

大規模言語モデルによる電子カルテからの患者適格性評価フレームワークの提案

宮本太郎 川本章太 園部真也
東北大学病院 医療 AI センター
taro.miyamoto.b3@tohoku.ac.jp

概要

大規模言語モデル (LLM) は電子カルテの非構造化テキストから臨床情報を抽出する新たなアプローチとして注目されている。本研究では、LLM を用いて心不全患者の外来心臓リハビリテーション (CR) 適格性を自動評価するフレームワークを提案し、その実現可能性を検証した。単施設において心不全関連診断で入院し、入院中 CR 実施かつ自宅退院した 387 名を対象とした。LLM (Gemini 2.5 Pro) により各患者の外来 CR 適格性と推奨スコア (1-5) を評価し、高推奨群・条件付き推奨群・非推奨群の 3 群に分類した。高推奨群では外来 CR 参加者の 60 日再入院率が非参加者より低い傾向を認め (4.7% vs 17.0%, OR 0.24), LLM による層別化が臨床的に意義のある患者選別に寄与する可能性が示唆された。判定不能例の分析から、退院時 ADL や歩行能力の記載不足が主要因であることが明らかとなり、LLM 活用における医療文書の質の重要性が示された。

1 はじめに

電子カルテ (EHR) には診療録や退院サマリなど患者の臨床状態を詳細に記述した非構造化テキストが蓄積されている。しかし、これらから特定の臨床基準に基づく患者適格性を評価することは、従来の自然言語処理 (NLP) 手法では教師データ作成コストや領域適応の問題から困難であった [1]。

大規模言語モデル (LLM) の登場により、タスク固有の訓練データを必要としないゼロショット学習による臨床テキスト解析が可能となった [2, 3]。LLM は複数の臨床的観点を同時に考慮した総合的判断と、その根拠の言語化が可能である点で従来手法と異なる。

1.1 心臓リハビリテーションと適格性評価の課題

心臓リハビリテーション (CR) は心不全患者の運動耐容能を改善し、再入院を減少させ、生存率を向上させることが多くの研究で示されている [4, 5]。日本循環器学会ガイドラインでも CR は心不全患者に対して Class I (エビデンスレベル A) で推奨されている。CR の有効性のメカニズムには、運動耐容能および骨格筋機能の改善、副交感神経緊張の亢進と交感神経過活動の抑制による自律神経バランスの改善、冠動脈内皮機能の改善が含まれる [6]。

しかし、外来 CR の利用率は日本で約 5-10%、欧米でも 10-30% と低い水準にとどまっている [7, 8]。この低い利用率の一因として、適格患者の系統的な同定が困難であることが挙げられる。入院中に蓄積される膨大な臨床記録から外来 CR 適格性を評価するには、専門的知識と多大な時間を要する。

本研究では、LLM を用いて電子カルテの非構造化テキストから外来 CR 適格性を自動評価するフレームワークを提案し、その実現可能性と臨床的妥当性を検証する。

2 提案手法

2.1 対象と研究デザイン

単施設 (東北大学病院) において、2014 年 4 月から 2024 年 3 月までに心不全関連診断 (急性・慢性心不全, 心筋症等) で入院した 1,680 名を対象とした後方視的観察研究である。このうち、入院中 CR を受け自宅退院した 398 名を一次解析対象とし、LLM 評価に十分な診療録を有する 387 名 (平均年齢 64.4 歳, 男性 64.3%, 在院日数中央値 16 日) を最終解析対象とした。

2.2 LLM による適格性判定の自動化

本研究では、外来 CR への参加が推奨される患者を電子カルテから自動的に抽出する仕組みを構築した。従来、この判定は専門医が個別にカルテを精査して行っていたが、膨大な患者数に対応することは現実的でなかった。LLM を活用することで、構造化データと非構造化テキストの両方を統合的に解析し、ガイドラインに基づく適格性判定を自動化することを目指した。

入力として構造化データ（年齢、性別、入院診断名、在院日数、入院 CR 実施回数）と非構造化テキスト（SOAP ノートおよび退院サマリ、最大 4,000 文字）を用いた。SOAP カルテの統合により、従来の退院サマリのみ（21%）と比較して 93% の患者が LLM 判定対象となった。

