

バイオインフォマティクス特講

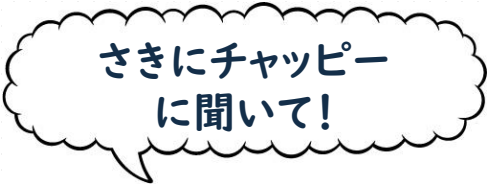
Application of Machine Learning to Healthcare

医学分野における AI・機械学習の実装

横浜市立大学 データサイエンス研究科ヘルスデータサイエンス専攻 清水沙友里



『研究はAIにサポートされながら実行する時代に』



さきにチャッピー
に聞いて!

クラウド型LLMとは

- ChatGPT, Claude, Gemini, 企業向けマネージドLLMなど
- 推論、コード生成、多言語対応など高性能
- 面倒なインフラ構築が不要、ハードウェア初期投資が不要
- 自動でモデルアップデート・最適化が行われる
- モデルは外部事業者のデータセンタ上に設置されるため、データがサーバに送信される
- 利用規約上、学習・ログ解析に使われる場合もあり、海外データセンタへの越境転移の可能性
- 個人情報はいない前提



Local LLMとは

インターネットを経由せず、PCやスマホ、オンプレミス環境のサーバで動く大規模言語モデル



[Hugging Face](#)や[GitHub](#)、 [Civitai](#)経由でダウンロードして利用

メリット

- データが外部に出ないため、クラウドに上げられないデータも利用可能
- 学内のコード・テンプレ・ドキュメントを組み合わせたRAGが可能
- ネットが利用できない環境でも動作
- タスクに合わせたモデルのカスタマイズが可能
- 動作のログが取れる
- APIコールごとにかかる利用料が不要

デメリット

- クラウドのモデルより性能が落ちる
- 高性能GPUなどハードウェアの初期費用
- 運用やシステム構築の知識
- 継続的な利用にはメンテナンスが必要
- 現段階ではクラウドLLMとの実力差が大きい

2025年に 技術的ブレイクスルー

オープンウェイトモデル：ダウンロードできるモデルの精度向上
ハードウェアの進歩：コンシューマGPUの性能向上。VRAMがUP、Apple シリコン
PC内蔵NPUの普及：いわゆる“AI PC”の登場
量子化・最適化ツールの成熟：VRAMの小さい環境でも推論・微調整が現実的に



Local LLMのチューニング

Local LLMを“自分仕様”に！



- ① プロンプトエンジニアリング(よい質問)
- ② システムプロンプト (ルール固定)
- ③ RAG (Retrieval-Augmented Generation : 外部知識の活用)
- ④ LoRA / QLoRA(軽量ファインチューニング：モデル内部のパラメータの調整)

分析での利用シーン

研究計画：先行研究の把握、研究仮説の洗練、研究計画/SAPドラフト

RAG：ガイドライン、過去のプロトコル、IRBひな形、用語集、参考文献、DB定義書

テンプレ出力：研究計画書の各項目のドラフト作成、マスタ作成

チェックリスト生成：IRBや執筆ガイドラインのレビュー



PECOに基づいて主要/副次評価項目と適格基準の妥当性をチェックして、バイアスとその対策を列挙して

研究計画を踏まえて、欠測メカニズム仮定と多重代入の設定を含むSAP案をマークダウン形式で作って

SAPから、対象患者のデータ抽出と前処理のスク립トを書いて。
エラーはデバッグして。

Local LLMの利用シーン

教育利用：医療合成データの作成、学生向け分析教育ツールの開発



『このサンプルのDPCデータをもとに、
85歳以上の救急車搬送患者の合成DPC様式1 データを作成して』
(特徴空間において、特徴量の分布が真のデータ分布に近似する合成データを生成する)


大学院生向けに、レセプト/DPCを用いた課題を対話的に行います。
マークダウン形式で提出された回答に対し、pythonで書いた自動
採点スクリプトを呼び出し、コード演習にフィードバックを与え
て。


“アルゴリズムの時代の到来”

AIは、博士号レベルのサイエンスの知識を持ち、ポーカーをマスターし、実験データから物理原則を学び、皮肉を理解しながら世界的ゲームプレイヤーになり、独創的な作曲家にもなっている。これらは少し前は不可能だと思われていた。これらと並行して、機械学習を医療に応用させる動きがみられている。臨床においての有用性と結果の解釈については懐疑的な意見が残るものの、病態の解明、創薬候補の同定、心停止の予測やコンピュータ診断支援など様々な研究やアプリケーションに採用されてきた。

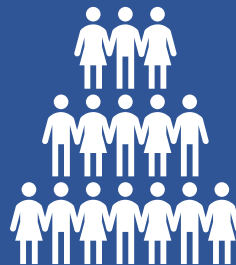
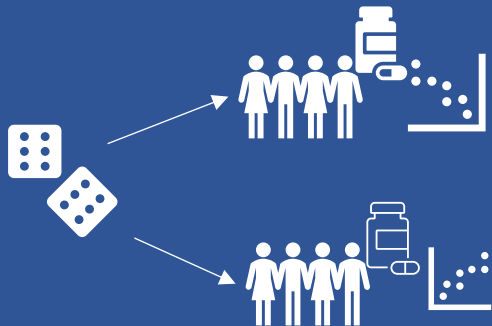
AI・機械学習を用いた医療分野の応用と課題、有望性について概説します

Table 1. Example questions and solutions from six MIT courses (18.01, 18.02, 18.03, 18.05, 18.06, 6.042), one Columbia University course (COMS3251), and six topics from the MATH dataset. The solutions can contain numerical answers, equations, plots, or other modalities

ID	Course	Question	Solution
1	18.01 Single Variable Calculus	A bacteria population is 4,000 at time $t = 0$ and its rate of growth is $1,000 + 2t$ bacteria per hour after t h. What is the population after 1 h?	$4000 + 1000 \log(2)$
2	18.02 Multivariable Calculus	Describe the graph of the function f : $f(x, y) = 10 - \sqrt{x^2 + y^2}$.	
3	18.03 Differential Equations	Find general solutions of the differential equations. If an initial condition is given, find the corresponding particular solution. Throughout, primes denote derivatives with respect to x . $y' + y = 2$, $y(0) = 0$.	$y(x) = 2(1 - e^{-x})$

- 
1. 臨床研究のゴールドスタンダードとリアルワールドデータ
 2. 日本の医療DB
 3. 分析手法の発展
 4. 医療分野における応用例

研究目的のデータとリアルワールドデータ



研究目的で収集したデータ
RCT、レジストリ、
患者・住民コホート



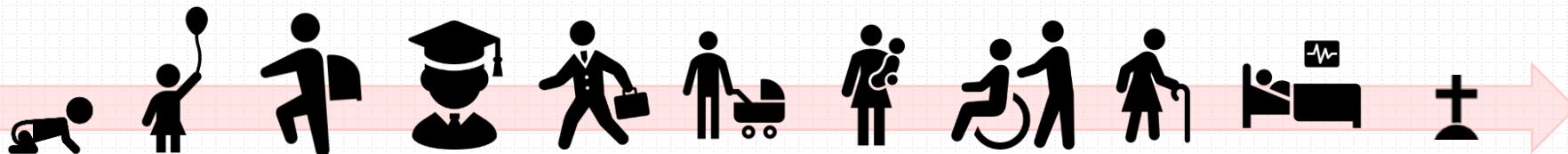
現実世界から生成される健康データ
Real World Data

Routinely-collected Health Data(RCD)



