

厚生労働科学研究費補助金

政策科学総合研究事業（臨床研究等 I C T 基盤構築・人工知能実装研究事業） 分担研究報告書

医療現場の A I 実装に向けた諸外国における保健医療分野の A I 開発及びその 利活用状況等についての調査研究

研究分担者 山口 類 愛知県がんセンター研究所 分野長

研究要旨

本研究の目的は、諸外国における AI 技術の保健医療分野における開発および利活用状況を調査し、日進月歩の現況の理解を進めると共に未来のトレンドの予測を目指し、我が国が抱える保健医療における課題の克服に向けた、AI 技術の開発および社会実装方策の立案に資する情報をまとめ提言することである。本年度は、機械学習、深層学習技術を含む当該分野の研究開発活動度の推移について、網羅的文献情報の探索に基づく情報の抽出と推計により、トレンドの概観を得ることを試みた。その結果、各国の研究開発状況の推移の傾向や差異を知ることができた。また文献に付与された研究内容を反映したカテゴリ情報を基に、どのような内容の研究が、当該分野で進みつつあるか概観をえることができた。また COVID-19 を契機とする研究開発の状況についても考察を行った。これらの情報は、次の保健行政の政策立案へに対して役立つことが期待される。

A. 研究目的

本研究の目的は、諸外国における人工知能技術の保健医療分野における開発および利活用状況を調査し、日進月歩の現況の理解を進めると共に未来のトレンドの予測を目指し、我が国が抱える保健医療における課題の克服に向けた、AI 技術の開発および社会実装方策の立案に資する情報をまとめ提言することである。

本年度は昨年度に引き続き、未来へ向けた技術開発および保健医療分野での AI の活用のトレンドを予測するために、直近の、諸外国における当該分野の研究開発活動度の推移について、網羅的文献情報の探索に基づく、有

用情報の抽出と推計により、上記のトレンドの概観を得ることを試みた。

B. 研究方法

本年度は、諸外国における保健医療分野における AI の開発およびその利活用状況について、主に文献情報に基づき網羅的な文献調査を行った。

調査対象とした文献の出版期間は、2015 年から 2020 年とした。昨年度の調査では 2015 年から 2019 年の間に出版されたものを対象としていた。

具体的には、近年の当該分野での各国における、研究開発状況の概観を得るために、文献データベース（Web of Science (Clarivate Analytics 社)）に対して、[“Artificial

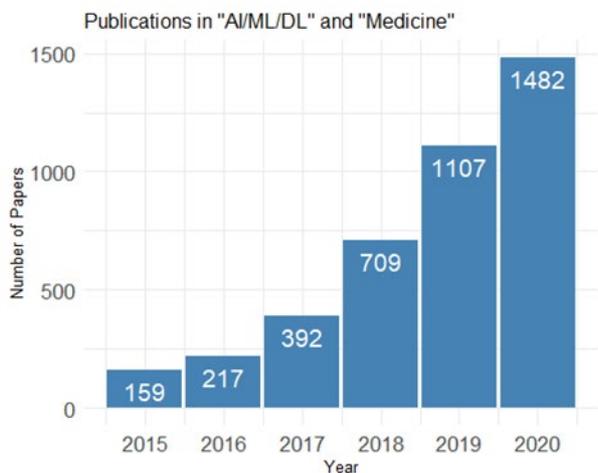


図 1 AI・ML・DL および医療関連論文の年次出版数

Intelligence” and “Medicine”]、[“Machine Learning” and “Medicine”]、または、[“Deep Learning” and “Medicine”] という組み合わせの検索ワードで文献の検索を行った。そして得られた文献の情報を解析し、いつ、どの国で行われた研究開発であるか、また、どのようなトピックにカテゴリ分けされる文献であるかの情報を抽出し、集計結果を可視化した。

