第53回
日本Mテクノロジー学会大会
in 美し国みえ

大会テーマ：

Beyond Digital Health

Mでつなぐ医療・介護・地域の未来

～ 美し国から始まる実装と共創 ～

講演論文集

会期　：2025年10月10日（金）～11日（土）

会場　：三重大学講堂 三翠ホール
三重県津市栗真町屋町1577番地

大会長：川中 普晴　（三重大学大学院工学研究科）

主催：日本Mテクノロジー学会
共催：三重大学
後援：日本医療情報学会 課題研究会 FHIR研究会
日本医療情報学会 関西支部会
関西医療情報懇談会（KMI）

大会長 ご挨拶

Beyond Digital Health：
Mでつなぐ医療・介護・地域の未来
～ 美し国から始まる実装と共創

このたび、10月10日と11日の2日間にわたり、三重大学講堂 三翠ホールにて第53回大会を開催させていただくことになりました。第53回大会では、近年の急速な技術革新と社会的課題の変化を背景に、「Beyond Digital Health：Mでつなぐ医療・介護・地域の未来 ― 三重から始まる実装と共創」をテーマとして掲げました。

開催地となる三重県は、地域医療や包括的ケアにおける情報利活用の先進的な取り組みが進む地であり、「みえメディカルバレー構想」に象徴されるように、地域と情報が有機的に連携する実装環境の実現に向けたさまざまな取り組みが進められてきました。一方、Mテクノロジー学会では、MUMPSを中心として、長年にわたり医療・福祉分野における情報システムの実装や運用、さらにはFHIR等の標準化に関する知見を積み重ねてきました。これらの知見は、国内において電子カルテをはじめとする診療支援システムの根幹を担うとともに、多くの現場でその実績と信頼性が実証されています。

現在、医療情報システムは、FHIRに代表される国際標準の普及やAPI、クラウドによる柔軟な連携基盤の構築、さらには近年急速に発展を続けるAIやIoTを含む多領域的技術との融合によって、大きな転換期を迎えつつあります。このような中にあって、MUMPS技術の堅牢さと柔軟性、そして現場実装に即した思考は、むしろ今後の情報基盤設計において再評価されるべき価値を含んでいると考えられます。システムの設計と運用においても、単なる機能性の追求にとどまらず、User Experience や Human-Centric Design の重要性について改めて認識され始めており、「現場で使える」「現場とともに育つ」情報技術こそが、真に持続可能な社会基盤となることを、今こそ広く共有すべき時機にあると感じております。

このような激動する社会の中にあって、本大会では「地域に限定された事例共有」にとどまらず、「全国に共通する課題」――すなわち高齢化や地域医療資源の偏在、医療・介護連携の制度的課題、さらには災害時の情報流通など――に対して、いかに応えるべきかを改めて問い直し、広域的かつ長期的な視座から議論を深める機会を提供したいと考えております。また、全国の医療・福祉・介護・行政・産業・学術の多様な立場の方々の知見を交差させ、Mテクノロジーの今後の在り方を共に考える貴重な機会となることを、心より願っております。

三重は古くから豊かな自然に恵まれており、「御食国」と称された伊勢志摩地域を抱えるなど、「美し国（うましくに）」として多くの人の心を惹きつけ、文化の交差点としても重要な役割を果たしてきました。本大会が学術交流の場にとどまらず、人や文化の交流の場としても意義あるものとなれば幸いです。今年度は台風の影響なども考慮し、10月上旬での開催となりますが、「美し国」にて皆様とお会いできることを、心より楽しみにしております。

2025年10月10日

第53回大会・大会長
川中 普晴（三重大学大学院工学研究科）

第51回 日本Mテクノロジー大会 プログラム

大会1日目：10月10日（金）

9:00　開場・受付開始

9:30 ～ 10:00　大会企画１　座長：土井 俊祐（千葉大学病院 企画情報部）

IPCIでRAGを作る（チュートリアル）

鳥飼 幸太(群馬大学医学部附属病院システム統合センター)

10:00 ～ 10:10　休憩

10:10 11:40　大会企画２

IPCIでRAGを作る（ハンズオン）

鳥飼 幸太(群馬大学医学部附属病院システム統合センター)

11:40 ～ 12:50　昼食休憩

12:50 ～ 13:00　開会式

13:00 ～ 14:20　一般講演（１）　座長：竹村匡正（兵庫県立大学大学院情報科学研究科）

(1)-1　オンプレミス型生成AI環境の医療機関への展開と地域連携への応用

重冨 徳 (株式会社ナレッジピース)

(1)-2　バイオマーカーによる患者急変予知モデルの検討

下川 忠弘(兵庫県立大学大学院情報科学研究科／公益社団法人京都保健会

(1)-3　病名予測タスクにおけるオープンソースの生成AIの評価と特徴の分析

福岡 知隆(群馬大学医学部附属病院)

(1)-4　IPCIをRAG機能として応用したローカルサーバFHIRマッピング推測の高精度化

鳥飼 幸太(群馬大学医学部附属病院システム統合センター)

(1)-5　統計検定のすすめ～ＤＳ（データサイエンス）系について

本多 正幸(千葉大学病院・企画情報部)

14:20 ～ 14:30　休憩

14:30 ～ 15:15　学会長講演　座長：川中普晴（三重大学大学院工学研究科）

AI時代のMTAと市民開発

鈴木隆弘（千葉大学病院・企画情報部）

15:15 ～ 16:00　協賛講演　座長：鳥飼 幸太(群馬大学医学部附属病院システム統合センター)

MUMPSからIRIS、そしてHealthShareへ：海外の地域連携事例と日本への展望

林雅音・上中慎太郎（インターシステムズジャパン株式会社）

16:00 ～ 16:10　休憩

16:10 ～ 17:10　大会長特別企画　座長：川中普晴（三重大学大学院工学研究科）

①招待講演

三重大学と三重県内における医療DXの歩みと展望

佐久間 肇（三重大学病院・病院長，理事・副学長）

②パネルディスカッション

佐久間 肇（三重大学病院・病院長，理事・副学長）
鳥飼 幸太（群馬大学医学部附属病院システム統合センター）
竹村 匡正（兵庫県立大学大学院情報科学研究科）
土井 俊祐（千葉大学病院 企画情報部）

17:10 ～ 17:25 **論文誌編集委員会**

本多 正幸(日本Mテクノロジー学会理事　論文誌編集委員長)

18:00 ～ 20:00 **情報交換会**

会場：食道園（〒514-0009 三重県津市羽所町347　近鉄/JR 津駅より徒歩1分）
https://shokudouen.gorp.jp/，050-5493-1904

大会1日目：10月11日（土）

8:30 開場・受付開始

9:00 10:30 **一般講演(2)　　山ノ内 祥訓（熊本大学病院）**

(2)-1 起床時体重と就寝時体重の症例報告その後
山本 和子（元島根医療情報学医療情報学講座）

(2)-2 オンプレミス型大規模言語モデル(LLM)を用いた高齢者施設入所者のDBD13推
横谷 升美（兵庫県立大学大学院情報科学研究科）

(2)-3 オンプレミス環境下における複数LLMを用いたサマリ生成の比較
門野 勇介（兵庫県立大学大学院情報科学研究科）

(2)-4 大規模言語モデルの検索拡張生成における医薬品情報の活用と評価に関する研
世良 庄司（武蔵野大学薬学部）

(2)-5 仮想データを用いたLLMによるICFコード推定の検証
中谷貫太郎（兵庫県立大学大学院情報科学研究科）

(2)-6 大規模言語モデルを用いた家族看護実践の類型化に関する基礎的検討
松本賢典（兵庫県立大学大学院情報科学研究科）

10:30 10:45 休憩

10:45 11:45 **SDMコンソーシア　座長：本多 正幸(千葉大学病院・企画情報部)**

(S)-1 地域一体型仮想バイオバンクネットワークを志向したSDMデータウェアハウス
中村恵宣（神戸大学医学研究科）

(S)-2 PHR、EHRにおける共通データベースモデル
鈴木英夫（SDMコンソーシアム）

(S)-3 がまごおりデジタル健康プラットフォーム」事業におけるIRIS/Pythonを用いた
飯田 征昌（蒲郡市民病院デジタル医療推進室）

11:45 12:00 **閉会式**

12:00 13:00 **社員総会**

第51回 日本Mテクノロジー学会大会　運営組織

大会長　川中 普晴　（三重大学大学院工学研究科）

会場・運営担当　北島 巧海　（三重大学大学院工学研究科）
木村 倫人　（千葉大学病院 企画情報部）
竹村 匡正　（兵庫県立大学大学院情報科学研究科）
山ノ内 祥訓（熊本大学病院）

財務・協賛担当　土井 俊介　（千葉大学病院企画情報部）（ 学会事務局長）

学会企画担当　鳥飼 幸太　（群馬大学医学部附属病院 システム統合センター）
本多 正幸　（千葉大学病院 企画情報部）
鈴木 英夫　（SDMコンソーシアム）
林　雅音　（インターシステムズジャパン株式会社）

医療機関における生成 AI の活用

オンプレミス型生成 AI 環境の医療機関への展開と地域連携への応用

重冨　徳
株式会社ナレッジピース　執行役員　シニアアドバイザー
ヘルスケア、AI/IoT 担当

【キーワード】
オンプレミス型生成 AI 基盤
　― 医療機関内でデータを保持し、安全性と自律性を両立する生成 AI 環境。
RAG（Retrieval-Augmented Generation）
　― 医療知識データベースを参照し、根拠に基づいた高精度な AI 応答を生成する技術。
FHIR（Fast Healthcare Interoperability Resources）
　― 医療文書（退院サマリや同意書など）の標準化とシステム間連携を支える国際規格。
医療安全とハルシネーション防止
　― 誤生成を抑止し、根拠提示とリスク検出により安全な AI 活用を実現。
地域医療連携クラウド
　― 中核病院と中小病院・在宅医療機関をつなぐ、安全なデータ共有・要約連携の基盤。

1.【背景】 医療 DX と生成 AI の新たな関係

医療現場では、診療の複雑化と人手不足により、医師・看護師の業務負担が増加している。特に文書作成や記録管理など、非医療行為に費やされる時間が長く、医療の質や安全性の維持が課題となっている。

本研究では、こうした背景のもとで生成 AI を活用した医療業務効率化の実現と、オンプレミス型 AI 基盤と地域医療連携との高度化を目的とし、その導入構想と効果を予測する。

2.【目的】 生成 AI による医療業務効率化の実際

生成 AI は自然言語処理や文脈理解を通じ、医療従事者が日常的に行う文書作成業務を大幅に自動化する。退院サマリ、紹介状、看護記録、カンファレンス議事録の初稿作成などに適用し、作業時間を 60 分から 10 分に短縮（約 83%削減）が期待できる。

また、疾患名・薬剤・検査・処置の記載抜け自動検出、同意説明文や返書骨子の自動生成など、文書品質の均一化にも寄与する。

さらに 2024 年の医師の働き方改革に対応し、AI 支援による夜間・休日の初期対応や若手医師の教育を支援することで、人的リソースの制約を補完し、持続可能な医療提供体制の確立に貢献する。

3.生成 AI システム構成と運用アーキテクチャ

提案するシステムは、以下の 4 層構成で運用される。

- アプリケーション層：医師・看護師が操作する生成 AI ツール群（退院サマリ支援、画像診断補助等）
- オーケストレーション層：情報の流れとワークフローを統制する AI 制御基盤
- RAG データベース層：診療ガイドライン、過去症例、院内手順書などを格納

した知識ベース

- LLM 層：医療専門モデルによる文書生成・要約・推論処理

これらを FHIR 準拠で連携させることで、病院情報システムとの安全な相互接続を実現する。

特に中核病院では完全オンプレミス構成を採用し、中規模病院ではクラウド同期型ハイブリッド構成により、医療データ主権を保持しつつも効率的な運用を可能にする。

4.【成果】地域医療連携への展開と医療安全対策

本システムは、地域医療情報連携基盤と統合され、在宅医療・訪問看護ステーションを含む多機関連携を支援する。

退院サマリや同意書、アクセス記録は FHIR の DocumentReference 形式で統一管理し、生成 AI による自動要約とアクセス制御を行うことで、スムーズな地域連携を実現する。

さらに、医療 AI に求められる安全性確保のため、以下の機能を実装する。

- ハルシネーション防止：誤生成の抑制と内容検証機能
- エビデンス参照：文献・ガイドライン根拠の明示
- リスク判定：危険な診断・治療提案の自動検出・警告

これにより、汎用生成 AI では実現できない医療特化型の”精度・信頼性・説明可能性（Explainability）”を担保する。

5.【考察】導入効果と今後の展望

導入効果として、大規模病院では医師 1 人あたり 1 日 2〜3 時間の業務削減、中規模病院では診断時間の 30〜40％短縮を実現する。 また、薬剤有害事象は 40〜50％削減、稀少疾患の早期発見率は 15〜20％向上するなど、安全性と医療精度の両面で成果が得られる。

地域全体では、医療機器の共同利用や情報共有の促進により、コスト最適化・患者負担軽減・地域医療の質向上が見込まれる。

今後は、生成 AI 基盤の標準化と自治体・大学病院との実証実験などを通じて、医療 AI の社会実装をさらに加速させ、「高品質・効率的・安全な医療」の実現を目指す。

図 1

病名予測タスクにおけるオープンソースの生成 AI の評価と特徴の分析

Evaluation and Analysis of the Characteristics of Open-Source Generative AI for Disease-Name Prediction

福岡 知隆 [1)], 野口 怜 [1), 2)], 鳥飼幸太 [1)]
Tomotaka Fukuoka[1)], Rei Noguchi[2)], Kota Torikai[1)]
群馬大学医学部附属病院 システム統合センター[1)], 東京理科大学 [2)]
System Integration Center, Gunma University Hospital[1)]
Tokyo University of Science[2)]

