

博士論文

Artificial intelligence in respiratory medicine
(呼吸器診療を支援する医療 AI 技術の開発)

令和 5 年 6 月

濱田 和希

山口大学大学院医学系研究科 医学専攻
システムバイオインフォマティクス講座

目次

第1章 序論	1
1.1 医療全体の課題	1
1.2 デジタル社会におけるデータ駆動型医療	4
1.3 AI技術の医療応用	10
1.4 呼吸器領域の課題改善に向けたAI技術の開発	15
1.5 本論文の構成	18
1.6 参考文献	19
第2章 ベイズ推定に基づく副作用の原因薬推定技術	23
2.1 序論	23
2.2 研究方法	24
2.2.1 学習データベース	26
2.2.2 ベイズ推定モデルの構築	26
2.2.3 検証方法	28
2.2.4 不均衡分析に基づく統計的手法	28
2.2.5 Bayesian Confidence Neural Network (BCPNN)	29
2.2.6 Precision-recall (PR) 解析	29
2.3 解析結果	30
2.3.1 検証に用いた副作用報告	30
2.3.2 ベイズ推定モデル適用例	34
2.3.3 性能評価	36
2.4 考察	41
2.5 結語	45
2.6 参考文献	46
第3章 喘息患者の気流閉塞の急速進行を予測する技術	50
3.1 序論	50
3.2 研究方法	51

目次

3.2.1	研究デザイン	51
3.2.2	学習データベース	53
3.2.3	全説明変数を用いた rapid decline 予測モデル	53
3.2.4	Permutation importance 解析	56
3.2.5	選択した説明変数を用いた rapid decline 予測モデル	57
3.3	解析結果	57
3.3.1	全説明変数を用いた機械学習モデルの予測性能	57
3.3.2	Permutation importance 解析結果	58
3.3.3	選択した説明変数を用いた機械学習モデルの予測性能	62
3.4	考察	65
3.5	結語	70
3.6	参考文献	70
第 4 章	喘息症状から病態特性を推定する技術	75
4.1	序論	75
4.2	研究方法	76
4.2.1	研究デザイン	76
4.2.2	Asthma Control Questionnaire-5 (ACQ-5)	79
4.2.3	学習データベース	79
4.2.4	次元削減	80
4.2.5	教師あり機械学習	80
4.2.6	教師なし機械学習	81
4.2.7	統計解析	81
4.3	解析結果	82
4.3.1	次元削減の結果	82
4.3.2	教師あり機械学習の結果	86
4.3.3	教師なし機械学習の結果	88
4.3.4	喘息症状と Treatable traits の関連性	93
4.4	考察	95

目次

4.5 結語	103
4.6 参考文献.....	103
第5章 総括	108
5.1 本論文のまとめ.....	108
5.2 今後の展望	111
5.3 参考文献.....	112
謝辞	114
付録	116

論文要旨

現在の日本は少子高齢化・医療高度化を背景に、要介護者数の増加、医療費増加、人材不足、医療格差などの医療課題に直面している。近年、人工知能（Artificial intelligence：AI）技術やシステム医学に基づいたデータ駆動型医療の登場によって、これら課題を解決できる可能性が広がった。筆者自身は呼吸器診療に携わる立場でもあることから、呼吸器診療を支援する以下3つの医療AI技術を開発した。

1つ目の技術では、副作用ビッグデータ（Japanese Adverse Drug Event Report：JADER）に基づきベイズ推定を行うことで、AUC 0.93の精度で副作用の原因薬を推定できた。呼吸器領域では薬剤性肺障害など重篤な副作用があるが、本技術の臨床応用によって、副作用による健康被害の最小化と副作用管理の効率化が期待できる。

2つ目の技術では、喘息患者の臨床データ（年齢、BMI、血中好酸球数、呼気NO値、増悪回数）をもとに教師あり機械学習を行うことで、喘息患者の気流閉塞の急速進行（1秒量低下）をAUC 0.85の精度で予測できた。実用化によって、早期の治療介入が必要な喘息患者を同定でき、重症化を防ぐための先制治療につながる。

3つ目の技術では、喘息質問票（Asthma Control Questionnaire-5: ACQ-5）のデータを用いて教師なし機械学習を行うことで、症状から喘息病態である気流閉塞、2型気道炎症、増悪リスクを推定できた。従来、治療選択のため病態評価には専門的検査が必要だが、本技術ではACQ-5に含まれる喘息症状の評価のみから、個々の喘息の病態に応じた治療選択（個別化治療）につなげることができる。本技術は、発展途上国、過疎地域、プライマリケアの現場など、医療環境が不十分な地域で、適切な喘息治療薬の選択を支援できる。ひいては、医療格差の是正につながる可能性が期待できる。

本研究で開発したAI技術の実用化によって、臨床現場における副作用管理、先制医療、個別化治療を支援する。これによって、健康寿命の延伸、医療費抑制を目指す。同時に、開発したAI技術によって専門医療の一部を補完することで、医療従事者の業務負担軽減と、医療格差の是正（均てん化）につながることを期待できる。

第1章

序論

1.1 医療全体の課題

現在、直面している医療全体の課題を図1にまとめて示す。

第一に健康寿命の喪失が課題に挙げられる。日本では世界に先駆けて少子高齢化が進んでいる。2022年度時点での日本の高齢化率（65歳以上の割合）は28.9%である。今後も高齢化率はさらに上昇する見込みである[1]。高齢化によって疾病構造は変化した。悪性腫瘍、慢性呼吸器疾患、心血管疾患をはじめとする生活習慣病の有病率および死亡率が増加した[2]。これらの生活習慣病は、健康寿命の喪失（要介護状態）の大きな要因となる。医療技術の高度化により平均寿命は延伸したが、平均寿命と健康寿命の間には年齢差がある[3]。平均寿命と健康寿命の差は、2019年度時点で、男性で8.7年、女性で12.1年であった。このような日常生活に制限のある「不健康な期間」では生活の質（Quality of Life：QOL）の低下が問題となる。

第二に医療費増加が挙げられる。健康寿命「不健康な期間」においては、高度医療が必要な機会が増える。高度医療は高額であるため、医療費が増加する。これによって、日本の医療費は増加し続けている[4]。最近、日本で行われたシミュレーション研究では、今後、認知症やフレイルによる疾病負荷や経済的負担のさらなる増加が予想されている[5]。

第三に、働き方改革、少子高齢化、医療の細分化などによる医療従事者の人材不足が懸念されている。

これに加え、医療の細分化や高度化によって医療格差が顕在化している[6]。医療技術の進歩とともに、医療従事者に求められる医療水準は高度化し、細分化した。専門医療の提供のためには、それを担う人材、高額な医療用機器の導入、医療費などが必

第1章 序論

要である。しかし、発展途上国では、専門医療を担う人材の不足や、経済格差のために、専門医療を享受することが困難である[7]。発展途上国だけでなく先進国の国々においても、プライマリアケアの現場や医療過疎地域では、医療提供体制が十分ではない。特に日本では都市部に医師が偏り、過疎地域では無医村や無医地区人口が増加しているため、地域間で医療提供体制に格差が生じている[8,9]。同時に、高度医療施設とプライマリアケアの間に、専門医療技術についての格差があり、施設間格差が生じている[10,11]。このような、様々な医療格差は、疾患の診断や治療の機会損失につながる。その結果、多くの患者で健康状態の格差（健康格差）が生じてしまう[12]。

これらの医療全体の課題の解決にむけて、健康寿命の延伸、医療費抑制、業務負担軽減、医療格差の是正が求められている。

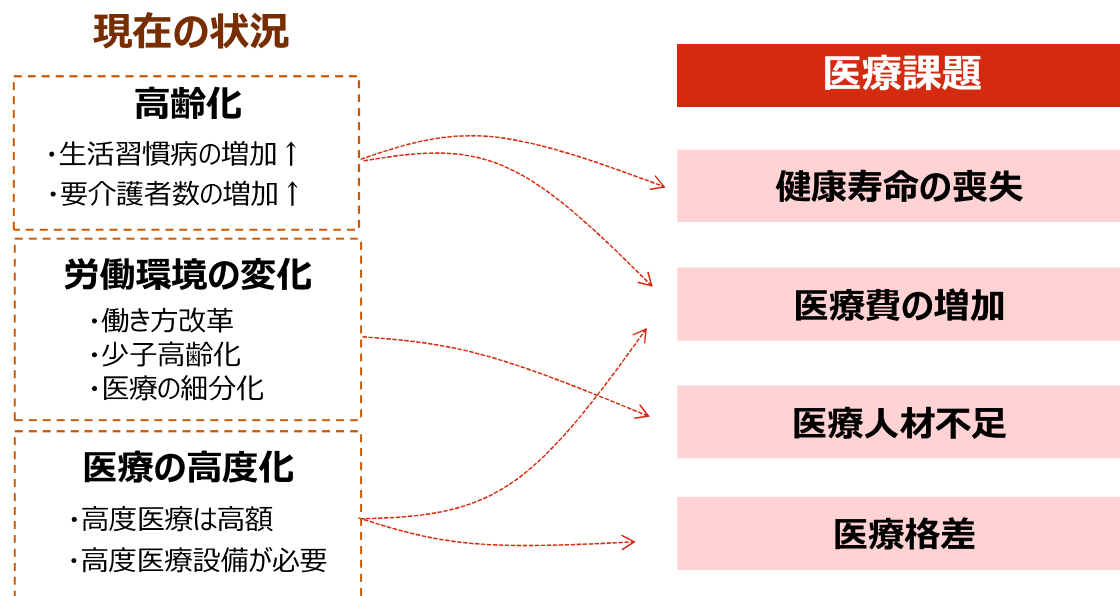


図1 医療全体が直面している課題について示す。高齢化、労働環境の変化、医療の高度化を背景に、日本の医療は、健康寿命の喪失、医療費の増加、医療人材不足、医療格差などの課題に直面している。

1.2 デジタル社会におけるデータ駆動型医療

情報通信技術（Information and communication technology：ICT）や人工知能（Artificial intelligence：AI）の解析技術の進歩を背景に、日本ではスマート社会（Society 5.0）への移行が進んでいる。このような社会構造の変革と同時に、前述の医療全体の課題（健康寿命と平均寿命の差、医療人材不足、医療費の増加、医療格差）の解決に向けて ICT や AI などを活用したデジタルヘルスへの期待が高まっている [13,14]。

スマート社会への移行を背景に医療 DX（デジタルトランスフォーメーション）が進み始めた。これによって、医療現場の膨大なデータがデジタルで取得されるようになった [15-18]。電子カルテの普及による医療データの電子化に始まり、近年ではスマートフォンやウェアラブルデバイスを用いた生理学的データ・行動データ（デジタルバイオマーカー）の取得や [19,20]、電子デバイスを用いた患者の主観的評価の取得（electronic patient-reported outcome：ePRO）などがある（図2） [21]。このようにして蓄積される医療ビッグデータは膨大である。医療ビッグデータは極めて膨大であるため [15-18]、人間の認識では詳細な特徴パターンを理解することは難しい。

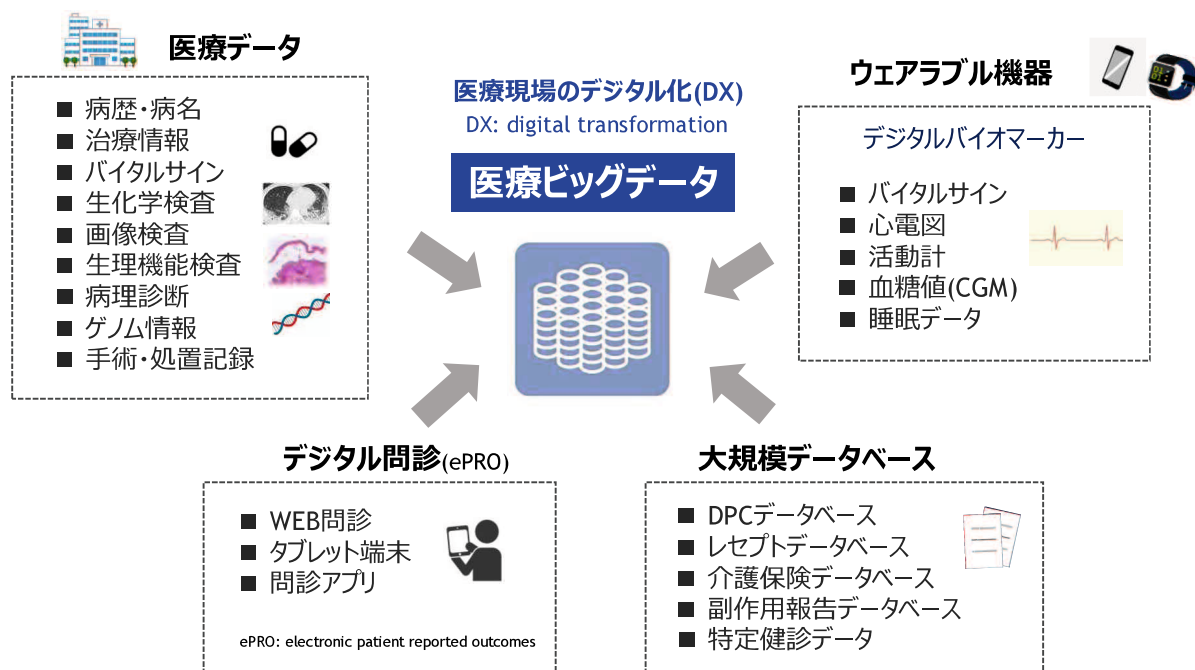


図2 医療のデジタル化（医療 DX）について示す。医療のデジタル化によって、病院内の医療情報だけでなく、生理学的データ・行動データ（デジタルバイオマーカー）[19,20]や、患者の主観的評価（ePRO）のデータ[21]がデジタル形式で取得される。これによって膨大な医療ビッグデータが蓄積されるようになった。DX, digital transformation; ePRO, electronic patient-reported outcome

第1章 序論

医療現場で取得されるデータは、データ項目数が膨大である点で極めて高次元なデータである。病歴、症状、バイタルサイン、身体所見、血液・生化学検査、細菌検査、生理機能検査、放射線検査、病理組織検査など医療データの項目は膨大で多岐にわたる。その上、それぞれの医療データ項目ごとでも変数の種類は膨大である。例えば血液検査だけでも項目数は何千種類とある。このような高次元データではデータ分布を可視化できないため、人間の目でみてデータの特徴パターンを認識することができない。多重検定の問題があるため、高次元データにおいては統計仮説検定の結果から適切に結論を導くことが困難である。

医療データは非線形である。その理由は、生命現象自体がそうだからである。非線形とは、一次式で表されるような単純な比例関係ではないことをいう。例えば、BMIと死亡率の関係は非線形である[22]。多くの疾患の発症や重症化には多因子が関わるため、単独の遺伝子や分子に原因を求めることはできない[15,16]。病態形成の過程は、遺伝子・分子・細胞・組織・臓器・個体と多階層に渡る。無数の因子間の非線形相互作用が影響する。遺伝子レベルや分子レベルなど各階層における相互作用と、複数の階層に渡った相互作用がある。階層自体は解析のための便宜的な概念に過ぎないが、このような3次元のネットワーク構造は、病態の複雑性や非線形性につながっている(図3)。加えて、このような個体要因だけでなく環境要因(生活習慣、運動習慣など)・社会要因(学歴、所得、家族構成、職業など)との複雑な相互作用も病態形成に関わる。

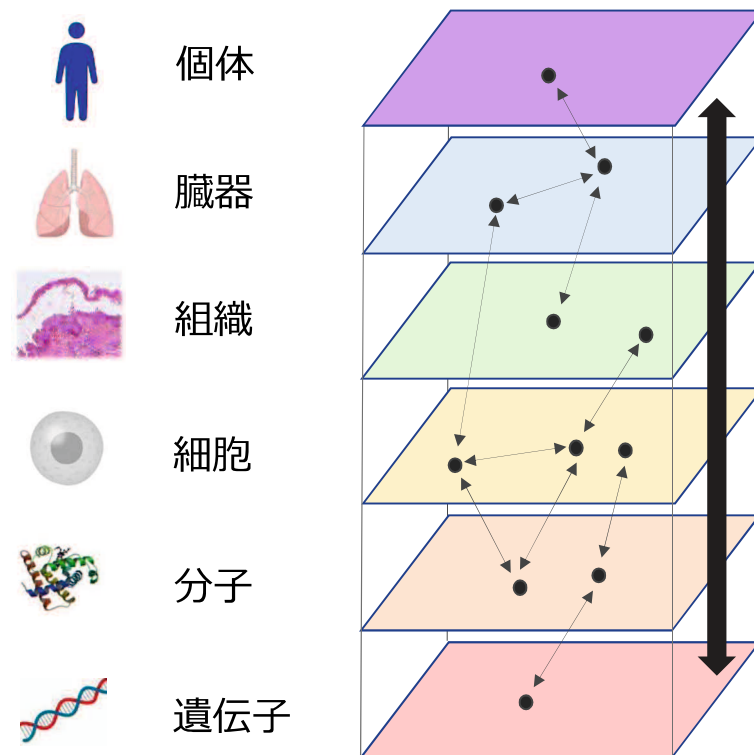


図3 病態形成における多階層性を示す。病態形成の過程は、遺伝子・分子・細胞・組織・臓器、個体と多階層に渡り、無数の因子間の非線形相互作用が関与する。遺伝子や分子など各階層における「水平方向」と、複数の階層に渡る「垂直方向」の相互作用がある。このような3次元ネットワークが病態の複雑性をもたらしている。

第1章 序論

非線形性や高次元性を有し、多階層に渡る医療ビッグデータを詳細に解析するための方法論として、近年、データ駆動型アプローチが注目を集めている。データ駆動型アプローチとは、データサイエンスの解析技術を用いて膨大な医療データから潜在的な特徴パターンを見出し判断や予測を行う、あるいは有益な知識を発見する方法論である。データ駆動型アプローチでは、従来の統計解析では考慮することが困難であった非線形性や高次元性を考慮できる[15-18]。データ駆動型アプローチには、人間による仮説や専門知識のフレームワークに囚われない長所がある。近年のデータサイエンス解析技術の急速な発展を背景とし、AI技術やシステムバイオロジーの医学研究や医療への応用が始まっている。

AIとは、人間の知的な情報処理、判断をコンピューターが自動で行う技術のことである[23]。AI技術の中核を担う機械学習では、膨大なデータをもとにパターン・法則性をコンピューターが自動学習し、高精度に判別・予測を行うことができる[24,25]。データ同士の類似性を学習することによって、探索的に知識を発見することも可能である。多くの機械学習は、非線形性なデータ、高次元なデータを解析することが可能である[26]。近年、機械学習の解析手法は発展著しく、医療分野への応用が始まっている。しかし、機械学習をはじめとするAI技術はあくまで変数間の数値的な関係にフォーカスした数学的方法論であり、医学における病態生理メカニズムを考慮していないという課題がある。

一方、システムバイオロジーは生命現象をシステムと捉え、生体メカニズムの統合的な理解を目指す新たな学問領域である[27]。個体単位での病態生理の統合的な理解を目指す。システムバイオロジーでは、蓄積した膨大なデータに基づき数理モデルを構築するアプローチをとられることが多い。システムバイオロジーとAI技術の協働によって、AI技術のみでは考慮できない病態生理メカニズムを捉えた上で、医学研究や医療へ応用できることが期待される(図4)。しかし、システムバイオロジーとAI技術の融合によるアプローチは、それぞれの研究領域への深い理解と長年の研究キャリアが不可欠である。そこで、本博士論文では、システムバイオロジーとAI技術の融合的アプローチに向けた第一歩として、まずはAI技術を用いた医学研究を行う。

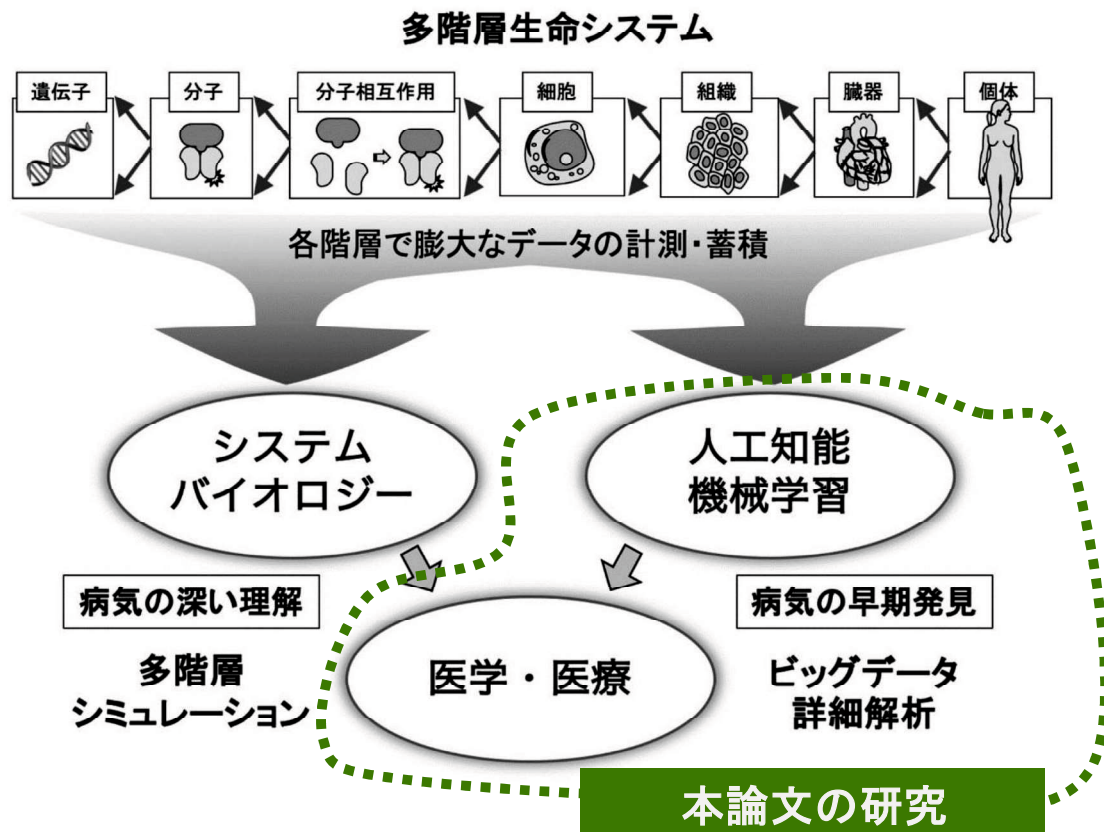


図4 システムバイオロジーと人工知能の融合と、その医学や医療への応用の模式図を示す。文献[27]より一部を改変した。本論文では、システムバイオロジーとAI技術の融合的アプローチに向けた第一歩として、まずはAI技術を用いた医学研究を行う。

1.3 AI技術の医療応用

近年、臨床現場へのAI技術の応用が進んでいる。AI技術に基づく先制治療、個別化治療の実現や専門医療の補完によって、健康寿命の延伸、医療費抑制、医療格差の是正、業務負担軽減につながることを期待できる。

AI技術の1つである教師あり機械学習では、既にある一連の入力データとそれに対する既存の応答データ（正解ラベル）を用いてモデルを訓練する（図5）。これによって、新規にデータが観測された場合に未知の応答データを分類あるいは予測できるようになる[24]。例えば応答データを「疾患の有/無」とした学習モデルを構築することで、新たな患者における疾患の有無を分類できる。応答データを「1年以内の重症化の有/無」にすると、患者の1年以内の重症化を予測できる。疾患の発症や重症化を予測できれば、発症や重症化の前に先制治療を行うことができる（図6）。

別のAI技術に教師なし機械学習がある。教師なし機械学習は、応答データ（正解）を与えずに学習させる方法である。教師なし機械学習の代表的な方法にクラスタリングがある[28]。クラスタリングでは、データ間の類似性（ベクトル間の距離）に基づいて、多様な集団をいくつかのグループへ分類する（図5）。これにより、同一疾患の中にも臨床的背景や病態の異なる疾患サブタイプを見出すことが出来る。疾患サブタイプそれぞれに有効な治療を選択することで、疾患の多様性に応じた個別化治療を実現できる（図7）[29]。

AI技術の予測に基づく先制医療は健康寿命の延伸につながる可能性がある。健康寿命の喪失（要介護状態）の要因となる生活習慣病は、適切な生活指導、早期治療、介護予防支援等によって発症や重症化を予防できる。AI技術の活用で疾患の発症や重症化を予測できれば、従来よりも早期に治療介入できるようになる[30]。これによって、疾患の発症や重症化の予防につながり、ひいては健康寿命の延伸に貢献する。AI技術の活用で個別化治療を実現できれば、個々の病態や背景因子に応じた治療介入で、より治療効果が期待できる。同時に、副作用リスクのある無効治療の回避や医療資源の適正配分にもつながる。先制医療・個別化治療で健康寿命が延伸すれば、救急

第1章 序論

受診、入院、要介護状態等に至るリスクが減少し、医療費抑制にもつながる。

AI技術の医療応用によって、医療格差の是正、業務負担軽減につながることも期待できる。専門医療には高度な医療設備や専門医療を担う人材が必要であるため、先進国・都市部など一部の医療施設でしか専門医療を提供できない。これは、医療格差をもたらす。しかし、熟練の専門医の知識・経験に基づく診療をAI技術で再現できれば、格差のない医療の実現につながる。実際に、放射線画像や内視鏡画像などの画像診断は、専門医の水準に迫る診断精度を達成している。このようなAI技術を遠隔医療で活用すれば、発展途上国や過疎地域など医療環境が十分でない地域においても専門医療を実現できる可能性がある[31,32]。これによって、国際間格差、地域間格差、施設間格差など様々な医療格差の是正につながる。AI技術はデータを学習することで自動的に判断できる。医療の一部をAI技術が補完できれば、医療従事者の業務負担軽減にもつながる。

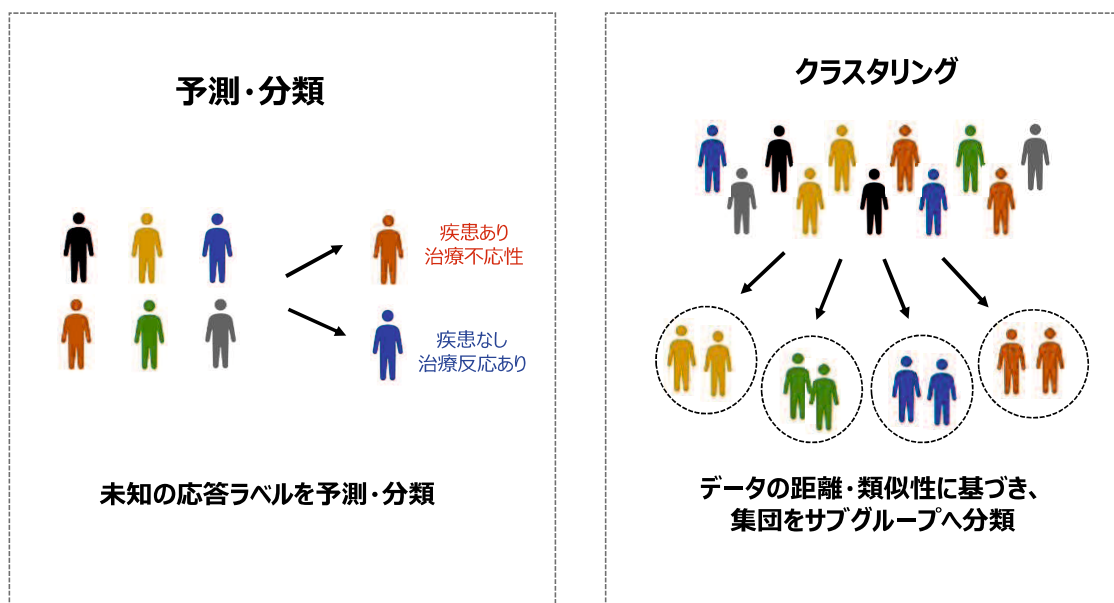


図5 AI技術の中核を担う機械学習について示す。AI技術の1つである教師あり機械学習では、既にある一連の入力データとそれに対する既存の応答データ（正解ラベル）を用いてモデルを訓練する。これによって、新規にデータが観測された場合に未知の応答データを判断・予測できるようにする。教師なし機械学習では、データ間の距離・類似性に基づいて、多様な全体集団をいくつかのグループへ分類する。これによって、データ内に存在する未知のパターンや構造を見出すことが可能となる。

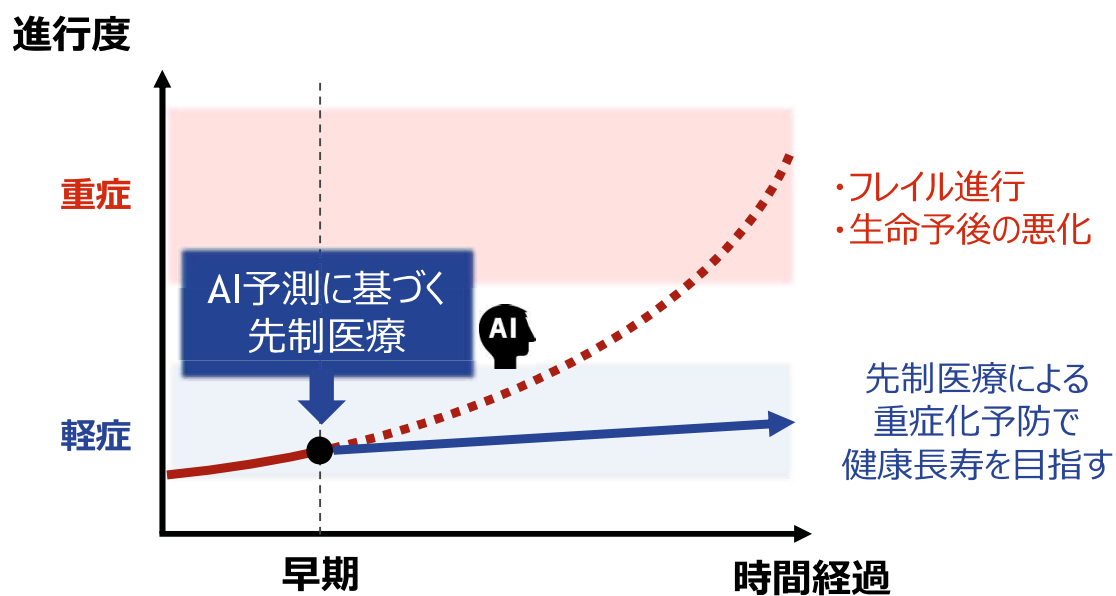


図6 AI技術の活用で疾患の発症や重症化を予測できれば、従来よりも早期に治療介入できるようになる。AI予測に基づく先制医療によって、疾患の発症や重症化の予防につながり、ひいては健康寿命の延伸に貢献することが期待できる。

