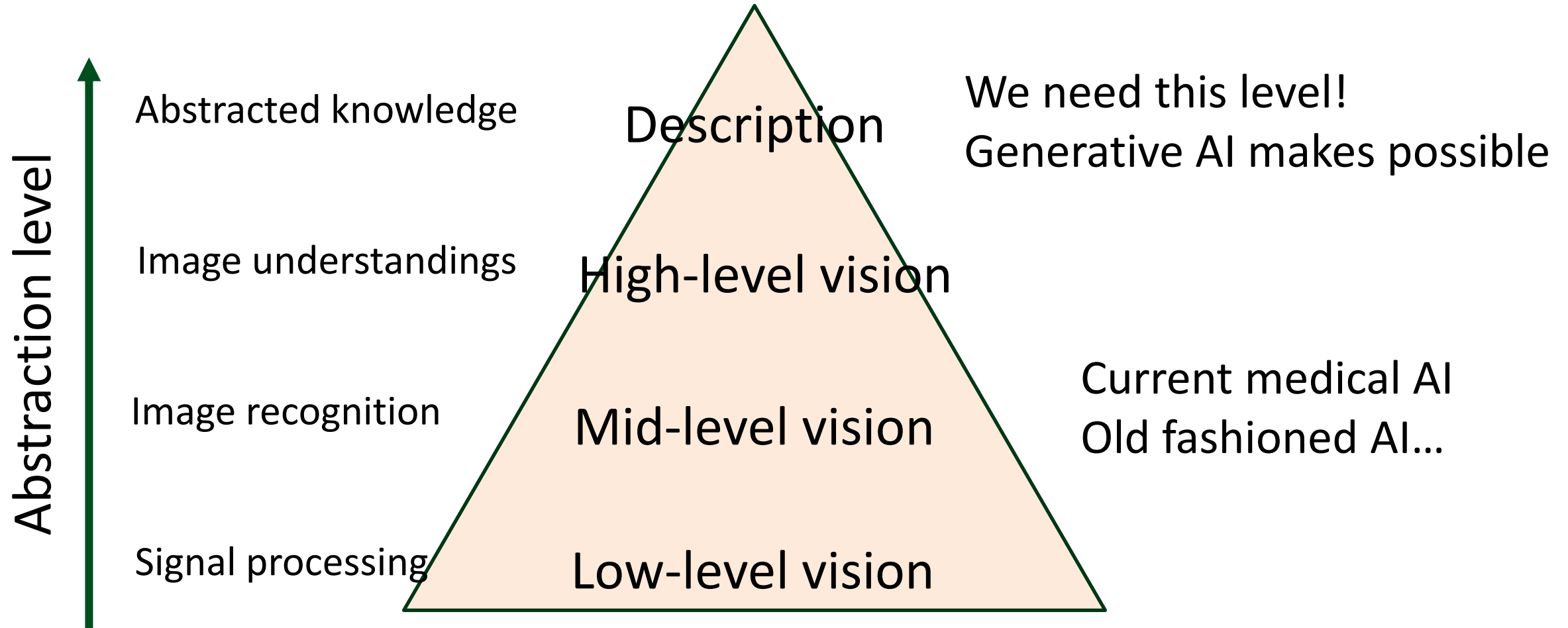
- (4)の形態が完全自動化



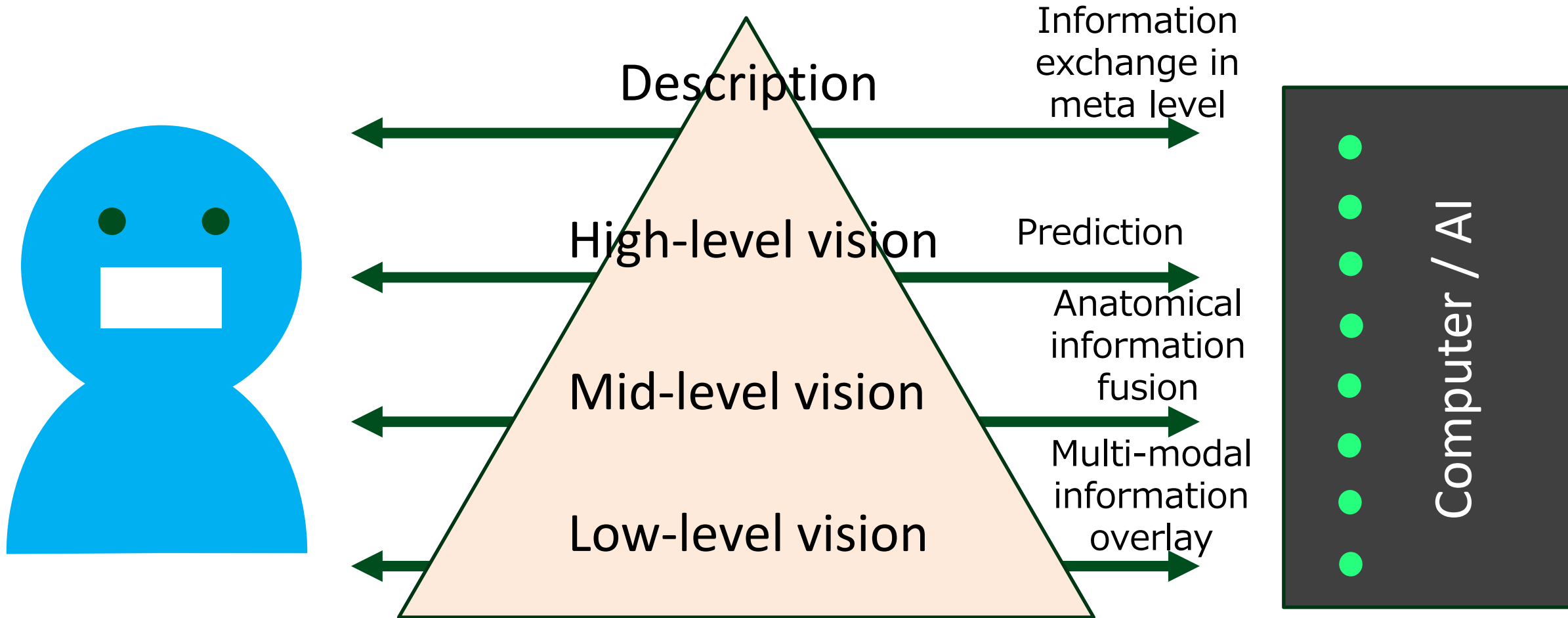
大腸内視鏡診断支援 ポリープ検出と質的診断



Abstraction level in medical assistance



Abstraction level in medical assistance



Computer-aided diagnosis (CAD) has become a practical clinical approach in diagnostic radiology, although at present only in the area of detection of breast cancer in mammograms. Current research efforts have been focused on detection and classification of images of many different types of lesions in a number of organs, obtained with various imaging modalities. It is likely that the present results of CAD are only at the tip of the iceberg. Although automated computer diagnosis is a concept based on computer algorithms only, CAD is a concept established by taking into account equally the roles of physicians and computers. The effect of CAD on differential diagnosis has already indicated that the performance level is high, and that CAD would be ready for clinical trials and commercialization efforts. The presentation of images similar to those of an unknown case may be useful as a supplemental tool for CAD in the differential diagnosis.

Semin Ultrasound CT MRI 25:404-410 © 2004 Elsevier Inc. All rights reserved.

KEYWORDS computer-aided diagnosis, computed tomography, magnetic resonance imaging

コンピュータ支援画像診断

- 当初は自動診断
 - Early studies on quantitative analysis of medical images by computer^{18, 19, 20, 21, 22, 23} were reported in the 1960s. At that time, it was generally assumed that computers could replace radiologists in detecting abnormalities, because computers and machines are better at performing certain tasks than are human beings. Thus, **the concept of computer diagnosis or automated diagnosis in radiology was established at that time.**
- 途中で少し路線変更し、コンピュータは医師のセカンドオピニオンとして機能する「コンピュータ支援診断」へと
 - *In the 1980s, however, another approach emerged which assumed that **the computer output could be utilized by radiologists, but did not replace them.** [K. Doi 2004]*

Kunio Doi, Overview on research and development of computer-aided diagnostic schemes, Seminars in Ultrasound, CT and MRI, Volume 25, Issue 5, 2004, 404-410, <https://doi.org/10.1053/j.sult.2004.02.006>.

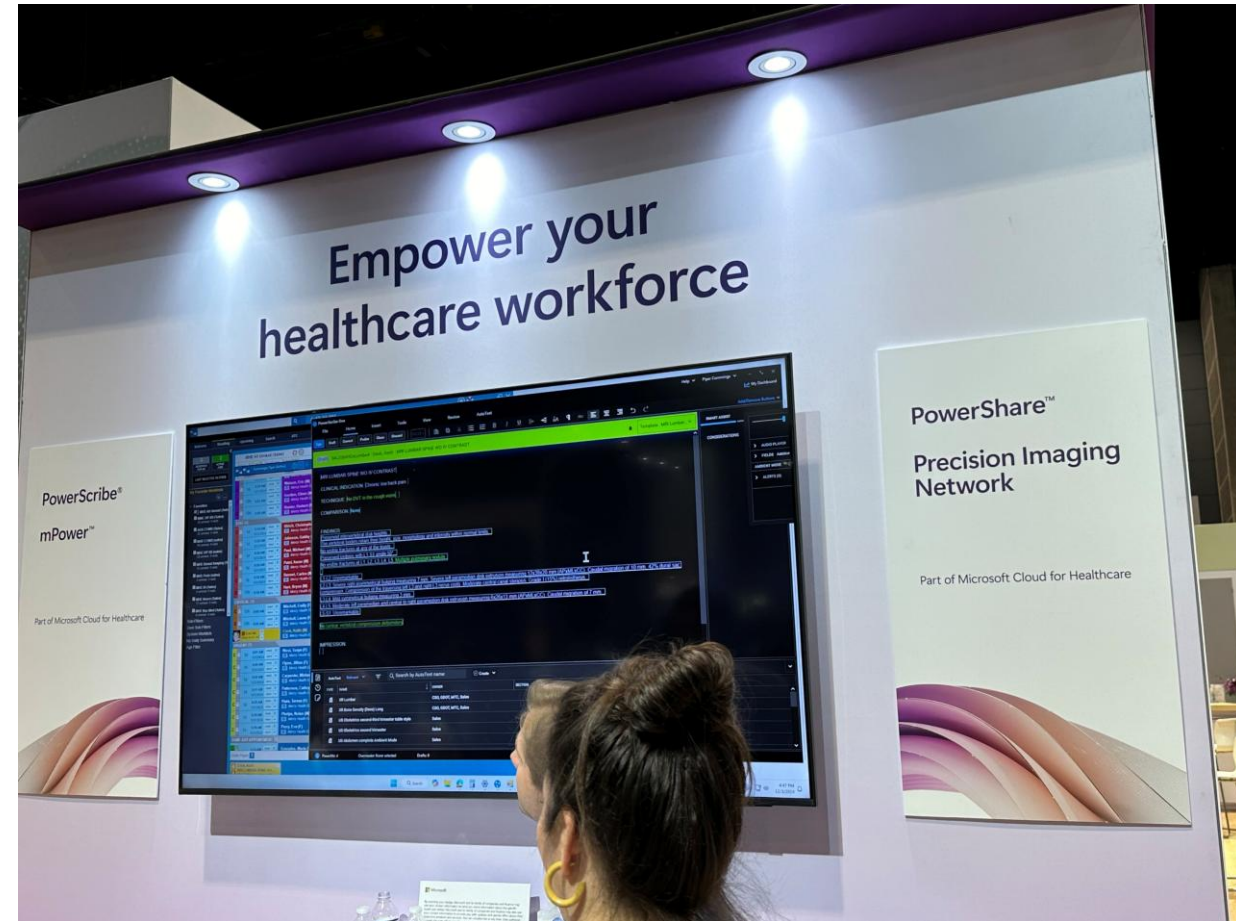
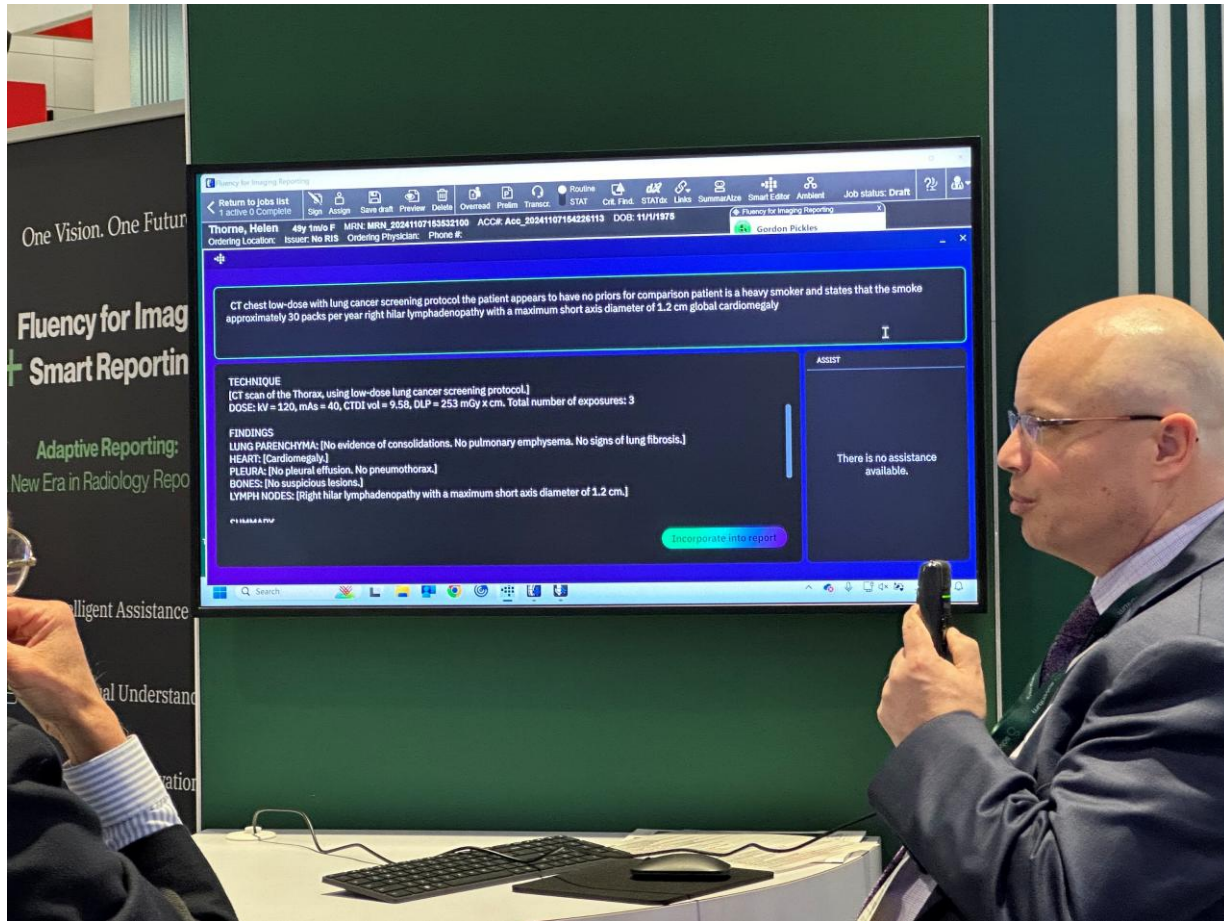
Large Language Model (LLM)による総合的な診断

- 大規模言語モデル (LLM) とは
 - 人間の言語を理解し、生成するように設計されたAIモデル
 - 主なアーキテクチャは Transformer
- 「大規模 (Large)」の意味
 - モデル内部に 数十億規模のパラメータ を持つ
 - 大量のデータと計算資源を必要とする
- 事前学習 (Pre-training)
 - 書籍・記事・ウェブサイトなどの大規模テキストで学習
 - 学習期間は 数週間～数か月
- 学習の仕組み
 - 「次に来る単語」を予測しながら言語を学習
 - 文法、事実、一般知識、文体などを獲得
- 学習の段階
 - 自己教師あり学習 (Self-supervised learning) → 言語の基本構造を理解
 - 命令調整 (Instruction tuning) → 質問応答や指示への応答能力を獲得
- 画像特徴をLLMに入力するとLLMが画像に関する**網羅的な解釈**を文章として返す

