医用画像診断支援システム

（人工知能技術を利用するものを含む）

開発ガイドライン２０１９（手引き）

令和元年１２月

経済産業省／国立研究開発法人日本医療研究開発

目次

1. 本ガイドライン（手引き）の目的 1
2. 用語の定義と適用範囲 2
2.1 用語定義 2
2.2 適用範囲及び想定読者 4
3. CAD の設計・開発管理 5
3.1 設計・開発管理が要求する項目 5
3.2 適用すべき規格・基準 7
4. 設計・開発管理が要求する項目の各プロセスへの適用方法 7
4.1 ソフトウェア安全クラス分類 7
4.2 設計開発プロセスへの適用 7
5. 設計開発・保守プロセスの実施 14
5.1 設計開発プロセスの実施 14
5.2 保守プロセスの実施 14
5.3 設計開発・保守プロセスにおけるデータの管理 14
6. 評価手法及び留意事項 15
6.1 性能評価に影響するデータの特性 15
6.2 性能評価の基準 16
7. 開発に関するその他の留意事項 18
8. 関連する規格及び参考資料 19
APPENDIX 21
Appendix 1. 旧版の Appendix から本ガイドラインに引き継がれなかった項目 21
Appendix 2. 性能評価法の基礎 22
Appendix 3. テストデータにおける画像データベースとゴールド・スタンダード 24
Appendix 4. 性能評価のための読影実験における注意点 26
Appendix 5. 性能評価手法 27
Appendix 6. 観察者特性の評価 29
Appendix 7. テストデータの量 31
Appendix 8. 市販後学習による性能変化の課題 32
Appendix 9. 個人情報保護法概要と対応 33
Appendix 10. 次世代医療基盤法・臨床研究法の概要と対応 35

医用画像診断支援システム（人工知能技術を利用するものを含む）
開発ガイドライン２０１９（手引き）

1. 本ガイドライン（手引き）の目的

本ガイドライン（手引き）は、医用画像診断支援システム（Computer Aided Diagnosis:以下CAD）の研究・開発者を対象に、その開発時のポイントをまとめたものである。本ガイドラインは 2010 年代前半に発行された CAD に関する開発ガイドラインの改訂版であり、既存の二つの CAD 開発ガイドラインを一つにまとめ、人工知能技術（以下、AI 技術）に関する記述を強化した。

従前の開発ガイドラインは、開発者の視点、特に単体ソフトウェアの医療機器化に際して新規参入する者も含んだ視点から読みやすいように記載されていたが、今回の開発ガイドラインは拡大・加速する AI 技術の応用が期待される CAD の開発を活性化することを目指した。特にこれまでの開発ガイドラインでは想定していなかった、市販後学習による性能変化が見込まれる CAD については、その注意点について慎重に議論をして、とりまとめている。

その他、本ガイドラインの策定作業では、AI 技術による大量データの学習、臨床に於ける様々な分野での応用研究開発が進められている現状に鑑み、これまでの主として放射線分野を想定した CAD に限定せず、内視鏡・超音波分野での運用も含めて検討を行った。更に、今回対象に含めなかった分野についても、その特性に鑑みた開発ガイドラインを今後策定していく予定である。

AI 技術の利用については法的・制度的な整備が進められている途上である。読者は法的・制度面の整備に留意することが望まれる。例えば、AI 技術の医療応用と責任の所在に関しては、「AI を用いた診断・治療支援を行うプログラムを利用して診療を行う場合についても、診断、治療等を行う主体は医師であり、医師はその最終的な判断の責任を負うこととなり、当該診療は医師法（昭和 23 年法律第 201 号）第 17 条の医業として行われるものであるので、十分ご留意をいただきたい。」旨が示されている[1]。

また、研究や開発に用いる臨床画像データの取扱いに関連する種々の新法、法改正（個人情報保護法の改正、次世代医療基盤法や臨床研究法）が施行され、データの取扱いについて適切な対応が必要になっている。これらの法の主旨と、研究・開発時の留意点を Appendix 9 及び 10 に記載しているので、研究・開発が適切に実施されるよう検討されたい。

なお、AI 技術を用いた CAD の開発に本ガイドラインを適用した事例を平成 30 年度報告書に収めている[2]。また、CAD の製造販売承認申請に当たっては、「コンピュータ診断支援装置に関する評価指標」（薬食機発 1207 第 1 号別添 3 平成 23 年 12 月 7 日）を、AI 技術を応用した CAD については「人工知能技術を利用した医用画像診断支援システムに関する評価

[1] 厚生労働省医政局医事課長通知「人工知能（AI）を用いた診断、治療等の支援を行うプログラムの利用と医師法第 17 条の規定との関係について」（医政医発 1219 第 1 号平成 30 年 12 月 19 日）
[2] https://www.aist.go.jp/pdf/aist_j/iryoukiki/2018/techrep_AI_fy2018.pdf

指標」（薬生機審発 0523 第 2 号別添 4 令和元年 5 月 23 日、以下、「AI-CAD 評価指標」と称す）を参考にされたい。

上述のとおり、本ガイドラインは、以下 2 通の開発ガイドラインを統合・整理の上、AI 技術に関する事項を追加したものである。

1) コンピュータ診断支援装置におけるソフトウェア設計・開発管理開発ガイドライン 2012（以下、CAD 開発ガイドライン 2012）
2) コンピュータ診断支援装置の性能評価開発ガイドライン 2015（以下、CAD 開発ガイドライン 2015）

改訂のポイントを以下に示す。

- AI 技術の発達に対応した。
- 放射線画像以外の画像モダリティ（内視鏡画像、超音波画像を含む）への CAD の適用範囲を拡大した。
- 平成 29 年の改正個人情報保護法施行、次世代医療基盤法、臨床研究法の制定等個人情報の扱いについて Appendix 9 及び 10 に含めた。
- 平成 27 年の改正医薬品医療機器等法施行を受け、医療機器プログラムとしての CAD を中心に記述を変更したほか、適用すべき規格類を更新した。
- その他、記載全体を見直して強制事項を含まないガイダンスであることを明確化した。

2. 用語の定義と適用範囲

2.1 用語定義

1) 医用画像診断支援システム（Computer Aided Diagnosis, CAD）

臨床現場において、医師が様々な画像撮影装置を用いて画像診断（読影）を行う際、参考となる情報を提供するソフトウェア又はシステムであり、CADe や CADx の機能を持つ。支援の方式としては使用者（CAD を実際に臨床現場で使用する医師等）に対する支援形式に応じ、first reader、second reader、concurrent reader 等が存在する。

（出典：AI-CAD 評価指標を一部改変）

1-1) CADe （Computer-Aided Detection）

画像上で病変の疑いのある部位をコンピュータが自動検出し、その位置をマーキングする機能を有する単体ソフトウェア又は当該ソフトウェアが組み込まれている装置。コンピュータにより医用画像データのみ又は医用画像データと検査データの両方を処理し、病変又は異常値の検出を支援する。

（出典：AI-CAD 評価指標）

1-2) CADx （Computer-Aided Diagnosis）

病変の疑いのある部位の検出に加え、病変候補に関する良悪性鑑別や疾病の進行度、悪性度の確率等の定量的なデータを数値やグラフ等として出力する機能を有する単体ソフトウェア又は当該ソフトウェアが組み込まれている装置。診断結果の候補やリスク評価に関する情報の提供等により診断支援を行うものを含む。

