

一般口演 | 知識工学

一般口演9

自然言語処理・テキストマイニング

2021年11月20日(土) 09:10 ~ 11:10 E会場 (2号館2階222+223)

[3-E-1-02] 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

*西岡 諭史¹、渡邊 知生¹、浅野 真輝¹、川上 和宜²、矢田 竣太郎³、荒牧 英治³、矢島 弘士⁴、木崎 速人¹、堀 里子¹
(1. 慶應義塾大学薬学部, 2. がん研有明病院, 3. 奈良先端科学技術大学院大学, 4. (株)メディエイド)

*Satoshi Nishioka¹, Tomomi Watanabe¹, Masaki Asano¹, Kazuyoshi Kawakami², Shuntaro Yada³, Eiji Aramaki³, Hiroshi Yajima⁴, Hayato Kizaki¹, Satoko Hori¹ (1. 慶應義塾大学薬学部, 2. がん研有明病院, 3. 奈良先端科学技術大学院大学, 4. (株)メディエイド)

キーワード : deep-learning, BERT, hand-foot syndrome, adverse drug reaction, internet patient community

【目的】手足症候群は抗がん薬の中止につながりうる副作用であり、自覚症状が主な発症シグナルとなる。患者は生活にきた支障をブログとして記録・発信することがあり、そこには診療時には表面化しない副作用のシグナルが含まれる可能性がある。本研究では、患者ブログを用いて、手足症候群疑いのあるユーザーを抽出する手法を構築することを目的とした。

【材料・手法】患者闘病記サイト Life Palette におけるがん患者のブログ10,646記事を対象とした。前処理を行い、手足部位情報を含む文を取得した後、研究者3名（内1名のがん専門薬剤師含む）によりアノテーションを行い、手足症候群の教師データ（Positive文）を取得した。データは訓練用：テスト用=4：1に分け、訓練用データを用いて3つの機械学習モデル（LSTM、双方向 LSTM、BERT）で学習を行い、手足症候群ユーザーの抽出器を構築した。その後、テスト用データで抽出器の性能を評価した。

【結果】手足部位情報を含む文を5,492件、さらにアノテーションにより手足症候群の Positive文を149件取得した。訓練用データによる学習後、テスト用データで抽出器の性能を評価したところ、BERTにおいて手足症候群ユーザーの抽出タスクの Precision、Recall、F値は、それぞれ0.63、0.82、0.66であった。

【考察】手足症候群 Positiveのユーザーは、関連文を複数件投稿する傾向があるため、いずれかの Positive文を拾うことにより、ユーザーの特定・抽出が可能になったと考える。今後、患者ブログで手足症候群疑いと判定されたユーザーに“気づき”を与えられれば、受診勧奨やセルフケア情報の提供を通して重症化予防に寄与できる可能性がある。

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

西岡諭史^{*1}、渡邊知生^{*1}、浅野真輝^{*1}、川上和宜^{*2}、矢田竣太郎^{*3}、
荒牧英治^{*3}、矢島弘士^{*4}、木崎速人^{*1}、堀里子^{*1}

*1 慶應義塾大学薬学部、*2 がん研有明病院、
*3 奈良先端科学技術大学院大学、*4 (株)メディエイド

Method to Extract Users with Hand-Foot Syndrome from Patient Blog through Natural Language Processing

Satoshi Nishioka^{*1}, Tomomi Watanabe^{*1}, Masaki Asano^{*1}, Kazuyoshi Kawakami^{*2}, Shuntaro Yada^{*3},
Eiji Aramaki^{*3}, Hiroshi Yajima^{*4}, Hayato Kizaki^{*1}, Satoko Hori^{*1}

*1 Keio University Faculty of Pharmacy, *2 Cancer Institute Hospital, Japanese Foundation for Cancer Research,
*3 Nara Institute of Science and Technology, *4 Mediad Corporation

Objective: Hand-foot syndrome (HFS) is one of the most problematic adverse drug reactions (ADRs) that can lead to discontinuation of anticancer drugs. Recently, an increasing number of patients post their daily experiences to internet community, where potential ADR signals not captured through routine clinic visits can be described. Therefore, this study aimed to identify patients with potential ADRs, focusing on HFS, from internet blogs by using natural language processing deep-learning methods

Method and materials: 10,646 blogs posted by cancer patients on Life Palette, an internet patient community in Japan, were utilized. After pre-processing, three researchers (including one certified oncology pharmacist) conducted annotation, obtaining HFS positive sentences. The dataset was divided into a 4 to 1 ratio to train and evaluate three deep-learning models: long short-term memory (LSTM), bidirectional LSTM, and bidirectional encoder representations from transformers (BERT).