日常診療から生成される臨床データ
レセプト、DPC、電子カルテなど

- レセプトや電子カルテなど、日常臨床で生み出されているデータを指す
- 日々の業務で生み出されたデータが集積した結果、二次利用として研究や政策立案などに応用
- 病院管理上必要な定められた項目について、膨大なN数を保有
- 高齢者、多疾患併存、試験を行うことが倫理的に難しい介入などについても、現実世界に即したセッティングで医療的介入の実態や有用性についての研究を行うことができる
- 保険請求の対象となる患者全体が母集団であることから、集団としての法則性をある程度見つけられる



生成される医療関連データ



レセプト

- 医科（入・外）
- DPC
- 歯科
- 調剤
- 特定健診等
- 医療機関由来と保険者由来



DPC

- 急性期医療機関ベース



医療情報システム

- 電子カルテ
- 部門システム
- 地域連携システム



政府統計

- 人口動態調査
- 生命表
- 患者調査
- 病院報告
- 医療施設調査
- 3師調査
- 国民医療費
- 受領行動調査
- 伝染病統計
- 社会福祉施設等調査
- 介護サービス・事業所調査
- 社会医療診療行為別調査
- 国民生活基礎調査
- …etc



調査

- 疾患レジストリ
- 住民コホート
- バイオバンク
- 臨床試験
- 質問紙等調査



その他

- ウェアラブルデバイス
- 健康アプリ

RCTとリアルワールドデータの違い

Routinely collected
Health Data

	RCT	リアルワールド
患者	組入/除外基準 (RWDと比し標準化)	現場判断 (保険適応、現場判断)
割付	ランダム割付	患者毎の判断 (患者嗜好、社会環境)
介入	事前に規定された用法・容量	様々な使用法 (低用量・低服薬遵守)
評価	事前に規定 (評価法・タイミング)	雑多な評価 (限定された評価項目・様々なタイミング)

n数の多いリアルワールドデータを活用する動き！バイアスに対する対処が必要。

RCTにおける対象者のイメージ

全人口

対象疾患（例：糖尿病）

Hight risk patient(年齢,性別,リスク因子)

薬剤の禁忌除外

コンプライアンス良好

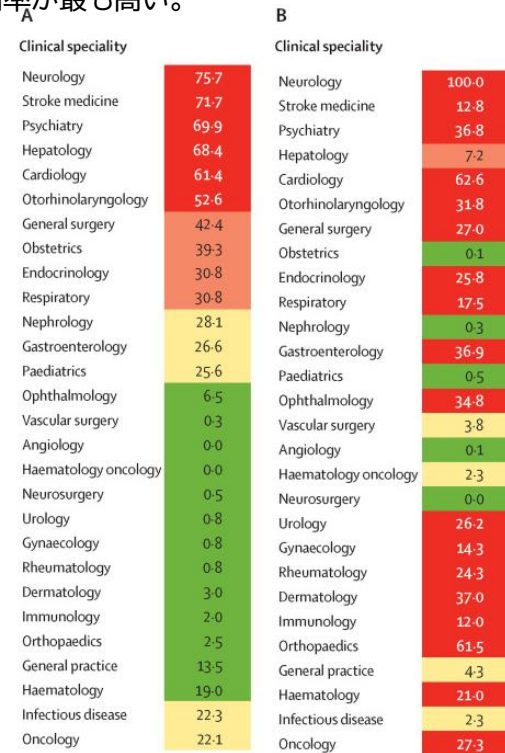
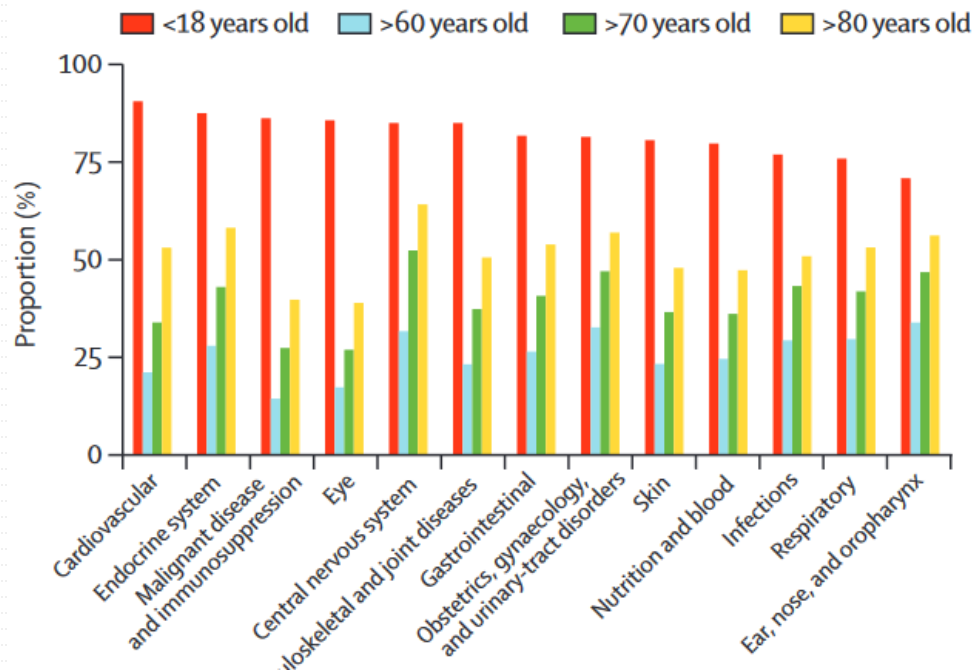


ランダム割付
介入群 vs. 対照群

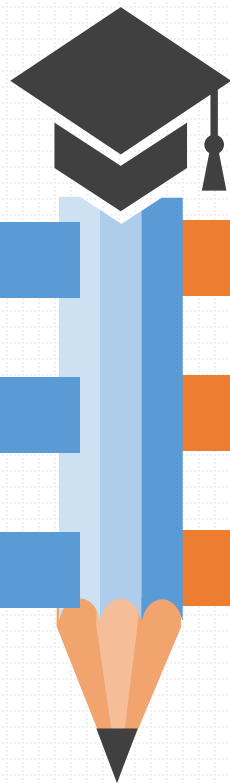
RCTで除外された“脆弱な集団”に対するリアルワールドの処方は？

-不十分な治療決定情報のリスクにさらされる可能性のある集団を特定-

43,895件のRCTから疾患分類毎に除外されている年齢を算出。右図は抗認知症薬RCTの除外割合と人口比の比較。RCTから大きく除外されている患者もリアルワールドでは多い。循環器科と精神科は、対応する除外割合と人口有病率が最も高い。



データベース研究のPros & Cons



現実世界を反映した多様な患者集団

RCTは管理された集団

比較的長期の追跡

長期の観察項目についてデータが取得できる

1研究あたりの実施コストが安い

費用の急上昇とエビデンスの量が見合わない*1

決定的な因果関係は不明

治療選択に影響を及ぼす要因がアウトカムにも影響

不正確な記録・データの欠落・未測定

研究のためのデータではない

再生性・再現性*2,3に課題

手法論的課題・**研究の手続き**

再生性：再解析で同様の結果（方法の再現可能性）
再現性：追試で同じ結果（結果・推論の再現可能性）

医療分野における機械学習

医療画像からの臓器認識、間質性肺疾患の分類、肺結節の検出、医療画像再構成、脳腫瘍セグメンテーションなどの汎用的なタスクで優れた結果を示している。ソフトウェアが放射線科医や医師の診療を支援し、MLが医学研究と診療に革命を起こすことが大いに期待されている

- 臨床医学はML/DLモデルのエキサイティングな応用分野として浮上
- 病理学、放射線学、眼科学、皮膚科学などで既に人間レベルの性能を達成したモデルがある
- 院内導入された敗血症の早期発見ツールを2年間利用した5病院において、実際に敗血症の高い感度(82%を同定)と高い利用率が示された。またシステムから警告があった患者は、最初の抗生物質処方までの中央値が1.85時間短縮された