注意点としては、昨年度の調査では、[“Artificial Intelligence” and “Medicine”]というキーワードの組み合わせに限定していたが、本調査では、人工知能 (AI: Artificial Intelligence)に加え、機械学習 (ML: Machine Learning)および深層学習 (DL: Deep learning)も含む、より広い観点での研究の動向を探った。また本検索は Web of Science データベース中の“Web of Science Core Collection”に含まれる文献に絞った検索であり、査読を受け出版済みの論文に集めた精度の高い結果である。一方、Web of Science データベースにおいては、上記の Core Collection 以外の Collection も含めた横断検索も可能であり、どのオプションを選ぶかによって得られる論文数も内容も変動することにも注意を要する。

以下に結果を示す。

(倫理面への配慮)

本研究の情報源は公開情報、文献情報であるため倫理面での特段の問題は無い。

C. 研究結果

まず、網羅的文献情報リストの抽出を行った。Web of Science データベースに対して、前述の三組の検索ワードの組み合わせを与え、2015年から2020年に出版された文献 (Journal paper, Review paper, Conference proceedings)の情報を含むテキストファイルを取得した。

結果、87か国で行われた4,066報の文献の情報が得られた。ここで、どの国で行われた研究開発であるかは、論文別刷り (reprint) 請求先となっている責任著者の所属機関の所在地の国名から判断した。もし一報の論文に対して、責任著者が複数存在

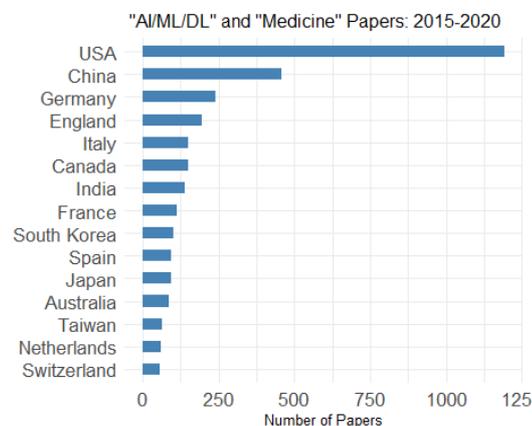


図 2 国別出版数 (2015~2020 年)。上位 15 国

し、またそれらの所属機関の国名が違う場合は、1を責任著者数で割った比率を国別に案分した。

図 1 に、世界における、医療と AI、ML、もしくは DL に関わる文献の出版数について、年次別に国を区別せずに集計した結果を示す。2015年から2020年にかけて、毎年、前年比で1.3倍から1.8倍の増加を示している。結果、2015年に比べて2020年には9.3倍の出版数となっており。当該分野の研究開発の盛り上がりが見て取れる。

図 2 に、対象期間中の国別の出版数を、出版数の多かった15か国について示す。一見してわかるように、米国(1194報;30%)、中国(456報;11.5%)の二か国が突出して、出版数が多いことがわかる。あとは10%以下の国が続き、日本(95報;2.3%)は、十位となっており、上位二か国との差は大きい。昨年度行った、[“Artificial Intelligence” and “Medicine”]に絞った文献検索では、日本は三位であったが、上位二

か国が突出している傾向に変わりはない。

図3は、国別（上位10か国）の出版数時系列である。上位二か国は、出版数その物

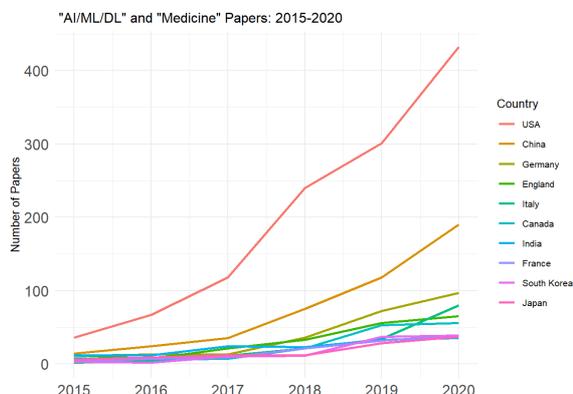


図3 国別年次出版数（2015~2020年）。上位10か国。

も多いが、近年の伸び率も大きい。2020年の出版数および前年比は、米国が、432報、1.44倍、中国が、190報、1.61倍となっている。一方、日本も、2020年の出版数37報、前年比1.32倍となっている。上位10か国の同伸び率の平均値は1.36倍、中央値は1.27倍であった。最も2020年における伸び率が大きい国は、イタリアであり2.35倍（2020年出版数：80報）であり、本分野での研究開発が活発化していることが分かる。

