

進行がん患者の治療期の専門的緩和ケアニーズ
を予測する機械学習モデルの構築

名古屋大学大学院医学系研究科

総合保健学専攻

川島 有沙

2024 年度 学位申請論文

進行がん患者の治療期の専門的緩和ケアニーズ
を予測する機械学習モデルの構築

名古屋大学大学院医学系研究科

総合保健学専攻

(指導：佐藤 一樹)

川島 有沙

目次

学位申請論文の要旨.....	
Abstract	
1.1. 背景.....	1
1.2. 目的.....	4
2. 方法.....	5
2.1. 対象.....	5
2.2. 研究デザインとデータソース.....	5
2.3. 正解定義.....	6
2.4. 調査手順.....	7
2.4.1. 専門的緩和ケアニーズに関連する記載の診療録調査.....	7
2.4.2. 専門的緩和ケアニーズに関連する記載がない場合の診療録調査.....	7
2.4.3. がん・緩和ケアの専門家による調査.....	8
2.4.4. 正解データの信頼性.....	8
2.5. 調査項目.....	9
2.6. 統計解析.....	11
2.7. 機械学習アルゴリズム.....	12
2.8. 倫理的配慮.....	15
3. 結果.....	15
3.1. 対象者背景.....	15
3.2. 正解データ.....	15
3.2.1. 苦痛スクリーニング.....	15
3.2.2. 専門的緩和ケアニーズのラベル付けと信頼性.....	16
3.2.3. 専門的緩和ケアニーズの内訳.....	16
3.3. 苦痛スクリーニングモデル.....	17
3.4. 機械学習モデル.....	17

3.4.1. 全変数投入モデル「ALL-model」	17
3.4.2 苦痛スクリーニング以外のモデル「NotSCR-model」	18
3.4.3 痛み以外の3項目の苦痛スクリーニングを除いた「NotSCR+Pain-model」	18
3.4.4 苦痛スクリーニングのみのモデル「SCR-model」	18
3.4.5 特徴量選択後の5項目の「Simple-model」	19
3.5. 特徴量の重要度	19
4. 考察	19
4.1. 主な知見	20
4.2. 専門的緩和ケアニーズの有病率と分布	20
4.3. 機械学習モデルの予測性能	21
4.4. ALL-model と、NotSCR/NotSCR+Pain/SCR-model の比較	22
4.5. 臨床への示唆	24
4.6. 看護への示唆	26
4.7. 研究の強み	26
4.8. 研究の限界	27
5. 結論	30
6. 謝辞	30
7. 図表付録	32
8. 引用文献	50

学位申請論文の要旨

進行がん患者の治療期の専門的緩和ケアニーズを予測する機械学習モデルの構築

1. 緒言

緩和ケアとは、重篤な疾患を持つ患者の心と体のつらさを軽減し、生活の質を向上させるためのケアである。進行がん診断後の早期緩和ケアにより生命予後が延長する可能性が明らかにされているが、専門的緩和ケアが必要となる時期やニーズの予測は困難であり、緩和ケアへの紹介は遅れやすい。専門的緩和ケアを必要とするがん患者を特定するための苦痛スクリーニングは、人員や時間の制約から普及が進まず、全てのがん患者での評価は困難な現状がある。そこで本研究では、化学療法を受ける進行がん患者を対象に、診療録データを用いてがん患者の治療期の専門的緩和ケアニーズを予測する機械学習モデルを構築し、その予測精度を評価することを目的とする。また、苦痛スクリーニング結果を含めた機械学習モデルと、スクリーニング結果を含まない機械学習モデル、スクリーニング結果のみの機械学習モデルで予測性能と変数重要度を比較し、臨床実装を考察する。

2. 対象及び方法

2018年4月1日から2023年3月31日までに名古屋大学医学部附属病院で外来化学

療法と苦痛スクリーニングを受けた、18 歳以上の進行固形がん患者を対象とした。進行がんの定義は、遠隔転移の存在とステージIVの診断とした。化学療法のうち、根治手術の術前術後に行う補助化学療法、化学放射線療法、治験は除外した。また、専門的緩和ケアニーズを予測するための情報は、外来診療録のみからは十分に得られないと予想されたため、苦痛スクリーニングの実施前 3 か月間に入院歴がない患者は除外した。

研究デザインは、診療録の既存情報を用いたレトロスペクティブコホート調査である。過去の診療録の回顧的な診療録調査と緩和ケア専門家による評価を行い、専門的緩和ケアニーズを正解データとした、教師あり機械学習の手法を用いた。専門的緩和ケアニーズの有無の評価では、ニーズの内容の内訳も調査した。データセットの 80% を学習用データ、20% を検証用データに分割し、モデルの学習と検証を行った。学習用データにおいては 5 分割交差検証を行い、グリッドサーチによるハイパーパラメータの調整を行った。不均衡データへの対処のため、専門的緩和ケアニーズなし群からランダムにアンダーサンプリングを行い、専門的緩和ケアニーズあり群と同数にした均衡なデータセットでの機械学習モデルも開発した。予測に用いる情報は、院内がん登録、診断群分類包括評価とレセプトデータ、入院時看護記録などであり、モデル構築には機械学習アルゴリズムの勾配ブースティング決定木：XGBoost を用いた。また、診療録データから構築された機械学習モデルの性能と、苦痛スクリーニングの性能と

を比較し、機械学習モデルにおける予測の変数重要度を算出した。

3. 結果

対象期間内に化学療法を受けた進行がん患者 1,878 名のうち、適格基準に該当した 561 名を分析対象とした。対象者の平均年齢は 65.7 歳、女性が 39.0%であり、主ながん原発部位は消化管や肝臓、胆道、膵臓であった。分析対象者のうち、専門的緩和ケアニードは 114 名(20.3%)で陽性であり、頻度が高かったニードの種類は「心理的苦痛」(87.7%)、「身体的苦痛」(57.0%)、「患者本人の緩和ケアの希望」(26.3%)であった。苦痛スクリーニングの結果を含む 12 の変数で作成したモデルの評価指標は、ROC 曲線下面積 (AUC) 0.89 [95%信頼区間; 0.80, 0.96]、感度 95.8% [79.8, 99.3]、特異度 71.9% [61.8, 80.2]と最も良い性能を示した。苦痛スクリーニングの性能との比較に関して、苦痛スクリーニングの評価指標は、感度 86.8%、特異度 65.6%、精度 39.1%であった。苦痛スクリーニングの結果を除き、患者報告による疼痛スコアを予測変数に含めた機械学習モデルの評価指標は、感度 79.2%、特異度 79.8%、精度 51.4%、AUC 0.83 [95%信頼区間: 0.74, 0.92]であり、各指標のバランスが良く予測性能も高かった。逐次変数選択後に、疼痛スコアを含む 5 つの変数で作成したモデルでも、AUC 0.82 [0.72, 0.91]であった。機械学習モデルで重要な変数は、疼痛スコア、カルシウム値、転倒歴、年齢、性別の順であった。

4. 考察

主な知見は、3点であった。1点目は、治療期の専門的緩和ケアニーズの有病率は約2割であった。森田らの報告では、進行がん患者の外来化学療法開始時に苦痛スクリーニングを実施し、18%に専門的緩和ケアが必要と判別していたことから、本研究でのニーズの有病率と同等であった。2点目は、機械学習モデルにより専門的緩和ケアニーズはAUC 0.89程度で予測可能である。先行研究で専門的緩和ケアへの紹介確率を予測した機械学習モデルの性能は、AUC 0.90、特異度 82%、精度 19%であり、同等の性能を示した。3点目は、患者報告による疼痛スコアを機械学習モデルの予測変数に含めることで苦痛スクリーニングを代替できる可能性であった。年齢と性別に加え、症状負担を示す疼痛スコアや、がんの進行度や全身状態を反映するカルシウム値、運動機能を反映する転倒歴を用いることで、専門的緩和ケアニーズを高い性能で評価できる可能性が示唆された。

5. 結語

化学療法を受ける進行がん患者の専門的緩和ケアニーズは、機械学習モデルにより予測できる可能性がある。機械学習モデルの臨床実装は苦痛スクリーニングの代替になり、全てのがん患者における緩和ケアニーズ評価が、高い性能で自動化できる可能性がある。これにより、早期緩和ケアの推進が可能になると考えられる。

Abstract

Predictive Models for Palliative Care Needs of Advanced Cancer Patients Receiving Chemotherapy

進行がん患者の治療期の専門的緩和ケアニーズを予測する機械学習モデルの構築

Context: Early palliative care is recommended within eight-week of diagnosing advanced cancer. Although guidelines suggest routine screening to identify cancer patients who could benefit from palliative care, implementing screening can be challenging due to understaffing and time constraints.

Objectives: To develop and evaluate machine learning models for predicting specialist palliative care needs in advanced cancer patients undergoing chemotherapy, and to investigate if predictive models could substitute screening tools.

Methods: We conducted a retrospective cohort study using supervised machine learning. The study included patients aged 18 or older, diagnosed with metastatic or stage IV cancer, who underwent chemotherapy and distress screening at a designated cancer hospital in Japan from April 1, 2018, to March 31, 2023. Specialist palliative care needs were assessed based on distress screening scores and expert evaluations. Data sources were hospital's cancer registry, health claims database, and nursing admission records. The predictive model was developed using XGBoost, a machine learning algorithm.

Results: Out of the 1878 included patients, 561 were analyzed. Among them, 114 (20.3%) exhibited needs for specialist palliative care. After under-sampling to address data imbalance, the models achieved an Area Under the Curve (AUC) of 0.89 with 95.8% sensitivity and a specificity of 71.9%. After feature selection, the model retained five variables, including the patient-reported pain score, and showcased an 0.82 AUC.

Conclusion: Our models could forecast specialist palliative care needs for advanced cancer patients on chemotherapy. Using five variables as predictors could replace screening tools and has the potential to contribute to earlier palliative care.

1. 緒言

1.1. 背景

緩和ケアは、生命を脅かす重篤な疾患を持つ患者とその家族および介護者の Quality of life (QOL) の向上を目的とした全人的ケアである¹⁾。緩和ケアの理念に基づくケアは、疾患や年齢を問わず必要とするすべての患者と家族に提供されるものであり、これを基本的緩和ケアと言う²⁾。基本的緩和ケアでは対応が難しい複雑な症状や状態に対して提供するケアを専門的緩和ケアと言い³⁾、看護師を含む緩和ケア専門家で構成される院内チームや、緩和ケア病棟・外来、在宅緩和ケアの場で提供される。特にがん患者は、疾患による苦痛や治療に伴う副作用症状を多く経験する⁴⁻⁶⁾ため、専門的緩和ケアの必要性が高いと言われている。

近年、進行がん患者に対する早期からの緩和ケアが推奨されている⁷⁾。早期からの緩和ケアとは、疾患の診断後の早い段階から苦痛やその他の問題を同定し、的確な評価と介入によって苦痛を予防し緩和するためのケアである⁸⁾。がん治療と並行した早期の専門的緩和ケアは、進行がん患者の症状負担を減らし、QOL や患者満足度の向上、そして生存期間を延長する可能性が示されている^{9,10)}。また、進行がん診断から1年間の苦痛の推移を調べた研究によると、診断時の苦痛が最も高いことが分かっている¹¹⁾。米国腫瘍学会のガイドラインにおいても、進行がんの診断から2ヶ月以内の専門的緩和

和ケアを推奨しており、緩和ケアへの早期紹介が求められている⁷⁾。このように、すべての進行がん患者に対する緩和ケアが推奨されているものの、がん患者の人数に比して専門的緩和ケアの資源は不足しており、全がん患者への専門的緩和ケアの提供は現実的でない。

早期緩和ケアの必要性は認識されつつあるが、緩和ケアへの紹介時期は遅れやすい現状がある。その理由としては、緩和ケアが必要となる時期の予測やニーズの把握が困難であることが挙げられる¹²⁾。また、紹介の判断は緩和ケア専門家でない主治医が担っており、非専門家の経験や直感に依存していることも、適時の紹介を困難にし、終末期や死亡直前での紹介につながっている¹³⁾。緩和ケアの導入が遅くなることで、長期間の苦痛を抱え、専門的緩和ケアを十分な期間受けることができないがん患者が多い。

緩和ケアの遅い導入を防ぐには、緩和ケアへの早期紹介を実現する、進行がん診断後の緩和ケアの必要性の評価と標準化された紹介システムが必要である¹⁴⁾。これまでの研究では、緩和ケアニーズの評価ツールや、がん患者の専門的緩和ケアへの紹介基準が開発されてきた^{15,16)}。また、本邦の施策でも、がん診療連携拠点病院の「がん患者の苦痛のスクリーニングを行う」という要件により、緩和ケアの必要性の評価が求められている。がん患者の苦痛スクリーニングは看護師を中心に行われており、患者

の身体症状、精神症状や心理社会的なニーズを包括的に評価する役割を担っている。

しかし、国内の全国調査の結果によると、苦痛のスクリーニングの実施について「人員が不足している」という回答が多く、人員や時間の制約から、緩和ケアの必要性の評価や紹介基準の活用が十分に行われていない課題がある¹⁷⁾。