ビッグデータの特徴①

①単体の計算機の演算能力を超えるデータのこと。データの**容量**が大きい

Q. なぜデータが大きくなったのか？

A. 企業活動や社会活動のデジタル化により、
ログや業務で記録されるイベントや観測値が増加した

- 大容量データのひとつの計算処理を、ネットワークで繋がった複数の計算機で同時並列処理をする技術(古くは2004年のMapReduce, Hadoop, Spark, 分散アルゴリズム, 商用クラウドサービスなど)の向上
- データ集約型の企業は、集められたデータを利用可能な形に加工するパイプラインとデータエンジニアが存在
- 世界最大の小売向け電子決済処理ネットワークを持つVISAは、毎日平均1億5千万件の取引を処理し、24000tps(トランザクション/秒)以上の処理をすることが可能
- いわゆる“データサイエンティスト”はこのパイプの専門家ではなく、データエンジニアと協働する

ビッグデータの特徴②

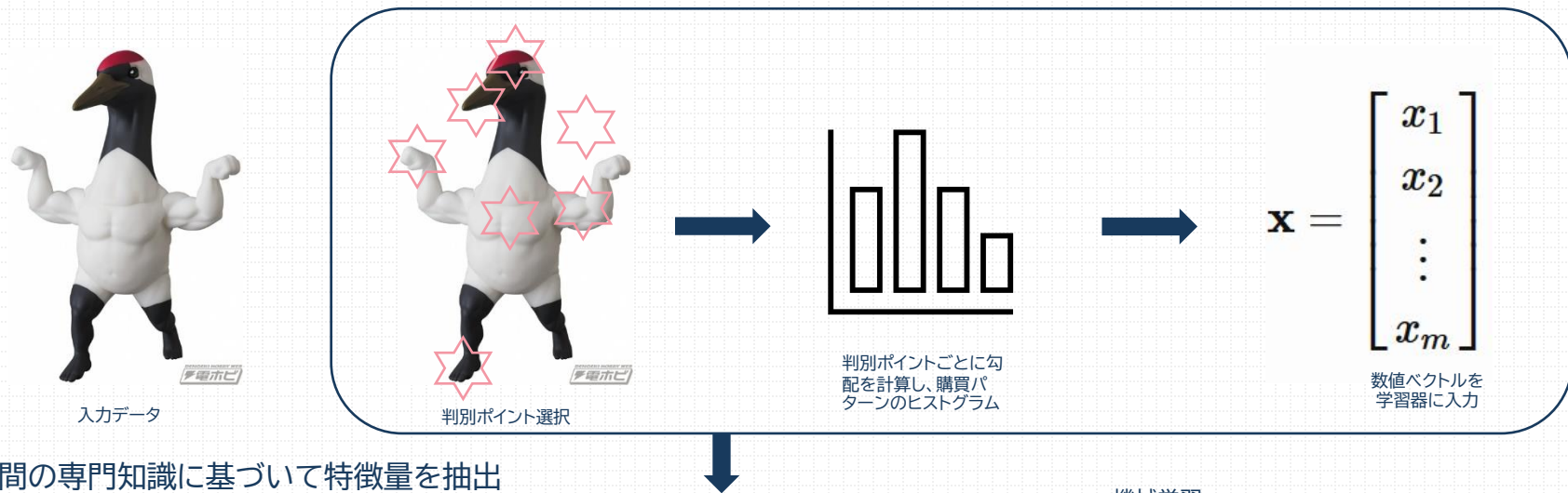
②データの分析段階において重視されることは複雑さ。データの**次元**が大きい

Q. なぜデータが複雑になったのか？

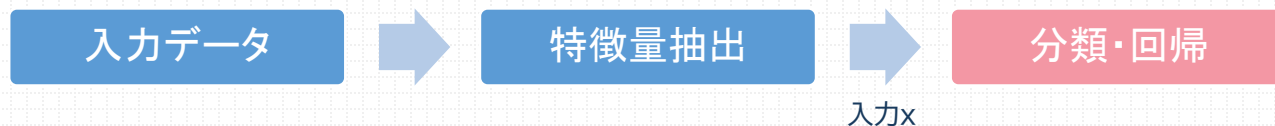
A. ログ、自然言語テキスト、画像、動画、音声、センサデータ
など構造化されていないデータが増加した

- データエンジニアが目指すのは、構造化されていないデータを正規化し、標準的な形式でデータテーブルに格納すること
- 変動するデータを構造化することは難しいため、半構造化(定義されているが個数が膨大)データとして保管
- 単純で次元の低いデータを分析するならば複雑なモデルは不要
- 従来は、画像は画像工学、音声は音響工学というようにデータタイプにより縦割りで専門家が縦割りで分析す手法を開発し、それぞれのデータから、有用な情報である特徴量の抽出の技術が研究されていた
- **特徴量ベクトル**に変換されていれば、分野が異なっても汎用的なアプローチで解析可能！ ⇒機械学習

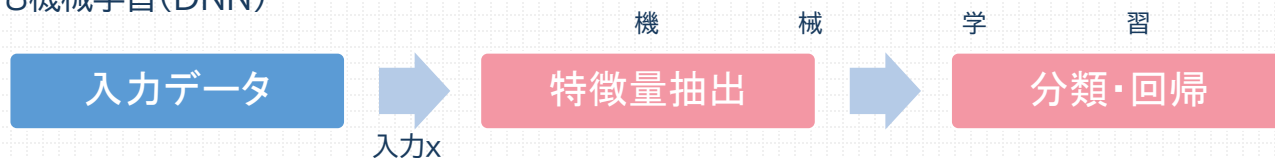
特徴量抽出を自動化



- 人間の専門知識に基づいて特徴量を抽出



- 特徴量抽出も機械学習(DNN)



統計学:モデル推定(回帰分析で個々の係数を推定し、モデルのパラメータを明らかにしたい)

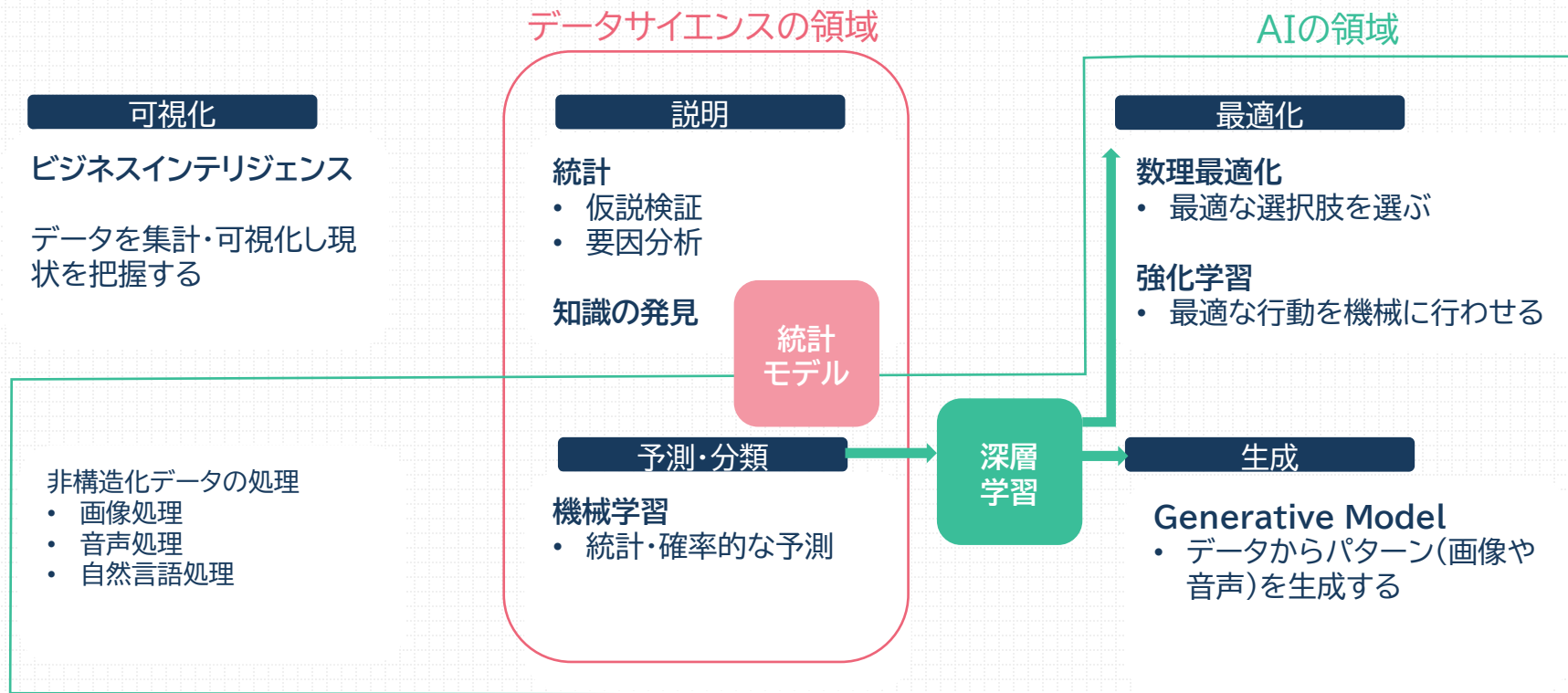
機械学習:予測精度(あるデータセットに基づいて訓練されたモデルが新たなデータをどれだけ精度よく予測できるか)