（出典：AI-CAD 評価指標を一部改変）

本ガイドラインでは、CADe 及び CADx を総称して、CAD と呼ぶ。

解説：開発ガイドライン 2015 では、CADx を次のように定義していた。

「画像を解析して内蔵する基準に基づいて病変の候補部位をコンピュータが自動的に分析し、医学的に広く臨床で用いられている診断基準に基づく質的診断に関する情報を提供するソフトウェア、あるいはそれを具備する装置。」

また開発ガイドライン 2015 の適用範囲では次のように述べていた。

「本ガイドラインでは、質的診断に関する情報として、良悪性等に関するクラス分類の結果を出力する場合を対象とする。従って、クラス分類結果を出力しないもの（たとえば、腫瘍のサイズ等の特徴量、単一あるいは複数の特徴量から導かれた悪性度や進行度等の連続値、類似症例の検索結果等を出力するだけのもの）は本ガイドラインの対象には含めない。」

2) 人工知能と機械学習

一般に AI とは、人の高度な知能によって行われている推論、学習等を模倣するコンピュータ装置あるいはソフトウェアを指す。

機械学習は、AI 技術の一分野として大きく発展した技術であり、人間が行っている学習等の機能をコンピュータ（ソフトウェア）で実現するための手法。アルゴリズムは様々であり、判別分析法等の古典的手法から、決定木、ニューラルネットワーク、サポートベクターマシン等、多数の方法が提案されている。代表的な応用としてはクラス分類や回帰があるが、これらは CAD の目的と一致していることから、多くの CAD で利用されてきている。

（出典：AI-CAD 評価指標を一部改変）

3) 深層学習

機械学習の一つであり、最近特に注目されている学習方法。多層化したニューラルネットワークを大量のデータを用いて学習させることに特徴がある。ネットワークには、階層型ネットワークのような確定的モデルと、ボルツマンマシンのような確率的モデルがある。学習アルゴリズムには、教師あり、教師なし、半教師学習以外に強化学習等もある。

（出典：AI-CAD 評価指標）

4) バリデーション[3]

機械学習アルゴリズムのハイパーパラメータ（例：サポートベクターマシンの目的関数を決定するパラメータ、深層学習のネットワークの層数、特徴マップ数、畳み込みのフィルタサイズ、学習回数等）を決定すること。

（出典：AI-CAD 評価指標を一部改変）

5) トレーニング（又は学習）

学習データにより、AI の内部状態を変化させること。教師あり学習では、学習データと正解のペアを用いて学習を行う。

注 1：本ガイドラインでは、使用者等に対する教育・訓練の意味にトレーニングの語を用いない。

注２：トレーニングは「望ましい出力を得るように（分類器を）変化させること」を指す限定的な意味、学習は知識獲得等より知的で広い概念をさす意味に区別して用いることがある。

6) テスト

トレーニングが完了した AI を、テストデータにより、性能評価すること。

テストデータは、学習データ、バリデーションデータと完全に分離することが望ましいが、開発時に必要な条件を備えたデータ数を確保できない場合は、Appendix 5 に示す評価手法を選択することができる。

2.2 適用範囲及び想定読者

本ガイドラインは、放射線画像や MRI 画像、超音波画像、内視鏡画像等の臨床画像（静止画像及び動画像）を用いる CAD を対象とする。本ガイドラインは AI 技術を利用しない CAD にも適用可能である。

一方、病理診断医の行う病理診断の支援を目的とする CAD は、確定診断に直結することが予想される等、上記の臨床画像に対する診断支援とは異なる側面を有することが否定できないことから今回の検討では対象外とした。また CAD における診断支援よりも広義の医療支援を指す Clinical and patient decision support software[4]及び computer aided triage[5]も対象外とした。これらに対しても本ガイドラインの多くの事項が適用可能であるが、臨床画像に対する CAD との相違を踏まえた検討が必要である。

---

[3] AI 技術の分野では用語「バリデーション」が JIS T82304-1:2018 が定義する用語「バリデーション」とは異なる意味で用いられてきた。このため、本ガイドラインでは、バリデーションを AI 技術分野の用語として用いるが、JIS T82304-1 の定義を置き換えるものではない。

参考：JIS Q9000:2015 3.8.13 妥当性確認（validation）

[4] Clinical and Patient Decision Support Software Draft Guidance; 2017 年 12 月米国 FDA 発行。

[5] Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data - Premarket Notification [510(k)] Submissions Guidance; 2012 年 7 月米国 FDA 発行

市販後学習により生じる性能変化に伴う品質管理を製造業者[6]が行うことが容易でない AI 技術を利用した CAD、例えば、使用者による市販後学習により施設ごとに別々の性能変化が可能となるものに関しては、薬機法上の課題等解決すべき事項が数多く存在することから、AI-CAD 評価指標においては別添に取りまとめられている。本ガイドラインでは、Appendix 8「市販後学習による性能変化の課題」において技術的課題を指摘するにとどめた。開発を行う場合にはその時点での最新の評価指標等を確認する事を推奨する。

本ガイドラインの想定する読者は、CAD を開発する者及び市販後学習を行わせる者とする。

3. CAD の設計・開発管理

3.1 設計・開発管理が要求する項目

CAD の設計・開発管理が要求する項目として以下の項目が挙げられる。

1) 開発及び性能評価に用いるデータ

CAD の設計開発プロセスで使用するデータは、開発に用いるデータ（学習データ、バリデーションデータ）と、性能評価に用いるデータ（テストデータ）に大別される。

基本的考え方

- CAD の開発及び性能評価の際には、妥当な性質並びに数量の臨床画像データの他、シミュレーションやファントム等を用いて生成された人工画像、転移学習の場合には自然画像や別の医用画像等を使用することができる。
- 性能評価を適切に行うため、開発（パラメータ調整や動作確認など）に用いるデータと性能評価に用いるテストデータは厳格に区別しなければならない。
  AI 技術を用いる CAD の場合は、テストデータを学習しないための方法又はその影響を十分小さくする方法が必要である。

2) 解析アルゴリズム・学習アルゴリズム

CAD の中核をなす解析アルゴリズムは、設計・開発管理が要求する主要な項目である。AI 技術を用いる CAD の場合､解析アルゴリズムのみならず学習アルゴリズム(教師あり、半教師あり、強化学習、自己学習、転移学習、市販後学習等）も管理すべき設計要素に含めて明確化する。学習の対象（ネットワークの結線（結合）の重み、ハイパーパラメータ等)、使用するデータ（学習データ、バリデーションデータ、テストデータ）も合わせて明確にする。

これらについては、AI-CAD 評価指標の 6.(2)「品目の検出・診断の原理、学習、情報セキュリティ等に関する基本的事項」に述べられている。

[6] この文書では「製造業者」「使用者」の用語を JIS T14971:2012 の定義に準じて用いる。「製造業者」は日本国内では薬機法の「製造販売業者」に近いが厳密には同義でない。

3) 組み合わせるハードウェアとソフトウェア

CAD は、医用画像装置や医療用画像管理システム（Picture Archiving and Communication Systems : PACS）等の他のハードウェアやソフトウェアと組み合わせて使用されることが想定される。この組合せについても考慮する。

3.2 適用すべき規格・基準

CAD に対する安全性・品質管理に関する法令・JIS 規格等の規格・基準を表 1 に示す。

表 1　CAD の安全性・品質管理に関連する基準

CAD の安全性・品質管理事項	適用すべき規格・基準
ソフトウェア安全性	JIS T82304-1:2018[7]
品質マネジメント	QMS 省令（厚生労働省令第 169 号）
リスクマネジメント	JIS T14971:2012
ソフトウェアのライフサイクルマネジメント	JIS T2304:2017

