

早稲田大学大学院 先進理工学研究科

# 博士論文概要

## 論文題目

日米における人工知能を用いたコンピュータ支援検出/診断システムの法規制及び性能評価方法に関する研究

Study on the Regulation and Performance Evaluation Method of the AI-based Computer-aided Detection/Diagnosis Systems in Japan and the United States

申請者

弓場 充  
Mitsuru YUBA

共同先端生命医科学専攻 循環器医工学研究

2022年9月

本研究の目的は、近年開発が盛んな人工知能（Artificial Intelligence: AI）/機械学習（Machine Learning: ML）を用いた医療機器の日米における規制の現状および求められる性能評価方法を調査・分析することにより、AI医療機器の開発促進に資する知見を得ることである。本研究は6章で構成されている。

第1章では、AIおよびAIを用いた医療機器開発の歴史について述べた上で、深層学習（Deep Learning: DL）の登場によって実用化が加速するAI医療機器に対する期待及び規制の課題について調査した。AIは1956年に開催されたダートマス会議において提唱された概念であり、「人間の知能と同じ特徴を持つ複雑なマシンを構築すること」を目的に研究が進められてきた。その中で、AIを医療に活用した代表例が「エキスパートシステム」である。1972年に開発された「Mycin」は細菌の専門医の経験則をプログラミングしたもので、細菌感染の診断を目的としたシステムとして注目された。しかし、エキスパートシステムを構築するためには専門家の知見を手動でデータベース化していく必要があること、さらに、人間が使う曖昧な表現をエキスパートシステムでは表現できないことが課題となり、AIを活用した医療機器開発は衰退していった。その後もAI開発は栄枯盛衰を繰り返してきたが、2010年以降DLの登場により急速に発展する。DLは前述のエキスパートシステムとは異なり、ルールや特徴をAI自身が学習できるという点で革新的なアルゴリズムであり、特定のタスクに対しては既に人間を超えるような事例も報告されるようになった。医療分野においては、医師の診断または病変の検出を補助するコンピュータ診断支援装置（Computer-aided detection/diagnosis: CADe/CADx）、さらにトリアージを目的としたCAST（Computer-aided simple triage）の開発が活発化している。しかし、AI医療機器はライフサイクルが従来の医療機器と比べて短いことや使用目的が多様であることを踏まえ、各規制当局が新たな医療機器として規制改革を推進している段階であり、開発者にとって予見性の高い規制や性能評価方法の確立が急務である。

第2章では、日米におけるAI医療機器規制の現状分析を目的として、厚生労働省が発出している18件の通知およびThe Food and Drug Administration（FDA）が公開している21件のガイダンスを調査した。米国では2016年に成立したThe 21<sup>st</sup> Century Cures Act以降、Digital Healthに関するガイダンスを12件公開しており、ソフトウェアの該当性判断に関するものが6件、性能評価に関するものが4件であった。性能評価についてはソフトウェア単体で評価する場合とAIによる補助の有無が使用者に与える影響を評価するReader Studyを実施する場合に留意すべき点を示したガイダンスも公表されていた。また、2件のガイダンスにおいてCADe/CADxだけでなくCASTについても言及されており、米国ではCASTの機能を有する製品も医療機器として扱われていた。一方、日本では2014年の薬事法改正により、ソフトウェアが単体で医療機器として扱われるよう

になったことを契機に規制整備が行われ、ソフトウェアの医療機器該当性に関するものが4件、性能評価に関するものが1件であった。CADe/CADxは医療機器に該当すると明記されていたが、CASTについては言及されていなかった。以上より、日本では現状、CASTに分類されるような医師に先んじて出力結果を出す製品の取扱いについては、CADが「第二の意見」として位置づけられている関係上、成立しえないものと考えられた。

