

多施設電子医療情報を用いた 肺がん患者の薬物治療効果へのAIモデルの適用

荒木 賢二³、松元 信弘^{1,2}、東郷 香苗⁴、米本 直裕⁴、大木 恵美子⁴、徐 凌華⁴、
長谷川 義行⁵、竹本 涼太⁵、邵 帥⁵、宮崎 泰可¹

¹ 宮崎大学医学部内科学講座 呼吸器・膠原病・感染症・脳神経内科学分野

² 独立行政法人国立病院機構 宮崎東病院 呼吸器科

³ 宮崎大学名誉教授

⁴ ファイザー株式会社

⁵ 株式会社NTTデータ

研究資金及び利益相反

- 本研究は、ファイザー株式会社からの研究資金で実施した。
- 本研究の研究担当者は、「宮崎大学医学部等における臨床研究等利益相反マネジメント規程」の規定に従って、宮崎大学医学部等における臨床研究等利益相反マネジメント委員会に必要事項を申告し、その審査と承認を得た。
- 次世代医療基盤法に基づき、認定匿名加工医療情報作成事業者である（一社）ライフデータイニシアティブの審査委員会において必要事項を申請し、審査、承認を得た。

研究の背景と目的

- 前向きにデータ収集する臨床研究に比べて、電子カルテデータベース(DB)を用いた研究は大規模な患者のデータを迅速に解析できるため、研究課題に対して適時にエビデンスが得られる
- 多施設電子医療情報に基づく電子カルテDB(Electronic Health Record: EHR)の構築が国内外で進んでいる一方で、治療効果などの臨床アウトカムに関わる多くのデータは経過記録や画像検査レポート等のテキストに含まれており、これらテキストの非構造化データの構造化を行わなければ解析が困難という課題がある
- 昨年度の本学会にて、宮崎大学の電子カルテDBの経過記録等のテキストから肺がん患者の薬物治療効果を自然言語処理の手法を用いて抽出した後ろ向き研究において、キーワードの文脈情報が治療効果の評価に重要であることを発表した

目的

本研究では宮崎大学の電子カルテDBおよび多施設EHRを用い、肺がん患者の薬物治療効果についてAIモデルであるBERTモデルを利用して評価する方法を検討した

研究デザイン

宮崎大学医学部附属病院(宮大)研究

- 宮大の電子カルテDBを用いた臨床アウトカムを評価する手法の検討



LDIの千年カルテDB(千年カルテ)研究

- 宮大の電子カルテDBを用いて検討した手法を多施設EHRデータに適用して評価



方法: 対象患者と評価方法・評価項目

対象患者

宮大

2018年4月～2020年9月に宮大に通院又は入院して、宮大で初めて肺がんの薬物治療を受けた18歳以上の肺がん患者

千年カルテ

2017年10月～2021年1月に千年カルテに含まれる6施設に通院又は入院して、当該病院で初めて肺がんの薬物治療を受けた18歳以上の肺がん患者

評価方法と評価項目

- 宮大の電子カルテDBを用いて宮大BERTモデルを構築し、そのモデルを千年カルテに適用した
- 宮大の電子カルテDBを用いて構築した宮大BERTモデルと千年カルテに適用した改良版BERTモデルの精度について、異なるAIモデル(LongformerモデルおよびNaïve Bayesモデル)と比較した
- 精度の評価にはAccuracyとPrecision、Recall、F1値を用いた
- 改良版BERTモデルにより抽出された薬物治療効果を用いて薬物治療ラインごとの無増悪期間(Time-to-Progression: TTP)を算出し、人が抽出した薬物治療効果(教師データ)で算出したTTPと比較した

方法: BERTモデル

- BERT: 2018年10月にGoogleが発表した革新的な自然言語処理向けの事前学習手法 & 汎用言語モデル

タスク	概要	前SOTA	BERT
GLUE	8種の言語理解タスク	75.2	81.9
1. MNLI	2入力文の含意/矛盾/中立を判定	82.1	86.7
2. QQP	2質問文が意味的に等価か判定	70.3	72.1
3. QNLI	SQuADの改変. 陳述文が質問文の解答を含むか判定	88.1	91.1
4. SST-2	映画レビューの入力文のネガポジを判定	91.3	94.9
5. CoLA	入力文が言語的に正しいか判定	45.4	60.5
6. STS-B	ニュース見出しの2入力文の意味的類似性をスコア付け	80.0	86.5
7. MRPC	ニュース記事の2入力文の意味的等価性を判定	82.3	89.3
8. RTE	2入力文の含意を判定	56.0	70.1
SQuAD	質疑応答タスク. 陳述文から質問文の解答を抽出	91.7	93.2
CoNLL	固有表現抽出タスク. 単語に人物/組織/位置のタグ付け	92.6	92.8
SWAG	入力文に後続する文を4つの候補文から選択	59.2	86.3