Results: Out of 5,492 sentences with hand-foot-relevant terms, 149 HFS positive sentences were obtained. The BERT model gave the best performance with precision 0.63, recall 0.82 and $f_{0.5}$ score 0.66 in HFS identification task.

Discussion: Our results demonstrate that this deep-learning model can successfully identify HFS patients from patient blog. It should be feasible to utilize patient-generated text data to improve ADR management for individual patients.

Keywords: deep-learning, BERT, hand-foot syndrome, adverse drug reaction, internet patient community

1 緒論

がん罹患率は世界的に右肩上がりであり、その臨床的及び経済的負担は大きい¹⁾。がん治療で期待する効果を得るために、抗がん剤の副作用をマネジメントすることはコンプライアンス維持の観点から重要である²⁾。そのため、抗がん剤副作用を早期に検知し、重症化を予防する取り組みは必要不可欠となる^{3,4)}。

手足症候群は根治療法のない典型的な抗がん剤副作用の一つであり、早期からの症状ケアが治療の核となる^{3,5)}。手足症候群は患者 QOL への影響度が大きく、抗がん剤の中止・中断の要因となり得ることから、問題視されている^{5,6)}。その発現シグナルは患者の自覚症状に依存しているが、日常診療下では制限ある診療時間及び患者自身が抗がん剤治療を優先する心情の影響から、患者が発現シグナルを医療者へ正確に報告していない事例も認められる⁷⁾。

日常診療にて医療者に語られない副作用の訴えは、時にインターネットコミュニティに投稿されることがある⁸⁾。そこで筆者らは、インターネット上の患者コミュニティのテキストデータが、副作用発現シグナルの検知に活用できるのではないかと考えた。実際、インターネットコミュニティが患者の安全性情報の早期収集に役立つとする研究報告は存在する⁹⁻¹²⁾。患者訴えをインターネットコミュニティから検知することは、副作用をより積極的にマネジメントする上で新たなアプローチにな

り得る。

インターネットコミュニティから必要な情報を効率よく収集する上で、自然言語処理技術は有用なツールとなる。この技術は、電子カルテからの情報抽出などで多用されてきたが¹³⁾⁻¹⁵⁾、近年では市販後調査などにおいて患者自身が記述するテキストへの応用も進んでいる¹⁶⁾。しかし、最新の深層学習手法を患者記述テキストに応用した例は未だ少ない。さらに、副作用情報を原因薬剤の情報無しに、患者表現のみから抽出しようとする試みはこれまでに報告がない。

2 目的

手足症候群に焦点を絞り、最新の深層学習手法が抗がん剤副作用疑い事象を発現した患者を、患者ブログ内の情報から特定できるか否かを確かめることを目的とした。

3 方法

3.1 全体像

この研究は2つのパートから構成される。前半は、ブログ記事を前処理し、手足症候群の訴えと思われる表現・文を抽出するパートである(図 1)。後半は、特定した手足症候群の文を基に深層学習を行い、モデルの構築と性能評価を行うパートである(図 2)。

