

せん妄の発声特徴量を用いた予測に関する文献調査報告書

1. はじめに

せん妄の概要と早期診断の重要性

せん妄は、集中治療室(ICU)の患者や高齢者に特に多く見られる、急性の神経精神医学的症候群であり、その発生率はICU環境では最大87%に達することもあります¹。この状態は、注意、意識、認知の急激な変化を特徴とし、しばしば変動する経過をたどります¹。患者は精神的な混乱、注意力の持続困難、見当識障害、さらには幻覚や著しい興奮を示すことがあります⁹。

せん妄の発生は、入院期間の延長、死亡率の増加、そしてその後のアルツハイマー病(AD)やAD関連認知症(AD/ADRD)の発症リスクの著しい上昇といった深刻な合併症と関連しています¹。これらの有害な転帰を考慮すると、せん妄の早期発見と積極的な管理は、その影響を軽減し、患者の全体的な予後を改善するために極めて重要です¹。

しかし、現在のせん妄の検出方法は、主に医師による主観的な臨床評価に依存しています。このアプローチは、特に経験の浅い臨床医にとっては大きな課題を提示し、病院環境におけるせん妄症例の75%以上が見過ごされているという憂慮すべき統計につながっています²。このような診断上のギャップは、臨床現場における満たされていない重要なニーズを浮き彫りにしています。この診断上の課題は、客観的でスケーラブルな非侵襲的ツールの開発を強く求めています。せん妄の診断基準の一つである「言語障害」を直接活用し、認知症などの関連する神経認知障害における音声バイオマーカーの成功事例を基盤とすることで、自動音声分析は極めて有望な解決策として浮上しています。その広範な採用は、せん妄を「隠れた流行病」から、より一貫して正確に特定できる状態へと変革し、タイムリーな介入を可能にし、患者の転帰を潜在的に改善する可能性があります。

発声特徴量を用いた予測研究の背景と意義

興味深いことに、「言語の障害」はせん妄の主要な診断基準の一つとして明確に認識されています¹¹。しかし、この直接的な関連性にもかかわらず、音声はこれまで、動的な臨床環境での標準化された信頼性の高い評価を達成することの難しさから、診断または予後ツールとして十分に探求されてきませんでした¹¹。

自動音声分析の登場は、この状況を変革する解決策を提供します。これは、音響的およびテキストの音声特徴量を客観的に定量化するための正確で非侵襲的かつ効率的な方法です¹¹。これらの計算音声特徴量は客観的なマーカーであり、すでにさまざまな形態の認知症など、他の神経認知疾患の診断と予測において成功を収めています³。

人工知能(AI)技術、特に機械学習(ML)と深層学習(DL)を適用してせん妄患者の大規模データを分析することで、微妙ながらも関連性の高いせん妄マーカーの特定と定量化が可能になります²。AIと音声分析のこの統合は、診断実践に革命をもたらし、喉頭系に影響する疾患から精神のおよび神経変性疾患まで、幅広い疾患の診断精度を大幅に向上させる計り知れない可能性を秘めています¹⁶。

この技術の進歩は、せん妄管理のパラダイムにおける重要な変化を示唆しています。目標は、単にせん妄が発症するかどうかを予測するだけでなく、患者のリスクと状態を継続的にリアルタイムで評価することにあります。この能力により、微妙な音声パターンの変化が、本格的な臨床症状が明らかになる前、または人間の観察によって認識される前に現れた場合でも、ケア計画の動的な調整と、より早期で的確な介入が可能になります。この継続的なモニタリング能力は、変動する経過を特徴とするせん妄のような状態にとって特に価値があります。

2. せん妄における発話・言語障害の特性

せん妄の診断基準と発話関連症状

せん妄は、注意、意識、認知機能の障害によって根本的に特徴付けられます²。中核的な特徴には、精神状態の障害、不注意、思考のまとまりのなさ、意識の変容が含まれます²。

特に重要なのは、「言語の障害」がせん妄の主要な診断基準として明確に挙げられている点です¹¹。この直接的な関連性は、音声をバイオマーカーとして調査する臨床的根拠を提供します。臨床現場からの定性的な観察では、これらの障害の性質がさらに詳しく記述されています。せん妄患者は、「会話上の困難感・違和感」、「声掛けに対する反応のワテンポの遅れ」、「ゆっくりとした反応」、「コミュニケーションのつじつまが合わない」、「言っていることと違う返答」、そして全体的な「会話が成立しない」といった症状を示すことがあります¹⁹。

せん妄の臨床的表現は多様であり、過活動型（落ち着きのなさ、興奮、幻覚、妄想を特徴とする）、低活動型（異常な眠気、無気力、ぼんやりした外見を特徴とする）、または混合型（過活動型と低活動型の症状が交互に現れる）の異なるサブタイプで現れます⁹。これらの異なる行動的表現は、特定の識別可能な音声パターンと関連している可能性が高いと考えられます。

せん妄は単一の疾患ではなく、過活動型、低活動型、混合型という異なる形態で現れることが示されています⁹。¹⁹で述べられている「反応がゆっくり」や「コミュニケーションのつじつまが合わない」といった症状は、低活動型せん妄とより一致する可能性があり、一方で「興奮」は過活動型せん妄におけるより速く、まとまりのない発話と相関する可能性があります。²⁰で言及されている、痛みやうつ病が音声のテンポや強度に影響を与えるという原則は、せん妄のサブタイプによって異なる内部状態が、異なる音声パターンとして現れることを示唆しています。このことは、単一の普遍的な音声特徴量セットや単一の予測モデルが、すべてのせん妄サブタイプに対して最適に有効であるとは限らない可能性を示唆しています。むしろ、将来の研究では、これらの異なる表現型に対応する特定の音声バイオマーカープロファイルを探索する必要があります。サブタイプ固有の音声シグネチャを特定することで、より繊細で正確な診断モデルにつながり、患者が経験しているせん妄の特定のタイプに基づいて、個別化された介入が可能になる可能性があります。これは、せん妄の多様な臨床像を完全に捉える音声データ収集の重要性も強調しています。

