

た、「一定の基準を満たす場合、韓国に拠点を持つ国内代理人を指定することが義務付けられる」のですが、具体的な基準はまだ示されていません。36条では、「利用者数、売上高などが大統領令で定める基準に該当する者は」としています。罰則については最高で3000万ウォン（約300万円）であり、日本の個人情報保護法違反の上限額と同程度であるため、厳しい制裁とは言い難いです。

技術講演会第1回 医療用マルチモーダルモデル  
(東京大学 原田 達也)

## 戦略的イノベーション創造プログラム (第3期) 補正予算「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」

- テーマ1 医療LLM基盤の研究開発・実装
- **テーマ2 医療LMM基盤の研究開発・実装**
  - 高信頼・高性能医療LMMの研究開発と社会実装
  - 研究開発責任者：原田達也（東京大学）
- テーマ3 ソリューション/アプリケーション開発
- テーマ4 医療データ基盤の構築と運営手法の検討

医療用のマルチモーダルモデルについて、簡単にお話しをさせていただきたいと思っています。

# 1. 医療 LLM 基盤の研究開発

## Gemini


<https://deepmind.google/technologies/gemini/>

Gemini is a new family of multi modal models and exhibits remarkable capabilities across image, audio, video, and text understanding.

**Prompt**

Here's a solution to a Physics problem by a student.

23. A skier slides down the frictionless slope as shown. What is the skier's speed at the bottom?



Solution:

The conservation of energy means that the energy of the skier is same at the beginning and at the end.

Start:  $E = mgh$

End:  $E = \frac{1}{2}mv^2$

$\Rightarrow mgh = \frac{1}{2}mv^2$

$\Rightarrow v = \sqrt{2gh} = \sqrt{2 \times 9.81 \times 40} = 39.6 \text{ m s}^{-1}$

Try to reason about the question step by step.

Did the student get the correct answer? If the solution is wrong, please explain what is wrong and solve the problem.

Make sure to use LaTeX for math and round off the final answer to two decimal places.

**Response**

The student did not get the correct answer. The student made a mistake in the calculation of the potential energy at the start of the slope. The potential energy at the start of the slope is given by  $mgh$ , where  $m$  is the mass of the skier,  $g$  is the acceleration due to gravity, and  $h$  is the height of the slope. The student used the length of the slope,  $L$ , instead of the height,  $H$ , in the calculation of the potential energy.

The correct solution is:

The conservation of energy means that the total energy of the skier is the same at the beginning and at the end of the slope.

At the start of the slope, the skier has only potential energy, which is given by  $mgh$ .

At the end of the slope, the skier has only kinetic energy, which is given by  $\frac{1}{2}mv^2$ .

Therefore, we can write:

$$mgh = \frac{1}{2}mv^2$$

Canceling out the mass  $m$  on both sides, we get:

$$gh = \frac{1}{2}v^2$$

Solving for  $v$ , we get:

$$v = \sqrt{2gh}$$

Substituting the given values, we get:

$$v = \sqrt{2 \times 9.81 \times 40} = 28.01 \text{ m/s}$$

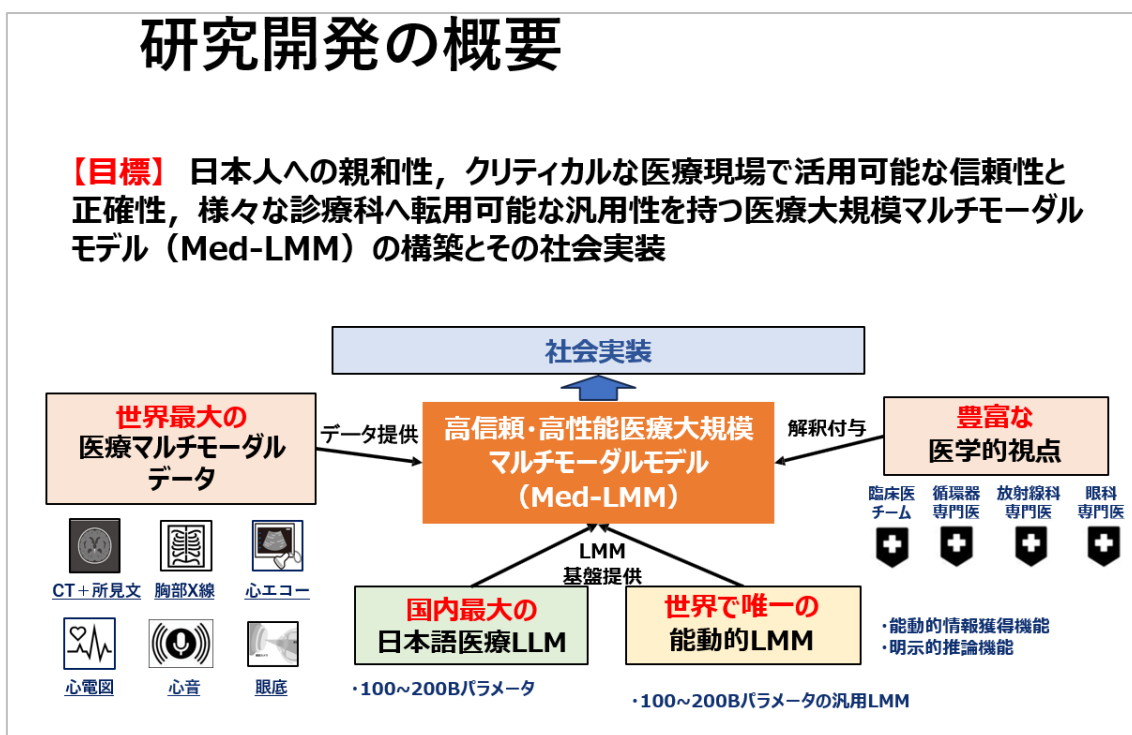
Therefore, the skier's speed at the bottom of the slope is 28.01 m/s.

