

スマートフォンによるデジタルバイオマーカーを用いた 摂食障害予測モデルの構築と分類精度の評価

中村 真穂^{1,a)} Panote Siriaraya^{2,b)} 中島 伸介^{1,c)}

概要：摂食障害の従来の診断手法は、自己申告や臨床観察に依存しており、想起バイアスや過少申告といった主観的要因の影響を受けやすい。そのため、日常生活における行動パターンから摂食障害に特徴的な行動シグネチャーを抽出し、早期発見のための予測手法を確立することが重要な課題となっている。本研究では、スマートフォンのセンサデータを用いて、摂食障害の有無を識別する機械学習モデルを構築し評価することを目的とする。具体的には、摂食障害群および健常対照群から収集した加速度、位置情報、アプリ使用履歴などのデータを用いて、XGBoost, SVM, ロジスティック回帰, ランダムフォレストの各手法による分類モデルを構築し、その分類精度を比較する。将来的には、本研究で得られた知見とモデルをもとに、予測機能を搭載したスマートフォンアプリケーションを開発し、日常的な使用を通じて摂食障害の兆候を検出し、通知できるシステムの実現を目指す。

キーワード：デジタルバイオマーカー, 摂食障害, 機械学習, スマートフォン

Construction of a prediction model for eating disorders using digital biomarkers on smartphones and evaluation of classification accuracy

Abstract: Conventional diagnostic methods for eating disorders rely on self-reports and clinical observations, making them susceptible to subjective factors such as recall bias and underreporting. Therefore, it is an important task to establish predictive methods for early detection by extracting behavioral signatures characteristic of eating disorders from behavioral patterns in daily life. This study aims to construct and evaluate a machine learning model that uses smartphone sensor data to identify the presence or absence of eating disorders. Specifically, we will construct classification models using XGBoost, SVM, logistic regression, and random forest methods, based on data such as acceleration, location information, and app usage history collected from eating disorder groups and healthy control groups, and compare their classification accuracy. In the future, based on the findings and models obtained in this study, we aim to develop a smartphone application with predictive functionality that can detect signs of eating disorders through daily use and notify users of them.

Keywords: digital biomarkers, eating disorders, machine learning, smartphones

1. はじめに

摂食障害は中枢性摂食異常症とも呼ばれ厚生労働省の特定疾患（難病）に指定される精神疾患である。主に拒食症と過食症に分けられ、どちらも食行動異常とそれに伴う認知や情動の障害を主徴としている。日本では、年間約2万

¹ 京都産業大学 情報理工学部 〒603-8555 京都府京都市北区上賀茂本山

² 京都工芸繊維大学 情報工学・人間科学系 〒606-8585 京都府京都市左京区松ヶ崎橋上町

a) g2254261@cc.kyoto-su.ac.jp

b) spanote@kit.ac.jp

c) nakajima@cc.kyoto-su.ac.jp

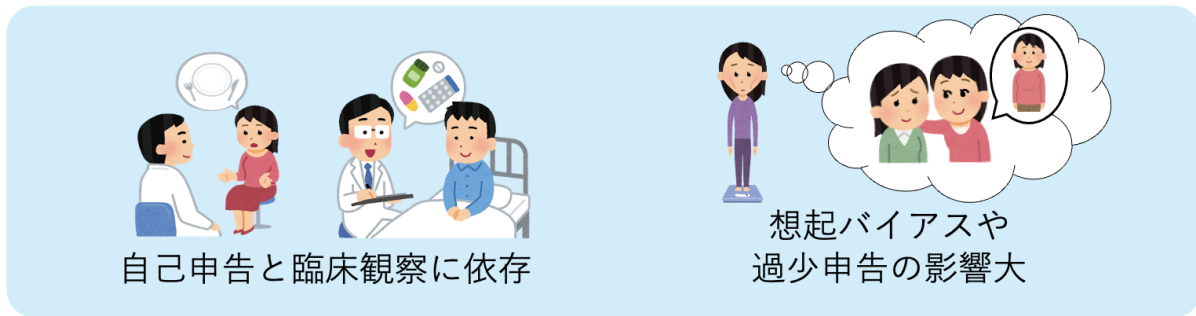


図 1 従来の摂食障害診断の問題点

3千人の患者がおり^{*1}、摂食障害の罹患率は増加傾向にあるため、その診断と治療の重要性が高まっている [1]。

従来の摂食障害の診断では、臨床医による観察や患者の自己申告が主に用いられており、これらの方法は主観に依存するため、記憶の偏りや過少申告といった問題が指摘されている (図 1)。すなわち、従来の方法では摂食障害の早期の発見および治療の実現は容易ではない。このような状況に対し、治療法の一つとして、食事や感情の記録を通じて自己理解や行動変容を支援するモバイルヘルスアプリの開発が進められている [2]。ただし、興味深い取り組みではあるものの、行動に着目した摂食障害の早期の発見および治療に対する効果的なシステムの開発には至っていない。