キーワード:生成 AI、病名予測、オープンソース、退院サマリ

１．はじめに

近年、生成 AI は急速な発展を遂げており、問診の代替や、サマリの自動要約など医療分野においても様々なサービスの提供が開始されている。しかし、それらの多くはクラウドサーバ上に展開されている。これは生成 AI の高性能化に伴い、必要とする計算資源が増大したことに加え、AWS や Google Cloud などのプラットフォームの発展により、クラウド資源が利用しやすく、オンプレミス環境に比べてコストを抑えやすい点が理由として考えられる。利用者がクラウド上のサービスを利用するには最低限ネットワークに接続可能な端末を用意し、必要な情報を送信するだけとなり、負担も少ない。

サービスのクラウド化は開発者、利用者双方に有益な場合が多い。しかし、医療情報分野では診療情報などの個人情報を取り扱っており、セキュリティの観点から、クラウドサーバなどの医療機関外への患者カルテなどの送信を制限される場合がある。そのため、医療機関のローカルネットワーク内でのみでも利用可能な生成 AI を用いたサービスが必要とされる場合がある。

また、近年の医療分野における DX が推進され電子カルテや電子処方箋の導入などが行われている[1]。その中で、業務効率化生成 AI を含めた情報処理サービスの医療機関への導入も進められているが、多くの医療機関の経営状況は厳しく[2]、サービスの導入にかかるコストが課題の一つとなっている。導入コストの問題に対する一つ解決方法として、サービスやシステムを購入するのではなく、組織内でのサービス開発があげられる。

サービス開発においては、無償で公開されているオープンソースのツールを組み合わせて必要な機能を実装する場合があり、生成 AI においても無償で利用可能な生成 AI モデルが公開されている。生成 AI が必要とする CPU や GPU メモリなどの計算資源はモデルのアーキテクチャやパラメータ数などによって異なるが、小規模なモデルはノートパソコンでも動作可能であり、手軽に利用できる。また、近年はプログラミング言語によるコーディングを行わずに GUI ベースのシステム作成ツールも公開されており、Java、Python などのプログラミング言語を習得していない職員でもシステム作成が可能である。このように、近年はシステム開発に要求される人材や資源、ツールなどのハードルは低くなっている。

一方で医療分野における日本語を対象とした生成 AI の定量的な評価は十分されておらず、どの生成 AI を利用すればよいか判断する根拠は少ない。本稿では異なる生成 AI の定量的な評価のため、ローカルネットワーク内で利用可能な複数の生成 AI モデルを用いて患者情報を入力とした病名予測タスクを行い、評価比較した結果を報告する。

２．方法

本稿で用いる生成 AI モデルについて概説する。必要とする GPU メモリのサイズが 10GB 以下の小さいモデルとして Qwen-3 8B と Gemma-3n E4B を、同 10GB 以上のモデルとして gpt-oss-20b と gpt-oss-120b、Llama4 Scout (meta-llama/Llama-4-Scout-17B-16E)を用いる。また、日本語の医学情報を用いたモデルとして、医学教科書や論文など SIP プロジェクトで収集した 0.3T 汎用コーパスにより llm-jp-3-8x13b に対して事前学習を実施した SIP-jmed-llm (SIP-med-LLM/SIP-jmed-llm-2-8x13b-OP-instruct)を用いる。また、このモデルとの比較のため。Wikipedia や Dolma データセットなどの非医療情報コーパスにより llm-jp-3-8x13b に対して事前学習を実施した llm-jp-3-8x13b (llm-jp/llm-jp-3-8x13b-instruct3)を用いる。評価においては汎用的な性能評価のため、公開されているモデルデータをそのまま利用する。

評価データには群馬大学医学部附属病院における循環器内科などの Word 形式で記録された退院サマリ約 10 年分（17,065 件）を用いる。このデータでは主病名における頻度の上位が循環器系の疾患と腫瘍であり、本稿では頻度の多い循環器系の病名が主病名であるの心疾患（発作性心房細動/ 持続性心房細動/ 虚血性心疾患/ 急性心筋梗塞/ 慢性心不全/ うっ血性心不全/ 心房粗動）の上位 7 位、1,573 症例データを対象とする。各症例データの件数を表 1 に示す。

表 1　評価用データの主病名ごとの件数

主病名	件数
発作性心房細動	267
持続性心房細動	192
虚血性心疾患	194
急性心筋梗塞	171
慢性心不全	329
うっ血性心不全	305
心房粗動	115

評価データには前処理として以下の処理を実施している[3]。

まず、退院サマリに対して、患者情報の匿名化、全角と半角、小文字と大文字などの表記統一によるテキストクレンジングを実施する。次に 1 症例が 1 レコードとなるように構造化データに変換する。また、本データでは病名も自由テキストで記述されているため、正規表現により複数病名を 1 病名ずつに分割後、「万病辞書（ver.201907）」を用いて標準病名に変換し名寄せした。

本稿で評価する生成 AI はテキストベースで処理内容を指示するプロンプトを受け取ることで病名予測を行う。病名予測の根拠とするデータは前処理を行った退院サマリデータの中から患者主訴、既往歴、家族歴とし、生成 AI は一つのプロンプトにおいてこれらの情報から最も該当する可能性が高い病名を日本語で出力する。プロンプトの生成にあたっては正確な出力がなされるように Llama3.3 70B を用いて文面を調整した。また、本稿では生成 AI による病名予測のプロンプトには設定した 7 つの心疾患から一つ選択するという制約を与える。

３．結果

表 2 にそれぞれのモデルにおける病名予測結果の正解率を示す。

表 2　生成 AI ごとの病名予測正解率

モデル分類	モデル名	正解率
小型モデル (＜GPU10GB)	Qwen-3 8B	59%
	Gemma-3n E4B	57%
中型モデル (≧GPU10GB)	gpt-oss-20b	64%
	gpt-oss-120b	62%
	Llama4 Scout	63%
医学知識事前学習	SIP_jmed_llm	43%
非医学知識事前学習	llm-jp-3-8x13b	55%

この結果から、今回評価に用いた生成 AI による病名予測結果で最も予測正解率が高いモデルは gpt-oss-20b であることが示された。また、Llama4 Scout と gpt-oss-120 もそれぞれ正解率 63%、62%であり、近しい予測結果となった。

表 3～5 にこれらのモデルの病名ごとの予測結果を示す。適合率はモデルが予測した結果のうち、正しく病名を予測できた割合を示す。再現率はモデルが予測した結果は用意された症例数に対してどれだけ病名を正確に予測できたかの割合を示す。F 値は適合率と再現率の調和平均を示す。Llama4 Scout、gpt-oss-120b において、慢性心不全の予測結果の F 値が他の病名と比べて 39%以上低い結果が示された。一方で他の病名の予測結果を比較した結果、6 病名中、5 病名において gpt-oss-20b の結果よりも高い F 値を示している。本稿の評価データにおいて慢性心不全は最も件数の多い病名であり、この予測結果が低いことがモデル全体の病名予測の正解率を引き下げている。また、これらの結果からモデルには病名ごとの退院カルテのテキストにより、予測正解率に差があることが示された。

次に図 1～3 にそれぞれのモデルの結果の混同行列を示す。Llama4 Scout と gpt-oss-120b は慢性心不全をうっ血性心不全とする誤った予測をしがちである。この結果から類似した症例を含む病名予測では生成 AI モデルのパラメータ数が多いだけでは正確な予測が困難であると考えられる。

SIP-jmed-llm-2-8x13b と llm-jp-3-8x13b を比較した場合、SIP-jmed-llm-2-8x13b は 12%低い F 値を示している。医学論文などで再学習を行った

表 3　gpt-oss-20b の主病名ごとの予測結果

主病名	適合率	再現率	F 値
発作性心房細動	75%	66%	70%
持続性心房細動	66%	67%	67%
虚血性心疾患	75%	70%	72%
急性心筋梗塞	74%	80%	77%
慢性心不全	54%	51%	52%
うっ血性心不全	50%	61%	55%
心房粗動	76%	63%	69%

表 4　gpt-oss-120b の主病名ごとの予測結果

主病名	適合率	再現率	F 値
発作性心房細動	85%	57%	68%
持続性心房細動	57%	84%	68%
虚血性心疾患	81%	74%	77%
急性心筋梗塞	82%	82%	82%
慢性心不全	40%	12%	19%
うっ血性心不全	45%	81%	58%
心房粗動	80%	80%	80%

表 5　Llama4 Scout の主病名ごとの予測結果

主病名	適合率	再現率	F 値
発作性心房細動	81%	71%	76%
持続性心房細動	52%	83%	64%
虚血性心疾患	81%	79%	80%
急性心筋梗塞	88%	77%	82%
慢性心不全	76%	7%	12%
うっ血性心不全	46%	85%	59%
心房粗動	83%	70%	76%

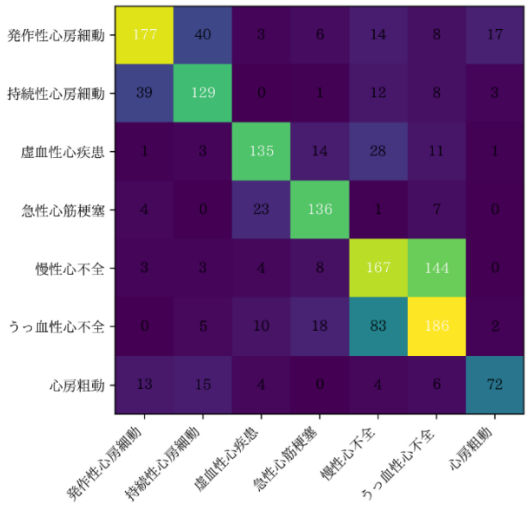


図 1　gpt-oss-20b による予測結果の混同行列

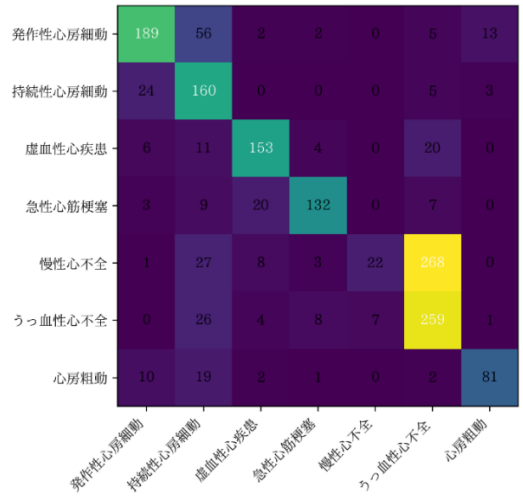


図 2　llama4 Scout による予測結果の混同行列

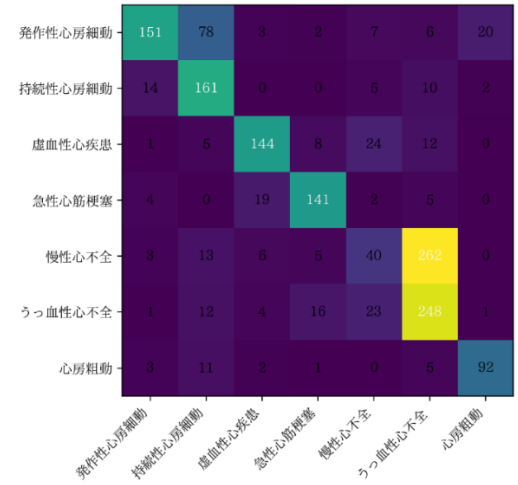


図 3　gpt-oss-120b による予測結果の混同行列

SIP-jmed-llm が低い結果を示した原因の一つとして日本語の指示に対する生成 AI の理解力不足が考えられる。生成 AI に予測の根拠も生成させたところ、SIP-jmed-llm の生成結果は一つの既往のみを根拠とする、すべての病名の予測根拠を同一とする場合などが多く、プロンプトを理解できていない可能性が示唆された。

Qwen-3 8B は評価したモデルで最も必要な GPU メモリサイズが小さいモデルだが、Gemma-3n E4B や llm-jp-3-8x13b より高い予測性能を示した。一方で、このモデルは出力生成に失敗する場合があった。指定した出力とは異なり英語で同じ解説文を繰り返し生成するため、対応が必要である。

Gemma-3n E4B モデルは Qwen-3 8B よりわずかに予測精度は低い結果であったが、英語で出力を生成することはなかった。一方で単語を間違える場合があったため、こちらも対応が必要である。

５．まとめ

本稿ではローカルネットワーク内で利用可能なオープンソースの生成 AI を用いた病名予測評価を実施し、生成 AI の予測性能における客観的な評価結果を示した。7 つのモデルを比較した結果、病名によって同一モデルでも結果が異なること、モデルのパラメータだけでは予測性能が決定されてないことを示した。今後はより客観的な評価指標を示すため、より多くの生成 AI モデルを用いるとともに、対象症例を心疾患に限定せずに評価を実施する。

参考文献

1. 厚生労働省，医療 DX について，[https://www.mhlw.go.jp/stf/iryoudx.html（cited 2025-09-30）]
2. 厚生労働省，医療機関等をとりまく状況（経営状況・人材確保等），[https://www.mhlw.go.jp/stf/iryoudx.html （cited 2025-09-30）]
3. 野口怜，鳥飼幸太，齋藤勇一郎，電子カルテデータを構造化した症例マトリクスによる疾患判別モデルの構築と表記揺れ集約の効果，第 24 回日本医療情報学会春季学術大会; pp. 252-253, WEB, 2020.

