令和６年度運営評議会

AI栄養統括研究室

背景・社会的意義等

- ○ データベース、AI技術を融合し、健康・栄養研究のDXと効率化に貢献
- ○ JST/NEDOなどの国策に合わせたプロジェクトへの参画と推進
- ○ データやAIモデル共有に基づく、社会ニーズに応じた産学官連携と人材育成の推進
- ○ 情報科学と生物学・化学・栄養学との異分野融合的な学術基盤を構築

目標・令和6年度実績・成果・課題

- ○ AI技術を利用したデータ解析フロー開発と循環器系疾患の予測に向けた重要指標抽出
- ○ 健康・栄養と腸内マイクロバイオームの関連解析
- ○ 代謝モデル、酵素探索技術と予測ツールの開発（NEDO PJ）
- ○ 産官学連携・共同研究を推進（JST・共創の場PJ）
- ○ 大規模言語モデル（LLM）を活用した健康食品の安全性情報の抽出

ポイント

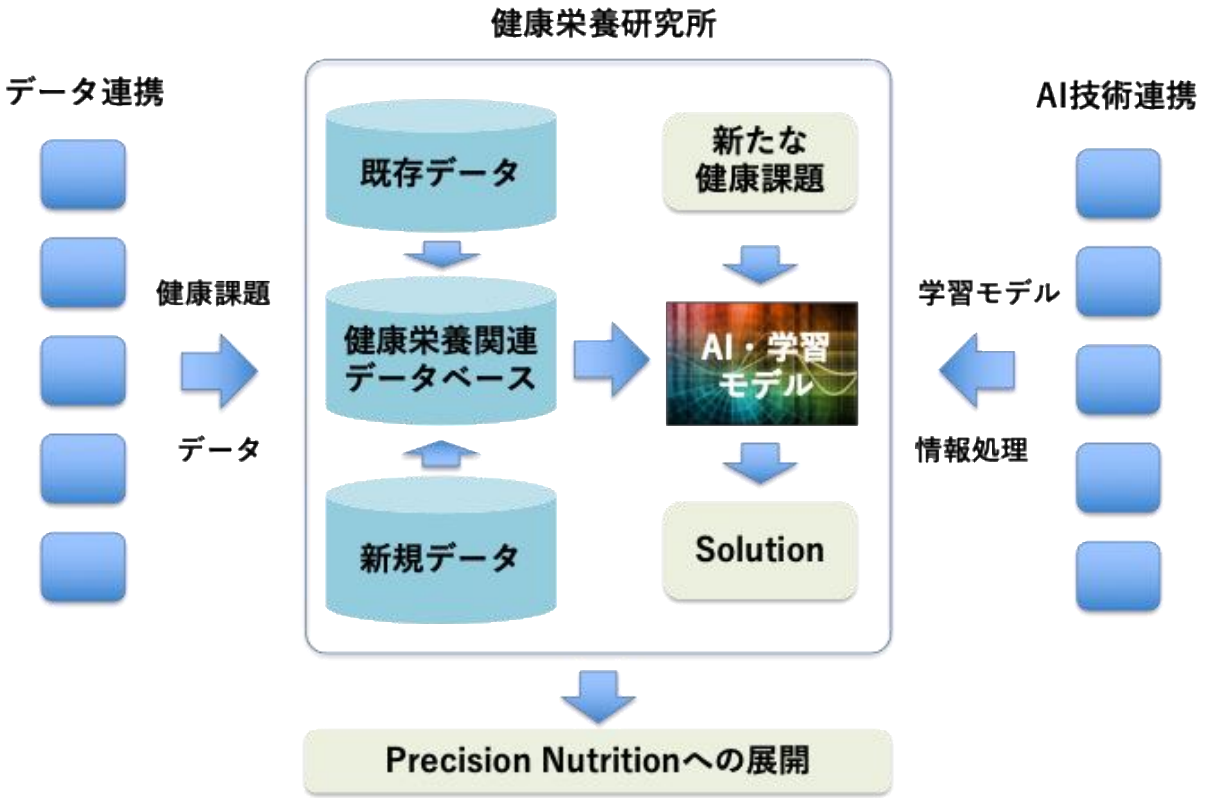
- ○ 循環器系疾患やフレイル予防・予測に向けた新たな指標の提案
- ○ 健康・栄養と腸内マイクロバイオームの代謝物の関係性の理解
- ○ 代謝予測・検索ツール公開による国内外の大学・企業との連携
- ○ 大規模言語モデル（LLM）を活用した健康食品の相互作用、安全性情報の抽出

