

自己教師あり学習を用いた

自由会話音声からの早期アルツハイマー病の予測

桑山芳明¹ 坪倉和哉² 入部百合絵^{1,2} 横井克典³ 中村昭範⁴ 北岡教英⁵ 勝野雅央⁶

¹愛知県立大学 情報科学部 ²愛知県立大学 大学院情報科学研究科

³国立研究開発法人 国立長寿医療研究センター 脳神経内科部

⁴国立研究開発法人 国立長寿医療研究センター 認知症先進医療開発センター
バイオマーカー開発研究部

⁵豊橋技術科学大学 情報・知能工学系 ⁶名古屋大学 大学院医学系研究科

{is211027, id231001}@cis.aichi-pu.ac.jp iribe@ist.aichi-pu.ac.jp

概要

本研究は、アミロイド PET やタウ PET など経済的・身体的負担の大きい早期アルツハイマー病の検査の代替として、雑談を含む自由会話音声から早期アルツハイマー病を識別する手法を提案する。具体的には、自由会話音声とそれを書き起こしたテキストを HuBERT と RoBERTa を活用して音声・言語特徴量を特徴抽出器として用い、識別を試みた。その結果、特徴抽出器に高齢者音声や認知障害特有の音声を学習させることにより、従来手法である音響的特徴量を用いた手法と比較して、F 値で 8%、正解率で 7% 上回る精度で正常群とプレクリニカル AD 群を識別できることが示され、自由会話音声からの早期アルツハイマー病の予測への応用が期待できる。

1 はじめに

近年、認知症者は増加し、介護や病院などの逼迫による社会問題を引き起こしている。認知症全体の 60-70% を占めるアルツハイマー型認知症 (Alzheimer's Disease : AD) は早期治療により、症状の緩和や進行を遅らせることが可能である。一昨年、日本でも承認されたレカネマブ[1]も投与対象を AD の早期段階としており、依然として AD の早期発見が求められている。

AD の早期発見を目的とした研究の多くは神経細胞死が起こり始めた軽度認知症 (Mild Cognitive Impairment : MCI) 期を対象としている[2,3]。しかし死滅した神経細胞は回復しないため、MCI 期より前の段階での検出が必要である。そこで本研究では健常者 (Healthy Control : HC) でありつつも、AD の原因物質の蓄積が認められるプレクリニカル AD に着目する。現在プレクリニカル AD の検出方法には、

画像診断であるアミロイド・イメージング (A β PET) とタウ・イメージング (タウ PET) がある。これらの PET 検査はプレクリニカル AD の検出に有効な一方で、経済的・身体的負担も大きい。そのため、プレクリニカル AD を検出できる安価かつ負担の少ない検査が求められている。そこで、AD の症状が表出されるモダリティの中で、日常生活から容易に収集が可能であり、収集する際に受検者への負担の少ない自由会話音声を用いてプレクリニカル AD の検出を目指す。

AD の検出に自由会話音声を利用した先行研究は多い。しかしそれらの多くは、ボストン失語症のクッキー・セフト絵[2,3]の説明など、一定のタスク実行時の音声を自由会話音声として使用している。つまり、AD の発見のためには受検者が特定のタスクを実行する必要がある。また、特定のタスクに特化した会話内容であるため、プレクリニカル AD の特徴が現れにくい可能性もある。プレクリニカル AD の識別を考えると、一定のタスク実行時よりも日常生活の中で交わす雑談対話に近い自由会話音声の方が微細な変化まで捉えられる可能性がある。

上記背景の下、本研究では自由会話音声から得たテキストと録音音声を用いてプレクリニカル AD の識別を行う。

2 収集データとラベルの付与

本研究では、共同研究「認知症関連疾患の会話への影響—自然言語処理を用いた検討 [4]」により収集したデータを用いる。複数の認知機能検査の結果と検査時の音声、および専門医が診断した A β PET とタウ PET の結果を収集した。実施した認知機能検査は Alzheimer's Disease Assessment Scale (ADAS)[5] を含む 10 種類である。

