

特集 2 計算論的精神医学の新展開：診断，病態解明，そして早期予測への革新的アプローチ

2. 機械学習によるうつ病の早期予測

陳 冲*, 中川 伸*

抄録：うつ病は労働損失や生活習慣病，自殺リスクの増加といった深刻な影響を与え，次世代にまで影響が及ぶ可能性も指摘されていることから，早期予測と介入の重要性が高まっている。機械学習の進展に伴い，予測精度の向上が期待され，うつ病の鑑別，治療反応，再発リスク予測への応用が進んでいる。本稿では，うつ病の早期予測に関する機械学習研究の最新動向と課題をレビューする。これまで多くの研究が実施されているが，新規発症予測の精度は依然として低い。また，うつ病の発症メカニズムにかかわる情動や報酬系の異常を評価する指標と，従来の人口統計情報や心理社会要因を統合的に活用するアプローチが不足しており，少数の特徴量で高精度な予測を行うことが今後の課題となっている。さらに，AI チューニングも十分とはいえず，効率的なベイズ最適化の導入が求められる。加えて，うつ病の罹患率による不均衡データの問題があり，それに対応した PR 曲線や F1 スコアの活用・最適化が必要である。こうした課題を踏まえ，実用的な予測モデル構築の展望を考察する。

日本生物学的精神医学会誌 36 (1) : 31-39, 2025

Key words : machine learning, artificial intelligence, depressive disorders, postpartum depression, early prediction, disease onset

うつ病の生涯有病率は6%と推定されており，これは約16～17人に1人が一生のうちうつ病を経験することを示している²²⁾。特に若年層の有病率が高く，たとえば出産を経験する女性の産後うつ病の罹患率は12～24%にのぼり，一般成人の2～4倍にも達することが報告されている^{9, 39)}。うつ病が発症すると，心理的苦痛に加え，動機づけや認知機能の低下により欠勤や生産性の低下といった社会的損失が生じやすくなる⁸⁾。うつ病により年間約38日分の労働損失が発生すると推定されている³⁴⁾。また，うつ病は糖尿病や高血圧といった生活習慣病のリスクを高めるだけでなく¹⁷⁾，自殺の最大リスク因子でもあり¹²⁾，患者の健康全体や生命に深刻な影響を及ぼす。さらに，産後うつ病は母子ボンディング障害や児童虐待のリスクを高め⁷⁾，育児行動や児童の発達に深刻な影響を与える可能性があり，次世代への波及も懸念されている⁵⁾。

こうした深刻な影響にもかかわらず，うつ病患者の約半数は初期治療に反応せず，治療で一度寛解し

ても再発を繰り返すケースが多いと報告されている^{23, 33)}。そのため，うつ病の発症前段階でリスクを正確に予測し，ハイリスクの個人に早期介入を行うことが重要である。多くの研究で，ハイリスク者への早期介入が発症予防に有効であることが示されている¹¹⁾。しかし，これを実現するためには，リスク予測の精度をさらに高める必要があり，そこで機械学習の利用が注目されている。

機械学習は優れた分類性能をもち，近年の進展により，精神医学の分野でもうつ病患者と健常者の識別（バイオマーカー探索研究を含む）や，抗うつ薬に対する治療反応性の予測，さらには予後や再発，自殺のリスク予測など，幅広い応用が進んできた^{31, 42)}。本稿では，これまでの機械学習アルゴリズムを用いたうつ病の早期予測に関する研究について，最新の動向や課題をレビューし，今後の展望を考察する。

Early prediction of depression via machine learning

*山口大学大学院医学系研究科 高次脳機能病態学講座 (〒755-8505 山口県宇部市南小串1-1-1) Chong Chen, Shin Nakagawa : Division of Neuropsychiatry, Department of Neuroscience, Yamaguchi University Graduate School of Medicine, 1-1-1 Minami-Kogushi, Ube, Yamaguchi 755-8505, Japan

【陳 冲 E-mail : cchen@yamaguchi-u.ac.jp】

1. 機械学習の基礎

機械学習はデータからモデルを学習し、新しいデータの予測に利用する技術である。主な流れは以下のとおりである。

1. データの収集と前処理：データを収集し、欠損値の補完や異常値の除去、標準化などの前処理を行い、データ品質を向上させる。
2. データ分割：データを訓練データ（学習用）とテストデータ（評価用）に分ける。必要に応じて、検証データを用いて交差検証を行い、モデルの性能を確認する。
3. 特徴量エンジニアリングと選択：予測に役立つ特徴量を抽出・生成し、重要な特徴量のみを選んで計算負荷や過学習を抑える。
4. モデル選択：ロジスティック回帰や決定木、ランダムフォレスト、サポートベクターマシン (SVM)、ニューラルネットワークなどさまざまなモデルから、データの特性や目的に適したものを選ぶ。複数のモデルを比較検証するのが一般的である。
5. モデルチューニング（ハイパーパラメータの調整）：グリッドサーチやランダムサーチ、ベイズ最適化などを使い、学習率や木の深さなど、モデル構造や学習過程に影響するハイパーパラメータを調整してモデルの性能を向上させる。
6. モデル学習：最適なハイパーパラメータを用いて訓練データを使ってモデルのパラメータ（特徴量の重みなど）を調整し、予測精度を最適化する。
7. モデル評価と改善：テストデータで性能を評価し、必要に応じて特徴量やパラメータを調整してさらなる改善を行う。

2. 文献検索

これまでの機械学習アルゴリズムを用いたうつ病の早期予測研究について、PubMed を用いて文献検索を実施した。検索は 2024 年 10 月に行い、以下のキーワードを用いた。また、総説論文^{29, 31, 42, 46)}も参考にし、関連文献を抽出している。