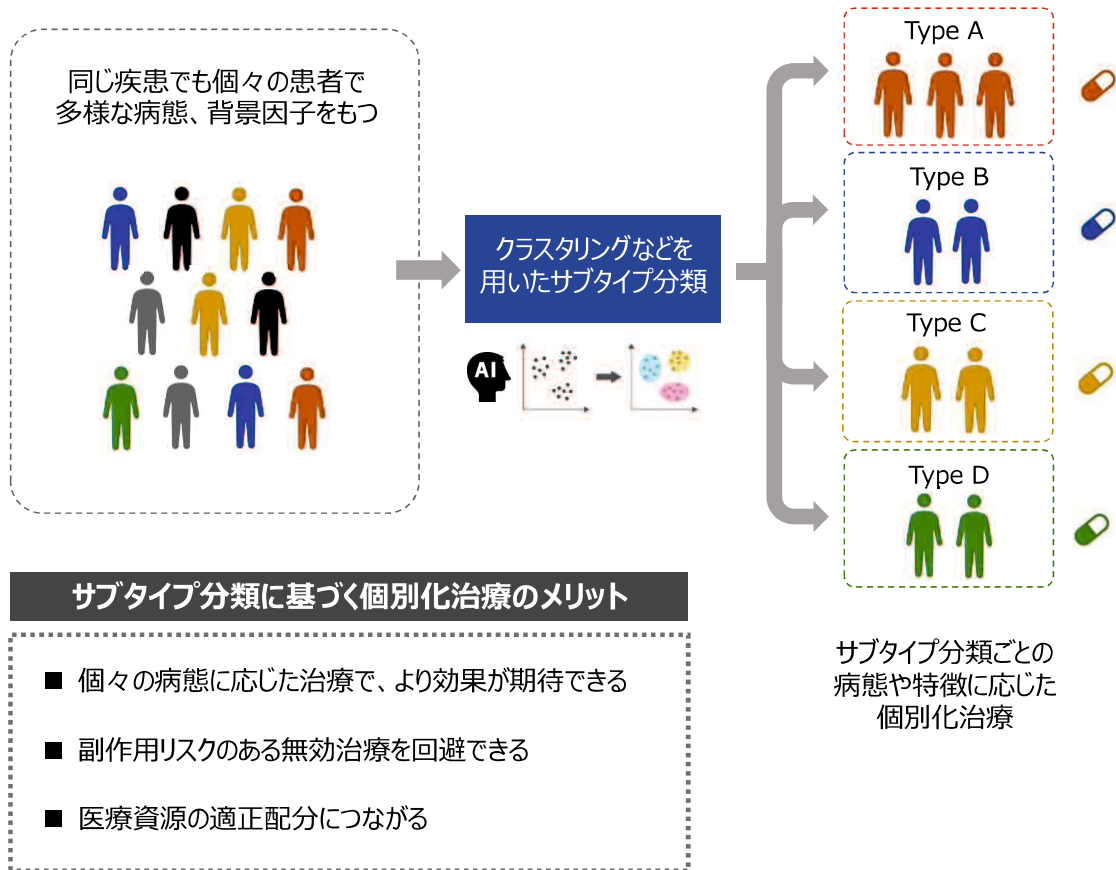


図7 クラスタリングなどの AI 技術の活用で個別化治療を実現できれば、個々の病態に応じた治療で、より高い治療効果が期待できる。同時に、治療効果が見込まれる患者へ治療を行うことで、副作用リスクのある無効治療の回避や、医療資源の適正配分にもつながる。

1.4 呼吸器領域の課題改善に向けた AI 技術の開発

筆者自身は呼吸器診療に携わる立場でもあることから、本論文では、呼吸器領域の医療課題と、その改善に向けた AI 技術の開発について考えた (図 8)。

第一に、難治性呼吸器疾患の不可逆的な進行が挙げられる。喘息、COPD、間質性肺疾患をはじめ慢性呼吸器疾患が進行すると、不可逆的な肺・気道構造の破壊や呼吸機能低下を生じる。呼吸器系の不可逆的な構造異常や機能異常は、生活の質低下やフレイルならびに要介護状態につながる。この状態に至ってしまうと、最善の治療介入を行っても元の正常構造や正常機能に復するのは困難なことが多い。死亡リスクの増加にもつながる。しかし、AI 技術の進行予測に基づいた早期段階での治療介入、すなわち先制医療を実現できれば、疾患の発症や重症化を未然に防ぐことができる。これによって健康寿命の延伸につながることを期待できる。

第二に分子標的薬や肺癌の免疫療法など新規治療の登場など医療高度化による医療費の増加も課題である。しかし、AI 技術に基づく先制医療で健康寿命を延伸できれば、高額な救急医療や高度医療が必要な機会が減少し、医療費抑制につながることを期待できる。同時に、AI 技術による個別化治療の支援が実現すれば、治療効果が期待できる患者への医療資源の適正配分が可能となり、これによっても医療費抑制につながる。

第三の課題に、背景因子や病態が多様な患者への一律な治療選択がある。患者ごとで病態や背景因子は多様であるにも関わらず、それを考慮せずに同じ一律の治療を行ってしまうという課題である。AI 技術によって個別化治療を支援できれば、個々の患者の病態や背景因子に適した、より有効な治療を行うことが可能となる。

4 番目に、高度化する副作用管理や、重篤な副作用の存在が挙げられる。呼吸器領域では薬剤性肺障害などの致死的な重篤な副作用が存在する。分子標的治療や免疫療法など新規治療の登場によって、免疫関連有害事象 irAE など新たな副作用もみられるようになった。副作用管理は年々、高度化している。AI 技術によって副作用管理を支援することができれば、副作用による健康被害を最小限に抑えることが期待できる。

第1章 序論

5 番目に呼吸器診療における医療格差を課題に挙げた。全国的な呼吸器専門医の不足、医療体制の施設間格差、地域格差などに起因する医療格差が存在する。AI 技術によって専門医療を補完することができれば、医療格差の是正につながる。

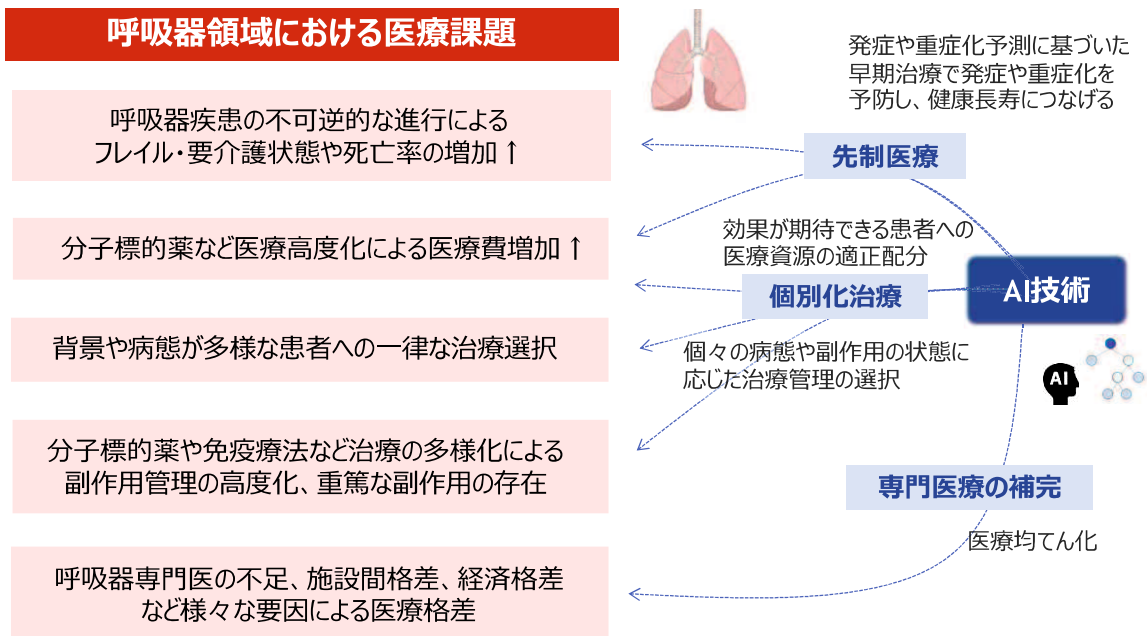


図8 呼吸器領域における医療課題および、課題改善にむけたAI技術の活用について示す。

1.5 本論文の構成

近年技術革新の著しいAI技術を医療現場へ応用することによって、健康寿命の延伸、医療費抑制、業務負担軽減、医療格差の是正に貢献できる可能性が広がった。

そこで、本論文では、以下の通り、呼吸器診療を支援する3つの医療AI技術を開発した。

- ・ベイズ推定に基づく副作用原因薬推定技術
- ・喘息患者の急速な呼吸機能低下を予測するAI技術
- ・喘息症状から治療標的となる病態特性を推定するAI技術

本論文の構成は以下の通りである。

第1章（本章）は序論である。医療全体の様々な課題と、その改善に向けたデータ駆動型アプローチの医療応用について概説した。

第2章以降では、診療場面ごとで、呼吸器診療を支援するAI技術について述べる。

第2章では、副作用の原因薬を推定する技術について述べる。呼吸器疾患の治療薬には様々な副作用がある。また、呼吸器内科学分野が治療管理に携わる副作用には薬剤性肺障害をはじめ様々なものがある。副作用管理には原因薬剤の同定が不可欠だが、副作用の原因薬推定は時に困難である。そこで、副作用ビッグデータである日本の有害事象報告データベース（JADER: Japanese Adverse Drug Event Report）[33]を用いて、副作用の原因薬を推定するベイズ統計モデルを構築した。この技術を用い、副作用の原因薬を推定することで、副作用管理を向上させ、効率化することができる。

第3章では、喘息患者における気流閉塞の急速な進行を予測する技術について述べる。気流閉塞の急速な進行は喘息重症化だけでなく、心血管イベントリスク上昇、死亡リスク上昇と関連するため、進行抑制が重要である。気流閉塞の急速な低下リスクが高いと予測される患者では、早期に治療強化を行うことで、呼吸機能の喪失を予防できる。これによって、喘息患者の重症化予防につながり、ひいては、健康寿命の延

伸に貢献することが期待できる。

第4章では、取得に専門的な検査機器を要さない喘息質問票（ACQ-5: Asthma Control Questionnaire-5）[34,35]の評価のみを用いて、治療標的となる病態特性を推定する技術について述べる。より優れた治療効果を得て、医療資源を効率的に利用するためには、喘息患者の病態の個別性に基づいた治療選択（個別化治療）を行うことが重要である。しかし、検査設備の導入費用や、検査技師の雇用の問題から、医療環境が不十分な国々や地域の医療機関では病態の評価が難しい[36,37]。喘息症状に含まれる情報のみから治療標的となる病態特性を推定することで、発展途上国、過疎地域、プライマリケアなどでも、喘息の個別化治療を支援することができる。これによって、グローバルな医療格差の是正に貢献する。

第5章では、本研究の成果をまとめ、今後の展望について述べる。

1.6 参考文献

1. 内閣府「令和4年版高齢社会白書（全体版）」
https://www8.cao.go.jp/kourei/whitepaper/w-2022/gaiyou/04pdf_indexg.html
2. 厚生労働省「令和3年(2021)人口動態統計月報年計（概数）の概況」
<https://www.mhlw.go.jp/toukei/saikin/hw/jinkou/geppo/nengai21/index.html>
3. 厚生労働省「令和3年簡易生命表の概況」
<https://www.mhlw.go.jp/toukei/saikin/hw/life/life21/index.html>
4. 国立社会保障・人口問題研究所「令和2年度社会保障費用統計」
https://www.ipss.go.jp/site-ad/index_Japanese/security.html
5. Kasajima M, Eggleston K, Kusaka S, Tnaka T, Son BK, Iijima K, et al. Projecting prevalence of frailty and dementia and the economic cost of care in Japan from 2016 to 2024: a microsimulation modelling study. *Lancet Public Health* 2022; 7: e458-68
6. Braveman, P., & Gruskin, S. (2003). Defining equity in health. *J Epidemiol Community Health*, 57 (4), 254-258.

第1章 序論

7. Stolbrink M, Thomson H, Hadfield RM, et al. The availability, cost, and affordability of essential medicines for asthma and COPD in low-income and middle-income countries: a systematic review. *Lancet Glob Health* 2022; 10: e1423–42
8. 厚生労働省「令和元年度無医地区等及び無歯科医地区等調査」令和2年5月
<https://www.mhlw.go.jp/toukei/list/76-16b/R2.html>
9. Otsubo T, Goto E, Morishima T. Regional variations in in-hospital mortality, care processes, and spending in acute ischemic stroke patients in Japan. *J Stroke Cerebrovasc Dis.* 2015; 24(1):239-51.
10. Redmond C, Heaney LG, Chaudhuri R, et al. Benefits of specialist severe asthma management: demographic and geographic disparities. *Eur Respir J* 2022
11. 川淵孝一、杉原茂「高度医療技術の有効性の施設間格差：経皮的冠動脈インターベンションのケース」*医療と社会* 2005年15巻1号 111-127
12. 近藤克則「健康格差社会 第2版: 何が心と健康を蝕むのか」医学書院 2022年6月20日
13. 佐々木隆之「デジタルテクノロジーの進展と医療ヘルスケアのパラダイムシフト-データ駆動型ヘルスケアの実現に向けて-」*医薬産業政策研究所リサーチペーパー・シリーズ(Web)* 2021年07月
14. ヒューマンサイエンス振興財団「医療分野におけるビッグデータ並びに ICT・AI の利活用の最新動向—創薬並びに個別化医療・先制医療への貢献の道を探る」2017年3月
15. 永井良三、宮野悟、大江和彦「ビッグデータ変革する生命科学・医療」*実験医学* 2016年
16. 田中博著「もっとよくわかる！医療ビッグデータ オミックス、リアルワールドデータ、AI医療、創薬」羊土社 2020年3月30日
17. Dash, S., Shakyawar, S.K., Sharma, M. & Kaushik, S. Big data in healthcare: management, analysis and future prospects. *Journal of Big Data.* 6 (2019).
18. Proserpi M, Min JS, Bian J, Modave F, Big data hurdles in precision medicine and

- precision public health. BMC Medical Informatics and Decision Making; 2018; 18:139
19. 日本製薬工業協会 医薬品評価委員会 臨床評価部会「医薬品開発におけるデジタルバイオマーカー」(dBM)の利活用と要件 2022年4月
 20. Sim I, Mobile Devices and Health. N Engl J Med 2019; 381(10): 956-968.
 21. 日本製薬工業協会 医薬品評価委員会 電子化情報部会「ePRO の普及に向けて」2016年3月
 22. Bhaskaran K, Dos-Santos-Silva I, Leon DA, Douglas IJ, Smeeth L. Association of BMI with overall and cause-specific mortality: a population-based cohort study of 3.6 million adults in the UK. Lancet Diabetes Endocrinol. 2018; 6(12): 944-953.
 23. 大江和彦「医療における人工知能の活用と将来展望」日腎会誌 2017; 59 (7) : 1060-1063.
 24. Sebastian Raschka (著), Vahid Mirjalili (著), 福島真太郎 (監修), 株式会社クイープ (翻訳)「Python 機械学習プログラミング 達人データサイエンティストによる理論と実践 (impress top gear)」インプレス 2020
 25. Liu Y, Chen PHC, Krause J, Peng L. How to read articles that use machine learning: users' guides to the medical literature. JAMA. 2019;322(18):1806-1816.
 26. Rajkomar A, Dean J, Kohane I, Machine Learning in Medicine. N Engl J Med 2019; 380: 1347-58.
 27. 浅井義之「医療 AI とシステム医学の融合に向けて」山口医学 2018年 第67巻第4号 187-192
 28. Jain AK, Murty MN, Flynn PJ, Data Clustering: A Review, ACM computing surveys (CSUR), 1999.
 29. Toh TS, Dondelinger F, Wang D. Looking beyond the hype: Applied AI and machine learning in translational medicine. EBioMedicine 47 (2019) 607-615.
 30. 井村裕夫、稲垣暢也「先制医療 実現のための医学研究」実験医学 2015年
 31. Guo L, Li B. The Application of Medical Artificial Intelligence Technology in Rural Areas of Developing Countries. Health Equity 2018; 2:1

32. Hosny A, Aerts HJWL, Artificial intelligence for global health. *Science* 2019; 366 (6468) : 955–956.
33. Pharmaceuticals and Medical Devices Agency. Japanese adverse drug event report database. <<https://www.pmda.go.jp/safety/info-services/drugs/adr-info/suspected-adr/0003.html>> Accessed 26 April 2022.
34. Juniper EF, O'Byrne PM, Guyatt GH, Ferrie PJ, King DR, Development and validation of a questionnaire to measure asthma control. *Eur Respir J* 1999; 14: 902-907.
35. Juniper EF, Bousquetb J, Abetz L, Bateman ED, The GOAL Committee Identifying ‘well-controlled’ and ‘not well-controlled’ asthma using the Asthma Control Questionnaire. *Respir Med* 2006; 100: 616–621.
36. Onyedum CC, Ukwaja KN, Desalu OO, Ezeudo C. Challenges in the management of bronchial asthma among adults in Nigeria: A systematic review. *Ann Med Health Sci Res* 2013; 3: 324-9.
37. Mortimer K, Reddel HK, Pitrez PM, Bateman ED, Asthma management in low and middle income countries: case for change. *Eur Respir J* 2022; 60: 2103179.

第2章

ベイズ推定に基づく副作用の原因薬推定技術の開発

2.1 序論

副作用とは、医薬品による望ましくない有害な作用のことである[1]。副作用による院内死亡率は0.4-2.7%とされる[2]。副作用は入院患者数の約3.5%を占めている[3,4]。臨床薬学の一つの分野としてファーマコビジランスがある。ファーマコビジランスとは、副作用の検出、リスク評価、予防を目的とした臨床薬学である。ファーマコビジランスは医薬品安全性の向上に重要である。副作用が生じた場合には、原因薬剤の減量か中止が必要となる[5]。副作用が生じた際に、原因薬剤を特定することは副作用による悪影響を最小限に抑えるために最も重要な過程の1つである。

しかし、複数の候補薬の存在や時間的制約などのため、複数の投与薬剤から原因薬剤を特定することは困難である[6-8]。医療従事者の31%が「副作用の原因薬剤を特定するのは非常に困難であることが多い」と報告している[7]。Naranjo法、Jones法、Karch-Lasagna法などの方法論が、副作用の原因薬剤の特定に使用されてきた[9-11]。しかし、これらの方法は臨床現場における実用化に耐える精度を示さなかった。例えば、Naranjo法は正答率45%程度に留まる[12]。特に、ポリファーマシー（5剤/日以上）の服薬で定義されることが多い[13]）の場合には、候補が多いため原因薬剤を特定することはさらに難しい[13,14]。

ビッグデータ時代の現在、人工知能を用いたデータ駆動型研究が医療分野で広く実施されている[15]。ファーマコビジランスも、その例外ではない。大規模副作用報告データベースのデータマイニングによる副作用シグナル検出は、副作用管理の向上に有効である[17,18]。Reporting Odds Ratio (ROR)、Proportional Reporting Ratio (PRR)、Likelihood Ratio Test (LRT) 法など頻度主義の統計手法がファーマコビジランスに利用されている[19,20]。これらの統計的手法は不均衡分析 (Disproportionality analysis) と

呼ばれ、副作用と医薬品の関係性について評価する有効な方法として用いられてきた。近年、ベイズ統計が注目を集めている[21-23]。ベイズ統計では、ベイズの定理を用いることで確率論的な推論が可能である。日本の医薬品医療機器総合機構（Pharmaceuticals and Medical Devices Agency: PMDA）は、Japan Adverse Drug Event Report（JADER）と呼ばれる大規模な有害事象データベースを提供している。JADERは、これまで様々な研究で活用されてきた[13,25,26]。

本研究では、副作用ビッグデータである JADER データベースを活用し、副作用の原因薬を推定するためのベイズ統計モデルを開発した。最尤推定（Maximum Likelihood Estimation: MLE）の概念を用いて[27]、各投与薬剤が原因薬である可能性を表す尤度関数を定義した。尤度の高い薬剤が、原因薬剤の可能性が高いと判断した。同時に、山口大学医学部附属病院の電子カルテ記録における副作用報告データセットを用いて、開発したベイズ推定モデルの精度を検証した。

2.2 研究方法

本研究の解析概要を図1に示す。データ前処理とベイズ推定モデル開発には、Python 3.9.10 を使用した。解析において、Python パッケージである pandas 1.4.0、numpy 1.22.1、および scipy 1.5.0 を用いた。Receiver operation characteristics（ROC）曲線解析を含むすべての統計解析は、GraphPad Prism 8.4.3、GraphPad Software（米国）を用いて実行した。

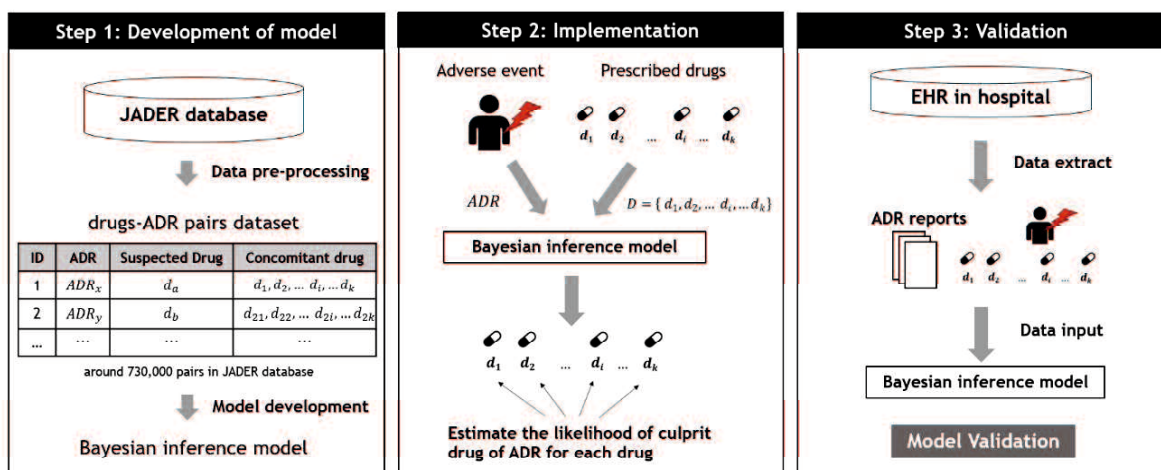


図1 ベイズ推定に基づく副作用の原因薬剤推定のワークフローを示す。ステップ1: JADER データベースを用いた原因薬剤を推定するためのベイズ統計モデルの開発の流れを示す。ステップ2: ベイズ統計モデルの適用方法を表す。副作用名と投与薬剤名の情報をベイズ推定モデルへ入力することで、それぞれの投与薬剤について原因薬である可能性を推定する。ステップ3: 病院の電子カルテから抽出された副作用報告を用いて、ベイズ推定モデルを検証した。

ADR, Adverse Drug Reactions; EMR, Electronic Medical Record; JADER, Japanese Adverse Drug Event Report。

2.2.1 学習データベース

モデルのパラメーター推定には PMDA が提供する JADER 副作用報告データベースを使用した。このデータベースでは 2004 年 4 月以降の日本国内における 73 万件以上の副作用報告が登録されている [24]。JADER データベースは完全匿名化されており、PMDA の Web サイト (<https://www.pmda.go.jp/safety/info-services/drugs/adr-info/suspected-adr/0003.html>) から無償でダウンロード可能である。JADER データベースは、(A) 患者背景テーブル; (B) 薬剤テーブル; (C) 有害事象テーブル; (D) 原疾患テーブルの 4 種類のテーブルから構成されている。すべてのテーブル (A-D) では、副作用報告ごとに一貫した ID 番号が付与されている。各テーブルは共通の ID 番号をもとに連結可能である。患者背景テーブル (A) では、性別、年齢、医薬品名が記録されている。薬剤テーブル (B) では、医薬品に関する情報が含まれている。薬剤・副作用間の因果関係の有・無に基づき、各薬剤は「被疑薬」、「相互作用」、「併用薬」の 3 種類のカテゴリへ分類される。本研究では、JADER データベースで「被疑薬」へ分類された薬剤を、副作用報告の原因薬剤と定義した。「併用薬」と分類された薬剤は、そのまま併用薬として用いた。JADER データベースにおいて「相互作用」へ分類された薬剤はほとんどなかったため (0.3%)、本研究では相互作用へ分類された薬剤は用いなかった。有害事象テーブル (C) では、副作用名と臨床転帰が登録されている。ID 番号をもとにテーブル B・C を結合することで、薬剤と副作用が対となったデータセットを作成した。この薬剤—副作用ペアを、学習データベースとして用いた。剤型 (経口、注射、吸入、貼付) を問わず、全ての薬剤を解析に使用した。

2.2.2 ベイズ推定モデルの構築

投与薬剤が副作用の原因である可能性を推定するベイズ統計モデルを開発した。JADER データベースをもとに作成した薬剤・副作用データセットに基づいて学習を行った。特定の副作用名を ADR と表記する。副作用報告における投与薬剤の組み合わせを $D = \{d_1, d_2, \dots, d_i, \dots, d_k\}$ と表記する。 d_i は i 番目の投与薬剤の名称、 k は投与さ

れた薬剤の数を表す。 $d_i = c$ は投与薬剤 d_i が原因薬剤であることを意味する。副作用 ADR が生じ、その患者の投与薬剤の組み合わせが $D = \{d_1, d_2, \dots, d_i, \dots, d_k\}$ であった場合、投与薬剤 d_i が副作用の原因薬剤である条件付き確率は次のように表される。

$$P(d_i = c | ADR, d_1 \in D, d_2 \in D, \dots, d_i \in D, \dots, d_k \in D)$$

最尤推定の概念の中では、原因薬剤は次のように推定される。

$$\hat{d}_i = \arg \max_{d_i} [P(d_i = c | ADR, d_1 \in D, d_2 \in D, \dots, d_i \in D, \dots, d_k \in D)]$$

上記の条件付き確率を、JADER データベースから導き出すことはできない。上記の条件付き確率を直接計算するためには、特定の副作用の発生率および各薬剤が投与される確率の情報が必要である。しかし、これらの周辺確率を JADER データベースから求めることはできない。課題設定を単純化するため薬剤相互作用を無視し、ベイズの定理を適用することで以下の近似式を得た（詳細については付録を参照）。

$$\begin{aligned} \arg \max_{d_i} [P(d_i = c | ADR, d_1 \in D, d_2 \in D, \dots, d_i \in D, \dots, d_k \in D)] &= \\ &\approx \arg \max_{d_i} [|(d_i = c | ADR)| / |(d_i \in D | ADR)|] \end{aligned}$$

ここで $|(d_i = c | ADR)|$ は、JADER データベース内において、 d_i が特定の副作用 ADR の原因薬剤と判定された副作用報告の件数を示している。 $|(d_i \in D | ADR)|$ は、JADER データベース内で特定の副作用 ADR において d_i が投与された副作用報告の全件数を表しており、原因薬剤か併用薬剤かは問わない。

比率 $|(d_i = c | ADR)| / |(d_i \in D | ADR)|$ を計算することで、各投与薬 d_i が原因薬である可能性を推定する。この比率を原因薬確率スコア（CPS: Culprit Probability Score）と定義した。CPS の降順で、投与薬剤 d_i ごとに原因薬剤である可能性の順位（以下、

RLC: Rank of the Likelihood for being a Culprit drug) を割り当てた。最も CPS が高い薬剤が RLC 1 位となる。CPS が同じ値の薬剤には、同じ RLC が割り当てられた。

2.2.3 検証方法

山口大学医学部附属病院の電子カルテから抽出した副作用報告を用いて、ベイズ推定モデルを検証した。この検証は、単施設で後方視的に実施した。2019年5月から2020年10月の期間の副作用報告を抽出した。検証のための副作用報告データの適格基準は以下の通りとした。(i) 副作用報告から6か月以内に少なくとも2つの薬剤を投与されていた; (ii) 副作用名、原因薬剤名の両方が JADER データベースに含まれていた; (iii) 原因薬剤が医師および薬剤師によって臨床的に特定された。(i) - (iii) のすべての適格基準を満たす副作用報告を精度検証に使用した。

検証データセット内に含まれる投与薬剤は「原因薬剤」か「併用薬剤」のいずれかに分類できる。しかし、本モデルを用いて原因薬剤の可能性を推定する際には「原因薬剤」か「併用薬剤」かのラベルは用いず、単に薬剤名の情報だけを用いた。

本モデルの推定精度を検証する際には、「原因薬剤」、「併用薬剤」のラベル情報を用いた。両群間で CPS と RLC の分布を比較した。Mann-Whitney 検定を用いて統計仮説検定を行った。原因薬剤の推定精度の評価のために ROC 曲線分析を行い、ROC 曲線下面積 (Area Under the Curve: AUC) を計算した。本研究は、山口大学医学部附属病院の倫理委員会によって承認され (承認番号 # 2021-144)、ヘルシンキ宣言の勧告に従って実施した。

2.2.4 不均衡分析に基づく統計的手法

ファーマコビジランスでは、LRT、PRR、ROR など不均衡分析に基づく統計的方法が良く用いられる。これらは、通常、未知の副作用シグナルを検出するために用いられる。一方、LRT、PRR、ROR を投与された薬剤間で比較することによって、副作用が生じた際の原因薬剤を推定し得るのではないかと考えた。検証データセットにおい

て薬剤ごとに、PRR と LLR (Log-Likelihood Ratio) の2つの比率統計量を計算した (算出方法は付録を参照のこと)。これら比率統計量の値が大きいほど、原因薬剤である可能性が高くなる。各副作用報告において、投与薬剤ごとの PRR 値、LLR 値に基づき順位を割り当てた。LLR 値と PRR 値を用いて原因薬剤の推定するために ROC 解析を実行し、われわれが提案するベイズ推定モデルの精度と比較した。ROR に関しては、PRR と定義が類似していること、並びに JADER データベース内に薬剤または副作用の報告が無い場合には ROR を定義できないことから、使用しなかった。

2.2.5 Bayesian Confidence Neural Network (BCPNN)

ファーマコビジランスでは上述の LRT、PRR、ROR など不均衡分析に基づく統計的方法に加えて、Bayesian Confidence Neural Network (BCPNN) というベイズ主義に基づく統計的手法も提案されている。BCPNN も、LRT、PRR、ROR と同様に、未知の副作用シグナル検出のための方法論である。不均衡分析に基づく統計的手法と同様に、BCPNN 法の統計量を投与薬剤間で比較することで、副作用の原因薬剤を推定できるのではないかと考えた。BCPNN では情報コンポーネント (Information component: IC) と呼ばれる統計量を考慮する (IC の算出方法は付録を参照のこと)。IC 値が大きいほど原因薬剤である可能性が高くなる。本研究では、IC の期待値を用いて、副作用の原因薬剤を推定する ROC 解析を実行し、われわれが提案するベイズ推定モデルの精度と比較した。

2.2.6 Precision-Recall (PR) 解析

本研究では、副作用の原因薬剤の推定精度を評価するため、ROC 解析に加えて、Precision-Recall (PR) 解析を実施した。Precision は陽性的中率、Recall は感度のことである。本研究で PR 解析を実施した理由を以下に述べる。本研究で検証用に用いた計 67 件の副作用報告-薬剤データセットは、合計で 67 剤の原因薬剤、合計で 1,320