表 1 被験者の内訳

ラベル	性別	50代	60代	70代	80代	合計	
AD	男	1	4	6	5	15	
	女	7	2	7	5	21	
MCI	男	0	1	8	6	15	
	女	0	1	2	3	6	
HC	A β +tau+	男	0	0	24	9	33
		女	0	3	11	5	19
	A β +tau-	男	0	2	10	2	14
		女	0	0	3	1	4
	A β -tau-	男	2	21	23	11	57
		女	0	4	20	5	29
合計		10	38	114	51	213	

実施した検査の1つである ADAS は自由会話を通して口語言語能力を評価する項目がある。この評価項目では、仕事や趣味などに関する会話が交わされており、雑談対話と内容が近い。そのため本研究では、この雑談対話中の自由会話音声を手書き起こし、被験者の発話部分のみを自由会話テキストとして用いた。また、A β PET 陽性から AD 発症までは 20 年以上の歳月がかかるため、50 歳以上を分析対象とした。さらに、AD 以外の神経変性疾患の可能性のあるデータは除外している。

収集データに対して、AD ラベルとプレクリニカル AD ラベルを付与した。AD ラベルは、複数の認知機能検査の結果をもとに AD, MCI, HC のいずれかを付与した。プレクリニカル AD ラベルは、HC の被験者に対して、PET 検査の結果をもとにプレクリニカル AD 後期、プレクリニカル AD 前期、正常のいずれかを付与した。被験者の年代、性別、ラベルの分布を表 1 に示す。

3 特徴量の抽出

AD の症状の多くは、記憶障害から症状が現れ始め、徐々に時間や場所などが認識できなくなる見当識障害が現れる。先行研究[6]により、認知機能の変化が主観的に経験される前のプレクリニカル AD の状態でも言語機能に関連があることが明らかになっている。そのため、本研究では事前学習済みモデル RoBERTa を用いた埋め込みベクトルの抽出や、GiNZA[7]を用いて算出した言語特徴量(13 品詞による総語数割合, 総発話数割合, TTR(Type-Token Ratio),

最大係受け距離の平均と分散などの 67 次元)を用いる。加えて、HuBERT から抽出された特徴ベクトルや eGeMAPSv02 を用いて抽出した音響特徴量 (Loudness, Harmonic-to-Noise Ratio, Pitch などに関連する 88 次元)を用いる。このような音響的特徴は、認知機能低下の早期段階における発話リズムや音響特性の変化を捉えることができ、早期アルツハイマー病の検出において重要な手がかりとなると考えられる[8]。また、言語的な内容に依存しないため、多言語環境での適用が可能であることも大きな利点である。特徴量の抽出手法は以下の①から⑨である。ここでは早期アルツハイマー病検出タスクに特化した事前学習済みモデルの有効性を確認することを目的とし、既存手法を①から⑤、提案手法を⑥から⑨を比較した。提案手法のモデルでは、高齢者音声に適応したモデルが構築されることを期待し、高齢者音声コーパス EARS[9],S-JNAS[10]を用いて、rinna 株式会社が公開している "rinna/japanese-hubert-base[11]" に対して図 1 のようなファインチューニングを実施した。ここでは、図 2 のように HuBERT からの層による Cross Entropy 損失を最小化するものを評価し、10 層目を選択した。また、識別対象の学習データを用いた対照学習については、RoBERTa モデルと HuBERT モデルから抽出し結合した特徴量に対して次元削減を行い、図 3 のように全結合層のパラメータを更新している。以上の提案手法の有用性を評価するため、既存手法①,②による高次元特徴量、あるいは既存手法③,④,⑤による事前学習モデルと比較することで、早期アルツハイマー病の音声言語特徴を学習した提案手法のモデルの有意性を示す。

[既存手法]

- 手法① 言語特徴量(語彙)のみを使用
- 手法② 音響特徴量(eGeMAPSv02)のみを使用
- 手法③ "rinna/japanese-roberta-base モデル"を使用
- 手法④ "rinna/japanese-hubert-base モデル"を使用
- 手法⑤ "facebook/hubert-base-ls960 モデル"を使用

[提案手法]