検索キーワード：depress * AND (cohort OR longitudinal OR prospective OR follow-up OR predict) AND (machine learning OR artificial intelligence OR decision tree OR random forest OR gradient boost OR xgboost OR support vector machine OR neural network)

論文の選択基準としては、以下の条件を満たすものとした。1) 特定の集団を対象とした追跡研究であり、前向き研究と後ろ向き研究のどちらでもよい。2) 前調査時点でのデータを用い、将来のうつ病あるいはうつ症状を予測すること。3) うつ病やうつ症状が明確に定義されていること。4) 機械学習アルゴリズムが用いられていること。

3. 予測研究の概要

文献検索の結果、合計 23 本の論文が選定された。各研究の概要（掲載年度、実施国、対象者の特徴、うつ病評価方法、予測因子、使用アルゴリズム、予測性能）は表 1 に示している。

研究数は 2019 年から 2024 年の間で 1 本、4 本、6 本、3 本、4 本、5 本と推移し、国別では米国が最多の 10 本、英国と中国がそれぞれ 3 本、日本、韓国、カナダ、スウェーデン、イスラエルが各 1 本、複数のヨーロッパ諸国を対象としたものが 2 本であった。対象者は妊婦が 12 本、青年期の若者が 5 本、一般成人が 4 本（高齢者を含む）、研修医が 1 本、そして心筋梗塞、乳がん、肝硬変患者を対象としたものが 1 本含まれる。うつ病やうつ症状の評価方法としては、自己評価（心理尺度が中心）、臨床診断、抗うつ薬の使用状況、非薬物療法の利用の 4 種類が用いられていた。

特徴量の数は 3 つから 60,000 以上と幅広く、データの種類もアンケート調査、電子健康記録 (electronic health record : EHR)、血液検査、血液細胞における遺伝子発現、睡眠ポリグラフ、脳画像、ウェアラブルデバイス計測と多岐にわたる。特に使用頻度が高かったアルゴリズムは、ランダムフォレスト (15 本)、ペナルティ付きロジスティック回帰 (14 本)、SVM (10 本)、決定木 (6 本)、XGBoost (6 本) であった。

4. 妊娠中および産後うつ病予測

妊婦を対象にした 12 本の研究のうち、1 本は妊娠 8 週前後の初回健診データを基に、エジンバラ産後うつ病質問票 (EPDS) で評価した妊娠中のうつ症状の予測を行った²⁴⁾。この米国の研究では、944 名のうつ病既往歴のない妊婦を出産まで追跡し、人口統計や心理社会要因など 14 の特徴量を用いて、PC-KCI 法（変数間の因果関係を条件付き独立性テストに基づいて推定する）を用いたグラフ学習により、受信者動作特性曲線下の面積 (ROC AUC)