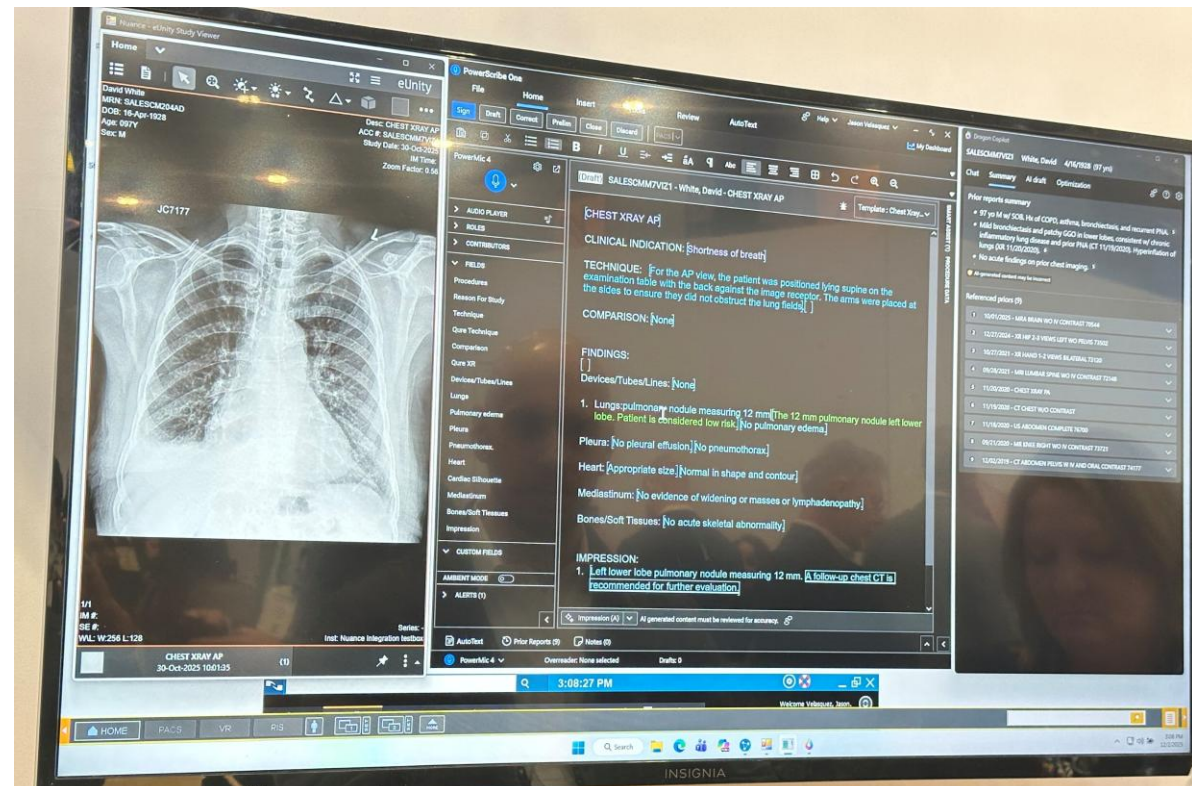
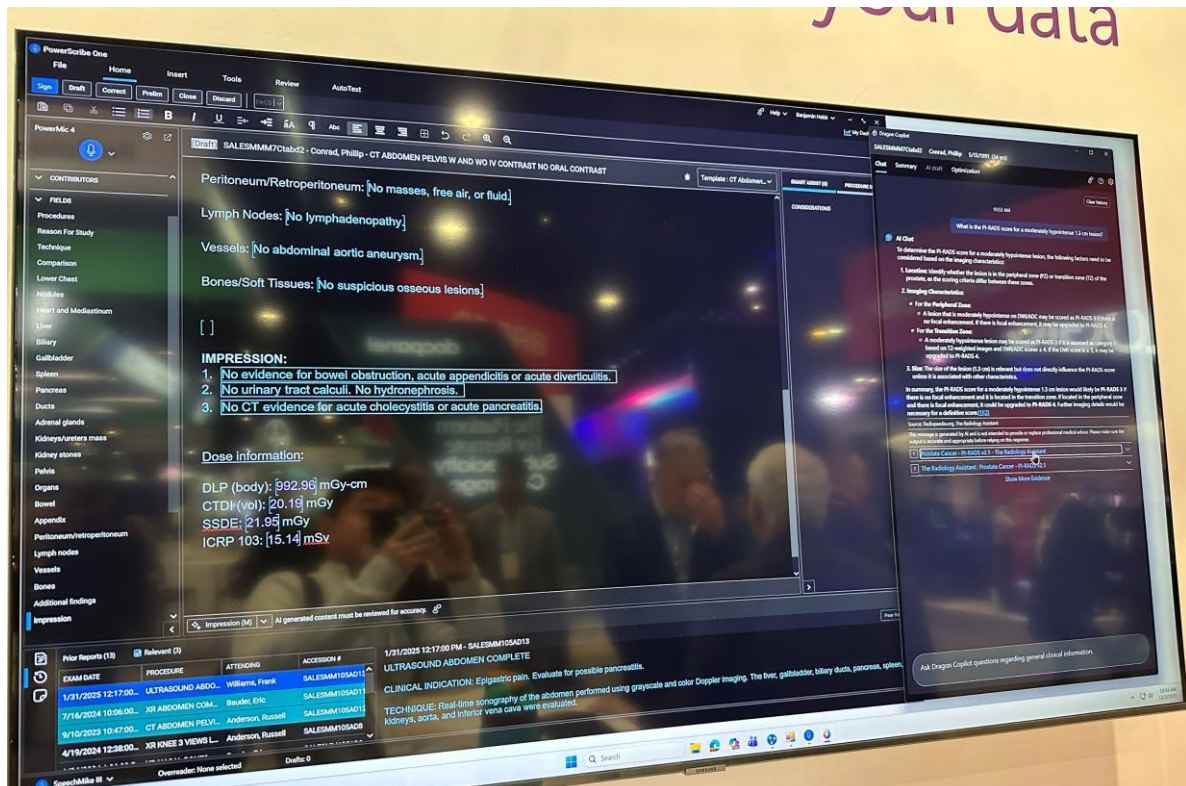
代表的なLLMの例：

- GPTシリーズ (OpenAI)
- Claude (Anthropic)
- Gemini (Google DeepMind)
- LLaMA (Meta)
- Mistral (Mistral AI)
- Qwen (Mistral AI)

言語モデルを利用した読影レポート作成支援 RSNA2024

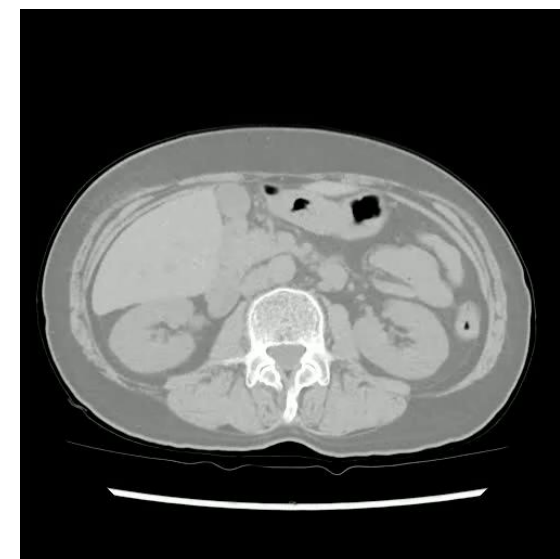


DictationされたFindingsからImpressionを自動生成



放射線分野での自動診断

- 放射線診断レポートの作成
 - 放射線診断の過程で放射線レポートが作成
 - FindingsとImpressions
- 大規模言語モデル (LLM) の発展
 - 自然言語処理のためのLLM
 - AIエージェントシステムの基盤
- Vision – Language Model (VLM)
 - 視覚と言語を統合的に扱うモデル
 - VLM は、画像とそれに関連する質問を入力として受け取り、質問に対する回答を出力



(Findings) Bilateral pleural effusion is unchanged from last time. There is mild collapse of the left lung base. A nodule in the central left lower lobe is unchanged but gradually increasing. A small localized ground-glass opacity in the left lower lobe is also unchanged. Right pleural thickening shows no significant change. Small scattered nodules in both lungs show no growth and are likely benign post-inflammatory. There is no significant lymph node enlargement. The right mediastinal soft tissue shadow is unchanged. Bilateral renal cysts are present, with no hydronephrosis or ascites

ARRG – Automated Radiology Report Generation

368

IEEE REVIEWS IN BIOMEDICAL ENGINEERING, VOL. 18, 2025



Automated Radiology Report Generation: A Review of Recent Advances

Phillip Sloan , Philip Clatworthy , Edwin Simpson , and Majid Mirmehdi 

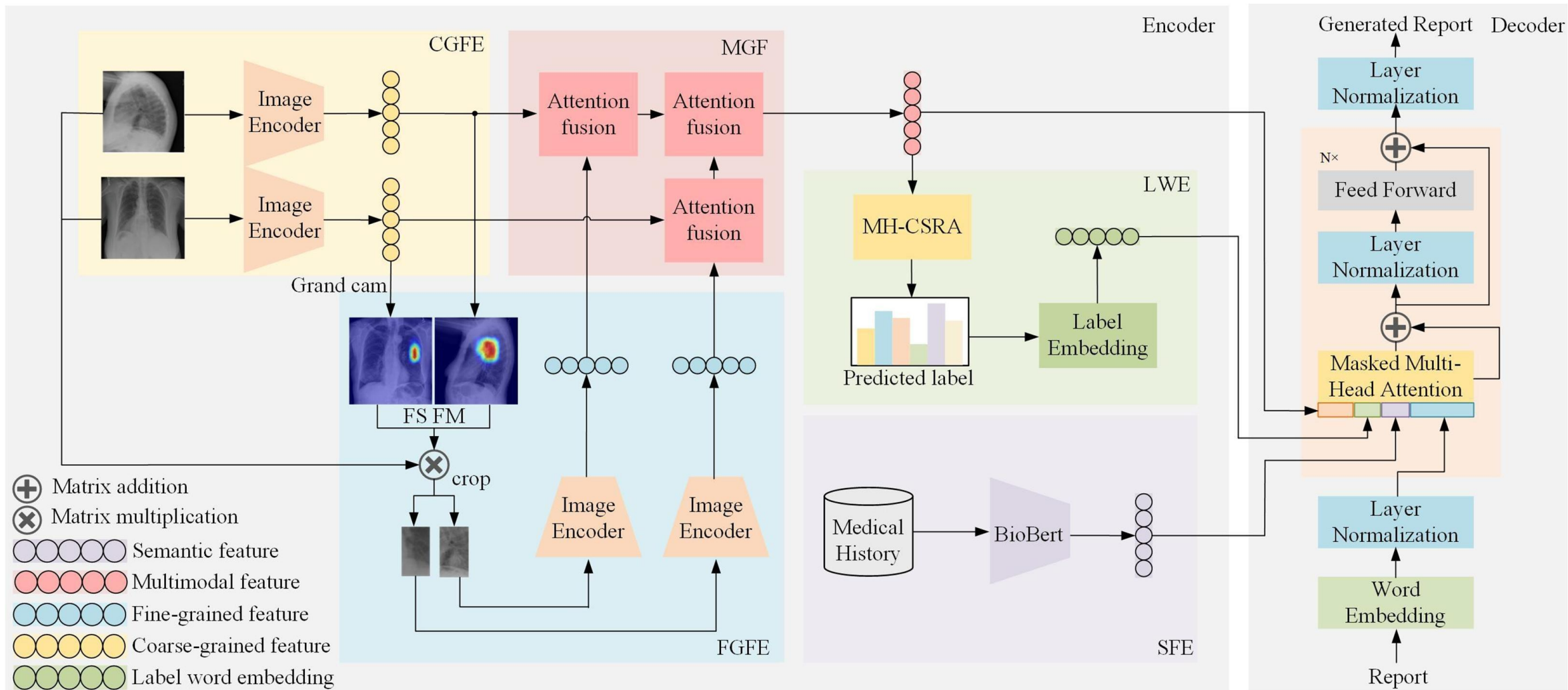
Abstract—Increasing demands on medical imaging departments are taking a toll on the radiologist’s ability to deliver timely and accurate reports. Recent technological advances in artificial intelligence have demonstrated great potential for automatic radiology report generation (ARRG), sparking an explosion of research. This survey paper conducts a methodological review of contemporary ARRG approaches by way of (i) assessing datasets based on characteristics, such as availability, size, and adoption rate, (ii) examining deep learning training methods, such as contrastive learning and reinforcement learning, (iii) exploring state-of-the-art model architectures, including variations of CNN and transformer models, (iv) outlining techniques integrating clinical knowledge through multimodal inputs

TABLE I

RECENT ARRG REVIEW ARTICLES AND THEIR COVERAGE INCLUDING TOTAL NUMBER OF CITATIONS AND THE TOTAL NUMBER OF ARTICLES RELATED TO ARRG OR MEDICAL REPORT GENERATION

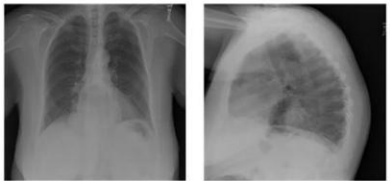
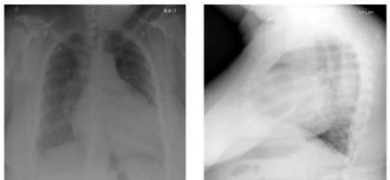
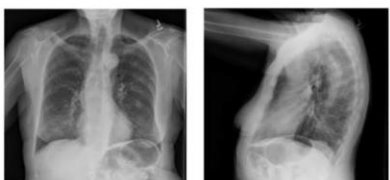

Author	Year	Coverage	# Cited	# ARRG
Allaouzi et al. [16]	2018	2015-2018	50	10
Pavlopoulos et al. [17]	2019	2015-2019	65	7
Monshi et al. [12]	2020	2015-2019	105	14
Kaur et al. [14]	2022	2015-2021	83	19
Messina et al. [15]	2022	2016-2021	167	40
Liao et al. [13]	2023	2016-2021	154	41
Our Review	2023	2020-2023	154	83

X線画像からの診断レポート自動生成



Junze Fang, Suxia Xing, Kexian Li, Zheng Guo, Ge Li, Chongchong Yu, Automated generation of chest X-ray imaging diagnostic reports by multimodal and multi granularity features fusion, Biomedical Signal Processing and Control, 2025, <https://doi.org/10.1016/j.bspc.2025.107562>

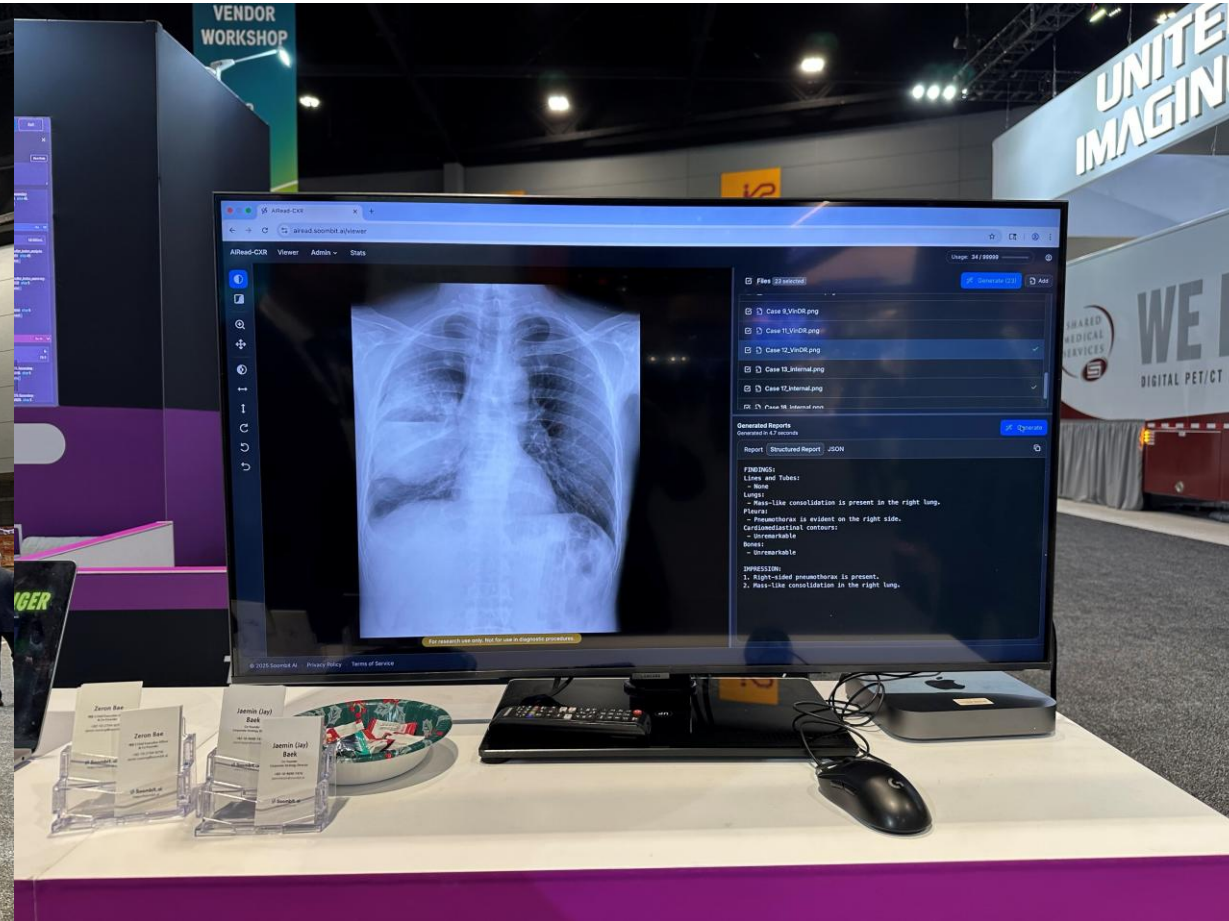
X線画像からの診断レポート自動生成

Image	Ground Truth	Base	MMG
	the cardiomeastinal silhouette is within normal limits. the lungs are normally inflated without evidence of focal airspace disease, pleural effusion or pneumothorax. no acute osseus abnormality large hiatal hernia.	the cardiomeastinal contours are within normal limits. the lungs are clear bilaterally specifically. no evidence of focal consolidation pneumothorax or pleural effusion. no acute bone abnormality.	the cardiomeastinal silhouette is within normal limits for size and contour. the lungs are normally inflated without evidence of focal airspace disease, pleural effusion or pneumothorax. no acute bone abnormality.
	there is moderate cardiomegaly, there are bilateral interstitial opacities increased since the previous exam, no focal airspace consolidation pleural effusions or pneumothorax, no acute bony abnormalities.	there is moderate cardiomegaly, no pleural effusion, pneumothorax or focal airspace disease, no acute bony abnormality.	there is moderate cardiomegaly, there is no focal airspace consolidation pleural effusions. no visualized pneumothorax. no acute bony abnormality.
	normal heart size. stable tortuous thoracic aorta. prior granulomatous disease healed rib fractures appear. stable focal opacity is noted in the left midlung overlying the 9th posterior rib which represents healing rib callus. no pneumothorax or pleural effusion.	the heart is normal in size. the mediastinum is stable. there is no pleural effusion or pneumothorax. aorta are tortuous. focal opacity is noted. arthritic changes of the skeletal structures are noted.	heart size within normal limits. stable mediastinal contours. the mediastinum is stable. no pleural effusion or pneumothorax. there are stable tortuous of the aorta. focal opacity is noted in the left.
	heart size and pulmonary vascularity appear within normal limits. there is mild tortuosity to the descending thoracic aorta. the lungs are free of focal airspace disease. no pleural effusion or pneumothorax is seen. no discrete nodules or adenopathy. degenerative changes are present in the spine.	the heart size is normal. the lungs are free of focal airspace disease. no pleural effusion or pneumothorax. degenerative changes are in the spine.	heart size and pulmonary vascularity appear within normal limits. the lungs are free of focal airspace disease. no pleural effusion or pneumothorax is seen. no discrete nodules. degenerative changes of the thoracic spine.