AI栄養統括研究室のミッション

AIを活用した健康・栄養研究

- 循環器系疾患を中心とした生活習慣病やフレイル等に関わる新たな指標抽出に向けて、マイクロバイオーム、栄養・食生活、身体活動、健康食品に関連する各種データについて、生成AI技術の活用やAI・情報解析技術の開発をもとにデータ解析を実施する

- 所内外の研究組織、機関と連携して、上記に関する横断的な研究連携・技術的サポートを推進する



令和６年度の研究実績

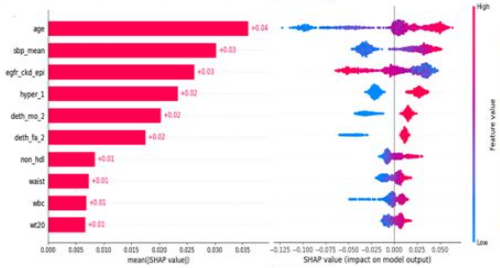
- 吹田研究データのAI解析と吹田研究NEXTデータの画像解析を中心に、AI技術の開発並びにデータ解析を実施、生活習慣病・心疾患に関わる重要指標の抽出（JST共創の場）
- 健康・栄養と腸内細菌に関するデータ解析（共同研究）
- 代謝モデル、酵素探索技術と予測ツールの開発（NEDO）
- 大規模言語モデル（LLM）を活用した知識抽出技術を開発する：健康被害データベース（消費者庁事業）

『AI栄養統括研究室』：AI解析フローの開発

成果

吹田研究（循環器病研究センター）について、AI解析フローを構築し、循環器系疾患（脳卒中、冠動脈疾患、心房細動）に関わる新たな重要指標、リスクスコアを抽出　（Stroke 2024, JCDD 2024）
→ 健都エリアの地域連携と共に疾患予防・予測に向けた新たな指標の提案につながることが期待される。

	ACC	AUC	Recall	Precision	F1-score
LR	0.66	0.71	0.66	0.66	0.66
RF	0.68	0.71	0.68	0.69	0.68
SVM	0.67	0.71	0.67	0.67	0.67
XGBoost	0.67	0.70	0.67	0.67	0.67
LightGBM	0.66	0.70	0.66	0.67	0.66



説明変数の重要度解析

ﾍﾞｰｽﾗｲﾝﾃﾞｰﾀｾｯﾄ

被験者数: 7672 人
説明変数: 169 個
目的変数: 脳卒中の発症

ﾃﾞｰﾀの前処理: ｸﾘｰﾆﾝｸﾞ, 構造化, ｴﾝﾘｯﾁﾝｸﾞ

- 新しい変数の作成 (eGFR, Non HDLc, BMI, 高血圧, 糖尿病, ﾒﾀﾎﾞ)
- 欠損値の処理 (欠損率30%未満を許容)
  + ｶﾃｺﾞﾘｰ変数は最頻値で, 数値変数はﾗﾝﾀﾞﾑﾌｫﾚｽﾄを用いて補完
- 重複/余剰な変数の削除

前処理後: 被験者数 7389 人 (女=4012, 男=3377), 説明変数 57個

ﾈｶﾞﾃｨﾌﾞｻﾝﾌﾟﾙのｱﾝﾀﾞｰｻﾝﾌﾟﾘﾝｸﾞ

- 正例(脳卒中発症)に対し、負例が極端に多い
- ﾓﾃﾞﾙが一方のｸﾗｽに偏るの防ぐ

学習用とﾃｽﾄ用にﾃﾞｰﾀを分割
(train:test=80:20)