なお、適用すべき規格・基準は国際整合の進展に伴って今後も変更が予定されていることから、最新の情報を確認するべきである。

4. 設計・開発管理が要求する項目の各プロセスへの適用方法

4.1 ソフトウェア安全クラス分類

JIS T2304 に従って CAD のソフトウェア安全クラス分類を決定する。決定にあたっては、CAD が患者にもたらす危害のリスクに応じてこれを適切に決定する。クラスに応じた JIS T2304 の一連の設計開発プロセスを実行する。

解説

旧版の CAD 開発ガイドライン 2012 では、次のように記載されていた。

… 診断そのものは医師による行為であり、CAD は JIS T2304 のソフトウェア安全クラスにおけるクラスＢに相当すると考えられる。…

しかし、JIS T2304:2017（IEC 62304:2005, Amd.1:2015）の細分箇条 4.3 ソフトウェア安全クラス分類の記載が変更になったことに伴い、本ガイドラインの記載を変更した。

4.2 設計開発プロセスへの適用

設計開発のプロセスについては、JIS T2304 及び QMS 省令に規定されている。

① 製品に関連する一般要求事項の明確化（意図する用途）[8]

下記に基づいて製品に対する一般要求事項を明確化する。

[7] 専用ハードウェアと組み合わせて出荷される CAD ソフトウェアの場合は、JIS T82304 は適用されない。JIS T0601-1:2017 又は個別規格（例：超音波画像診断装置の場合は JIS T0601-2-37：2018 医用電気機器－第 2-37 部：医用超音波診断装置及びモニタ機器の基礎安全及び基本性能に関する個別要求事項）が適用される。

[8] 保守サイクルにおいては、“変更の要求事項の明確化”となる

i) 意図する使用目的、使用方法

CAD の意図する使用目的、使用方法を明確化する。この際、first reader、second reader、concurrent reader の区別を明らかにすることも重要である。

例：病変と想定される部位の抽出（CADe の場合）

例：異常部位の良悪性の鑑別（CADx の場合）

AI 技術を用いる CAD の場合、市販後学習の有無を明確化する。市販後学習を行う場合は、それに伴う性能変化の有無とタイミング、市販後学習と性能変化の目的も合わせて明確化する。

市販後学習と性能変化については、Appendix 8「市販後学習による性能変化の課題」にて解説している。

ii) 解析対象の画像、適用部位、解析対象の病変

CAD の対象とする画像の種類と仕様（画素数、階調等）、適用部位及び解析対象病変を特定する。

例１：マンモグラフィ用 CAD の場合

対象画像：画素数（4728×5928）、階調 14bits

対象部位：乳房

解析対象病変：マンモグラフィの乳がん所見のうち、腫瘤病変、石灰化病変、構築の乱れ、FAD（Focal Asymmetric Density 局所的非対称陰影）（CADe の場合）

例２：胸部 CT 用 CAD の場合

対象画像：画素数（512×512）、階調 12bit

対象部位：胸部

解析対象病変：3mm 以上の肺結節（種類：Solid、GGO、Part-Solid）

例３：腹部超音波用 CAD の場合

対象画像：画素数（512×512）、階調 12bit

対象部位：腹部臓器（肝臓、胆嚢、膵臓）

解析対象病変：肝血管腫、肝細胞癌、転移性肝癌、胆嚢ポリープ

iii) 意図する使用者

意図する使用者を明確化する。使用者には医師等の他に、医療機関に所属する CAD のメンテナンスを行うものが含まれる。

- 専門外の医師等が CAD の出力を過信することの影響を含めて検討することが望ましい。
- 市販後学習を含む CAD の場合、これを行わせる者も使用者に含まれる。

iv) 意図する使用環境

使用者がソフトウェアを利用する場所だけでなく、ソフトウェアを実行する実体の計算機の存在する環境、ネットワーク環境等を含むことができる。

市販後学習を含む CAD の場合、市販後学習を行わせる環境はソフトウェアを利用する環境と同一とは限らない。

例：放射線科読影室、診察室

例：〇〇クラウドサーバ

② 設計開発の指針とソフトウェア要求事項の明示

4.2 ①「製品に関連する一般要求事項の明確化（意図する用途）」で明確化した製品に対する一般要求事項に基づいて、設計開発の仕様書における開発の指針やソフトウェア要求事項を具体化する。以下に CAD に関係する特徴的なソフトウェア開発要求を示す。その他の事項（例：附属文書への記載要求）については、JIS T2304 の 5.1 ソフトウェア開発計画及び 5.2 ソフトウェア要求事項分析を参照する。

ソフトウェア開発の初期には、必ずしも、これらの要求事項の全てが明らかになっているとは限らない。

i) 患者画像の仕様、接続可能な画像診断装置の特定

CAD への入力源となる患者画像を提供する、CAD に接続可能な画像診断装置を特定する。

接続可能な画像診断装置への要求事項は、患者画像の具備すべき仕様（例：解像度、信号対雑音比、ダイナミックレンジ等；iii)「使用するデータの明確化」を参照）に基づいて設定する。

- CAD の接続対象が、画像・接続インターフェイス（DICOM 等）の仕様（スペック）で特定可能な場合はスペックで指定する事も可能である。
- AI-CAD 評価指標の 6.(2)「品目の検出・診断の原理、学習、情報セキュリティ等に関する基本的事項」で併用する画像撮影装置等の条件に言及している。

患者画像の具備すべき仕様は、一連の開発プロセスを経たのちに目標性能を達成するために必要事項として特定されることが多い。

ii) ソフトウェア実行環境の明確化

4.2 ①「製品に関連する一般要求事項の明確化（意図する用途）／iv) 意図する使用環境」に基づいて、ソフトウェア要求事項の具体化に必要なソフトウェア実行環境を明確化する。以下を特定する。

- 組み合わせるハードウェア環境の特定
- 組み合わせる他のソフトウェアの特定

例：最小限必要なハードウェア及びソフトウェア構成、オプションとして追加

可能な構成、ネットワーク接続、VPN 等

iii) 使用するデータの明確化

CAD の設計開発プロセスで使用するデータに対して、実装工程（AI 技術を用いる CAD の場合は学習プロセスが含まれる）及びシステム試験の詳細計画が実施可能になるように、使用するデータを特徴付ける事項を明確化して記述する。

例：

- データソースの特性、データソースの元となる患者群の持つ傾向、データの取得・管理方法、画像データ取得時の撮影パラメータ、紐付けした臨床データの種類、臨床画像データ以外のデータ使用の有無
- アノテーションの取得方法、アノテーションを行なった医師等
- テストデータ及びテストデータに関する最終判断（テストデータの選択や確定診断等）を行った医師等を開発のプロセスから完全に切り離して管理するための方策
- データソースの持つ制限（商用利用の可否、二次加工の可否等）、データソースの具備すべき法令上、倫理上の制約

データソースの法令上、倫理上の制約については、我が国では個人情報保護法、次世代医療基盤法、臨床研究法の規定に留意する（Appendix 9, 10 参照）。さらに海外で取得するデータを使用する際は、それぞれの国・地域での類似の法令に留意する。例えば、米国の HIPAA 及び HITECH といった連邦法、EU における一般データ保護規則(GDPR)である。