Junze Fang, Suxia Xing, Kexian Li, Zheng Guo, Ge Li, Chongchong Yu, Automated generation of chest X-ray imaging diagnostic reports by multimodal and multi granularity features fusion,

Biomedical Signal Processing and Control, 2025, <https://doi.org/10.1016/j.bspc.2025.107562>

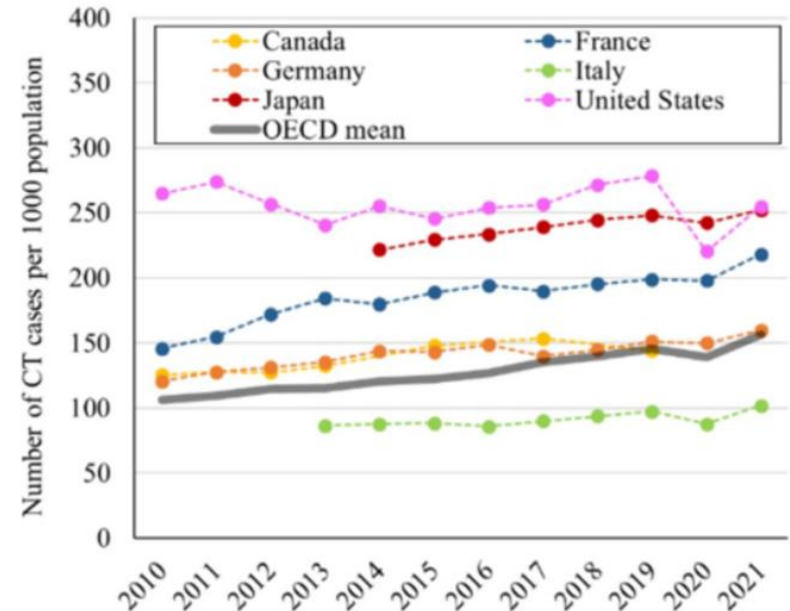
胸部 X 線写真の自動診断



Soombit.ai

CT検査数の増大

- 毎年何百万件ものCT検査が実施
 - CT検査の自動解析
- 経時解析は臨床上極めて重要
 - 以前のCTと比較治療効果の評価や経過観察
 - 複数CTの読影は時間と労力がかかる
 - 認知負荷・疲労・ストレス増大



人口千人当たりのCT検査数[1]

自動レポート生成システムの必要性が高い

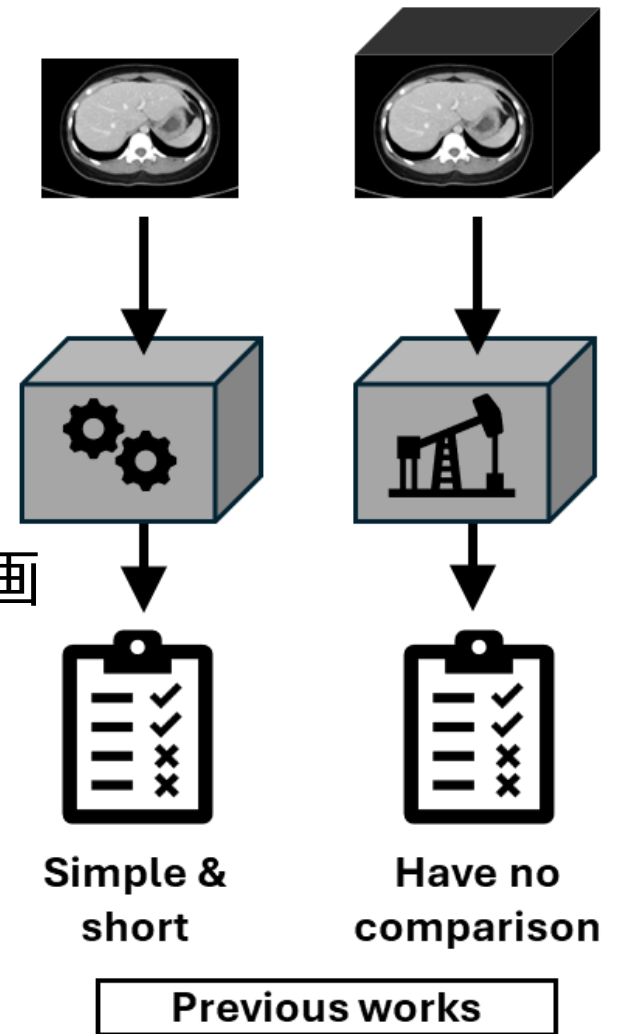
CT画像と所見文



- [腹部CT] 肝：pneumobiliaあり。脾：n.p. 胆嚢：摘出後。膵：膵頭十二指腸切除術後。残膵はやや萎縮。腎：両腎に嚢胞を認めます。副腎：n.p. リンパ節腫大なし。腹水なし。

医用画像LMM

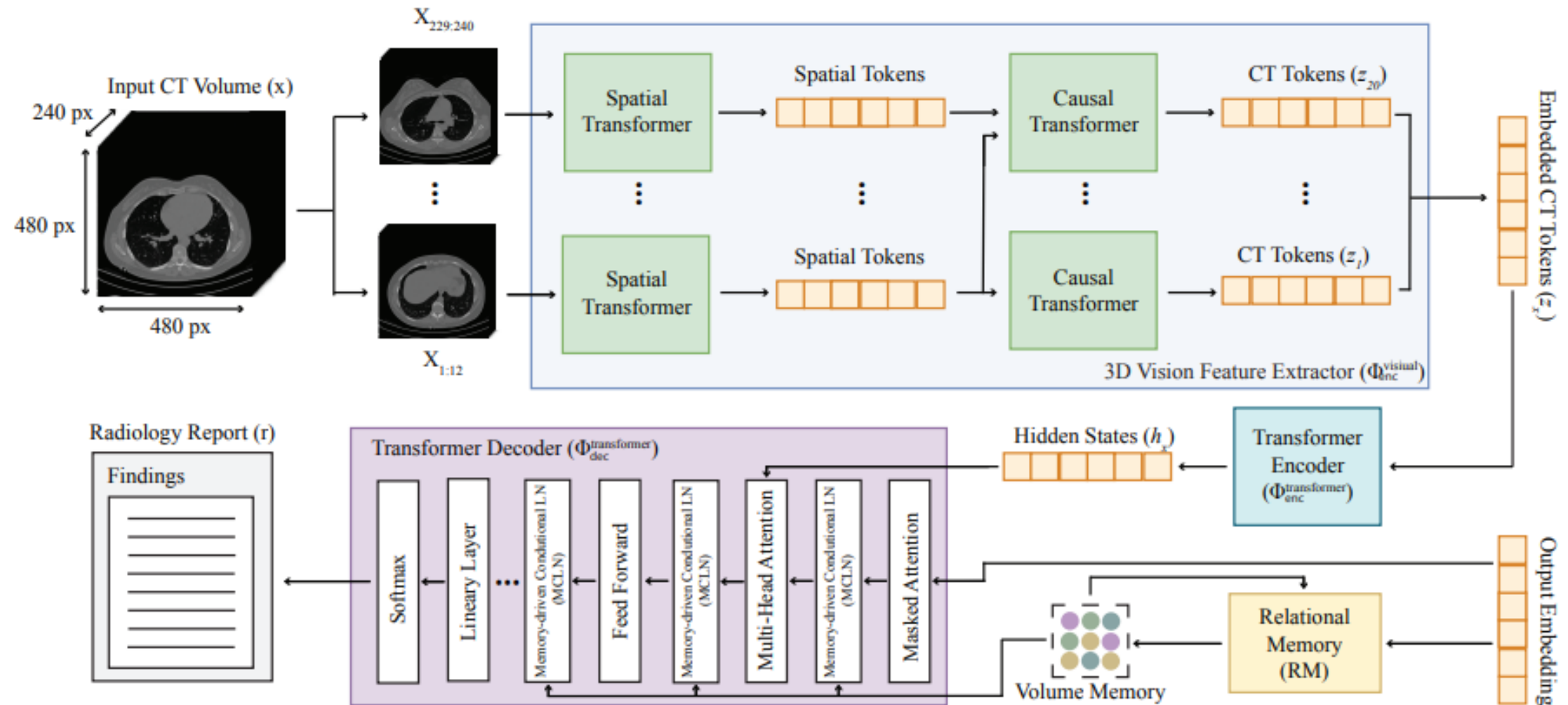
- 2次元画像中心
 - VLMの多くが2D画像に特化
 - ボリュームとして情報を捉えられない
 - 2次元モデルの3次元モデルへの拡張は不十分
- 3次元画像を対象としたものは少ない
 - 過去画像の情報を用いるものは極めて少ない
 - 一部研究 (CT2RepLong[3]) は関連性記憶を用いて経時画像に対応
 - 日本語所見文等を用いて学習させたものはない
 - 日本での所見文と他国での所見文にも差
 - **学習用データも大量に必要**



[2] Blankemeier, L., et al.: Merlin: A vision language foundation model for 3d computed tomography (2024)

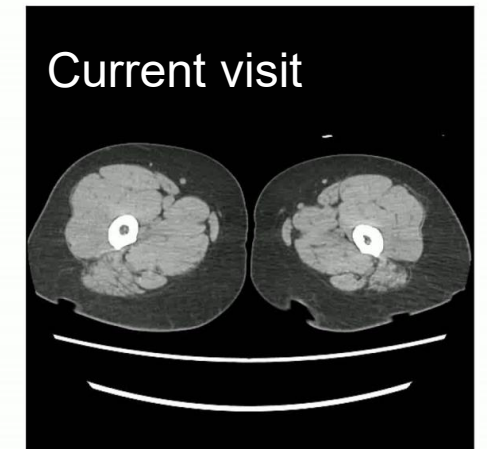
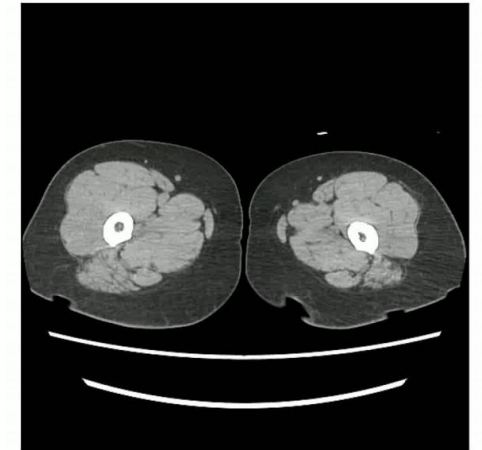
[3] Hamamci, I.E., et al.: CT2Rep: Automated radiology report generation for 3d medical imaging (2024)

Various-vision language models CT2Rep⁴⁾



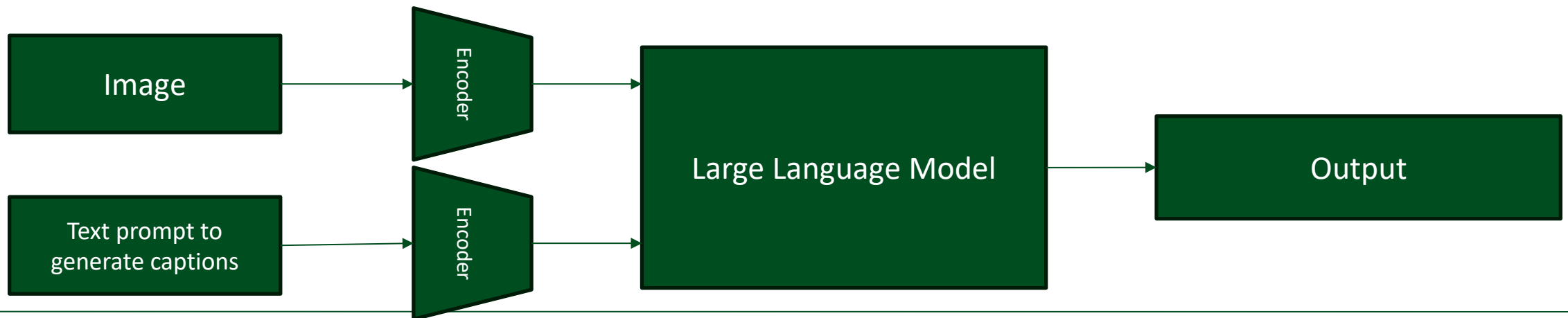
CT image interpretation

- Single-shot image interpretation
 - Diagnosis has been made only from one volume
- Longitudinal image interpretation
 - Diagnosis has been made from longitudinal volumes to find changes



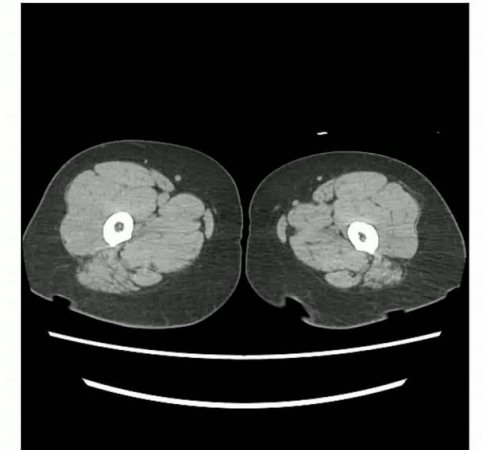
Vision-Language Model (VLM)

- 画像と言語（テキスト）の両方を扱うAIモデル
- 画像からキャプション（説明文）を生成可能
 - 説明文の代わりにロボット操作コマンドも生成可能（Vision Language Action）
- 放射線診断レポート作成補助、CT/MR画像の自動生成支援に応用可能
- 一般的にVLMは以下の要素で構成
 1. 画像エンコーダ (Image Encoder)
 2. テキストエンコーダ (Text Encoder)
 3. 言語モデル (Vision-Language Model)
 4. デコーダ (Decoder)

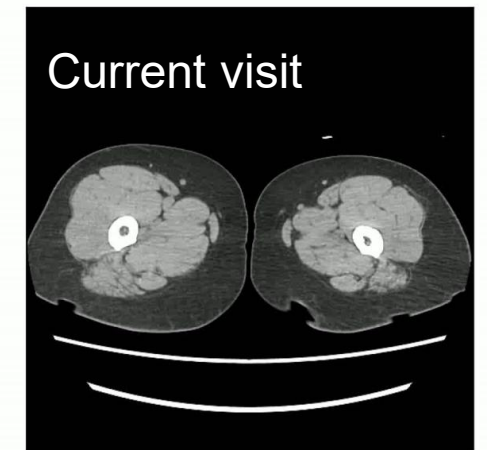
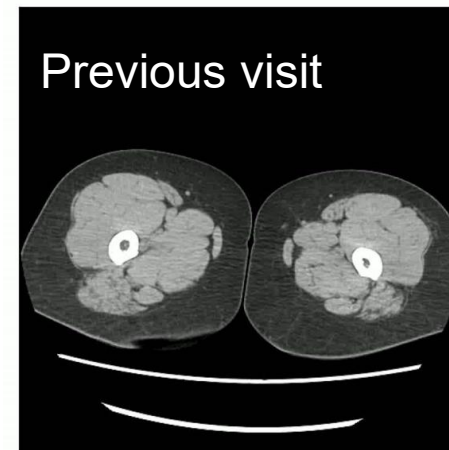


CT image interpretation

- Single-shot image interpretation
 - Diagnosis has been made only from one volume



- Longitudinal image interpretation
 - Diagnosis has been made from longitudinal volumes to find changes



CT report generation by VLM

- **Using a 3D encoder to directly process 3D CT volumes**

- Visual Encoder

- pretrained 3D Swin Transformer is used to extract features from 3D CT volumes

- **Feature Fusion**

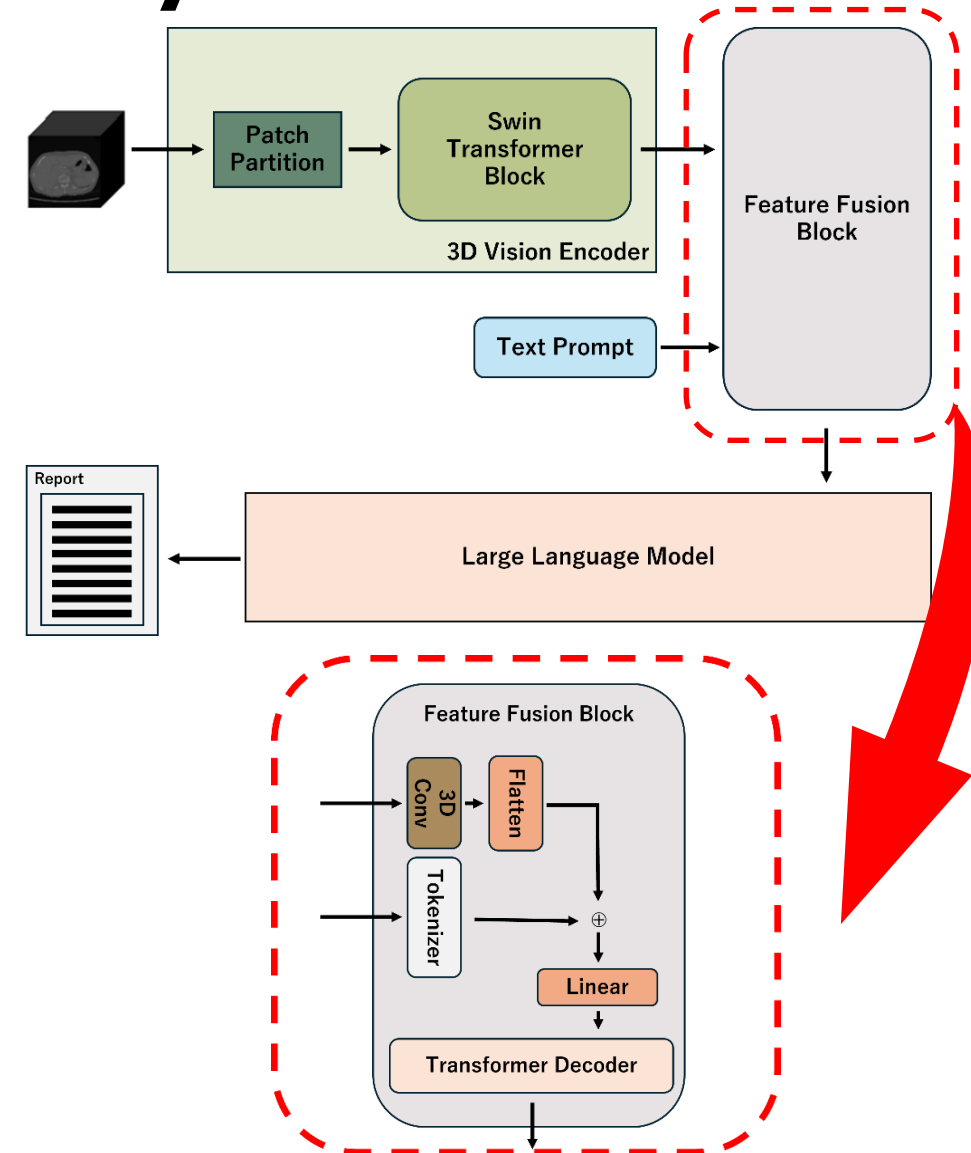
- A 3D convolutional layer followed by flattening merges spatial information

- Language Model

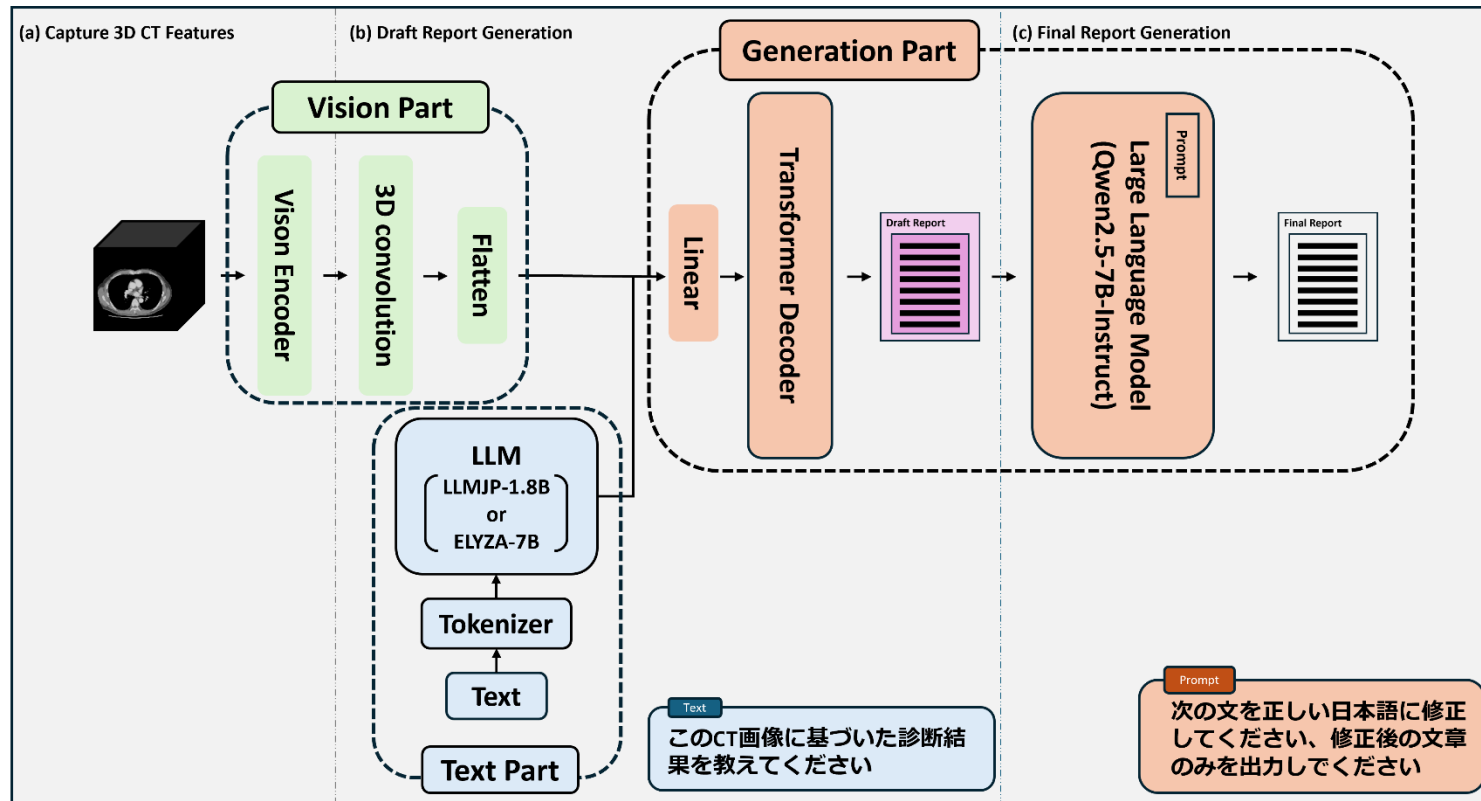
- A Japanese LLM (ELYZA-japanese-Llama-2-7b) processes prompts and generates the final report

- End-to-End Flow

- 3D CT volume → Feature extraction → Transformer Decoder → Prompted LLM → Japanese report



CT report generation by VLM



(a) Capture 3D CT Features

- Utilize a pretrained 3D Swin Transformer to capture CT features from 3D CT volumes.