そこで我々は、スマートフォンを利用したデジタルバイオマーカーに着目した。デジタルバイオマーカーとは、埋め込み型環境センサやウェアラブルなどのデジタルデバイスによって収集および測定される、客観的で定量化可能な生理学および行動的データである [3]。これらは行動データをリアルタイムで収集し、従来の方法では得られない特徴を捉えることができる。たとえば、過食症の患者は、不規則な食事時間、頻繁なファストフード店の訪問、または身体活動レベルの低下といった特異な行動を示す可能性がある。したがって本稿では、スマートフォンのセンサデータを活用し、摂食障害に関連する行動パターンを検出し、機械学習を用いた予測モデルの構築を目的としている。そこで、本稿ではスマートフォンのセンサデータを使用して機械学習を行い、構築した各モデルの分類精度を評価したため報告する。

2. 関連研究

摂食障害の症状の客観的な定量化についての研究は、ソーシャルメディアを活用したもののが主に挙げられる。

Chancellor らは Instagram^{*2} を使用し、ソーシャルメディアにおける精神疾患の重症度 (MIS) を定量化した。Latent Dirichlet Allocation (LDA) を利用して投稿内容をトピック

モデル化し、専門家と非専門家によるトピック評価を行い、トピックごとの MIS をスコア付けた。また、過去 7 か月間の MIS データを基に、翌月の MIS を予測する回帰モデルを構築している [4]。

また、Wang らは Twitter (現 X)^{*3} を使用し、摂食障害に苦しむ個人のコミュニティ構造と相互作用について調査している。サンプリングされた摂食障害ユーザと、非摂食障害ユーザに対して収集された 2 つの参照データセットとの社会的ステータス、行動パターン、心理測定特性における差異を定量化し、摂食障害ユーザと非摂食障害ユーザを分類する予測モデルを構築している [5]。

摂食障害と関連する精神疾患として議論される、うつ病については以下のような研究がある。

阪口らはアンサンブル学習を用いて、スマートフォンの会話、位置情報、ロックのなどのデータから 9 つの特徴量を抽出し、うつ状態を判別するモデルを構築している。結果、会話、位置情報、スマートフォンのロックから抽出した特徴量はうつ状態の判別に有用であるとしている [6]。

田近らは開発したスマートフォンアプリとウェアラブルデバイスによる日々の記録のデータから、うつ病の再発予測モデルを構築している [7]。

Moshe らはスマートフォンとウェアラブルデバイスを用いてうつ病と不安障害の症状を予測可能か評価する研究を行った。GPS とウェアラブルデバイスのデータでメンタルヘルスを予測できること、GPS データのみの場合と比較して、GPS とウェアラブルデータを組み合わせたモデルでは、うつ病や不安の症状を予測する能力が大幅に向上したという結果であった [8]。

上述の通り、摂食障害の診断について客観的指標を用いた定量化は行われつつあるが、まだ一部のソーシャルメディアの活用に限られている。また、うつ病などの他の精神疾患ではデジタルバイオマーカーを用いた予測モデル開発が研究されており、これは摂食障害にも応用できると考える。

^{*1} 厚生労働科学研究成果データベース 摂食障害の診療体制整備に関する研究 (2016) <https://mhlw-grants.niph.go.jp/project/26089>

