

機械学習モデルによる軽度認知障害と健常の自動判別手法の開発

Development of an automatic classification method between mild cognitive impairment and normal by a machine learning model

安田朝子[†]、浅野敬幸[†]、木之下節夫^{†,††}、田中俊郎[†]、佐原徹^{†††}、田中淑満^{†††}、
本間昭^{††††}、繁田雅弘^{††}

Asako Yasuda, Takayuki Asano, Setsuo Kinoshita, Toshiro Tanaka, Toru Sahara, Toshimitsu
Tanaka, Akira Homma, Masahiro Shigeta

[†]日本テクトシステムズ株式会社, ^{††}東京慈恵会医科大学精神医学講座,

^{†††}エヌ・ティ・ティ・コミュニケーションズ株式会社, ^{††††}お多福もの忘れクリニック

Nipontect Systems Corporation, NTT Communications Corporation,

Otafuku Memory Clinic, The Jikei University School of Medicine

yasuda@nipontect.co.jp

概要

我々は、AIを活用し、誰もがMCIにおける僅かな認知機能の低下を発話のみ且つ短時間でセルフチェックできるツールの社会実装を目指している。本研究では、被検者の年齢と性別、設問に対する被検者の回答の採点結果とその音声特徴量を用いて、健常とMCIを判別する機械学習モデルの構築とその性能評価を行った。

キーワード：機械学習、音声特徴量、セルフチェック、軽度認知障害 (MCI)

1. はじめに

我が国では、2025年には65歳以上の人口が約3600万人に達し、そのうち認知症の人の数は700万人に達するとの推計がなされている。認知症は、65歳以上の人の約5人に1人が罹患する、ますます身近な疾患となった。現在、厚生労働省が主導する「認知症施策推進総合戦略(新オレンジプラン)」のもと、認知症の早期発見・早期治療が推進されている。また、軽度認知障害(Mild Cognitive Impairment: MCI)は、何らかの認知機能障害を呈するが日常生活の支障は来していない、正常加齢と認知症の境界領域に該当する概念である。MCIはアルツハイマー病などの認知症の前駆状態としてとらえられることが多く、認知症における早期診断・早期治療の重要性という観点から注目されている。

神経心理検査は、適時に実施することによって認知機能低下の兆候を早期に発見し、より良いセルフケアと将来への備えを可能とする。また、早期発見・早期介入によって認知機能の低下を遅らせ、認知症発症のリスクを低減できることが報告されている[1, 2]。

近年、デジタル技術と機械学習技術の進歩を受け、デジタル機器を用いてより簡便に認知症を検出するツールの開発が行われている。これらのツールは一般的に既存の認知機能検査に基づいて開発されているが、そ

の多くが、検査者を必要としたり、被検者個人に煩雑な操作を求めたりするため、導入障壁が高いという問題がある。

以上を受けて我々は、時間見当識(今日の日付)の質問に対する被検者の回答、音声特徴量、および被検者の年齢から、認知機能に一定以上の変化があるかどうかを20秒程度で判別する[3]ソフトウェア(ONSEI)を開発した。これまでに、90名を対象とした臨床的妥当性研究において、臨床的認知症尺度(Clinical Dementia Rating: CDR)による分類に対するONSEIの判別精度を評価し、CDR1群(軽度認知症)の正分類率が93%、CDR0群(健常)の正分類率が92%であることを明らかにした。これにより、ONSEIは軽度認知症と健常を高精度に判別できることが示唆された。一方、CDR0.5群(認知症疑いまたはMCI相当)の正分類率は69%であり、ONSEIは、MCI段階におけるより早期の認知機能低下の検出には十分な精度がないと考えられた。

そこで、本研究では、検査者を必要とせず、誰もが簡単に、MCI段階における認知機能低下を発話のみで検知できる新たな認知機能チェックツールを開発するため、健常とMCIを判別する機械学習モデルの構築を目指した。設問の設計にあたっては、音声テキスト変換技術を用いて、被検者の回答音声からより正確に自動採点できるものとし、健常とMCIの判別に適した設問を過去の文献情報や我々の臨床研究の結果等に基づいて選別、適宜変更を行った。その上でアプリケーションを開発し、設問に対する被検者の回答の自動採点結果とその音声特徴量を収集し、健常とMCIを判別する機械学習モデルの構築とその性能評価を行った。

2. 方法

2-1. 参加者

港区シルバー人材センターと三鷹市シルバー人材センターを通じた公募にて参加した50歳以上の人を対象とした。本研究は浅井皮膚科治験審査委員会において承認を得た後に、ヘルシンキ宣言に従い、参加者に対して研究内容、方法などについて十分な説明を行い、文書による同意を得て実施した。

