



Title	構造化・非構造化臨床データを活用した診断支援AI手法の開発と評価
Author(s)	韓, 豊
Description	配架番号 : 2986
Degree Grantor	北海道大学
Degree Name	博士(医学)
Dissertation Number	甲第16893号
Issue Date	2026-03-25
DOI	https://doi.org/10.14943/doctoral.k16893
Doc URL	https://hdl.handle.net/2115/99738
Type	doctoral thesis
File Information	HAN_Feng.pdf, 全文



学位論文

構造化・非構造化臨床データを活用した診断 支援 AI 手法の開発と評価

(Development and Evaluation of AI Methods for Diagnostic Support
Using Structured and Unstructured Clinical Data)

2026 年 3 月

北 海 道 大 学

韓 豊

(HAN FENG)

学位論文

構造化・非構造化臨床データを活用した診断 支援 AI 手法の開発と評価

(Development and Evaluation of AI Methods for Diagnostic Support
Using Structured and Unstructured Clinical Data)

2026 年 3 月

北 海 道 大 学

韓 豊

(HAN FENG)

目次

発表論文目録および学会発表目録.....	1 頁
要旨.....	2 頁
略語表.....	5 頁
全体の緒言.....	7 頁
1. 研究の背景.....	7 頁
1.1 人工知能と機械学習について、.....	7 頁
1.2 自然言語処理、固有表現抽出と大規模言語モデルについて、...7 頁	
1.3 構造化データと非構造化データについて、.....	8 頁
2. 構造化データ解析の現状と課題.....	8 頁
3. 構造化データ解析課題の解決策.....	9 頁
4. 非構造化データ解析の現状と課題.....	9 頁
5. 非構造化データ解析課題の解決策.....	11 頁
6. 研究課題と本研究の目的.....	12 頁
第一章 構造化データに基づく入浴後体温予測：ファジー推論システムの有効性検証.....	13 頁
1.1 緒言.....	13 頁
1.1.1 現状と課題(入浴と体温予測).....	13 頁
1.1.3 ファジー理論について.....	13 頁
1.1.4 ファジー推論システム(FIS)について.....	14 頁
1.1.5 目的.....	14 頁
1.2 方法.....	14 頁
1.2.1 データセット.....	15 頁
1.3.2 前処理.....	16 頁
1.3.3 特徴選択.....	16 頁
1.3.3.1 Lasso 正則化(L1).....	16 頁
1.3.3.2 Ridge 正則化(L2).....	16 頁
1.3.3.3 Random Forest(RF：特徴重要度の算出).....	17 頁
1.3.3.4 逐次後退法(RFE).....	17 頁
1.3.4 モデルの学習と推論.....	17 頁
1.3.4.1 Adaptive Neuro-Fuzzy Inference System(ANFIS).....	18 頁

1.3.4.2 Evolving Takagi–Sugeno System(eTS)	19 頁
1.3.4.3 Evolving Fuzzy Inference System(EVOFIS)	19 頁
1.3.4 評価指標と統計解析	19 頁
1.4 結果	21 頁
1.4.1 前処理の結果.....	21 頁
1.4.2 特徴選択の結果	21 頁
1.4.3 モデル学習の結果	25 頁
1.4.4 予測結果.....	26 頁
1.4.5 可視化の結果.....	28 頁
1.4.6 評価・統計解析の結果	28 頁
1.5 考察	29 頁
1.5.1 主要知見の解釈	29 頁
1.5.2 非接触測定の限界	30 頁
1.5.3 安全性・応用上の意義	30 頁
1.5.4 限界と今後の課題	30 頁
1.5.5 位置づけと第二章への接続	31 頁
第二章：日本語 EHR 呼吸器 SOAP の病名予測：多面的対照群による比較研究	32 頁
2.1 緒言	32 頁
2.1.1 現状と課題	32 頁
2.1.5 目的	33 頁
2.2 方法	33 頁
2.2.2 前処理.....	36 頁
2.2.3 ラベル付与(アノテーション).....	36 頁
2.2.4 NER モデル学習	37 頁
2.2.5 LLM タスク設計と予測	37 頁
2.2.6 評価指標と統計解析	43 頁
2.3 結果	44 頁
2.3.1 ラベル付与(アノテーション)の結果	44 頁
2.3.2 NER モデル学習	47 頁
2.3.3 LLM タスク設計と予測	47 頁
2.4 考察	53 頁

2.4.1 限界.....	54 頁
2.4.2 今後の展望	55 頁
2.4.3 本章の位置づけと全体結論との接続.....	55 頁
全体の結論.....	56 頁
1. 本研究から得られた新知見	56 頁
2. 新知見の意義.....	57 頁
3. 本研究で得られた新知見から今後展開されうる研究	57 頁
4. 今後の課題	57 頁
謝辞.....	58 頁
利益相反	59 頁
引用文献	60 頁

発表論文目録および学会発表目録

本研究の一部は、以下の論文に発表した。

Han F, Tang M, Zhang Z, Hirata K, Okugawa Y, Matsuda Y, Nakaya J, Ogasawara K, Kudo K. Prediction of Post-Bath Body Temperature Using Fuzzy Inference Systems with Hydrotherapy Data. Healthcare. 2025; 13(9):972. <https://doi.org/10.3390/healthcare13090972>

本研究の一部は、以下の学会に発表した。

FENG HAN, ZIHENG ZHANG, MINGHUI TANG, KENJI HIRATA, GENKI HATA, KATSUHIKO OGASAWARA, KOHSUKE KUDO, JUN NAKAYA, Analysis and prediction of the effect of balneotherapeutic factors on metabolism: a pilot study using the ANFIS approach, 6th FHS International Conference, 20 October, 2023, Sapporo

韓 豊、張 孜恒、平田 健司、松田 悠之介、小笠原 克彦、工藤 與亮、唐 明輝、フuzzy推論システムを用いた入浴療法における入浴後の深部体温予測、第7回日本メディカル AI 学会学術集会, 2025.6.27, 京都

韓 豊、張 孜恒、唐 明輝、平田 健司、秦 元気、松田 悠之介、小笠原 克彦、工藤 與亮、中谷 純、入浴療法の効果分析と予測:Fuzzy 推論を用いた検討、第6回日本メディカル AI 学会学術集会, 2024.6.21, 名古屋

韓 豊、張 孜恒、張 洪健、唐 明輝、平田 健司、小笠原 克彦、工藤 與亮、中谷 純、Fuzzy-c-means法を用いた医療テキストからの程度を表すことばの抽出と数値化。第5回日本メディカル AI 学会, 学術集会, 2023.6.18, 東京

要旨

【背景と目的】

医療現場の電子カルテ（electronic health record: EHR）には日々多様なデータが蓄積されている。大半は数値データ(検査値などの構造化データ)と自由記述テキストデータ(診療録などの非構造化データ)に大別される。

構造化データは体温・各種検査値などの規格化された定量指標で、数値中心の連続量として分布する。これらの指標は年齢や基礎疾患、治療介入と相互作用し、非線形かつ時系列依存を伴う。従来の利活用手法はこうした特性に十分対応していない。体温や血圧のような定量指標などに対して臨床便宜上の閾値二値化は運用上有用だが、情報損失や閾値近傍の不安定性を招く。また、機械学習手法であるランダムフォレスト(random forest : RF)は非線形を表せる一方、分割の累積で境界が段差化し連続変化や寄与の解釈が難しくなる。ニューラルネットワークやサポートベクターマシン(support vector machine : SVM)は非線形に強いがブラックボックス化しやすい。従って、二値化に頼らず連続量を扱い、相互作用・時系列を捉えつつ、根拠を人が理解できる形で提示できる手法が求められる。

一方、非構造化データは医師の自由記述や所見などのテキストで、豊富な臨床文脈を含むが、施設方針や記載者の癖、略語・誤記、日本語特有の分かち書き欠如や全角・半角混在など表記揺れが大きい。また、過去情報の持ち回りやコピー＆ペーストで冗長化し、重要情報がノイズに埋もれやすい。それを解決するため、従来では固有表現認識(named entity recognition : NER)を利用し、疾患・症状・薬剤等を抽出・正規化して冗長性を低減する。また、近年の汎用型の大規模言語モデル(large language model : LLM) が医療にも応用され、内蔵の正規化・サブワード分割を備え、同一 LLM で推論まで一気通貫に扱えるため期待が高い。ただし、このような前処理の妥当性が保証できなく、特に医療に対し、前処理の正確さが求められるため、NER の結果を用いた推論との検証が必要である。さらに、非構造化医療データの研究及びモデルの公開は英語テキストが中心であり、日本語医療データに対して LLM モデルの妥当性・再現性・前処理の有効性を体系的に検証する基盤が不足している。したがって、日本語 EHR においては、タスク設計(入力様式・評価指標)とモデル設計(汎用 LLM の指示調整)を比較検討し、前処理の寄与と限界を明らかにする必要がある。

以上のように、構造化データと非構造化データの特性が根本的に異なるため、人工知能(artificial intelligence, AI)を用い、データ利活用する際に、各データの特性に対応できるアルゴリズムの選択が重要である。

本研究は、以上の課題を踏まえ、構造化データと非構造化データという二つのデータ様式に対する診断支援 AI の実現可能性を、二つの実証研究を通じて検討すること

を目的とする。第一章では、構造化データとして連続的生理量を扱う推論アプローチの有効性を、第二章では、非構造化データとして日本語 EHR におけるタスク設計およびモデル性能の妥当性を比較検討する。

【対象と方法】

第一章では、1991～1997 年に収集された入浴関連データ(左手深部、左手非接皮表、右手皮膚、前額部皮膚の 4 部位×前値+10/20/30/60 分、浴温・入浴時間ほか計 25 変数)を整備し、欠損・外れ値除去と表記正規化後、215 セッション中 213 セッションを解析対象とした。目的変数は各部位の入浴 60 分後体温、説明変数は前値・各中間時点温度・浴温・入浴時間等である。L1 正則化/L2 正則化/RF およびそれぞれと逐次後退法(recursive feature elimination : RFE)の併用で特徴を 6 項目まで選抜し、三つのファジィ推論システム(ANFIS(Adaptive Neuro-Fuzzy Inference System)/eTS(evolutionary Takagi-Sugeno モデル)/EVOFIS(Evolving Fuzzy Inference System))と機械学習手法の RF・SVM を学習し、平均二乗誤差(MSE)と決定係数(R^2)を主要指標に比較した。また、メンバーシップ関数とルールの可視化により説明可能性を評価した。

第二章では、北海道大学病院呼吸器内科の電子カルテ記録(SOAP: Subjective, Objective, Assessment, Plan)を使用した。(a) 自由記述テキストの全文入力と(b) NER による構造化要約テキストの 2 種類の入力データを作成し、(1) 自由予測(病名名を上位 3 候補まで生成)と(2) 選択予測(55 の候補疾患リストから最大 3 件を選択)の 2 種類のタスクを設定した。対象症例数は初診 200 例と再診 200 例の計 400 例である。比較には GPT-5(Generative Pre-trained Transformer)、GPT-4.1、Llama-3.3-70B(Large Language Model Meta AI)の 3 種類の大規模言語モデルを用いた。各モデルには温度パラメータを 0.2 に設定し、出力フォーマットを固定、同一の prompt 設定で推論を実行した。同一症例で二つの入力条件(非構造化の全文/構造化の要約)を厳密に対応させ、単段階と二段階の prompt 設計も比較した。評価は Top-1 正解率(最上位 1 件が正解である割合)と Top-3 包含率(上位 3 件のいずれかに正解が含まれる割合)を用い、二項比率の両側検定(p 値)により群間差を検証し、多重比較は偽発見率(FDR)で補正した。

【結果】

第一章では、入浴後 60 分の体温予測においてファジィ推論システム(fuzzy inference system : FIS)が有効であることを示した。特に深部体温では EVOFIS が最良の誤差を達成し(例：左手深部温 MSE=0.038°C)、非線形性と時系列依存の強い現象への適合性を確認した。一方、表在温では ANFIS と SVM が高い精度を示し(右手皮膚：MSE=0.028°C、 R^2 =0.92、前額部：MSE=0.047°C、 R^2 =0.90)、部位・指標により最適手法が異なることを示した。併せて、ルール可視化により予測根拠を提示でき、説明可能性を担保できることを示した。

第二章では、日本語 EHR の診断推論において選択予測が一貫して自由予測を上回

った。Top-1(非構造化 40.54% vs 25.08%、構造化 40.29% vs 25.83%)および Top-3(非構造化 61.08% vs 40.58%、構造化 60.25% vs 41.25%、すべて $p<0.001$)。受診別では初診が非初診より高精度で(自由：32.17% vs 19.50%、選択：48.17% vs 32.92%、すべて $p<0.001$)、入力形式(構造化/非構造化)の影響は小さく有意差はなかった。prompt の効果はタスク依存で、自由予測では有意差がなく(23.91% vs 26.25%、25.25% vs 26.42%)、選択予測では指示追従調整(prompt tuning)が有意に改善(+6.08%、 $p<0.01$ ；+9.92%、 $p<0.001$)。同一モデル内(例：Llama)でも選択予測が自由予測を大きく上回る傾向が確認された。

【考察】

第一章と第二章の結果から、データ形式の違いによって最適なアルゴリズム選択および運用上の要件が規定されることが明らかとなった。第一章では、FIS の中でも EVOFIS が深部体温において最小誤差(左手深部温 $MSE=0.038^{\circ}C$)を達成した。表在温においては、ANFIS や SVM が高精度を確認した(右手皮膚 ANFIS: $MSE=0.028^{\circ}C$ 、前額部 SVM: $MSE=0.047^{\circ}C$)。また、FIS はルール可視化により根拠提示も可能である。第二章では、選択予測タスクが自由予測タスクを Top-1 で約 15%、Top-3 で約 20%上回り、入力の違い(構造化及び非構造化)による主効果は小さく、選択予測タスクにおける prompt tuning が+6~10%の上乗せを示した。これは、モデル種別よりも「候補制約×Top-k 提示×出力スキーマ整合×段階的推論」という設計が性能を規定することを意味する。総じて、数値データは FIS+可視化、テキストデータは候補制約+Top-k +リッチ構造化+適切な prompt という設計原理が、背景で指摘した二値化による情報損失と自由記述の冗長・表記揺れを同時に緩和し、説明可能性と実装容易性の両立に資する。

【結論】

本研究は、構造化・非構造化の二様式について、二つの実証研究で診断支援 AI の実現可能性を検証し、妥当性と限界を示した。今後は、第一章の FIS 体温予測をウェアラブルや介護の安全ダッシュボードに実装し、湯温・入浴時間・クーリング等の個別推奨を出す。第二章で有効性を示した選択式予測を核に、疾患カテゴリー別の候補診断提示と症例・タスクに応じた prompt/モデル選択を整備する。構造化と非構造化を統合する意思決定支援基盤を構築し、精度・説明可能性・処理効率の最適化を図る。

(2991 字)

略語表

本文中および図中で使用した略語は以下のとおりである。

略語	英語正式名
AI	artificial intelligence
AMIE	Articulate Medical Intelligence Explorer
ANFIS	Adaptive Neuro-Fuzzy Inference System
DE	Differential Evolution
eTS	evolving Takagi–Sugeno system
EHR	electronic health record
EMR	electronic medical record
EVOFIS	Evolving Fuzzy Inference System
FDR	false discovery rate
FIS	fuzzy inference system
GPT	Generative Pre-trained Transformer
ICD-10	International Classification of Diseases, 10th Revision
JSON	JavaScript Object Notation
L1	Lasso (least absolute shrinkage and selection operator)
L2	ridge regression
LLaMA	Large Language Model Meta AI
LLM	large language model
MAE	mean absolute error
ML	machine learning
MSE	mean squared error
NER	named entity recognition

NLP	natural language processing
PR	pulse rate
RBF	radial basis function
RF	random forest
RFE	recursive feature elimination
SOAP	Subjective, Objective, Assessment, Plan
SVM	support vector machine
SVR	support vector regression
TS	Takagi–Sugeno
XLM-RoBERTa	cross-lingual RoBERTa

全体の緒言

1. 研究の背景

1.1 人工知能と機械学習について、

電子カルテシステムの普及に伴い、医療現場には医用画像及び診療記録テキストなども含め、多様で膨大な臨床データが継続的に蓄積されている(Ross, M. K., et al., 2015)。このようなデータの利活用において、人手のみでこの情報から有用な知見を抽出するには限界がある。近年、人工知能(artificial intelligence: AI)の進みにより、このようなデータの利活用が期待されており、疾患の早期発見・重症化予防、治療効果の評価、業務負担の軽減、患者ケアの質向上を実現するには、AIを含む先進的なデータ解析の導入が不可欠である(Olusegun, J., et al., 2024; Moafa KMY, et al., 2024)。人工知能とは、人間の知的活動を模倣し、学習・推論・判断・予測といった複雑な処理をコンピュータが自律的に実行する技術の総称である。その中、コンピュータが大量のデータをもとに規則やパターンを学習し、その知見に基づいて新たなデータを予測・分類する機械学習(machine learning: ML)という方法がある。代表的な手法には、決定木やランダムフォレスト(random Forest: RF)やサポートベクターマシン(support vector machine: SVM)やニューラルネットワークを用いたディープラーニングが含まれる。医療分野では、がんの予後予測、再入院リスク評価、敗血症の早期検知など多様な課題に適用されてきた(Kourou et al., 2015)。具体的な応用面は、医用画像を対象とするディープラーニングと診療記録等のテキストを対象とする自然言語処理(natural language processing: NLP)に大別できる。医用画像分野では、ディープラーニングにより胸部 X 線や CT から病変を高感度に検出するシステムが実装され、肺がんの胸部 X 線診断で AI の感度が 56.4~95.7%と放射線科医の 23.2~76%を上回ったとの報告がある(Megat Ramli, P. N., et al., 2025)。一方、テキスト領域では、情報抽出(Ohno, Y., et al., 2024)・要約(Fraile Navarro, D., et al., 2025)などの NLP 応用が進むものの、テキストには検査値やバイタルなどの構造化データと自由記述等の非構造化データが混在するため、その取り扱いが性能と汎用性を左右する。

