

令和 4 年 5 月 4 日現在

機関番号：12501

研究種目：若手研究

研究期間：2018～2021

課題番号：18K15476

研究課題名（和文）精神医学・疫学におけるビッグデータの利活用法の探索

研究課題名（英文）Exploring the usability of bigdata in psychiatric epidemiology

研究代表者

須藤 千尋 (Sutoh, Chihiro)

千葉大学・大学院医学研究院・講師

研究者番号：30612650

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：本研究では、「精神疾患の病因研究におけるビッグデータ解析の利用可能性」を探索するために、3つの研究課題を行った。

うち2課題では、データ量不足のために適切な結果を得ることができなかった。

「機械学習を用いた画像認識技術による性周期判定ソフトの開発」では、必要なデータ（約3300枚の雌マウスの膣スメア画像）の獲得と、その画像から細胞診を行うコンピュータプログラムを開発し、英文学術論文に掲載された。また、GitHubにて開発したプログラムを、千葉大学リポジトリCURATORにて学習に用いた画像データセットを公開した。

研究成果の学術的意義や社会的意義

精神疾患の多くで有病率などに性差があることが知られている。しかし、従来の動物実験の知見の多くはオス個体を中心に得られており、その結果を男女を含めたヒト全体へ適用することに限界が指摘されてきた。動物実験にメス個体が使われにくい理由の一つは、性周期判定のハードルの高さである。本課題は、実験動物メスの性周期判定を容易にすることで、今後メスを対象に含めた研究数を世界規模で増やすのに貢献することを企図し、細胞診を行うコンピュータプログラムを開発した。テスト用画像による判定精度比較で、プログラムは熟練研究者に並ぶ高精度での判定ができることがわかった。

研究成果の概要（英文）：To explore the possibility of big data analysis in the cause study of psychiatric disorders, we conducted three research subjects. In two of these tasks, no good result was obtained due to the lack of enough data. we developed a computer program to perform cytodiagnosis of sexual cycle from the female mouse vaginal smear images. The result was published in an academic paper, and the acquired program and data set were released on GitHub and at the Chiba University repository CURATOR.

研究分野：画像解析

キーワード：機械学習

1. 研究開始当初の背景

本研究では、「精神疾患の病因研究におけるビッグデータ解析の利用可能性」を探索するために、後述する3つの研究課題を行った。

疾患研究において病因の特定はすなわち予防や治療へとつながる重要な発見となる。しかし、精神疾患の多くは、様々な遺伝的素因(ゲノム、エピゲノム)と環境因(出生前後の栄養・薬物・感染症・放射線、社会的ストレス、疲労、親の年齢、生活習慣など)の寄与が示唆され、単独の原因を特定しがたい状況で、疾患概念や分類も揺らぎ続けている。このような分野では、既知の因子だけでなく、多数の因子群が複合的に寄与する疾患モデルや未知の因子の存在を考慮することで大きな進展が見こまれる。その解明にはビッグデータの利用が必須である。

学术界に無数に生まれてくるアイデアの全てを仮説検証型研究にかけることは、理想ではあるが、得られる結果に対して人的・経済的・時間のコストが大きく、現実的ではない。そのため研究者は経験と勘と先行研究を頼りに仮説を絞り込み、その仮説の正否を調べる。これに対し、データサイエンス技術では、非常に多数の因子(変数)や例数を含むために、経験や勘に基づく仮説を立てることができないような規模のデータを取り扱うことができる。そして、その技術はビッグデータ解析による仮説検証だけでなく、データマイニングという別名が示すように、データの山を網羅的に採掘して、未知の仮説を発見する役目も果たす。例えば、従来の仮説検証型研究で降圧薬Aの効果の評価したいとき、投与群と非投与群で数百例ずつの血圧を、既知の数十個の交絡因子の影響を除いて比較すれば、「薬物Aは血圧を下げる」と言えるだろう。だが、もしここに数万人規模のビッグデータがあれば、既知の交絡因子の影響を除去した2群比較の解析でより確からしい結果を得る(仮説検証)ばかりでなく、血圧と一見無関係に見える無数の因子(例:前日見たTV番組、服の好み、遠隔地の天候、最寄り駅の乗降客数など)の中からも、薬物Aの効果に影響する新規の交絡因子(仮説)を発見しうるのである。