この課題を解決するために、近年、人工知能（AI: Artificial Intelligence）への期待が高まっている¹⁸⁾。先行研究では、緩和ケアの必要性を予後の推定期間より判断するため、機械学習を用いた予後予測を行っている^{19, 20)}が、緩和ケアの必要性の判断は、予後予測などの時間指標ではなく、身体的・精神的苦痛などのニーズ指標を用いるのが望ましいと言われている。そのため、予後だけでなくニーズ指標を反映する研究も行われた²¹⁾。その研究では、緩和ケアの必要性を過去の緩和ケア紹介歴と定義し、緩和ケア紹介例を正解データとして、機械学習を用いた紹介確率の予測を行っている。しかし、がんに限らない過去の全ての疾患の紹介例と非紹介例から学習しているため、実際の専門的緩和ケアニーズの有無や、紹介の是非が考慮されていないことが限界の一つとされている。

本研究の対象施設では、入院化学療法から外来化学療法に移行する前の患者全例で、つらさと支障の寒暖計（DIT: Distress and Impact Thermometer）を用いた苦痛スクリーニングを行っており、患者自身が用紙に自己記入する形式で回答される（Fig 1）。

また、苦痛スクリーニングの結果をもとにがん看護の経験が豊富な看護師が問診を行い、必要であれば緩和ケアチームなどの専門家に連絡する役割を担っているため、早期緩和ケア提供の契機となっている。しかし、がん患者数の多さや苦痛スクリーニングと問診の業務負担から、化学療法を行わないがん患者や入院化学療法患者には苦痛スクリーニングが実施されていない。さらに、苦痛スクリーニング陽性時の対応やその後のフォローアップ体制が確立されていないという課題もある。

そこで本研究では、診療録に記載される苦痛スクリーニングの結果とがん・緩和ケア専門家の評価により、対象患者の専門的緩和ケアニーズの有無を事前に評価し、専門的緩和ケアニーズを持つ例を正解データとすることで、進行がん治療期の専門的緩和ケアニーズを予測する機械学習モデルを構築する。さらに、苦痛スクリーニング結果を含まない機械学習モデルや苦痛スクリーニング結果のみでの機械学習モデルの予測性能を比較することにより、臨床実装の意義のある専門的緩和ケアへの自動紹介モデルを構築できると考える。

1.2. 目的

名古屋大学医学部附属病院で化学療法を受けた進行がん患者を対象に、診療録データを用いた機械学習により、がん治療期の専門的緩和ケアニーズの予測モデルを開発し、予測精度を評価する。また、苦痛スクリーニング結果を含めた機械学習モデルと、

スクリーニング結果を含めない機械学習モデル、スクリーニング結果のみの機械学習モデルで予測性能と変数重要度を比較し、臨床実装を考察する。

2. 方法

2.1. 対象

2018年4月1日から2023年3月31日の間に名古屋大学医学部附属病院で外来化学療法を受けた18歳以上の進行固形がん患者で、外来化学療法導入前に苦痛スクリーニングを受けたものを対象とした。対象者には、入院化学療法を経てから外来化学療法を導入する化学療法中の患者と、初回から外来で化学療法を行う化学療法導入前の患者が含まれる。進行がんの定義は、遠隔転移の存在とステージIVの診断とした。化学療法のうち、根治手術の術前術後に行う補助化学療法、化学放射線療法、治験は除外した。また、専門的緩和ケアニードを予測するための情報は、外来診療録のみからは十分に得られないと予想されたため、苦痛スクリーニング時点の苦痛スクリーニングの実施前3か月間に入院歴がない患者は除外した。

2.2. 研究デザインとデータソース

研究デザインは、診療録の既存情報を用いたレトロスペクティブコホート調査である。過去の診療録の専門的緩和ケアニードを正解データとした、教師あり機械学習の手法を用いた。専門的緩和ケアニードの調査手順については、専門的緩和ケアニード

の有無を判断するための情報を以下のデータベースと診療録から収集し、研究者と臨床家による診療録調査で同定した。データソースは主に、院内がん登録情報、診療報酬情報、入院時看護テンプレート記録、院内データウェアハウス (CLISTA!) である。診療報酬情報では、患者の属性や入退院、診断、手術、重症度情報などを含む Diagnosis Procedure Combination (DPC) の様式 1、診療行為や処方された薬剤などの保険請求レセプト情報を含む入院 EF ファイルと外来 EF ファイル、処置加算情報などを含む DPC レセプトデータを使用した。

2.3. 正解定義

医療におけるニードとは「医療から利益を受ける状態」を示す。専門的緩和ケアニードの標準的な定義はないが、「専門的緩和ケアから利益を受ける、複雑性の高いニード」という文脈で一般的に使用されている。本研究での専門的緩和ケアニードの定義は、「緩和ケア・がん医療の専門家が苦痛スクリーニングと問診から判断した専門的緩和ケアのニード」とした。正解データとなる専門的緩和ケアニードの有無は、苦痛スクリーニング結果、診療録調査、専門家評価の三段階で評価した (Supplementary Fig 1)。

2.4. 調査手順

2.4.1. 専門的緩和ケアニーズに関連する記載の診療録調査

調査担当の研究者が、苦痛スクリーニング結果の問診記録等のレビューを行い、臨床家による専門的緩和ケアニーズの判断に関連する記載（例、患者への緩和ケアの提案、症状緩和が必要というアセスメント）の調査を行った。専門的緩和ケアニーズに関連する記載があった場合は、専門的緩和ケアニーズあり（正解）とラベル付けを行い、専門的緩和ケアニーズの内訳を調査した。

2.4.2. 専門的緩和ケアニーズに関連する記載がない場合の診療録調査

専門的緩和ケアニーズの有無に関連する記載がなかった場合は、専門的緩和ケアニーズなしと記録未記載との判断ができないため、苦痛スクリーニング結果からニーズの判断を行った。苦痛スクリーニング陰性（疼痛、疼痛以外の症状、きもちのつらさが NRS 4 未満、きもちのつらさの支障が NRS 3 未満）であった場合は、専門的緩和ケアニーズなし（不正解）とラベル付けを行った。苦痛スクリーニング陽性（疼痛、疼痛以外の症状、きもちのつらさが NRS 4 以上、きもちのつらさの支障が NRS 3 以上）であった場合は、研究者が診療録情報からニーズの有無の判断ができるか（苦痛スクリーニングが陽性であることの根拠となる情報の記載があるか）を調査した。研究者が専門的緩和ケアニーズありと判断できる記載の例は、主治医が薬物調整を行っても改善されない身体・精神症状の存在や、生活の支障に対する患者自身の専門家介入へ

の希望である。記載からニードありと判断した場合は、専門的緩和ケアニードあり（正解）とラベル付けを行い、専門的緩和ケアニードの内訳を調査した。ニードの判断の根拠となる記載がない場合は、本研究チームの専門家による調査を行った。

2.4.3. がん・緩和ケアの専門家による調査

専門家は、長年の臨床経験と専門知識を有し、当該患者の苦痛スクリーニングを担当しているがん・緩和ケアの専門家である。緩和ケアチーム専従のがん看護専門看護師 1 名とがん化学療法認定看護師 2 名の計 3 名で構成され、日常業務として化学療法患者の症状や社会的課題を聴取し、専門的緩和ケアニードを評価する役割を持つ。専門家は対象者の苦痛スクリーニング実施日の前後約 1 か月間の診療録をレビューし、緩和ケアや化学療法ケアの経験と知識に基づき専門的緩和ケアニードの有無の直感的な評価を行った。ニードありと判断された場合は、身体的苦痛、心理的苦痛、意思決定支援の必要性などからニードの内訳を調査した。

2.4.4. 正解データの信頼性

研究者による診療録調査の評価者間信頼性の確認のために、研究者によるラベル付けがされた対象者（専門家による調査の対象外であった患者）の 10%をランダム選択し、専門家による評価を行った。研究者と専門家による評価との一致率と κ 係数を確認した。なお、苦痛スクリーニング時の問診を担当した看護師とは異なる評価者とな

るよう、独立した専門家評価を実施した。研究者評価と専門家評価が一致しなかった患者については、双方で 2 度目の診療録調査を行い、協議の上で専門的緩和ケアニーズの有無の最終判定を行った。

2.5. 調査項目

苦痛スクリーニングは 0-10 点で NRS と同様の評価方法を取り、疼痛が 4 以上、疼痛以外の身体的なつらさが 4 以上、気持ちのつらさが 4 以上、気持ちのつらさによる日常生活の支障が 3 以上の場合には、スクリーニング陽性と判断される。1-2 分程度で実施可能な簡便なスクリーニングツールとして国内で広く用いられ、良好な感度と特異度、妥当性の検証もされている^{22,23)}。

専門的緩和ケアニーズの分類には、専門的緩和ケア外来への紹介基準として開発された Hui らの 11 基準を用いた²⁴⁾。Hui らが開発した基準はデルファイ法で国際的に合意が確認されており、身体的苦痛、心理的苦痛、早く死にたいという希望、精神的または実存的危機、意思決定支援の必要性、患者自身の希望、せん妄、脳または浸潤性髄膜転移、脊髄圧迫または馬尾症候群の存在、という 9 つのニーズ基準と、生存中央値 1 年以内の進行・難治性がんの診断から 3 ヶ月以内、進行性がんの診断で二次治療後の病状の進行、という 2 つの時間基準から成り、本研究ではニーズ基準を分類に使用した。

機械学習での予測に使用する項目について、以下の調査項目を収集した。苦痛スクリーニングが行われた時点（化学療法前もしくは化学療法中）での専門的緩和ケアニーズを予測するために、スクリーニング実施日に最も近い入院情報と外来情報から調査項目を収集した。括弧内に主なデータソースと調査ウィンドウを示す；

(1) 患者属性：診断時年齢、化学療法オリエンテーション時年齢、性別等（院内がん登録情報より）(2) 疾患：がん原発部位、がんステージ、転移の部位、併存疾患スコア、ECOG PS (ECOG; Eastern Cooperative Oncology Group Performance Status)等（苦痛スクリーニング前 90 日以内の入院情報と、問診時の記録より）、(3) 処方：化学療法レジメン、オピオイド性鎮痛薬処方等（苦痛スクリーニング前 90 日以内の処方オーダーより）、(4) 採血：アルブミン、ビリルビン、トロポニン、好中球、AST、ALT、アニオンギャップ、乳酸、PT-INR、クレアチニン、ヘマトクリット等（苦痛スクリーニング前 30 日以内の採血のうち直近のデータより）、(5) 栄養：BMI (Body Mass Index)、過去 6 カ月の体重減少、食欲低下、嚥下障害の有無、褥瘡リスクスコア（苦痛スクリーニング前 90 日以内の入院情報より）、(6) 医療利用：入院診療科、転入元、過去 1 年間の入院歴の有無、過去 1 年間の ICU 利用歴の有無等（苦痛スクリーニング前 90 日以内の入院情報と、苦痛スクリーニング前 1 年間の医療利用歴より）、(7) リテラシー：使用言語、学習障害の有無等（苦痛スクリーニング前 90 日以内の入院情報より）、(8) 社会経済的背景：職業の有無、同居者の有無、キーパーソンとの続柄、通院手段・通

院時間、居住地域、喫煙歴、アルコール多飲歴、宗教の有無等（苦痛スクリーニング前 90 日以内の入院情報と、問診時の記録より）、(9) 症状、ADL、認知機能：疼痛 NRS、呼吸困難感など症状の有無、日常生活自立度スコア、起居動作スコア、転倒リスクスコア、意識レベル、認知機能スコア、せん妄リスクスコア、精神科受診歴等（苦痛スクリーニング前 90 日以内の入院情報より）。調査項目の概要を、Fig 2.として示す。

2.6. 統計解析

診療録調査の結果より、専門的緩和ケアニーズありと評価された患者と、なしと判断された患者の背景を示す記述統計量の算出を行った。また、既存のスクリーニングと機械学習モデルの性能比較のために、一般的にがん患者の適応障害やうつ病のスクリーニングとして用いられる 2 項目の「つらさと支障の寒暖計」でのスクリーニングと、対象施設の外来化学療法導入前に用いられる 4 項目の苦痛スクリーニングの、正解率、精度、感度、特異度を算出した。欠損補完について、20%以上の欠測が生じている変数については、欠測が生じる要因がランダムでない（MNAR: Missing Not At Random）可能性が高いと判断し²⁵⁾、20%以上の欠測が生じている変数全てを欠損の有無で二値変数に変換して予測モデルに使用した。欠損率 20%未満の変数については、XGBoost に組み込まれている訓練データ内の欠損値を自動的に補完するアルゴリズムを用いて欠損補完を行った。XGBoost は決定木モデルであるため、このアルゴリズムは欠損値に対する分岐方向を学習して補完する方法である。他にも、欠損補完方法の

感度分析として、scikit-learn の IterativeImputer ライブラリを用いた多重代入法を実施して欠損補完を行う方法と、欠損値自体を 1 つのカテゴリ変数として投入する方法で、それぞれ機械学習モデルを作成した (Supplementary table 1)。IterativeImputer は、ベイズアンリッジを用いた回帰により、他の変数から欠損データを予測して多重代入補完を行う手法で、機械学習モデルで多く用いられている。統計解析や予測モデル構築、欠損補完には Python 3.10.12 を使用し、有意水準は 5%未満とした。

