慶應義塾大学学術情報リポジトリ
Keio Associated Repository of Academic resouces

Title	ソーシャルメディアからの患者の悩み・実践知の抽出技術と活用基盤の確立
Sub Title	Development of technology and infrastructure for extracting and utilizing patient concerns and practical knowledge from social media
Author	堀, 里子(Hori, Satoko) 荒牧, 英治(Aramaki, Eiji)
Publisher	
Publication year	2024
Jtitle	科学研究費補助金研究成果報告書 (2023.)
JaLC DOI	
Abstract	本研究では，ソーシャルメディア上のがん患者の記述や語りを対象として，自然言語処理技術を用いて疾患や治療に関連する自覚症状や悩みのエピソードを抽出する手法を開発した．具体的には，"身体の苦痛"や"就労・経済的負担"など複合的に発生する悩みを捉える高性能なマルチラベル分類器，実践知の把握や感情分析を行う手法，手足症候群シグナルや日常生活に支障を及ぼす有害事象シグナルの抽出器の構築に成功した．さらに，患者症状訴えを含む異なるテキストソース（薬局薬歴等）に対してもこれらの手法を活用し，その臨床的有用性を示した． In this study, episodes of subjective symptoms and concerns related to diseases and treatments were effectively extracted from narratives of cancer patients on social media, using natural language processing (NLP) technology. The research specifically targeted unique concerns of cancer patients, such as "physical concerns" and "work/financial burdens," which led to the development of a high-performance multi-label classifier. Moreover, an innovative extractor was designed to identify signals for hand-foot syndrome and adverse events with a significant impact on daily-life activities. The clinical utility of this extractor was demonstrated through its application to various text sources, including pharmaceutical care records containing detailed patients' concerns. The significance of this research lies in its development of a method to extract data from non-medical texts using cutting-edge NLP technology, capturing concerns and practical knowledge from the patients' perspective.
Notes	研究種目：基盤研究 (B) (一般) 研究期間：2021～2023 課題番号：21H03170 研究分野：健康情報学
Genre	Research Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KAKEN_21H03170seika

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the KeiO Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　６ 年　 ６ 月　 ４ 日現在

機関番号：３２６１２
研究種目：基盤研究(B)（一般）
研究期間：2021～2023
課題番号：２１Ｈ０３１７０

研究課題名（和文）ソーシャルメディアからの患者の悩み・実践知の抽出技術と活用基盤の確立

研究課題名（英文）Development of technology and infrastructure for extracting and utilizing patient concerns and practical knowledge from social media

研究代表者

堀　里子（Hori, Satoko）

慶應義塾大学・薬学部（芝共立）・教授

研究者番号：７０３１３１４５

交付決定額（研究期間全体）：（直接経費）　　12,800,000円

研究成果の概要（和文）：本研究では，ソーシャルメディア上のがん患者の記述や語りを対象として，自然言語処理技術を用いて疾患や治療に関連する自覚症状や悩みのエピソードを抽出する手法を開発した．具体的には，“身体の苦痛”や“就労・経済的負担”など複合的に発生する悩みを捉える高性能なマルチラベル分類器，実践知の把握や感情分析を行う手法，手足症候群シグナルや日常生活に支障を及ぼす有害事象シグナルの抽出器の構築に成功した．さらに，患者症状訴えを含む異なるテキストソース（薬局薬歴等）に対してもこれらの手法を活用し，その臨床的有用性を示した．

研究成果の学術的意義や社会的意義
非医療者が生成したテキストから，自然言語処理技術を用いて患者視点で悩みや症状を捉える抽出手法を確立した本研究の意義は大きい．本研究により，医療介入が必要な事象を早期に捉えることが可能となり，患者の治療生活の質を大きく向上させることが期待される．さらに，この技術を応用することで，病気治療中に患者の記述から患者の自覚症状や悩み，実践知を可視化し，医療者とのコミュニケーションを促進し，患者のQOL（Quality of Life）を向上させることが期待される．

研究成果の概要（英文）：In this study, episodes of subjective symptoms and concerns related to diseases and treatments were effectively extracted from narratives of cancer patients on social media, using natural language processing (NLP) technology. The research specifically targeted unique concerns of cancer patients, such as "physical concerns" and "work/financial burdens," which led to the development of a high-performance multi-label classifier. Moreover, an innovative extractor was designed to identify signals for hand-foot syndrome and adverse events with a significant impact on daily-life activities. The clinical utility of this extractor was demonstrated through its application to various text sources, including pharmaceutical care records containing detailed patients’ concerns. The significance of this research lies in its development of a method to extract data from non-medical texts using cutting-edge NLP technology, capturing concerns and practical knowledge from the patients' perspective.