^{*2} Instagram <https://www.instagram.com/>

^{*3} X <https://x.com/>

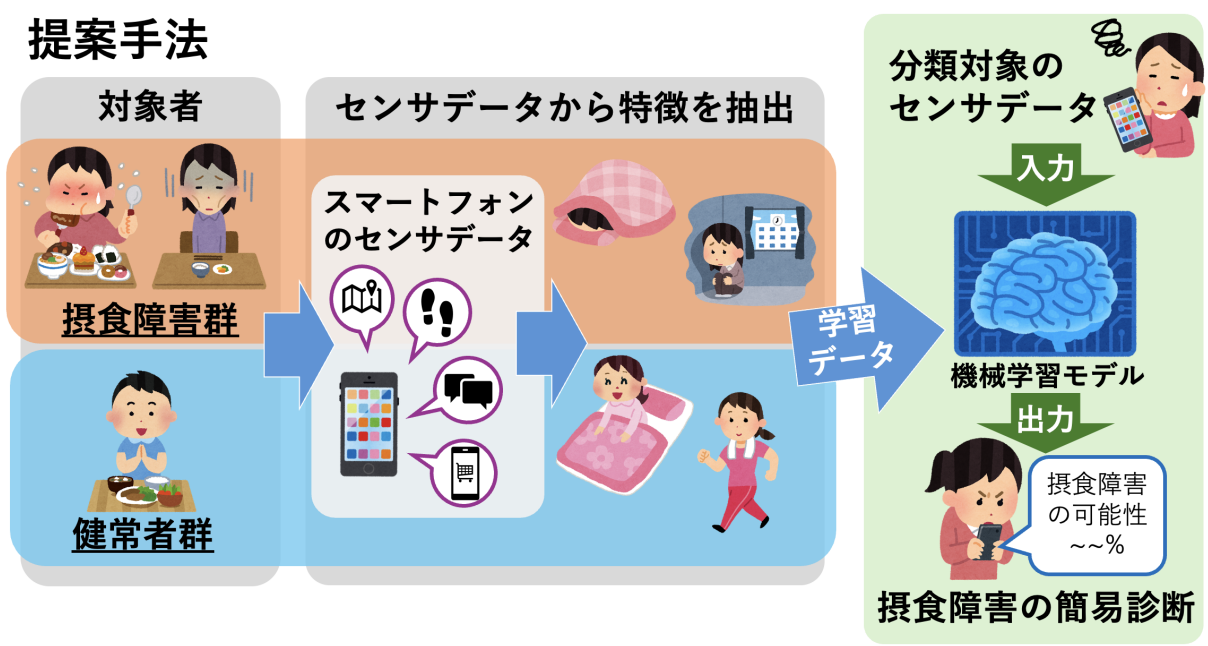


図 2 提案手法の概要

3. デジタルバイオマーカーを用いた摂食障害の予測モデルの構築

本節では、デジタルバイオマーカーを用いた摂食障害の予測モデルの構築方法について説明する。

3.1 予測モデル構築に利用するデータ

摂食障害群と、正常な食事をしている対照群から、スマートフォンを使用し、加速度計の測定値、GPS の位置情報、アプリケーションの使用状況などのデータを収集する。これらのデータを統合して機械学習モデルに組み込むことにより、高度な行動パターンの分類が可能となるのではないかと考える (図 2)。

予測モデルの開発に使用するデータは、スマートフォンアプリ、AWARE^{*4}を使用して収集されたものである。収集されたデータの内容を表 1 に示す。

位置情報からは行動パターンや訪問場所の分析が可能である。例えば、過食症患者の特異な食事行動を特定することができる可能性がある。また、加速度センサデータを活用することで、身体活動の低下や活動リズムの変調を検出することができる。スマートフォンアプリをどのように使用し、どのくらいの頻度で使用しているかは、その人の精神的健康や行動について貴重な手掛かりになると予想する。例えば深夜に SNS を長時間見る、過剰な頻度でスマートフォンを使用するといった行動が関連する可能性があるのではないかと考える。

表 1 収集したセンサデータ一覧
スマートフォンのセンサデータ

ジャイロセンサ
加速度計
位置情報
スクリーンのオン/オフ時間
フォアグラウンドでのアプリの使用状況
アプリの通知
Bluetooth
バッテリーの充電状況
バッテリーの放電状況
バッテリーの状態
画面のタッチ情報
画面の明るさ
キーボード入力情報

3.2 機械学習を用いた予測モデルの構築

分析を行うために、カテゴリ項目はベクトル化し、機械学習モデルが処理できるように前処理を行う。そして、XGBoost、ランダムフォレスト、SVM、ロジスティック回帰の 4 つの機械学習モデルを使用し、比較実験を行う。それぞれのモデルについて、5 分割の交差検証を行うことでモデルに対する評価を行い、摂食障害の予測が適切であるかを考察する。