IPCI を RAG 機能として応用した
ローカルサーバ FHIR マッピング推測の高精度化

Improving the Accuracy of Local Server FHIR Mapping Inference
by Applying IPCI as a RAG Function

鳥飼 幸太 [1)], 福岡 和隆 [1)]
Kota Totikai[1)] ,Kazutaka Fukuoka[1)]
群馬大学医学部附属病院システム統合センター[1)]
Gunma University Hospital[1)]

Node-RED、LLM、RAG、HL7 FHIR, Database Mapping, IPCI

1．背景

生成 AI に分類される技術分野における Large Language Model(LLM)の性能は急速に向上しており、自然文での指示による多様な情報処理の可能性が提供されている。ハルシネーション問題についての対策として、確度の低い回答の際に無出力とする側を、何かを必ず出力する側よりも高くスコアリングするようプロンプト記述に追加する方法などが提言されている[1]。生成 AI のチャットボットとしての利用が進む一方、生成 AI のシステム機能としての利用に対して関心が高まっている。

2022 年より M テクノロジー学会から継続提供している医療 IT 向けプラットフォームである In-Process Clinical Intelligence(IPCI)は、生成 AI の急速な浸透以前に提唱され、主にオープンソースソフトウェアによって構成された仮想マシンであり、Agent 構成において、医療 DX を行うために実装されるソフトウェア機能を結合するための基盤として活用されることを目的としている[2]。結合機能を提供するプラットフォームは、一度導入してしまうと、複数のソフトウェアを結合していく性質上、個々のソフトウェア機能のような交換容易性が損なわれやすい。また、医療 IT は 10 年を超える期間で使用し続ける必要が生じることが多く、産業 IT の中でも成熟し安定した資源を基礎とすることが必須である。同様の理由により、中核となる機能が医療機関ごとに作成される場合、機能の保守性を担保することが不可欠である。これらの目的を達するために、IPCI は①仮想マシン/OS を交換可能要素として位置づける、②保守に必要な言語の種類を可能な限り削減する、③医療 IT における積年の課題（クラス肥大化による保守困難度の時間的増大、アジャイル性の確保）への対応などを構成時に満たすべき目標として、機能配置ならびにコード実装箇所（アーキテクチャ）が設計されている（図 1）。

IPCI は複数の企業からも高い関心をもって受け入れられており、利用の拡大が進んでいる。

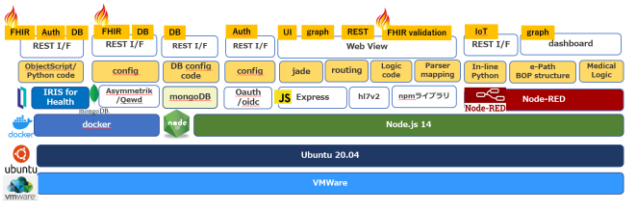


図 1　IPCI の機能実装

社会における LLM を含む生成 AI の活用は急激に進行しており、生成 AI を活用したサービスも進展している。日本 M テクノロジー学会では、主に医療情報学会を通じて IPCI チュートリアルならびに

OpenAI API や電子処方箋フォーマットの実装など、今日現在必要とされる機能の「実行可能なテンプレート」を提供してきた。FHIR の実装に関するチュートリアルでは、バージョンの変化に対応する方法を確立する必要性、ならびに LLM の精度では必ずしも FHIR のフォーマットが正確に行えない課題が認識されている。

２．目的

近年、最新の状況情報を取り込んだ LLM 活用の仕方として、LLM の前段にデータ検索検索を備えた Retrieval-Augmented Generation(RAG)機能が普及しつつあり、IPCI の FHIR サーバ等への活用に対してより有効である可能性がある。そこで本研究では、IPCI 内に RAG 機能を実装し、LLM 単体および LLM に RAG 機能を結合したモデルプログラムにおいて FHIR 変換フォーマットの作成の精度や充実度について比較することで、RAG 活用の有用性を調査することを目的とする。

３．方法

基礎となるプラットフォームとして、IPCI 内の Node-RED [3] を使用する。RAG 機能提供装置としてオープンソースソフトウェアである Dify [4] を用いる。Dify は Docket コンポーネントとして提供されているため、IPCI 内 Web サーバとして REST API による通信を行う。LLM は OpenAI API を選択した。LLM の利用では OpenAI API のライセンスキーを取得し REST API により通信を行う。

ソフトウェア機能の比較条件を以下のように設定した。LLM については A:OpenAI API(4o-mini)、B:Google LLM(gemma3.5)を比較対象とした。RAG については、参照サイトを HL7 FHIR の Observation が掲載されたページ(https://www.hl7.org/fhir/observation.html)、このページを pdf として取得したデータを Dify が備える機能コンポーネントを用いてベクトル化し、ナレッジベースを構成する。生成される FHIR フォーマットとして、A/B×1: (RAG 有/ 2:無)の 4 種を比較した。

４．結果

上述の方法により HL7 FHIR Observation フォーマットを生成し比較した結果を表 1 に示す。過去のある時点までの情報を元に学習されており、また医療情報以外の分野にも学習がされている LLM では、医療情報規格として利用可能な質を担保することがいずれも困難であったが、RAG を導入することによって、情報選択の精度が向上した。

表 1 LLM 種類/RAG 有無による FHIR Observation フォーマットの比較

	4o-mini	gemma3.5
有	◎	○
無	△	▲

５．考察

検証結果は LLM の質が必ずしも十分でない場合でも RAG 機能により選択/抽出操作をすべき情報の精度を向上させることで実用性を担保する FHIR マッピングが実現する可能性を示唆している。汎用 LLM の利点は使用者が自然文を用いてプログラミングできる点であり、AI Agent としてのインターフェースとして汎用 LLM は不可欠である。RAG 機能は汎用 LLM の短所を補完する機能として有用であり、RAG に相当するナレッジセットの高精度化により、医療情報のような精度ならびに再現性の高いフォーマット出力機能についても実用レベルで実装可能であることが強く期待される。

参考文献

1. Adam Tauman Kalai, Ofir Nachum, Santosh S. Vempala, Edwin Zhang, Why Language Models Hallucinate, Computer Science, 2005 https://arxiv.org/abs/2509.04664
2. 鳥飼幸太、IPCI（In-Process Clinical Intelligence）のコンセプト、第 50 回日本 M テクノロジー学会年会抄録
3. https://nodered.org/
4. https://dify.ai/

統計検定のすすめ～ＤＳ（データサイエンス）系について

The Recommendation of "Statistical Tests" – Focusing on the series of Data Science

本多正幸

Masayuki Honda)

長崎大学病院／千葉大学病院

Nagasaki University Hospital/ Chiba University Hospital

キーワード:データサイエンティスト、統計検定、データサイエンス

１．はじめに

データサイエンティストが備えておくべき知識や技術、ノウハウは、データベース関連、ネットワーク関連、プログラム（ソフトウエア）関連など多岐にわたると思われる。データサイエンティストとして、自分自身の力量を判定し、スキルアップへの啓発をする手段の一つとして、日本統計学会が公式に認定している全国統一試験である「統計検定」がある。

「統計検定」では、１級（ＰＢＴ）とＣＢＴがあり、ＣＢＴは、準１級から９種類の試験が準備されている。本報告では、ＭＴＡ大会参加者にとって関心が高いであろう「データサイエンスＤＳ（基礎・発展・エキスパート)」について、その概要や試験内容を紹介し、ＤＳのスキルアップへの啓発を行いたい。

本報告では、「統計検定」全体について概観した上で、ＤＳ系について焦点を絞る。

２．統計検定の勧め

統計学を学び自身の能力を図る仕組みとして、統計学会が主催する「統計に関する知識や活用力を評価する」全国統一試験（https://www.toukei-kentei.jp/　）である。対象は、中高生、大学生、職業人を対象としている。図１は、種類と内容を示している。図２は、紙の試験は１級のみでその他の級はすべて CBT であることを示している。

近年、データサイエンスの重要性が叫ばれ、いくつかの大学にもデータサイエンスに関する学部などが創設されている。DX、ＡＩなどに関心か高まる中、図１に示されているが「統計検定」にデータサイエンス・初級、発展、エキスパートの３つの級が設置されている。初級のテキストを拝読すると、中身は EXCEL を活用した実例に基づく分析であり、そこに統計学の知識が融合された問題となっている。

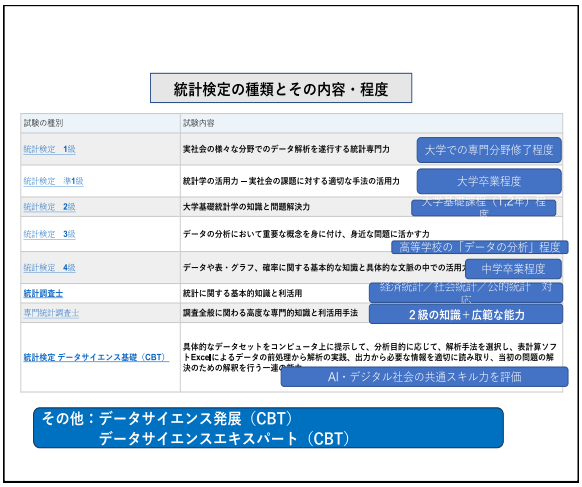


図１

図２

３．ＤＳ（データサイエンス）系について

ＤＳ（基礎）に焦点を当てて、ＨＰからの抜粋を紹介する。ＤＳ（発展）ＤＳエキスパートについては「統計検定」ＨＰを参照してください。

３.１　出題要素（Ｄ Ｓ基礎）

評価するキーコンピテンシーをデータアナリティクス基礎とし、（1）データハンドリング技能、（2）データ解析技能、（3）解析結果の適切な解釈　の３つの観点が出題となっている。（図３）

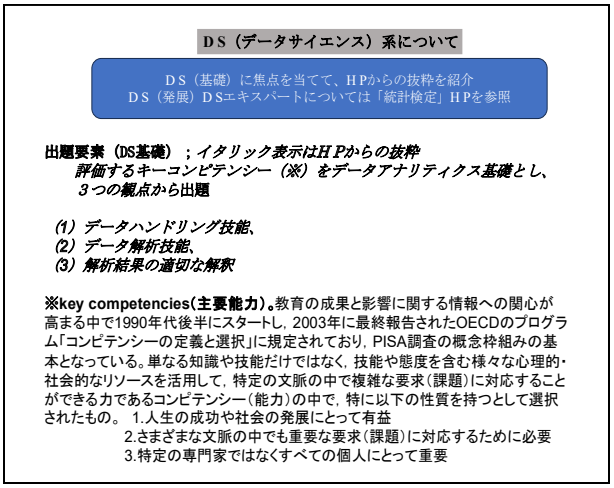


図３　DS（基礎）における出題要素

具体的内容としては、図４に示す通り、データ処理技術と統計学的処理の２面性となっており、また問われる能力及びキーワードなども示されている。

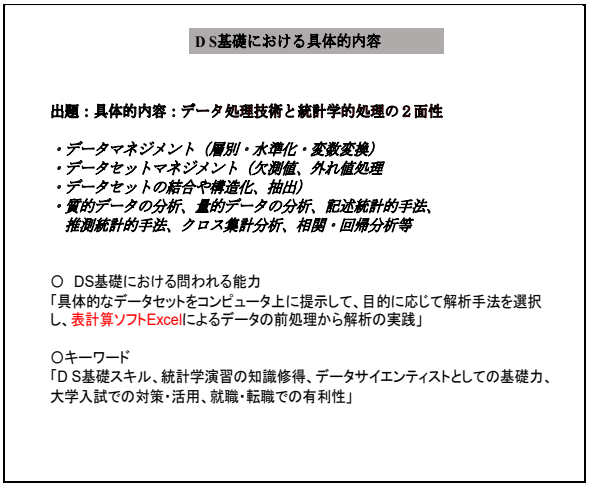


図４　DS（基礎）における具体的内容

テキストである、データアナリティックス基礎（日本統計学会編）およびHPには、出題範囲表が示されている。小項目として、以下のキーワードである。

・データベースマネージメント
・データマネージメント
・データの可視化
・1変量の質的・量的データの分析
・2変量以上の質的・量的データの分析
・確率と確率分布
・推定、検定
・時系列データの分析
・テキストマイニング

３.２　練習問題（テキストから抜粋）

図５は、職場のストレスに関するデータを用いた演習問題である。Excelの PIVOT テーブル機能をうまく使えるのかを問うている。

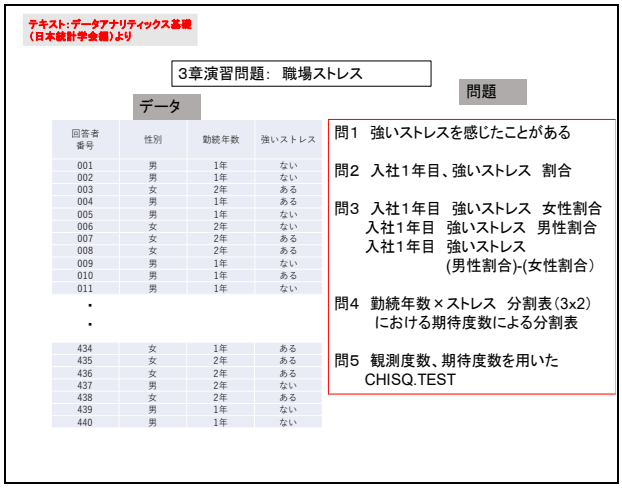


図５　３章演習問題：職場ストレスデータ

４．最後に

DX 推進が叫ばれる中、データサイエンティストの必要性が増している。医療情報部の職員、病院管理業務の職員など、患者データに触れることができる方々が、データサイエンティストとしてのスキルアップを目指され、データ活用により業務の効率化、研究支援等の業務活性化に期待したいと思います。発表者は、関連する大学病院の診療情報管理士に対し、ＤＳ基礎に関する演習体験があり彼女らにとっても良い経験となった。

付録：　DS のみならず、統計学分野に関する「統計検定」のうち、特に馴染みやすいのが、2 級と 3 級ではないでしょうか。そこで、以下に具体的な 2 級と 3 級の内容を示した。また、練習問題も軽視したので、力試しにチャレンジしていただきたい。

さらに、各級に関する教科書が発行されているので、興味がある方は参考にしていただき、是非チャレンジ受験をご検討ください。

統計検定２級の内容

試験内容
大学基礎課程（１・２年次学部共通）で習得すべきことについて検定を行います。

(1) 現状についての問題の発見、その解決のためのデータの収集
(2) 仮説の構築と検証を行える統計力
(3) 新知見獲得の契機を見出すという統計的問題解決力
について試験します。