2.7. 機械学習アルゴリズム

機械学習モデルとして、(1) 126 項目の全変数投入モデル「ALL-model」、(2) 4 項目の苦痛スクリーニング結果以外の変数投入モデル「NotSCR-model」、(3) 痛み以外の 3 項目の苦痛スクリーニングを除外した「NotSCR+Pain-model」、(4) 苦痛スクリーニング結果のみ投入モデル「SCR-model」、(5) 変数選択後の 5 項目のみ使用した「Simple-model」を作成した。(1) 全変数投入モデル「ALL-model」は、専門的緩和ケアニードの予測に寄与し得る可能な限りの変数を投入し、機械学習モデル内の変数選択を行うことで、予測性能の最大化を目指したモデルである。(2) 4 項目の苦痛スクリーニング結果以外の変数投入モデル「NotSCR-model」は、臨床で日常的に取得される可能性の低い苦痛スクリーニングの 4 変数を除いた上で、実現可能性の高さを目指したモデルである。(3) 痛み以外の 3 項目の苦痛スクリーニングを除外した「NotSCR+Pain-model」では、4 項目の苦痛スクリーニングのうち、日常臨床で取得されやすい疼痛

NRSのみ投入した上で、4項目の苦痛スクリーニング全てを除いた「NotSCR-model」と比較して予測性能がどの程度変化するかを評価するためのモデルである。(4) 苦痛スクリーニング結果のみ投入モデル「SCR-model」では、苦痛スクリーニング4項目のみで機械学習を行うことで、苦痛スクリーニング項目を含めたモデルの比較対象としてのモデルである。(5) 変数選択後の5項目のみ使用した「Simple-model」では、他医療機関でも一般的に取得可能と考えられ、データ取得者間の主観が入りにくい客観的な変数の中から、より予測性能が高い変数の組み合わせを探することで、社会実装につなげることを見据えたモデルである。

データセットは、80%を訓練用データ、20%をテストデータに分割し、モデルの学習と検証を行った。訓練用データにおいては5分割交差検証を行い、グリッドサーチによるハイパーパラメータの調整を行った。分類器は、勾配ブースティング回帰木(XGBoost)を選択した。XGBoostを選択した理由の一点目は、本研究で機械学習での予測に使用する項目の多くがテキストデータを基にしたカテゴリ変数であるためノイズが大きいと予想された点、二点目はモデルの学習時にリッジ回帰とラッソ回帰の要素を組み込んでいることから、過学習のリスクを減らすことができるためである。未知データに対する予測精度については、テスト用データを予測モデルに投入して、正解率、精度、感度、特異度、Fスコアを求め、ROC曲線(Receiver Operating Characteristics Curve)を作成し、AUC(Area Under the ROC Curve)を算出して評価

した。正解率 (Accuracy) は、予測モデルが正しく予測したデータの割合を示し、正解した予測の数を全体の予測数で割った割合である。精度 (Precision) は、モデルが「専門的緩和ケアニードあり」と予測したデータのうち、実際に「専門的緩和ケアニードあり」であったデータの割合である。感度 (Sensitivity) は、実際のデータが「専門的緩和ケアニードあり」であったものの中で、モデルが正しく「専門的緩和ケアニードあり」と予測できた割合である。特異度 (Specificity) は、実際のデータが「専門的緩和ケアニードなし」であったものの中で、モデルが正しく「専門的緩和ケアニードなし」と予測できた割合である。F スコアは、精度と感度の調和平均であり 0 から 1 の範囲でモデルの性能のバランスの良さを要約するために使用される。ROC 曲線は、感度を縦軸に、偽陽性率 (1-特異度) を横軸にプロットした曲線であり、その曲線下の面積である AUC は 1 に近いほどモデルの性能が高いことを示し、0.5 に近いほどランダムな予測と同等であることを示す。また、モデルの臨床的な解釈のために各変数の重要度を算出した。

専門的緩和ケアニードの有病率の低さから、アウトカムの各クラス数 (専門的緩和ケアニードありの少数派クラスと、なしの多数派クラス) の比率に偏りが生じる可能性が予想されたため、専門的緩和ケアニードなし群からランダムサンプリング (以下、アンダーサンプリング) を行い、専門的緩和ケアニードあり群と同数にした均衡なデータセットでの機械学習モデルと、アンダーサンプリングを行わないデータセットで

の機械学習モデルを、アンダーサンプリングを行ったデータセットで作成した機械学習モデルとの結果の比較のために作成した (Supplementary table 4)。

2.8. 倫理的配慮

既存情報のみを収集して行う疫学研究であるため、研究対象者の同意は取得せず、情報公開文書による拒否機会の保証とオプトアウトを行った。なお、名古屋大学生命倫理審査委員会の承認を得て行った (承認番号: 2022-0494)。

3. 結果

3.1. 対象者背景

2018年4月1日から2023年3月31日に外来化学療法と苦痛スクリーニングを実施した患者1,878名の適格基準を確認した。適格基準に該当した829名のうち、院内がん登録情報が入手できなかった245名と、すでに緩和ケアチームによる介入がされていた23名を除外した後の561名を解析対象者とした (Fig 3.)。対象者の平均年齢は65.7歳、女性が39.0%であり、主ながん原発部位は消化管や肝・胆・膵がんであった。

3.2. 正解データ

3.2.1. 苦痛スクリーニング

解析対象者561名のうち、疼痛NRSが4以上であったものが86名 (15%)、欠損1

名であった。疼痛以外の身体症状のNRSが4以上であったものが7名（1%）であり、欠損が109名であった。きもちのつらさのNRSが4以上であったものが207名（37%）、欠損2名であり、きもちのつらさによる生活の支障がNRS3以上であったものが134名（24%）欠損2名であった。

3.2.2. 専門的緩和ケアニーズのラベル付けと信頼性

研究者による診療録調査の結果、専門家による評価が必要となった患者は561名中51名であった（9.1%）。専門家評価を必要とした理由の多くは、気持ちのつらさのNRSが高かったがその理由の記載がないものや、専門的緩和ケアニーズに関連した記載はあるものの詳細な記載ではないためにラベル付けが困難と判断された例であった。また、専門家による評価が不要と判断された510名から10%のランダムサンプリングで抽出された51名を追加し、計102名（561名のうち18.2%）が専門家評価の対象者となった。研究者と専門家の判断の一致率は97.1%（102名中99名の一致）であり、 κ 係数は0.91であった。

3.2.3. 専門的緩和ケアニーズの内訳

解析対象者561名のうち、最終的に専門的緩和ケアニーズありと判断されたのは114名（20.3%）であった。専門的緩和ケアニーズの内訳を、Huiの基準を参考に分類した結果、頻度が高かったニーズの種類は「心理的苦痛」（114名中100名、87.7%）、

「身体的苦痛」(65 名、57.0%)、「患者本人の緩和ケアの希望」(30 名、26.3%)、「脳転移の存在」(12 名、10.5%)、「意思決定支援の必要性」(10 名、8.8%)であり、「実存的苦痛」は 1 名のみであった。Hui の基準以外では、「社会的サポートの必要性」(18 名、15.8%) や、「がん治療に伴う経済負担」(7 名、5.8%) のニーズがあった。

Table 1. に、専門的緩和ケアニーズの有無別の対象者背景を示す。

3.3. 苦痛スクリーニングモデル

2 項目の「つらさと支障の寒暖計」の性能は、感度 80.7%、特異度 71.8%であった (Table 2)。4 項目の対象施設の運用に従った苦痛スクリーニングの性能は、感度 86.8%、特異度 65.6%であった。

3.4. 機械学習モデル

5 つの機械学習モデル：全変数投入モデル「ALL-model」、4 項目の苦痛スクリーニング結果以外の変数投入モデル「NotSCR-model」、痛み以外の 3 項目の苦痛スクリーニングを除外した「NotSCR+Pain-model」、苦痛スクリーニング結果のみ投入モデル「SCR-model」、変数選択後の 5 項目の「Simple-model」の結果の ROC 曲線を Supplementary Fig 2. に示す。予測性能の評価指標の結果と信頼区間を Table 2 に示す。

3.4.1. 全変数投入モデル「ALL-model」

アンダーサンプリング後の XGBoost による ALL-model モデルの予測性能は、正解

率 (accuracy) 77%、精度 (precision) 47.9%、感度 (sensitivity) 95.8%、特異度 (specificity) 71.9%、F スコア 63.9%、AUC 0.89 であった (Fig 4.)。アンダーサンプリング前後の患者背景は表 1 に示す。予測に使用された変数は 16 項目であった。

3.4.2 苦痛スクリーニング以外のモデル「NotSCR-model」

XGBoost による NotSCR-model では、正解率 (accuracy) 71.7%、精度 (precision) は 40.5%、感度 (sensitivity) は 70.8%、特異度 (specificity) 71.9%、F スコアは 51.5%、AUC は 0.74 であり、ALL-model と比較すると、すべての評価指標で予測性能は悪化した。

3.4.3 痛み以外の 3 項目の苦痛スクリーニングを除外した「NotSCR+Pain-model」

XGBoost による NotSCR+Pain-model では、正解率 (accuracy) 79.7%、精度 (precision) は 51.4%、感度 (sensitivity) は 79.2%、特異度 (specificity) 79.8%、F スコアは 62.3%、AUC は 0.83 であり、NotSCR-model と比較すると、全体的に予測性能は向上した。

3.4.4 苦痛スクリーニングのみのモデル「SCR-model」

XGBoost による SCR-model では、正解率 (accuracy) 89.4%、精度 (precision) は 68.8%、感度 (sensitivity) は 91.7%、特異度 (specificity) 88.8%、F スコアは 78.6%、AUC は 0.92 であり、5 つの機械学習モデルの中で AUC は最も良好であった。

3.4.5 変数選択後の 5 項目の「Simple-model」

XGBoost による痛みの項目を含めた 5 項目の Simple-model では、正解率 (accuracy) 76.1%、精度 (precision) は 46.2%、感度 (sensitivity) は 75.0%、特異度 (specificity) 76.4%は、F スコアは 57.1%、AUC は 0.82 であった。

3.5. 変数の重要度

ALL-model において、予測に使用された 16 項目は、疼痛 NRS、きもちのつらさによる支障の寒暖計、転倒歴、HbA1c、きもちのつらさの寒暖計、年齢、プロトロンビン時間、疼痛以外の身体症状の NRS、総タンパク質、病院への通院手段、不安、化学療法レジメンの催吐性リスク分類、性別、総コレステロール、および γ -GTP であり、苦痛スクリーニングの結果が専門的緩和ケアニードの予測に特に重要であった (Fig 5)。

Simple-model において、逐次変数選択後に予測に使用された 5 項目は、疼痛 NRS、血清カルシウム値、年齢、性別、転倒歴であり、特に疼痛 NRS が専門的緩和ケアニードの予測に高く寄与していた (Fig 6)。

4. 考察

本研究は、診療録に記載される苦痛スクリーニングの結果と専門家の判断により、専門的緩和ケアニードを持つ例を正解データとすることで、進行がん治療期の専門的緩和ケアニードを予測する機械学習モデルを構築した。進行がん診断後の治療期の専

専門的緩和ケアニーズは、AUC 0.89 程度で予測可能であることが明らかにされた。予測モデル構築により、これまで専門的緩和ケアニーズが高いと考えられていなかった患者にも、適時の専門的緩和ケア介入の検討が可能になると考えられる。

4.1. 主な知見

主な知見は、3点ある。1点目は治療期の専門的緩和ケアニーズの有病率は約2割であること、2点目は機械学習モデルにより専門的緩和ケアニーズはAUC 0.89程度で予測可能であること、3点目は疼痛NRSを組み込んだモデルは、苦痛スクリーニングの代用となり得る可能性である。

4.2. 専門的緩和ケアニーズの有病率と分布

1点目の知見として、進行固形がん患者の治療期の専門的緩和ケアニーズの有病率は約2割であった。専門的緩和ケアニーズは、一般的にSPICT (Supportive and Palliative Care Indicator Tool)などのアセスメントツールを用いて評価される。HamanoらはSPICT-JPを用いて専門的緩和ケアを必要とする在宅患者を同定しており、65歳以上の患者のうち17.3%に専門的緩和ケアが必要であると評価していた²⁵⁾。またMoritaらの報告では、進行がん患者の外来化学療法開始時に苦痛スクリーニングを実施し、18%に専門的緩和ケアが必要と判別した²⁶⁾。本研究でのニーズの有病率が2割程度と同等であったことを考慮すると、苦痛スクリーニング結果と専門家による

評価で専門的緩和ケアニーズを同定可能であったと考えられる。

専門的緩和ケアニーズの内訳の多くは、心理的苦痛や身体的苦痛、患者自身の緩和ケアの希望、意思決定支援の必要性で構成されていた。進行がん診断後の緩和ケアニーズは、身体的苦痛より精神的苦痛が多いという Vogt らの報告と一致する¹¹⁾。本研究で精神的ニーズや身体的ニーズの割合が高かった理由としては、苦痛スクリーニングでの精神・身体症状評価による症状の訴えやすさの影響が大きいと考えられる。反対に、社会的なニーズや情報提供のニーズ、スピリチュアルニーズの評価は十分に行われずに有病率が過小評価されている可能性がある。

4.3. 機械学習モデルの予測性能

2点目の知見として、機械学習モデルにより専門的緩和ケアニーズは AUC 0.89 程度で予測可能であった。Murphee らの専門的緩和ケアへの紹介確率を予測した機械学習モデル²¹⁾の性能は、AUC = 0.90、特異度 (specificity) 82%、精度 (precision) 19%であった。本研究の ALL-model では、AUC = 0.89、特異度 71.9%、精度 47.9%であったため同程度の性能を示している。評価指標のなかでも、特に感度を重視したため、感度 95.8%という結果はモデルの良好な性能を示していると言える。

専門的緩和ケアニーズの予測に重要な変数は、苦痛スクリーニングを構成する支障の NRS、きもちのつらさの NRS、痛みの NRS、痛み以外の身体症状の NRS であった。