なお、使用するデータの具備すべき特徴は、6.1「開発及び性能評価に用いるデータ」、AI-CAD 評価指標の 6.(2)2)「学習」で述べられている。また、本ガイドライン 4.2 ③ i) 「SOUP アイテムへの対応」も参照のこと。

iv) 機能の明確化

4.2 ②「設計開発の指針とソフトウェア要求事項の明示」の前項までにおいて明確化した設計開発の仕様書における要求事項に対して、次工程のアーキテクチャ設計やシステム試験の詳細計画が実施可能になるようにソフトウェアの機能をさらに展開し、その展開した機能単位で要求事項を記述する。

例：

- 画像取得方法、着目する特徴、特徴抽出の方法、統計解析の方法、出力方法

AI 技術を用いる CAD の場合の例：

- 学習方法を特徴付ける事項（教師あり、半教師あり、強化学習、自己学習、転移学習の有無、市販後学習機能の有無）

市販後学習機能を含む CAD の場合、この段階でその機能につき明確化する。性能変化によるリスクを防ぐための機能・手順を含む場合、同様にこの段階で明確化す

る。

v) 性能の明確化

- 診断支援性能

  意図する用途に見合う性能を設定する。

  例：真陽性率 x%、及び偽陽性率 y%（CADe の場合）

  例：2 クラス分類又は多クラス分類の正分類率 z%（CADx の場合）

- 処理時間

  意図する用途に見合う解析処理時間を設定する。

- 性能変化の許容範囲

  市販後学習を行う CAD の場合は、市販後学習における診断支援性能変化の許容範囲を明確化する。

③ アーキテクチャ設計

下記に基づいてアーキテクチャを設計することが望ましい。

※ 深層学習等の AI 技術を用いる CAD の場合、ii)「解析対象病変の特徴の明示」、iii)「解析対象病変の特徴量の明示」は必ずしも常に可能又は必要とは限らない。その場合、これらのステップは解析アルゴリズム及びネットワーク構造や学習アルゴリズム等の明示により代えることができる。

i) SOUP[9]アイテムへの対応

- SOUP アイテム（開発過程が不明なソフトウェア）に対して、医療機器プログラムに対する要求事項である JIS T2304 に従った対応を行う。
- AI 技術を用いる CAD の場合、学習データ及び学習済みモデルを、SOUP アイテムと同様の条件下に扱うことができる。その場合も AI-CAD 評価指標 6.（2）の 2)「学習」で規定される要求事項が求められる。

ii) 解析対象病変の特徴の明示

病変の特徴を明示する。

例：大きさ、濃度、辺縁部の特徴等（CADe、CADx）

iii) 解析対象病変の特徴量の明示

病変の特徴量を選定し、これらを明示する。

[9] 開発過程が不明なソフトウェア（Software Of Unknown Provenance）
既に開発されていて一般に利用できるが、医療機器に組み込むことを目的に開発したものではないソフトウェアアイテム（市販品（off-the-shelf）として知られているソフトウェア）または以前において開発されたソフトウェアでその開発プロセスに関する十分な記録が利用できないもの。[JIS T2304 用語および定義より]

例：大きさ、濃度、辺縁部の特徴等を定量化する（CADe、CADx）

※ 病変の特徴量は、十分な経験を積んだ医師による診断が下されている画像データに適用した結果に基づいて決定されていることが望ましい。

iv) 解析アルゴリズム・処理フローの明示

- 処理のフロー図等により示すことができる。
- 輪郭抽出、領域分割等のアルゴリズムは具体的なアルゴリズム名で示すことが望ましい。

  例：Canny エッジ検出、Voronoi 分割、ウェーブレット変換等
- 深層学習型の AI 技術を用いる CAD の場合は、開発設計時のネットワーク構造等も示す。

  例：2D/3D-畳み込みニューラルネットワーク、モデル ResNet-160 等

v) 学習方法の明示（AI 技術を用いる場合）

- 学習方法、用いるデータや学習済みモデルを明示する。
- 前項にて明示したアルゴリズム（ネットワーク等）に対して、収集したデータや学習済みモデルのどれを使って学習を行わせたのか､学習の順番､強化学習、自己学習、転移学習等の組合せ方等の具体的手順を記録する。

vi) 解析結果の出力方法の明示

例：円や四角で囲む、矢印、中心の位置等（CADe の場合）

例：異常部位の質的診断に関する情報をテキスト情報として出力する（CADx の場合）

④ システム試験

4.2 ②「設計開発の指針とソフトウェア要求事項の明示」で具体化したソフトウェア要求事項に対応して以下を含む検証を行う。

i) ソフトウェア実行環境の検証

4.2 ②「設計開発の指針とソフトウェア要求事項の明示／ii) ソフトウェア実行環境の明確化」にて明確化した以下の 2 項目に対してシステム試験を実施する。

- 組み合わせるハードウェア環境
- 組み合わせる他のソフトウェア

少なくとも最小限の構成について、実際の実行環境又はこれを模した環境でシステム試験を実施する。必要に応じて、オプションとして追加可能な構成についても試験する。この段階で行うべき試験項目は、個別の結合試験では十分に検証できない複雑なハードウェア、ソフトウェア構成の確認である。試験項目に CAD の機能、性能の

検証を含む必要はない（次項以降にて実施する）。

ii) 機能的要求事項の検証

4.2 ②「設計開発の指針とソフトウェア要求事項の明示／iv) 機能の明確化」にて明確化した機能的要求事項に対してシステム試験を実施する。

市販後学習機能を有する CAD の場合、市販後学習機能、学習後の性能評価法（承認時の性能評価法と同一であれば省略可）及び、性能変化によるリスクを回避する為の機能についても検証すること。

iii) 性能的要求事項の検証

4.2 ②「設計開発の指針とソフトウェア要求事項の明示／v) 性能の明確化」にて明確化した下記の性能的要求事項に対してシステム試験を実施する。

- 診断支援性能
- 処理時間
- 性能変化の許容範囲

性能的要求事項のシステム試験については、6.2「性能評価の基準」でも述べている。

市販後学習機能を有する CAD の場合は、性能変化の結果をシステムに反映させる前に検証する。

⑤ リスクマネジメント

JIS T2304 の箇条 7「ソフトウェアリスクマネジメントプロセス」によってソフトウェアリスクマネジメントを実施する。CAD の場合に留意すべき主な事項を以下に示す。

1）リスク分析

リスク分析を行う。ハザードと危険状態を特定する。

a) 学習データの特徴に起因する不適当な性能又は挙動。例えば、

- 希少事例のデータの不足等による挙動の不安定
- 偏った疾患又は正常例を過学習することによる性能低下（特に市販後学習時）

b) 合理的に予見可能な誤使用。例えば、

- （使用条件に含まない場合の）first reader、解析対象外の画像、非適用部位への利用、（専門医の使用を適用範囲とする場合の）専門外の医師による使用
- 必要な講習又は訓練を受けていない使用者による市販後学習の実施