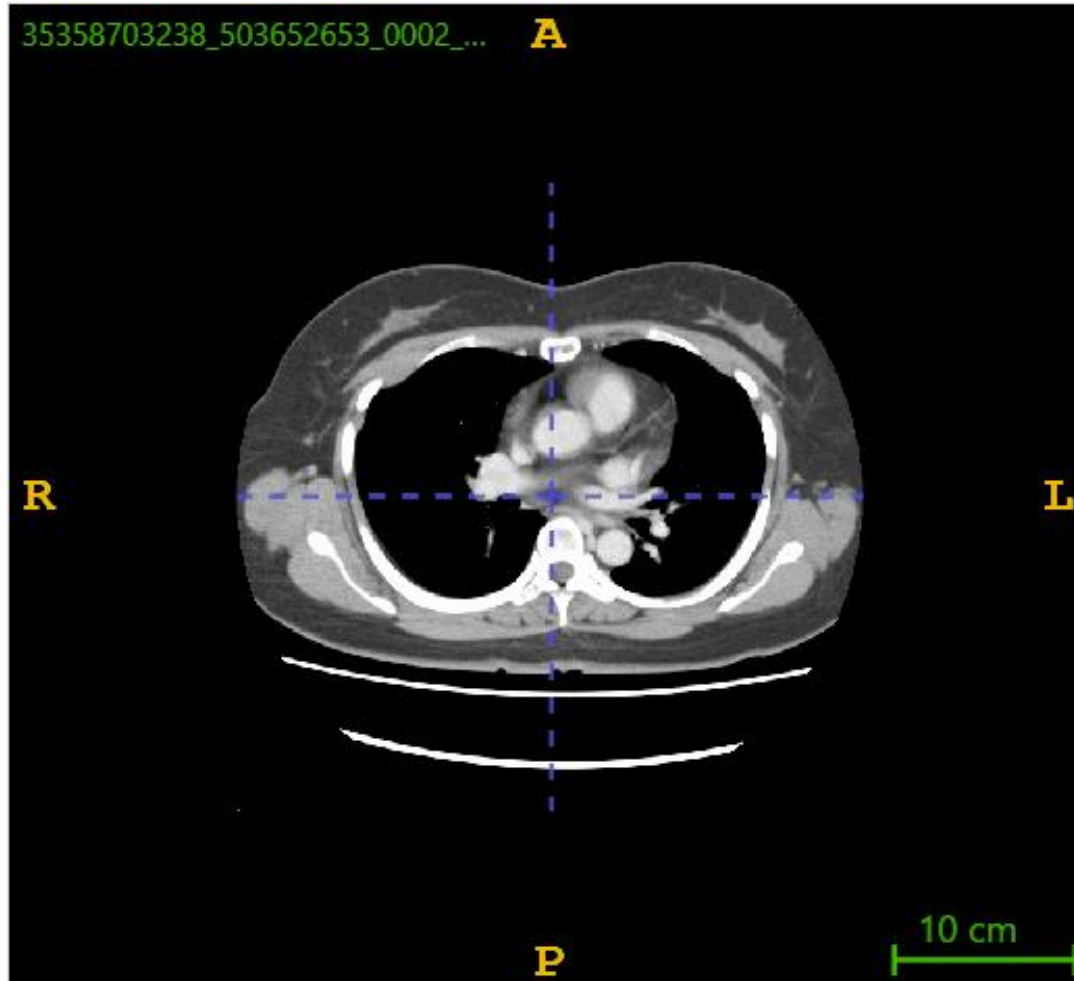
(b) Draft Report Generation (Stage 1)

- CT features are first transformed into the text embedding space using a 3D convolutional layer followed by a flattening layer.
- The text prompt is transformed into text embeddings through a pre trained LLM (ELYZA-7B and LLMJP-1.8B).
- Transformed CT features are concatenated with the text embeddings and the concatenated features are sent to the Transformer Decoder Layer to generate a draft report.

(c) Final Report Generation (Stage 2)

- The LLM (Qwen2.5-7B-Instruct) refines the draft report guided by prompts to produce the final report.

Result



Output	Ground Truth
<p>腎に嚢胞が認められるが、肝、脾、胆嚢、膵、副腎には異常なし。リンパ節腫大も認められず、腹水はなし。</p> <p>(Cysts are observed in the kidney, but no abnormalities are found in the liver, spleen, gallbladder, pancreas, or adrenal glands. There is no lymphadenopathy, and no ascites is present.)</p>	<p>両腎に嚢胞を認めるが、肝、脾、胆嚢、膵、副腎には異常なし。リンパ節腫大も見られず、腹水はなし。</p> <p>(Cysts are observed in both kidneys, but no abnormalities are found in the liver, spleen, gallbladder, pancreas, or adrenal glands. There is no lymphadenopathy, and no ascites is present.)</p>

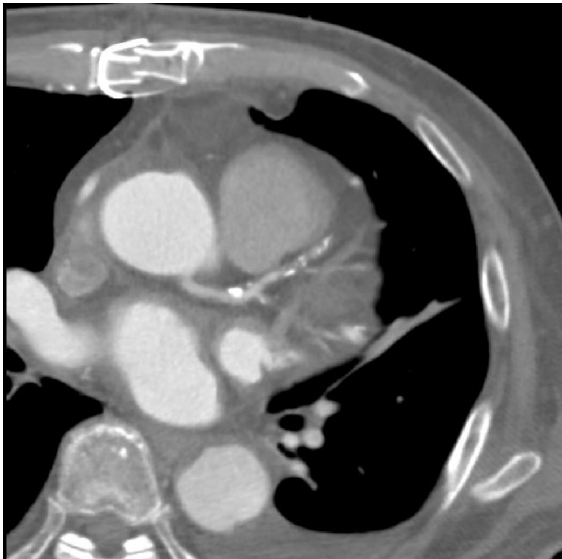
E-2 冠動脈造影3D CT像 (目的特化3D CT)に対するレポート生成

症例1 (入力は3次元CT)



(AI生成) 気管は開いています。縦隔内に病的なリンパ節腫脹は認められませんでした。心臓の検査では、左前降支 (LAD) と右冠動脈が石灰化した粥状硬化斑点で塞栓性狭窄を示しています。心拡大あり。両側胸腔において胸水や肥厚は認められませんでした。上腹部の評価では、肝実質疾患に関連する可能性のある低密度領域が観察されました。骨構造には変形があります。椎体後部要素および脊髄圧迫症候群が見られます。

症例2 (入力は3次元CT)

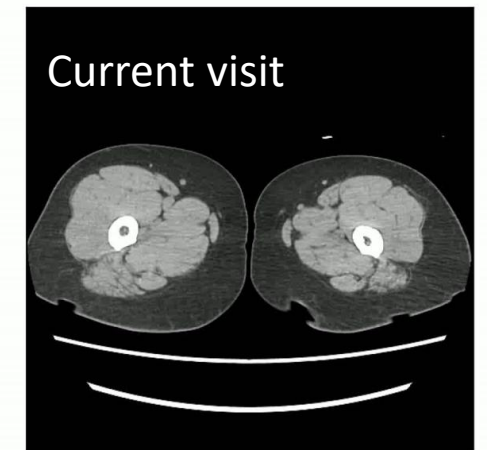
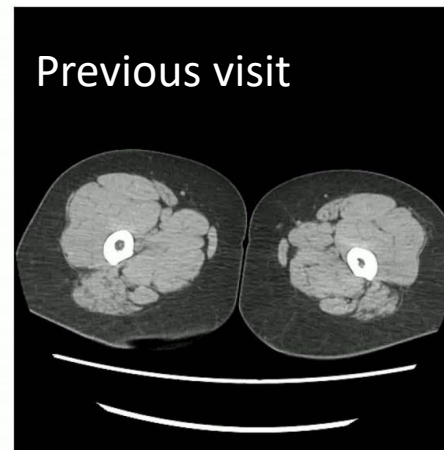
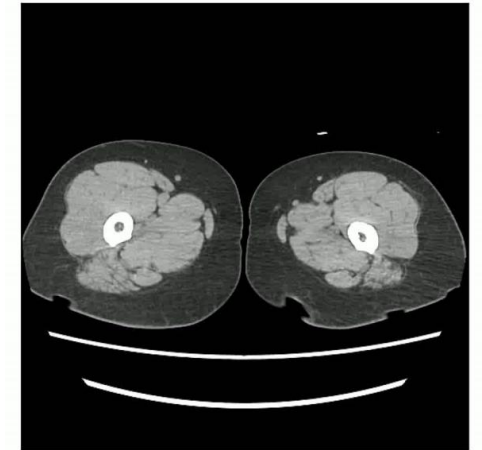


(AI生成) 気管、主支気管は開いています。右肺上葉に軽度の線状無気肺が見られます。両側肺下部には胸膜-実質後遺症があります。左肺下葉前内基底セグメントでは、直径約10mmの結節性病変 (非特異的) と疑わしい腫瘤様所見があり、その周囲でわずかなすりガラス密度が観察されます。この外観から、感染や炎症過程による可能性があると考えられました。臨床および検査との相関が必要です。心拡大あり。大動脈及び冠動脈には石灰化したアテロームプラークが認められます。縦隔リンパ節群において病理学的に拡大したリンパ節は確認されませんでした。食道壁厚の増加も認められず、胃内容物は自然なものです。胆嚢腔内には石像のような高密度領域が存在します。骨構造には破壊的な溶骨性または硬化性占位病変は認められませんでした。

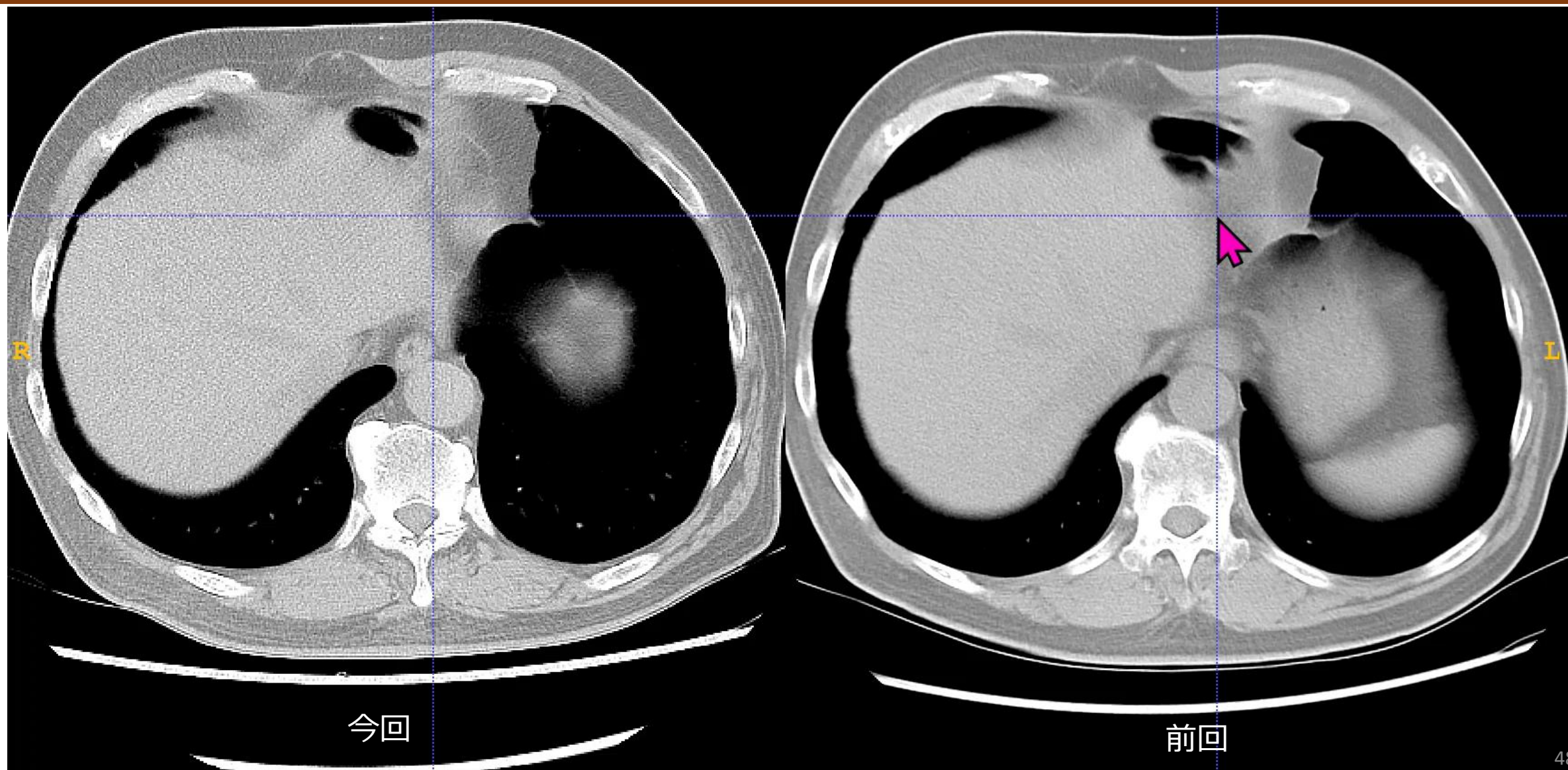
CT image interpretation

- Single-shot image interpretation
 - Diagnosis has been made only from one volume

- Longitudinal image interpretation
 - Diagnosis has been made from longitudinal volumes to find changes

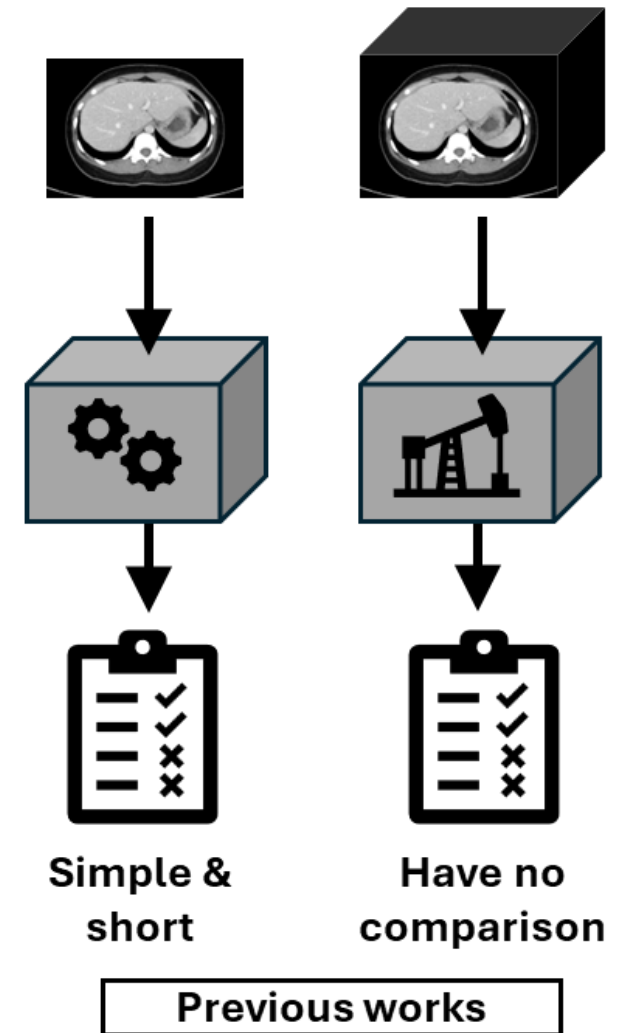


比較読影 - 2つの画像を比較しながら画像を読む



医用画像LMM

- 2次元画像中心
 - VLMの多くが2D画像に特化
 - ボリュームとして情報を捉えられない
 - 2次元モデルの3次元モデルへの拡張は不十分
- 3次元画像を対象としたものは少ない
 - 過去画像の情報をを用いるものは極めて少ない
 - 一部研究 (CT2RepLong[3]) は関連性記憶を用いて経時画像に対応
 - 日本語所見文等を用いて学習させたものはない
 - 日本での所見文と他国での所見文にも差
 - **学習用データも大量に必要**

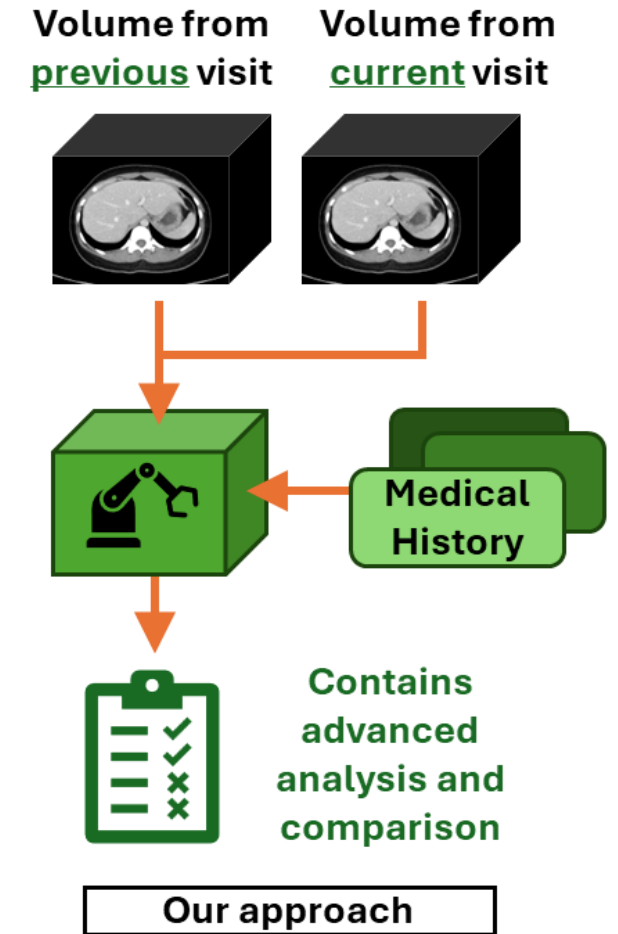


[2] Blankemeier, L., et al.: Merlin: A vision language foundation model for 3d computed tomography (2024)