2.3 適格・非適格の判定基準

LLM に心臓リハビリテーション専門医としての役割を付与し、外来 CR 適格性の判定を指示した。判定基準は以下のように設計した。

適格と判定される条件：入院治療により急性期を脱し、外来通院でリハビリテーションを安全に継続できる状態まで回復した患者を対象とする。具体的には、(1) 急性期心不全後の血行動態安定、(2) 独立した通院能力（歩行可能、日常生活動作〔ADL〕自立）、(3) 運動療法の安全な実施が可能、の 3 条件を満たすことが求められる。

非適格と判定される条件：運動療法のリスクが高い、または外来通院自体が困難な状態を示す。具体的には、(1) 寝たきりまたは高度 ADL 障害、(2) 持続的強心薬投与、(3) コントロール不良の不整脈、(4) 重度認知症、(5) 重篤な併存症（活動性感染症等）が該当する。

本研究の対象は心不全で入院し CR を必要とした患者に限定されているため、「健康で CR を必要としない」患者は対象外であり、適格基準は「回復した患者」を選定するものである。

2.4 出力形式と運動時の注意事項

LLM は以下の JSON 形式で出力する：適格性判定（適格／非適格／判定不能）、判定根拠（2-3 文）、運動時注意事項、推奨スコア（1-5）。

運動時注意事項の項目は、日本循環器学会の運動療法ガイドラインに基づき、運動療法の開始や継続

にあたって慎重な評価を要する医学的状態（不安定狭心症、コントロール不良の心不全、重症不整脈、重症弁膜症等）を指す。LLM がカルテテキストからこれらの状態を検出した場合に記載される。推奨スコアにより、二値判定では表現できない連続的な適格性の程度を捉えることを意図した。

3 実験

3.1 LLM 評価の実施

LLM（Gemini 2.5 Pro）により各患者の外来 CR 適格性を評価した。LLM は適格判定（適格あり／適格なし／判定困難）と推奨度（1-5、1=推奨しない、3=条件付き推奨、5=強く推奨）を出力した。結果に基づき 3 群に分類した：高推奨群（適格かつスコア 4-5）、条件付き推奨群（判定不能または適格かつスコア 1-3）、非推奨群（非適格）。

3.2 評価指標

主要評価項目は外来 CR 継続（退院後に 1 回以上の外来 CR を実施）とした。副次評価項目は退院後 60 日以内の非予定心不全再入院および全原因再入院とした。非予定心不全再入院は、予定入院（弁膜症手術、デバイス植込み、アブレーション等）および非心血管疾患による入院を除外した心不全関連再入院と定義した。

4 結果

4.1 LLM による評価結果と群分類

387 名に対する LLM の評価結果は、適格 224 名（57.9%）、非適格 109 名（28.2%）、判定不能 54 名（14.0%）であった。推奨度の分布は、推奨度 5（強く推奨）162 名（41.9%）、推奨度 4（推奨）25 名（6.5%）、推奨度 3（条件付き推奨）74 名（19.1%）、推奨度 2 が 10 名（2.6%）、推奨度 1（推奨しない）116 名（30.0%）であった。

アウトカム解析対象 380 名の 3 群分類では、高推奨群 184 名（48.4%）、条件付き推奨群 88 名（23.2%）、非推奨群 108 名（28.4%）となった。

4.2 3 群の臨床的特徴と転帰

表 1 に 3 群の臨床的特徴を示す。高推奨群は最も若く（ 62.2 ± 15.9 歳）、在院日数が短かった（中央値 15 日）。条件付き推奨群は高推奨群より 5.4 歳高齢

であった (67.6±16.8 歳)。非推奨群は在院日数が最長 (中央値 28 日) で、入院中 CR 単位数も最多であった (中央値 764)。これは入院中のリハビリは必要であったが、ADL 低下・認知機能障害・重篤な併存疾患などの理由で外来 CR 非適格と判定された群である。

表 1 3 群分類別のベースライン特性と臨床転帰 (N=380)