マルチモーダルモデルとは何かという話ですが、これは Google が作った Gemini というマルチモーダルモデルです。このマルチモーダルモデルは、自然言語の入力だけでなく、画像や音声など、いろいろなものを入力して、モデル自体が知的に振る舞っているというものになります。

例えば、この手書きの物理の問題ですが、これはノートに殴り書きをしたようなもので、画像なのです。絵があって、手書きでいろいろ書いてあります。人が滑っていくと最終的にはどのくらいの速度になりますか、というのを生徒が解いた解答で、これが合っているかきちんとチェックしてくださいと、Gemini というマルチモーダルモデルに聞きます。もし間違っていたら、どう修正すればいいですか、それも教えてくださいと聞きました。すると、きちんと、この図の中の絵や手書きの文字を理解して、

運動エネルギーの計算が間違っているから、きちんと解くとこのようになります、というところまでやってくれます。

これは一般的なマルチモーダルモデルですが、これを医療に应用することを考えています。この真ん中は物理の問題の図になっていますが、これが CT の画像、MRI の画像であり、医師が書かれた手書きの文字など、そこまでは想定はしていませんが、様々な画像や音声のデータを入力して、これに関する質問を投げかけると、適切に答えてくれるものを作りたいということです。



そのためには何が必要かという、言語モデルがベースになります。これをベースとして、さらに医療用のマルチモーダルデータ、例えば CT の画像と所見文がペアになっているもの、ここでは 5 億ペアくらいのを今は想定していますが、その画像を使って、自然言語での医療用のモデルなどを拡張することによって、大規模なマルチモーダルモデルを作ることになっています。

# 研究開発テーマ

- 予後予測により時系列情報を正しく活用した世界初のMed-LMMを実現する。
- 循環器読影試験で、専門医がMed-LMMを活用することで有意にスコアが改善することを目指す。
- 複数の企業と連携により、得られた収益の一部をMed-LMM運用へ活用する。
- 日本人三大死亡原因の病気を診る：(3-1) 循環器診療、(3-2) がん診療、(3-3) 保健行政としての健診

## 研究開発テーマ (1) 汎用LMM基盤の開発

日本語大規模言語モデルと、大規模画像エンコーダを統合し、汎用LMMを開発する。この汎用LMMに明示的推論機能や能動的知識獲得機能を実装することで、信頼性を担保しつつ、従来の受動的LMMを上回る質問応答性能を実現する。