[3] Hamamci, I.E., et al.: CT2Rep: Automated radiology report generation for 3d medical imaging (2024)

経時CT画像の読影レポート生成

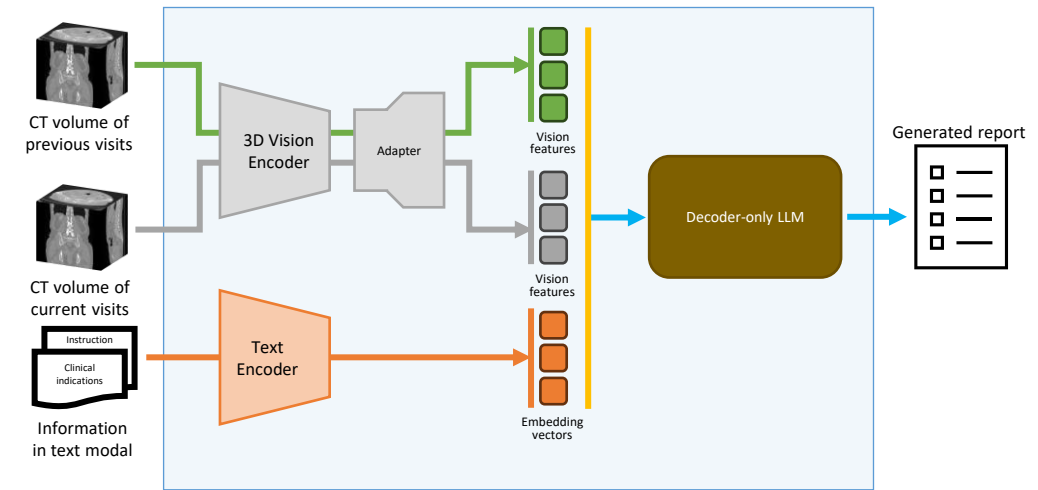
- 経時CT画像を対象とした読影レポート生成
 - 2つの3次元CT画像から読影レポートを生成
- 視覚言語モデル (VLM) により実現
 - 経時3次元CT画像間の変化を記述
 - 以前のデータと介入情報(例: 胃半摘出術後など)の利用
 - 日本語でのレポート生成
 - 日本の臨床現場で記述されているレポートと同様なレポート



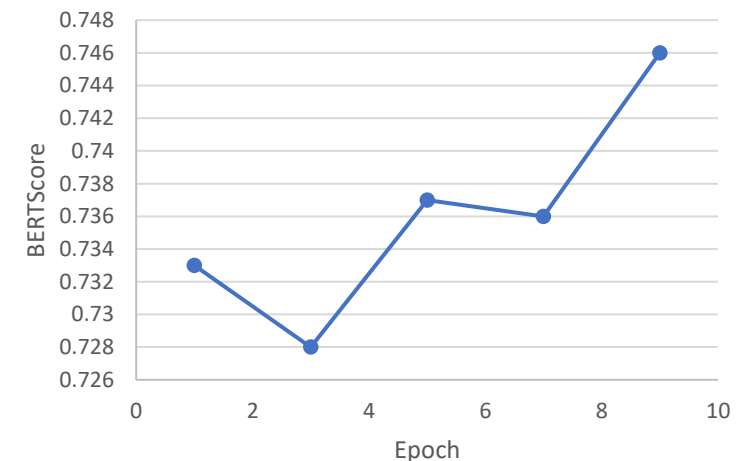
LMMによる経時画像読影レポート自動生成

名古屋大学

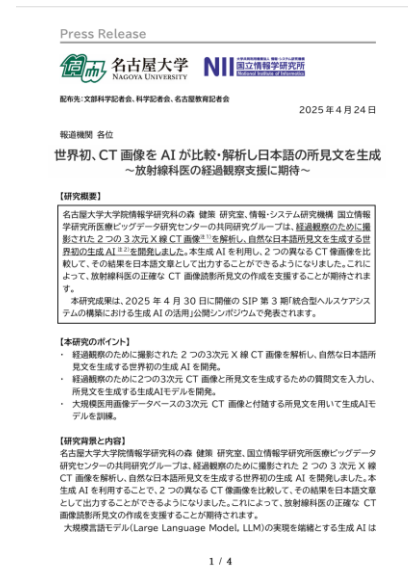
- 経過観察のために撮影された2つの3次元X線CT画像を解析し、自然な日本語所見文を生成する世界初の生成AI
- 経過観察のために2つの3次元CT画像と所見文を生成するための質問文を入力すると所見文を生成
- 大規模医用画像データベースの3次元CT画像と付随所見文を用いて訓練
 - 3.5万ペア (7万件) の3次元CT画像 (約700万断面像)
 - 1万4千人の経時画像
- 追加質問にも対応可能
- 2025/4/24にプレスリリース



モデルの概要

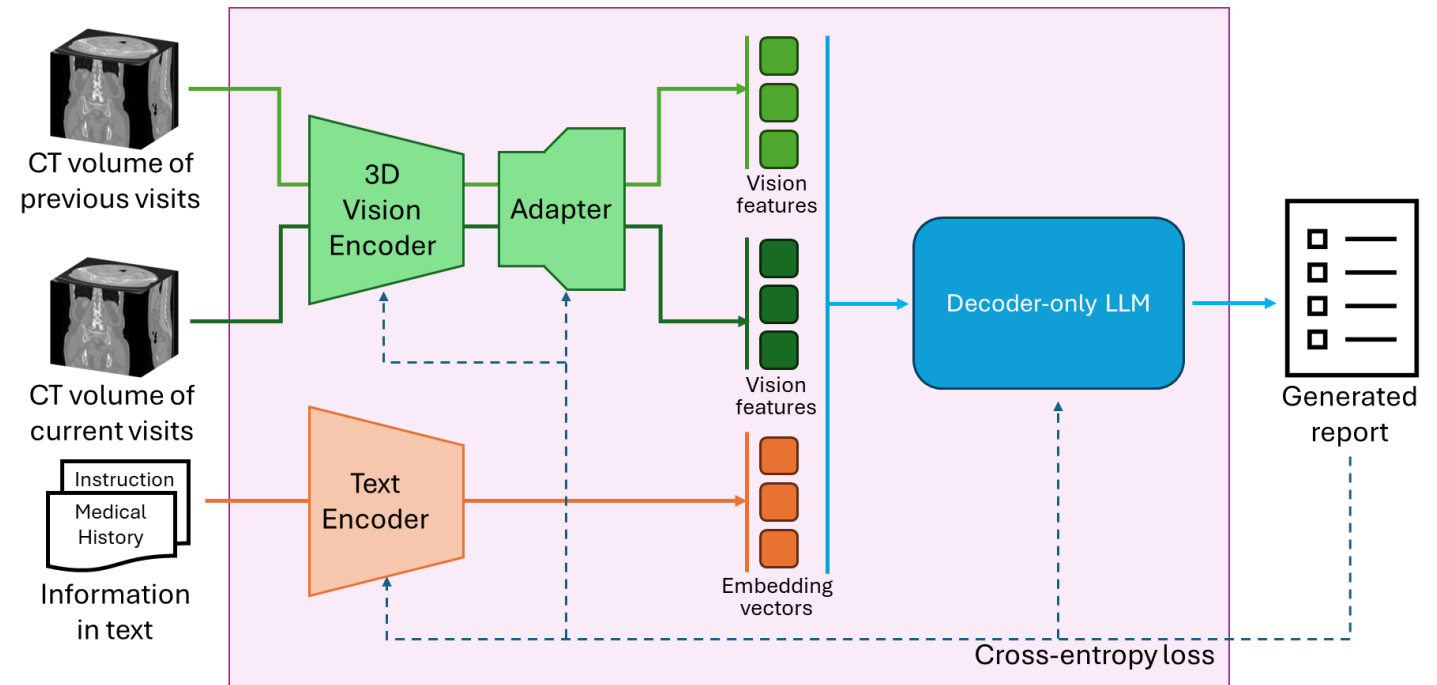


学習進行に伴うBERT Scoreの変化



モデル構造

- モデル入力
 - 過去と現在のCT画像
 - 比較読影レポート
- モデル出力
 - 所見文
- カギとなるコンポーネント
 - 3D Vision Encoder
 - Adapter
 - Text Encoder
 - Vision-Language Decoder



データセットとモデル

- 国立情報学研究所医療ビッグデータ研究センターDBを利用

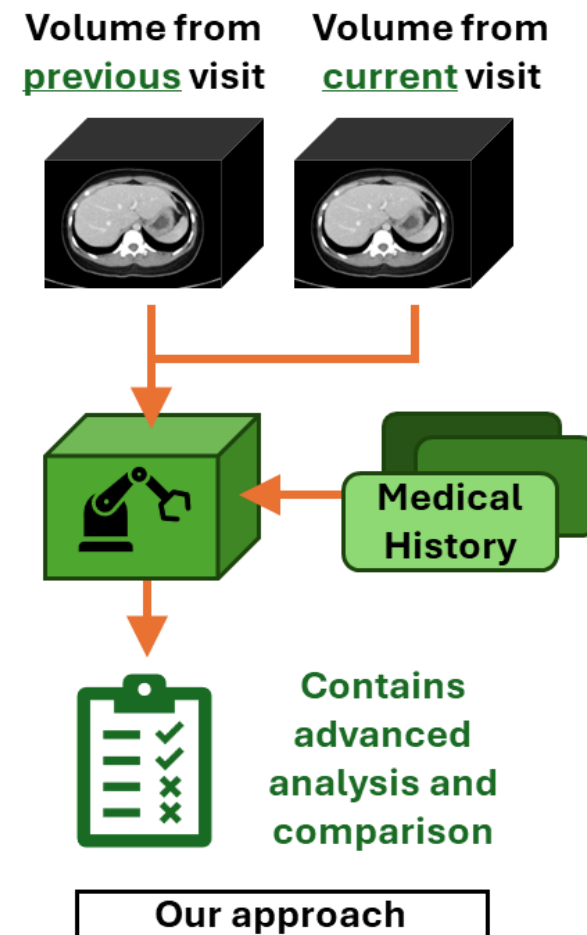
- 14,000 症例 (M:F=8:6)
- 50,000 3D-CT 画像
 - Size on axial plane 512×512 pixel
- 対応所見文

- 35,000 経時画像ペア

- 2つの3次元画像と所見文を入力

- モデル

- Vision-encoder SwinTransformer
- Decoder-only LLM: Gemma-2-9B-it



経時3次元CT像の自動レポート生成

Press Release



配布先: 文部科学記者会、科学記者会、名古屋教育記者会

2025年4月24日

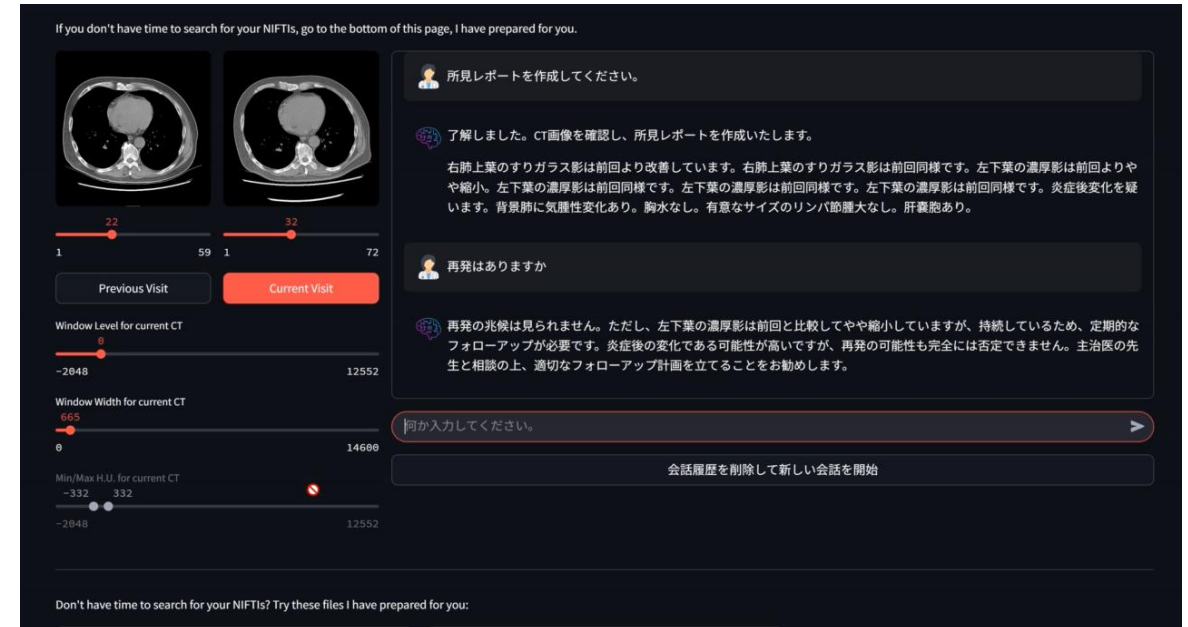
報道機関 各位

世界初、CT 画像を AI が比較・解析し日本語の所見文を生成 ～放射線科医の経過観察支援に期待～

【研究概要】

名古屋大学大学院情報学研究科の森 健策 研究室、情報・システム研究機構 国立情報学研究所医療ビッグデータ研究センターの共同研究グループは、経過観察のために撮影された2つの3次元X線CT画像^{注1)}を解析し、自然な日本語所見文を生成する世界初の生成AI^{注2)}を開発しました。本生成AIを利用し、2つの異なるCT像画像を比較して、その結果を日本語文章として出力することができるようになりました。これによって、放射線科医の正確なCT画像読影所見文の作成を支援することが期待されます。

本研究成果は、2025年4月30日に開催のSIP第3期「統合型ヘルスケアシステムの構築における生成AIの活用」公開シンポジウムで発表されます。



経時画像読影レポート自動生成

If you don't have time to search for your NIFTIs, go to the bottom of this page, I have prepared for you.

22 32
1 59 1 72

Previous Visit Current Visit

Window Level for current CT
0
-2048 12552

Window Width for current CT
665
0 14600

Min/Max H.U. for current CT
-332 332
-2048 12552

所見レポートを作成してください。

了解しました。CT画像を確認し、所見レポートを作成いたします。

右肺上葉のすりガラス影は前回より改善しています。右肺上葉のすりガラス影は前回同様です。左下葉の濃厚影は前回よりやや縮小。左下葉の濃厚影は前回同様です。左下葉の濃厚影は前回同様です。左下葉の濃厚影は前回同様です。炎症後変化を疑います。背景肺に気腫性変化あり。胸水なし。有意なサイズのリンパ節腫大なし。肝嚢胞あり。

再発はありますか

再発の兆候は見られません。ただし、左下葉の濃厚影は前回と比較してやや縮小していますが、持続しているため、定期的なフォローアップが必要です。炎症後の変化である可能性が高いですが、再発の可能性も完全には否定できません。主治医の先生と相談の上、適切なフォローアップ計画を立てることをお勧めします。

何か入力してください。

会話履歴を削除して新しい会話を開始

Don't have time to search for your NIFTIs? Try these files I have prepared for you:

単なる所見文生成だけでなく指示に応じたレポート整形

> RUNNING... Stop

1 1
1 59 1 72

Previous Visit Current Visit

📘 ボタン「Previous Visit」または「Current Visit」をクリックすると WL/WW が調整できます。

Findings:

- **Right Upper Lobe:** Known primary lung cancer has decreased in size compared to the previous scan.
- **Right Lung Apex:** A small nodule is present, similar to the previous scan, suggesting post-inflammatory changes.
- **Intrapulmonary Metastasis:** No lesions suggestive of intrapulmonary metastasis are observed.
- **Lung:** Emphysema is present.
- **Pleura:** No pleural effusion is noted.
- **Lymph Nodes:** No lymph nodes of significant size are enlarged.
- **Liver:** A hepatic cyst is present.

Impression:

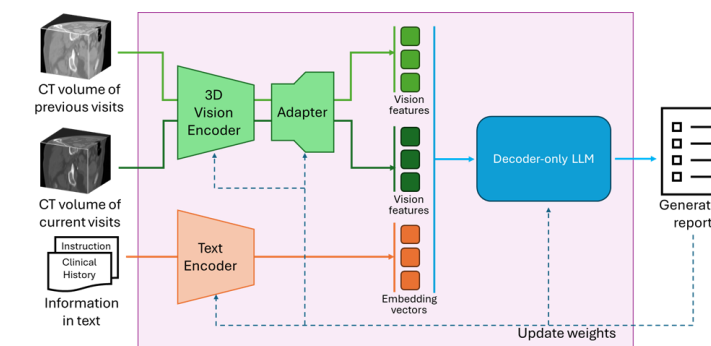
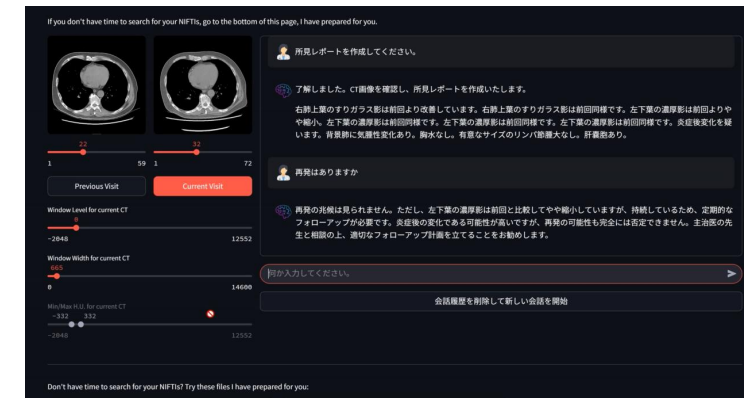
- Decreased size of the known primary lung cancer in the right upper lobe.
- Small nodule |

何か入力してください。

Clear conversation history and start a new one

経時3D CT画像所見文生成モデルベンチマーク

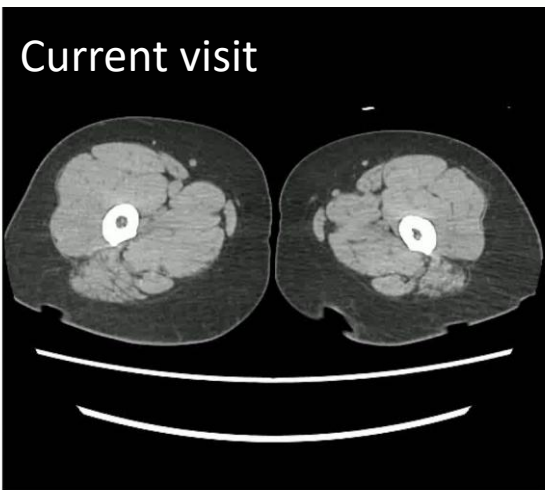
- SIGMA+
 - SIP E-2 名大Gを中心に開発された経時3D CT画像所見文生成モデル
- CT2RepLong^[1]と比較
 - CT2RepLongはNII DBを用いて名大にて再学習
 - 経時3D CT画像所見文生成モデルとして唯一比較可能なモデル
- 全てのスコアにおいてCT2RepLong^[1]を上回る



