

厚生労働科学研究費補助金

政策科学総合研究事業

(臨床研究等ICI基盤構築・人工知能実装研究事業)

様々なデータを用いた AI 解析によるうつ病の診断、  
重症度、反応性、層別化に関する実証研究

平成29年度～30年度 総合研究報告書

研究代表者 岡本 泰昌

令和元（2019）年 5月

# 目 次

## I. 総合研究報告

- 様々なデータを用いたA I解析によるうつ病の診断、重症度、反応性、層別化に関する実証研究 1  
岡本泰昌

## II. 分担研究報告

1. 脳MRIデータとA I解析によるうつ病診断の汎用性向上とデータ駆動型サブタイプ分類に関する研究 13  
岡本泰昌 吉本潤一郎 山脇成人 池田和隆 橋本 亮太
2. 表情、音声などのセンシングデータの人工知能解析 21  
岸本泰士郎
3. ライフログデータに基づく寛解期うつ病患者の再燃再発予測 26  
古川壽亮 吉本潤一郎
4. 抗うつ薬治験データの人工知能解析に関する研究 29  
丸尾和司 古川壽亮

平成29年度～30年度厚生労働科学研究費補助金（政策科学総合研究事業  
（臨床研究等 ICT 基盤構築・人工知能実装研究事業））総合研究報告書

様々なデータを用いたAI解析によるうつ病の診断、重症度、反応性、層別化に関する実証研究

研究代表者 岡本泰昌 広島大学大学院医歯薬保健学研究科教授

研究要旨

本研究は、脳画像、表情、音声などのバイオデータ、プラセボ対照の抗うつ薬臨床試験データといった様々な比較的大規模なデータセットを、複数のAIアルゴリズムを用いて解析することによって、各データセットに対する最適なAIアルゴリズムを特定し、解析パイプラインの提案を行うこと、また、一部のデータについては外部データに対する汎化性能を確認することで、AIを用いた解析の有用性の検証を目的とした。安静時fMRIデータを用いて、解析パイプライン用パラメータ、判別のための機械学習手法、施設間の測定バイアスの検討を行った。MRIデータに関する教師あり学習によるうつ病判別と外部データへの汎化性能の検証についてはうつ病患者と健常者から収集したデータを対象に、うつ病の判別器の作成を行い、うつ病と健常対照者を判別することができ、独立した外部データにおいて汎化性能が確認できた。教師なし学習によるデータ駆動的うつ病サブタイプ推定については、うつ病患者と健常者から収集したデータを対象に、多重ベイズ共クラスタリングを適用し、データ駆動的なうつ病サブタイプ分類を試み、抗うつ剤に対する治療反応性の良し悪しと対応付けられる3つのクラスター（サブタイプ）を発見した。うつ病と認知症の音声および表情データを用いた判別は、うつ病あるいは認知症患者のデータセットを用い、音響学的特徴および表情特徴を解析対象とし、判別器の生成を行い、うつ病と認知症を判別精度は89.9%を得た。12ヶ月にわたって採取したライフログデータおよびウェアラブルデバイスによって記録されるデータから、うつ病の増悪を予測するモデルを作成し、最大2週間前まで偶然を越える予測能を示すAUC 0.7を越えるモデルを作成できた。日本で行われた7本の抗うつ剤のプラセボ対照試験データをプールして、機械学習手法を用いて治療効果の異なるサブグループを探索した結果、うつ病初発から1年以上を経過しているかどうかと性別とにより、質的な差異を示すサブグループが同定された。以上、本研究により、うつ病の診断、再燃・再発、反応性、層別化などについて、いくつかの解決すべき課題はあるものの、様々なデータセットに対するAIを用いた解析の有用性を示すことができた。

研究分担者

岸本泰士郎 慶應義塾大学医学部 専任講師  
吉本潤一郎 奈良先端科学技術大学院大学情報科学研究科 准教授  
橋本亮太 大阪大学大学院大阪大学・金沢大学・浜松医科大学・千葉大学・福井大学連合小児発達学研究科 准教授  
古川壽亮 京都大学大学院医学研究科 教授  
丸尾和司 筑波大学医学医療系 准教授  
池田和隆 東京都医学総合研究所分子精神医学分野長  
山脇成人 広島大学大学院医歯薬保健学研究科

A. 研究目的

わが国のうつ病を含む気分障害患者の受診はこの10年間に2.4倍増加し100万人を超え、抗うつ薬の売上は年10%ずつ増加し1300億円にも達する。Global Burden of Disease研究によれば、2030年においてもうつ病は自殺・休職の主要因であり、わが国のみならず人類共通の苦悩の最大原因であることが示されている。うつ病はコモディティーズで、診断内の異質性が非常に高い。この異質性のため、医師は長い時間をかけて試行錯誤

しながら最適治療を探ることを余儀なくされている。

これまでにわれわれは、AIを用いた診断、治療反応予測に関して、数十例のうつ病と健常者の臨床データと課題遂行時や安静時の脳活動といったバイオデータを組み合わせて探索的な検討を行ってきた (Shimizu et al, 2015; Yoshida et al, 2017)。しかし、脳機能画像以外のバイオデータは検討していないこと、サンプル数が十分でないこと、外部テストデータでの汎化性能は検証していないなどの課題が残存していた。

そこで、本研究では、脳機能画像に加え、脳構造画像、表情、音声などのバイオデータ、プラセボ対照の抗うつ薬臨床試験データといった様々な比較的大規模なデータセットを、複数のAIアルゴリズムを用いて解析することによって、各データセットに対する最適なAIアルゴリズムを特定し、解析パイプラインの提案を行う。また、一部のデータに関して、外部データに対する汎化性能を確認することで、うつ病の診断、再燃・再発、反応性、層別化などについて、AIを用いた解析の有用性を検証する。

## B. 研究方法

### B-1. 脳機能画像などを用いた検討

#### B-1-1 判別のための機械学習手法の検討

安静時脳機能画像を用いたうつ病患者と健常者の判別（診断）に関しては、高次元データを低次元の部分空間に射影することで、分類精度を向上させることが期待されたため、いくつかの分類方法について検討した。すなわち、ナイーブな線形判別分析法 (Linear Discriminant Analysis ; LDA)、サポートベクターマシン (Support Vector Machine; SVM)、最小二乗回帰法 (ordinary least squares regression; OLS)、部分最小二乗回帰法 (partial least squares; PLS)、さらに 2 次多項式カーネル (KPLS-Poly(2))、3 次多項式カーネル (KPLS-Poly(3))、Gaussian カーネル (KPLS-Gauss) を用いた部分最小二乗回帰法などの、精度、感度、

特異度を比較検討した。

#### B-1-2 施設間の測定バイアスの検討

トラベリングサブジェクト 9 名を用いて異なる施設の MRI スキャナーで測定がおこない、施設間の測定バイアスとうつ病と健常対照者の差異と比較した。

#### B-1-3 解析パイプライン用パラメータの網羅的探索

広島大学および連携医療機関を受診した 148 名のうつ病患者群、および、広島大学で募集した 269 名の対象健常者群から、MRI を用いて安静時脳活動データを収集し、解析パイプラインのパラメータを網羅的に検討した。

#### B-1-4 教師あり学習によるうつ病判別と外部データへの汎化性能の検証

安静時の脳機能結合を用いたバイオマーカー作成に関して、広島大学 4 施設より集められた MRI データの内、躁病、薬物依存、アルコール依存、精神病性障害、パーソナリティ障害等の併存疾患を除外した後の症例 93 名と年齢性別を合わせた健常者 93 名の合計 186 名が解析対象とされた。

#### B-1-5 教師なし学習によるデータ駆動的うつ病サブタイプ推定

前節で述べた研究参加者の部分集合にあたるうつ病患者群 67 名と健常者群 67 名については、安静時 fMRI データに加えて、うつ病の重症度を評価する臨床指標 (HRSD, BDI) や幼児期トラウマ体験を指標化した CATS (Child Abuse Trauma Scale)、血液サンプルより測定した遺伝子多型や BDNF メチル化レベルなどの生理指標も計測・取得し、参加者数 134 人×特徴量 2948 次元のデータ行列を構成した。

#### B-1-6. 脳構造画像などを用いた検討

脳構造画像に関しては、Voxel-based morphometry (VBM) 解析を行い、サポートベクターマシンを用いて、うつ病患者と健常者の判別の感度と特異度とともに、測定を行ったスキャナーの寄与度もあわせて検討した。

### B-2. 音声・音響指標などを用いた検討

診療場面で医師と患者が行う通常の会話の様子を10-30分間、録音、録画したデータセットの内、うつ病や認知症の症状が明確に存在するうつ病、認知症それぞれ64名、74名分のデータを抽出し、データセットとした。音声解析のため Praat を用いて、話者の Formant、Pitch、Intensity、Pulse 等を解析することが可能である。これらのうち、うつ病重症度等の解析に有用であった要素を利用し、特徴エンジニアリングを行い、次に Lasso (least absolute shrinkage and selection operator) を用いて特徴選択を行った。うつ病あるいは認知症と標識したデータでサポートベクターマシンを用いた分類器の生成を試みた。また、表情の解析に際しては、オムロン社の OKAO Vision を用いた。OKAO Vision は顔検出・顔器官検出を自動で行い、表情(喜び・驚き・怒り・悲しみ・無表情)、瞬目などを定量することが可能である<sup>3,4)</sup>。種々の施行から、うつ病重症度等の解析に有用であった要素を利用し、特徴エンジニアリングを行い、次に Lasso を用いて特徴選択を行った。うつ病あるいは認知症と標識したデータで support vector machine with cubic kernel による分類器の生成を行った。

### B-3. ライフログなどを用いた検討

名古屋市立大学、高知大学、広島大学、東邦大学の4つの大学病院と、それらの関連の精神科病院とクリニックの外来に通院する寛解期の大うつ病患者から89名が本研究に参加した。被験者自身の iPhone に「くらしアプリ」をインストールすると共に、ウェアブルデバイス「シルミー」を装着してもらい、ライフログ情報を収集した。また、うつ状態診断を臨床試験コーディネーターの電話インタビューによって収集した。

発予測モデルの構築にあたり、PHQ 値の増減に着目し、PHQ が前回の測定から5ポイント増加した場合を「再発有り」、それ以外を「再発無し」と定義した。再発予測に有用な特徴量選択のための予備解析として、PHQ 測定日を基点とした event-triggered average 法によって、「再発有り」

の場合と「再発無し」の場合の時系列を統計的に比較した。具体的には、すべての特徴量の時系列データを PHQ の測定日から30日前までを一区切りとして抽出した。次に、PHQ 測定日から15日前まで、日ごとに各特徴量の2群間差を U-test (有意水準 0.05) で評価した。さらに、有意な違いが連続して起きる日数が偶然では説明できないものを permutation test (有意水準 0.01) で評価することで、再発予測モデルのための説明変数として抽出した(図1)。

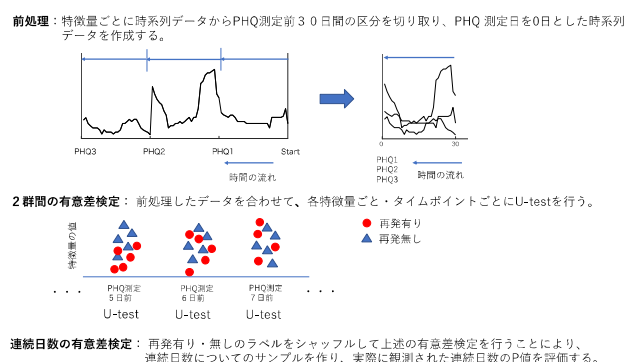


図1：時系列データ分析の概略

得られた説明変数から従属変数である再発有無のラベルをどの程度予測できるかについて、コックス比例ハザードモデルを用いて検討した。モデルパラメータは最尤推定法により決定し、モデルの出力である生存確率が0.5以上の時は「再発無し」、0.5未満の時は「再発有り」を予測結果とした。あらゆる説明変数の組合せの中で、最良の汎化性能が期待できるものを決定するために、Leave-one-out 交差検証法によって得られたテストサンプルの予測結果と正解の比較を精度、感度、特異度、および、ROC (Receiver Operating Characteristic) 曲線の AUC (Area Under the Curve) スコアを各モデルで計算した。最後に、最良のモデルが、ライフログデータを用いて何日前に再発を予測できるか統計的に検証した。

### B-4. プラセボ対照抗うつ薬治療データを用いた検討

日本で行われた7本、2399人分の、抗うつ剤のプラセボ対照治療データをプールして、

Qualitative Interaction Trees (QUINT)という新しい機械学習手法を用いて治療効果の異なるサブグループを探索した。主解析では、QUINT法を適用した。この方法は、2つの治療の群間差が異なる2つの集団に逐次分割し、治療との交互作用に関して特徴的な部分集団を抽出できる決定木ベースのクラスタリング法である。内部整合性と外部妥当性の検討のために、4研究のそれぞれで、主解析で抽出された各リーフ(以降、各リーフ)でのCohen's dとそのSEを推定した。外部妥当性の検討のために、4研究被験者番号が後半の症例データについて、各リーフでのCohen's dとそのSEを推定した。外部妥当性の検討のために、残りの3研究試の各試験で、各リーフでのCohen's dとそのSEを推定した。

## C. 研究成果

### C-1. 安静時脳機能画像などを用いた検討

#### C-1-1 判別のための機械学習手法の検討

安静時脳機能画像を用いていくつかの分類方法を組み合わせて、精度、感度、特異度を用いて、その仮説を検証(図2)した。

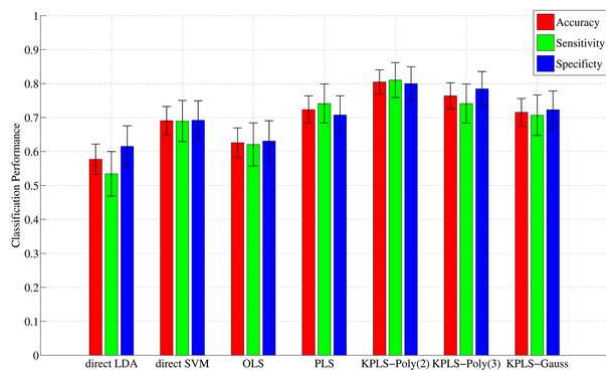


図2: うつ病患者と健常者の判別成績

ナイーブな線形判別分析法 (Linear Discriminant Analysis ; LDA) (精度 57.7%、感度 53.4%、特異度 61.5%) およびサポートベクターマシン (Support Vector Machine; SVM) (精度 69.1%、感度 69.0%、特異度 69.2%) と比べて、2次多項式カーネルを用いた部分最小二乗回帰法 (Kernel Partial Least Squares regression with the 2nd order Polynomial; KPLS-Poly (2)) は有意に優れた精度

を示した(精度 80.5%、感度 81.0%、特異度 80.0%)。この結果は、分類モデルを構築するために臨床測定予測モデルを利用することの有用性を示唆している。さらに、KPLS-Poly (2) は、通常最小二乗回帰法 (Ordinary Least Squares Regression; OLS) (精度 62.6%、感度 62.1%および特異度 63.1%) と比べて有意に優れた精度を示し、回帰モデルの潜在空間を考慮することの有用性が示唆された。

#### C-1-2 施設間の測定バイアスの検討

測定バイアスに関しては、異なる施設スキャナーで安静時脳機能画像データを収集した場合、施設間の測定バイアスはうつ病などの精神疾患と健常者の違いと同程度かさらに大きかった。

#### C-1-3 解析パイプライン用パラメータの網羅的探索

あらゆるパラメータ設定の組合せに対して診断モデルの予測精度を評価した。その結果、開眼状態での安静時 fMRI データに対して、BAL<sup>6)</sup> の脳領域分割法による機能結合性を定義したものを特徴量として、診断モデルへの入力前に一般線形モデルによる攪乱変数除去と有意水準 1%の二群検定法による特徴量選択の前処理を施した時が最適となり、その時の精度は 0.75、感度は 61%となった(表1)。