教師ありﾓﾃﾞﾙ

ﾛｼﾞｽﾃｯｸ回帰(LR), ﾗﾝﾀﾞﾑﾌｫﾚｽﾄ(RF), SVM, XGBoost, LightGBMを脳卒中予測に適用

最適ﾓﾃﾞﾙの選定

各ﾓﾃﾞﾙのACC, AUC, Recall, Precision, F1-scoreを比較

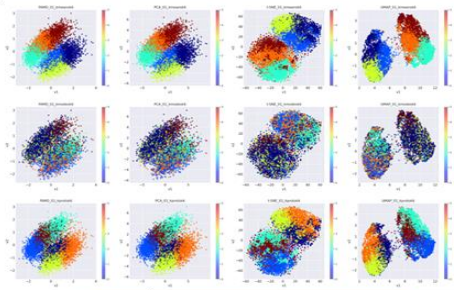
重要変数の抽出

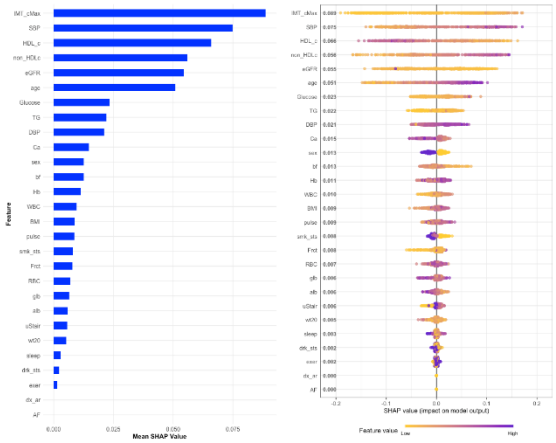
SHAPを用いて、脳卒中発症に最も重要な変数を抽出

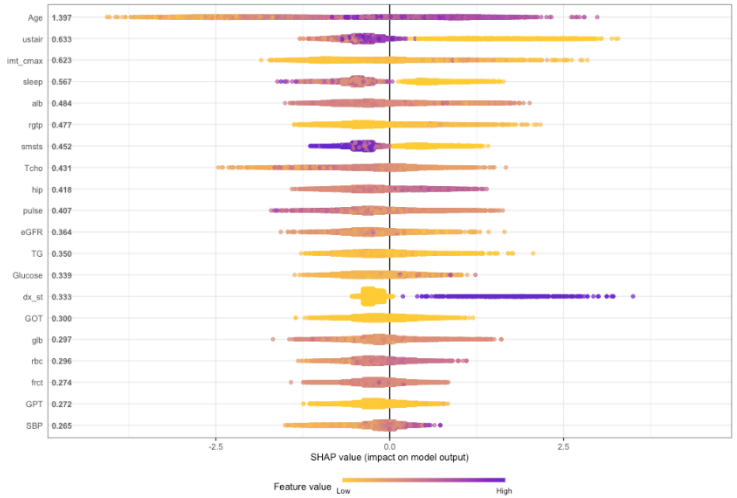
教師なしﾓﾃﾞﾙ

ｸﾗｽﾀﾘﾝｸﾞ法を用いて、脳卒中の発症をﾍﾞｰｽとしたｸﾗｽﾀﾘﾝｸﾞによるｸﾞﾙｰﾌﾟ分けを行った。ｸﾗｽﾀﾘﾝｸﾞ数は、ｴﾙﾎﾞｰ法とｼﾙｴｯﾄ法に基づいて決定した。本ﾃﾞｰﾀは数値ﾃﾞｰﾀとｶﾃｺﾞﾘｰﾃﾞｰﾀの両方から構成されているため、k-meansとk-modesを組み合わせたk-prototypesｸﾗｽﾀﾘﾝｸﾞを適用した。