図4は、文献の研究内容を表現するカテゴリの集計情報の時系列を示す。各論文には、データベースにより、研究カテゴリのタグが付与されている¹。また、一つの文献に対して、複数のカテゴリが付与されていることがある。ここでは、タグ付け数の多かったカテゴリの上位12個の時系列を示している。これにより、保健医療におけるAIの活用が、どのような内容の研究においてなされているかを概観することができる。

まず上位のカテゴリには、“Computer Science”という単語を含む、カテゴリが並び、多くの研究が情報科学系の研究の文脈で行われている様子がわかる。より詳細を見ると、コンピュータ科学の中でも、“Computer Science, Artificial Intelligence”が上位に来ている。また2020年の段階では、“Medical Informatics”が最上位となっている。より具体的な医療に関わるカテゴリとして、“Radiology, Nuclear Medicine & Medical Imaging”の総数が一貫して伸びが大きい。これは、CTやMRI画像に対するDeep

Neural Networkモデルを適用した研究が盛んになっているためと思われる。また、直近の伸びが最も大きいカテゴリは、Oncology（腫瘍学）であり、2020年の前年比は2.53倍であり（平均：1.26倍、中央値：1.18倍）。がんの分野での、AI、MLおよびDLを用いた研究が活発化していることが見て取れる。一方、“Computer Science, Theory & Methods”のカテゴリが、2020年には減少傾向にあり、これはAIの理論的研究が落ち着きつつあり、それに代わって、より具体的な医療への応用が盛んになっている様子が見て取れ興味深い。

D. 考察

本研究では、直近6年間の、医療において人工知能(AI)、機械学習(ML)および深層

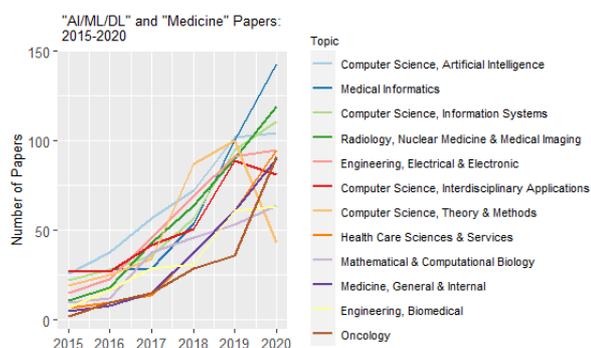


図4 AI・医療関連論文に付与された論文カテゴリワードの年次推移。上位12カテゴリ

学習(DL)を対象とした網羅の文献情報の解析により、人工知能を活用した医療分野における研究の開発のトレンドと現状の概観を試みた。

まず当該分野の論文の出版数の集計の推移について検討した。出版数は研究の活動度を反映する一つの指標であると考えられることから、各国の出版数とその推移の傾向を比較することで、当該分野における研究開発活動度の比較をすることができる。その結果、特に米国および中国の活動度が突出して高いことが分かり、またその順位関係は現状変化の兆しは見当たらない。これは、昨年度に行った、医療および人工知能にキーワードを絞った解析と同様の結果であった。一方、昨年度の解析では、日本も、全体の三位につけていたが、今回の解析では十位となっていた。これは、本調査では文献の検索キーワードを広げたことの影響や、検索データベースを“Web of Science Core Collection”に絞ったことが考

えられるが、全体の傾向としては、上位二か国が突出した状況にあることに変わりはない。しかしその中でも、イタリアのように大きく伸びを示している国もあり、他国の研究活動度の比較の結果を元に、各国の、医療・保健政策、産業育成政策、産官民の連携構造を参考にし、日本における保健行政を考える必要がある。