【具体的な内容】
統計検定2級では、統計検定3・4級の内容に加え、以下の内容を含みます。
•１変数データ（中心傾向の指標、散らばりの指標、中心と散らばりの活用、時系列データの処理）
•２変数以上のデータ（散布図と相関、カテゴリカルデータの解析、単回帰と予測）
•推測のためのデータ収集法（観察研究と実験研究、各種の標本調査法、フィッシャーの３原則）
•確率（統計的推測の基礎となる確率、ベイズの定理）
•確率分布（各種の確率分布とその平均・分散）
•標本分布（標本平均・標本比率の分布、二項分布の正規近似、ｔ分布・カイ二乗分布、Ｆ分布）
•推定（推定量の一致性・不偏性、区間推定、母平均・母比率・母分散の区間推定）
•仮説検定（ｐ値、２種類の過誤、母平均・母比率・母分散の検定[１標本、２標本]）
•カイ二乗検定（適合度検定、独立性の検定）
•線形モデル（回帰分析、実験計画）

統計検定３級の内容

試験内容
大学基礎統計学の知識として求められる統計活用力を評価し、認証するために検定を行います。

(1) 基本的な用語や概念の定義を問う問題（統計リテラシー）
(2) 不確実な事象の理解、２つ以上の用語や概念の関連性を問う問題（統計的推論）
(3) 具体的な文脈に基づいて統計の活用を問う問題（統計的思考）
を出題します。

【具体的な内容】
統計検定3級では、統計検定4級の内容に加え、以下の内容を含みます。
•データの種類（量的変数、質的変数、名義尺度、順序尺度、間隔尺度、比例尺度）
•標本調査と実験（母集団と標本、実験の基本的な考え方、国勢調査）
•統計グラフとデータの集計（１変数データ、２変数データ）
•時系列データ（時系列グラフ、指数（指標）、移動平均）
•データの散らばりの指標（四分位数、四分位範囲、分散、標準偏差、変動係数）
•データの散らばりのグラフ表現（箱ひげ図、はずれ値）
•相関と回帰（散布図、擬相関、相関係数、相関と因果、回帰直線）
•確率（独立な試行、条件付き確率）
•確率分布（確率変数の平均・分散、二項分布、正規分布、二項分布の正規近似）
•統計的な推測（母平均・母比率の標本分布、区間推定、仮説検定）

統計検定３級の試験問題（2019.11.24実施）

問４　次の表は，ある高校の定期試験における英語と数学の結果である。

教科	満点	平均点	標準偏差
英語	200	112	16
数学	100	48	10

(1) 全員の数学の点数に 10 点を加算することとした。その際，100 点を超えた人は いないものとする。このときの数学の点数の平均点と標準偏差の組合せとして，次の ①～⑤ のうちから適切なものを一つ選べ。
① 平均点：48 標準偏差：10　② 平均点：48 標準偏差：20　③ 平均点：58 標準偏差：10
④ 平均点：58 標準偏差：20　⑤ 平均点：58 標準偏差：110

統計検定２級の試験問題（2019.11.24実施）

問８　ある検定試験の対策講座が開講され，その対策講座を受講すれば 70% の確率で検 定試験に合格し，受講しなければ 30% の確率で合格するものとする。検定試験の受 験者が対策講座を受講する確率は 20% であるとする。
(1) 検定試験を受験した人から無作為に 1 人選んだとき，その人が対策講座を受講 した合格者である確率はいくらか。次の ①～⑤ のうちから最も適切なものを一 つ選べ。
① 0.14　② 0.20　③ 0.24　④ 0.30　⑤ 0.70

(2) 検定試験を受験した人から無作為に 1 人選んだとき，その人が合格者であるこ とが判明した。このとき，その人が対策講座の受講生である確率はいくらか。次の ①～⑤ のうちから最も適切なものを一つ選べ。
① 0.02　② 0.15　③ 0.37　④ 0.48　⑤ 0.59

問 13　ある選挙において，100 人の投票者に出口調査を行ったところ，A 候補に投票し た人は 54 人であった。出口調査は単純無作為抽出に基づくとし，二項分布は近似的 に正規分布に従うとする。A 候補の得票率の 95% 信頼区間として，次の ①～⑤ の うちから最も適切なものを一つ選べ。

① 0.54 ± 0.005　② 0.54 ± 0.008　③ 0.54 ± 0.049　④ 0.54 ± 0.082　⑤ 0.54 ± 0.098

起床時体重と就寝時体重の症例報告その後

山本和子（個人）
（元島根大学医学部医療情報学講座所属）

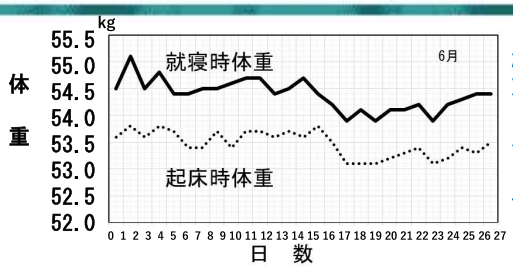
目次（測定時期で冬編と夏編に分けています）

冬編の説明

2023年10月9日から2024年3月31日までの6ヶ月、計175日間測定したから冬編である。

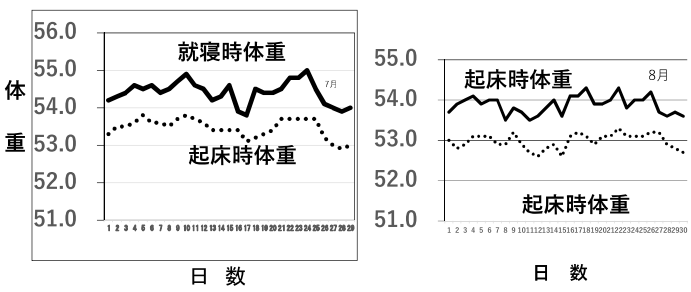
起床時体重と就寝時体重の変動に関する解析
ー夜間の水分損失における脳梗塞発症リスクへの警鐘ー
と題してあらゆる角度から詳細に解析した。
学会誌 Mumps32 に投稿、
別冊はご希望の先生に謹呈いたします。

夏編の分布（2025年6月）

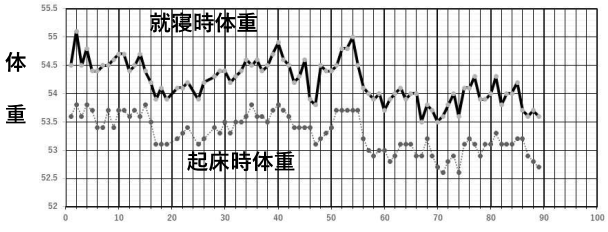


起床時体重と就寝時体重は一定の間隔が保たれているしかし前半が高いのは６月のみ。

夏編の分布（2025年7月と8月）



夏編の分布（2025年6月7月8月合計）



個別の差昼間と夜間（合計）の計算

個別の差（昼間）

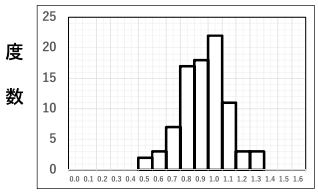
例数	起床時体重	就寝時体重	差
1	53.6	54.5	0.9
2	53.8	55.1	1.3
3	53.6	54.5	0.9
｡	｡	｡	｡
｡	｡	｡	｡
｡	｡	｡	｡
87	52.9	53.6	0.7
88	52.8	53.7	0.9
89	52.7	53.6	0.9
平均	53.3	54.2	0.9

個別の差（夜間）

例数	起床時体重	就寝時体重	翌日起床時体重	差
1	53.6	54.5	53.8	−0.7
2	53.8	55.1	53.6	−1.5
3	53.6	54.5	53.8	−0.7
｡	｡	｡	｡	｡
｡	｡	｡	｡	｡
｡	｡	｡	｡	｡
87	52.9	53.6	52.8	−0.8
88	52.8	53.7	52.7	−1.0
89	52.7	53.6		
	平均値	54.2	53.3	−0.9

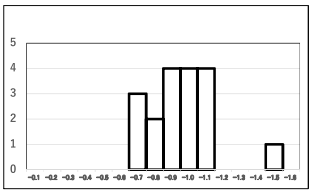
個別の差昼間と夜間の度数分布

個別の差（昼間）の度数分布



0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4

個別の差（夜間）の度数分布



-0.2 -0.4 -0.6 -0.8 -1.0 -1.2 -1.4 -1.6

総数のまとめ（冬編と夏編の比較）

分類　例数	区分	平均値	水分損失
冬	起床時体重	53.9	
175	就寝時体重	54.6	0.71
174	翌日起床時体重	53.9	−0.71
夏	起床時体重	53.3	
89	就寝時体重	54.2	0.9
88	翌日起床時体重	53.3	−0.9

終りに

ご静聴ありがとうございました。
信じられないような成績です。

例数を増やしたいと思います。
ご協力よろしくお願い申し上げます。

オンプレミス型大規模言語モデル（LLM）を用いた高齢者施設入所者の DBD13 推測の可能性の検証

Verification of the feasibility of predicting DBD13 among elderly care facility residents using an on-premise Large Language Model (LLM)

横谷 升美[1)], 竹村 匡正[1)], 松川 未来[1)], 門野 勇介[1)], 山下 晃平[1)], 山口 裕子[3)], グライナー 智恵子[3)], 龍野 洋慶[2)]

Mami Yokotani[1)] Tadamasa Takemura[1)] Miku Matsukawa[1)] Yusuke Kadono[1)] Kohei Yamashita[1)] Yuko Yamaguchi[3)] Chieko Greiner[3)] Hirochika Ryuno[2)]

兵庫県立大学大学院 情報科学研究科[1)], 滋賀医科大学大学院 医学系研究科[2)], 神戸大学大学院 保健学研究科[3)]

Graduate School of Information Science, University of Hyogo[1)]

Graduate School of Medicine, Shiga University of Medical Science[2)]

Graduate School of Health Science, Kobe University[3)]

キーワード：高齢者施設、大規模言語モデル、認知症、DBD13

1. はじめに

昨今、要介護高齢者の増加に伴い、高齢者施設では質の高い介護ケアの提供が求められている。しかし、介護人材が不足していることから、高齢者施設における介護ケアの質の維持や向上が課題となっている。特に認知症ケアは、認知症の行動・心理症状（BPSD）による多様な症状によって、介護者の大きな負担となっている[1)]。介護負担の大きい認知症の入所者に対して質の高いケアを実現するには、その症状を客観的かつ継続的に把握する必要がある。現在、高齢者施設では BPSD を評価する尺度として認知症行動障害尺度（DBD13）が用いられている。この尺度は、BPSD を評価する上で有用な尺度であり、施設ケアの負担軽減や効率化のために推進されている科学的介護情報システム（LIFE）の項目の 1 つとなっている。しかし、13 個の質問項目を持つ DBD13 を人手によって評価することは、職員にとって大きな負担となっていることが現状である。

また、高齢者施設では、入所者の日々の行動や発言などの入所者に関する内容が電子カルテに記載されている。これらの記載の中には DBD13 に関連する記載も多く含まれるが、これらを全て把握することは非常に手間がかかる。

一方で、電子カルテのようなテキストデータを用いた患者の状態の評価として、大規模言語モデル（LLM）が有効であると言われている[2)]。LLM は、電子カルテのような非構造化された情報を処理することや、データの重要な部分を自動で理解することができる。このことから、LLM には電子カルテから DBD13 に関連する記載を認識し、入所者の DBD13 の点数を自動で推測できる可能性がある。

よって、本研究は LLM が電子カルテから高齢者施設入所者の DBD13 を推測できるか検討することを目的とする。

2. 方法

LLM による DBD13 の推測の方法は以下の図 1 の通りである。

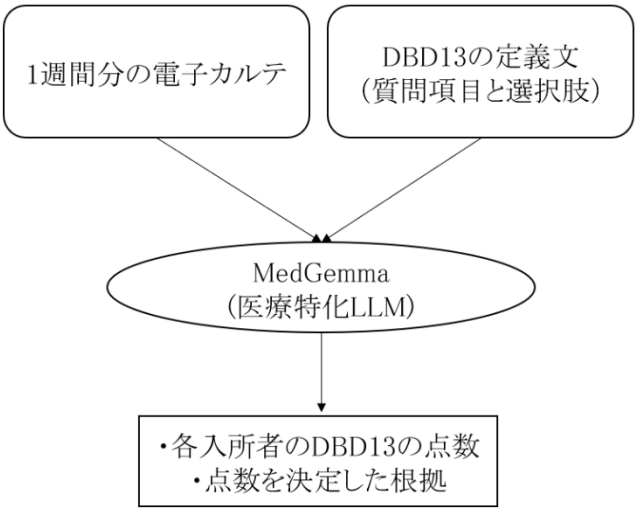


図 1　LLM による DBD13 の推測の全体図

医療分野に特化した LLM の MedGemma に、DBD13 の定義指示文と無作為に抽出した入所者 30 人の DBD13 測定日を最終日とする 1 週間分の電子カルテデータをプロンプトとして入力し、各入所者の DBD13 の各点数とそれらの点数を決定した根拠を出力させる。なお、DBD13 の質問項目は以下の表 1 の通りである。

表 1　DBD13 の質問項目

(1)同じことを何度も何度も聞く
(2)よく物をなくしたり、置場所を間違えたり、隠したりしている
(3)日常的な物事に関心を示さない
(4)特別な理由がないのに夜中起き出す
(5)特別な根拠もないのに人に言いがかりをつける
(6)昼間、寝てばかりいる
(7)やたらに歩き回る
(8)同じ動作をいつまでも繰り返す
(9)口汚くののしる
(10)場違いあるいは季節に合わない不適切な服装をする
(11)世話されるのを拒否する
(12)明らかな理由なしに物を貯め込む
(13)引き出しやたんすの中身を全部だしてしまう

これらの 13 個の質問項目に対して、回答の選択肢と各選択肢の点数はそれぞれ、0 点が「該当する行動が全く見られない」、1 点が「該当する行動が時々見られるが、問題になる頻度ではない」、2 点が「該当する行動がかなりの頻度で見られるが、重篤ではない」、3 点が「該当する行動が頻繁に見られ、介護において問題となる」、4 点が「該当する行動が非常に頻繁に見られ、介護において重大な問題となる」の 5 つである。