1.2 自然言語処理、固有表現抽出と大規模言語モデルについて、

自然言語処理とは、人間の言語から意味・関係・意図を機械的に理解させる技術であり、テキスト要約、情報検索、文章分類、文章生成、対話などを可能にする基盤技術である(Lopez, I., et al., 2025)。医療領域における現行の NLP 応用例は、電子カルテ

記録の構造化(自由記述テキストからのデータ抽出)、診療ラベルの自動付与、匿名化(個人情報の検出と置換)、臨床コーディング支援、サマリーや退院時要約の自動生成などと多岐にわたる(Z. Kraljevic et al., 2023)。これらの中核技術は、テキストから医学的概念を抽出する固有表現抽出(named entity recognition: NER)と、大規模言語モデル(large language model: LLM)による推論・生成である(Shimizu, S., et al., 2025 ; Tu, T., et al., 2024)。

1.3 構造化データと非構造化データについて、

電子カルテ(electronic medical record: EMR)や電子健康記録(electronic health record: EHR)は、処方情報、臨床報告書、検査データ、自由記述テキストなど、多様な情報を包含するリッチなデータ基盤である(Ross, M. K., et al., 2015)。これらを統合し、臨床意思決定に還元することは医療の質の向上と業務負担の軽減に直結する。また、EMR/EHR に蓄積されるデータは、構造化データと、非構造化データに区分できる。構造化データは、検査値、体温、血圧、脈拍などの定量的・規格化された情報であり、集計や統計解析に適する。これに対し、非構造化データは、医師の自由記述、経過記録、所見記述といったテキストなどの形式で保存される情報を指し、豊富な臨床文脈を含む。特性が異なるため、解析の方法及び課題が異なっている。

2. 構造化データ解析の現状と課題

構造化データは、医療情報の中でも取り扱いやすく、診断支援や予後予測などで広く活用されている。臨床で扱う多くの生体指標(体温、SpO₂、乳酸値、血圧など)は典型的な構造化データである。従来の機械学習手法ではこのようなデータを解析できるが、データの二値化による問題、非線形のデータの対応性及び結果の解釈性に課題が残されている。

臨床現場では指示・トリアージ・運用の単純化のために、連続量を閾値で区切って「発熱/平熱」「陽性/陰性」等の二値に整理して扱うことが多い。そのため二値化(0/1)は実務上の便宜ではあるが、情報損失や閾値近傍の不安定性を招く(Shawn Hemelstrand, et al., 2025; Royston, P., et al., 2006)。特に、決定木やランダムフォレストなどの機械学習手法が、二値化複数回を行い、行うたびに境界が段差状になりやすく、その結果、連続的な臨床状態の把握や説明可能性を損なうことがある(Durán JM, et al., 2021)。一方、SVM やニューラルネットワークなどの手法は二値化を利用しないため、数値データの説明可能性を生かせることができる。

実際の解析応用では年齢・基礎疾患・バイタルサイン・検査値・治療介入など多数の因子が相互に作用し、その効果は単純な加算では説明できないため、問題は本質的

に非線形である。特に、決定木やランダムフォレストなどの機械学習手法では線形モデルであり、解釈性は優れたが、複雑な医療データに対し、限界がある。SVM とニューラルネットワークなどの手法では非線形データを対応できるが、解釈性においては SVM が高次元特徴空間への写像とサポートベクトルに依存するため、意思決定過程の解釈が難しい。また、ニューラルネットワークの内部表現がブラックボックス化しやすく、各因子が最終出力にどのように寄与したかを直観的に把握するのは容易ではない。

従って、数値データに対して単純の二値化を利用しなく、非線形性が対応できる、解釈性があることを同時に満たす手法が必要である。

3. 構造化データ解析課題の解決策

このような背景から、本研究ではファジー推論システム(fuzzy inference system: FIS)に着目する。FIS は「もし A ならば B(IF-THEN)」というルールとメンバーシップ関数を用いて、対象がある概念に属する度合い($\mu \in [0,1]$)を表現する。これは人間が「微熱」「中等度の痛み」といった曖昧な表現で程度を認識する方法に近く、従来の硬い二値を緩和した意思決定を可能にする(Akira Suzuki ,et al.,2024)。例えば「微熱」「発熱」など臨床的に用いられるあいまいな区分を連続量として扱えるため、閾値をまたぐ際の不安定性を低減しつつ情報を保持できる。また、FIS は各ルールが局所的に適用されるため、システム全体として複雑な非線形関係を柔軟に近似できる。さらに、ルールやメンバーシップ関数自体が人間に理解しやすい形で可視化・提示できるため、モデルの判断根拠を説明可能な形で示せるのも利点である (Gupta, Nisha ,et al.,2022)。

本研究の第一章では、入浴関連保健データを用い、入浴後の体温変化という連続的な生理現象を予測対象とし、提案される FIS 手法(Adaptive Neuro-Fuzzy Inference System(ANFIS)、evolving Takagi-Sugeno System(eTS)、Evolving Fuzzy Inference System(EVOFIS))の有効性を検証し、従来機械学習(RF、SVM)との性能差に加え、精度と説明可能性の観点から評価を行う。

4. 非構造化データ解析の現状と課題

SOAP(Subjective, Objective, Assessment, Plan)形式の EMR/EHR はテキストであり、患者の状態や経過に関する有用な情報を豊富に含む。しかし、その自由記述形式ゆえに形式的及び内容的に多くの問題点があり、自動解析が困難である。形式的に、まず、医療機関ごとの記録方針や記載者の癖の違いにより、同じ内容でも表現が多様化することが多い。また、手入力による誤字脱字や略語の乱用も頻繁に見られ、先行研究

(Yang, Z., et al.,2020)では電子カルテ中の低頻度トークンの多くがスペルミスや多様な省略形によるものであったと報告されている。さらに、日本語の診療文書では、単語間に空白がなく全角と半角文字が混在するため、単語境界の認識や表記ゆれの統一といった前処理が不可欠である(Kawazoe Y., et al.,2021)。内容的に、SOAP 形式の経過記録には診療経過や既往歴、処方箋、検査結果など過去から引き継がれた情報が大量に含まれる。このように過剰な情報が繰り返し記載された診療ノートの冗長化は広く指摘されており(Adam Rule., et al.,2021)、実際に近年の調査では外来経過記録の平均長が10年間で約60%伸び、前回記録からの重複部分が約10%増加したと報告されている(Adam Rule., et al.,2021)。冗長な記述は重要情報を埋もれさせ、記録の理解や利用を妨げる可能性がある(Adam Rule., et al.,2021)。そのため、テキスト中の情報冗長性が推論モデルの性能に与える影響を評価し、適切な前処理や要約による対策を検討する必要がある。

前処理の一手法として臨床固有表現認識(NER)は、診療テキストから病名・薬剤名・症状・検査・処置などを抽出し、用語集やIDへ写像して構造化データに変換するアプローチが広く用いられている。これにより、一定の情報冗長性を低減できる。代表例としてMedExは、臨床記録から薬品名・用量・頻度・投与経路などの用薬要素を高精度に抽出した(Hua Xu., et al., 2010)。日本語でも、MedNLP、Real-MedNLP、臨床日本語BERT(UTH-BERT)などのNER手法が開発され(Yada, Shuntaro., et al., 2022 ; Kawazoe, Y., et al., 2021)、MedNER-Jを用いた日本語の疾患情報抽出の実応用報告もある(Ohno, Y., et al., 2024)。一方で、自然言語処理分野ではGPTシリーズやLLaMAに代表される大規模言語モデル(LLM)には入力文字列に対する正規化、サブワード分割、必要に応じた前トークン化などの内蔵前処理があり、一般文書を対象とし、とりわけトークナイザーの選択は下流性能や計算効率に有意な影響を与えることが報告されている(Iaroslav Chelombitko., et al., 2024)。このようなLLMの内蔵前処理が便利であるが、前処理の妥当性が保証できなく、特に医療に対し、前処理の正確さが求められるため、検証が必要である。

推論においては、LLMが近年飛躍的な性能向上を遂げ、人間に匹敵する成果を挙げている。代表例として、MultiMedQA/Med-PaLM系列は医療試験型QAや消費者質問まで横断する評価枠組みを整備し、Med-PaLM 2はMedQA(USMLE形式)で最大86.5%を達成するなど、長文回答の推論力と根拠提示を大幅に引き上げたと報告されている(Singhal, K., et al.,2023)。また、AMIE(Articulate Medical Intelligence Explorer)は診療会話に最適化された研究用LLMで、問診から鑑別診断、方針提案までを対話的に推論する設計が示され、診断推論の精度向上や会話品質の改善が示唆された(Tu, T., et al.,2025)。さらに、ChatGPTの米国医師国家試験(USMLE)に対する性能評価では、特別な医療微調整なしでも全ステップで合格ライン付近に到達し、解答の一貫性や洞察

の質も定量化され、汎用 LLM が医療推論課題にゼロ/少数ショットで適応し得ることを示す初期根拠となった(Kung TH., et al.,2021)。しかし、非構造化医療データの研究及びモデルの公開は英語テキストが中心で、日本語の医療領域に目を向けると、この種のモデルや手法の適用例はまだ少ない(Kawazoe Y., et al.,2021)。最近の研究においても、日本語 LLM の開発は汎用分野が中心でバイオ医療分野での進展は遅れており、比較評価のための包括的ベンチマークも欠如していると指摘されている (Junfeng Jiang., et al.,2025)。そのため、日本語医療テキストに対する LLM の性能検証、モデル間の比較研究や適用可能性については、体系的な評価が今後必要である。

LLM を医療タスクに適用するにあたっては、単に汎用モデルをそのまま用いるだけでなく、医療領域データで追加学習を行い、指示に従った応答生成ができるようモデルを指示追従調整(prompt tuning)といった工夫が有効と考えられる。実際、LLM に医学分野の指示データを与えて微調整することで、質問応答や情報抽出、テキスト生成といった BioNLP タスクの性能が大きく向上したと報告されている (Tran, H., et al.,2024)。また、モデルの継続的な改良や専門知識の付与も診断支援に有望である。例えば、Meta 社の LLaMA2 を基に性能強化が図られた後継モデルでは、症例報告を用いた鑑別診断リスト作成の正答率が前世代モデルに比べ約 1.5 倍に向上したとされる (Hirosawa T., et al.,2024)。このようなデータセット特化調整や prompt tuning の手法が日本語の医療データに対して有効に機能し、診断支援など下流タスクの精度向上に寄与するかどうかは、検証すべき重要な課題である。

テキストを NER で構造化表現へ変換し、その出力を LLM に与えて推論するアプローチ(NER 前処理を組み合わせた手法)の効果と、医療データ・同一評価系で厳密に比較した研究は十分に蓄積されておらず、両者の貢献差を客観的に位置づけるには追加検証が必要である。

5. 非構造化データ解析課題の解決策

以上のように、非構造化である医療テキストの解析には形式を修正する前処理や情報冗長性への対処、さらに汎用モデルの活用と適応が不可欠であるが、日本語の医療テキストに着目した研究はまだ十分とは言えない。本研究では、これらの課題に対して大規模言語モデル技術を診療記録に適用し、その有効性と問題点を検証する。

本研究の第二章では、北海道大学病院の SOAP 形式電子カルテの S(主観情報)と O(客観情報)を用い、診断名予測をタスクとし、汎用 LLM モデルである GPT-5、GPT-4.1、Llama 3.3-70B の3種類それぞれで推論を行った。その際、前処理は NER 及び LLM の内蔵前処理のみをそれぞれ行い、推論の精度を比較する。また、プロンプト設計の工夫として、単段階と二段階の prompt tuning 手法を比較し、診断予測における効

果の違いを評価する。ノートの冗長化が診断精度に与える影響を検討するため、入力データは初診群と再診群に分けて、予測精度の比較を行う。

6. 研究課題と本研究の目的

本研究の目的は、性質の異なる医療情報である構造化データと非構造化データに対し、それぞれに適した AI 手法を実装・比較し、医療 AI の実運用における有効性と課題を明らかにすることである。具体的には、構造化データには連続量の表現・非線形性の捕捉・説明可能性を兼ね備えるファジー推論システム(FIS)の適用可能性を検証した。非構造化データには自然言語処理による前処理(NER)と大規模言語モデル(LLM)を組み合わせた診断予測の有効性と限界を評価する。さらに、モデル間比較、プロンプト設計(単段階/二段階)、初診と再診の違いの三つの観点から、医療テキストにおける LLM の精度を比較検証する。構造化・非構造化の両データの比較を通じて、データ特性(構造化・非構造化)と AI モデルの適合性に関する探索的知見の抽出を目指す。第一章では、入浴に関連した生理データを用いて FIS(ANFIS・eTS・EVOFIS)と従来手法(RF・SVM)を比較し、非線形性への適応や説明可能性の観点から FIS の有効性を検討する。第二章では、SOAP 形式の診療記録を用いて NER によりテキストを構造化し、GPT-5・GPT-4.1・LLaMA 3.3 70B といった LLM に入力することで病名予測を実施した。さらに、モデル間比較、プロンプト設計(単段階 vs 二段階)、および初診と再診の違いに基づく情報冗長性や処理効率への影響を評価する。これらを通じて、医療データの特性に応じた AI モデル選定の重要性とその運用指針を明確化し、医療現場への AI 導入に向けた基盤構築に寄与することを目指す。

第一章

構造化データに基づく入浴後体温予測：ファジー推論システムの有効性検証

1.1 緒言

1.1.1 現状と課題(入浴と体温予測)

入浴(ハイドロセラピー)は、日本の高齢社会における代表的なセルフケアであり、循環・代謝・体温調節へ作用して一次予防や QOL 維持に資する一方(Bender, et al., 2005)、個人差や環境要因との相互作用により入浴後の体温推移が大きく変動する。体温は皮膚温(表在)と深部体温(コア)に大別され、皮膚温は外界の影響を受けやすく、深部体温は恒常性維持の観点から厳密に制御されるため、高齢者・要介護者などの体温調節能が低下した人ではわずかな変動でも重篤な危険につながり得る(Kenny, et al., 2018)。深部体温の測定にはさまざまな方法があるが、多くは侵襲的または専用機器を要し、非侵襲法が求められる。近年、赤外サーマル像の回帰($\pm 0.285^{\circ}\text{C}$ /平均絶対値誤差 0.240°C)、皮膚温+熱流束の推定(RMSD $0.28\text{--}0.34^{\circ}\text{C}$)、外耳道センサ(95%一致限界 $\pm 0.49^{\circ}\text{C}$)など非侵襲推定の精度向上が報告されているものの(Limpabandhu, et al., 2022; Niedermann, et al., 2014; Olson, et al., 2023)、既存研究の多くは現在値の推定に着目しており、入浴によるリスク管理上重要な入浴後(例：入浴後 10–60 分)の温度予測は十分に検討されていない。

従来の機械学習(RF 及び決定木など)では二値化が便宜上用いられるが、生体の状態は本来連続的であり、二値化は情報損失や閾値近傍の不安定性を生む。さらに、入浴後の温度ダイナミクスは非線形・時変で個体差が大きく、従来の機械学習だけでは説明可能性の保持・非線形性の表現・説明可能性を同時に満たすことが難しい場面がある。したがって、本研究では、連続量の「程度」を保ちつつ非線形関係を表現し、IF-THEN ルールにより根拠(解釈性)を提示できる枠組みとしてファジー推論に着目する。

1.1.3 ファジー理論について

ファジー理論 (L.A. Zadeh, et al., 1965) は、曖昧な概念を 0~1 の“程度”で表す枠組みであり、ファジー集合(例：「高い水温」「長い入浴時間」)とメンバーシップ関数、「高い/低い・長い/短い」といった言語変数、そして「もし水温が高く入浴時間が長いなら 60 分後体温は上がりやすい」といった IF-THEN ルールを用いて推論する。メン

メンバーシップ関数の典型形状には三角形・台形・ガウス型などがあり、ガウス型は中心 μ (平均値)と幅 σ (標準偏差)で滑らかな境界(連続微分可能)を与え、入力に対する出力の説明可能性を担保する。形状・パラメータは「典型値(中心)」「許容幅(ばらつき)」を直観的に表し、専門家知識の可視的パラメータ化が可能となる。これにより、個々のルールが担う局所的な非線形関係を保ちながら、全体としては滑らかで連続的な写像が得られ、閾値付近の不安定性や二値化による情報損失を回避できる。さらに、メンバーシップ関数とルールは人間可読であり、どの入力領域でどの規則がどの程度効いたかを可視化できるため、モデル判断の説明可能性が担保される。結果として、入浴後体温のように非線形・時変・個体差の大きい現象に対しても、説明可能性(程度)・非線形性・説明可能性を同時に満たす推定が可能となる。

1.1.4 ファジー推論システム(FIS)について

FIS はファジー理論に基づく推論システムである。本研究では三つの FIS を使用した。ANFIS は神経ネットワークと FIS のハイブリッドで、局所非線形の精緻な表現と可視化に強みがある一方、初期化や学習率に敏感である。eTS はデータ流入に合わせてルールを逐次生成・更新でき、時変性や概念ドリフトに適応する反面、ノイズ由来のルール過剰生成問題がある。EVOFIS は進化計算によりルール構造とパラメータを広域探索し、複雑な相互作用でも高い表現力を示すが、計算コストとハイパーパラメータ依存への配慮が求められる。各手法はいずれも背景で指摘した課題に異なる角度から応えるものであり、詳細な実装は 1.3.4 予測モデルに記す(詳細は 1.3.4 参照)。

1.1.5 目的

以上を踏まえ、入浴後の体温を対象に、説明可能性(程度)・非線形性・説明可能の予測モデルを確立するために FIS の適用可能性を検証することが本章の狙いである。具体的には、ANFIS、eTS、EVOFIS という代表的バリエーションを取り上げ、従来法(RF、SVM)と比較して性能差を評価する。