せん妄患者にみられる具体的な発声特徴量

自動音声分析は、幅広い音響的およびテキスト的音声特徴量を客観的に定量化する手段を提供します¹¹。せん妄（および関連する神経認知障害）の調査対象となる特徴には、以下のものが含まれます。

- テンポ（発話速度）：発話の速度を指します¹¹。²⁰では痛みやうつ病（テンポが急速に増加したり減少したりする）について議論していますが、テンポが内部状態の敏感なバイオマーカーであることを確立しています。せん妄では、覚醒度の変動が不規則なまたは遅い発話テンポとして現れる可能性があります。
- プロソディ（イントネーションとリズム）：イントネーションパターン、強調、基本周波数（F0）

のフロア、F0の変動と範囲、文の輪郭など、発話の旋律的およびリズムミカルな側面を網羅します¹¹。統合失調症のような状態では、非定型的なプロソディが単調さにつながるものが観察されており²¹、これはせん妄と一部の認知・精神医学的症状の重複があります。発話におけるスペクトル強調と強度の変動の減少は、コミュニケーションへの関与の欠如を示唆する可能性があります²¹。せん妄における思考のまとまりのなさは、「不規則な発話リズムと異常なイントネーションパターン」をもたらすこともあります¹⁶。

- 構成(発話の流れと構造): 発話がどのように構成されているか、発話境界や情報フローを含みます¹¹。認知的な構成に影響を与える状態(統合失調症など²¹)の患者は、より短く、構造が乏しい情報のかたまりで構成された発話を示し、しばしば沈黙や長い休止が散見されます。
- 語彙的特徴(単語の選択と多様性): 使用される特定の単語、語彙の多様性(語彙の豊富さ)、および単語使用の特性(例: 代名詞や副詞の使用の増加、心理的完了状態を反映する単語の減少)が含まれます³。これらの特徴は、AD/ADRDのリスクが高いことと関連しており¹⁴、せん妄の認知への影響を考えると関連性があります。
- 非流暢性(発話の中断): 発話の流れの中断を指し、言い直し現象、ためらい、沈黙、休止などが含まれます¹¹。非流暢性の増加は、AD/ADRDのリスクが高いことと関連する言語的特徴であり¹⁴、せん妄における認知的なまとまりのなさのマーカーとなる可能性があります。
- 音響的特徴(声質): F0以外に、強度(声の大きさ)、ハーモニック・トゥ・ノイズ比(声の明瞭度を反映)、高周波エネルギーなどの側面が含まれます³。声の大きさは、肺から供給される声門下圧と声帯振動の振幅によって決まります¹⁶。声帯に影響を与える病理学的状態は、声のかすれや息切れ声につながる可能性があります¹⁶、せん妄が発話の運動制御に影響を与える場合に重要となる可能性があります。

これらの特徴は通常、患者が特定の言語タスクを実行中に音声録音されたサンプルを逐語的に転写し、その後自動でコンピュータ処理を行うことで抽出されます¹¹。このプロセスでは、音響的特徴の専用分析ツールに加えて、言語的特徴のために自然言語処理(NLP)や音声認識が活用されることがよくあります³。

研究文献は、音声バイオマーカーが認知症³や統合失調症²¹といった他の神経認知・精神疾患に成功裏に適用されていることを繰り返し示しています。さらに、せん妄自体がAD/ADRDと強く関連しており、認知機能低下を加速させる可能性があります³。分析されている特定の音声特徴量(例: テンポ、プロソディ、非流暢性、語彙的特徴)は、これらの神経認知・精神疾患全体で共通しています³。

この重複は、機会と課題の両方をもたらします。一方で、認知症や統合失調症における音声バイオマーカー研究から得られた既存の方法論、特徴抽出技術、および知見をせん妄に応用し、活用できるという機会があります。他方で、診断の特異性に関して重大な課題を提起します。モデルは、せん妄に特異的な音声変化と、基礎疾患または併存疾患(例: 既存の認知症、うつ病、精神疾患)によって引き起こされる変化をどのように確実に区別できるのでしょうか。こ

のことは、音声特徴量だけでは複雑な臨床状況において十分な特異性を提供できない可能性があるため、音声特徴量に加えて包括的な臨床的背景(例:急性発症と慢性進行、病歴、投薬状況)を統合することの極めて重要な必要性を強調しています。

Table 1: せん妄に関連する主要な発声特徴量

特徴カテゴリ	特定の特徴	せん妄との関連性/説明	関連文献ID
音響的特徴	ピッチ(基本周波数 F0)	覚醒度の変化を示す変動や範囲の減少、単調さ ²⁰	20
	強度(声の大きさ)	覚醒度の変化やコミュニケーションへの関与の欠如を示す変動 ²⁰	20
	ハーモニック・トゥ・ノイズ比	声の明瞭度の低下、声帯振動の障害の可能性 ¹⁶	16
	高周波エネルギー	声質の変化、コミュニケーションへの関与の欠如 ²⁰	20
プロソディ的特徴	発話速度(テンポ)	認知機能の低下や思考のまとまりのなさを反映する遅延または不規則なテンポ ¹¹	11
	プロソディの変動性	イントネーションパターンの平坦化、単調さ、思考のまとまりのなさ ¹⁶	16
	文の輪郭	異常なイントネーションパターン、認知機能障害 ¹⁶	16
言語的特徴	語彙の多様性	語彙の減少、認知機能障害の指標 ³	3
	代名詞・副詞の使用	特定の単語選択パターン	14