また、今後は時系列を考慮するため LSTM を使用する。初めは 1 日単位でのデータの変化を分析し、その後は午前と午後、6 時間ごと、自宅にいる時間と学校にいる時間など、様々な時間で分けて比較を行う。これにより、夜間に SNS を見る頻度が高いといった特徴や、平日と休日の行動の違いなどが分析でき、摂食障害かどうかの予測に役立つのではないかと考える。

^{*4} AWARE <https://awareframework.com/>

4. 各モデルの分類精度検証実験

3.2 節で説明した予測モデルを実際に構築し、その分類精度について評価実験を行った。

4.1 実験内容

データは 2024 年 5 月 6 日から 2024 年 5 月 12 日の 7 日間でスマートフォンアプリ AWARE を使用してイギリスで収集された 40 人分のデータに対して分析を行う。40 人のうち摂食障害患者は 24 人であった。センサデータはジャイロセンサ、加速度計、スクリーンのオン/オフ時間、フォアグラウンドでのアプリの使用状況、アプリの通知、バッテリーの充電状況、バッテリーの放電状況、バッテリーの状態、画面のタッチ情報、画面の明るさキーボード入力情報を使用した。また、一人当たり 1 レコードのデータとなるように、それぞれ 7 日分のデータを平均値、標準偏差、最大、最小の値となるように計算した。

3.2 節で説明した、XGBoost、ランダムフォレスト、SVM、ロジスティック回帰の 4 つのモデルを使用し、5-fold 交差検証を行う。また、特徴量の寄与度を可視化し、どのセンサデータが摂食障害の判別に寄与しているかを確認する。それぞれのモデルで寄与度が高い特徴量を使用し、特徴量の数を変更することで精度が向上するかを検証する。

4.2 実験結果

4.2.1 分類精度の評価結果

XGBoost、ランダムフォレスト、SVM、ロジスティック回帰の 5-fold 交差検証の結果を表 2 として示す。

表 2 5-fold 交差検証の結果

モデル	Accuracy	F1 Score	Precision	Recall
XGBoost	0.63	0.60	0.63	0.61
RandomForest	0.68	0.66	0.72	0.68
SVM	0.60	0.42	0.40	0.51
LogisticRegression	0.68	0.65	0.71	0.66

ランダムフォレストおよびロジスティック回帰が他のモデルよりも一貫して高い精度を示した。特にランダムフォレストは、Precision および Recall において最も高い値であった。

XGBoost は全体的に中程度の性能を示し、F1 スコアは 0.60 と比較的良好であるが、Accuracy は 0.63 とやや劣る結果となった。

SVM は最も低い F1 スコアおよび Precision を記録し、全体的に他モデルに比べて劣る結果となった。

4.2.2 それぞれのモデルに対する特徴量の寄与度の結果

本研究では、各モデルがどの特徴量に注目して予測を行っているかを可視化するため、SHAP 値 (SHapley Additive

exPlanations) を用いた解析を行った。以下に、XGBoost、ランダムフォレスト、ロジスティック回帰、SVM、それぞれのモデルにおける waterfall plot および heatmap の結果を示す。