開眼/閉眼	脳領域分割	変数選択	攪乱変数除去	精度	感度
開眼	BAL	二群検定法 ( $\alpha=0.01$ )	一般線形モデル	0.7483	0.6081
閉眼	BAL	SCCA	SCCA	0.7196	0.6028
開眼	Stanfordx90	二群検定法 ( $\alpha=0.01$ )	一般線形モデル	0.6962	0.4918
閉眼	BAL	SCCA+二群検定法 ( $\alpha=0.03$ )	SCCA	0.6869	0.5354
開眼	Stanfordx90	SCCA+二群検定法 ( $\alpha=0.05$ )	SCCA	0.6801	0.4621
開眼	Stanfordx90	SCCA	SCCA	0.6747	0.4874
開眼	BAL	二群検定法 ( $\alpha=0.03$ )	一般線形モデル	0.6737	0.5068
閉眼	BAL	SCCA+二群検定法 ( $\alpha=0.01$ )	SCCA	0.6729	0.5181
開眼	BAL	SCCA+二群検定法 ( $\alpha=0.01$ )	SCCA	0.6643	0.4589
閉眼	BAL	SCCA+二群検定法 ( $\alpha=0.05$ )	SCCA	0.6589	0.5272
開眼	BAL	SCCA+二群検定法 ( $\alpha=0.05$ )	SCCA	0.6573	0.4797
閉眼	BAL	二群検定法 ( $\alpha=0.03$ )	一般線形モデル	0.6542	0.5068
閉眼	Stanfordx90	SCCA	SCCA	0.6434	0.4394
開眼	Stanfordx499	SCCA+二群検定法 ( $\alpha=0.05$ )	SCCA	0.641	0.4058
閉眼	Stanfordx90	SCCA+二群検定法 ( $\alpha=0.05$ )	SCCA	0.6381	0.4347
閉眼	Stanfordx90	二群検定法 ( $\alpha=0.05$ )	一般線形モデル	0.6381	0.459
開眼	BAL	SCCA+二群検定法 ( $\alpha=0.03$ )	SCCA	0.627	0.4604
閉眼	Stanfordx499	SCCA+二群検定法 ( $\alpha=0.05$ )	SCCA	0.6247	0.438
閉眼	Stanfordx499	SCCA	SCCA	0.62	0.4072
開眼	Stanfordx499	二群検定法 ( $\alpha=0.05$ )	一般線形モデル	0.6084	0.4392

表1: 解析パイプラインのための網羅的パラメータ探索結果 (精度に関して上位20位までを抜粋)

#### C-1-4 教師あり学習によるうつ病判別と外部デ

ータへの汎化性能の検証

全ての対象では判別率 51%、AUC 0.52 にとどまったため、うつ病の中でも生物学的要因の影響が大きいとされるサブタイプ、メランコリー型の特徴を有する一群に絞った検討を行った。結果、判別率 70%、AUC:0.77 まで成績が向上した。このバイオマーカーを、完全に独立な施設のデータセットに適用した結果、判別率は 65%(AUC 0.62)であり、汎化性能が確認された。判別器として抽出された 12 の脳機能結合の中には、うつ病の先行研究において重要であることが示唆されている脳領域が多数含まれていた (図 3)。

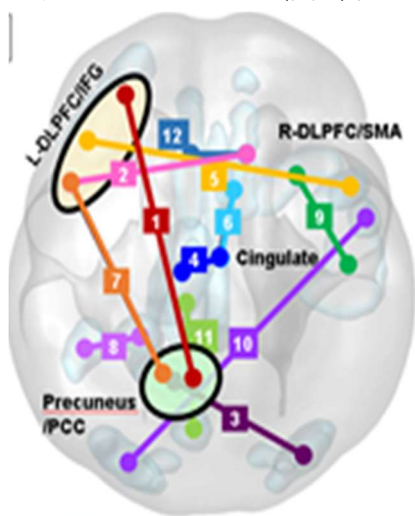


図 3: 完全な独立コホートに汎化するメランコリア特徴を有するうつ病の安静時機能結合の判別器

さらに、このバイオマーカーの臨床的意義や有用性を明らかにするために、判別器の値 (Weighted Linear Sum: WLS) とうつ病重症度との関連について検討した結果、この値は抑うつ症状の質問紙得点 (BDI) と相関を示し、6 週間の抗うつ薬治療により健常方向に変化したことから、うつ状態を反映するバイオマーカーである可能性が示唆された。

C-1-5 教師なし学習によるデータ駆動的うつ病サブタイプ推定

B.節で述べたデータ行列に対して、多重ベイズ共クラスタリングを適用した結果、15 種類の共クラスタ構造が得られた (図 4)。

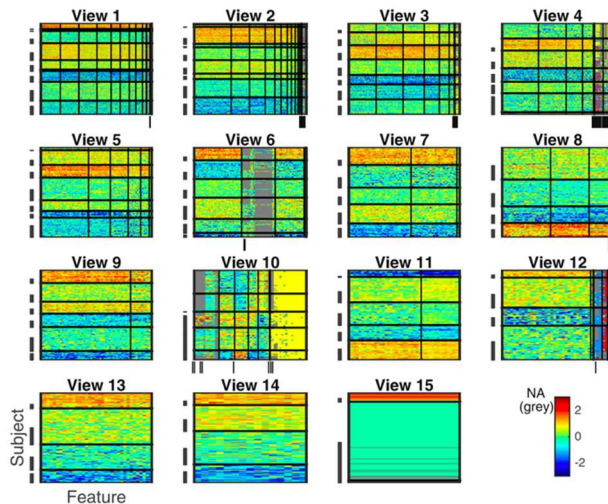


図 4: 多重ベイズ共クラスタリングの適用結果 (縦軸は参加者、横軸は特徴量を表しており、参加者軸に付随したダッシュ記号はうつ病患者を表している。ヒートマップ中の太線はクラスタの境界を示している。)

このうち、View 10 で示されるクラスタ構造は、他の View に比べても特にうつ病患者群と健常者群の分類とも一致しており、最初の 2 つのクラスタは健常者群のみで構成されている一方で、残りの 3 つのクラスタに属する参加者のほとんどがうつ病患者となった。そこで、View 10 で示されるクラスタ構造について、より詳細に調べた。まず、このクラスタ構造に含まれる安静時機能結合性を抽出したところ、右角回を中心とするスポークアンドハブ構造の機能ネットワーク (図 5) がこのクラスタを特徴づけていることが分かった。

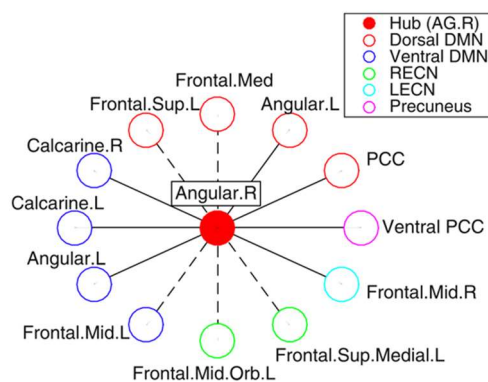


図 5: 図 3 の View 10 のクラスタ構造に含まれる安静時機能結合ネットワーク

また、ほとんどがうつ病患者から構成される3つのクラスターの違いを調べたところ、BDIやHRSDによって評価されたうつ病重症度の初診時とSSRI投薬開始後6週間（または6ヶ月）後の差、すなわち、SSRIに対する治療反応性がこれらのクラスターの特徴を反映していることが分かった。また、これら3つのクラスターの違いには、幼児期トラウマ指標であるCATSも関連していることがわかった。そこで、View10に含まれる機能結合性の第1主成分とCATSの2次元でデータの分布を可視化したところ、右角回を中心とした機能ネットワークにより、D3クラスター（治療反応性が良いクラスター）への割当が決まり、残りの2つのクラスターについては、CATSが小さければD2クラスター（治療反応性が良いクラスター）へ、CATSが大きければD1クラスター（治療反応性が悪いクラスター）へと割り当てが決まる構造があることが分かった（図6）。

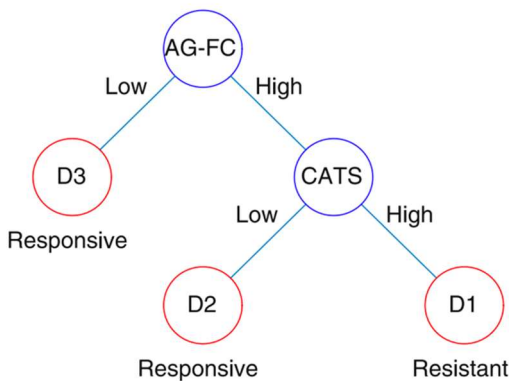


図 6: 右角回を中心とした機能ネットワークの安静時機能結合性と幼児期トラウマ指標に基づく治療反応予測モデル（AG-FC は右角回を中心とした12個の安静時機能結合性のスコアであり、CATSは幼児期トラウマ指標である。）

#### C-1-6 脳構造画像などを用いた検討

脳構造画像を用いたうつ病患者と健常者の判別（診断）に関しては、予備的な検討ではあるが、精度75.9%（感度78.1%、特異度72.9%）で、異

なる施設のスキャナー（Scan10,21,22,23,30,40,41）で測定した影響は小さかった（図7）。

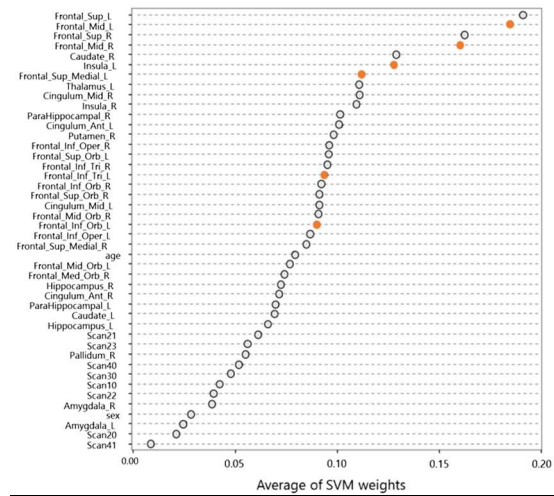


図 7: SVM を用いたうつ病診断への寄与度

#### C-2. 音声・音響指標などを用いた検討

一個抜き交差検証（Leave one out cross validation）によってうつ病、認知症の分類の精度は82.2%であった。

Support Vector Machine with cubic kernel による分類器の生成を行った結果、うつ病と認知症を分類する音声および表情データを用いた精度、感度、特異度は以下の様であった（10-fold cross validation による）。

	精度	感度	特異度
音声データのみ	0.884	0.919	0.844
表情データのみ	0.826	0.813	0.838
音声・表情データの組み合わせ	0.899	0.919	0.875



### C-3. ライフログを用いた検討

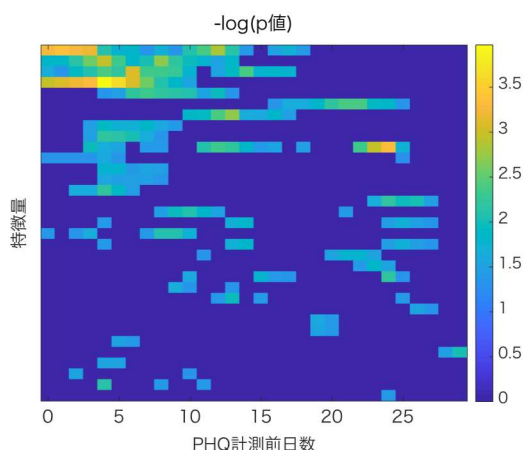


図 8: 2 群 (再発有り、無し) の違いに関する U-test の P 値 (濃い青は有意差なし)。特徴量は連続した有意日数が多いに並び替えた。上位 5 つの特徴量は順に、ゴロゴロ時間、走行回数、過去 1 週間のエネルギー消費量平均、睡眠時間である。

予備解析における特徴量選択の結果、「ゴロゴロ時間」、「走行回数」、「通勤・通学時間」、「睡眠時間」、「過去 1 週間のエネルギー消費量平均」の 5 つの特徴量が抽出された (図 8)。抽出された特徴量のすべての組み合わせに対して、コックス比例ハザードモデルを当てはめ、交差検定を行った結果 (表 2)、AUC が 0.70 以上の組み合わせは 10 通りであった。このうち、予測モデルに基づいて被験者にアラートを発することを想定し、感度 (再発する真陽性率) が 0.60 以上、特異度が 0.50 以上という条件を満たす組み合わせは 2 通りであった。

特徴量	精度	感度	特異度	AUC
走	0.25	1.00	0.17	0.74
ゴ+走+エ	0.25	0.95	0.17	0.72
走+通+睡+エ	0.25	0.91	0.18	0.72
ゴ+走	0.27	0.96	0.19	0.71
ゴ+走+通	0.28	0.92	0.21	0.71
ゴ+走+睡+エ	0.24	0.91	0.17	0.71
<b>ゴ+通</b>	<b>0.60</b>	<b>0.64</b>	<b>0.59</b>	<b>0.70</b>
<b>ゴ+通+睡</b>	<b>0.60</b>	<b>0.64</b>	<b>0.59</b>	<b>0.70</b>
ゴ+走+通+睡	0.33	0.88	0.27	0.70
すべて	0.27	0.91	0.19	0.70

表 2: コックス比例ハザードモデルを用いた交差検定結果<sup>7</sup>  
(AUC 0.70 以上)

最後に、この 2 つの特徴量の組み合わせに対して、生存関数を用いて PHQ 測定前の AUC を評価した。その結果、いずれのモデルも 2 週間前に AUC がランダム化した場合の上限値 0.60 を超えることがわかり、2 週間前の再発予測の可能性が示唆された (図 9)。

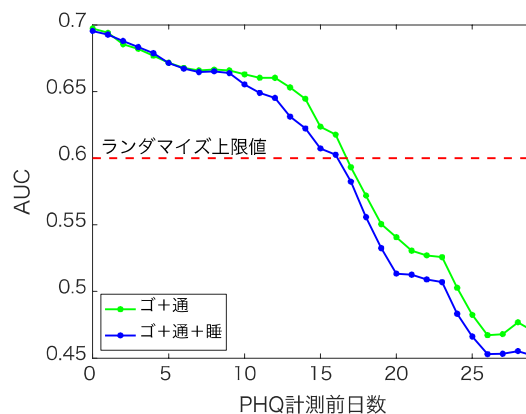
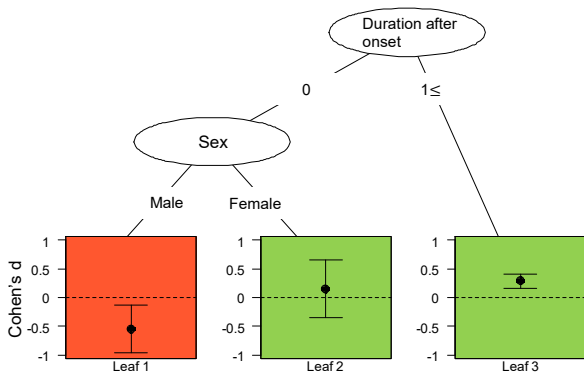


図 9: PHQ 計測前の AUC 値の推移

### C-4. プラセボ対照抗うつ薬治療データを用いた検討

QUINT 法による樹状図を作成した。効果修飾因子として、発症年齢、性別、体重、ベースラインの HAMD 総得点の 4 個の因子が残ったが、リーフの刈込後は最初の 3 因子のみが残った。発症年齢が 30 歳未満では実薬が効果サイズ 0.34 (95%信頼区間: 0.22 から 0.46) で有効であり、一方発症年齢が 30 歳以上になると男性では効果サイズ -0.15 (95%信頼区間: -0.29 から -0.01) でプラセボの方が有効という結果になった。30 歳以上の女性においては、体重が 54 キロ以上ならば実薬の方が有効 (0.46, 95%CI: 0.08 から 0.84) で、54 キロ未満の者では有意差はなかった。QUINT 法による最終的な樹状図および各部分集団での Cohen's d を作成した。発症からの経過年数 (duration after onset) と性別を以下に示す



リーフ		1	2	3
プラセボ	N	35	21	153
	Mean	-12.54	-10.19	-8.77
	SD	6.05	8.03	6.51
抗うつ剤	N	67	67	361
	Mean	-9.25	-11.13	-10.67
	SD	6.04	5.56	6.72
Cohen's d	d	-0.54	0.15	0.29
	SE	0.21	0.25	0.06