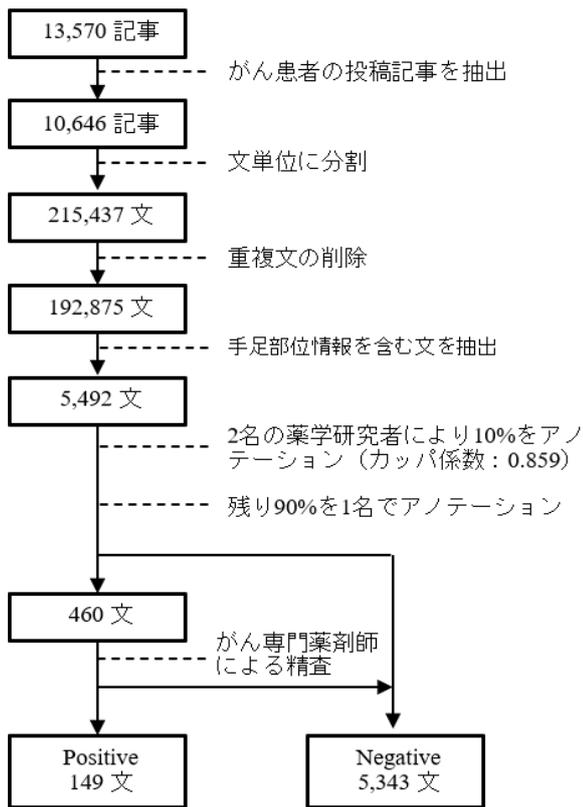


図1 前処理・アノテーションのフロー

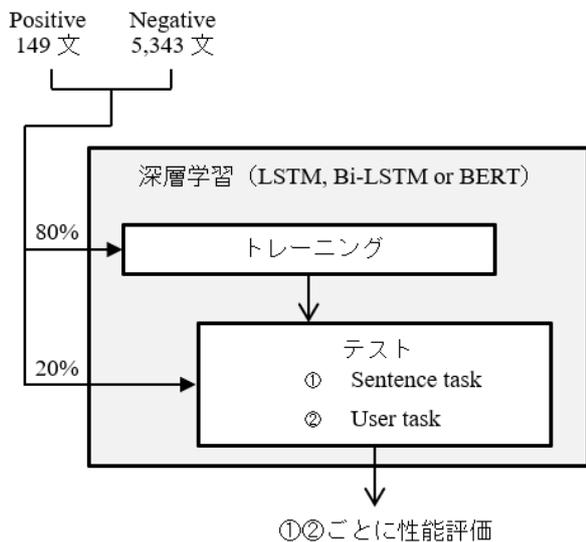


図2 深層学習と2種類の予測タスク

3.2 データ

3.2.1 前処理

患者コミュニティ Life Palette¹⁷⁾に投稿された日本語のブログ記事を対象とした。データソースは、2008年3月から2014年11月までに投稿された13,570記事(289ユーザー)から構成される。

最初に、がん患者によって書かれたブログ記事 10,646 記事を抽出した。その後、GitHub に公開されている“ja_sentence_segmenter”¹⁸⁾を用いて記事を文単位に分割し、215,437 文を得た。このうち、重複する文を削除した結果、独立した 192,875 文を取得した。さらに、手足症候群の記述が含まれると予想される、手足部位情報(「手」「足」「指」「爪」など)を少なくとも一単語含む文を抽出した(図1)。

3.2.2 手足症候群 positive/negative 文の取得

手足症候群の positive 表現・文を特定するにあたり、複数のアノテーターが共通の基準を持てるよう、文献^{3),5)}を参照してアノテーションガイドラインを作成した(表1)。アノテーションガイドラインに従い、2名の薬学研究者が独立して、手足部位情報を含む文のうち10%をアノテーションした。カッパ係数^{19),20)}を用いて高い一致率を確認した後、残り90%を1名の薬学研究者がアノテーションした(図1)。

続いて、がん専門病院に勤務する日本医療薬学会認定がん専門薬剤師1名が、薬学研究者によりアノテーションされた手足症候群疑い文を精査した。その結果で手足症候群の記述と判断された文を手足症候群 positive 文、その他全ての手足情報部位を含む文を手足症候群 negative 文とした(図1)。

表1 アノテーションガイドライン

定義	抗がん剤が誘引したと思われる手足症候群様の症状
Positive 基準	手足部位に関わる下記表現を含む文 <ul style="list-style-type: none"> ・ 痛み ・ 痺れ ・ 腫れ ・ 手が赤い ・ 湿疹 ・ 違和感 ・ 感覚異常 ・ 爪の変形 ・ その他、患者特有表現
除外基準	<ul style="list-style-type: none"> ・ 現在進行形ではない症状の記述 (例：既に回復済み、可能性の話、症状の説明、記事の引用、想像) ・ 不定愁訴様な表現 (例：だるい、しっくりこない) ・ 薬剤以外の原因が明らか (例：歩きすぎで足がしびれた、採血で手が痛む) ・ 「腫れ」のうち、明らかに水分貯留が要因のもの

3.3 モデル

3.3.1 深層学習モデルの構築

前半パートで得られた手足症候群 positive または negative 文は、深層学習のトレーニングとテスト用に4対1の割合で分割した(図2)。

本研究に用いた自然言語処理モデルは、深層学習手法に分類される long short-term memory (LSTM)、bidirectional LSTM (Bi-LSTM)、bidirectional encoder representations from transformers (BERT) の3つである(図2)。効率的に日本語表現を学習できるよう、トレーニング済みデータセットである“fastText”²¹⁾を LSTM 及び Bi-LSTM に、東北大学公開日本

