

理学療法士間の知識共有に向けた 臨床推論テキストのICF/ICDへの分類

Classifying Clinical Reasoning Texts Using ICF/ICD Frameworks for Facilitating Knowledge Sharing among Physiotherapists

櫛 力輔^{1*}
Rikisuke Ichii¹

松下 光範¹
Mitsunori Matsushita¹

堀 寛史²
Hirofumi Hori²

¹ 関西大学

¹ Kansai University

² 甲南女子大学

² Konan Women's University

Abstract: 理学療法における臨床推論は、患者の問題を特定し管理について判断を下すための推論プロセスであり、重要な技能である。診療録などの分析や活用が臨床推論教育に有用だと考えられるが、記述内容には理学療法士の思考過程が反映される一方で、自由記述も含まれ、表現の差異が知識共有の障害となる。本研究の目的は、そのような非構造データに医療分野の標準的枠組みである国際生活機能分類 (ICF) や国際疾病分類 (ICD) を取り入れて構造化し、理学療法士の共通理解のもと計算機処理を可能にすることである。本稿ではその端緒として、模擬症例における推論文章の各文が ICF, ICD に該当するか判別する分類器を BERT を用いて作成した。

1 はじめに

理学療法は「身体に障害のある者に対し、主としてその基本的動作能力の回復を図るため、治療体操その他の運動を行なわせ、及び電気刺激、マッサージ、温熱その他の物理的手段を加えることをいう」と定義されており¹、理学療法士はそれら医学的リハビリテーションを提供する専門職である。昨今の高齢化問題を背景に、社会保障の観点から健康寿命を延伸させる必要性が説かれている²。健康寿命とは「健康上の問題で日常生活が制限されることなく生活できる期間」を指し、疾病や障害による身体機能の回復や維持、予防を行うことで日常における活動や社会参加の維持・向上を支援する理学療法士は健康寿命の延伸に寄与する重要な役割を担っており、その育成が重要視されている。

理学療法士が治療立案をするにあたって重要な要素の一つに臨床推論が挙げられる。臨床推論は世界理学療法連盟によって「理学療法士やその他医療従事者が、データを収集・評価し、患者の問題特定や管理に関する判断を下すために用いる推論のプロセス」と定義されている³。効果的な臨床推論は患者の転帰を良好にするための不可欠なスキルであるとも言われており [5]、

臨床推論教育は理学療法教育において重要な位置を占めている。芳野ら [22] の研究では職員指導経験がある理学療法士 15 名に対して、指導している理学療法士がどのような能力を獲得したときに理学療法士として自立したと感じるかをインタビューした結果、臨床思考能力が 24.0% と最も多く記録されたと報告されている。しかし、2019 年に日本理学療法士協会が 3942 人の理学療法士を対象に行ったアンケート調査 [17] によれば、「臨床推論を知っていますか?」という問いに対して「知っていて実践している」と回答したのは 47% に留まり、36% が「知っているがあまり実践できていない」、16% が「知らない」と回答し、アンケートに回答した理学療法士の半数以上が普段から臨床推論を実践できていないことが明らかとなった。以上のことから、臨床推論が理学療法士にとって不可欠なスキルである一方で、その教育や実践には依然として課題が残っていると見える。臨床推論は、単純な知識の習得で身につくものではなく、患者の個別性や多様な状況に応じて変化する思考プロセスであり、その教育は容易では

¹理学療法士及び作業療法士法: <https://laws.e-gov.go.jp/law/340AC000000137> (2024/12/14 確認)。

²令和 4 年版厚生労働白書—社会保障を支える人材の確保—: <https://www.mhlw.go.jp/wp/hakusyo/kousei/21/d1/zentai.pdf> (2024/12/14 確認)

³World Physiotherapy: <https://world.physio/resources/glossary> (2024/12/14 確認)

*連絡先: 関西大学総合情報学部
〒 569-1095 大阪府高槻市霊仙寺町 2-1-1
E-mail: k244081@kansai-u.ac.jp

構成要素	第1部：生活機能と障害		第2部：背景因子	
	心身機能・ 身体構造	活動・参加	環境因子	個人因子
領域	心身機能 身体構造	生活・人生領域 (課題、行為)	生活機能と障害へ の外的影響	生活機能と障害へ の内的影響
構成概念	心身機能の変化 (生理的) 身体構造の変化 (解剖学的)	能力 標準的環境におけ る課題の遂行 実行状況 現在の環境におけ る課題の遂行	物的環境や社会的 環境、人々の社会 的な態度による環 境の特徴がもつ促 進的あるいは阻害 的な影響力	個人的な特徴の 影響力
肯定的側面	機能的・構造的 統合性	活動 参加	促進因子	非該当
	生活機能			
否定的側面	機能障害 (構造障害を含 む)	活動制限 参加制約	阻害因子	非該当
	障害			

