

自記式質問紙を用いた抗精神病薬関連遅発性 ジスキネジアの機械学習予測モデルの開発

布袋病院
林將人

2025/6/24版

研究目的

- 抗精神病薬の長期使用に関連する遅発性ジスキネジアは、治療が難しいことが多く、症状が進行すると日常生活に大きな支障をきたす。そのため早期発見が重要である
- AI技術・機械学習の進歩により、医療分野においても、様々なデータから高精度に臨床的評価・予測が行えると報告されている
 - (severe mental distress予測(高GAD-7, 高PHQ-9得点者予測)[1]、医療関連感染症予測[2])
- 本研究では先行調査で取得された自記式質問紙を用いることにより、ジスキネジア有症状者を予測できる機械学習モデルの開発を行った

先行調査

- 2023年10月～12月に外来を受診し、抗精神病薬を投与されている患者1070名に自記式アンケートを配布し、862名分を回収
- 回答拒否8名、回答不備27名を除外した827名
- 質問該当項目がある392名
 - 医師・看護師がDIEPSS評価
- ジスキネジア症状1点以上：38名
0点：354名

記入日, 年齢, 性別の選択, 16のチェック式質問項目, **自由記載欄**から構成

ヤンセンファーマ社が作成した「もしかして遅発性ジスキネジアではありませんか？」(2023)という冊子を参考に作成した。

16の質問

Q1: 話しにくい

Q2: 口から音が出る

Q3: 歩きにくい

Q4: 食べ物が噛みにくい

Q5: 食べ物が飲み込みにくい

Q6: 歯の噛み合わせがよくない

Q7: 入れ歯（義歯）が合っていないと感じる

Q8: 無意識な動きによって痛みを感じる

Q9: 物（鉛筆、ジッパー、箸、スプーン、歯ブラシなど）をつかみにくい

Q10: 文字を書く・パソコンや携帯電話で字が打ちにくい

Q11: 体の動きが気になってしかたがない

Q12: 家族や他人から見られているような気がする

Q13: 公共の場所（スーパーマーケット、電車・バスなど）に行きづらい

Q14: 友人や知人に会いにくい

Q15: 学校・職場・デイケアなどに行きにくい

Q16: 家族や他人から止めるように注意を受ける

自由記載:

先行調査の結果

表1. ジスキネジア有症状群, 無症状群の患者背景

Variable		Participants (n=392)	Dyskinesia group (n=38)	Non-Dyskinesia group (n=354)	p-value	Odds Ratio
		n(%)	n(%)	n(%)		
Age Group	0-9	5 (1.3)	1 (2.6)	4 (1.1)	0.687	
	10-19	26 (6.6)	1 (2.6)	25 (7.1)		
	20-29	49 (12.5)	6 (15.8)	43 (12.1)		
	30-39	86 (21.9)	10 (26.3)	76 (21.5)		
	40-49	101 (25.8)	12 (31.6)	89 (25.1)		
	50-59	57 (14.5)	2 (5.3)	55 (15.5)		
	60-69	43 (11.0)	4 (10.5)	39 (11.0)		
	70-79	23 (5.9)	2 (5.3)	21 (5.9)		
	80-89	2 (0.5)	0 (0.0)	2 (0.6)		
Sex	Male	184 (46.9)	16 (42.1)	168 (47.5)	0.609	0.81
Questions	Q1	91 (23.2)	13 (34.2)	78 (22.0)	0.106	1.84
	Q2	33 (8.4)	4 (10.5)	29 (8.2)	0.545	1.32
	Q3	74 (18.9)	5 (13.2)	69 (19.5)	0.512	0.63
	Q4	51 (13.0)	7 (18.4)	44 (12.4)	0.309	1.59
	Q5	52 (13.3)	4 (10.5)	48 (13.6)	0.802	0.75
	Q6	71 (18.1)	7 (18.4)	64 (18.1)	1.000	1.02
	Q7	26 (6.6)	0 (0.0)	26 (7.3)	0.158	0.00
	Q8	66 (16.8)	6 (15.8)	60 (16.9)	1.000	0.92
	Q9	40 (10.2)	11 (28.9)	29 (8.2)	0.001	4.57
	Q10	67 (17.1)	10 (26.3)	57 (16.1)	0.116	1.86
	Q11	66 (16.8)	8 (21.1)	58 (16.4)	0.493	1.36
	Q12	136 (34.7)	14 (36.8)	122 (34.5)	0.858	1.11
	Q13	123 (31.4)	14 (36.8)	109 (30.8)	0.464	1.31
	Q14	120 (30.6)	10 (26.3)	110 (31.1)	0.711	0.79
	Q15	104 (26.5)	13 (34.2)	91 (25.7)	0.253	1.50
	Q16	66 (16.8)	9 (23.7)	57 (16.1)	0.254	1.62
Notes	Existed	144 (36.7)	9 (23.7)	135 (38.1)	0.110	0.50

