

表情特徴量を用いたせん妄判別モデルに関する探索的研究

Exploratory study on delirium classification model using facial expression features

仁保 貴耀^{*1}

Takaaki Niho

湯口 彰重^{*1}

Akishige Yuguchi

松本 吉央^{*1}

Yoshio Matsumoto

大谷 清子^{*2}

Kiyoko Otani

小川 朝生^{*2}

Asao Ogawa

岡留 有哉^{*1}

Yuya Okadome

^{*1}東京理科大学

Tokyo University of Science

^{*2}国立がん研究センター

National Cancer Center

近年の機械学習技術の性能の向上により、医療の現場での診断支援等に機械学習を活用する研究が進められている。その中には、動画から表情情報を定量化し、患者が疾患に罹患しているかを判別する研究が行われている。この技術により、せん妄のような専門家でも判断の難しい疾患に対する診断補助手法の開発が期待されている。本研究では、医療関係者と患者の会話を収録した動画から、特徴量設計と様々な機械学習手法の組み合わせによって、高精度でせん妄を判別できる手法を探索的に検討する。実験結果より、 k 近傍法を用いることや先頭 30 フレームの標準偏差を特徴量とすることがせん妄を判別する上で精度が高くなる傾向があることがわかった。

1. はじめに

近年の機械学習技術の性能の向上により、医療の現場での診断支援等に機械学習を活用する研究が進められている。その一例として、表情から病気の初期兆候が見られるパーキンソン病において、表情の分析を用いた診断支援システムの開発が進められている [Lim 22]。このような技術により、パーキンソン病だけに限らず、専門家でも判断の難しい疾患に対する診断補助手法の開発が期待されている。

せん妄は入院患者の 20~30%に見られる意識障害である。せん妄を患うことで、カテーテルや点滴の自己抜去や転倒などによるケガの原因、認知症の進行、入院の長期化、コミュニケーションの困難などのリスクが発生する。また、現在の日本の医療現場では、せん妄を診断できる精神科医が少なく、臨床診断が大切なせん妄にとっては人手不足は重要な問題である。実際に、経験に基づいた看護師の判断でも、せん妄患者の 8 割近くを見落していると報告されている [小川 21]。

一方で、せん妄患者は非せん妄患者と比べて思考している時の表情に違いが出やすいことが指摘されている [生田 23]。そこで、その表情の違いや変化を定量化し、機械学習手法を用いてせん妄を分類できるモデルを検討する。それにより、せん妄の診断の補助になり、症状の見落としや人手不足の問題の解決につながる可能性がある。

従来のせん妄に関する研究では、電子健康記録データなどを特徴量として使い、ロジスティック回帰によるせん妄の予測精度が曲線下面積 0.87 という結果が得られている [Boogaard 12, Kalisvaart 06]。しかし、この研究は ICU や高齢患者などの患者が限定的かつ、せん妄になり得る可能性を予測するものである。せん妄は一過性の症状であることが多いため、予測タスクでは突発的に発症した場合への対処が困難である。また、様々な臨床場面でせん妄の発症頻度が高いことから、通常の臨床現場において、リアルタイムでせん妄を判別できるシステムが必要である。そこで、表情特徴を特徴量として勾配ブースティング木 (XGBoost) による分類モデルを作成した結果、分類精度が 91% という結果が得られている [生田 23]。この研究はリアルタイムでのせん妄判別の可能性を示唆している。しかし、機械学習手法が XGBoost のみであるため、データの構造

が XGBoost に適しているか確証はない。また、患者のデータ数が少なく、せん妄患者の判別が可能かが定かではないという問題点がある。

本研究では、医療関係者と患者の会話を収録した対話動画を用いて、特徴量設計と様々な機械学習手法の組み合わせから、高精度でせん妄を判別できる手法を探索的に検討する。まず、収録した動画に対して、表情情報を抽出する。その後、医療関係者と患者の会話場面の患者の応答した部分 (応答部分) の区間にラベル付けを行う。そして、4 種類の統計量を算出し、算出した統計量を特徴量として 7 種類の機械学習手法を適用し、せん妄の分類精度を比較する。

実験結果から、機械学習の観点では、ロジスティック回帰、サポートベクターマシン、ランダムフォレスト、XGBoost を用いるよりも、 k 近傍法を用いることでせん妄を判別する上で精度が高くなることがわかった。また、特徴量に標準偏差を使用することでせん妄を判別する上で精度が高くなる傾向があることがわかった。