表 1 各研究の概要

	国	サンプルサイズ	追跡期間	うつ病評価	特微量の数	特微量の種類	特微量選択	アルゴリズム	データ分割	不均衡データ対応	チューニング	特微量重要度	ベストモデル評価結果
産後うつ病・妊娠中うつ病													
Wang et al 2019	米国	9,980	産後 1 年間	臨床診断	98	人口統計情報, 診断, 薬使用	単変量ロジスティック回帰	ridge logistic regression, decision tree, random forest, XGBoost, naive bayes, SVM	10 分割交差検証	オーバーサンプリング	10 分割交差検証	—	SVM + 98 特微量: ROC AUC 0.79 感度 0.894 特異度 0.580 SVM + 26 特微量: ROC AUC 0.76 感度 0.892 特異度 0.588 SVM + 人口統計情報: ROC AUC 0.60 感度 0.551 特異度 0.609; SVM + 妊娠初期情報: ROC AUC 0.66 感度 0.855 特異度 0.428; SVM + 妊娠初期+中期情報: ROC AUC 0.69 感度 0.908 特異度 0.307
Payne et al 2020	米国	185	産後 1 ヶ月間	EPDS	3	妊娠中 EPDS, DNA メチル化のバイオマーカー	—	SVM	学習 51 名分, 残りテスト	—	LOOCV	—	妊娠後期データ: ROC AUC 0.78 (0.64, 0.78) 精神科既往歴がない人で妊娠後期データ: ROC AUC 0.72 (0.56, 0.72)
Zhang et al 2020	中国	508	産後 6 週間	EPDS	25	人口統計情報, 社会環境, メンタルヘルス	専門家選択, random forest に基づくフィルタ特徴選択 (FFS-RF)	random forest, SVM	学習 75%, テスト 25%	—	—	—	SVM + 専門家選択 17 特微量: ROC AUC 0.81 感度 0.68 特異度 0.87 適合率 0.72 陰性的中率 0.84 SVM + FFS-RF 選択 7 特微量: ROC AUC 0.78 感度 0.69 特異度 0.83 適合率 0.68 陰性的中率 0.84
Amit et al 2021	英国	266,544	産後 1 年間	臨床診断, 抗うつ薬使用, 非薬物療法使用	69	EHR	—	XGBoost	学習 2/3, テスト 1/3	—	—	Shap, 上位 20 項目	AUC 0.732 (0.729, 0.735) 感度 0.535 (0.53, 0.541) 特異度 0.80 精神疾患の既往歴のない女性で AUC 0.666 (0.662, 0.67) 感度 0.413 (0.406, 0.42) 特異度 0.80
Andersson et al 2021	スウェーデン	4,313	産後 6 週間	EPDS	—	人口統計情報, ライフスタイル, 健康, 妊娠, 出産関連情報	—	ridge logistic regression, lasso logistic regression, gradient boosting, distributed random forests, extreme randomized trees, naive bayes, stacked ensemble	割合不明	オーバーサンプリング	交差検証	MDI, 上記 25 項目	extremely randomized trees: ROC AUC 0.81 感度 0.72 特異度 0.75 適合率 0.33 陰性的中率 0.94 うつ病既往歴のない人: ROC AUC 0.725 感度 0.518 特異度 0.765 適合率 0.130 陰性的中率 0.966
Hochman et al 2021	イスラエル	214,359	産後 1 年間	新規発症: 臨床診断, 抗うつ薬使用	156	EHR	—	XGBoost	学習 2008-2014 データ, テスト 2015 データ	—	学習データの 20% を検証用	Shap	65 特微量: ROC AUC 0.712 (0.69, 0.733) 9 特微量: ROC AUC 0.686 (0.662, 0.709)
Park et al 2021	米国	532,802	産後 2 ヶ月間	臨床診断, 抗うつ薬使用	71	人口統計情報, 妊娠, 健康関連情報	—	logistic regression, random forest, XGBoost	学習 50%, 検証 30%, テスト 20%	—	—	F score	XGBoost + 10 特微量: ROC AUC 0.730 感度 0.639 適合率 0.781 バランス正解率 0.730
Zhang et al 2021	米国	69,169	産後 1 年間	臨床診断, 抗うつ薬使用	32	EHR	SFS	regularized logistic regression, decision tree, XGBoost, multilayer perceptron	学習 15,197 (1 施設) テスト 53,972 (多施設)	—	10 分割交差検証, grid search	—	MLP + 出産までのデータ: ROC AUC 0.887 (0.880, 0.894) 感度 0.66 特異度 0.88 適合率 0.28 陰性的中率 0.97 MLP + 妊娠 12 週目までのデータ: ROC AUC 0.812 (0.803, 0.821) 感度 0.65 特異度 0.87 適合率 0.26 陰性的中率 0.97
Matsuo et al 2022	日本	10,013	産後 1 ヶ月間	EPDS	41	人口統計情報, 妊娠, 出産, 新生児, 産後情報	—	logistic regression, ridge logistic regression, elastic net, SVM, random forest	学習 70%, テスト 30%	—	10 分割交差検証	coefficients and odds ratios	ridge logistic regression + 退院まで 38 特微量: ROC AUC 0.630 感度 0.727 特異度 0.493 適合率 0.078 陰性的中率 0.968 ridge logistic regression + 産後 2 週間検診まで 41 特微量: ROC AUC 0.702 感度 0.506 特異度 0.810 適合率 0.136 陰性的中率 0.965
Garbaza et al 2024	イタリア, スイス	376	妊娠中期から産後半年間	EPDS	124	人口統計情報, 血液データ, 医療データ, 睡眠検査 (polysomnography)	permutation importance	penalized logistic regression, random forest, SVM	3 分割交差検証, 5 回実施	—	入れ子構造の 4 分割交差検証 (nested 4-fold CV), grid search	Shap	SVM + 妊娠初期の 10 特微量: ROC AUC 0.774 ± 0.053 PRAUC 0.388 ± 0.084 感度 0.528 ± 0.121 特異度 0.826 ± 0.023