2）リスクコントロール手段

AI 技術を用いた CAD に特有のリスクコントロール手段は、AI-CAD 評価指標の 6.(4)「リスクマネジメント」に例示されている。

⑥ 構成管理プロセス

i) 組み合わせるソフトウェア（SOUP アイテム含む）を識別する仕組みを確立する。

ii) SOUP アイテムの特定

名称／製造業者／識別子（バージョン等）

⑦ 設計開発の検証

4.2 ②「設計開発の指針とソフトウェア要求事項の明示」で設定した要求事項のうち、4.2 ④「システム試験」で検証されなかった項目を文書の確認等により検証する。

一般的には以下の項目が該当する。

i) 画像診断装置との接続の検証

ii) 附属文書への記載内容の検証

⑧ 設計開発の妥当性確認

CAD を実際の使用環境又は使用環境を模擬した環境で実行し、顧客ニーズ（意図する用途）に対する妥当性確認を行う。附属文書の妥当性も確認する。

5. 設計開発・保守プロセスの実施

5.1 設計開発プロセスの実施

上記 4.「設計・開発管理が要求する項目の各プロセスへの適用方法」を適用し、CAD の設計開発を JIS T82304 等の定める開発プロセスに基づき実施する。

5.2 保守プロセスの実施

市場リリース後の保守サイクルにおいても、設計開発プロセス同様に上記 4.「設計・開発管理が要求する項目の各プロセスへの適用方法」の項目を必要に応じ適用し、変更要求に対する設計変更開発を行う。

5.3 設計開発・保守プロセスにおけるデータの管理

(1) データの保管期限

① テストデータ

他の評価エビデンスと同様に、医療機器に要求される QMS 省令による設計管理において求められる期間の保管が必要。さらに性能変化後の再評価が必要になる事を想定すると、少なくとも評価対象の当該バージョンが市場で利用されている期間は保管することが必要になる。

② 学習データ、バリデーションデータ

どのような素性のデータで学習を行ったかという記録は残す。一方、学習データ全てを学習時の状態で維持する事は現実的でない場合も多い。このような場合は、テストデータで品質を担保する等の学習データ保管が不要な理由を明確する。学習データの保管が困難な例としては、ビッグデータを利用する場合や、医療機関や海外で学習済みのコンポーネント等を利用する場合等が挙げられる。

(2) テストデータを開発プロセスから切り離して管理するための方策

テストデータが開発プロセスで使用されないように管理する合理的に実施可能な方法が望まれる。

機械学習による開発プロセスでは、テストデータを学習データ及びバリデーションデータから切り離して管理する。なお、テストデータの学習結果への影響を完全に排除することは、容易でない[10]。テストデータ数が十分に多い場合、学習データが極めて多い場合等、テストデータを開発プロセスから切り離した場合と実質的に等価であると判断できる場合もありうることから、テストデータを開発プロセスから切り離して管理するための方策については、PMDA と相談することを推奨する。

6. 評価手法及び留意事項

本章では、設計開発プロセスで実施する性能評価において考慮すべき事項について説明する。

6.1 性能評価に影響するデータの特性

性能評価に影響するデータの主な特性を示す。これらは、設計開発プロセスの 4.2 ①「製品に関連する一般要求事項の明確化（意図する用途）」及び 4.2 ②「設計開発の指針とソフトウェア要求事項の明示」において明確化、明示する事項の検討に反映されることが望まれる。その他、テストデータ量に関する留意事項については Appendix 7 「テストデータの量」を参照されたい。

(1) データソースの特性

データの取得・管理方法を明確化する。データの元となる患者群の持つバイアス（偏り）等にも留意する。ただし非常に大規模なデータベースの場合、母集団に対するバイアスは小さくなることが期待される。

(2) 画像データ取得時の撮影パラメータ

CAD が有する性能を正しく評価するために、適用できるモダリティ（検査機器の種類）、入力となる医用画像データの収集条件（撮影パラメータ等）、入力画像の仕様（経時画像、造影剤の有無等）、画像の空間的な歪みや背景雑音に関して分析することが期待される。

撮像条件の例

- X 線 CT の場合：スライス厚、管電流や管電圧、再構成カーネル、造影剤の使用

[10] 例えば、条件（学習データやモデル）を変えて 10 通りの学習を行い、学習後の 10 通りのモデルをテストデータを用いて評価し、その中から下限を上回るモデルを選択したり、もっとも性能が良いモデルを学習後のモデルとして選択することは、開発過程でしばしば行われているが、モデルの選択にテストデータを利用している（＝学習や設計にテストデータを利用している）ことになる。

の有無等

- MRI の場合：静磁場強度、撮像シーケンスとその代表的パラメータ（TR、TE、フリップ角、積算回数、位相エンコード等）、造影剤の使用の有無等
- 内視鏡の場合：視野における対象の捉え方、動き等の条件、拡大や染色法の有無、特殊なスペクトルの照明の利用等
- 超音波の場合：プローブの種類や周波数、フレームレート、PRF、走査部位（心窩部、右肋間等）、造影剤の使用の有無等

(3) 正解・臨床画像データ以外のデータ

CAD の開発時において使用した画像データに対応する正解や、臨床画像データ以外のデータについてその背景事項を分析することが期待される。例えば、

- 正解には、病理診断結果その他の各種の診断結果、アノテーション等のほか、その正解に関する最終的な判定を行った医師等に関する情報（選択の方法、診療歴、専門医であるか等）も含まれうる。
- 臨床画像データ以外のデータには、シミュレーションやファントム等を用いて生成された人工画像、画像以外の医療情報が含まれうる。

これらを用いる場合の留意事項については AI-CAD 評価指標の 6.(2)2)「学習」の注 4)に記載されている。

6.2 性能評価の基準

性能評価の目的は、CAD の使用目的（意図した使用）としての診断支援が実施できることを示すことにある。

CAD の性能を十分な症例数を含む適切なテストデータ（Appendix 3 「テストデータにおける画像データベースとゴールド・スタンダード」参照）を利用して統計評価する（Appendix 2 「性能評価法の基礎」参照）。症例数、比較対照及び海外で収集されたデータの扱いについては、「コンピュータ診断支援装置に関する評価指標」（薬食機発 1207 第 1 号（別添 3）平成 23 年 12 月 7 日）に述べられている。

医師等による成績の単純比較による優劣の評価が困難な場合（例：CADe の検出性能を評価したい場合。CADe では検出漏れを防ぐため感度を高く設定するが、特異度が下がってしまい、医師の成績に対抗することができない）は、臨床的に妥当な評価基準によって判定する。その評価基準の医学的妥当性を示さねばならない。

性能評価の基準の例：

- 「臨床において医師が行っている診断精度」よりも、「CAD を用いた医師の診断精度」が、統計的に有意に優れているか同等である。

  （注：比較対象とする医師の資格、経歴等の要件は、CAD の意図する使用目的

によって変わりうる）

- 同一製品のバージョンアップ等の場合は、「既承認品の CAD の分類精度」よりも統計的に有意に優れているか同等である。

医師等を被験者とする CAD の性能評価の際には、CAD の使用による医師への影響、例えば、読影時間や疲労等について考慮する必要がある（Appendix 4 「性能評価のための読影実験における注意点」参照）。

7. 開発に関するその他の留意事項

ソフトウェア、データセットをダウンロードすることで医師が AI 技術を使った CAD プログラムを自作することも技術的には可能となっている。国内未承認の CAD プログラムを診療で自ら用いることは、医師法における医師の裁量権の範囲内とされる。しかし、設計開発プロセス管理が十分になされないソフトウェア、評価の定まっていない医療技術やソフトウェアを通常診療で使用することは近年の臨床倫理の観点では受け容れられないとの懸念がある。研究である場合は平成 30 年 4 月に施行された臨床研究法を遵守する必要がある。医師が CAD を自作して自ら用いる場合は、これらの臨床倫理、法令の趣旨等に沿うべきであり、更に、本ガイドラインを参考に設計開発プロセスを実施・記録し、CAD の開発プロセスの妥当性を担保することが望まれる。