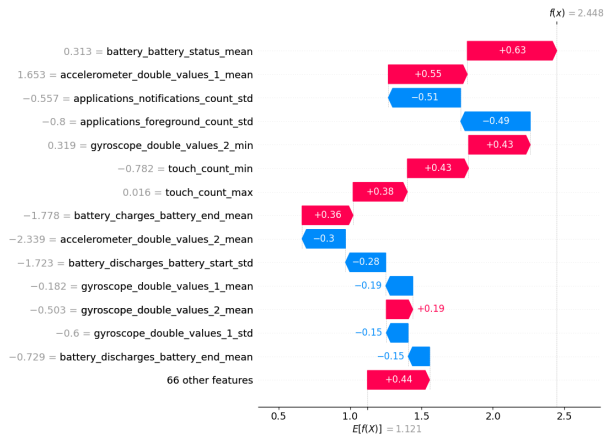


図 3 XGBoost の waterfall plot

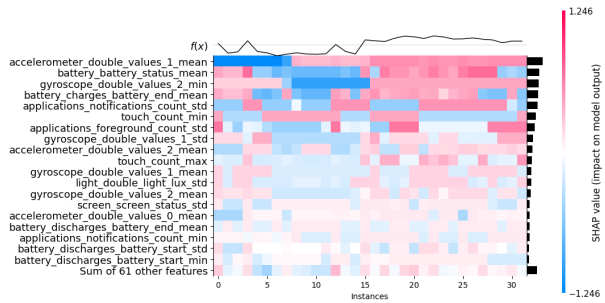


図 4 XGBoost の heatmap

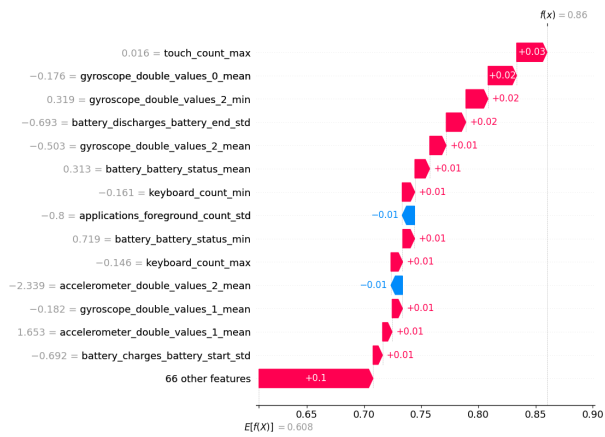


図 5 ランダムフォレストの waterfall plot

全体として、加速度・バッテリー・ジャイロ系の特徴量が高い寄与を示し、これらがモデルの予測に大きな影響を与えていることが分かった。

XGBoost やランダムフォレストでは、複数の重要特徴量が個々のインスタンスに対して柔軟に寄与しており、非

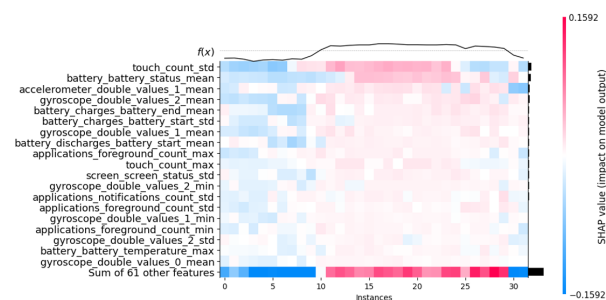


図 6 ランダムフォレストの heatmap

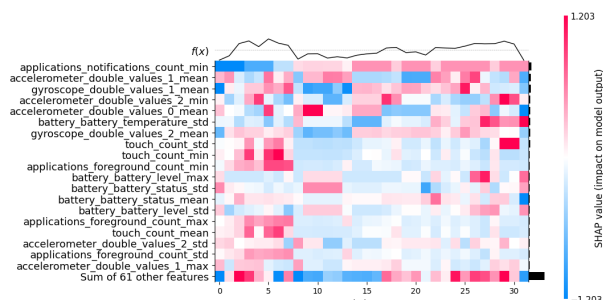


図 10 ロジスティック回帰の heatmap

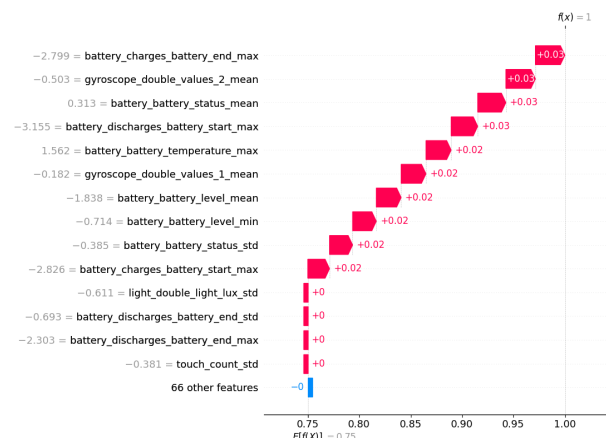


図 7 SVM の waterfall plot

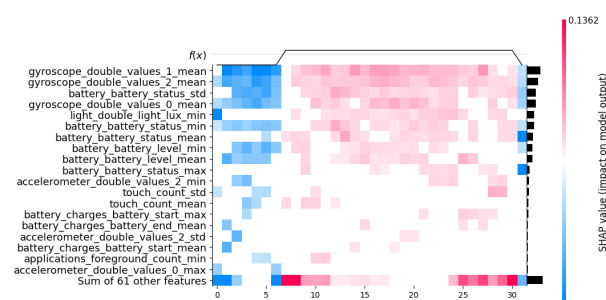


図 8 SVM の heatmap

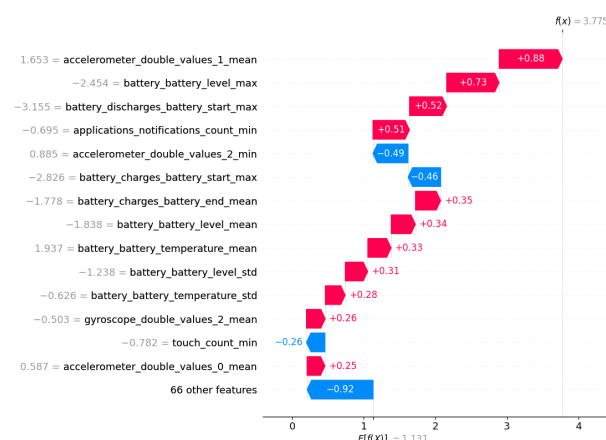


図 9 ロジスティック回帰の waterfall plot

線形モデルならではの複雑な予測根拠が確認された。一方で、ロジスティック回帰は、重要特徴量の寄与方向が一貫

しており、予測理由が明確で解釈しやすい特性が見られた。SVM は全体的に SHAP 値の変動幅が小さく、他モデルに比べて寄与の明瞭さや予測根拠の可視性がやや劣る傾向があった。

4.2.3 特徴量を変更した場合の分類精度の評価結果

特徴量の数は 10, 20, 30, 40 に変更し、4.2.1 節と同じく 5-fold 交差検証を行った。Accuracy, F1 Score, Precision, Recall のそれぞれの値の結果を表 3, 表 4, 表 5, 表 6 に示す。

表 3 特徴量数の変更による各モデルの Accuracy の比較

特徴量数	10	20	30	40	80(全件)
XGBoost	0.70	0.68	0.68	0.65	0.63
Random Forest	0.65	0.70	0.70	0.70	0.68
SVM	0.70	0.63	0.60	0.60	0.60
Logistic Regression	0.78	0.70	0.70	0.65	0.68

表 4 特徴量数の変更による各モデルの F1 スコアの比較

特徴量数	10	20	30	40	80(全件)
XGBoost	0.68	0.66	0.66	0.63	0.60
Random Forest	0.62	0.67	0.69	0.69	0.66