	国	サンプルサイズ	追跡期間	うつ病評価	特徴量の数	特徴量の種類	特徴量選択	アルゴリズム	データ分割	不均衡データ対応	チューニング	特徴量重要度	ベストモデル評価結果
Krishnamurti et al 2024	米国	944 (うつ病既往歴なし)	妊娠中	EPDS	55	アプリによって収集したデータ: 人口統計情報, 既往歴, 心理社会, 妊娠関連情報	—	lasso logistic regression, forward stepwise selection, shallow decision tree, random forest, XGBoost, kernel-based conditional independence (PC-KCI)	学習 80%, テスト 20%	オーバーサンプリング	5 分割交差検証	—	PC-KCI + 初回産前健診 (妊娠 8 週 ± 2 週) のうつ病: ROC AUC 0.89, 感度 0.81, 特異度 0.83
Wang et al 2024	中国	3,174	産後 1-6 カ月間	EPDS	—	EHR	—	random forest, k-nearest neighbor, SVM, decision tree	学習 70%, テスト 30%	—	—	Gini importance, 上位 15 項目	random forest : 産後 1 ヶ月 AUC 0.856 F1 0.610 適合率 0.667 産後 3 ヶ月 AUC 0.909 F1 0.741 適合率 0.731 産後 6 ヶ月 AUC 0.917 F1 0.769 適合率 0.757
青年期の若者													
Ho et al 2022	米国	7,995	1 年間	保護者による報告: Child behavior checklist	—	人口統計情報, 環境, 構造と安静時 fMRI	—	elastic net regression, gradient boosting	学習 75%, テスト 25%	—	10 分割交差検証, grid search	Shap, 上位 10 項目	elastic net regression : MAE 4.255 R ² 0.103 gradient boosting : MAE 4.262 R ² 0.089
Toenders et al 2022	フランス, ドイツ, アイルランド, 英国	544	5 年間	新規発症: 自己評価 DAWBA	145	臨床, 認知, 環境, 構造 MRI	—	penalized logistic regression	学習 75%, テスト 25%	—	10 分割交差検証	—	4 特徴量: ROC AUC 0.71 感度 0.50 特異度 0.81 5 特徴量: ROC AUC 0.68 感度 0.49 特異度 0.77 8 特徴量: ROC AUC 0.72 感度 0.51 特異度 0.83
Hawes et al 2023	米国	374	3-12 年間	診断面接 (DSM-4/5) at age 15	429	自己および保護者による報告: 気質・人格, 環境要因, 生物学的・神経認知的要因 (例: コルチゾール, 情動課題時脳波, 実行機能), 臨床的特徴	—	logistic regression with and without ridge regularization, random forest, neural networks	10 分割交差検証	—	10 分割交差検証	—	3 歳時データ: random forest ROC AUC 0.599 6 歳時データ: いずれもチャンネルレベル 9 歳時データ: neural networks, ROC AUC 0.675 12 歳時データ: ridge logistic regression, ROC AUC 0.745
Gracia-Tabuenca et al 2024	米国	2,658	2 年間	診断面接 KSADS-5	構造 MRI1,196 特徴量, DTTI,140 特徴量, task-fMRI2,548 特徴量, 安静時 fMRI61,776 特徴量	脳画像データ: 構造 MRI, diffusion tensor imaging (DTI), and task and rest fMRI	principle components analysis (75 主成分); ANOVA	elastic net regression	LOOCV	ROSE	Bootstrap	Cohen's d	各種脳画像データ ROC AUC 0.58-0.62 ハイリスク者 (親がうつ病既往歴あり): 安静時 fMRI + ANOVA による特徴量 ROC AUC 0.72
Yoo et al 2024	英国	1,799	2-8 年間	自己評価 Short mood and feelings questionnaire	266	認知, 社会要因, 感情, 生物学的要因	SBFS	decision tree, gaussian naive bayes, multinomial naive bayes, SVM, adaboost, random forest, MLP	学習 80%, テスト	SMOTE	5 分割交差検証, grid search	—	adaboost + 14 特徴量: F1 0.538 感度 0.8 特異度 0.374 適合率 0.405 陰性的中率 0.779
一般成人													
Na et al 2020	韓国	6,588	2 年間	CES-D	24	人口統計情報, 健康, 経済, 家庭関連情報	—	random forest	学習 70%, テスト 30%	SMOTE	10 分割交差検証	RF based feature importance	ROC AUC 0.870 感度 0.730 特異度 0.866 適合率 0.152 陰性的中率 0.990
Gahiouline et al 2023	英国	906	3 年間	新規発症: 臨床診断, 抗うつ薬使用, 自己評価	3,025	安静時 fMRI	—	logistic regression, decision tree, random forest, gradient boosting, adaboost, naive bayes, k-nearest neighbors, SVM, neural network, quadratic discriminant analysis	入れ子構造の交差検証	—	デフォルト設定または grid search	Shap	SVM + 動的因果モデリングによる指標: AUC 0.64 F1 0.60 感度 0.56 適合率 0.63
Song et al 2023	カナダ	19,024	3 年間	新規発症: 自己申告した診断, CES-D, 抗うつ薬使用	2,529	人口統計情報, 心理尺度	—	ridge logistic regression, random forest, balanced random forest	学習 80%, テスト 20%	—	—	Shap	balanced random forest : ROC AUC 0.791 ± 0.016 感度 0.750 ± 0.031 特異度 0.689 ± 0.010 バランス正解率 0.720 ± 0.014
Su et al 2021	中国	1,538	2 年間	自己評価	18	人口統計情報, 健康リスク因子情報	—	logistic regression with and without lasso regularization, random forest, gradient boosting, SVM, deep neural network, recurrent neural network (RNN) based on long short-term memory	学習 70%, テスト 30%	—	—	—	logistic regression with lasso regularization : ROC AUC 0.629 感度 0.623 特異度 0.680 適合率 0.281 陰性的中率 0.900
研修医													
Horwitz et al 2023	米国	2,459	92 日間	PHQ-9	33	気分日記, Fitbit 測定	—	elastic net regression, random forest	入れ子構造の交差検証	—	3 × 3 grid search	—	最初 2 週間データ + elastic net regression : AUC > 0.70

	国	サンプルサイズ	追跡期間	うつ病評価	特徴量の数	特徴量の種類	特徴量選択	アルゴリズム	データ分割	不均衡データ対応	チューニング	特徴量重要度	ベストモデル評価結果
患者 (心筋梗塞, 乳がん, 肝硬変)													
Meng et al 2020	米国	10,148	1年間	新規発症: 診断面接, 抗うつ薬使用	9,285	EHR, 人口統計情報, 医療メモ (トピック分析による自然言語処理)	—	lasso logistic regression, SVM, MLP, random forest (with or without variational auto encoder), MiME, HCET	学習 70%, 検証 10%, テスト 20%	—	Adam	attention weights	HCET + 2 週間前までのデータ: ROC AUC 0.81 PR AUC 0.73 1 年前までのデータ: ROC AUC 0.78 PR AUC 0.67