## 研究開発テーマ (2) 汎用Med-LMM基盤の開発

汎用LMM基盤を、世界最大の医療マルチモーダルデータで追加学習し、汎用Med-LMMへと進化させる。



## 研究開発テーマ (3) 診療、保健行政、臨床研究支援を行う医療用LMMソリューションの開発

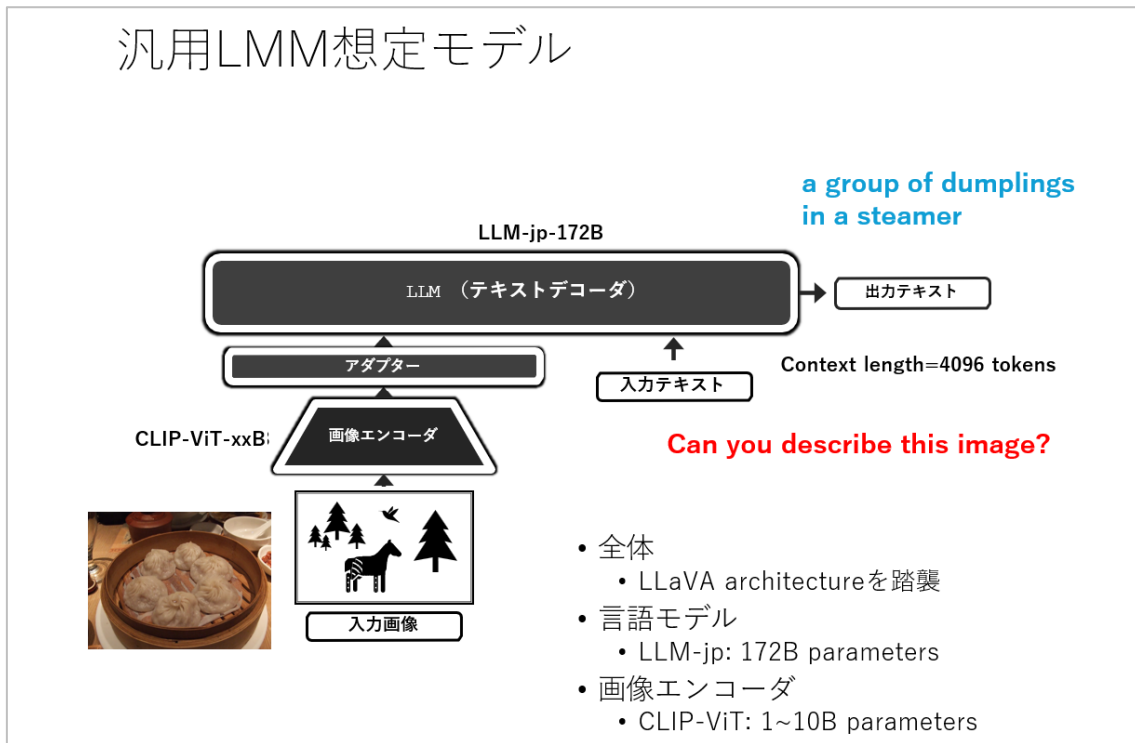
具体的医療シナリオに沿った特化型Med-LMMソリューションへ展開する。

(3-1) 循環器診療、(3-2) がん診療、(3-3) 保健行政としての健診、の三つの診療業務を社会実装を見据えたテーマとして設定する。

これを開発する順番はややこしくて、医学部の学生の教育課程に似ていると思っています。例えば、18歳くらいまでは、学校でジェネラルな勉強をするわけです。物理、化学、生物、数学もやることで、汎用的な基盤を持っているわけです。18歳で医学部に入ると、医学の汎用的な知識を勉強するわけです。卒業間近になると、どこの専門医になるかを選び、そこで特化した知識を勉強していくことになります。このマルチモーダルモデルもまったく同じような手順を踏んで、学習させることが必要になります。

ですので、このテーマ(2)の開発手順としての1番目は、汎用的なマルチモーダルモデルの基盤を作ることを決めます。2番目のステップとして、医師であればこのくらい知っていないとまずいなというようなかたちで、汎用的な医療用のマルチモーダルモデルを作ります。そのあとに、「特化型」と呼んでいます、それぞれの診療科に適合したマルチモーダルモデルにするということを行います。具体的には循環器やがん診療、保健行政としての健診などでの特化型のマルチモーダルモデルを作っていきます。