可視化解析による探索的データ分析



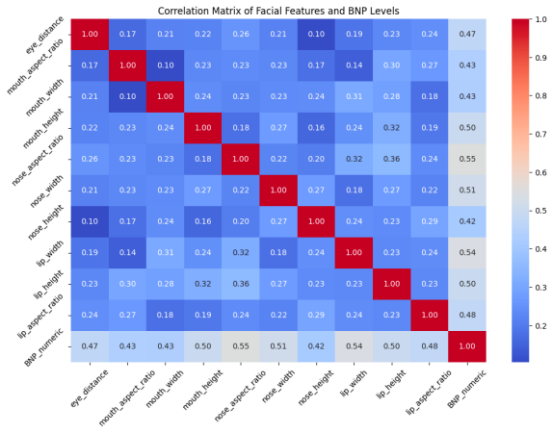


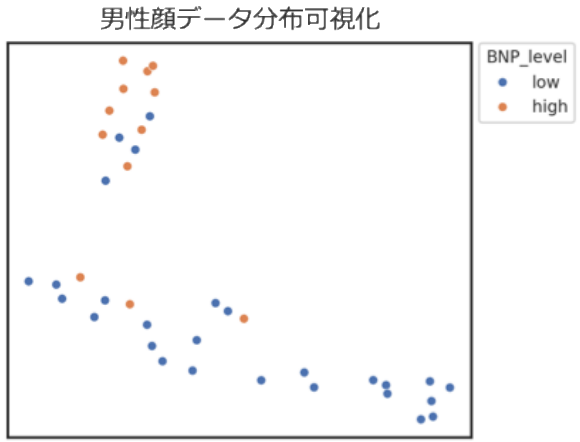


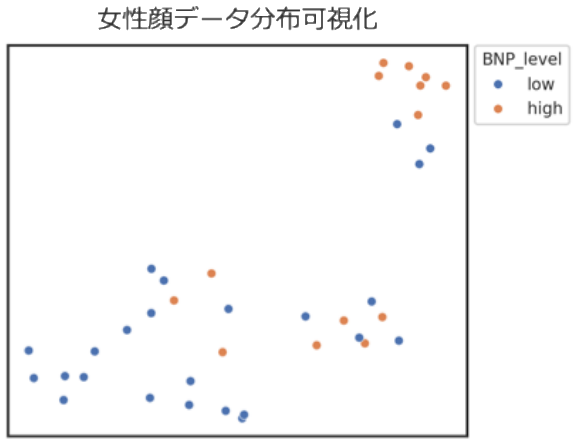
特に、機械学習のアプローチにより、既存の危険因子を確認するとともに、脳卒中モデルの新しい予測因子を探索することができます。

『AI栄養統括研究室』：画像データ解析フローの開発

吹田研究NEXT （循環器病研究センター）を利用したAI解析において、特に顔・表情画像と心疾患・認知症の関連解析を実施
→ AI解析に向けた利用データの拡張





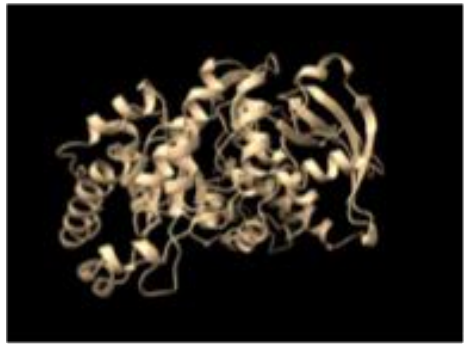


心不全指標（BNP）と顔画像データの相関解析と2次元マッピング

腸内細菌ゲノムデータの活用

理研・富岳と連携して、腸内細菌ゲノムにコードされた酵素配列情報を利用した網羅的なタンパク構造予測(AlphaFold)を実施 （医薬健栄研・理研・京大共同研究）
→ 腸内細菌代謝のより詳細なメカニズム解明が期待される