1.2 方法

方法の全体像は図 1 に示す。この部分は(1) データの前処理、(2) 特徴選択、(3)モデルの学習と推論(4) 評価指標と統計解析、という流れで進む。

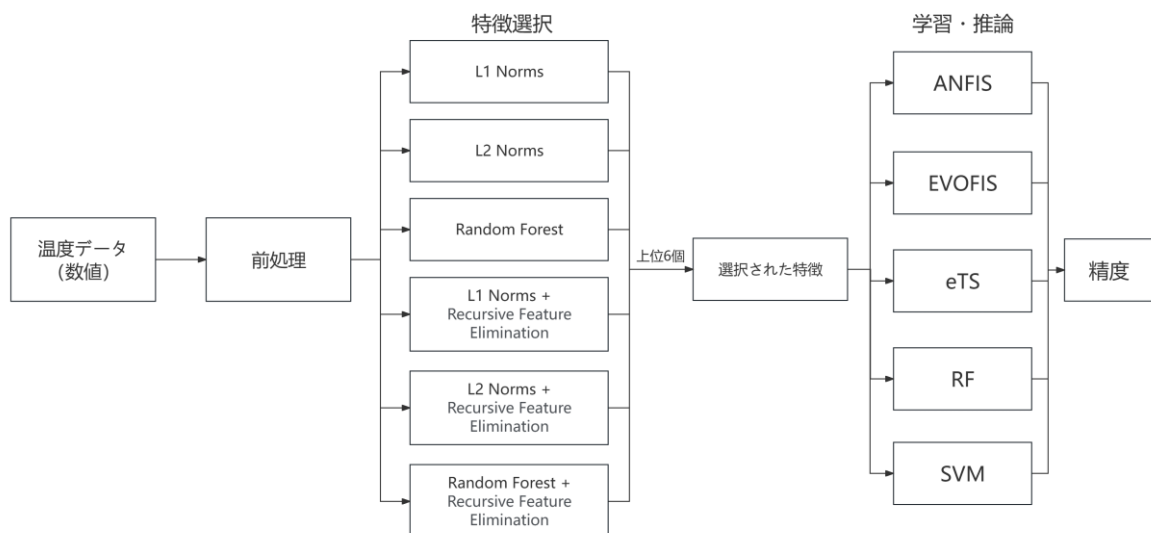


図 1.フローチャート

1.2.1 データセット

本研究では、1991～1997 年に株式会社バスクリンが実施した入浴介入試験に由来する構造化データを用いた。データベースは 215 の観測からなり、各観測は一回の入浴セッションに対応する。記録変数は 25 項目で、被験者属性(例：性別)、入浴条件(浴温 [°C]、入浴時間 [分])および体温計測系列を含む。

体温計測は入浴直前(前値)ならびに入浴後 10、20、30、60 分に実施した。計測部位は右手皮膚温、左手深部温、前額部皮膚温、左手非接触皮表温の 4 部位である。左手皮表温は非接触センサーにより取得され、いずれの温度も摂氏(°C)で記録された。

表 1.データ情報

入浴条件

浴温	41°C
入浴時間	5 分
性別	M/F

測定室条件

室温	25°C
湿度	50%

温度データ

浴後時間	左手深部温	左手非接皮 表温	右手皮膚温	前額部皮膚 温
前値	32.9	32.7	34	34.3
浴後 10 分	34	33.9	35.1	33.5
浴後 20 分	34.1	33.5	35	33.8
浴後 30 分	33.9	33.5	34.9	33.7
浴後 60 分	33.6	33.1	34.7	33.5

目的変数は各部位の 60 分後体温であり、説明変数として 前値、10/20/30 分後体

温、浴温、入浴時間 などを用いる。部位間の生理学的性質の差異(表在 vs. 深部、接触 vs. 非接触)を考慮し、解析は部位別に実施した。

本研究はヘルシンキ宣言の原則に則り、北海道大学病院生命・医学系研究院倫理審査委員会の承認(承認番号 024-0475)を得て実施した。すべての手順は関連する指針および規則に従って行った。データの利用にあたっては、インフォームド・コンセントに基づき、書面による同意を取得した。

1.3.2 前処理

本研究の前処理では、測定誤差や記録上の不整合を最小化するため、(1) 欠損値および外れ値を含むデータの削除、(2) 単位・尺度の統一(表記ゆれの正規化)を実施した。具体的には、正規表現を用いて記号・全半角・単位表記等のゆれを正規化し、データスキーマに準拠した変数名・カテゴリーの統一を行った。その後、欠損や外れ値を含むデータは一括削除した。元の 215 セッションから 2 セッション削除し、最終的な解析対象は 213 セッションである。

1.3.3 特徴選択

ノイズを最小化し、冗長性を減らし、計算効率を高めるため、25 の特徴から主要な予測特徴を 6 個に選抜した。採用した特徴選択法は、Lasso 正則化(L1 ノルム)、Ridge 正則化(L2 ノルム)、random forest(RF)に加え、これらを逐次後退法 RFE(recursive feature elimination)と組み合わせた L1+RFE/L2+RFE/RF+RFE である。

1.3.3.1 Lasso 正則化(L1)

Lasso は係数ベクトルの L1 ノルムを罰則として加えることで、不要な係数を 0 へと直接押し込み、スパースな解を与える。二乗誤差損失を用いる回帰の場合、目的関数は次式で表される。

$$\min_{w,b} \frac{1}{n} \sum_{k=1}^n (y^{(k)} - w^T x^{(k)} - b)^2 + \lambda |w|_1, \quad \text{where } |w|_1 = \sum_{j=1}^d |w_j|$$

ここで、 n は標本数、 d は特徴量の次元数、 $x^{(k)}$ は第 k 標本の特徴ベクトル、 $y^{(k)}$ は目的変数、 $w = (w_1, \dots, w_d)^T$ は回帰係数ベクトル、 b は切片、 $\lambda \geq 0$ は正則化強度、 $|w|_1$ は L1 ノルムである。係数の絶対値 $|w_j|$ を重要度として用いることで、特徴の自動選択が可能となる。

1.3.3.2 Ridge 正則化(L2)

Ridge は係数の L2 ノルム平方を罰則とし、係数の振れを縮小(shrinkage)して多重共

線性に頑健な推定を与える。目的関数は次式である。

$$\min_{w,b} \frac{1}{n} \sum_{k=1}^n (y^{(k)} - w^T x^{(k)} - b)^2 + \lambda |w|_2^2, \quad \text{where } |w|_2^2 = \sum_{j=1}^d w_j^2$$

ここで、記号の意味は L1 と同じであり、 $|w|_2^2$ を L2 ノルム平方とする。単独では係数が完全に 0 になりにくい、 $|w_j|$ に基づく順位付けや RFE と組み合わせることで確定的な選抜に利用できる。

1.3.3.3 Random forest(RF：特徴重要度の算出)

RF はブートストラップ標本から学習した多数の決定木をアンサンブルし、各分割でランダムにサブセット化した説明変数候補から MSE の減少が最大となる分割を選ぶ。回帰木の不純度(分割後の誤差)の減少量を全木・全ノードで加算した指標(mean decrease in impurity)を重要度として用いる。RF は非線形・高次の相互作用・スケール差に比較的頑健で、L1/L2 とは異なる視点から有用特徴を抽出できる(本研究では樹数 100、乱数種 42 を使用)。

1.3.3.4 逐次後退法(RFE)

基礎推定器(L1/L2/RF)を学習し、その重要度 π_j に基づいて最も寄与の小さい特徴を 1 つ除去し、再学習を繰り返すことで、所定数(本研究では 6)まで段階的に削減する手法である。RFE により、L2 のように単独では 0 化しにくい推定器でも明示的な選抜が可能になる。選抜はいずれも学習データ内で実施し、外部テストへの情報漏洩を避けた。

以上により、各部位ごとに 6 特徴を確定した後、後続のモデル学習へと進んだ。

1.3.4 モデルの学習と推論

本研究では、ファジー論理に基づく ANFIS、eTS、EVOFIS を実装し、RF および SVM を比較対象として学習・評価した。学習・検証は、特徴選択後の 180 件を学習、残余 33 件をテストとして実施した。説明可能性と学習ダイナミクスの解釈のため、ガウス型メンバーシップ関数の形状および収束曲線を取得・可視化した。

FIS 系：ANFIS では、各特徴量に 2 個のガウス型メンバーシップ関数(学習可能パラメータ μ, σ)を割り当て、初期値は $\mu=[0.0, 1.0]$, $\sigma=[1.0, 1.0]$ とした。損失関数は平均絶対誤差(L1 Loss)、最適化は Adam(学習率 0.005)を用い、1,000 エポック学習した。eTS では、学習率 0.1、初期重み 1000 とし、データ分布の変化に応じてルールおよびパラメータを逐次更新した。EVOFIS では、Optuna(Preferred Networks, Tokyo, Japan, v3.6.1)を用いて差分進化のハイパーパラメータを探索した。探索範囲は mutation rate 0.0–1.0、crossover probability 0.0–1.0、population size 30–60、generations 1500–2000 とした。

従来法：RF は決定木 100 本、random-state=42 で学習した。SVM(SVR, RBF カーネル)は scikit-learn 1.3.0 の GridSearchCV を用い、5-fold 交差検証で負の MSE を最小化するように $C \in \{0.1, 1, 10, 100\}$ 、 $\gamma \in \{1e-4, 1e-3, 1e-2, 1e-1, \text{scale}\}$ 、 $\epsilon \in \{0.1, 0.2, 0.5, 1\}$ のグリッドを探索し、最適組合せで最終モデルを学習してテストに適用した。

計算環境は Windows 11(23H2)/Python 3.11.4/12th Gen Intel Core i7-12700K(3.60 GHz)/RAM 16 GB/NVIDIA GeForce RTX 3060 Ti である。

1.3.4.1 Adaptive Neuro-Fuzzy Inference System(ANFIS)

本研究の ANFIS は TS 型 FIS を基盤とし、入力 $x \in R^d$ に対する出力を

$$\hat{y}(x) = \frac{\sum_{i=1}^R w_i f_i(x)}{\sum_{i=1}^R w_i},$$

$$w_i = \prod_{j=1}^d \mu_{A_{ij}}(x_j),$$

$$f_i(x) = a_i^T x + b_i$$

と定める。ここで、 x は入力ベクトル、 d は特徴量の次元数、 R はルール数、 w_i はルール i の発火度、 $f_i(x) = a_i^T x + b_i$ は結論部である。 a_i は、第 i ルールの線形係数ベクトル ($a_i = [a_{i1}, \dots, a_{id}]^T$) であり、各要素は入力ベクトル x_j に対応する係数を表す。 b_i は同ルールの切片である。なお、 a_i と b_i は学習により推定されるパラメータである。

各前件部はガウス型メンバーシップ関数

$$\mu_{A_{ij}}(x_j) = \exp\left(-\frac{(x_j - c_{ij})^2}{2\sigma_{ij}^2}\right)$$

で表し、各特徴につき 2 個(低/高)を割り当てた。ここで、 $\mu_{A_{ij}}(x_j)$ はルール i の特徴 j に対応するメンバーシップ、 A_{ij} はそのファジィ集合、 c_{ij} は平均値、 σ_{ij} は標準偏差である。

損失関数は L1(平均絶対値誤差)とし、

$$\mathcal{L} = \frac{1}{n} \sum_{k=1}^n |y^{(k)} - \hat{y}(x^{(k)})|$$

を最小化する。ここで、 n は標本数、 $(x^{(k)}, y^{(k)})$ は第 k 標本、 $\hat{y}(x^{(k)})$ は上記 ANFIS の出力である。

訓練設定は 1.3.4 に準ずる(損失関数=L1 [平均絶対値誤差]、最適化手法=Adam 学習率=0.005, epochs=1000)。再現性のため乱数種を固定した。学習時はメンバーシップ曲線と規則発火強度を可視化して挙動を確認した。

1.3.4.2 Evolving Takagi–Sugeno System(eTS)

ETS は逐次学習でルール集合を進化させる TS 型である。各入力 x に対し既存ルールの発火強度 w_i を計算し、新規性指標 $novelty(x)$ が閾値を超える場合は新規ルールを生成、そうでない場合は既存ルールの前件・後件を更新する。前件パラメータ(平均値 c_i , 標準偏差 σ_i)は, 例えば

$$c_i \leftarrow (1 - \eta) c_i + \eta x \quad (\eta = 0.1),$$

$$c_i \leftarrow updata(\sigma_i, x)$$

で更新し, 後件は逐次最小二乗で回帰係数を推定する。ルール刈り込み/統合は重複度や寄与度に基づき実施し, ノイズ由来の過剰生成を抑制する。ルール数・パラメータの発散を防ぐために正則化と上限制約を設ける。出力の集約は ANFIS と同じく

$$\hat{y}(x) = \frac{\sum_i w_i f_i(x)}{\sum_i w_i}$$

を用いる。初期状態では初期重み 1000 を付与した。

1.3.4.3 Evolving Fuzzy Inference System (EVOFIS)

EVOFIS は差分進化で FIS の構造(ルール数)とパラメータ(c, σ, a, b)を同時最適化する。個体 x は全ルールの前件・後件パラメータを連結したベクトルで表し, 適応度は検証 MSE とした。差分進化の突然変異(他個体の差分を足して試行解を作る操作)と交叉(試行解と親個体を成分ごとに混ぜて新個体を作る操作)は

$$v = x_{r_1} + F(x_{r_2} - x_{r_3}),$$

$$u_j = \begin{cases} v_j & \text{if } rand_j \leq CR \text{ or } j = j_{rand} \\ x_j & \text{otherwise} \end{cases}$$

に従い, 選択で適応度の高い個体を次世代へ残す。ここで, x はターゲット個体, $x_{r_1}, x_{r_2}, x_{r_3}$ は現在の母集団から相異なるインデックス r_1, r_2, r_3 を無作為抽出して得た個体, v は突然変異ベクトル(mutant), u は試行ベクトル(trial), F は変異スケール, CR は交叉率, $rand_j$ は成分 j ごとの一様乱数, j_{rand} は少なくとも 1 成分が v から継承されることを保証するためのランダムな成分インデックスである。訓練設定は 1.3.4 に準じて, Optuna v3.6.1 により $F \in [0,1]$, $CR \in [0,1]$, $pop \in [30,60]$, $gens \in [1500,2000]$ を探索し, 早期打ち切りと交差検証で過適合を抑制した。

1.3.4 評価指標と統計解析

本研究の主要評価指標は平均二乗誤差(MSE)と決定係数(R^2)とした。定義は以下の通りである。

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

ここで、 n は評価に用いた標本数、 y_i は第 i 標本の観測値、 $\hat{y}_i$ はその予測値、 $\bar{y}$ は観測値の平均値である。なお、 R^2 の分子 $\sum_{i=1}^n (y_i - \hat{y}_i)^2$ は残差平方和、分母 $\sum_{i=1}^n (y_i - \bar{y})^2$ は全変動である。

左手深部・左手非接皮表・右手皮膚・前額部皮膚の各部位ごとに、5モデル(ANFIS、eTS、EVOFIS、RF、SVM)でテストセットのMSEと R^2 を算出し、部位別の結果を提示するとともに、総合比較にはモデル別の平均MSEを用いた。多群同時比較はFriedman検定で実施し(帰無仮説：各モデルの中央値は等しい)、有意であれば両側のWilcoxon符号付順位検定によるペア比較を行い、Bonferroni補正で多重比較を制御した。有意水準はp値=0.05とした。統計解析ソフトウェアはSPSS(Statistical Package for the Social Sciences) v29.0.02を用いた。

1.4 結果

1.4.1 前処理の結果

本研究の前処理は、表記ゆれの正規化と、欠損値・外れ値を含む観測の listwise deletion を中核として実施した。記号・全半角・単位表記の統一後、欠損や生理学的に不自然な記録を削除し、最終的な解析対象は 213 セッションであった。単位はいずれも摂氏(°C)で整合した。以上により、後段の特徴選択・学習過程に影響しうる前処理起因のバイアスは抑制された。

異常例(前処理で対応したケース)

表 2.欠損値(例)：

	深部温 左手	非接皮表温 左手	皮膚温 A 右手	皮膚温 B 前額部
前値	欠損	36.9	32.1	33.2
10 分	欠損	33.6	33.9	33.6
20 分	欠損	33.3	34.2	33.6
30 分	欠損	33.2	34.0	33.5
60 分	欠損	33.5	34.0	34.0

表 3.異常値(例)：

	深部温 左手	非接皮表温 左手	皮膚温 A 右手	皮膚温 B 前額部
前値	31.6	28.3(異常)	28.8(異常)	33.6
10 分	33.6	33.4	33.6	33.9
20 分	33.5	33.5	34.2	33.9
30 分	33.3	33.3	34.0	33.9
60 分	33.1	32.3	34.0	34.0

1.4.2 特徴選択の結果

表 4 に各アルゴリズム別の選抜結果、表 5 に選抜回数の要約を示す。前値(入浴直前)と 20 分・30 分の温度は、多くの部位で反復して上位に選抜された。左手深部温で