この三つのステップを踏んでいくというところが、ランゲージモデルとは違います。



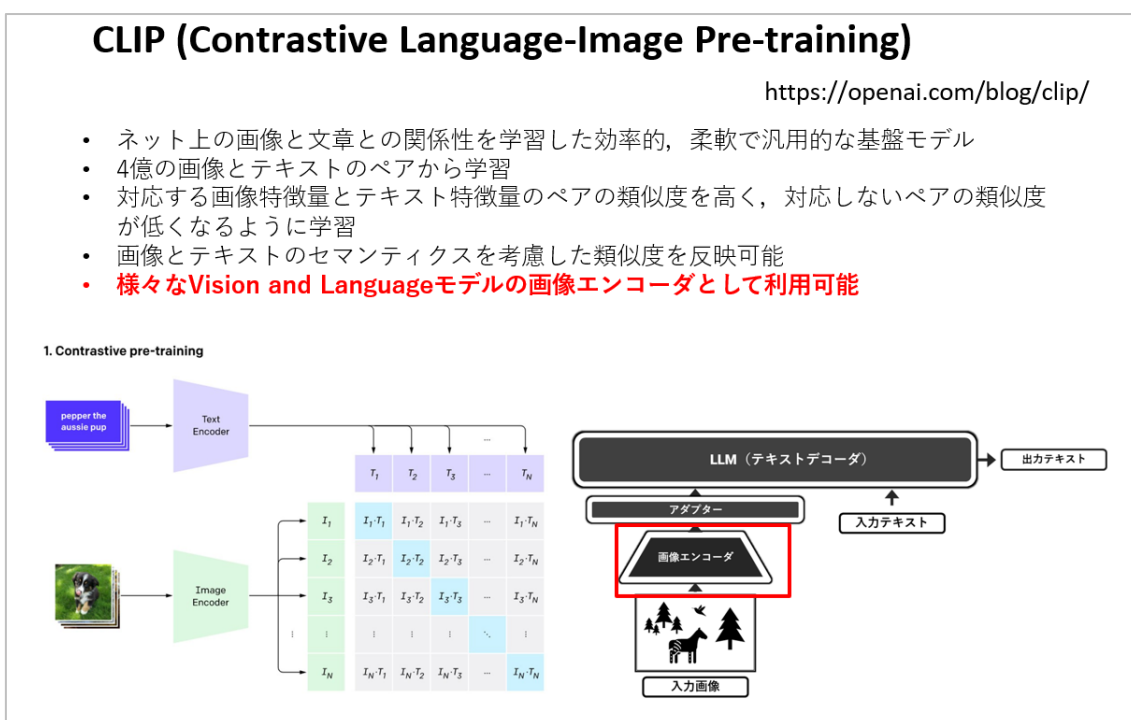
このマルチモーダルモデルはどう作るのかという話です。基本的には、言語モデルが基盤となっています。モデルに対し、テキストを入れると、テキストが返ってくるというモデルが言語モデルですが、これをマルチモーダルモデルにするには、画像と質問を入れて、それに対する答えを出すようなモデルを考えてみます。

画像をそのまま入れても、違うモダリティなので言語モデルには直接的には入らないのですが、画像のエンコーダと呼ばれる画像の内容を理解するようなモデルを、1個用意します。そして、そこで画像の内容を理解します。さらに、アダプターというところがあり、このアダプターを通じて、それを言語モデルに入力できるような形式に変換します。

画像の内容を、単語のように変換するのが、アダプターです。そうすることにより、画像の内容をあたかも言語、単語のように扱います。質問は言語ですので、そのまま言語モデルに入ります。そうすると、入力されたテキストの内容と画像の内容を同列

に扱うことができるため、その後は言語モデルがよしに扱ってくれるということになります。

ここで重要になってくるのは、画像のエンコーダを医療用に作らないといけないです。これが一つ重要なことです。後は、このアダプターにどうやって学習させるかということも重要です。また、この言語モデルと画像エンコーダとアダプターを 1 個のモデルとするのですが、1 個のモデルにしたときに、このままだと一般的過ぎるため、これを医療用にチューニングするためのデータが必要になってきます。



まず、画像のエンコーダを学習させるためには、画像の内容を言語寄りに変換しないといけないです。そのために、画像とテキストのペアを大量に集めてくる必要があります。

## 2. 医療 LLM の社会実装

### 一般画像-日本語テキストペア

- Webをクローリングして50億ペアを収集予定

参考：LAIONデータセット

The screenshot shows the LAION CLIP retrieval interface. At the top, there is a search bar with the query 'Cool Robot'. Below the search bar, there are 24 image results arranged in a grid. Each result consists of a small image and a text caption. The captions are in English and describe various types of robots, such as 'Recycled Metal Sculptures by Brian Mock', 'Vintage tin space toy inspired robot by Greg Guede...', 'how to make an astronaut recycle | How to make a r...', 'wonderful Lego robot by Yul B Kanel', 'Futuristic robot concept with electrical wire hair...', 'Clockwork robot', 'Happy smiley robot on gray blue background. Creati...', 'Robô de Sucata - Como fazer um para as crianças br...', 'dmitar-minkov-soldier-bot', 'Proto | Factoria de Androides by Sâtrapa', 'Ю3 Георгий. Робот', 'Are you seeing near duplicates? KNN search are good at', and 'Ю3 Георгий. Робот'. The interface also includes a sidebar with various filters and options like 'Clip retrieval works by converting the text query to a CLIP embedding...', 'Display captions', 'Display full captions', 'Display similarities', 'Safe mode', 'Remove violence', 'Hide duplicate urls', 'Hide (near) duplicate images', 'Enable aesthetic scoring', 'Aesthetic score', 'Aesthetic weight', 'Search over', 'Search with multilingual clip', and 'This UI may contain results with nudity and is best used by adults. The images are under their own copyright.' At the bottom of the interface, there are two URLs: 'https://laion.ai/' and 'https://atmarkit.itmedia.co.jp/ait/articles/2301/18/news011.html'.