8. 関連する規格及び参考資料

- 「人工知能技術を利用した医用画像診断支援システムに関する評価指標」（薬生機審発 0523 第 2 号別紙 4 令和元年 5 月 23 日）
- 「プログラムの医療機器への該当性に関する基本的な考え方について」の一部改正について（薬生監麻発 1228 第 2 号平成 30 年 12 月 28 日）
- 医療機器プログラムの取扱いについて（プログラム基本通知）（薬食機参発 1121 第 33 号、薬食安発 1121 第 1 号、薬食監麻発 1121 第 29 号平成 26 年 11 月 21 日）
- 医療機器プログラムの承認申請に関するガイダンスの公表について（事務連絡平成 28 年 3 月 31 日）
- JIS T82304-1:2018「ヘルスソフトウェアー第 1 部：製品安全に関する一般要求事項」
- IEC 82304-1:2016：Health software - Part 1: General requirements for product safety
- JIS T2304:2017「医療機器ソフトウェアーソフトウェアライフサイクルプロセス」
- IEC 62304,:2006+AMD1:2015 CSV, Medical device software - Software life cycle processes（医療機器ソフトウェア—ソフトウェアライフサイクルプロセス）
- 医療機器におけるサイバーセキュリティの確保について（薬食機参発 0428 第 1 号、薬食安発 0428 第 1 号平成 27 年 4 月 28 日）
- 医療機器のサイバーセキュリティの確保に関するガイダンスについて（薬生機審発 0724 第 1 号、薬生安発 0724 第 1 号平成 30 年 7 月 24 日）
- JIS T14971:2012「医療機器ーリスクマネジメントの医療機器への適用」
- IEC/TR 80002-1:2009（Medical device software - Part 1: Guidance on the application of ISO 14971 to medical device software; 医療機器ソフトウェアー第 1 部：医療機器ソフトウェアへの ISO 14971 の適用の手引き）
- JIS Q13485:2018「医療機器ー品質マネジメントシステムー規制目的のための要求事項」
- ISO 9001:2015 Quality management systems -- Requirements（品質マネジメントシステムー要求事項）
- QMS 省令：「医療機器及び体外診断用医薬品の製造管理及び品質管理の基準に関する省令」（厚生労働省令第 169 号平成 16 年 12 月 17 日）
- GCP 省令：「医療機器の臨床試験の実施基準に関する省令」（厚生労働省令第 36 号平成 17 年 3 月 23 日）
- 医療機器の迅速かつ的確な承認及び開発のための治験ガイダンスの公表について（事務連絡平成 29 年 11 月 17 日）
- GVP 省令：「医薬品、医薬部外品、化粧品、医療機器及び再生医療等製品の製造販売後の安全管理の基準に関する省令」（厚生労働省令第 135 号平成 16 年 9 月 22 日）
- 独立行政法人医薬品医療機器総合機構に対する機械器具等に係る治験不具合等報告について（薬食発 0329 第 14 号平成 25 年 3 月 29 日）
- 人を対象とする医学系研究に関する倫理指針（平成 26 年文部科学省・厚生労働省告示

第 3 号）

- 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（平成 29 年 5 月 29 日一部改訂）
- 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（附則編）（平成 29 年 3 月 8 日）
- 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（附則編）別添 経過措置に関するＱ＆Ａ集（平成 29 年 3 月 8 日）
- 医療・介護関係事業者における個人情報の適切な取扱いのためのガイダンス（個人情報保護委員会 厚生労働省 平成 29 年 4 月 14 日）
- 「医療・介護関係事業者における個人情報の適切な取扱いのためのガイダンス」に関する Q&A（事例集）（個人情報保護委員会事務局 厚生労働省 平成 29 年 5 月 30 日）

APPENDIX

Appendix 1. 旧版の Appendix から本ガイドラインに引き継がれなかった項目

旧版の CAD 開発ガイドライン（CAD 開発ガイドライン 2012 及び 2015）に掲載していた Appendix のうち、いくつかは本ガイドラインでは削除又は記載を簡略化し本文に統合されている。それらを掲載しなかった主な理由を以下に示す。

CAD 開発ガイドライン 2012

Appendix 「汎用ハードウェアで動作する医療用ソフトウェアの設計評価における技術的な裏付け」

以下の 4 項目につき解説していた。

1. 「医療機器の基本要件基準」への適合
2. IEC62304 適合のエビデンス
3. IEC62304 適合性検証レポート
4. 汎用ハードウェアの選定条件

しかし、平成 26 年の医薬品医療機器等法施行により（汎用ハードウェアで動作する）医療機器プログラムが制度化され、また関連する JIS 等の整備が進み、関連通知等にて上記項目も含めて明確化されたことから、本ガイドラインでは削除した。これらについては以下を参照されたい。

1. 「医療機器の基本要件基準」への適合
   →「医療機器の基本要件基準第 12 条第 2 項の適用について」（薬生機審発 0517 第 1 号平成 29 年 5 月 17 日）に、JIS T2304 への適合性確認により行うこと等を明記してある。
2. IEC62304 適合のエビデンス
   →同通知別添記載事例 1、2 として例示されている。
3. IEC62304 適合性検証レポート
   →同通知添付資料 1「プログラムのライフサイクルへの適合に関する概要報告書」として例示されている。
4. 汎用ハードウェアの選定条件
   →「医療機器プログラムの取扱いについて」（平成 26 年 11 月 21 日付け薬食機参発 1121 第 33 号、薬食安発 1121 第 1 号、薬食監麻発 1121 第 29 号厚生労働省大臣官房参事官（医療機器・再生 医療等製品審査管理担当）、厚生労働省医薬食品局安全対策課長、厚生労働省医 薬食品局監視指導・麻薬対策課長連名通知、以下「プログラム基本通知」）に、製造販売承認申請書の形状、構造及び原理欄の記載要領として明記されている。

CAD 開発ガイドライン 2015

A.1 「CAD の分類と CADx の位置付け」

CADe と CADx の定義、単体ソフトウェアの場合とソフトウェアを搭載した装置の 2 通りがありうること、適用するモダリティ等について解説していたが、本ガイドラインでは第 2 章で扱うこととなったため削除した。

A.3 「CAD に対する QMS （Quality Management System）」
CAD に対する QMS につき解説していた。ソフトウェアの場合、QMS と関連して IEC 62304（JIS T2304）、IEC TR 80002-1 等を適用することが望ましいとしていた。しかし、本ガイドラインでは JIS T2304 の内容を含むものとしたこと、プログラム基本通知等によりプログラム医療機器の QMS 調査に関しても明記されたことから削除した。

A.8 「データ収集する施設数」
A.9 「性能評価に対して収集しなければならないデータ数」
本ガイドラインでは、Appendix 7 「テストデータの量」として統合した。

A.10 「入力する画像診断装置における収集条件の明確化」
簡略化して、6.1②(2)「画像データ取得時の撮影パラメータ」に記載した。

Appendix 2. 性能評価法の基礎
CAD の性能評価に対しては、以下に示す分類率や、それを用いた ROC （Receiver Operating Characteristics）解析のいずれかを用いることを推奨する。どちらを使うかは、臨床利用の観点から決定してよい。すなわち、臨床的に一組の分類率による評価が妥当な場合には分類率を用いた評価で良く、様々な分類率の場合を総合的に評価しなければならない場合には ROC 解析を実施するのが一般的である。
分類率や ROC の計算には下記の表（クラス分類実験結果の集計表）の数値を用いる。