また、実際に高齢者施設で測定された DBD13 の点数や電子カルテの記載と、LLM が推測した点数や根拠を比較する。

なお、倫理的配慮として、神戸大学大学院研保健学研究科保健学倫理委員会第 1273 号で承認された。

3. 結果

正解率は約 57.7%、適合率は約 44.0%、再現率は約 21.5%となった。

具体的に見てみると、DBD13 の各質問項目の正解率については表 2 の通りとなった。

表 2　DBD13 の各質問項目の正解率

質問項目	正解率(%)
(1)同じことを何度も何度も聞く	27
(2)よく物をなくしたり、置場所を間違えたり、隠したりしている	40
(3)日常的な物事に関心を示さない	33
(4)特別な理由がないのに夜中起き出す	60
(5)特別な根拠もないのに人に言いがかりをつける	67
(6)昼間、寝てばかりいる	53
(7)やたらに歩き回る	67
(8)同じ動作をいつまでも繰り返す	67
(9)口汚くののしる	67
(10)場違いあるいは季節に合わない不適切な服装をする	73
(11)世話されるのを拒否する	47
(12)明らかな理由なしに物を貯め込む	70
(13)引き出しやたんすの中身を全部だしてしまう	80

また、実際に高齢者施設で測定された DBD13 の点数と LLM が推測した DBD13 の点数における各質問項目の分布は以下の表 3 の通りとなった。

表 3　実際の点数と LLM が推測した点数の各質問項目の分布

(1)同じことを何度も何度も聞く

	実測値	LLM の推測値
0	7	25
1	12	5
2	3	0
3	4	0
4	4	0

(2)よく物をなくしたり、置場所を間違えたり、隠したりしている

	実測値	LLM の推測値
0	13	24
1	10	2
2	2	3
3	3	1
4	2	0

(3)日常的な物事に関心を示さない

	実測値	LLM の推測値
0	8	1
1	9	28
2	6	1
3	5	0
4	2	0

(4)特別な理由がないのに夜中起き出す

	実測値	LLM の推測値
0	18	20
1	8	2
2	1	2
3	2	3
4	1	3

(5)特別な根拠もないのに人に言いがかりをつける

	実測値	LLM の推測値
0	18	26
1	8	1
2	1	2
3	3	1
4	0	0

(6)昼間、寝てばかりいる

	実測値	LLM の推測値
0	15	16
1	5	9
2	7	5
3	3	0
4	0	0

(7)やたらに歩き回る

	実測値	LLM の推測値
0	21	25
1	7	1
2	1	2
3	0	2
4	1	0

(8)同じ動作をいつまでも繰り返す

	実測値	LLM の推測値
0	19	28
1	8	2
2	0	0
3	3	0
4	0	0

(9)口汚くののしる

	実測値	LLM の推測値
0	19	28
1	6	0
2	3	1
3	2	1
4	0	0

(10)場違いあるいは季節に合わない不適切な服装をする

	実測値	LLM の推測値
0	23	28
1	5	0
2	1	2
3	1	0
4	0	0

(11)世話されるのを拒否する

	実測値	LLM の推測値
0	20	12
1	6	17
2	3	1
3	0	0
4	1	0

(12)明らかな理由なしに物を貯め込む

	実測値	LLM の推測値
0	21	30
1	4	0
2	2	0
3	3	0
4	0	0

(13)引き出しやたんすの中身を全部だしてしまう

	実測値	LLM の推測値
0	24	30
1	5	0
2	0	0
3	0	0
4	1	0

4. 考察

LLM は電子カルテに記載のある DBD13 の質問項目の多くに対して、症状の有無を正しく判定した。「同じことを何度も何度も聞く」という質問項目では、同じ内

容のやり取りを複数行った様子や夜間に時刻を何度も尋ねているといった記載があった入所者に対して、LLM はこれらの記録を根拠に、同じことを何度も聞く行動が見られることを正しく判定した。同様に、「よく物をなくしたり、置場所を間違えたり、隠したりしている」という質問項目では、物を紛失したことや義歯ケースを不適切な場所に持ち帰っていたといった記載があった入所者に対して、LLM はこれらの記録を根拠に、物の紛失や置場所の間違いがあることを正しく判定した。また、「日常的な物事に関心を示さない」という質問項目において、実際に高齢者施設で測定された点数は、30 人中 8 人が 0 点の「該当する行動が全く見られない」であった。しかし、電子カルテに記載の多いレクリエーションの参加状況や日中の様子についての記載を根拠に点数を判断していたため、LLM は 30 人中 1 人にのみ 0 点と判断し、他の 29 人に対しては 1 点の「該当する行動が時々見られるが、問題になる頻度ではない」、または 2 点の「該当する行動がかなりの頻度で見られるが、重篤ではない」と判断した。また、点数の判断根拠となった「穏やか」や「マイペース」という記載を、LLM は「積極的ではない」と解釈していた。「特別な理由がないのに夜中起き出す」という質問項目では、「良眠」という記載や夜間のナースコールの様子などの記載があった入所者に対して、LLM はこれらの記録を根拠に、特別な理由なく夜中に起き出していることを正しく判定した。一方、トイレ希望や脱衣があったなど、理由があって夜中に起き出した場合も根拠に含まれていることがあった。「特別な根拠もないのに人に言いがかりをつける」と「口汚くののしる」の 2 つの質問項目では、「暴言あり」や「感情失禁あり」、具体的な発言の内容などの記載があった入所者に対して、LLM はこれらの記録を根拠に、言いがかりやののしる行動が見られることを正しく判定した。「昼間、寝てばかりいる」という質問項目では、日中に臥床していることや臥床が原因でレクリエーションが中止となったことなどの記載があった入所者に対して、LLM はこれらの記録を根拠に、昼間に寝ていることを正しく判定した。「やたらに歩き回る」という質問項目では、事務所内の移動や部屋の出入り、「所在確認を行う」などの記載があった入所者に対して、LLM はこれらの記録を根拠に、歩き回る行動が見られることを正しく判定した。「同じ動作をいつまでも繰り返す」という質問項目では、夜中に何度もゴソゴソする様子や同じやり取りを数回繰り返した様子などの記載があった入所者に対して、LLM はこれらの記録を根拠に、同じ動作を繰り返す行動が見られることを正しく判定した。この質問項目における実際に高齢者施設で測定された点数は、30 人中 19 人が 0 点の「該当する行動が全く見られない」であった。しかし、電子カルテ内に同じ動作を繰り返すという内容に関連する記載が少なかったため、LLM は 30 人中 28 人に対して 0 点と判断した。「場違いあるいは季節に合わない不適切な服装をする」という質問項目では、靴を片方のみ履いている様子やコートを着て廊下で話す様子などの記載があった入所者に対して、LLM はこれらの記録を根拠に、不適切な服装をすることがあることを正しく判定した。「世話されるのを拒否する」という質問項目では、リハビリへの受け入れ不良であることや薬の服用を拒否した様子などの記載があった入所者に対して、LLM はこれらの記録を根拠に、介護を拒否する行動が見られることを正しく判定した。しかし、別の場所で過ごしていたことによりレクリエーションが中止となったという記載を根拠に、LLM が 1 点の「該当する行動が時々見られるが、問題になる頻度ではない」と判断したなどのケースもあった。最後に、「明らかな理由なしに物を貯め込む」と「引き出しやたんすの中身を全部だしてしまう」の 2 つの質問項目では、全入所者の電子カルテに関連記載が全くなかったため、LLM は全入所者に対して 0 点の「該当する行動が全く見られない」と判断した。

このことから、電子カルテに記載のあるものについて、LLM は DBD13 の点数の推測が概ね可能であった。特に、「暴言あり」といった記載や具体的な発言や行動の記録など、質問項目に関連する明確なキーワードや直接的な記述が存在する項目においては、LLM は症状の有無を非常に正確に判定できた。また、「所在確認を行う」という記載から移動が多いことを判断したことや、「本日拒否なし」という記載から普段は世話を拒否する

ことがあることを判断するなど、間接的な記述についても根拠として適切に読み取り、点数に反映できているケースが多かった。よって、LLM が電子カルテから高齢者施設入所者の DBD13 を推測できる可能性があることが示唆された。

一方、正解率や適合率、再現率は低い数値となった。これらが低かった原因として、主に 3 つあると考えられる。1 つ目は、電子カルテに記載のないものがあったことである。「明らかな理由なしに物を貯め込む」と「引き出しやたんすの中身を全部だしてしまう」の 2 つの質問項目では関連する記載が全くなかったため、LLM は全入所者に対して 0 点と判断した。よって、電子カルテに記載のないものは推測が困難であった。このことから、LLM による DBD13 の推測の精度は、入力するデータに DBD13 に関連する記録が詳細かつ網羅的に記述されているかに大きく依存していると考えられる。

2 つ目は、症状の深刻度を 5 段階で正確に分類することが難しかったことである。電子カルテに記載のあるものについて、LLM は症状の有無は正確に判定できていたものの、症状の深刻度や頻度の判定は正確でなかった。そのため、高齢者施設で測定した点数と LLM が推測した点数が一致せず、正解率や適合率、再現率が低い数値となったと考えられる。

3 つ目は、行動の背景にある意図や文脈の解釈が本来の意味と異なることがあったことである。LLM は「穏やか」や「マイペース」という記載に対して「積極的ではない」と判断したように、本来の意味と異なった解釈をし、その解釈を基に DBD13 の点数を推測してしまうことがあった。

このように、電子カルテに記載のないものは推測が難しかったことや、症状の深刻度の正確な判定が難しかったこと、行動の背景にある意図や文脈の解釈が本来の意味と異なることがあったことから、今後は DBD13 に関連する記載がある他のデータの利用やプロンプトの改善、ファインチューニングなどを検討する。

5. おわりに

本研究では、LLM が電子カルテから高齢者施設入所者の DBD13 を推測できるかを検討するために、医療分野に特化した LLM の MedGemma に、DBD13 の定義指示文と無作為に抽出した入所者 30 人の DBD13 測定日を最終日とする 1 週間分の電子カルテデータをプロンプトとして入力し、各入所者の DBD13 の点数とその根拠を出力させた。また、実際に高齢者施設で測定された DBD13 の点数や電子カルテの記載と、LLM が推測した点数や根拠を比較した。その結果、正解率は約 57.7%、適合率は約 44.0%、再現率は約 21.5%となった。また、LLM は電子カルテに記載のある項目の多くは症状を正しく判定したが、記載のない項目は全入所者で 0 点の「該当する行動が全く見られない」と判断した。これらの結果より、電子カルテに記載のあるものは概ね推測が可能であったことから、LLM が電子カルテから高齢者施設入所者の DBD13 を推測できる可能性があることが示唆された。一方、電子カルテに記載のないものは推測が難しかったことや、症状の深刻度の正確な判定が難しかったこと、行動の背景にある意図や文脈の解釈が本来の意味と異なることがあったことから、今後は DBD13 に関連する記載がある他のデータの利用やプロンプトの改善、ファインチューニングなどを検討する。

参考文献

1. 山口晴保, 中島智子, 内田成香 他. 認知症疾患医療センター外来の BPSD の傾向 : NPI による検討. 認知症ケア研究誌, 2017, 1, p3-10

2. Zhangfeifan Yang. (2024). Comparing Traditional Machine Learning and Large Language Models: An Application to Mental Health Text Classification [Doctoral or Master's thesis, University of California]. eScholarship. https://escholarship.org/content/qt0d63p0jj/qt0d63p0jj_noSplash_26cdde9d26cb664cd033aa68206c96f3.pdf?t=sol1

オンプレミス環境下における複数 LLM を用いたサマリ生成の比較

A Comparative Study of Summarization Using Multiple LLMs in an On-Premises Environment

門野勇介 [1)]，山本剛 [2)]，竹村匡正 [1)]
Yusuke KADONO[1)], Tsuyoshi YAMAMOTO[2)], Tadamasa TAKEMURA[1)]
兵庫県立大学大学院 情報科学研究科 [1)]，大阪けいさつ病院 [2)]
Graduate School of Information Science University of Hyogo[1)],
Osaka Police Hospital[2)]

キーワード：LLM、サマリ生成、オンプレミス、電子カルテ

1. はじめに

看護師の病棟業務において記録作成や確認は重要である反面、業務負担が大きいことが問題となっている。特に、いわゆる退院時サマリや逐次サマリ、中間サマリなどは患者のケアの状況を包括的に把握することや、迅速な対応、引き継ぎ等の業務において重要であるが、その作成作業は看護師にとって大きな負担となってきた。これらのサマリ作成は、実際にはニーズに応じて作成されるものであり、看護師同士の情報共有に用いられる場合もあれば、診療情報管理士などの他職種の診療プロセスの把握などにも用いられる。その場合、厳密には求められるサマリはその都度質的・量的に異なるものであると考えられる。

一方で、大規模言語モデル（LLM）は自然で一貫性のある文章を生成できるまでに発展しており、オンプレミス型の LLM でも高い性能で文書生成が可能である。しかし、先述したような質的・量的に最適化できるかどうかは明らかではなく、またモデルによる違いがサマリ作成にどのような影響があるのかは明らかではない。

2. 目的

本研究では、看護記録を入力として、日々のサマリである中間サマリ（＝逐次サマリ）の作成を想定し、オンプレミス環境で実装した複数の LLM がどのようなサマリ作成を行うのかについて検証を試みる。特に、量的・質的な変更が可能かを検証するために、「サマリの長さ」の制御可能性およびその内容の特性について比較検証を試みる。

3. 方法

本研究では、大阪府内にある大規模急性期病院において、2022 年 1 月～2022 年 12 月までに入院した患者 10 名についての電子カルテのうち、看護師が記載した看護記録をデータセットとして用いる。看護記録には Unicode 標準化などの前処理を行う。

また、使用モデルとして Qwen3-30B-A3B-Instruct-2507[1)]（以下、Qwen3）、medgemma-27b-text-it[2)]（以下、medgemma）、gpt-oss-120b[3)]（以下、gpt-oss）の 3 種類の LLM を用いる。Qwen3 は汎用的な大規模モデルであり、幅広い自然言語処理タスクへの適用を目的として設計されている。medgemma は医療データで追加学習された特化型モデルである。gpt-oss は本研究で用いた中では最大規模のモデルであり、多様なタスクにおいて高い性能を示す。これらの LLM をオンプレミス環境に実装し、看護記録を用いて 1 日ごとの中間サマリを生成させる。実験環境は NVIDIA RTX A6000 GPU（48GB VRAM）を 2 基搭載した Linux サーバである。