剤の併用薬剤を含んでいた。原因薬剤の割合は 4.8 %、併用薬剤の割合は 95.2 %と、各投与薬剤のクラスラベル（原因薬剤または併用薬剤）の割合が極端に不均衡（imbalanced）になっている。このように、クラスラベルの割合が不均衡なデータセット（以下、不均衡データ）においては、ROC-AUC だけではモデルの精度評価は十分でない。不均衡データにおいては、陽性的中率が低くても陰性的中率さえ高ければ ROC-AUC の値は高くなり、見かけ上、優れたモデルであると評価されてしまうからである。以上より、本研究では ROC 解析だけでなく PR 解析を実施し、原因薬剤の推定精度を手法（LRT、PRR、BCPNN、CPS）ごとに比較するための PR 曲線を作成した。

2.3 解析結果

2.3.1 検証に用いた副作用報告

モデルを検証するために、山口大学医学部附属病院の電子カルテから合計 159 件の副作用報告を抽出した。内 89 件は、投与薬剤が 1 剤しかなかったため除外した。3 件の副作用報告に関しては、実際の副作用の原因薬剤が JADER データベースに含まれていなかったために除外した。最終的に、67 件の副作用報告がモデル検証のために登録された（図 2）。合計で 1,387 剤の薬剤（67 剤の原因薬、1,320 剤の併用薬剤）が含まれていた。副作用報告ごとの投与薬剤数のヒストグラムを図 3 に示した。投与薬剤数の中央値は、副作用報告 1 件あたり 11.0 剤（四分位範囲: 6.0-35.0 剤）であった。副作用内訳は表 1 に示す。副作用内訳は、System Organ Classification（World Health Organization Adverse Reaction Terminology）に基づいて分類した。呼吸器系障害が最も多い有害事象であり（50.7%）、次いで皮膚障害（8.9%）、消化器障害（7.5%）となっていた。

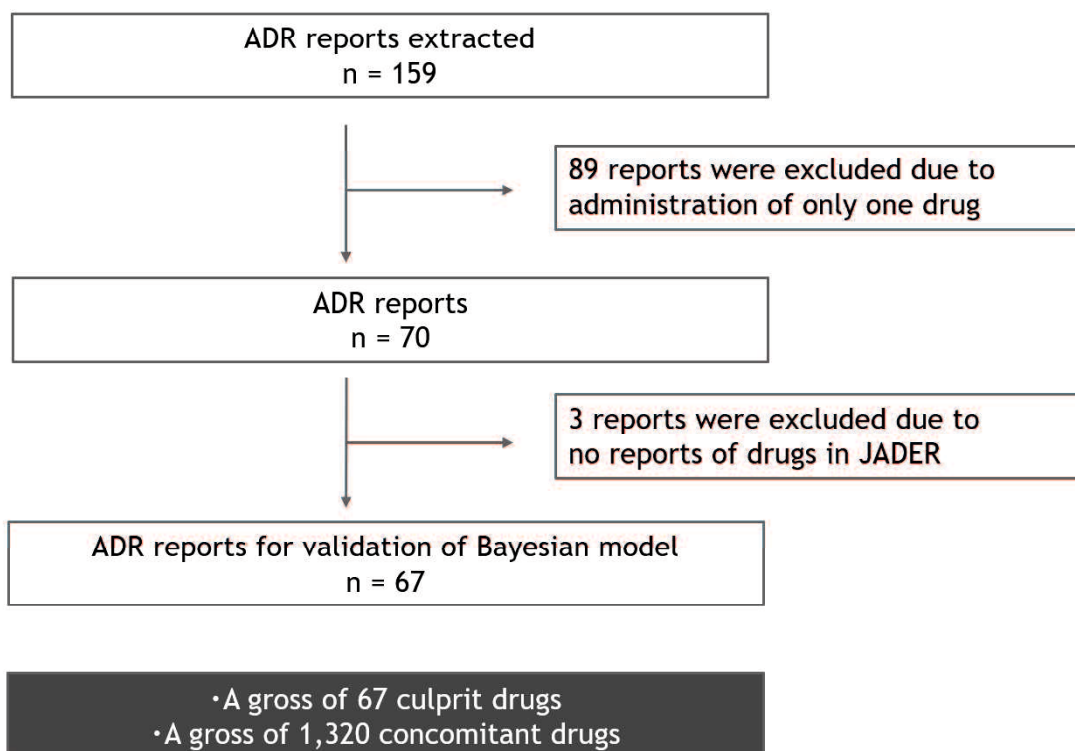


図2 ベイズ推定モデル検証のための副作用報告のフローチャートを示す。モデル検証のために、計159件のレポートが山口大学医学部附属病院の電子カルテシステムから抽出された。最終的に、副作用67件、総計1,387剤（67剤の原因薬剤および1,320剤の併用薬剤）がベイズ推定モデルの検証のために登録された。ADR, Adverse Drug Reactions; EMR, Electronic Medical Record。JADER, Japanese Adverse Drug Event Report。

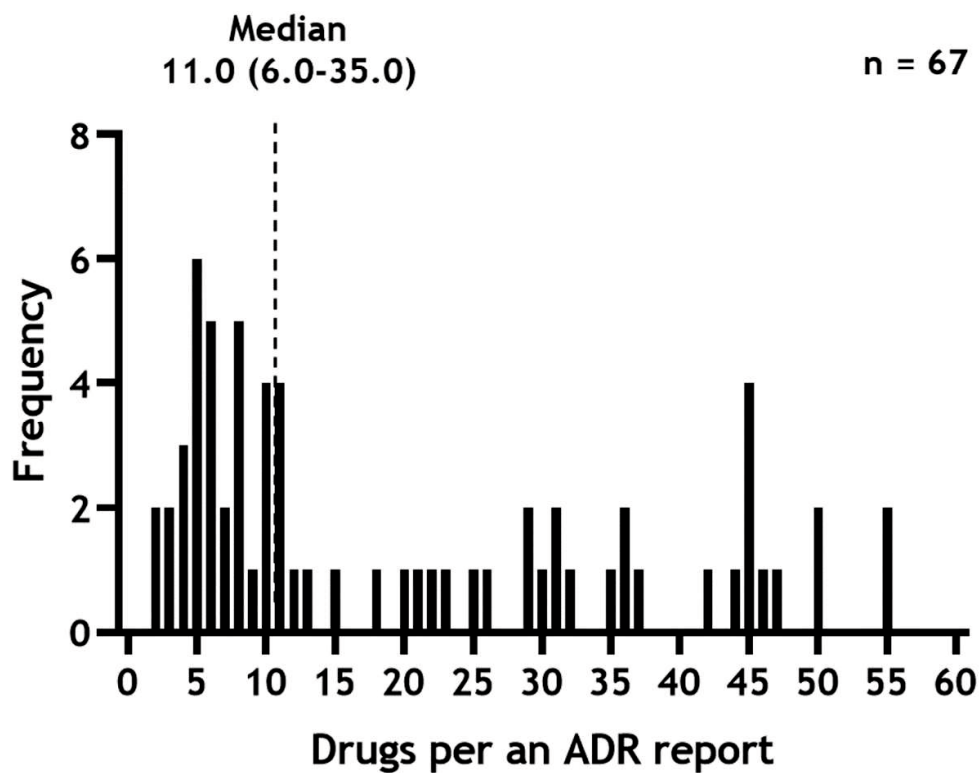


図3 副作用報告ごとの投与薬剤数のヒストグラムを示す。副作用報告1件あたりの投与薬剤数の中央値は11.0（四分位範囲: 6.0-35.0）であった。ADR, Adverse Drug Reaction。

Table 1 Clinical manifestation of ADR reports

WHO-ART SOC	ADR reports (%) <i>n</i> = 67
Skin	6 (8.9)
Anaphylaxis	1 (1.5)
Severe cutaneous adverse reactions	2 (3.0)
Rash	3 (4.5)
Gastrointestinal	5 (7.5)
Hepatobiliary	2 (3.0)
Nervous system	1 (1.5)
Blood and lymphatic system	2 (3.0)
Endocrine	4 (6.0)
Respiratory	34 (50.7)
Heart-related	4 (6.0)
Renal and urinary	2 (3.0)
Others	7 (10.4)

ADR, Adverse Drug Reaction; SOC, System Organ Classification; WHO-ART, World Health Organization Adverse Reaction Terminology.

表1 検証データセットにおける副作用の内訳を示す。副作用内訳は、System Organ Classification (World Health Organization Adverse Reaction Terminology) に基づいて分類した。呼吸器系障害が最も多い有害事象であり (50.7%)、次いで皮膚障害 (8.9%)、消化器障害 (7.5%) となっていた。ADR, Adverse Drug Reaction; SOC, System Organ Classification; WHO-ART, World Health Organization Adverse Reaction Terminology

2.3.2 ベイズ推定モデル適用例

山口大学医学部附属病院において、開発したベイズ推定モデルを適用した2症例の副作用報告の結果を表2に示す。

症例1は薬剤性肺障害の症例である。この患者は7種類の薬剤（エベロリムス、ロスバスタチン、エソメプラゾール、ウルソデオキシコール酸、インスリンデグルデク、レバミピド、インスリンアスパルト）を投与されていた。ベイズ推定モデルはRLCの1位はエベロリムスであると判断した。次いで、ロスバスタチン、エソメプラゾール、ウルソデオキシコール酸、その他の薬剤の順となっていた。この副作用報告では、実際にエベロリムスが原因薬剤であった。すなわち、RLCが1位の薬剤と、臨床的に原因薬剤と判断された薬剤は一致していた。この症例では、RLCが1位の薬剤のCPSは、RLCが2位の薬剤のCPSよりもはるかに高かった。

症例2は薬剤性肝障害の症例である。症例2では、11種類の薬剤が投与されていた。RLCが2位の薬剤が実際の原因薬剤であった。RLCが1位の薬剤と、臨床的に原因薬剤と判断された薬剤は一致しなかった。それでも、RLCが1位の薬剤（プラスグレル）と2位の薬剤（ロスバスタチン）のCPSの差は非常に小さかった。

Table 2 An application example of the Bayesian Inference model to the ADR reports

A name of administered drugs	Probabilistic score	Rank	Actual involvement
Case 1 (drug-induced lung injury)			
Everolimus	0.995	1	Causative
Rosuvastatin	0.166	2	Concomitant
Esomeprazole	0.094	3	Concomitant
Ursodeoxycholic acid	0.092	4	Concomitant
Insulin degludec	0.063	5	Concomitant
Rebamipide	0.045	6	Concomitant
Insulin aspart	0.032	7	Concomitant
Case 2 (drug-induced liver injury)			
Prasugrel	0.350	1	Concomitant
Rosuvastatin	0.347	2	Causative
Ezetimibe	0.278	3	Concomitant
Esomeprazole	0.253	4	Concomitant
Aspirin	0.187	5	Concomitant
Carvedilol	0.180	6	Concomitant
Lubiprostone	0.170	7	Concomitant
Enalapril	0.127	8	Concomitant
Potassium chloride	0.050	9	Concomitant
Magnesium oxide	0.047	10	Concomitant
Glycerin	0.000	11	Concomitant

表2 ベイズ推定モデルによって算出された確率スコアは、副作用の原因薬剤である可能性の大きさを表す。症例1は薬剤性肺障害の症例である。ベイズ推定モデルは、エベロリムスである可能性が最も高いと出力した。この症例では実際にエベロリムスが原因薬剤であった。症例2は薬剤性肝障害の症例である。症例2では、11種類の投与薬剤があり、RLCが2位の薬剤が実際の原因薬であった。それでも、RLCが1位の薬剤と2位の薬剤でCPSの差は非常に小さかった。ADR, Adverse Drug Reaction; CPS, Culprit Probabilistic Score。

2.3.3 性能評価

検証データセットに関してベイズ推定モデルを適用した。図 4A と 4B は、それぞれ、実際の原因薬剤と併用薬剤の CPS のヒストグラムを示している。原因薬剤において、より高い CPS を示した。原因薬剤群における CPS の中央値（四分位範囲）は 0.95（0.67-0.99）であった。一方、併用薬剤における CPS は低かった。併用薬剤における CPS の中央値（四分位範囲）は 0.03（0.00-0.14）であった。原因薬剤群と併用薬剤群の間で、CPS に有意差がみられた（ $P < 0.0001$ ）（図 4C）。

RLC についても検討した。実際の原因薬剤の 76.1%は RLC が 1 位であった（図 4D）。原因薬剤群における RLC の中央値（四分位範囲）は 1.0（1.0-4.0）であった。一方で、ベイズ推定モデルが併用薬剤を RLC 1 位と評価したのは、わずか 1.2%であった（図 4E）。併用薬剤群における RLC の中央値（四分位範囲）は 16.0（6.0-24.0）であった。RLC についても、原因薬剤群と併用薬剤群の間に統計的な有意差がみられた（ $P < 0.0001$ ）（図 4F）。

ROC 曲線を作成し、ベイズ推定モデルと不均衡分析に基づいた他の統計的手法（PRR、LRT）とで推定精度を比較した。ベイズ推定モデル、LLR 値、および PRR 値、それぞれの ROC-AUC（95%信頼区間）は、0.93（0.88-0.97）、0.87（0.81-0.93）、0.83（0.77-0.89）であった（図 5A）。ベイズ推定モデルによって原因薬剤を推定する感度は、LLR および PRR における感度よりも一貫して高かった（図 5B）。全副作用報告（67 件）において、RLC が上位 3 位までに原因薬剤の 88.1%が含まれていた（図 5B）。

同時に、BCPNN の統計量である IC の期待値を用いて、副作用の原因薬剤を推定する ROC 曲線を作成し、我々が提案するベイズ推定モデルと推定精度を比較した（図 6a）。IC 値による副作用の原因薬剤の推定精度は、ROC-AUC（95%信頼区間）0.86（0.80-0.92）であった。本研究で開発したベイズ推定モデルが出力する CPS の ROC-AUC（95%信頼区間）は 0.93（0.88-0.97）であり、IC を上回っていた。

第2章 ベイズ推定に基づく副作用原因薬推定技術の開発

本研究で用いたデータセットは不均衡データであったことから、ROC解析に加えて原因薬剤の推定するためのPR解析を実施した。図6bに手法ごとのPR曲線を示す。PR曲線においても、本モデルが出力するCPSが最も優れていた。

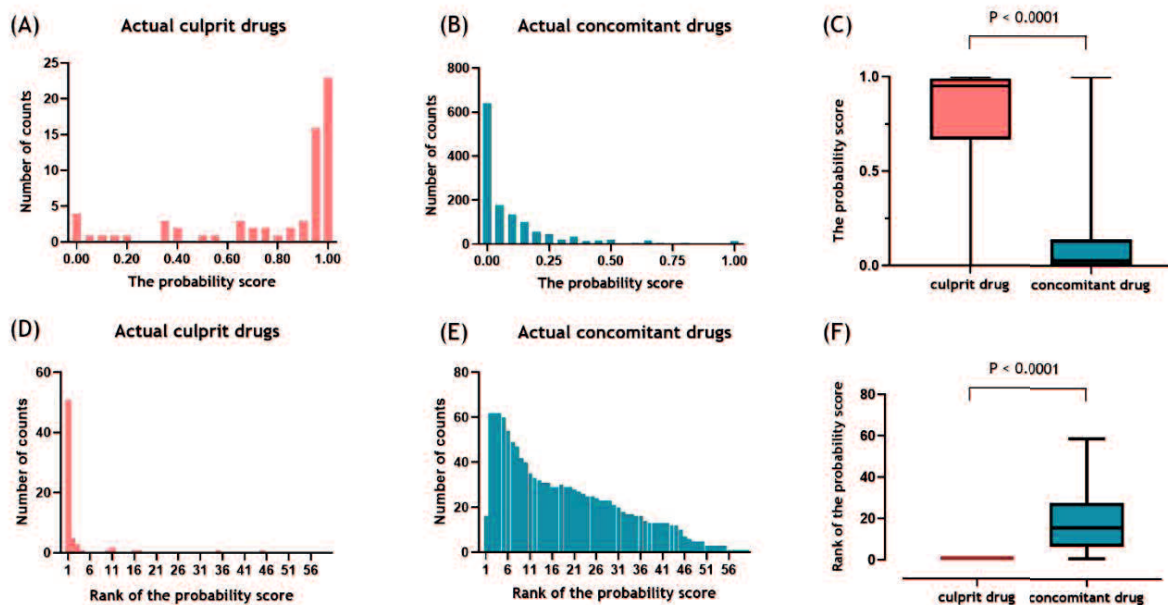


図4 検証においてモデルが出力した確率スコアとランクについて示す。実際の原因薬剤のグループ (A)、実際の併用薬のグループ (B) それぞれの群において、ベイズ推定モデルが出力した確率的スコアのヒストグラムを示す。(C) 実際の原因薬グループにおける確率スコアの中央値 (四分位範囲) は 0.95 (0.67–0.99) である一方、実際の併用薬群では 0.03 (0.00–0.14) であった ($p < 0.0001$)。 (D) 実際の原因薬剤のグループ (E) 実際の併用薬のグループそれぞれにおいて、ベイズ推定モデルが出力した順位のヒストグラムを示す。(F) 実際の原因薬剤グループにおける順位の中央値 (四分位範囲) は 1.0 (1.0–4.0) である一方、実際の併用薬群では 16.0 (6.0–24.0) であった ($p < 0.0001$)。統計仮説検定は Mann-Whitney 検定を用いた。

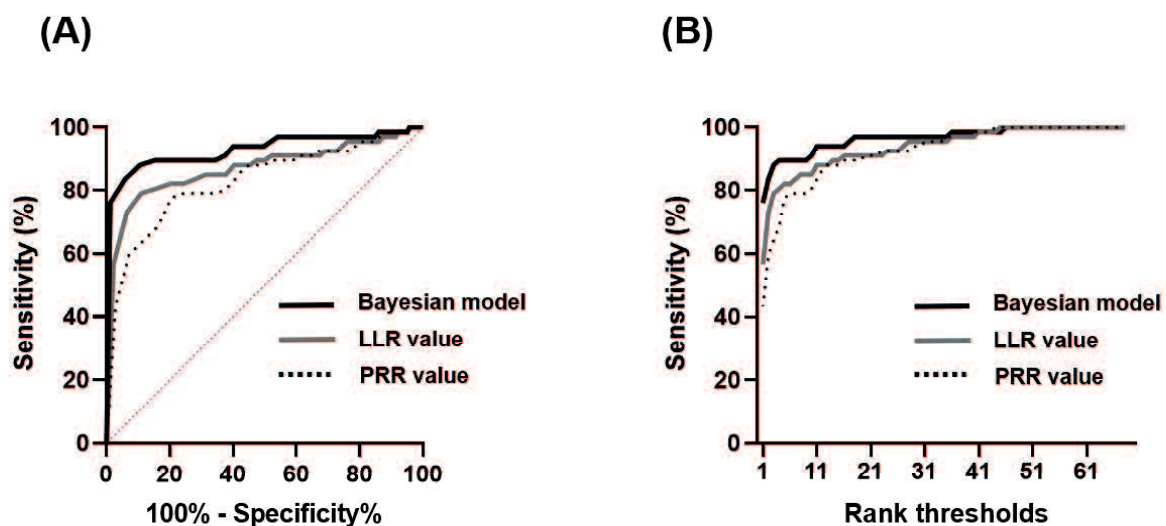


図5 副作用の原因薬の推定するためのROC解析を行った (A)。ベイズ推定モデル、LLRスコア、およびPRRスコアの順位を用いたROC解析では、それぞれAUC (95%信頼区間) 0.93 (0.88–0.97)、0.87 (0.81–0.92)、および0.83 (0.77–0.89)の精度で原因薬剤を推定した。ベイズ推定モデル、LLR、PRRそれぞれにおいて、順位閾値ごとの原因薬剤の推定感度を示す (B)。LLR, Log-likelihood ratio; PRR, proportional reporting ratio; ROC, receiver operating characteristic。

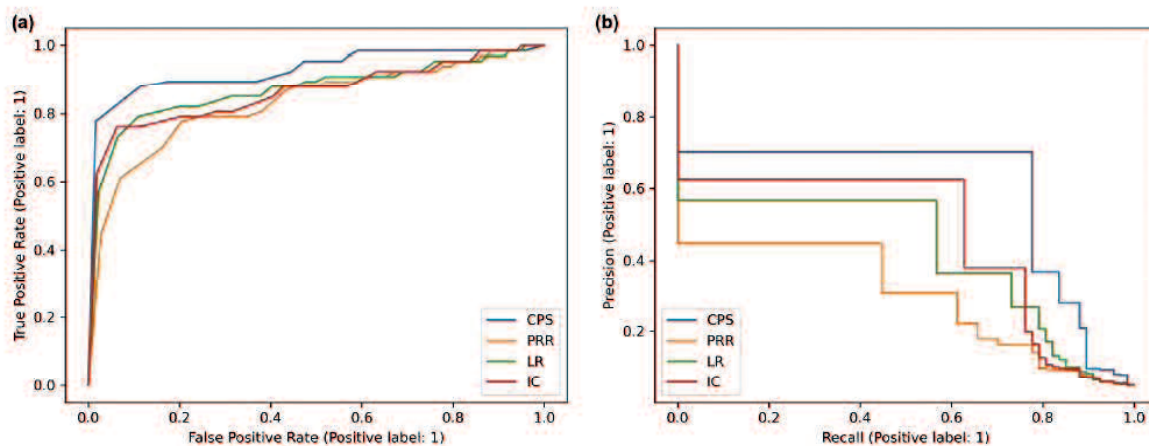


図6 (a) モデルが出力する順位を用いて原因薬剤を推定する ROC 解析を示す。CPS、PRR、LR、IC による原因薬剤推定の精度は、それぞれ AUC (95%信頼区間) は 0.93 (0.88–0.97) , 0.83 (0.77–0.89)、0.87 (0.81–0.92)、0.86 (0.80–0.92) であった。(b) モデルが出力する順位を用いて原因薬剤を推定する PR 解析について示す。青線は CPS、オレンジ線は PRR、緑線は LR、赤線は IC を示す。CPS, culprit probability score; PRR, proportional reporting ratio; LR, log-likelihood ratio; IC, information component ; PR, precision recall。

2.4 考察

本研究では、JADER 副作用報告データベースに基づいて、副作用の原因薬剤の同定を支援するベイズ統計モデルを開発した。モデル精度の検証によって、我々のモデルは、臨床現場での実用に耐えうる高精度（AUC 0.93：95% CI 0.88-0.97）で副作用の原因薬剤を推定できた。

過去の研究では、副作用の原因薬を特定するための免疫学的方法やデータ駆動型アプローチに基づく方法が報告されている。投与薬剤に対する末梢血液中の免疫細胞の活性化を評価する *in vitro* の方法（例：リンパ球形質転換試験）が、原因薬剤の推定に有用であることが示されている[28]。しかし、この方法は免疫機序で生じる副作用のみにおいて有用な方法である。免疫機序以外のメカニズムで生じる副作用については適用できない。データ駆動型アプローチとして、医薬品成分の化学構造を深層学習し、副作用の原因薬剤を推定するニューラルネットワーク技術が開発された[25]。しかし、このモデルは特定の1種類の副作用（重篤な皮膚障害）のみを対象としていた。実際の臨床現場においては、副作用の種類が数多くあること、投与薬剤の組み合わせが無数にあることを鑑みると、投与薬剤の中から副作用の原因薬剤を特定する方法論が必要となる。各投与薬剤について副作用原因薬剤の可能性を推定することで、薬剤の減量または中止の手がかりとする。我々の知る限り、本研究では、副作用の原因薬剤を推定するベイズ推定モデルを初めて開発した。大規模な副作用報告データベースに基づき、副作用の原因薬剤である可能性を推定できる。

本研究で開発したベイズ推定モデルと、LRT 法、PRR 法、BCPNN 法とで、副作用の原因薬推定の精度を、ROC-AUC を用いて比較した。検証の結果、本研究で開発したベイズ推定モデルは、より優れた精度を示した。LRT 法、PRR 法、BCPNN 法など従来のファーマコビジランスにおける方法論は、未知の副作用シグナルを検出することを目的としている。これらの方法は、個別の副作用報告において、原因薬剤を特定することは意図していない。我々のベイズ推定モデルは、薬剤の副作用への関与の有無を表す「被疑薬」と「併用薬」の2種類のラベルを使用しているという利点があ

る。確率計算過程で2種類のラベルの違いを考慮している。一方、LRT法、PRR法、BCPNN法では、計算過程で「被疑薬」と「併用薬」の2種類のラベルが区別されていない。以上より、投与薬剤の中から未知の原因薬剤を推定する場合には、過去に観測された副作用報告データベースの「被疑薬」と「併用薬」のラベル情報を考慮することが有効であると考えた。

本研究で原因薬剤の推定精度の検証に用いた副作用-薬剤データセットは、不均衡データ（被疑薬が4.8%、併用薬が95.2%）であったことから、ROC-AUCのみでは推定精度の評価は不十分である可能性があった。ROC-AUCでは評価できない陽性的中率を含めた精度を評価するために、原因薬剤を推定するPR解析を実施した。PR解析においても、ROC解析と同様、本モデルが出力するCPSが最も高精度を示した。しかし、統計量に基づいて被疑薬の可能性について順位付けを行う際には、期待される推定精度のベースライン確率が、投与薬剤数によって変動することに留意する必要がある[29]。投与薬剤数が増加するにつれて原因薬剤である可能性のチャンスレベル（事象が偶然生じる確率）が低下するためである。以上から、本研究においてPR曲線を解釈する際には注意が必要である。

本研究は、過去の研究で報告された、副作用との因果関係評価を行うベイジアンネットワーク法の限界の克服につながる[23]。この先行研究では、専門家の事前知識や、既知の副作用・医薬品間の因果関係の情報をもとに、ベイジアンネットワークの構造を決定した。ベイジアンネットワークは変数間の因果関係を検討するための強力なツールである。しかし、データ量や変数の数が多い場合、ベイジアンネットワークの学習には膨大な時間がかかり、CPUを高度に使用する。現実的な時間内に学習を完了するためには、複数の仮定、専門家の事前知識、高度な計算能力が必要となる。一方、本研究のベイズ推定モデルで使用されるすべての条件付き確率は、データのみに基づいて計算され、人間の主観的判断が入る余地はない。本研究におけるベイズ推定モデルは、730,000件を超える副作用報告に基づいて開発された。Intel Core i7 CPUを内蔵したノートPCで2分以内に学習が完了する。

本研究で提案するベイズ推定モデルは、副作用名、投与薬剤名、原因薬剤名の情報

を含む副作用報告データベースがあれば開発することができる。すなわち、米国 FDA の有害事象報告システム (FAERS: FDA Adverse Event Report system) や、欧州連合 (EU) が運営する有害事象報告システム EudraVigilance データベースなど、JADER 以外の大規模副作用報告データベースを用いてベイズ推定モデルを開発することも可能である[30,31]。JADER データベースに基づく原因薬剤の推定モデルでは、副作用発症の感受性の人種差によるバイアスが生じる可能性がある[32]。日本人以外の患者には本研究で構築した推定モデルを応用できない可能性がある。しかし、JADER 以外の適切なデータベースを用いてモデルを開発することで、人種差によるバイアスが軽減され得る。例えば FAERS を用いたベイズ推定モデルは、米国患者における副作用の原因薬剤の推定において、有用であるかもしれない。

本研究で開発したベイズ推定モデルは、投与薬剤を「原因薬剤」か「併用薬剤」の2種類のクラスに分類するのではなく、すべての投与薬剤について原因薬剤である可能性 (Culprit Probability Score: CPS) を推定する。続いて、CPS の値に従って全投与薬剤を順位付けし、その結果を医療従事者に提供する。医療従事者は、臨床情報 (発症までの期間、薬剤の中止・再投与後の副作用の状態、薬剤の添付文書など) を追加で考慮することで、より効果的に原因薬剤を特定することができる。通常、副作用が生じた場合には、医薬品安全性情報を注意深く調査し、原因薬剤の可能性を検討する必要がある。しかし、臨床現場では時間的制約から、全投与薬剤について医薬品安全性情報を参照するのは現実的に難しい場合がある[33-35]。

本モデルは、投与薬剤の中から1、2剤の原因薬候補を提案するだけでなく、全投与薬剤について原因薬剤である確率を推定する。これは、原因薬剤の誤推定を防ぐのに役立つかもしれない。表2の症例2に示すように、ベイズ推定モデルが複数の投与薬剤に対して同様の高いスコアを返した場合、それら複数の薬剤を原因薬剤の有力な候補と見なすことができる。表2の症例2では、原因薬剤である可能性の順位 (RLC) が2位の薬剤が実際の原因薬剤であった。RLC 1位以外が原因薬剤である可能性について留意する必要がある。それでも、この症例では RLC が1位と2位の薬剤の CPS は、ほとんど同程度の値であった。

我々のベイズ推定モデルを副作用の原因薬剤の推定に用いる際には、以下3つの点について留意することが望ましい。第1に、RLCとCPSの両方を参照することが望ましい。第2に、モデルの出力に加えて、発症までの時間、相互作用、患者の臨床的特徴（年齢、性別）などの他の臨床情報を考慮することが有用である。複数の薬剤が同様に高いCPSを示す場合や、モデル出力結果と臨床情報が大きく乖離する場合は、特に臨床情報が重要になる。例えば、RLCが1位の薬剤に関して、投与開始からADR発症までの期間が、好発時期（薬剤・副作用の組み合わせごとで多様）と大きく乖離する場合には、RLC1位の薬剤を即座に原因薬剤と特定することは合理的でない。このような場合、臨床情報も考慮し、総合的に原因薬剤を判断する必要がある。第3に、稀な副作用については、学習データのサンプルサイズが小さくなる。ベイズ推定モデルの精度が低下する可能性があるため、稀な副作用についてはモデル出力結果を慎重に解釈する必要がある。

この研究の限界は以下の通りである。第1に、承認されたばかりの医薬品は、JADERデータベース内の医薬品の報告件数が少ないため、原因薬剤の可能性が過小評価される可能性がある。しかし、ベイズ推定モデルとJADERデータベースの定期的な更新によって、この限界は改善できる。第2に、自発的な報告データベースを使用しているため、報告バイアスを排除することはできない。JADERデータベースには、より重篤な副作用報告のレポートが含まれやすいことが知られている[36]。同時に、副作用報告データベースに内在するWeber効果、Notoriety効果など報告バイアスの影響も無視できない。一方で、付録に要約する通り、我々が提案する予測推論モデルは、別のデータベースから得られた副作用報告に基づいて事後分布を更新することが出来る。ある特定の医療施設でのデータベースをもとに、ベイズ更新を適用することで、JADERなどの副作用報告データベースに本質的に存在するバイアスの影響を軽減できるかもしれない。第3に、本モデルは、薬剤投与量、副作用発症までの期間、相互作用を考慮していない。薬剤は、規定の薬剤投与量レベルで患者に投与される。薬剤投与量が多いほど、副作用が発生しやすい傾向にある。この研究では、ベイズ推定モデルは薬剤投与下での副作用の条件付き確率に焦点を当てている。本モデルでは副作用

発生率は考慮していない。したがって、薬剤の投与量が、副作用の原因薬剤推定の精度に与える影響は限定的であると考えられる。副作用発症までの期間、相互作用、および薬剤投与量に関する情報は、JADER データベースに記録されている。今後、これらの情報をベイズ推定モデルのパラメーターとして含めることで、推定精度の向上につながるかもしれない。第4に、本モデルは JADER データベースに基づいている。日本人以外の患者に対して、JADER データベースに基づく本モデルをそのまま適用することは難しいかもしれない。

2.5 結語

副作用の原因薬である可能性を確率論的に推定するベイズ推定モデルを開発した。精度検証では、臨床現場での実用に耐える高い精度（AUC 0.93：95% CI 0.88-0.97）を示した。このモデルを臨床現場へ応用することで、副作用による健康被害を最小限に抑え、副作用管理を向上・効率化できる。今後、投与開始から副作用発症までの期間や、相互作用、患者の臨床的特徴（年齢・性別）などの、他の臨床情報を考慮しモデルを改良することで、さらに高い推定精度を得ることができるとも思われる。

第2章は、Hamada K, Abe T, Kouda K, Nakatsui M, Kitahara T, Matsunaga K, Asai Y, Estimating Culprit Drugs for Adverse Drug Reactions Based on Bayesian Inference, Clin Pharmacol Ther. 2023 May;113(5):1117-1124. doi: 10.1002/cpt.2867. Epub 2023 Feb 22. に原著論文として発表したもの、および Hamada K, Abe T, Kouda K, Nakatsui M, Kitahara T, Matsunaga K, Asai Y. Reply to “Questioning Bayesian inference-based models for estimating Culprit Drugs for Adverse Drug Reactions.” Clin Pharmacol Ther. 2023 (in press) に letter to the editor として発表したものからデータを引用して作成した。

本技術は現在、特許出願中である（特願 2021-151548 副作用の被疑薬推定支援装置と被疑薬推定支援プログラムと被疑薬推定支援方法：濱田和希、安部武志、幸田恭治、

浅井義之)。

2.6 参考文献

1. Edwards, I.R. & Aronson, J.K. Adverse drug reactions: definitions, diagnosis, and management. *The Lancet*. 356, 1255-1259 (2000).
2. Khalil, H. & Huang, C. Adverse drug reactions in primary care: a scoping review. *BMC Health Serv Res*. 20, 5 (2020).
3. Aagaard, L., Strandell, J., Melskens, L., Petersen, P.S. & Holme Hansen, E. Global patterns of adverse drug reactions over a decade: analyses of spontaneous reports to Vigibase. *Drug Saf*. 35, 1171-1182 (2012).
4. Stausberg, J. International prevalence of adverse drug events in hospitals: an analysis of routine data from England, Germany, and the USA. *BMC Health Serv Res*. 14, 125 (2014).
5. Beninger, P. Pharmacovigilance: An Overview. *Clin Ther*. 40, 1991-2004 (2018).
6. Bose, R. et al. Methods for identifying culprit drugs in cutaneous drug eruptions: A scoping review. *J Cutan Med Surg* 12034754211027509 (2021)
7. Hughes, M.L. & Weiss, M. Adverse drug reaction reporting by community pharmacists- The barriers and facilitators. *Pharmacoepidemiol Drug Saf*. 28, 1552-1559 (2019).
8. Mascolo, A. et al. Can causality assessment fulfill the new European definition of adverse drug reaction? A review of methods used in spontaneous reporting. *Pharmacol Res*. 123, 122-129 (2017).
9. Jones, J.K. Adverse drug reactions in the community health setting: approaches to recognizing, counseling, and reporting. *Fam Community Health*. 5, 58-67 (1982).
10. Karch, F.E. & Lasagna, L. Toward the operational identification of adverse drug reactions. *Clin Pharmacol Ther*. 21, 247-254 (1977).
11. Naranjo, C.A. et al. A method for estimating the probability of adverse drug reactions. *Clin Pharmacol Ther*. 30, 239-245 (1981).