は 10 分が上位に位置した。右手皮膚温では浴温および入浴時間が上位に位置した。
左手非接触皮表温では前値・20 分・30 分が上位に含まれた。

目的変数 60 分右手皮膚温												60 分左手深部温													
特徴選択																									
アルゴリ		L1		L1+RFE		L2		L2+RFE		RF		RF+RFE		L1		L1+RFE		L2		L2+RFE		RF		RF+RFE	
ズム																									
30 分右		30 分右		30 分右		前値左手		前値左手		前値左手		30 分左		30 分左		30 分左		30 分左		前値左手		前値左手		前値左手	
手皮膚温		手皮膚温		手皮膚温		深部温		深部温		深部温		手深部温		手深部温		手深部温		手深部温		深部温		深部温		深部温	
前値前額		30 分前		20 分右		前値右手		前値前額		前値前額		前値左手		20 分左		20 分左		20 分左		前値右手		前値前額		10 分左	
部皮膚温		額部皮膚温		手皮膚温		皮膚温		皮膚温		部皮膚温		非接皮表		深部温		手深部温		手深部温		手深部温		皮膚温		部皮膚温	
部皮膚温		温		温		部皮膚温		部皮膚温		部皮膚温		温		温		部皮膚温		部皮膚温		部皮膚温		部皮膚温		手深部温	
選択され																									
浴温		前値前額		前値右手		前値前額		20 分前		前値右手		20 分左		前値左手		前値前額		10 分左		前値前額		10 分前		額部皮膚温	
部皮膚温		部皮膚温		皮膚温		部皮膚温		額部皮膚温		皮膚温		手深部温		深部温		深部温		深部温		部皮膚温		部皮膚温		手深部温	
た特徴量																									
前値右手		前値左手		前値左手		30 分右		30 分左		20 分右		前値前額		前値前額		10 分左		20 分左		20 分左		20 分左		20 分左	
皮膚温		深部温		深部温		手皮膚温		手深部温		手皮膚温		部皮膚温		部皮膚温		手深部温		手深部温		手深部温		手深部温		手深部温	
前値左手		20 分前		30 分左		30 分右		30 分右		30 分左		30 分前		10 分前		30 分左		30 分左		30 分左		30 分左		30 分左	
深部温		額部皮膚温		手非接皮		浴温		手皮膚温		手非接皮		額部皮膚温		額部皮膚温		手非接皮		浴温		手皮膚温		手非接皮		手非接皮	
温		温		表温		温		表温		温		温		表温		温		表温		手皮膚温		表温		表温	
入浴時間		30 分左		前値左手		30 分前		30 分前		30 分右		10 分左		10 分左		10 分前		30 分左		30 分左		30 分左		20 分左	
手深部温		温		非接皮表		入浴時間		額部皮膚温		手皮膚温		手非接皮		手深部温		手深部温		手深部温		手深部温		手深部温		手非接皮	

表 4. 特徴選択の結果

目的変数 60 分左手非接皮表温													60 分前額部皮膚温																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																							
特徴選択																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
アルゴリ													アルゴリ																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																							
L1													L1+RFE													L2													L2+RFE													RF													RF+RFE													L1													L1+RFE													L2													L2+RFE													RF													RF+RFE																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					
ズム																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				

L1：L1に基づく選択；L1+RFE：L1と再帰的特徴削減（RFE）の併用；L2：L2に基づく選択；L2+RFE：L2とRFEの併用；RF：ランダムフォレストに基づく選択；RF+RFE：ランダムフォレストとRFEの併用

表 5. 特徴選択の選択回数

60 分右手皮膚温		60 分左手深部温		60 分左手非接皮表温		60 分前額部皮膚温	
特徴量	回数	特徴量	回数	特徴量	回数	特徴量	回数
30 分右手皮膚温	6	30 分左手深部温	6	30 分左手非接皮表温	6	30 分前額部皮膚温	6
前値左手深部温	6	前値左手深部温	6	前値左手非接皮表温	6	20 分前額部皮膚温	6
前値前額部皮膚温	4	20 分左手深部温	6	20 分左手非接皮表温	6	前値左手非接皮表温	5
前値右手皮膚温	4	前値前額部皮膚温	4	30 分右手皮膚温	4	前値左手深部温	5
浴温	2	30 分前額部皮膚温	4	前値右手皮膚温	4	入浴時間	2
入浴時間	2	10 分左手深部温	4	10 分左手非接皮表温	4	30 分左手非接皮表温	2
30 分前額部皮膚温	2	10 分前額部皮膚温	2	30 分前額部皮膚温	2	20 分左手非接皮表温	2
20 分前額部皮膚温	2	10 分左手非接皮表温	1	30 分左手深部温	2	10 分前額部皮膚温	2
30 分左手深部温	2	前値左手非接皮表温	1	20 分左手非接皮表温	1	前値前額部皮膚温	2
20 分右手皮膚温	2	前値右手皮膚温	1	30 分左手深部温	1	30 分左手深部温	1
30 分左手非接皮表温	2	20 分左手非接皮表温	1	10 分前額部皮膚温	1	20 分右手皮膚温	1
前値左手非接皮表温	2					30 分左手深部温	1
						20 分左手深部温	1

1.4.3 モデル学習の結果

学習ダイナミクス(図 2)では、ANFIS(図 2a・2c)が学習初期に損失を急速に低下させ、終盤に安定化するのに対し、EVOFIS(図 2b)は差分進化による世代更新に伴って適合度が緩やかかつ持続的に改善した。なお、eTS は学習率 0.1・初期重み 1000 の設定下で逐次更新によりルール集合を調整し、収束が確認された。

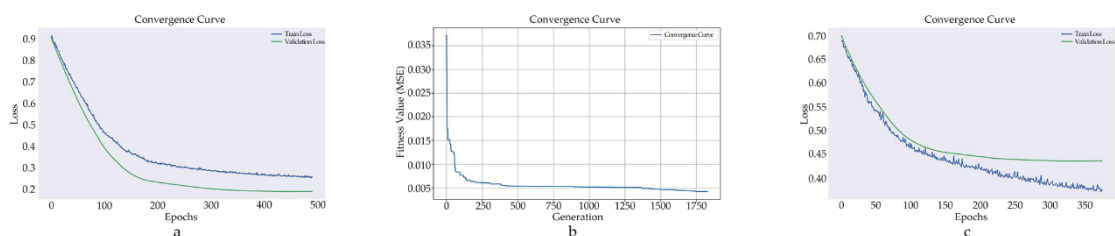


図 2. FIS モデルの収束曲線。左(a)および右(c)は、ANFIS で学習したモデルの学習損失・検証損失の曲線を示す。中央(b)は、EVOFIS モデルにおける世代を通じた適応度(または誤差)の減少を表す

1.4.4 予測結果

表 6 に各モデル×特徴選択法の MSE を示す。部位別の最良組合せは次のとおりである：右手皮膚温は ANFIS + RF+RFE(MSE=0.028°C)、左手深部温は EVOFIS + L2(MSE=0.038°C)、左手非接触皮表温は ANFIS + L2(MSE=0.091°C)、前額部皮膚温は SVM+L2(MSE=0.047°C)。図 3 の散布図(予測値 vs.実測値)でも、 $R^2=0.92/0.85/0.67/0.90$ と高い直線性が確認され、MSE の傾向と整合した(図 3a-d)。

表 6. 平均二乗誤差(MSE)の結果

60 分右手皮膚温							60 分左手深部温					
	L1	L1+R FE	L2	L2+R FE	RF	RF+ RFE	L1	L1+R FE	L2	L2+R FE	RF	RF+ RFE
ANFIS	0.040	0.036	0.040	0.061	0.035	0.028	0.058	0.051	0.070	0.057	0.069	0.051
eTS	0.041	0.040	0.046	0.036	0.036	0.036	0.070	0.082	0.074	0.048	0.082	0.082
EVOFIS	0.034	0.033	0.034	0.035	0.033	0.035	0.045	0.044	0.038	0.044	0.044	0.045
RF	0.036	0.036	0.042	0.040	0.035	0.035	0.045	0.047	0.055	0.057	0.059	0.060
SVM	0.037	0.036	0.038	0.039	0.037	0.037	0.043	0.043	0.048	0.044	0.049	0.049

60 分左手非接皮表温							60 分前額部皮膚温					
	L1	L1+R FE	L2	L2+R FE	RF	RF+ RFE	L1	L1+R FE	L2	L2+R FE	RF	RF+ RFE
ANFIS	0.099	0.101	0.091	0.099	0.111	0.106	0.060	0.056	0.069	0.060	0.091	0.156
eTS	0.109	0.106	0.109	0.109	0.117	0.117	0.063	0.063	0.053	0.052	0.069	0.064
EVOFIS	0.102	0.099	0.094	0.096	0.105	0.098	0.049	0.050	0.050	0.052	0.049	0.049
RF	0.120	0.119	0.118	0.119	0.132	0.132	0.073	0.074	0.070	0.068	0.064	0.069
SVM	0.102	0.107	0.102	0.102	0.098	0.098	0.052	0.052	0.047	0.049	0.049	0.049

L1=L1 に基づく選択；L1+RFE=L1 と再帰的特徴削減(RFE)の併用；L2=L2 に基づく選択；L2+RFE=L2 と RFE の併用；RF = ランダムフォレストに基づく選択；RF+RFE = ランダムフォレストと RFE の併用。ANFIS = Adaptive Neuro-Fuzzy Inference System；eTS = Evolving Takagi-Sugeno System；EVOFIS = Evolving Fuzzy Inference System；SVM = Support Vector Machine。

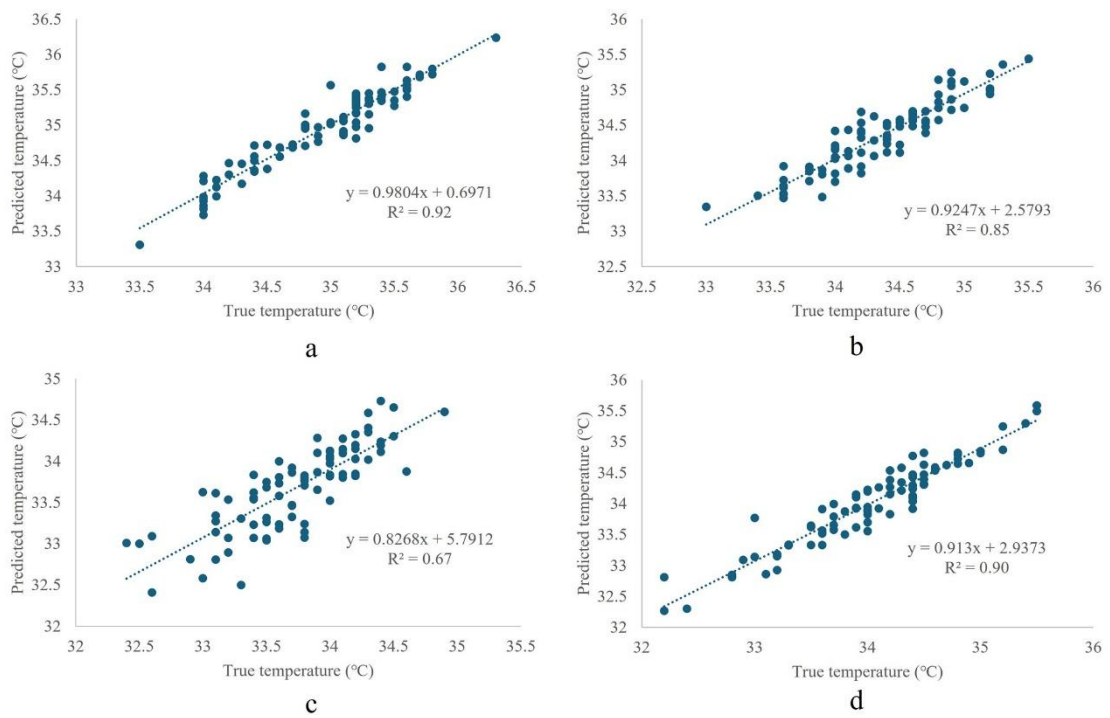


図 3. 予測温度と真値の比較

(a) 右手皮膚温：RF+RFE を用いた ANFIS による予測；(b) 左手深部温：L2 に基づく特徴選択を用いた EVOFIS による予測；
(c) 左手非接皮表温：L2 に基づく特徴選択を用いた ANFIS による予測；(d) 前額部皮膚温：L2 に基づく特徴選択を用いた SVM による予測。

1.4.5 可視化の結果

図4は、表6の太字で示した最高性能モデルと対応する特徴選択手法に基づくメンバーシップ関数の可視化である。ANFIS(図4a・4c)は各特徴軸に対して2つのガウス型メンバーシップ関数で入力空間を分割し、各グラフが個別の特徴量を表して明確な領域を形成する。一方、EVOFIS(図4b)はルール単位で全特徴を統合する全体的アプローチを取り、各グラフが全特徴間の相互作用を内包する単一のファジィルールに対応する。

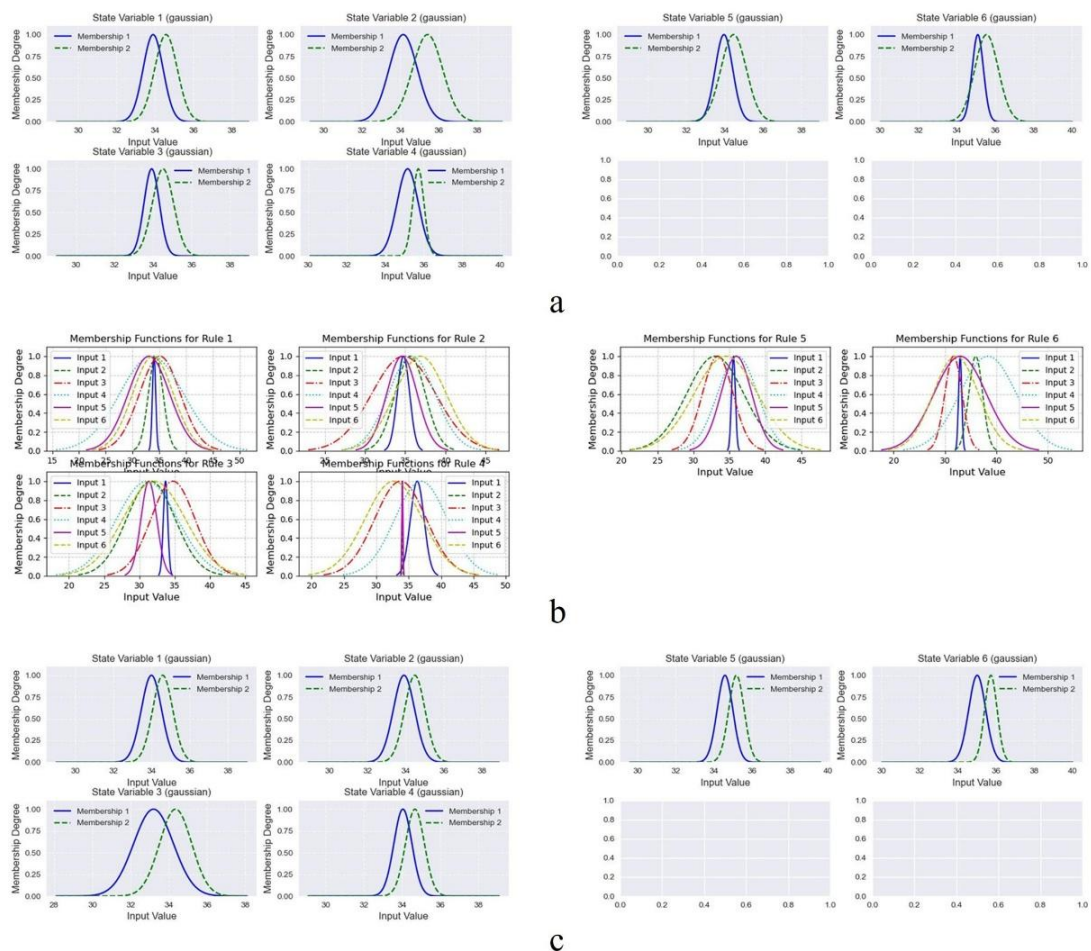


図4. FIS モデルにおけるメンバーシップ関数グラフ。

(a)右手皮膚温(RF+RFE+ANFIS)；(b)左手深部温(L2+EVOFIS)；(c)左手非接皮表温(L2+ANFIS)

1.4.6 評価・統計解析の結果

平均 MSE に基づく統計比較(図5)では、EVOFIS は ANFIS($p=0.001$)、eTS($p<0.001$)、RF($p<0.001$)より有意に低かった。SVM は eTS($p=0.001$)および RF($p=0.002$)より有意に低かったが、EVOFIS および ANFIS との差は有意ではなかった。

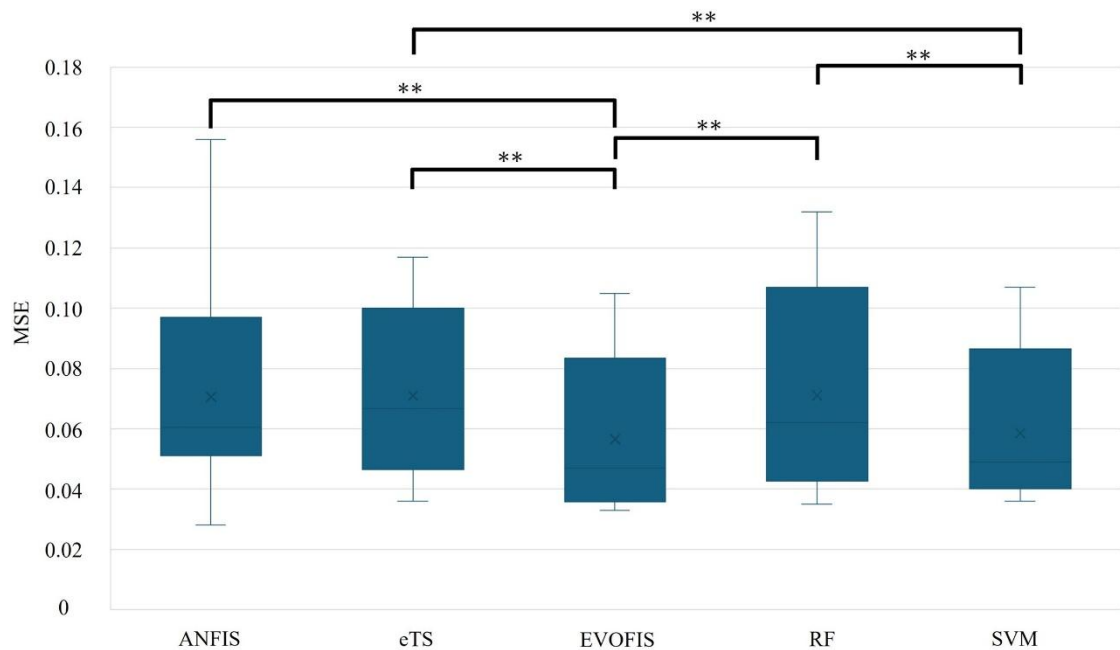


図 5. 予測モデル間の MSE 比較(ボックスプロット)