この結果は、苦痛スクリーニング実施の重要性を強調するものである。しかし、苦痛スクリーニングの結果自体が専門的緩和ケアニーズの判定の一部となっているため、苦痛スクリーニングを構成する変数を含めたモデルの評価指標は、統計上の過大評価がされている可能性がある点には注意が必要である。ALL-model でも NotSCR-model においても、共通して重要度が高かった変数は、性別、年齢、疼痛 NRS、転倒歴であった。性別・年齢が緩和ケアニーズの有無に影響を与えることは、先行研究の結果と同様である^{27,28)}。また、転倒歴は身体機能状態を示すために緩和ケアニーズに影響したと考えられる。

4.4. ALL-model と、NotSCR/NotSCR+Pain/SCR-model の比較

3点目の知見として、疼痛 NRS を組み込んだモデルは、苦痛スクリーニングの代用となり得る可能性が明らかになった。NotSCR-model の評価指標を見ると、感度だけではなく、正解率や AUC などモデル全体として予測性能が低下している。本研究で投入した既存のカルテ情報のみでは、緩和ケアニーズの予測は困難であり、NotSCR-model でスクリーニングを代替する自動紹介システムには適していないモデルであると考えられる。しかし、NotSCR-model に疼痛 NRS 項目を追加して NotSCR+Pain-model とすると予測性能は改善された。本研究の対象者の専門的緩和ケアニーズの内訳は「心理的苦痛」が約 88%、「身体的苦痛」が 57.0%、「患者本人の緩和ケアの希望」が 26.3%であり、「痛み」は主要なニーズではなかったものの、変数重要度において「痛

み」は専門的緩和ケアニーズに高く寄与していた。疼痛は、身体的苦痛の1つであるだけでなく、心理社会的苦痛、スピリチュアルペインとも相互に関連しており、QOL への影響は大きい。そのため、疼痛は体温、呼吸、脈拍、血圧に続く「第5のバイタルサイン(the 5th vital sign)」と呼ばれ、痛みの数値を定期的経過表に記録することが推奨されている。本研究での疼痛の寄与率の高さを考慮すると、疼痛が心理的苦痛や患者自身の希望などの他のニーズと相互に関連していた可能性がある。疼痛 NRS のデータを取得し機械学習モデルに投入できる場合、スクリーニングに代わる予測モデルとなり得る可能性がある。

さらに、SCR-model でも、ALL-model と同等のレベルでニーズを予測できることが明らかになった。苦痛スクリーニングは心理社会的ニーズを持つ者を同定し、患者と医療者間のコミュニケーションを改善する効果があり、精神や心理の専門家に紹介するためのツールとして使用されている。そのため心理的ニーズが多くを占める専門的緩和ケアニーズの同定にも寄与した可能性がある²⁹⁾。SCR-model では、感度以外の評価指標において ALL モデルの性能を上回っている。本研究では、専門的緩和ケアニーズのある患者を見逃さずに陽性と判定することを重視しているために感度に重きを置いているが、陽性的中率や陰性的中率も同様に重要である。陽性的中率が低い場合は偽陽性が増え、専門的緩和ケアニーズの低い患者に緩和ケア資源を配分することになり、陰性的中率が低い場合は偽陰性が増え、専門的緩和ケアニーズの高い患者に、

緩和ケア提供がされないことにつながる。専門的緩和ケア資源の制約などで、感度以外の評価指標を重視すべき状況においては ALL-model より SCR-model が適している可能性がある。

4.5. 臨床への示唆

緩和ケアに関する NCCN 腫瘍学臨床診療ガイドライン (NCCN Guidelines for Palliative Care) では、すべてのがん患者の緩和ケアニーズをスクリーニングし、紹介基準を満たした場合には専門的緩和ケアへの紹介を行うことを推奨している⁷⁾。しかし、臨床現場では、苦痛スクリーニングの業務負担が大きく全患者のスクリーニングが行えていない。さらに、スクリーニング実施後のアルゴリズムが確立されていないため、スクリーニング陽性となった患者に統一した対応がなされていないことが課題である。

この課題に対して、本研究では ALL-model、NotSCR-model、NotSCR+Pain、SCR-model の比較を行い、予測モデルが苦痛スクリーニングを代替し得るかを考察した結果、疼痛 NRS を組み込んだモデルは、苦痛スクリーニングの代用となり得る可能性が明らかになった。疼痛 NRS は簡易に評価可能な数値スケールであり、看護師による定期評価を推奨している医療機関も多く、日本全国の医療機関で専門的緩和ケアニーズの予測に活用できる可能性の高い変数である。しかし、これまでの研究では、スクリーニング後のアルゴリズムのアドヒアランスの不足が患者アウトカム改善を妨げるこ

とが分かっている³⁰⁾。苦痛スクリーニングや、疼痛 NRS を含む機械学習モデルでもニードの予測は可能だが、ALL-model のように他の定期的なアセスメント項目を組み合わせることで、全体的に良い指標のモデルとなり、スクリーニング後のアルゴリズムの徹底にもつながるために、ニードを持つ可能性の高い患者の抽出が行える可能性がある。

本機械学習モデルを臨床実装する上での課題は大きく3点あると考える。1点目は、緩和ケア啓蒙の不足である。機械学習モデルで陽性と判定された場合には、緩和ケアについて患者への説明が求められる。機械学習モデルを使用し、患者に緩和ケアを提案するのは緩和ケアを専門としない主治医となる可能性が高く、緩和ケアに関する基本知識や説明のためのコミュニケーション技術を要する。緩和ケアは終末期に提供されるものであるという一般的な誤解は根強いため、緩和ケアの情報提供自体が患者に侵襲を与える可能性がある。緩和ケアに関するパンフレットを合わせて提供するなど、モデルでの陽性判定後の説明にも支援が必要であると考え。2点目は、モデルでニード陰性と判定された患者に緩和ケアの対応がされない可能性がある点である。複雑な緩和ケアニードが低くとも、すべての進行がん患者に基本的な緩和ケアニードはありと考えられ、モデルでの陰性の判定は緩和ケアが不要であることを意味していない。全てのがん患者に緩和ケアは必要であり、モデルの判定により緩和ケア提供が妨げられることのないよう、情報提供が必要である。また、全てのがん患者に基本的緩和

ケアが提供されるよう、緩和ケア教育を推進していくことも重要であると考え。3点目は、機械学習モデルを使用する際のデータ入力の負担が実装の障壁となり得る点である。すべての項目が電子カルテより自動収集されると、データ入力の負担なく緩和ケアニーズのタイムリーな評価が行われるが、現在はデータ入力への労力を要する変数を使用したモデルである。苦痛スクリーニングにおいても、評価における業務負担が普及の障壁となっているため、すべて自動化したシステムへの発展が求められる。

4.6. 看護への示唆

本研究では、専門・認定看護師による臨床的判断を基にして専門的緩和ケアニーズの評価を行った点と、入院時看護アセスメント情報を予測情報に多く使用した点が特色である。重要度の高い変数の中でも、不安や転倒歴、病院への通院にかかる負担は看護師がアセスメントする機会が多く、定期的にあセスメントを行った方が良い可能性が示唆された。また、予測モデルを臨床実装した場合、患者をアセスメントする機会の増加につながり、看護の質向上に寄与できる可能性がある。

4.7. 研究の強み

これまで緩和ケアニーズの予測モデルを構築した研究は知る限りでは存在しておらず、多くが緩和ケアニーズの代替として予後予測指標を使用していた。本研究では緩和ケア専門家との協議により、専門的緩和ケアニーズを持つ例を正解データとするこ

とで、進行がん治療期の専門的緩和ケアニーズを予測する機械学習モデルを構築することが可能となった。本研究は、臨床実装の意義のある専門的緩和ケアへの自動紹介モデルの基礎となると考えられる。

4.8. 研究の限界

1点目はアウトカムである専門的緩和ケアニーズの妥当性と信頼性の課題である。本研究では専門家による評価を正解定義としたが、IPOS(Integrated Palliative Outcome Scale)等の妥当性が検証されたツールで前方視的に患者の専門的緩和ケアニーズの測定をして正解データとすることが望ましい。もしくは、妥当性が検証されたツールを用いた評価と、本研究のような専門家による評価の一致率を確認することが望ましいと考えられる。また、対象施設は大学病院、かつ、がん拠点病院であり、一般病院と比較して各診療科の行う基本的緩和ケアのレベルは高い可能性があり、専門家による専門的緩和ケアニーズの判断は基本的緩和ケアのレベルや緩和ケアチームのリソースにも依存した判断となっている可能性もある。しかし、専門的緩和ケアの明確な定義はなく、日本ホスピス緩和ケア協会が専門的緩和ケアを「基本的緩和ケアで対応が困難」と定義しているように、専門的緩和ケアの必要性は各診療科の緩和ケアの能力に依存した判断になってしまうため、評価ツールでの定義は本質的に不可能であるとも考えられる。本研究では、専門家と研究者との評価の一致率も高かったことから、評価者間信頼性は高く、正解データの信頼性は十分であると考えられる。

2 点目は不均衡データの課題である。本研究の対象者のうち、専門的緩和ケアニーズありと判断されたのは2割であり、軽度～中度の不均衡データとなった。そのため、全ての解析対象者から作成した予測モデルでは、正解率が高く評価されやすく、適合率や感度が低い結果となった。サンプルサイズの減少やそれに伴う対象者背景のバイアスのリスクにはなるが、不均衡に対して適切にアンダーサンプリングの統計的処理を行うことにより、感度が高く性能の良いモデルが構築できたと考える。アンダーサンプリング前後の対象者背景も大きな違いはなく、サンプリングによるバイアスの影響は大きくないと考えられる。

3 点目は予測に用いるテキスト情報や患者報告型の症状情報の不足、社会的・霊的ニーズに関する情報の不足である。本研究では自然言語処理を用いた予測を行っていないため、テキスト情報として格納されやすい患者の症状や苦痛情報を十分に使用することができていない。また、症状は医療者評価では過小評価になりやすいため患者報告による評価が好ましいとされているが、患者報告型の症状データはルチン収集されていないため、疼痛や呼吸困難以外の症状の程度や推移や、患者が考える疼痛のゴール (personalized pain goal :PPG) 情報の使用ができていない。症状や苦痛は緩和ケアニーズを構成する大きな要素であるため、Masukawa らの自然言語処理を用いた社会的苦痛やスピリチュアルペインの収集³¹⁾や、Shimada らの言語化が難しい患者の自覚症状の予測を他覚症状から行う手法³²⁾を組み合わせることで、より包括的な情報を

用いた予測が可能になると考える。

4点目は、データの精度の課題である。院内がん登録情報やDPC様式1は各項目の定義が定められており、診療情報管理士をはじめとする診療録の有資格者が入力している。看護テンプレートの入力項目は評価者により判断が異なり、入力データにばらつきが生じていた可能性がある。

5点目に、調査時期は2018年から2023年であり、新型コロナウイルス感染症によるパンデミック前の2018年4月～2019年12月と、第一波が到来した2020年1月以降の期間を含んでいる。そのため、パンデミックに伴う不安や病院の面会制限などが精神症状や身体症状などのニーズに影響を与えた可能性がある。

6点目に、交互作用の影響があった可能性である。これまでの調査から、がんの原発部位は緩和ケアニーズの程度に影響を与えることが分かっている。しかし、本研究で構築した機械学習モデルでは、がん部位は重要度の高い変数とされておらず、性別による交互作用の影響があった可能性が考えられる。アンダーサンプリングを行った後のサンプルサイズが小さかったために、男女別でのモデル構築など交互作用を考慮した解析が行えていないことは機械学習モデルの解釈性に影響を与えた可能性がある。

最後に、予測モデルの一般化可能性である。研究対象者は単施設の入院患者に限定されており、外来で化学療法を開始しやすいレジメンや乳がん、若年で身体機能の良い患者、

全身状態が不良で化学療法の実施が困難な患者、血液がんの患者には、予測モデルの適用に制限がある可能性がある。今後は外来での化学療法導入が増える可能性があるが、専門的緩和ケアニーズが高い患者は入院治療を選択する可能性も高く、入院患者に絞って予測を行うことの影響は大きくはないと予想される。

5. 結論

化学療法を受けた進行がん患者を対象に、診療録データを用いた機械学習による専門的緩和ケアニーズの予測モデルを開発した結果、治療期の専門的緩和ケアニーズの予測が可能であることが明らかにされた。今後は多施設で予測モデルを当てはめ、予測モデル適用による患者アウトカムへの影響を評価することにより、より性能が高く臨床現場で活用できるモデルに改良できると考えられる。

6. 謝辞

本研究の遂行にあたり、指導教官として終始多大なご指導を賜った、佐藤一樹教授に深く感謝致します。また、メディカルITセンター古川大記先生をはじめとする AI-MAILs の先生方には、機械学習に関する専門的なご助言を賜りましたのでここに深謝の意を表します。また、緩和ケアニーズの評価に多大なるご協力を賜りました、看護師の原万里子主任、山田里美様、濱昌代様、川口綾師長に深く感謝いたします。そして、データ抽出にご尽力いただいた、名古屋大学医学部附属病院先端医療開発部の諸

橋朱美様と今泉貴広先生、病歴管理室の小田祐人様、医事課診療録管理系の松田健矢様、医事課医療システム管理系の井上裕太様に、心より感謝申し上げます。また、研究計画の草案にご助言を賜りました未来社会創造機構 Mohammed Hajjaj 先生にも感謝申し上げます。最後に、主査をご担当いただいた玉腰浩司先生、並びに副査をご担当いただいた中析昌弘先生に感謝いたします。