- 手法⑥ "rinna-hubert モデル"に対して対照学習を実施
- 手法⑦ "rinna-hubert モデル"+語彙に対して対照学習を実施
- 手法⑧ "rinna-hubert モデル"+"rinna-roberta モデル"に対して対照学習を実施
- 手法⑨ "rinna-hubert モデル"+"rinna-roberta モデル"+語彙に対して対照学習を実施

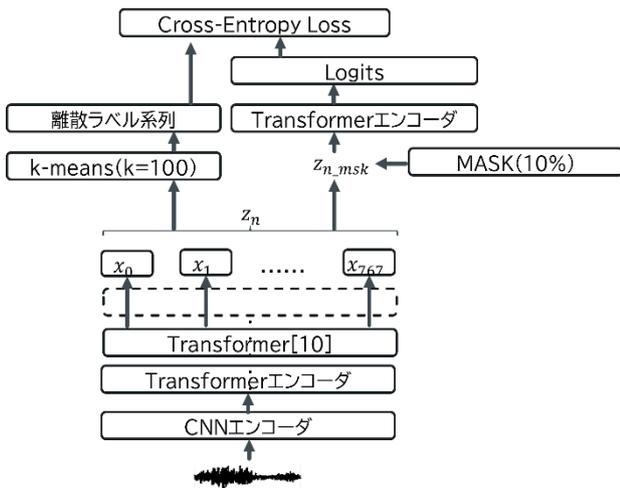


図 1 高齢者音声コーパスを用いた“rinna-hubert モデル”のファインチューニング手法

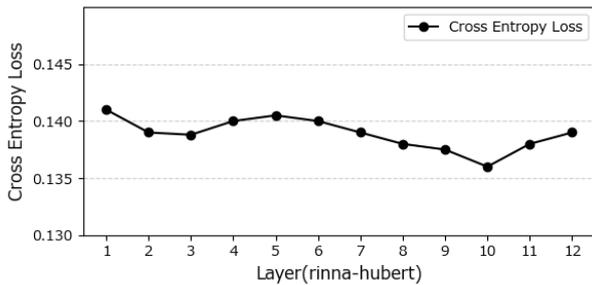


図 2 ファインチューニング時の各層における損失関数

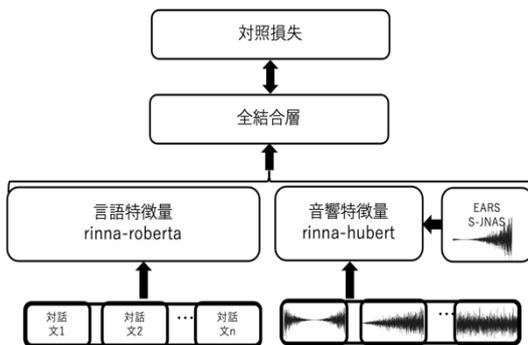


図 3 手法⑧の概要図

4 早期アルツハイマー病の識別

前章までに説明した通り、収集データに AD ラベルとして、HC と MCI と AD を付与し 3 群に分け、プレクリニカル AD ラベルとして、正常群とプレクリニカル AD 前期/後期群の 3 群に分けた。プレクリニカル AD の検出に関する先行研究は非常に限られており、特に軽度認知症 (MCI) より前の段階にお

ける言語 (音響) 的特徴を用いた研究はほとんど行われていないため、本研究ではまず正常群とプレクリニカル AD 群(プレクリカル AD 前期と後期を合わせたもの)の 2 群間に焦点を当て、その識別精度を向上させることを優先する。一方で、プレクリニカル AD 前期群は記憶力や認知機能といった点で正常及びプレクリニカル AD 後期に似た特徴を持つため、現時点ではその段階のみを検出する優先度を下げている。本研究の段階的アプローチにより、軽度認知障害以前の AD リスクの高い対象者を早期に識別するための基盤を構築することを目指す。

本研究では、識別にあたり群間のデータ数に偏りがあるためアンダーサンプリング法として NearMiss1 を適用し 2 群間の多数派クラスのデータ数を少数派クラスのデータ数に合わせた。また識別器には、SupportVectorClassification(SVC), LogisticRegression, XGBClassifier を採用した。データの分割には、5 分割交差検証法を用い、ハイパーパラメータ最適化は Optuna[12]を使用して評価基準を F 値とした。