研究分野：健康情報学

キーワード：ソーシャルメディア　自然言語処理　深層学習　患者の悩み　患者主訴

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

1. 研究開始当初の背景

疾患治療の適正化には，患者が症状・副作用や悩み（ニーズ）に対して適切に対処できることが重要である．しかし，日常診療では患者が体験する心身の不調がしばしば過小評価されている可能性が指摘されている．患者は治療を通じて，例えば副作用の対処方法等の医療的知識や，治療に伴う生活上の悩みの解消方法といった社会的知識を徐々に習得する．患者会・団体は患者同士の経験の共有において重要な役割を果たしているが，これらの情報にアクセスできていない患者が大半であり，活用されないままの患者の記録や記述が極めて多い．近年，情報化社会の進展により個人での情報発信が容易になり，患者やその家族がインターネット上のソーシャルメディア（ブログ，インタビュー配信等）で疾病体験について発信する機会が増えている．これらのプラットフォームには医療従事者には伝わりづらいが，患者の治療の適正化や QOL 向上に役立つ可能性のある情報が含まれている．患者は医療者だけでなく，インターネットを情報源として利用しているが，インターネット上の情報はその種類と量が膨大であり，適切な情報を選択することが困難である．また，膨大な情報を人手で整理し，患者に適切で正しい情報を提示することには多大な労力を要する．このような状況のなか，申請者らはソーシャルメディアの膨大な患者テキストから，医療的かつ患者にとって価値ある知見を見つけ出し，活かすことを着想した．

2. 研究の目的

本課題では，ソーシャルメディアから得られる患者による記述から，自然言語処理（Natural Language Processing, NLP）技術を用いて，疾患・治療に関連する自覚症状や悩みのエピソードを抽出することを目的とした．特に，がん患者に焦点を当て，彼らが長期にわたる治療過程で直面する様々な悩みや有害事象を捉えるための手法の開発を目指した．さらに，がん患者のケアの質向上のための基盤構築に向けて，開発した NLP モデルの活用可能性を検証した．

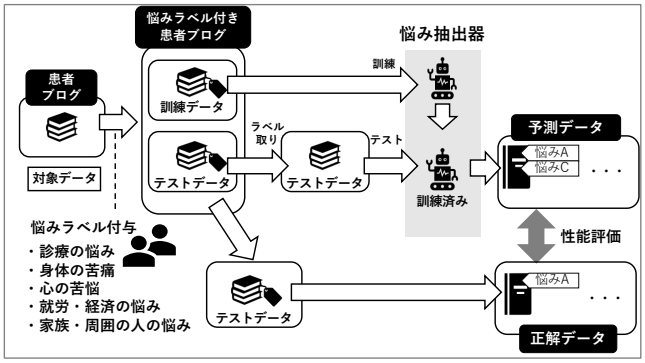
3. 研究の方法

（1）がん患者の悩み・感情抽出手法の開発

①　がん患者における悩みの特徴理解とマルチラベル悩み分類器の開発 [1-3]

患者ブログ「ライフパレット」（メディエイドより提供）に投稿された乳がん患者による記事 2,272 件（55 名），乳がん患者の語り（DIPEx Japan よりデータシェアリングにより提供）のテキスト 508 件（52 名）を解析対象とした．まず，乳がん患者ブログの記事内容の特徴理解を試みた．Doc2vec を用いて記事文章をベクトルに変換し，k-means 法による非階層クラスタリングを行った．分かれたクラスタ毎の特徴を頻出語，ワードクラウド，共起ネットワーク，悩みラベル付与割合を用いて分析した．また，ユーザーの発信内容の変化をクラスタ分類に基づき検討した．

つづいて，患者の悩みとして 5 種の悩みラベル（「診療の悩み」「身体の苦痛」「心の苦悩」「就労・経済の悩み」「家族・周囲の人の悩み」）を設定し，各記事に対し複数ラベルを許容して付与した．対象記事のテキストを入力とし，事前学習済み Bidirectional Encoder Representations from Transformers（BERT）に分類層を付け足したモデルをファインチューニングすることで，5 種の悩みを予測するマルチラベル分類器を構築した．また，ファインチューニングの前に，乳がん患者ブログデータあるいは乳がん患者の語りデータを使用してドメイン適応を行った分類器を構築した．5 分割交差検証を行い，これらの手順で構築した分類器の性能を評価した（**右図**）．