ここで、本研究の限界を指摘しておきたい。まず本研究では、出版済みの文献の情報のみを用いていることにある。しかし、近年の研究発表方法の動向として、論文が査読を経て出版される前に、草稿（プレプリント）を arXiv や、bioRxiv 等のプレプリントサーバ上で公開することが多い。機械学習の分野では、研究の進展が早く、特にその傾向が顕著であり、新しいアルゴリズムおよび解析手法の研究は出版前の論文を参考に進んでいくことも通常である。故に、より正確な研究開発の動向をつかむためには、上記のプレプリントサーバの情報も加味する必要がある。しかしプレプリントサーバも複数あり、また玉石混淆の出版前の草稿であり、取り扱いが難しく、本研究では除外した。

また COVID-19 関連の研究開発についても触れておきたい。本年度は、世界中が COVID-19 パンデミックに襲われるという未曾有の状況であった。その中で、画像診断をはじめ、様々な研究開発が進められた。今回抽出した文献の中にも、40 報を超える COVID-19 をキーワードに含む論文が含まれている (Alsharif et al., 2020; PMID: 33275275 等)。日々状況が変化する中で研究の速報性がかつて無いほど求められるようになり、結果、査読前のものも含めて大量の文献が産生されている。論文の大波に研究者が飲まれる様子が *Science* 誌の記事でも取り上げられている²。これらの COVID-19 関連の論文は、その緊急性を鑑みて本文も含めてデジタル化されたものが、データベース (CORD-19)³ として公開されている。2020 年 12 月 24 日の段階で、395,751 報の論文 (プレプリントを含む) が収録されており、2020 年だけでも 300,227 報の論文が収録されている。その中で、AI、ML、DL をタイトルもしくは要旨に含む論文数は、5,018 報 (1.6%) であった。COVID-19 を契機に、AI 技術に代表される情報科学技術を駆使した、医療機器やソフトウェアの開発が加速するものと思われる。

また、これらの大量の文献から、有用な情報を抽出するために、自然言語処理技術を含む AI 技術を活用する試みも進められている⁴。大量の文献を AI に読ませ、知識化す

る試みは、がんの分野では IBM Watson for Genomics 等で実用化されていたが、大量の文献情報が本文データも含めて公開されることにより、新たな技術開発および製品開発につながることを期待される。

E. 結論

本研究では、直近 6 年間の、医療および人工知能を対象とした網羅的文献情報の解析により、人工知能を活用した医療分野における研究開発のトレンドと現状を概観した。その結果、各国の研究のアクティビティ、特に米国および中国の活動度が突出して高いことが分かり、またその順位関係は現状変化の兆しは見当たらない。また当該分野においてどのような内容の研究活動がなされているかを、文献に付与されたカテゴリタグの集計により、その推移を概観した。更に COVID-19 を契機に急速に進展する AI 技術を含む研究の動向と、それに伴って産生される大量の文献から有用な情報抽出を行うための技術開発の動向について述べた。

今後、研究のトレンドの予測と諸外国の活動度の推移を元に、その背後にどのような、各国の保健行政、産業育成政策、産官民連携体制があるかを調べ、今後の政策立案等に役立てる必要がある。

F. 研究発表

1. 論文発表

Moriyama T, Imoto S, Miyano S, Yamaguchi R. Theoretical Foundation of the Performance of Phylogeny-Based Somatic Variant Detection. *Lecture Notes in Computer Science*. 2020;12508:87-101.

2. 学会発表

Moriyama T, Imoto S, Miyano S, Yamaguchi R. Theoretical Foundation of the Performance of Phylogeny-Based Somatic Variant Detection. ISMCO 2020 2020 年 10 月 8 日 オンライン開催 (ポスター)

Yamaguchi R.

Data Science and Artificial Intelligence Toward Genomic Precision Medicine. The Power of Data Science to Accelerate Health Medical Research 2021 年 3 月 26 日 オンライン開催 (口頭)

G. 知的財産権の出願・登録状況

1. 特許取得
なし。

2. 実用新案登録
なし。

3. その他
なし。

参考 URL

1. https://images.webofknowledge.com/images/help/WOS/hp_subject_category_terms_tasca.html
2. <https://www.sciencemag.org/news/2020/05/scientists-are-drowning-covid-19-papers-can-new-tools-keep-them-afloat>
3. <https://www.semanticscholar.org/cord19>
4. <https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge/tasks>