図 1: ICF の概観 (文献 [16] より引用)

ない。特に、経験の浅い理学療法士や学生にとっては、患者情報の収集や解釈、治療計画の立案といった一連のプロセスにおける適切な考え方を学ぶことは困難であり、理学療法士育成の大きな課題となっている。

臨床推論教育を進める上では、診療録や症例レポートといった実際の記録を分析することが有用であると考えられている [20]。上記に挙げた文書には、患者の主訴、患者の現在の状態や推移、今後の治療方針、それらに対する理学療法士の思考（アセスメント）などが記録されており、これを活用することで、実際に行われている臨床推論の具体的なプロセスを教育に取り入れることが期待される。このときアセスメントにおける複雑な判断の背景などを記載する際には、自由記述形式が採用されていることが多い。しかし、同様の意味を表す記述であっても個人間での表現に差異が生まれ、理学療法士間での知識の共有が困難である。診療録や症例レポートなどを効果的に利用するためには、内容を統一的な形式に整理し、その上で共通の理解を得るための工夫が必要である。

本研究では、診療録や症例レポートに含まれる自由記述文のような非構造データを効率的かつ効果的に活用するために、国際生活機能分類 (ICF) や国際疾病分類 (ICD) といった国際的に医療分野で広く用いられている分類体系を利用し、これらのデータを構造化することを目指す。ICF や ICD に基づいて整理することが可能となれば、理学療法士間での共通理解が深まるだけでなく、そこに含まれる情報の分析や比較が容易になり、類似症例の検索や症例内容の概観などを迅速に行うことが可能となる。これにより、教育現場では、理学療法士が症例を効率的に探索・参照することが期

待される。また、研究分野では、構造化データによる大規模解析や傾向分析を通して新たな知見の発見につながる。

本稿では、大腿骨頸部骨折の模擬症例を対象として記述された臨床推論テキストの各文が ICF の心身機能、身体構造、活動・参加、また ICD に該当するかを判別する分類器を BERT [7] を用いて作成した。

2 医療標準化

WHO (世界保健機関) が定める ICF 及び ICD は、医療の標準化を推進し、医療の質向上や効率的な組織運営、さらには病態や治療方針に関わる説明の透明性向上など、公平な医療提供を実現するための基盤としても機能している。その意義は多岐にわたるが、特筆すべきは疾患や健康状態に関する統一的な診断・記録の枠組みを提供し、国際的な情報共有や比較を可能にする点である。これらの分類を活用した研究を進めることで疾患の実態把握が進むとともに、医療教育においても世界的に標準化された知識体系を構築することにつながる。また、WHO によると ICD と ICF は補完関係にあり、ICD は、疾病・障害・傷害などの健康状態の病因論的枠組みを提供する一方で、これらの健康状態に関連する機能や障害は ICF で分類されるとしており、ユーザにこの 2 つの分類を併用することを推奨している [13]。そのため、本研究においてもこれら 2 つを用いて、診療録や症例レポートに含まれている臨床推論に関わる自由記述文の構造化を試みる。

2.1 国際生活機能分類 (ICF)

国際生活機能分類 (International Classification of Functioning, Disability and Health) とは、人間のあらゆる健康状態に関係した生活機能状態から、その人を取りまく社会制度や社会資源までをアルファベットと数字を組み合わせた方式で分類し、記述・表現をしようとするものである [16]。図 1 に示すように、ICF には 2 つの部門があり、それぞれは 2 つの構成要素からなる。以下に本稿で使用する「心身機能」、「身体構造」「活動」「参加」の 4 つの定義を記す。

「心身機能」とは、身体系の生理的機能 (心理的機能を含む) であると定義されており、例えば、手足の動き、精神の働き、視覚・聴覚、内臓の働きが含まれる⁴。「身体構造」とは、器官・肢体とその構成部分などの、身体の解剖学的部分であると定義されており、手

⁴厚生労働省: ICF (国際生活機能分類) — 「生きることの全体像」についての「共通言語」— <https://www.mhlw.go.jp/stf/shingi/2r9852000002ksqi-att/2r9852000002kswh.pdf> (2024/12/14 確認)