		ンの変化、AD/ADRDリス スクとの関連 ¹⁴	
	思考の分析性	思考のまとまりのなさ、 認知機能の低下 ¹⁴	14
非流暢性特徴	ポーズ・沈黙	発話の流れの中断、思 考プロセスの困難さ ¹⁴	14
	言い直し現象	発話の修正、認知的な 負荷の増加 ¹⁴	14
	発話のまとまりのなさ	会話のつじつまが合わ ない、コミュニケーション の困難さ ¹⁹	19

3. 発声特徴量を用いたせん妄予測モデルの現状

機械学習・深層学習技術の応用

人工知能(AI)技術、特に機械学習(ML)と深層学習(DL)の応用は、せん妄の予測、診断、および管理のための大規模な患者データ分析において、ますます重要性を増しています²。MLモデルは、人間の観察では知覚できない複雑なデータ内の微妙なパターンを識別する独自の能力を備えており、これにより幅広い疾患においてより早期かつ正確な診断を可能にします¹⁶。DLはMLの専門的なサブセットであり、複数の隠れ層と畳み込みなどの操作を持つニューラルネットワークを使用することで、生の発話波形や抽出された特徴量などの高次元かつ非構造化データをより効果的に処理するのに特に適しています¹⁵。

せん妄予測研究では、様々なMLアルゴリズムが採用されています。これには、ロジスティック回帰(LR)、サポートベクターマシン(SVM)、K近傍法(KNN)、ニューラルネットワーク、ランダムフォレスト(RF)、アダプティブブースティング(AdaBoost)、勾配ブースティングマシン(GBM)、Extra-Trees Classifier、Light Gradient Boosting Machine(LightGBM)が含まれます⁵。これらの複雑なモデルの性能を最適化するために、GridSearchCVなどの手法を用いた効果

的なハイパーパラメータチューニングが重要なステップとして一貫して強調されています⁵。

主要なモデルアーキテクチャと特徴量抽出

自動音声分析パイプライン: 一般的なワークフローには、音声録音された発話サンプルの収集、逐語的な転写(テキスト分析に適用可能な場合)、および自動コンピュータ処理が含まれます。この処理により、前述の様々な音響的およびテキスト的特徴量が定量化されます¹¹。このプロセスでは、音声特徴量のための専用の音響分析ツールに加えて、言語的特徴のために自然言語処理(NLP)および音声認識が活用されることがよくあります³。

統合モデル(マルチモーダルデータ): せん妄予測における重要な傾向は、複数のデータソースの統合です。モデルはしばしば、静的な患者特徴(例: 年齢、APACHE-IIスコア、併存疾患レベル)と、動的な時系列生理学的データ(例: 心拍数、血圧、呼吸数などのバイタルサイン)を組み合わせて、患者の健康状態の包括的かつ全体的な表現を構築します¹。¹と⁶は音声特徴量を明示的に含んでいませんが、これらはせん妄予測におけるマルチモーダルデータ統合の確立された原則を示しており、他の臨床データと並行して音声特徴量を組み込むための非常に適切なアプローチです。

時系列データのための深層学習アーキテクチャ: 特筆すべき進展は、ICUせん妄予測のために特別に設計された、注意機構を強化した単純化された長・短期記憶(LSTM)ネットワークなどの深層学習モデルの導入です¹。LSTMネットワークは、シーケンシャルな時系列データを処理するのに特に適しており、動的な測定値(例: 患者の生理学的状態の進化や連続的な発話ストリーム)内の時間的依存関係を効果的に捉えることができます。注意機構の統合により、モデルは、せん妄の予測に特に関連性の高いデータ内の特定のタイムステップや変動に異なる重要度を割り当てることができ、これにより解釈可能性と予測精度の両方が向上します¹。発話の場合、これは診断上重要な特定の単語、フレーズ、または非流暢性の瞬間に焦点を当てることを意味する可能性があります。

特徴量重要度分析: 一部の複雑なMLモデルの「ブラックボックス」的性質に対処するため、Shapley additive explanations(SHAP)などの手法が採用されています。これらの方法は、モデルの予測に対する個々の変数の貢献を特定し、定量化することで、モデルの解釈可能性を高めます⁵。この解釈可能性は、臨床医間の信頼を育み、AI駆動型ツールの臨床導入を促進するために極めて重要です²²。

MLおよびDLモデルは、せん妄予測において印象的な精度と可能性を示していますが¹、研究文献は重要な懸念を明確に提起しています。ニューラルネットワークは「不透明」であり、しばしば「ブラックボックス」として扱われることがあります²²。この透明性の欠如は、臨床導入にお

ける障壁と直接的に関連しており、「説明可能性」と「正しい推論連鎖」の必要性が、医師と患者双方の信頼を促進するために不可欠であると強調されています²²。これは、高い予測精度を達成するだけでは、AIツールを臨床実践に成功裏に統合するには不十分であることを示しています。臨床医がこれらのモデル、特に音声特徴量のような複雑で微妙なバイオマーカーを扱う際に、信頼して効果的に利用するためには、特定の予測や診断が

なぜ行われたのかを理解する必要があります。これは、モデルの不透明性が臨床的有用性を直接妨げるという因果関係を示唆しています。したがって、将来の研究では、予測性能と並行して、本質的に解釈可能なモデルまたは堅牢な説明可能性技術の開発を優先し、「説明可能なAI(XAI)」を臨床応用における重要な領域と位置付ける必要があります。