本研究は、東海国立大学機構融合フロンティア次世代研究事業の JST 次世代研究者挑戦的研究プログラム JPMJSP2125、公益財団法人安田記念医学財団（癌看護研究助成 A）、名古屋大学基金（Nagoya University Research Fund）の財政支援を受けたものです。この場を借りて御礼申し上げます。

7. 図表付録

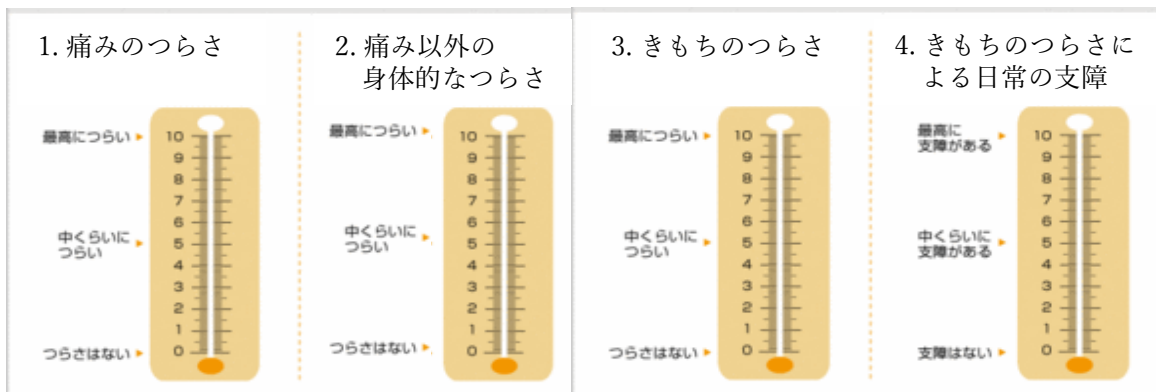


Fig 1. 苦痛スクリーニングの概要

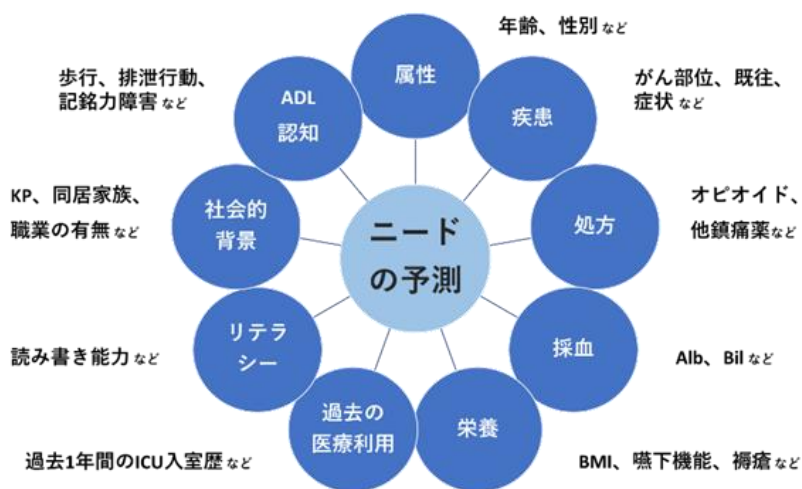


Fig 2. 機械学習で使用するための調査項目の概要

Table 1. Patient characteristics

	Imbalanced data (n=561)				Rebalanced data (n=190) ¹			
	High SPC needs (n=114)		Low SPC needs (n=447)		High SPC needs (n=95)		Low SPC needs (n=95)	
	N or mean	<i>SD or (%)</i>	N or mean	<i>SD or (%)</i>	N or mean	<i>SD or (%)</i>	N or mean	<i>SD or (%)</i>
Age (years)	61.2	12.7	66.8	10.8	60.3	12.7	66.9	12.0
Sex assigned at birth (female)	63	(28.8)	156	(71.2)	57	(59.4)	39	(40.6)
Cancer types								
Gastrointestinal	23	(13.5)	147	(86.5)	37	(53.6)	32	(46.4)
Hepatobiliary and pancreatic	45	(29.8)	106	(70.2)	17	(40.4)	25	(59.6)
Lung	18	(14.9)	103	(85.1)	16	(40.0)	24	(60.0)
Urological	7	(14.9)	40	(85.1)	6	(54.5)	5	(45.5)
Gynecological	10	(32.3)	21	(67.7)	11	(64.7)	6	(35.3)
Head and neck	5	(33.3)	10	(66.7)	5	(83.3)	1	(16.7)
Skin	4	(36.4)	7	(63.6)	3	(100)	0	(0)
Others	0	(0)	7	(100)	0	(0)	2	(100)
Multiple cancers	9	(13.2)	59	(86.8)	7	(38.9)	11	(61.1)
Recurrence	35	(18.0)	159	(82.0)	29	(43.3)	38	(56.7)
Chemotherapy								
1 st line	82	(19.0)	350	(81.0)	69	(46.3)	80	(53.7)
2 nd line	24	(27.3)	64	(72.7)	20	(64.5)	11	(35.5)

3 rd line	7	(20.6)	27	(79.4)	5	(71.4)	2	(28.6)
4 th line	1	(14.3)	6	(85.7)	1	(33.3)	2	(66.7)
ECOG PS								
0	77	(16.9)	379	(83.1)	68	(45.6)	81	(54.4)
1	24	(34.3)	46	(65.7)	19	(65.6)	10	(34.4)
2	8	(50.0)	8	(50.0)	4	(66.7)	2	(33.3)
3	3	(60.0)	2	(40.0)	2	(100)	0	(0)
4	0	(0)	1	(100)	0	(0)	0	(0)
Metastasis								
Liver	46	(22.4)	159	(77.6)	36	(49.3)	37	(50.7)
Lung	31	(16.8)	153	(83.2)	23	(43.4)	30	(56.6)
Bone	26	(21.8)	93	(78.2)	23	(65.7)	12	(34.3)
Brain	12	(30.8)	27	(69.2)	10	(66.7)	5	(33.3)
Consciousness level ²								
Clear	114	(20.7)	437	(79.3)	95	(50.5)	93	(49.5)
1-digit	0	(0)	8	(100)	0	(0)	1	(100)
2-digit	0	(0)	1	(100)	0	(0)	1	(100)
Distress screening ³								
NRS \geq 4 (pain)	38	(44.2)	48	(55.8)	31	(83.8)	6	(16.2)
NRS \geq 4 (other symptoms)	5	(71.4)	2	(28.6)	4	(80.0)	1	(20.0)

Distress Thermometer ≥ 4	87	(42.0)	120	(58.0)	71	(70.3)	30	(29.7)
Impact Thermometer ≥ 3	69	(51.9)	64	(48.1)	53	(80.3)	13	(19.7)

Notes: N; number, SD; standard deviation, JCS; Japan Coma Scale, NRS; Numerical Rating Scale, ECOG PS; The Eastern Cooperative Oncology Group Performance Status.

¹ Rebalanced data is a dataset that has had under-sampling applied.

² Consciousness level was assessed by the Japan Coma Scale at the time of hospitalization. A 1-digit score indicates that the patient is awake without any stimuli. A 2-digit score indicates that the patient can be aroused and then revert to the previous state after cessation of stimulation

³ Four items were assessed as part of the distress screening at the hospital: numerical rating scale of pain, numerical rating scale of physical symptoms other than pain, Distress Thermometer (DT), and Impact thermometer (IT) for psychological distress. NRS for physical symptoms excluding pain had 109 missing values, and DT score and IT score had 2 missing values, respectively.

Table 2. Summary of predictive performance of each model

(95% Confidence Interval)	Accuracy	Precision	Sensitivity	Specificity	F score	AUC
ALL-model	77.0% (68.4-83.8)	47.9% (34.5-61.7)	95.8% (79.8-99.3)	71.9% (61.8-80.2)	63.9% (49.3-75.6)	0.89 (0.80-0.96)
NotSCR-model	71.7% (62.8-79.2)	40.5% (27.0-55.5)	70.8% (50.8-85.1)	71.9% (61.8-80.2)	51.5% (35.7-64.3)	0.74 (0.64-0.83)
NotSCR+Pain-model	79.7% (71.3-86.0)	51.4% (35.9-66.6)	79.2% (59.5-90.8)	79.8% (70.3-86.8)	62.3% (47.1-75.9)	0.83 (0.74-0.92)
SCR-model	89.4% (82.4-93.8)	68.8% (51.4-82.1)	91.7% (74.2-97.7)	88.8% (80.5-93.8)	78.6% (64.9-89.4)	0.92 (0.82-0.98)
Simple-model	76.1% (67.5-83.0)	46.2% (31.6-61.4)	75.0% (55.1-88.0)	76.4% (66.6-84.0)	57.1% (40.0-70.3)	0.82 (0.72-0.91)
Distress screening (4-item) ¹	69.9%	39.1%	86.8%	65.6%	54.0%	NA
Distress screening (2-item) ²	73.6%	42.2%	80.7%	71.8%	55.4%	NA

Note: AUC; Area Under the Curve.

ALL-model; XGBoost, which included 126 input features, NotSCR-model XGBoost, which excluded the 4-item distress screening from input features, NotSCR+Pain-model; XGBoost, which excluded the 3-item distress screening other than pain NRS from input features, SCR-model; XGBoost, solely used the 4-item distress screening, Simple-model; XGBoost, designed for enhanced generalization which included five features. The 4-item comprises the Distress Thermometer (DT), the Impact Thermometer (IT), the Numerical Rating Scale (NRS) for pain, and NRS for symptoms other than pain.

¹ The predictive performance of distress screening following the hospital's protocol, if a patient scored above the threshold on any of the four distress screening criteria, patients are treated as screening positive.

² The predictive performance of distress screening following the Distress Thermometer and Impact thermometer, if a patient scored above the threshold on either of the two items, they are treated as screening positive.