Q. モデル推定と予測、どちらかが簡単でしょうか？

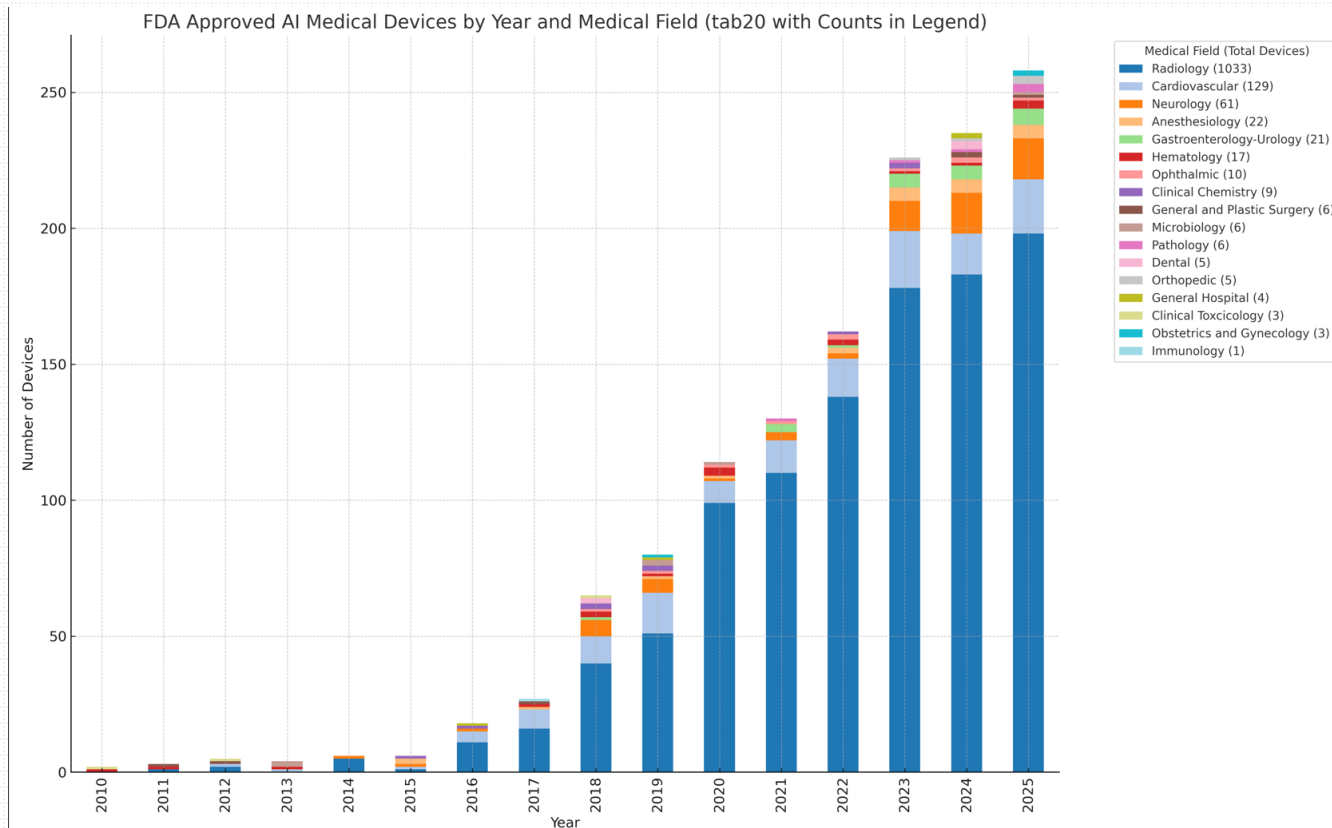
- 多くの医学研究では、ラベル付きのトレーニングデータが十分にあれば、どんな方法でもよい結果を出せる
- そのようなデータがなければどのような方法でもうまくいかない
- 機械学習は、データがこれまでに想定されていないほど複雑であっても、頑健な予測を導出できる
- 機械学習における重要なパーツ(ラッソや木や森など)は多くが統計学で生み出されたもの
- 説明可能性などの研究はかなり進展したが、学習済みモデルのパラメータに構造的な意味を持たせない

※ただし近年は統計的因果探索も流行っています

データから何らかの価値をもたらすことが目的



FDAにおけるAI対応の医療機器の承認

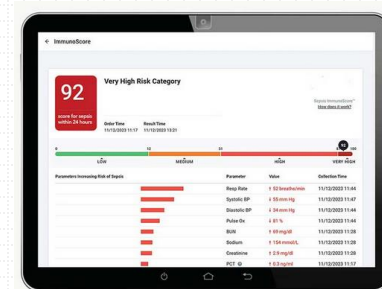


2025年12月1日現在、FDAは1357のAI/ML搭載医療機器を承認。もっとも多いのが放射線(1,033)、次いで循環器(129)、神経(61)、麻酔(22)、消化器・泌尿器(21)、血液学(17)。

2024年以降だけで500件程度が追加されている。医療分野別では、放射線領域(CT/MRI再構成、CADなど)が依然として最多。次いで 心血管／循環器、消化器内視鏡、神経・睡眠、遠隔医療・バイタルモニタリング が拡大。

救急・集中治療系:Prenosis社のSepsis ImmunoScore(敗血症診断ツール),2024年

- 血液検査やバイタルなど22項目程度の指標をAIで解析し、敗血症／重症化リスクをリアルタイムで4段階に層別化する最初のAI診断機器。入院時の死亡率、退院前の滞在期間、ICU入院、機械的換気の必要性、および血管圧迫機能の使用を同時に予測する
- 米国病院での実装データでは、抗菌薬投与までの時間短縮(1時間)やSEP-1コンプライアンスが17項目で改善するなどの成果が報告
- AI診断補助の多くが”擬陽性でアラートがなりすぎる”ことが多いが、意思決定に必要なタイミングを正確に把握し情報を提供する



心不全・心房細動リスク予測Eko Low Ejection Fraction Tool (ELEFT)、2024年

- 電子聴診器やECGから取得した心音・心電図をAIで解析し、左室駆出率LVEF $\leq 40\%$ (潜在的な心不全)を疑う患者を絞り込む。検査中にELEFTを利用することで、心不全指標である低EFを15秒で検出可能。