語モデル²²⁾を BERT に適用した。トークナイザーは、Janome²³⁾を LSTM 及び Bi-LSTM に、McCab²⁴⁾を BERT に使用した。学習手法は、5 分割交差検証法を採用した。

3.3.2 深層学習モデルの評価

二種類の予測タスクを設定した。一つは、それぞれの文が手足症候群 positive か negative かを予測する「sentence task」であり、もう一つはユーザー単位で手足症候群 positive か negative かを予測する「user task」である(図 2)。後者の user task は、本研究の目的である手足症候群発症患者の特定に重点を置いたタスクである。各予測タスクにおける構築モデルの性能は、以下に示す適合率 (precision)、再現率 (recall)、F₁ 及び F_{0.5} score で評価した。F₁ score は適合率と再現率を均等に調平均した指標であり、一方の F_{0.5} score は適合率重視で調平均した指標である。

$$Precision = \frac{\# \text{ of True Positive}}{\# \text{ of True Positive} + \# \text{ of False Positive}}$$

$$Recall = \frac{\# \text{ of True Positive}}{\# \text{ of True Positive} + \# \text{ of False Negative}}$$

$$F_1 \text{ score} = \frac{2 * precision * recall}{precision + recall}$$

$$F_{0.5} \text{ score} = \frac{1.25 * precision * recall}{0.25 * precision + recall}$$

本研究における主要評価項目は、F_{0.5} score と設定した。これは、患者自身に副作用疑い事象の検知を知らせる実運用を見越した際に、不要な心理的負荷を避けるために false positive を少なく出力できる適合率重視のモデルが望ましいと考えたためである。

3.4 倫理的配慮

Life Palette の利用規約に従い、「運営会社が第三者に記事を委託することを認める」「第三者によるデータの二次利用を制限する」旨を、サービス利用時点で投稿ユーザーから同意を得ている。これに基づき、当研究室は Life Palette を運営するメディエイド社と共同研究契約を締結し、慶應義塾大学薬学部倫理委員会から承認を得た上で、研究を実施した。

4 結果

4.1 データセット

前処理により、手足部位情報を少なくとも一つ含む 5,492 文を抽出した(図 1)。一文に含まれる文字数の平均及び中央値(最小-最大)は、68.80 及び 44 (3-1328) 語であった。

アノテーションガイドライン(表 1)に従い、2 名の薬学研究者が 5,492 文の 10%をアノテーションしたところ、高い一致率を確認できた(カッパ係数:0.859)。残り 90%を 1 名がアノテーションし、手足症候群疑い 460 文を取得した(図 1)。

続いて、がん専門薬剤師が自身の臨床経験を基に 460 文をレビューし、手足症候群として確度の高い表現のみを選択した。結果、手足症候群 positive として 149 文を取得した(図 1)。これは、アノテーション前の集団(手足部位情報ありの 5,492 文)に対して、約 2.7%を占める割合である。

4.2 手足症候群 positive 文及びユーザー

手足症候群 positive 文の例を表 2 に示す。「足に痛み」「スプーン爪」といった手足症候群に典型的な表現に加え、「手の皮がべろべろに剥けている。」「足のブヨブヨ」のように、患者特有のオノマトペ表現が含まれていることを確認した。手足

症候群 positive 文の文字数の平均及び中央値(最小-最大)は、52.06 及び 37(11 - 235) 語であった。また、ユーザー 1 名あたりの手足症候群 positive 文の投稿数の平均及び中央値(最小-最大)は、3.55 及び 2(1-24) 文であった。

表 2 手足症候群 positive 文 (例示)

ところが、4 月後半になると階段を上るときに足に痛みを感じるようになった。
どうも僕の指の爪は、スプーン爪と呼ばれる状態になっているらしい。
ちょっと爪先がどこかに当たっただけではがれて 2 枚爪になったりするし、ポロポロと欠けることも多い。
朝から、手のひら、足裏が真っ赤。
手の皮がべろべろに剥けている。
足のブヨブヨは、すっかり固くなってる感じ。

