



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_abstract.pdf, 論文内容の要旨 https://creativecommons.org/licenses/by/4.0/



学位論文内容の要旨

博士の専攻分野の名称 博士（医 学） 氏 名 韓 豊

学 位 論 文 題 名

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)

【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 で推論まで一貫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、三つのファジー推論システム(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ー推論システム(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℃)、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℃、 R^2 =0.92、前額部：MSE=0.047℃、 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℃)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℃、前額部 SVM: MSE=0.047℃)。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。