BERT: Bidirectional Encoder Representations from Transformers
SOTA (state of the art): the highest level of algorithm at a particular time

事前学習 (pre-training)

- ▶ 教師なし学習
文書をそのまま学習可能
- ▶ 双方向Transformer
離れた単語の関係性(文脈情報)を考慮可能

BERTモデル(汎用言語モデル)

タスクへの適用 (fine-tuning)

- ▶ 転移学習
上記モデルをベースに少量の学習データでタスクに適用可能となり、従来技術より高い精度を実現

タスク

分類

情報抽出

検索

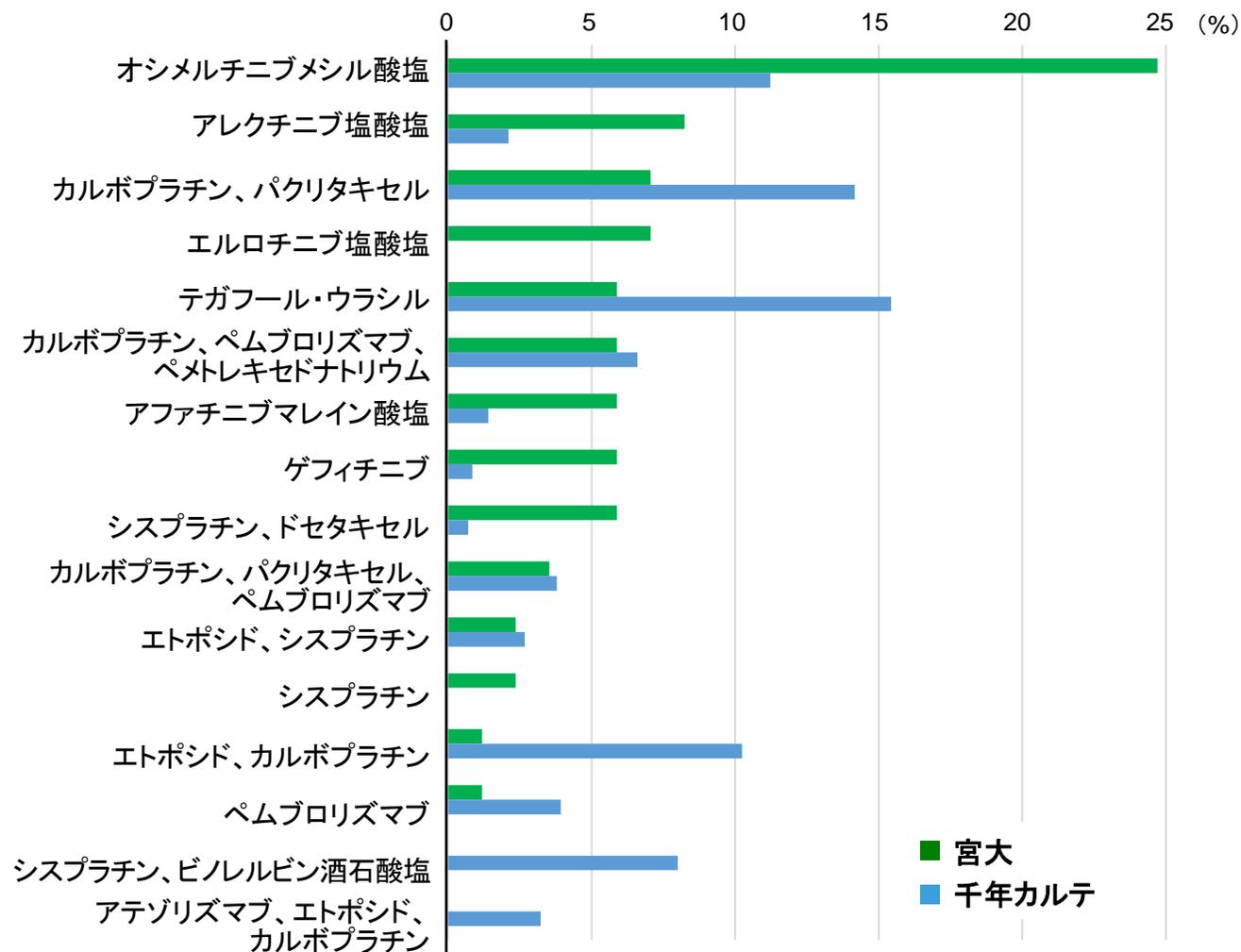
要約

結果:患者背景

患者基本情報

		宮大 (n=85)	千年カルテ (n=713)
年齢 (平均±SD)		67.3±10.4	68.2±10.2
性別	男性	37 (43.5%)	482 (67.6%)
	女性	48 (56.5%)	231 (32.4%)
身長 (平均±SD)		159.6±9.1	162.3±9.3
体重 (平均±SD)		57.7±10.9	58.7±12.8
原疾患の ステージ	1	4 (4.7%)	62 (8.7%)
	2	3 (3.5%)	79 (11.1%)
	3	9 (10.6%)	93 (13.0%)
	4	47 (55.3%)	167 (23.4%)
	不明	22 (25.9%)	312 (43.8%)
原疾患 手術歴	あり	6 (7.1%)	31 (4.3%)
	なし	0 (0%)	559 (78.4%)
	不明	79 (92.9%)	123 (17.3%)
原疾患 放射線治療歴	あり	8 (9.4%)	200 (28.1%)
	なし	5 (5.9%)	513 (71.9%)
	不明	72 (84.7%)	0 (0%)
原疾患の再発	あり	1 (1.2%)	8 (1.1%)
	なし	84 (98.8%)	705 (98.9%)

1次治療で用いた薬物



※いずれかの群で2%以上の患者に投与された薬剤のみを記載
各施設での最初の肺がん薬物治療を1次治療と定義

結果：誤りパターンの解析