足の一部、心臓の一部（弁など）などの、体の部分のことを指す。「活動」とは、課題や行為の個人による遂行のことであると定義されており、実用歩行やその他のADL（日常生活行為）だけでなく、調理・掃除などの家事行為・職業上の行為・余暇活動（趣味やスポーツなど）に必要な行為・趣味・社会生活上必要な行為を指す。「参加」とは、生活・人生場面（life situation）への関わりのことであると定義されており、社会参加だけではなく、家庭内役割であるとか、働くこと、職場での役割、あるいは趣味にしても趣味の会に参加する、スポーツに参加する、地域組織のなかで役割を果たす、文化的・政治的・宗教的などの集まりに参加する、などの広い範囲のものが含まれる。

ICFは分類コード、分離点、評価点で構成される。具体例として「s7402.181」を挙げる。分類コードは「s7402」にあたり、分離点はピリオドで表される。分離点以降の数字列「181」が評価点にあたる。構成要素それぞれは分類コード内のアルファベットに対応しており、心身機能は「b」、身体構造は「s」、活動・参加は「d」で表される。分類コード内の数字は構成要素の分類を示している。この分類は階層化されており、数字の位が階層の深さに対応している。具体例に示した「s7402」は「s：身体構造」、「7：運動に関連した構造」、「40：骨盤部の構造」、「2：筋肉」を表し、全てをまとめると「骨盤部の筋肉」を表すコードとなる。評価点が示すものはそれぞれの構成要素毎に異なる。主に機能障害の程度や性質、障害部位などを表すために用いられる。

2.2 国際疾病分類 (ICD)

国際疾病分類 (International Statistical Classification of Diseases and Related Health Problems) とは、正式には「疾病及び関連保健問題の国際統計分類」といい、異なる国や地域から、異なる時点で集計された死因や疾病のデータの記録、分析、比較を行うために国際的に統一した基準で設けられた分類であり [15]、ICD-11では疾病・障害・死因が 27 章構成で分類されている。ICD も ICF と同様にアルファベットと数字を組み合わせることで表現され、コードによって疾病や障害部位、原因などを表すことができる。例えば、「S72.00」というコードは「大腿骨頸部骨折、閉鎖性」を表し、「S72.01」というコードは「大腿骨頸部骨折、開放性」を示す。この例では、S72 が大腿骨頸部骨折を表し、その後の 00・01 で閉鎖骨折か開放骨折かを分類することができる。

3 関連研究

臨床推論の過程やその思考方法を明らかにしようとする試みやそれらを支援しようとする試みは盛んに行

われている。Wainwright[11] らは、初学者と経験豊富な理学療法士の間で、臨床意思決定能力とそのプロセスにどのような違いがあるかを明らかにし、臨床意思決定に影響を与える要因のモデルを構築することを目的とした。脳血管障害 (CVA) 患者を治療する初学者理学療法士と経験豊富な理学療法士 1 名ずつのペア 3 組を対象に、評価・治療の過程をビデオ録画及び半構造化インタビューを通じてデータを収集し、分析を行った。その結果、臨床意思決定に影響を与える要因は、情動的因子と指示的因子の 2 つに分類され、初学者は情動的因子に頼る傾向が強く、一方で経験豊富な理学療法士は指示的因子に依存する傾向が見られた。

また、以下のように ICF の枠組みを取り入れ支援を試みている例も存在する。Hage ら [6] は筋骨格理学療法における臨床推論の学習を支援する新しいカードゲーム「Hypothesis Family Game」を提案している。ゲームを通して、患者の症状に関連する情報を分析し、仮説を構築する過程を楽しみながら学ぶことができ、学生の知識とスキルを統合的に高めることが期待されている。ここでは、ICF に基づいた仮説カテゴリーの枠組みがゲームの構造に組み込まれており、学生が患者の状態を分析する際に、ICF の視点を適切に反映させることを促すような仕組みになっている。

Atkinson ら [1] は、臨床推論スキルの向上を促すための手段として、振り返りと議論が重要が指摘されているものの、それらの重要なニーズを満たすためのツールが提供されていないことに目を向け、あらゆるスキルレベルの理学療法士が臨床推論スキルを育成するための振り返りと議論を深められるツールを開発した。ツールには「理学療法士実践ガイド」と「ICF」といった既存の概念的枠組みが統合されている。

上記に述べた研究では、様々な方法で臨床推論過程の解明や推論支援が行われているが、定性的な分析も多く計算機による分析を行った例は多くない。また、臨床推論に関する研究は小規模なデータを用いた例が多く、大規模データを分析することによって行われた例は少ない。このような状況の中で、近年の電子カルテの普及に伴い、紙媒体では困難であった大規模な医療情報の活用が期待されており、医療分野における情報抽出技術を評価・洗練させる目的で Medical Records Track[10] などの共有タスクが国際的に取り組まれている。このような取り組みは医療上の意思決定支援を行うツールの開発を促進することにもつながる。また、Medical Records Track が英語の医療文書を対象としたタスクであるのに対して、日本語を対象とした取り組みも行われており、情報アクセス技術向上のためのプロジェクト NTCIR のコアタスクの一つとして MedNLPDoc[3] が取り組まれている。MedNLPDoc のタスクに取り組んでいる具体例として、Sakishita ら [9] がルールベースによる ICD コードの推定を行っている。この例では、