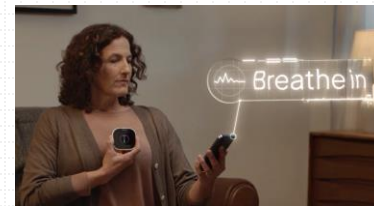


X-ray InterpreterTempus ECG-AF(Tempus AI, Inc.)2024年

- 心房細動の既往のない65歳以上の患者を対象とし、誘導心電図をAIで解析し、12ヶ月以内に心房細動(AF)/心房粗動(AFL)を発症するリスクが高い患者を検出するソフトウェア。Tempus ECG-AFの結果を、患者の元の心電図記録、他の検査、症状、病歴などの他の診断情報と並行して考慮することが推奨されている。

「診断確定」よりも 将来リスク予測・スクリーニングにフォーカスしたAIが
心血管分野で増加しているのが特徴。

近年のAI医療機器の動向③ 遠隔診療・モニタリング系



Tyto Insights for Crackles Detection(Tyto Care)2024年

- デバイスを使うことで、患者がどこにいても専門医による検査が受けられ、心臓や肺の音を聴取したり、耳道、喉、血圧、血中酸素濃度などの診断が遠隔で可能。デジタル聴診器の音声をAIが解析し、ラ音(crackles)や喘鳴を自動検出するなど。



Oxevision Sleep Device(Oxehealth) 2024年

- 病室に固定設置したカメラ映像から、体動や生理的データに基づいて睡眠状態を推定し、夜間見守りや転倒リスク管理などに活用。特に急性期の精神病床などでの活用を期待。

高品質なバーチャル/ハイブリッド診療所の構築し、遠隔診療が他の医療提供プロセスと統合されることで、単なる“遠隔診療”ではなく、アクセスの公平性、迅速な診断、個別ケア、アウトカムの向上など対面と同等以上の質を提供が求められるフェーズに。

予後予測への機械学習の応用

予後予測とは、臨床の現場で予想される病気の進行を予測し、特定の疾患に関する症状や徴候を特定し、それらが時間とともに悪化するか、改善するか、安定したままであるか、関連する健康問題、合併症、日常生活の能力、再発や死亡を特定することも含まれる。

- ゲノム、プロテオミクス、病理学的検査結果、医療画像などのマルチモーダルな患者データが収集され、機械学習モデルが疾患の予後、診断、治療を支援する
- 特に、今後期待されている個別化医療に向けて、バイオインフォマティクスなどの隣接分野と融合しなら予後予測モデルの実証的な応用が期待されている

例) 脳腫瘍の一種であるグリオーマ(神経膠腫)を、97.54%の精度で低グレード又は高グレードに分類(低グレード:グレードIの毛様細胞性星状細胞腫やグレードIIの低グレードグリオーマ 高グレード:グレードIIIの悪性神経膠腫やグレードIVの多形性膠芽腫が含まれた)

MRI画像から得られるデータの中には、腫瘍の形状、テクスチャ、画像強度の詳細など、肉眼では検出できないものも多く含まれるが、機械学習アルゴリズムにより、このようなデータの抽出を支援することができる

電子健康記録(EHRs)を用いた応用

医療機関やその他の医療サービス提供者は、日々膨大な数の電子健康記録を作成しており、患者の完全な投薬履歴を含む構造化および非構造化データが存在

- 診断プロセスを容易にするための臨床的特徴の抽出や、電子カルテの自動記載、死亡率予測や医療費推定など様々な分野で研究が進んでいる

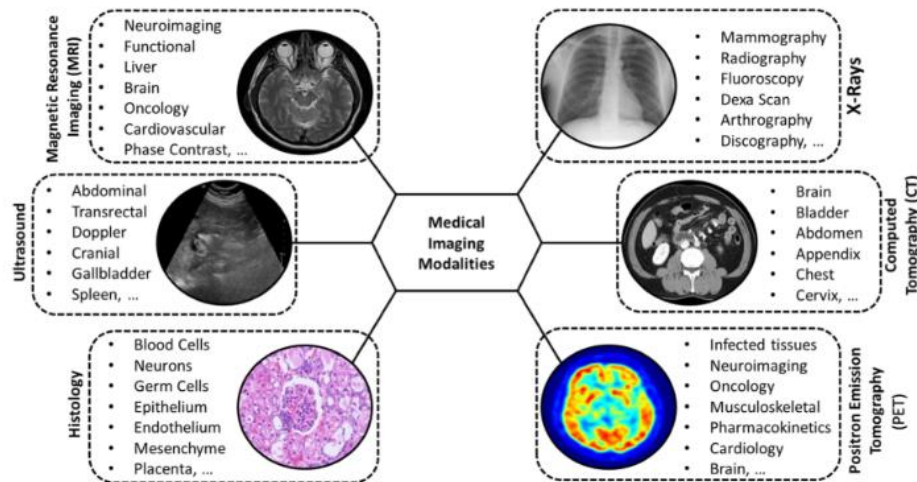
例)

- 電子カルテを用いた臨床予測モデルにおいて、導入後に予測性能が低下することが知られている。ある年に学習したモデルが長期にわたって予測能を維持するための緩和策を開発。mimicデータは2008年にデータ構造が大きく変更され、個人識別子が削除された。2002年から2012年のmimicデータを用いて、死亡とLOS予測のモデルを作成し、次年度以降のデータでテストした結果、死亡予測のAUROCが0.29低下、ランダムフォレストでの入院期間予測が0.1低下した。
- 千年カルテに参加している宮崎大学の電子カルテのテキスト情報(肺がん患者の遺伝子検査結果)を形態素解析し、ルールベースでアノテーションし、テキストを抽出して分析ができる自然言語処理システムを構築。

医用画像への機械学習の応用

機械学習は、MRI、CT、超音波、PETなど、様々な画像診断法を用いて得られた医用画像から、効率的かつ効果的に情報を抽出するために使用されており、完全に自動化された医用画像診断システムは、次世代ヘルスケアシステムの一部として期待されている。

- 医用画像解析の主な目的は、臨床医や放射線技師が効率的に病気の診断や予後を判断できるよう支援すること
- 医用画像解析の主なタスクは、検出、分類、セグメンテーション、検索、再構成、画像登録など

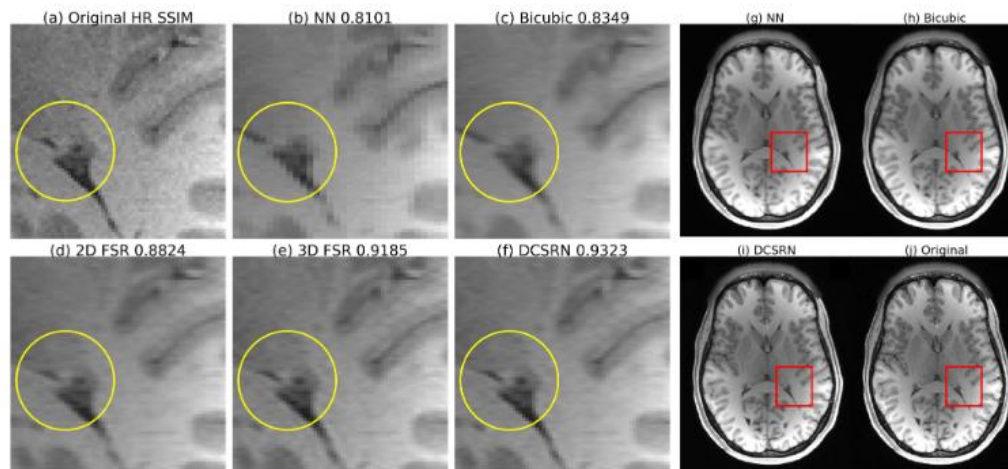


劣化した医用画像の強調処理

医用画像の取得プロセスでは、多くの機械的ノイズや外乱が発生し、画像の品質や重要性を低下させるため、MRI の際には、詳細な軟組織のコントラストを取得するために、患者はできるだけ静止してじっとしている必要がある。そのため低解像度の画像を高解像度に復元する様々なモデルが開発されている。