2-2. データ取得

参加者の年齢や性別（背景情報）は本人から入手した。臨床的情報はMini Mental State Examination-Japanese (MMSE-J) 精神状態短時間検査 改訂日本版[4]とCDR[5]により入手した。CDRは、訓練を受けた臨床心理士等が構造化面接を通じて見当識、記憶、判断力と問題解決、地域社会、家庭生活および趣味・関心、介護状況の6領域について評価し、判定を行った。参加者は日常生活において自立しており、シルバー人材センターを通じた公募により単身で研究に参加している状況を踏まえ、本人との面接により得られた情報のみに基づいて評価した。CDRの判定結果については、その定義及びDSM-5によるMCIの診断基準に準じ、CDR0=健常、CDR0.5=MCI相当とした。

機械学習モデルの構築に際して使用する設問としては、時間見当識、文章の記憶課題（直後再生および遅延再生）、スパン復唱課題を採用した。時間見当識課題では、今日の西暦年月日と曜日を答えさせた。文章の記憶課題では、9つの文節からなる買物に関する文章（記銘材料10個。文の構成は一定で、記銘材料となる内容はあらかじめ用意されたリストの中からランダム提示）の直後再生を3回行って記銘させ、スパン復唱課題実施後（記銘からおよそ3分後）に遅延再生を行った。スパン復唱課題では、3個～7個の2桁のランダムな数字を順に提示し、最後の3つの数字を回答させた内容をスコア化した。以上の設問から構成される検査を音声のみで実施するアプリケーションを作成し、設問に対する参加者の回答の自動採点結果とその音声特徴量を収集した。音声特徴量は、音響特徴量である基本周波数

(F0)と韻律特徴量である沈黙率を用いた。沈黙率は、発話開始から発話終了までの間に発話していない時間の割合とした。

2-3. データ分析および機械学習モデルの構築

参加者の背景情報、臨床的情報および検査で取得したデータの平均値や標準偏差を健常とMCI相当の群ごとに算出した。健常とMCI相当を判別する機械学習モデルは、参加者の年齢と性別、設問に対する回答の採点結果とその音声特徴量を説明変数として、サポートベクターマシン (SVM)、Light Gradient Boosting Machine (LightGBM)、Gaussian ナイーブベイズ (GaussianNB)、ロジスティック回帰または多層パーセプトロン (MLP) を分類器に用いて構築した。機械学習モデルの性能はLeave-One-Out Cross-Validationにより評価した。評価指標はAccuracy、AUC、感度、特異度とした。なお、データ分析および機械学習モデルの構築はpython (ver 3.10) 環境で行った。

3. 結果

総計344名が研究に参加した。そのうち、欠損のあるデータセットを除き、333名分のデータセットを解析に用いた。CDRによる判定の結果、333名のうち、193名(58.0%)が健常、140名(42.0%)がMCI相当と分類された。各群の人数、年齢、教育年数およびMMSE得点の平均値を表1に示す。

小海らや杉下ら[4,6]によれば、MMSE得点が27点と28点は健常とMCIが混在する得点層と考えられる。本研究の参加者のうち27点と28点の人は全体の35.0%を占めていたため(図1)、本研究の参加者には健常とMCIの境界域にいる人が多く含まれている可能性が考えられた。機械学習モデルの性能を正確に評価するためには、これらの境界域にいる人を除外し、確信度の高い健常と確信度の高いMCI相当の人の判別結果によって行う必要があると考え、CDRの判定結果とMMSE得点を用いて、CDR0かつMMSE29点以上の者を「確信度の高い健常」、CDR0かつMMSE28点以下の者を「確信度の低い健常」、CDR0.5かつMMSE27点以上の者を

表1 研究参加者の基本属性

	確信度の高い健常	確信度の低い健常	確信度の低いMCI相当	確信度の高いMCI相当
n (男性/女性)	118 (65/53)	75 (36/39)	98 (50/48)	42 (23/19)
年齢 (歳)	69.57 ± 5.37	72.2 ± 6.04	72.78 ± 5.48	75.55 ± 4.73
教育年数	14.86 ± 1.83	14.0 ± 2.0	14.5 ± 2.11	14.12 ± 2.47
MMSE得点	29.47 ± 0.5	27.25 ± 1.09	28.42 ± 1.01	24.64 ± 1.65