2. せん妄患者表情データセット

本研究では、医療関係者と入院患者の会話場面の表情情報からせん妄を判別する機械学習モデルの作成を行う。そのため、会話場면을収録した対話動画を提供していただき、それを基に、特徴量の設計を行う。

2.1 収録データ

本研究で使用する対話動画は、国立がん研究センター東病院にて、医療関係者と入院患者との会話場면을収録したものである。図 1 にデータセットのサンプル画像を示す。収録した動画には、せん妄患者 12 名、非せん妄患者 19 名の合計 31 名分の入院患者の会話場面が収録されている。各動画の記録時間は 3 分程度で収められている。収録は全方位集音マイクを搭載し、手振れ補正や日常の様々な照明条件で約 30 fps で撮影可能なキヤノン株式会社製の Powershot v10^{*1} を用いた。収録した会話場面の例を図 1 に示す。

収録した会話場面のうち使用する区間を検討した。本研究では患者の発話区間の先頭 10 フレーム (0.3 秒) もしくは 30 フレーム (1.0 秒) を使用する。これは患者の発話区間が最小

連絡先: 岡留有哉, 東京理科大学工学部情報工学科, 東京都葛飾区新宿 6-3-1, 03-5876-1681, okadome@rs.tus.ac.jp

^{*1} <https://personal.canon.jp/product/compact-digital-camera/powershot/v10>



図 1: 会話場面のサンプル画像

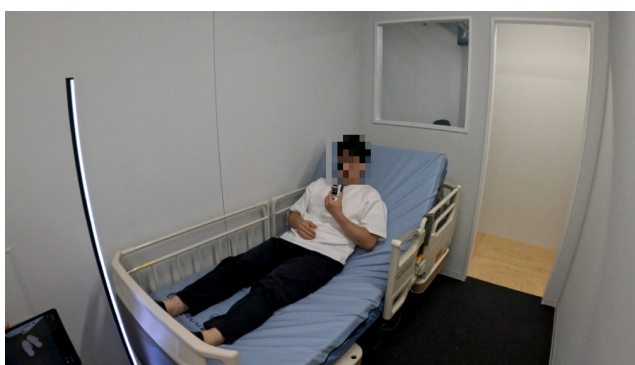


図 2: リビングラボ内の患者用ベッドで照明条件を確認する様子

で約 30 フレームであったためである。また、対話場面において、せん妄患者の反応潜在時間が延長することを考慮し、先頭 10 フレームを採用している。発話区間はアノテーションツールを使用してラベル付けする。

なお、国立がん研究センター東病院の病棟での会話場面の照明環境下で表情を視認できるかを東京理科大学葛飾キャンパスのリビングラボ内の患者用ベッド上にて、様々な照明環境下で確認した (図 2)。実際の病棟での撮影環境である、窓側、通路側での明るさをカバーする 17.5 lx から 520.7 lx の範囲でかつ手持ち撮影で表情を視認できることを確認した。

2.2 特徴量設計

次に、せん妄判別モデルを作成するための特徴量を設計する。医療関係者と患者の会話場面の動画に対し Py-Feat [Cheong 23] を用いて表情情報である Facial Action Unit (AU) [Ekman 02] を抽出する。また、患者一人に対して抽出した AU の分布を図 3 に示す。Py-Feat は 20 種類の目元周りや口元周りを中心とした AU が抽出できるが、図 3 に示すように、その中の番号 7, 11, 20 は値が本来の仕様と違い、常に二値であったため今回は使用せず、表 1 に内訳を示す 17 種類の AU (AU1, 2, 4, 5, 6, 9, 10, 12, 14, 15, 17, 23, 24, 25, 26, 28, 43) を使用した。

次に、算出する統計量は、患者の発話区間の先頭 10 フレームもしくは 30 フレームに対しての AU 毎の平均 (μ)、標準偏差 (σ) である。生田らの研究 [生田 23] では患者の発話区間の 3 秒間を使用し、その 3 秒間の分散を特徴量としている。今回の学習に使用する特徴量は、フレーム数と統計量の算出方法に応じた以下の式 (1) から (4) の 4 通りを設計する (x_i : i 番目のフレームの値)。