P-value; Age Group: Chi-square test, Others: fisher exact test

- 促進因子 (5%水準で有意差)
 - 「Q9: 物をつかみにくい」 ($p < 0.001$)
- 促進因子である可能性 (15%水準)
 - Q1 ($p=0.106$), Q10($p=0.116$)
- 抑制因子である可能性
 - 自由記載欄に記入 ($p=0.110$)

• 機械学習を用いることで、有症状者を高精度に予測可能か？

方法

- DIEPSS被験者392名のデータ
- 性別、年齢層、二択（有症状、無症状）の回答項目16個、「その他」自由記載項目
 - これらの特徴量をLLMなどを用い加工(特徴量エンジニアリング)
- 機械学習アルゴリズム: Light Gradient Boosting Machine
- 学習評価: 層化 k 分割交差検証($k=5$)
- マッシュューズ相関係数(MCC)で高得点のものをモデルに採用
 - 医学分野ではCohenの k 係数を用いて評価されることが多いが、不均衡データの二値分類では不安定性が指摘されているため、MCCを採用 [3]
- 作成したモデルの正答率、感度等のパフォーマンスを示す
- SHAP値を図示しモデルの構造について考察

基礎理論 LLM / SHAP

- 大規模言語モデル (Large Language Model; LLM)
 - ChatGPT, Gemini, Copilot, Llamaなどの言語を扱う**生成AIの技術**
 - 本研究では自由記載項目を解釈し、数値化するために活用した
 - ある文章 x が、単語 x_1, x_2, \dots, x_T で構成されているとする。パラメータ θ で文章全体が生成される同時確率 $P(x; \theta)$ は以下のように表せる。LLMの仕事はこの条件付き確率を計算することである。

$$P(x; \theta) = \prod_{t=1}^T P(x_t | x_1, x_2, \dots, x_{t-1}; \theta)$$

- SHAP (SHapley AdditivePlanation)
 - 機械学習モデルの予測根拠を明らかにする技術
 - 一つ一つの特徴量が「どの程度」「どちらの方向」に影響を与えたか可視化
 - 本研究では**正の寄与**で**ジスキネジア症状があると予測**しやすくなる。
 - ある特徴量 j 以外のすべての特徴量の組み合わせについて、予測モデルの出力 f が特徴量 j を含むことで生じる予測結果の差を評価することにより、特徴量 j の貢献度(重要性)が分かる

$$\phi_j = \sum_{S \subseteq J \setminus \{j\}} \frac{|S|! (|J| - |S| - 1)!}{|J|!} (f(S \cup \{j\}) - f(S))$$

$f(S)$ は特徴量の集合 S の元での予測結果の期待値
 $J = \{1, \dots, J\}$ は集合を表す
 $|J|$ は集合 J の要素の数 (ここでは $|J| = J$)
集合 S は集合 J から j を除いた集合 ($J \setminus \{j\} = \{1, \dots, j-1, j+1, \dots, J\}$) の部分集合

特徴量

• 特徴量エンジニアリング

- AGE_GROUP: 年齢層 (離散): 20代以下, 30代, ... 70代, 80代以上
- SEX: 性別 (2値): 男性であれば1
- Qj: j 番目の質問の回答結果 (2値): 該当すれば1
- **COUNT**: 質問の総回答数 (離散)
- **Qj_AG_DIFF**: j 番目の質問の回答結果と年齢層の各回答項目の出現確率との差(平均センタリング) (連続)
- **LLM**: **自由記載項目をLLMによる数値化** (連続)