* $p < 0.05$, ** $p < 0.01$. なお「x」は平均値を示す。

1.5 考察

1.5.1 主要知見の解釈

本研究は、FIS が入浴後体温(60 分値)予測に有効であることを示した。とくに深部体温においては EVOFIS が最良の MSE(例：左手深部温 0.038°C)を達成し、非線形性と時系列依存の強い現象に対する適合性を示した(1.4.4、1.4.5)。一方、表在温(右手皮膚・前額部)では ANFIS や SVM が高い性能(例：右手皮膚 ANFIS+RF+RFE, $\text{MSE}=0.028^{\circ}\text{C}$; $R^2=0.92$ /前額部 SVM+L2, $\text{MSE}=0.047^{\circ}\text{C}$; $R^2=0.90$)を示した。統計比較では、EVOFIS は ANFIS/eTS/RF に対して有意に低い平均 MSE($p=0.001$, $p<0.001$, $p<0.001$)を示し、SVM は eTS/RF より有意に低い($p=0.001$, $p=0.002$)が、EVOFIS/ANFIS との差は有意ではなかった(1.4.5)。加えて、図 4 の可視化結果から、FIS が予測根拠を提示し可説明性を担保できることがより明確になった。ANFIS(図 4a・4c)は各入力軸を 2 本のガウス型メンバーシップ関数で連続的に分割し、ピーク位置は当該特徴の代表的な値域、幅は感度、重なりは境界の滑らかさを示す。この構造は、少数の変数の単調な影響が強い表在温の特性と整合し、右手皮膚などで良好な性能を示したことから噛み合う。一方、EVOFIS(図 4b)はルール単位で全入力を同時に可視化し、各パネルが「IF...AND...THEN...」という条件の組合せを表すため、多時点・多部位の相互作用を直接読み取れる。これは深部体温が複合要因と時系列依存で決まる機序と一致し、

最良誤差(例：左手深部温 $MSE=0.038^{\circ}C$)を得た結果と符合する。さらに、特徴選択で頻出した前値・20～30分といった時点は、図中でもピーク集中や帯域の狭さ(高感度)として現れ、「前値が高く20～30分も高いと60分後が上がりやすい」といった介入に直結する解釈的ルール抽出を可能にする。この結果からFISはメンバーシップ関数やルール図の可視化により、予測根拠の提示と説明可能性の確保が可能である。

部位別の最良組合せは、右手皮膚温がANFIS+RF+RFE、左手深部温がEVOFIS+L2、左手非接触皮表温がANFIS+L2、前額部皮膚温がSVM+L2であった(1.4.4)。特徴選択の結果(1.4.2)では、前値・20分・30分が反復して上位に入り、深部温では10分が上位に位置した。これらは、初期の再昇温から再冷却にかけての過渡応答の違いが60分値に影響し得ることを示唆する。

1.5.2 非接触測定の限界

左手非接触皮表温の精度($MSE \cdot R^2$)が相対的に低かった背景として、センサーの皮膚間距離・角度、室内対流・環境温などの影響が考えられる。改善策として、(i)測定プロトコルの標準化(距離・角度・計測高さ)、(ii)環境補正(室温・湿度の同時計測による補正項導入)の併用が有効である。

1.5.3 安全性・応用上の意義

入浴後の体温応答を事前に予測できることは、高齢者・循環器疾患・自律神経不安定といった人における安全管理に資する。予測された60分値とメンバーシップ曲線から、浴温・入浴時間の推奨レンジやクーリング介入(扇風機・水分補給・冷却時間)を、説明可能な根拠とともに提示できる。地域保健の文脈では、在宅や介護施設におけるリスク層別化、入浴指導のパーソナライズ、ウェアラブルによる支援への実装が想定される。

1.5.4 限界と今後の課題

入力特徴の時点：本データの制約から、10・20・30分など入浴後の中間時点を説明変数として用いた。実運用に向けては、入浴前のみで取得可能な特徴(例：年齢、体組成、既往、服薬、睡眠、屋外気象、浴剤・温泉成分、予定入浴時間・浴温)を統合し、事前時点での60分値予測を目指す。

特徴数の最適化は、解釈性の観点から6特徴に固定したが、最適数は依然として課題である。特徴量の安定性(データや分割が変わっても選ばれやすさ)と相関冗長の抑制(似た特徴のダブリを減らすこと)を両立しつつ、感度分析でロバスト性を評価する。

ハイパーパラメータと学習データ選択は設定・分割の影響を受け得るため、厳密なクロスバリデーションと早期打ち切り、反復実験の分位点報告を徹底する。

データセット規模は学習 180/テスト 33 の小規模単施設データであり、外部検証・前向き検証・施設間データセット差の評価を通じて汎化性を高める。

測定の信頼度では非接触計測における距離・角度やラベルノイズが課題であり、品質ログの一体運用や品質フラグによる重み付け学習が有効である。

1.5.5 位置づけと第二章への接続

本章は、構造化データに対する FIS の有効性と、部位別に異なる至適モデルの存在を示した。第二章では、電子カルテ由来のテキスト(SOAP 記述)を対象に、医学 NER により「病名・症状・身体所見・既往・検査・処方」等を部分構造化し、大規模言語モデル(LLM)へ入力して病名予測を行う。両章を通じて、データ形式(構造化 vs. 非構造化)の違いが予測性能・説明可能性・運用性に及ぼす影響を横断的に比較し、保健・医療実装における意思決定支援の有効性と限界を比較検討する。

第二章：日本語 EHR 呼吸器 SOAP の病名予測：多面的対照群による比較研究

2.1 緒言

2.1.1 現状と課題

医療現場では、診療録の中核をなす自由記述には診断・治療の根拠が埋め込まれている。EHR データの約 80% は非構造化のテキストで占められており (Au Yeung, J., et al., 2024)、症状・診断・病歴・社会的要因・患者の訴えなど、構造化データには含まれない豊富な情報が記録されており (Lopez, I., et al., 2025)、データ分析などの利活用が期待されている。特に、呼吸器系の疾患は世界的にも罹患率・死亡率の主要な要因となっており、毎年約 400 万人が呼吸器疾患で命を落とすと報告されており (Ferkol T, et al., 2014)、従って呼吸器に関する電子カルテの分析が意義深いと考えられる。

日本の電子カルテでは経過記録の多くが SOAP 形式で、見出しはあるが本文は自由記述の「見出し付き非構造テキスト」として蓄積される (中西, 2023)。このため、記載者・施設差、誤記や略語 (Yang, Z., et al., 2020)、日本語特有の無空白や全角・半角混在による表記揺れが自動解析を難しくする (Kawazoe Y., et al., 2021)。さらに既往や処方の方の再掲により冗長化が進み、外来記録の平均長は 10 年で約 60%、重複は約 10% 増加した (Adam Rule., et al., 2021)。重要情報の埋没とモデル性能低下を避けるには、表記正規化・重複削減・要約などの前処理と、その効果の定量評価が不可欠である。

代表的な前処理である臨床 NER は、疾患・薬剤・症状・検査・処置を抽出し語彙や ID へ写像して構造化する (Hua Xu., et al., 2010; Yada, S., et al., 2022; Kawazoe, Y., et al., 2021; Ohno, Y., et al., 2024)。一方、LLM は正規化やサブワード分割などの内蔵前処理を持ち、トークナイザーの選択は性能・効率に影響する (Iaroslav Chelombitko., et al., 2024)。このような LLM の内蔵前処理が便利であるが、前処理の妥当性が保証できなく、特に医療に対し、前処理の正確さが求められるため、検証が必要である。

推論面では MultiMedQA/Med-PaLM (Singhal, K., et al., 2023)、AMIE (Tu, T., et al., 2025) などの LLM モデルの性能が飛躍し、人手に匹敵する成果が報告されている。また、ChatGPT は特別な医療微調整なしでも米国医師国家試験全ステップで合格域付近に到達し、解答一貫性や洞察が定量化された (Kung TH., et al., 2021)。しかし、非構造化医療データの研究・モデル公開は英語中心で、日本語医療領域の適用例・比較研究・専用ベンチマークは不足している (Kawazoe, 2021; Junfeng Jiang, 2025)。また、医療デー

タでの追加学習や指示追従調整(prompt tuning)といった工夫が有効と考えられ(Tran, H., et al., 2024, Hirosawa T., et al., 2024)、prompt tuning の手法の検討も必要である。このため、日本語医療テキストに対する LLM の性能検証、prompt tuning の手法による性能検証、モデル間比較、適用可能性の体系的評価が必要である。

非構造化であるテキストを NER で構造化し、その出力を LLM に与える併用アプローチの効果は、同一データ・同一評価系での厳密比較がまだ不足している。LLM 単体と NER 併用の貢献差を客観化するには、追加検証が求められる。

2.1.5 目的

以上を踏まえ、本研究では、日本語で記載された呼吸器疾患領域の電子カルテ自由記述(SOAP)を対象に、NER によりテキストの構造化をまず行った。LLM 前処理の妥当性を評価するため、NER により構造化されたデータと構造化されていない元データをそれぞれ GPT-5・GPT-4.1・LLaMA 3.3 70B といった LLM に入力することで病名予測を実施した。さらに、モデル間比較、プロンプト設計(単段階 vs 二段階)、および初診と再診の違いに基づく情報冗長性や処理効率への影響を評価する。

2.2 方法

方法の全体像は図 6 に示す。この部分は(1)前処理、(2)アノテーション、(3)NER モデルファインチューニング、(4)LLM プロンプト設計と予測、(5)評価指標と統計解析、という流れで進む。

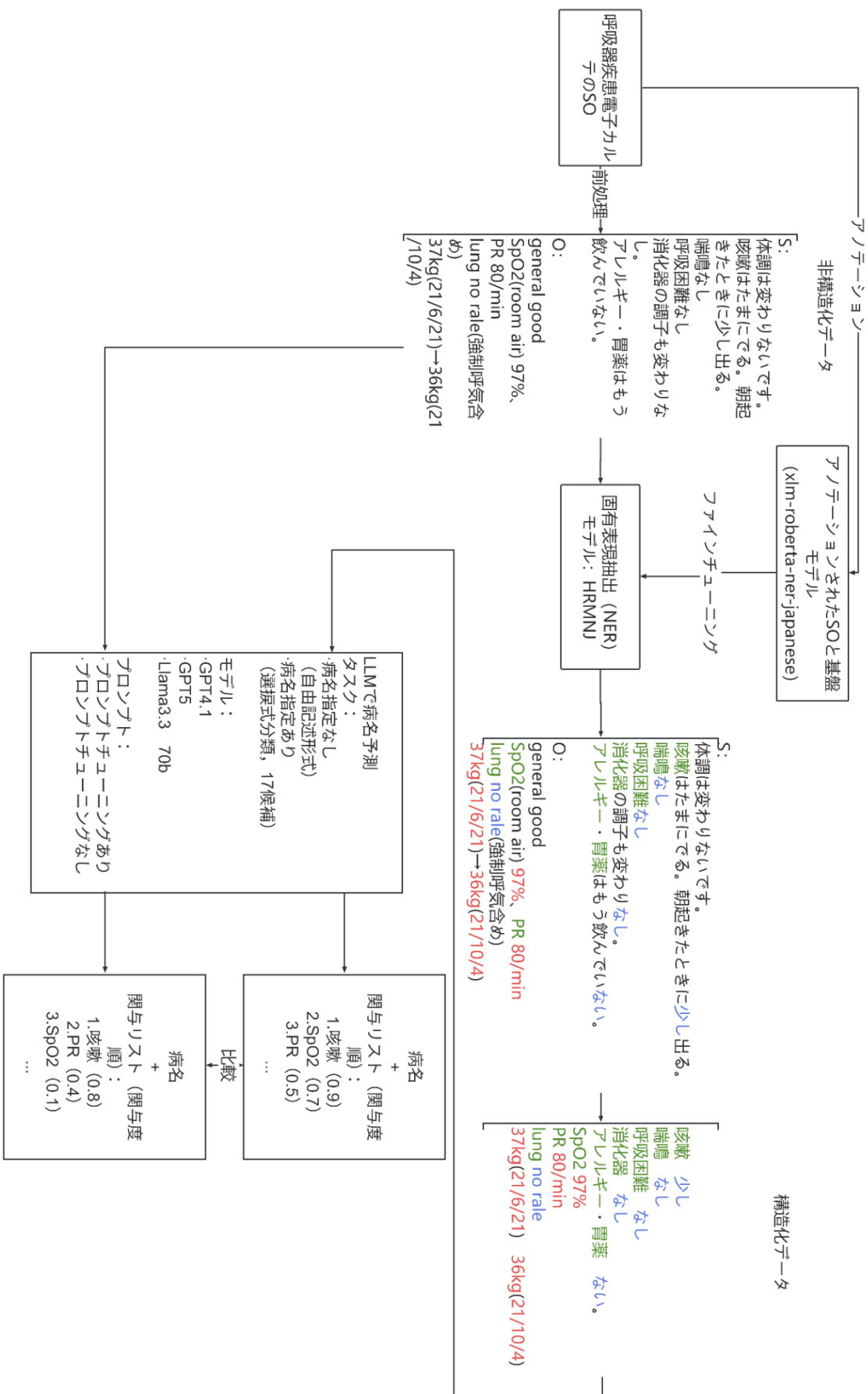


図 6.フローチャート

2.2.1 データセット

対象は北海道大学病院における呼吸器疾患の 7455 症例である。対象期間は 2008 年 8 月 12 日から 2024 年 2 月 24 日までである。先行研究(Cai, Tianrun, et al., 2022)のデータ量と手作業によるアノテーションの負担を考慮し、初診 200 例と非初診 200 例をランダムに選択し、約 400 例という規模のデータセットとし、対照比較に用いた。データセットは電子カルテの SOAP 形式テキストで、同一患者・同一受診日の S)と O)を統合し、患者番号・日付・自由記述本文から成るレコードを作成した(S/O の分離・同日マージ、本文結合を含む)。全体データセットのうち、①NER の学習・開発・評価用データセット、②LLM で下流の病名推定の比較実験用サブセットに分けて運用した。比較実験用サブセットでは、同一症例に対して二系統の入力を準備した。すなわち、自由記述そのままの S/O 非構造化全文と、注釈・正規化に基づく構造化要約データである。本章の LLM 実験では、両系統を完全に同一条件・パラメータ(モデル、temperature、出力スキーマ)で推論し、入力形態の効果を検証できるようにした。下流の病名推定タスクにおけるゴールドスタンダードとしては、呼吸器内科医による付与ラベルを用いて予測タスクを構成した。さらに、病名の上位概念評価のため、候補疾患を肺腫瘍、肺気腫、気管支喘息、肺線維症とアレルギー性肺炎の 5 つの ICD-10 大分類に整理し、その配下に 55 の小分類を整備した。これにより、自由予測タスクと階層的選択予測タスクの双方を同一データ上で評価可能とした。病名カテゴリーを表 7 に示す。患者 ID のみにアクセスし、患者氏名は記録していない。

本研究はヘルシンキ宣言の原則に則り、北海道大学病院生命・医学系研究院倫理審査委員会の承認（承認番号：024-0206）を得て実施した。すべての手順は関連する指針および規則に従って行った。インフォームド・コンセントはオプトアウトで行った。研究の詳細は北海道大学病院のホームページに掲載し、この手続きについても倫理審査委員会の承認を受け、すべての対象者に参加を辞退する機会を提供した。

表 7.病名カテゴリー(55 の小分類)

びまん性間質性肺炎	特発性間質性肺炎	慢性肺気腫	多発性肺腫瘍	通常型間質性肺炎
呼吸細気管支炎関連性間質性肺疾患	肺線維症の急性増悪	炎症後肺線維症	巨大気腫性肺のう胞感染	難治性気管支喘息
右肺腫瘍	アレルギー性肺炎	特発性間質性肺炎の急性増悪	気管支喘息	多発気腫性肺のう胞
運動誘発性喘息	咳喘息	肺腫瘍	特発性器質化肺炎	急性好酸球性肺炎

非特異性間質性肺炎	肺線維症	ブラ性肺気腫	特発性非特異性 間質性肺炎	気管支喘息合 併妊娠
多発肺腫瘍	遷延性肺好酸球 増加症	喘息性気管支喘息	若年性肺気腫	特発性肺線維 症
急性間質性肺炎	非特異性間質性 肺炎の増悪	気腫性肺嚢胞	リンパ球性間質 性肺炎	肺好酸球増加 症
気管支腺腫	好酸球増加性喘 息	重症気管支喘息	喘息性気管支炎	転移性肺腫瘍
特発性肺線維症の急 性増悪	慢性好酸球性肺 炎	非特異性間質性肺 炎の急性増悪	巨大気腫性肺の う胞	単純性肺好酸 球増加症
肺好酸球増加症の再 発	気腫合併肺線維 症	好酸球性気管支炎	閉塞性肺気腫	右肺中葉肺腫 瘍
肺気腫	気腫性肺のう胞	好酸球性肺炎	間質性肺線維症	難治性喘息

2.2.2 前処理

個人情報(氏名・住所・連絡先・ID・具体的日付等)は除去した。具体的に、アノテーション段階で氏名・住所・連絡先・IDなどの個人情報をそもそも注釈対象から除外する方針とし、注釈データ自体に個人情報が含まれないよう設計した。SOAP テキストの整形に特化し、同一患者・同一受診日に対応する S)と O)を連結して1本の文書としたうえで、日付は必要に応じて標準化した。表記上のノイズを減らすため、全角/半角・改行や余分な空白の削除、コロン等の記号の統一を施した。この整形後テキストを以降の情報抽出と推論に供した。

2.2.3 ラベル付与(アノテーション)

ラベルスキーマは先行研究(Mattupalli, S., et al., 2024)により設定し、Label Studio を用いてラベルし、呼吸器内科医の協力・指導のもと臨床記述の構造を反映したカテゴリーを基本とした。具体的なラベルを表8に示す注釈はガイドラインに基づきスパン単位で手動付与し、複合語(例：右上葉肺癌)の分割規則、略語展開、否定のスコップ境界、および時間・持続の適用範囲を明示した。完成したエンティティは標準語彙へ正規化して同義語を統合し、後段の検索性・再利用性を確保した。注釈データはJSON形式で保存し、注釈JSONにはテキスト上のスパン、カテゴリー、作業時間などのメタ情報を保持し、学習・評価・監査に利用できるよう設計した。

