

共同企画

共同企画5

日本医学放射線学会／日本放射線技術学会：臨床研究における画像データの取り扱いについて

2019年11月22日(金) 14:50～16:50 C会場 (国際会議場 2階国際会議室)

[2-C-2] 臨床研究における画像データの取り扱いについて

奥田 保男¹ (1. 国立研究開発法人量子科学技術研究開発機構 放射線医学総合研究所)

キーワード：Artificial Intelligence, DICOM, De-identification

電子カルテなどの診療情報の電子化により、ここ数年、医療分野では急速に「AI」という言葉が使われるようになった。厚生労働省の「保健医療分野におけるAI活用推進懇談会」では、AI開発を進める重点6領域として、ゲノム医療、画像診断支援、診断・治療支援、医薬品開発、介護・認知症、手術支援が挙げられている。画像診断支援はAIの実用化が比較的早いと考えられる領域であると報告書にまとめられている。画像診断支援領域では、画像情報だけでなく臨床情報や他検査などの情報が必要であり、より精度の高いラーニングをするためには大量の良質な信頼性の高いデータが求められている。しかしながら、現場から発生している情報は良質なデータでない場合がある。画像データはDICOM規格が普及しており、タグ情報についても規格化が行われているが、施設やメーカーごとの運用によってばらつきが生じてしまう。そこで、画像データや所見データにおける現状とその問題点について報告したいと考えている。また、医療機関内にて研究などの二次利用目的で臨床データを収集する場合は大きな問題は生じなかったが、今後は学術機関などが臨床データを収集する場合など、様々な点において注意する必要がある。医療機関におけるデータ抽出と匿名化、データを送信するまでの安全かつ適切なガイドラインについて紹介する。本企画では現状の問題点を整理して今後AIを利用していくために必要な技術的要件について議論したいと考えている。

臨床研究における画像データの取り扱いについて

木村 通男^{*1}、坂本 博^{*2}、谷川原 綾子^{*3}、
川眞田 実^{*4}、奥田 保男^{*5}

*1 浜松医科大学、*2 東北大学病院、*3 北海道科学大学、
*4 大阪国際がんセンター、*5 量子科学技術研究開発機構

Handling of Medical Imaging in Clinical Research

Kimura Michio^{*1}, Sakamoto Hiroshi^{*2}, Yagahara Ayako^{*3}, Kawamata Minoru^{*4}, Okuda Yasuo^{*5}

*1 Hamamatsu University hospital, *2 Tohoku University Hospital, *3 Hokkaido University of Science,

*4 Osaka International Cancer Institute, *5 National Institutes for Quantum and Radiological Science and Technology

In recent years, “Artificial Intelligence” has been rapidly used in the medical field due to the influence of electronic medical information such as electronic medical records. The image diagnosis support area requires not only image information but also clinical information, and a large amount of data and highly reliable data are required to obtain a more accurate learning result. However, the information generated from the clinical practice may not be good quality data. The DICOM standard is widely used for image data, and tag information is also standardized. However, variations occur depending on the operation of each facility or manufacturer. Therefore, We would like to discuss the current status and problems of image data and findings data in clinical research. In addition, we will introduce data extraction and de-identification in medical institutions, and safe and appropriate guidelines for data transmission.

Keywords: Artificial Intelligence, DICOM, De-identification

1. はじめに

2017年のLancetにおいて、保健医療の質の高さやアクセスの容易さを世界195カ国で調査した結果が掲載されている。我が国におけるランキングは11位であり、アメリカは35位、イギリスは30位である。また、昨今は医療ツーリズムとして海外から多くの患者が訪れている状況をふまえると、我が国は非常に高い医療水準であることがわかる。これらの高い医療水準を今後も担保していくためには、いかに正確なデータを蓄積し、利活用できる環境を整えるかが重要である。

医用画像分野では世界的なデファクトスタンダードであるDICOM規格と標準規格の実装ガイドであるIHE(Integrating the Healthcare Enterprise)が普及している。これらは厚生労働省が指定している標準規格であり、広く臨床に用いられており、医療機関の業務効率化のために貢献してきた。特に、患者紹介の際に診療情報を提供するIHE統合プロフィールは現代医療において不可欠なものとなっている。また、2020年4月より診療用放射線にかかる安全管理体制に関する規定が改正されることから、線量情報管理を担うオブジェクトであるDICOM規格であるRDSRが注目されている。情報の1次利用の観点から臨床現場において必要不可欠であり、我が国の医療水準を担保するためのものであることがわかる。2次利用においては業務改善や医療安全などの様々な分野に用いられている。また、臨床研究としても2次利用されるケースが多く見られる。臨床研究には単施設研究が中心であったが、多施設共同研究も増加傾向にある。医療分野研究開発推進計画²⁾(2017年2月健康・医療戦略本部決定)では、「医療・介護等のデジタルデータの利活用基盤の構築を進めるとともに、革新的な人工知能の基盤技術を構築し、収集されたビッグデータを基に人工知能技術を活用することで、診療支援や新たな医薬品・医療技術の創出に資する研究開発を進める。」とされている。