4.3 モデル構築と性能評価

深層学習モデルのトレーニング用に、3 つのデータセットを用意した - 1.オリジナルの集団と同様に約 2.7%の割合で手足症候群 positive を含むデータセット(“Original”とする)、2.手足症候群 negative 文をアンダーサンプリングし、positive: negative = 1:1 の均等割合としたデータセット(“Balanced”とする)、3.同じくアンダーサンプリングにより、positive: negative = 1:20 の割合としたデータセット(“Under-sampling”とする)。アンダーサンプリングとは、多数派データを過学習してしまうリスクのある非均衡データに対する学習テクニックの一つである。

テスト用データを用いて評価した性能値を表 3 に示す。主要評価項目に設定した F_{0.5} score において、sentence task での最高値は“Under-sampling”でトレーニングした BERT の 0.48 であり、user task での最高値は同じく“Under-sampling”でトレーニングした BERT の 0.66 であった。“Balanced”によるトレーニングは再現率を大幅に改善した一方で、適合率には負の影響があり、結果的に F score における改善には繋がらなかった。

表 3 各学習モデルの性能値

a. LSTM	適合率	再現率	F ₁	F _{0.5}
Sentence task				
Original	0.28	0.20	0.23	0.26
Balanced	0.10	0.96	0.19	0.12
Under-sampling	0.41	0.33	0.37	0.39
User task				
Original	0.66	0.52	0.58	0.63
Balanced	0.23	1.00	0.37	0.27
Under-sampling	0.57	0.57	0.57	0.57
b. Bi-LSTM	適合率	再現率	F ₁	F _{0.5}
Sentence task				
Original	0.35	0.33	0.34	0.35
Balanced	0.15	0.86	0.26	0.18
Under-sampling	0.33	0.46	0.38	0.35

User task				
Original	0.65	0.68	0.66	0.65
Balanced	0.30	1.00	0.46	0.35
Under-sampling	0.50	0.68	0.57	0.52

c. BERT

	適合率	再現率	F ₁	F _{0.5}
Sentence task				
Original	0.43	0.23	0.30	0.37
Balanced	0.03	0.56	0.07	0.04
Under-sampling	0.45	0.66	0.54	0.48
User task				
Original	0.53	0.36	0.43	0.49
Balanced	0.13	0.93	0.23	0.16
Under-sampling	0.63	0.82	0.71	0.66

5 考察

本試験の結果は、最新の自然言語処理技術がブログ記事からの手足症候群患者の特定を可能とすることを示している。今回用いた深層学習手法が、患者特有のオノマトペ表現を含めて正しく学習できていたことは注目すべきである。さらに、原因薬剤の情報無しに、患者記述の文章のみから副作用シグナルを抽出した本手法は、応用性の高いアプローチであると考えられる。

本試験では、LSTM、Bi-LSTM 及び BERT という 3 つの深層学習モデルを手足症候群の表現抽出に用いた。LSTM 及び Bi-LSTM は 1997 年以降に開発が発展してきたリカレントニューラルネットワークの 1 種であり²⁵⁾⁻²⁷⁾、BERT は 2018 年に Google が発表した最新の深層学習手法である²⁸⁾。LSTM と Bi-LSTM の特徴を比較したとき、前者は前から後ろ方向の文脈しか記憶できない一方、後者は両方向に記憶を行い、予測結果に反映できるという違いがある。患者記述の文章は構文が必ずしも正確ではないため、双方向記憶に強みを Bi-LSTM で性能向上が認められると期待したが、実験結果は必ずしもそうとは言えない結果であった。トレーニングデータセットの構成とモデルの組み合わせを網羅的に探索することで、より良い性能を示せる可能性はある。LSTM ベースのモデルと BERT を比較した場合、再現率では LSTM ベースのモデルで優位性が認められる一方、適合率及び F score は BERT で高い傾向を認めた。実運用で重視すべき性能次第では、どちらのモデルにも活かし方があると想像するが、少なくとも筆者が想定する実運用(患者自身への副作用シグナルの検知アラート)を念頭に置くと、BERT の方がバランスの取れた抽出器であると考えられる。

過去、ヘルスケア関連ソーシャルメディアまたは Twitter から副作用情報を抽出した研究では、F₁ score は概ね 0.5 から 0.7 程度と報告されている²⁹⁾⁻³¹⁾。別の報告では、ソーシャルネットワークから皮膚関連副作用情報を検知するシステムにおいて、micro average F₁ score は 0.74 であった⁸⁾。筆者らの構築モデルでは、より狭義の副作用である手足症候群に絞った条件において、best F₁ score が BERT の 0.71 であった。各種コンディションの調整により、更に性能向上できる可能性は残しつつも、患者ブログから自動的に手足症候群患者を特定する上で、BERT は相性の良い深層学習手法であると考えられる。