表8.設定されたラベル

症状・徴候	個人歴	検査値・検査結果
-------	-----	----------

症状の重症度	薬剤名、投与方法	数値
疾患・障害	投与量	単位
否定詞	頻度	部位
臨床イベント	診断・検査	画像所見
具体的な時間	持続時間	

2.2.4 NER モデル学習

NER モデルは xlm-roberta-base を用いた。ラベルは開始、内部、外部の表記方式で整備し、Label Studio 由来の注釈 JSON から文字単位のタグ列を構築した後、以下の手順でサブワード単位に整合させた。まず各文書のテキストを文字列配列としてトークナイザーに入力し、「元の文字インデックス → サブワードトークンインデックス」の対応を求めた。1つのサブワードに複数の文字ラベルが写像される場合は、最大値アグリゲーションにより代表ラベルを決定し、マッピング不能文字(空白等)は損失計算から除外した。これにより、文字単位の注釈をサブワード単位へ安定して移行できるようにした。データセットは文書単位で 9:1 にランダム分割した。トークン化後のデータは、計算効率を保つためバッチ単位で動的にパディングして長さを揃えた。学習設定は epochs=1, train batch size=12, eval batch size=12 である。評価指標は f1 score を採用し、評価用データに対してエンティティ単位の F1, precision, recall を算出した。エンティティ単位で集計した TP(true positive)、FP(false positive)、FN(false negative) に基づき、各指標を以下で定義する：

$$\text{Precision} = \frac{TP}{TP + FP}, \text{Recall} = \frac{TP}{TP + FN}, F1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Precision は「モデルが抽出したエンティティのうち、正解と一致していた割合」を表し、誤検出 (false positive) が少ないほど高くなる。Recall は「正解として存在するエンティティのうち、モデルが取り出せた割合」を表し、見逃し (false negative) が少ないほど高くなる。F1 は Precision と Recall の調和平均であり、両者のバランスを単一指標として評価するために用いる。

2.2.5 LLM タスク設計と予測

下流推論は、(a)自由予測タスクと(b)選択予測タスクの二つのタスクで構成した。

(a)自由予測タスクでは、病名候補を事前に制約せず、システム役割を「呼吸器内科の臨床医」と設定し、S/O に基づいて呼吸器疾患の鑑別診断候補(最大 3 件)を尤度順に出力し、各候補の診断根拠を併記させた。温度は 0.2 に固定し、二つの入力条件(自由記述テキスト/NER に基づく構造化サマリー)は同一設定で推論した。

(b)選択予測タスクでは、病名をデータセット全体から抽出された 55 の候補疾患リストに限定し、その中から最大 3 件を選択して根拠を出力させた。温度は 0.2 に固定し、自由記述テキスト/構造化サマリーの両入力条件を同一設定で推論した。

さらに両タスクに対して三つの対照比較を設けた。

- (i)初診群と非初診群(それぞれ該当する S/O を使用)、
- (ii)使用モデルの比較(GPT-5、LLaMA 3.3 70B、GPT-4.1)、
- (iii)プロンプト設計の比較(単段階 vs 二段階)。具体的なプロンプトは以下に示す。

- 自由予測タスクにおける単段階プロンプト(直接に病名予測)

あなたは呼吸器疾患の医者です。与えられた情報から、考えられる呼吸器疾患の病名を最大 3 つまで予測してください。

【カルテ情報】

{snippet}

- 二段階プロンプト(ICD を予測してから病名予測)

SYSTEM_PROMPT = (

"あなたは臨床経験が豊富な呼吸器内科医です。"

"必ず与えられた初診情報から、呼吸器領域を中心に鑑別診断を挙げます。"

"指定フォーマットで回答してください。"

)

USER_TEMPLATE = ""以下の初診 S/O から、呼吸器疾患のみの診断候補を最大 3 件、尤度の高い順に提案してください。

【カルテ情報】

{snippet}

出力形式(厳守):

病気名 所属 ICD1: <コード or N/A> <名称>(XX%)

理由: <主要根拠>

病気名 所属 ICD2: <コード or N/A> <名称>(XX%)

理由: <主要根拠>

病気名 所属 ICD3: <コード or N/A> <名称>(XX%)

理由: <主要根拠>

注意:

- 病名を予測し最大 3 つ、確率の高い順に列挙してください。
- まず病名を記載し、次にその病名に対応する ICD-10 コードと名称を記載。

- ICD-10 コードが不明な場合は <コード or N/A> の部分を N/A とする(名称は疾患群名で可)。
- %は推定確率、簡潔な根拠(症状・所見・既往・検査の示唆など)を1行で。

● 選択予測タスクにおける単段階プロンプト(直接に候補疾患リストから選択予測)

あなたは呼吸器疾患の医者です。与えられた情報から、病名を正確に分類してください。

あなたは呼吸器専門医です。以下の情報をもとに、考えられる呼吸器疾患の病名を次の選択肢の中から3つ選んでください。

【カルテ情報】

{snippet}

【選択肢】

- びまん性間質性肺炎
- アレルギー性肺炎
- ブラ性肺気腫
- リンパ球性間質性肺炎
- 単純性肺好酸球増加症
- 右肺中葉肺腫瘍
- 右肺腫瘍
- 呼吸細気管支炎関連性間質性肺疾患
- 咳喘息
- 喘息性気管支喘息
- 喘息性気管支炎
- 多発性肺腫瘍
- 多発気腫性肺のう胞
- 多発肺腫瘍
- 好酸球増加性喘息
- 好酸球性気管支炎
- 好酸球性肺炎
- 巨大気腫性肺のう胞
- 巨大気腫性肺のう胞感染
- 急性好酸球性肺炎
- 急性間質性肺炎
- 慢性好酸球性肺炎

- 慢性肺気腫
- 気管支喘息
- 気管支喘息合併妊娠
- 気管支腺腫
- 気腫合併肺線維症
- 気腫性肺のう胞
- 気腫性肺嚢胞
- 炎症後肺線維症
- 特発性器質化肺炎
- 特発性肺線維症
- 特発性肺線維症の急性増悪
- 特発性間質性肺炎
- 特発性間質性肺炎の急性増悪
- 特発性非特異性間質性肺炎
- 肺好酸球増加症
- 肺好酸球増加症の再発
- 肺気腫
- 肺線維症
- 肺線維症の急性増悪
- 肺腫瘍
- 若年性肺気腫
- 転移性肺腫瘍
- 通常型間質性肺炎
- 運動誘発性喘息
- 遷延性肺好酸球増加症
- 重症気管支喘息
- 閉塞性肺気腫
- 間質性肺線維症
- 難治性喘息
- 難治性気管支喘息
- 非特異性間質性肺炎
- 非特異性間質性肺炎の増悪
- 非特異性間質性肺炎の急性増悪

(複数該当する場合は、もっとも代表的なものを3つ選んでください。)

- 二段階プロンプト(ICD を選択予測し、候補疾患リストから選択予測)

#ICD カテゴリー

```
ICD_CLASSES = [
    ("D381", "肺腫瘍"),
    ("J439", "肺気腫(含ブラ/嚢胞)"),
```

```

("J459", "喘息"),
("J82", "好酸球性肺炎/好酸球関連"),
("J841", "びまん性間質性肺疾患(線維化系/IIP)",
]

# 55 の候補疾患リスト
DISEASE_CLASSES = [
    "びまん性間質性肺炎", "アレルギー性肺炎", "ブラ性肺気腫", "リンパ球性間質性肺炎",
    "単純性肺好酸球増加症",
    "右肺中葉肺腫瘍", "右肺腫瘍", "呼吸細気管支炎関連性間質性肺疾患", "咳喘息", "喘息性気管支喘息", "喘息性気管支炎",
    "多発性肺腫瘍", "多発気腫性肺のう胞", "多発肺腫瘍", "好酸球増加性喘息", "好酸球性気管支炎", "好酸球性肺炎",
    "巨大気腫性肺のう胞", "巨大気腫性肺のう胞感染", "急性好酸球性肺炎", "急性間質性肺炎", "慢性好酸球性肺炎",
    "慢性肺気腫", "気管支喘息", "気管支喘息合併妊娠", "気管支腺腫", "気腫合併肺線維症", "気腫性肺のう胞", "気腫性肺嚢胞",
    "炎症後肺線維症", "特発性器質化肺炎", "特発性肺線維症", "特発性肺線維症の急性増悪", "特発性間質性肺炎",
    "特発性間質性肺炎の急性増悪", "特発性非特異性間質性肺炎", "肺好酸球増加症", "肺好酸球増加症の再発", "肺気腫",
    "肺線維症", "肺線維症の急性増悪", "肺腫瘍", "若年性肺気腫", "転移性肺腫瘍", "通常型間質性肺炎", "運動誘発性喘息",
    "遷延性肺好酸球増加症", "重症気管支喘息", "閉塞性肺気腫", "間質性肺線維症", "難治性喘息", "難治性気管支喘息",
    "非特異性間質性肺炎", "非特異性間質性肺炎の増悪", "非特異性間質性肺炎の急性増悪",
]

# ICD と 55 の候補疾患リストの関係定義
DX_TO_ICD = {
    # —— D381 肺腫瘍 ——
    "肺腫瘍": "D381", "右肺腫瘍": "D381", "右肺中葉肺腫瘍": "D381", "多発肺腫瘍": "D381",
    "多発性肺腫瘍": "D381", "転移性肺腫瘍": "D381", "気管支腺腫": "D381",

    # —— J439 肺気腫/嚢胞 ——
    "肺気腫": "J439", "慢性肺気腫": "J439", "若年性肺気腫": "J439", "閉塞性肺気腫": "J439",
    "ブラ性肺気腫": "J439",
    "気腫性肺嚢胞": "J439", "気腫性肺のう胞": "J439", "多発気腫性肺のう胞": "J439",

```

"巨大気腫性肺のう胞":"J439","巨大気腫性肺のう胞感染":"J439","気腫合併肺線維症":"J439",

—— J459 喘息 ——

"気管支喘息":"J459","咳喘息":"J459","重症気管支喘息":"J459","難治性気管支喘息":"J459","難治性喘息":"J459",

"運動誘発性喘息":"J459","好酸球増加性喘息":"J459","気管支喘息合併妊娠":"J459","喘息性気管支喘息":"J459",

"喘息性気管支炎":"J459",

—— J82 好酸球関連 ——

"好酸球性肺炎":"J82","急性好酸球性肺炎":"J82","慢性好酸球性肺炎":"J82","好酸球性気管支炎":"J82",

"肺好酸球増加症":"J82","単純性肺好酸球増加症":"J82","遷延性肺好酸球増加症":"J82","肺好酸球増加症の再発":"J82",

—— J841 ILD/線維化 ——

"特発性間質性肺炎":"J841","特発性間質性肺炎の急性増悪":"J841","特発性肺線維症":"J841",

"特発性肺線維症の急性増悪":"J841","非特異性間質性肺炎":"J841","非特異性間質性肺炎の増悪":"J841",

"非特異性間質性肺炎の急性増悪":"J841","呼吸細気管支炎関連性間質性肺疾患":"J841","リンパ球性間質性肺炎":"J841",

"急性間質性肺炎":"J841","間質性肺線維症":"J841","炎症後肺線維症":"J841","肺線維症":"J841",

"肺線維症の急性増悪":"J841","びまん性間質性肺炎":"J841","通常型間質性肺炎":"J841",

"特発性器質化肺炎":"J841","アレルギー性肺炎":"J841",

}

prompt = f"""あなたは呼吸器専門医です。以下の初診 S/O から、まず 5 つの ICD-10 大分類の中から **最大 3 つ** を選び、つづいて **選んだ各 ICD に対応する小分類(下記の候補)** から **最大 3 つ** を選んでください。

(ICD と小分類は必ず対応させてください。小分類は、その ICD に属するもののみを選んでください)

【カルテ情報】

{input_text}

【ICD-10 大分類の候補(先にここから最大 3 つ選ぶ)】

{icd_choices}

【各 ICD に属する小分類(上の ICD を選んでから、この中から最大 3 つ選ぶ)】

{dx_choices}

出力形式(厳守):

ICD1: <コード> <名称>(XX%)

小分類:

- <小分類 1>(xx%) 理由: ~~~

- <小分類 2>(xx%) 理由: ~~~

- <小分類 3>(xx%) 理由: ~~~

ICD2: <コード> <名称>(XX%)

小分類:

- <小分類 1>(xx%) 理由: ~~~

- <小分類 2>(xx%) 理由: ~~~

- <小分類 3>(xx%) 理由: ~~~

ICD3: <コード> <名称>(XX%)

小分類:

- <小分類 1>(xx%) 理由: ~~~

- <小分類 2>(xx%) 理由: ~~~

- <小分類 3>(xx%) 理由: ~~~

注意:

- まず ICD(大分類)を最大 3 つ、確率の高い順に列挙してください。
- 各 ICD ごとに、その ICD 配下の小分類のみを最大 3 つ挙げてください。
- %は推定確率、簡潔な根拠(症状・所見・既往・検査の示唆など)を 1 行で。

すべての対照で温度 0.2、出力様式および候補集合を固定し、同一症例について二つの入力条件(自由記述テキスト/構造化サマリー)を厳密に対応させた。

2.2.6 評価指標と統計解析

本研究では、自由記述テキスト入力と構造化サマリー入力における病名予測性能を比較するため、以下の 2 つの指標を用いた。

- Top-1 正解率(Top-1 Accuracy)

モデルが出力した第 1 候補の病名が正解病名と一致した割合。

$$\text{TOP - 1 Accuracy} = \frac{N_{\text{correct,top1}}}{N_{\text{total}}}$$

ここで、 $N_{\text{correct, top1}}$ は Top-1 が正解した症例数、 N_{total} は全症例数である。

- Top-3 包含率(Top-3 Inclusion Rate)

正解病名がモデルの上位 3 候補のいずれかに含まれる割合。

$$\text{Top-3 Inclusion} = \frac{N_{\text{included,top3,top3}}}{N_{\text{total}}}$$

ここで、 $N_{\text{included, top3}}$ は正解病名が Top-3 候補に含まれた症例数である。

2.3 結果

2.3.1 ラベル付与(アノテーション)の結果

方法で設定したラベルに基づき、データにアノテーションを付与した。結果図を以下に示す。

症状・徴候	例：咳嗽	1	症状の重症度	例：ひどい	2	疾患・障害	例：肺炎	3	否定詞	例：なし	4	臨床イベント	例：手術	q
個人歴	例：喫煙歴あり	w	薬剤名	e	投与方法	例：経鼻	r	投与量	t	頻度	y	診断・検査	例：X線	a
検査値・検査結果	例：CRP	s	数値	例：38.5	d	単位	例：mg/dL	f	部位	例：肺	z	画像所見	例：結節影	x
具体的な時間 例：2024年3月 c 持続時間 例：2週間 v														

【主訴】 息切れ! 症状・徴候 例：咳嗽 **【現病歴】** 北海道健診センタークリニック(齋藤和雄Dr)より紹介。一週間前から 2: 持続時間 例：2週間 息切れ 3: 症状・徴候 例：咳嗽 を自覚。 2020/6/8 4: 具体的な時間 例：2024年3月 前医受診 5: 臨床イベント 例：手術。 左下肺野 6: 部位 例：肺 の肺炎像 7: 画像所見 例：結節影 指摘され、当科紹介初診 8: 臨床イベント 例：手術。 3週間前 9: 具体的な時間 例：2024年3月 に大丸の地下食品売り場に行った。 公共交通機関は使用していない 10: 個人歴 例：喫煙歴あり。

【既往歴】 54歳 11: 具体的な時間 例：2024年3月 右自然気胸 12: 疾患・障害 例：肺炎、2008年 13: 具体的な時間 例：2024年3月 頸椎症手術 14: 臨床イベント 例：手術 (当院整形外科)、高血圧症 15: 疾患・障害 例：肺炎 **【家族歴】** 父: 胃癌 (生活歴) 飲酒歴 (+) 16: 個人歴 例：喫煙歴あり; ビール 2本/日 17: 頻度 アルコール不耐 18: 個人歴 例：喫煙歴あり 19: 否定詞 例：なし 喫煙歴 (+) 20: 個人歴 例：喫煙歴あり; 20本/日×20年 21: 頻度 (20歳~40歳) 22: 持続時間 例：2週間、BI 23: 検査値・検査結果 例：CRP = 400 24: 数値 例：38.5 **アレルギー** 歴 25: 個人歴 例：喫煙歴あり (-) 26: 否定詞 例：なし 薬剤アレルギー 27: 個人歴 例：喫煙歴あり (-) 28: 否定詞 例：なし 吸入歴 29: 個人歴 例：喫煙歴あり (-) 30: 否定詞 例：なし