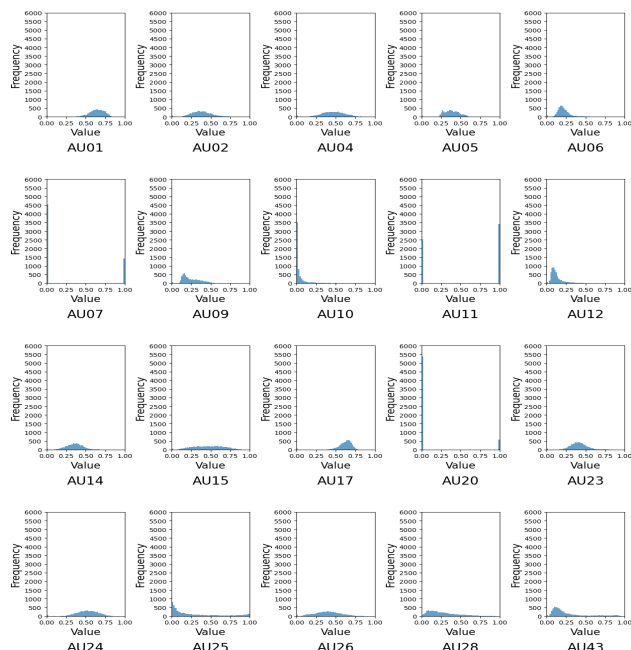


図 3: AU の分布例

表 1: Py-Feat で抽出し本研究で用いる AU の内訳

AU 番号	AU の部位・動作	AU 番号	AU の部位・動作
1	眉の内側を上げる	15	唇両端を下げる
2	眉の外側を上げる	17	おとがいを上げる
4	眉を下げる	23	唇を固く閉じる
5	上脣を上げる	24	唇を押し付ける
6	頬を持ち上げる	25	顎を下げて唇を開く
9	鼻に皺を寄せる	26	顎を下げて唇を開く
10	上唇を上げる	28	唇を吸い込む
12	鼻唇溝を深める	43	閉眼
14	えくぼを作る		

$$\mu_{30} = \frac{1}{30} \sum_{i=1}^{30} x_i \quad (1)$$

$$\mu_{10} = \frac{1}{10} \sum_{i=1}^{10} x_i \quad (2)$$

$$\sigma_{30} = \sqrt{\frac{1}{30} \sum_{i=1}^{30} (x_i - \mu_{30})^2} \quad (3)$$

$$\sigma_{10} = \sqrt{\frac{1}{10} \sum_{i=1}^{10} (x_i - \mu_{10})^2} \quad (4)$$

ここで、 μ_{10}, μ_{30} とは各動画に対して、10, 30 フレームの各 AU 値の平均を算出し、 σ_{10}, σ_{30} とは各動画に対して、10, 30 フレームの各 AU 値の標準偏差を算出することを意味する。

3. せん妄判別モデルの性能評価実験

本実験では、せん妄判別のために複数の機械学習手法を使用して分類モデルを構築し、モデルの性能を比較する。モデルの学習では、入院患者 31 名分のデータのうち、30 名分を学習データとして学習させ、残り 1 名分をテストデータとして評価する、leave one person out 検証を行う [原嶋 21]。

本実験で使用した機械学習手法について、本データセットに対する性能を探索的に調査するために、以下の基本的な機械学

表 2: 7 種類の機械学習手法によるせん妄判別モデルの推定結果 (%)

	特徴量							
	μ_{30}		μ_{10}		σ_{30}		σ_{10}	
	感度	特異度	感度	特異度	感度	特異度	感度	特異度
線形回帰	33	79	33	84	25	79	42	84
ロジスティック回帰	8	100	8	89	8	95	8	95
サポートベクターマシン	8	84	8	79	8	95	0	95
決定木	25	47	33	68	42	74	25	89
ランダムフォレスト	0	95	8	100	33	89	42	100
XGBoost	25	79	25	95	0	100	8	100
k 近傍法	42	74	42	79	42	95	33	100

習手法と生田らの研究 [生田 23] で用いられた勾配ブースティング木 (XGBoost) の 7 通りの手法を検討する。

- 線形回帰：

線形回帰は目的変数と説明変数の線形関係をモデル化する手法である [Hope 20]。

- ロジスティック回帰：

ロジスティック回帰は複数の説明変数から線形結合によって予測値を算出し、シグモイド関数で求めた確率値に基づいて分類する手法である [Kumar 24]。

- サポートベクターマシン (SVM)：

サポートベクターマシンは異なるカテゴリのデータ点間に最大のマージンを持つ境界線を引くことで分類を行う手法である [Chandra 21]。

- 決定木：

決定木は木構造を持ち、各ノードで特徴量の条件分岐を行い、最終的に葉ノードでクラスを決定する手法である [Myles 04]。

- ランダムフォレスト：

ランダムフォレストはブートストラップ法を使用して複数のデータ群を生成し、それぞれのデータ群に対して決定木から予測結果を算出、そして、それらの予測結果の平均をとる手法である [Rigatti 17]。