これは英語ですが、目標としては、日本語の 50 億の画像と言語のペアを収集する予定です。例えば、これも有名なデータセットですが、資料の画像はビンテージロボットという単語がついていますが、このような感じの画像があったときに、こんな感じの画像とテキストのペアをわっとクローリングし、データを集めてきて、この画像を言語寄りに変換するようなモデルを作ります。

次に、この全体のモデルに医療の画像を入れ、その医療の画像に関する質問を入れて、答えを出すということをやらないといけないため、医療画像に対する質問と答えのデータセットを作らないといけません。

# 日本語VQAデータセット

• 公開されているデータセットを日本語訳を使って利用

参考：Visual Genomeデータセット

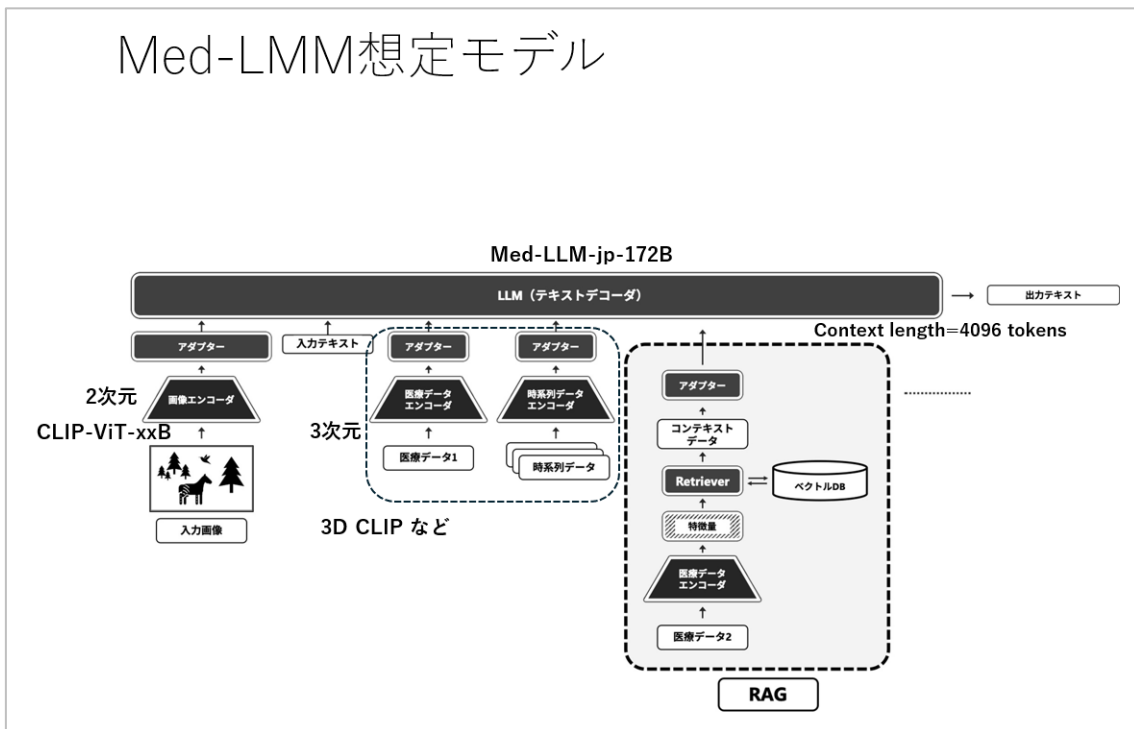


Fig. 29 Example QA pairs in the Visual Genome dataset. Our QA pairs cover a spectrum of visual tasks from recognition to high-level reasoning

Krishna, R., Zhu, Y., Groth, O. et al. Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations. *Int J Comput Vis* 123, 32–73 (2017). <https://doi.org/10.1007/s11263-016-0981-7>