職業歴: 医師(元北大病院院長、形成外科名誉教授) **【内服薬】** ユニシア 31: 薬剤名、ヌイスリー 32: 薬剤名 **【全身状態】** 意識障害 33: 疾患・障害 例：肺炎 (-) 34: 否定詞 例：なし SpO2 35: 検査値・検査結果 例：CRP 91 36: 数値 例：38.5 % 37: 単位 例：mg/dL (RA) 38: 投与方法 例：経鼻、脈拍 39: 検査値・検査結果 例：CRP 120 40: 数値 例：38.5 回/分 41: 単位 例：mg/dL、体温 42: 検査値・検査結果 例：CRP 36.7 43: 数値 例：38.5 °C 44: 単位 例：mg/dL **【局所状態】** 45: 診断・検査 例：X線 心音異常 46: 疾患・障害 例：肺炎 (-) 47: 否定詞 例：なし、頻脈 49: 疾患・障害 例：肺炎 あり 呼吸音異常 (+) 48: 疾患・障害 例：肺炎 左下肺野 50: 部位 例：肺 呼吸音低下 51: 検査値・検査結果 例：CRP **【血液検査】** 52: 診断・検査 例：X線 20/06/08 53: 具体的な時間 例：2024年3月 13: 55 54: 具体的な時間 例：2024年3月 ----- **白血球数** 55: 検査値・検査結果 例：CRP (×1000/ μ l) 56: 単位 例：mg/dL) 11.8 57: 数値 例：38.5 H 58: 検査値・検査結果 例：CRP (×10⁶/ μ l) 60: 単位 例：mg/dL) 5.13 61: 数値 例：38.5 **ヘモグロビン** 62: 検査値・検査結果 例：CRP (g/dl) 63: 単位 例：mg/dL) 15.9 64: 数値 例：38.5 **ヘマトクリット** 65: 検査値・検査結果 例：CRP (%) 66: 単位 例：mg/dL) 46.8 67: 数値 例：38.5 **血小板** 68: 検査値・検査結果 例：CRP (×1000/ μ l) 69: 単位 例：mg/dL) 280 70: 数値 例：38.5 **neutro** 71: 検査値・検査結果 例：CRP (%) 72: 単位 例：mg/dL) 76.9 73: 数値 例：38.5 **lympho** 74: 検査値・検査結果 例：CRP (%) 75: 単位 例：mg/dL) 14.4 % 76: 数値 例：38.5 **L** 77: 検査値・検査結果 例：CRP (%) 78: 検査値・検査結果 例：CRP (%) 79: 単位 例：mg/dL) 6.8 80: 数値 例：38.5 **eosino** 81: 検査値・検査結果 例：CRP (%) 82: 単位 例：mg/dL) 1.4 83: 数値 例：38.5 **baso** 84: 検査値・検査結果 例：CRP (%) 85: 単位 例：mg/dL) 0.5 86: 数値 例：38.5 **尿中レジオネラ抗原** 87: 検査値・検査結果 例：CRP **尿中肺炎球菌抗原** 89: 検査値・検査結果 例：CRP **尿中** 90: 否定詞 例：なし **【画像検査】** 91: 診断・検査 例：X線 **胸部X線** 92: 診断・検査 例：X線 **左胸水貯留** 93: 画像所見 例：結節影、右中下肺野 94: 部位 例：肺 も 薄い陰影 95: 画像所見 例：結節影 認める

図 7. アノテーション結果

さらに、アノテーション済みデータから NER 学習用の JSON ファイルを生成した。
JSON ファイルの例を以下に示す。

```
{
  "id": 621827,
  "annotations": [
    {
      "id": 831,
      "completed_by": 1,
      "result": [
        {
          "value": {
            "start": 4,
            "end": 9,
            "text": "呼吸困難感",
            "labels": [
              "症状・徴候 例：咳嗽"
            ]
          },
          "id": "PsViIm-jMM",
          "from_name": "label",
          "to_name": "text",
          "type": "labels",
          "origin": "manual"
        },
        {
          "value": {
            "start": 15,
            "end": 20,
            "text": "2009 年",
            "labels": [
              "具体的な時間 例：2024 年 3 月"
            ]
          },
          "id": "sguXx7ySl_",
          "from_name": "label",
          "to_name": "text",
          "type": "labels",
          "origin": "manual"
        }
      ]
    }
  ]
}
```

2.3.2 NER モデル学習

全 11,722 件に対する評価指標は F1 スコア 0.687、Precision 0.681、Recall 0.694 であった。なお、構造化前の原文と、アノテーションに基づいて抽出・正規化した構造化結果の対応は表 9 に示す。

表 9.入力データと出力データの結果例

自由記述テキスト	構造化サマリー	
20 日ごろから鼻汁あり。その後咳もでてきた。今朝 KT37.1 度。軽い寒気はあり。肺炎球菌ワクチンは受けたが、インフルエンザのワクチンはうけていない。熱低くてもインフルな人がいるときいた。インフルエンザの検査希望。SpO2 98%、PR 92/min、。胸部聴診:ラ音聴取せず。インフルチェック:陰性	20 日 鼻汁 咳 今朝 KT 37.1 度 軽い 寒気 肺炎球菌ワクチン インフルエンザ ワクチン うけていない	SpO2 98 % PR 92 /Min 胸部聴診 ラ音 聴取せず インフルチェック 陰性

2.3.3 LLM タスク設計と予測

本節では、自由予測(自由記述テキストから鑑別候補を生成)と選択予測(55 の候補疾患リストから最大 3 件を選ぶ)の 2 タスクについて、入力形態(自由記述テキスト/NER に基づく構造化サマリー)、受診別(初診相当 n=200、非初診相当 n=200)、モデル(GPT-4.1/GPT-5/Llama)、プロンプト(tuning/no-tuning)を要因として集計し、二項比率の 2 群検定(両側、多重検定で FDR 補正)により差の有意性を評価した。各モデル(GPT-4.1, GPT-5, Llama)、データ種別(初診・非初診)、データ構造(非構造化・構造化)、プロンプトチューニング有無、タスク種別(自由予測・選択予測)の各条件における TOP1/TOP3 の正解数(正解/全体)および正解率を表 10 に示す。

表 10. 各条件別における TOP1/TOP3 の正解数(正解/全体)および正解率

task	非構造化 TOP1	構造化 TOP1	非構造化 TOP3	構造化 TOP3	非構造化 TOP1	構造化 TOP1	非構造化 TOP3	構造化 TOP3
------	--------------	-------------	--------------	-------------	--------------	-------------	--------------	-------------

	正解/全体	正解/全体	正解/全体	正解/全体	正解率	正解率	正解率	正解率
GPT4.1_初診データ prompt-tuning 自由予測タスク	68/200	72/200	105/200	102/200	34.00%	36.00%	52.50%	51.00%
GPT4.1_初診データ prompt-tuning 選択予測タスク	95/200	96/200	149/200	142/200	47.50%	48.00%	74.50%	71.00%
GPT4.1_初診データ no-prompt-tuning 自由予測タスク	68/200	71/200	98/200	102/200	34.00%	35.50%	49.00%	51.00%
GPT4.1_初診データ no-prompt-tuning 選択予測タスク	84/200	91/200	129/200	133/200	42.00%	45.50%	64.50%	66.50%
GPT4.1_非初診データ prompt-tuning 自由予測タスク	46/200	49/200	68/200	70/200	23.00%	24.50%	34.00%	35.00%
GPT4.1_非初診データ prompt-tuning 選択予測タスク	48/200	49/200	94/200	94/200	24.00%	24.50%	47.00%	47.00%
GPT4.1_非初診データ no-prompt-tuning 自由予測タスク	49/200	47/200	67/200	71/200	24.50%	23.50%	33.50%	35.50%
GPT4.1_非初診データ no-prompt-tuning 選択予測タスク	60/200	48/200	92/200	90/200	30.00%	24.00%	46.00%	45.00%
GPT5_初診データ prompt-tuning 自由予測タスク	70/200	77/200	97/200	107/200	35.00%	38.50%	48.50%	53.50%
GPT5_初診データ prompt-tuning 選択予測タスク	101/200	102/200	151/200	150/200	50.50%	51.00%	75.50%	75.00%
GPT5_初診データ no-prompt-tuning 自由予測タスク	72/200	76/200	96/200	103/200	36.00%	38.00%	48.00%	51.50%
GPT5_初診データ no-prompt-tuning 選択予測タスク	100/200	102/200	151/200	147/200	50.00%	51.00%	75.50%	73.50%
GPT5_非初診データ prompt-tuning 自由予測タスク	43/200	45/200	72/200	72/200	21.50%	22.50%	36.00%	36.00%
GPT5_非初診データ prompt-tuning	72/200	81/200	116/200	117/200	36.00%	40.50%	58.00%	58.50%

選択予測タスク								
GPT5_非初診データ								
no-prompt-tuning	40/200	44/200	58/200	62/200	20.00%	22.00%	29.00%	31.00%
自由予測タスク								
GPT5_非初診データ								
no-prompt-tuning	69/200	70/200	105/200	101/200	34.50%	35.00%	52.50%	50.50%
選択予測タスク								
Llama_初診データ								
prompt-tuning	44/200	42/200	97/200	95/200	22.00%	21.00%	48.50%	47.50%
自由予測タスク								
Llama_初診データ								
prompt-tuning	115/200	120/200	131/200	134/200	57.50%	60.00%	65.50%	67.00%
選択予測タスク								
Llama_初診データ								
no-prompt-tuning	57/200	48/200	104/200	93/200	28.50%	24.00%	52.00%	46.50%
自由予測タスク								
Llama_初診データ								
no-prompt-tuning	83/200	67/200	139/200	139/200	41.50%	33.50%	69.50%	69.50%
選択予測タスク								
Llama_非初診データ								
prompt-tuning	16/200	18/200	53/200	50/200	8.00%	9.00%	26.50%	25.00%
自由予測タスク								
Llama_非初診データ								
prompt-tuning	92/200	95/200	101/200	98/200	46.00%	47.50%	50.50%	49.00%
選択予測タスク								
Llama_非初診データ								
no-prompt-tuning	29/200	31/200	59/200	63/200	14.50%	15.50%	29.50%	31.50%
自由予測タスク								
Llama_非初診データ								
no-prompt-tuning	54/200	46/200	108/200	101/200	27.00%	23.00%	54.00%	50.50%
選択予測タスク								

まず、タスク効果は一貫して大きかった。非構造化入力の Top-1 は自由予測タスク 25.08%に対して選択予測タスク 40.54%(+15.46%、 $p<0.001$)、構造化入力の Top-1 は自由予測タスク 25.83%に対して選択予測タスク 40.29%(+14.46%、 $p<0.001$)でいずれも有意に選択予測が高かった。Top-3 でも非構造化 40.58%→61.08%(+20.50%、 $p<0.001$)、構造化 41.25%→60.25%(+19.00%、 $p<0.001$)と同様の傾向であった。

表 11. 選択予測タスクと自由予測タスクの比較

比較組	非構造化 TOP1	構造化 TOP1	非構造化 TOP3	構造化 TOP3
正解数(選択タスク)	973	967	1466	1446
正解数(自由予測タスク)	602	620	974	990
全体数				
(選択予測タスク/自由タスク)	2400	2400	2400	2400
正解率(選択予測タスク)	40.54%	40.29%	61.08%	60.25%
正解率(自由予測タスク)	25.08%	25.83%	40.58%	41.25%
差(%)	15.46	14.46	20.50	19.00
P 値	<0.001	<0.001	<0.001	<0.001

受診別の差も明瞭で、初診データは非初診データより有意に高かった。Top-1 では、自由予測タスク-非構造化：31.58% vs 18.58%(+13.00%、 $p<0.001$)、自由予測タスク-構造化：32.17% vs 19.50%(+12.67%、 $p<0.001$)、選択予測タスク-非構造化：48.17% vs 32.92%(+15.25%、 $p<0.001$)、選択予測タスク-構造化：48.17% vs 32.42%(+15.75%、 $p<0.001$)。

表 12. 初診データと非初診データの比較(TOP1)

比較組	自由予測タスク 非構造化 TOP1	自由予測タスク 構造化 TOP1	選択予測タスク 非構造化 TOP1	選択予測タスク 構造化 TOP1
正解数(初診データ)	379	386	578	578
正解数(非初診データ)	223	234	395	389
全体数(初診データ/非 初診データ)	1200	1200	1200	1200
正解率(初診データ)	31.58%	32.17%	48.17%	48.17%
正解率(非初診データ)	18.58%	19.50%	32.92%	32.42%
差(%)	13.00	12.67	15.25	15.75
P 値	<0.001	<0.001	<0.001	<0.001

入力形態の主効果は小さく、有意差は認めなかった。自由予測タスクでは構造化 25.83%に対し非構造化 25.08%(+0.75%、 $p=0.551$)、選択予測タスクでは構造化 40.29%に対し非構造化 40.54%(-0.25%、 $p=0.860$)。

表 13. 構造化と非構造化の比較(TOP1)

比較組	自由予測タスク TOP1	選択予測タスク TOP1
正解数(構造化)	620	967
正解数(非構造化)	602	973
全体数(構造化/非構造化)	2400	2400
正解率(構造化)	25.83%	40.29%
正解率(非構造化)	25.08%	40.54%
差(%)	0.75	-0.25
P 値	0.551	0.860

プロンプトの効果はタスク依存で、自由予測タスクでは差が非有意(非構造化：prompt-tuning 23.91% vs no-prompt-tuning 26.25%、 $p=0.187$ ；構造化：prompt-tuning 25.25% vs no-prompt-tuning 26.42%、 $p=0.514$)だった一方、選択予測タスクでは有意に prompt-tuning の方が高かった(非構造化：prompt-tuning 43.58% vs no-prompt-tuning 37.50%、+6.08%、 $p<0.01$ 、FDR=0.0024；構造化：prompt-tuning 45.25% vs no-prompt-tuning 35.33%、+9.92%、 $p<0.001$ 、FDR= 7.33×10^{-7})。

表 14. prompt-tuning と no-prompt-tuning の比較(TOP1)

比較組	自由予測タスク 非構造化	自由予測タスク 構造化	選択予測タスク 非構造化	選択予測タスク 構造化
正解数(prompt-tuning)	287	303	523	543
正解数(no-prompt-tuning)	315	317	450	424
全体数 (prompt/ no-prompt-tuning)	1200	1200	1200	1200
正解率(prompt-tuning)	23.91%	25.25%	43.58%	45.25%
正解率(no-prompt-tuning)	26.25%	26.42%	37.50%	35.33%
差(%)	-2.33	-1.16	6.08	9.92
P 値	0.187	0.514	<0.01	<0.001

モデル差について、GPT-5 と GPT-4.1 の差は小さかった。一方、選択予測タスク Top-1 では、GPT-5 が GPT-4.1 を有意に上回り(非構造化：42.75% vs 35.88%、 $p<0.01$ 、FDR=0.008；構造化：44.38% vs 35.50%、 $p<0.001$ 、FDR=0.0006)、Llama は GPT-4.1 より高い条件もあった(非構造化：43.00% vs 35.88%、 $p<0.01$ 、FDR=0.006；構造化：41.00% vs 35.50%、 $p=0.024$ 、FDR=0.034)。ただし GPT-5 と Llama の差は選択予測タスクでは非有意(非構造化： $p=0.920$ 、構造化： $p=0.172$)。

表 15. GPT5 と GPT4.1 の比較(TOP1)

比較組	選択予測タスク 非構造化 TOP1	選択予測タスク 構造化 TOP1
正解数(GPT5)	342	355
正解数(GPT4.1)	287	284
全体数(GPT5/GPT4.1)	800	800
正解率(GPT5)	42.75%	44.38%
正解率(GPT4.1)	35.88%	35.50%
差(%)	6.88	8.88
P 値	<0.01	<0.001

表 16. GPT5 と Llama の比較(TOP1)

比較組	選択予測タスク 非構造化 TOP1	選択予測タスク 構造化 TOP1
正解数(GPT5)	342	355
正解数(Llama)	344	328
全体数(GPT5/Llama)	800	800
正解率(GPT5)	42.75%	44.38%
正解率(Llama)	43.00%	41.00%
差(%)	-0.25	3.38
P 値	0.920	0.172

表 17. GPT4.1 と Llama の比較(TOP1)

比較組	選択予測タスク 非構造化 TOP1	選択予測タスク 構造化 TOP1
正解数(GPT4.1)	287	284
正解数(Llama)	344	328
全体数(GPT4.1/Llama)	800	800
正解率(GPT4.1)	35.88%	35.50%
正解率(Llama)	43.00%	41.00%
差(%)	-7.13	-5.50
P 値	<0.01	0.024

以上より、「タスク(選択予測の優位)」と「受診別(初診相当の優位)」は統計学的にも支持された一方、入力形態の主効果は有意でないこと、プロンプトとモデルの効果はタスクと条件に依存して現れることが確認された。

2.4 考察

選択予測タスクは自由予測タスクに対して Top-1 正解率で一貫して優位であった非構造化条件では 40.54%に対して 25.08%となり、差は 15.46%で、p 値は 0.001 未満である。構造化条件でも 40.29%に対して 25.83%となり、差は 14.46%で、p 値は 0.001 未満である。この差は、離散ラベル空間を明示的に提示することで探索空間が縮退し、モデルの確率質量が候補内に集中する「評価-出力形式の整合性」によって確率較正が働きやすくなるため、語彙選択や言い換えに起因する幻覚・近縁疾患への漂流を抑制できたことに由来すると解釈できる。とりわけ呼吸器領域のように近縁疾患が密集する多クラス設定では、候補制約と出力スキーマが推論の指揮線として作用し、自由生成の自由度がもたらすノイズを体系的に減じたと考えられる。

同様に Top-3 でも、選択予測タスクが自由予測タスクを大きく上回った。非構造化条件では 61.08%に対して 40.58%となり、差は 20.50%で、p 値は 0.001 未満である。構造化条件でも 60.25%に対して 41.25%となり、差は 19.00%で、p 値は 0.001 未満である。Top-1 に比した Top-3 の上昇幅が大きい事実は、モデルが「正解近傍の候補」を高順位に列挙できていることを示している。「Top-k を前提とした候補提示+根拠の添え書き」という運用設計が有効であることを裏づける。

受診別の結果では、初診データが非初診データよりも Top-1 正解率で明確に高かった。自由予測タスク・非構造化では 31.58%に対して 18.58%で、差は 13.00%、p 値は 0.001 未満である。自由予測タスク・構造化では 32.17%に対して 19.50%で、差は 12.67%、p 値は 0.001 未満である。選択予測タスク・非構造化では 48.17%に対して 32.92%で、差は 15.25%、p 値は 0.001 未満である。選択予測タスク・構造化では 48.17%に対して 32.42%で、差は 15.75%、p 値は 0.001 未満である。この差は、初診記録が主訴・現症・初期鑑別といった診断決定に直結する情報を高密度に含み、再診記録に少ないためである。

入力形態の主効果は小さく、Top-1 正解率の差はいずれも有意ではなかった。自由予測タスクでは、構造化が 25.83%、非構造化が 25.08%で、差は 0.75%にとどまり、 $p=0.551$ である。選択予測タスクでは、構造化が 40.29%、非構造化が 40.54%で、差は -0.25% とわずかで、 $p=0.860$ である。この結果は、概念正規化や関係抽出を併用した構造化の表現がテキストより高性能になるとする先行研究(Luo, Z., et al., 2011)の傾向と異なる。主因として、本研究で用いた NER が単純なエンティティ列挙にとどまり、原文の否定のスコップ、時相・時間範囲、確信度(アサーション)、および症状一部位といった関係情報を保持できていないため、診断決定に効く微細な手掛かりが失われた可能性が高いと考えられる。