2. 研究の目的

研究課題1 国内最大規模のSNPデータベースを用いた精神疾患への寄与パターンの探索

本課題の目的は、様々な精神疾患に対して一塩基多型(SNP)*の組み合わせがどのような寄与をするのか探索し、モデル化することであった。

*一塩基多型(Single Nucleotide Polymorphism, SNP)..... DNAの塩基配列に1~少数塩基の変異として存在する多型で、集団内の頻度が1%(一人が2コピーをもつので50人に一人)を超えるもの。

精神疾患の発症因子は、複数の遺伝的変異の組み合わせと環境因の双方が関与する(疾患ごとに遺伝的素因と環境因の寄与度は異なる)と考えられている。遺伝的素因研究においては、これまでに特定遺伝子の変異が指摘されてきたが十分に再現性を持って示されている決定的な変異は非常に少ない。こうした経緯から、特定の遺伝子が決定的に疾患発症に関与するというよりも、それぞれの寄与度は低くても、遺伝子変異(例えばSNP)の総数やその組み合わせが重要なタンパク質機能や分子ネットワークに異常をもたらし、総体として発症脆弱性や治療反応性などに影響を与えている可能性が推測されてきた。

本課題では、精神疾患への寄与を探索するSNPデータベースとして、東北メディカル・メガバンク計画による統合データベース『dbTMM』の利用を検討した。dbTMMは、東日本大震災からの復興事業として開始され、東北大学・岩手医科大学により推進中の複数の長期健康コホート調査データで構築されている。2015年より研究目的でのデータ分譲を開始し、現在、東北地域の健康調査事業により集積された2.3万人分の生活習慣、診療情報、ゲノム・オミックス情報、血液・尿検査情報を含む統合データベースが分譲可能である。この事業の中で開発された、66万個のSNP情報と遺伝子型インピュテーション技術により最大2,000万SNPを取得可能なSNPアレイ『ジャポニカアレイ』を採用することで、日本人に最適化されたSNPデータベースとなっている。

研究課題2 安静時機能的脳MRI画像を用いた発達障害型の行動的傾向の神経学的基盤の探索

本課題の目的は、発達障害の行動的傾向を裏付ける脳機能的所見の探索であった。障害の種類や有無で群間比較するのではなく、各被験者の行動的特徴を表す多数の尺度を変数に入力した回帰モデルを用いる。発達障害が、明確な線引きによる診断のイメージとは異なり、病的な行動的傾向が人口集団の中にスペクトラムで存在しており、一人の「発達障害者」も複数の行動的傾向をまちまちの程度で内在しているという実態を反映でき、個別の行動的傾向の背景にある神経基盤を抽出することができる。これはとくに高い演算処理能力を要する方法である。

自閉スペクトラム症(ASD)は社会的コミュニケーションの障害と限定された興味を特徴とし、生涯有病率は1%前後である。注意欠陥多動性障害(ADHD)は不注意、多動性、衝動性が特徴

で、有病率は小児の5%、成人の1~3%とされる。どちらも近年増加が指摘されている。ところが、発達障害者と定型発達者との境目は、障害あり・なしのように明確に区分されるものではない。これらの病的な行動的傾向を弱く持つ者から強く持つ者まで、集団の中にスペクトラム(分布、連続体)として存在し、言うなれば「たまたま診断される機会があったかどうか」によるのである。ゆえに、発達障害の近年の増加も、病的な行動的傾向を診断の境界域程度の強さで有している者が、発達障害の知名度が高まりによって新たに診断されるケースが増えたためというのが有力な説である。また、一人の「発達障害者」を見ると、複数の発達障害の診断名に属するような行動的傾向を、副診断になるレベルからそうでないものまでまちまちの程度で内在している。発達障害の原因研究にこうした実態を反映するには、診断の有無で少数の対象者を群間比較するよりも、大人数の行動的傾向の指標との回帰モデルで探索する必要がある。