	BLEU-1	BLEU-2	BLEU-3	BLEU-4	METEOR	ROUGE-L
CT2RepLong	0.298	0.216	0.174	0.144	0.276	0.235
SIGMA	0.331	0.248	0.199	0.162	0.312	0.253

[1] Hamamci, I.E., et al.: CT2Rep: Automated radiology report generation for 3D medical imaging (2024)

レポート生成例



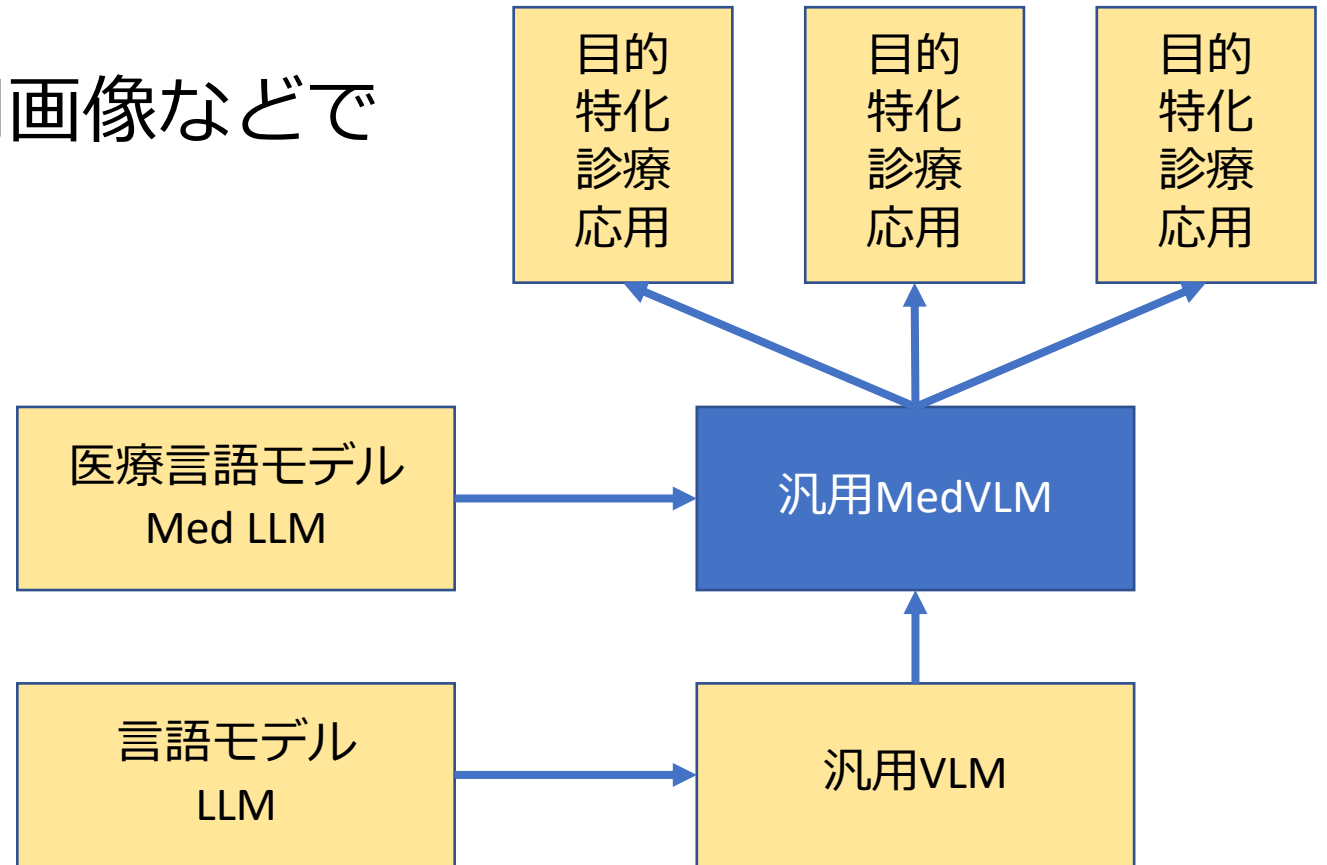
入力3次元画像

	Text	Text (Translated to English)
プロンプト	所見文を作成してください。直腸癌術後。	Please create a report. Postoperative rectal cancer.
AI出力	明らかな局所再発は指摘できません。有意なサイズのリンパ節腫大は指摘できません。明らかな肝転移は指摘できません。肝嚢胞あり。水腎症なし。腹水なし。明らかな肺転移は指摘できません。胸水なし。	No obvious local recurrence can be identified. No significant lymph node enlargement can be identified. No clear liver metastasis can be identified. Liver cysts are present. No hydronephrosis. No ascites. No clear lung metastasis can be identified. No pleural effusion.
放射線科医によるレポート	明らかな局所再発なし。両肺に多発小結節影あり。サイズの増大ありません。一部の結節はリング状に内部欠損みられます。新規結節は指摘されません。有意なサイズのリンパ節腫大はみられません。胸水なし。甲状腺両葉のLDAに大きな変化なし。明らかな肝転移なし。肝嚢胞あり。CVポートデバイスあり。その他、特記所見なし。	No obvious local recurrence. Multiple small nodules in both lungs, with no increase in size. Some nodules show ring-like internal defects. No new nodules identified. No significant lymph node enlargement. No pleural effusion. No significant changes in the low-density areas (LDA) of both thyroid lobes. No clear liver metastasis. Liver cysts are present. CV port device is present. No other notable findings.

Similar to the findings written by radiologist

医用画像VLM位置づけ

- 医用画像を解釈して文章などを生成できる
- 汎用的なモデルをベースに医用画像などでさらに訓練を行う
- 言語モデルも重要な役割
 - 言語モデルは誰かが構築したものを再学習させることが多い
 - 海外製モデルが現在主流



医療画像ビッグデータクラウド基盤 - 蓄積されるのは日々のデータ

大量の医療画像データの収集・蓄積・解析

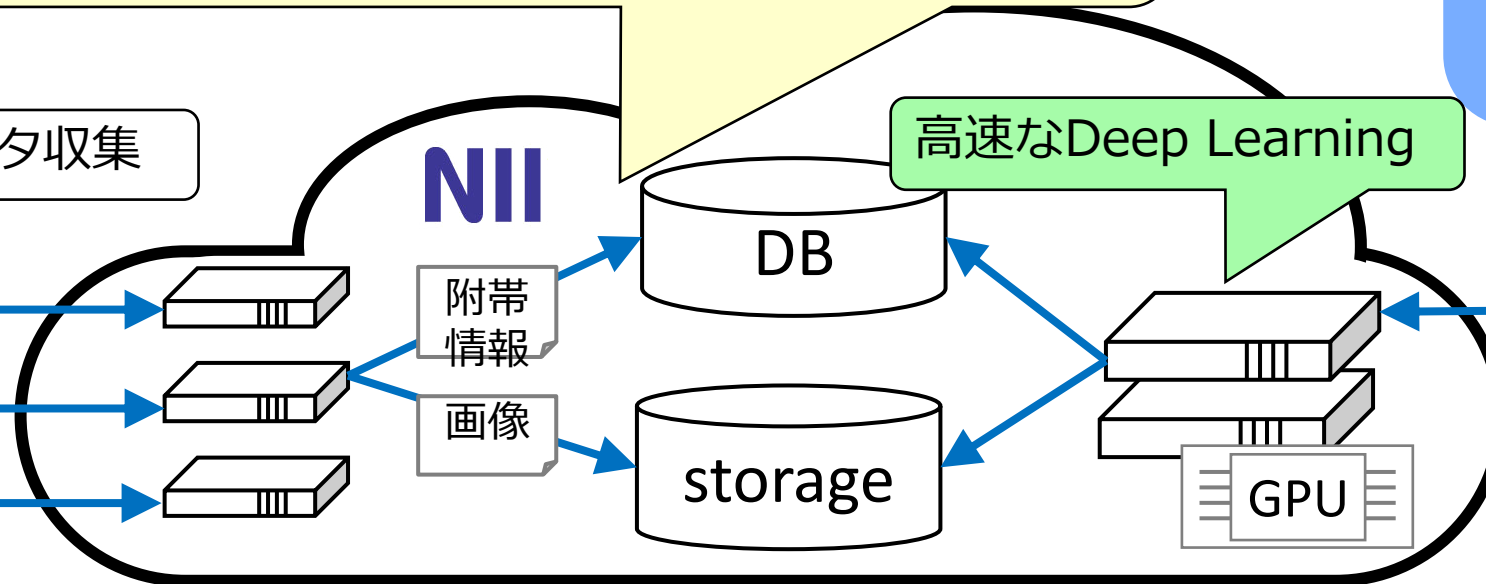
分野（学会）毎・対象疾患毎のデータ形式やファイル構成の違いを吸収
研究者が統一的なインターフェースで医療画像データへアクセスすることを実現

安全かつ高速なデータ収集



医療画像データ
(匿名化)

これまでの画像提供機関
日本医学放射線学会
日本消化器内視鏡学会
日本病理学会
日本眼科学会
日本皮膚科学会
日本超音波医学会
日本心療内科学会



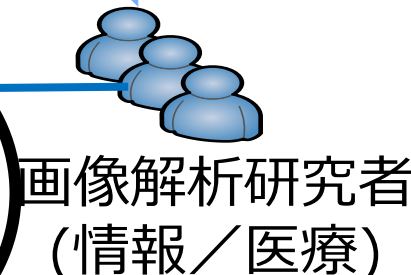
医療画像ビッグデータクラウド基盤 (NII MIDB)

画像検索

前処理

機械学習
(深層学習)

結果確認

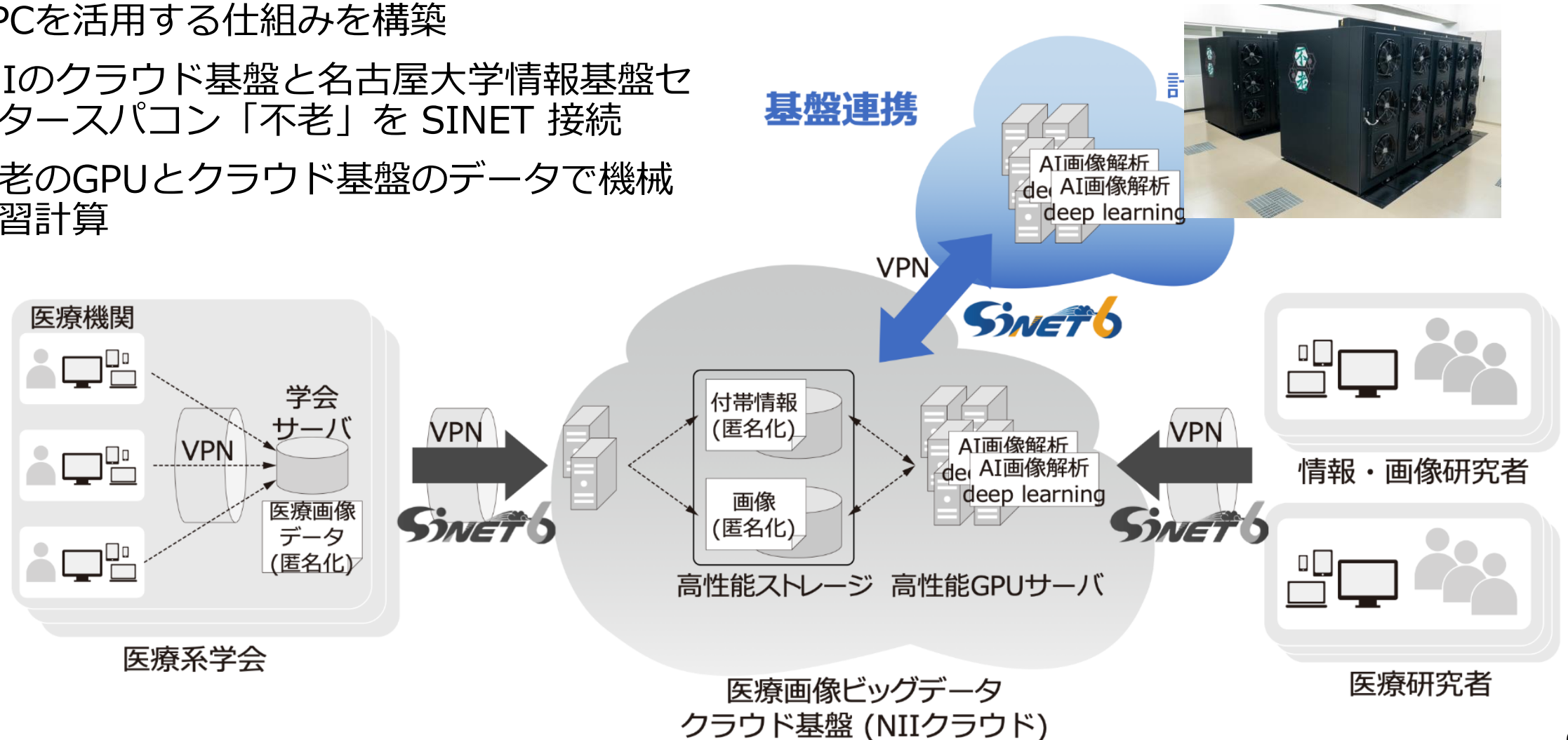


画像解析研究者
(情報/医療)

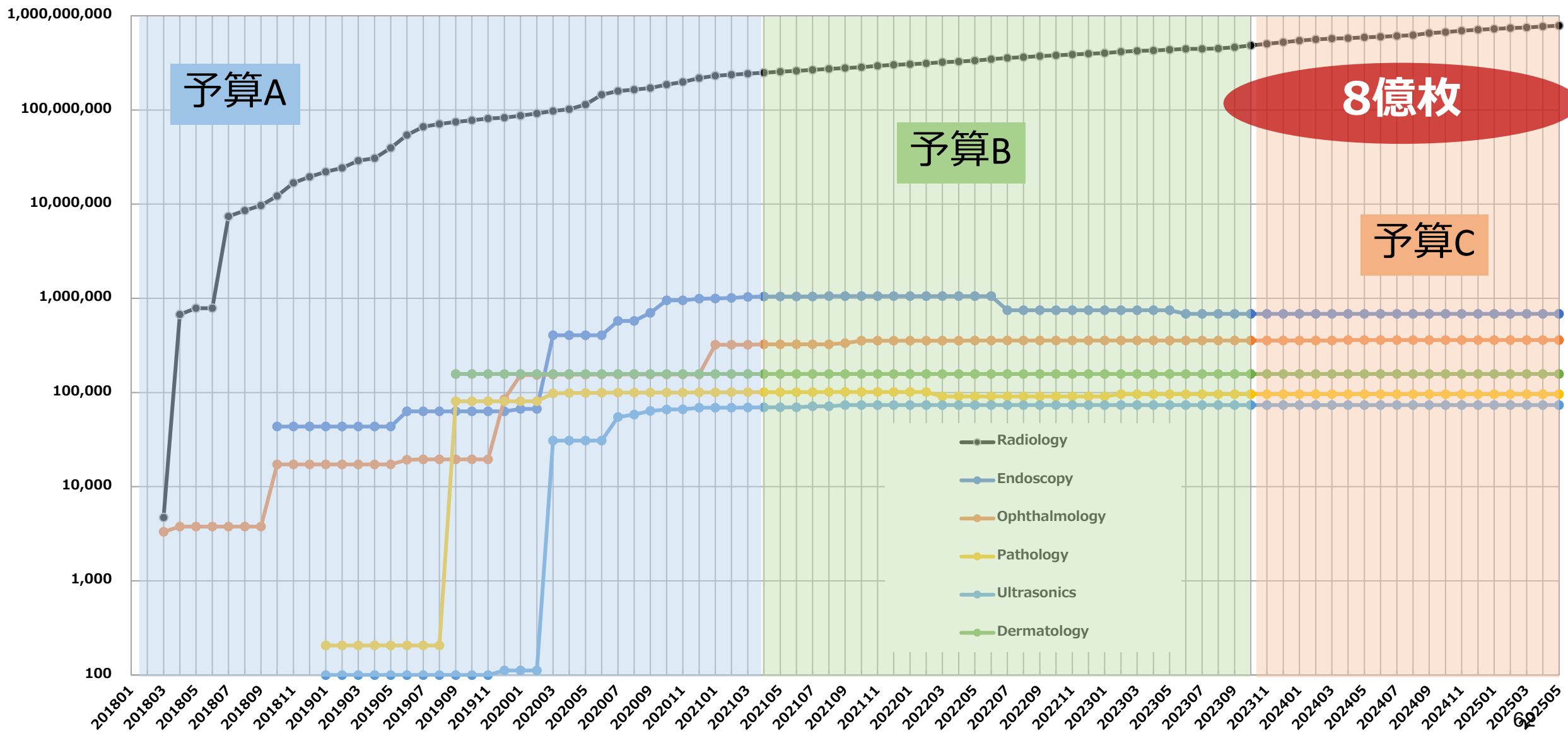
2017年にSINETを活用した画像収集クラウド基盤を構築

DB-GPU基盤連携

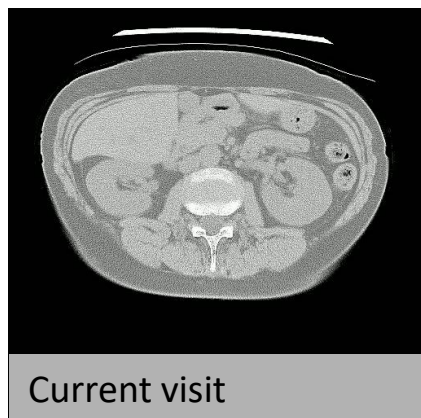
- 計算資源の逼迫に対処するために、外部のHPCを活用する仕組みを構築
- NIIのクラウド基盤と名古屋大学情報基盤センタースパコン「不老」をSINET接続
- 不老のGPUとクラウド基盤のデータで機械学習計算



クラウド基盤に格納された画像数 2025年5月末現在



生成AIによる診断文をどのように評価するか？



Input volumes

Expert: 両側胸水は前回同様。左肺底部の肺虚脱軽度あり。左下葉中枢側の結節は前回同様、経時的に増大あり。左下葉に小さな限局性すりガラス影は前回同様です。右胸膜肥厚に著変ありません。その他、両肺に散在する小結節影に増大傾向なく、炎症後変化など良性結節疑い。有意なサイズのリンパ節腫大なし。右縦隔の軟部影は前回同様です。両腎嚢胞あり。水腎症なし。腹水なし。

(Bilateral pleural effusion is unchanged from last time. There is mild collapse of the left lung base. A nodule in the central left lower lobe is unchanged but gradually increasing. A small localized ground-glass opacity in the left lower lobe is also unchanged. Right pleural thickening shows no significant change. Small scattered nodules in both lungs show no growth and are likely benign post-inflammatory. There is no significant lymph node enlargement. The right mediastinal soft tissue shadow is unchanged. Bilateral renal cysts are present, with no hydronephrosis or ascites.)

Input: 所見文を作成してください。
(Please generate a report.)

SIGMA: 右肺上葉の結節影は前回同様です。両肺に慢性気道炎症疑いあり。前回同様です。胸水なし。有意サイズのリンパ節腫大なし。

(Nodules in the right upper lobe are unchanged. Both lungs show findings suggestive of chronic airway inflammation, unchanged from last time. No pleural effusion. No significant lymph node enlargement.)