- 勾配ブースティング木 (XGBoost)：

勾配ブースティング木である XGBoost は決定木を複数構築し、それらを組み合わせて全体として強力な予測モデルを作り上げる [Chen 16]。

- k 近傍法：

k 近傍法は新しいデータ点を予測するときに、学習データから最も近い k 個の近傍点を探し、その近傍点のなかで最も多く出現するクラスに新しいデータ点を割り当てる手法である [Zhang 17]。

線形回帰以外の各機械学習手法のパラメータチューニングはグリッドサーチを行っている。グリッドサーチとは探索空間内のすべての値の組み合わせを試行し、最適なパラメータを見つける手法であり、本研究では、学習データ 30 名分に対して 3 分割の交差検証で最適なハイパーパラメータを決定する。グリッドサーチ後の最適なハイパーパラメータを使用してモデル

を再訓練させ、モデルをテストデータに適用して、せん妄の判別の予測を行う。使用したハイパーパラメータは表 3 のとおりである。

本実験での評価方法は感度、特異度を用いる。感度、特異度は、それぞれ患者の疾患の有無を判定できる割合を示す重要な評価指標である。ここで、感度とは、真に疾患を有する者を検査によって陽性と正しく判定できた割合のことであり、特異度とは真に疾患を有さない者を検査によって陰性と正しく判定できた割合のことである。

4. せん妄判別モデルの推定結果と考察

表 2 に 7 種類の機械学習のせん妄の判別結果を示す。表の太字は感度と特異度ともに精度が高い結果を表している。まず、特異度の結果に着目すると、7 種類の機械学習手法のほとんどで 80% 以上となる特徴量との組み合わせがある。次に、感度の結果に着目すると、どのような特徴量を用いても決定木やランダムフォレスト、XGBoost、ロジスティック回帰、サポートベクターマシンはほとんどが 30% 以下である。一方、線形回帰や k 近傍法は感度が 30% 以上のものがほとんどである。また、特徴量で比較すると、感度では平均と標準偏差でそこまで差はないが、特異度では標準偏差を用いる方が高い結果を示す傾向がある。

以上の結果から、木構造に基づく複雑なアルゴリズムの機械学習手法を使用するよりも、 k 近傍法や線形回帰といった距離関数に基づくアルゴリズムで精度良くせん妄の判別が可能であることが示唆された。これは、木構造アルゴリズムが今回使用したデータに対して適していない可能性が考えられる。また、 k 近傍法と比較した際にロジスティック回帰や線形回帰の精度が低下した理由として、データが非線形に分布していることが影響している可能性も考えられる。さらに、平均と標準偏差を比較した際に特異度に差が出たことについては、標準偏差が非せん妄患者のデータの変動を平均よりも反映しやすく、個人差の影響も受けづらいということから、標準偏差の方が今回のデータセットには向いている可能性が考えられる。

5. 結論

本研究では、せん妄を高い確率で分類できるせん妄予測モデルの構築をめざし、4 種類の特徴量に対し、7 種類の機械学習手法を調査した。まず、医療関係者と患者の会話場面を収録したデータセットを提供していただき、表情特徴量として患者の発話区間から先頭 10 フレームもしくは先頭 30 フレームの AU を抽出した。次に、抽出した AU に対して 4 種類の特徴量設計を行い、4 種類それぞれを特徴量として 7 種類の機械学習

表 3: 各学習手法で使用したパラメータ

学習手法	パラメータ	設定値
ロジスティック回帰	正則化強度 (C) ペナルティ (penalty) ソルバー (solver)	{0.01, 0.1, 1, 10, 100} L2 lbfgs
SVM	正則化強度 (C) ガンマ (gamma) カーネル (kernel)	{0.1, 1, 10, 100} {1, 0.1, 0.01, 0.001} rbf
ランダムフォレスト	決定木の本数 (n_estimators) 木の最大深さ (max_depth) 分割の最小サンプル数 (min_samples_split) 最大特徴量数 (max_features)	{100, 150} {10, 15, 20} {2, 5} sqrt
勾配ブースティング (XGBoost)	木の最大深さ (max_depth) 学習率 (learning_rate) 決定木の数 (n_estimators) サブサンプル比率 (subsample) 特徴量のサンプリング比率 (colsample_bytree)	{3, 5, 7} {0.1, 0.01} {50, 100, 200} {0.8, 1.0} {0.8, 1.0}
k 近傍法 (kNN)	近傍数 (n_neighbors) 重み (weights) 距離の種類 (metric)	{3, 5, 7, 9} {uniform, distance} {euclidean, manhattan, minkowski}