②　がん患者のマルチクラス・マルチラベル感情分類手法の開発 [4]

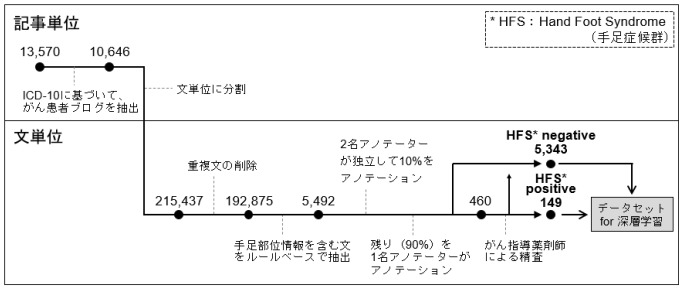
がん経験者が自身のペイシェントジャーニーについて語ったインタビュー（NPO 法人がんノート）の逐語録を対象とした．インタビュー逐語録から 1,998 件（15 名）のデータセットを作成した．それらのデータに対して 2 つの方法でアノテーションを行った．方法 1 では Positive/Neutral/Negative のいずれかのラベルを付与し，方法 2 では Plutchik の 8 感情 [5] を複数ラベルを許容して付与した．ラベル付与したテキストを訓練データとして，事前学習済み BERT 及び LUKE モデルをそれぞれファインチューニングし，3 感情のマルチクラス分類器と Plutchik の 8 感情のマルチラベル分類器を構築した．がん経験者インタビューをテストデータとして性能を評価した．同時に，既存の感情ラベル付き SNS データセット WRIME[6] を訓練データとした分類器を同様に構築しこれらの性能を比較した．

（2）がん患者の有害事象訴え抽出手法の開発

①　手足症候群シグナル抽出器の開発 [7]

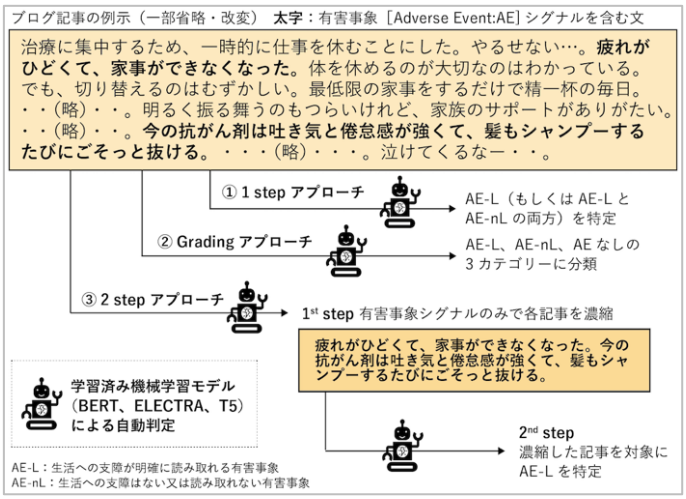
患者ブログ「ライフパレット」に投稿されたがん患者によるブログ記事 13,570 件（287 名分）を対象とした．対象記事から手足部位情報を含む文を絞り込んだ後，手足症候群の発現が疑われ

る文をアノテーションした．がん専門病院勤務のがん指導薬剤師による精査を経て，手足症候群 positive 文を特定し，それ以外を negative 文とした（**右図**）．得られた positive/negative 文の 8 割を用いて深層学習モデルの訓練を行い，残りの 2 割で性能評価（適合率，再現率，F1 スコア）を実施した．手足症候群シグナルを文及びユーザー単位で特定する抽出器を構築した．本実験では，文脈理解が可能な深層学習 3 モデル（Long Short-Term Memory ［ LSTM ］，Bidirectional LSTM ［ Bi-LSTM ］，BERT）を使用した．



② 日常生活への支障度に応じた有害事象シグナル抽出器の開発 [8]

方法(1)①で対象とした乳がん患者のブログ記事 2,272 件を用いて，我々の先行研究で「身体トピック」ラベルが付与された記事を対象に，患者表現から日常生活動作への支障が明らかに読み取れる有害事象を AE-L，それ以外の有害事象を AE-nL，有害事象のない場合を no-AE と定義しアノテーションした．このデータの 8 割を用いて訓練を行い，残り 2 割で性能を評価し，日常生活への支障度に応じた有害事象シグナル抽出器を構築した．本実験では文脈理解が可能な深層学習 3 モデル（BERT，Efficiently Learning an Encoder that Classifies Token Replacements Accurately［ELECTRA］，Text-To-Text Transfer Transformer［T5］）を使用した．さらに，本実験では AE-L を効率よく抽出するために 3 つの異なるアプローチ（1 step，Grading，2 step）の性能を比較した（**右図**）．また，All AE の症状種別も解析した．