(1) 診断を含む文の選別, (2) 「ライフサイエンス辞書」を用いた英語医療用語の日本語変換, (3) アルツハイマー型認知症やアルツハイマー病といった表記ゆれへの対応, (4) 病名と部位の組み合わせを検出しコードを推定, (5) XML タグ情報を活用し, 特定のタグ推定という5つのルールベース手法を組み合わせたコード推定が行われており, MedNLPDoc タスクにおいて最も厳しい評価基準である「SURE」で最高の F 値を記録し, 他チームを上回る性能を示した。

臨床推論のプロセスや教育支援に関する様々な試みが行われ, ICF の枠組みや振り返り手法などが有用であることが示されてきた。臨床推論教育を支援していく上では, これまでのようにインタビューなどを通じた定性的な研究で詳細を把握していくと同時に, 大規模データを活用して統計的な知見を導き出すなど大きな視点での分析やそれに基づいた支援が必要となると考える。そうした状況を整えるためには, 医療文書の構造化や情報抽出を進め, 理学療法文書に計算機が適用可能な状況を整える必要がある。MedNLPDoc タスクのように日本語の医療文書においてもそのような取り組みは進められており, これらを理学療法分野に取り入れることで新たな知見の発見やそれらを取り入れた支援システムの開発につながると考える。そのため, 本研究では理学療法の推論テキストを ICF や ICD を取り入れて構造化し, 理学療法士の共通理解のもと計算機処理を可能にすることを目指す。

4 データセット

臨床推論テキストの各文に対して, ICF の「心身機能」, 「身体構造」, 「活動・参加」, ICD に該当する記述が含まれているかを 0,1 のラベルによってアノテーションを行うことでデータセットを作成した。

4.1 臨床推論テキスト

臨床推論テキストには, 吉田ら [21] 及び畠山ら [18] が行った先行研究で収集されたテキストを使用している。以下にそれらの研究概要及びデータ取得について簡潔に述べる。

吉田らは理学療法初学者が必要な医療情報や理学療法検査の実測値を見落とさないことを目的として情報収集の段階で患者の重要な医療情報や理学療法検査の着目点を提示するシステムを構築した。システムの有用性を評価するために, 急性期病院に所属する1年目から5年目の理学療法士計20名に対し, 模擬症例(大腿骨頸部骨折症例)に対する理学療法プロセスの論理的文章記述と模擬症例の問題点を箇条書きで抽出する課題を与えた。課題は2回与えられ, 1回目は20名全

員が同様条件下で課題を行い, 2回目は被験者を10名ずつに分け片群は吉田らの提案システム使用下で同一課題を行い, 片群はシステムを使用せずに同一課題を行った。

畠山らは理学療法分野のテキスト解析を促進するため, 理学療法分野のテキストにおけるアノテーション基準の構築とその妥当性を検証することを目的とした。事前に収集した大腿骨頸部骨折術後の模擬症例に対する臨床推論文章15件に, アノテーションの付与作業を行い, 作成したアノテーション基準の妥当性を検証するため, 1~8年目の理学療法士5名に, 作成したアノテーション基準を提示した上で, 実際にアノテーションの付与作業を行い, 付与されたアノテーションの一致度を検証した。

吉田らの論理的文章記述課題において取得された文章計40件と畠山らが取得した臨床推論文章15件を組み合わせて本稿で取り扱う計55件の臨床推論テキストとした。臨床推論テキストは全て大腿骨頸部骨折患者の模擬症例を基に作成されている。

4.2 アノテーション

臨床推論テキストを句点を基に1文に分け, 計816文をアノテーションの対象とした。表1に示すように各文が ICF の心身機能 (b), 身体構造 (s), 活動・参加 (d) また ICD に該当する記述を含んでいない場合に0を含んでいる場合に1をそれぞれラベル付けした。