医用画像領域の人工知能に関する研究は画像認識(自動認識)や読影レポートの作成補助など多岐にわたり進められている。国立研究開発法人日本医療研究開発機構(AMED)の臨床研究等ICT基盤構築・人工知能実装研究事業では、

医療研究開発のために2016度から医療画像データベースを学会主導で構築する事業を支援しており、医学放射線学会を含む6学会が画像データとメタ情報としての臨床データの収集を多施設から行っている。人工知能に関する研究は国レベル・学会レベルで取り組みが行われている。

こういった背景のもと、個人情報保護法改正をはじめ、臨床研究法や次世代医療基盤法など様々な法整備が進められている。

2. データ収集時における注意点

一般的に臨床研究用の元データは病院情報システムに保管されている。保管されているシステムからデータを抽出し、必要な変換と匿名化を行なった上で収集するのが一般的である。単施設研究であれば匿名化したデータを用いて研究を行うが、多施設共同研究の場合はオフラインもしくはオンラインにてデータを収集した上で研究を行う。

2.1 匿名化について

2017年に全面施行された改正個人情報保護法では、個人識別符号が明確となり、個人情報の定義が明確化された。また、病歴情報は要配慮個人情報に指定され臨床研究に用いる画像データにおいては慎重な取り扱いが求められている。研究などのデータ利活用とデータの安全管理を両立させるための対策として「匿名化」がある。「匿名化」とは完全に誰かわからなくする意味での匿名化が代表例であるが、「人を対象とする医学系研究に関する倫理指針ガイダンス」によれば、氏名・顔画像・保険証番号などの個人識別番号などの単体記述で個人が識別できる情報の削除により匿名化と定義されている。検査データや画像データなどについては、特異的な記述であっても、その記述単体で直ちに判別できる記述には該当しないと記載がある(第5章第12(3)アウ)また、個人情報保護法では個人を識別することができないように個人情報を加工して得られる情報であって、当該個人情報を復元できない匿名加工情報が定義されている。こういった状況からも、匿名化の定義が曖昧であることがわかる。しかしながら匿名加工情報への加工や非個人情報への匿名化処理は医療施設

が責任を持って実施する必要がある。

また、DICOM 規格における匿名化については、DICOM 規格書 15 巻「セキュリティおよびシステム管理プロファイル」の付属書 3)に、教育用データ作成などを目的とした匿名化処理として推奨している基本アプリケーションレベル機密性プロファイルをベースとして、画像の加工や利用目的を考慮して特定の情報削除あるいは利用目的を考慮して特定の情報を保持する 12 の基本アプリケーションレベルオプションを定義している。匿名化を行う際に参考にできる。また、Supplement 142 ではそれぞれのタグ情報のアプリケーションレベルオプションについて定義されている。(図1)

属性	タグ値	リタイア	処理方法
患者のID	(0010,0020)		Z
患者の名前	(0010,0010)		Z
患者の生年月日	(0010,0030)		Z
患者の年齢	(0010,1010)		X
患者の性別	(0010,0040)		Z
検査日付	(0008,0020)		Z
実行番号	(0008,0050)		Z
検査ID	(0008,0010)		Z
検査内容	(0008,1030)		Z
検査コメント	(0010,4000)		Z
検査コメント	(0032,4000)		Z
検査コメント	(0018,4000)		Z
オーバーレイデータ	(000a,3000)		Z
施設名	(0008,0080)		Z
装置名前	(0008,1010)		Z
操作者名	(0008,1070)		Z
検査インスタンスUID	(0020,000D)		U
SOP インスタンスUID	(0008,0018)		U
プライベート属性	グループ番号が奇数		X

15巻付属書Eの表E.1-1から、主な属性データ要素の処理方法の抜粋
「医療情報利活用における匿名化技術ガイド」
(JIRA画像医療IT産業推進WG SWG3)

図1 DICOM 規格における匿名化

2.2 データ収集について

単施設研究では画像データを第三者へ提供することはないが、多施設共同研究などを行う場合はデータ提供の必要性が生じてくる。2017 年の個人情報保護法改定により、病歴情報は全て要配慮個人情報となり、オプトアウト手続きによる第三者提供は認められていない。しかしながら、人を対象とする医学研究系研究に関する倫理指針(第5章 第12の7)4)では「社会的に重要性の高い研究として認められる」ものに関してはオプトアウトでの収集が可能であるとされている。これにより、先述した AMED が中心となって進めている広域な医用画像データベースを構築する事業が可能となる。その一方で、医療情報部などの情報管理部門による情報漏洩の監督も必要であることから、医療情報学会ではAMEDの委託研究において「診療画像など医療情報抽出・管理のあり方に関する研究」を担務している。