医療用のものを学習する前に、一般的な常識、こう聞いたらこう答えなければいけないという常識も学習させなければいけないのですが、これも公開されている様々な英語のデータセットがあるため、それを日本語に訳して扱っていかうと考えています。

# Med-LMM想定モデル



マルチモーダルモデルが難しいところは、今は画像とテキストの対応関係について話でしたが、扱うデータについては、それ以外にも 3次元 CT 画像というものがあります。人間を輪切りにしたようなデータで、その1枚、1枚は2次元ですが、それを重ねると3次元のデータになります。これはこれで、2次元の画像とは別物なので、3次元のデータを言葉の空間に変換するようなエンコーダをさらに学習させる必要があります。そして、このエンコーダを学習させたあとに、アダプターを通じて理解した内容をあたかも単語のように変換して、言語モデルに入力するというを行わないといけません。

# 日本語医療VQAデータセット

• J-MID (日本医用画像データベース) の5億ペアから作成

参考: VQA-RADより

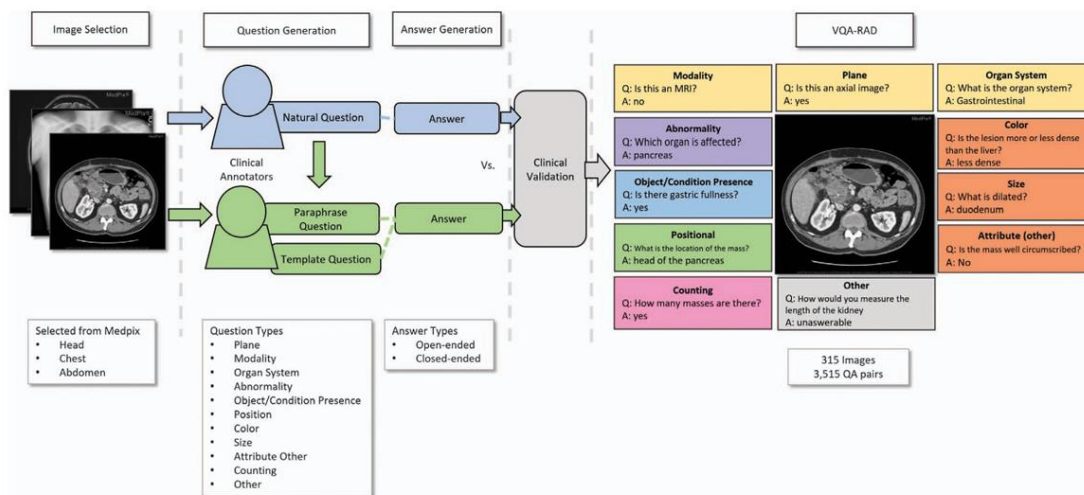


Figure 1. Flow Diagram of VQA-RAD build.

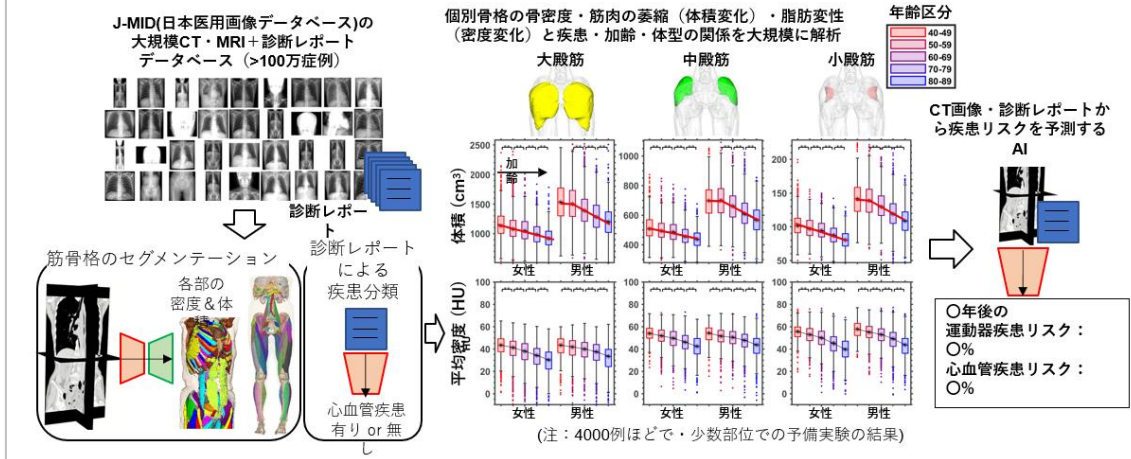
<https://www.nature.com/articles/sdata2018251.pdf>