発症からの経過年数が1年以上のリーフ3では抗うつ剤の効果が有意にプラセボよりも高かった。経過年数が1年未満の場合において、男性(リーフ1)ではプラセボに有意に劣っており、女性(リーフ2)では治療群間差が小さかった。

#### D. 考察

##### D-1. 安静時脳機能画像などを用いた検討

安静時脳機能画像に関しては、臨床測定の予測モデルを媒介すること、低次元の特徴空間を考慮することで、比較的小サンプルでも分類性能の向上が期待できることを明らかにした。また、安静時脳機能画像は測定バイアスの影響が大きく、スパース推定による特徴選択、施設効果の線形回帰による regression-out、アンサンブル学習、traveling subjects を用いた測定バイアスの分離と推定といった数理統計手法の利用が測定バイアスの補正に必要であると考えられた。パラメータの網羅的探索の結果から、安静時 fMRI 撮像時では開眼状態の方が、最適な予測精度が実現できるという意味

では望ましいことが示唆される。しかしながら、他のパラメータに依存して閉眼時の方が予測精度が良くなる場合もあり、実用上、開眼・閉眼のどちらが良いかについては一貫性のある結果は得られなかった。また、脳領域分割法については BAL の結果が総じて良かった。変数選択法と攪乱変数除去法については、一般線形モデルによる攪乱変数除去と二群検定法(有意水準 1%) の組合せが最適な結果となったが、この組合せを変えても、BAL による脳領域分割法を用いている限りは精度が大きく落ちるということにはなかった。したがって、現状の結果からは、BAL を用いた脳領域分割が特に推奨される。一方で、昨年度までの結果から入力変数の変数選択(または低次元化)と測定バイアス除去のための攪乱変数除去法の導入は必須であるものの、その方法については大きく精度の改善や悪化につながるようなものは同定できなかった。

教師あり学習の結果から、完全な独立コホートに汎化するメランコリア特徴を有するうつ病の安静時機能結合の判別器を作成した。この判別器の12結合の線形加重和は、抑うつ症状と有意な正の相関があることから、状態を反映するバイオマーカーであると考えられた。教師なし学習のサブタイプ分類法の開発においては、データ駆動的に治療反応性の違いに応じてうつ病患者群が3つのグループに分類できることを見出した。また、この3つのグループは右角回を中心とした12個の安静時機能結合性と幼児期トラウマ経験の大小によって特徴づけられていた。現在のところ、まだ少サンプルの解析結果であるため、その再現性は確認できていないが、これらの結果は、SSRI に対する治療反応性予測するためのモデルとして図6の決定木が有望であることを示唆している。

脳構造画像に関しては、SVM に適用した結果からは異なる施設のスキャナーで測定した影響(測定バイアス)は小さいことが明らかになった。

##### D-2. 音声・音響指標などを用いた検討

音声あるいは表情の指標によるうつ病、認知症

の比較も行ったがそれぞれを明確に分類するような単一の指標は存在せず、特徴エンジニアリングおよび機械学習によってうつ病と認知症患者の分類が90%近い精度で可能であった。今後、さらなるデータ収集や学習モデルの改善を通じて、分類の精度をあげることでより臨床上的有用性は増すものと考えられる。また、現在は10分程度の撮影データを元に解析を行っているが、より短時間でも可能かどうかの検証も有用と考える。将来は治療反応を予測するような臨床的により有用な予測技術の開発が求められる。なお、本検証結果は前述の研究に基づくデータセットのみを対象としており、外部データによる検証を行う必要がある。

#### D-3. ライフログを用いた検討

今回の解析から、交差妥当性がAUCで0.7を超えるモデルが得られた。かつ、再発の2週間前から、偶然を超える予想が可能であることが示された。実際の運用においては、特異度を高めるカットオフを設定することによって、陽性適中率を高めることが出来ると予想される。これにより、最大2週間前に「悪化の兆しがあります。きちんと服薬しましょう」あるいは「認知行動療法を復習しましょう」というようなワーニングを出すことが可能である。ワーニングであって、診断ではないので、陽性適中率が80%や90%と言うような数字になる必要はない。ワーニング疲れを引き起こさない程度でワーニングを出せば、予防効果の向上に繋がることが予想される。次の段階の研究としては、同じライフログデータから、本解析で用いた以外の特徴量を抽出することが出来るので、追加の特徴量をモデルに加えることで、AUCをさらに高めることが出来るかを検討したい。そして、そのモデルを利用して、実際にワーニングを出すことが再発減少に繋がるかどうかを検証するためには、RCTが必要であると考えている。

#### D-4. プラセボ対照抗うつ薬治療データを用いた検討

日本で行われた7本、2399人分の、抗うつ剤のプラセボ対照治療データをプールして、QUINTと

いう新しい機械学習手法を用いて治療効果の異なるサブグループを探索したところ、うつ病初発から1年以上を経過しているかどうかと、性別により、質的な差異を示すサブグループが同定された。このサブグルーピングは、内的妥当性および同じ試験の後半データセットにおける時間妥当性は概ね確認されたが、まったく別個の試験における外的妥当性は一定しなかった。

#### E. 結論

安静時脳機能画像に関しては、スパース推定L1-Sparse Canonical Correlation Analysis(L1-SCCA) & Sparse Logistic Regression (SLR)による特徴選択、施設効果の線形回帰による regression-out、あるいは traveling subjects を用いた施設効果の測定バイアスの分離と推定といった数理統計手法の利用がデータ解析に際しては必要と考えられた。解析パイプライン用パラメータに関しては、開眼状態での安静時fMRIデータに対して、BALの脳領域分割法による機能結合性を定義したものが最適と考えられた。また、測定バイアスに対する分類器の頑健性はアンサンブル学習の導入によって改善されることが期待された。一方、脳構造画像に関しては、これらの工夫や補正の必要性が小さいものと考えられた。

安静時脳機能MRIデータに関する教師あり学習によるうつ病判別と外部データへの汎化性能の検証についてはうつ病患者と健常者から収集したデータを対象に、うつ病の判別器の作成を行い、うつ病と健常対照者を判別することができ、独立した外部データにおいても汎化性能も確認できた。教師なし学習によるデータ駆動的うつ病サブタイプ推定については、うつ病患者と健常者から収集したデータを対象に、多重ベイズ共クラスタリングを適用し、データ駆動的なうつ病サブタイプ分類を試みた。その結果、抗うつ剤に対する治療反応性の良し悪しと対応付けられる3つのクラスター(サブタイプ)を発見した。

診療場面で医師と患者が行う通常の会話の様子を録音・録画した音声・表情データを用いて90%

程度の精度でうつ病と認知症の分類が可能であった。さらなるデータ収集や学習モデルの改善を通じて、分類の精度をあげることでより臨床上の有用性は増すものと考えられる。また、現在は10分程度の撮影データを元に解析を行っているが、より短時間でも分別が可能かどうかの検証も有用と考えられた。

ライフログに関しては、スマートフォンを利用して半自動的に採取される活動記録およびウェアラブルデバイスによる記録から、再発再燃予測モデルを構築し、それをさらにブラッシュアップすることで、AUCで0.7を越え、かつ2週間前から予測が可能なモデルを作成した。

プラセボ対照抗うつ薬治療データについてQUINT法は、個別化医療においてとくに重要な質的な交互作用を検出するために有用な手法である。しかし、安定した外的妥当性を得るためにはさらなるデータと解析が必要と考えられた。

以上のように、本研究により、うつ病の診断、再燃・再発、反応性、層別化などについて、いくつかの解決すべき課題はあるものの、様々なデータセットに対するAIを用いた解析の有用性を示すことができた。

## F. 健康危険情報 なし

## G. 研究発表

### 1. 学会発表・招致講演等

Ichikawa N, Lisi G, Yahata N, Okada G, Takamura M, Hashimoto R, Yamada T, Yamada M, Suhara T, Moriguchi S, Mimura M, Yoshihara Y, Takahashi H, Kasai K, Kato N, Yamawaki S, Seymour B, Kawato M, Morimoto J, Okamoto Y. A classifier of melancholic depression with whole-brain resting-state connectivity. rtFIN2017, Nara, Japan, 2017.12.1,  
岸本泰士郎. AIを用いた精神疾患臨床症状定量化の試み, 平成29年度医薬品評価委員会臨床評価部会総会, 東京, 2018.2.21

岸本 泰士郎. 機械学習による精神運動制止の評価の試み, 第114回日本精神神経学会学術総会, 神戸, 2018.6.21

岸本 泰士郎. 情報通信技術や機械学習を活用した臨床症状評価, 第114回日本精神神経学会学術総会, 神戸, 2018.6.21

岸本 泰士郎. 情報通信技術 (ICT) や人工知能 (AI) の活用という観点で30年後の精神科医療を展望する, 第114回日本精神神経学会学術総会, 神戸, 2018.6.23

澤田 恭助, 高宮 彰紘, 岸本 泰士郎, 三村 将. うつ病患者に対する音声データの臨床的有用性の検討: システムティックレビューとメタ解析, 第114回日本精神神経学会学術総会, 神戸, 2018.6.22

岸本 泰士郎. 遠隔医療は国民に寄り添った医療になるか? 本邦における遠隔精神科医療の展望と課題, 日本精神神経科診療所協会 第24回学術研究会, 兵庫, 2018.6.24

市川奈穂, 岡本泰昌. 安静時fMRI活動を用いたうつ病の判別, 第15回日本うつ病学会, 東京, 2018.7.27

岡本泰昌, 市川奈穂. 脳機能画像研究からみたうつ病の異種性, 第15回日本うつ病学会, 東京, 2018.7.28

岡本泰昌. 神経回路病態に基づくうつ病の診断・治療法の開発, 第26回脳の世紀シンポジウム 『AIと脳』, 東京, 2018.9.12

岸本泰士郎. 情報通信技術や機械学習を用いた精神症状定量化の試み, 日本線維筋痛症学会第10回学術集会, 東京, 2018.9.29

岸本 泰士郎. 遠隔医療や機械学習を活用した認知症診療の展望, 第37回日本認知症学会学術集会, 札幌, 2018.10.13

工藤弘毅, 岸本泰士郎. 「メンタルヘルスリテラシーと人工知能 (AI)」, 第38回日本精神科診断学会, 埼玉, 2018.10.19

岸本泰士郎. 精神科医療の遠隔化は診療所・精神病院・総合・大学病院に普及し得るか, 第38回医療情報学連合大会 (第19回日本医療情報学会学術大

会), 福岡, 2018.11.23

岸本泰士郎. 情報通信技術や機械学習を活用した精神科領域の展望, 第2回デジタルヘルス学会, 東京, 2018.12.23

岸本泰士郎. 人工知能技術を用いた精神疾患症状定量化の試み, 第1回日本メディカルAI学会, 東京, 2019.1.26

Ichikawa N, Lisi G, Yahata N, Okada G, Takamura M, Hashimoto R, Yamada T, Yamada M, Suhara T, Moriguchi S, Mimura M, Yoshihara Y, Takahashi H, Kasai K, Kato N, Yamawaki S, Seymour B, Kawato M, Morimoto J, Okamoto Y. Melancholic depression biomarker of resting-state functional connectivity. AsCNP-ASEAN2019, Yogyakarta, Indonesia, 2019.3.2

#### H. 知的財産権の出願・登録状況

知的財産の内容(うつ症状の判別方法、うつ症状のレベルの判定方法、うつ病患者の層別化方法、うつ症状の治療効果の判定方法及び脳活動訓練装置)、種類・番号 PCT/JP2018/36952、出願年月日 2018.10.02、取得年月日、権利者国立大学法人広島大学、国際電気通信基礎技術研究所(ATR)

#### 参考文献

- 1) Ichikawa N, Lisi G, Yahata N, Okada G, Takamura M, Yamada M, Suhara T, Hashimoto R, Yamada T, Yoshihara Y, Takahashi H, Kasai K, Kato N, Yamawaki S, Kawato M, Morimoto J, Okamoto Y. Identifying melancholic depression biomarker using whole-brain functional connectivity. ArXiv 2017
- 2) Yoshida K, Shimizu Y, Yoshimoto J, Takamura M, Okada G, Okamoto Y, Yamawaki S, Doya K. Prediction of clinical depression scores and detection of changes in whole-brain using resting-state functional MRI data with partial least squares regression. PLoS One. 2017, 12: e0179638.
- 3) Tokuda T, Yoshimoto J, Shimizu Y, Okada G, Takamura M, Okamoto Y, Yamawaki S, Doya K. Multiple co-clustering based on nonparametric mixture models with heterogeneous marginal distributions. PLoS One 2017 12: e0186566,
- 4) Yahata N, Morimoto J, Hashimoto R, Lisi G, Shibata K, Kawakubo Y, Kuwabara H, Kuroda M, Yamada T, Megumi F, Imamizu H, Nández J, Takahashi H, Okamoto Y, Kasai K, Kato N, Sasaki Y, Watanabe T, Kawato M, A small number of abnormal brain connections predicts adult autism spectrum disorder. Nature Communications 2016, 7, 11254.
- 5) Shimizu Y, Yoshimoto J, Toki S, Takamura M, Yoshimura S, Okamoto Y, Yamawaki S, Doya K. Toward Probabilistic Diagnosis and Understanding of Depression Based on Functional MRI Data Analysis with Logistic Group LASSO. PLoS One. 2015, 10: e0123524.
- 6) World Health Organization. (2017). Depression and other common mental disorders: global health estimates. World Health Organization.
- 7) American Psychiatric Association. (2000). Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision (DSM-IV-TR).
- 8) Shimizu, Y., Yoshimoto, J., Toki, S., Takamura, M., Yoshimura, S., Okamoto, Y., Yamawaki S., Doya, K. (2015). Toward Probabilistic Diagnosis and Understanding of Depression Based on Functional MRI Data Analysis with Logistic Group LASSO. PLOS ONE, 10(5), e0123524.
- 9) Ichikawa N, Lisi G, Yahata N, Okada G, Takamura M, Yamada M, Suhara T, Hashimoto R, Yamada T, Yoshihara Y, Takahashi H, Kasai K, Kato N, Yamawaki S, Kawato M, Morimoto J, Okamoto Y. (2017). Identifying melancholic depression biomarker using whole-brain functional connectivity. arXiv:1704.01039.
- 10) Beck AT, Steer RA, Ball R, Ranieri WF. (1996). Comparison of Beck Depression Inventories-IA and-II in psychiatric outpatients. Journal of Personality Assessment 67(3):588-597.
- 11) Watson D, Clark LA, Tellegen A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. Journal of Personality and Social Psychology 54(6):1063.
- 12) Snaith R, Hamilton M, Morley S, Humayan A, Hargreaves D, Trigwell P. (1995). A scale for the assessment of hedonic tone the Snaith-Hamilton Pleasure Scale. The British Journal of Psychiatry 167(1):99- 103.
- 13) Wold, H. (1975). Soft Modelling by Latent Variables: The Non-Linear Iterative Partial Least Squares (NIPALS) Approach. Journal of Applied Probability, 12(S1), 117-142.
- 14) Rosipal R, Trejo LJ. (2002). Kernel partial least squares regression in reproducing kernel hilbert space. Journal of Machine Learning Research 2:97-123.
- 15) Breiman, L. (1996). Bagging Predictors. Machine Learning 24(2): 123-140.
- 16) Boersma P and Weenink D. Praat: doing phonetics by computer [Computer

- program].Version 6.0.37, retrieved 14 March 2018 from <http://www.praat.org/>
- 17) Boersma P and Weenink D. Praat, a system for doing phonetics by computer. *Glott International* 5(9/10): 341-345.200114; 33(2): 219-37.
  - 18) Higuchi T, Murasaki M, Kamijima K. Clinical evaluation of duloxetine in the treatment of major depressive disorder: placebo- and paroxetine-controlled double-blind comparative study. *Japanese Journal of Clinical Psychopharmacology [in Japanese]* 2009; **12**: 1613-34.
  - 19) Hirayasu Y. A dose-response study of escitalopram in patients with major depressive disorder: a placebo-controlled, double-blind study. *Rinsho Seishin Yakuri (Japanese Journal of Clinical Psychopharmacology)* 2011; **14**(5): 871-82.
  - 20) Hirayasu Y. A dose-response and non-inferiority study evaluating the efficacy and safety of escitalopram in patients with major depressive disorder: a placebo- and paroxetine-controlled, double-blind, comparative study. *Rinsho Seishin Yakuri (Japanese Journal of Clinical Psychopharmacology)* 2011; **14**(5): 883-99.
  - 21) Kinoshita T. A double-blind, placebo-controlled study of a new antidepressant, mirtazapine, in depressed patients. *Japanese Journal of Clinical Psychopharmacology [in Japanese]* 2009; **12**: 289-306.
  - 22) Dusseldorp E, Van Mechelen I. Qualitative interaction trees: a tool to identify qualitative treatment-subgroup interactions. *Stat Med* 2014; **33**(2): 219-37.