	高推奨群 (n=184)	条件付き (n=88)	非推奨群 (n=108)
年齢 (歳)	62.2±15.9	67.6±16.8	65.6±17.8
男性 (%)	64.7	63.6	63.9
在院日数 (中央値)	15	13	28
入院中 CR 単位数	358	507	764
外来 CR 継続 (%)	21.2	13.6	20.4
心血管再入院 (%)	7.1	6.8	7.4

外来 CR 継続率は高推奨群 21.2%、条件付き推奨群 13.6%、非推奨群 20.4%であり、群間で有意差を認めなかった ($p=0.31$)。60 日心血管再入院率は高推奨群 7.1%、条件付き推奨群 6.8%、非推奨群 7.4%で群間差を認めなかった ($p=0.99$)。

興味深いことに、非推奨群の外来 CR 実施率は高推奨群と同等であった (20.4% vs 21.2%)。これは LLM 判定と実際の臨床判断に乖離があること、または家族支援等により通院が可能となった症例の存在を示唆する。

4.3 高推奨群における外来 CR 参加と再入院率

高推奨群に限定した解析では、外来 CR 参加者 (n=39) の 60 日再入院率は 4.7%であったのに対し、非参加者 (n=145) では 17.0%であり、外来 CR 参加群でオッズ比 0.24 (95%CI: 0.05–1.08) と再入院リスクの低下傾向を認めた。

この結果の解釈には慎重を要する。本研究は後方視的観察研究であり、外来 CR 参加者は参加可能な状態であった患者集団であるため、因果関係の主張には限界がある。CR を受けた患者の再入院率が低いのは、CR の効果というよりも、もともと状態が良好で CR に参加できた患者群であった可能性がある。しかしこの結果は、LLM が「高推奨」と判定した患者群において外来 CR への積極的な紹介が有益である可能性を示唆しており、前向き研究による検証が望まれる。

4.4 判定不能例の分析：退院時情報の不足

54 名の判定不能症例について、LLM が出力した判定根拠を分析した。主要因は以下の通りで

あった：

1. 退院時 ADL・歩行能力に関する記載不足 (最多, 67.9%が推奨度 3) : 「退院時の機能状態, ADL, 歩行能力が不明」「自力での通院が可能かどうかの情報がカルテに記載されていない」等
2. 退院時情報の不足 : 「カルテ記録が入院治療中で終了しており, 退院時の状態が不明」「強心薬の持続投与が退院までに中止されているか不明」等
3. 複合的な判断困難 : 「重症大動脈弁狭窄症と冠動脈有意狭窄が存在し, 運動療法の安全性が担保できない」等

特に (1) が最多であったことから、退院サマリにおける ADL・歩行能力・通院可否の記載標準化が LLM 精度向上に寄与すると考えられる。

4.5 年齢層別解析

年齢層別の LLM 判定分布と外来 CR 実施率を分析した。65 歳未満 (n=159) では適格あり 59.7%、外来 CR 実施率 29.6%であった。65–74 歳 (n=120) では適格あり 64.2%、外来 CR 実施率 19.2%であった。75 歳以上 (n=101) では適格あり 46.5%、適格なし 36.6%と高齢者で非適格の割合が高く、外来 CR 実施率は 3.0%と著しく低かった。

75 歳以上の高齢者では外来 CR 実施率が著しく低く、LLM 判定は高齢者の身体機能低下を反映していると考えられる。この年齢層に対しては在宅リハビリや訪問リハビリなど、外来 CR 以外の代替的選択肢を検討する必要がある。

5 考察

5.1 LLM 判定の臨床的妥当性

本研究は、LLM がゼロショットで臨床的に妥当な適格性判定と推奨スコアを出力可能であることを示した。14.0%という判定不能率は退院時 ADL や歩行能力の記載不足に起因しており、LLM 活用において入力となる医療文書の質が重要であることを示す。一方で、情報不足時に「判定不能」を出力する LLM の振る舞いは、不確実性の適切な表現として臨床的に望ましい。

3 群間で心血管再入院率に有意差を認めなかった点については以下のように解釈できる：(1) 入院中 CR を受け自宅退院した患者は比較的前後良好な集

団である、(2)60日という短期間では群間差が顕在化しなかった可能性がある、(3)全体として外来CR継続率が約19%と低く、CRの効果を検出する検出力が不足していた可能性がある。