LLM に入力するプロンプトには、まず中間サマリの

定義と目的を提示した上で、前日の中間サマリがない場合には当日の看護記録のみで作成すること、前日の中間サマリがある場合には当日の看護記録と前日の中間サマリの情報を統合して中間サマリを更新することを指示する。さらに、患者情報・看護上の注意点・検査や薬剤情報などを含めることを明示し、出力文字数を200 字または 400 字で指定する。これを共通のプロンプトとして、3 つのモデルを用いて生成された中間サマリについて、プロンプトの指示内容にある文字数指定に対する適合度や看護記録の記載をどれくらい忠実に反映できているかという観点から比較を行う。

4. 倫理的配慮

本研究は、大阪警察病院倫理審査委員会の承認を受けて実施した(承認番号 1817 号)。

5. 結果

3 つのモデルを用いて、10 人の患者に対して 100 個の中間サマリが生成された。生成された中間サマリの平均文字数を表 1 に示す。200 文字指定では、Qwen3 と gpt-oss がほぼ指示通りの文字数で生成されたのに対し、medgemma は平均 265 字と超過傾向を示した。400 文字指定では、gpt-oss が比較的近い値を維持した一方、Qwen3 はやや不足し、medgemma は大幅に超過した。内容面では、いずれのモデルも検査値などの数値情報を正確に抽出していた。一方で、Qwen3 や gpt-oss が生成したサマリには、モデルによる解釈や看護記録にない記述が含まれる例が見られた。

表 1 生成された中間サマリの平均文字数

model	200 words (mean ± 95%CI, n=100)	400 words (mean ± 95%CI, n=100)
Qwen3	200.8 ± 8.4	331.8 ± 8.8
medgemma	265.6 ± 12.8	722.7 ± 40.7
gpt-oss	200.4 ± 2.0	426.7 ± 14.0

6. 考察

文字数指定への適合度については、Qwen3 と gpt-oss が比較的高い適合度を示した。Qwen3 は 200 字指定では安定していたものの、400 字指定では不足傾向を示しており、入力の中から必要と判断した情報が限定されていた可能性がある。gpt-oss は両方の指定で安定しており、大規模なモデルの制御性能の高さが影響したと考えられる。一方、medgemma は 200 字・400 字のいずれでも大幅に超過しており、文字数の制御に関する学習が十分ではない可能性がある。中間サマリの内容面では、3 モデルとも検査値などの数値データは正確に抽出できていた。一方で、Qwen3 では「改善」や「軽減」など看護記録に明示されていない推測的な表現が散見された。gpt-oss においても、看護記録に存在しない情報が含まれる出力が確認された。これらは、汎用的なモデルにおいて、入力に対して文脈的な補完や説明を加える傾向が影響した可能性がある。これに対し、medgemma は看護記録の記載に基づいた中間サマリを生成する傾向を示し、3 モデルの中では最も高い忠実性を示した。これは、医療データでの追加学習によって、臨床記録を直接反映するような出力が促されていた可能性がある。ただし、いずれのモデルにおいても、前日の中間サマリの内容が翌日以降にも残り、時系列的な誤りが生じるケースが確認された。

7. 結論

本研究では、3 種類の LLM を用いて看護記録から日次の中間サマリを生成し、文字数の制御可能性および内容の忠実性を比較した。Qwen3 と gpt-oss は文字数指定に比較的適合しており、サマリの長さの制御可能性が示唆された。内容面では、3 モデルとも数値情報の抽出は正確であったが、Qwen3 では推測的な表現が、gpt-oss では入力に存在しない情報が含まれる例が確認された。一方で、medgemma は看護記録への忠実性が高い傾向を示した。以上より、LLM による中間サマリ生成にはモデルごとに量的・質的な特性の差異が存在することが示唆され、これらの特性を改善・補完することで、より実用的なサマリ生成が期待される。

参考文献

1. Qwen Team, Qwen3 Technical Report, arXiv:2505.09388, 2025.

2. Andrew Sellergren, Sahar Kazemzadeh, Tiam Jaroensri, et al., MedGemma Technical Report, arXiv:2507.05201, 2025.

3. OpenAI, gpt-oss-120b & gpt-oss-20b Model Card, arXiv:2508.10925, 2025.

大規模言語モデルの検索拡張生成における医薬品情報の活用と評価に関する研究

Application and Evaluation of Retrieval Augmented Generation for Drug Information Service in Large Language Models

世良庄司 [1,2)], 木下 隼仁 [1)], 岡田 章 [1,2)], 永井尚美 [1,2)]
Shoji Sera[1,2)] Hayato Kinoshita[1)] Akira Okada[1,2)] Naomi Nagai[1,2)]
武蔵野大学 薬学部 レギュラトリーサイエンス研究室 [1)], 武蔵野大学 薬学研究所 [2)]
Laboratory of Regulatory Science, Faculty of Pharmacy, Musashino University[1)]
Research Institute of Pharmaceutical Sciences, Musashino University[2)]

キーワード:生成 AI、大規模言語モデル、ChatGPT、検索拡張生成、医薬品情報

1. はじめに

各種の医療情報がデジタル化される中、それらを活用し医療従事者の判断をサポートするシステム・データベースの開発が進められている。近年、急速に発展した生成 AI である大規模言語モデル（LLM）についても各種医療関係業務への活用が拡大されているが、インターネット上の情報を基に学習されたことによる情報の誤り（ハルシネーション）や学習時点の情報が出力される等の鮮度が問題となっている。

一方、これらの AI 技術におけるニューラルネットワークは各情報の関連性について高度に探索、結果を導き出すことに優れている。本技術の活用により病態、処方状況、検査値等、複雑な背景情報から、その患者に特に必要な留意事項等を提示、医療従事者に気付きを促す高度なサポートシステムを構築することが可能であると考えられる。

2. 目的

我々はこれまでに生成 AI を活用した医薬品情報システムの実現形を模索するため、構造化した医療用医薬品データを用い、LLM の強化学習（ファインチューニング）用データとして使用、質問に対する回答の精度向上を確認している[1]。そこで本研究では、蓄積したデータベースの内容を元に回答を生成する技術である拡張検索生成（RAG）を使用し、医薬品情報データベース（DB）を情報源として回答するチャットシステムを構築（図 1）、質問に対する関連情報の取得状況と回答精度を定量的に評価することで医薬品情報業務への活用について検討した。

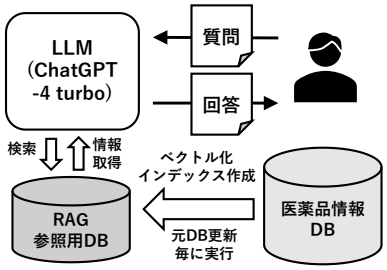
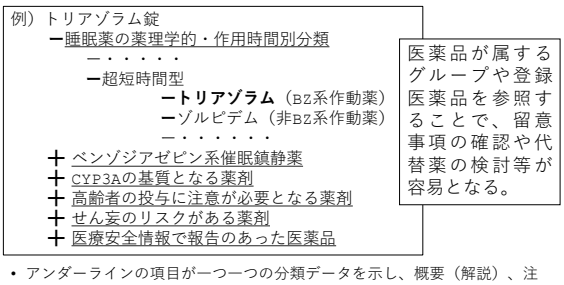


図 1　構築したチャットシステムの概要

3. 方法

医薬品情報 DB には、医療用医薬品添付文書や各種文献を参考に、薬効群や目的ごとに医薬品を紐づけ作成したデータ（分類データ，図 2）を使用した。



・アンダーラインの項目が一つ一つの分類データを示し、概要（解説）、注意事項、一般名、キーワード、参照文献等の情報も保持している。

図2　使用した医薬品情報（分類データ）の概要

チャットシステムは Microsoft Azure の App Service 上に、GitHub で公開されているサンプルコード(Python)[2]を元に表 1 に記載した条件で構築した。

表1 構築システムの条件等

使用LLM(ハイパーパラメータ)
ChatGPT-4 turbo (temperature = 0 , top_p = 0.7)
システムプロンプト
あなたは、日本の医師、薬剤師、看護師等の医療従事者向けの医薬品情報に関する質問に答えるアシスタントです。特に医薬品の安全性に関する情報等、医薬品を使用する際に注意すること等を提示してください。回答は日本語で、長くても250文字以下にしてください。
制限事項
回答の際には与えられたコンテキストの内容を参考にしてください。 ［中略 (DBの各フィールドにどのような情報が記載されているか指示)］ コンテキスト内に該当する情報が見つからなかった場合は、あなたが保持している情報を元に回答してください。
RAG参照用DB格納情報, 検索モード
医薬品情報データ(分類データ)2129件, Hybrid検索 (vector＋text)

回答精度の評価は、Azure プロンプトフローを用い、構築したチャットシステムと同じ条件で RAG 実装有りと無しモデルのフローを作成、前者については全分類データ(2129 件)を参照するモデルと、質問生成に用いた分類データを除いた 2079 件を参照するモデルを作成、計三種類を用意した。評価用の質問と模範回答は、オープンソース RAG 評価ライブラリ Ragas[3]を用い分類データから 50 件をランダムに抽出し 30 問を生成(表2)、各モデルが生成した回答と比較し、F1-score (検索結果の適合率と再現率の調和平均; 0～1)、Similarity (用意した回答との類似性; 1～5)、Relevance (質問と回答の関連性; 1～5)、及び Groundedness (RAG で取得した情報との一致性; 1～5) を用い評価した。なお、フロー自体はモデル毎に3回実行、各評価算出の成功率の合計値が最も高いフローの評価値を採用した。

表2 Ragas で生成した質問・回答例(抜粋)

Q	不穏に対する薬剤選択において、どのような注意が必要ですか？
A	不穏に対する薬剤選択では、せん妄の促進因子である不眠や過活動型せん妄に対する処方を明確にし、治療中はせん妄に対する治療以外で使用している薬剤との相互作用や腎・肝機能、既往歴にも注意を要します。特に、糖尿病やパーキンソン病に禁忌の薬剤があるため、適切な選択が求められます。

4. 結果及び考察

プロンプトフローにより出力した 30 問の質問と模範解答に対する、各モデルの回答の評価値を表3に示す。RAG の実装により、F1-score は 0.30 から 0.40 に、Similarity は 3.04 から 3.66 に向上した。これは、模範解答に対し余分な情報(ノイズ)が無く且つ必要な情報が欠落していないこと、及び、模範解答との類似性も向上していることを示しており、RAG により質問の回答に必要な情報源を適切に検索・取得していることが分かる。また、Relevance については、いずれも 5 段階中 4.5 以上と高評価であるが、ChatGPT が自力で生成した回答との関連性を評価しているため、RAG により生成された回答では評価値が若干低下したものと思われる。さらに、質問生成に用いた分類データを除いた RAG 実装モデルにおいても、F1-score は 0.30 から 0.31 に、Similarity は 3.04 から 3.08 に若干ではあるが向上した。これは、Groundedness が比較的高値 (3.67) を示していることからも、類似した情報や関連性の高いデータからも情報を抽出し、必要な回答を生成することができる可能性を示唆している。

表3 各モデルの評価結果

F1-score	Similarity	Relevance	Groundedness
RAG実装モデル；2129件分類データ投入			
0.40	**3.66**	**4.80**	**4.33**
RAG実装モデル；2079件分類データ投入(質問生成用除く)			
0.31	**3.08**	**4.63**	**3.67**
RAG実装無しモデル(分類データ参照無し)			
0.30	**3.04**	**4.83**	—※

※RAGで情報を取得してないので評価無し

以上のことから、構造化した医薬品情報を RAG に活用することで、回答精度の向上が可能であった。エビデンスの明らかな情報を参照し LLM に回答を生成させることで、医療情報等の専門的な内容にも対応可能と考えられる。また本技術は、医薬品情報のみならず患者の診断/検査/処方情報を同時に参照させることでより高度なサポートシステムの構築に利用可能と考えられる。必要なトークン量は増加するものの、対応可能な LLM はもちろん、インターネットを介さずローカル PC 上で動作する SLM (小規模言語モデル) の開発も進んでおり、効率的なデータ構造や検索手法の検討も併せることで近い将来実現可能と考える。

参考文献

1. 世良庄司ら, 医薬品情報を用いた大規模言語モデルの強化学習による推論結果の精度向上について,日本 M テクノロジー学会大会論文集(第 51 回大会), pp21-22, 2023
2. Azure Samples, https://github.com/Azure-Samples/ (2025/8/10 参照)
3. Ragas, https://docs.ragas.io/en/stable/ (2025/8/10 参照)

仮想データを用いたLLMによるICFコード推定の検証

Validation of LLM-based ICF Code Estimation Using Synthetic Data

中谷 貫太郎, 門野 勇介, 山下 晃平, 竹村 匡正

Kantaro Nakaya　Yusuke Kadono　Kohei Yamashita Tadamasa Takemura

兵庫県立大学大学院情報科学研究科

Graduate school of Information Science University of Hyogo

キーワード:ICF, 仮想データ, LLM

1．はじめに

近年、ウェアラブルデバイスの普及により歩数や心拍、睡眠データに加えて位置情報を継続的に取得することができるようになり、健康増進や生活機能の把握への応用が期待されている。生活機能を標準的に表現する仕組みとして国際生活機能分類（International Classification of Functioning, Disability and Health : ICF）が用いられており、健康状態や心身機能、活動、参加などを体系的にコード化できる。ウェアラブルデータをICFで取り扱うことで、日常生活の具体的な活動や参加を包括的に表現できる可能性がある。例えば、歩行データや位置情報から分かる「銀行に歩いて行き、帰りはバスに乗った。」という事実は、「d450 歩行」だけでなく「d470 移動手段の利用」や「d860 基本的な経済的取引」など複数の側面に対応づけられる。しかし、これらを人手でICFコードに付与するのは現実的ではなく、これまで我々は大規模言語モデル（Large Language Model: LLM）による自動推測の方法を検討してきた[1)]。その一方で、自動推測の正確性を検証するには多様なプロファイルや状況を含む実データが必要不可欠であるが、実データを収集することは困難である。