表1を用いて, ラベル付与について詳しく説明する。ID1のテキストでは, 「圧痛」「伸長時間」「収縮時痛」といった記述が心身機能にあたるため, bの項目に1が付与される。加えて, 「右大腿近位外側部」といった記述が身体構造にあたるため, sの項目に1が付与される。その他の項目には該当する記述が存在しないため, 0が付与される。ID2のテキストでは, 「短距離であれば杖歩行自立」や「長距離歩行では歩行器を使用し自立レベル」といった記述が活動・参加にあたるため, dの項目に1が付与される。ID3のテキストでは, 「歩行時の荷重痛」といった記述が心身機能にあたるため, bの項目に1が付与され, 加えて, 「歩行困難を来している」といった記述が活動・参加にあたるため, dの項目に1が付与される。ID4のテキストでは, 「転倒」や「大腿骨頸部骨折」といった記述が ICD 項目にあたるため, 1が付与される。ID1やID3の様に記述内容によっては複数のラベルに1が付与される場合も存在する。

816文のうち1が付与されたテキスト数は心身機能 (b) が413文 (50.6%), 身体構造 (s) が274文 (33.6%), 活動・参加 (d) が186文 (22.8%), ICD が77文 (9.4%) であった。ラベル付けデータ作成後に理学療法士2名の精査を通じて, 妥当性の担保を行った。

表 1: アノテーションの例

ID	臨床推論テキスト	b	s	d	ICD
1	本症例では著名な疼痛は認めないものの、右大腿近位外側部に圧痛、伸長時間痛、収縮時痛を認める。	1	1	0	0
2	現在術後 14 日目であり、トイレまでの短距離であれば杖歩行自立、長距離歩行では歩行器使用し自立レベルとなっている。	0	0	1	0
3	歩行について見てみると、まず主訴として歩行時の荷重痛を訴えており、それによる歩行困難を来していると考える。	1	0	1	0
4	本症例は転倒契機に大腿骨頸部骨折 (Garden IV) を受傷し、人工骨頭置換術 (後方アプローチ) を施行された症例である。	0	0	0	1

表 2: 評価結果 (データ拡張前)

分類項目	正解率	適合率	再現率	F1-score
心身機能 (b)	0.867	0.837	0.923	0.875
身体構造 (s)	0.934	0.850	0.976	0.908
活動・参加 (d)	0.887	0.811	0.665	0.727
ICD	0.937	0.838	0.690	0.757

表 3: 該当記述数と各評価指標の相関

評価指標	相関係数	p 値
正解率	-0.663	0.337
適合率	0.236	0.764
再現率	0.777	0.223
F1-score	0.739	0.261

5 評価・考察

入力された 1 文に ICF の心身機能 (b)、身体構造 (s)、活動・参加 (d)、ICD に該当する記述が含まれているか否かをそれぞれ 2 値で判定する分類器を BERT によって作成し、計 4 つの分類器を作成した。BERT の事前学習済モデルには東北大学乾研究室が公開している日本語 BERT モデル⁵を用いた。事前学習モデルに対して、4 章で構築したデータセットを用いて、ファインチューニングを行うことにより、分類器を作成した。ファインチューニングにおけるバッチサイズは 8、エポック数は 1 とした。学習及び評価は 4 章で構築したデータセットの規模を考慮し、5 分割交差検証にて行い、4 つの分類器それぞれの正解率 (Accuracy)、適合率 (Precision)、再現率 (Recall)、F1-score を算出した。

表 2 に評価結果を示す。心身機能 (b) 並びに身体構造 (s) の分類器については心身機能の適合率を除く全ての項目で 0.85 以上を記録し、良好な結果となった。一方で、活動・参加 (d)、ICD の分類器については、正

解率を除く全ての項目で 0.85 を下回る結果となり、高い精度での分類には至らなかった。分類精度低下につながった要因の一つとして、データセット内に含まれる該当記述数が少ないことが考えられる。4.2 節で示した通り、分類精度が高かった心身機能 (b) や身体構造 (s) に関してはデータセット内に 1 のラベルが付与された該当記述数がそれぞれ心身機能: 413 文、身体構造: 274 文と多く存在する。しかし、分類精度が低かった活動・参加 (d)、ICD に関しては活動・参加 (d): 186 文、ICD: 77 文と分類精度が高かったものと比べると該当記述数が少ない傾向が見られた。そのため、それぞれの分類器作成における該当記述数を独立変数、各評価指標を従属変数として相関係数を算出した。結果を表 3 に示す。再現率・F1-score に関しては、p 値が低く統計的有意性は確認できなかったものの、中程度の正の相関が見られ、該当する記述データを増やすことで分類精度が改善する可能性が示唆された。

6 データ拡張による分類精度向上

評価結果を踏まえた考察からデータセットの規模を拡大することで分類精度が向上する可能性が示唆された。そのため、本章では既存のデータセットに対して、データ拡張を実施することで分類精度向上を図ることとした。

6.1 データ拡張 (Data Augmentation)