質問の主成分分析による変換、全サンプル・年齢層における希少性スコア、cosine類似度、等の特徴量も検討

• Prompt Engineering (LLM)

- モデルにはLlama-3-ELYZA-JP-8Bを用いた
- 厚生労働省の文書[4]を参考にDynamic Knowledge Injection (+Few-shot prompting), Role prompting, Chain of Thought等の技術を用いプロンプトを作成した。
- 「その他」の文章を入力し、10回出力させ、「症状が存在する」「症状が存在しない」と答えた数を割合として算出した。
(文章の記載がないものは-1とした)

#背景

ジスキネジアは、自分の意思とは無関係に身体が勝手に動いてしまう不随意運動の一種です。(省略)

#役割

あなたに与えられたタスクは (省略) 症状が存在する可能性があれば「ジスキネジア症状が存在する」を書いてください。
症状が存在する可能性が無かったり、関係のない文章であれば「ジスキネジア症状が存在しない」を書いてください。

$$q_{LLM} = \frac{N_{Positive}}{N_{Positive} + N_{Negative}}$$
$$q_{LLM} \in \{-1\} \cup [0,1]$$

結果

表2. モデルのパフォーマンス

Evaluation metrics	point (95% CI)
MCC	0.497 (0.362, 0.622)
Accuracy	0.929 (0.906, 0.949)
Recall (Sensitivity)	0.343 (0.219, 0.474)
Precision (PPV)	0.814 (0.636, 1.000)
Specificity	0.992 (0.983, 1.000)
ROC AUC	0.667 (0.604, 0.733)
PR AUC	0.610 (0.484, 0.716)
Kappa coefficient	0.447 (0.303, 0.583)
Balanced accuracy	0.667 (0.604, 0.733)
F1 score	0.478 (0.333, 0.612)
F0.5 score	0.631 (0.482, 0.759)
F2 score	0.386 (0.254, 0.519)
G-mean	0.579 (0.465, 0.685)
Brier score	0.071 (0.051, 0.094)