そこで本研究では、個人プロファイルを含むオープンデータセット利用して、LLM によるウェアラブルデバイスや位置情報等の仮想データを生成し、ICF コードの自動推測を検証することを試みる。

2. 方法

本研究では、LLM を用いて個人のウェアラブルデータや位置情報の仮想データを生成した。その上で、仮想データに対して LLM を用いて ICF コードの自動推測を行った。仮想データの生成および ICF コードの推測には、OpenAI 社が公開しているオープンウェイトモデル gpt-oss-20b を用いた。本研究における手法の全体像を Fig.1 に示す。

2.1 仮想データの生成

仮想データの生成には nvidia/Nemotron-Personas-Japan を使用し、このデータセットに含まれるペルソナとプロファイルを組み合わせてプロンプトとして LLM に入力、1 日の行動ログ及びその際の歩数、心拍数、睡眠時間を出力させた。

出力データの形式は、先行研究において位置情報の取得に用いた GoogleMap タイムライン の構造を参考に設計した。GoogleMap タイムラインでは「〇〇時～〇〇時 : □□に滞在」という時間区間ベースの形式で位置履歴が記録される。本研究でもこの形式を踏襲し、各時間区間に対して歩数・心拍数・睡眠データを付与することで、実際の行動履歴に近い構造を再現した。

生成された仮想データの妥当性は、人手により内容の一貫性、時間的連続性、プロファイルとの整合性の観点から検証した。

2.2　LLM を用いた ICF コード自動推測

得られた仮想データのうち、妥当と判断されたものを対象として ICF コードの自動推測を行った。仮想データ、データセットに含まれるプロファイル（性別・年齢・職業）に加え、ICF コードおよびその説明を記載した json ファイルを 4 分割したものを組み合わせ、プロンプトとして LLM に入力した。出力として、各行動に対応する ICF コードに加え、その ICF コードを選定した理由も同時に生成させた。自動推測された ICF コードは人手で確認し、その妥当性を評価した。

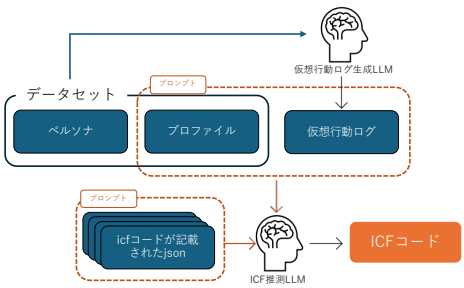


Fig.1 Process of ICF code estimation with virtual data

3.結果

3.1 仮想データの検証

本研究ではデータセットから計 100 例の仮想データを生成した。出力された仮想データの 1 例を Table.1 に示す。生成した 100 例の内、問題なく出力できていたのは 41 例、生成したログが途中で打ち切られていたのは 9 例、ログは生成されていたが、歩数や心拍、位置情報からわかり得ない情報が含まれていたのは 50 件であった。

Table.1 An example of generated virtual data

時刻	行動内容	歩数	平均心拍	最小心拍	最大心拍	睡眠時間
00:00–07:00	睡眠	約30歩	57	49	68	7時間（深い睡眠1.8h）
07:00–08:00	自宅滞在	約280歩	74	58	92	–
08:00–09:00	自宅から公園へ移動（徒歩）	約1,150歩	102	88	119	–
09:00–10:00	公園散策	約850歩	95	82	110	–
10:00–11:00	公園から自宅へ移動（徒歩）	約1,000歩	98	85	112	–
11:00–12:00	自宅滞在	約200歩	70	55	88	–
12:00–12:30	自宅からカフェへ移動（徒歩）	約700歩	105	90	123	–
12:30–14:00	カフェ滞在	約400歩	80	65	100	–
14:00–15:00	カフェから自宅へ移動（徒歩）	約1,050歩	97	84	115	–
15:00–18:00	自宅滞在	約500歩	72	57	90	–
18:00–18:30	自宅からスーパーへ移動（徒歩）	約650歩	104	89	120	–
18:30–20:00	スーパー滞在	約350歩	78	63	105	–
20:00–21:00	スーパーから自宅へ移動（徒歩）	約1,100歩	99	86	118	–
21:00–22:00	自宅滞在	約250歩	68	53	88	–
22:00–23:00	自宅滞在	約200歩	65	50	85	–
23:00–00:00	自宅滞在	約180歩	62	48	82	–

3.2　ICF コードの検証

妥当であると判断した仮想データに対して付与された ICF コードは合計 250 件であり、そのうち存在しないコードが 47 件、人手による確認で妥当と判断されたコードは 167 件、存在はしているが間違っているコードは 20 件、コード自体はあっているが妥当かどうかを判断できなかったコードは 16 件であった。

4.考察

4.1　仮想データの考察

ウェアラブルデバイスや GPS ログから得られるデータのみを想定していたが場所だけでなく、そこで行ったことなどが含まれていたログが半数であった。これらはデータセットに含まれているペルソナを LLM が過剰に反映した可能性がある。また 1 日のログを出力するようにしたが途中で生成が打ち切られていたログもあったが、LLM が与えられたプロンプトに対して十分に回答したと誤認した可能性がある。

4.2　ICF コードの考察

存在しない ICF コードが 47 件あったがこれは LLM が与えた json ではなく LLM 自身の知識に基づいて出力した可能性がある。そのうち ICF コードを付与した理由は妥当であるにも関わらず付与されている ICF コードが間違っていた例が 35 件存在していた。これは LLM がコードの選定段階で誤りを生じている可能性を示しており RAG の適用やプロンプトエンジニアリングによって自動推測の精度向上が求められる。

5.参考文献

1.　中谷 貫太郎, 門野 勇介, 山下 晃平, 竹村 匡正, ウェアラブルデバイス及び LLM を用いた ICF 分類コード自動推測の試み, 生体医工学シンポジウム予稿集, 甲府, 202

大規模言語モデルを用いた
家族看護実践の類型化に関する基礎的検討

A Fundamental Study on the Typology of Family Nursing Practice
Using Large Language Models

松本賢典 [1)], 本田順子 [2)], 築田誠 [3)], 野島敬祐 [4)], 竹村匡正 [1)]
Kensuke Matsumoto[1)] Junko Honda[2)] Makoto Tsukuda[3)] Keisuke Nojima[4)] Tadamasa Takemura[1)]
兵庫県立大学大学院 情報科学研究科 [1)], 兵庫県立大学 地域ケア開発研究所 [2)],
兵庫医科大学 看護学部 [3)] 京都橘大学 看護学部 [4)]
Graduate School of Information Science, University of Hyogo[1)]
Research Institute of Nursing Care for People and Community, University of Hyogo[2)]
Faculty of Nursing, Hyogo Medical University[3)]
Faculty of Nursing, Kyoto Tachibana University[4)]

キーワード：家族看護、電子カルテ、自然言語処理、大規模言語モデル、類型化

１．はじめに

患者の家族全体をケア対象とする「家族看護」は、今後もその重要性がますます高まるとされている [1)]。家族看護分野では、日々実践されている家族看護の記録が電子カルテの自由記述として蓄積されつつあり、そこから家族看護実践に関する記載の実態を把握することが求められている。その中で、自由記述内に含まれている家族看護実践を家族看護学の観点から類型化することは、記載内容を適切に整理するために必要である。特に、こうした類型化を大規模に行うことは一般性のある実態の把握につながるが、膨大な自由記述全体に対して人手で行うことは現実的ではない。

一方で、大規模言語モデル（Large Language Models; LLM）は、近年多くの自然言語処理タスクで画期的な性能を示しており、医療分野では臨床記録分析、診断コード付与を含む分類タスクへの応用も期待されている。特に、従来の機械学習と比較して大規模なラベル付きデータを必要とせず、少数の例示あるいは例示なしでも良好な性能を発揮できる点が特徴である。そのため、理論に基づき意味づけられたカテゴリをLLM の入力内で与えることで、その概念を反映しつつ自動的に類型化を行える可能性がある。したがって、家族看護実践の類型化を大規模に実施するうえで、LLM は有効な手段である可能性がある。

２．目的

本研究では、LLM を用いた家族看護実践に関する記載の類型化の有効性について検討する。

３．方法

本研究の概略を図 1 に示す。

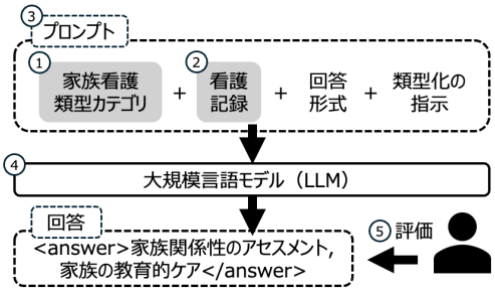


図 1 本研究の概略

① 家族看護類型カテゴリの作成

法橋らが提唱した家族同心球環境理論（CSFET）を参考に、便宜的に家族看護実践を類型化するための 9 種類のカテゴリを作成し、専門家の協力のもと調整を行った。実際に類型化に用いたカテゴリを以下に示す。
[家族構造のアセスメント，家族関係性のアセスメント，

家族のパワー構造のアセスメント，家族の機能のアセスメント，家族の外部環境のアセスメント，家族の問題や課題の診断，家族への直接的ケア，家族への教育的ケア，環境調整・多職種連携 ］

② データセット

大阪府内の大規模急性期病院で 2007 年から 2023 年にかけて蓄積された電子カルテのうち、患者 ID と記載日時が同一の記録群を 1 データとして、専門家が家族看護実践を含むと判断した 812 件を分析対象とした。

③ プロンプト

9 種類の類型カテゴリの定義と、対象の看護記録、回答形式を提示し、その上で看護記録内の各家族看護実践に該当するカテゴリをまとめて回答させた。また、どのカテゴリにも該当しないと判断された場合は「対象外」のみを回答するように指示した。

④ 使用モデル

日本語での応答性能が高い QwQ-Bakeneko-32B を採用し、これを推論時に使用した。

⑤ 評価方法

専門家 1 名による定性的評価を行った。具体的には、LLM の思考過程を参照しつつ、9 種類のカテゴリと LLM の回答を照合してカテゴリの過不足を評価した。

4. 倫理的配慮

大阪警察病院倫理委員会の承認を受けて実施した（承認番号 1817 号）。

5．結果

全 812 件の看護記録に対し、LLM は回答形式に則ってカテゴリを出力した。カテゴリの分布を図 2 に示す。

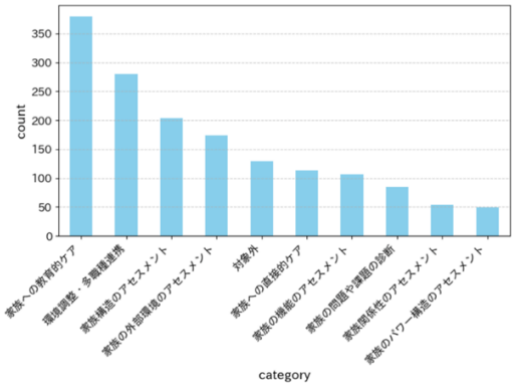


図 2 カテゴリの分布

LLM へ入力した看護記録（概要）と、それに対する LLM の回答の例を示す。

【オリエンテーション】術前オリエンテーション実施¥n 入院オリエンテーション実施¥n 説明を聞いている¥n キーパーソン:妹
家族のパワー構造のアセスメント,家族への教育的ケア

上述した具体例に対する定性的評価を示す。

“キーパーソン: 妹” から「家族のパワー構造のアセスメント」を導いた点は妥当である。一方で、この記録全体を通して家族の同席が明示されていないにもかかわらず「家族への教育的ケア」を導いた点については、家族が説明に加わっていたと解釈する根拠が乏しい。

6．考察

家族看護実践の類型化に関して、LLM は与えられた形式的制約に則った出力を生成でき、家族看護実践の整理に資することが確認された。また、LLM が専門家と類似した判断を伴って、特定の記載に対して専門家と一致したカテゴリを導いた事例が複数確認された。以上より、LLM は提示された家族看護実践の類型カテゴリと形式的制約に基づいて類型化を行える可能性があり、大規模な類型化への有効性が示唆された。一方で、家族の同席が記録上確認できないにもかかわらず、「家族への教育的ケア」など主に家族に対する実践のカテゴリを回答に含める事例も確認された。そのため、LLM が真に家族の存在を認識したうえで判断を下しているのか、またそれが可能かどうかについては検証が必要である。

7．おわりに

本研究では、LLM を用いた家族看護実践の大規模な類型化の有効性が示唆された。今後は、家族の存在を判断するプロセス等を含めた条件分岐の明確化や Few-shot 等の適用を通じて、類型化の妥当性を高めていく。

参考文献

1. 影山 葉子, 矢郷 哲志, 門間 晶子, 野々山 友, 加藤 明美, 小林 裕美, 目 麻里子, 浅野 みどり, 家族支援専門看護師の活動に関する実態調査（第 1 報）—Web 調査報告—, 家族看護学研究, 2025, 30 巻, p. 122-133

地域一体型仮想バイオバンクネットワークを志向した Semantic Data Model(SDM)データウェアハウスの設計と構築

Pilot Design and Construction of a Data Warehouse Based on a Semantic Data Model (SDM) for Regional Integration of Virtual Biobank Networks

中村恵宣 [1),2)], **鈴木英夫** [3),4)], **森本耕平** [1),2)], **岡野隆一** [1),2)], **佐藤伊都子** [5)], **藤原明子** [1),2)], **小林定利** [6)], **植木博文** [6)], **山口智子** [7)], **前田英一** [6)], **村垣善浩** [3),7)], **松岡 広** [1),2)]

Yoshinori Nakamura[1),2)], Hideo Suzuki[3),4)], Kohei Morimoto[1),2)], Takaichi Okano[1),2)], Itsuko Sato, Akiko Fujiwara[1),2)], Sadatoshi Kobayashi[5)], Hirofumi Ueki[5)], Tomoko Yamaguchi[7)], Eiichi Maeda[5)], Yoshihiro Muragaki[3),7)], Hiroshi Matsuoka[1),2)]