研究課題3 機械学習を用いた画像認識技術による性周期判定ソフトの開発

本課題の目的は、マウス膣スミア画像による性周期自動判定ソフトの開発であった。精神疾患研究の基盤研究においてメス個体の利用拡大は世界的な潮流である。うつ病、心的外傷後ストレス障害(PTSD)、不安症では約2倍、摂食障害では約10倍、それぞれ女性の有病率が男性に比べ高いとされている。しかし、従来の動物実験の知見の多くはオス個体を中心に得られており、それをヒト全体へ適用することに限界が指摘されてきた。メス個体が使われにくい理由の一つは、性周期判定のハードルの高さである。近年、米国NIHが、生物学研究において雌雄両方を用いるようにと勧告(2015, Notice #NOT-OD-15-102)を発表したため、性周期判定の重要度は世界的に高まっている。本課題は、今後メスを対象に含めた研究数を世界規模で増やすことに貢献する基盤プロジェクトである。精神科医療では、女性患者に精神症状の変動と妊娠状態や性周期の関係を尋ねることがある。動物の成熟メス個体は性周期にともなって、例えばホルモンバランスや精神神経活動など、多様なデータに変動が現れる。ヒトにおいて性周期の精神面への影響は大きく、正常月経周期をもつ女性の5~10%が、月経前に緊張・不安・抑うつ・睡眠障害などの症状により、月経前不快気分障害(PMDD)または月経前症候群(PMS)の診断基準を満たすとされる。これと同様に、実験動物のメス個体が性周期のどの段階にあるのか把握することは、メス個体のデータを解釈するために必須の技術である。げっ歯類の性周期判定は、膣内の細胞サンプル(膣スミア)を光学顕微鏡で観察することで行う。しかしながら、性周期判定は訓練を受けていない者には困難であり、熟練者にとっては手間のかかる作業である。

3. 研究の方法

研究課題1 国内最大規模のSNPデータベースを用いた精神疾患への寄与パターンの探索

本課題では、dbTMMから抽出した精神疾患の既往症情報(右表)に対し、全ゲノム内に存在するSNPの組合せがどのような寄与をするのか探索し、モデル化を行うことを目指した。1人につき2000万箇所ある各SNP情報だけでなく、全SNPの変異出現頻度の総計的指標や、精神疾患への寄与が指摘された遺伝子に限定したSNP情報などの変数を設定し、疾患の有無に寄与するSNPパターンを網羅的に探索するものであった。

研究課題2 安静時機能的脳MRI画像を用いた発達障害型の行動的傾向の神経学的基盤の探索

申請者らは7大学(大阪大学、金沢大学、浜松医科大学、福井大学、千葉大学、弘前大学、鳥取大学)による連合大学院・連合小児発達学研究所の枠組みの中で、安静時機能的脳MRI画像データおよび構造的脳MRI画像データを集積している。発達障害研究に主眼が置かれたデータライブラリであり、主な対象者はASD、ADHD、および定型発達者である。本課題では、MRIデータから算出される多数の脳領域間の相関的機能連携指標を組み合わせ、対象者の行動的傾向をMRIデータから説明するような回帰式を導き出すことを目指した。