以降、ラベル毎に各手法の F 値の最も高い識別結果を表 2-4 示す。

表 2 HC vs AD の識別結果

手法	識別器	F 値	正解率	感度	特異度
1	XGB	0.79	0.79	0.78	0.81
2	Logistic	0.76	0.78	0.69	0.86
3	Logistic	0.73	0.78	0.61	0.94
4	SVC	0.76	0.78	0.69	0.86
5	Logistic	0.69	0.71	0.64	0.78
6	SVC	0.70	0.71	0.69	0.72
7	Logistic	0.71	0.72	0.69	0.75
8	SVC	0.85	0.85	0.86	0.83
9	Logistic	0.74	0.74	0.75	0.72

表 3 HC vs MCI の識別結果

手法	識別器	F 値	正解率	感度	特異度
1	XGB	0.74	0.76	0.67	0.86
2	Logistic	0.65	0.57	0.81	0.33
3	Logistic	0.76	0.79	0.67	0.90
4	Logistic	0.70	0.71	0.67	0.76
5	Logistic	0.68	0.69	0.67	0.71
6	XGB	0.74	0.72	0.67	0.81
7	XGB	0.77	0.79	0.71	0.86
8	Logistic	0.78	0.79	0.76	0.81
9	SVC	0.67	0.67	0.67	0.67

表 4 正常群 vs プレ AD 群の識別結果

手法	識別器	F 値	正解率	感度	特異度
1	XGB	0.71	0.70	0.74	0.66
2	XGB	0.57	0.52	0.64	0.40
3	SVC	0.61	0.56	0.67	0.46
4	SVC	0.60	0.59	0.60	0.59
5	XGB	0.61	0.59	0.64	0.54
6	SVC	0.59	0.52	0.70	0.34
7	Logistic	0.59	0.59	0.59	0.60
8	Logistic	0.65	0.59	0.76	0.41
9	Logistic	0.63	0.63	0.64	0.61

5 考察

表 2 から 4 より、音響特徴量のみを用いた既存手法②と自己教師あり学習を組み合わせて学習を行った提案手法⑧を比較すると、F 値が 8-13%、正解率は 7-22%に向上していることが分かり、従来行われていた音響特徴量のみを用いた早期アルツハイマー病の識別に比べ、本研究で提案した手法が有効的だと言える。提案手法⑧は、言語モデルの RoBERTa と音響モデル HuBERT から抽出し結合した特徴量を用いているが、加えて本データによりパラメータ調整し適応化したモデルであるため、自己教師あり学習の有用性を示す結果となった。また、音響だけではなく言語特徴も組み合わせたことの有用性も示している。

表 2 と 3 より、アルツハイマー病や軽度認知症患者を健常者と識別するにあたり、提案手法⑧が既存手法①から⑤を比較し、最も高い F 値である 0.85 と 0.78 を得ていることがわかる。この結果から、高齢者や認知障害特有の音響変化を大規模言語・音声モデルに学習させることで、患者を見落とす確率が低下する可能性がある。

表 4 の正常群とプレクリニカル AD 群の識別についても、提案手法⑧の感度が 0.76 と既存手法①から⑤に比べて 0.02-0.16%高いことから、自己教師あり学習を組み合わせ、患者の発話や言語能力の傾向を学習することで、プレクリニカル AD の症状の特徴を捉え、軽度認知症よりもさらに前の段階の患者を見逃す確率を低下させることが期待できる。