年齢、教育年数、MMSE得点は平均±標準偏差で示す。

「確信度の低いMCI相当」、CDR0.5 かつ MMSE26 点以下の者を「確信度の高いMCI相当」に分類した(表1)。

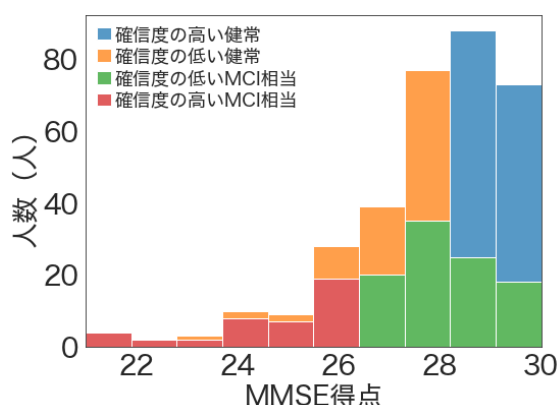


図1 MMSE 得点の分布

以上を踏まえ、参加者の年齢と性別、設問に対する回答の採点結果とその音声特徴量を説明変数とし、SVM、LightGBM、GaussianNB、ロジスティック回帰またはMLPを分類器として構築した機械学習モデルの性能を比較した結果、GaussianNBを用いて構築した機械学習モデルの判別精度が最も高く(データ示さず)、確信度の高い健常と確信度の高いMCI相当の判別精度はaccuracy 0.788、AUC 0.868、感度 0.81、特異度 0.78であった(図2)。なお、すべてのデータセットを用いて健常とMCI相当を判別した場合、精度はaccuracy 0.637、AUC 0.657、感度 0.586、特異度 0.674であった。

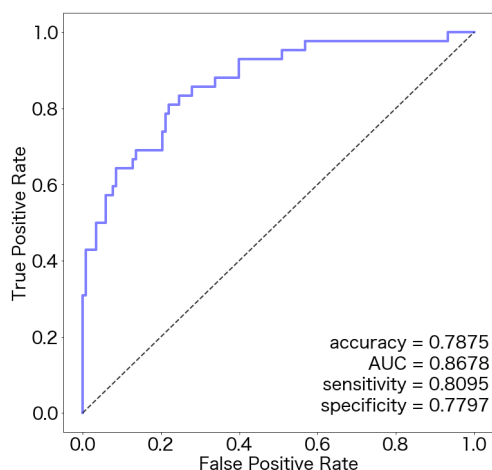


図2 機械学習モデルのROC (receiver operating characteristic) 曲線

4. 考察

我々はAIを活用し、誰もがMCIにおける僅かな認知機能の低下を発話のみ且つ短時間でセルフチェックできるツールの社会実装を目指している。本研究ではシルバー人材センターを通じて自立した生活ができていない人に対して音声による検査を実施し、収集した性別、年齢、質問に対する回答の採点結果とその音声特徴量を用いて機械学習モデルを構築し、その性能を評価した。

本研究の参加者のうちMCI相当と判定されたのは42.0%であった。我が国のMCIの有病率は10-20%程度[7,8]であることから、本研究でMCI相当と判定された人の中には機能障害が比較的軽い人が多く含まれていた可能性がある。その要因としては、コロナ禍によるコミュニケーション量の低下の影響が考えられた。

我々は音声による検査の設問を設計するにあたり、発話のみで実施可能な課題とした。すなわち、記述や描画を必要とする課題など、画面入力を行うものは取り扱わないこととした。短時間で健常とMCIの判別を試みる場合、記憶課題の中でも特に遅延再生は両者の識別に優れるという知見[9]などから、難易度を高めに設定し、単語セットや文章を記録させて数分後に再生させる課題が行われることが多い。一方、最近の研究ではMCIにおけるワーキングメモリと実行機能の低下が注目されている[10]。以上を参照し、本検査では、最終的に時間見当識、文章の記憶課題(直後再生および遅延再生)、スパン復唱課題を選択し、各課題に即した設問を設計した。時間見当識課題は時間見当識の能力を、文章の記憶課題・直後再生は即時記憶の能力を、文章の記憶課題・遅延再生は近時記憶の能力を、スパン復唱課題は作動記憶の更新、単純な注意の範囲や強度等の能力をそれぞれ評価する。

MMSE等の既存のスクリーニング検査は内容が概ね固定されており、繰り返し実施すると学習効果が生じることが報告されている[11]。また、MMSEは認知症スクリーニング検査であるため、健常群の多くが満点となること(天井効果)が懸念される[12]。そこで、我々は、学習効果を抑制する仕組みとして、文章の記憶課題やスパン復唱課題で毎回異なる課題内容を提示するように設計した。具体的には、記憶課題で出題する文章は複数の単語から構成されるが、そのパターンは16億種以上あるため、ほぼ同じ課題を出題しない仕組みとなっている。スパン復唱課題についても、3個~7個の2