ユーザーの質問は特に音声特徴量に焦点を当てていますが、研究文献の包括的な調査は、せん妄予測のための多くの高性能モデル(音声は明示的に使用されていない場合でも)が複数のデータソースを統合していることを明らかにしています。これには、静的な患者特徴、動的な生理学的データ、および電子医療記録が含まれます¹。IDOSAプロジェクト¹¹では、音声特徴量が人口統計学的データや疾患の重症度単独よりも正確なせん妄状態の分類を可能にすることが示されましたが、これはそれらを組み合わせることによる潜在的な利益を否定するものではありません。せん妄の多因子的な性質(例:²³の環境因子)も、多様な入力の実用性を示唆しています。このことは、音声特徴量が単独のバイオマーカーとして強力である一方で、他の臨床データと統合された場合に、その予測および診断の可能性を最大限に発揮する可能性が高いという強いテーマが浮上していることを示唆しています。バイタルサイン、検査結果、患者人口統計、投薬歴、さらには環境因子と音声特徴量を組み合わせるこのマルチモーダルアプローチは、モデルの堅牢性を大幅に向上させ、偽陽性および偽陰性を減らし、より包括的なリスク評価を提供できます。せん妄は複数の寄与因子を持つ複雑な症候群であるため、これは極めて重要です。このような統合されたアプローチは、患者の状態をより全体的に理解し、より正確な予測につながり、最終的に臨床意思決定と患者転帰を向上させる可能性があります。

研究事例の紹介

IDOSAプロジェクト(自動音声分析を用いた入院高齢者のせん妄診断バイオマーカーの特定と認知転帰の予測):これは、ユーザーの質問に直接関連する重要なプロジェクトです。

- 目的: 高齢者(75歳以上)におけるせん妄の診断と、退院後の認知機能低下の探索的予測のためのデジタル音声バイオマーカーを開発することを目的としており、最終的には持続的な認知機能障害およびAD/ADRDに対する脆弱性を検出することを目指しています¹¹。

- 方法論: 研究デザインには、既存の認知機能障害を持つかなりの割合の患者を含む210人の入院高齢者を募集することが含まれています。入院中の2つの異なる時点で音声録音された発話サンプルが収集され、参加者は一連の言語タスクを完了します。これらの発話サンプルは、逐語的な転写と自動コンピュータ処理を受け、テンポ、プロソディ、構成、語彙的特徴、および非流暢性に関連する音声特徴量が定量化されます。機械学習手法が分析の中心となります¹¹。
- パイロット研究の結果: 予備的なパイロット研究では、計算音声特徴量が病院環境で実現可能に収集できること、そして、人口統計学的データや疾患の重症度単独よりも高齢者のせん妄状態のより正確な分類を可能にすることが示されました¹¹。ただし、このプロジェクトはさらなる検証が必要であることを認識しています。
- 期待される成果: プロジェクトの成功は、入院中のせん妄診断の適時性と精度を大幅に向上させ、AD/ADRDおよび持続的な認知機能低下に対する脆弱性のある患者を特定することが期待されています¹¹。

一般的なせん妄予測研究(ML中心、音声に特化しないが方法論に関連):

- 様々な機械学習モデルを、バイタルサイン、血液データ、電子医療記録などの多様な臨床データで訓練し、ICUせん妄を予測する複数の研究が存在します⁵。
- 熱傷患者に焦点を当てたある研究では、バイタルサインや血液検査結果を含む70の変数を用いて、様々なMLモデルで高いAUC値が報告されており、ロジスティック回帰は平均AUC 0.906を達成しました⁸。
- 別の研究では、連続的な生理学的データを用いてランダムフォレスト(RF)モデルを開発し、内部検証でAUROC 0.82、時間的検証で0.73、外部検証で0.84という堅牢な性能を示しました⁷。この研究はまた、信頼性の高いCAM-ICU評価結果を示すCohen κ係数0.81を報告しています⁷。
- 意思決定木分析を用いた興味深い例外的な研究では、日中の離床時間と在床時間、日中の照度、夜間の騒音といった環境因子がせん妄発症の有意な予測因子として特定され、精度0.81、AUC 0.84を達成しました²³。これは音声特徴量に直接関連するものではありませんが、せん妄の多因子的な性質と、予測における多様なデータソースの可能性を浮き彫りにしています。

4. 予測モデルの評価指標と性能分析

主要な評価指標(AUC、精度、感度、特異度など)の説明

予測モデルの性能を評価するためには、いくつかの主要な指標が用いられます。

- **精度 (Accuracy):** モデルが作成した正しい予測 (真陽性と真陰性の両方) の全体的な割合を表します。
- **受信者操作特性曲線下面積 (ROC-AUC):** この指標は、すべての可能な分類閾値において、モデルが陽性 (せん妄状態) と陰性 (非せん妄状態) のケースを区別する能力を定量化します。0 から 1 までの高い AUC 値は、優れた識別能力を示します¹。
- **適合率-再現率曲線下面積 (PR-AUC):** 不均衡なデータセット (せん妄のように一方のクラスが他方よりもはるかに稀な場合) において特に価値があり、PR-AUC は適合率 (すべての陽性予測のうち正しく識別された陽性予測の割合) と再現率 (すべての実際の陽性ケースのうち正しく識別された陽性予測の割合) のトレードオフを評価します¹。
- **適合率 (Precision)、再現率 (Recall)、F1スコア:** これらは包括的な分類レポートの構成要素であり、単純な精度を超えたモデル性能のより詳細な評価を提供します。適合率は陽性予測の正確さに焦点を当て、再現率 (または感度) はモデルがすべての陽性ケースを検出する能力に焦点を当て、F1スコアは適合率と再現率の調和平均であり、バランスの取れた尺度を提供します¹。
- **感度 (Sensitivity):** 真陽性率として定義され、モデルによって正しく識別された実際の陽性ケース (せん妄患者) の割合を測定します¹⁴。スクリーニングツールにおいては、真のケースを見逃すリスクを最小限に抑えるために、高い感度がしばしば重要となります。
- **特異度 (Specificity):** 真陰性率として定義され、モデルによって正しく識別された実際の陰性ケース (せん妄ではない患者) の割合を測定します¹⁴。偽陽性や不必要なフォローアップ介入を避けるために、高い特異度が重要です。
- **Cohen の κ 係数 (Cohen's Kappa):** この統計的尺度は、2 人の評価者または方法間の偶然による期待を超える一致度を評価します。せん妄予測の文脈では、せん妄判定記録の信頼性 (例: 看護師と研究者の間) を確認するために使用できます⁷。