推定値, 95% CIはN=10000の
ブートストラップ法から導出

正答率(Accuracy): 92.9%

感度(Sensitivity): 34.3%

陽性的中率(Precision): 81.4%

特異度(Specificity): 99.2%

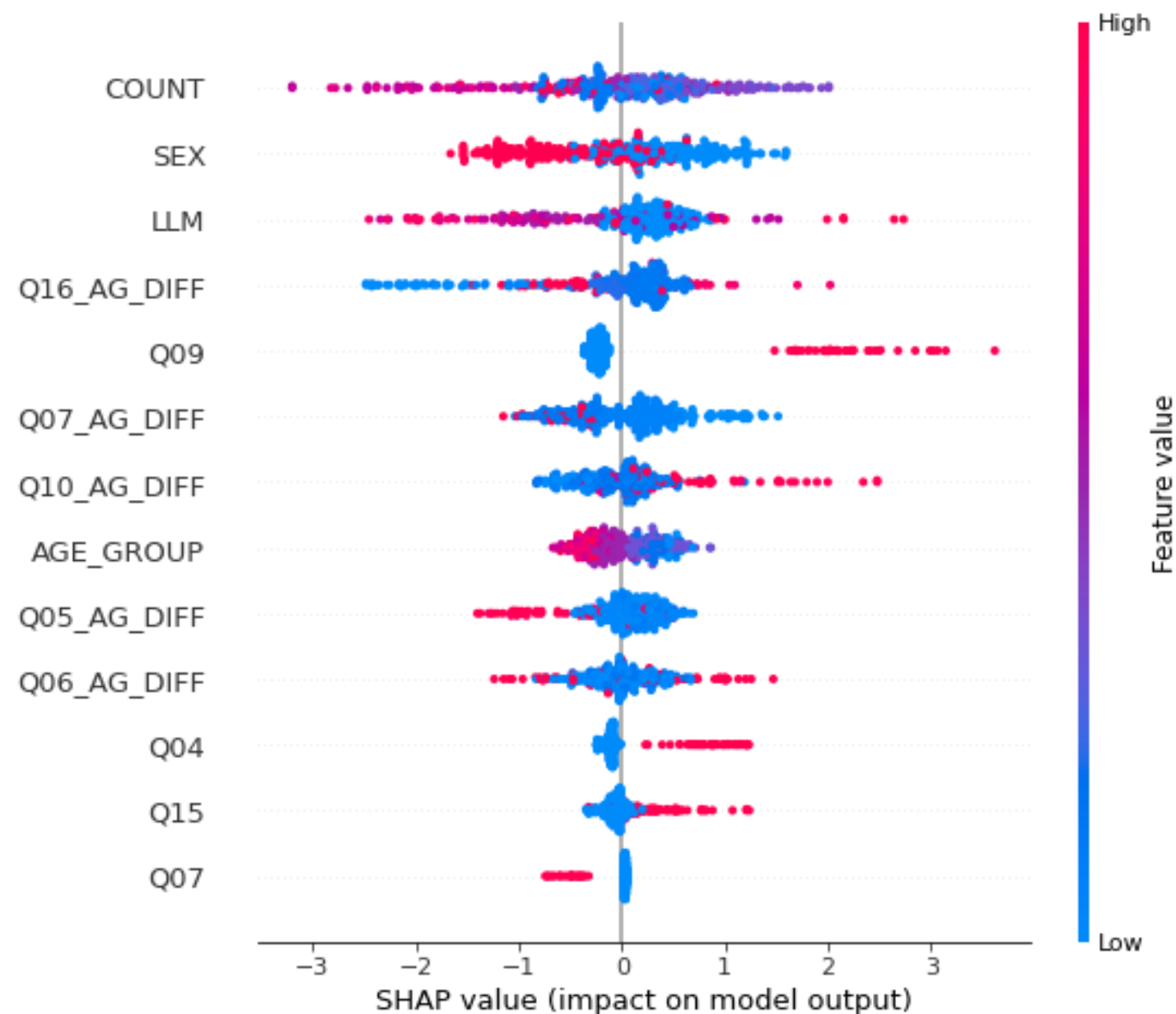


図1. Beeswarm Plot. (全てのデータのSHAP値をプロット)

Beeswarm Plot: 特徴量の重要性について一目で知ることができる
上位に位置する特徴量ほど重要性が高い
横軸がSHAP値: 正であるほどジスキネジア症状を有すると予測する
カラーバー: 赤色~青色は特徴量の値 (大~小)

表3. 自由記載項目のLLM解釈による数値化（抜粋）

自由記載項目	LLM
舌の癖がある（舌が動くときがある）	1.000
足踏みしてしまう	1.000
無意識に体、手足が動く。ストップがかからない	1.000
口をもぐもぐ	1.000
舌の出し入れをしてしまう。手がクネクネする	1.000
足がピクピクする	0.900
口が動かしにくい/外出できない	0.900
口、頬が知らないうちに軽く動いていることがある。	0.900
手が震える。足がムズムズ	0.900
足のムズムズ。動かしたくなる	0.900
舌が気になる	0.800
食べ物を口からこぼしてしまう	0.778
自転車の乗り始めがふらつく	0.778
考えがまとまらないから歩きにくい	0.700
ふらふらすることがある	0.600
幻聴するとそのストレスで顎が動く	0.600
骨折の後の左足が重たい。	0.333
左マヒ、杖使用のため	0.222
電車の音が怖いです	0.200
手先のしびれ	0.200
時々目がまわる	0.111
他人から噂されている気がする	0.111
体のこわばり、頭痛、肩こり、背中痛み	0.000
歯が全然ないので話しにくい	0.000
ペットボトルを強く締めすぎる（強迫症）	0.000
頭が回らない	0.000
両ひざの痛み。	0.000
体がしんどい	0.000
汗が気になる	0.000
感情的になる	0.000