— : 不明; Adam : adaptive moment estimation, adaboost : adaptive boosting, ANOVA : analysis of variance, CES-D : Center for epidemiologic studies depression scale, DAWBA : development and well-being assessment, EHR : electronic health record, EPDS : Edinburgh postnatal depression scale, HCET : hierarchical clinical embeddings combined with topic modeling, KSADS-5 : Kiddie schedule for affective disorders and schizophrenia, LOOCV : leave one out cross validation, MAE : mean absolute error, MDI : mean decrease in impurity, MiME : multilevel medical embedding, MLP : multilayer perceptron, PHQ-9 : Patient health questionnaire-9, ROSE : randomly over sampling examples, SBFS : sequential backward feature selection, SFS : sequential forward selection, Shap : shapley additive explanation, SMOTE : synthetic minority over-sampling technique, SVM : support vector machine, XGBoost : extreme gradient boosting (Amit G, et al : BMC Pregnancy and Childbirth, 21 : 630. 2021²⁾, Andersson S, et al : Sci Rep, 11 (1) : 7877. 2021³⁾, Galiouline H, et al : Neuroimage, 273 : 119986. 2023¹³⁾, Garbaza C, et al : Psychiatry Res, 337 : 115957. 2024¹⁴⁾, Gracia-Tabuenca Z, et al : Neuroimage Clin, 42 : 103604¹⁵⁾, Hawes MT, et al : Psychol Med, 53 (13) : 6205-6211. 2023¹⁸⁾, Ho TC, et al : J Child Psychol Psychiatry, 63 (12) : 1523-1533. 2022¹⁹⁾, Hochman E, et al : Depress Anxiety, 38 (4) : 400-411. 2021²⁰⁾, Horwitz AG, et al : Psychol Med, 53 (12) : 5778-5785. 2023²¹⁾, Krishnamurti T, et al : Arch Womens Ment Health, 27 (6) : 1019-1031. 2024²⁴⁾, Matsuo S, et al : J Obst Gynaecol Res, 48 (7) : 1775-1785. 2022²⁶⁾, Meng Y, et al : IEEE J Biomed Health Inform, 25 (4) : 1265-1272. 2020²⁷⁾, Na KS, et al : Neurosci Lett, 721 : 134804. 2020²⁸⁾, Park Y, et al : JAMA Netw Open, 4 (4) : e213909. 2021³⁰⁾, Payne JL, et al : Psychiatry Res, 285 : 112711. 2020³²⁾, Saito T, et al : PLoS one, 10 (3) : e0118432. 2015³⁵⁾, Song Y, et al : J Affect Disord, 339 : 52-57. 2023³⁶⁾, Su D, et al : J Affect Disord, 282 : 289-298. 2021³⁷⁾, Toenders YJ, et al : Biol Psychiatry Cogn Neurosci Neuroimaging, 7 (4) : 376-384. 2022³⁸⁾, Wang S, et al : Stud Health Technol Inform, 264 : 888-892. 2019⁴⁰⁾, Wang Y, et al : J Affect Disord, 365 : 185-192. 2024⁴¹⁾, Yoo A, et al : Sci Rep, 14 (1) : 23282. 2024⁴³⁾, Zhang W, et al : JMIR Med Inform, 8 (4) : e15516. 2020⁴⁴⁾, Zhang Y, et al : J Affect Disord, 279 : 1-8. 2021⁴⁵⁾ を元に作成)

0.89, 感度 (再現率ともよばれる) 0.81, 特異度 0.83 の精度を達成した。

残りの 11 本は産後 1 ヶ月から 1 年以内のうつ病予測を対象とし, 妊娠中や出産時に収集されたデータを活用している。4 本は EHR データ^{2, 20, 41, 45)}, 2 本は血液由来のデータ (例: DNA メチル化)^{14, 32)}, 1 本は睡眠データ¹⁴⁾, 残りはアンケート調査^{3, 24, 26, 30, 40, 44)}によるものであった。もっとも高い予測精度は, 中国の 3,174 名を対象とした EHR データによる研究⁴¹⁾で, ランダムフォレストを用いて ROC AUC 0.917, F1 0.769, 適合率 0.757 を達成した (使用する特徴量の数は不明)。また, 米国の 69,169 名のデータを用いた研究⁴⁵⁾では, 32 の特徴量を使った多層パーセプトロン (順伝播型ニューラルネットワークの一つ)により, AUC 0.887, 感度 0.66, 特異度 0.88, 適合率 0.28, 陰性的中率 0.97 を記録している。

一方, うつ病や精神疾患の既往歴がない人や, 産後の新規発症 (妊娠中の評価で症状なし) のみを対象とした研究では予測精度は低く^{2, 3, 32)}, たとえば, イスラエルの 214,359 名を対象にした EHR データを使った研究²⁰⁾では, XGBoost により 65 特徴量で ROC AUC 0.712, 9 特徴量で ROC AUC 0.686 という結果であった。

5. 青年期のうつ病予測

青年期を対象とした 5 本の研究では, 追跡期間が 1 ~ 15 年に及び, 4 本が脳画像データ (構造 MRI や fMRI, 脳波)^{15, 18, 19, 38)}, 2 本が血中コルチゾールなどの生物学的データを使用していた^{18, 43)}。脳画像データのみの予測精度は低く, たとえば, 米国 2,658 名の青年期若者を 2 年間追跡した研究¹⁵⁾では, 構造 MRI や fMRI とエラスティックネット (ラッソ回帰とリッジ回帰の L1 正則化と L2 正則化を組み合わせたペナルティ項付きロジスティック回帰)を使った予測で ROC AUC 0.58 ~ 0.62 に留まった。一方, 親にうつ病歴があるハイリスク群に限定すると, 予測精度が向上し, ROC AUC は 0.72 となった。

予測精度は追跡期間が長くなるほど低下する傾向がある。米国 374 名青年期若者を対象とした研究¹⁸⁾では, 15 歳時のうつ病を予測するために, 自己や保護者から収集した気質・パーソナリティ (例: 行動抑制, 反すう思考の傾向), 環境要因 (例: ライフイベント, 学校でのいじめ経験), 生物学的要因 (例: 血中コルチゾール), 神経認知的要因 (例: 情動課題時の脳波, 実行機能, 注意バイアス), 臨床的特徴 (精神疾患の症状評価), および親のメンタルヘルスやパーソナリティなど, 幅広い情報が使用された。分析には, ペナルティ付きおよびなしのロジスティック回帰, ランダムフォレスト, ニューラ

ルネットワークが用いられている。なお、3歳と6歳時のデータのみを使用した場合、ROC AUCは0.6以下あるいは偶然と同等のレベルに留まったが、9歳と12歳のデータを使用するとROC AUCがそれぞれ0.675と0.745に向上した。