5.2 臨床実装への示唆

本研究の結果は、LLMベースのCR適格性スクリーニングの臨床実装に向けた以下の示唆を提供する：(1)推奨度5の患者の優先紹介：リソースが限られた環境では、推奨度5の患者を最優先で外来CRに紹介することが効率的である、(2)判定困難例への追加評価：ADL・歩行能力の情報を追加収集することで、判定困難例の多くを適格判定に移行できる可能性がある、(3)高齢者への代替介入：75歳以上で外来CR非適格と判定された患者には、在宅リハビリや訪問リハビリの検討が推奨される、(4)カルテ記載の標準化：退院サマリにADL・歩行能力・通院可否を記載することで、LLM判定の精度向上が期待できる。

現状では外来CR紹介の判断は主治医の経験に依存しており、適格患者の見落としが生じうる。LLMによるスクリーニングは、膨大なカルテ情報から適格候補を効率的に抽出し、専門家による最終判断を支援するツールとして機能しうる。特に、高推奨と判定された患者への積極的な介入により、CR利用率向上と予後改善に寄与する可能性がある。

5.3 研究の限界

本研究の限界として、(1)単施設データであり一般化可能性に限界があること、(2)専門家による判定との比較検証を行っていないこと、(3)プロンプト設計の最適化が十分でないこと、(4)後方視的デザインのため因果推論は困難であること、(5)心不全重症度データ(LVEF, BNP値, NYHA分類)が入手できずLLM判定との関連を検討できなかったこと、(6)60日の追跡期間は心血管アウトカム評価には短い可能性があることが挙げられる。今後は専門家アノテーションによる精度検証と他の臨床適格性評価タスクへの汎化可能性の検討が必要である。

6 おわりに

本研究では、LLMによる電子カルテからの患者適格性評価フレームワークを提案し、心不全患者の外来CR適格性判定への応用可能性を示した。LLMは387名中57.9%を適格あり、28.2%を適格なし、

14.0%を判定困難と判定し、推奨度(1-5スケール)を含む評価を出力した。高推奨群でのCR参加者における再入院率低下傾向は、LLMによる層別化の臨床的意義を示唆する。判定困難例の分析により退院時情報の記載標準化の重要性が明らかとなった。75歳以上の高齢者では外来CR実施率が著しく低く(3.0%)、代替介入の検討が必要である。本手法は心臓リハビリテーションに限らず、様々な臨床的適格性評価に応用可能な汎用的フレームワークとして期待される。

参考文献

- [1] Kory Kreimeyer, et al. Natural Language Processing Systems for Capturing and Standardizing Unstructured Clinical Information: A Systematic Review. *Journal of Biomedical Informatics*, Vol. 73, pp. 14–29, 2017.
- [2] Arun James Thirunavukarasu, et al. Large Language Models in Medicine. *Nature Medicine*, Vol. 29, pp. 1930–1940, 2023.
- [3] Karan Singhal, et al. Large Language Models Encode Clinical Knowledge. *Nature*, Vol. 620, pp. 172–180, 2023.
- [4] Rod S. Taylor, et al. Exercise-based Rehabilitation for Heart Failure: Cochrane Systematic Review, Meta-analysis, and Trial Sequential Analysis. *JACC: Heart Failure*, Vol. 7, No. 8, pp. 691–705, 2019.
- [5] Lindsey Anderson, et al. Exercise-Based Cardiac Rehabilitation for Coronary Heart Disease: Cochrane Systematic Review and Meta-Analysis. *Journal of the American College of Cardiology*, Vol. 67, No. 1, pp. 1–12, 2016.
- [6] Fabio Roveda, et al. The Effects of Exercise Training on Sympathetic Neural Activation in Advanced Heart Failure: A Randomized Controlled Trial. *Journal of the American College of Cardiology*, Vol. 42, No. 5, pp. 854–860, 2003.
- [7] Yoichi Goto. Current State of Cardiac Rehabilitation in Japan. *Progress in Cardiovascular Diseases*, Vol. 56, pp. 557–562, 2014.
- [8] Karam Turk-Adawi, et al. Cardiac Rehabilitation Availability and Density Around the Globe. *EClinicalMedicine*, Vol. 13, pp. 31–45, 2019.