データ拡張 (Data Augmentation) は、機械学習においてモデルの性能向上を目的として、既存のデータセットを加工・変換することで新たな学習データを生成することを指し、自然言語処理のみならず、画像認識や音声認識などの多様な分野で活用されている。特に収集可能なデータが限られている状況においてこの手法は効果的であり、小規模なデータセットであっても追加のデータを作成することで、モデルの過学習を防ぎ、汎化性能を向上させることが可能である。本稿

⁵tohoku-nlp/bert-base-japanese-v2: <https://huggingface.co/tohoku-nlp/bert-base-japanese-v2> (2024/12/14 確認)

表 4: データ拡張により生成されたテキストの例

	テキスト例
原文	また、FRT が 15cm とカットオフ値を下回っており股関節伸展制限の影響を受けている。
同義語置換後	また、FRT が 15cm とカットオフ値を下回っており股関節伸展制限の影響を聴き入れる て御出でなされる。
ChatGPT 修正後	また、FRT(Functional Reach Test) が 15cm でカットオフ値を下回っており、股関節伸 展制限の影響が認められます。

で扱っているような医療ドメインのテキストデータはプライバシー保護などの観点から入手が困難なこともあり、データ量が不足してしまうことが今回のようにモデル性能の低下を招く大きな要因となることもある。また、医療などの専門性が高いドメインに対するアノテーションは多大な労力を要し、加えて、その妥当性を担保することも必要となる。データ拡張によって生成されたデータには拡張前と同じラベルを付与することが可能なため、そうした観点からもデータ拡張を行う意義は大きいと考える。

6.2 手法

本節ではデータ拡張に用いた具体的な手法を述べる。自然言語処理分野におけるデータ拡張手法として Wei ら [12] による EDA (Easy Data Augmentation) が提案されている。この手法では、同義語置換 (Synonym Replacement)、ランダムな単語の挿入 (Random Insertion)、ランダムな単語の入れ替え (Random Swap)、ランダムな単語の削除 (Random Deletion) の 4 つを組み合わせて拡張テキストを生成する。本稿では、EDA 内の同義語置換と ChatGPT を用いたテキスト修正を行うことでデータ拡張を実施した。

まず、同義語置換について述べる。EDA における同義語置換は文からストップワード以外の n 個の単語をランダムに選び、これらの単語をランダムに選ばれた同義語で置き換えるというものである。本稿では、MeCab⁶ によりテキストを形態素に分割した後、置き換える単語数 n を 2 に設定し、日本語 WordNet[4] により同義語を検索・置換した。ストップワードには、ライブラリ SlothLib[14] に搭載されているストップワードリストを用いた。この際、テキストを特徴づける医療専門用語が置き換わることを防止するために、置き換える単語の品詞を動詞に限定した。生成されたテキストの具体例を表 4 の同義語置換後に示す。原文の「受け」の部分が「聴き入れる」に置換され、「いる」の部分が「御出でなされる」に置換されていることが見て取れる。一般に用いられる表現とはかけ離れた表現のテキストが同義語置換により生成されたため、診療録などで用い

られていても遜色のないテキストに修正する必要があると考えた。そのため、同義語置換の工程を経た後に、ChatGPT によってテキストを修正した。GPT のバージョンには GPT-4 Omni を用いた。プロンプトには、「以下のテキストを正しい日本語に修正してください。」と入力し、そのプロンプトの下部に同義語置換後のテキストを記述した。生成されたテキストの具体例を表 4 の ChatGPT 修正後に示す。修正後は「聴き入れる」や「御出でなされる」といった表現が「認められます」といったものに変化し、一般に用いられる表現に近づいたことが見て取れる。また、FRT という略称の記述に対して「Functional Reach Test」などの正式名称が補足されるといった副次的な効果も見てとれた。

上記に述べた同義語置換並びに ChatGPT による修正の工程を生成元のテキスト 1 文につき 1 度実施し、新たに 816 文を生成した。新たに生成された 816 文の意味が生成元のテキストと大きく異なることを理療士 2 名に確認してもらった上で、生成元のテキストと同じラベルを付与した。ラベルを付与した 816 文を新たにデータセットに加えることでデータ拡張を実施し、データ数を 2 倍の 1632 文とした。

6.3 データ拡張後の評価・考察

拡張したデータセットを用いて、新たに 4 つの分類器を作成し、拡張前と同様に 5 分割交差検証を行った。このとき、学習パラメータなどはデータ拡張前と同様である。評価結果を表 5 に示す。全 16 項目のうち、12 項目で精度の向上が確認された。最も大きく上がったものは、活動・参加における再現率で 0.665 から 0.759 となり、0.094 (14.1%) の精度向上が認められた。加えて、活動・参加の F1-score や ICD の再現率に関して、およそ 0.05 (活動参加/F1-score: 8.1%, ICD/再現率: 7.3%) の精度向上が認められ、前回の考察の通り、該当記述数が少ないものに対して、データ拡張によってデータ量を補うことで一定程度の精度向上が可能であった。一方で、最も精度が低下したのは身体構造の再現率で 0.976 から 0.942 となり、0.034 (3.4%) の精度低下が認められた。