例)

1,113人の被験者のデータセットを用いて、4倍解像度低下させた画像の復元において、従来の手法よりも優れたMRIのノイズ除去に対するモデルを構築した。



画像検出

医用画像から特定の疾患パターンや異常(腫瘍、癌)を識別するプロセスは、検出と呼ばれる。従来の臨床診療では、このような異常は放射線技師や医師によって識別され、多くの時間と労力を必要とすることが多い。機械学習・深層学習ベースを用いたこのタスクに対する様々な研究が発表されている

例)

がんの病理組織診断は、病理医の経験値に基づいて行われている。深層学習技術を用いることにより、病理画像の特徴を数値化する技術を開発。数値と病理医の評価を比較し検証を行い、病理組織の評価に最適なモデルが構築できた。これをwebならびにアプリとして一般公開している。

<https://luigi-pathology.com/>

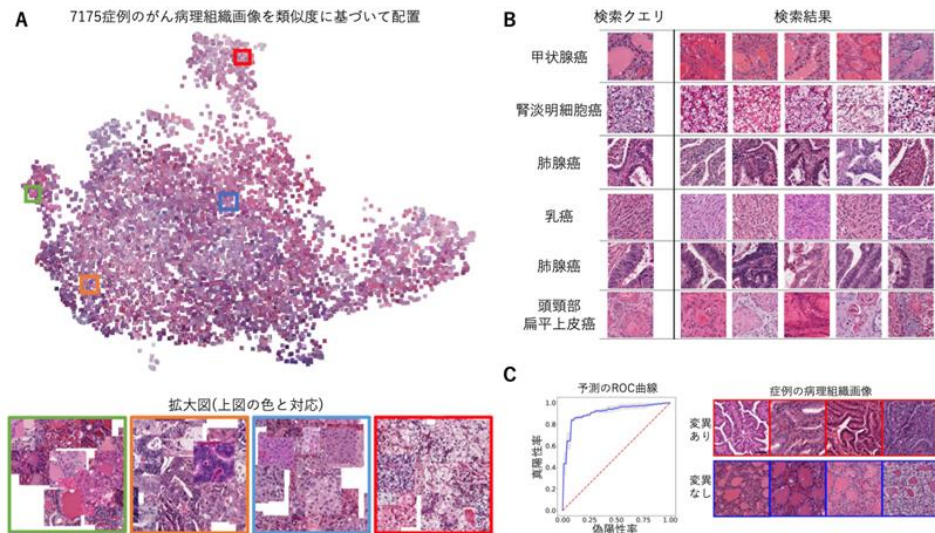
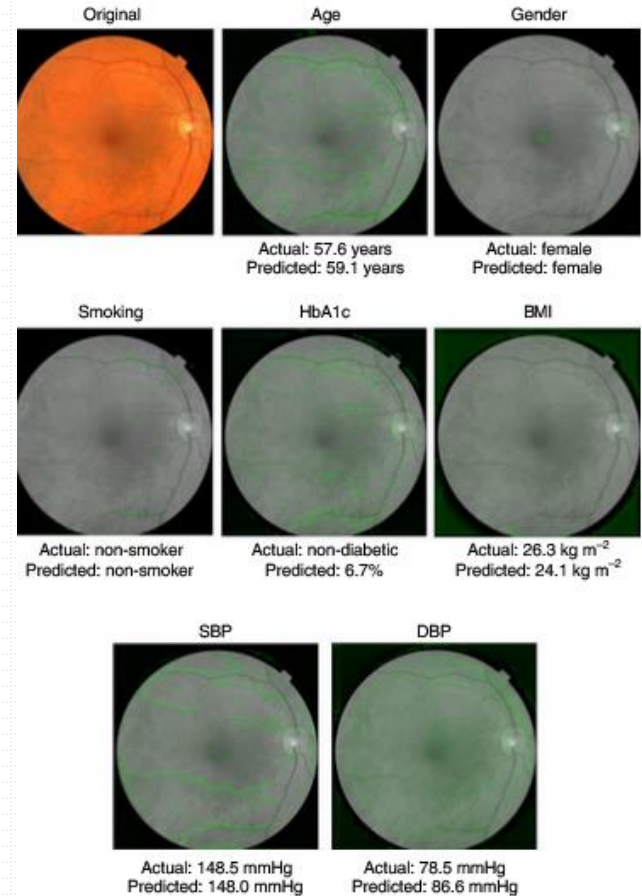


図2 A. 32がん種7175症例の病理組織画像の可視化。病理組織画像のディープテクスチャが類似している症例を近くなるように配置している。拡大図：組織学的に類似した症例が近くに配置されている。B. 類似症例検索の一例。様々ながん種で類似した組織画像が検索できている。C. ディープテクスチャに基づく甲状腺癌のBRAF変異の予測。AUC (注5) が0.91の精度で変異の有無を予測できている。

- 眼底写真⇒眼科疾患や糖尿病はもちろんのこと
- 眼底写真から心血管疾患リスクを同定する(コレステロール値からの予測と同等のレベル)
- 喫煙者と非喫煙者を71%の精度で分類
- アテンションを使って、画像のどこに注目して予測したのかということを説明可能になった
- 例) 血圧⇒目の血管部分に着目
- 眼底写真ではなく、眼の画像からHbA1c、や血中脂肪を頂いていた
- より痛みを与えず(侵襲性が低く)、正確、安価で、より容易に利用できる新しい病気の検出やモニタリングの方法が必要



糖尿病性網膜症の自動検出AIシステム

LumineticsCore

(IDx-DR 2021 update)

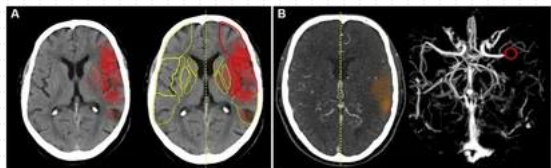
- カメラ対応拡大
- 診断精度向上
- HERとの連携強化
- インターフェース改善
- データセキュリティ強化



<https://www.youtube.com/watch?v=8AhzDfmtSEE>

激戦区！ 脳卒中診断、大血管閉塞(LVO)の同定

- まだAI研究の初期である2014-2019年の時点でMLを用いた研究が20以上出版
- Rapid LVO、Viz LVO、Aidoc、CINA ICH、Accipio Ix・Accipio Ax(撤退)、e-ASPECTS、Cercare Medical Suite など多数の製品