12. Garcia-Cortes, M. et al. Evaluation of naranjo adverse drug reactions probability scale in causality assessment of drug-induced liver injury. *Aliment Pharmacol Ther.* 27, 780-789 (2008).
13. Mabuchi, T., Hosomi, K., Yokoyama, S. & Takada, M. Polypharmacy in elderly patients in Japan: Analysis of Japanese real-world databases. *J Clin Pharm Ther.* 45, 991-996 (2020).
14. Masnoon, N., Shakib, S., Kalisch-Ellett, L. & Caughey, G.E. What is polypharmacy? A systematic review of definitions. *BMC Geriatr.* 17, 230 (2017).
15. Hussain, R. Big data, medicines safety and pharmacovigilance. *J Pharm Policy Pract.* 14, 48 (2021).
16. Ventola, C.L. Big Data and Pharmacovigilance: Data Mining for Adverse Drug Events and Interactions. *P T.* 43, 340-351 (2018).
17. Chen, J.H. & Asch, S.M. Machine Learning and Prediction in Medicine - Beyond the Peak of Inflated Expectations. *N Engl J Med.* 376, 2507-2509 (2017).
18. Dash, S., Shakyawar, S.K., Sharma, M. & Kaushik, S. Big data in healthcare: management, analysis and future prospects. *Journal of Big Data.* 6 (2019).
19. Huang, L., Guo, T., Zalkikar J.N., & Tiwari, R.C. A Review of Statistical Methods for Safety Surveillance. *Ther Innov Regul Sci.* 48, 98-108 (2014)
20. Huang, L., Zalkikar J.N., & Tiwari, R.C. A Likelihood Ratio Test Based Method for Signal Detection With Application to FDA's Drug Safety Data. *J Am Stat Assoc.* 106, 1230-1241 (2011)
21. Barmaz, Y. & Menard, T. Bayesian Modeling for the Detection of Adverse Events Underreporting in Clinical Trials. *Drug Saf.* 44, 949-955 (2021).
22. Cai, R. et al. Identification of adverse drug-drug interactions through causal association rule discovery from spontaneous adverse event reports. *Artif Intell Med.* 76, 7-15 (2017).
23. Rodrigues, P.P., Ferreira-Santos, D., Silva, A., Polonia, J. & Ribeiro-Vaz, I. Causality assessment of adverse drug reaction reports using an expert-defined Bayesian network. *Artif Intell Med.* 91, 12-22 (2018).

24. Pharmaceuticals and Medical Devices Agency. Japanese adverse drug event report database. <<https://www.pmda.go.jp/safety/info-services/drugs/adr-info/suspected-adr/0003.html>>. Accessed 26 April 2022.
25. Ambe, K., Ohya, K., Takada, W., Suzuki, M. & Tohkin, M. In Silico Approach to Predict Severe Cutaneous Adverse Reactions Using the Japanese Adverse Drug Event Report Database. *Clin Transl Sci.* 14, 756-763 (2021).
26. Tsuchiya, M., Obara, T., Sakai, T., Nomura, K., Takamura, C. & Mano, N. Quality evaluation of the Japanese Adverse Drug Event Report database (JADER). *Pharmacoepidemiol Drug Saf.* 29, 173-181 (2020).
27. Myung, I.J. Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology.* 47, 90-100 (2003).
28. Srinoulprasert, Y. Lymphocyte transformation test and cytokine detection assays: Determination of read out parameters for delayed-type drug hypersensitivity reactions. *J Immunol Methods.* 496, 113098 (2021).
29. Saito T, Rehmsmeier M (2015) The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLOS ONE* 10(3): e0118432. <https://doi.org/10.1371/journal.pone.0118432>.
30. U.S. Food and Drug Administration (FDA). <<http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/default.htm>>. Accessed 26 April 2022.
31. European Medicines Agency. EudraVigilance. <<https://www.ema.europa.eu/en/human-regulatory/research-development/pharmacovigilance/eudravigilance>>. Accessed 26 April 2022.
32. McDowell, S.E., Coleman, J.J. & Ferner, R.E. Systematic review and meta-analysis of ethnic differences in risks of adverse reactions to drugs used in cardiovascular medicine. *BMJ.* 332, 1177-1181 (2006).
33. Campbell, C., Braund, R. & Morris, C. A mixed methods study on medicines information

- needs and challenges in New Zealand general practice. *BMC Fam Pract.* 22, 150 (2021).
34. Rambaran, K.A., Huynh, H.A., Zhang, Z. & Robles, J. The Gap in Electronic Drug Information Resources: A Systematic Review. *Cureus.* 10, e2860 (2018).
35. Reppe, L.A. et al. Factors associated with time consumption when answering drug-related queries to Scandinavian drug information centres: a multi-centre study. *Eur J Clin Pharmacol.* 70, 1395-1401 (2014).
36. Matsuda, S. et al. Bias in spontaneous reporting of adverse drug reactions in Japan. *PLoS One.* 10, e0126413 (2015).

第3章

喘息患者の気流閉塞の急速進行を予測する技術

3.1 序論

喘息は気道炎症で特徴づけられる呼吸器疾患である[1]。多くの喘息患者の気道炎症は、吸入ステロイド治療によって制御されうる。しかし、適切な吸入治療にも関わらず、気道炎症が持続する患者集団が存在する。気道炎症の持続は気道リモデリングにつながる。その結果、不可逆的な気流閉塞をもたす。気流閉塞は対標準1秒量（percent predicted FEV₁ : %FEV₁） $\leq 80\%$ で定義され、重症喘息の基準の1つである[6]。不可逆的な気流閉塞が起きると、1秒量（Forced Expiratory Volume in 1 Second : FEV₁）の低下が観測される。FEV₁が急速に低下する患者集団をここでは Rapid decliner と呼ぶ [2-4]。FEV₁の急速な低下（Rapid decline）は、不可逆的な気流閉塞のリスクファクターとして知られている[5]。

喘息患者においてはFEV₁の低下と、症状コントロールの悪化や喘息増悪リスク上昇との間に関連があることが報告されている [7,8]。また、FEV₁と努力性肺活量（FVC）などの呼吸機能の Rapid decline は、喘息の重症化だけでなく、心血管イベント発症率や死亡率の増加にもつながる[10-12]。喘息の重症化は、疾病負荷だけでなく、社会的・経済的な負担の増加にもつながる[9]。これらのことから、喘息国際ガイドラインであるGINA（Global Initiative for Asthma）report では、喘息患者の現在の喘息コントロール状態と、FEV₁経年低下を含めた将来のリスクを評価し、総合的な評価に基づいて治療管理を行うことが重要と述べられている[13]。

喘息におけるFEV₁の推移は患者ごとで多様であるが、少なくとも喘息患者におけるFEV₁の rapid decline を予測することができれば、気流閉塞の進行を抑制するために早期治療介入を行うことが可能となる。これまでに、気流閉塞の進行の予測因子として、血中好酸球数、呼気一酸化窒素濃度（Fraction of exhaled Nitric Oxide : FeNO）、年

年齢、Body mass index (BMI)、喘息増悪などが個別の研究から報告されている[14-19]。これらの予測因子を組み合わせることで、個々の喘息患者の気流閉塞の進行を予測できる可能性がある。しかし、我々が知る限り、これらの予測因子を組み合わせる喘息患者の FEV₁ の rapid decline を予測した研究はこれまでにない。

近年の機械学習の解析技術の進歩に伴い、データの非線形性や複数の説明変数間の複雑な相互作用を考慮した上で、個々の患者で高精度に分類・予測するモデルを構築することが可能になった[20,21]。過去には、喘息増悪の予測因子を組み合わせる機械学習モデルを構築することで、喘息増悪を高精度に予測した研究が報告されている[23]。

本研究では、気流閉塞の予測因子を組み合わせる教師あり機械学習を実施することで、喘息患者において3年後の FEV₁ の rapid decline を高精度に予測するモデルを構築した。機械学習を用いて未来の FEV₁ の低下を高精度に予測できれば、不可逆的な気流閉塞を生じる前の早期に治療強化を行うことができる。これによって、喘息患者の重症化予防や健康寿命の延伸につながることを期待できる。

3.2 研究方法

3.2.1 研究デザイン

本研究の解析概要を図1に示す。喘息患者の3年後の rapid decline を予測する教師あり機械学習モデルを構築した。全てのデータ解析においてプログラム言語である Python3 (version 3.8.2) を用いた。Python パッケージの pandas (version 1.4.0)、numpy (version 1.22.1)、seaborn (version 0.11.1)、scipy (version 1.7.3) を適用した。機械学習は、Python3 のライブラリーである scikit-learn (version 0.23.1) を用いて実施した。本研究は、山口大学医学部附属病院の倫理審査委員会によって承認され (承認番号: 2022-126)、ヘルシンキ宣言の勧告に従って実施された。

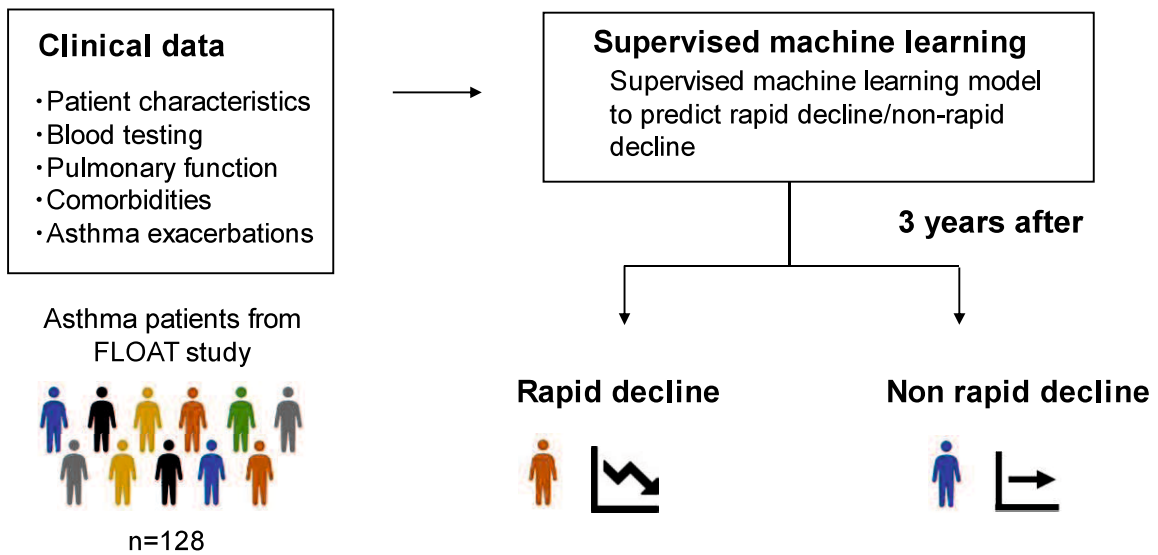


図1. 教師あり機械学習のワークフローを示す。喘息患者を対象とした前向き観察研究：

Factors affecting the LOng-term Asthma Therapy (FLOAT) study へ参加した 128 名の患者の臨床データを用いて、3 年後の rapid decline の有・無を予測する教師あり機械学習モデルを構築した。説明変数として、患者の臨床的背景、血液検査、呼吸機能検査、併存症、喘息増悪の回数などを考慮した。

3.2.2 学習データベース

本研究では、喘息患者を対象とした前向き観察研究：Factors affecting the Long-term asthma therapy (FLOAT study) のデータベースを機械学習に用いた[2-4]。FLOAT study では、標準治療（吸入ステロイド薬等）でコントロール維持されている喘息患者が対象とされた。対象期間は2010年から2013年の合計3年間であった。和歌山県立医科大学附属病院に通院中の喘息患者がリクルートされた。最終的に合計128名の喘息患者が研究を完遂した。対象患者において、ベースラインの喘息治療内容、併存症（慢性副鼻腔炎、アレルギー性鼻炎、逆流性食道炎）、血中好酸球数、血清免疫グロブリンE（Immunoglobulin E: IgE）の値、アトピー（抗原特異的IgE陽性）のデータが収集された。3ヵ月ごとの受診おきに、喘息コントロールテスト（Asthma Control Test：ACT）スコア、呼吸機能検査データ、FeNO値が取得された。観察期間中の増悪回数も取得された。ベースラインから3年後のFEV₁減少量が120mL以上（FEV₁年間平均減少量が40mL/年以上）をrapid declineと定義し、そうでない場合をnon-rapid declineと定義した。Non-rapid decline群、rapid decline群それぞれにおける各臨床データのベースラインを表1に示す。

3.2.3 全説明変数を用いた rapid decline 予測モデル

FLOAT studyに参加した128名の喘息患者データを用いて、3年後のrapid declineを予測する機械学習モデルを構築した。機械学習アルゴリズムとして、ランダムフォレスト、勾配ブースティング決定木、ナイーブベイズ、ロジスティック回帰を用いた[24,25]。対象患者のベースラインにおける以下のデータ項目を説明変数として考慮した：年齢、性別、BMI、喫煙歴、併存症（慢性副鼻腔炎、アレルギー性鼻炎、逆流性食道炎）、ACTスコア、血中好酸球数、血清IgE値、アトピー、呼吸機能検査の項目である努力性肺活量対標準予測値（%FVC: Percent Predicted Forced Vital Capacity）、1秒量対標準予測値（%FEV₁）、FeNO値、過去の増悪回数。応答変数は、喘息患者の3年後のrapid decliner、non-rapid decline（2値変数）とした。

機械学習モデルの予測精度は交差検証法を用いて評価した[26]。交差検証法では、データセットをランダムに k 個に分割する。そのうちの $k-1$ 個を学習に用い、1個を性能評価に用いる。この手順を k 回繰り返すことで、 k 通りの学習モデルを取得し性能を推定する。本研究では $k = 10$ とした（10分割交差検証法）。解析データのうち9割を学習用データとして用いて機械学習を行い、残り1割を機械学習モデルの精度検証に用いた。全データが、一度は学習と検証の両方に使われるため、10分割交差検証法では10通りの学習モデルが構築される。この10通りのモデルについて rapid decline 予測のための Receiver Operating Curve (ROC) 解析を行った。10通りの ROC 曲線の Area Under the Curve (AUC) の平均値および95%信頼区間を求め、予測精度の推定値とした。

Variables	Non-Rapid decliners (n = 95)	Rapid decliners (n = 33)
Age (yrs)	42.3 ± 13.8	47.8 ± 12.3
Gender (female/male)	64/31	22/11
Body mass index (kg/m ²)	21.8 ± 3.7	22.8 ± 5.1
Smoking status (never/Ex), n	72/23	22/11
Atopy, n (%)	75 (78.9)	27 (81.8)
Allergic rhinitis, n (%)	69 (72.6)	24 (72.7)
Chronic rhinosinusitis, n (%)	11 (11.6)	8 (24.2)
Gastro-esophageal reflux disease, n (%)	12 (12.6)	5 (15.2)
Asthma Control Test (points)	23.2 ± 1.0	23.3 ± 1.1
%FVC (%)	103.3 ± 9.4	103.9 ± 10.6
%FEV ₁ (%)	96.9 ± 7.7	97.6 ± 8.2
Blood eosinophil counts (cells/mL)	222 ± 187	299 ± 270
FeNO (ppb)	26.3 ± 15.8	35.9 ± 15.9
Annual exacerbation rate (/yr)	0.05 ± 0.14	0.49 ± 0.62

表1 Non-rapid decliners 群、rapid decliners 群それぞれにおけるベースラインの臨床的背景を表す。表内の数値は平均値±標準偏差または数（割合%）を示す。%FVC, percent predicted forced vital capacity; %FEV₁, percent predicted forced expiratory volume in 1 second; FeNO, fraction of exhaled nitric oxide

3.2.4 Permutation importance 解析

Permutation importance を用いて rapid decline の予測に重要な説明変数を選択した。Permutation importance は、機械学習モデルにおいて各説明変数が予測や判別にどのくらい重要かを評価する解析手法である[24, 27]。

Permutation importance では、解析データベースの中で特定の説明変数をランダムに並び替える（シャッフルする）。ある説明変数を並び替えた後のモデルの予測・判別精度を計算することで、モデルにおける説明変数の重要度を評価する。ある説明変数を並び替えて予測・判別精度が低下した場合、モデルはその説明変数に依存した予測・判別をしている。すなわち、その説明変数は予測・判別に重要であると判断できる。逆に予測・判別精度が変わらない場合、その説明変数は予測・判別に重要な役割を果たしていないと考えられる。

ここで、 j を各説明変数、 s をシャッフルする前の機械学習モデルの予測・判別精度スコア、 K を合計の実行回数、 k をある1回の実行と表記する。 $s_{k,j}$ は、説明変数 j をランダムに並び替えた1回の実行 k におけるモデルの予測・判別精度スコアを表す。Permutation importance (i_j) は以下の通り定義される。

$$i_j = s - \frac{1}{K} \sum_{k=1}^K s_{k,j}$$

説明変数の重要度が高いほどシャッフル後の予測・判別精度 ($s_{k,j}$) が低下するため i_j が高くなる。本研究では、rapid-decline を予測するランダムフォレストモデルにおいて、permutation importance は変数をシャッフルするときの乱数の影響を受け実施回毎に結果の重要度の値が多少変動する。この影響を考慮するため permutation importance を求めるプロセスを100回繰り返し、以下の各説明変数の permutation importance (i_j) の中央値と四分位範囲を得た：年齢、性別、BMI、喫煙歴、ACT、%FVC、%FEV₁、FeNO、白血球数、血中好酸球数、血清IgE値、増悪回数、

アレルギー性鼻炎の有無、副鼻腔炎の有無、逆流性食道炎の有無。Permutation importance (i_j) が高い変数を rapid decliner の予測に必要な説明変数として選択した。Permutation importance の解析は scikit-learn (version 0.23.1) を用いて実行した。

3.2.5 選択した説明変数を用いた Rapid decline 予測モデル

Permutation importance の解析結果に基づいて選択した説明変数をもとに、rapid decline を予測する教師あり機械学習モデルを行った。学習アルゴリズムは、同様に、ランダムフォレスト、勾配ブースティング決定木、ナイーブベイズ、ロジスティック回帰を考慮した。前述と同様に、構築した機械学習モデルそれぞれに関し、10分割交差検証法を実行し、rapid decline の予測精度を ROC-AUC で評価した。同時に、選択した説明変数それぞれについて単変量で ROC 解析を行い、rapid decline を予測する精度を AUC で評価した。これによって、機械学習モデルの予測精度と単変量解析での予測精度を比較した。

3.3 解析結果

3.3.1 全説明変数を用いた機械学習モデルの予測性能

ランダムフォレスト、勾配ブースティング決定木、ナイーブベイズ、ロジスティック回帰を用いて、全説明変数をもとに rapid decline を予測する教師あり機械学習モデルを構築した。それぞれの教師あり機械学習アルゴリズムによる予測精度を 10分割交差検証法で検証した。勾配ブースティング決定木、ランダムフォレスト、ナイーブベイズ、ロジスティック回帰の rapid decline 予測精度を表す AUC 平均値 (95%信頼区間) は、それぞれ 0.69 (0.57-0.81)、0.70 (0.61-0.79)、0.71 (0.50-0.92)、0.58 (0.41-0.75) であった (表2 左列)。

3.3.2 Permutation importance 解析結果

ベースラインの全説明変数を用いて、rapid decliner を予測するランダムフォレストモデルを構築した。このランダムフォレストモデルにおいて、全説明変数についての permutation importance を解析したところ、喘息の増悪回数の permutation importance (i_j) の値が最も高かった。次いで、年齢、血中好酸球数、BMI、FeNO の順で高かった (図2)。これらの5種類の説明変数を、rapid decline の予測において重要な説明変数として選択した。これら5種類の説明変数の値の分布を Rapid decline 群と non-rapid decline 群のそれぞれにおいて示す (図3)。

Machine learning model	ROC AUC (95%CI)		
	All variables	Selected five variables (exacerbation: history)	Selected five variables (number of exacerbations)
Gradient boosting trees	0.69 (0.51-0.87)	0.73 (0.65-0.81)	0.77 (0.62-0.92)
Random Forest	0.70 (0.61-0.79)	0.80 (0.67-0.93)	0.79 (0.60-0.98)
Naive bayes	0.71 (0.50-0.92)	0.80 (0.69-0.91)	0.85 (0.65-1.00)
Logistic regression	0.58 (0.41-0.75)	0.79 (0.67-0.91)	0.83 (0.67-1.00)

表2 教師あり機械学習モデルにおける rapid decline の予測精度を示す。表内の数値は、10分割交差検証における AUC 平均値（95%信頼区間）を表す。左列は全説明変数を用いたモデル、真ん中列（exacerbation: history）は、増悪歴の有・無、年齢、血中好酸球数、BMI、FeNO の5つを説明変数として用いた機械学習モデル、右列（number of exacerbations）は、増悪回数そのままの数値、年齢、血中好酸球数、BMI、FeNO の5つを説明変数として用いた機械学習モデルの AUC 平均値（95%信頼区間）を示す。ROC, Receiver operating characteristic; AUC, area under the curve; BMI, body mass index; FeNO, fraction of exhaled nitric oxide。

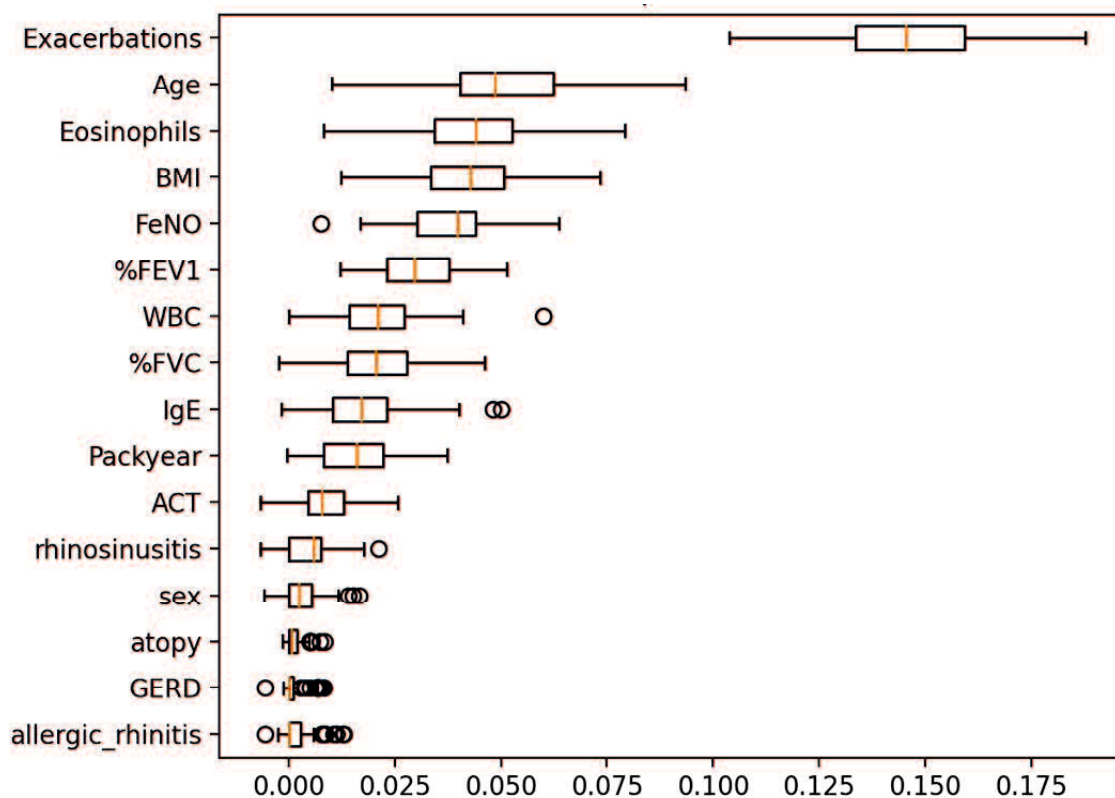


図2 Permutation importance 解析の結果を示す。Rapid decline を予測するランダムフォレストモデルにおいて、各説明変数をランダムに並び変え、permutation importance (i_j) を計算した（合計 100 回反復試行）。説明変数ごとの permutation importance (i_j) の値の分布を箱ひげ図に表す。グラフの横軸は Permutation importance を示す。Permutation importance (i_j) の値が高いほど rapid decline の予測に重要である。増悪回数の Permutation importance (i_j) の値が最も高く、次いで、年齢、血中好酸球数、BMI、FeNO の順となっていた。BMI, body mass index; FeNO, fraction of exhaled nitric oxide; %FEV1, percent predicted expiratory volume in 1 second; WBC, white blood cell; %FVC, percent predicted FVC; IgE, Immunoglobulin E; ACT, asthma control test; GERD, gastro-esophageal reflux disease

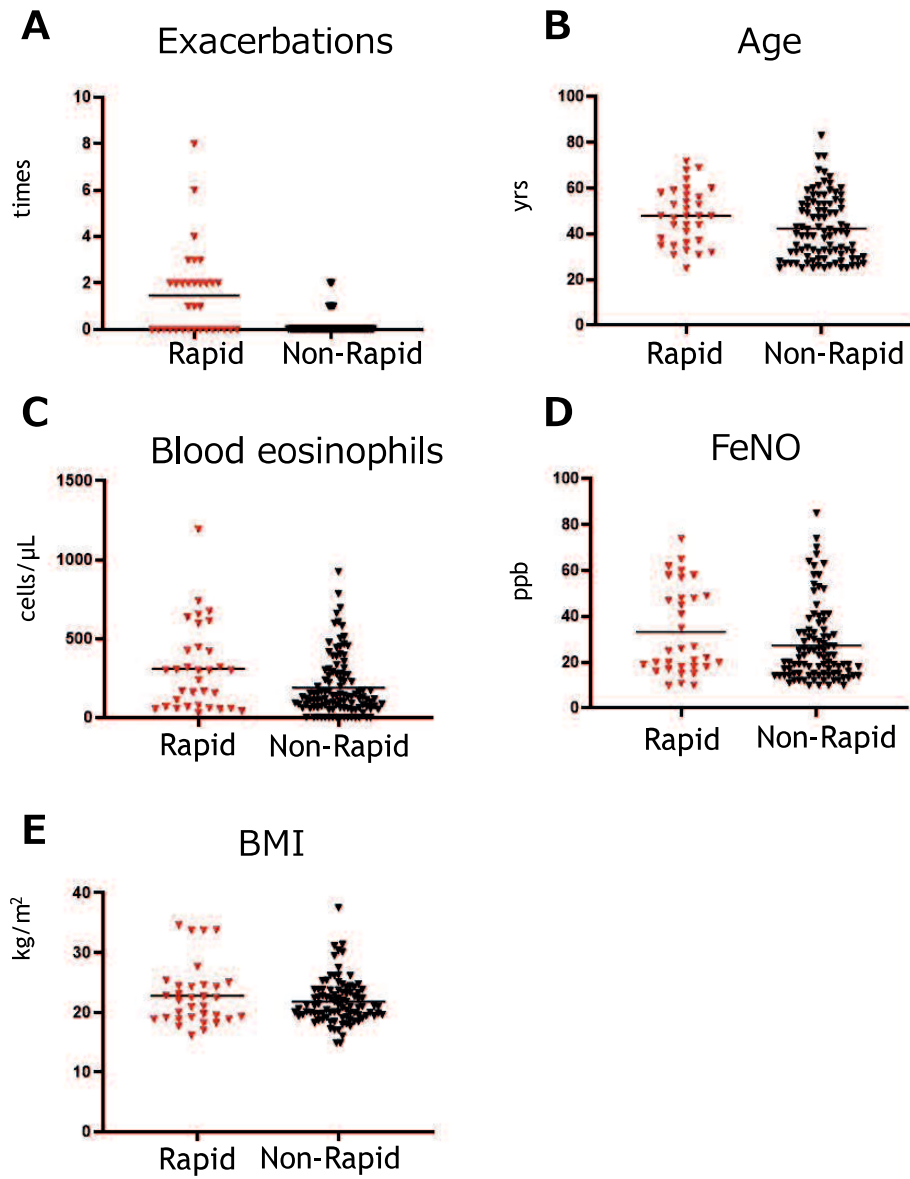


図3. Permutation importance が上位の説明変数の値の分布を Rapid decline 群と non-rapid decline 群それぞれで示す。Rapid decline 群と non-rapid decline 群のそれぞれにおいて (A) 増悪回数 (B) 年齢 (C) 血中好酸球数 (D) FeNO (E) BMI の分布を示す。BMI, body mass index; FeNO, fraction of exhaled nitric oxide