既存研究におけるモデル性能の比較

主に生理学的/臨床データを使用するモデル (音声特徴量を明示的に使用しないものも含む) :

- 深層学習モデル (注意機構付き LSTM) は、合成データにおいて「ほぼ完璧な性能」を示し、ROC および PR 曲線で AUC 1.00 を達成し、適合率、再現率、F1スコアのすべてで 100% を記録しました¹。ただし、これらの結果が合成データから実際の臨床設定に一般化できるかについては、著者自身が明示的に限界を述べていることに注意が必要です。
- 熱傷患者のせん妄予測において、バイタルサインや血液検査結果などの臨床データを利

用した様々な機械学習モデルを用いた研究では、ロジスティック回帰が平均AUC 0.906を達成するなど、強力な平均AUCが報告されています。その他、サポートベクターマシン(0.897)、K近傍法(0.894)、ニューラルネットワーク(0.857)、ランダムフォレスト(0.850)、アダプティブブースティング(0.832)といったモデルが挙げられます⁸。

- 連続的な生理学的データで訓練されたランダムフォレストモデルは、内部検証でAUROC 0.82、時間的検証で0.73、外部検証で0.84という堅牢な性能を示しました。AUPRC値は、内部検証で0.62、時間的検証で0.85、外部検証で0.77でした⁷。この研究では、信頼性の高いCAM-ICU評価結果を示すCohen κ係数0.81も報告されています⁷。
- 意思決定木モデルは、日中の離床時間と在床時間、日中の照度、夜間の騒音といった環境因子に焦点を当て、せん妄発症予測において精度0.81、AUC 0.84を達成しました²³。

音声特徴量を使用するモデル(主に認知症/ADRD向けだが、せん妄への可能性を示唆):

- 軽度認知障害(MCI)の検出において、音声から導出された語彙的意味特徴量(AUC = 0.80)および音響的特徴量(AUC = 0.77)は、標準的な神経心理学的検査(ボストン命名テスト、AUC = 0.66)と比較して高い診断性能を示しました³。
- アルツハイマー病および関連認知症(AD/ADRD)の分類において、語彙的特徴量に適用されたランダムフォレストモデルは、DementiaBankデータセットで平均感度69.4%、特異度83.3%を達成しました。実際のパイロットデータで評価された場合、その性能は感度70.0%、特異度52.5%でした¹⁴。
- 神経学的、喉頭学的、精神疾患における発話パターンのML応用に関するシステマティックレビューでは、パーキンソン病、うつ病、統合失調症、MCI、アルツハイマー型認知症などの疾患において「高い診断精度」が一般的に結論付けられましたが、研究間のばらつきも認められています¹⁶。

研究文献は、対照的な結果を示しています。¹と⁶は、深層学習モデルが

合成データにおいて「非現実的なほど完璧な」性能(AUC 1.00、適合率/再現率/F1スコア 100%)を達成したと報告していますが、著者自身が「臨床設定の完全な複雑さを捉える上での合成データの限界」について明確に警告しています。これとは対照的に、実際の臨床データを使用した研究⁷では、依然として良好ではあるものの、一貫して低いAUC(例:0.73から0.906の範囲)が報告されています。AD/ADRD予測において、キュレーションされたデータセットでの特異度が83.3%から実際のパイロットデータでの52.5%に低下した例¹⁴は、このギャップをさらに明確に示しています。このパターンは、医療応用におけるAI研究における重大な「合成データトラップ」を示しています。合成データは、初期モデル開発、迅速なプロトタイプニング、理論的可能性のデモンストレーションには非常に価値がありますが、実際のシナリオでの性能を過大評価してしまいます。モデルが合成データから実際の臨床データに移行する際に観察される大幅な性能低下は、実際の患者データに内在する計り知れない複雑さ、変動性、ノイズ、および微妙なニュアンスを浮き彫りにしています。合成データへの依存は、厳密な実世界検証がなければ、期待を過度に膨らませ、研究者や臨床医を誤解させる可能性があります。

す。したがって、合成データのみから導き出された結果の過度な楽観的解釈に対して強く警告し、真の臨床的有用性のためには、多様な実世界臨床データセットに対する大規模で綿密な作業が不可欠であることを強調する必要があります。

研究文献は、感度と特異度を主要な性能指標として導入しています¹⁴。例えば、¹⁴は、パイロットデータにおける音声ベースのAD/ADRD予測の具体的な感度(70.0%)と特異度(52.5%)の値を提供しています。²⁴は、せん妄

スケール(音声バイオマーカーではない)の適応について議論していますが、プールされた感度0.93、特異度0.94というベンチマークを提供しています。これは、異なる臨床ツールが異なる性能プロファイルを持つことを示しています。感度(可能な限り多くの真のケースを特定して、重要な診断を見逃すリスクを最小限に抑える)と特異度(誤報を最小限に抑えて、不必要な介入や不安を避ける)のどちらを優先するかという臨床的決定は、状況に依存します。スクリーニングツールの場合、ケースを見逃すこと(感度が低いこと)は、誤報よりも有害である可能性があります。このことは、せん妄の音声バイオマーカーの臨床応用において、決定的に重要な考慮事項を明らかにしています。感度と特異度の最適なバランスは普遍的ではなく、意図された臨床応用によって異なります。幅広い集団(例:すべての入院高齢者)における早期検出を目的としたスクリーニングツールの場合、偽陽性率が高くても、できるだけ多くのリスクのある患者を特定するために、より高い感度が好まれる可能性があります。逆に、明確な診断ツールや、即座の侵襲的介入を導くツールの場合、誤診や不必要な処置を避けるために、より高い特異度が最も重要となる可能性があります。これは、せん妄の音声バイオマーカーに関する将来の研究が、これらの指標を包括的に報告するだけでなく、特定のユースケース(例:早期警告システム対確定診断)との関連でその臨床的有用性を明確に議論する必要があることを示唆しています。