また、新規発症の予測を行った研究³⁸⁾では、ヨーロッパの4カ国から544名を5年間追跡し、ペナルティ付きロジスティック回帰を用いて分析した。この研究では、ライフイベントや神経症傾向、学校で受けたいじめ、言語理解力、性別、構造MRIによる縁上回表面積など8つの特徴量を用いて、ROC AUC 0.72、感度0.51、特異度0.83の精度が得られた。

6. 一般成人のうつ病予測

高齢者を含む成人を対象としたうつ病予測研究は、2～3年の追跡期間で実施されている。3本の研究はアンケートデータから18～2,529の特徴量を使用し^{28, 36, 37)}、1本は安静時fMRIから3,025の特徴量を用いていた¹³⁾。

韓国の6,588名成人を対象とした研究²⁸⁾では、24の特徴量を用いたランダムフォレストでROC AUC 0.870、感度0.730、特異度0.866、適合率0.152、陰性的中率0.990の精度を達成した。一方、中国の1,538名高齢者を対象とした研究³⁷⁾では、18の特徴量とペナルティ付きロジスティック回帰を用い、ROC AUC 0.629、感度0.623、特異度0.680、適合率0.281、陰性的中率0.900を得ている。

カナダの研究³⁶⁾では、19,024名の中高齢者を対象に人口統計情報や心理尺度から抽出した2,529の特徴量 balanced random forest で新規発症うつ病を予測し、ROC AUC 0.791 ± 0.016 、感度 0.750 ± 0.031 、特異度 0.689 ± 0.010 、バランス精度 0.720 ± 0.014 を記録した。英国の研究¹³⁾は、906名成人を対象に安静時fMRIデータのみを用いて3年間追跡し、新規発症うつ病の予測においてAUC 0.64、F1スコア0.60、感度0.56、適合率0.63を得ている。

7. 特殊集団におけるうつ病予測

研修医と特定の患者を対象としたうつ病予測研究がそれぞれ1本ずつあった。米国の初年度研修医2,459名を92日間追跡した研究²¹⁾では、最初の2週間で収集した気分日記やFitbitデータから33の特徴量を抽出し、エラスティックネットでこころとからだの質問票 (Patient health questionnaire-9 :

PHQ-9) によるうつ症状をAUC > 0.70の精度で予測した。

一方、米国の心筋梗塞、乳がん、肝硬変患者10,148名を対象とした研究²⁷⁾では、EHRや人口統計情報、医療メモから抽出した9,285の特徴量を使用し、HCETとよばれる階層型臨床エンベディングとトピック分析による自然言語処理の組み合わせモデルで新規発症のうつ病を予測した。診断の1年前のデータでROC AUC 0.78、PR AUC 0.67、2週間前のデータでROC AUC 0.81、PR AUC 0.73を達成した。

8. 考察

多くの予測研究が実施されているが、新規発症の早期予測精度は依然として低く (例: 大半の研究でROC AUC ≤ 0.72)、数千の特徴量を使用してもわずかな向上に留まり (ROC AUC 0.791³⁶⁾)、公衆衛生現場での実用性には課題が残る。また、家族歴や慢性疾患をもつハイリスク群を対象にすると予測精度が向上することが示されており^{15, 27)}、こうしたハイリスク者に向けた実用的な予測モデルの開発が求められる。さらに、追跡期間が長くなるほど精度が低下する傾向があり¹⁸⁾、定期的なモニタリングの導入が必要とされている。

以下では、これまでの研究における課題、特に機械学習の観点から考察する。

9. 課題

a. 特徴量の質

これまで、人口統計や既往歴、心理社会要因、生物学的要因などさまざまなデータが用いられてきたが、これらを総合的に考慮したアプローチは依然少ないのが現状である^{18, 43)}。予測性能はデータの質に依存するため、うつ病の病態や発症メカニズムに関するデータの収集が重要である。たとえば、うつ病は情動や報酬系の異常との関連が示されており^{10, 16)}、情動課題²⁵⁾や行動経済学的意思決定課題、計算論的モデリング⁶⁾を通じた指標が予測精度向上に寄与する可能性がある。今回レビューした論文の中では、青年期のうつ病予測を行ったHawes et al (2023)¹⁸⁾のみが情動課題を用いており、このような指標を取り入れる研究は限定的である。また、今後は、データ収集のコストも考慮し、少数の特徴量で高精度な予測が可能なモデルの開発が求められる。

b. AI チューニング

レビュー対象の 23 研究のうち、AI チューニング手法を明示していたのは 7 本 (30%) のみで、そのうち 6 本はグリッドサーチ^{13, 14, 19, 21, 43, 45)}、1 本は適応的モーメント推定 (Adam 法) を使用していた²⁷⁾。グリッドサーチはハイパーパラメータを網羅的に探索する方法であるが、多くの研究で組み合わせの範囲が限られており (例: Horwitz et al, 2023²¹⁾ では 9 通り)、効果が限定的であった。今後は、探索効率の高いベイズ最適化の利用が期待される。たとえば、わが国の Preferred Networks 社が開発した OptunaTM は、探索履歴を基に tree-structured parzen estimator というアルゴリズムにより最適解に近いパラメータを効率的に見つけることができる¹⁾。

c. 不均衡データと性能評価

うつ病の罹患率データが不均衡で、ポジティブクラスが少数派となる状況がみられる。このような場合、モデルの最適化指標として ROC AUC を主に用いると限界がある。たとえば ROC AUC や感度、特異度が高くても、適合率が低いケースが多くみられる。レビューした 16 モデルのうち 7 モデルでは適合率が 0.3 以下であり、たとえば Matsuo et al (2022) の産後 1 ヶ月のうつ病予測では適合率が 0.078 と低く、予測で陽性と判断された 13 人のうち実際に陽性であったのは 1 人だけであった²⁶⁾。偽陽性が心理社会的および医療的に悪影響を及ぼす可能性があること⁴⁾ を考慮し、今後は適合率-再現率曲線 (precision-recall curve: PR 曲線) や F1 スコア (適合率と再現率の調和平均) といった指標の活用³⁵⁾、あるいは従来の ROC AUC 指標との併用が望まれる。