Input: 所見文を作成してください。右上葉肺癌術後、特発性乳び胸術後
(Please generate a report. Post-right upper lobe lung cancer surgery, post-idiopathic chylothorax surgery)

SIGMA+: 右肺尖部には浸潤影、すりガラス影あり。前回より増大。肺炎の疑い。右肺上葉の結節影は前回同様です。右肺尖部の濃厚影は前回同様です。背景肺に気腫性変化あり。左胸水は前回同様です。有意なサイズのリンパ節腫大なし。

(In the right lung apex, there is an infiltrative and ground-glass shadow, increased since last time, suggesting possible pneumonia. Nodules in the right upper lobe are unchanged. Dense shadow in the right lung apex is unchanged. Emphysematous changes are seen in the background lung. Left pleural effusion is unchanged. No significant lymph node enlargement.)

Note: Original input and output are in Japanese, the English version was translated by authors, written in the parentheses.

Green for findings match with the expert
Red for findings differ from the expert

医療機器評価における生成AI利用

- 自然言語処理の場合の代表的なメトリック
 - BLEU-n: Based on n-gram matches
 - METEOR: Considers synonyms, stemming, and word order
 - ROUGE-L: Measures longest common subsequence
 - BertScore: Measures semantic similarity
 - r_m (Matching Rate): Fraction of output findings that match the expert's findings
 - GREEN: Evaluate factual correctness and clinical relevance
- 上記の評価メトリックは生成された診断文の正しさを表すか？
 - GPTを用いた評価も行われる
 - プロンプトエンジニアリングで評価

評価結果の例

Method	Natural Language Generation							Clinical Efficacy	
	BLEU-1	BLEU-2	BLEU-3	BLEU-4	METEOR	ROUGE-L	BERTScore	r_m	GREEN
W/o clinical history (SIGMA)	0.313	0.243	0.183	0.147	0.230	0.253	0.730	0.306	0.272
W/ clinical history (SIGMA+)	0.318	0.247	0.196	0.153	0.322	0.287	0.746	0.353	0.361
Below shows performance shown in Ref. [3] as reference. c.f. CT2RepLong is English model, and our model is Japanese model, direct comparison is impossible.									
CT2RepLong[3]	0.374	0.327	0.304	0.401	0.285	0.263	-	-	-

- BLEU-n: Based on n-gram matches
- METEOR: Considers synonyms, stemming, and word order
- ROUGE-L: Measures longest common subsequence
- BertScore: Measures semantic similarity
- r_m (Matching Rate): Fraction of output findings that match the expert's findings
- GREEN: Evaluate factual correctness and clinical relevance

BLEUスコアの仕組み (機械翻訳の評価指標)

1. 参考訳 (正解) と候補訳を比較

参考訳 (Reference):

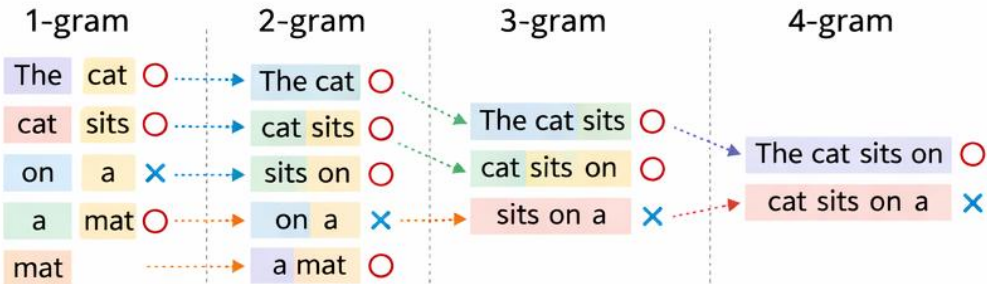
“The cat is sitting on the mat.”

候補訳 (Translation):

“The cat sits on a mat.”



2. n-gramの一致度を計算



3. n-gram適合率 (Precision) を計算

$$1\text{-gram適合率} = \frac{\text{一致数}}{\text{総n-gram数}} = \frac{4}{6}$$

$$2\text{-gram適合率} = \frac{3}{5}$$

$$3\text{-gram適合率} = \frac{2}{3}$$

$$4\text{-gram適合率} = \frac{1}{2}$$

幾何平均で統合

$$\text{BLEU} = \text{BP} \times \exp\left(\sum w_n \log P_n\right)$$

(w_n : 重み, P_n : n-gram適合率)

4. 長さペナルティ (BP)

候補訳が短すぎる場合に調整

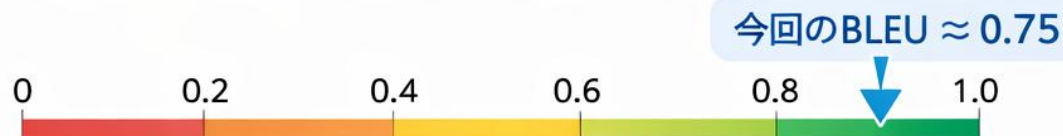
参考訳: “The cat is sitting on the mat.”

候補訳: “The cat sits on a mat.”

短い → ペナルティ

→ $\text{BP} < 1$ でスコア調整

BLEUスコア = 0 ~ 1
(1に近いほど良い翻訳)



文章の評価例 BLEUスコア

• 文例1

- A) 明日雨が降るので傘を持っていくとよいです。
- B) 明日は雨なので傘を持っていきましょう。
- BLEU-4 0.3763



ほぼ同じ意味でも
スコア低い

• 文例2

- A) **明日**雨が降るので傘を持っていくとよいです。
- B) **今日**は雨なので傘を持っていきましょう。
- BLEU-4 0.3561



今日か明日が違うし
スコアも低い

• 文例3

- A) **右**肺上葉に結節影が認められます
- B) **左**肺上葉に結節影が認められます
- BLEU-4 0.9254



左と右を間違える
致命的誤りにも関わらず
スコアが高い

評価における生成AIの利用

単語レベルだけの評価では難しい部分も評価

プロンプト

以下はコンピュータによって生成された診断文と専門医が書いた診断文です。2つの診断文の一致度を評価して0-100のスコアで返してください。

専門医 胃体中部後壁の既知の胃癌は縮小を維持しているようです。以前に腫大を認めた胃小弯リンパ節、傍大動脈リンパ節はいずれも縮小を維持しています。両肺に5mm大までのすりガラス結節や粒状影を複数認めます。前回とほぼ著変ありません。多発肝嚢胞あり。右腎嚢胞あり。胆道系、脾臓、副腎、膵臓に明らかな異常なし。子宮筋腫と思われる腫瘤を複数認めます。骨盤底に腹水貯留あり。骨転移を疑う変化は指摘できません。

コンピュータ 胃壁の肥厚は改善しています。胃幽門部周囲のリンパ節や播種は縮小維持しています。他、明らかに増大傾向なリンパ節は指摘できません。多発肝転移は前回より全体に縮小しています。腹水は減少しています。両肺に散見される小結節影は前回同様です。胸水なし。

一致度スコア：45 / 100

胃病変の評価専門医：胃体中部後壁の既知の胃癌は「縮小を維持」コンピュータ：胃壁の肥厚は「改善」→どちらも「病変の縮小・改善」という方向性で一致。

リンパ節の評価専門医：胃小弯リンパ節・傍大動脈リンパ節「縮小を維持」コンピュータ：胃幽門部周囲リンパ節や播種は「縮小維持」、増大リンパ節なし

→ 部位の表現は少し違いますが、「リンパ節は縮小維持で、新たな増大なし」でほぼ一致。

FDA Request for Public Comment: Measuring and Evaluating AI-enabled Medical Device Performance in the Real-World

Measuring and Evaluating Artificial Intelligence-enabled Medical Device Performance in the Real-World

Request For Public Comment

This Request for Public Comment is intended for discussion purposes only and does not represent draft or final guidance. It is not intended to propose or implement policy changes regarding the evaluation of devices which integrate artificial intelligence (AI), including generative AI (GenAI)-enabled technology. This document is not intended to communicate the FDA's proposed (or final) regulatory expectations but is instead meant to seek early feedback from groups and individuals outside the Agency and advance a broader discussion among the AI healthcare ecosystem on this topic.

The objective of this 'Request for Public Comment' is to obtain comment and feedback from interested parties and the public on a series of questions related to the current, practical approaches to measuring and evaluating the performance of AI-enabled medical devices in the real-world, including strategies for identifying and managing performance drift, such as detecting changes in input and output. Please submit your comments to <https://www.regulations.gov>, Docket No. FDA-2025-N-4203 for 'Measuring and Evaluating Artificial Intelligence-enabled Medical Device Performance in the Real World; Request for Public Comment.' FDA intends to consider all comments timely submitted to this docket (FDA-2025-N-4203) by December 1, 2025, related to this topic.

Background

AI, including GenAI, presents opportunities to improve patient outcomes, advance public health, and accelerate medical innovation. At the same time, these technologies introduce new considerations as it relates to assuring the maintained safety and effectiveness of AI-enabled medical devices across the total product life cycle, and particularly with respect to assessing their performance, safety, and reliability after deployment in real-world settings.

The U.S. Food and Drug Administration (FDA or the Agency) is seeking information from interested parties and the public on best practices, methodologies, and approaches for measuring and evaluating real-world performance of AI-enabled medical devices. This includes

パフォーマンス評価の重要性

AI医療機器の安全性、効果、信頼性を評価するための指標やメトリクスの定義と重み付けが求められています。

実世界での使用におけるパフォーマンス評価の時間枠についても意見を求めています。

評価方法とインフラ

AI医療機器の展開後のパフォーマンスを監視するためのツールや方法論についての情報を収集しています。

人間の専門家によるレビューと自動監視のバランスを取る方法、及びそれぞれの利点と欠点についても意見を求めています。

データソースと品質管理

パフォーマンス評価に使用するデータソース（電子健康記録、デバイスログなど）や、データの質、完全性、相互運用性の課題への対処方法についての意見を求めています。

臨床結果やユーザーフィードバックをモデル更新に組み込む効果的な方法についても関心があります。

モニタリングのトリガーと対応策

追加評価が必要となるトリガーや、実世界でのパフォーマンス劣化に対する定義と対応方法についての意見を求めています。

人間とAIの相互作用

臨床使用パターンやユーザーの相互作用がAI医療機器のパフォーマンスに与える影響についての観察結果を求めています。

安全で効果的な使用を維持するためのデザイン機能やユーザートレーニングの戦略についても意見を求めています。

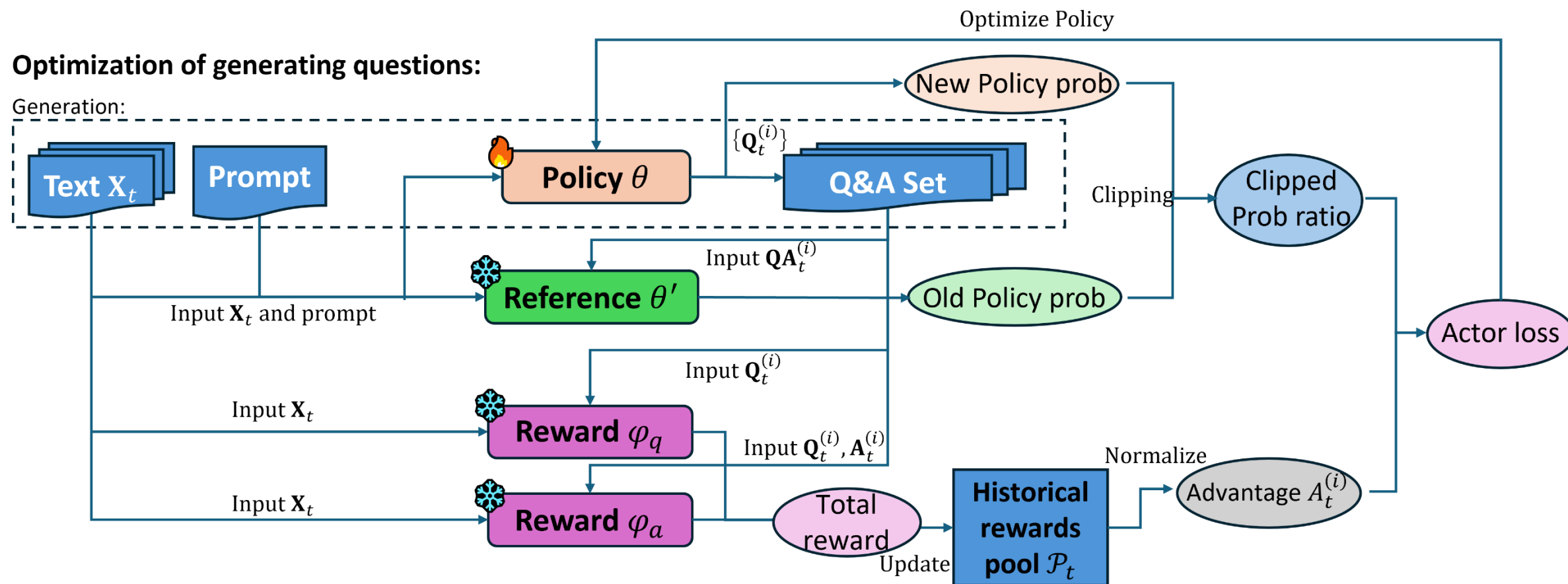
追加の考慮事項とベストプラクティス

実世界での検証システムの開発と実施における他の考慮事項やベストプラクティスについての意見を求めています。

実施上の障壁や患者のプライバシー保護に関するアプローチについても関心があります。

<https://www.regulations.gov/document/FDA-2025-N-4203-0001>

QA pair generation from finding texts



Generated QA pairs and scores

finding1: 腹部造影CT: 前日CTと比較します。膵尾部仮性嚢胞に対してEUS-CD後。外瘻チューブ留置後。前回認めた胃と脾の間の液貯留は消失しています。その他、明らかなcavityを認めません。膵頭部癌に対してGnP併用RT後。膵頭部の原発巣のサイズは著変ありません。総胆管は径約12mmで前回より拡張、腫瘍の総胆管浸潤によるものと思われます。SMA周囲の軟部影は大差ありません。主膵管拡張や膵腫大は前回より改善、急性膵炎の経過と思われます。肝:n.p.胆:胆石あり。脾:n.p.副腎:n.p.腎:左腎に小嚢胞あり。腹水なし。有意なサイズのリンパ節腫大なし。

response:

Q: 膵臓にどのような病変がみられましたか? score: 8.5625

A: 膵尾部に仮性嚢胞が認められています。score: 0.51171875

Q: 胃と脾の間に液貯留があったのですが、その状態は現在どうなっていますか? score: 2.09375

A: 現在は消失しています。score: 2.25

Q: 総胆管の直径はどれくらいですか? score: 15.0625

A: 約12mmで前回よりも拡張していることが分かります。score: 5.625

Q: 腹膜やお腹の中に異常はみられますか? score: 10.0625

A: 腹水はなく、有意なサイズのリンパ節腫大も認められていません。score: -5.125

Q: 膵頭部癌とは何を指すのでしょうか? score: 2.859375

A: 膵頭部に存在する癌です。score: 0.98828125

Q: 主膵管や膵臓自体の大きさはいかがでしょうか? score: 2.1875

A: 前回より改善しており、急性膵炎の経過であると考えられます。score: 0.349609375

Abdominal contrast-enhanced CT: Compared with the previous day's CT. After EUS-CD for a pancreatic tail pseudocyst. After placement of an external drainage tube. The fluid retention between the stomach and spleen seen previously has disappeared. No other obvious cavity is observed. After GnP with RT for pancreatic head cancer. The size of the primary lesion in the pancreatic head shows no remarkable change. The common bile duct measures about 12 mm in diameter and is more dilated than previously, thought to be due to tumor invasion of the common bile duct. The soft tissue shadow around the SMA shows no significant change. Dilatation of the main pancreatic duct and pancreatic swelling have improved compared to the previous study, considered to reflect the course of acute pancreatitis. Liver: n.p. Gallbladder: gallstones present. Spleen: n.p. Adrenal glands: n.p. Kidneys: a small cyst in the left kidney. No ascites. No lymph node enlargement of significant size.

Responses:

Q: What kind of lesion was observed in the pancreas?

A: A pseudocyst was observed in the pancreatic tail.

Q: There was fluid retention between the stomach and spleen; what is its status now?

A: It has now disappeared.

Q: What is the diameter of the common bile duct?

A: About 12 mm, and it is more dilated than before.

Q: Were there any abnormalities in the peritoneum or abdomen?

A: There is no ascites, and no enlarged lymph nodes of significant size are observed.

Q: What does pancreatic head cancer refer to?

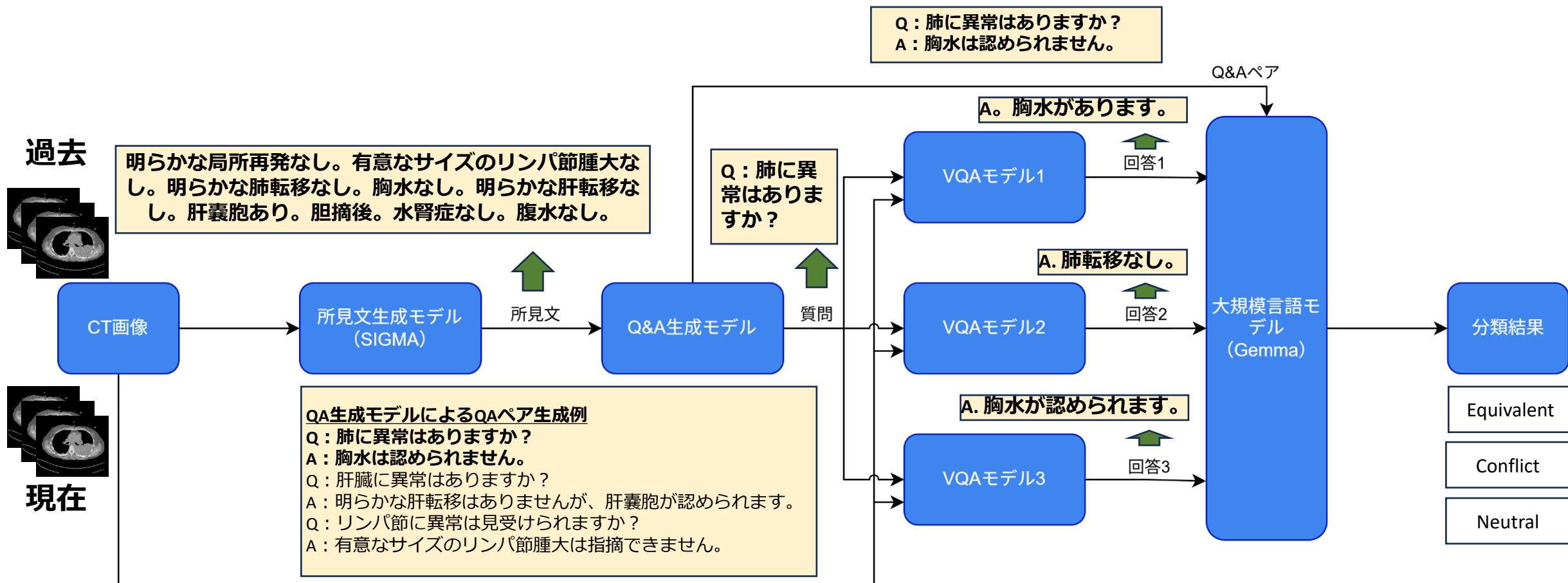
A: It refers to cancer located in the pancreatic head.