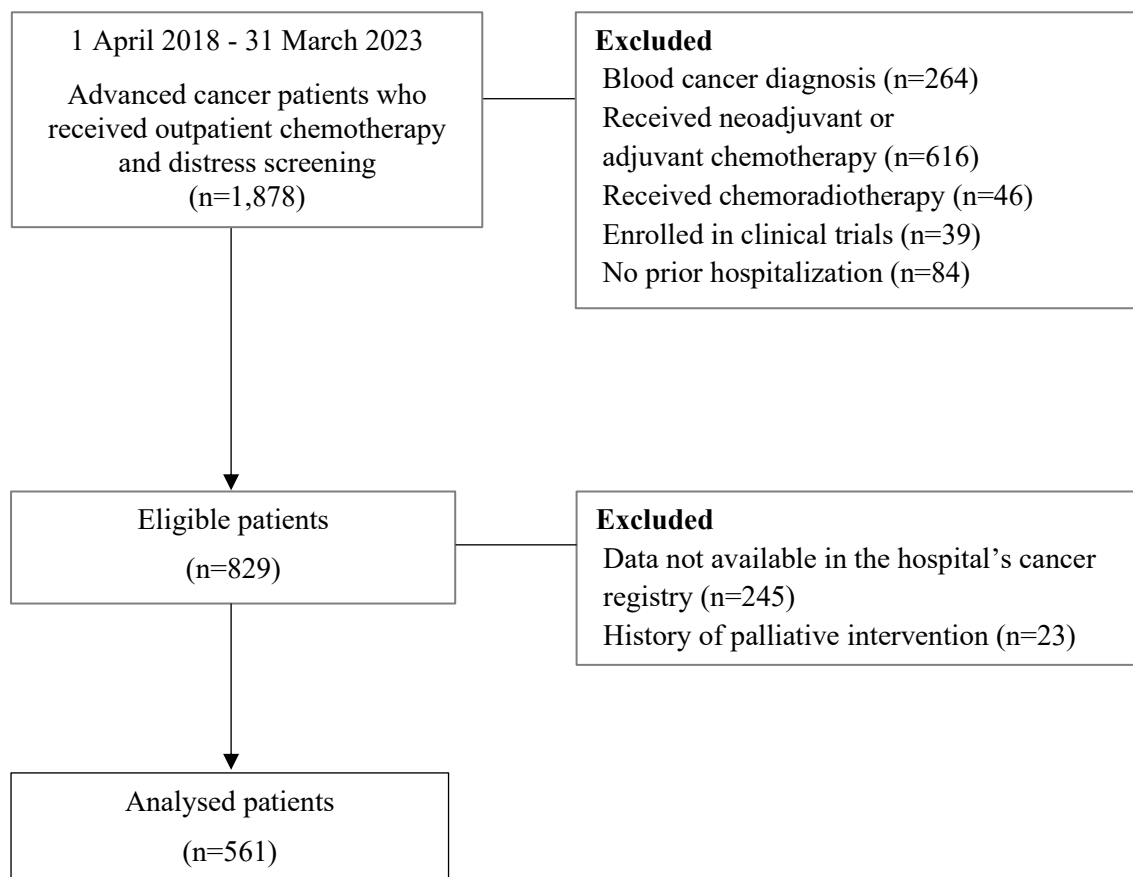


Figure 3. Participant flow

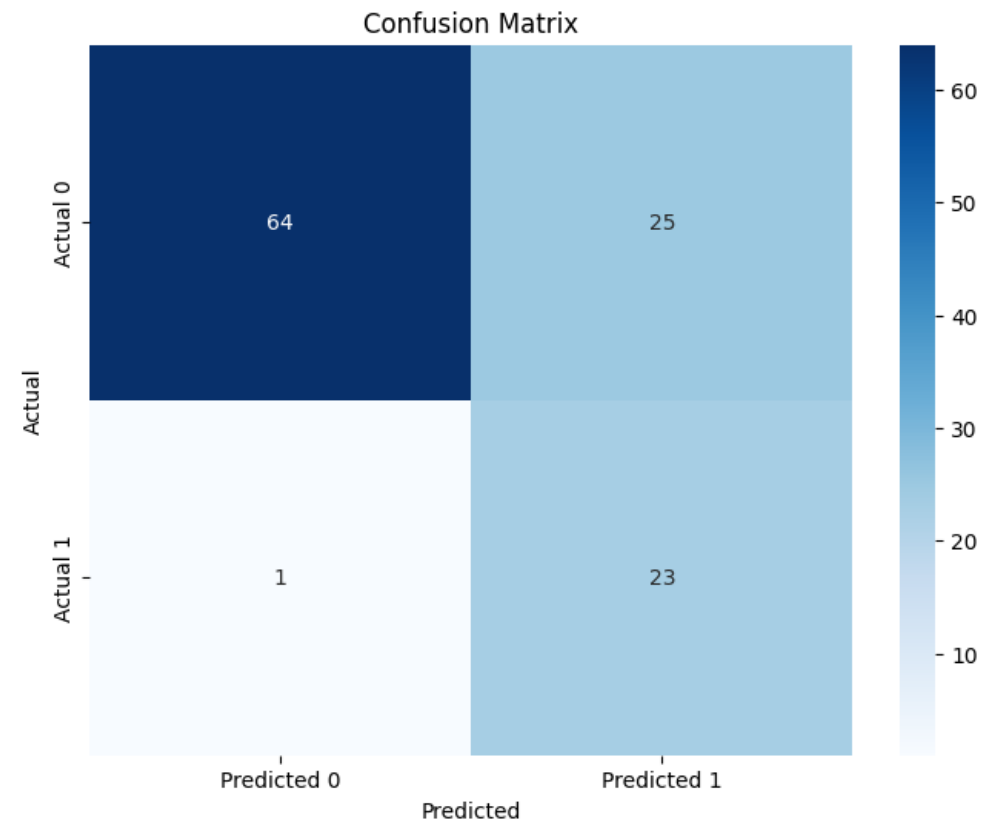
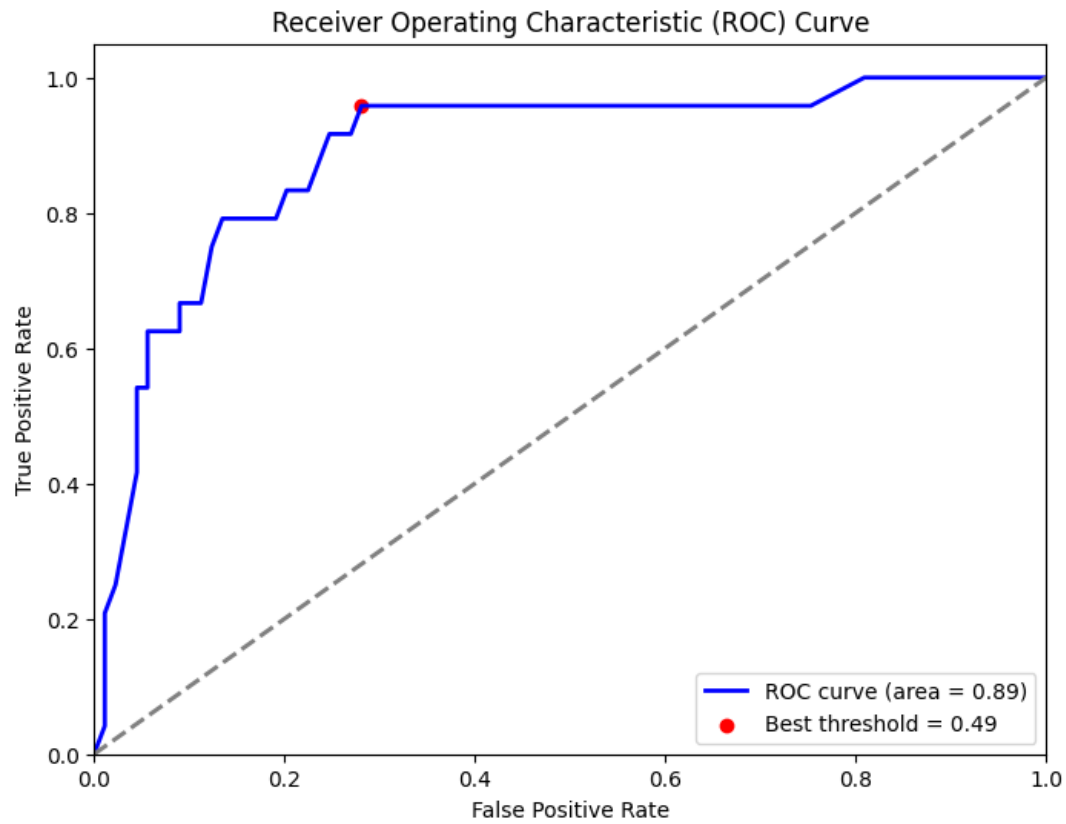


Fig 4. ROC curve and confusion matrix (ALL-model)

Note: Predictive performance (95% Confidence Interval): Accuracy; 77.0% (68.4-83.8), Precision; 47.9% (34.5-61.7), Sensitivity; 95.8% (79.8-99.3), Specificity; 71.9% (61.8-80.2), F1 score; 63.9% (49.3-75.6).

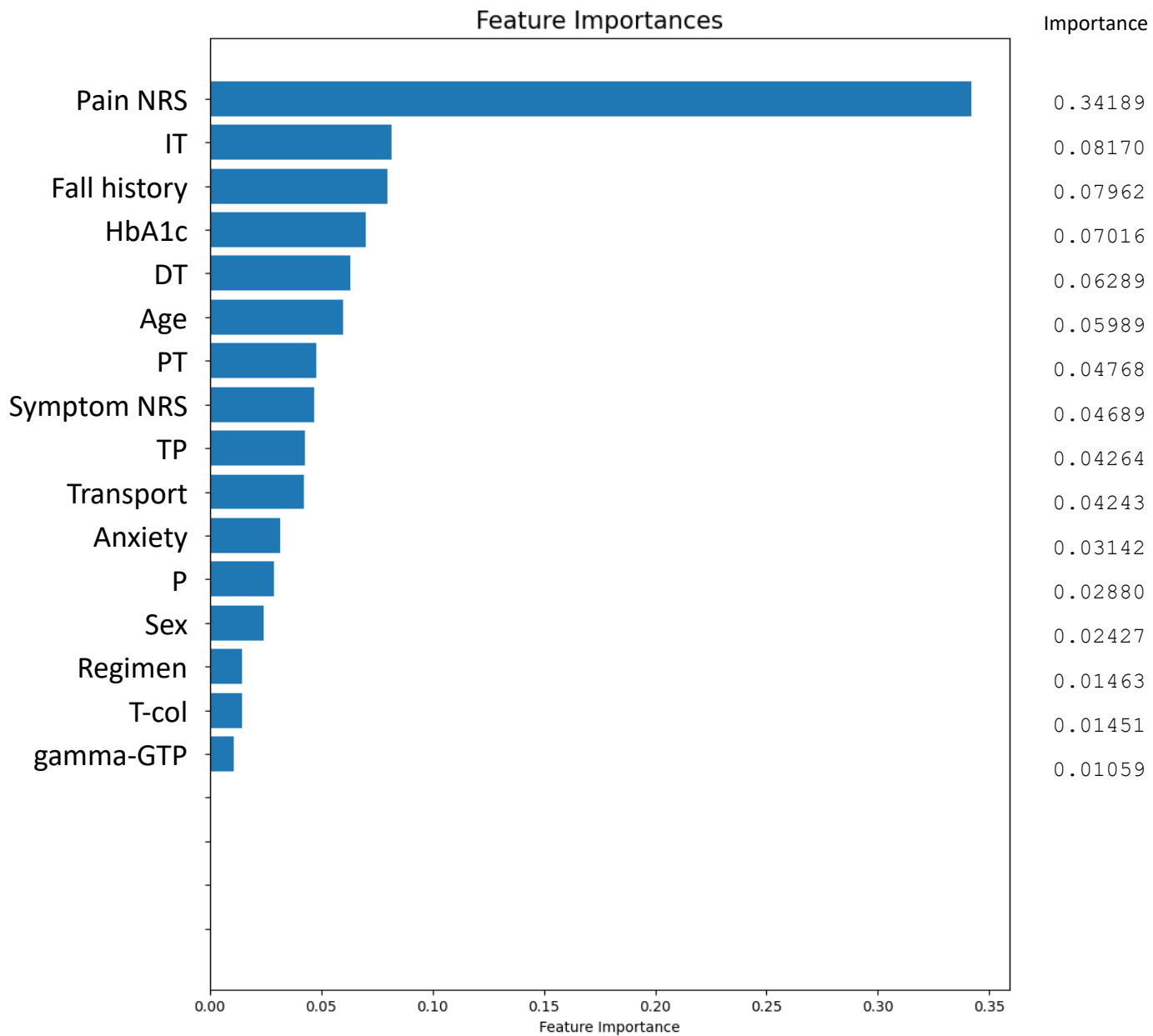


Fig 5. Top 16 features by importance (ALL-model)

Note: NRS; Numerical rating scale, IT; Impact Thermometer, DT; Distress Thermometer, PT; prothrombin time, Symptom NRS; symptom NRS other than pain, Transport; means of transport to a hospital, Anxiety; nursing assessment of anxious facial expressions, P; Phosphorus, Regimen; emetogenic risk of the chemotherapy regimen