先ほど、一般的な知識として、一般的な画像にこういう質問をしたら、こういう答えが返ってくるようなデータセットが既にあり、それをうまく日本語に訳して使うというお話をしました。しかし、医療用で、質問に対してきちんと答えを返してくれる、きっちりとしたデータセットが存在しますが、多くないです。ですので、このプロジェクトの中では、「J-MID」と呼ばれている日本医用画像データベース、これは画像とテキストで5億ペアがありますが、これから、何か医療画像に対して質問をして、それに返ってきた答えというようなペアを自前で作って、このモデルに学習させることを考えています。

# 汎用Med-LMM (筋骨格)

大竹先生 (奈良先端大) のテーマ2キックオフのスライドより

- 大規模医用画像+診断レポートデータベースの層別解析に基づく筋骨格の加齢変化解析と、心血管疾患のリスク予測AIの構築



ほかにも、このモデルで、3次元のエンコーダというものを作らないといけないというお話をしましたが、この3次元のエンコーダも、先ほどのJ-MIDのデータセットを使って構築する予定です。


3次元のエンコーダができると、資料の画像は筋骨格ですが、どこに大臀筋があり、どこに小臀筋があるかというものを、パーツごとに分解します。そうすると各筋肉量が定量的に測れますので、それを経時的な変化としてうまく捉えることにより、運動器の疾患リスクや心血管の疾患リスクなど、様々なリスクを推定することが可能になります。

# 特化型Med-LMM（循環器）

小寺先生（東大）のテーマ2キックオフのスライドより

**心電図のLMM開発**

心電図 + 心エコー



心電図から従来は困難であった左室収縮機能低下を検出するAIを開発し、PMDA申請の臨床


社会実装 ↓ LLM

**心電図読影LMM**

結果は陽性です。本当ですか？あまり異常に見えません。今回の心電図では、左側胸部のS波の延長を認めます。わかりました。専門医に紹介したほうが良いですか？専門医で心エコー検査をお勧めします。

**胸部レントゲンのLMM開発**

レントゲン + 読影レポート



胸部レントゲンから循環器疾患を判読する深層学習モデルはすでに開発済みである。

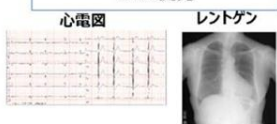
社会実装 ↓ LLM

**レントゲン読影LMM**

結果は陽性です。息切れはありません。本当ですか？今回のレントゲンでは、心拡大、肺うっ血を認めます。わかりました。利尿剤を処方したほうが良いですか？専門医での精査をお勧めします。

**心電図とレントゲンを融合したLMM開発**

心電図 + レントゲン




心電図の深層学習モデル、レントゲンの深層学習モデルはすでに開発済み。LMMを用いて

社会実装 ↓ LLM

**マルチモーダル診療支援AI**

心電図から左室収縮機能低下が疑われます。レントゲンから心拡大、肺うっ血が疑われます。循環器外来受診を検討してください。



**心電図のLMM、レントゲンのLMM、さらに心電図とレントゲンを融合したLMMの開発を目指す**

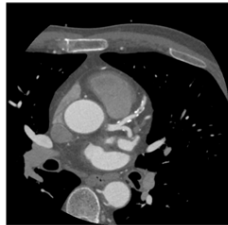
このようなかたちで、汎用的な医療用のマルチモーダルモデルを作りますが、それを循環器、保険診療などに特化型で学習しています。特化型の例の一つとして、循環器のものを説明します。例えば、どのようなことができるようになるかという、循環器の場合ですと、心電図と心エコーを入力して、これに関する質問をマルチモーダルモデルにすると、答えが返ってくるものができて、心電図とレントゲンの情報をマルチモーダルモデルに入れて、「何か問題はありますか」と聞くと、「心機能の低下が疑われます。外来受診を検討してください」というような自然言語でのアウトプットが返ってくるものができます。