3.3.3 選択した説明変数を用いた機械学習モデルの予測性能

Permutation importance 解析結果をもとに選択した5種類の説明変数（増悪、BMI、年齢、FeNO、血中好酸球数）を用いて、rapid decliner を予測する機械学習モデル（勾配ブースティング決定木、ランダムフォレスト、ナイーブベイズ、ロジスティック回帰）を構築した。5種類の説明変数のうち、増悪に関しては増悪歴の有無を説明変数として用いた場合と増悪回数そのままの数値を説明変数として用いた場合の機械学習モデルを構築した。それぞれの機械学習モデルについて10分割交差検証法を行った。ROC 解析を実施し rapid decline を予測する精度を検証した。

増悪については、増悪歴の有無（二値変数）を説明変数として用いた場合の勾配ブースティング決定木、ランダムフォレスト、ナイーブベイズ、ロジスティック回帰の予測精度を表す AUC 平均値（95%信頼区間）は、それぞれで、0.73（0.65-0.81）、0.80（0.67-0.93）、0.80（0.69-0.91）、0.79（0.67-0.91）であった（表2 真ん中列）。

同時に、増悪については、増悪回数そのままの数値を説明変数として用いた場合の機械学習モデルも構築した。この機械学習モデルの ROC 曲線を作成した（図4）。勾配ブースティング決定木、ランダムフォレスト、ナイーブベイズ、ロジスティック回帰の予測精度を表す AUC 平均値（95%信頼区間）は、それぞれ、0.79（0.60-0.98）、0.77（0.62-0.85）、0.85（0.65-1.00）、0.83（0.67-0.100）であった（表2 右列）。増悪に関して増悪歴の有無を説明変数として用いた場合よりも、増悪回数そのままの数値を説明変数として用いた場合のモデルの方が、rapid decliner を予測する精度が高い傾向にあった。ナイーブベイズで、rapid decliner の予測精度が最も高かった。

選択した5つの説明変数（増悪回数、年齢、血中好酸球数、BMI、FeNO）について、それぞれ単変量で rapid decliner を予測する ROC 解析を行った。増悪回数、年齢、血中好酸球数、BMI、FeNO による予測精度を示す AUC（95%信頼区間）は、それぞれ、0.73（0.62-0.85）、0.63（0.53-0.74）、0.57（0.45-0.70）、0.62（0.48-0.71）、0.53（0.40-0.65）であった（表3）。各説明変数を用いた単変量解析よりも、これら5種類の説明変数を組み合わせた機械学習モデルの方が、rapid decline 予測精度が高かった。

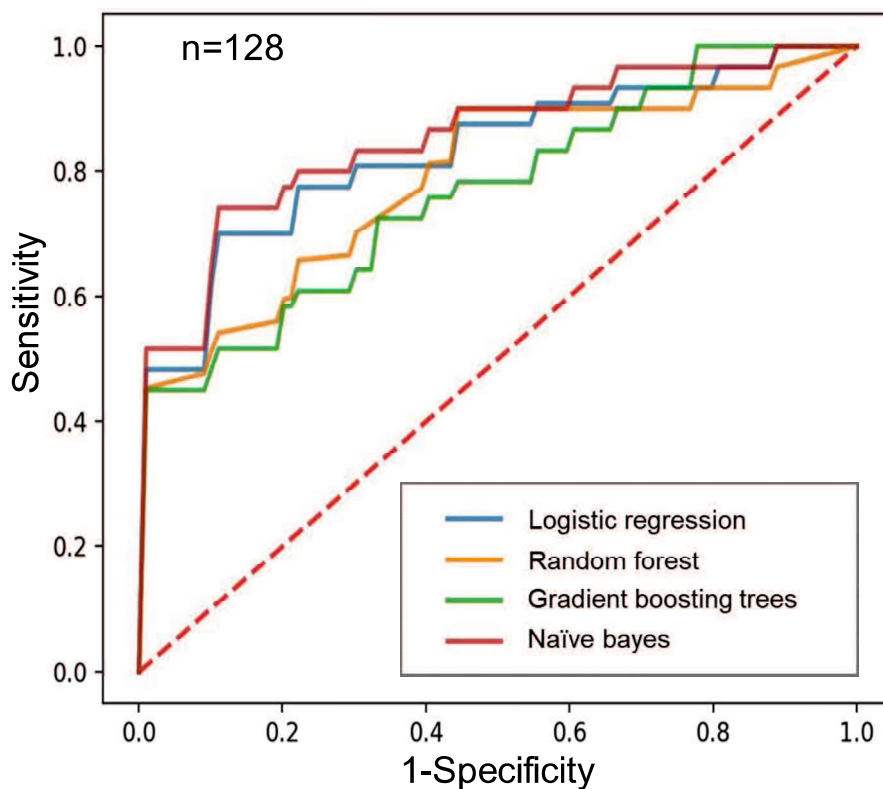


図4. Rapid decline を予測する教師あり機械学習モデルのROC 曲線を示す。説明変数には選択した5変数（増悪回数、BMI、年齢、FeNO、血中好酸球数）を用いた。ランダムフォレスト、勾配ブースティング決定木、ナイーブベイイズモデル、ロジスティック回帰モデルにおけるrapid decline 予測精度を示すAUC 平均値（95%信頼区間）は、それぞれ0.79（0.60-0.98）、0.77（0.62-0.85）、0.85（0.65-1.00）、0.83（0.67-0.100）であった。青線がロジスティック回帰、オレンジ線がランダムフォレスト、緑線が勾配ブースティング決定木、赤線がナイーブベイイズを示す。BMI, body mass index; FeNO, fraction of nitric oxide; ROC, receiver operating characteristic; AUC, area under the curve。

Univariate analysis	ROC AUC (95%CI)
Exacerbations	0.73 (0.62-0.85)
Age	0.63 (0.53-0.74)
Blood eosinophils	0.58 (0.45-0.70)
BMI	0.53 (0.40-0.65)
FeNO	0.62 (0.48-0.71)

表3 単変量それぞれのROC解析におけるrapid decline予測精度を示す。ROC, Receiver operating characteristic; AUC, area under the curve; BMI, body mass index; FeNO, fraction of exhaled nitric oxide。

3.4 考察

解析結果の要約

本研究では、FEV₁の rapid decline の予測に重要な説明変数を組み合わせて教師あり機械学習を行うことで、喘息患者における FEV₁の rapid decline を高精度に予測できた。教師あり機械学習アルゴリズムの中で、ナイーブベイズモデルが最も高い精度を示し、その AUC (95%信頼区間) は 0.85 (0.65-1.00) であった。Permutation importance の結果に基づき説明変数を選択することで、rapid decline 予測精度が向上した。

機械学習の解析結果についての考察

全ての説明変数を用いて rapid decline を予測する教師あり機械学習を行った際には、予測精度は最大でも AUC 約 0.7 程度であった。その後、Permutation importance の結果に基づいて、全ての説明変数の中から重要度の高い説明変数（増悪回数、年齢、血中好酸球数、BMI、FeNO）を選択した。これらの選択した説明変数を用いて教師あり機械学習を行った場合には、予測精度が最大で AUC 0.85 まで向上した。Rapid decline の予測に重要な説明変数を選択することで、喘息患者の FEV₁の rapid decline を高精度に予測することが出来た。

教師あり機械学習では、説明変数が増加すると、データが冗長となり「次元の呪い」が生じる場合がある[28]。「次元の呪い」とは、説明変数が増えることで、そのデータを分析する際の計算量や必要サンプルサイズが指数関数的に増大する現象を指す。これによって、機械学習モデルが高精度に予測・判別することができない。また、説明変数が多すぎると、説明変数間の多重共線性が生じる可能性が増し、予測や判別がうまくできないこともある。予測や判別に重要な説明変数を選択することで、これらの学習上の問題を解決し、学習モデルの予測精度を向上させることができる。

Permutation importance は重要な説明変数を選択する上で有用である。Permutation

importance は、機械学習モデルにおいて、各説明変数が予測・判別にどの程度重要かを定量する解析手法である[27]。重要度の高い説明変数を選択することによって、学習モデルにおける冗長性を回避し、より高精度な機械学習モデルの構築につながる可能性がある。実際に本研究でも、permutation importance の結果に基づき説明変数の選択は、モデルの予測精度の向上に寄与した。従来の統計モデルでは、解析者の主観や仮説をもとに説明変数を選択する。しかし、本研究では permutation importance の解析結果に基づきデータ駆動的に説明変数を選択した。この選択方法は、解析者の仮説や主観に依存しないという長所がある。Permutation importance は「説明可能な AI (Explainable AI: XAI)」の技術としても注目されている。どの説明変数が予測・判別に重要であるかを評価できるため、構築した機械学習モデルの予測・判別根拠として解釈することもできる[29]。

本研究の解析結果からは、説明変数のデータを二値化するのではなく、そのものの数値を用いたほうが、機械学習モデルの精度が優れることが示唆された。FLOAT study の先行研究の報告では喘息患者を増悪歴の有無で層別化し、それぞれの群の FEV₁ 値の推移について解析した。これによって、喘息増悪と気流閉塞の進行 (FEV₁ 低下) の関連性について示した[4]。一方、この先行研究では、増悪歴の有無で層別化したことで、増悪回数について、数値の大小関係の情報が喪失した。しかし、本研究では増悪歴の有無 (二値変数) だけでなく、増悪回数そのものの数値を用いた場合の機械学習モデルも構築し、それぞれで、FEV₁ の rapid decline 予測モデルの精度を比較した (表 2)。増悪回数を二値化した予測モデルよりも、増悪回数そのものの数値を用いた予測モデルの方が rapid decline の予測精度が高かった。これは、機械学習に変数そのものの数値を用いることで、数値の大小関係の情報を保持できたからかもしれない。

Rapid decline 予測モデルで選択した説明変数に関する考察

Permutation importance を尺度としたとき、増悪回数、年齢、血中好酸球数、BMI、FeNO が rapid decline 予測において重要度が高いことを示した。興味深いことに、これ

らの説明変数は、いずれも気流閉塞の予測因子であることが過去の研究で示されている。喘息増悪によって過剰な気管支収縮や気道への物理的ストレスが生じ、気道のリモデリングが進行する。これによって、気流閉塞につながると考えられている[4]。2型炎症バイオマーカー（血中好酸球数と FeNO）の高値は喘息の将来リスクである[30]。血中好酸球数高値、FeNO 高値は、FEV₁ の rapid decline と関連することが過去に報告されており、本研究の結果と合致している[3,19]。肥満（BMI≥25）は気道炎症の増悪に寄与し、気流閉塞の進行と関わるということが報告されている[14,15]。加齢による胸郭コンプライアンス低下や呼吸筋力低下で、FEV₁ の低下につながることが報告されている[18]。

これら既知の気流閉塞の予測因子を用いた単変量解析で rapid decline を予測した際には、十分な精度で予測することが出来なかった（表2）。しかし、これら気流閉塞の予測因子を組み合わせた機械学習モデルでは、rapid decline を高精度に予測できた。

本技術の臨床現場への応用

本機械学習モデルの活用に必要なデータは、血中好酸球数、FeNO、BMI、年齢、増悪回数の5種類の説明変数の数値のみである。これらの説明変数のデータを入力することで、喘息患者が3年後に rapid decline に至るか否かを高精度に予測できる。これらの説明変数は、日常診療の範囲内で取得するデータ項目である。このように日常診療の範囲内の臨床データに基づいたAI技術は、臨床判断支援システム（Clinical decision support system: CDSS）として臨床現場へ円滑に導入することができる[31,32]。説明変数の数値を、患者ごとに電子カルテ（electronic health record : EHR）にデジタルデータとして取得し、電子カルテシステムと連携すれば機械学習モデルが予測結果を出力できる。本技術は、Intel Core i7 CPU を内蔵したノート PC でも数秒以内に予測結果を出力することができるため、リアルタイムに診療を支援することが可能である。将来的には、遠隔医療において、本研究で開発した機械学習モデルを実装することで、発展途上国や過疎地域などの医療環境体制が不十分な地域でも活用できる[33-

37]。これによって、医療格差の是正につながる可能性がある。

過去の機械学習を用いた喘息研究との比較

Zein JG らの研究では、EHR データベースを用いて喘息増悪を予測する教師あり機械学習が行われた。全身性ステロイドを要する増悪、救急受診を要する増悪、入院を要する増悪を、それぞれ AUC (95%信頼区間) 0.71 (0.70-0.72)、0.88 (0.86-0.89)、0.85 (0.82-0.88) の精度で予測した[23]。しかし、この研究では、喘息コントロール状態、呼吸機能検査などの重要な臨床情報が不足していた。例えば、呼吸機能検査の取得率はわずか 15.7%であった。他にも、EHR のデータを用いて喘息増悪を予測した研究がいくつか報告されているが、同様に臨床情報の不足が課題となっていた[38,39]。これらの過去の研究での重要な臨床情報の不足は、予測性能に悪影響を与えていた可能性がある。一方、本研究で解析データベースとして用いた FLOAT study では、すべての喘息患者において症状コントロール評価や呼吸機能検査が実施されているという長所がある。本研究では、上記の喘息における重要な臨床情報を考慮し、教師あり機械学習を行うことで、喘息患者における FEV₁ の rapid decline を高精度に予測できた。

本研究の有用性

喘息における FEV₁ の急速な低下は、不可逆的な気流閉塞のリスク因子である。気流閉塞は、症状コントロール悪化、増悪リスクの増加につながり、喘息の重症化をもたらす[7,8]。開発した AI 技術を活用することで FEV₁ の急速な低下を予測し、早期治療介入の手掛かりにする。これによって、喘息増悪リスクの抑制や症状コントロールの改善につなげる。

近年、喘息の分子標的治療薬（生物学的製剤）の登場によって、喘息コントロールは飛躍的に向上した。これによって、喘息寛解も実現可能な達成目標となった[40,41]。早期の治療強化が喘息の重症化予防や寛解導入につながる可能性が最近示されている。生物学的製剤の早期導入が、喘息寛解の予測因子であることが示されてい

る[41,42]。本研究で開発した AI の予測結果に基づいて、早期の治療強化を行うことで、重症化予防や喘息寛解につながる可能性が期待できる。

長期間にわたる喘息患者の FEV₁ 値の自然経過 (trajectory) をみた観察研究では、長期にわたり rapid decline が続いた症例では、不可逆的な気流閉塞を生じ慢性閉塞性肺疾患の新規合併や重症化につながる事が報告されている[43]。本研究で開発した機械学習モデルは AUC 0.80 以上の高精度で3年後の FEV₁ の rapid decline を予測できる。本モデルの予測結果に基づいて早期治療介入を行うことで、喘息患者の長期的な FEV₁ の自然経過を修飾し、安定した FEV₁ 値を長期にわたって維持できるかもしれない。これによって、特に高齢者の喘息患者においては、喘息の重症化予防だけでなく心血管イベントリスクを予防し、ひいては、身体的フレイルの予防や健康寿命の延伸にもつながる可能性が期待できる。

従来、喘息における気流閉塞の進行 (FEV₁ 低下) の予測には、十分な喘息診療の経験や専門知識が必要であった。なおかつ、この予測は専門家の主観的判断であるという限界があった。一方、本研究で構築した機械学習モデルを活用することで、経験の浅い修練医や非専門の一般内科医においても、気流閉塞の進行を客観的に予測できる可能性がある。本研究で開発した学習モデルを応用することによって、早期の治療強化や専門施設への紹介のための手掛かりを提供する。

今後の展望

今後、本研究で用いた学習データベースとは別の喘息患者コホートを用いて外的検証を行うことが望ましい。将来的には、本 AI 技術の予測に基づいて早期治療介入を行い、喘息の重症化予防の効果を前向きに検証することが重要である。

本研究の限界

学習データベースが単施設のみから収集されていること、機械学習を行う上では比較的サンプル数が小さいことの2点から本 AI 技術の汎化性能について懸念がある。今

後、別の喘息患者コホートを用いて外的検証（external validation）を実施する必要がある。

学習データベースとして用いた FLOAT study の喘息患者は、喘息コントロールが安定した患者に限定されている。重症喘息患者対しても同様の予測精度を示すかに関しては、同様に外的検証が必要である。

3.5 結語

教師あり機械学習を行い、複数の予測因子を組み合わせた予測モデルを構築することで、急速に FEV₁ が低下する喘息患者（rapid decline）を高精度で予測することができた。本機械学習モデルの活用によって、FEV₁ の rapid decline のリスクが高い患者を同定できる。本 AI モデルの予測結果に基づき先制治療を行うことによって、喘息の重症化予防、喘息寛解、喘息の疾患経過の修飾、ひいては、喘息患者の健康寿命の延伸に寄与することが期待できる。

本技術は現在、特許出願中である（特願 2021-167363 病気の重症化予測システム：濱田和希、松永和人、浅井義之）。

3.6 参考文献

1. Reddel HK, Bacharier LB, Bateman ED, et al. Global Initiative for Asthma Strategy 2021: executive summary and rationale for key changes. *Eur Respir J* 2022; 59: 2102730.
2. Matsunaga K, Ichikawa T, Oka A, et al. Changes in forced expiratory volume in 1 second overtime in patients with controlled asthma at baseline. *Respir Med*. 2014; 108: 976-982.
3. Matsunaga K, Hirano T, Oka A, Ito K, Edakuni N. Persistently high exhaled nitric oxide and loss of lung function in controlled asthma. *Allergol Int* 2016; 65: 266-271.

4. Matsunaga K, Hirano T, Oka A, et al. Progression of Irreversible Airflow Limitation in Asthma: Correlation with Severe Exacerbations. *J Allergy Clin Immunol Pract* 2015; 3: 759-64.
5. Lange P, Celli MB, Agusti A, et al. Lung-Function Trajectories Leading to Chronic Obstructive Pulmonary Disease. *N Engl J Med* 2015; 373:111-22.
6. Chung KF, Wenzel SE, Brozek JL, et al. International ERS/ATS guidelines on definition, evaluation and treatment of severe asthma. *Eur Respir J* 2014; 43: 343–373.
7. Contoli M, Baraldo S, Marku B, et al. Fixed airflow obstruction due to asthma or chronic obstructive pulmonary disease. *J Allergy Clin Immunol* 2010; 125: 830-7.
8. Haselkorn T, Szeffler SJ, Chipps BE, et al. Disease Burden and Long-Term Risk of Persistent Very Poorly Controlled Asthma. *J Allergy Clin Immunol Pract* 2020; 8: 2243-53.
9. Nagase H, Adachi M, Matsunaga K, Prevalence, disease burden, and treatment reality of patients with severe, uncontrolled asthma in Japan. *Allergol Int.* 2020;69(1): 53-60.
10. Silvestre OM, Nadruz W, Roca QG, et al. Declining Lung Function and Cardiovascular Risk. *J Am Coll Cardiol.* 2018;72(10):1109-1122.
11. Huang S, Vasquez MM, Halonen M, Martinez FD, Guerra S. Asthma, airflow limitation, and mortality risk in the general population. *European Respiratory Journal* 2015 45: 338-346.
12. Okayama Y, Kawayama T, Kinoshita T, Impact of airflow obstruction on long-term mortality in patients with asthma in Japan. *Allergology International* 2019;68: 462-469.
13. Global Initiative for Asthma. 2022 GINA Report, Global Strategy for Asthma Management and Prevention, Updated 2020. Available online: <http://ginasthma.org> (accessed on 2 April 2023).
14. Sood A, Ford ES, Camargo CA. Association between leptin and asthma in adults. *Thorax* 2006; 61: 300–305.
15. Peters MC, Fahy JV. Metabolic consequences of obesity as an “outside in” mechanism of disease severity in asthma. *Eur Respir J* 2016; 48: 291–293.

16. Backmana H, Lindberg A, Hedmana L, et al. FEV1 decline in relation to blood eosinophils and neutrophils in a population-based asthma cohort. *World Allergy Organ J* 2020; 13(3): 100110.
17. Hancox RJ, Pavord ID, Sears MR. Associations between blood eosinophils and decline in lung function among adults with and without asthma. *Eur Respir J* 2018; 51: 1702536.
18. Kenneth P. Hough, Miranda L. Curtiss, Blain TJ, Liu R, Trevor J, Deshane JS, Thannickal VJ. Airway Remodeling in Asthma. *Front Med (Lausanne)*. 2020; 21(7):191.
19. Coumou H, Westerhof GA, de Nijs SB, et al. Predictors of accelerated decline in lung function in adult-onset asthma. *Eur Respir J* 2018; 51: 1701785.
20. Rajkomar A, Dean J, Kohane I, Machine Learning in Medicine. *N Engl J Med* 2019; 380: 1347-58.
21. Liu Y, Chen PHC, Krause J, Peng L. How to read articles that use machine learning: users' guides to the medical literature. *JAMA*. 2019;322(18):1806-1816.
22. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. *Nat Methods*. 2018;15(4):233-234.
23. Zein JG, Wu CP, Attaway AH, Zhang P, Nazha A, Novel Machine Learning Can Predict Acute Asthma Exacerbation. *CHEST* 2021; 159(5):1747-1757.
24. Breiman L. Random forests. *Mach Learn* 2001; 45: 5–32.
25. Friedman J. Greedy boosting approximation: a gradient boosting machine. *The Annals of Statistics* 2001; 29(5): 1189-232.
26. Kohavi R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence* 1995;14(12): 1137-1145.
27. Altmann A, Tološi L, Sander O, Lengauer T. Permutation importance: a corrected feature importance measure. *Bioinformatics* 2010; 15: 26(10): 1340-7
28. Berisha V, Krantsevich C, Hahn PR, Hahn S, Dasarathy G, Turaga P, Liss J. Digital medicine and the curse of dimensionality. *NPJ Digit Med*. 2021; 28; 4(1): 153.

29. Molnar C. Interpretable Machine Learning. A Guide for Making Black Box Models Explainable. March 2, 2023 <<https://christophm.github.io/interpretable-ml-book/>>accessed in May 2, 2023.
30. Couillard S, Laugerud A, Maisha Jabeen M, et al. Derivation of a prototype asthma attack risk scale centred on blood eosinophils and exhaled nitric oxide. *Thorax* 2022; 77: 199–202.
31. Murphy ME, Fahey T, Smith SM. Computer-based clinical decision support for general practitioners. *Fam Pract*. 2014;31(5):497-8.
32. Bousquet J. Electronic clinical decision support system (eCDSS) in the management of asthma: from theory to practice. *Eur Respir J* 2019; 53: 1900339.
33. Kaplan A, Cao H, FitzGerald JM, Iannotti N et al. Artificial intelligence/machine learning in respiratory medicine and potential role in asthma and COPD diagnosis. *J Allergy Clin Immunol Pract*. 2021; 9(6): 2255-2261.
34. Exarchos KP, Beltsiou M, Votti C-A, et al. Artificial intelligence techniques in asthma: a systematic review and critical appraisal of the existing literature. *Eur Respir J* 2020; 56: 2000521.
35. Guo L, Li B. The Application of Medical Artificial Intelligence Technology in Rural Areas of Developing Countries. *Health Equity* 2018, 2.1.
36. Kaplan A, Cao H, FitzGerald M et al., Artificial intelligence/machine learning in respiratory medicine and potential role in asthma and COPD diagnosis. *J Allergy Clin Immunol Pract* 2021; 9: 2255-61.
37. Khoury P, Srinivasan R, Kakumanu S, et al. A Framework for Augmented Intelligence in Allergy and Immunology Practice and Research. *J Allergy Clin Immunol Pract* 2022; 10: 1178-88.
38. Goto T, Camargo CA, Faridi MK, Yun BJ, Hasegawa K. Machine learning approaches for predicting disposition of asthma and COPD exacerbations in the ED. *Am J Emerg Med*. 2018; 36(9): 1650-1654.

39. Loymans RJB, Debray TP, Honkoop PJ, et al. Exacerbations in Adults with Asthma: A Systematic Review and External Validation of Prediction Models. *J Allergy Clin Immunol Pract.* 2018 Nov-Dec;6(6):1942-1952.e15.
40. Hamada K, Oishi K, Murata Y, Hirano T, Matsunaga K: Feasibility of Discontinuing Biologics in Severe Asthma: An Algorithmic Approach. *J Asthma Allergy* 2021; 14: 1463-1471.
41. Eger K, Kroes JA, Brinke AT, Bel EH. Long-Term Therapy Response to AntiIL-5 Biologics in Severe Asthma—A Real-Life Evaluation. *J Allergy Clin Immunol Pract.* 2021; 9(3): 1194-1200.
42. Oishi K, Hamada K, Murata Y, Matsuda K, Ohata S, Yamaji Y, Asami-Noyama, M, Edakuni, N, Kakugawa, T, Hirano T. A Real-World Study of Achievement Rate and Predictive Factors of Clinical and Deep Remission to Biologics in Patients with Severe Asthma. *J. Clin. Med.* 2023, 12, 2900.
43. McGeachie MJ, Yates KP, Zhou X, et al. Patterns of Growth and Decline in Lung Function in Persistent Childhood Asthma. *N Engl J Med* 2016; 374:1842-52.

第4章

喘息症状から病態特性を推定する技術

4.1 序論

喘息の病態解明や治療法進歩に伴い、喘息病態の多様性に応じた個別化治療が進められている[1]。喘息の個別化治療には、Treatable Traits (TTs) の評価が不可欠である[2-6]。TTs は、治療標的となる臨床的特徴と定義され、病態生理や合併疾患等によって特徴づけられる[6]。喘息や慢性閉塞性肺疾患などの閉塞性肺疾患における代表的なTTs に2型気道炎症、気流閉塞、頻回増悪などがある[7]。2型気道炎症とは、Th2細胞・2型自然リンパ球などが産生する2型サイトカイン (interleukin (IL) -4, IL-5, IL-13, Thymic Stromal Lymphopoietin (TSLP) など) によって生じる気道炎症のことをいう。気流閉塞は、気道炎症によって気道平滑筋収縮や気道リモデリングが生じ、気管支内腔が狭くなった状態をいう。TTs に基づいた治療薬の選択は、喘息の増悪リスク抑制や症状コントロールの改善につながる[8]。例えば、気流閉塞に対しては気管支拡張薬の吸入、ステロイド抵抗性の2型気道炎症に対しては生物学的治療が有効である。

TTs の中でも、気流閉塞や2型気道炎症の評価は、特に治療選択に直結する[7]。気流閉塞や2型気道炎症を評価するためには、呼吸機能検査での1秒量や1秒率の測定、呼気一酸化窒素濃度測定などの専門的な検査が必要である。専門的な検査の実施には、検査機器の導入、検査技師の雇用が必要である。しかし、発展途上国のような医療環境が十分でない地域では、専門的検査の実施が困難であるため、最適な喘息治療薬の選択が難しい[9,10]。先進国においても、過疎地域やプライマリケアの現場など医療体制が十分でない地域・臨床場面では、同様の課題を抱えている[11]。

この課題は、喘息における医療格差につながる。特に、発展途上国や過疎地域では、喘息患者の死亡率増加、社会経済的損失につながっている[12-14]。発展途上国に

おける喘息の疾病負荷は、世界全体の約90%を占めている[12]。先進国の国々においても、都市部・過疎地域間の地域格差や、呼吸器専門施設とプライマリケア間の施設間格差が指摘されている。日本の疫学調査では、都道府県間での喘息増悪率の格差が示されている[15]。英国で行われた研究では、呼吸器専門施設・プライマリケア間での症状コントロール状態の格差が示された[16]。喘息国際ガイドラインを発刊している Global Initiative for Asthma management (GINA) は、このような国家間格差、地域間格差、施設間格差など、様々な医療格差の是正を目標に掲げている[17]。

このように医療環境が十分でない地域であっても、患者報告アウトカム (Patient-Reported Outcome: PRO) は取得することが出来る。PRO とは患者の症状の主観的評価のことである[18]。PRO の取得には専門的な検査機器を必要としない。近年、PRO データを、機械学習をはじめ AI 解析に用いることで、診療を支援する試みが注目を集めている[19-22]。PRO データを用いた AI 解析によって、疾患サブタイプ分類、病態生理、リスク層別化などに関する新たな知見を発見することが期待されている。これによって、疾患コントロールにむけた臨床的判断を支援できる可能性がある。

本研究では、喘息で最もよく用いられる PRO の1つである喘息管理質問票 (ACQ-5: asthma control questionnaire-5) [23,24]に基づいて機械学習を行い、TTs の中でも、喘息の代表的な TTs である2型気道炎症、気流閉塞、頻回増悪の有・無を、ACQ-5に含まれる喘息症状の情報のみから推定することを目指した。これによって、発展途上国や過疎地域など、専門的な検査の実施が難しい地域においても、患者個別の TTs に応じた治療薬選択 (個別化治療) を支援できる可能性がある。

4.2 研究方法

4.2.1 研究デザイン

本研究の解析概要を図1に示す。多施設から収集された喘息患者 (1,697名) の ACQ-5 データに含まれる5種類の喘息症状のデータを用いて機械学習を行い、代表的

な TTs である 2 型気道炎症、気流閉塞、頻回増悪を推定した。まず、5 つの症状（夜間覚醒症状、起床時症状、日常活動の活動制限、息切れ症状、喘鳴）から構成される ACQ-5 データを用いて次元削減を行い、2 次元平面に射影した。これによって、喘息症状の特徴的な分布パターンを可視化した[25]。次に、教師あり機械学習アルゴリズムのランダムフォレストを用いて[26]、ACQ-5 のデータから気流閉塞、2 型気道炎症、頻回増悪（Exacerbation prone）それぞれの有・無を判別した。続いて、ACQ-5 のデータに教師なし学習である階層的クラスタリング[27]を適用して、対象患者集団を 5 クラスタに分類した。各クラスタに所属する患者の ACQ-5 データや臨床的背景とそれぞれの検査値を精査することで症状サブタイプを同定し、各症状サブタイプの 2 型気道炎症、気流閉塞、頻回増悪について解析した。全てのデータ解析にプログラム言語である Python3（version 3.8.2）を用いた。Python パッケージの pandas（version 1.4.0）、numpy（version 1.22.1）、seaborn（version 0.11.1）、scipy（version 1.7.3）、scikit-learn（version 0.23.1）、umap（version 0.5.3）を適用した。