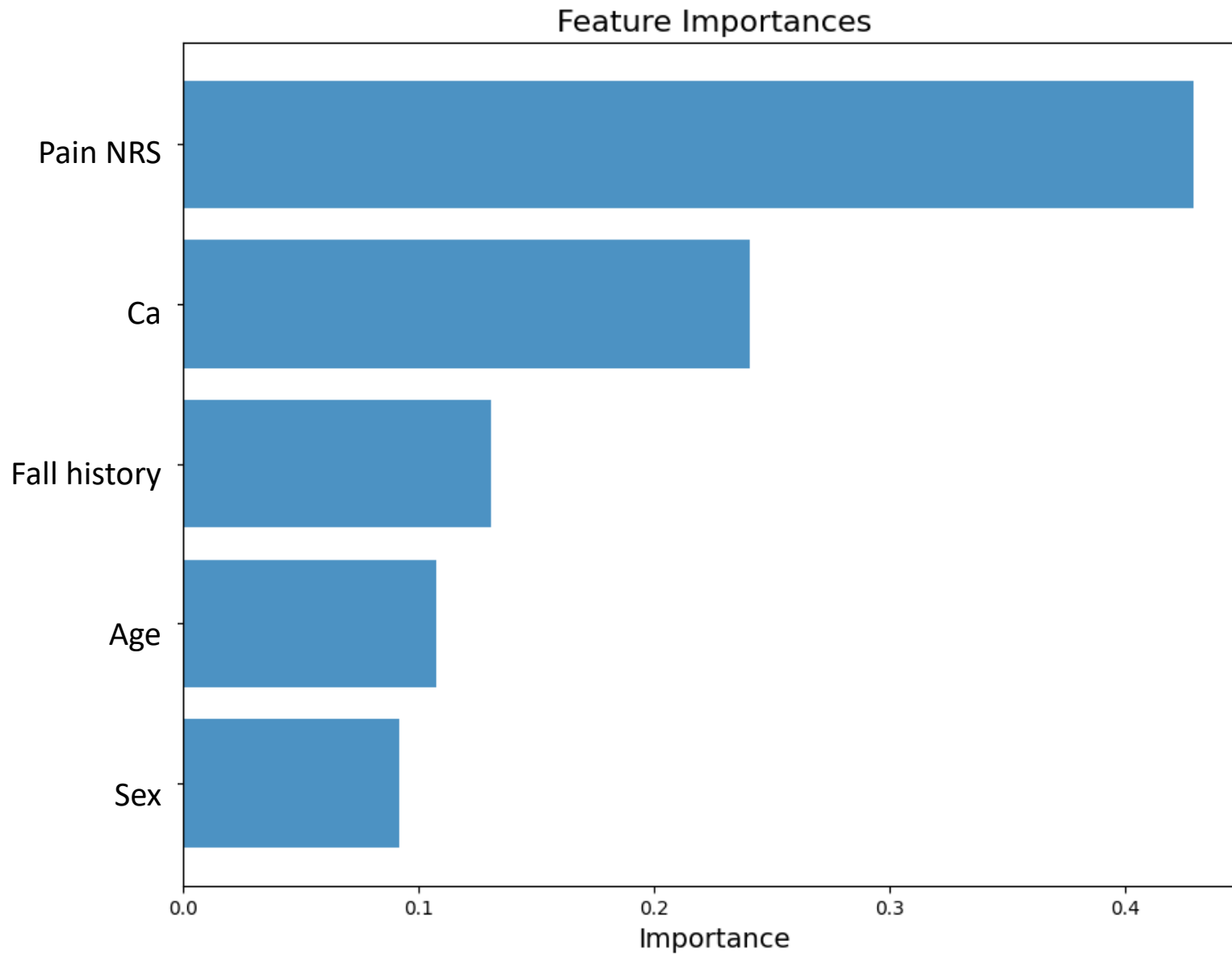
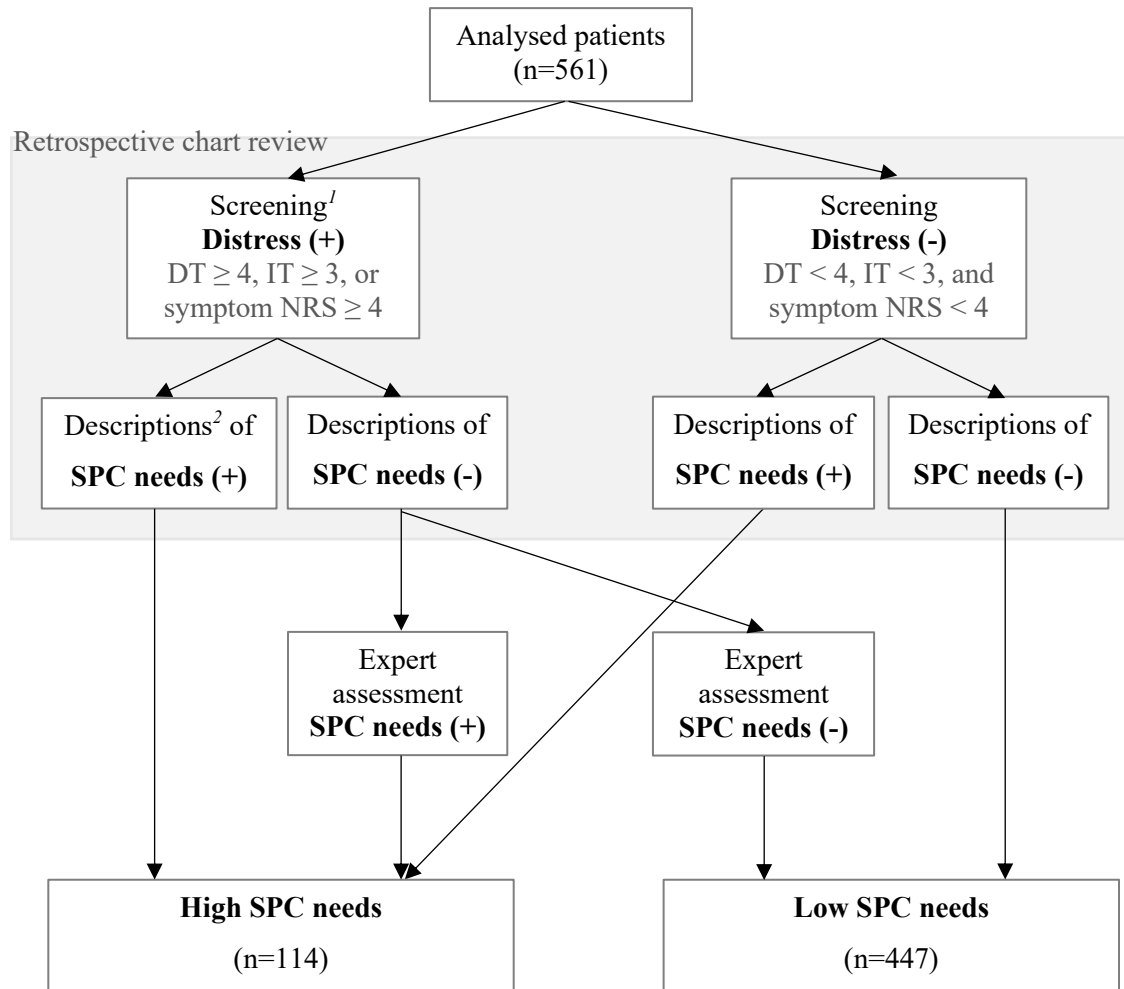


Fig 6. Top features after feature selection (Simple-model)

Note: NRS; Numerical rating scale, Ca; Calcium. After ALL-developed, feature selection was conducted for clinical implication. Predictive performance: Accuracy; 76.1%, Precision; 46.2%, Sensitivity; 75.0%, Specificity; 76.4%, F1 Score; 57.1%, AUC; 0.82.

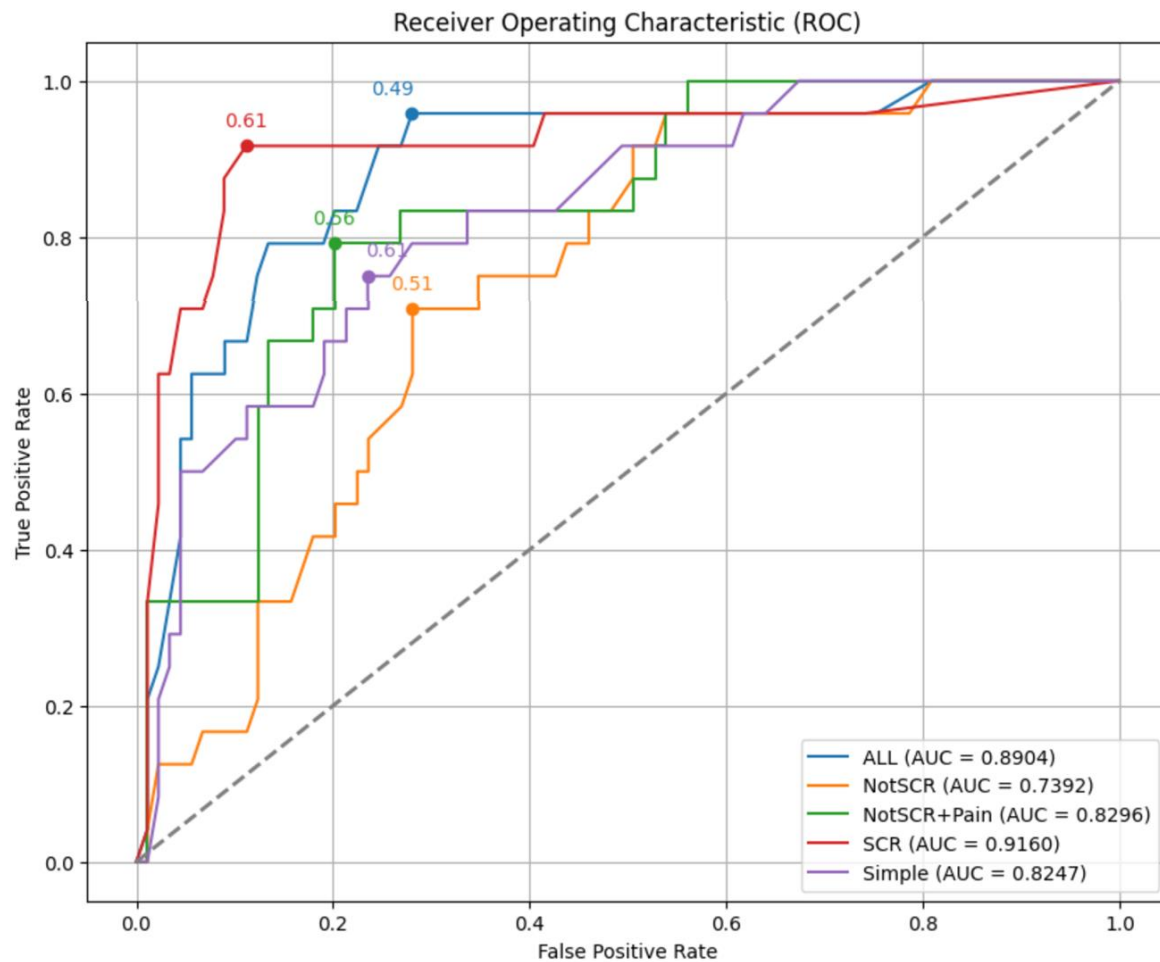


Supplementary Fig 1. Flow chart of specialist palliative care needs assessment

Notes: DT; Distress Thermometer, IT; Impact Thermometer, NRS; Numerical Rating Scale, SPC; Specialist Palliative Care

¹ Distress screening was conducted either during hospitalization or after discharge, but prior to transitioning to outpatient chemotherapy. This screening was carried out in person by a nurse, involving both the patient and their family, in an outpatient chemotherapy room.

² Descriptions of SPC needs; Descriptions related to the nurse's assessment of SPC needs in nursing records.



Supplementary Fig 2. Comparison of ROC curves

Note: ALL-model; XGBoost, which included 126 input features, NotSCR-model XGBoost, which excluded the 4-item distress screening from input features, NotSCR+Pain-model; XGBoost, which excluded the 3-item distress screening other than pain from input features, SCR-model; XGBoost, solely used the 4-item distress screening, Simple-model; XGBoost, designed for enhanced generalization which included five features. The 4-item comprises the Distress Thermometer (DT), the Impact Thermometer (IT), the Numerical Rating Scale (NRS) for pain, and NRS for symptoms other than pain.

	Accuracy	Precision	Sensitivity	F score	AUC
IterativeImputer	75.2%	40.0%	94.7%	56.3%	0.83
Another category	75.2%	40.4%	100.0%	57.6%	0.85
XGBoost	77.9%	43.2%	100.00%	60.3%	0.87

Supplementary Table 1. Sensitivity analysis for missing data imputation in ALL-model

Note: IterativeImputer; For variables with less than a 20% missing rate, we implemented multiple imputation using the IterativeImputer library from scikit-learn and completed the missing data. Using Bayesian Ridge regression, missing data was predicted from other variables including the outcome. Imputation was carried out with a cap of 100 iterations. For variables with more than 20% missing data, it was determined that the cause of the missing data might not be random (MNAR; Missing Not At Random), and therefore, they were not included in the model. Another category (Treat missing data as another category); We created a variable that treats missing data as its own category for use in the prediction model. (Example in blood test data categories; 0=below normal value, 1=within normal value, 2=above normal value, 3=missing). If the missing data was related to a particular year, it was not included in the model. (Example. No data obtained in 2018, but data was obtained from 2019 onwards). Handling of missing data by XGBoost; XGBoost supports missing values by default. In tree algorithms, branch directions for missing values are learned during training.

Supplementary table 2. List of variables for machine learning prediction

Demographic	Age at diagnosis, age at chemotherapy, sex, nationality, residential area, history of smoking, smoking index, history of heavy alcohol consumption, smoking habits.
Disease	Primary cancer site, cancer stage, site of metastasis, days since diagnosis, Charlson comorbidity score, performance status, medical department, presence of multiple cancers, history of psychiatric visits.
Prescription	Chemotherapy regimen, opioid analgesia, presence of cancer recurrence, blood transfusion, line of chemotherapy (e.g., first-line), presence of oxygen administration, emetogenic risk score of chemotherapy regimen.
Blood testing	Albumin, CRP; C-reactive Protein, total protein, White Blood Cells, Red Blood Cells, platelets, neutrophils, direct bilirubin, total bilirubin, AST; Aspartate Aminotransferase, ALT; Alanine Aminotransferase, gamma-Glutamyl Transferase, alkaline phosphatase, amylase, cholinesterase, creatinine, Blood Urea Nitrogen, Creatine Kinase, hemoglobin, haematocrit, hemoglobin A1c, Low-Density Lipoprotein cholesterol, total cholesterol, blood sugar, calcium, sodium, potassium, chloride, magnesium, phosphorus, PT; Prothrombin Time, PT-INR; International Normalized Ratio, APTT; Activated Partial Thromboplastin Time, partial pressure of oxygen, partial pressure of carbon dioxide, Base Excess, Anion Gap, lactate, troponin.
Nutrition	Body Mass Index (BMI), weight loss in the past 6 months, lack of appetite, dysphagia score, pressure sore risk score, presence of current pressure sores, history of pressure sores, presence of bony prominences, presence of pressure ulcer procedure, pressure sore classification by the NPUAP (National Pressure Ulcer Advisory Panel).
Healthcare utilisation	Hospitalization history in the past year, ICU utilization in the past year, utilization of home-visit nursing, outpatient or inpatient procedure, admission from emergency room, utilization of home care, emergency hospital admission, unplanned hospital admission, reason for hospital admission.
Literacy	Presence of literacy deficiencies, learning disabilities, language.
Socioeconomic	Employment status, housemate, number of housemates, relationship with key person, travel time to hospital, mode of transportation to hospital, religion, insurance type, hobby, desire for intervention, desired medical/social service type.
Symptoms and functional status	Numerical Rating Scale (NRS) of pain, NRS of physical symptoms other than pain, the Distress Thermometer, the Impact Thermometer, respiratory rate, blood pressure, heart rate, body temperature, oxygen saturation presence of anxious expression, presence of dyspnea, presence of dizziness, presence of numbness or tingling, independence degree in daily living, physical function score, risk score of falling, fall history in the past year, presence of muscle weakness, presence of joint contracture, presence of joint disorders, type of walking aid, unsteady gait, use of diapers, level of consciousness, cognitive function score, delirium risk score, care dependency, independence in; eating, bathing, grooming, dressing, walking, stair climbing, urinary management, bowel management.

Note: All features, except binary ones, were categorized and then input into XGBoost. Blood testing features were divided into three categories; below the normal value, within the normal range, and above the normal value. For the prediction of SPC needs at the time of distress screening, we extracted information from the patient's most recent admission within the 3-month prior to the distress screening date.