プロンプトの効果はタスク依存であり、自由予測タスクでは差は有意ではなかった。

非構造化条件では 23.91%に対して 26.25%となり、p 値は 0.187 である。構造化条件では 25.25%に対して 26.42%となり、p 値は 0.514 である。これに対して選択タスク予測では、prompt tuning が有意に優れていた。非構造化条件では 43.58%に対して 37.50%となり、差は 6.08%で、p 値は 0.01 未満である。構造化条件では 45.25%に対して 35.33%となり、差は 9.92%で、p 値は 0.001 未満である。この違いは、選択予測タスクにおいて候補提示と段階的手順(大分類→小分類→根拠出力)をプロンプトで形式拘束でき、意思決定過程に直接介入できるのに対し、自由予測タスクでは巨大な生成空間にプロンプト差分が拡散して効果が希薄化するためと解釈できる。実運用では、評価指標と一致した出力スキーマをプロンプトに明記し、思考の外化を促すテンプレート化が効果的である。

モデル内比較の具体例としても、同一モデル条件で選択予測タスクが自由予測タスクを大幅に上回る傾向が再確認された。たとえば Llama では、初診で選択予測タスク 57.50%~67.00%に対し自由予測タスク 24.00%~52.00%、非初診で選択 23.00%~50.50%に対し自由予測タスク 8.00%~31.50%と、いずれも選択予測タスクが優越した。この結果は、モデルの規模や事前学習コーパスと無関係に、出力拘束と候補設計そのものが性能決定因子として強く作用することを示唆し、コストやレイテンシの制約下でもプロンプトと UI の工学的設計によって性能を底上げできる可能性を示している。

以上の結果パターン(選択予測タスク>自由予測タスク、初診>非初診、入力主効果は小、プロンプト効果は選択予測タスクで顕著)は相互に整合しており、制約の設計と記録の性質に適応した前処理・提示様式が臨床意思決定の安定化に直結することを示す。臨床導入に向けては、候補集合の標準化と段階選択、再診向けの時相・否定・確信度の明示化、リッチ構造化による文脈保持、そして評価指標と一致した出力スキーマの採用を優先課題とすべきである。これにより、正解近傍を高順位に集約する性質を Top-k 運用で最大化しつつ、誤りの方向性を臨床的に許容可能な近縁範囲へ制御できる土台が整う。

2.4.1 限界

今回のスキーマは概念抽出(NER)に主眼を置き、関係・イベント・タイムラインの表現が簡素であったため、自由記述テキストに含まれる微妙な臨床文脈(例：否定の作用域、症状→部位→程度→発現時刻の連結)が十分に LLM へ伝搬しなかった可能性がある。

また、単一施設・呼吸器内科の日本語 EHR に限定したため外的妥当性(一般化可能性)に限界がある。今後は施設・機器・記載慣行の異なる多施設データでの外部検証が必要である。

さらに、解析は各 200 例の初診・非初診サブセットを中心に行っており、条件別の試行数に偏りがあるうえ検出力が限定的である。症例レベルの対応を保持した前向きデータ収集と層別設計が望まれる。

2.4.2 今後の展望

今後は、否定・時相・程度のスコープを明示的にリンクとして保持する拡張スキーマ、イベント・タイムライン化、などを併記して LLM に与えるハイブリッド入出力に加え、ICD-10 /部位語彙を結節点にした知識グラフとの連携を探索する。これにより、“情報損失の少ない構造化”と“高い二次利用性”の両立を図る。

2.4.3 本章の位置づけと全体結論との接続

第一章が“連続量を扱う構造化データ×ファジィ推論”の適合性を示したのに対し、本章は“非構造化データ×LLM”の最適設計(選択型タスク、階層プロンプト、症例属性依存の運用)を実証した。データ形式の違いが、求められるアルゴリズム特性(精度・説明可能性・処理効率)と設計選好を規定するという、本研究の横断的主張を補強している。

全体の結論

1. 本研究から得られた新知見

● 研究全体

本研究は、構造化データと非構造化データという性質の異なる医療データに対し、データ形式に適合する AI 手法を選定・実装し、予測性能・説明可能性・実運用性の観点から比較検証した。結果として、データの形式やタスク設計が、最適なアルゴリズム選択と到達しうる精度に大きく影響することを示した。

● 第一章

入浴に伴う生理学的指標から入浴後体温を予測した。ファジィ推論系 (FIS) を中心に多手法を比較し、測定部位ごとの最良モデルを同定した。

- 右手皮膚温：ANFIS+RF+RFE が最高精度 ($MSE=0.028^{\circ}C$)。
- 左手深部温：EVOFIS+L2 正則化が最適 ($MSE=0.038^{\circ}C$)。
- 左手非接触皮表温：ANFIS+L2 正則化 ($MSE=0.091^{\circ}C$)。
- 前額部皮膚温：SVM+L2 正則化 ($MSE=0.047^{\circ}C$)。

いずれも予測値と実測値の線形性が高く (R^2 高値)、高精度を確認した。平均 MSE の群間比較では、EVOFIS は ANFIS・eTS・RF より有意に低誤差であり、SVM は eTS・RF より有意に低誤差である。ただし EVOFIS と ANFIS の差は有意ではない。

● 第二章

電子カルテの SOAP 記録に対して NER を実施し、抽出結果を LLM へ入力して診断予測を評価した。タスクは (1) 自由記述予測 と (2) 55 疾患からの選択予測 を比較した。

- 選択予測が一貫して優位で、Top-1 精度差は約 14~15% (いずれも $p < 0.001$)。
- 構造化入力 vs 非構造化入力の主効果は小さく、本データでは有意差なし。
- 受診形態による差を観察：初診が再診より有意に高精度 (約 $^{**}+12.67 \sim 15.75\%^{**}$, $p < 0.001$)。
- モデル依存性：条件によっては選択予測タスクの一部で Llama が GPT-4.1 を有意に上回る。

以上より、日本語医療テキストではタスク設計 (出力空間の制約) が精度に強く影響し、入力の事前構造化の効果は限定的である可能性が示唆された。

2. 新知見の意義

本研究で得られた知見は、個別化医療や臨床 AI の分野において重要な意義を持つ。第一章では、FIS が「程度」を扱うルールベース手法として温度応答の非線形変化を捉えられることを示した。これにより、湯温・入浴時間・入浴後のクーリングなどの介入で、個別化推奨の根拠をモデルのルールとして提示できる可能性がある。第二章の EHR テキスト分析では、候補診断カテゴリーを先に提示して探索空間を絞る選択式予測が有用であると確認した。入力 of 構造化は平均精度に大きな影響はなかったが、ノイズ低減、入力の短縮、不要な固有名詞の除去によるプライバシー強化など、運用上の利点が期待できる。

3. 本研究で得られた新知見から今後展開されうる研究

本研究の新知見に基づき、今後さらに幾つかの研究展開が考えられる。保健・在宅領域では、第一章の FIS による体温予測をウェアラブルや介護現場の安全管理ダッシュボードに実装し、湯温・入浴時間・クーリングなどの個別推奨を提示する実運用へ発展させる。臨床テキスト AI では、第二章で有効性を示した選択式予測を核に、疾患カテゴリー別に候補診断を提示する鑑別診断支援へ拡張する。さらに、症例属性やタスク特性に応じた prompt/モデル選択戦略を整備し、予測精度と説明可能性を高める。加えて、構造化データと非構造化データを統合する推論枠組みを備えた意思決定支援の構築も有望である。数値とテキストを併用することで、精度・説明性・処理効率の最適化を図り、臨床・在宅の双方で高度な支援につなげられる。

4. 今後の課題

本研究にはいくつかの限界と課題が残されている。規模が小さく特定条件のデータに依存するため、知見の一般化可能性を検証すべきである。今後は、環境・機器・記録法の異なる複数施設で外部／前向き検証を行い、施設間で異なるデータ特性の影響を体系的に評価する。第 1 章の構造化データ解析では、データセットと対象条件を拡充し、予測モデルの有効性を広く確認する必要がある。第 2 章の EHR テキスト分析では、単一診療科や日本語 EHR 特有の形式に依存しない検証設計とし、他領域データを加えて汎用的な診断支援を目指す。また、複数の専門特化モデルの統合運用も課題である。非構造化データ前処理は現状、関連語と否定語の抽出に留まり、症状・部位・程度・時間・否定の関係を扱えていない。今後はこれらの属性関係を抽出する工程を追加し、情報損失を抑えつつモデルの理解精度を高めたい。

謝辞

本稿を結ぶに当たり、終始懇切なる指導を賜りました指導教官の北海道大学大学院医学研究院放射線科学分野画像診断学教室 工藤與亮 教授に深い感謝の意を表します。

併せて、本研究全般に渡って直接的なご指導を賜りました北海道大学大学院医学研究院放射線科学分野画像診断学教室 唐明輝 特任講師に深い感謝の意を表します。

加えて、研究にご協力を頂いた、平田健司 准教授、小笠原克彦 教授、谷川原綾子 准教授、北井秀典 助教、佐藤夏季さんに深く感謝申し上げます。

また、データに関して多大なご協力下さいました、アース製薬株式会社 奥川洋司 様および松田悠之介 様に深く感謝いたします。

さらに、本論文の最終審査に際し、貴重なご指導ならびにご助言を賜りました古元重和 教授、朝倉聡 教授、セポソ・サークセス・テソロ 准教授に謹んで御礼申し上げます。

最後に、本論文は臨床データを使用させていただいた患者の皆様、当該分野内外の先生方、その他多くの方々のご指導とご協力に支えられて完成したものであり、研究に携わったすべての方々への心からの感謝の意を表します。

本研究(の一部)は、JST 次世代研究者挑戦的研究プログラム JPMJSP2119 の支援を受けたものです。

利益相反

開示すべき利益相反状態はない。

引用文献

Au Yeung J, Shek A, Searle T et al (2024) Natural language processing data services for healthcare providers. *BMC Medical Informatics and Decision Making* 24, 356.

Bender T, Karagülle Z, Bálint GP et al (2005) Hydrotherapy, balneotherapy, and spa treatment in pain management. *Rheumatology International* 25, 220–224.

Busch F, Hoffmann L, Rueger C et al (2025) Current applications and challenges in large language models for patient care: a systematic review. *Communications Medicine* 5, 26.

Cai T, He Z, Hong C, Zhang Y, Ho YL, Honerlaw J, Geva A, Panickan VA, King A, Gagnon DR, et al (2022) Scalable relevance ranking algorithm via semantic similarity assessment improves efficiency of medical chart review. *Journal of Biomedical Informatics* 132, 104109.

Chelombitko I, Safronov E and Komissarov A (2024) Qtok: A comprehensive framework for evaluating multilingual tokenizer quality in large language models. *arXiv* doi:10.48550/arXiv.2410.12989 (PREPRINT).

Durán JM and Jongsma KR (2021) Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI. *Journal of Medical Ethics* 47, 329–335.

Ferkol T and Schraufnagel D (2014) The global burden of respiratory disease. *Annals of the American Thoracic Society* 11, 404–406.

Fraile Navarro D, Coiera E, Hambly TW et al (2025) Expert evaluation of large language models for clinical dialogue summarization. *Scientific Reports* 15, 1195.

Gupta N, Singh H and Singla J (2022) Fuzzy logic-based systems for medical diagnosis – A review. pp. 1058–1062.

Hemelstrand S and Inoue T (2025) Stop splitting hairs: The problems with dichotomizing continuous data in language research. *Research Methods in Applied Linguistics* 4, 100272.

Hirosawa T et al (2024) Comparative analysis of diagnostic performance: differential diagnosis lists by LLaMA3 versus LLaMA2 for case reports. 8, e64844.

Jiang J, Huang J and Aizawa A (2025) JMedBench: A benchmark for evaluating Japanese biomedical large language models. In *Proceedings of the 31st International Conference on*

Computational Linguistics, 5918–5935.

Jung K, LePendur P, Iyer S, Bauer-Mehren A, Percha B and Shah NH (2015) Functional evaluation of out-of-the-box text-mining tools for data-mining tasks. *Journal of the American Medical Informatics Association* 22, 121–131.

Kawazoe Y, Shibata D, Shinohara E, Aramaki E and Ohe K (2021) A clinical specific BERT developed using a huge Japanese clinical text corpus. *PLOS ONE* 16, e0259763.

Kenny GP, Wilson TE, Flouris AD and Fujii N (2018) Heat exhaustion. *Handbook of Clinical Neurology* 157, 505–529.

Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV and Fotiadis DI (2015) Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal* 13, 8–17.

Kraljevic Z, et al (2023) Validating transformers for redaction of text from electronic health records in real-world healthcare. In *2023 IEEE 11th International Conference on Healthcare Informatics (ICHI)*, 544–549.

Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al (2023) Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digital Health* 2, e0000198.

Lample G, Ballesteros M, Subramanian S, Kawakami K and Dyer C (2016) Neural architectures for named entity recognition. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 260–270.

Liao KP, Cai T, Savova GK, Murphy SN, Karlson EW, Ananthakrishnan AN, Gainer VS, Shaw SY, Xia Z, Szolovits P, et al (2015) Development of phenotype algorithms using electronic medical records and incorporating natural language processing. *BMJ* 350, h1885.

Limpabandhu C, Hooper FSW, Li R and Tse Z (2022) Regression model for predicting core body temperature in infrared thermal mass screening. *IPEM-Translation* 3–4, 100006.

Lopez I, Swaminathan A, Vedula K et al (2025) Clinical entity augmented retrieval for clinical information extraction. *npj Digital Medicine* 8, 45.

Luo Z, Yetisgen-Yildiz M and Weng C (2011) Dynamic categorization of clinical research eligibility criteria by hierarchical clustering. *Journal of Biomedical Informatics* 44, 927–935.

Martinis MC, Zucco C and Cannataro M (2024) Negation detection in medical texts. In *Computational Science – ICCS 2024 (Lecture Notes in Computer Science, vol 14835)*. Springer, Cham.

Mattupalli S (2024) Medical-NER [Model card]. In Hugging Face, <https://huggingface.co/blaze999/Medical-NER> (2025 年 11 月 07 日参照).

Megat Ramli PN, Aizuddin AN, Ahmad N, Abdul Hamid Z and Ismail KI (2025) A systematic review: The role of artificial intelligence in lung cancer screening in detecting lung nodules on chest X-rays. *Diagnostics* 15, 246.

Moafa KMY, et al (2024) Artificial intelligence for improved health management: application, uses, opportunities, and challenges — a systematic review. *Egyptian Journal of Chemistry* 67, 865–880.

Niedermann R, Wyss E, Annaheim S, Psikuta A, Davey S and Rossi RM (2014) Prediction of human core body temperature using non-invasive measurement methods. *International Journal of Biometeorology* 58, 7–15.

Ohno Y, Kato R, Ishikawa H, Nishiyama T, Isawa M, Mochizuki M, Aramaki E and Aomori T (2024) Using the Natural Language Processing System Medical Named Entity Recognition-Japanese to analyze pharmaceutical care records: Natural language processing analysis. *JMIR Formative Research* 8, e55798.

Olson KD, O’Brien P, Lin AS, Fabry DA, Hanke S and Schroeder MJ (2023) A continuously worn dual temperature sensor system for accurate monitoring of core body temperature from the ear canal. *Sensors* 23, 7323.

Olusegun J, et al (2024) Integration of AI with electronic health records: enhancing clinical workflows.

Park Y, Son G and Rho M (2024) Biomedical flat and nested named entity recognition: methods, challenges, and advances. *Applied Sciences* 14, 9302.

Ross MK, Wei W and Ohno-Machado L (2014) “Big data” and the electronic health record.

Yearbook of Medical Informatics 9, 97–104.

Royston P, Altman DG and Sauerbrei W (2006) Dichotomizing continuous predictors in multiple regression: a bad idea. *Statistics in Medicine* 25, 127–141.

Rule A, Bedrick S, Chiang MF et al (2021) Length and redundancy of outpatient progress notes across a decade at an academic medical center. *JAMA Network Open* 4, e2115334.

Shimizu S, Nishiyama T, Nagai H, Wakamiya S and Aramaki E (2025) Toward cross-hospital deployment of natural language processing systems: Model development and validation of fine-tuned large language models for disease name recognition in Japanese. *JMIR Medical Informatics* 13, e76773.

Singhal K, Azizi S, Tu T et al (2023) Large language models encode clinical knowledge. *Nature* 620, 172–180.

Suzuki A and Negishi E (2024) Fuzzy logic systems for healthcare applications. *Journal of Biomedical and Sustainable Healthcare Applications* 4, 1–9.

Tran H, Yang Z, Yao Z and Yu H (2024) BioInstruct: instruction tuning of large language models for biomedical natural language processing. *Journal of the American Medical Informatics Association* 31, 1821–1832.

Tu T, et al (2024) Towards conversational diagnostic AI. arXiv doi:10.48550/arXiv.2401.05654 (PREPRINT).

Tu T, Schaekermann M, Palepu A et al (2025) Towards conversational diagnostic artificial intelligence. *Nature* 642, 442–450.

Xu H, Stenner SP, Doan S, Johnson KB, Waitman LR and Denny JC (2010) MedEx: A medication information extraction system for clinical narratives. *Journal of the American Medical Informatics Association* 17, 19–24.

Yada S, Nakamura Y, Wakamiya S and Aramaki E (2022) Real-MedNLP: Overview of REAL document-based MEDical Natural Language Processing Task SUBTASKS.

Yang Z, Dehmer M, Yli-Harja O et al (2020) Combining deep learning with token selection for patient phenotyping from electronic health records. *Scientific Reports* 10, 1432.

Zadeh LA (1965) Fuzzy sets. *Information and Control* 8, 338–353.

中西永子，高見美樹，石垣恭子(2022)ベテラン看護師と新人看護師の勤務前情報収集に関する電子カルテ利用による情報探索行動の差異—視線計測機を用いた注視エリアからとらえた情報収集時間，範囲，量の違い—．*医療情報学* 42, 249–262.