こうした課題に対応し、実用的かつ精度の高い予測モデルの開発により、公衆衛生現場でのうつ病の発症予防への貢献が期待される。

本論文に記載した筆者らの研究に関してすべて倫理的配慮を行っている。開示すべき利益相反は存在しない。

文 献

1) Akiba T, Sano S, Yanase T, et al (2019) Optuna: a next-generation hyperparameter optimization framework. KDD'19: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery and data mining, 2623-2631.

2) Amit G, Girshovitz I, Marcus K, et al (2021) Estimation of postpartum depression risk from electronic

health records using machine learning. BMC Pregnancy and Childbirth, 21: 630.

3) Andersson S, Bathula DR, Iliadis SI, et al (2021) Predicting women with depressive symptoms postpartum with machine learning methods. Sci Rep, 11 (1): 7877.

4) Belsher BE, Smolenski DJ, Pruitt LD, et al (2019) Prediction models for suicide attempts and deaths: a systematic review and simulation. JAMA psychiatry, 76 (6): 642-651.

5) Bhutta ZA, Bhavnani S, Betancourt TS, et al (2023) Adverse childhood experiences and lifelong health. Nat Med, 29 (7): 1639-1648.

6) Chen C, Mochizuki Y, Hagiwara K, et al (2022) Computational markers of experience-but not description-based decision-making are associated with future depressive symptoms in young adults. J Psychiatr Res, 154: 307-314.

7) Chen C, Mochizuki Y, Okawa S, et al (2024) Postpartum loneliness predicts future depressive symptoms: a nationwide Japanese longitudinal study. Arch Womens Ment Health, 27 (3) 447-457.

8) Chen C, Okubo R, Hagiwara K, et al (2024) The association of positive emotions with absenteeism and presenteeism in Japanese workers. J Affect Disord, 344: 319-324.

9) Chen C, Okubo R, Okawa S, et al (2023) The prevalence and risk factors of suicidal ideation in women with and without postpartum depression. J Affect Disord, 340: 427-434.

10) Chen C, Takahashi T, Nakagawa S, et al (2015) Reinforcement learning in depression: a review of computational research. Neurosci Biobehav Rev, 55: 247-267.

11) Cuijpers P, Pineda BS, Quero S, et al (2021) Psychological interventions to prevent the onset of depressive disorders: a meta-analysis of randomized controlled trials. Clin Psychol Rev, 83: 101955.

12) Favril L, Yu R, Uyar A, et al (2022) Risk factors for suicide in adults: systematic review and meta-analysis of psychological autopsy studies. Evid Based Ment Health, 25 (4): 148-155.

13) Galioulline H, Frässle S, Harrison SJ, et al (2023) Predicting future depressive episodes from resting-state fMRI with generative embedding. Neuroimage, 273: 119986.

14) Garbazza C, Mangili F, D'Onofrio TA, et al (2024) A

- machine learning model to predict the risk of perinatal depression : psychosocial and sleep-related factors in the Life-ON study cohort. *Psychiatry Res*, 337 : 115957.
- 15) Gracia-Tabuenca Z, Barbeau EB, Xia Y, et al (2024) Predicting depression risk in early adolescence via multimodal brain imaging. *Neuroimage Clin*, 42 : 103604.
- 16) 萩原康輔, 陳冲, 中川伸 (2020) 情報処理理論からみたうつ病における認知障害 : hot vs. cold cognition. *精神科 = Psychiatry*, 37 (3) : 291-296.
- 17) 萩原康輔, 陳冲, 中川伸 (2023) うつ病患者における生活習慣病の予防と改善について. *精神科治療学*, 38 (9) : 1029-1034.
- 18) Hawes MT, Schwartz HA, Son Y, et al (2023) Predicting adolescent depression and anxiety from multi-wave longitudinal data using machine learning. *Psychol Med*, 53 (13) : 6205-6211.
- 19) Ho TC, Shah R, Mishra J, et al (2022) Multi - level predictors of depression symptoms in the adolescent brain cognitive development (ABCD) study. *J Child Psychol Psychiatry*, 63 (12) : 1523-1533.
- 20) Hochman E, Feldman B, Weizman A, et al (2021) Development and validation of a machine learning - based postpartum depression prediction model : a nationwide cohort study. *Depress Anxiety*, 38 (4) : 400-411.
- 21) Horwitz AG, Kentopp SD, Cleary J, et al (2023) Using machine learning with intensive longitudinal data to predict depression and suicidal ideation among medical interns over time. *Psychol Med*, 53 (12) : 5778-5785.
- 22) Ishikawa H, Tachimori H, Takeshima T, et al (2018) Prevalence, treatment, and the correlates of common mental disorders in the mid 2010's in Japan : the results of the world mental health Japan 2nd survey. *J Affect Disord*, 241 : 554-562.
- 23) Kessler RC and Bromet EJ (2013) The epidemiology of depression across cultures. *Annu Rev Public Health*. 34 (1) : 119-138.
- 24) Krishnamurti T, Rodriguez S, Wilder B, et al (2024) Predicting first time depression onset in pregnancy : applying machine learning methods to patient-reported data. *Arch Womens Ment Health*, 27 (6) : 1019-1031.
- 25) Matsubara T, Chen C, Hirotsu M, et al (2022) Prefrontal cortex activities during verbal fluency and emotional words tasks in major depressive, adjustment, and bipolar disorders with depressive states. *J Affect Disord*, 316 : 109-117.
- 26) Matsuo S, Ushida T, Emoto R, et al (2022) Machine learning prediction models for postpartum depression : a multicenter study in Japan. *J Obstet Gynaecol Res*, 48 (7) : 1775-1785.
- 27) Meng Y, Speier W, Ong M, et al (2020) HCET : hierarchical clinical embedding with topic modeling on electronic health records for predicting future depression. *IEEE J Biomed Health Inform*, 25 (4) : 1265-1272.
- 28) Na KS, Cho SE, Geem ZW, et al (2020) Predicting future onset of depression among community dwelling adults in the Republic of Korea using a machine learning algorithm. *Neurosci Lett*, 721 : 134804.
- 29) Nickson D, Meyer C, Walasek L, et al (2023) Prediction and diagnosis of depression using machine learning with electronic health records data : a systematic review. *BMC Med Inform Decis Mak*, 23 (1) : 271.
- 30) Park Y, Hu J, Singh M, et al (2021) Comparison of methods to reduce bias from clinical prediction models of postpartum depression. *JAMA Netw Open*, 4 (4) : e213909.
- 31) Park Y, Park S, and Lee M (2024) Effectiveness of artificial intelligence in detecting and managing depressive disorders : Systematic review. *J Affect Disord*, 361 : 445-456.
- 32) Payne JL, Osborne LM, Cox O, et al (2020) DNA methylation biomarkers prospectively predict both antenatal and postpartum depression. *Psychiatry Res*, 285 : 112711.
- 33) Rush AJ, Trivedi MH, Wisniewski SR, et al (2006) Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps : a STAR* D report. *Am J Psychiatry*, 163 (11) : 1905-1917.
- 34) Sado M, Yamauchi K, Kawakami N, et al (2011) Cost of depression among adults in Japan in 2005. *Psychiatry Clin Neurosci*, 65 (5) : 442-450.
- 35) Saito T and Rehmsmeier M (2015) The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, 10 (3) : e0118432.
- 36) Song Y, Qian L, Sui J, et al (2023) Prediction of depression onset risk among middle-aged and elderly