表 A1　クラス分類率に基づく性能評価（クラス分類実験の集計表）

CADx の出力 / 正解(Gold standard)	クラス１	クラス２	…	クラスｎ	合計※１
クラス１	a_{11}	a_{21}	…	a_{n1}	$S_1=a_{11}+…+a_{n1}$
クラス２	a_{12}	a_{22}		a_{n2}	$S_2=a_{12}+…+a_{n2}$
:	:	:	:	:	:
クラスｎ	a_{1n}	a_{2n}	…	a_{nn}	$S_n=a_{1n}+…+a_{nn}$
合計※２	$a_{11}+…+a_{1n}$	$a_{21}+…+a_{2n}$	…	$a_{n1}+…+an_{1n}$	$S=S_1+…+S_n$

※１正しく分類された数についての正解クラス別の合計

※２正しく分類された数についての出力別の合計

これらの数値は、CAD の実際の利用を想定して作成したデータベースを用いて、Leave one out 法、k-fold Cross Validation 法、Hold Out 法のいずれかによって求めることが望ましい

（Appendix 5「性能評価手法」参照）。ここで、評価に用いたデータベースが実際の利用を想定しているためには、ランダム標本か階層標本等、学術的に妥当な方法を用いてデータ収集を行えば良い（Appendix 3「テストデータにおける画像データベースとゴールド・スタンダード」参照）。
また、正解は病理診断結果や合理的な診断結果から導かれていなければならない。学術的に裏付けのないクラスを独自に設定して CAD の性能評価に用いても、裏付けのない性能評価にしかならない。ここで、「合理的」とは、学術的な客観性や再現性があることを指す。したがって、手術や生検で得られた病理所見、画像検査で決定した診断、経過観察等を含む画像検査で下された診断、上記の検査の結果に基づく総合的な判定等がありうる。ただし、様々な要因によって検査結果に大きなばらつきやバイアスが含まれる事例では、複数の医師による診断結果の平均や合意等により差異を小さくするか、より精密な他の検査に置き換える必要がある。なお、ファントム等の人工データを利用する場合には、妥当な方法で定義された正解を用いねばならない。

分類率による評価

正分類率は、下記のように、表 A1 中の数値から計算される。

$$\text{クラス i の正分類率} = a_{ii} / S_i \quad (i=1...n)$$

ここで、あるクラスの診断精度は従来よりも良くなったが、他のクラスの診断精度が悪くなったのでは意味が無いことに注意をしなければならない。十分な数のデータを用いた上で、すべてのクラスに対する精度が同等か、あるいは、精度が統計的に有意に向上したクラス（クラス名を明確にすること）が一つ以上であり、かつ、その他のクラスに関する性能は統計的に同等であることを示す必要がある。

ROC （Receiver Operating Characteristics）解析による評価

2 クラス分類の場合は各クラスの正分類率の間に存在するトレードオフの関係を表した ROC 曲線（Appendix 6「観察者特性の評価」参照）を用いることにより統計的に証明することができる。ここで、精度が同等であることを示す場合には、統計的に十分な数のデータに基づいていなければならない。また、ROC の軸は、上記の各クラスの正分類率に基づいて定義する。例えば、良悪性鑑別の場合には、縦軸は感度、横軸は偽陽性率となる。ただし、対象によっては、LROC（Localized response ROC）、FROC（Free response ROC）、AFROC（Alternative FROC）、及び JAFROC（Jackknife AFROC）等の他の評価法が適していることがある[1)]。その場合には、上記の分類率を適切な性能指標、例えば FROC の場合には、偽陽性率を一症例あたりの偽陽性数等に置換して用いることが必要である。
良悪性の判定のみならず、鑑別診断結果をリストアップする場合や、疾患の程度の分類を行う場合等は多クラスの分類が必要となる。この場合は、ある一つの注目するクラスと、残り

の(n-1)のクラスを一つにまとめたクラスとの 2 クラスについて行う ROC 解析を、注目するクラスを替えながら n 通りについて行えばよい。一部のクラス群のみに注目して ROC 解析を行う方法も考えられるが、その場合には、一部のみに注目することの臨床的妥当性や、評価の統計的妥当性を学術文献等により確認し、根拠を明確にする必要がある。

【参考文献】

1) 尾川ら編：医用画像工学ハンドブック（Part II, §3.3〜§3.5），日本医用画像工学会, 2012.

統計的検定に関する注意事項

全ての検定は、統計学上の手続きを踏まえたものでなければならない。仮説検定における P 値等の有意水準としては統計的に妥当なものを用いる。P 値については慣例的に用いられてきた 0.05 では不十分とする論文が発表されている。読影実験の際の施設数や医師数はそれぞれ複数が望ましいが、1 施設でも良いと自己判断する場合は科学的に妥当な根拠を示すことが不可欠である（Appendix 7 「テストデータの量」）。ただし、多クラスの場合にはクラス毎のデータ数に偏りがあり、クラス間のデータ数の比率が実際の臨床における症例数とかけ離れる可能性が高くなる。そのため、特に注意をして十分なデータ数及び施設数を確保する必要がある。なお、クラスごとの分類率にコスト（重み）を導入して再定義し、それに基づいて統計的検定を実施する場合には、検定を実施するより前に、そのコストが臨床的・学術的に正当であることを、査読を受けた論文等で確認して根拠を明確化しなければならない。また、各クラスのデータ数については臨床適用可能な状態で統計解析の結果において有意な差を示すために十分な数を用いることが重要となる。

【参考文献】

1) Wagner et al., Assessment of medical imaging systems and computer aids: a tutorial review. Acad Radiol., 14(6):723-748, 2007.
2) 藤田ら（監修），実践医用画像解析ハンドブック（§6.3.3），オーム社, 2012.
3) Wasserstein, Lazar., The ASA's Statement on p-Values: Context, Process, and Purpose., The American Statistician., 70(2):129-33, 2016.
4) Benjamin, Berger, Johnson et. al., Redefine statistical significance., Nature Human Behaviourvolume 2:6–10, 2018.

Appendix 3. テストデータにおける画像データベースとゴールド・スタンダード

CAD 研究において、CAD のトレーニング（又は学習）及びテストに使用される画像データベースの特性は非常に重要であり、その特性によって CAD 研究の結果が大きく左右されるといっても過言ではない [1)]。CAD 研究に用いられる画像データベースには、多くの場合、画像だけではなく、様々な付帯情報が含まれる必要がある。その情報としては、細胞診や組織

診等の病理診断により得られた確定診断結果や専門医による客観的な診断の難易度、患者の性別や年齢等が挙げられる。こういった画像データベースの構築においては大量の画像と幅広い症例の情報を収集し、それらを整理するための工夫が必要となる。しかしながら、単一の施設では収集できる症例数が限られるため、大規模な画像データベースの構築が困難な場合が多い[2)]。そのため、デジタル画像が医用の世界に普及し、CAD 研究が盛んに行われるようになった頃から、誰もが使用可能な公共の画像データベースの構築に関する研究が国内外で行われてきた[2-5)]。現在では、いくつかの公共での使用が可能な画像データベースが公開されており、それらには、日本放射線技術学会で構築された結節影あり／なしの胸部単純 X 線写真のデータベース[2)]や、特に読影学習を目的としたマンモグラフィのデータベース[3)]、米国の National Cancer Institute（NCI）の研究班が構築した肺結節の CT 画像のデータベースである Lung Image Database Consortium（LIDC）[4)]、南フロリダ大学のデジタルマンモグラフィのデータベース[5)]等がある。このような画像データベースの存在により、画像データベースの構築が困難な研究者でも量的及び質的に充実した研究用データを使用し、CAD の研究を行うことが可能となり、また同じ画像データベースを使用しているものであれば、開発された CAD の性能の相互比較が可能となった。新しいモダリティや検査法に応じた CAD の開発においては、各研究施設において研究目的に応じた画像データベースを構築することも想定されるが、その際には構築したデータベースによって、CAD の評価がバイアスを受けることのないように注意する必要がある。