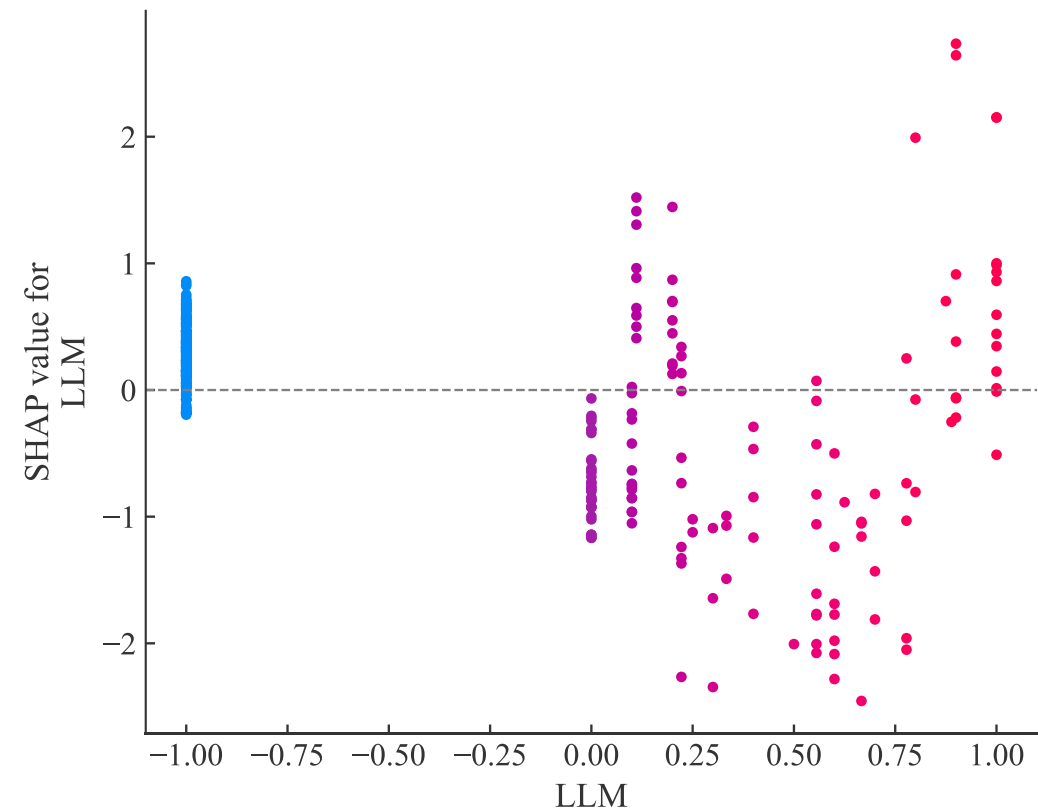


図2. LLMで症状を有すると解釈した割合, SHAP値の関係

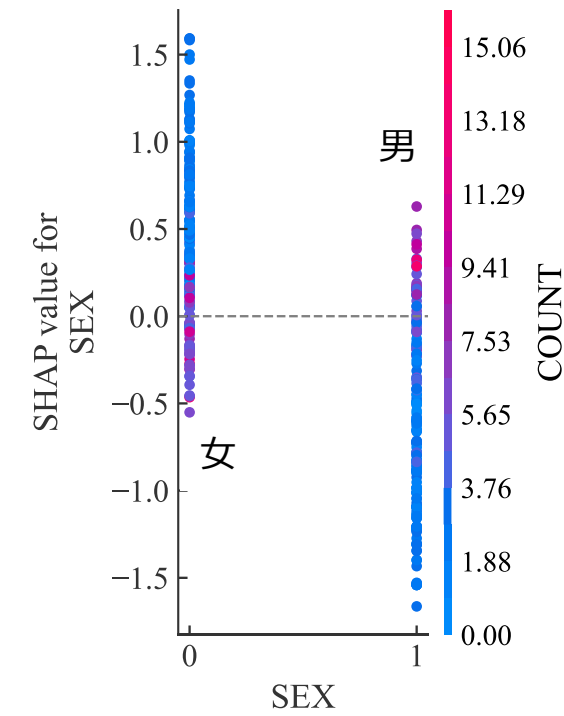
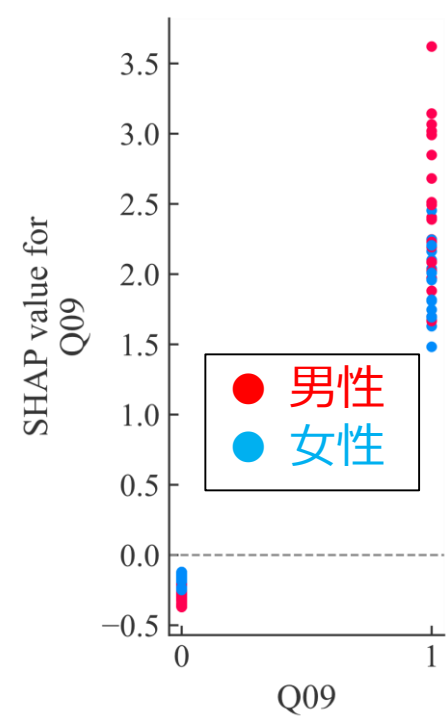