平成29・30年度厚生労働科学研究費補助金(政策科学総合研究事業)  
分担研究報告書

脳 MRI データとAI解析によるうつ病診断の汎用性向上とデータ駆動型サブタイプ分類に関する研究

研究代表者 岡本泰昌 広島大学医歯薬保健学研究科 精神神経医科学 教授  
研究分担者 吉本潤一郎 奈良先端科学技術大学院大学先端科学技術研究科情報科学領域 准教授  
研究協力者 徳田智磯 奈良先端科学技術大学院大学先端科学技術研究科情報科学領域 博士研究員  
研究分担者 山脇成人 広島大学大学院医歯薬保健学研究科特任教授  
研究分担者 池田和隆 東京都医学総合研究所分子精神医学分野長  
研究分担者 橋本 亮太 大阪大学大学院大阪大学・金沢大学・浜松医科大学・千葉大学・福井大学連合  
小児発達学研究科附属子どものこころの分子統御機構研究センター 准教授

研究要旨

脳機能画像および脳構造画像の比較的大規模なデータセットを、複数の AI アルゴリズムを用いて解析することによって、各データセットに対する最適な AI アルゴリズムを特定するとともに、それぞれのデータセットに対する測定バイアスについても検討した。また、安静時脳活動データから AI 技術を活用したうつ病診断の実用化と医師の診断を用いたうつ病判別技術の確立および診断を用いないデータ駆動的なうつ病関連予測技術の確立を目指して、解析パイプラインのためのパラメータ最適化、教師あり学習によるうつ病判別法および教師なしクラスタリングによるうつ病サブタイプ分類法の開発を行った。教師あり学習に関しては、うつ病患者群 93 名と健常者群 93 名の研究参加者から収集したデータを対象に、うつ病の判別器の作成を行い、うつ病と健常対照者を判別することができ、独立した外部データにおいても汎化性能も確認できた。教師なし学習に関しては、うつ病患者群 67 名と健常者群 67 の研究参加者から収集したデータを対象に、多重ベイズ共クラスタリングを適用し、データ駆動的なうつ病サブタイプ分類を試みた。その結果、抗うつ剤(セロトニン再取り込み阻害剤; SSRI)に対する治療反応性の良し悪しと対応付けられる 3 つのクラスタ(サブタイプ)が発見され、これらのクラスタは、直角回を中心とした 12 個の安静時機能結合性と幼児期トラウマ経験の大小によって特徴づけられていることを見出した。

A. 研究目的

うつ病は抑うつ気分と意欲低下に特徴づけられる精神疾患である。世界的に見ても全人口の 4.4%にあたる 3 億人以上がうつ病に苦しんでいるという高い罹患率に加えて、自殺の要因ともなりうることから<sup>1)</sup>、うつ病に対する適切な診断と治療が不可欠である。しかしながら、がんや脳卒中などでは、血液検査や画像診断などの客観的な診断法が確立されているのに対して、うつ病では、客観的な診断法がまだ確立されておらず、面談や質問紙の内容を DSM<sup>2)</sup>に代表される診断基準や医師の経験と照らし合わせて診断しているのが現状である。

これまでにわれわれは、AI を用いた診断、治療反応予測に関して、数十例のうつ病と健常者の臨床データと課題遂行時や安静時の脳活動といったバイオデータを組み合わせて探索的な検討を行ってきた(Shimizu et al, 2015; Yoshida et al, 2017; Ichikawa et al, 2017)。しかし、脳機能画像以外のバイオデータは検討していないこと、サンプル数が十分でないこと、一部は外部テストデータでの汎化性能は検証していないなどの課題が残存している。

また、客観的な診断指標を確立しようと、我々はこ

れまで課題遂行時や安静時の fMRI 画像から、人工知能の一分野である機械学習法を活用して、DSM に基づく医師の診断結果を予測するアルゴリズムの探索的検討を行ってきた<sup>3,4)</sup>。その結果、言語流暢性課題時の脳活動を用いたうつ病診断については、90%を超える予測精度を、安静時脳活動を用いたうつ病診断については、機能結合性を特徴量としたアンサンブル学習法の導入によって約 70%の予測精度を達成することができた。一方で、fMRI の原画像から特徴量となる機能結合性を算出し、診断モデルを構築する過程では、全脳をどのように領域分割するか、多施設でデータ取得する際に生じる測定バイアスをどのように除去するかなどのパラメータ設定によって、汎化性能が大きく異なる。このパラメータ設定を最適化することは、診断のための解析パイプラインを確立する上で、重要な問題であるものの、これまで網羅的なパラメータ探索は行ってこなかった。

それに加え、うつ病は診断内の異質性が非常に高い。例えば、うつ病治療の臨床現場での第一選択肢は抗うつ剤による薬物療法であるが、薬物療法のみで 2-4 ヶ月の急性期に寛解に達する者は 50%未満であるという現状がある<sup>5)</sup>。したがって、単純に診断だけ

にとどまらず、うつ病内の異質性(サブタイプ)を分類し、治療効果予測につながる特徴量がデータ駆動的に発見できるのが理想である。

以上の問題解決に向けて、0)各データセットに対する最適な AI アルゴリズムの特定と、それぞれのデータセットに対する測定バイアスの検討、1)解析パイプラインの最適化に向けたパラメータの網羅的探索、2)医師の診断情報を用いた教師あり学習法に基づくうつ病分類法の開発と外部汎化性能の検証および、3)医師の診断情報を用いない教師なし学習法に基づくデータ駆動的なうつ病サブタイプ分類法の開発に取り組み、その有用性を検証した。

## B. 研究方法

### 【検討1:各 MRI データセットに対する最適な AI アルゴリズムの特定と測定バイアスの検討】

安静時脳機能画像を用いたうつ病患者と健常者の判別(診断)に関しては、高次元データを低次元の部分空間に射影することで、分類精度を向上させることが期待されたため、いくつかの分類方法について検討した。すなわち、ナイーブな線形判別分析法(Linear Discriminant Analysis ; LDA)、サポートベクターマシン(Support Vector Machine; SVM)、最小二乗回帰法(ordinary least squares regression; OLS)、部分最小二乗回帰法(partial least squares; PLS)、さらに2次多項式カーネル(KPLS-Poly(2))、3次多項式カーネル(KPLS-Poly(3))、Gaussianカーネル(KPLS-Gauss)を用いた部分最小二乗回帰法などの、精度、感度、特異度を比較検討した。また9名のトラベリングサブジェクトを用いて異なる施設のMRIスキャナーで測定がおこなない、施設間の測定バイアスとうつ病と健常対照者の差異と比較した。

脳構造画像に関しては、広島大データにVoxel-based morphometry (VBM)解析を行い、サポートベクターマシンを用いて、うつ病患者と健常者の判別の感度と特異度とともに、測定を行ったスキャナーの寄与度もあわせて検討した。

### 【検討2:解析パイプライン用パラメータの網羅的探索】

広島大学および連携医療機関を受診した148名のうつ病患者群、および、広島大学で募集した269名の対象健常者群から、MRIを用いて安静時脳活動データを収集した。

データ収集は、異なる場所に設置された、撮像条件も異なる4つのfMRIスキャナを用いて実施した(表1)。各研究参加者のデータはこのうちの1つのfMRIスキャナを用いて収集されたものである。MRI撮像時、研究参加者には、眠らざりリラックスした状態で、できる限り何も考えないように教示した。合わせて、セッションごとに開眼、閉眼のいずれかの状態であるよう教示した。

Parameter	Site1	Site2	Site3	Site4
Participants (Patients / HC)	67 / 68	14 / 49	23 / 34	44 / 118
MRI scanner	GE Signa HDxt	GE Signa HDxt	Siemens Magnetom	Siemens Verio
Magnetic field (T)	3.0	3.0	3.0	3.0
Field of view (mm)	256	256	192	212
Matrix	64x64	64x64	64x64	64x64
Number of slices	32	32	38	40
Number of volumes	150	150	112	244
In-plane resolution (mm)	4.0 x 4.0	4.0 x 4.0	3.0 x 3.0	3.3 x 3.3
Slice thickness (mm)	4.0	4.0	3.0	3.2
Slice gap (mm)	0	0	0	0
TR (ms)	2,000	2,000	2,700	2,500
TE (ms)	27	27	31	30
Total scan time (mm:ss)	5:00	5:00	5:03	10:10
Flip angle (deg)	90	90	90	80
Slice acquisition order	Ascending (Interleaved)	Ascending (Interleaved)	Ascending (Interleaved)	Ascending (Interleaved)

表 1: 安静時脳活動データ収集に用いた4種類のfMRI撮像条件

撮像された各fMRI画像に対して、まず、スライスタイミング補正、体動補正、標準脳への位置合わせ、空間フィルタによる平滑化などの一般的な前処理を行うことによって、各ボクセル単位のBOLD信号時系列を算出した。その後、後述するいずれかの脳領域分割法によって定義される各脳領域内の空間平均を取ることで、各脳領域の活動時系列を得て、2領域間の時間変動に対する相互相関係数をその2領域間の安静時機能結合性として計算した(例えば、全脳を90脳領域に分割する場合には、 $90 \times 89 / 2 = 4005$ 組の安静時機能結合性が得られる)。その後、安静時機能結合性を入力データ、DSM基準に基づく医師の診断結果(うつ病罹患の有無)を出力教師データとして、アンサンブル学習法の一つであるRandom Forestにより診断モデルを構築した。

以上の手続きにおいて、探索したパラメータ種類とそれぞれの設定値とは以下の通りである。

- 計測時の開眼・閉眼条件
  - ① 開眼優先: 開眼時のセッションのみのデータのみを利用する
  - ② 閉眼優先: 閉眼時のセッションのみのデータのみを利用する
- 脳領域分割法
  - ① BAL: 文献<sup>6)</sup>と同じ方法を用いる
  - ② Stanfordx90: Stanford大学が公開している90領域に分割した機能脳領域<sup>7,8)</sup>を用いる
  - ③ Stanfordx499: Stanford大学が公開している499領域に分割した機能脳領域<sup>7)</sup>を用いる
- 攪乱変数除去
  - ① SCCA(スパース正準相関分析法): 文献<sup>6)</sup>と同じ方法を用いる
  - ② 一般線形モデル: 最小二乗線形回帰法により攪乱変数による影響を除去する
- 変数選択法
  - ① SCCA: 文献<sup>6)</sup>と同じ方法を用いる
  - ② 二群検定法: 各特徴量ごとに学習データ



内の健常者群とうつ病患者群の間に統計的有意差が認められたもののみを診断モデルの入力変数として用いる

以上のパラメータ設定の各組合せに対して、診断モデルの予測性能は、leave-one-site-out 交差検証法を利用して、精度 (accuracy) と感度 (sensitivity) により評価した。

### 【検討3: SLR+SCAA 法によるうつ病判別と外部データへの汎化性能の検証】

安静時の脳機能結合を用いたバイオマーカー作成に関して、広島大学 4 施設より集められた MRI データの内、躁病、薬物依存、アルコール依存、精神病性障害、パーソナリティ障害等の併存疾患を除外した後の症例 93 名と年齢性別を合わせた健常者 93 名の合計 186 名が解析対象とされた。被験者は、広島市周辺地域のクリニック (うつ病患者) と地方紙による広告 (健常者) から集め、Yahata et al (2016) の判別器作成用のアルゴリズムと同様な方法を用いて、うつ病バイオマーカーの作成を行った。具体的には、脳を 140 個の小領域に分割し、1 人 1 人について各領域における機能的 MRI 信号の時間波形を取り出し、それらが任意の 2 領域間でどの程度似ているか相関係数として数値化した。140 個の小領域の全てのペア (9, 730 個) について相関係数 (機能的結合) を計算することで、個人の脳全体の回路を定量できて、全脳の回路図 (=9, 730 個の数値からなるベクトル) が作成される。これを研究参加者全員分について求め、開発した人工知能技術 (SCAA+SLR) を適用した。外部独立したデータセットは放医研のデータを利用した。

### 【検討4: ベイズ共クラスタリング法によるデータ駆動的うつ病サブタイプ推定】

前節で述べた研究参加者の部分集合にあたるうつ病患者群 67 名と健常者群 67 名については、安静時 fMRI データに加えて、うつ病の重症度を評価する臨床指標 (HRSD, BDI) や幼児期トラウマ体験を指標化した CATS (Child Abuse Trauma Scale)、血液サンプルより測定した遺伝子多型や BDNF メチル化レベルなどの生理指標も計測・取得し、参加者数 134 人 × 特徴量 2948 次元のデータ行列を構成した (安静時 MRI データ以外の特徴量については表 2 を参照)。

Numerical features	
age	age when first depressive symptoms show up.
	The number of days elapsed for current episode.
	BAS* (Behavioral Activation Scale).
	BDNF (Quantity of brain-derived neurotrophic factor in blood).
	BDI* (Beck Depression Inventory).
	BIS* (Behavioral Inhibition Scale).
	CATS (Child Abuse and Trauma Scale).
	CATS:total for all items
	CATS:N for items on neglect
	CATS:S for items on sexual abuse
	CATS:P for items on punishment
	CATS:E for items on emotion
	Cortisol (Quantity of cortisol in blood).
	CpG., S., SU., SD. (Methylation probability).
	FC1-2701 (Functional connectivity).
	GAF* (Global Assessment of Functioning).
	PHQ9* (Patient Health Questionnaire).
	HRSD17* (17-item Hamilton Rating Scale for Depression).
	HRSD21* (21-item Hamilton Rating Scale for Depression).
	HRSDchange (Increment rate of HRSD17 scores defined as (HRSD17 <sub>6w</sub> - HRSD17)/HRSD17).
	JART* (Adult reading test).
	LES (Life Experiences Survey).
	LES:total for all events
	LES:P for positive events
	LES:N for negative events
	PANASP* (Positive Affect Schedule).
	PANASN* (Negative Affect Schedule).
	SHAPS* (Snaith-Hamilton Pleasure Scale).
	STAI* (State-Trait Anxiety Inventory).
	N*, E*, O*, A*, C* (Five factors in revised NEO Personality Inventory)
Categorical features	
	BDI*_i1_ (Items of BDI initially).
	BDI*_i2_ (Items of BDI after six weeks of treatment).
	drug* (states of dosing of lexapro).
	HRSD*_i1_ (Items of HRSD initially)*.
	HRSD*_i2_ (Items of HRSD after six weeks of treatment)*.
	Melancholic* (Melancholic depression or not).
	MINI* (Mini-International Neuropsychiatric Interview): the numbering corresponds to the following psychiatric symptoms.
	Major depressive disorder (1).
	Dysthymia (2), Suicide risk (3).
	Mania (4), Panic disorder (5), Agoraphobia (6).
	Social phobia (7).
	Obsessive compulsive disorder (8).
	PTSD (9), Alcohol dependence and abuse (10).
	Drug dependence and abuse (11).
	Psychotic disorder (12), Anorexia (13), Bulimia (14).
	Generalized anxiety disorder (15).
	Antisocial personality disorder (16).
	Recurrent* (Recurrent depression or not).
	Response* (whether there is response to the treatment based on HRSD17).
	Remission* (whether a patient is remitted after the tremens).
	Sex
	SNPs* 1-8: Single Nucleotide Polymorphisms that are located in the following genome sites, respectively.
	(in parenthesis are the relevant gene functions)
	rs1187323 (NTRK2), rs34118353 (5HT1a receptor), rs3756318 (NTRK2), rs3813929 (5HT2c receptor),
	rs4554739 (NTRK2), rs56384968 (SLC6A4), rs6265 (BDNF), rs6294 (5HT1a receptor)
Integer features	
	Episode (the number of past experiences of depression).
	RecNum (the number of times of recurrent depression)