誤りパターンとその対応

誤りパターン		文章イメージ	対応
未学習の表現	学習データ不足により、未学習となっている表現の予測を誤る	【正解】奏効【予測結果】評価不可 【文章】…右上肺野腫瘍径は若干小さくなった… 未学習の表現	学習データ追加による、モデルが学習する表現の拡充
否定表現	学習データ不足により、未学習となっている否定表現の予測を誤る	【正解】安定【予測結果】進行 【文章】…CT: 明らかな病変増大なかった… 増大の否定表現	否定表現の学習データ追加による、モデルが学習する表現の拡充
表現の混在	肺癌以外の疾患や薬物治療以外の治療・複数部位についての表現があると予測が困難となる	【正解】進行【予測結果】奏効 【文章】…肺門部癌: 増大 …中葉肺癌はやや縮小… 複数部位についての表現の混在	セクション分割ルールの見直し等

※「表現の混在」への対応は今後の課題と整理しており、本研究内での誤りパターン解析を受けた対応は未学習の表現・否定表現に対する学習データ追加となる

誤りパターン対応前後のBERTモデルの精度

	F1値(平均±SD)	
	学習データ追加前	学習データ追加後
奏効	0.59±0.08	0.61±0.05
安定	0.45±0.05	0.51±0.05
進行	0.32±0.18	0.37±0.13
評価不可	0.57±0.11	0.51±0.14
3種平均	0.45±0.10	0.50±0.08

※ 3種平均: 奏効、安定、進行の3つの平均

結果:BERTモデルの精度

教師データ数

	宮大		千年カルテ	
	文書数	患者数	文書数	患者数
奏効	153	27	191	109
安定	98	22	81	60
進行	75	17	140	79
評価不可	703	31	412	215
合計	1,029	31	824	322

宮大BERTモデルおよび千年カルテ研究のモデル精度

		宮大	千年カルテ		
		最終宮大BERTモデル	改良版BERTモデル	Longformerモデル	Naïve Bayesモデル
3 種 平 均	Precision	0.42	0.40	0.36	0.43
	Recall	0.63	0.54	0.26	0.38
	F1値	0.50	0.45	0.27	0.40
Accuracy		0.67	0.58	0.30	0.44