# 特化型Med-LMM (循環器)

的場先生 (九大) の  
テーマ2キックオフの  
スライドより

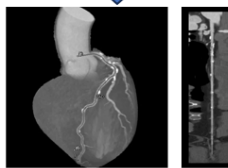


## LMM2: 画像から治療方針・予後を予測するLMM



横断面画像

- 予測する臨床予後 R7-
1. カテーテル冠動脈造影を受ける
  2. PCIを受ける
  3. PCI後長期予後



画像ws前処理画像

- 予測する臨床予後 R6
1. PCIでデバイス不通過
  2. PCIで切削デバイス必要
  3. PCI合併症

### 実施体制

- データ: 九州大学・東京大学・熊本大学・自治医科大学
- 九州大学システム情報科学: 4名
- 九州大学循環器内科: 4名

### データ収集

- 画像(A): IHD診断冠動脈CT、CLIDAS-PCI (予後データあり) **1500例/1000例以上** (九州大学378/PCI前のみ196例、東京大学333/216例、熊本大学265/194例、自治医大230例、自治医大さいたま〇例)
- 画像(B): IHD診断冠動脈CT→CAG実施、PCIなし **1000例** (九州大学1177例、熊本大学〇例、東京大学〇例、自治医大〇例、自治医大さいたま〇例)
- 画像(C): IHD診断冠動脈CT→CAGなし、PCIなし **1000例** (九州大学2596例、東京大学)

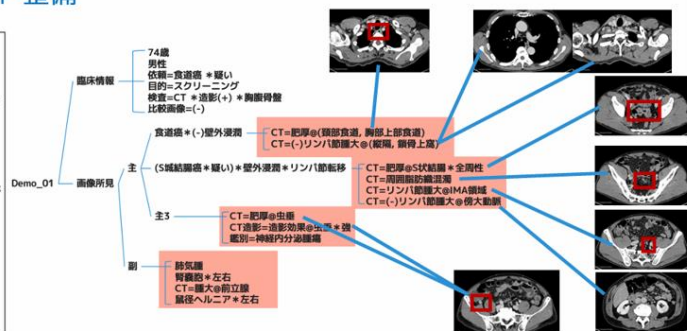
# 特化型Med-LMM (がん病変)

菊地先生 (自治医大) のテーマ2  
キックオフのスライドより

## ①LMM開発・検証に資する日本語での所見-3Dキー画像ペア作成

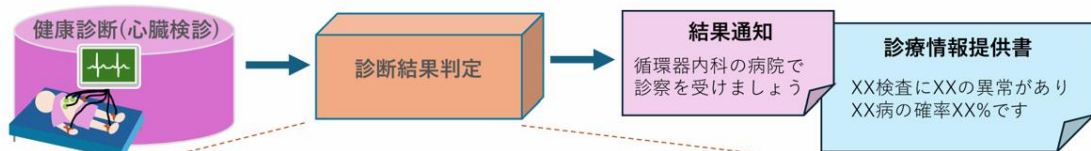
- 自然文レポートを構造化レポートへ変換 (J-CaseMapの模倣)
- 構造化レポートの単位構造とキー画像 (bbox) を紐づけたデータセットの作成
- がん関連画像診断文脈でVQAとして利用し、回答精度検証やチューニング  
体幹部CTを入力して自然文レポートを出力するLMMの構築
- がん以外でのデータセット整備

Demo\_01、74歳、男性  
病名: 食道癌の疑い  
目的: スクリーニング  
所見  
食道/胸部上部食道に軽度の壁肥厚があるように見えます。内視鏡所見を優先してください。明らかな壁外浸潤は指摘できません。縦隔、縦膈上段リンパ節に有意な腫大は認めません。  
胃腸筋に気腫性変化を認めます。  
肺に活動性炎症や腫瘍性病変は指摘できません。  
肝臓、両側腎臓、前立腺腫大、両側鼠径ヘルニアを認めます。  
S状結腸に全周性壁肥厚を認めます。周囲脂肪織の毛羽立ちを認め、壁外浸潤が疑われます。  
虫垂先端部に壁肥厚、明瞭な造影効果を認めます。NETを含めた腫瘍性病変の除外も必要です。  
TMA: 肺臓にリンパ節腫大を認め、転移と考えます。  
腸動脈リンパ節に有意な腫大は認めません。  
腸部実質臓器に転移所見は認めません。  
病的腹水貯留はありません。  
骨に顕著な病変は指摘できません。  
診断  
食道癌 ESD 前: リンパ節転移・遠隔転移所見なし。  
S状結腸腫瘍: リンパ節転移疑い。  
虫垂先端壁肥厚: r/o NET



# 特化型Med-LMM（健康診断） 牧元先生（自治医大）のテーマ2 キックオフのスライドより

## モデルの目的：心臓検診支援プログラム（全体像）



あと、これは健康診断ですが、心電図、心音、レントゲン、血液検査など、さまざまなマルチモーダルデータがあります。そういったものをモデルに入れることによって、われわれが健診に行ったときは、その内容を総合的に判断して、分かりやすく、平易な言葉で、現状について説明してくれるようなマルチモーダルモデルなどができるのではないかと考えています。