表 2: ベイズ共クラスタリング解析に用いた安静時 fMRI データ以外の特徴量のリスト

このデータ行列に対して、ベイズ推定の原理を用いて、参加者と特徴量を同時にクラスタリングし、クラスタ数も自動決定できる多重ベイズ共クラスタリング<sup>9)</sup>を適用した。クラスタリング後、患者群と健常者群の区別が最も良くできるクラスタに注目し、そのクラスタに含まれる安静時機能結合性を同定した。また、各クラスタと抗うつ剤 (セロトニン再取り込み阻害剤; SSRI) に対する治療反応性の関係性について統計的に解析した。

(倫理面への配慮)

本研究は、広島大学研究倫理委員会の承認を得て実施し、すべての参加者から書面による同意を得ている。

## C. 研究成果

### 【検討1: 各データセットに対する最適な AI アルゴリズムの特定と測定バイアスの検討】

安静時脳機能画像を用いていくつかの分類方法を組み合わせて、精度、感度、特異度を用いて、その仮説を検証した (図1)。

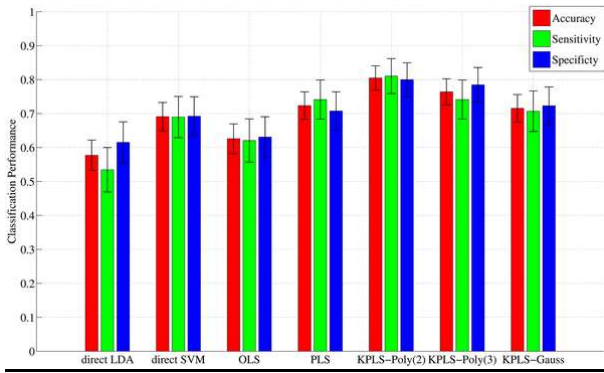


図1 うつ病患者と健常者の判別成績

ナイーブな線形判別分析法 (Linear Discriminant Analysis ; LDA) (精度 57.7%、感度 53.4%、特異度 61.5%) およびサポートベクターマシン (Support Vector Machine; SVM) (精度 69.1%、感度 69.0%、特異度 69.2%) と比べて、2 次多項式カーネルを用いた部分最小二乗回帰法 (Kernel Partial Least Squares regression with the 2nd order Polynomial; KPLS-Poly (2)) は有意に優れた精度を示した (精度 80.5%、感度 81.0%、特異度 80.0%)。この結果は、分類モデルを構築するために臨床測定の予測モデルを利用することの有用性を示唆している。さらに、KPLS-Poly (2) は、通常の最小二乗回帰法 (Ordinary Least Squares Regression; OLS) (精度 62.6%、感度 62.1% および特異度 63.1%) と比べて有意に優れた精度を示し、回帰モデルの潜在空間を考慮することの有用性が示唆された。

安静時脳機能画像の測定バイアスに関しては、異なる施設スキャナーで安静時脳機能画像データを収集した場合、施設間の測定バイアスはうつ病などの精神疾患と健常者の違いと同程度かさらに大きかった。

脳構造画像を用いたうつ病患者と健常者の判別 (診断) に関しては、予備的な検討ではあるが、精度 75.9% (感度 78.1%、特異度 72.9%) で、異なる施設の スキャナー (Scan10,21,22, 23, 30,40,41) で測定した影響は小さかった (図2)。

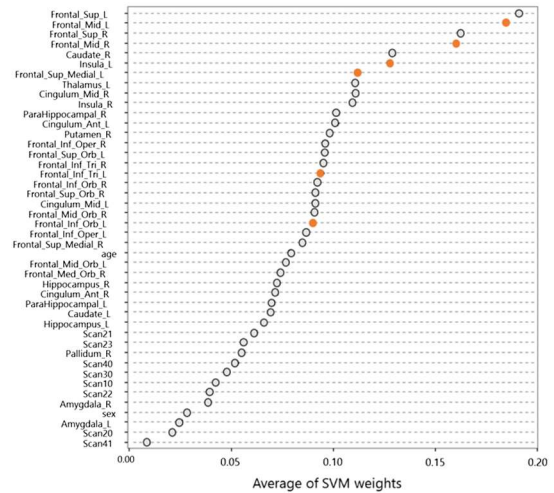


図2 SVM を用いたうつ病診断への寄与度

### 【検討2; 解析パイプライン用パラメータの網羅的探索】

B.節で述べたあらゆるパラメータ設定の組合せに対して診断モデルの予測精度を評価した。その結果、開眼状態での安静時 fMRI データに対して、BAL<sup>6)</sup> の脳領域分割法による機能結合性を定義したものを特徴量として、診断モデルへの入力前に一般線形モデルによる攪乱変数除去と有意水準 1% の二群検定法による特徴量選択の前処理を施した時が最適となり、その時の精度は 0.75、感度は 61% となった (表3)。

開眼/閉眼	脳領域分割	変数選択	攪乱変数除去	精度	感度
開眼	BAL	二群検定法 ( $\alpha=0.01$ )	一般線形モデル	0.7483	0.6081
閉眼	BAL	SCCA	SCCA	0.7196	0.6028
開眼	Stanfordx90	二群検定法 ( $\alpha=0.01$ )	一般線形モデル	0.6962	0.4918
閉眼	BAL	SCCA+二群検定法 ( $\alpha=0.03$ )	SCCA	0.6869	0.5354
開眼	Stanfordx90	SCCA+二群検定法 ( $\alpha=0.05$ )	SCCA	0.6801	0.4621
閉眼	Stanfordx90	SCCA	SCCA	0.6747	0.4874
開眼	BAL	二群検定法 ( $\alpha=0.03$ )	一般線形モデル	0.6737	0.5068
閉眼	BAL	SCCA+二群検定法 ( $\alpha=0.01$ )	SCCA	0.6729	0.5181
開眼	BAL	SCCA+二群検定法 ( $\alpha=0.01$ )	SCCA	0.6643	0.4589
閉眼	BAL	SCCA+二群検定法 ( $\alpha=0.05$ )	SCCA	0.6589	0.5272
開眼	BAL	SCCA+二群検定法 ( $\alpha=0.05$ )	SCCA	0.6573	0.4797
閉眼	BAL	二群検定法 ( $\alpha=0.03$ )	一般線形モデル	0.6542	0.5068
閉眼	Stanfordx90	SCCA	SCCA	0.6434	0.4394
閉眼	Stanfordx499	SCCA+二群検定法 ( $\alpha=0.05$ )	SCCA	0.641	0.4058
閉眼	Stanfordx90	SCCA+二群検定法 ( $\alpha=0.05$ )	SCCA	0.6381	0.4347
閉眼	Stanfordx90	二群検定法 ( $\alpha=0.05$ )	一般線形モデル	0.6381	0.459
閉眼	BAL	SCCA+二群検定法 ( $\alpha=0.03$ )	SCCA	0.627	0.4604
閉眼	Stanfordx499	SCCA+二群検定法 ( $\alpha=0.05$ )	SCCA	0.6247	0.438
閉眼	Stanfordx499	SCCA	SCCA	0.62	0.4072
開眼	Stanfordx499	二群検定法 ( $\alpha=0.05$ )	一般線形モデル	0.6084	0.4392

表 3: 解析パイプラインのための網羅的パラメータ探索結果 (精度に関して上位 20 位までを抜粋)

### 【検討3; SLR+SCAA 法によるうつ病判別と外部データへの汎化性能の検証】

全ての対象では判別率 51%、AUC 0.52 にとどまったため、うつ病の中でも生物学的要因の影響が大きいとされるサブタイプ、メランコリー型の特徴を有する一群に絞った検討を行った。メランコリー型うつ病 66 名と同数の健常者を合わせた計 132 名の安静時 fMRI データを用いて判別器の作成を行い、一個抜き交差検証 (LOOCV) を行った

結果、判別率 70%、AUC:0.77 まで成績が向上した。このバイオマーカーを、完全に独立な施設のデータセットに適用した結果、判別率は 65%(AUC 0.62)であり、汎化性能が確認された。判別器として抽出された 12 の脳機能結合の中には、うつ病の先行研究において重要であることが示唆されている脳領域が多数含まれていた。さらに、このバイオマーカーの臨床的意義や有用性を明らかにするために、判別器の値 (Weighted Linear Sum: WLS) とうつ病重症度との関連について検討した結果、この値は抑うつ症状の質問紙得点 (BDI) と相関を示し、6 週間の抗うつ薬治療により健常方向に変化したことから、うつ状態を反映するバイオマーカーである可能性が示唆された。

#### 【検討4: ベイズ共クラスタリング法によるデータ駆動的うつ病サブタイプ推定】

B.節で述べたデータ行列に対して、多重ベイズ共クラスタリングを適用した結果、15 種類の共クラスタ構造が得られた(図 3)。

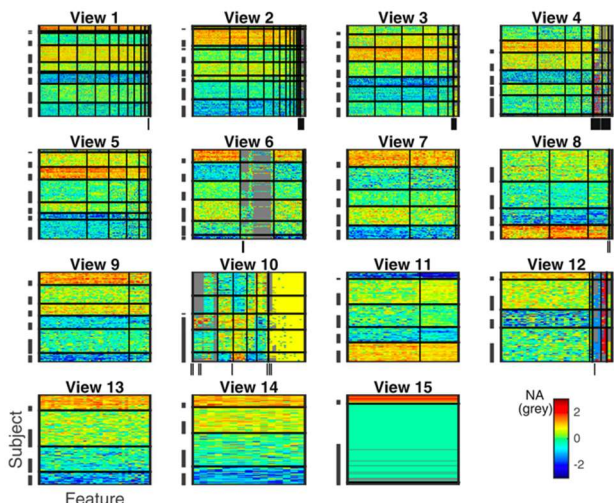


図 3: 多重ベイズ共クラスタリングの適用結果(縦軸は参加者、横軸は特徴量を表しており、参加者軸に付随したダッシュ記号はうつ病患者を表している。ヒートマップ中の太線はクラスタの境界を示している。)

このうち、View 10 で示されるクラスタ構造は、他の View に比べても特にうつ病患者群と健常者群の分類とも一致しており、最初の 2 つのクラスタは健常者群のみで構成されている一方で、残りの 3 つのクラスタに属する参加者のほとんどがうつ病患者となった。

そこで、View 10 で示されるクラスタ構造について、より詳細に調べた。まず、このクラスタ構造に含まれる安静時機能結合性を抽出したところ、右角回を中心とするスポークアンドハブ構造の機能ネットワーク(図 4)がこのクラスタを特徴づけていることが分かった。

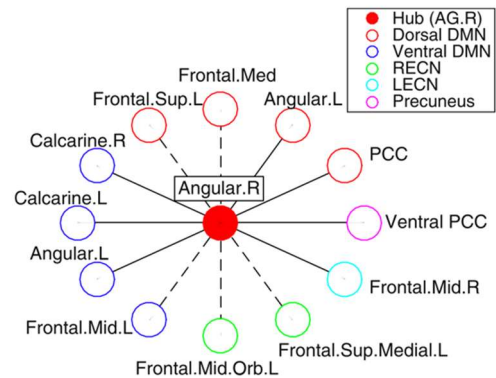


図 4: 図 3 の View 10 のクラスタ構造に含まれる安静時機能結合ネットワーク

また、ほとんどがうつ病患者から構成される 3 つのクラスタの違いを調べたところ、BDI や HRSD によって評価されたうつ病重症度の初診時と SSRI 投薬開始後 6 週間(または 6 ヶ月)後の差、すなわち、SSRI に対する治療反応性がこれらのクラスタの特徴を反映していることが分かった。また、これら 3 つのクラスタの違いには、幼児期トラウマ指標である CATS も関連していることがわかった。そこで、View 10 に含まれる機能結合性の第 1 主成分と CATS の 2 次元でデータの分布を可視化したところ、図 5 のようになった。すなわち、右角回を中心とした機能ネットワークにより、D3 クラスタ(治療反応性が良いクラスタ)への割当てが決まり、残りの 2 つのクラスタについては、CATS が小さければ D2 クラスタ(治療反応性が良いクラスタ)へ、CATS が大きければ D1 クラスタ(治療反応性が悪いクラスタ)へと割り当てが決まる構造があることが分かった。

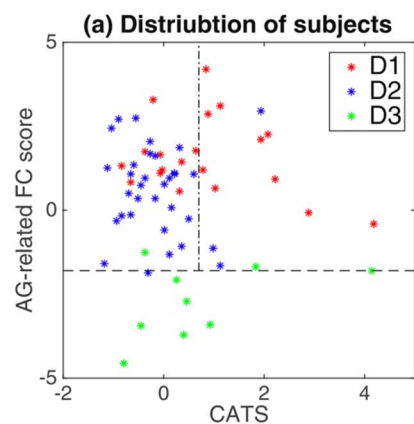


図 5: 右角回を中心とした機能ネットワークの第 1 主成分と幼児期トラウマ指標 (CATS) の散布図(各点の色の違いは、View 10 内で所属するクラスタの違いを表している)

#### D. 考察

各データセットに対する最適な AI アルゴリズムの特定と測定バイアスの検討、解析パイプラインの最適化

に向けたパラメータの網羅的探索、および、医師の診断情報を用いない教師なし学習法に基づくデータ駆動的なうつ病サブタイプ分類法の開発に取り組んだ。

検討1の最適なAIアルゴリズムの特定と測定バイアスの検討の結果から、安静時脳機能画像に関しては、臨床測定の予測モデルを媒介すること、低次元の特徴空間を考慮することで、比較的小サンプルでも分類性能の向上が期待できることを明らかにした。また、安静時脳機能画像は測定バイアスの影響が大きく、スパース推定による特徴選択、施設効果の線形回帰による regression-out、アンサンブル学習、traveling subjects を用いた測定バイアスの分離と推定といった数理統計手法の利用が測定バイアスの補正に必要であると考えられた。一方、構造画像に関しては、SVM に適用した結果からは異なる施設のスキナーで測定した影響(測定バイアス)は小さいことが明らかになった。

検討2のパラメータの網羅的探索の結果から、安静時 fMRI 撮像時では開眼状態の方が、最適な予測精度が実現できるという意味では望ましいことが示唆される。しかしながら、他のパラメータに依存して閉眼時の方が予測精度が良くなる場合もあり、実用上、開眼・閉眼のどちらが良いかについては一貫性のある結果は得られなかった。また、脳領域分割法については BAL の結果が総じて良かった。変数選択法と攪乱変数除去法については、一般線形モデルによる攪乱変数除去と二群検定法(有意水準 1%)の組合せが最適な結果となったが、この組合せを変えても、BAL による脳領域分割法を用いている限りは精度が大きく落ちるということにはなかった。したがって、現状の結果からは、BAL を用いた脳領域分割が特に推奨される。一方で、昨年度までの結果から入力変数の変数選択(または低次元化)と測定バイアス除去のための攪乱変数除去法の導入は必須であるものの、その方法については大きく精度の改善や悪化につながるようなものは同定できなかった。

検討3の結果から、完全な独立コホートに汎化するメランコリア特徴を有するうつ病の安静時機能結合の判別器(図6)を作成した。この判別器の12結合の線形加重和は、抑うつ症状と有意な正の相関があることから、状態を反映するバイオマーカーであると考えられた。

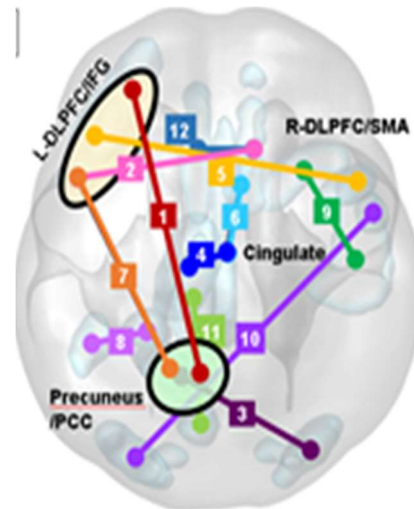


図 6: 完全な独立コホートに汎化するメランコリア特徴を有するうつ病の安静時機能結合の判別器

検討4のサブタイプ分類法の開発においては、データ駆動的に治療反応性の違いに応じてうつ病患者群が3つのグループに分類できることを見出した。また、この3つのグループは右角回を中心とした12個の安静時機能結合性と幼児期トラウマ経験の大小によって特徴づけられていた。現在のところ、まだ少サンプルの解析結果であるため、その再現性は確認できていないが、これらの結果は、SSRI に対する治療反応性予測するためのモデルとして図7の決定木が有望であることを示唆している。