（3）開発手法の有用性評価 [9]

方法（2）で構築した 2 つの異なる特徴を持つ有害事象シグナル抽出器の有用性を，他のテキストデータへの適用を通じて評価した．適用先として，テキスト化された乳がん患者のインタビュー内容（DIPEx Japan）および SOAP 形式で記録された薬局薬歴データ（ナカジマ薬局）を使用した．特に薬歴データでは，S 記録に記載された患者の主訴と，その他の項目に含まれる医療情報を利用し，抽出された有害事象シグナルに基づく医療介入の内容やタイミング，処方された抗がん剤の種類等を解析した．

なお，本研究は，慶應義塾大学薬学部倫理委員会の承認を得て実施した（(1) ①：承 201120-3，承 210914-1，(2) ①②：承 201120-3，(3)：承 210914-1，承 230217-1）．

4. 研究成果

（1）がん患者の悩み・感情抽出手法の開発

① がん患者における悩みの特徴理解とマルチラベル悩み分類器の開発 [1-3]

乳がん患者ブログのクラスタ分析により，ブログ記事は 7 つのクラスタ（“病院外でのポジティブな話題”“診察の内容や悩み等のシビアな話題”“マイナス感情の話題”“友達，仲間，家族等周囲の人との話題，病気に立ち向かう前向きでポジティブな話題”“治療法やその効果・苦痛の話題”等）に分類された．これにより，ブログには患者 QOL に影響を及ぼすものの医療従事者に伝わりづらい話題が多く含まれていることが明らかになった．各ユーザーの記事分析からは，治療時期によって変化する患者の悩みや心情が可視化され，多様なペイシェントジャーニーの一端を読み取ることができた．

がん患者のブログテキストから構築したマルチラベル悩み分類器は，5 種類の悩みラベルに対していずれも高い適合率を示し，特に「身体の苦痛」の適合率（0.82）が高かった．一方で，がん患者の語りテキストを用いて構築したマルチラベル悩み分類器では，「身体の苦痛」に加えて，「就労・経済的負担」も高い適合率（0.80 超）を示した．その一方，「診療の悩み」の抽出は難しかったが（適合率 0.25），ドメイン適応を行うことで，分類器の全体の性能が顕著に向上した．ドメイン適応は，患者の語りテキストからの悩み抽出精度を向上させる手段として有用であることが明らかになった．本研究によって，BERT を用いて患者テキストから 5 種の悩みを抽出するマルチラベル分類器が構築できた．

② がん患者のマルチクラス・マルチラベル感情分類手法の開発[4]

がん経験者インタビューテキストを用いて LUKE をファインチューニングした結果，3 感情（ポジティブ, ニュートラル, ネガティブ）のマルチクラス分類器が均等な性能を示した．一方，8 感情（喜び，信用，恐れ，驚き，悲しみ，嫌悪，怒り，期待）の分類では，大部分の感情において良好な性能を示したものの，一部の感情（特に「驚き」や「怒り」）の抽出が困難であった．感情の出現頻度や表現の多様性，インタビューのテーマ性の影響が原因であると考えられる．これらの感情の識別と理解を深めることは，個別化された患者支援において重要な役割を果たすことが期待される．

（2）がん患者の有害事象訴え抽出手法の開発

① 手足症候群シグナル抽出器の構築[7]