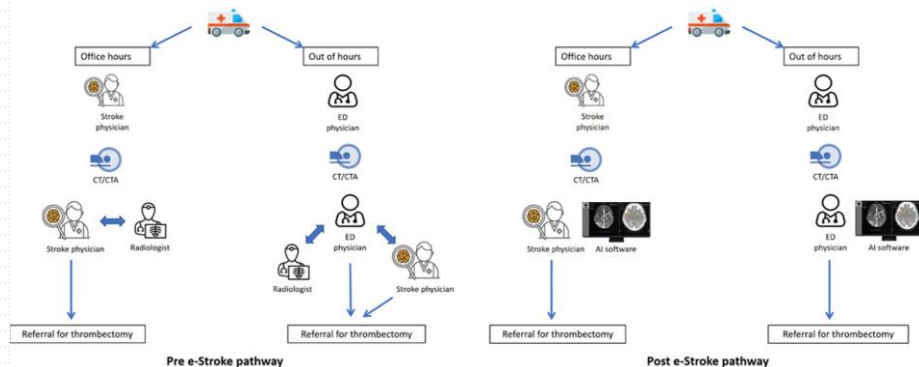
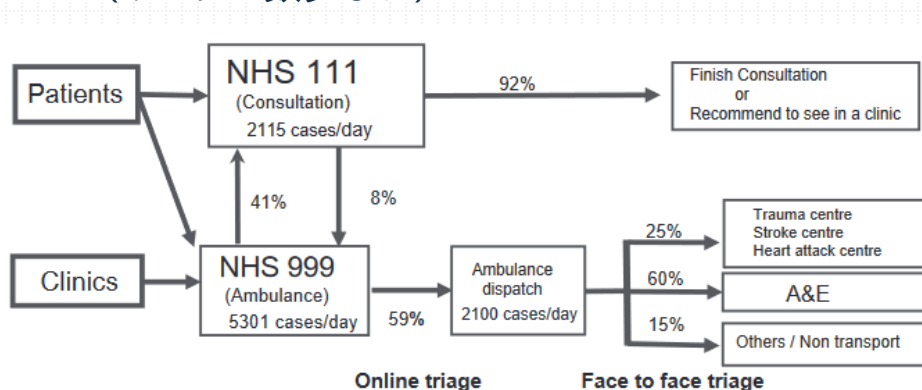


Brainomix 360 Strokeの市販後研究

脳卒中CT画像をAIで分析し、臨床的に検証済みの特定の画像特徴を自動的に検出

①レトロスペクティブな英国の脳卒中センターにおける単施設調査

- CSCは日中は専門医、時間外と週末は救急医が勤務、画像は専門医と救急医の全員が精通しているわけではなく、所見の確認を放射線専門医に求めることがよくあった(その後CSCでの受け入れが決定)
- 導入前後(2019年~2021年)のEVTを受けた全ての患者が対象
- 導入後、平均DIDO時間が141分から79分へと、62分間短縮、主に紹介までにかかる時間の改善
- EVT後に機能的自立(90日後のmRS 0-2)を達成した脳卒中患者の数が16%から48%へと3倍に増加(サンプル数少ない)



Brainomix 360 Strokeの市販後研究

脳卒中CT画像をAIで分析し、臨床的に検証済みの特定の画像特徴を自動的に検出

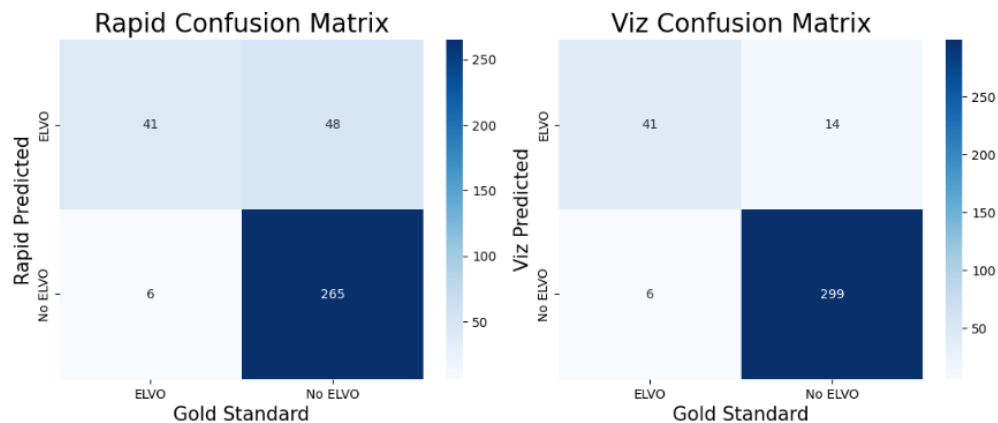
②AI画像診断の最大規模の前向き研究(欧州脳卒中学会2024の発表)



- 調査期間3年、26施設、83,000人
- 機械的血栓回収療法を受ける患者数が50%増加
- ドア・イン・ドア・アウト(DIDO)時間を49分(27%)短縮

検証! Viz LVO v.s. Rapid LVO

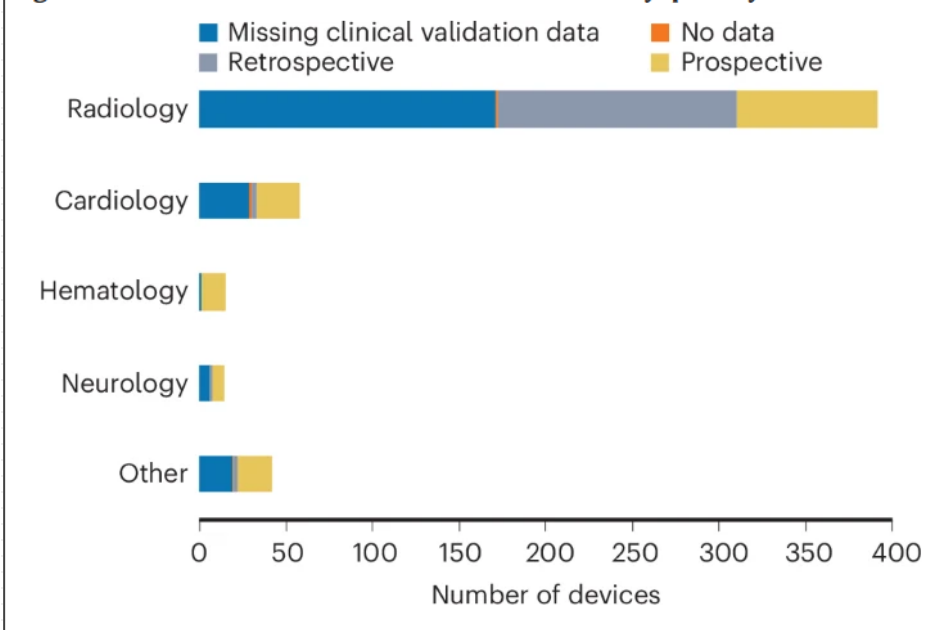
- システムにより、8分未満で画像を処理でき、介入医への早期通知により、door to needleまでの時間短縮、その他のワークフロー指標時間の短縮、転送時間の短縮につながることを示されていた
- 2022～2023年に、ヒューストンの2つのCSCに導入されたViz LVOとRapid LVOの診断精度を比較
- 神経放射線科医の解釈をゴールドスタンダードとして、Rapid LVOとViz LVOの読影結果から要約統計量を算出
- 100件のスキャン中、Rapid LVOとViz LVOはそれぞれ11件のLVOを正確に検出したが、6件のLVOを見逃した。Rapid LVOは13件のスキャンをLVOと誤認し、Viz LVOは4件のスキャンをLVOと誤認
- 双方ともに、擬陽性による医療従事者のアラーム疲労を防ぐため、偽陰性値を防ぐことを優先し、偽陽性値を犠牲にしていた