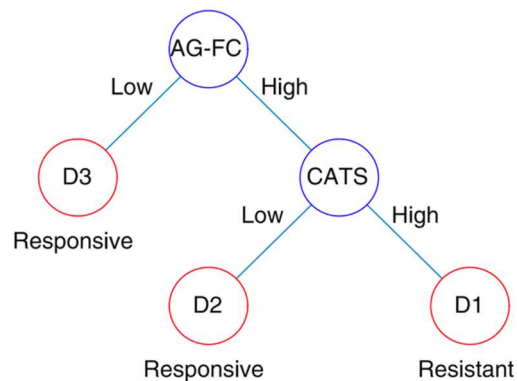


図 7: 右角回を中心とした機能ネットワークの安静時機能結合性と幼児期トラウマ指標に基づく治療反応予測モデル(AG-FC は右角回を中心とした12個の安静時機能結合性のスコアであり、CATS は幼児期トラウマ指標である。)

## E. 結論

本研究では、まず、検討1において、安静時脳機能画像に関しては、スパース推定 L1-Sparse Canonical Correlation Analysis(L1-SCCA) & Sparse Logistic Regression (SLR)による特徴選択、施設効果の線形回帰による regression-out、あるいは traveling subjects を用いた施設効果の測定バイアスの

分離と推定といった数理統計手法の利用がデータ解析に際しては必要と考えられた。また、測定バイアスに対する分類器の頑健性はアンサンブル学習の導入によって改善されることが期待された。一方、脳構造画像に関しては、これらの工夫や補正の必要性が小さいものと考えられた。

検討2では、解析パイプラインの最適化に向けたパラメータの網羅的探索を行った。その結果、脳領域分割法の選択が最も予測精度に良し悪しに影響し、現在利用できるものの中では BAL<sup>®</sup>が推奨されることが分かった。fMRI 撮像時の眼の開閉状態については、どちらを採用しても大きな違いはなく、変数選択法と攪乱変数除去法の導入は必須ながらも、その具体的な方法については、合理的なものであれば顕著な違いはなかった。

検討3では、安静時脳機能結合を用いて脳全体にわたる機能的結合のうち、ごく一部(わずか 0.2%)の結合から個人の『うつ度』(バイオマーカー)を測り、その大小でうつ病と健常対照者を判別することが可能となった。これらは独立した外部データにおいて汎化性能も確認できた。

検討4では、DSMに基づく医師の診断結果を明に用いない、教師なし学習法に基づくデータ駆動的なうつ病のサブタイプ分類をベイズ共クラスタリング手法を用いて試みた。その結果、治療反応性の違いに応じてうつ病患者群が3つのグループに分類できることを見出し、そのグループの違いは右角回を中心とした12個の安静時機能結合性と幼児期トラウマ経験の大小によって特徴づけられていることを見出した。

## F. 健康危険情報

なし

## G. 研究発表

### 1. 学会発表・招致講演等

Ichikawa N, Lisi G, Yahata N, Okada G, Takamura M, Hashimoto R, Yamada T, Yamada M, Suhara T, Moriguchi S, Mimura M, Yoshihara Y, Takahashi H, Kasai K, Kato N, Yamawaki S, Seymour B, Kawato M, Morimoto J, Okamoto Y. A classifier of melancholic depression with whole-brain resting-state connectivity. rtFIN2017, Nara, Japan, 2017.12.1, 市川奈穂, 岡本泰昌. 安静時fMRI活動を用いたうつ病の判別, 第15回日本うつ病学会, 東京, 2018.7.27 岡本泰昌, 市川奈穂. 脳機能画像研究からみたうつ

病の異種性, 第15回日本うつ病学会, 東京, 2018.7.28

岡本泰昌. 神経回路病態に基づくうつ病の診断・治療法の開発, 第26回脳の世紀シンポジウム『AIと脳』, 東京, 2018.9.12

Ichikawa N, Lisi G, Yahata N, Okada G, Takamura M, Hashimoto R, Yamada T, Yamada M, Suhara T, Moriguchi S, Mimura M, Yoshihara Y, Takahashi H, Kasai K, Kato N, Yamawaki S, Seymour B, Kawato M, Morimoto J, Okamoto Y. Melancholic depression biomarker of resting-state functional connectivity. AsCNP-ASEAN2019, Yogyakarta, Indonesia, 2019.3.2

## H. 知的財産権の出願・登録状況

知的財産の内容(うつ症状の判別方法、うつ症状のレベルの判定方法、うつ病患者の層別化方法、うつ症状の治療効果の判定方法及び脳活動訓練装置)、種類・番号 PCT/JP2018/36952、出願年月日 2018.10.02、取得年月日、権利者国立大学法人広島大学、国際電気通信基礎技術研究所(ATR)

## 参考文献

- 1) World Health Organization. (2017). Depression and other common mental disorders: global health estimates. World Health Organization.
- 2) American Psychiatric Association. (2000). Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision (DSM-IV-TR).
- 3) Shimizu, Y., et al. (2015). Toward Probabilistic Diagnosis and Understanding of Depression Based on Functional MRI Data Analysis with Logistic Group LASSO. PLOS ONE, 10(5), e0123524.
- 4) Ichikawa N., et al. (2017). Identifying melancholic depression biomarker using whole-brain functional connectivity. arXiv:1704.01039.
- 5) Trivedi, M. H., et al. (2006). Evaluation of outcomes with citalopram for depression using measurement-based care in STAR\*D: Implications for clinical practice. American Journal of Psychiatry, 163(1), 28-40.
- 6) Yahata, N., et al. (2016). A small number of abnormal brain connections predicts adult autism spectrum disorder. Nature Communications, 7, 11254.
- 7) Neuropsychiatric Disorders Lab at Stanford University. Functional ROIs, [https://findlab.stanford.edu/functional\\_ROIs.html](https://findlab.stanford.edu/functional_ROIs.html).
- 8) Shirer W. R., et al. (2012). Decoding Subject-

Driven Cognitive States with Whole-Brain Connectivity Patterns. *Cerebral Cortex*, 22(1), 158–165.

- 9) Tokuda, T., et al. (2017). Multiple co-clustering based on nonparametric mixture models with heterogeneous marginal distributions. *PLOS ONE*, 12(10), e0186566.
- 10) Yoshida K, Shimizu Y, Yoshimoto J, Takamura M, Okada G, Okamoto Y, Yamawaki S, Doya K. Prediction of clinical depression scores and detection of changes in whole-brain using resting-state functional MRI data with partial least squares regression. *PLoS One*. 2017, 12: e0179638.

平成29・30年度厚生労働科学研究費補助金  
(政策科学総合研究事業 (臨床研究等 ICT 基盤構築・人工知能実装研究事業))  
分担研究報告書

表情、音声などのセンシングデータの人工知能解析

研究分担者 岸本泰士郎 慶應義塾大学医学部精神・神経科学教室 専任講師

## 研究要旨

うつ病と認知症はそれぞれへの治療アプローチが異なるため、鑑別が極めて重要である一方で、両者の鑑別に有効かつ簡便なバイオマーカーは存在しない。このため、両者の鑑別は、記憶障害の程度に対する機能障害の程度、自身の記憶障害への自覚の程度など臨床的な判断に頼らざるを得ないのが現状である。人工知能技術を利用して両者の鑑別が可能になれば有用であるし、また人工知能に判断を委ねるのではなくとも、どのような指標が両者の分別に（機械学習の特徴量として）有用なのかが明らかになることで、病態の理解につながり、臨床的に有用な示唆となる。

本分担研究課題においては、学習モデル構築および精度検証に用いるデータセットとして、日本医療研究開発機構（AMED）の委託研究「表情・音声・日常生活活動の定量化から精神症状の客観的評価をリアルタイムで届けるデバイスの開発（研究代表者：岸本泰士郎）」のデータを用いた。この研究では、診療場面で医師と患者が行う通常の会話の様子を録音・録画している。

初年度は、うつ病や認知症の症状が明確に存在するデータセットをそれぞれ 51、56 データセット抽出し、音響学的特徴を解析対象とした。音声を Formant frequency 等の指標に変化した後、特徴エンジニアリングを行い、support vector machine による分類器の生成を行った。結果、うつ病と認知症の分類精度は 82.2%であった（一個抜き交差検証による）。次年度では、年齢を揃えたうつ病あるいは認知症患者それぞれ 64 名、74 名分のデータセットを用い、音響学的特徴および表情特徴を解析対象とした。support vector machine with cubic kernel による分類器の生成を行った結果、うつ病と認知症を分類する精度は 89.9%であった（10-fold cross validation による）。本検証結果は前述の研究に基づくデータセットを対象にしており、外部データでの検証を行う必要があるが、高い精度での分別が可能だったことから、今後、実用化に向けての開発が望まれる。

## A. 研究目的

うつ病と認知症の鑑別は臨床上、しばしば困難である。両者の合併例も存在するが「仮面認知症」のように臨床表現型は認知症様であってもうつ病の治療に反応する一群の患者も存在する。うつ病と認知症では治療アプローチが異なるため、両者の鑑別は極めて重要である。一方で、両者の鑑別に有効かつ簡便なバイオマーカーは存在せず、画像検査においても特に初期の認知症の判断は困難である。その結果、両者の鑑別は、ライフイベントの関与の程度、記憶障害の程度に対する機能障害

の程度、自身の記憶障害への自覚の程度、さらには抗うつ薬への反応性など臨床的な判断に頼らざるを得ないのが現状である。人工知能技術を利用して両者の鑑別が可能になれば有用である。あるいは人工に判断を委ねなくとも、こういった指標が両者の分別に（機械学習の特徴量として）有用なのかが明らかになることで、病態の理解につながり、臨床的に有用な示唆となる。

## B. 研究方法

学習および判別能力の精度検証に用いたデータ

セットとして、日本医療研究開発機構 (AMED) の委託研究で慶應義塾大学が代表機関を務める「表情・音声・日常生活活動の定量化から精神症状の客観的評価をリアルタイムで届けるデバイスの開発(研究代表者:岸本泰士郎)」のデータを用いた。この研究では、診療場面で医師と患者が行う通常の会話の様子を 10-30 分間、録音、録画し (いわば診療の様子を機械に観察させ)、音声や画像データを種々のソフトウェアを用いて数値化し、機械学習に用い、重症度分類等を試みている。現在までに、種々の重症度のうつ病・双極性障害患者 172 名、834 データセット、認知症 60 名、218 データセットが蓄積されている。

初年度の研究開発においては、前述の目的に向けての初めのステップとして、典型的なうつ病患者および認知症患者の分類を行う学習モデルを生成し、その精度を検証した (うつ病 51 データセット、認知症 56 データセットを利用)。初年度は診療観察データのうち、音声音響学的要素のみを対象にした。患者音声に解析には Praat を用いた。Praat は音声音響学的分析にしばしば用いられるソフトウェアで、話者の Formant、Pitch、Intensity、Pulse 等を解析することが可能である<sup>1,2)</sup>。

また、2 年度の研究開発においては、以下のように年齢を揃えたデータセットを用意した。

- ・ 57-81 歳 (データセット数が確保でき、2 群に年齢差が生じない範囲として任意に設定)
  - ・ 10 分以上の自由会話の動画データがあるもの
  - ・ 認知症患者の場合：Mini Mental State Examination のスコアが 23 点以下、かつ、Geriatric Depression Scale (Long Form) において 10 点未満のもの
  - ・ うつ病患者の場合：Hamilton Rating Scale for Depression 17 項目版のうち 8 点以上のもの
- 結果、上述を満たした、うつ病、認知症それぞれ 64 名、74 名分のデータを抽出し、研究用のデータセットとした。また、2 年度の研究開発においては、初年度同様、Praat を用いた音響学的な特徴量を利用しただけでなく、表情データから得られる

特徴量を利用した。表情の解析に際しては、オムロン社の OKAO Vision を用いた。OKAO Vision は顔検出・顔器官検出を自動で行い、表情 (喜び・驚き・怒り・悲しみ・無表情)、瞬目などを定量することが可能である<sup>3,4)</sup>。これらのうち、当研究グループによる種々の施行から、うつ病重症度等の解析に有用であった要素を利用し、特徴エンジニアリングを行い、次に Lasso (least absolute shrinkage and selection operator) を用いて特徴選択を行った。うつ病あるいは認知症と標識したデータで support vector machine with cubic kernel による分類器の生成を行った。

(倫理面への配慮)

本研究の施行に際しては、人を対象とする医学系研究の倫理指針に準拠した倫理委員会に諮り、承認を得て行った。また、全ての被験者から同意を取得して行った。

### C. 研究成果

初年度の音声のみを用いたうつ病・認知症の判別 (認知症を同定する) の精度、感度、特異度は一個抜き交差検証 (Leave one out cross validation) で以下の様であった。

	精度	感度	特異度
音声データのみ	0.822	0.824	0.821

また、2 年度に取り組んだ音声および表情データを用いた精度、感度、特異度を以下に示す (10-fold cross validation による) 音声、表情それぞれ単独のモダリティを用いた検証は以下の様であった。

	精度	感度	特異度
音声データのみ	0.884	0.919	0.844
表情データのみ	0.826	0.813	0.838

音声・表情データ両者を用いた分類器の精度、感度、特異度を以下の様であった

	精度	感度	特異度
音声・表情データの組み合わせ	0.899	0.919	0.875



## D. 考察

診療場面から得られる音声および表情データを用いたうつ病と認知症の分類器の精度は 89.9%に達した。なお、本検証結果は前述の研究に基づくデータセットのみを対象にしており、外部データによる検証を行う必要がある。

今後、さらなるデータ収集や学習モデルの改善を通じて、分類の精度をあげることでより臨床上の有用性は増すものと考えられる。また、現在は 10 分程度の撮影データを元に解析を行っているが、より短時間でも可能かどうかの検証も有用と考える。将来は治療反応を予測するような臨床的により有用な予測技術の開発が求められる。

## E. 結論

診療場面で医師と患者が行う通常の会話の様子を録音・録画したデータを用いて 90%程度の精度でうつ病と認知症の分類が可能であった。さらなるデータ収集や学習モデルの改善を通じて、分類の精度をあげることでより臨床上の有用性は増すものと考えられる。また、現在は 10 分程度の撮影データを元に解析を行っているが、より短時間でも分別が可能かどうかの検証も有用と考える。さらに、本研究では、うつ病で認知機能の低下を認めない症例と、認知症でうつ症状を認めない症例とで構成されるデータセットを用いて分類器の生成を行ったが、今後は、鑑別が実際に困難だったケースなどでの検討が必要である。

上述のような課題はあるものの、90%程度での分別能力は実臨床での使用を考えた際にも有用であり、今後は実用化に向けての開発が望まれる。

## F. 健康危険情報

なし

## G. 研究発表

### 1. 学会発表・招致講演等

岸本泰士郎. 情報通信技術や人工知能を用いた精神科領域における新たな試み,第38回多摩精神科

臨床研究会, 東京, 2017.11.8

岸本泰士郎. ICTや機械学習を用いた精神科領域における新しい試み,慶應義塾大学ウェルビーイング研究センターリサーチコンプレックス推進事業設置講座, 神奈川県, 2017.11.30

岸本泰士郎. 情報通信技術によって開かれる精神科診療の展望と課題. 第55回兵庫県精神医療学術講演会, 神戸, 2018.1.20

岸本泰士郎. ICTやAIを活用した今後の精神科医療の展望& 3つの試験デザインに基づくLAIの有効性の検討. 第129回福井県神経科精神科医会, 福井, 2018.2.2