第3章では、個人情報保護法および次世代医療基盤法を調査し、AI医療機器開発に必要な患者情報収集に関する規制の現状を分析した。個人情報保護法は2005年に施行された法律であり、事業者が取得した個人情報の取扱いに関するルールを定めた法律である。当該法律は2017年に改正され、病歴などの医療情報が「要配慮個人情報」となった。これに伴い、匿名加工情報にしなければオプトアウト方式で第三者に情報提供することができなくなった。しかし、医療機関における医療情報の匿名化は現実には困難であり、医学研究や医薬品医療機器開発の発展に悪影響を与える恐れがあった。そこでオプトアウト方式で、医療機関が医療情報を「認定匿名加工医療情報作成事業者」に提供できるようにしたのが2017年に施行された次世代医療基盤法である。次世代医療基盤法によるメリットは、①認定業者が匿名加工の責任を負うこと（医療機関に責任はない）、②多数の医療機関の情報を収集できること（個別医療機関単位の匿名化は不要）、③国が匿名加工の能力を有する事業者を認定すること（医療機関が業者を選定しなくてよい）である。以上より患者情報の第三者機関への提供が次世代医療基盤法によりオプトアウト方式でできるようになった。しかし、AI医療機器の審査において論点となる教師データのアノテーションの完全性、アノテーターの妥当性、データの網羅性などの信頼性担保の方法については未だ合理的な対応方法は定められておらず、開発を促進するために今後取り組むべき課題である。

第4章では、AIを応用したCAD製品に対してどのような評価が行われているかを明らかにするため、日米において承認されたAI CAD製品を網羅的に抽出し、市販前の性能評価方法を調査・分析した。日本医療機器センターが公開しているデータベースを対象に日本で承認されたAI CAD製品を抽出し、12製品を分析対象とした。米国において承認されたAI CAD製品の抽出はFDAが公開しているプロダクトコードで抽出し、45製品を分析対象とした。米国で前向き試験を実施して承認された3製品を除くすべての製品（54/57）が、後ろ向きで評価されていた。日本では、承認された製品の内、半数の6製品が大腸内視鏡関連であり、米国と比較しても開発が先行している領域であることが明らかとなった。さらに、日本では大腸ポリープを検出するCADe製品の性能評価に、画像ではなく動画が用いられており、実臨床を踏まえた評価が受け入れられていた。したがって、後ろ向きの評価であるが合理的な審査がされていた。一方、米国ではCADe/CADx

だけでなく 31 製品の CAST 製品が承認されており、日本と比較して多様な製品が開発されていた。さらに、AI CAD 製品の評価方法として、ソフトウェア単体での評価と Reader Study の実施の有無について分析した結果、米国では明確に CAST 製品はソフトウェア単体での評価、CADe/CADx 製品は Reader Study での評価と明確に分類されていたが、日本では CAST 製品が存在せず、明確な判断基準を設けていないことが示唆された。したがって、本論文の第 2 章で考察した CAST が医療機器として扱われていない現状を裏付けるものであった。

第 5 章では、市販後の改良に対して求められる性能評価方法を分析することを目的として、第 4 章で抽出した米国で承認された AI CAD 製品の内、市販後に改良した 7 件の再申請における性能評価方法を分析した。2018 年 10 月に承認された脳内出血を検出する Accipiolx (MaxQ-AI Ltd.) は、2020 年 8 月に ML ベースから DL ベースのアルゴリズムへの変更に伴う申請をしていた。当該変更により、Accipiolx の感度は 92% から 97% に、特異度は 86% から 93% に向上していた。同じく脳内出血を検出する Viz ICH (Viz ai Inc.) は、2020 年 3 月に承認取得後、2021 年 3 月に GE 社の CT を解析可能なモダリティーとして追加するための申請をしていた。当該変更により、感度は 93% から 95% に、特異度は 90% から 96% に向上していた。両製品ともソフトウェア単体での評価で承認されていた。乳がんの検出/診断を目的とした Transpara V1.3 (ScreenOint Neducak BV) は、2018 年 11 月に承認取得後、V1.5 (2019 年 12 月)、V1.6 (2020 年 3 月)、V1.7 (2021 年 6 月) と数回の改良を実施しその都度申請を行っていた。V1.3 から V1.5 への変更は、併用可能なモダリティーとして FUJIFILM 社製の Mammogram を追加するもので、初回申請時には Reader Study を実施していたが、ソフトウェア単体での同等性評価で承認を取得していた。V1.5 から V1.6 への変更は、Mammogram だけでなく Digital Breast Tomosynthesis (DBT) を解析可能なモダリティーとして追加するものであり、Reader Study を実施して承認を取得していた。V1.6 から V1.7 への変更は、併用可能なモダリティーとして FUJIFILM 社製の DBT を追加するもので、ソフトウェア単体での同等性評価で承認されていた。以上より、米国では使用目的に影響を及ぼさない改良であれば、同等性評価としてソフトウェア単体での評価で承認されていることが明らかとなった。