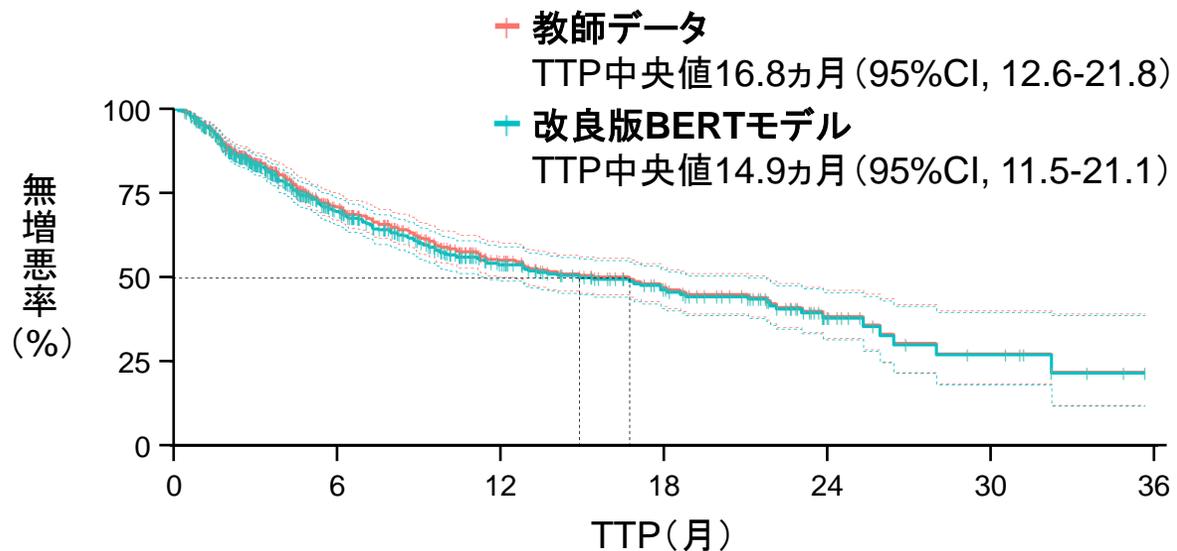
※3種平均:奏効、安定、進行の3つの平均

- Precision(適合率):モデルが奏効、安定、進行と予測したもののうち、実際に奏効、安定、進行であった割合
- Recall(感度):実際に奏効、安定、進行であったもののうち、モデルが奏効、安定、進行と予測したものの割合
- F1値:PrecisionとRecallの調和平均
- Accuracy(正解率):モデルが正しく治療効果を予測できた割合

結果:1次治療と2次治療の無増悪期間(TTP)

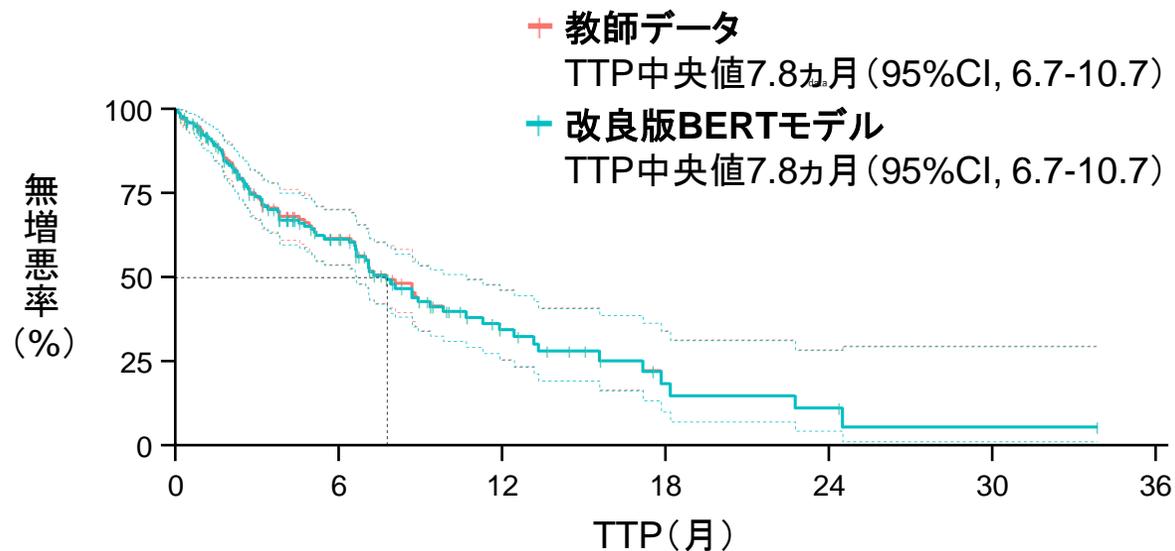
人が抽出した結果(教師データ)と改良版BERTモデル予測結果

1次治療 Kaplan-Meier曲線



Number at risk		0	6	12	18	24	30	36
教師データ	713	264	125	73	22	8	0	
改良版BERTモデル	713	260	124	73	22	8	0	

2次治療 Kaplan-Meier曲線



Number at risk		0	6	12	18	24	30	36
教師データ	209	61	17	5	3	**	0	
改良版BERTモデル	209	61	17	5	3	**	0	

※患者数が1~2名の場合「**」でマスクした。

- 肺がん患者の薬物治療効果についてAIモデルであるBERTモデルを宮大の電子カルテDBおよび多施設EHRに適用して評価し、いずれも肺がん患者の薬物治療効果が得られた
- BERTモデルは他のモデルと比較して高い精度を示した一方で、BERTモデル自体の精度には改善の余地があった
- BERTモデルで抽出した薬物治療効果から算出したTTPは、人が抽出した結果と同様の傾向を示した
- 多施設EHRにAIモデルを適用することで、前向きにデータ収集する臨床研究に比べて、より迅速かつ大規模な肺がん患者の薬物治療効果のエビデンスが得られる可能性が示唆された