岸本泰士郎. AI を用いた精神疾患臨床症状定量化の試み, 平成 29 年度 医薬品評価委員会 臨床評価部会総会, 東京, 2018.2.21

岸本泰士郎. 精神疾患・認知症診断におけるテクノロジー利活用の課題と展望, 東海大学認知症研修会プログラム, 神奈川, 2018.2.27

岸本泰士郎.情報通信技術 (ICT) や人工知能 (AI) を活用した診断支援の試み. 第30回東京都医師会医療とITシンポジウム, 2018.3.10

岸本泰士郎. AI はこころの病を診ることができるか? -人工知能を用いたうつ症状定量化の試み- . ぐんま・脳とこころのアカデミー, 群馬, 2018.03.24

岸本 泰士郎. 機械学習による精神運動制止の評価の試み, 第114回日本精神神経学会学術総会, 神戸, 2018.6.21

岸本 泰士郎. 情報通信技術や機械学習を活用した臨床症状評価, 第114回日本精神神経学会学術

総会, 神戸, 2018.6.21

岸本 泰士郎. 情報通信技術 (ICT) や人工知能 (AI) の活用という観点で30年後の精神科医療を展望する, 第114回日本精神神経学会学術総会, 神戸, 2018.6.23

岸本 泰士郎. 情報通信技術や人工知能がもたらす近未来の精神医学の展望, 第114回日本精神神経学会学術総会, 神戸, 2018.6.23

澤田 恭助, 高宮 彰紘, 岸本 泰士郎, 三村 将. うつ病患者に対する音声データの臨床的有用性の検討: システムティックレビューとメタ解析, 第114回日本精神神経学会学術総会, 神戸, 2018.6.22

岸本 泰士郎. 遠隔医療は国民に寄り添った医療になるか? 本邦における遠隔精神科医療の展望と課題, 日本精神神経科診療所協会 第24回学術研究会, 兵庫, 2018.6.24

岸本 泰士郎. 情報通信技術 (ICT) や機械学習を用いたうつ病診療の展望, 第3回CNSサミット, 東京, 2018.8.5

岸本泰士郎. 情報通信技術や機械学習を用いた精神症状定量化の試み, 日本線維筋痛症学会第10回学術集会, 東京, 2018.9.29

岸本 泰士郎. 遠隔医療や機械学習を活用した認知症診療の展望, 第37回日本認知症学会学術集会, 札幌, 2018.10.13

工藤弘毅, 岸本泰士郎. 「メンタルヘルスリテラシーと人工知能 (AI)」, 第38回日本精神科診断学会, 埼玉, 2018.10.19

岸本泰士郎. 情報通信技術や人工知能を活用した

近未来の精神科医療の展望, 第3回 Venlafaxine研究会, 神奈川, 2018.10.24

岸本泰士郎. 情報通信技術や人工知能を活用した近未来の医療の展望, 平成30年度女性医師の勤務環境に関する病院長等との懇談会, 沖縄, 2018.10.30

岸本泰士郎. 精神科医療の遠隔化は診療所、精神病院、総合・大学病院に普及し得るか, 第38回医療情報学連合大会 (第19回日本医療情報学会学術大会), 福岡, 2018.11.23

岸本泰士郎. 情報通信技術や機械学習を活用した精神科領域の展望, 第2回デジタルヘルス学会, 東京, 2018.12.23

岸本泰士郎. 人工知能技術を用いた精神疾患症状定量化の試み, 第1回日本メディカルAI学会, 東京, 2019.1.26

H. 知的財産権の出願・登録状況  
出願準備中

#### 参考文献

1. Boersma P and Weenink D. Praat: doing phonetics by computer [Computer program]. Version 6.0.37, retrieved 14 March 2018 from <http://www.praat.org/>
2. Boersma P and Weenink D. Praat, a system for doing phonetics by computer. *Glott International* 5(9/10): 341-345. 200114; 33(2): 219-37.
3. Konishi Y, Kinoshita K, Kawade M. Sparse bayesian regression for head pose estimation. 18th International Conference on Pattern Recognition (ICPR'06). Vol 0, Hong Kong, China, 20-24, August 2006:507-510.
4. Kitazawa M, Yoshimura M, Kuo-Ching

L, Wada S, Mimura M, Tsubota K, Kishimoto T.  
Utilization of Facial Image Analysis Technology  
for Blink Detection: A Validation Study. Eye  
Contact Lens. 2018. Nov;44 Suppl 2:S297-S301.

平成29・30年度厚生労働科学研究費補助金(政策科学総合研究事業)  
分担研究報告書

ライフログデータに基づく寛解期うつ病患者の再燃再発予測

研究分担者 古川壽亮 京都大学大学院医学研究科健康増進・行動学分野 教授  
研究分担者 吉本潤一郎 奈良先端科学技術大学院大学先端科学技術研究科情報科学領域 准教授  
研究協力者 田近亜蘭 京都大学医学部附属病院精神科神経科 助教  
研究協力者 徳田智磯 奈良先端科学技術大学院大学先端科学技術研究科情報科学領域 博士研究員

研究要旨

うつ病で維持治療中の患者89名を対象に、スマートフォンを利用して半自動的に採取される活動記録ライフログおよびウェアラブルデバイスによって記録されるデータを精神症状の評価とともに、12ヶ月間にわたり採取した。この諸変数を予測因子として変数選択した上で、うつ病の増悪を予測するモデルを構築したところ、ゴロゴロ時間、通勤・通学時間、および睡眠時間という、活動量に関する変数の組み合わせで、AUCが0.7を越えるモデルを作成できた。これらのモデルは最大2週間前まで偶然を越える予測能を示した。特異度を高く維持するカットオフを利用することにより、ワーニング疲れを引き起こさずに、患者に再燃を予測警告できるシステムが構築できる可能性が示唆された。

A. 研究目的

うつ病は治療により一旦寛解したとしても再燃・再発が多い疾患として知られる。再燃・再発に対して早く対応することで治療効果が上がるため、その早期発見が大きな課題となっている。再燃・再発が起こると、抑うつ気分のために活動性が低下することは周知の事実ではあるものの、具体的に再燃・再発に先だって活動パターンがどのように変化するかを今までは詳しく調べることはできなかった。これは、簡便に活動を記録できるツールがなかったことに加え、うつ病患者が継続的に自身の活動を記録し続ける難しさによることが大きかった。

そこで、われわれは患者の活動パターンを負担少なく記録できる装置を開発してきた。具体的には、近年普及著しいスマートフォン自体に搭載されたセンサーから取得した位置情報や歩数情報などのライフログ情報を得ることができる。これらの情報を適宜サーバにアップロードすると、このライフログ情報と、それまでに入力された活動とを照合し、当日の活動内容をコンピューターが機械学習アルゴリズムにより予測を行い、大まかな活動記録表を自動的に作成することができる。コンピューターによる予測は時に外れるため、スマートフォン上で自動的に作成された活動記録表を患者各自がチェックして、修正する作業は必要であるものの、患者はスマートフォン上に表示された、あらかじめ推測された活動を承認、または微修正するだけで良いので、日々の負担はかなり軽減することができる。我々はこのような機能を持った活動記録表作成支援アプリケーション(“くらしアプリ”)を開発し、2015年度

には健常成人42名を対象としたオープンパイロット研究を行った。その結果、行動の予測精度は70%近くあり、従来式の紙媒体での記載に比べて、アプリの方が負担が少ないと回答した人の割合は85.7%に達し、この方法の実施可能性が検証された。

加えて、近年、日々の活動を簡単に記録するツールとして、様々なウェアラブルデバイスが市販されている。東芝(その後TDK)の開発したウェアラブルデバイス、シルミーW20は、リストバンド型の装置で、これを着用しているだけで、活動量や睡眠時間などのデータに加え、くらしアプリでは取得できない日々の会話量や紫外線量などのデータが自動測定される。

本研究では、うつ病の維持治療中で再燃再燃の可能性が高い患者を対象に、くらしアプリとシルミーW20を継続記録し、これにより2~4週間後のうつ病再燃再燃を予測可能かを検証した。

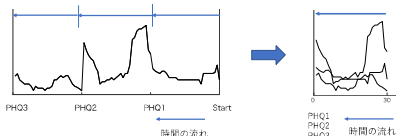
B. 研究方法

名古屋市立大学、高知大学、広島大学、東邦大学の4つの大学病院と、それらの関連の精神科病院とクリニックの外来に通院する寛解期の大うつ病患者から89名が本研究に参加した。被験者自身のiPhoneに“くらしアプリ”をインストールすると共に、ウェアラブルデバイス“シルミー”を装着してもらい、ライフログ情報を収集した。得られたライフログ情報は、歩数、移動距離、仕事・勉強時間、睡眠時間等の日々の活動に関連した60個の特徴量からなる。また、うつ状態を診断する指標としてK6(1週間ごと)、BDI-II(4週間ごと)を被験者の自己記入により、PHQ-9(4週間ごと)を臨

床試験コーディネーターの電話インタビューによって収集した。

再発予測モデルの構築にあたり、PHQ 値の増減に着目し、PHQ が前回の測定から5ポイント増加した場合を「再発有り」、それ以外を「再発無し」と定義した。再発予測に有用な特徴量選択のための予備解析として、PHQ 測定日を基点とした event-triggered average 法によって、「再発有り」の場合と「再発無し」の場合の時系列を統計的に比較した。具体的には、すべての特徴量の時系列データを PHQ の測定日から30日前までを一区切りとして抽出した。次に、PHQ 測定日から15日前まで、日ごとに各特徴量の2群間差を U-test (有意水準 0.05) で評価した。さらに、有意な違いが連続して起きる日数が偶然では説明できないものを permutation test (有意水準 0.01) で評価することで、再発予測モデルのための説明変数として抽出した (図 1)。

前処理：特徴量ごとに時系列データからPHQ測定前30日間の区分を切り取り、PHQ測定日を0日とした時系列データを作成する。



2群間の有意差検定：前処理したデータを合わせて、各特徴量ごと・タイムポイントごとにU-testを行う。



連続日数の有意差検定：再発有り・無しのラベルをシャッフルして上述の有意差検定を行うことにより、連続日数についてのサンプルを作り、実際に観測された連続日数のP値を評価する。

図 1 データ分析の概略

本解析では、以上で得られた説明変数から従属変数である再発有無のラベルをどの程度予測できるかについて、コックス比例ハザードモデルを用いて検討した。モデルパラメータは最尤推定法により決定し、モデルの出力である生存確率が 0.5 以上の時は「再発無し」、0.5 未満の時は「再発有り」を予測結果とした。

あらゆる説明変数の組合せの中で、最良の汎化性能が期待できるものを決定するために、Leave-one-out 交差検証法によって得られたテストサンプルの予測結果と正解の比較を精度、感度、特異度、および、ROC (Receiver Operating Characteristic) 曲線の AUC (Area Under the Curve) スコアを各モデルで計算した。最後に、最良のモデルが、ライフログデータを用いて何日前に再発を予測できるか統計的に検証した。

(倫理面への配慮)

本研究は、京都大学大学院医学研究科医の倫理委員会の承認を得て、すべての参加者から書面による同意を得ている。

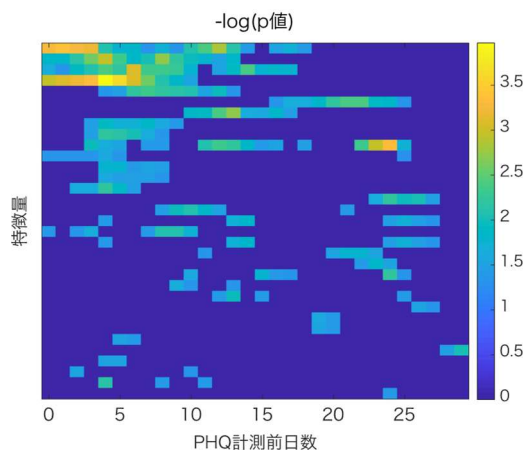


図 2 2群 (再発有り、無し) の違いに関する U-test の P 値 (濃い青は有意差なし)。特徴量は連続した有意日数が多いに並び替えた。上位 5 つの特徴量は順に、ゴロゴロ時間、走行回数、過去1週間のエネルギー消費量平均、睡眠時間である。

### C. 研究成果

予備解析における特徴量選択の結果、「ゴロゴロ時間」、「走行回数」、「通勤・通学時間」、「睡眠時間」、「過去1週間のエネルギー消費量平均」の5つの特徴量が抽出された (図 2)。

抽出された特徴量のすべての組み合わせに対して、コックス比例ハザードモデルを当てはめ、交差検定を行った結果 (表 1)、AUC が 0.70 以上の組み合わせは以下の通りであった。

- ・ 走行回数 (0.74)
- ・ ゴロゴロ時間 + 走行回数 + 過去1週間のエネルギー消費量平均 (0.72)
- ・ 走行回数 + 通勤・通学時間 + 睡眠時間 + 過去1週間のエネルギー消費量平均 (0.72)
- ・ ゴロゴロ時間 + 走行回数 (0.71)

特徴量	精度	感度	特異度	AUC
走	0.25	1.00	0.17	0.74
ゴ+走+エ	0.25	0.95	0.17	0.72
走+通+睡+エ	0.25	0.91	0.18	0.72
ゴ+走	0.27	0.96	0.19	0.71
ゴ+走+通	0.28	0.92	0.21	0.71
ゴ+走+睡+エ	0.24	0.91	0.17	0.71
<b>ゴ+通</b>	<b>0.60</b>	<b>0.64</b>	<b>0.59</b>	<b>0.70</b>
<b>ゴ+通+睡</b>	<b>0.60</b>	<b>0.64</b>	<b>0.59</b>	<b>0.70</b>
ゴ+走+通+睡	0.33	0.88	0.27	0.70
すべて	0.27	0.91	0.19	0.70

表 1 コックス比例ハザードモデルを用いた交差検定結果 (AUC 0.70 以上)

- ・ ゴロゴロ時間+走行回数+通勤・通学時間 (0.71)
- ・ ゴロゴロ時間+走行回数+睡眠時間+過去1週間のエネルギー消費量平均(0.71)
- ・ ゴロゴロ時間+通勤・通学時間 (0.70)
- ・ ゴロゴロ時間+通勤・通学時間+睡眠時間 (0.70)
- ・ ゴロゴロ時間+走行回数+通勤・通学時間+睡眠時間 (0.70)
- ・ すべての特徴量 (0.70)

このうち、予測モデルに基づいて被験者にアラートを発することを想定し、感度(再発する真陽性率)が0.60以上、特異度が0.50以上という条件を満たす組み合わせは、

- ・ ゴロゴロ時間+通勤・通学時間
- ・ ゴロゴロ時間+通勤・通学時間+睡眠時間

であった。

最後に、この2つの特徴量の組み合わせに対して、生存関数を用いて PHQ 測定前の AUC を評価した。その結果、いずれのモデルも2週間前に AUC がランダム化した場合の上限値 0.60 を超えることがわかり、2週間前の再発予測の可能性が示唆された(図3)。

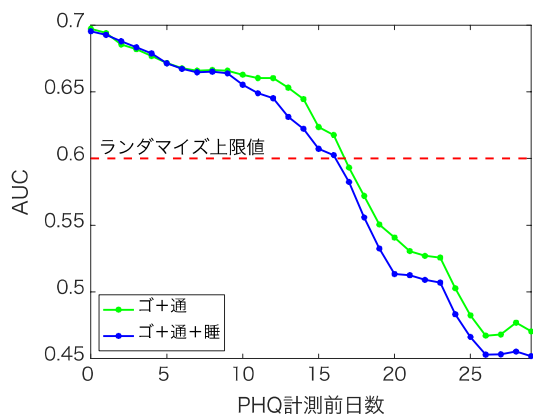


図3 PHQ 計測前の AUC 値の推移

#### D. 考察

活動記録ライフログおよびウェアラブルデバイスによって採取された諸変数の中から、ゴロゴロ時間、通勤・通学時間、および睡眠時間という、活動量に関する変数が、うつ病の増悪を予測する因子として抽出されてきた。これは臨床経験に合致する所見である。

さらに、これらを組み合わせることによって、交差妥当性が AUC で 0.7 を越えるモデルが得られた。かつ、再発の 2 週間前から、偶然を越える予想が可能であることが示された。

実際の運用においては、特異度を高めるカットオフを設定することによって、陽性適中率を高めることが出来る。これにより、最大 2 週間前に「悪化の兆しがあ

ります。きちんと服薬しましょう」あるいは「認知行動療法を復習しましょう」というようなワーニングを出すことが可能である。ワーニングであって、診断ではないので、陽性適中率が 80% や 90% というような数字になる必要はない。ワーニング疲れを引き起こさない程度でワーニングを出せば、予防効果の向上に繋がることが予想される。