第 6 章では、本研究を総括した。本研究は、(1) 本邦において CAST に分類される AI 医療機器が承認されておらず、製品の多様性が損なわれていること、(2) 本邦において AI CAD 製品の性能評価方法としてソフトウェア単体での評価または Reader Study による評価が必要か判断基準を明確化する必要があること、(3) 同じ疾患を対象とした製品であっても、日米で臨床的な位置づけが異なる場合は、性能評価方法に違いがあることを明らかにした。本研究で得られた知見が、AI 医療機器における規制の最適化および開発促進に寄与することが期待される。

## 早稲田大学 博士（生命医科学） 学位申請 研究業績書

氏名： 弓場 充 印

(2023年 2月 現在)

| 種類別 | 題名、 発表・発行掲載誌名、 発表・発行年月、 連名者（申請者含む）   |
|-----|--|
| ○論文 | 1. Systematic analysis of the test design and performance of AI/ML-based approved medical devices for triage/detection/diagnosis in the United States and Japan, Scientific Reports, 2022;12(1):16874. Epub 2022/10/08. doi: 10.1038/s41598-022-21426-7, <b>Mitsuru Yuba</b> , Kiyotaka Iwasaki. |
| 発表  | 1. 心外膜下脂肪量および脂肪肝を用いた虚血性心疾患診断に用いる機械学習モデルの構築に関する基礎的研究、第60回日本人工臓器学会大会、2022年11月、 <b>弓場充</b> 、山本匡、岩崎清隆  |
| 発表  | 2. 米国におけるAI/ML-based CADの市販後改良に対する評価方法の特徴に関する研究、第12回レギュラトリーサイエンス学会学術大会、2022年9月、 <b>弓場充</b> 、岩崎清隆   |
| 発表  | 3. 米国における市販後に改良を実施したAI/ML-based CAD製品の評価方法に関する研究、日本生体医工学会専門別研究会、第24回医療機器に関するレギュラトリーサイエンス研究会、2022年8月、 <b>弓場充</b> 、岩崎清隆  |
| 発表  | 4. Construction of Machine Learning Model for Diagnosis of Ischemic Heart Disease Using Epicardial Adipose Tissue Volume, IUPESM WC2022, June 2022, <b>Mitsuru Yuba</b> , Tadashi Yamamoto, Kiyotaka Iwasaki.  |
| 発表  | 5. 心外膜下脂肪組織量を用いた虚血性心疾患診断に用いる機械学習モデルの構築に関する基礎的研究、第59回日本人工臓器学会大会、2021年11月、 <b>弓場充</b> 、山本匡、岩崎清隆  |
| 発表  | 6. 米国におけるTriage/Notification/Detection/Diagnosisを目的としたAI/ML医療機器の性能評価試験の特徴に関する研究、第11回レギュラトリーサイエンス学会学術大会、2021年9月、 <b>弓場充</b> 、岩崎清隆   |
| 発表  | 7. 大腸ポリープ検出用AI CAD製品の性能評価に関する日米比較、日本生体医工学会専門別研究会、第22回医療機器に関するレギュラトリーサイエンス研究会、2021年8月、 <b>弓場充</b> 、岩崎清隆   |
| 発表  | 8. ソフトウェア医療機器の医療機器該当性に関する日米比較、日本生体医工学会専門別研究会、第21回医療機器に関するレギュラトリーサイエンス研究会、2021年2月、 <b>弓場充</b> 、岩崎清隆   |