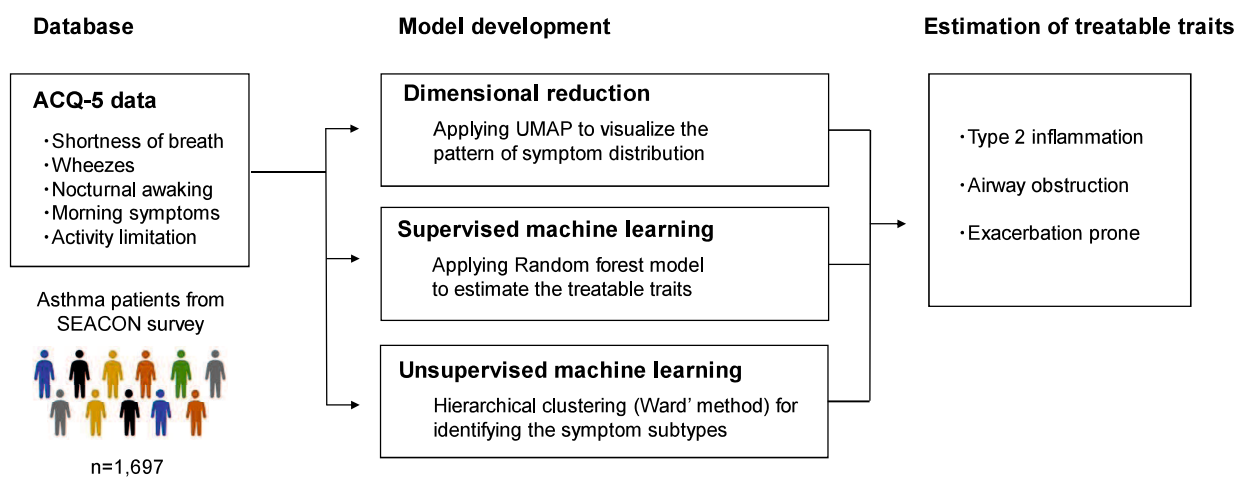


図1. 本研究における機械学習の解析概要を示す。1,697名の喘息患者から取得したACQ-5に含まれる各症状スコアをもとに、次元削減、教師あり機械学習、教師なし機械学習を実施した。これによって、代表的な treatable traits である2型気道炎症、気流閉塞、頻回増悪の有・無を推定することを目指した。ACQ-5, Asthma control questionnaire-5; SEACON survey, Subjective evaluation of asthma control by physician survey; UMAP, Uniform manifold approximation and projection

4.2.2 Asthma Control Questionnaire-5 (ACQ-5)

喘息管理質問票 (Asthma Control Questionnaire-5: ACQ-5) は、喘息症状の評価のための患者報告アウトカム (PRO) である。ACQ-5 は、喘息診療において世界的に最もよく用いられる PRO の1つである。ACQ-5 は、夜間覚醒症状、起床時症状、活動制限、息切れ症状、喘鳴の5種類の喘息症状から構成される[23,24]。各喘息症状の強度に従って、0-6点の整数スコアが昇順で割り当てられる。すなわち、ACQ-5の各喘息症状スコアは順序尺度変数である。ACQ-5は、従来、5つの症状の点数の平均値 (スカラー値) を喘息コントロール状態の評価に用いてきた。一方、ACQ-5は5種類の喘息症状のスコアから構成されることから、5次元ベクトルとして表現できる。本研究では ACQ-5 平均値 (スカラー値) ではなく、ACQ-5 に含まれる5種類の個別の喘息症状スコアのデータ (5次元ベクトル) を用いて機械学習を行った。総合的な症状コントロール状態の評価については、従来の ACQ-5 基準値に沿って、ACQ-5 平均値が 0.75 点未満をコントロール良好、0.75-1.5 点をコントロール不十分、1.5 点以上をコントロール不良と判定した[24]。

4.2.3 学習データベース

本研究では喘息診療実態調査である SEACON (Subjective Evaluation of Asthma CONTROL by the physicians) survey を解析データベースとして用いた [28,29]。SEACON survey は、山口県内の29施設の診療所、病院を対象とした横断的な多施設共同観察研究である。研究期間は2017年10月から12月であった。1,697名の喘息患者が解析対象とされた。SEACON survey では、ほとんどの喘息患者が吸入ステロイド治療を含む標準的治療を受けている[28]。SEACON survey についての詳細な情報は、Matsunaga K らの過去の論文に記載されている[28,29]。SEACON survey の全喘息患者において ACQ-5 データが取得されている。全ての喘息患者において過去1年以内の喘息増悪の回数を取得している。1,697名中964名で呼吸機能検査を実施し、1秒量対標準予測値(percent predict

Forced Expiratory Volume in 1 second : %FEV₁) を測定された。%FEV₁ は気流閉塞の指標である。1,697 名中 848 名で呼気一酸化窒素濃度 (Fraction of exhaled Nitric Oxide : FeNO) を測定された。FeNO は 2 型気道炎症の指標である。SEACON survey は、山口大学医学部附属病院の倫理審査委員会で承認され (承認番号 : H28-108-4)、ヘルシンキ宣言の勧告に従って実施された。

4.2.4 次元削減

ACQ-5 データは 5 次元ベクトルであるため、その分布を人間の目で認識することが出来ない。この 5 次元ベクトルの ACQ-5 データの次元削減を行い、二次元平面に写像することで、喘息症状の分布を可視化することができる。本研究では Uniform Manifold Approximation and Projection (UMAP) を用いて次元削減を実施した[25]。投影した UMAP plot (n=1,697) を 5 種類の喘息症状のスコアの数値に基づいて色付けし、各喘息症状スコアの分布を可視化した。また、2 型気道炎症 (FeNO \geq 35 parts per billion (ppb))、重度の気流閉塞 (%FEV₁ \leq 60%)、頻回増悪 (過去 1 年の喘息増悪 \geq 2 回) それぞれの有・無についても色付けし示した。UMAP の解析には Python3 のパッケージである umap (version 0.5.3) を用いた。

4.2.5 教師あり機械学習

教師あり機械学習アルゴリズムの 1 つであるランダムフォレストを用いて[26]、ACQ-5 データから、喘息患者における各々の重度な気流閉塞、2 型気道炎症、頻回増悪の有・無を判別する学習モデルを構築した。ACQ-5 に含まれる 5 種類の喘息症状 (夜間覚醒症状、起床時症状、活動制限、息切れ症状、喘鳴) の各スコアを説明変数とした。目的変数は、重度の気流閉塞 (%FEV₁ \leq 60%)、2 型気道炎症 (FeNO \geq 35ppb)、頻回増悪 (過去 1 年間の増悪回数 \geq 2 回) の有・無の 3 通りを考慮した。ランダムフォレストモデルにおける各々の目的変数 (2 型気道炎症、重度の気流閉塞、頻回増悪) の判別精度を、交差検証法を用いて検証した[30]。交差検証法ではデータセットをランダムに k 個に分

割する。そのうちの $k-1$ 個を学習に用い、1 個を性能評価に用いる。この手順を k 回繰り返すことで、 k 通りの学習モデルを取得し、それぞれの性能を推定する。本研究では $k=10$ とした (10 分割交差検証法)。解析データのうち 9 割を学習用データとして用いて機械学習を行い、残り 1 割を機械学習モデルの精度検証に用いた。全データが、一度は学習と検証の両方に使われるため、10 分割交差検証法では 10 通りの学習モデルが構築される。この 10 通りのモデルについて、2 型気道炎症、重度の気流閉塞、頻回増悪リスクを推定するための Receiver Operating Curve (ROC) 解析を行った。10 通りの ROC 曲線の Area Under the Curve (AUC) の平均値と 95% 信頼区間を求め、予測精度の推定値とした。教師あり機械学習の解析には Python パッケージの scikit-learn (version 0.23.1) を用いた。

4.2.6 教師なし機械学習

ACQ-5 の 5 種類の喘息症状 (喘鳴、息切れ症状、夜間覚醒症状、起床時症状、日常生活の活動制限) のスコアを用いて階層的クラスタリング解析を実施した。階層的クラスタリング解析には Python パッケージの scikit-learn (version 0.23.1) を用いた。ベクトル間の距離の計算には、ユークリッド距離を用いた。クラスター間の距離の測定方法には Ward 法を用いた [31]。非類似度 (dissimilarity) に基づいて、デンドログラムを作成した。クラスターの数は、デンドログラムおよび UMAP plot での症状分布パターンから決定した。以上のクラスタリング解析を行うことによって、喘息症状サブタイプを同定した。

4.2.7 統計解析

全ての統計検定は Python パッケージ scipy (version 1.7.3) の stats モジュールを用いて実行した。全ての検定において $p < 0.05$ を有意水準に設定した。各クラスターにおいて、記述統計量として、各変数の平均値と標準偏差 (Standard deviation : SD) を計算した。FeNO 値は正規分布しないことから、FeNO 値については刈り込み平均値

(Trimmed mean) を用い、FeNO の分布の上下 2.5% を削除したデータに対して平均値を求めた。各クラスター間の%FEV₁ の分布の差については Student の t 検定を、FeNO 値の分布の差については Yuen の t 検定を用いて統計検定を行った。各クラスターの 2 型気道炎症 (FeNO \geq 35 ppb)、重度の気流閉塞 (%FEV₁ \leq 60%)、頻回増悪 (過去 1 年以内の喘息増悪 \geq 2 回) の比率を計算した。各クラスター間の比率の差の検定については、Fisher の正確性検定を用いた。

5 種類の喘息症状 (夜間覚醒症状、起床時症状、活動制限、息切れ症状、喘鳴) それぞれと、%FEV₁ および FeNO との間の関連性を評価するために、Spearman の相関係数を計算した。5 種類の喘息症状それぞれについて、%FEV₁ および FeNO との間の相関関係を評価し、計 10 通りの相関係数を計算した。多重検定となるため、相関係数の有意性検定においては Bonferroni 法を用いて有意水準の補正を行った。

4.3 解析結果

4.3.1 次元削減の結果

UMAP 解析によって 5 次元データの ACQ-5 (n=1,697) を次元縮約し、各喘息症例を二次元平面に投影した。2 次元平面に投影した次元削減後の UMAP plot を ACQ-5 平均値 (図 2A)、息切れ症状 (図 2B)、喘鳴 (図 2C)、起床時症状 (図 2D)、夜間覚醒症状 (図 2E)、日常活動制限 (図 2F) のそれぞれのスコアの値で色付けした。

UMAP plot における ACQ-5 平均値の分布を図 2A に示す。ACQ-5 平均値が同じ程度の症例は水平方向 (UMAP1 軸方向) に沿って分布し、等高線が水平になるような分布を示した。同時に、ACQ-5 平均値の分布は垂直方向 (UMAP2 軸方向) に沿って連続性を示した。ACQ-5 平均値がより低い症例は UMAP2 軸の負方向に、より高い症例は UMAP2 軸の正方向に分布していた。

UMAP plot において個別の喘息症状の分布をみた場合、息切れ症状・喘鳴については、症状強度が高い症例が図の右上 (UMAP1 軸の正方向かつ UMAP2 軸の正方向) に

第4章 喘息症状から病態特性を推定する技術

分布している（図 2B・C）。一方、夜間覚醒症状については、UMAP plot において症状強度が高い症例が左上（UMAP1 軸の負方向かつ UMAP2 軸の正方向）に多く分布している（図 2E）。

図 2A と、図 2B・C・E とを対比すると、ACQ-5 平均値が同程度の患者集団の中でも、個別の喘息症状パターンが異なることが分かる。例えば、図 2A において ACQ-5 平均値が約 2.0 点程度の症例集団の中でも、息切れ症状・喘鳴が優位な症例集団（図 2B・C の右側（UMAP1 軸の正方向））と、夜間覚醒症状が優位な症例集団（図 2E の左側（UMAP1 軸の負方向））に分かれていた。5次元ベクトルである ACQ-5 を次元削減し、その結果を写像することで、個別の喘息症状の特徴的な分布パターンを明らかにできた。

同時に、UMAP plot において、2型気道炎症（ $\text{FeNO} \geq 35 \text{ppb}$ ）（図 3A）、重度の気流閉塞（ $\% \text{FEV}_1 \leq 60\%$ ）（図 3B）、頻回増悪（過去1年の喘息増悪回数 ≥ 2 回）（図 3C）それぞれの有・無で色付けし示した。UMAP plot において、2型気道炎症（ $\text{FeNO} \geq 35 \text{ppb}$ ）を有する患者は図の左側（UMAP1 軸の負方向）へ（図 3A）、重度の気流閉塞（ $\% \text{FEV}_1 \leq 60\%$ ）を有する患者は図の左側（UMAP1 軸の正方向）へ（図 3B）、頻回増悪は図の上側（UMAP2 軸の正方向）へより多く分布している傾向がみられた（図 3C）。

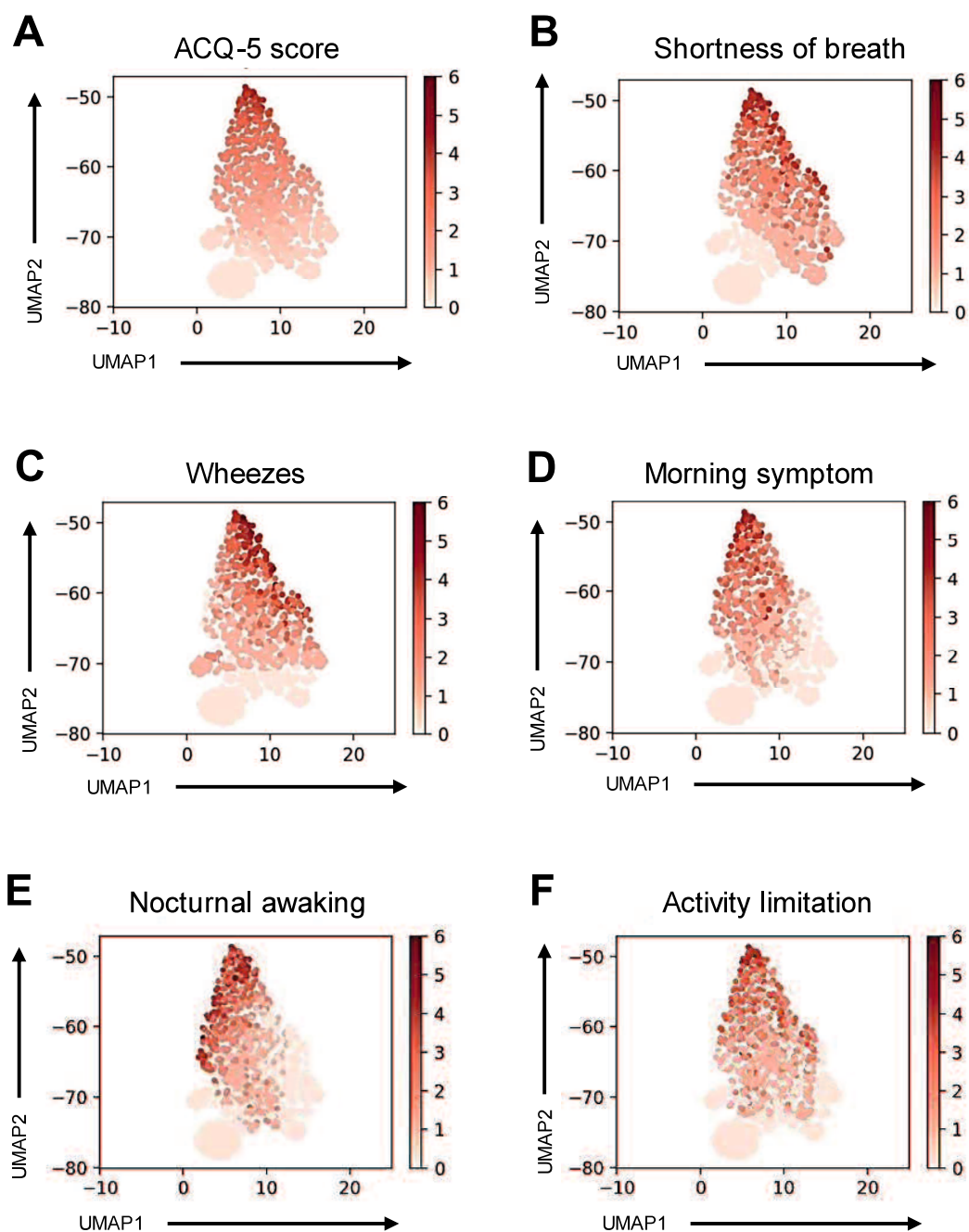


図2. UMAP plot における各喘息症状の分布パターンを示す。ACQ-5 に含まれる 5 種類の各症状スコアについて UMAP を行い、2次元平面に投影した。(A) ACQ-5 平均点、(B) 息切れ症状スコア、(C) 喘鳴スコア、(D) 起床時症状スコア、(E) 夜間覚醒症状スコア、(F) 日常活動制限スコアの値をそれぞれ用いて色付けした。ACQ-5 : Asthma control questionnaire-5; UMAP, Uniform manifold approximation and projection

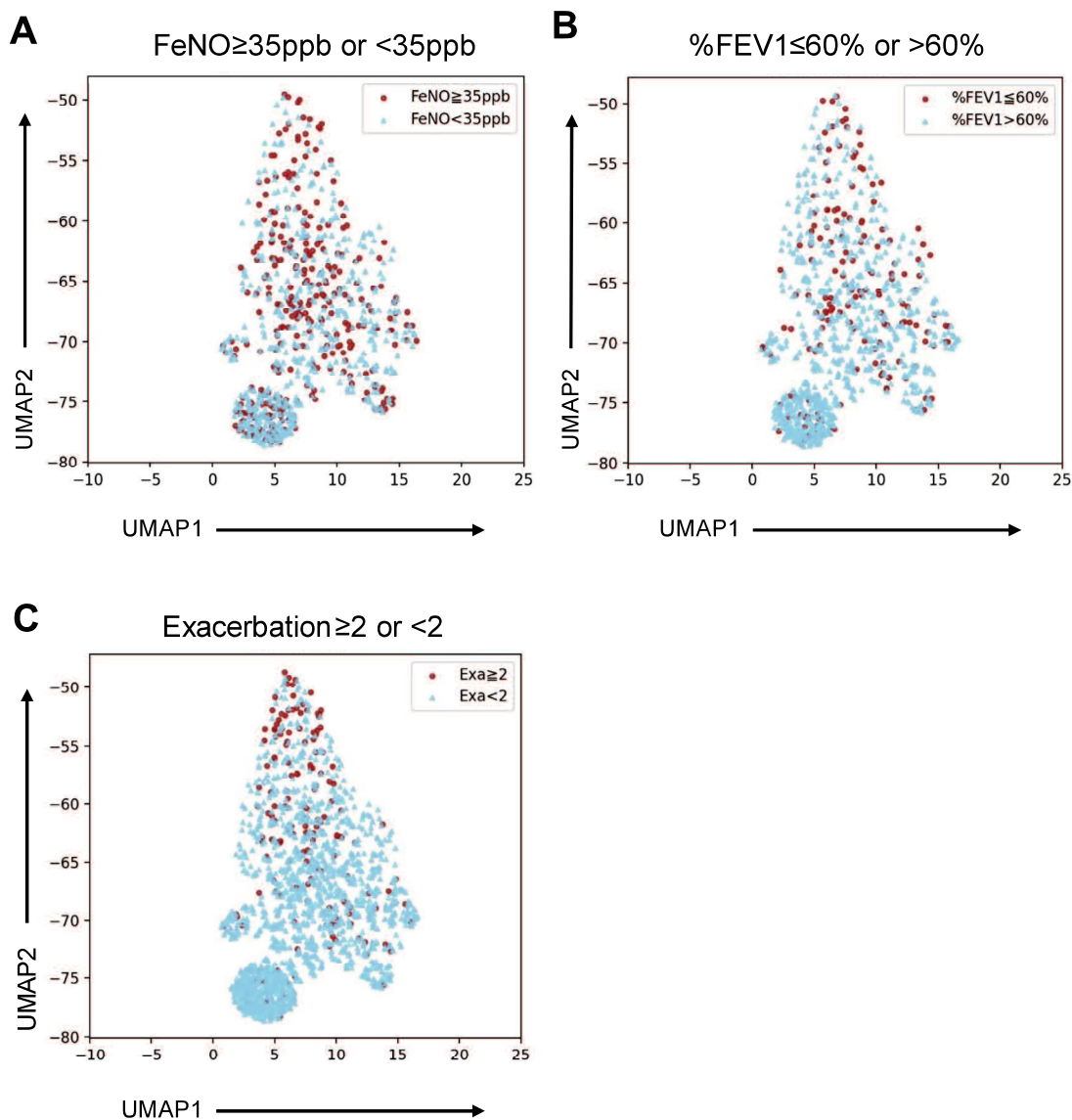


図3. UMAP plot における各々の2型気道炎症、重度の気流閉塞、頻回増悪の有・無の分布を示す。ACQ-5に含まれる5種類の各症状スコアを用いてUMAPを行い、2次元平面に投影した。(A) 2型気道炎症 (FeNO \geq 35ppb)、(B) 重度の気流閉塞 (%FEV₁≤60%)、(C) 頻回増悪 (過去1年の喘息増悪回数 \geq 2回) それぞれの有・無で色付けし示す。青色のプロットは「無」、赤色のプロットは「有」を示す。ACQ-5, asthma control questionnaire-5; FeNO, fraction of exhaled nitric oxide; %FEV₁, percent predicted expiratory volume in 1 second; UMAP, Uniform manifold approximation and projection

4.3.2 教師あり機械学習の結果

ACQ-5 に含まれる 5 種類の喘息症状（夜間覚醒症状、起床時症状、日常活動の活動制限、息切れ症状、喘鳴）のスコアを説明変数とし、重度の気流閉塞（ $\%FEV_1 \leq 60\%$ ）、2 型気道炎症（ $FeNO \geq 35\text{ppb}$ ）、頻回増悪（過去 1 年間の増悪回数 ≥ 2 回）の有・無を判別する教師あり機械学習モデル（ランダムフォレスト）を構築した。10 分割交差検証法を用いて判別精度を検証した。重度の気流閉塞、2 型気道炎症、頻回増悪の有・無の判別精度を示す AUC 平均値（95%信頼区間）は、それぞれ 0.56（0.36-0.76）、0.51（0.32-0.70）、0.68（0.48-0.88）であった（図 4A-C）。重度の気流閉塞、2 型気道炎症の判別精度と比較すると、頻回増悪の判別精度はやや優れていた。

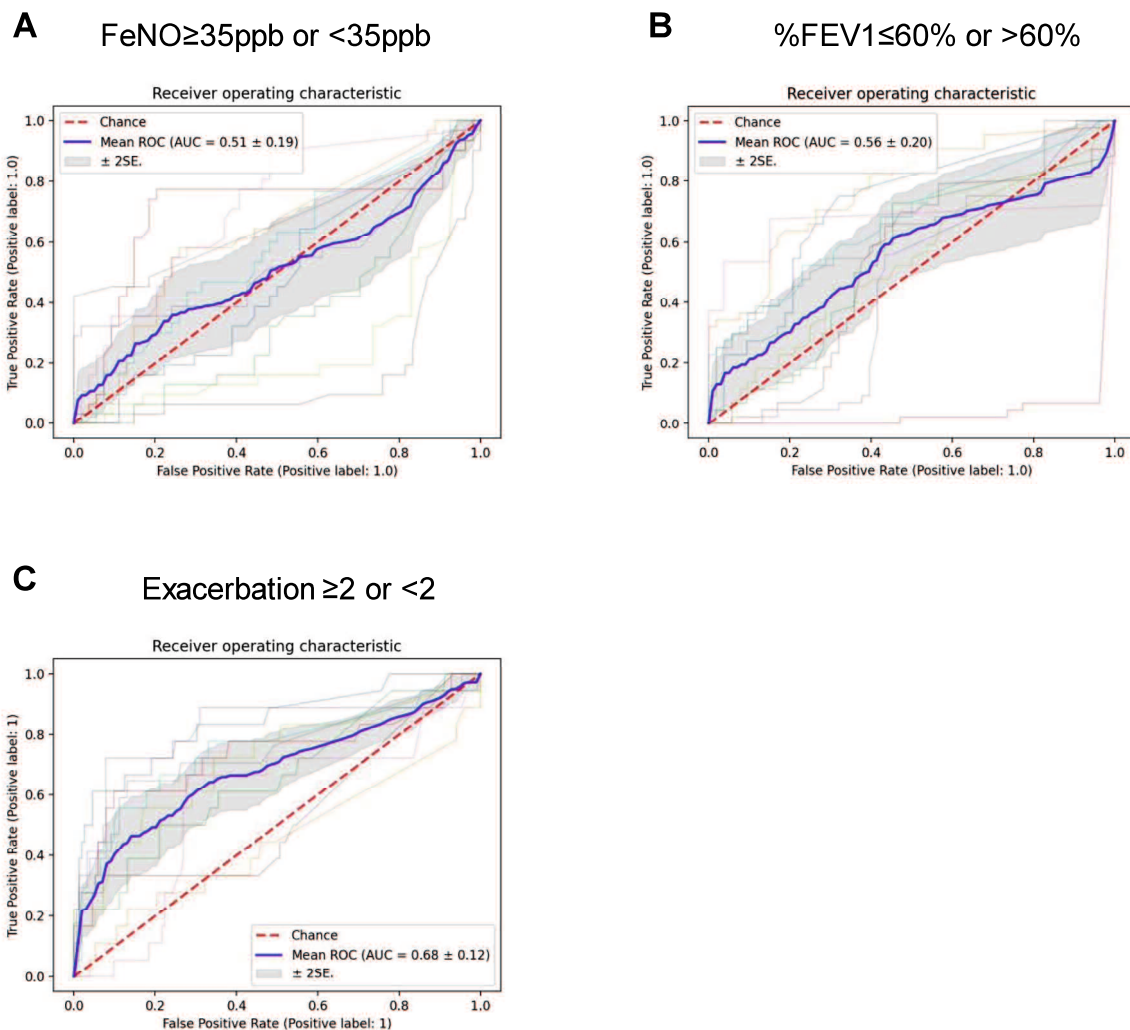


図4. 教師あり機械学習モデルのROC解析を示す。ACQ-5に含まれる5種類の各症状スコアを説明変数として用い、(A) 2型気道炎症 (FeNO \geq 35ppb)、(B) 重度の気流閉塞 (%FEV₁ \leq 60%)、(C) 頻回増悪 (過去1年の喘息増悪回数 \geq 2回) それぞれの有・無を判別する教師あり機械学習を実行した。重度の気流閉塞、2型気道炎症、頻回増悪の有・無の判別精度を示すAUC平均値(95%信頼区間)は、それぞれ、0.56(0.36-0.76)、0.51(0.32-0.70)、0.68(0.48-0.88)であった。機械学習アルゴリズムはランダムフォレストを用いた。判別精度の検証には10分割交差検証法を用いた。ROC, receiver operating characteristic; AUC, area under the curve; FeNO, fraction of exhaled nitric oxide; %FEV₁, percent predicted forced expiratory volume in 1 second。

4.3.3 教師なし機械学習の結果

ACQ-5 データ (n=1,697) を用いて階層的クラスタリングの Ward 法を実行し、デンドログラムを作成した (図 5A)。デンドログラムにおけるデータ間の非類似度 (Dissimilarity) や UMAP plot におけるクラスタの分布状況からクラスタ数を 5 に決定した。ACQ-5 の UMAP plot において、5 つの各クラスタ番号を色付けし示す (図 5B)。各クラスタの総合的な喘息コントロール状態 (コントロール良好/不十分/不良) の割合を図 5C に示す。クラスタ 1 では全ての症例 (100%) がコントロール良好であった (ACQ-5 平均値 \pm SD : 0.05 \pm 0.11)。クラスタ 2 では症状コントロール良好/不十分/不良の割合は、それぞれ 51.1%、47.5%、1.4%であった (ACQ-5 平均値 \pm SD : 0.73 \pm 0.35)。クラスタ 3 では症状コントロール良好/不十分/不良の割合は、それぞれ 0.4%、35.3%、64.3%であった (ACQ-5 平均値 \pm SD : 1.68 \pm 0.44)。クラスタ 4 においては、コントロール良好な症例は含まれず (0%)、コントロール不十分/不良の割合は、それぞれ 17.1%、82.9%であった (ACQ-5 平均値 \pm SD : 1.95 \pm 0.48)。クラスタ 5 は全患者 (100%) がコントロール不良であった (ACQ-5 平均値 \pm SD : 3.20 \pm 0.58)。

各クラスタにおける各喘息症状スコアの平均値を算出し、平均値に基づいてヒートマップを作成した (図 5D)。クラスタ 1 は、ほとんど喘息症状がみられない症状サブタイプであった。クラスタ 2 は、どの喘息症状も軽度な症状サブタイプであった。クラスタ 3 は、息切れ症状・喘鳴が優位な症状サブタイプであった。クラスタ 4 は起床時症状・夜間覚醒症状が優位な症状サブタイプであった。クラスタ 5 は全症状が高度にみられる症状サブタイプであった。UMAP plot でみられたような個別の喘息症状の分布パターンが異なる症例集団を、クラスタリングによって 5 種類の症状サブタイプへ分類することが出来た。

症状サブタイプと気流閉塞、2 型気道炎症、頻回増悪の関連性を解析するために、クラスタ 2 から 5 の %FEV₁、FeNO の値、%FEV₁ \leq 60% の割合、FeNO \geq 35ppb の割合、過去 1 年の喘息増悪回数 \geq 2 回、をクラスタ 1 に対して比較し、統計検定を行っ

第4章 喘息症状から病態特性を推定する技術

た。クラスター2・3・5で%FEV₁値が統計的に有意に低く（図6A）、重度の気流閉塞を呈する喘息患者（%FEV₁≤60%）の割合が統計的に有意に高かった（図6B）。クラスター4・5ではFeNO値が統計的に有意に高く（図6C）、2型気道炎症を有する喘息患者の割合（FeNO≥35ppb）が統計的に有意に高かった（図6D）。頻回増悪（過去1年の喘息増悪回数≥2回）の割合については、クラスター3・4・5で頻回増悪の割合がより高く、特にクラスター5で顕著に頻回増悪リスクの割合が高かった（図6E）。各症状クラスターにおける臨床的背景を表1に示す。

