

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、抗うつ薬の売上げは年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^{3,4)}。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)。

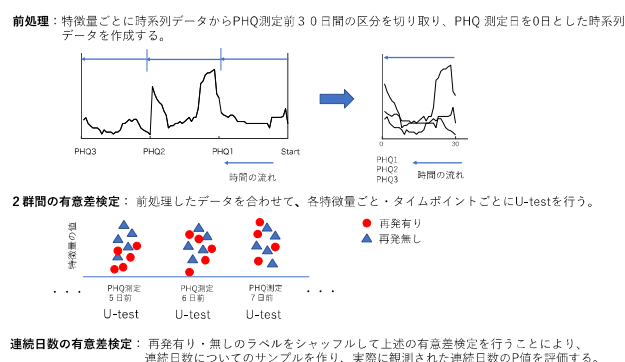


図1: 時系列データ分析の概略

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。

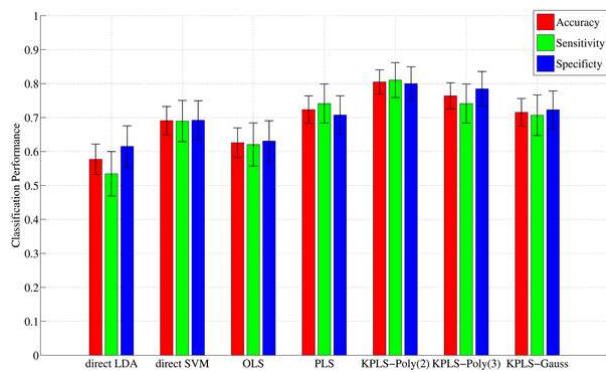


図2: うつ病患者と健常者の判別成績

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、BAL⁶⁾ 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)。

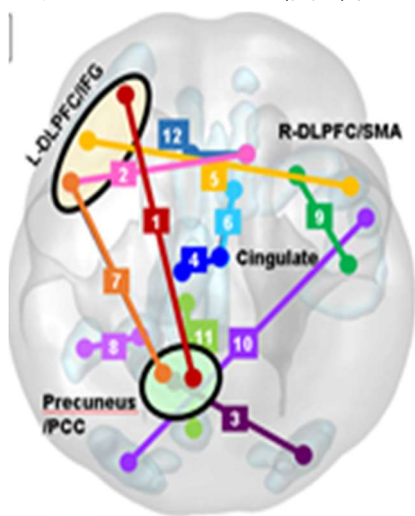


図 3: 完全な独立コホートに汎化するメランコリア特徴を有するうつ病の安静時機能結合の判別器

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)。

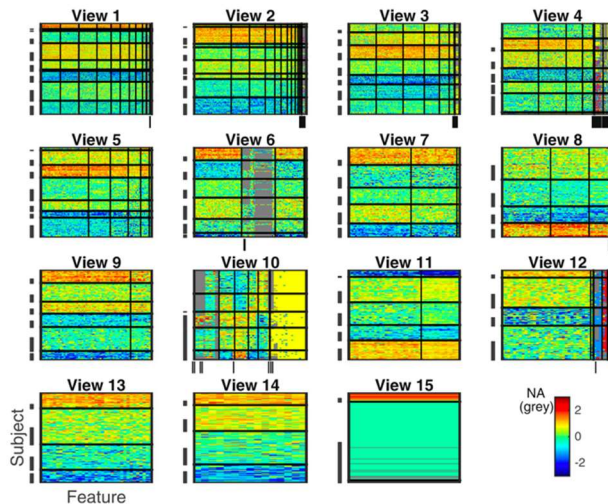


図 4: 多重ベイズ共クラスタリングの適用結果 (縦軸は参加者、横軸は特徴量を表しており、参加者軸に付随したダッシュ記号はうつ病患者を表している。ヒートマップ中の太線はクラスタの境界を示している。)