⁶<https://taku910.github.io/mecab/> (2024/12/14 確認)

表 5: 評価結果 (データ拡張後)

分類項目	正解率	適合率	再現率	F1-score
心身機能 (b)	0.875	0.840	0.934	0.882
身体構造 (s)	0.927	0.859	0.942	0.897
活動・参加 (d)	0.906	0.818	0.759	0.786
ICD	0.962	0.821	0.741	0.779

7 課題・展望

本稿で作成した分類器によって臨床推論テキスト内に ICF の心身機能 (b), 身体構造 (s), 活動・参加 (d) ICD に該当する記述が含まれているかを判別することが可能となった。しかしながら, 今回の分類では 2 章に示した ICF や ICD の非常に浅い階層への分類を達成したのみであり, より深い階層への分類や評価点の推定には至っていない。ICF のコードは評価点があつてはじめて完全なものとなる [16], とも述べられており, 臨床推論テキストデータを活用していくためにはより深い階層のコード特定や評価点推定を実現する必要がある。その際, 深い階層へ進むにつれ分類項目が増加するため, 今回のようにそれぞれの項目に対して 2 値分類手法を用いるとモデル数が膨大になるという問題が生じる。そのため, 今後のモデル構築にはマルチレベル分類手法などの導入を検討する必要がある。また, 評価値推定に関しては, テキスト内に含まれる検査内容や検査結果, 加えてそれらの経時的変化などを捉えることで算出が可能であると考えられるが, 現状の 1 文単位という処理単位の粒度はそれに適さない可能性もある。これに対しては, Ando ら [2] や宮本ら [19] が行っているように取り扱うテキストにおける構造化の最適な粒度を設定し, 必要に応じてより詳細な処理を実現する必要があると考える。また, 本稿で取り扱った分類器や今後のモデル構築においては, 特定ドメインのテキストで事前学習されたモデルを用いることで分類精度が向上する可能性がある。日本語の医療テキストに関するものでは, Kawazoe ら [8] によって UTH-BERT モデルが公開されており, このようなモデルを用いて推定を行うことで更なる精度向上を図ることができる。

本研究では, テキスト内に含まれる情報を構造化することで臨床推論への活用を試みている。しかし, 判断に使用している非常に細かな情報や個人が持つ独自の解釈などは理学療法士によっては記載しない可能性がある。そのため, 実際に理学療法士が診療録などに記載しない情報をインタビューなどを通して明らかにし, そのような暗黙的に用いられる知見を引き揚げる仕組み作りが必要であると考えられる。

8 おわりに

本研究では, 診療録や症例レポートに含まれる自由記述文のような非構造データを効率的かつ効果的に活用するために, 国際生活機能分類 (ICF) や国際疾病分類 (ICD) といった医療分野で広く用いられている分類体系を利用し, これらのデータを構造化することで理学療法士の共通理解のもと計算機処理を可能にすることを旨とする。本稿では, その端緒として, 臨床推論テキスト 55 件の 816 文を対象として, 1 文に ICF の心身機能 (b), 身体構造 (s), 活動・参加 (d), ICD に該当する記述が含まれているか否かをそれぞれ 2 値で判定する分類器を BERT によって作成し, 計 4 つの分類器を作成した。5 分割交差検証による分類精度評価の結果, それぞれの F1-score は心身機能 (b) : 0.875, 身体構造 (s) : 0.908, 活動・参加 (d) : 0.727, ICD : 0.757 となった。その後, 分類器の精度向上を企図して, 同義語置換と ChatGPT によるテキスト修正を組み合わせたデータ拡張を行い, データ量を倍増した。データ拡張実施後, 5 分割交差検証による再評価を行い, 評価項目全 16 項目のうち, 12 項目で精度向上を達成した。今後はより深い階層への分類や評価点推定を目指す。