研究課題3 機械学習を用いた画像認識技術による性周期判定ソフトの開発

申請者らのグループでは、情動と行動に焦点をあて、性差を加味した動物研究を続けてきた。既に約3000枚のマウス膣スミア画像と性周期判定情報を保有し、さらに追加収集していく予定である。本課題では、それらを訓練データとした物体認識問題をコンピュータに「教師(正解)あり学習」させ、アルゴリズムを最適に調整することで予測式(自動判定ソフト)を開発する。すなわち、新規の膣スミア画像を入力して、経験豊かな研究者と同じ判定結果を得られるものを目指す。このソフトにより、熟練者・非熟練者を問わず、信頼性の高いワールドスタンダードの判定を手軽に得られるようになるものである。

4. 研究成果

研究課題1 国内最大規模のSNPデータベースを用いた精神疾患への寄与パターンの探索

主に学習させるデータ量が不足しているために発生する機械学習の過学習の問題によって適切な結果を得ることができなかった。

研究課題2 安静時機能的脳 MRI 画像を用いた発達障害型の行動的傾向の神経学的基盤の探索

新型コロナウイルス感染症流行に伴う実験環境の障害によって、研究に妥当な規模のデータベース取得を達成できなかった。

研究課題3 機械学習を用いた画像認識技術による性周期判定ソフトの開発

必要なデータ（約 3300 枚の雌マウスの膣スミア画像）の獲得と、その画像から細胞診を行うコンピュータプログラムの開発という進捗を得られた。各画像が示す性周期の phase 情報をあらかじめ熟練研究者が判定し、画像データとは別の「正答」データとした。そのうち約 2000 枚を訓練画像として Python ベースの畳み込みニューラルネットワークプログラムに投入、より高精度で正答にたどり着く"読影法"を学習させた。訓練に用いていない残りの画像から、約 500 枚の評価用画像セット、約 700 枚のテスト用画像セットを作成した。評価用画像およびテスト用画像の示す性周期 phase をプログラムに判定させたところ、各 phase で 90% 前後の判定制度を得た。テスト用画像による判定精度比較で、プログラムは熟練研究者に並ぶ高精度での判定ができることがわかった。成果は英文学術論文に掲載された。また、開発したプログラムをソフトウェア開発プラットフォームである GitHub にて、学習に用いた画像データセットを千葉大学学術成果リポジトリ CURATOR にて、それぞれ公開した。

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件 / うち国際共著 0件 / うちオープンアクセス 1件）

| | |
|---|---------------------|
| 1. 著者名 Sano K, Matsuda S, Tohyama S, Komura D, Shimizu E, Sutoh C | 4. 巻 10 |
| 2. 論文標題 Deep learning-based classification of the mouse estrous cycle stages | 5. 発行年 2020年 |
| 3. 雑誌名 Scientific Reports | 6. 最初と最後の頁 11714 |
| 掲載論文のDOI（デジタルオブジェクト識別子） 10.1038/s41598-020-68611-0 | 査読の有無 有 |
| オープンアクセス オープンアクセスとしている（また、その予定である） | 国際共著 - |

〔学会発表〕 計1件（うち招待講演 0件 / うち国際学会 0件）

| |
|---|
| 1. 発表者名 佐野恭平, 松田真悟, 遠山 卓, 河村大輔, 清水栄司, 須藤千尋 |
| 2. 発表標題 機械学習によるマウス性周期判定技術の開発 |
| 3. 学会等名 第249回 生理学東京談話会 |
| 4. 発表年 2019年 |

〔図書〕 計0件

〔産業財産権〕

〔その他〕

| |
|--|
| Secreit https://github.com/SanoKyohei/Secreit Cytology images for machine learning https://opac.ll.chiba-u.jp/da/curator/108041/ Deep learning-based classification of the mouse... https://opac.ll.chiba-u.jp/da/curator/108101/ |
|--|

6. 研究組織

| 氏名 （ローマ字氏名） （研究者番号） | 所属研究機関・部局・職 （機関番号） | 備考 |
|---------------------------|-----------------------|----|
|---------------------------|-----------------------|----|

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

| 共同研究相手国 | 相手方研究機関 |
|---------|---------|
|---------|---------|