図3. Q9の回答, 性別, SHAP値 図4. 性別, 該当数, SHAP値の関係

図3, 図4; 各変数間の相互作用によりどのように予測に影響を与えているか分かる。
横軸が第1特徴量, (カラーバーが第2特徴量,)縦軸がSHAP値（予測への影響）
正の寄与でジスキネジア症状と予測しやすくなる

考察

- 表2: モデルのパフォーマンス

- 正答率: 92.9%, 感度: 34.3%, 陽性的中率: 81.4%, 特異度: 99.2%有する優れたパフォーマンス

- 図1: Beeswarm Plot

- 様々な特徴量が予測に影響
- Count, LLM, Q16_AG_DIFFなどは複雑な影響, SEX, Q09, Q10_AG_DIFFは分かりやすい
- 他の多くの質問、主成分分析、希少性スコア、cosine類似度を用いた特徴量は精度向上に寄与しなかった
 - 情報量などを考慮したものを使用したら寄与できる可能性

- 表3: 自由記載項目のLLM解釈による数値化

- LLM > 0.75 の文章は臨床的感覚としても**ジスキネジア症状を有している**と推察できる
 - 訴えに対する大まかなスクリーニングとして利用可能性

- 図2: **大規模言語モデル**で自由記載欄を解釈

- 0 ~ 0.75: **負の寄与** → ジスキネジア症状と関係ある記載は、
- 0.75 ~ 1: **正の寄与** ジスキネジア症状があると予測しやすい

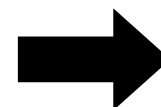
- 図3:

- 質問9に該当 × 男性: 強い**正の寄与**
- 質問9に該当 × 女性: 弱い**正の寄与**
- 質問9に非該当 × 男性・女性: 弱い**負の寄与**

複雑な相互作用があるとわかる。
解釈は困難。

- 図4:

- 男性 × 該当項目数 多: 弱い**正の寄与**
- 男性 × 該当項目数 少: 強い**負の寄与**
- 女性 × 該当項目数 多: 弱い**負の寄与**
- 女性 × 該当項目数 小: 強い**正の寄与**



女性の方が健康状態を低く評価する傾向[5]などを反映していると考えられる

まとめ・今後の展望

■まとめ

- 自記式アンケートからジスキネジア症状を予測する機械学習モデルを開発した
- 特徴量エンジニアリング、大規模言語モデルを用いることにより高精度を達成した
- SHAPによるモデル解釈性も十分有している
- 当院PCに本モデルの導入を行い、臨床に役立てている
 - オフライン中古PCでも高速に動作する(1件あたり2分以内)。ソフトのライセンス料なども不要

■今後の展望

- 本モデルの更なる検証・改良を行う
- 今後、他施設においても本モデルの有効性について検証する

■参考文献

- [1] L. Zhang *et al.*, BMC Psychiatry **24**:581 (2024)
- [2] M. Blankers *et al.*, BMC Med. Inform. Decis. Mak. **20**:332 (2020)
- [3] D. Chicco *et al.*, IEEE Access **9**:78368 (2021)
- [4] 厚生労働省, <https://www.mhlw.go.jp/topics/2006/11/dl/tp1122-1c22.pdf>, (2025/06/03参照)
- [5] 男女共同参画局, 令和5年度男女の健康意識に関する調査報告書

■謝辞

量子物性、感性工学の専門家であるヤンマー(株)吉岡直人博士には評価指標の工夫等について有意義なアドバイスをいただきました。

■連絡先

林 将人

mdhayashim@gmail.com

<https://github.com/mdhayashim>



mail



website

(本研究・発表は布袋病院倫理委員会の承認を得ている。本発表は、当院で既に作成されている匿名加工情報のみを利用しており、特に倫理的配慮を要しない。)

(COIなし)