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。

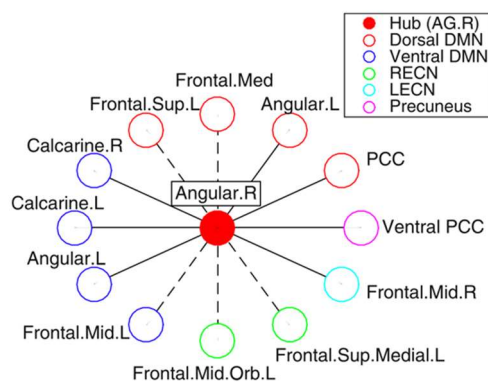


図 5: 図 3 の View 10 のクラスタ構造に含まれる安静時機能結合ネットワーク

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）。

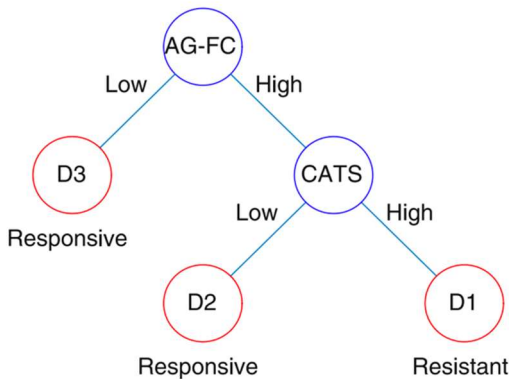


図 6: 右角回を中心とした機能ネットワークの安静時機能結合性と幼児期トラウマ指標に基づく治療反応予測モデル（AG-FC は右角回を中心とした12個の安静時機能結合性のスコアであり、CATSは幼児期トラウマ指標である。）

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）。

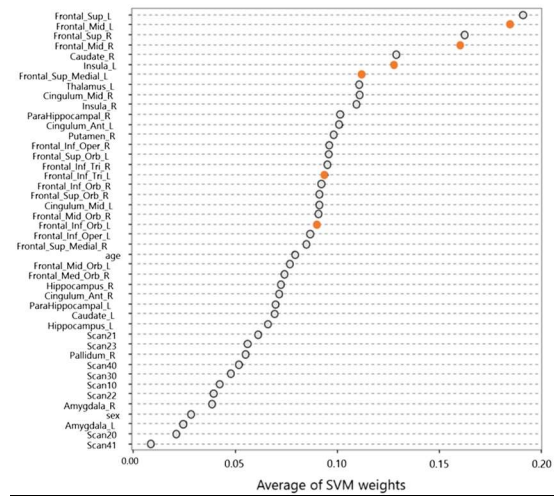


図 7: SVM を用いたうつ病診断への寄与度

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

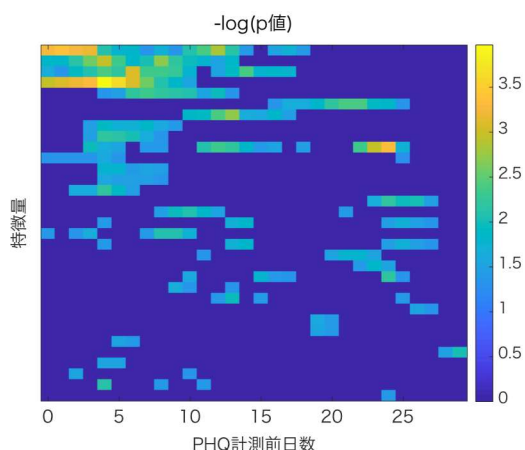


図 8: 2 群 (再発有り、無し) の違いに関する U-test の P 値 (濃い青は有意差なし)。特徴量は連続した有意日数が多いに並び替えた。上位 5 つの特徴量は順に、ゴロゴロ時間、走行回数、過去 1 週間のエネルギー消費量平均、睡眠時間である。

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
ゴ+通	0.60	0.64	0.59	0.70
ゴ+通+睡	0.60	0.64	0.59	0.70
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⁷
(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)。

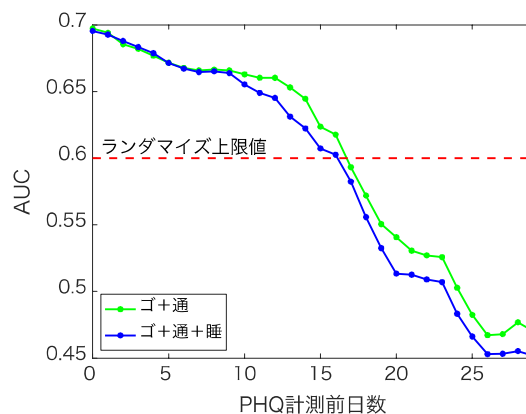
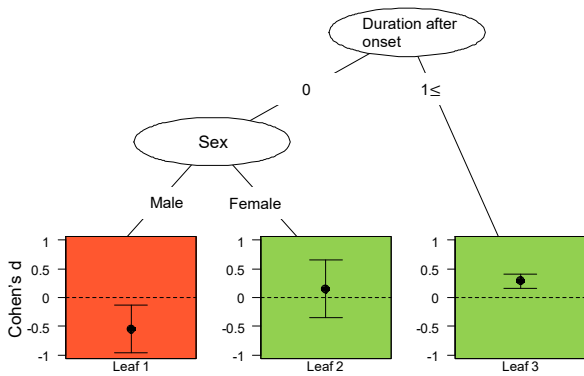


図 9: PHQ 計測前の AUC 値の推移

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.