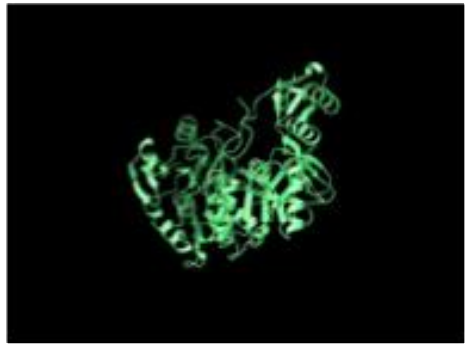
IGC1152131



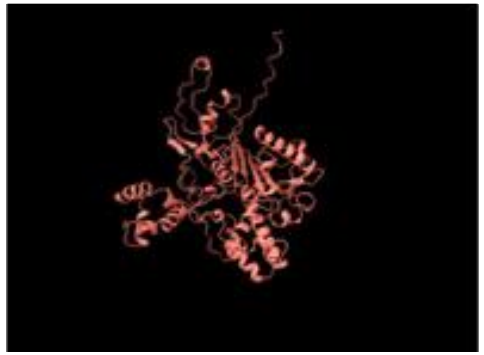
IGC1152136

IGC1152152

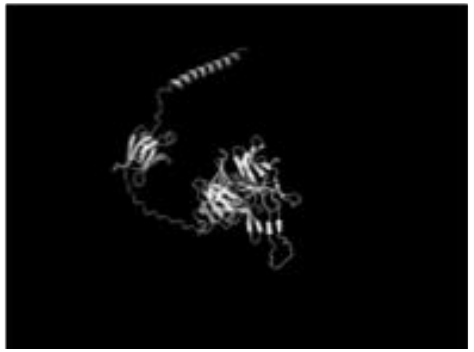
IGC1152155



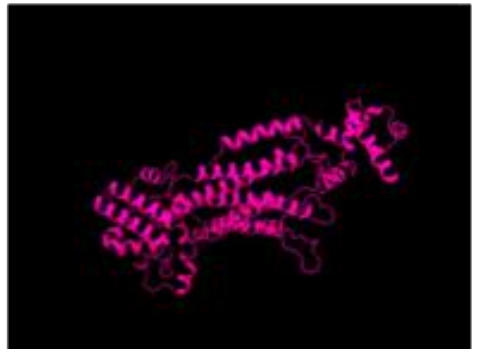
IGC1152176



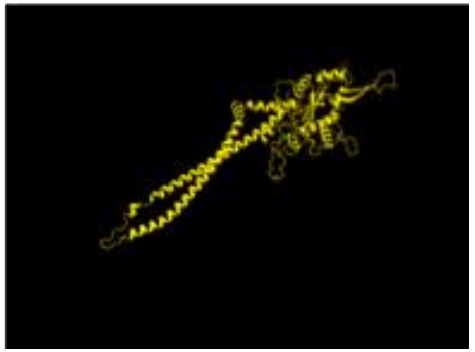
IGC1152181



IGC1152203



IGC1152212



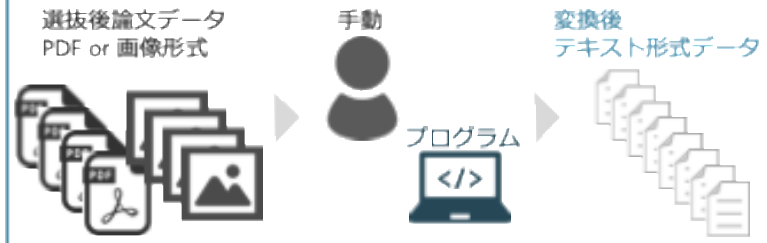
AlphaFold2によるタンパク質構造の大量推定（約10万配列）による代表的な構造

知識抽出技術の開発：健康被害データベース

1. 対象論文の検索

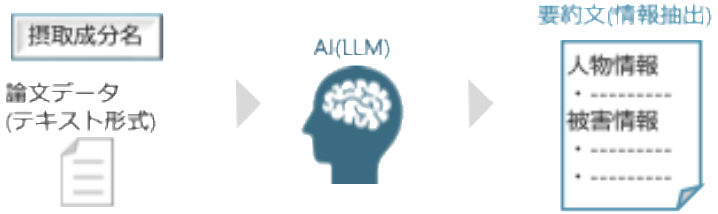
PubMedの検索エンジンを用いて、
膨大な論文データから対象論文を選抜。
※検索式は手動で決定。

2. テキストデータへの変換



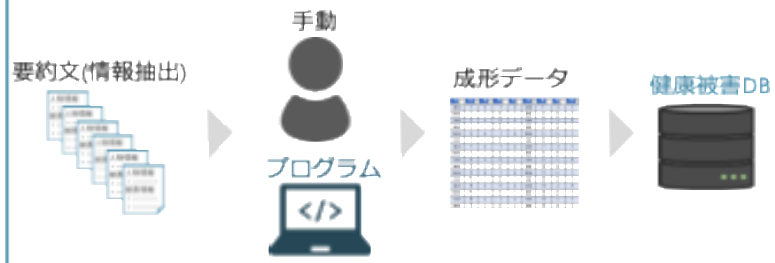
これ以降の処理のために
手動またはプログラムを用いて、
論文データ(PDFや画像形式)をテキスト形式に変換

3. 論文のAI要約



AI(LLM: Large Language Model)を用いて
摂取成分名と論文データから
人物や健康被害情報などを抽出し要約文を作成

4. 要約結果の成形とDB構築



手動またはプログラムを用いて
要約結果をDB用に成形。
成形したデータを蓄積しDBを構築。

今後の取組

- 吹田研究NEXTデータの画像解析を中心に、AI技術の開発並びにデータ解析を実施、生活習慣病・心疾患に関わる重要指標の抽出を行う（JST共創の場）
- 健康・栄養と腸内細菌に関するデータ解析を行う（共同研究）
- 大規模言語モデル（LLM）を活用した知識抽出技術を開発する：健康被害データベース（消費者庁事業）