Table 2: 発声特徴量を用いたせん妄予測研究の概要と性能

研究/プロジェクト名	主要データタイプ	使用された主要な音声特徴量(該当する場合)	機械学習モデル	性能指標(AUC, 精度, 感度, 特異度など)	データ特性	関連文献ID
深層学習モデル (LSTM with Attention)	生理学的時系列データ, 静的特徴	N/A (生理学的データ)	LSTM with Attention	ROC-AUC: 1.00, PR-AUC: 1.00, 適合率: 100%,	合成データ	¹

)				再現率: 100%, F1 スコア: 100%		
熱傷患者の せん妄予測 研究	バイタルサイン, 血液 データ (70 変数)	N/A (臨床 データ)	ロジスティッ ク回帰, SVM, KNN, ニューラル ネットワー ク, RF, AdaBoost	平均AUC: LR 0.906, SVM 0.897, KNN 0.894, NN 0.857, RF 0.850, AdaBoost 0.832	実世界デー タ (82患者)	8
連続生理学的 データを 用いたせん 妄予測	連続生理学的 データ	N/A (生理 学的データ)	ランダム フォレスト (RF)	AUROC: 内 部0.82, 時 間的0.73, 外部0.84; AUPRC: 内 部0.62, 時 間的0.85, 外部0.77; Cohen κ: 0.81	実世界デー タ (開発コ ホート: 5478記録, 外部検証: 670患者)	7
せん妄発症 予測因子研 究	昼間の離床 時間・在床 時間, 昼間 の照度, 夜 間の騒音	N/A (環境 因子)	意思決定木	精度: 0.81, AUC: 0.84	実世界デー タ	23
IDOSAプロ ジェクト	音声特徴量 (テンポ, プ ロソディ, 構 成, 語彙的 特徴, 非流 暢性), 人口 統計, 疾患 重症度	テンポ, プロ ソディ, 構 成, 語彙的 特徴, 非流 暢性	機械学習	パイロット 研究で人口 統計・疾患 重症度単独 より正確な 分類	実世界デー タ (パイロ ット研究, 計 画: 210入 院高齢者)	11
AD/ADRD 音声バイオ マーカー研 究	音響的特徴 量, 語彙的 意味特徴 量, 接続発 話	音響的特徴 量, 語彙的 意味特徴量	機械学習	MCI検出: 語彙的意味 AUC 0.80, 音響AUC 0.77;	Dementia Bank (N=291), 実世界パイ ロットデータ	3

				AD/ADRD 分類 (RF): 感度 69.4% (Dementia Bank), 70.0% (パ イロット); 特 異度 83.3% (Dementia Bank), 52.5% (パ イロット)	(N=22)	
--	--	--	--	--	--------	--

5. 課題と今後の展望

データ収集と標準化の課題

主要な課題の一つは、多様な臨床現場における発話サンプルの標準化された信頼性の高い収集と評価を達成することに内在する困難さです¹¹。この実践的な課題は、これまで音声が発断モダリティとして十分に探求されてこなかった一因となっています。せん妄患者の、大規模で多様かつ綿密に注釈付けされたデータセットの作成が喫緊の課題です。せん妄におけるAIに関する現在のシステムティックレビューは、データの包括性に関する限界をしばしば指摘しており、多くの研究が単一のデータタイプに焦点を当てているか、異なるデータソース間の体系的な分類が欠如しています²。

せん妄の提示の著しい変動性（低活動型から過活動型、混合型まで多岐にわたる）⁹は、異なる症状が異なる誘発方法や捕捉環境を必要とする可能性があるため、データ収集プロトコルの標準化の取り組みをさらに複雑にしています。医療環境で機密性の高い音声データを収集する際には、患者のプライバシーとデータセキュリティに関する倫理的考慮が最重要であり、堅牢なプロトコルと明確な同意プロセスが必要です。

「臨床設定の完全な複雑さを捉える上での合成データの限界」¹が認識されていることは、モデルの適用可能性と堅牢性を確保するために、広範な実世界データ収集と検証が不可欠で

あることを強く示しています。これらの研究文献は、臨床現場におけるMLモデルの成功が、相互に関連するいくつかのデータ関連の課題を克服することにかかっていることを集合的に強調しています。発話評価の標準化の難しさ¹¹と、実世界の複雑さを捉えきれない合成データの固有の限界¹が明確に強調されています。同時に、MLの有効性は、大規模で多様かつ代表的なデータセットの入手可能性に直接結びついています¹⁴。この複雑さに加えて、せん妄自体の変動性や異質性⁹があり、これは「より多くのデータ」だけでは不十分であり、より

質が高く、文脈的に豊かなデータが必要であることを意味します。この複雑な相互作用は、最大の課題であり、同時に最大の機会が、実際の臨床環境でせん妄患者から高品質で多様な音声データを収集するための堅牢で標準化されたプロトコルを確立することにあることを示唆しています。これには、患者の異質性、せん妄の動的な経過、および交絡因子を考慮に入れる必要があります。データ品質や代表性の低さが、アルゴリズムの洗練度に関わらず、モデルの一般化可能性や実世界での性能を直接制限するという因果関係は明確です。したがって、多施設共同研究などを通じて、音声の臨床データ収集基準を確立するための大規模で協調的な努力が、この分野を有望なパイロット研究の段階からさらに進展させるために不可欠です。