SVM	0.61	0.48	0.42	0.46	0.42
Logistic Regression	0.75	0.66	0.66	0.60	0.65

表 5 特徴量数の変更による各モデルの Precision の比較

特徴量数	10	20	30	40	80(全件)
XGBoost	0.71	0.69	0.69	0.70	0.63
Random Forest	0.67	0.74	0.73	0.71	0.72
SVM	0.65	0.51	0.40	0.46	0.40
Logistic Regression	0.81	0.64	0.64	0.60	0.71

表 6 特徴量数の変更による各モデルの Recall の比較

特徴量数	10	20	30	40	80(全件)
XGBoost	0.70	0.68	0.68	0.66	0.61
Random Forest	0.64	0.69	0.70	0.70	0.68
SVM	0.65	0.55	0.51	0.53	0.51
Logistic Regression	0.78	0.70	0.70	0.63	0.66

XGBoost では、特徴量数 10 件時に最も高い Accuracy および F1 スコアを示したが、特徴量を増加させるにつれて徐々に性能が低下した。

ランダムフォレストでは、特徴量数 30 前後で最も安定した高性能を示し、重要特徴量を適切に選択することで精度と汎化性能のバランスが取れることが確認された。

SVM においては、全体的に性能が低く、特に F1 スコアは特徴量数が 20 件を超えると著しく低下した。特徴量数の増加による性能向上は見られなかった。

ロジスティック回帰では、特徴量数 10 件のときに最も高い Accuracy および F1 スコアを示し、その後は特徴量数が増加するにつれて性能が若干低下した。

4.3 考察

5-fold 交差検証の結果、ランダムフォレストおよびロジスティック回帰が最も安定した性能を示し、いずれの評価指標においても高い数値を記録した。特にランダムフォレストは Precision, Recall ともに最も高く、誤判定を抑えつつ幅広く正例を捉えるバランスのとれたモデルであることが示されたと考える。一方、SVM は F1 スコアが 0.42 と極端に低い結果となった。そのため、今回のタスクに対して適さない可能性があると考えられる。

heatmap および waterfall plot の可視化からは、共通して加速度センサやジャイロセンサ、バッテリー関連の指標が寄与度の高い特徴量として現れた。これは、摂食障害のユーザの日常行動パターンや端末の使用状況と密接に関連しており、予測モデルが日常生活における無意識的な行動変化を捉えている可能性があると考えられる。特に XGBoost やランダムフォレストの waterfall plot では、これらの特徴量が正負いずれの方向にも強く寄与しており、モデルが状況に応じた柔軟な判断をしていると考えられる。