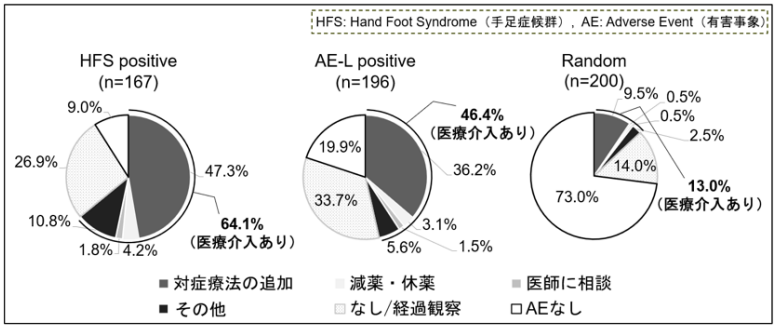
207 名分のがん患者による 10,646 記事から，手足部位情報を含む 5,492 文をルールベースで抽出後，アノテーションと精査により 149 文の手足症候群 positive を特定した．この中には，「足に痛み」「スプーン爪」といった疾患特有の表現の他，「べろべろに剥けている」「足のブヨブヨ」といった患者独自のオノマトペ表現も認めた．過学習を避ける目的でアンダーサンプリングを行った上で 3 つの深層学習モデルに訓練を行った．性能評価の結果，手足症候群 positive を文またはユーザー単位で検出するタスクにおいて，BERT が最も高い F1 スコアを示した（0.54 及び 0.72）．なお，偽陽性の大半は，手足症候群以外の皮膚関連有害事象の記述だった．以上から，重症化予防の必要性が高い手足症候群のシグナルを検出する抽出器を構築できた．患者テキストから手足症候群という狭義の副作用シグナルの検出に成功した例は，本研究が初めてである．

② 日常生活支障ありの有害事象シグナル抽出器の構築[8]

乳がん患者のブログ 2,272 記事から，AE-L positive を 191 記事/428 文，AE-nL positive を 702 記事/2,415 文としてアノテーションした．本データを用いて 3 つの NLP モデルを訓練し，複数のアプローチを比較した．その結果，記事単位で AE-L を抽出する 1 step アプローチにおいて，T5 が最も高い F1 スコアを示した（0.56）．なお，偽陽性の大半は，AE-nL 関連の記述（有害事象ではあるものの生活支障までは来たさない，もしくは文章内から生活支障を明確に読み取れない）であった．また，同じ 1 step アプローチで All AE を抽出対象とした場合，T5 は 0.81 と最も高い F1 スコアを達成した．AE-L として記述された有害事象シグナルの中で最も多かったのは“痛み・痺れ（56%）”，“倦怠感（30%）”，“発熱（24%）”，“吐き気（22%）”であった（頻度 20%以上のみ記載）．以上から，日常生活に支障を来たす有害事象シグナルの抽出器の構築に成功した．患者視点で生活への支障度を指標に重症度の高い事象を捉えるこのアプローチは，意義ある有害事象シグナルの抽出手法である．

（3）構築済み抽出器の有用性評価[9]

構築した 2 種の有害事象シグナル抽出器（手足症候群と AE-L）の性能を，異なるテキストソースで検証した．乳がん患者の語り 508 件の記事から，手足症候群抽出器は 1 件，AE-L 抽出器は 18 件を positive と判断し，これらはすべて実際に有害事象シグナルを含む記事であった．特に“痛み・痺れ”が最も頻繁に語られる症状であり（33%），“発熱”と“吐き気”が続いた（ともに 22%）．抗がん剤処方歴を持つ患者の薬歴 S 記録 30,784 件に抽出器を適用した結果，手足症候群で 167 件，AE-L で 196 件が positive として抽出され，そのうち 91%と 80%が実際に有害事象を含む S 記録であった．偽陽性の例は，いずれも薬剤師による症状の否定や軽快状態の確認記録であった．抽出器によって特定された有害事象シグナルに対する医療介入状況を分析した結果，医療介入が行われた S 記録の割合は，比較対象としたランダム抽出の 13%から，手足症候群で 64%，AE-L で 46%に大幅に上昇した（**下図**）．手足症候群シグナルが抽出された患者の 7 割以上に, 手足症候群が好発する抗がん剤が処方されていた. AE-L シグナルの症状としては，“痛み・痺れ”が最も多く（36%），“発熱（29%）”，“吐き気（26%）”が続いた．そのうち，“発熱”と“吐き気”では医療介入を伴うシグナルの割合が高かったのに対し，“痛み・痺れ”では医療介入の割合が低くなる傾向を認めた．これらは，本研究で構築した有害事象シグナル抽出器が，患者訴えに基づく抗がん剤副作用シグナルの自動抽出手法として，早期介入/重症化予防すべき医療上重要な事象を拾い上げることができることを支持する結果である．