Supplementary table 3. P-values between before and after sampling

	Imbalanced data (n=561)		Rebalanced data (n=190) ¹		<i>p-value</i>
	Low SPC needs (n=447)		Low SPC needs (n=95)		
	N or mean	<i>SD or (%)</i>	N or mean	<i>SD or (%)</i>	
Age (years)	66.8	10.8	66.9	12.0	0.935
Sex assigned at birth (female)	156	(80.0)	39	(20.0)	0.309
Cancer types					0.355
Gastrointestinal	147	(82.1)	32	(17.9)	
Hepatobiliary and pancreatic	106	(66.8)	25	(33.2)	
Lung	103	(81.1)	24	(18.9)	
Urological	40	(88.9)	5	(11.1)	
Gynecological	21	(77.3)	6	(22.7)	
Head and neck	10	(90.9)	1	(9.1)	
Skin	7	(100)	0	(0)	
Others	7	(77.8)	2	(22.2)	
Multiple cancers	59	(84.3)	11	(15.7)	0.795
Recurrence	159	(80.7)	38	(19.3)	0.820
Chemotherapy					0.343
1 st line	350	(81.4)	80	(18.6)	
2 nd line	64	(85.3)	11	(14.7)	

3 rd line	27	(93.0)	2	(7.0)	
4 th line	6	(75.0)	2	(25.0)	
ECOG PS					0.953
0	379	(82.4)	81	(17.6)	
1	46	(82.1)	10	(17.9)	
2	8	(80.0)	2	(20.0)	
3	2	(100)	0	(0)	
4	1	(100)	0	(0)	
Metastasis					
Liver	159	(81.0)	37	(19.0)	0.618
Lung	153	(83.6)	30	(16.4)	0.699
Bone	93	(88.6)	12	(11.4)	0.090
Brain	27	(84.4)	5	(15.6)	0.956
Consciousness level ²					0.588
Clear	437	(89.9)	93	(10.1)	
1-digit	8	(88.9)	1	(11.1)	
2-digit	1	(50)	1	(50)	
Distress screening ³					
NRS \geq 4 (pain)	48	(88.9)	6	(11.1)	0.261
NRS \geq 4 (other symptoms)	2	(66.7)	1	(33.3)	1.0

Distress Thermometer ≥ 4	120	(80.0)	30	(20.0)	0.425
Impact Thermometer ≥ 3	64	(83.1)	13	(16.9)	0.995

Notes: N; number, SD; standard deviation, JCS; Japan Coma Scale, NRS; Numerical Rating Scale, ECOG PS; The Eastern Cooperative Oncology Group Performance Status.

¹ Rebalanced data is a dataset that has had under-sampling applied.

² Consciousness level was assessed by the Japan Coma Scale at the time of hospitalization. A 1-digit score indicates that the patient is awake without any stimuli. A 2-digit score indicates that the patient can be aroused and then revert to the previous state after cessation of stimulation

³ Four items were assessed as part of the distress screening at the hospital: numerical rating scale of pain, numerical rating scale of physical symptoms other than pain, Distress Thermometer (DT), and Impact thermometer (IT) for psychological distress. NRS for physical symptoms excluding pain had 109 missing values, and DT score and IT score had 2 missing values, respectively.

Supplementary Table 4. Sensitivity analysis for imbalanced data

	Accuracy	Precision	Sensitivity	Specificity	F score	AUC
ALL-model in imbalanced data	78.8%	0%	0%	100%	0%	0.50
ALL-model in rebalanced data ¹	77.0%	47.9%	95.8%	71.9%	63.9%	0.89
Simple-model in imbalanced data	81.4%	80.0%	16.7%	98.9%	27.6%	0.58
Simple-model in rebalanced data ¹	76.1%	46.2%	75.0%	76.4%	57.1%	0.82

Note: ALL-model; XGBoost, which included 126 input features, Simple-model; XGBoost, designed for enhanced generalization which included five features.

¹ Rebalanced data is a dataset that has had under-sampling applied.

Supplementary Table 5. TRIPOD Checklist: Prediction Model Development

Section/Topic		Checklist Item	Page
Title and abstract			
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	2
Introduction			
Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	3,4
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	3,4
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	5
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	5
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	5
	5b	Describe eligibility criteria for participants.	5
	5c	Give details of treatments received, if relevant.	5
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	5,6
	6b	Report any actions to blind assessment of the outcome to be predicted.	6
Predictors	7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	6,7, suppl.
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	7
Sample size	8	Explain how the study size was arrived at.	NA
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	8
Statistical analysis methods	10a	Describe how predictors were handled in the analyses.	7,8
	10b	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	7
	10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	7
Risk groups	11	Provide details on how risk groups were created, if done.	8
Results			
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	Fig 1
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	9
Model development	14a	Specify the number of participants and outcome events in each analysis.	9, 10
	14b	If done, report the unadjusted association between each candidate predictor and outcome.	NA
Model specification	15a	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	10
	15b	Explain how to use the prediction model.	10
Model performance	16	Report performance measures (with CIs) for the prediction model.	9
Discussion			
Limitations	18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	13,14
Interpretation	19b	Give an overall interpretation of the results, considering objectives, limitations, and results from similar studies, and other relevant evidence.	11-13
Implications	20	Discuss the potential clinical use of the model and implications for future research.	12,13
Other information			
Supplementary information	21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	15
Funding	22	Give the source of funding and the role of the funders for the present study.	14

8. 引用文献

1. Knaul FM, Farmer PE, Krakauer EL, et al. Alleviating the access abyss in palliative care and pain relief-an imperative of universal health coverage: the Lancet Commission report. *Lancet* 2017; 391(10128): 1391-454.
2. Munday D, Boyd K, Jeba J, Kimani K, Moine S, Grant L, Murray S. Defining primary palliative care for universal health coverage. *Lancet*. 2019;394(10199):621-622.
3. Palliative Care Australia. Palliative Care Service Development Guidelines. Australia: Palliative Care Australia incorporated, 2018.
4. Chang VT, Hwang SS, Feuerman M, Kasimis BS. Symptom and quality of life survey of medical oncology patients at a veterans affairs medical center: a role for symptom assessment. *Cancer* 2000;88(5):1175–83.
5. Downey L, Engelberg RA. Quality-of-life trajectories at the end of life: assessments over time by patients with and without cancer. *J Am Geriatr Soc*. 2010;58(3):472-479.
6. Giesinger JM, Wintner LM, Oberguggenberger AS, et al.. Quality of life trajectory in patients with advanced cancer during the last year of life. *J Palliat Med*. 2011;14(8):904-912.
7. Levy MH, Smith T, Alvarez-Perez A, Back A, Baker JN, Beck A, et al. NCCN Clinical Practice Guidelines in Oncology. Palliative Care.
URL: http://www.nccn.org/professionals/physician_gls/f_guidelines.asp [accessed Sep 10, 2023].
8. Ferrell BR, Temel JS, Temin S, Alesi ER, Balboni TA, Basch EM, et al. Integration of Palliative Care Into Standard Oncology Care: American Society of Clinical Oncology Clinical Practice Guideline Update. *J Clin Oncol* 2017;35(1):96–112
9. Temel JS, Greer JA, Muzikansky A, et al. Early palliative care for patients with metastatic non-small-cell lung cancer. *N Engl J Med*. 2010;363:733–742. [PubMed] [Google Scholar]

10. Zimmermann C, Swami N, Krzyzanowska M, et al. Early palliative care for patients with advanced cancer: A cluster-randomised controlled trial. *Lancet*. 2014;383:1721–1730. [PubMed] [Google Scholar]
11. Vogt J, Beyer F, Sistermanns J, Kuon J, Kahl C, Alt-Epping B, Stevens S, Ahlborn M, George C, Heider A, Tienken M, Loquai C, Stahlhut K, Ruellan A, Kubin T, Dietz A, Oechsle K, Mehnert-Theuerkauf A, van Oorschot B, Thomas M, Ortmann O, Engel C, Lordick F; Arbeitsgemeinschaft Palliativmedizin (APM) of the German Cancer Society (DKG). Symptom Burden and Palliative Care Needs of Patients with Incurable Cancer at Diagnosis and During the Disease Course. *Oncologist*. 2021;26(6):e1058-e1065.
12. Hui D, Kim SH, Kwon JH, Tanco KC, Zhang T, Kang JH, et al. Access to palliative care among patients treated at a comprehensive cancer center. *Oncologist* 2012;17(12):1574–80.
13. Hui D, Hannon BL, Zimmermann C, Bruera E. Improving patient and caregiver outcomes in oncology: Team-based, timely, and targeted palliative care. *CA Cancer J Clin*. 2018;68(5):356-376.
14. Hui D, Heung Y, Bruera E. Timely Palliative Care: Personalizing the Process of Referral. *Cancers (Basel)*. 2022.
15. Hui D, Mori M, Watanabe SM, Caraceni A, Strasser F, Saarto T, Cherny N, Glare P, Kaasa S, Bruera E. Referral criteria for outpatient specialty palliative cancer care: an international consensus. *Lancet Oncol*. 2016;17(12):e552-e559.
16. Solar S, Wieditz J, Lordick F, Mehnert-Theuerkauf A, Oechsle K, van Oorschot B, Thomas M, Asendorf T, Nauck F, Alt-Epping B. Screening versus multidimensional assessment of symptoms and psychosocial distress in cancer patients from the time of incurability. *Front Oncol*. 2023;13:1002499.
17. 二宮他. 外来がん化学療法患者を対象とした苦痛のスクリーニングの導入—苦痛対応に

必要なリソースに関する分析ー, 2019年14巻1号, Palliative Care Research, p. 15-21

18. Porter AS, Harman S, Lakin JR. Power and perils of prediction in palliative care. *Lancet*. 2020 Feb 29;395(10225):680-681. doi: 10.1016/S0140-6736(20)30318-4.

19. Storick V, O'Herlihy A, Abdelhafeez S, Ahmed R, May P. Improving palliative and end-of-life care with machine learning and routine data: a rapid review. *HRB Open Res*. 2019 Jul 15;2:13.

20. Bange EM, Courtright KR, Parikh RB. Implementing automated prognostic models to inform palliative care: more than just the algorithm. *BMJ Qual Saf*. 2021 Oct;30(10):775-778.

21. Murphree DH, Wilson PM, Asai SW, Quest DJ, Lin Y, Mukherjee P, Chhugani N, Strand JJ, Demuth G, Mead D, Wright B, Harrison A, Soleimani J, Herasevich V, Pickering BW, Storlie CB. Improving the delivery of palliative care through predictive modeling and healthcare informatics. *J Am Med Inform Assoc*. 2021 Jun 12;28(6):1065-1073.

22. Ownby KK. Use of the Distress Thermometer in Clinical Practice. *J Adv Pract Oncol*. 2019 Mar;10(2):175-179.

23. Cutillo A, O'Hea E, Person S, Lessard D, Harralson T, Boudreaux E. The Distress Thermometer: Cutoff Points and Clinical Use. *Oncol Nurs Forum*. 2017 May 1;44(3):329-336.

24. Hui D, Meng YC, Bruera S, Geng Y, Hutchins R, Mori M, Strasser F, Bruera E. Referral Criteria for Outpatient Palliative Cancer Care: A Systematic Review. *Oncologist*. 2016 Jul;21(7):895-901.

24. Wolgast A, Schwinger M, Hahnel C, Stiensmeier-Pelster J. Handling Missing Data in Structural Equation Models in R. A Replication Study for Applied Researchers, *Electronic Journal of Research in Educational Psychology*, 15(1), 5-47. ISSN: 1696-2095. 2017. no. 41.p11.

25. Hamano J, Oishi A, Kizawa Y. Identified Palliative Care Approach Needs with SPICT in Family Practice: A Preliminary Observational Study. *J Palliat Med*. 2018 Jul;21(7):992-998.

26. Morita T, Fujimoto K, Namba M, Sasaki N, Ito T, Yamada C, Ohba A, Hiroyoshi M, Niwa H, Yamada T, Noda T. Palliative care needs of cancer outpatients receiving chemotherapy: an audit of a clinical screening project. *Support Care Cancer*. 2008 Jan;16(1):101-7.
27. Hamdan Alshehri H, Olausson S, Öhlén J, Wolf A. Factors influencing the integration of a palliative approach in intensive care units: a systematic mixed-methods review. *BMC Palliat Care*. 2020 Jul 22;19(1):113.
28. de Oliveira Valentino TC, Paiva BSR, de Oliveira MA, Hui D, Paiva CE. Factors associated with palliative care referral among patients with advanced cancers: a retrospective analysis of a large Brazilian cohort. *Support Care Cancer*. 2018 Jun;26(6):1933-1941.
29. Carlson LE, Waller A, Mitchell AJ. Screening for distress and unmet needs in patients with cancer: review and recommendations. *J Clin Oncol*. 2012 Apr 10;30(11):1160-77.
30. Hollingworth W, Metcalfe C, Mancero S, et al. Are needs assessments cost effective in reducing distress among patients with cancer? A randomized controlled trial using the Distress Thermometer and Problem List. *J Clin Oncol*. 2013;31(29):3631–3638.
31. Masukawa K, Aoyama M, Yokota S, Nakamura J, Ishida R, Nakayama M, Miyashita M. Machine learning models to detect social distress, spiritual pain, and severe physical psychological symptoms in terminally ill patients with cancer from unstructured text data in electronic medical records. *Palliat Med*. 2022 Sep;36(8):1207-1216.
32. Shimada K, Tsuneto S. Novel method for predicting nonvisible symptoms using machine learning in cancer palliative care. *Sci Rep*. 2023 Jul 26;13(1):12088.