- adults using machine learning and Canadian longitudinal study on aging cohort. *J Affect Disord*, 339 : 52-57.
- 37) Su D, Zhang X, He K, et al (2021) Use of machine learning approach to predict depression in the elderly in China : a longitudinal study. *J Affect Disord*, 282 : 289-298.
- 38) Toenders YJ, Kottaram A, Dinga R, et al (2022) Predicting depression onset in young people based on clinical, cognitive, environmental, and neurobiological data. *Biol Psychiatry Cogn Neurosci Neuroimaging*, 7 (4) : 376-384.
- 39) Tokumitsu K, Sugawara N, Maruo K, et al (2020) Prevalence of perinatal depression among Japanese women : a meta-analysis. *Ann Gen Psychiatry*, 19 : 41.
- 40) Wang S, Pathak J and Zhang Y (2019) Using electronic health records and machine learning to predict postpartum depression. *Stud Health Technol Inform*, 264 : 888-892.
- 41) Wang Y, Yan P, Wang G, et al (2024) Trajectory on postpartum depression of Chinese women and the risk prediction models : a machine-learning based three-wave follow-up research. *J Affect Disord*, 365 : 185-192.
- 42) Winter NR, Blanke J, Leenings R, et al (2024) A systematic evaluation of machine learning-based biomarkers for major depressive disorder. *JAMA psychiatry*, 81 (4) : 386-395.
- 43) Yoo A, Li F, Youn J, et al (2024) Prediction of adolescent depression from prenatal and childhood data from ALSPAC using machine learning. *Sci Rep*, 14 (1) : 23282.
- 44) Zhang W, Liu H, Silenzio VMB, et al (2020) Machine learning models for the prediction of postpartum depression : application and comparison based on a cohort study. *JMIR Med Inform*, 8 (4) : e15516.
- 45) Zhang Y, Wang S, Hermann A, et al (2021) Development and validation of a machine learning algorithm for predicting the risk of postpartum depression among pregnant women. *J Affect Disord*, 279 : 1-8.
- 46) Zhong M, Zhang H, Yu C, et al (2022) Application of machine learning in predicting the risk of postpartum depression : a systematic review. *J Affect Disord*, 318 : 364-379.

■ ABSTRACT**Early prediction of depression via machine learning**

Chong Chen, Shin Nakagawa

Division of Neuropsychiatry, Department of Neuroscience, Yamaguchi University Graduate School of Medicine

Depression has severe impacts, such as causing labor loss, increasing the risk of lifestyle-related diseases, and raising suicide risk, with potential effects extending to future generations. This underscores the importance of early prediction and intervention. Advances in machine learning are expected to improve prediction accuracy, with applications expanding to classifying depression, predicting treatment response, and assessing recurrence risk. This paper reviews recent trends and challenges in machine learning research on early prediction of depression. Despite numerous studies, the accuracy of predicting new onset of depression remains low. There is also a lack of approaches that comprehensively utilize indicators related to abnormalities in emotional and reward systems—mechanisms linked to the onset of depression—alongside traditional demographic and psychosocial factors. Achieving high accuracy with fewer features remains a challenge. Moreover, AI tuning is often insufficient, highlighting the need for efficient optimization methods such as Bayesian optimization. Additionally, the imbalance in depression prevalence necessitates the use and optimization of PR curves and F1 scores as evaluation metrics. This paper discusses the future prospects for building practical prediction models in light of these challenges.