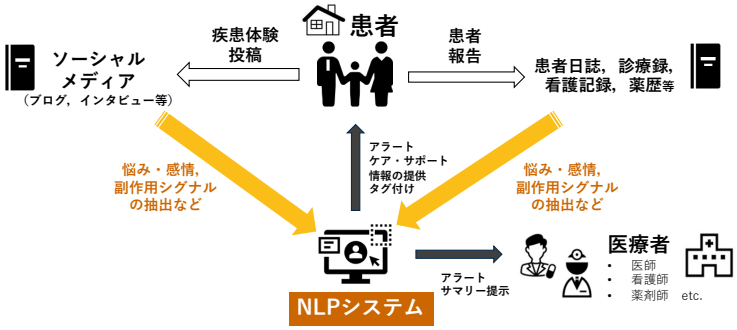


（4）本研究成果の意義と今後の展望

本研究では，ソーシャルメディア上のがん患者の記述や語りから，NLP 技術を用いて疾患や治療に関連する自覚症状や悩みのエピソードを抽出することに成功した．特に，“身体の苦痛”や“就労・経済的負担”といったがん患者特有の具体的な悩みに焦点をあて，高性能のマルチラベル分類器を開発した．このマルチラベル分類器は，複合的に発生する悩みを捉えることが可能

である点を特徴とする．さらに，本研究では手足症候群や日常生活に支障を及ぼす有害事象（AE-L）シグナルを特定する抽出器も構築し，異なる患者テキストに適用することでその汎用性と有用性を示した．

患者視点で悩みや有害事象を捉えるこれらの抽出器は，その新規性と意義が大きく，医療介入が必要な事象を早期に捉えることで，患者の治療生活の質に大きく向上させることが期待される．また，この技術を応用することにより，病気治療中に患者の記述から治療に必要な情報や感情を可視化し，医療者とのコミュニケーションを促進することが期待される．例えば，患者が自身の疾患体験を記録・発信するプラットフォームにこの技術を組み込むことで，患者への適切なケア情報の提供やソーシャルサポートの提案，さらには有害事象の重症度に応じた医療機関への受診勧奨が可能になると期待される．また，経験豊富な患者の蓄積された記述から新たに罹患した患者への実践的な知識の継承（患者間での共有・学び）への活用も期待される（**下図**：今後の展望）．



本研究で構築した悩みや有害事象シグナル抽出手法は，実社会で運用するに足る性能を示した一方，本格的に利活用するためには，モデル性能のさらなる向上や偽陽性/偽陰性シグナルへの対処等，さまざまな課題を解決していく必要がある．これらの課題に対処し，患者や医療者が受け入れ可能かつ付加価値の高い形での利活用を目指すためには，前向き研究を通じて本手法の有用性と求められる性能を慎重に検討していく必要があるだろう．

＜引用文献＞

1) Watanabe T, Yada S, **Aramaki E**, Yajima H, Kizaki H, **Hori S**. Extracting Multiple Worries From Breast Cancer Patient Blogs Using Multilabel Classification With the Natural Language Processing Model Bidirectional Encoder Representations From Transformers: Infodemiology Study of Blogs. ***JMIR Cancer***. 2022 Jun 3;8(2):e37840. doi: 10.2196/37840.
2) 淺野真輝，渡邊知生，矢田竣太郎，**荒牧英治**，矢島弘士，西岡諭史，木﨑速人，**堀里子**．自然言語処理を用いた乳がん患者ブログ記述内容の特徴分析．***日本医療薬学会第 5 回フレッシャーズ・カンファランス*** 2022 年 6 月．
3) Watabe S, Watanabe T, Yada S, **Aramaki E**, Yajima H, Kizaki H, **Hori S**. Exploring a method for extracting concerns of multiple breast cancer patients in the domain of patient narratives using BERT and its optimization by domain adaptation using masked language modeling. ***PLoS One. in press***
4) 久村颯海、佐山杏子、渡部哲、木﨑速人、今井俊吾、岸田徹、福岡奈津美、矢田竣太郎、**荒牧英治**、**堀里子**．がん経験者インタビューテキストを用いたがん患者の感情分析．***2024 年度人工知能学会全国大会（第 38 回）*** 2024 年 5 月．
5) Plutchik R. The Nature of Emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. ***American Scientist***, 2001, 89(4), 344–350.
6) 梶原智之, Chenhui Chu, 武村紀子, 中島悠太, 長原一.主観感情と客観感情の強度推定のための日本語データセット.***言語処理学会第 27 年次大会***, pp. 523-527, 2021.
7) Nishioka S, Watanabe T, Asano M, Yamamoto T, Kawakami K, Yada S, **Aramaki E**, Yajima H, Kizaki H, **Hori S**. Identification of hand-foot syndrome from cancer patients' blog posts: BERT-based deep-learning approach to detect potential adverse drug reaction symptoms. ***PLoS One***. 2022 May 4;17(5):e0267901. doi: 10.1371/journal.pone.0267901.
8) Nishioka S, Asano M, Yada S, **Aramaki E**, Yajima H, Yanagisawa Y, Sayama K, Kizaki H, **Hori S**. Adverse event signal extraction from cancer patients' narratives focusing on impact on their daily-life activities. ***Sci Rep.*** 2023 Sep 19;13(1):15516. doi: 10.1038/s41598-023-42496-1.
9) Nishioka S, Watabe S, Yanagisawa Y, Sayama K, Kizaki H, Imai S, Someya M, Taniguchi R, Yada S, **Aramaki E**, **Hori S**. Adverse Event Signal Detection Using Patients' Concerns in Pharmaceutical Care Records: Evaluation of Deep Learning Models. ***J Med Internet Res.*** 2024 Apr 16;26:e55794. doi: 10.2196/55794.