モデルの汎用性と臨床応用への障壁

機械学習モデルの有効性は、異なる状態間で大きく異なることが示されており、せん妄に対する一貫した性能を確保するためにはさらなる研究が必要です¹⁶。モデルの汎用性は、MLモデルの臨床適用性にとって依然として大きな課題です。これは、堅牢で普遍的に適用可能な予測モデルを訓練するために必要な、十分に大規模で多様な患者データセットへのアクセスが限られているために特に深刻です¹⁴。

一部の複雑なニューラルネットワークの「ブラックボックス」的性質は、臨床導入における重大な障壁となっており、医師と患者双方の信頼を妨げます。したがって、これらの技術を日常的なケアに統合するためには、説明可能なモデルの要求が極めて重要です²²。臨床実践への成功裏な統合には、シームレスなデータ取得メカニズム、リアルタイム処理能力、医療従事者向けの効率的なアラートシステムなど、様々な物流上およびインフラ上の課題を克服する必要があります。規制当局の承認も、広範な展開に対する重要な障壁となります。せん妄の早期診断は重要ですが、現在、せん妄に対する確立された特定の治療法がないため⁵、予測の

行動可能性を臨床医にとって明確に定義する必要があります。予測は、臨床的に価値があるために、明確で有益な介入経路につながる必要があります。

多角的アプローチと他疾患との関連性

有望な将来の方向性として、音声特徴量を他の日常的に収集される臨床データ(例:バイタルサイン、電子医療記録、検査結果)と組み合わせるマルチモーダルアプローチの採用が挙げられます。このような統合は、より堅牢で包括的かつ正確な予測モデルを生み出すと期待されています¹。この傾向は、より広範なせん妄予測研究においてすでに明らかになっています。

せん妄と他の神経認知障害、特にアルツハイマー病および関連認知症(AD/ADRD)との複雑な関係をさらに調査することは極めて重要です。これは、共通のリスク因子と、せん妄が認知機能低下を加速させる可能性が観察されていることに関連しています³。音声バイオマーカーは、せん妄エピソード後にAD/ADRDまたは持続的な認知機能障害に対する脆弱性のある患者を特定するための貴重なツールとなる可能性があります¹¹。

せん妄の根底にある病態生理学的メカニズム(例:全身性炎症、神経炎症、神経伝達物質異常)³と、音声特徴量に観察される変化を関連付けて探求することは、疾患プロセスへのより深い洞察を提供し、よりの絞ったバイオマーカーの開発に役立つ可能性があります。将来の研究では、せん妄発症に対する環境要因(例:騒音レベル、光曝露)の影響²³も考慮に入れるべきです。これらの要因は、音声パターンに間接的に影響を与えたり、マルチモーダルモデルに統合された場合に補完的な予測シグナルを提供したりする可能性があります。

IDOSAプロジェクトの目標は、単なるせん妄の診断を超えて、「退院後の認知機能低下の予測」を探求し、「AD/ADRDに対する脆弱性」を特定することにまで及んでいます¹¹。この野心は、せん妄とAD/ADRDとの確立された関連性³、および音声バイオマーカーがすでに認知症の進行予測において可能性を示しているという事実によって強化されています³。これは、これらのバイオマーカーの認識されている有用性が、単一の診断時点から、より長期的で予後的な役割へと変化していることを示しています。このことは、せん妄管理における音声バイオマーカーの強力な将来の軌跡を示唆しています。これらは急性期の診断ツールにとどまらず、せん妄エピソード後に持続的な認知機能障害やその後のAD/ADRD発症のリスクが高い患者を特定するための貴重な予後指標となる可能性を秘めています。これにより、個別化された退院後ケア、的を絞った認知リハビリテーション戦略、および長期的な認知機能低下を軽減することを目的とした早期介入のための全く新しい道が開かれます。音声バイオマーカーが、入院期間を超えた患者ケアへのより全体的で予測的なアプローチにどのように貢献できるかという、このより広範な長期的な影響を強調すべきです。

6. 結論

せん妄は、その高い発生率、重篤な合併症、そして驚くべきほど高い未診断率によって、臨床現場に大きな負担を課しています。患者の転帰を改善するためには、客観的で早期の検出方法が極めて重要であり、これは現在満たされていないニーズです。

自動音声分析と機械学習および深層学習技術の組み合わせは、この診断上のギャップに対処する上で、有望で急速に進化する役割を担っています。音響的、プロソディ的、言語的な幅広い音声特徴量が、その予測可能性について現在探求されています。

モデル開発と報告された性能指標において顕著な進歩が達成されており、特に制御された環境や合成データ環境ではその傾向が顕著です。しかし、これらの有望な初期結果と、実際の臨床データで観察されることが多い低い性能との間の重要な区別を強調することは不可欠であり、臨床応用における複雑さを浮き彫りにしています。

広範な臨床導入のために体系的に対処されなければならない主要な課題が再確認されます。それは、標準化された堅牢なデータ収集プロトコルの必要性、多様な患者集団や臨床設定におけるモデルの汎用性の確保、そして医療専門家間の信頼を育む解釈可能なAIモデルの開発の必要性です。

将来の研究に対する説得力のあるビジョンが示されています。これには、音声特徴量を他の包括的な臨床データと統合するマルチモーダルアプローチへの強い重点、音声変化に反映されるせん妄の根底にある病態生理学的メカニズムのより深い探求、そして音声バイオマーカーが急性期検出を超えて、長期的な予後評価と個別化された介入戦略のための貴重なツールへと進化し、最終的にせん妄ケアに革命をもたらす変革的潜在能力が含まれます。