学習モデルの信頼性評価のため、false positive 分析を行った。その結果、モデルが誤って positive と判定した文であっても、その中には true positive に類似した表現が含まれていたことが確認された(例:「指の先端に軽い痺れ」「足・・・右下腿にポツポツできてます」)。これら true positive に類似した false positive 文は、2 名の薬学研究者がアノテーションを行った後の手足症候群疑い 460 文には含まれる文であった。このことから、今回構築した手足症候群の抽出器は、より広義に皮膚関連有害事象を拾う上で、性能値が示す以上に有用となり得る可能性がある。

本試験で構築した学習器には、大きく 3 つの発展可能性がある。1 つ目は、トレーニングデータを増やすことで手足症候群の重症度評価を可能とすることである。トレーニングデータの小ささは本試験の Limitation であったが、それを増やすことにより、基礎性能の向上のみならず、重症度別の患者表現の違いを学習することもできるのではと期待する。2 つ目に、手足症候群のみならず、より広義の副作用表現の学習である。本試験の結果から、最新の深層学習手法を用いることで患者特有の副作用表現も抽出できたことから、他の副作用表現への展開も可能と考える。最後は、患者ブログ以外のソースデータへの適用である。今回構築した抽出器は 1 文表現から手足症候群患者を特定する仕様であるため、例えば Twitter のような短文の患者記述テキストにも適用可能であるため、運用イメージを広げて他のソースデータに適用を拡大することも一つの方向性と考えられる。

今回構築した抽出器をインターネット上の患者コミュニティに実装できれば、副作用事象に関して記述する患者にアラートを出すことにより、早期の受診を促すことができる。同時に、信頼性の高いセルフマネジメント方法を患者に紹介することでも、重症化予防に貢献できると考える。さらには、このような患者自身にメリットが生じる機能をインターネットコミュニティに実装することで、患者がより能動的に疾患体験を記録する効果も期待される。実運用において、今回構築した抽出器の性能が不十分であった場合は、手足症候群と他の疾患を鑑別するための追加質問を行うことで(例:症状発現部位の細部の確認)、モデルの性能を補完できると考える。

6 結論

自然言語処理の深層学習手法は、インターネット上の患者ブログから手足症候群発症ユーザーの特定を可能とする。本手法の実装は、個々の患者の副作用マネジメントの改善につながる可能性がある。

参考文献

- 1) Mattiuzzi C, Lippi G. Current Cancer Epidemiology glossary. *J Epidemiol Glob Health*. 2019;9(4):217–22.
- 2) Lin C, Clark R, Tu P, Bosworth HB, Zullig LL. Breast cancer oral anti-cancer medication adherence: a systematic review of psychosocial motivators and barriers. *Breast Cancer Res Treat*. 2017;165(2):247–60.
- 3) 厚生労働省医薬・生活衛生局医薬安全対策課. 重篤副作用疾患別対応マニュアル. Available from: https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryuu/yakuhin/topics/tp061122-1.html
- 4) Anderson W. Guidelines for the management of chemotherapy and systemic anticancer therapy induced toxicities within primary care. *Northen Cancer Alliace*. 2018;(June 2018):0–21.