５．主な発表論文等

〔雑誌論文〕　計5件（うち査読付論文　4件／うち国際共著　0件／うちオープンアクセス　4件）

１．著者名 Nishioka Satoshi、Watanabe Tomomi、Asano Masaki、Yamamoto Tatsunori、Kawakami Kazuyoshi、Yada Shuntaro、Aramaki Eiji、Yajima Hiroshi、Kizaki Hayato、Hori Satoko	４．巻 17
２．論文標題 Identification of hand-foot syndrome from cancer patients' blog posts: BERT-based deep-learning approach to detect potential adverse drug reaction symptoms	５．発行年 2022年
３．雑誌名 PLOS ONE	６．最初と最後の頁 e0267901
掲載論文のDOI（デジタルオブジェクト識別子） 10.1371/journal.pone.0267901	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 Watanabe Tomomi、Yada Shuntaro、Aramaki Eiji、Yajima Hiroshi、Kizaki Hayato、Hori Satoko	４．巻 8
２．論文標題 Extracting Multiple Worries From Breast Cancer Patient Blogs Using Multilabel Classification With the Natural Language Processing Model Bidirectional Encoder Representations From Transformers: Infodemiology Study of Blogs	５．発行年 2022年
３．雑誌名 JMIR Cancer	６．最初と最後の頁 e37840
掲載論文のDOI（デジタルオブジェクト識別子） 10.2196/37840	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 Nishiyama Tomohiro、Yada Shuntaro、Wakamiya Shoko、Hori Satoko、Aramaki Eiji	４．巻 25
２．論文標題 Transferability Based on Drug Structure Similarity in the Automatic Classification of Noncompliant Drug Use on Social Media: Natural Language Processing Approach	５．発行年 2023年
３．雑誌名 Journal of Medical Internet Research	６．最初と最後の頁 e44870
掲載論文のDOI（デジタルオブジェクト識別子） 10.2196/44870	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 Nishioka Satoshi、Asano Masaki、Yada Shuntaro、Aramaki Eiji、Yajima Hiroshi、Yanagisawa Yuki、Sayama Kyoko、Kizaki Hayato、Hori Satoko	４．巻 13
２．論文標題 Adverse event signal extraction from cancer patients' narratives focusing on impact on their daily-life activities	５．発行年 2023年
３．雑誌名 Scientific Reports	６．最初と最後の頁 15516
掲載論文のDOI（デジタルオブジェクト識別子） 10.1038/s41598-023-42496-1	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

1．著者名 Nishioka Satoshi、Watabe Satoshi、Yanagisawa Yuki、Sayama Kyoko、Kizaki Hayato、Imai Shungo、Someya Mitsuhiro、Taniguchi Ryoo、Yada Shuntaro、Aramaki Eiji、Hori Satoko	4．巻 26
2．論文標題 Adverse Event Signal Detection Using Patients' Concerns in Pharmaceutical Care Records: Evaluation of Deep Learning Models	5．発行年 2024年
3．雑誌名 Journal of Medical Internet Research	6．最初と最後の頁 e55794
掲載論文のDOI（デジタルオブジェクト識別子） 10.2196/55794	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

〔学会発表〕　計13件（うち招待講演　0件／うち国際学会　1件）

1．発表者名 淺野真輝，渡邊知生，矢田竣太郎，荒牧英治，矢島弘士，西岡諭史，木﨑速人，堀里子
2．発表標題 自然言語処理を用いた乳がん患者ブログ記述内容の特徴分析
3．学会等名 日本医療薬学会第5回フレッシャーズ・カンファランス
4．発表年 2022年