引用文献

1. Enhanced Predictive Modelling for Delirium in Intensive Care Using Simplified Deep Learning Architecture with Attention Mechanism - Scientific Research Publishing, 6月 20, 2025にアクセス、
<https://www.scirp.org/journal/paperinformation?paperid=140110>
2. (PDF) Artificial intelligence applications in delirium prediction, diagnosis, and management: a systematic review - ResearchGate, 6月 20, 2025にアクセス、
https://www.researchgate.net/publication/392440605_Artificial_intelligence_applications_in_delirium_prediction_diagnosis_and_management_a_systematic_review
3. Biomarkers of neurocognitive disorders: A ray of hope on the ..., 6月 20, 2025にアクセス、
https://www.researchgate.net/publication/378071891_Biomarkers_of_neurocognitive_disorders_A_ray_of_hope_on_the_horizon
4. Biomarkers of delirium risk in older adults: a systematic review and meta-analysis - Frontiers, 6月 20, 2025にアクセス、

<https://www.frontiersin.org/journals/aging-neuroscience/articles/10.3389/fnagi.2023.1174644/full>

5. Machine Learning–Based Prediction of Delirium and Risk Factor Identification in Intensive Care Unit Patients With Burns: Retrospective Observational Study, 6月 20, 2025にアクセス、<https://pmc.ncbi.nlm.nih.gov/articles/PMC11923481/>
6. Enhanced Predictive Modelling for Delirium in Intensive Care Using Simplified Deep Learning Architecture with Attention Mechanism - Open Access Library (OALib), 6月 20, 2025にアクセス、<https://www.oalib.com/articles/6846084>
7. Development and Validation of a Machine Learning Model for Early Prediction of Delirium in Intensive Care Units Using Continuous Physiological Data: Retrospective Study - PubMed, 6月 20, 2025にアクセス、<https://pubmed.ncbi.nlm.nih.gov/40173433/>
8. Machine Learning–Based Prediction of Delirium and Risk Factor Identification in Intensive Care Unit Patients With Burns: Retrospective Observational Study - JMIR Formative Research, 6月 20, 2025にアクセス、<https://formative.jmir.org/2025/1/e65190>
9. Delirium | Johns Hopkins Medicine, 6月 20, 2025にアクセス、<https://www.hopkinsmedicine.org/health/conditions-and-diseases/delirium>
10. Delirium and Dementia | PM&R KnowledgeNow, 6月 20, 2025にアクセス、<https://now.aapmr.org/delirium-and-dementia/>
11. Identifying diagnostic biomarkers for Delirium and predicting cognitive Outcomes in hospitalized older adults using automated Speech Analysis (IDOSA) - NIH RePORTER, 6月 20, 2025にアクセス、<https://reporter.nih.gov/project-details/10806491>
12. Identifying diagnostic biomarkers for Delirium and predicting cognitive Outcomes in hospitalized older adults using automated Speech Analysis (IDOSA) - NIH RePORTER, 6月 20, 2025にアクセス、<https://reporter.nih.gov/search/MX3tpaHFGUiOYSt-dYFEsg/project-details/10806491>
13. 発話音声からアルツハイマー型認知症を検出するアプリを開発, 6月 20, 2025にアクセス、<https://1post.jp/6982>
14. [2501.18731] Evaluating Spoken Language as a Biomarker for Automated Screening of Cognitive Impairment - arXiv, 6月 20, 2025にアクセス、<https://arxiv.org/abs/2501.18731>
15. Artificial Intelligence–Powered Quality Assurance: Transforming Diagnostics, Surgery, and Patient Care—Innovations, Limitations, and Future Directions, 6月 20, 2025にアクセス、<https://pmc.ncbi.nlm.nih.gov/articles/PMC12028931/>
16. a scoping review of state of the art machine learning methods for clinical speech classification - arXiv, 6月 20, 2025にアクセス、<https://arxiv.org/pdf/2503.04802>
17. (PDF) The order in speech disorder: a scoping review of state of the art machine learning methods for clinical speech classification - ResearchGate, 6月 20, 2025にアクセス、https://www.researchgate.net/publication/389693999_The_order_in_speech_disorder_a_scoping_review_of_state_of_the_art_machine_learning_methods_for_clinical_speech_classification

18. a scoping review of state of the art machine learning methods for clinical speech classification - arXiv, 6月 20, 2025にアクセス、<https://arxiv.org/html/2503.04802v1>
19. 外科病棟の看護師が認識する 術後せん妄発症予測に関する構成要素 - 札幌保健医療大学, 6月 20, 2025にアクセス、
<https://www.sapporo-hokeniryou-u.ac.jp/uploads/files/library/kiyo/vol.6/p.13-27.pdf>
20. Speech as a Biomarker - VOCAL Technologies, 6月 20, 2025にアクセス、
<https://vocal.com/voice-processing/speech-as-a-biomarker/>
21. Markers of schizophrenia at the prosody/pragmatics interface. Evidence from corpora of spontaneous speech interactions - Frontiers, 6月 20, 2025にアクセス、
<https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2023.1233176/full>
22. Predicting explainable dementia types with LLM-aided feature engineering | Bioinformatics | Oxford Academic, 6月 20, 2025にアクセス、
<https://academic.oup.com/bioinformatics/advance-article/doi/10.1093/bioinformatics/btaf156/8108827>
23. 博士論文 要約 高齢者におけるせん妄発症予測モデルの構築 - OPAC, 6月 20, 2025にアクセス、https://opac.ll.chiba-u.jp/da/curator/900123105/NKA_0254_Y.pdf
24. Delirium - ASHA, 6月 20, 2025にアクセス、
<https://apps.asha.org/EvidenceMaps/Maps/LandingPage/ef815b23-827c-ed11-8141-005056834e2b>
25. Biomarkers of Delirium Duration and Delirium Severity in the ICU - IU Indianapolis ScholarWorks, 6月 20, 2025にアクセス、
<https://scholarworks.indianapolis.iu.edu/server/api/core/bitstreams/3f68a742-5e83-4855-a871-b7204bb6985e/content>