AIを用いた医療機器はどのように検証すべきか？

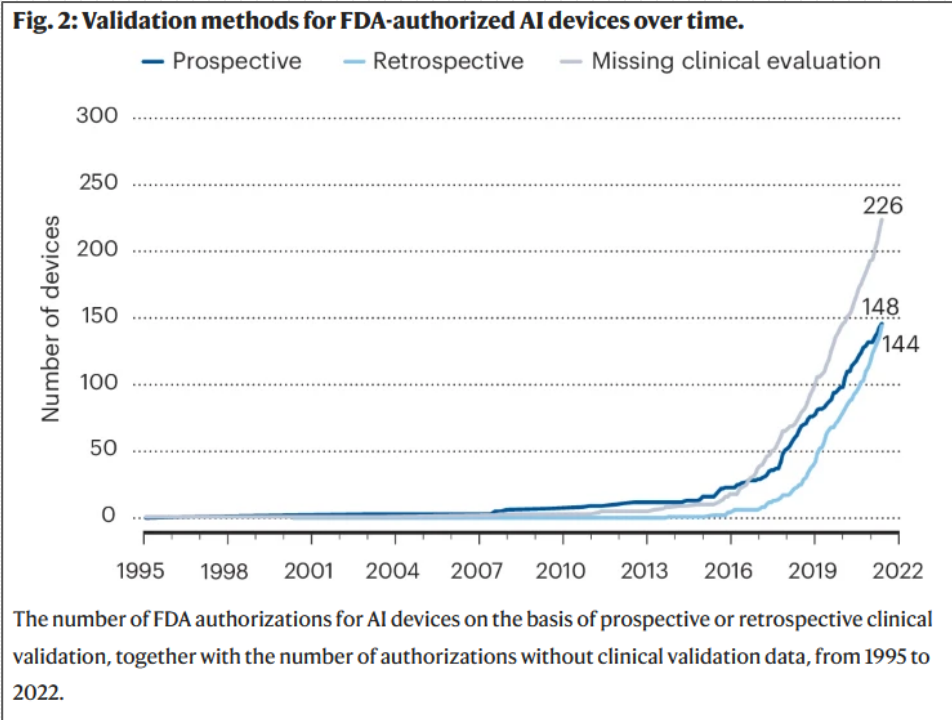
FDA承認を受けた521件の医療AIツールの臨床的有効性の検証状況を分析

Fig. 1: Validation methods for FDA-authorized AI devices by specialty.



- 最も多いのが放射線機器(392件、75%)であり、最も一般的な機器の機能は画像アーカイブおよび通信(123件、24%)であり、認証件数の大半(512件、98.3%)はクラスII機器
- 「RCT検証済み」、「プロスペクティブ検証済み」、「レトロスペクティブ検証済み」、または「臨床検証データなし」に分類
- FDAの2023年のガイダンスでは、メーカーへの推奨事項において、異なるタイプの臨床検証研究を明確に区別していない

AIを用いた医療機器はどのように検証すべきか？



- 292件(56%)が臨床的検証を報告しており、そのうち144件(27.6%)は「レトロスペクティブ検証」、148件(28.4%)は「プロスペクティブ検証」
- 226件(43.4%)については臨床検証データが欠落
- 2016年以降、臨床検証データが欠落している承認件数は、レトロスペクティブおよびプロスペクティブ(RCTを含む)検証済み医療機器の数を上回っている
- 一部のデバイスでは実際の患者データではなく、コンピューターで生成した架空の画像を使用していた

AIを用いた医療機器はどのように検証すべきか？

- 論文では、FDAと医療AI開発者は、より多くの臨床検証データを公開し、デバイス検証のための前向き研究を優先すべきであると主張
- FDAが後ろ向き調査のみによる検証を無条件に許容しているわけではない
- 実質的な同等性の原則やAI技術の急速な進化などを考慮すると、FDAは、限定的な状況下において、後ろ向き調査データによる検証をある程度許容しているともいえる



どのような検証が可能？

マクロ計量モデルに対するルーカス批判 (Lucas critique)

予測は絶対ではなく、機械学習アルゴリズムは過去と将来に影響を受ける。保険収載の可否、新薬の開発などが起こった場合には何が起こるかはわからない。ここに重大な限界がある。

ロバート・ルーカス(1937-)

1995年にノーベル経済学賞を受賞。数理的方法を用いてケインズ経済学を終わらせた人物。

1960-70年代には、変数の相関関係で政策立案を行うマクロ計量モデルによる政策決定が一般的に行われていた。例えば、インフレ率の低下は失業率の上昇に伴って生じるため、金利を引き下げて求人を高める、というようなことが行われていた。

ルーカスは、マクロ経済モデルを構造的に作らなければ、個々の方針の変化が全体に及ぼす影響を理解することはできないとした。政策というゲームのルールが変われば、パラメータも常に変わるはずである。



ロバート・ルーカス(1937-)
1995年にノーベル経済学賞

<https://www.nobelprize.org/prizes/economic-sciences/1995/lucas/biographical/>

問題の構造の理解があって、複雑な問題は多くの予測作業に分解される
よいデータサイエンスとは、他人に任せられる予測作業を見つけ、
難しい構造的な問題のための余力を残すためにある

2024年以降のAIを用いた医療機器の動向

AI医療機器開発のトレンド

- 放射線AIから「全診療科AI」への拡大:Radiology偏重は続くものの、敗血症診断、心不全・AFリスク予測、睡眠・精神医療、遠隔医療などへ急速に拡大
- リスク予測・トリアージ系AIの増加:Sepsis ImmunoScore や心血管リスクAIのように、「誰を優先的に精査・介入すべきか」を決めるツールが増加
- アップデートを前提としたAIプラットフォーム化:GI GeniusやSKOUTなど、データ追加に応じてアルゴリズムをアップデートし続ける内視鏡AIが普及。これに対応するため、「Predetermined Change Control Plan(PCCP)」が規制面で整備
- AI基盤モデル・LLMを組み込んだ医療機器の検討:FDAはAI搭載医療機器リストの中で、今後Foundation ModelやLLMを使う機器にタグ付けする方針を示しており、チャットボット型やマルチモーダルAIの医療機器への組み込みを見据えている

2024年以降のAIを用いた医薬品開発の動向

AIを用いた医薬品開発の動き

- AI薬物設計(Drug Discovery/Design): Insilico Medicine, ExscientiaなどによるAI支援
- 生成AIで標的探索から化合物設計まで行った薬剤がPhase 2まで進行し、初のAI設計薬の有望なPhase 2aが報告。
- Insilicoの腫瘍領域MAT2A阻害薬が2024年にFDAのINDクリアランスを取得し、2025年に臨床試験開始
- 「AI設計薬のFDA上市」は出ていないが、Phase I~IIの成功例が積み上がっており、AI設計薬の成功確率に関する解析論文も(Phase Iで80-90%の成功率など)。

規制当局公認の「AI開発ツール」の誕生

- 2025年12月にFDAはMASH(脂肪肝炎)臨床試験用のAI病理解析ツール「AIM-NASH」Drug Development Toolとして承認。肝生検スライド画像から脂肪化・炎症・線維化をAIでスコアリングし、複数病理医の判定を代替・標準化して試験のばらつきを減らすことが目的。医療機器としてのAIだけでなく、薬事審査に使う評価ツールとしてのAIを、FDAが積極的に認め始めた象徴的な事例です。
- AIによる臨床試験設計・シミュレーション: 特に希少疾患領域では、AIを用いたリアルワールドデータからの合成対照群(Synthetic Control Arms, SCA)が、倫理的・実務的な利点から注目。

まとめ

- 医療・健康分野の各種データベースが整備されつつある。課題は個人情報保護を担保しながらデータの縦と横のリンケージ(標準化)を行うこと
- 臨床医学においては、データが人間の活動から生み出される故の分析の難しさがある
- データ生成背景の理解は重要
- 国による医療制度の違いや人種・文化的背景などによりアウトカムが異なることがある
- ICT産業と違い、トライ&エラーはできない
- 機械学習・AIモデルの特性を活かして、様々な機械学習モデルが医療分野で活躍中
- 臨床医、データサイエンティスト、エンジニアなどの異なる専門家間のコラボレーションが不可欠
- 倫理的・法的・社会的課題(Ethical, Legal and Social Issues)への対処も必要
- 異常検知、時系列解析、転移学習、画像登録、創薬、パターン認識など紹介しきれていない分野が多数ありますので、分野ごとの総説が出ていますので、ご興味がある方はそちらをご覧ください