謝辞

本研究の実施にあたり, 吉田龍洋氏 (岸和田徳洲会病院), 畠山駿弥氏 (兵庫県立尼崎総合医療センター) の協力を得た。記して謝意を表す。

参考文献

- [1] Atkinson, H. L., Nixon-Cave, K.: A tool for clinical reasoning and reflection using the international classification of functioning, disability and health (ICF) framework and patient management model, *Physical Therapy*, Vol. 91, No. 3, pp. 416–430 (2011)
- [2] Ando, K., Okumura, T., Komachi, M., Horiguchi, H., Matsumoto, Y.: Exploring optimal granularity for extractive summarization of unstructured health records: Analysis of the largest multi-institutional archive of health records in Japan, *PLOS Digital Health*, Vol. 1, No. 9, pp. 1–19 (2022)
- [3] Aramaki, E., Kano, Y., Ohkuma, T., Morita, M.: MedNLPDoc: Japanese Shared Task for Clinical NLP, *Proceedings of the Clinical Natural Language Processing Workshop*, pp. 13–16 (2016)

- [4] Bond, F., Baldwin, T., Fothergill, R., Uchimoto, K.: Japanese SemCor: A Sense-tagged Corpus of Japanese, *Proceedings of the 6th global WordNet conference*, pp. 56–63 (2012)
- [5] Elvén, M., Boersma, K., Overmeer, T.: Assessing clinical reasoning in physical therapy: discriminative validity of the Reasoning 4 Change instrument, *Physiotherapy*, Vol. 117, pp. 8–15 (2022)
- [6] Hage, R., Fourré, A., Ramonfosse, L., Leteneur, S., Jones, M., Dierick, F.: Description and rules of a new card game to learn clinical reasoning in musculoskeletal physiotherapy, *Journal of Manual Manipulative Therapy*, Vol. 31, No. 4, pp. 287–296 (2023)
- [7] Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1, pp. 4171–4186 (2019)
- [8] Kawazoe, Y., Shibata, D., Shinohara, E., Aramaki, E., Ohe, K.: A clinical specific BERT developed using a huge Japanese clinical text corpus, *PLOS ONE*, Vol. 16, No. 11, pp. 1–11 (2021)
- [9] Sakishita, M., Kano, Y.: Inference of ICD Codes by Rule-Based Method from Medical Record in NTCIR-12 MedNLPDoc, *Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies*, pp. 80–84 (2016)
- [10] Voorhees, E. M., Hersh, W.: Overview of the TREC 2012 Medical Records Track, *21st Text REtrieval Conference, TREC 2012* (2013)
- [11] Wainwright, S. F., Shepard, F. K., Harman, B. L., Stephens J.: Factors that influence the clinical decision making of novice and experienced physical therapists, *Physical Therapy*, Vol. 91, No. 1, pp. 87–101 (2011)
- [12] Wei, J., Zou, K.: EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pp. 6382–6388 (2019)
- [13] World Health Organization: How to Use the ICF: A Practical Manual for Using the International Classification of Functioning, Disability and Health (ICF) : Exposure Draft for Comment, *WHO*, (2013)
- [14] 大島裕明, 中村聡史, 田中克己: Slothlib: webサーチ研究のためのプログラミングライブラリ, *DBSJ letters*, Vol. 6, No. 1, pp. 113–116 (2007)
- [15] 厚生労働省大臣官房統計情報部編: 疾病、傷害および死因統計分類提要 ICD-10 (2003年版) 準拠 第1巻 総論, 財団法人厚生統計協会 (2006)
- [16] 障害者福祉研究会: 国際生活機能分類 (ICF) —国際障害分類改定版—, 中央法規出版, (2002)
- [17] 日本理学療法士協会: 日本理学療法士協会ニュース (JPTA) , Vol. 323 (2020)
- [18] 畠山 駿弥, 中村 圭介, 松下 光範: 理学療法分野におけるテキスト解析のためのアノテーション基準構築と妥当性検証, 第35回兵庫県理学療法学会大会, No. P-022, pp. 40 (2024)
- [19] 宮本 誠人, 松下 光範, 高岡 良行, 堀 寛史: 理学療法初学者の支援を目的とした動作分析テキストの構造の可視化, 2022年度人工知能学会全国大会論文集, 1I1-OS-6-04 (2022)
- [20] 村瀬 雅敏: 症例報告の重要性を考える, 理学療法の歩み, Vol. 16, No. 1, pp. 5–13 (2005)
- [21] 吉田 龍洋, 畠山 駿弥, 杉本 明文, 堀 寛史, 佐々木 恭志郎, 高岡 良行, 松下 光範: 医療情報における着目点提示による理学療法初学者の臨床推論支援, 電子情報通信学会 HCG シンポジウム 2023 論文集, No. B-2-2 (2023)
- [22] 芳野 純, 二渡 玉江, 大谷 健, 白田 滋: 自立した理学療法士が獲得すべき能力に関する質的研究, 理学療法学, Vol. 37, No. 6, pp. 410–416 (2010)