次の段階の研究としては、同じライフログデータから、本解析で用いた以外の特徴量を抽出することが出来るので、追加の特徴量をモデルに加えることで、AUC をさらに高めることが出来るかを検討したい。そして、そのモデルを利用して、実際にワーニングを出すことが再発減少に繋がるかどうかを検証するためには、RCT が必要であると考えている。

#### E. 結論

スマートフォンを利用して半自動的に採取される活動記録およびウェアラブルデバイスによる記録から、最大 2 週間前から再発再燃の予測が出来る可能性が示唆された。

#### F. 健康危険情報

なし

#### G. 研究発表

なし

#### H. 知的財産権の出願・登録状況

該当なし

抗うつ薬治験データの人工知能解析に関する研究

研究分担者 丸尾和司 筑波大学医学医療系准教授  
研究分担者 古川壽亮 京都大学大学院医学研究科教授

研究要旨

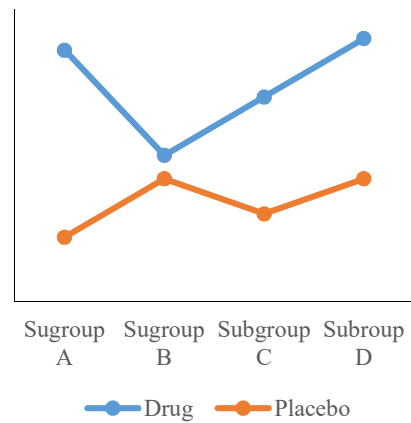
日本で行われた7本、2399人分の、抗うつ剤のプラセボ対照治験データをプールして、QUINTという新しい機械学習手法を用いて治療効果の異なるサブグループを探索した。結果、うつ病初発から1年以上を経過しているかどうかと、性別とにより、質的な差異を示すサブグループが同定された。発症からの経過年数が1年以上の群では抗うつ剤の効果が有意にプラセボよりも高かった。経過年数が1年未満の場合において、男性ではプラセボに有意に劣っており、女性では治療群間差が小さかった。

A. 研究目的

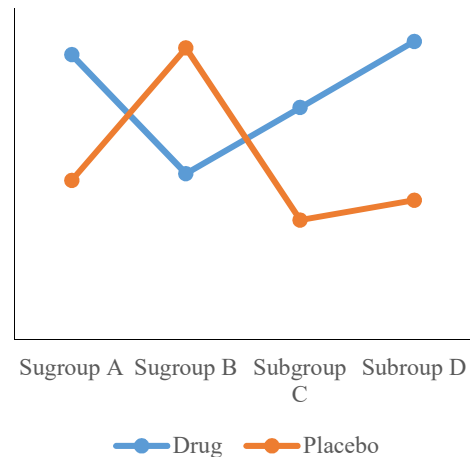
患者特性に合わせて治療を選択する個別化医療が喧伝されるが、精神医療においてはこれまで実現されていない。

本研究では、日本で行われたプラセボ対照の抗うつ剤の治験の個人レベルのデータを用いて、抗うつ剤の選択の個別化（どういう患者ではとくに抗うつ剤が効きやすく、どういう患者ではプラセボとの差が付かないあるいはむしろプラセボの方が良いか）を明らかにする。

このように治療効果の大きさ（つまり、実薬とプラセボの間の差）に影響する因子を効果修飾因子と呼ぶ。効果修飾は、モデル的には因子と治療効果の間の交互作用によって検出されるが、交互作用には2種類ある。下図のように、サブグループごとに治療効果（つまり実薬とプラセボの間の差）は異なるが、いずれのサブグループにおいても実薬がプラセボにまさっている場合と、



下図のようにサブグループによって優劣が逆転することがある場合である。前者を量的な交互作用、後者を質的な交互作用という。



個別化医療でとくに重要なのは、後者の、質的な交互作用である。本研究では、質的な交互作用に着目して、抗うつ薬の個別化医療を可能にする因子を探索する。

## B. 研究方法

本研究の対象とした 7 研究のデータ数の概要は以下の通りである。

試験	1 週	2 週	3 週	4 週	5 週	6 週	8 週
試験 1 <sup>1</sup>	438	423	413	405	0	391	0
試験 2 <sup>2</sup>	296	285	0	277	0	265	259
試験 3 <sup>3</sup>	483	468	455	451	0	437	423
試験 4 <sup>4</sup>	262	255	243	228	218	216	0
試験 5 <sup>5</sup>	399	395	393	379	0	367	360
試験 6 <sup>6</sup>	367	345	0	327	0	318	312
試験 7 <sup>7</sup>	521	516	502	499	0	489	530
計	2399	1874	1551	1788	218	1728	1572

試験 5~7 はポータルサイト上でのみ解析可能であるため、全試験の個人データの統合は困難であった。従って、試験 1~4 のデータを主要な試験とし、さらに試験 1~4 で被験者番号が前半の症例のデータを主要なデータと位置付け、主解析を実施した。試験 1~4 の被験者番号が後半の被験者のデータ及び試験 5~7 のデータは主解析の統計モデルの外部妥当性を検討するために用いた。

データ数も多く、治療効果も十分に検出される 6 週目を主要な時点として設定し、6 週目の HAMD のベースラインからの変化量を主要なアウトカムとして設定した。

主解析では、Qualitative Interaction Trees (QUINT) 法<sup>8</sup>を適用した。この方法は、2つの治療の群間差が異なる 2つの集団に逐次分割し、治療との交互作用に関して特徴的な部分集団を抽出できる決定木ベースのクラスタリング法である。

部分集団構成のための因子となり得る背景因子としては、少なくとも 9 割以上の症例で取られている以下の変数を利用した。

- i) 年齢
- ii) 性別

- iii) 体重
- iv) 大うつ病診断分 (単一性 or 反復性)
- v) 身体合併症の有無
- vi) 最初の大うつ病発症時年齢
- vii) 現在の大うつ病エピソードの持続期間
- viii) 最初の大うつ病発症からの経過年数
- ix) HAMD ベースライン値
- x) HAMD サブスケール : anhedonia retardation (HAMD1+2+3+16)
- xi) HAMD サブスケール : body symptoms (HAMD9+13+14)
- xii) HAMD サブスケール : sleep problems (HAMD6+7+8)
- xiii) HAMD サブスケール : appetite (HAMD4+5)
- xiv) HAMD サブスケール : guilt & agitation (HAMD10+11+12+15+17)

欠測値の取扱いについては、欠測のある被験者を除外した完全症例解析を行った。QUINT 法で抽出された部分集団(樹木の葉:リーフ)ごとに Cohen's d とその SE を推定した。

主解析の感度分析として、以下の解析を実施した。

- (i) 内部整合性と外部妥当性の検討のために、試験 1~4 のそれぞれで、主解析で抽出された各リーフ(以降、各リーフ)での Cohen's d とその SE を推定した。
- (ii) 外部妥当性の検討のために、試験 1~4 の被験者番号が後半の症例データについて、各リーフでの Cohen's d とその SE を推定した。
- (iii) 外部妥当性の検討のために、試験 5~7 の各試験で、各リーフでの Cohen's d とその SE を推定した。
- (iv) 上記の全データセット(主解析のデータ、感度分析(i)~(iii)のデータ)で線形仮説  $L^T \beta = 0$  に関する推測を行った。ここで、 $L$  は対比ベクトルであり、 $\beta$  はリーフ、群とこれらの交互作用を因子とした線形モデルの回帰係数ベクトルである。治療群間差に関する各リーフの任意の順位に対応し



た複数の対比ベクトルを設定し、選択された対比の主解析との類似性を評価した。また、対比の検定から、リーフ間の治療効果の差が有意に異なるかを検討した。

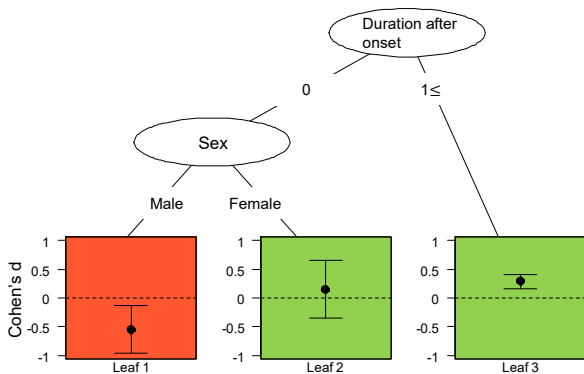
- (v) QUINT 法は単一の RCT のデータを対象としており、試験間差をランダム効果として組み入れられない。この点での QUINT 法の妥当性確認のために、リーフ、群、リーフと群の交互作用を固定効果、試験 ID を変量効果とした線形混合効果モデルを適用し、(iv)と同様の線形仮説の検定を実施した。そのうえで、主解析との類似性を検討した。

(倫理面への配慮)

本研究は、人を対象とする医学系研究の倫理指針の第 1 章第 3 の 1「適用される研究」においていうところの、「既に連結不可能匿名化されている情報」を用いた研究であるので、倫理委員会の審査を要しない。

C. 研究成果

QUINT 法による最終的な樹状図および各部分集団での Cohen's d を以下に示す



リーフ	1	2	3
プラセボ N	35	21	153
Mean	-12.54	-10.19	-8.77
SD	6.05	8.03	6.51
抗うつ剤 N	67	67	361

Mean	-9.25	-11.13	-10.67	
SD	6.04	5.56	6.72	
Cohen's d	d	-0.54	0.15	0.29
SE		0.21	0.25	0.06

発症からの経過年数(duration after onset)と性別がモデルに含まれた。発症からの経過年数が 1 年以上のリーフ 3 では抗うつ剤の効果が有意にプラセボよりも高かった。経過年数が 1 年未満の場合において、男性(リーフ 1)ではプラセボに有意に劣っており、女性(リーフ 2)では治療群間差が小さかった。

以下の表は、感度分析(i)~(iii)の結果である。表内の数値は Cohen's d でありカッコ内はその SE である。

感度分析   Leaf	1	2	3
(i) 試験 1	-0.10(0.25)	-0.42(0.34)	0.54(0.13)
(i) 試験 2	-0.22(0.28)	-0.07(0.33)	0.13(0.16)
(i) 試験 3	-0.33(0.30)	0.88(0.33)	0.09(0.13)
(i) 試験 4	-0.79(0.68)	0.45(0.62)	0.50(0.17)
(ii)	-0.01(0.22)	0.03(0.26)	0.17(0.11)
(iii) 試験 5	0.52(0.25)	-0.03(0.23)	0.25(0.14)
(iii) 試験 6	-0.44(0.31)	-0.51(0.31)	0.18(0.13)
(iii) 試験 7	0.15(0.22)	0.02(0.22)	0.13(0.11)

(i), (ii)については、主解析とおおむね同様の結果が得られた。(iii)については、全ての試験においてリーフ 3 で抗うつ剤がプラセボと比較して効果がある傾向がみられたものの、その他のリーフでの効果は一定した傾向がみられなかった。

以下に、感度分析(iv), (v)の結果を示す。表中の数値は対比検定の p 値であり、p 値が最小の対比が選択された対比である。二重下線が最小の p 値、下線が 2 番目に最小の p 値である。主解析ではリーフ 3, 2, 1 の順に治療効果が高く、そのリーフ間差は有意水準 0.05 で有意ということが示された。

Order	主	(i) 試験 1	(i) 試験 2	(i) 試験 3	(i) 試験 4

3 2 1	<u>0.001</u>	<u>0.027</u>	<u>0.250</u>	0.250	<u>0.129</u>
2 3 1	0.581	<u>0.004</u>	<u>0.636</u>	0.076	0.969
2 1 3	0.581	0.004	0.636	<u>0.076</u>	0.969
3 1 2	<u>0.049</u>	0.330	0.655	<u>0.025</u>	<u>0.266</u>
1 2 3	0.001	0.027	0.250	0.250	0.129
1 3 2	0.049	0.330	0.655	0.025	0.266
Order	(ii)	(iii) 試験 5	(iii) 試験 6	(iii) 試験 7	(v)
3 2 1	<u>0.451</u>	0.322	<u>0.111</u>	0.920	<u>0.002</u>
2 3 1	<u>0.577</u>	<u>0.298</u>	<u>0.064</u>	<u>0.650</u>	0.465
2 1 3	0.577	0.298	0.064	0.650	0.465
3 1 2	0.931	0.098	0.846	0.657	<u>0.109</u>
1 2 3	0.451	0.322	0.111	0.920	0.002
1 3 2	0.931	<u>0.098</u>	0.846	<u>0.657</u>	0.109

感度分析(i), (iii)で選択された対比は概ね主解析と同様であったが、(iv)については、一定の傾向がみられなかった。感度分析(v)では、主解析とほとんど同等の結果が得られ、変量効果を考慮しなかったことの影響は小さかったことが示唆された。

#### D. 考察

日本で行われた7本、2399人分の、抗うつ剤のプラセボ対照試験データをプールして、QUINTという新しい機械学習手法を用いて治療効果の異なるサブグループを探索した。

結果、うつ病初発から1年以上を経過しているかどうかと、性別とにより、質的な差異を示すサブグループが同定された。

このサブグループングは、内的妥当性および同じ試験の後半データセットにおける時間妥当性は概ね確認されたが、まったく新しい試験における外的妥当性は一定しなかった。

#### E. 結論

QUINTは、個別化医療においてとくに重要な質的な交互作用を検出するために有用な手法である。しかし、安定した外的妥当性を得るためにはさらなるデータと解析が必要である

#### F. 健康危険情報

なし

#### G. 研究発表

該当なし

#### H. 知的財産権の出願・登録状況

該当なし

#### 参考文献

- Higuchi T, Murasaki M, Kamijima K. Clinical evaluation of duloxetine in the treatment of major depressive disorder: placebo- and paroxetine-controlled double-blind comparative study. Japanese Journal of Clinical Psychopharmacology [in Japanese] 2009; 12: 1613-34.
- Hirayasu Y. A dose-response study of escitalopram in patients with major depressive disorder: a placebo-controlled, double-blind study. Rinsho Seishin Yakuri (Japanese Journal of Clinical Psychopharmacology) 2011; 14(5): 871-82.
- Hirayasu Y. A dose-response and non-inferiority study evaluating the efficacy and safety of escitalopram in patients with major depressive disorder: a placebo- and paroxetine-controlled, double-blind, comparative study. Rinsho Seishin Yakuri (Japanese Journal of Clinical Psychopharmacology) 2011; 14(5): 883-99.
- Kinoshita T. A double-blind, placebo-controlled study of a new antidepressant, mirtazapine, in depressed patients. Japanese Journal of Clinical Psychopharmacology [in Japanese] 2009; 12: 289-306.
- Higuchi T, Hong JP, Jung HY, Watanabe Y, Kunitomi T, Kamijima K. Paroxetine controlled-release formulation in the treatment of major depressive disorder: a randomized, double-blind, placebo-controlled study in Japan and Korea. Psychiatry Clin Neurosci 2011;65:655-663.
- Koshino Y, Bahk WM, Sakai H, Kobayashi T. The efficacy and safety of bupropion sustained-release formulation for the treatment of major depressive disorder: a multi-center, randomized, double-blind, placebo-controlled study in Asian patients. Neuropsychiatr Dis Treat 2013;9:1273-1280.
- Higuchi, T., Kamijima, K., Nakagome, K., Itamura, R., Asami, Y., Kuribayashi, K., Imaeda, T., 2016. A randomized, double-blinded, placebo-controlled study to evaluate the efficacy and safety of venlafaxine extended release and a long-term extension study for patients with major depressive disorder in Japan. Int. Clin. Psychopharmacol. 31, 8-19.
- Dusseldorp E, Van Mechelen I. Qualitative interaction trees: a tool to identify qualitative treatment-subgroup interactions. Stat Med 2014; 33(2): 219-37.

研究成果の刊行に関する一覧表

書籍

著者氏名	論文タイトル名	書籍全体の 編集者名	書 籍 名	出版社名	出版地	出版年	ページ
	(なし)						

雑誌

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
	(なし)				