1．発表者名 岡澤優太，木﨑速人，鈴木信行，堀里子
2．発表標題 インタビュー調査を用いた患者コミュニティへの参加が健康意識に及ぼす影響の探索
3．学会等名 日本医療薬学会第5回フレッシャーズ・カンファランス
4．発表年 2022年

1．発表者名 渡部哲，渡邊知生，矢田竣太郎，荒牧英治，木﨑速人，堀里子
2．発表標題 自然言語処理モデルBERTを用いた乳がん患者の悩みマルチラベル分類器の患者語りテキストへの応用
3．学会等名 医療薬学フォーラム2022/第30回クリニカルファーマシーシンポジウム
4．発表年 2022年

１．発表者名
西山智弘，矢田竣太郎，若宮翔子，堀里子，荒牧英治

２．発表標題
ソーシャルメディアを用いた一般用医薬品のファーマコビジランス

３．学会等名
言語処理学会第29回年次大会

４．発表年
2023年

１．発表者名
沖本 綾音，木﨑 速人，飯野 温，卯田 健太，今井 俊吾，田中 聡一，森 貴幸，堀 里子

２．発表標題
薬局での症状モニタリングにおける患者症状聞き取りアプリケーションの有用性評価

３．学会等名
日本薬学会第143年会

４．発表年
2023年

１．発表者名
渡邉知生，矢田竣太郎，荒牧英治，矢島弘士，木崎速人，堀里子

２．発表標題
自然言語処理モデル BERT を用いた乳がん患者ブログからの悩み抽出手法の検討

３．学会等名
第23回日本医薬品情報学会学術大会

４．発表年
2021年

１．発表者名
渡邊聖也，木﨑速人，堀里子

２．発表標題
患者の自覚症状聞き取りアプリケーションの開発と評価

３．学会等名
日本医療薬学会第４回フレッシャーズ・カンファランス

４．発表年
2021年

１．発表者名
西岡諭史，渡邉知生，淺野真輝，川上和宜，矢田竣太郎，荒牧英治，矢島弘士，木崎速人，堀里子

２．発表標題
自然言語処理を用いた患者ブログからの手足症候群発症ユーザー抽出手法の構築.

３．学会等名
第41回 医療情報学連合大会（第22回日本医療情報学会学術大会）

４．発表年
2021年

１．発表者名
西山智弘，矢田竣太郎，若宮翔子，堀里子，荒牧英治

２．発表標題
ソーシャルメディアにおける服薬ノンコンプライアンス発言の分類と化学構造類似度を用いたコーパス可搬性の検討

３．学会等名
言語処理学会第28回年次大会

４．発表年
2022年

１．発表者名
佐藤 隆哉，木﨑 速人，西山 智弘，矢田 竣太郎，荒牧 英治，今井 俊吾，堀 里子

２．発表標題
ソーシャルメディアへの医薬品のオーバードーズに関連した投稿の解析

３．学会等名
日本薬学会第144年会

４．発表年
2024年

１．発表者名
渡部哲，西岡愉史，柳澤友希，佐山杏子，木﨑速人，今井俊吾，染谷光洋，谷口亮央，矢田 竣太郎，荒牧英治，堀里子

２．発表標題
深層学習を用いた患者主訴からの有害事象シグナル検出器の開発およびその有用性評価

３．学会等名
2024年度 人工知能学会全国大会（第38回）

４．発表年
2024年

１．発表者名
久村颯海，佐山杏子，渡部哲，木﨑速人，今井俊吾，岸田徹，福岡奈津美，矢田竣太郎，荒牧英治，堀里子

２．発表標題
がん経験者インタビューテキストを用いたがん患者感情分類器の構築

３．学会等名
2024年度 人工知能学会全国大会（第38回）

４．発表年
2024年

１．発表者名
Nishioka S, Asano M, Yada S, Aramaki E, Yajima H, Kizaki H, Hori S

２．発表標題
Detection of adverse event signals with severity grade classification from cancer patient narrative.

３．学会等名
MEDINFO2023（国際学会）

４．発表年
2023年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

ウェブサイト（慶應義塾大学薬学部医薬品情報学講座）：https://keio-di.jp/project/post-406/
プレスリリース（慶應義塾大学，奈良先端科学技術大学院大学，2023.10.12）：https://www.keio.ac.jp/ja/press-releases/2023/10/12/28-153137/

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
研究分担者	荒牧　英治 (Aramaki Eiji) (70401073)	奈良先端科学技術大学院大学・先端科学技術研究科・教授 (14603)	

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関