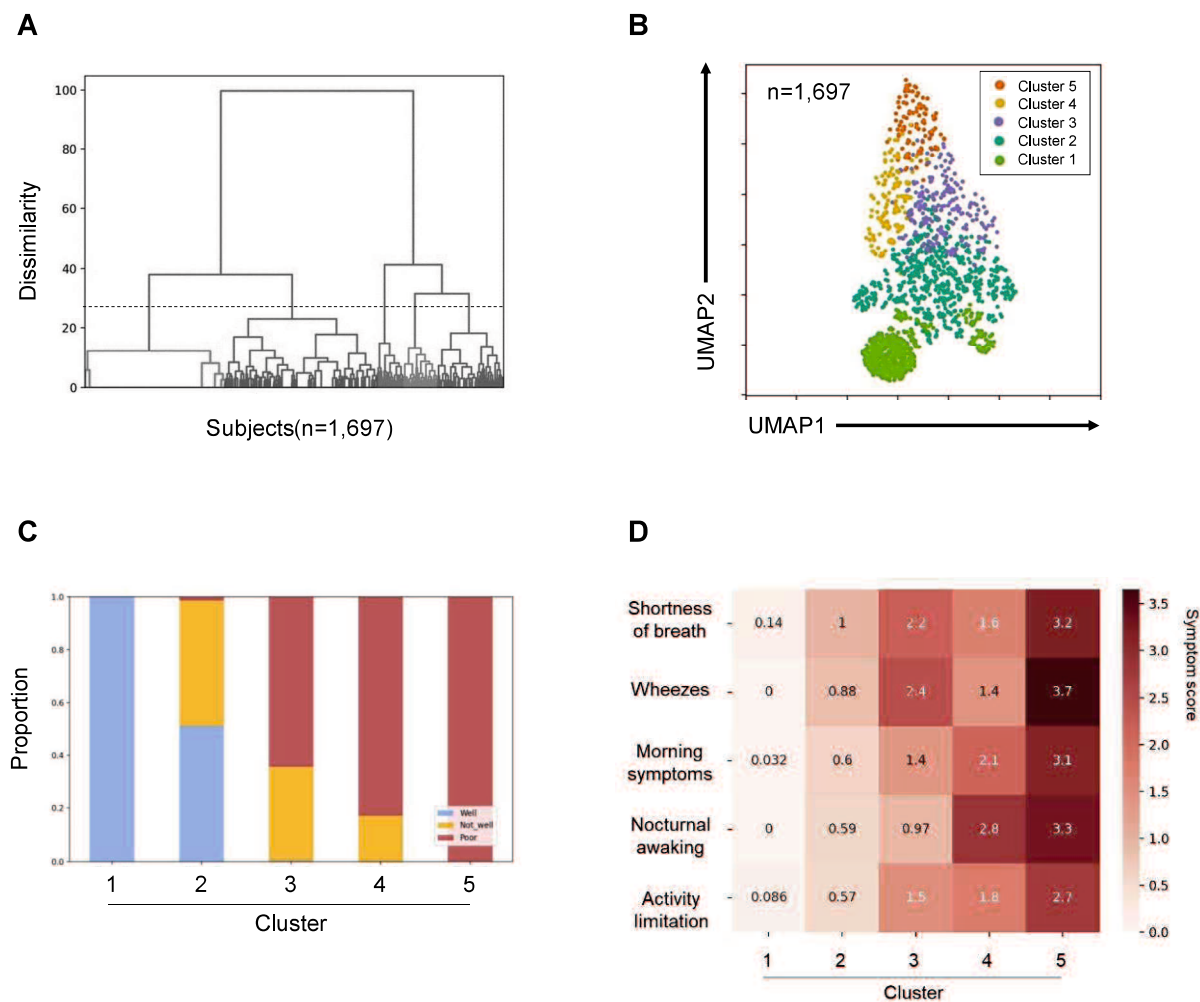


図5 クラスタリングの結果を示す。(A) 階層的クラスタリングを実行し、作成されたデンドログラムを示す。(B) UMAP plot において各クラスター番号で各症例を色付けした。(C) 症状クラスター毎の喘息コントロール状態（コントロール良好/不十分/不良）の割合を示す。青はコントロール良好（ACQ-5 < 0.75 点）、黄色はコントロール不十分（ACQ-5 0.75-1.5 点）、赤はコントロール不良（ACQ-5 > 1.5 点）を表す。(D) 各症状クラスターにおいて5種類の各喘息症状スコアの平均値を表したヒートマップである。

第4章 喘息症状から病態特性を推定する技術

Characteristics	Cluster 1 n=570	Cluster 2 n=617	Cluster 3 n=266	Cluster 4 n=140	Cluster 5 n=104
Gender (Female)	0.61	0.54	0.59	0.66	0.71
Age	60.9 (± 16.1)	63.2 (± 15.6)	63.3 (± 16.3)	65.3 (± 14.0)	62.2 (± 18.1)
BMI	23.0 (± 3.7)	23.6 (± 4.1)	24.2 (± 4.4)	23.0 (± 4.2)	23.7 (± 4.9)
Smoking history	0.30	0.37	0.35	0.34	0.37
Use of ICS	0.95	0.95	0.94	0.97	0.98
Use of LABA	0.71	0.80	0.85	0.85	0.94
Use of LAMA	0.04	0.07	0.11	0.09	0.09
Use of LTRA	0.36	0.36	0.42	0.41	0.41
Use of theophylline	0.08	0.16	0.21	0.16	0.26
Use of OCS	0.05	0.05	0.08	0.08	0.08
Use of omalizumab	0.01	0.00	0.01	0.01	0.03
Use of mepolizumab	0.00	0.00	0.00	0.01	0.02

表1 各症状クラスターの臨床的背景について示す。表内の数値は平均値 (±標準偏差) または割合を示す。BMI, body mass index; ICS, inhaled corticosteroids; LABA, long-acting beta agonist; LAMA, long-acting muscarinic antagonist; LTRA, leukotriene receptor antagonist; OCS, oral corticosteroids

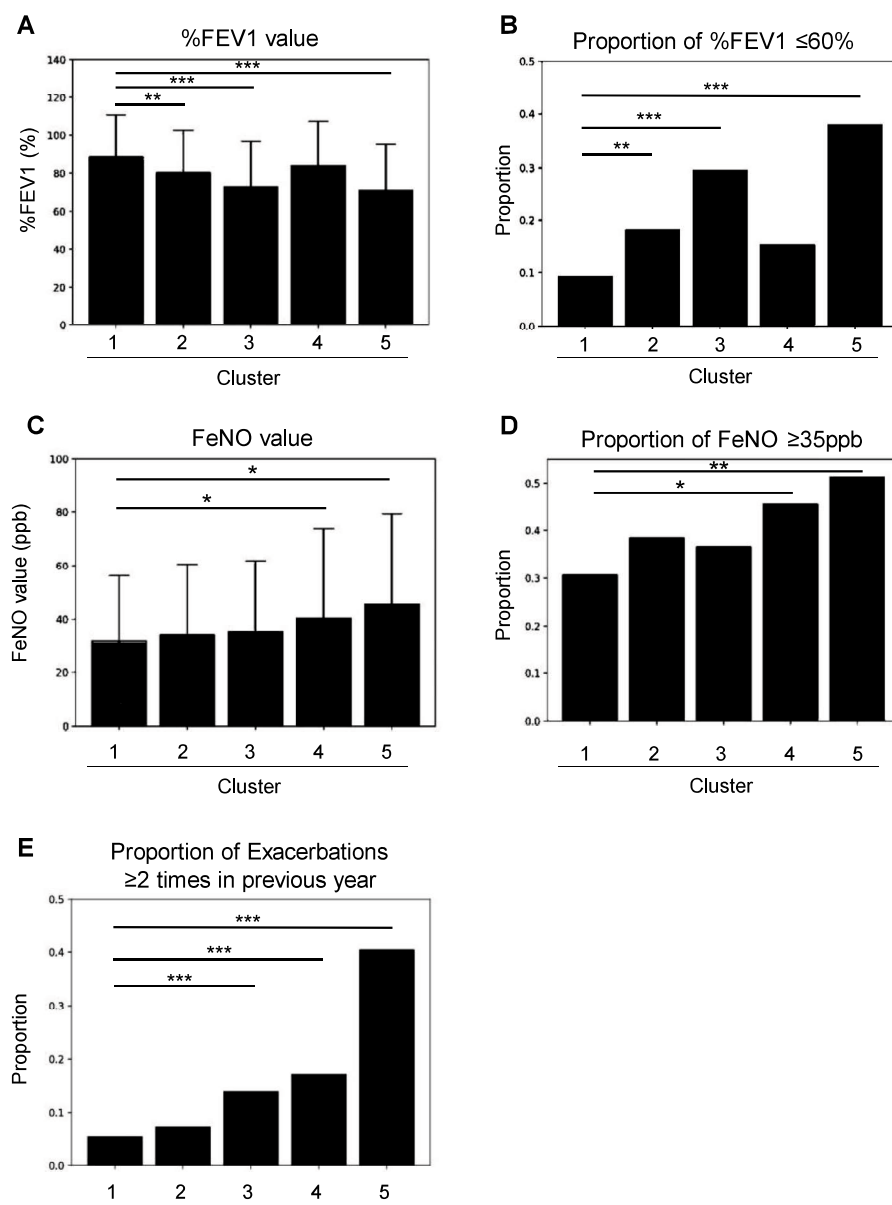


図6 各症状クラスターの重度の気流閉塞、2型気道炎症、頻回増悪について解析した。(A) は%FEV₁の平均値、(B) は%FEV₁≤60%の割合、(C) はFeNO値の平均値、(D) はFeNO≥35ppbの割合 (E) は過去1年間の喘息増悪が2回以上の割合を示す。平均値の差の検定には Student の t 検定、割合の差の検定には Fisher の正確性検定を用いた。%FEV₁, percent predicted forced expiratory volume in 1 second; FeNO, fraction of exhaled nitric oxide。* p<0.05、** p<0.01、*** p<0.0001

4.3.4 喘息症状と Treatable traits の関連性

5種類の各喘息症状（息切れ症状、喘鳴、起床時症状、夜間覚醒症状、日常生活の活動制限）と%FEV₁およびFeNO値との間でSpearman相関解析を行った。それぞれの組み合わせごとの相関係数を求めた。相関係数の値でヒートマップを作成した（図7）。息切れ症状スコアと喘鳴スコアは%FEV₁と負の相関を示し、夜間覚醒症状スコアと起床時症状スコアはFeNOと正の相関を示した。すなわち、息切れ症状・喘鳴は気流閉塞と関連し、喘鳴・夜間覚醒症状・起床時症状は2型気道炎症と関連していた。

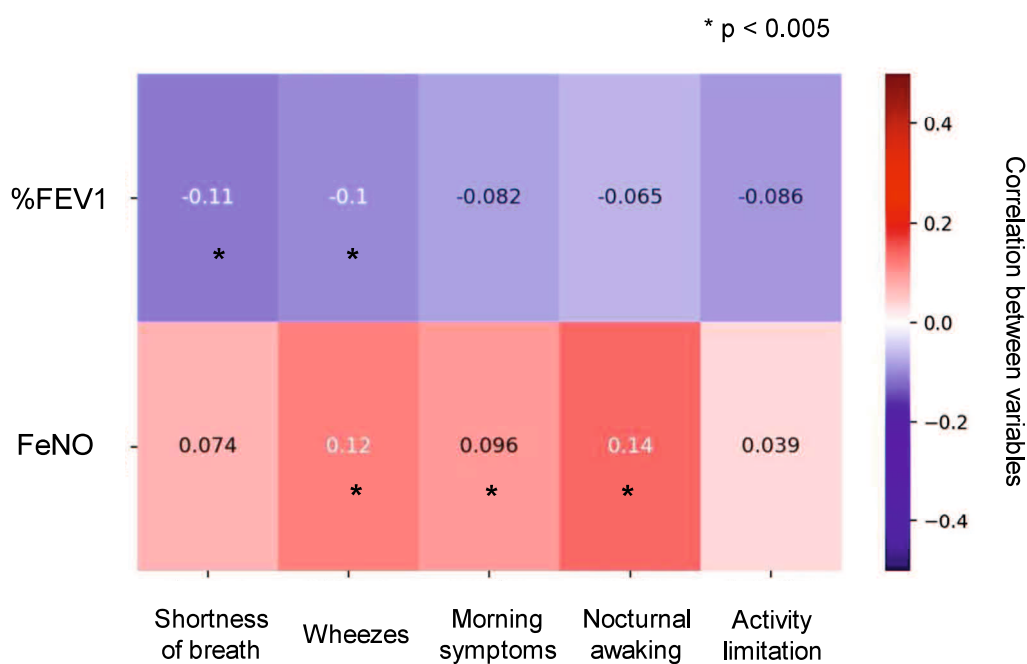


図7 5種類の喘息症状（息切れ症状、喘鳴、起床時症状、夜間覚醒症状、日常生活の活動制限）の症状スコアそれぞれと、%FEV₁、FeNOとのスピアマン相関係数を計算した。ヒートマップ内の数値は相関係数を示す。息切れ症状スコアと喘鳴スコアは、どちらも%FEV₁と負の相関を示し、夜間覚醒症状スコアと起床時症状スコアは、どちらもFeNOと正の相関を示した。すなわち、息切れ症状および喘鳴は気流閉塞と関連があり、喘鳴、夜間覚醒症状および起床時症状は、2型気道炎症と関連があった。多重検定のため、相関係数の有意性検定についてはBonferroni法を用いた補正を行った。%FEV₁, percent predicted forced expiratory volume in 1 second; FeNO, fraction of exhaled nitric oxide. *P<0.005

4.4 考察

機械学習の解析結果についての考察

本研究では、患者報告アウトカム（PRO）である ACQ-5 を用いて機械学習を行い、喘息症状から TTs である気流閉塞、2 型気道炎症、頻回増悪の有・無を推定した。まず、5 種類の喘息症状から構成される ACQ-5 の 5 次元ベクトルを、UMAP を用いて次元縮約し、二次元平面に射影した。これによって、人間の目では認識できない 5 次元の ACQ-5 データを可視化し、喘息患者の特徴的な症状パターンを明らかにした。続いて、ACQ-5 を用いて気流閉塞、2 型気道炎症、頻回増悪の有・無を推定するための教師あり学習を実施した。しかし、教師あり学習では気流閉塞、2 型気道炎症、頻回増悪の有・無を、十分な精度で推定することは出来なかった。続いて、教師なし学習である階層的クラスタリング（Ward 法）を実行し、ACQ-5 データをもとに 5 種類の特徴的な症状サブタイプ（無症状、軽度の症状、喘鳴・息切れ優位、夜間覚醒症状優位、全症状が重度）を同定した（図 8）。各々の症状サブタイプは、2 型気道炎症（FeNO ≥ 35 ppb）、重度の気流閉塞（%FEV1 $\leq 60\%$ ）、頻回増悪（過去 1 年の喘息増悪回数 ≥ 2 回）の有・無と対応していた（表 2）。

人間の目では ACQ-5 に含まれる 5 次元の症状分布パターンを認識できないため、従来、ACQ-5 に含まれる 5 種類の症状スコアの平均値を喘息評価に用いてきた。一方、本研究では、クラスタリングによって、同程度の ACQ-5 平均値の喘息患者でも症状パターンの異なる集団へ分類できた。例えば、症状クラスター 3・4 は同程度の ACQ-5 平均値を示したが、症状クラスター 3 は喘鳴・息切れ症状が優位であったのに対し、症状クラスター 4 は夜間覚醒症状・起床時症状が優位であった。

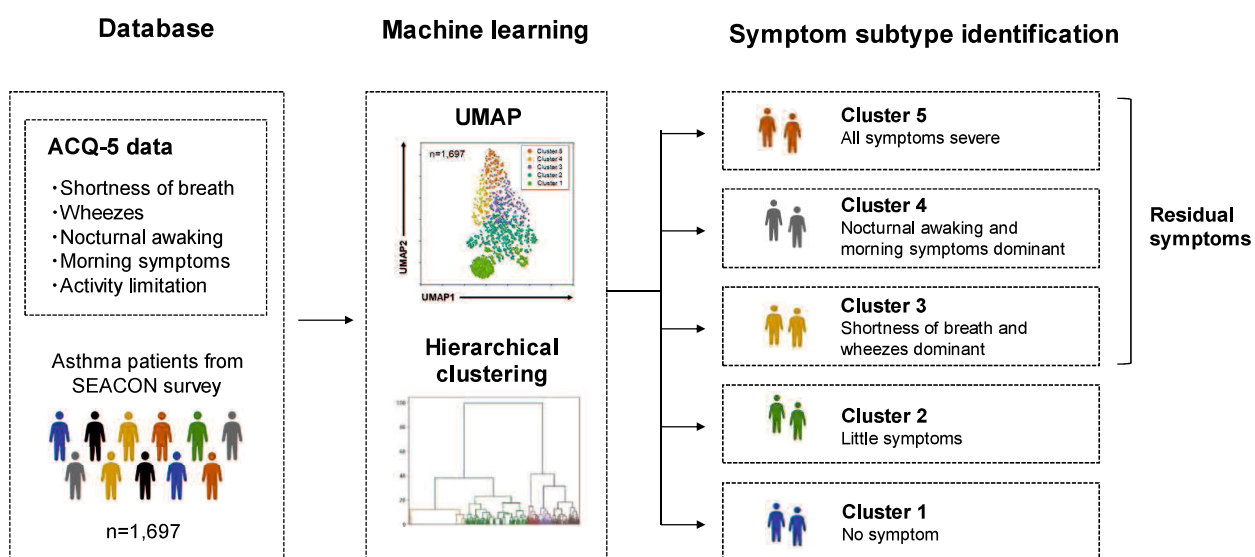


図8 機械学習の概要を示す。ACQ-5に含まれる喘息症状を用いた機械学習によって5種類の症状サブタイプ（クラスター1：無症状、クラスター2：軽度の症状、クラスター3：喘鳴・息切れ症状が優位、クラスター4：夜間覚醒症状・起床時症状が優位、クラスター5：全症状が重度）を同定した。クラスター3・4・5は喘息症状が残存している。ACQ-5, asthma control questionnaire-5; SEACON, Subjective evaluation of asthma control by physician survey; UMAP, uniform manifold approximation and projection; T2, type 2

第4章 喘息症状から病態特性を推定する技術

同定した5種類の症状サブタイプは、各々、TTsである気流閉塞、2型気道炎症、頻回増悪リスクの有・無と対応していた(表2)。クラスター3(息切れ症状・喘鳴が優位)では重度の気流閉塞($\%FEV_1 \leq 60\%$)と頻回増悪(過去1年の喘息増悪回数 ≥ 2 回)のリスクが高かった。クラスター4(夜間覚醒症状・起床時症状が優位)では2型気道炎症($FeNO \geq 35ppb$)と頻回増悪のリスクが高かった。クラスター5(全症状が重度)では、2型気道炎症、重度の気流閉塞、頻回増悪いずれのリスクも高かった。

第4章 喘息症状から病態特性を推定する技術

Symptom subtypes	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
	No symptoms	Little symptoms	Shortness of breath and wheezes dominant	Nocturnal awakening and morning symptoms dominant	All symptoms severe
Type 2 inflammation	—	—	—	✓	✓
Airway obstruction	—	✓	✓	—	✓
Exacerbation prone	—	—	✓	✓	✓

表2 5種類の症状サブタイプごとの2型気道炎症（FeNO \geq 35ppb）、重度の気流閉塞（%FEV1 \leq 60%）、頻回増悪（過去1年の喘息増悪回数 \geq 2回）の有（✓）・無（—）を示す。症状サブタイプは、各々、treatable traitsである2型気道炎症（FeNO \geq 35ppb）、重度の気流閉塞（%FEV1 \leq 60%）、頻回増悪（過去1年の喘息増悪回数 \geq 2回）の有・無と対応していた。クラスター3（息切れ症状・喘鳴が優位）では重度の気流閉塞と頻回増悪のリスクが高かった。クラスター4（夜間覚醒症状・起床時症状が優位）では2型気道炎症と頻回増悪のリスクが高かった。クラスター5（全症状が重度）では、2型気道炎症、重度の気流閉塞、頻回増悪いずれのリスクも高かった。

教師あり学習（ランダムフォレスト）では TTs である 2 型気道炎症、重度の気流閉塞、頻回増悪の有・無を十分な精度で推定できなかったが、教師なし学習（クラスタリング）では推定できた。この解析結果の違いは、教師あり学習と、教師なし学習の解析方法の違いを反映していると考えた。教師あり機械学習では、1 症例ごとに判別や予測を行う。一方、教師なし機械学習（クラスタリング）では集団ごとに特徴を分析する。教師あり機械学習で十分な精度が得られない場合でも、教師なし機械学習では臨床的意義のある結果が得られる可能性がある。

症状サブタイプと病態特性（Treatable traits）との関係性

喘鳴・息切れ症状が強くみられるクラスター3・5 では、 $\%FEV_1$ 値が低値であり高度な気流閉塞の割合（ $\%FEV_1 \leq 60\%$ ）が統計的に有意に高かった。夜間覚醒症状・起床時症状が強くみられるクラスター4・5 では、FeNO 値が高値であり 2 型気道炎症の割合が統計的に有意に高かった。喘鳴・息切れ症状と $\%FEV_1$ （気流閉塞の指標）の間には負の相関があった。喘鳴・起床時症状・夜間覚醒症状と FeNO（2 型気道炎症の指標）の間には正の相関がみられた。すなわち、それぞれの症状サブタイプ間の比較検討から示唆された喘鳴・息切れ症状と気流閉塞の関連性、並びに夜間覚醒症状・起床時症状と 2 型気道炎症の関連性は、相関分析でもみられた。これらによって、各喘息症状と、気流閉塞や 2 型気道炎症との関連性が裏付けられた。

これらの喘息症状と、気流閉塞、2 型気道炎症との関連性は、過去の研究結果とも一致する。喘鳴は、気流閉塞によって気流が乱流となり気道壁の振動が起こることで生じる [33]。息切れ症状は、気流閉塞によって肺胞内でのガス交換が障害されることで生じる [33]。夜間・早朝症状を示す喘息患者では、2 型気道炎症の指標である肺胞内 NO 濃度・喀痰好酸球比率が上昇していることが示された [33,34]。最近の報告では、夜間症状が慢性閉塞性肺疾患（Chronic Obstructive Pulmonary Disease : COPD）患者における 2 型気道炎症の併存と関連していた [35]。

頻回増悪の割合は、クラスター3・4・5 でやや高く、クラスター5 で顕著に高かつ

た。クラスター3・4はそれぞれ気流閉塞、2型気道炎症の割合が高い症状サブタイプであり、クラスター5は気流閉塞・2型気道炎症のどちらの割合も高い症状サブタイプであった。これは、気流閉塞・2型気道炎症が喘息増悪リスクを上昇させるという過去の報告と一貫した解析結果である[36,37]。

本技術の臨床現場での活用

本技術では ACQ-5 に含まれる 5 つの喘息症状スコアを入力することで、個々の喘息患者の症状サブタイプを判定し、各々の症状サブタイプにおいて、TTs である 2 型気道炎症、重度の気流閉塞、頻回増悪の有・無を推定できる。本研究で用いた SEACON survey では、ほとんどの喘息患者（95.7%）が吸入ステロイド治療を含む標準治療を受けていた[28]。すなわち、クラスター3・4・5へ分類された患者（全体の 47.8%）は、標準的治療にも関わらず症状が残存する患者である（図 5C）。本 AI 技術の活用によって、標準治療で症状が残存する喘息患者において、TTs である気流閉塞や 2 型気道炎症を標的とした追加治療薬の選択を支援できる。これにより、喘息増悪リスク抑制や症状コントロール改善につながる。

クラスター3（喘鳴・息切れ症状が優位）は高度な気流閉塞を有すると推定できるため、気流閉塞を改善させる気管支拡張薬（長時間作用性 β_2 刺激薬・長時間作用型抗コリン薬など）を追加することが有用である。クラスター4（夜間覚醒症状・起床時症状が優位）に分類される患者の TTs は 2 型気道炎症と推定できるため、2 型気道炎症を制御する治療の強化（例：吸入ステロイドの増量、生物学的治療の導入など）が有効である。クラスター5（全症状が重度）に分類される患者の TTs は 2 型気道炎症・気道閉塞・頻回増悪と推定できる。十分な患者指導と吸入治療を実施してもクラスター5に分類される患者に対しては、気管支拡張薬の追加および生物学的製剤の導入を考慮する（生物学的製剤の導入が難しい施設では呼吸器専門施設への紹介を考慮する）。

本技術の実用性

発展途上国や過疎地域などでは、専門的な検査による喘息病態の評価が困難であるため、TTsに応じた喘息治療薬の最適な選択が難しい[9,10]。これは、発展途上国や過疎地域などにおける、喘息の医療格差（死亡率増加、社会経済的損失）につながる[12-14]。本研究では機械学習を用いることで、ACQ-5に含まれる喘息状態の評価のみから、喘息のTTsを推定した。これによって、ACQ-5の評価のみから喘息治療薬の選択支援につながるため、医療格差の是正に貢献できることが期待できる。ACQ-5の取得には専門医療人材や専門的な検査機器を必要としないので、ACQ-5であれば発展途上国、過疎地域などでも取得できる。本技術をアプリケーション化あるいはクラウド化し臨床現場へ導入することで、発展途上国や過疎地域でも最適な喘息治療薬の選択を支援できる可能性がある。発展途上国や過疎地域では、近年、医療格差の是正に向けてAI技術を臨床現場へ応用する動きが進んでいる[20]。AI技術は自動で解析結果を出力できるため専門医療の一部を補完できる可能性がある。本研究で開発したAI技術を、発展途上国や過疎地域の医療機関へ導入することで、格差のない個別化治療の実現につながる。

PROを用いたAI解析

本研究では、喘息の代表的なPRO（患者の症状の主観的評価）であるACQ-5を用いて機械学習を行い、治療標的となる病態特性を推定した。本研究のように、PROはAI技術との協働によって、単なる症状評価ツールとしての役割を越え、臨床的判断を支援できる可能性を秘めている。しかし、これまでに報告されているPROを用いたAI研究においては、サンプルサイズや施設バイアスが課題として指摘されている[21]。例えば、近年、患者のPROデータから免疫チェックポイント阻害薬による免疫関連有害事象の発症を高精度に予測した研究が報告されている[38]。この研究の学習データのサンプルサイズは34症例と少なく、研究結果の一般化可能性について懸念があった。一方、本研究では、約1,700症例とサンプルサイズは十分であり、PROデータは

多施設（合計 29 施設）から収集された。以上より、本研究は一般化可能性が高いことが期待できる。

PRO は、従来、紙媒体で取得されていた。紙媒体の PRO ではデータがデジタルで入力されていないため、PRO データの利活用の障壁が高かった。一方で、近年、医療のデジタル化の潮流の中で、電子デバイスを用いた PRO の取得（electronic PRO: ePRO）が世界的に普及し始めている。ePRO では、PRO データをデジタルデータとして取得するため AI 技術を推進しやすい。今後、ePRO の普及によって、本技術を含めて PRO に基づく AI 技術の臨床現場への実用化が加速することが期待できる[20,39]。

今後の展望

本研究で用いた ACQ-5 は、横断的に取得されたスナップショットデータである。今後、縦断的に ACQ-5 データを取得し、時系列で解析を行うことで、症状や軌跡（trajectory）に関する知見を得られる可能性がある。これによって、喘息悪化の予兆を早期に捉え、先制治療につなげられるかもしれない。今後の研究で、本 AI 技術に基づき前向き研究を行うことで、本 AI 技術の臨床的有用性を検証することが望ましい。

本研究における限界

機械学習において学習データとして用いた SEACON survey では、副鼻腔炎・胃食道逆流症（Gastro-esophageal reflux disease: GERD）・精神疾患など併存症の状態についてアセスメントできていない。これら未評価の併存症の状態が、喘息症状を修飾し、TTs の推定精度に影響を与えている可能性もある。

本研究の学習データとして用いた SEACON survey では、喘息患者の平均年齢は 62.6 歳である。一般的な喘息研究コホート比べて高齢である。より若年の喘息患者を対象に同様の研究を行った際には、解析結果が部分的に変わる可能性も否定はできない。それでも、今後、学習データベースが更新され、より若年の喘息患者のデータが追加されることで、この潜在的な課題は克服し得る。

4.5 結語

喘息における PRO である ACQ-5 を用いた機械学習によって、5 種類の症状サブタイプを同定できた。それぞれの症状サブタイプは、各々、病態特性 (TTs) である気流閉塞、2 型気道炎症、頻回増悪の有・無と対応していた。本研究結果をもとに、標準治療で症状が残存する喘息患者において、TTs に応じた適切な追加治療薬の選択 (個別化治療) を支援できる。ACQ-5 データの取得には専門的な検査機器や専門技術を要さないため、医療環境が十分でない地域 (発展途上国や過疎地域など) でも治療薬選択を支援することができる。ひいては、医療格差を是正する有効手段の 1 つとなり得る。ACQ-5 をはじめとする PRO は、AI 技術との協働によって、単なる症状評価ツールとしての役割を越えて、新たな医学的知見の発見や臨床的判断を支援できる可能性を秘めている。

本研究は「日本アレルギー学会臨床研究支援プログラム」の研究費助成を受けて行った。本技術は現在、特許出願中である (特願 2022-200914 喘息病態推定サーバと喘息病態推定プログラムと喘息病態推定方法: 濱田和希、松永和人、浅井義之)。

4.6 参考文献

1. Kaur R, Chupp G. Phenotypes and endotypes of adult asthma: Moving toward precision medicine. *J Allergy Clin Immunol* 2019; 144: 1-12.
2. Agustí A, Barnes N, Cruz AA, et al. Moving towards a Treatable Traits model of care for the management of obstructive airways diseases. *Respir Med* 2021; 187: 106572.
3. Agustí A, Rapsomaniki E, Richard Beasley R, et al. Treatable traits in the NOVELTY study. *Respirology*. 2022; 27(11): 929-940.
4. McDonald VM, Fingleton J, Agustí A, et al. Treatable traits: a new paradigm for 21st

- century management of chronic airway diseases: Treatable Traits Down Under International Workshop report. *Eur Respir J* 2019; 53: 1802058.
5. Agustí A, Gibson PG, McDonald VM, Treatable Traits in Airway Disease: From Theory to Practice *J Allergy Clin Immunol Pract*. 2023 Mar;11(3):713-723.
 6. Agustí A, Bafadhel M, Beasley R, et al. Precision medicine in airway diseases: moving to clinical practice. *Eur Respir J* 2017; 50: 1701655
 7. Shaw DE, Heaney LG, Thomas M, Beasley R, Pavord ID. Balancing the needs of the many and the few: where next for adult asthma guidelines? *Lancet Respir Med* 2021;9: 786–94.
 8. McDonald VM, Clark VL, Cordova-Rivera L, et al. Targeting treatable traits in severe asthma: a randomised controlled trial. *Eur Respir J* 2020; 55: 1901509.
 9. Onyedum CC, Ukwaja KN, Desalu OO, Ezeudo C. Challenges in the management of bronchial asthma among adults in Nigeria: A systematic review. *Ann Med Health Sci Res* 2013; 3: 324-9.
 10. Cheong AT, Lee PY, Shariff-Ghazali S, et al., Implementing asthma management guidelines in public primary care clinics in Malaysia. *npj Prim Care Respir Med* 2021; 31:47.
 11. Choi JY, Yoon HK, Lee JH, et al. Nationwide pulmonary function test rates in South Korean asthma patients. *J Thorac Dis* 2018;10(7):4360-4367.
 12. Mortimer K, Reddel HK, Pitrez PM, Bateman ED, Asthma management in low and middle income countries: case for change. *Eur Respir J* 2022; 60: 2103179
 13. Stolbrink M, Thomson H, Hadfield RM, et al. The availability, cost, and affordability of essential medicines for asthma and COPD in low-income and middle-income countries: a systematic review. *Lancet Glob Health* 2022; 10: e1423–42.
 14. Meghji J, Mortimer K, Agustí A, et al. Improving lung health in low-income and middle-income countries: from challenges to solutions. *Lancet* 2021; 397: 928–40.
 15. Yokoyama A, Okazaki H, Makita N, et al. Regional differences in the incidence of asthma exacerbations in Japan: A heat map analysis of healthcare insurance claims data. *Allergol*

- Int 2022;71: 47-54.
16. Redmond C, Heaney LG, Chaudhuri R, et al. Benefits of specialist severe asthma management: demographic and geographic disparities. *Eur Respir J* 2022; 60(6):2200660.
 17. Closing Gaps in Asthma Care; World Asthma Day 2022 - Global Initiative for Asthma – GINA. < <https://ginasthma.org/closing-gaps-in-asthma-care-world-asthma-day-2022/>> accessed at May 2023.
 18. Kingsley C, Patel S. Patient-reported outcome measures and patient-reported experience measures. *BJA Education* 2017; 17 (4): 137–144.
 19. Verma D, Bach K, Mork PJ, Application of Machine Learning Methods on Patient Reported Outcome Measurements for Predicting Outcomes: A Literature Review. *Informatics* 2021, 8, 56.
 20. Guo L, Li B. The Application of Medical Artificial Intelligence Technology in Rural Areas of Developing Countries. *Health Equity* 2018, 2.1
 21. Rivera SC, Liu X, Hughes SE, Dunster H, Manna E, Denniston AK, Calvert MJ. Embedding patient-reported outcomes at the heart of artificial intelligence health-care technologies. *Lancet Digit Health* 2023; 5: e 168–73.
 22. Pearce FJ, Rivera SC, Liu X, Manna E, Denniston AK, Calvert MJ. The role of patient-reported outcome measures in trials of artificial intelligence health technologies: a systematic evaluation of ClinicalTrials.gov records (1997–2022). *Lancet Digit Health* 2023; 5: e160–67.
 23. Juniper EF, O'Byrne PM, Guyatt GH, Ferrie PJ, King DR, Development and validation of a questionnaire to measure asthma control. *Eur Respir J* 1999; 14: 902-907.
 24. Juniper EF, Bousquetb J, Abetz L, Batemand ED, The GOAL Committee Identifying ‘well-controlled’ and ‘not well-controlled’ asthma using the Asthma Control Questionnaire. *Respir Med* 2006; 100: 616–621.
 25. McInnes L, Healy J, Saul N, Großberger L. UMAP: Uniform Manifold Approximation and Projection. *J Open Source Softw* 2018; 3: 861.