SHAP に基づき寄与度の高い特徴量のみを用いることで、多くのモデルにおいて性能が向上することが確認された。XGBoost およびロジスティック回帰では、上位 10 件の特徴量のみを使用したときに最も高い F1 スコアと Accuracy を示し、全特徴量を用いた場合よりも良好な結果を得られた。ランダムフォレストでは、上位 20~40 件の特徴量で最高性能を記録した。一方、SVM では常に他モデルと比較して低いスコアに留まった。これは、SVM が今回のように多様で非線形なデータ構造を持つセンサデータにおいて、柔軟な適応が難しかったためと考えられる。以上のことから、Random Forest と Logistic Regression は、予測精度と説明可能性のバランスに優れており、実運用への応用可能性が高いと考える。

5. まとめ

本稿では、スマートフォンのセンサデータを用いて摂食障害の予測モデルの構築とその分類精度の評価を行った。

結果として、ランダムフォレストおよびロジスティック回帰が高い予測性能と安定性を示し、SHAP による特徴量選択が性能向上に有効であることが確認された。また、少数の重要な特徴量に限定することでモデルの精度が向上した。

今後は、SHAP で選定された特徴量の数をより細かく調整し（例：5 件刻み）、各モデルにおける最適な特徴量数を探索する予定である。これにより、性能向上とモデル簡素化の両立が期待される。また、LSTM などの時系列を考慮するモデルを検討し、活動量や端末使用の継続的な変化を捉え、予測の精度や早期発見の可能性を高めたい。

謝辞 本研究の一部は、京都産業大学先端科学技術研究所（人間情報学研究センター）共同研究プロジェクト（M2301）の助成を受けたものである。ここに記して謝意を表す。

参考文献

- [1] 田上 哲也 and 葛谷 英嗣. 中枢性摂食異常症. *医療*, 55(12):624–627, 2001.
- [2] Anjali Devakumar, Jay Modh, Bahador Saket, Eric P. S. Baumer, and Munmun De Choudhury. A review on strategies for data collection, reflection, and communication in eating disorder apps. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery.
- [3] Antoine Piau, Katherine Wild, Nora Mattek, and Jeffrey Kaye. Current state of digital biomarker technologies for real-life, home-based monitoring of cognitive function for mild cognitive impairment to mild alzheimer disease and implications for clinical care: Systematic review. *J Med Internet Res*, 21(8):e12785, Aug 2019.
- [4] Stevie Chancellor, Zhiyuan Lin, Erica L. Goodman, Stephanie Zerwas, and Munmun De Choudhury. Quantifying and predicting mental illness severity in online pro-eating disorder communities. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16, page 1171–1184, New York, NY, USA, 2016. Association for Computing Machinery.
- [5] Tao Wang, Markus Brede, Antonella Ianni, and Emmanouil Mentzakis. Detecting and characterizing eating-disorder communities on social media. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, WSDM '17, page 91–100, New York, NY, USA, 2017. Association for Computing Machinery.
- [6] 航太 阪口, 拓人 佐久間, and 昇平 加藤. アンサンブル学習を用いたスマートフォンログからのうつ状態判別モデル. In *第 85 回全国大会講演論文集*, volume 2023, pages 645–646, feb 2023.
- [7] 田近 亜蘭, 熊谷 成将, and 古川 壽亮. スマートフォンアプリとウェアラブルデバイスを用いた、寛解期のうつ病患者の再発予測. *予防精神医学*, 7(1):3–12, 2022.
- [8] Isaac Moshe, Yannik Terhorst, Kennedy Opoku Asare, Lasse Bosse Sander, Denzil Ferreira, Harald Baumeister, David C. Mohr, and Laura Pulkki-Råback. Predicting symptoms of depression and anxiety using smartphone and wearable data. *Frontiers in Psychiatry*, 12, 2021.