桁の数字を無作為に提示するため、ほぼ同じ課題を出題しない仕組みとなっている。また、天井効果や床効果を抑制するため、課題の難易度をMMSEより高めに設定した。その結果、健常群において、文章の記憶課題やスパン復唱課題の点数は満点付近に偏っておらず、天井効果は認められなかった(データ示さず)。また、MCI相当群の点数は0点付近に集中せず、床効果は認められなかった(データ示さず)。これらの結果から、本研究で採用した設問の難易度は適切に設定できていると考えられた。

本検査は検査者が不要で、利用者のみで実施することができる。検査者を要する従来の認知機能検査では検査者の技能や態度により評価結果が揺らぐ可能性があるが、本検査はこの懸念を回避することができる。また、検査開始から終了までの所要時間は約6分であり、CDRやMoCA-Jよりも短時間で実施することができる。以上から、本検査は日常的な場で使用可能な、健常とMCI相当の判別に有用なツールになりうると考える。

5. まとめと今後の展望

本研究では参加者の年齢、性別、設問に対する発話内容と音声特徴量を用いて健常とMCI相当を判別する機械学習モデルを構築し、その性能評価を行った。なお、本研究においては、本人との構造化面接により得られた情報のみに基づいて参加者を健常あるいはMCI相当に分類しているため、診断上のMCIとは異なる点に留意する必要がある。

本研究で構築した機械学習モデルは、発話内容と音声特徴量に基づいて健常とMCI相当を判別するため、電話サービスへの搭載が可能である。現在、社会実装に向けて機械学習モデルの判別精度の向上や検査時間の短縮化を試みている。本大会ではこれらの検討結果を含めて報告する予定である。

文献

- [1] Roberts R, Knopman DS. (2013) Classification and epidemiology of MCI, *Clin Geriatr Med*, Vol. 29, No. 4, pp. 753-772.
- [2] Dubois B, Padovani A, Scheltens P, Rossi A, Dell'Agnello G. (2016) Timely diagnosis for Alzheimer's disease: a literature review on benefits and challenges, *J Alzheimers Dis*, Vol. 49, No.3, pp. 617-631.
- [3] Maikusa N, Sonobe R, Kinoshita S, Kawada N, Yagishi S, Masuoka T, Kinoshita T, Takamichi S, Homma A. (2019) Automatic detection of Alzheimer's dementia using speech features of Hasegawa's Dementia Scale-Revised, *Geriatr Med*, Vol. 57, No. 2, pp. 1117-1125.

- [4] 杉下 守弘, 腰塚 洋介, 須藤 慎治, 杉下 和行, 逸見 功, 唐澤 秀治, 猪原 匡史, 朝田 隆, 美原 盤 (2018) MMSE-J (精神状態短時間検査-日本版) 原法の妥当性と信頼性, *認知神経科学*, Vol. 20, No. 2, pp. 91-110.
- [5] 目黒謙一 (2008) 認知症早期発見のための CDR 判定ハンドブック, 医学書院
- [6] 小海宏之, 前田明子, 山本愛, 加藤佑佳, 岡村香織, 園田薫, 安藤悦子, 岸川雄介 (2010) 日本語版 MMSE の検出力と特異性について, 花園大学社会福祉学部研究紀要, Vol. 18, pp. 91-95.
- [7] 朝田隆 (2012) 有病率: どこまで増える認知症, *臨床神経学*, Vol. 52, No. 11, pp.962-964.
- [8] 朝田隆 (2013) 都市部における認知症有病率と認知症の生活機能障害への対応, 厚生労働科学研究費補助金 疾病・障害対策研究分野 認知症対策総合研究報告書
- [9] 河野 直子, 梅垣 宏行, 鈴木 裕介, 山本 さやか, 茂木七香, 井口 昭久 (2007) ADAS 単語カードを用いた遅延再生課題の軽度認知機能低下者識別に対する有用性: 外来もの忘れ検査利用者を対象とした検討, *日本老年医学会雑誌*, Vol. 44, No. 4, pp. 490-496.
- [10] Kirova AM, Bays RB, Lagalwar S (2015) Working memory and executive function decline across normal aging, mild cognitive impairment, and Alzheimer's disease, *Biomed Res Int*, 748212.
- [11] Gavett BE, Gurnani AS, Saurman JL, Chapman KR, Steinberg EG, Martin B, Chaisson CE, Mez J, Tripodis Y, Stern RA (2016) Practice Effects on Story Memory and List Learning Tests in the Neuropsychological Assessment of Older Adults, *PLoS One*, Vol. 11, No.10, e0164492.
- [12] Spencer RJ, Wendell CR, Giggey PP, Katzel LI, Lefkowitz DM, Siegel EL, Waldstein SR. (2013) Psychometric limitations of the mini-mental state examination among nondemented older adults: an evaluation of neurocognitive and magnetic resonance imaging correlates, *Exp Aging Res*, Vol. 39, No.4 pp. 382-397.