- 5) Silva D, Gomes A, Ms Lobo J, Almeida V, Almeida IF. Management of skin adverse reactions in oncology. *J Oncol Pharm Pract Off Publ Int Soc Oncol Pharm Pract*. 2020 Oct;26(7):1703–14.
- 6) Miller KK, Gorcey L, McLellan BN. Chemotherapy-induced hand-foot syndrome and nail changes: a review of clinical presentation, etiology, pathogenesis, and management. *J Am Acad Dermatol*. 2014 Oct;71(4):787–94.
- 7) Fromme EK, Eilers KM, Mori M, Hsieh Y-C, Beer TM. How accurate is clinician reporting of chemotherapy adverse effects? A comparison with patient-reported symptoms from the Quality-of-Life Questionnaire C30. *J Clin Oncol Off J Am Soc Clin Oncol*. 2004 Sep;22(17):3485–90.
- 8) Nikfarjam A, Ransohoff JD, Callahan A, Jones E, Loew B, Kwong BY, et al. Early detection of adverse drug reactions in social health networks: A natural language processing pipeline for signal detection. *JMIR Public Heal Surveill*. 2019;5(2):1–11.
- 9) Lee J-Y, Lee Y-S, Kim DH, Lee HS, Yang BR, Kim MG. The Use of Social Media in Detecting Drug Safety-Related New Black Box Warnings, Labeling Changes, or Withdrawals: Scoping Review. *JMIR Public Heal Surveill*. 2021;7(6):e30137.
- 10) Kürzinger ML, Schück S, Texier N, Abdellaoui R, Faviez C, Pouget J, et al. Web-based signal detection using medical forums data in France: Comparative analysis. *J Med Internet Res*. 2018;20(11).
- 11) Basch E. The missing voice of patients in drug-safety reporting. *N Engl J Med*. 2010 Mar;362(10):865–9.
- 12) Basch E, Jia X, Heller G, Barz A, Sit L, Fruscione M, et al. Adverse symptom event reporting by patients vs clinicians: relationships with clinical outcomes. *J Natl Cancer Inst*. 2009 Dec;101(23):1624–32.
- 13) Sheikhalishahi S, Miotto R, Dudley JT, Lavelli A, Rinaldi F, Osmani V. Natural language processing of clinical notes on chronic diseases: systematic review. *JMIR Med informatics*. 2019 Apr;7(2):e12239.
- 14) Datta S, Bernstam E V, Roberts K. A frame semantic overview of NLP-based information extraction for cancer-related EHR notes. *J Biomed Inform*. 2019 Dec;100:103301.
- 15) Aramaki E, Miura Y, Tonoike M, Ohkuma T, Masuichi H, Waki K, et al. Extraction of adverse drug effects from clinical records. *Stud Health Technol Inform*. 2010;160(Pt 1):739–43.
- 16) Dreisbach C, Koleck TA, Bourne PE, Bakken S. A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data. *Int J Med Inform [Internet]*. 2019;125(February):37–46. Available from: <https://doi.org/10.1016/j.ijmedinf.2019.02.008>
- 17) LifePalette. URL: <https://mediaid.co.jp/service/lifepalette/> [accessed 2021-6-9].
- 18) ja_sentence_segmenter. URL: https://github.com/wwwcojp/ja_sentence_segmenter [accessed 2021-6-9].
- 19) Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas [Internet]*. 1960;20(1):37–46. Available from: <https://doi.org/10.1177/001316446002000104>
- 20) Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med*. 2005 May;37(5):360–3.
- 21) fastText. URL: <https://github.com/facebookresearch/fastText> [accessed 2021-6-9].
- 22) bert-japanese. URL: <https://github.com/cl-tohoku/bert-japanese> [accessed 2021-6-9].
- 23) Janome. URL: <https://mocobeta.github.io/janome> [accessed 2021-6-9].
- 24) MeCab. URL: <https://taku910.github.io/mecab/> [accessed 2021-6-9].
- 25) Hochreiter S. Long Short-Term Memory. 1997;1780:1735–80.
- 26) Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*. 2005;18(5–6):602–10.
- 27) Graves A, Mohamed AR, Hinton G. Speech recognition with deep recurrent neural networks. *ICASSP, IEEE Int Conf Acoust Speech Signal Process - Proc*. 2013;(3):6645–9.
- 28) Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conf North Am Chapter Assoc Comput Linguist Hum Lang Technol - Proc Conf*. 2019;1(Mlm):4171–86.
- 29) Cocos A, Fiks AG, Masino AJ. Deep learning for pharmacovigilance: Recurrent neural network architectures for labeling adverse drug reactions in Twitter posts. *J Am Med Informatics Assoc*. 2017 Jul 1;24(4):813–21.
- 30) Liu X, Chen H. A research framework for pharmacovigilance in health social media: Identification and evaluation of patient adverse drug event reports. *J Biomed Inform [Internet]*. 2015;58:268–79. Available from: <http://dx.doi.org/10.1016/j.jbi.2015.10.011>
- 31) Gupta S, Pawar S, Ramrakhiani N, Palshikar GK, Varma V. Semi-Supervised Recurrent Neural Network for Adverse Drug Reaction mention extraction. *BMC Bioinformatics [Internet]*. 2018;19(8). Available from: <https://doi.org/10.1186/s12859-018-2192-4>