26. Breiman L. Random forests. *Mach Learn* 2001; 45: 5–32.
27. Jain AK, Murty MN, Flynn PJ, Data Clustering: A Review, *ACM computing surveys (CSUR)*, 1999
28. Matsunaga K, Hamada K, Oishi K, Yano M, Yamaji Y, Hirano T. Factors Associated with Physician-Patient Discordance in the Perception of Asthma Control. *J Allergy Clin Immunol Pract* 2019; 7: 2634-41.
29. Chikumoto A, Matsunaga K, Oishi K, Hamada K, Yamaji Y, Hirano T. What are key symptoms to be captured for preventing a physician’s underestimation of asthma control? *J Allergy Clin Immunol Pract*. 2021; 9(1): 564-566.
30. Kohavi R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence* 1995;14(12): 1137-1145.
31. Joe H. Ward. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association* 1963; 58: 236-244
32. Jakeways N, McKeever T, Lewis SA, Weiss ST, Britton J. Relationship between FEV1 reduction and respiratory symptoms in the general population. *Eur Respir J* 2003; 21: 658–663.
33. Lehtimaki L, Kankaanranta H, Saarelainen S, Turjanmaa V, Moilanen E. Increased alveolar nitric oxide concentration in asthmatic patients with nocturnal symptoms. *Eur Respir J* 2002; 20: 841–845.
34. Durrington HJ, Gioan-Tavernier GO, Blaikley JF, Singh D, Smpson A, Ray DW. Time of Day Affects Eosinophil Biomarkers in Asthma: Implications for Diagnosis and Treatment. *Am J Respir Crit Care Med* 2018; 198(12):1578-1581.
35. Suzuki Y, Nagase H, Toyota H, et al. Questionnaire for diagnosing asthma-COPD overlap in COPD: Development of ACO screening questionnaire (ACO-Q). *Allergol Int* 2023 in press.
36. Kitch BT, Paltiel DA, Kuntz KM, et al. A Single Measure of FEV1 Is Associated with Risk

- of Asthma Attacks in Long-term Follow-up. *CHEST* 2004; 126:1875–1882.
37. Couillard S, Laugerud A, Maisha Jabeen M, et al. Derivation of a prototype asthma attack risk scale centred on blood eosinophils and exhaled nitric oxide. *Thorax* 2022; 77: 199–202.
 38. Iivanainen S, Ekstrom J, Virtanen H, Kataja VV, Koivunen JP. Electronic patient-reported outcomes and machine learning in predicting immune-related adverse events of immune checkpoint inhibitor therapies. *BMC Med Inform Decis Mak* (2021) 21:205.
 39. Knapp A, Harst L, Hager S, Schmitt J, Scheibe M. Use of Patient-Reported Outcome Measures and Patient-Reported Experience Measures Within Evaluation Studies of Telemedicine Applications: Systematic Review. *J Med Internet Res* 2021; 23(11): e30042.

第5章

総括

5.1 本論文のまとめ

本研究では、呼吸器診療を支援する AI 技術を開発した。

以下に各章で得られた成果について要約する。

第2章では、副作用報告ビッグデータを用いたベイズ推定モデルを構築することで、副作用の原因薬剤の推定技術を開発した。精度検証の結果、臨床現場への実用化に耐える高精度で副作用の原因薬を推定できた。本技術は、治療薬による副作用の管理向上や、医療者の業務時間短縮につながる。

第3章では、喘息患者の急速な FEV₁ 低下 (rapid decline) を高精度で予測する機械学習モデルの開発について述べた。この技術を応用することで、治療強化が必要な喘息患者を同定し、気流閉塞の進行を防ぐための早期の治療強化 (先制治療) を行う。これによって、喘息の寛解や重症化予防につながり、ひいては健康寿命の延伸に貢献する。

第4章では、喘息質問票 (ACQ-5) のデータを用いて教師なし機械学習を行うことで、喘息の症状サブタイプを同定し、その背景にある病態特性を推定する技術を開発した。これによって ACQ-5 に含まれる喘息状態の評価のみから、喘息の病態の多様性に応じた治療選択 (個別化治療) につなげることができる。ACQ-5 の取得には専門的な検査機器を必要としない。本技術は、発展途上国、過疎地域、プライマリケアの現場など、医療環境が不十分な地域で、適切な喘息治療薬の選択を支援できる可能性がある。ひいては、グローバルな医療格差の是正につながる。

本研究で開発した3つの AI 技術は、呼吸器診療の場面ごとに診療を支援する。治療管理の方針決定の場面では、第3章で述べた「喘息患者の急速な呼吸機能低下を予測する AI 技術」を活用する。重症化リスクの層別化を行い、治療強化が望ましい患者を

同定する。これによって重症化予防・寛解率向上のための先制医療につなげる。次に、治療選択の場面では、第4章で述べた「喘息症状から治療標的となる病態を推定するAI技術」を活用し、患者の病態の個別性に応じた治療薬の選択（個別化治療）を支援する。治療後に副作用が生じた際には、第2章で述べた「ベイズ推定に基づく副作用の原因薬推定技術」を活用し、副作用管理を支援する。このように、本研究で開発したAI技術を臨床現場で活用することで、一連の診療の流れの中でデータ駆動的医療を実現することができる（図1）。このように診療場面ごとに、患者の個別性に応じて治療方針を決定する概念は、動的治療計画（Dynamic Treatment Regime, DTR）として近年注目を浴びている[1]。

本研究で開発したAI技術は、いずれも、リアルワールドにおける観察研究のデータベースあるいは医療ビッグデータを学習データとして用いた。リアルワールドデータとは、臨床試験のような実験的環境ではなく、日常診療の環境で収集された医療データのことである[2]。第2章、第3章で学習データベースとして用いたFLOAT study、SEACON surveyは喘息患者のリアルワールドの観察研究である。第4章で用いたJADERデータベースは、日本国内のリアルワールドにおける副作用報告を集積したビッグデータである。

リアルワールドデータに基づくAI技術開発は、臨床現場への実装を進めやすいことが期待できる。その理由として、リアルワールドデータの一般化可能性が高いことが挙げられる。一般的に、リアルワールドの患者集団は、臨床試験のように厳格な基準で選択された均質な患者集団に比べ、多様であり、かつ集団規模も大きい。すなわち、リアルワールドデータは患者集団の真の母集団を反映している可能性が高い。このことから、臨床試験など厳格な基準で選択された患者を対象とした場合に比べ、リアルワールドデータに基づくAI技術は、汎化性能がより高いことが期待できる。ただし、各々のリアルワールドデータを用いたAI技術の汎化性能は、施設、地域、都道府県、国などデータを集積する場面に依存する点に注意が必要である。リアルワールドデータに基づいたAI技術は、日常診療へ円滑に導入できる。日常診療範囲内のみの臨床情報を入力することで、解析結果を出力できるからである。

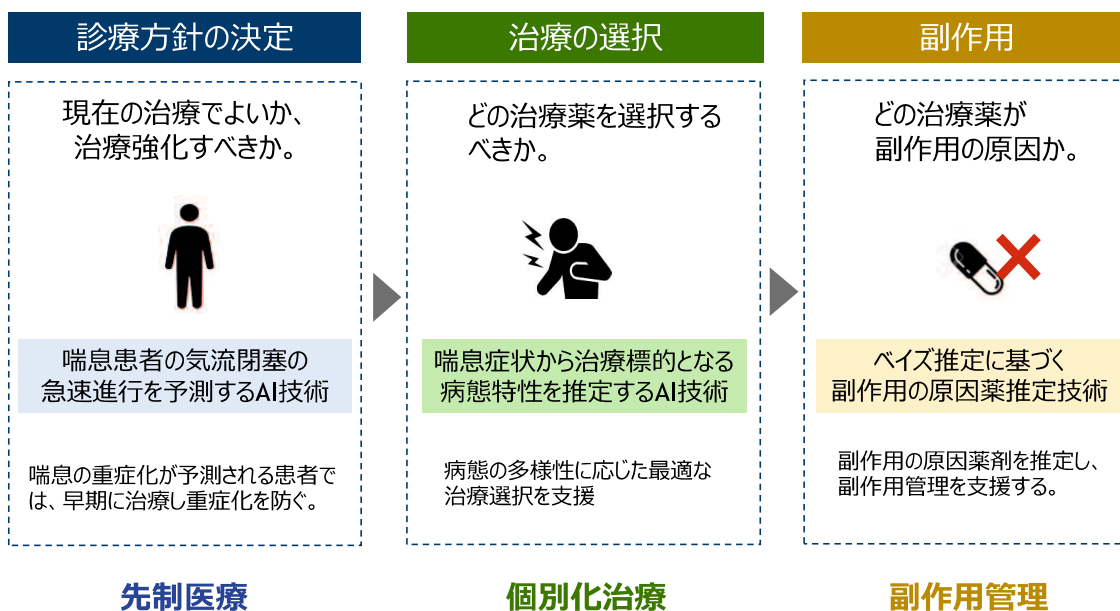


図1 診療場面ごとの AI 技術の活用を示す。診療方針の決定の場面においては、重症化リスクの層別化により早期治療が必要な患者を同定し、先制治療につなげる。治療選択の場面では、治療標的となる病態を推定し、病態の多様性に応じた最適な治療選択を支援する。治療によって副作用が生じた場面では、副作用の原因薬剤を推定し、副作用管理を支援する。本研究で開発した AI 技術を活用することで、一連の呼吸器診療の流れの中で、データ駆動的医療を実現することができる。

5.2 今後の展望

今後、本研究で開発した AI 技術を実用化し、臨床現場へ応用することで先制医療や個別化治療の実現を目指す。これによって、健康寿命の延伸、医療費抑制につながる。AI 技術では、データを与えることで、自動的に判断できる。専門医療の一部を AI 技術によって補完することで、医療者の業務負担軽減にもつながる。これらによって、医療人材不足の改善に貢献することも期待できる。AI 技術の実用化に向けた具体的な手段として、電子カルテシステムへの実装、アプリケーション開発、クラウドシステム開発などが挙げられる。

開発した AI 技術を遠隔医療として実装することで、医療格差の是正（医療の均てん化）につながる[3]。発展途上国や医療過疎地域では、専門医療を担う人材が不足している。一方、高齢者は多数の併存症を抱えており、複数の専門分野にまたがる場合が多い。専門医療を担う人材が不足する地域では、専門分野以外の診療についても意思決定をしていく必要がある。しかし、AI 技術によって専門医療を補完できるようになれば、現在は都市部の専門施設でしか受けられないような専門医療も、あらゆる地域の医療機関で実現できる可能性がある。これによって、医療の質が均てん化され、医療格差の是正につながる。

AI 技術の開発と、その応用を効率的に推進していくには、AI 技術の開発および応用を日常診療のワークフローの中に組み込むことが有効である。そのためには、様々な医療現場での医療情報を、デジタル形式で自動的に蓄積する必要がある。蓄積される医療データは構造化されていることが望ましい。構造化されていない医療データでは、データ前処理（クリーニング）に多大な時間・労力・コストを消費してしまうからである。

同時に、前向き研究によって開発した AI 技術の有用性を検証することが望ましい。AI 技術に基づいた介入研究を行う際には、CONSORT-AI 声明を遵守して研究計画を立案し、研究を遂行する必要がある[4]。

医療現場への AI 技術導入の際の課題として、法的責任が挙げられる。現在の日本に

においては、診療の中で AI 技術を用いた場合にも法的責任は医療者に求められる[5,6]。しかし、AI 技術の進歩によって、一部のタスクにおいては、AI が人間の判断能力を上回る場面が増加すると想定される。すなわち、人間による判断と AI 技術による判断が不一致となる場面の増加である。そのような中で、判断の誤りの責任の所在について、改めて議論が必要となる[7]。

最後に、第一章「序論」でも述べた通り、本論文では、システムバイオロジーと AI 技術の融合的アプローチに向けた第一歩として、まずは AI 技術を用いた医学研究を行った。今後は、システムバイオロジーと AI 技術の融合的アプローチを目指し、病態生理の深い理解を踏まえた上で、データ駆動型技術の医療応用につなげたいと考えている。また、本研究で得た3つの成果は、呼吸器診療における、ごく限られた臨床場面において活用されうる技術である。医療全体の課題を解決するためには、今後、専門分野や職種の垣根を越えた、さらなる研究開発が必要である。

5.3 参考文献

1. Mahar RK, McGuinness MB, Chakraborty B, Carlin JB, IJzerman JM, Simpson JA. A scoping review of studies using observational data to optimise dynamic treatment regimens. *BMC Medical Research Methodology* (2021) 21:39.
2. ヒューマンサイエンス振興財団「医療リアルワールドデータの利活用：安全性調査、臨床研究、製造販売承認申請にいかに関与するか」創薬技術調査報告書 Part 2 2019年3月
3. Guo L, Li B. The Application of Medical Artificial Intelligence Technology in Rural Areas of Developing Countries. *Health Equity* 2018, 2.1.
4. Liu X, Samantha Cruz Rivera, David Moher, Melanie J Calvert, Alastair K Denniston, and the SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension.

Lancet Digital Health 2020; 2: e537–548.

5. 厚生労働省「AI を用いた診断、治療等の支援を行うプログラムの利用と医師法第17条の規定との関係について」第4回 保健医療分野 AI 開発加速コンソーシアム資料 <https://www.mhlw.go.jp/stf/shingi2/0000148680_00005.html>平成31年1月16日
6. 厚生労働省「人工知能（AI）を用いた診断、治療等の支援を行うプログラムの利用と医師法第17条の規定との関係について」医政医発1219第1号 平成30年12月19日
7. 松尾剛行「健康医療分野における AI の民刑事責任に関する検討: AI 画像診断(支援)システムを中心に」(13):2019.9

謝辞

4年間にわたる博士課程での研究遂行にあたり、指導教員として、日頃から親身になっご指導いただいた山口大学大学院医学系研究科 システムバイオインフォマティクス講座/医学部附属病院 AI システム医学・医療研究教育センター (AISMEC) の浅井義之教授に深く御礼を申し上げます。データサイエンスやプログラミングの学問的な面や技術面だけではなく、研究を行う者としての姿勢や哲学についても多くのことをご教授頂きました。この場をお借りして厚く御礼申し上げます。

AISMEC の安部武志助教からは、常日頃から研究内容についてご助言・ご討議を頂き、データサイエンスやプログラミングについて、どんな些細なことでも丁寧に教えて頂きました。心より感謝申し上げます。AISMEC の中津井雅彦教授、同研究科システムバイオインフォマティクス講座の早野崇英講師からも、研究遂行にあたり日々貴重なご助言・ご討論を頂き、深謝申し上げます。

同研究科 呼吸器・感染症内科学講座の松永和人教授からは、呼吸器内科学の観点から貴重なご助言・ご討論を頂き、おかげさまで研究を円滑に進めることができました。この場をお借りして厚く御礼申し上げます。

また、副作用の原因薬剤推定技術の共同研究者として副作用データ集積へのご協力を頂き、臨床薬学の観点から貴重なご意見を賜りました、山口大学医学部附属病院薬剤部の北原隆志教授、防府消化器病センター薬剤部の幸田恭治先生に深く感謝いたします。

開発したシステム実用化に向けてご指導並びにご支援賜りました、山口大学医学部附属病院医療情報部の石田博教授、平野靖准教授、AISMEC の中津井雅彦教授に厚く御礼申し上げます。医療現場への技術の還元、橋渡しをしていただいたお陰で、臨床現場への実用化につなげることができました。

4年間の研究生活を送ることが出来たのは、博士課程の研究に関わって頂いたすべての皆さま方の温かいご支援、ご指導のお陰であります。心より感謝申し上げます。

謝辞

この4年間の博士課程で学んだことを糧に、今後も研究と日々の診療に励んで参ります。今後ともご指導ご鞭撻のほど宜しくお願い申し上げます。

付録

本付録は、第 2 章「ベイズ推定に基づく副作用原因薬推定技術の開発」の「2.2.2 ベイズ推定モデルの構築」の過程について詳細に述べたものである。

1. 副作用における原因薬剤の尤度推定

本節は、ベイズ統計を用いて副作用の原因薬剤の推定する過程について要約したものである。日本国内の医薬品副作用報告データベース（JADER）に基づいて、医薬品が副作用（ADR）のである可能性を推定するアルゴリズムを開発した。ベイズ統計では観測データから決定則を定義する。本研究においては、観測データとして JADER データベースから取得された 730,000 件以上の医薬品-副作用データセットを用いた。これらのデータを前処理し、副作用を生じやすい薬剤に関するサンプル標本が得ることができる。

ある特定の副作用名を ADR 、投与薬剤の集合を $D = \{d_1, d_2, \dots, d_i, \dots, d_k\}$ とする。 d_i は投与薬剤名、 k は投与された薬剤の数を表す。 $d_i = c$ は、 d_i が副作用 ADR の原因薬剤であることを示す。 ADR と投与薬剤の集合 $D = \{d_1, d_2, \dots, d_i, \dots, d_k\}$ が与えられた場合に、 d_i が副作用 ADR の原因である条件付き確率は次のように表される。

$$P(d_i = c | ADR, d_1 \in D, d_2 \in D, \dots, d_i \in D, \dots, d_k \in D) \quad (1.1)$$

最尤推定においては、以下の定義により推定原因薬剤 \hat{d}_i の決定則が導出される。

$$\hat{d}_i = \arg \max_{d_i} [P(d_i = c | ADR, d_1 \in D, d_2 \in D, \dots, d_i \in D, \dots, d_k \in D)] \quad (1.2)$$

ベイズの定理によって、(1.1) は以下のように変形できる。

$$\frac{P(d_i = c, d_1 \in D, d_2 \in D, \dots, d_i \in D, \dots, d_k \in D | ADR)}{P(d_1 \in D, d_2 \in D, \dots, d_i \in D, \dots, d_k \in D | ADR)} \quad (1.3)$$

条件付き確率 (1.3) の最大値を求めるためには、本来、結合確率 $P(d_1 \in D, d_2 \in D, \dots, d_i \in D, \dots, d_k \in D | ADR)$ を推定する必要がある。しかし、JADER データベースには、現実的に取り得る投与薬剤の組み合わせのごく一部しか含まれないため、この結合確率を直接推定することは困難である。ただし、いくつかの仮定を置くことで、観測データから尤度を推定することが可能である。詳細を以下に記す。

式 (1.3) の分母は投与薬剤同士の比較においては共通項となるため、(1.1) は以下の確率に比例する。

$$P(d_i = c, d_1 \in D, d_2 \in D, \dots, d_i \in D, \dots, d_k \in D | ADR) \quad (1.4)$$

ベイズの定理によって、結合確率 (1.4) は以下の数式に分解される。

$$\frac{P(d_1 \in D, d_2 \in D, \dots, d_k \in D | d_i \in D, d_i = c, ADR)}{P(d_i \in D | d_i = c, ADR)} \cdot P(d_i = c | ADR) \quad (1.5)$$

薬剤の相互作用を無視できると仮定すると（すなわち、特定の副作用 ADR が発生したもとの $i \neq j$ の場合）、 $d_i = c$ および $d_j \in D$ は条件付き独立となるため、(1.5) は以下のように単純化される。

$$\frac{P(d_i = c | ADR)}{P(d_i \in D | ADR)} \cdot \prod_{j=1}^k P(d_j \in D | ADR) \quad (1.6)$$

$\prod_{j=1}^k P(d_j \in D | ADR)$ は、各 i に関して共通項となるため、(1.6) は薬剤 d_i が特定の副作用 ADR の原因薬剤である確率に比例する。

$$\frac{P(d_i = c|ADR)}{P(d_i \in D|ADR)} \cdot \prod_{j=1}^k P(d_j \in D|ADR) \propto \frac{P(d_i = c|ADR)}{P(d_i \in D|ADR)} \quad (1.7)$$

JADER データベース内における薬剤ごとの副作用 ADR の報告頻度のデータに基づいて、(1.7) の右辺を推定できる。すなわち、以下の通りである。

$$\frac{P(d_i = c|ADR)}{P(d_i \in D|ADR)} \approx |(d_i = c|ADR)|/|(d_i \in D|ADR)| \quad (1.8)$$

ここで $|(d_i = c|ADR)|$ は JADER データベースで 薬剤 d_i が 特定の副作用 ADR の原因薬剤である (JADER データベース上で「被疑薬」とラベル付けされている) 報告件数を示す。 $|(d_i \in D|ADR)|$ は、データベースの特定の副作用 ADR において薬剤 d_i が報告された件数の総数を示す (JADER データベース上で、薬剤 d_i に対して「被疑薬」とラベル付けされるか否かは問わない)。薬剤相互作用を無視できると仮定すると、投与薬剤同士の事後確率分布の大小関係は、比率 $|(d_i = c|ADR)|/|(d_i \in D|ADR)|$ の大小関係に一致する。したがって、投与薬剤間の比較の課題においては、次の近似式を得ることができる。

$$\begin{aligned} P(d_i = c|ADR, d_1 \in D, d_2 \in D, \dots, d_i \in D, \dots, d_k \in D) \\ \approx |(d_i = c|ADR)|/|(d_i \in D|ADR)| \end{aligned} \quad (1.9)$$

(1.9) から次の近似式を得ることができる。

$$\begin{aligned} \arg \max_{d_i} [P(d_i = c|ADR, d_1 \in D, d_2 \in D, \dots, d_i \in D, \dots, d_k \in D)] \\ \approx \arg \max_{d_i} [|(d_i = c|ADR)|/|(d_i \in D|ADR)|] \end{aligned}$$

投与薬剤ごとに比率 $|(d_i = c|ADR)|/|(d_i \in D|ADR)|$ を計算する。最尤推定の概念に基

づいて、特定の副作用 *ADR* の原因薬剤を推定することができる。

2. 副作用原因薬についてのベイズ予測推定

本節では、副作用の原因薬剤である確率スコア (CPS) を推測する手法の背後にあるベイズ理論について記述する。まず、特定の副作用 *ADR* において投与薬剤の情報が与えられた条件下で、原因薬剤に関するディリクレカテゴリ変数モデルを定義する。

無情報事前分布を明示することによって、ベイズ推定モデルの予測推論が CPS に基づく順位付けを可能にすることが示されている [1]。投与薬剤を $d_i (i = 1, \dots, k)$ と表し、 k はその投与薬剤の総数を表す。原因薬剤を同定する課題を、以下の方程式の解を求めることに置き換える。

$$\operatorname{argmax}_i p_i$$

ここで、 p_i は d_i が副作用の原因薬剤である (未知の) 確率を表している。 $\sum_{i=1}^k p_i = 1$ となる。ベイズ法で p_i を推定するために、 d_i が特定の副作用 *ADR* の原因薬剤である確率 p_i がカテゴリ分布に従うと仮定する。

$$i \sim \text{categorical}(p_1, \dots, p_k).$$

つまり、 p_i はモデルのパラメーターとなる。計算を容易にするために、共役事前分布としてディリクレ分布を以下のように用いる。

$$(p_1, \dots, p_k) \sim \text{Dirichlet}(\alpha_1, \dots, \alpha_k) \quad (2.1)$$

ここで、 $\alpha_i (i = 1, \dots, k)$ はベイズ推定モデルのハイパーパラメーターである。特定の副作用 *ADR* の件数をカウントするための副作用データベースから取得したデータ X を

使用し事後分布を確立する。 $n_i (i = 1, \dots, k)$ は、薬剤 d_i が副作用の原因薬であるとラベル付けされた副作用報告の件数を示す。すると、事後分布は以下の通りとなる。

$$f(p_1, \dots, p_k | X) = \text{Dirichlet}(\alpha'_1, \dots, \alpha'_k)$$

ここで、 $\alpha'_i = \alpha_i + n_i$ である。したがって、副作用ADRが発生した際に原因薬剤である事後確率は、

$$p_i = P(d_i = c | X, \alpha_1, \dots, \alpha_k) = \frac{\alpha_i + n_i}{\sum_{j=1}^k (\alpha_j + n_j)} \quad (2.2)$$

となる。

本研究では、利用可能な事前知識がないため、すべての i に対して $\alpha_i \approx 0$ として同じ定数値を採用した。これは、無情報事前分布としたことを意味する。この場合、上記の事後確率分布予測によって定義される順位付けは、CPSの順位に一致する。

3. 不均衡分析 (Disproportionality analysis) のための統計手法

本節では、安全性シグナル検出のための不均衡分析で使用される、2種類の比率統計量についてまとめた。1つは log likelihood ratio (LLR) であり、もう1つは、Proportional reporting ratio (PRR) である。本研究では、ベイズ推定モデルが出力するCPSが原因薬剤の推定を精度するについて比較するために、投与薬剤間の順位付けを提供するために適用した。大規模な副作用報告データベース (JADER、FAERS) には、多数の医薬品と副作用が含まれている。これらのデータベースでは、I行 (医薬品) とJ列 (副作用) のデータ表を作成できる。 n_{ij} をデータ表内の i 番目の薬剤かつ j 番目の副作用のセルの件数とする。 $n_{i\cdot}$ と $n_{\cdot j}$ を、それぞれ表内の i 番目の薬剤の合計件数、 j 番目の副

作用の合計件数とする。表内の総件数は、 $n_{..}$ として表される。Log likelihood ratio test (LRT) 法では、 i 番目の薬剤と j 番目の副作用についての尤度関数 LR_{ij} を次式のよ
うに定義している。

$$LR_{ij} = \left(\frac{n_{ij}}{n_{i.}}\right)^{n_{ij}} \left(\frac{n_{.j} - n_{ij}}{n_{..} - n_{i.}}\right)^{n_{.j} - n_{ij}} / \left(\frac{n_{.j}}{n_{..}}\right)^{n_{.j}}$$

PRR は以下の比率として定義されている。

$$PRR_{ij} = \frac{n_{ij}/n_{i.}}{(n_{.j} - n_{ij})/(n_{..} - n_{i.})}$$

LRT 法では、過去の研究で定義されたように尤度関数 LR_{ij} の対数 LLR_{ij} を使用した。検証データセットにおいて、各投与薬剤 $d_i (i = 1, \dots, k)$ に関し LLR_{ij} と PRR_{ij} を計算し、それぞれの値に基づいて降順で順位付けを行った。 LLR_{ij} と PRR_{ij} の順位を用いて、ROC 解析を行い、検証データセットにおける副作用の原因薬剤を推定した。2種類の比率統計量 (LLR_{ij} , PRR_{ij}) の詳細については過去の研究を参照のこと [2,3]。

4. Bayesian Confidence Neural Network (BCPNN)

ファーマコビジランスでは LRT、PRR、ROR など不均衡分析に基づく統計的方法に加えて、Bayesian Confidence Neural Network (BCPNN) と呼ばれるベイズ主義に基づいた統計的手法も提案されている。BCPNN も、上述の LRT、PRR、ROR と同様、未知の副作用シグナル検出のための方法論である。BCPNN では、情報コンポーネント (Information component: IC) と呼ばれる統計量を考慮する。 i 番目の薬剤と j 番目の副作用についての情報コンポーネント IC_{ij} は、以下の通りに定義される[2]。

$$IC_{ij} = \log_2 \left(\frac{\theta_{ij}}{\theta_{i.} \times \theta_{.j}} \right)$$

ここで、 θ_{ij} は医薬品と副作用の大規模データベースの表内において、 i 番目の薬剤が投与される確率と j 番目の副作用の結合確率を表し、 θ_i は i 番目の薬剤が投与される確率、 θ_j は j 番目の副作用が生じる確率を表す。この定義に基づいて、以下の通りに IC の期待値である $E(IC_{ij})$ を計算し、BCPNN 法の統計量として得る（BCPNN 法における情報コンポーネント IC_{ij} の詳細については、文献[2]を参照のこと）。

$$E(IC_{ij}) = \log_2 \frac{(n_{ij} + 1)(n_{..} + 2)^2}{(n_{..} + \gamma)(n_{i.} + 1)(n_{.j} + 1)}$$

参考文献

1. Tu, S. The dirichlet-multinomial and dirichlet-categorical models for bayesian inference. Computer Science Division, UC Berkeley, 2 (2014).
<https://people.csail.mit.edu/stephentu/writeups/dirichlet-conjugate-prior.pdf>
2. Huang, L., Guo, T., Zalkikar J.N., & Tiwari, R.C. A Review of Statistical Methods for Safety Surveillance. Ther Innov Regul Sci. 48, 98-108 (2014)
3. Huang, L., Zalkikar J.N., & Tiwari, R.C. A Likelihood Ratio Test Based Method for Signal Detection With Application to FDA’s Drug Safety Data. J Am Stat Assoc. 106, 1230-1241 (2011)