手法を検討し、性能を比較する。その結果、k 近傍法と標準偏差を使用するモデルの分類精度が高いという結果が得られた。

今後、精度の向上を目指して、対話システムを用いた声かけの統制や、これらのモデルと併用して音声データのようなマルチモーダル情報を使用することを検討したい。これは、せん妄患者と非せん妄患者には発話において差異がみられたためである。また、特徴量の検討かつ重要度の可視化、特徴量の相関関係の調査、臨床の治験との整合性などを検討したい。

謝辞

本研究は、AMED の課題番号 JP24ym0126804 の支援、および国立がん研究センター橋渡し研究推進センター (CPOT) の支援、さらに JSPS 科研費 23K09731 の支援を受けて行われたものである。

参考文献

- [Boogaard 12] Boogaard, Van den M., Pickkers, P., Slooter, A., Kuiper, M., Spronk, P., Van Der Voort, P., Van Der Hoeven, J., Donders, R., Achterberg, van T., and Schoonhoven, L.: Development and validation of PRE-DELIRIC (PREdiction of DELIRium in ICu patients) delirium prediction model for intensive care patients: observational multicentre study, *BMJ*, Vol. 344, (2012)
- [Chandra 21] Chandra, M. A. and Bedi, S.: Survey on SVM and their application in image classification, *International Journal of Information Technology*, Vol. 13, No. 5, pp. 1–11 (2021)
- [Chen 16] Chen, T. and Guestrin, C.: Xgboost: A scalable tree boosting system, in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 785–794 (2016)
- [Cheong 23] Cheong, J. H., Jolly, E., Xie, T., Byrne, S., Kenney, M., and Chang, L. J.: Py-Feat: Python Facial Expression Analysis Toolbox, *Affective Science*, Vol. 4, No. 4, pp. 781–796 (2023)
- [Ekman 02] Ekman, P., Friesen, W. V., and Hager, J. C.: *Facial Action Coding System*, Research Nexus, Salt Lake City, USA, 2nd edition (2002)
- [Hope 20] Hope, T. M.: Linear regression, in *Machine learning*, pp. 67–81, Elsevier (2020)
- [Kalisvaart 06] Kalisvaart, K. J., Vreeswijk, R., De Jonghe, J. F., Van Der Ploeg, T., Van Gool, W. A., and Eikelenboom, P.: Risk factors and prediction of postoperative delirium in elderly hip-surgery patients: implementation and validation of a medical risk factor model, *Journal of the American Geriatrics Society*, Vol. 54, No. 5, pp. 817–822 (2006)
- [Kumar 24] Kumar, S.: Logistic Regression, in *Python for Accounting and Finance: An Integrative Approach to Using Python for Research*, pp. 319–327 (2024)
- [Lim 22] Lim, W. S., Chiu, S.-I., Wu, M.-C., Tsai, S.-F., Wang, P.-H., Lin, K.-P., Chen, Y.-M., Peng, P.-L., Chen, Y.-Y., Jang, J.-S. R., and Lin, C.-H.: An integrated biometric voice and facial features for early detection of Parkinson's disease, *npj Parkinson's Disease*, Vol. 8, No. 145 (2022)
- [Myles 04] Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., and Brown, S. D.: An introduction to decision tree modeling, *Journal of Chemometrics: A Journal of the Chemometrics Society*, Vol. 18, No. 6, pp. 275–285 (2004)
- [Rigatti 17] Rigatti, S. J.: Random forest, *Journal of Insurance Medicine*, Vol. 47, No. 1, pp. 31–39 (2017)
- [Zhang 17] Zhang, S., Li, X., Zong, M., Zhu, X., and Cheng, D.: Learning k for knn classification, *ACM Transactions on Intelligent Systems and Technology (TIST)*, Vol. 8, No. 3, pp. 1–19 (2017)
- [原嶋 21] 原嶋 春輝, 荒川 豊, 石田 繁巳, 中村 優吾: ウェアラブル心拍センサによるワーク・エンゲイジメントの推定, 第 29 回マルチメディア通信と分散処理ワークショップ論文集, pp. 115–122 (2021)
- [小川 21] 小川 朝生: せん妄と転倒, 日本転倒予防学会誌, Vol. 7, No. 3, pp. 19–21 (2021)
- [生田 23] 生田 光輝, 松本 吉央, 小川 朝生, 大谷 清子: 表情情報を用いた「せん妄」患者の識別モデル生成と特徴分析, 第 24 回計測自動制御学会システムインテグレーション部門講演会 (SI2023), pp. 1522–1525 (2023)