施設内に匿名化機能やデータ収集機能などを持ち合わせたキュレーションサーバーを設置し、データを外部のストレージサーバーにオンラインで送信する手法がある。また、病院情報システムネットワークを物理的に外部ネットワークから切り離している場合は厚生労働省標準規格に指定されているIHE PDIにて出力を行い、図2に示す netPDI という仕組みを用いてネットワーク経由で送付できる。医療機関はデータを netPDI サービスセンターに送信し、バーコードの印刷されたトークンシートをデータセンターに送付する。データセンターは受領したトークン票を用いて当該データをストレージサーバにダウンロードすれば良い。病院情報システムのネットワークが外部接続されていない場合は、データを一旦 USB などの物理媒体に移し、人手でアップロード装置にコピーしてからサービスセンターに送付すれば良い。

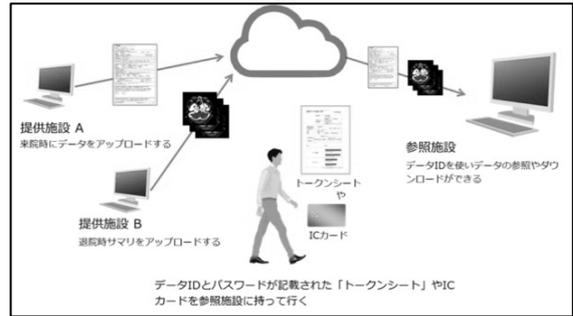


図2 net PDI の仕組み

3.今後の研究利用について

人工知能に関する研究は Deep learning の導入以来、構文解析、機械翻訳、要約など、様々な分野において進歩を遂げてきた。医療分野においても、副作用などの有害事象抽出や、症例検索が行われており、人工知能を活用した先制医療や予防医学の取り組みが行われている。画像領域では画像診断支援に用いられる物体検出技術、放射線治療などにおける自動輪郭抽出技術、MR 画像から CT 画像を生成するという画像生成技術、息止め不良画像を検出するクラス分類といった様々な研究が行われている。しかしながら、生成する画像種別は主に TIFF 画像であることから臨床応用への事例は非常に少ない。また、AI 研究を成功させるための鍵はできる限り多くの良質なデータが必要である。基本的には画像データそのものから AI 技術解析はできないことから適切なアノテーションづけしてはじめて有用なデータとして取り扱うことができる。専門医などがアノテーションづけを行うことが理想的であるが、時間的な制約により難しい状況である。今後はデータの構造化をどのようにして行うのかということも検討していくべきであると考えられる。教師データなしに学習できる敵対生成ネットワークにも注目が集まっているが、人間の検証なしに結果を鵜呑みにするのは危険である。常に人間が検証するシステムを実装していくことが望まれる。

4. おわりに

医療の質を担保していくためには医用画像領域に関する研究は非常に重要である。日本放射線医学学会 (JRS)、日本放射線技術学会 (JSRT)、日本医療情報学会 (JAMI) の共同によりデータの生成から結果まで安全に研究実践することが重要である。

参考文献

- 1) Healthcare Access and Quality Index based on mortality from causes amenable to personal health care in 195 countries and territories, 1990-2015: a novel analysis from the Global Burden of Disease Study 2015. Lancet 2017; 390: 231-266
- 2) 医療分野研究開発推進計画. 健康・医療戦略推進本部, 2017. [\[https://www.kantei.go.jp/jp/singi/kenkouiryu/senryaku/suishin_plan_henkou.pdf\]](https://www.kantei.go.jp/jp/singi/kenkouiryu/senryaku/suishin_plan_henkou.pdf) (cited 2019-Aug-25)
- 3) NEMA DICOM Standard [\[http://www.dicomstandard.org/\]](http://www.dicomstandard.org/) (cited 2019-Aug-25)
- 4) JIRA 人を対象とする医学研究系研究に関する倫理指針, 2014 [\[https://www.mhlw.go.jp/file/06-Seisakujouhou-12600000-Seisakutoukatsukan/0000168764.pdf\]](https://www.mhlw.go.jp/file/06-Seisakujouhou-12600000-Seisakutoukatsukan/0000168764.pdf) (cited 2019-Aug-25)
- 5) 厚生労働省標準規格 HS009 「IHE 統合プロファイル」可搬型医用画像」およびその運用指針 [\[http://helics.umin.ac.jp/files/HS009/HS009_rep20130705_JAMI.pdf\]](http://helics.umin.ac.jp/files/HS009/HS009_rep20130705_JAMI.pdf) (cited 2019-Aug-25)