Q: How are the main pancreatic duct and the pancreas itself in terms of size?

A: They have improved compared to the previous study, considered to reflect the course of acute pancreatitis.

E-2 所見文の誤り検出に向けた取り組み

- 自動生成された所見文からQAを自動生成し、3つのVQA-VLMが生成するAと自動生成された所見文からAとの一致を自動評価



誤り検出の例

<医師の所見文>

腹腔内に術後変化を疑うairや脂肪織濃度上昇あり。右側胸部～腹部にかけて皮下気腫あり。術後変化を疑います。肝転移なし。肝嚢胞あり。胆、膵、脾、副腎、腎に特記所見なし。肝周囲に少量腹水あり。有意な大きさのリンパ節腫大なし。両肺背側に無気肺あり。含気のある肺に明らかな活動性肺炎や転移は認めません。両側ごく少量の胸水あり。

<生成された所見文>

明らかな局所再発は指摘できません。有意なサイズのリンパ節腫大は指摘できません。明らかな肝転移は指摘できません。肝嚢胞あり。胆嚢結石あり。胆嚢腺筋腫症(s/o)。水腎症なし。腹水なし。明らかな肺転移は指摘できません。胸水なし。

Q&Aペア	VQAの回答		LLMによる分類	実際の判定
Q：胆嚢に異常はありますか？ A：胆嚢結石と胆嚢腺筋腫症が認められます。	VQA1	特に問題はないようです。	Conflict	Conflict
	VQA2	特に問題が指摘されていません。	Conflict	
	VQA3	著変が認められません。	Conflict	

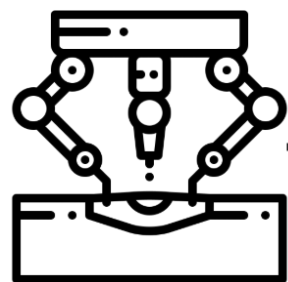
「s/o」は「suspicion of」の略記。日本語で「～の疑い」の意

腹腔鏡映像に対するVQA

Input: Image + Template Question
eg. Please describe this image.

Output : Description / Answer
eg. This image shows a scene XXX.

Surgery Image



User Query

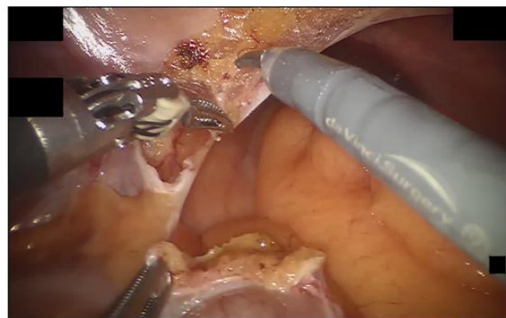


Description



Answer

腹腔鏡映像に対するVQA



Please describe this image.

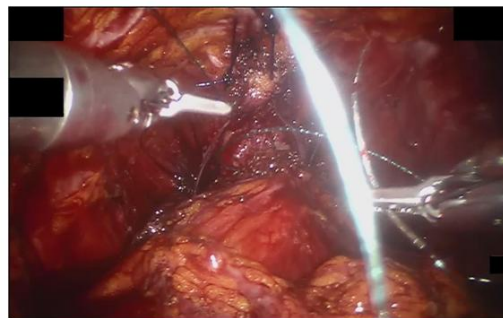


This image shows a scene from an endoscopic surgical procedure. There are 3 instruments in this picture. The Prograsp Forceps is holdstilling. The Bipolar Forceps is holdpulling. The Monopolar Curved Scissors is traveling.

How many instruments are there in this image?



3



Locate all visible surgical instruments in this image and return their bounding boxes.

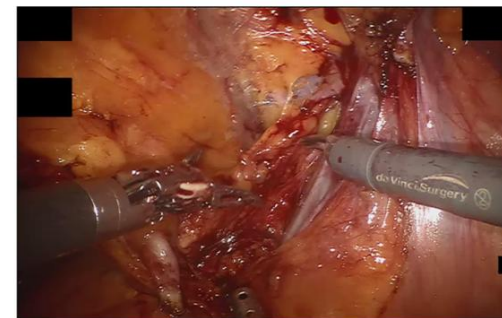


```
[{'bbox_2d': [918, 340, 1257, 566], 'label': 'Large Needle Driver'}, {'bbox_2d': [0, 50, 559, 366], 'label': 'Large Needle Driver'}]
```

Tell me the instrument's number.



There are two Large Needle Drivers in this image.



Locate the Monopolar Curved Scissors in this image and return its bounding box.



```
{'bbox_2d': [709, 318, 1259, 563], 'label': 'Monopolar Curved Scissors'}
```

Locate the Monopolar Curved Scissors in this image and return its bounding box.



```
{'bbox_2d': [528, 699, 636, 783], 'label': 'Suction Instrument'}
```

Vision Language Action

- 画像と言語からロボット動作コマンド

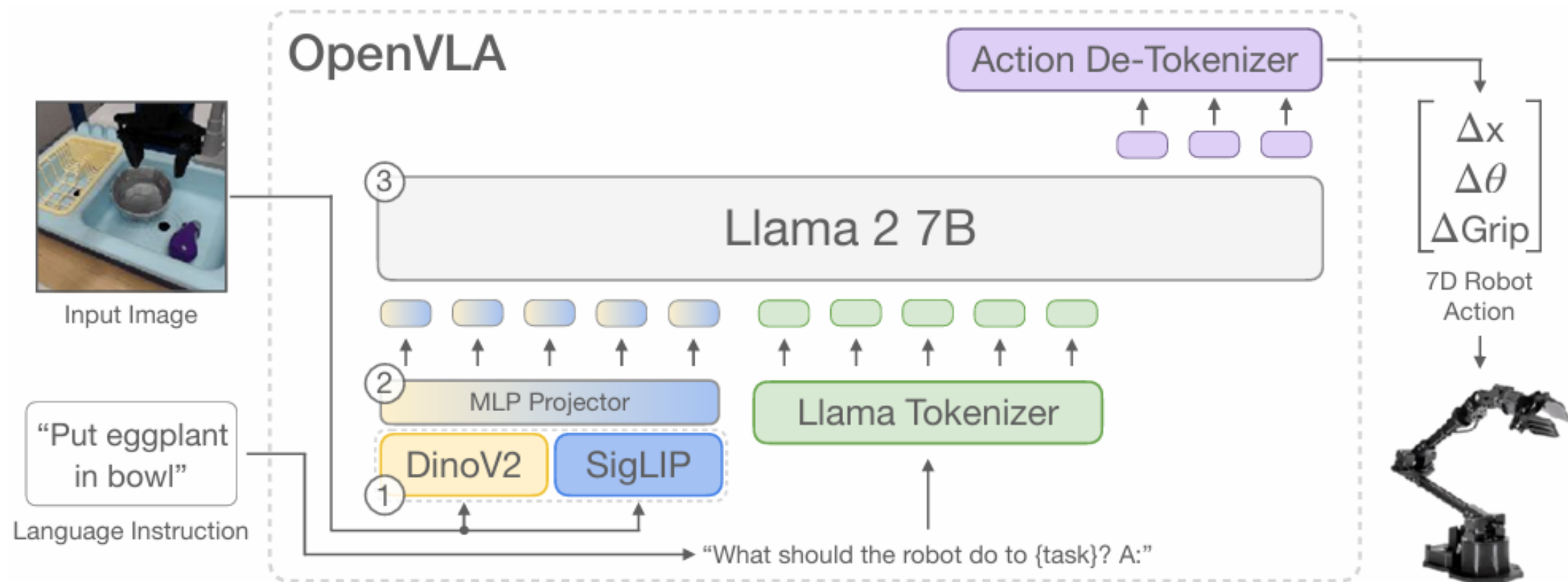
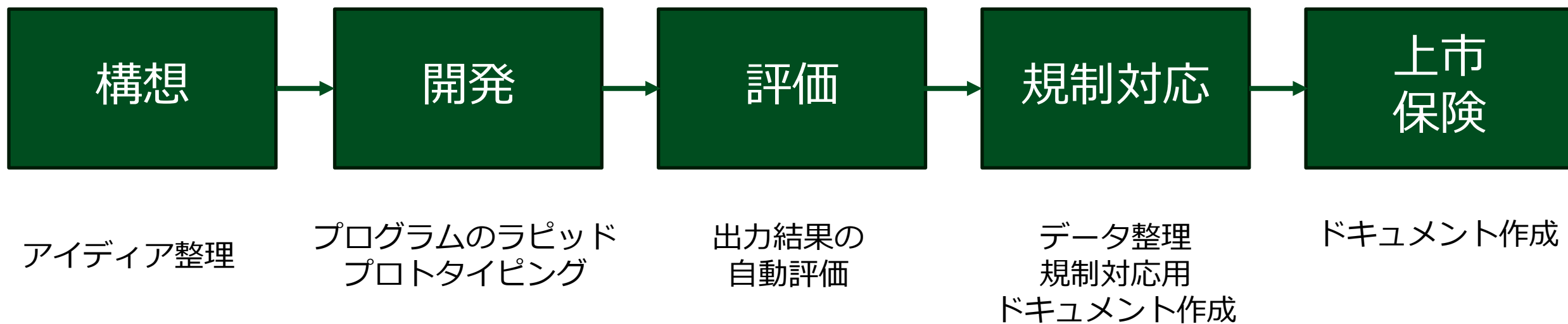


Figure 2: **OpenVLA model architecture.** Given an image observation and a language instruction, the model predicts 7-dimensional robot control actions. The architecture consists of three key components: (1) a **vision encoder** that concatenates Dino V2 [25] and SigLIP [79] features, (2) a **projector** that maps visual features to the language embedding space, and (3) the **LLM backbone**, a Llama 2 7B-parameter large language model [10].

生成AIによる医療機器開発



**生成AIを用いた医療機器開発は
 生成AIを用いた医療機器の開発だけでなく
 医療機器開発プロセスの迅速化などを目指した生成AIの利活用も含まれる**

生成AIによるプログラムのラピッドプロトタイピング

「眼底画像（2D RGB JPEG）から年齢を回帰推定する PyTorch コードを作成してください。ResNet50（torchvision が配布している ImageNet 事前学習モデル）を使用し、最終層のみ 1 出力の回帰に置き換えてください。train.csvならびにval.csvには image_path, age の列があり、test.csvには image_pathのみがあります。data/images/ に画像がある前提で、学習・検証・推論まで行い、テスト画像の推定結果を pred_test.csv に保存するフルスクリプトを作成してください。」（ChatGPTによるプロンプト修正を含む）

```
import os
from typing import Tuple
import pandas as pd
from PIL import Image
import torch
import torch.nn as nn
from torch.utils.data import Dataset, DataLoader
from torchvision import models, transforms
```

ChatGPTによるコード生成例

```
# =====
# Dataset 定義
# =====
class FundusAgeTrainValDataset(Dataset):
    """眼底画像 + 年齢(回帰) 用 Dataset"""
    def __init__(self, csv_path: str, img_root: str, transform=None):
        """
        Args:
            csv_path: image_path, age を含む CSV
            img_root: 画像のルートディレクトリ
            transform: torchvision.transforms など
        """
        self.df = pd.read_csv(csv_path)
        self.img_root = img_root
        self.transform = transform
```

```
if "image_path" not in self.df.columns or "age" not in self.df.columns:
    raise ValueError("CSVには 'image_path' と 'age' 列が必要です。")
```

```
# =====
# Train / Eval
# =====
def train_one_epoch(
    model: nn.Module,
    dataloader: DataLoader,
    criterion,
    optimizer,
    device: torch.device,
    epoch: int,
) -> Tuple[float, float]:
    model.train()
    running_mse = 0.0
    running_mae = 0.0
    total = 0
```

迅速にプロトタイピングが可能であり医療機器開発の加速に貢献

FDA Total Product Lifecycle Considerations for Generative AI Enabled DevicesにおけるAI Lifecycle

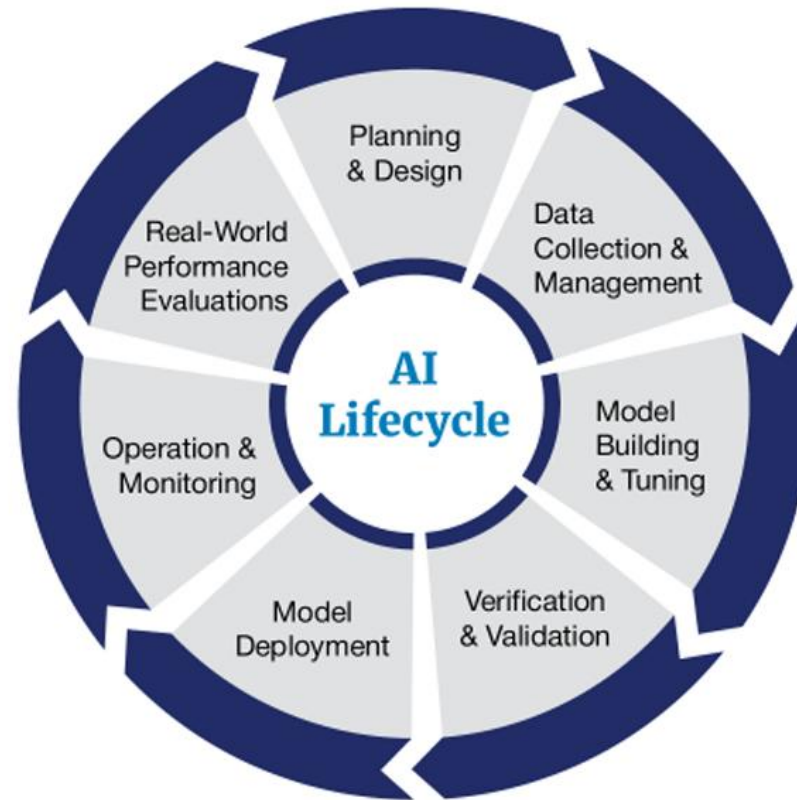
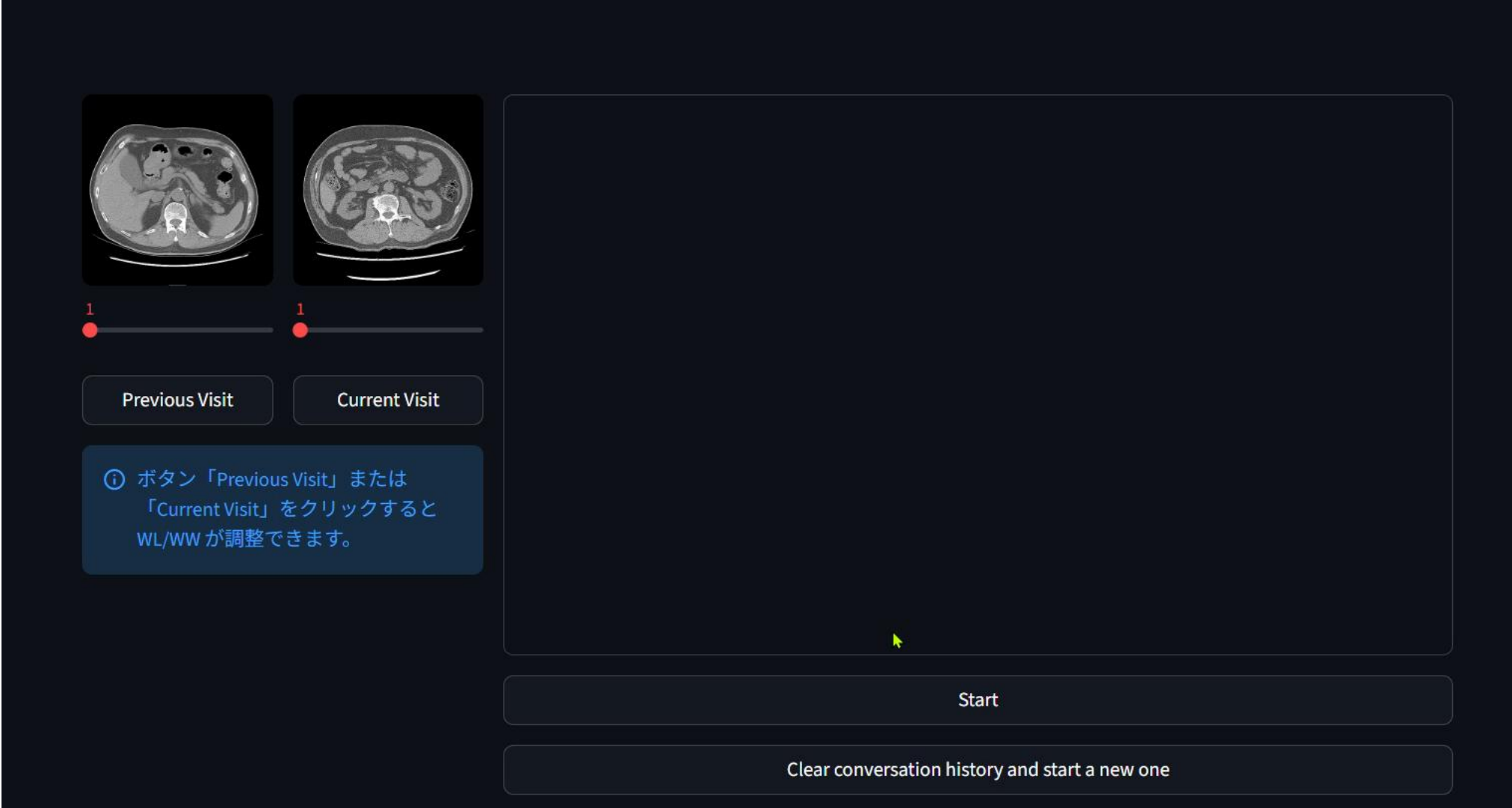


Figure 1. AI Lifecycle

AI vs AI (Qを作りAを返しQを作りAを返す)



The screenshot shows a dark-themed interface for medical image analysis. On the left, two axial CT scan images are displayed side-by-side. Each image has a red dot and the number '1' above it, with a horizontal slider below. Below the sliders are two buttons: 'Previous Visit' and 'Current Visit'. A blue information box contains the text: 'ボタン「Previous Visit」または「Current Visit」をクリックするとWL/WWが調整できます。' (Clicking the 'Previous Visit' or 'Current Visit' buttons allows you to adjust WL/WW). The main area on the right is a large, empty rectangular box. At the bottom of the interface are two buttons: 'Start' and 'Clear conversation history and start a new one'.

1960年代の対話Eliza (人工無能とも呼ばれる)

```
Linux taka 5.10.0-36-686-pae #1 SMP Debian 5.10.244-1 (2025-09-29) i686
```

```
1 updates could not be installed automatically. For more details,
see /var/log/unattended-upgrades/unattended-upgrades.log
```

```
The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.
```

```
Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
```

```
You have mail.
```

```
Last login: Fri Feb 20 17:40:00 2026 from 10.0.0.1
```

```
mori@taka:~$
```

```
mori@taka:~$
```

```
mori@taka:~$ emacs
```

```
mori@taka:~$ emacs
```

```
mori@taka:~$
```

言語を介したAIとの対話 -画像特徴言語化の時代-