一方、F 値や正解率を見ると精度向上の余地があることから、モデルの組み合わせや学習方法を改善することに取り組んでいきたい。

6 おわりに

本研究では、自由会話テキストを用いて、健常者と早期アルツハイマー病の識別を行った。自己教師あり学習を用いた複数の識別手法を提案した結果、従来用いられていた eGeMAPS の音響的特徴量を使った既存手法②やコーパスチューニング・対照学習を行っていない既存手法よりも高齢者や認知障害の特徴を学習した RoBERTa と HuBERT を特徴抽出器として組み合わせた提案手法⑧が、F 値や感度が高くなる場合が多く、早期アルツハイマー病の検出タスクにおいて有効的な手法である可能性が高いといえる。以上より、本研究の提案手法は受検者の負担の大きい PET 検査の必要性を判断する材料としての活用や、日常で交わす雑談からのアルツハイマー病の早期発見が期待できる。今後は、アップサンプリング法においても同程度の精度となることを検証する。また現状では感度 (Sensitivity) が十分に高いとは言えず、疾患を見逃すリスクが残されている。しかし、感度は疾患を正確に発見する重要な指標であり、早期診断や誤診の防止に寄与することから、感度が 1 に近づくようなモデルの最適化を図り精度向上を目指す。さらに、自動書き起こしを使用した人手を必要としない手法で同様の精度を得られることも目指す。

謝辞

本研究は JSPS 科研費 JP23H00493 の助成を受けたものです。また、本研究は国立研究開発法人産業技術総合研究所と行っている「高齢者を対象とした音声言語処理に関する研究開発」の一部として実施されたものです。

参考文献

1. エーザイ株式会社。「レケンビ® 点滴静注」(一般名:レカネマブ)について, 日本においてアルツハイマー病治療薬として製造販売承認を取得 — ニュースリリース :2023 年 . <https://www.eisai.co.jp/news/2023/news202359.html>, 2024-10-29 閲覧
2. Saturnino Luz, Fasih Haider, Sofia de la Fuente, Davida Fro-mm, Brian MacWhinney. Alzheimer's Dementia Recognition Through Spontaneous Speech: The ADReSS Challenge. *Proc. Int-erspeech 2020*, pp.2172-2176, doi: 10.21437/Interspeech.2020-2571, 2020.
3. Saturnino Luz, Fasih Haider, Sofia de la Fuente, Davida Fro-mm, Brian MacWhinney. Detecting Cognitive Decline Using S-peech Only: The ADReSSo Challenge. *Proc. Interspeech 2021*, pp.3780-3784, doi:10.21437/Interspeech.2021-1220, 2021.
4. 国立長寿医療研究センター, 受付番号 No.22TB8. <https://www.ncgg.go.jp/ncgg-kenkyu/ekigaku/22TB8.html>, 2024-10-29 閲覧.
5. 河月稔, 神経心理学的検査, 医学検査, Vol.66 No.J-STAGE-2 認知症予防のための検査特集 2017, pp.11-21, 2017.
6. Ihab Hajjar, et al. Development of digital voice biomarkers and associations with cognition, cerebrospinal biomarkers, and neural representation in early Alzheimer's disease, *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* 15(1), doi: 10.1002/dad2.12393, 2023.
7. 松田寛. (2020). GiNZA - Universal Dependencies による実用的日本語解析. *自然言語処理*, 27(3), 695-701. <https://doi.org/10.5715/jnlp.27.695>
8. Yamada, Y., Shinkawa, K., Nemoto, M., Nemoto, K., & Arai, T. (2023). A mobile application using automatic speech analysis for classifying Alzheimer's disease and mild cognitive impairment. *Computer Speech & Language*. <https://doi.org/10.1016/j.csl.2023.101514>
9. 北岡教英 (2023): 超高齢者音声コーパス (EARS). 国立情報学研究所 音声資源コンソーシアム. (データセット). <https://doi.org/10.32130/src.EARS> Meiko Fukuda, Ryota Nishimura, Hiromitsu Nishizaki, Koharu Horii, Yurie Iribe, Kazumasa Yamamoto, Norihide Kitaoka (2023): A new speech corpus of super-elderly Japanese for acoustic modeling. *Computer Speech and Language*, 77: 101424. <https://doi.org/10.1016/j.csl.2022.101424>
10. 鹿野清宏 (2007): 新聞記事読み上げ高齢者音声コーパス (S-JNAS). 国立情報学研究所 音声資源コンソーシアム. (データセット). <https://doi.org/10.32130/src.S-JNAS>
11. Hugging Face, rinna/japanese-hubert-base, 入手先 <https://huggingface.co/rinna/japanese-hubert-base>
12. Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. 2019. Optuna: A Next-generation Hyperparameter Optimization Framework. In KDD.