神戸大学医学研究科未来医学講座バイオリソース・ヘルスケア統合解析科学分野 [1)], **神戸大学医学部附属病院バイオリソースセンター**[2)], **SDM コンソーシアム** [3)], **株式会社 MoDeL**[4)], **神戸大学医学部附属病院検査部** [5)]、**神戸大学医学部附属病院医療情報部** [6)], **神戸大学医学研究科医学研究科医療創成工学専攻** [7)]

Kobe University Graduate School of Medicine, Department of Future Medical Sciences, Division of Integrated Analysis of Bioresource and Health Care[1)], Kobe University Hospital Bioresource Center[2)], SDM Consortium[3)], MoDeL Inc.[4)], Department of Clinical Laboratory, Kobe University Hospital[5)], Division of Medical Information Processing, Kobe University Hospital[6)], Kobe University Graduate School of Medicine, Department of Medical Device Engineering[7)]

キーワード: SDM、Semantic Data Model、診療情報、リアルワールドデータ、バイオバンク

1．はじめに

神戸大学医学部附属病院バイオリソースセンター（以下、BRC）は、病院併設型バイオバンクとして、研究開始前に試料・情報に対するニーズのヒアリングを実施し、研究目的に適合した収集手順を設定したうえで、前向きな収集および提供を行う「ニーズドリブン型バイオバンク」モデルを採用している。これは、あらかじめ試料と情報を保管・データベース化し、利活用の要望に応じてマッチするものを検索・提供する一般的なバイオバンクとは異なり、特に新鮮試料等の保管が困難なものや、他のバイオバンクでは保管されていないような試料に関する多岐にわたるニーズに対応することを目的としており、実際にこのモデルにより、従来の枠組みでは対応が難しかった研究ニーズに対して柔軟かつ的確な試料・情報提供が可能となっている。

このように、多様かつ高度なニーズに応える BRC を支えるデータ連携基盤としては、一般的なバイオバンクにおいて生体試料および関連する臨床・解析情報の一元管理に用いられる試料・情報管理システム（LIMS：Laboratory Information Management System）のみでは十分とは言えない。特に、新規研究に必要となる試料・情報の入手可能性の検討や症例検索を行うには、電子カルテと連携し、診療情報に対して複雑な条件検索を高速かつ半リアルタイムで実行可能なシステムが不可欠である。さらに、生体試料に対するニーズの複雑化により、BRC 単独での収集が困難となるケースも増加している。これに対応するため、複数の医療機関が連携・協力して試料・情報の収集を可能とする地域

一体型バイオバンクネットワーク構想の実現を目指している（図 1）。以上を可能とするシステム基盤として診療情報 DWH を構築することとし、その共通データモデルには、検索性能の高さに加え、カルテベンダーや解析ツールへの依存が少なく、連携先医療機関へのシステム導入が容易である点を評価し、SDM（Semantic Data Model）を採用した [1-2]。本研究では、プロトタイプとして当院に構築した SDM DWH の構築と利活用について論ずる。

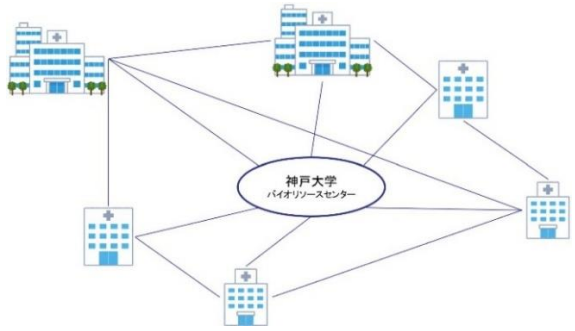


図 1　地域一体型仮想バイオバンクネットワーク

2．SDM DWH の設計・構築と考察

SDM は、患者プロファイル、病歴、処方歴、注射歴、検体検査歴など、臨床研究や症例検索に必要な情報を網羅的かつ構造化して保持する。条件検索のためのアプリケーションを開発して使用することが多いが、BRC では SQL ベースで条件検索を実行している。一般的な検索アプリでは、AND/OR 条件の数やテーブル結合方式、絞り込みの粒度などに制限があるが、SQL を自由に記述すれば、複雑なロジックや多段階の条件分岐を含む検索でも制約なく実行できるからである。加えて、高速処理性能にも優れており、研究設計や臨床現場での即時性が求められる場面でも有効に機能することも利点と考える。さらに、SDM 導入施設間で SQL クエリの再利用が可能であるので、BRC が作成した SQL クエリを他施設の SDM 環境にそのまま適用するだけで、データを移行・統合することなく各施設で同一条件に合致する症例・試料・情報を抽出し、個人情報を保護しながら必要最低限の情報のみを安全に共有する施設横断的な運用が実現できる。

DWH 構築に際しては、実験的に SDM の 11 テーブルを選定し、HIS の参照 DB から SDM の各テーブルに対応する全件データを、ETL（Extract, Transform, Load）処理を介して直接 SDM の DB へ移行した。また、一部 LIMS データに関しても ETL 処理を介して移行した。以降、日々追加される診療データは夜間バッチタスクにより差分抽出・移行を行っており、翌日には前日分までの全件を対象とした検索が可能となっている（図２）。

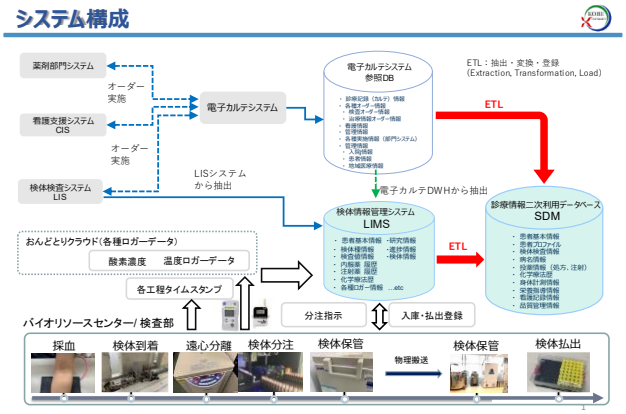


図２　BRC データ連携基盤

3．結語

BRC では LIMS と SDM を相補的に運用することで、迅速な検体払い出しや新規研究に必要となる試料・情報の入手可能性の検討および症例検索が可能となり、研究ニーズへの即応性を高める基盤を構築できた。SDM は単なるデータ格納モデルではなく、セキュアで簡便な施設間連携・検体管理・研究設計・リアルワールドデータ統合など、多面的な展開を可能にする基盤と位置づけている。今後はこの仕組みを連携病院へ展開し、分散保持のまま共通検索ロジックを活用することで、地域全体での臨床研究基盤の強化を図り、地域一体型仮想バイオバンクネットワークの実現を目指す。

参考文献

1. 鈴木英夫,診療情報の二次利用に適した共通データモデル SDM,Mumps Vol.31,pp11-28,2024
2. 鈴木英夫, Semantic Data Modeling, AI in Surgery, pp.43‐53, Springer, 2025

PHR, EHR における共通データベース・モデル

Common Database Model for PHR and EHR

鈴木英夫
Hideo Suzuki
SDM コンソーシアム
SDM Consortium

キーワード:Common Database Model, PHR, HER, SDM

１．はじめに

地域ヘルスケアは、地域住民に対して、保健、医療、福祉の各サービスを提供する社会モデルである。しかし各サービスは独立しており、同一サービス内であっても、異なる機関により運営され、システムも独立しているため、情報連携がなされていないのが現状である。さらに、保健、福祉は、市町村など行政単位で管理されるが、医療は医療圏単位の管轄であることも、情報連携を困難にしている。一方、個人が自身の健康、医療、介護情報を管理する PHR と、医療機関が患者の診療情報を管理する EHR においては、バイタルや健診結果など一部情報は共通であるにもかかわらず、粒度、頻度、精度が異なるため、情報共有がなされていない。本研究では、ヘルスケア共通データベース・モデル SDM を用いて開発した PHR データ定義のデータモデルと、それぞれのサービスにおける利用方法について論ずる。

２．地域ヘルスケア・モデル

地域におけるヘルスケア・サービスは、自治体を構成する市町村が行う地域保健を中心として、職域保健、医療保険者による保健、学校保健、環境保健、広域保健などの保健サービス、医療法、薬事法などで管理される医療サービス、および障害者、児童、介護者などに行われる福祉サービスを、地域住民に対して行うことである。地域ヘルスケア・サービスの社会モデルを図１に示す。

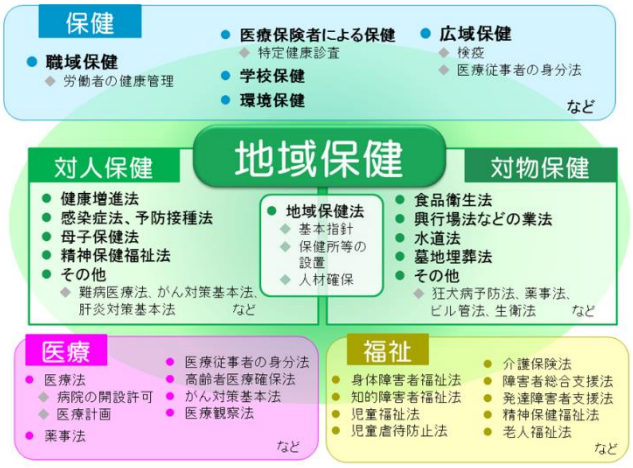


図１　地域保健に関連する施策[1]

３．個人の健康問題

サービス提供者側のモデルに対して、サービス利用者である地域住民の立場で考えてみると、それぞれの立場の違いで、利用形態は変わるが、健常者であれば健康の維持が目的となり、障害者や介護認定者であれば、状態の維持が目的となる。図２に現状におけるサービス利用者である個人を中心としたヘルスケア論理モデルを示す。図２に示すように、個人は自身の健康管理に責任を持ち、自分の判断で各サービスを受けることになる。そして一旦福祉サービス受給の認定を受けると、社会福祉士などが積極的に介入することができる。しかし、各サービス機関で発生した個人情報は共有されないため、個人が自分の情報を取得し、別のサービス機関へ伝達す

るか、各所で新たにインタビューを行わなければならないのが現状である。

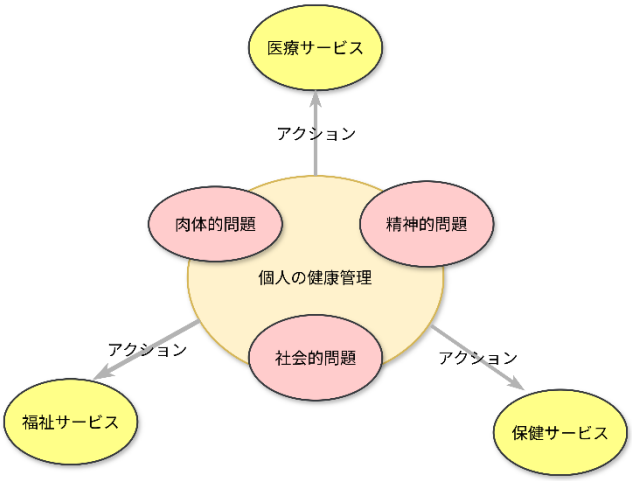


図２　個人を中心としたヘルスケア論理モデル

４．情報共有のための共通データモデル

各サービス機関で保有している個人データを、個人の健康管理のために積極的に利用することができれば、積極的な介入であるアラート機能を実現することができる。そして積極的な介入ができれば、癌や生活習慣病の早期発見、早期治療を可能にし、個人の出費を軽減するばかりでなく、就労からの離脱による収入減を抑止するので、個人にとってもメリットとなる。地域としても、医療費、介護費用の削減に寄与し、病気や介護による生産人口減を抑止し、所得税収入、医療保険、介護保険料の収入も確保できるなど、地方行政にとってもメリットがあり、健診事業など、様々な施策における投資対効果も期待できる。

この情報共有を実現するための方法として、一次利用に適した共通メッセージ交換モデルがある。しかし情報の二次利用のメリットを考えると、検索に適した共通データベース・モデルが望ましい。ヘルスケアデータの共通データベース・モデルとして開発した SDM（一般社団法人 SDM コンソーシアム）で公開されているテーブルは、電子カルテの機能をほぼカバーしているので、バイタルサインなどは、PHR にも適応可能となる[2]。SDM ヘルスケアの論理モデルを図３に示す。

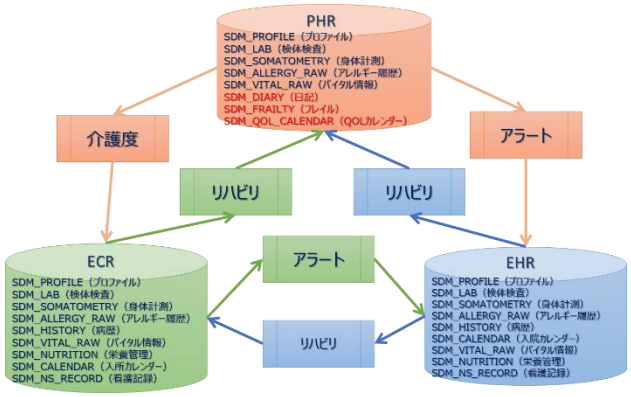


図３　SDM ヘルスケアの論理モデル

５．考察

図３に示す通り、公開している SDM のテーブルの多くが、PHR、ECR（Electronic Care Record）において共通に使用できるが、PHR には、SDM_DIARY（イベントの記録）、SDM_FRAILTY（心身の衰えを示すスコア）、SDM_QOL_CALENDAR（QOL の記録）を加えている。これは医療機関、介護施設などにとって、原因を知るための情報として活用できるだけではなく、自立生活を維持する、すなわち健康寿命を予測するための情報にもなる。

共通データベース・モデルを利用することにより、個人の健康情報を地域のサービス機関で共有し、適切な介入を行うことが可能となるばかりでなく、自立生活から介護生活への一方向を、介護生活から自立生活へ戻る可能性も考えられるようになる。ただし、各所で入力されたデータは、粒度、頻度、精度が違うため、安易な仮説を導かないように、分析には注意を要する。

参考文献

1. 厚生労働省 HP,地域保健, https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/tiiki/index.html
2.鈴木英夫,診療情報の二次利用に適した共通データモデル SDM,Mumps Vol.31,pp11-28,2024