画像データベースに収録される画像の収集を計画する場合、その母集団について考慮する必要がある。例えば、過去数年間にある医療機関を受診したすべての患者を対象とするのか、ある特定の検査を受けた患者のすべてを対象にするのかによって、収集する画像の母集団は違ってくる。CAD の開発を行う場合に、対象となる疾患の罹患率は重要な要素となるので、受診したすべての患者を母集団とすれば、およその罹患率を推定することが可能になる。しかし、よほど罹患率の高い疾患でない限り、CAD 開発のための画像データベースを構築するために必要な母集団の数は膨大なものになると予想される。そのため、多くの場合は、CAD が対象とする疾患の疑いがあり、対象とするモダリティで検査を受けたすべての症例を母集団として設定し、その母集団からランダム標本、又は階層標本によって研究に必要な症例数を確保する研究デザインが採用される。ここで、ランダム標本とは、母集団からランダムに標本を抽出し、データベースに収録する方法である。他方、階層標本とは、母集団の特性（年齢・性別構成、既往歴の有無等）を調べた上で、その特性に合わせて標本を抽出する方法である。一般に、収録される症例の数が多くなるにつれて、ランダム標本と階層標本の両者を用いた場合の差異が小さくなる。

画像の収集と同時に考慮しなければいけないのは、画像に含まれる病変のゴールド・スタンダード（Gold Standard; GS）の決定である。GS は Reference Standard や Ground Truth とも呼ばれ、臨床研究において「その症例が間違いなく CAD 研究の対象となる疾患である」又は「間違いなく疾患ではない」ということを証明するための証拠のことであり、GS の決定が不明であったり、明確でなかったりする場合は、CAD 研究そのものの真偽が問われるこ

とになる場合がある。
特にテストデータにおいては正解を明確にする事が重要である。ただし、CAD の使い方によっては、疑陽性のデータを付加した人工画像等をテストデータとして用いる場合も想定される（例えば見逃し防止に重点化した CAD の場合等）。
また、AI 技術を用いた CAD の場合には、学習データとしては GS 付きの臨床画像以外のデータ（人工画像等）を用いる場合も考えられる。
CAD 研究用の画像データベースに含まれる画像の GS の決定には、(1)手術又は生検で得られた組織・細胞の病理所見、(2)病理所見と臨床判断（経過観察）の組合せ、(3)臨床判断のみ、(4)上位の診断システムの結果（例えば、胸部単純 X 線像に対する CT 検査の所見）、(5)専門医によるコンセンサス、(6)ファントム実験やシミュレーション信号等既知のデータを用いる方法が挙げられる。悪性腫瘍の病変を対象として CAD を開発する場合には、すべての症例に関して病理所見で診断が確定していることが理想であるが、悪性が強く疑われない場合は、生検や手術なしで経過観察の臨床処置がとられる場合が多いので、病理所見と臨床判断との組合せで GS の決定が行われるのが一般的である。なお、規制当局への提出を想定している試験の場合、GS の作成を含む評価系全体で用いる情報の種類により、規制当局の当該試験に対する取扱いが変化する可能性があるので留意されたい。
画像データベースに収録される症例の数は、CAD において使用される様々な識別機構（Classifier）の性能に大きく影響する [3)]。一般には、症例数が多ければ多いほど CAD の性能の正当性は高くなるが、CAD 研究に必要な多くの症例を確保することは時として非常に困難であるので、限られた症例数で信頼性の高い CAD を開発するための工夫が必要となる。

【参考文献】

1) Nishikawa RM, Giger ML, Doi K, Metz CE, Yin F-F, Vyborny CJ, Schmidt RA: Med Phys, 21(2), 265-269, 1994.
2) Shiraishi J, Katsuragawa S, Ikezoe J, Matsumoto T, Kobayashi T, Komatsu K, Matsui M, Fujita H, Kodera Y, Doi K: AJR Am J Roentgenol, 174(1), 71-74, 2000.
3) Chan H-P, Sahiner B: Med Phys, 26(12), 2654-2668, 1999.
4) Li Q, Doi K: Med Phys, 34(3), 871-876, 2007.
5) ICRU Report 79. Receiver Operating Characteristic Analysis in Medical Imaging. Oxford University Press, Oxford, UK, 2008.

Appendix 4. 性能評価のための読影実験における注意点

6.2「性能評価の基準」要件を証明するための読影実験ではバイアス等が含まれているため、結果の解釈を誤ることを防ぐための代表的な注意点や、実験の再現性を担保するための注意点について述べる。その他の注意点については Appendix 7 「テストデータの量」を参照のこと。

○ 評価結果にバイアスが混入する恐れのある以下のような読影実験は避けなければならない。ただし、バイアスを混入させる因子はこれら以外にもあるので注意が必要である。
  - 同一医師群が同一症例群を用いて、CAD を利用しない場合と利用した場合の 2 回の実験を短期間で行ってしまうと、症例に対する記憶が 2 回目の読影結果に影響を与える恐れがある。
  - 読影の際の画像の提示順序は実際の臨床の場合と同様、原則としてランダムでなければならず、恣意的に決めてはならない。
○ CAD の利用により想定される不利益を全て記録し、その妥当性について評価する必要がある。例えば、CAD の出力の待ち時間や CAD の出力を参照することで増加する読影時間等である。これらを測定し、有効性を上回る不利益が無いことを確認しなければならない。
○ CAD を用いた読影実験の再現性についても注意を払わなければならない。すなわち、第三者による追試によって同等の性能が得られるよう、再現に必要な全ての実験条件を記録しなければならない。以下は放射線を例に条件を示したものであるが、モダリティの特性にあわせて、追加、削除し、CAD の性能を提示する際には必要なすべての条件を開示できるように準備しておくべきである。
  - 入力画像の仕様：画素サイズや濃度レベル数等（3 次元画像であればスライス厚やスライス間隔等のパラメータも含む）
  - 画像の撮影条件：6.1②(2)「画像データ取得時の撮影パラメータ」を参照のこと。
  - 対象疾病：疾病の種類や診断の難易度
  - 被検者情報：年齢、性別。必要に応じて過去の疾病や手術等の既往歴、体型
  - データ収集法：施設名、収集時期、画像枚数、及び具体的なデータのサンプリング法（ランダム標本化や階層標本化）
  - CAD を動作させたコンピュータ環境：CPU の性能、メモリサイズ、ディスプレイの解像度やγ特性
  - CAD の処理パラメータ：製品化後に使用者が変更可能な処理パラメータ。例えば、良悪性鑑別の場合は悪性度に対する閾値
  - CAD の利用形態：first reader や second reader、concurrent reader の区別
  - CAD を利用した医師に関する情報：専門分野、読影経験年数、CAD の利用法に関する事前説明、当該 CAD に関する習熟度、その他の CAD の利用経験

Appendix 5. 性能評価手法

CAD で使用される処理手法の多くは、原画像から候補領域を抽出する処理と、抽出された

候補領域を良性・悪性又は真陽性・偽陽性に識別する処理の二つに大別される。識別のための機構としては、線形判別分析（Linear Discriminant Analysis: LDA）、人工ニューラルネットワーク（Artificial Neural Network: ANN)、サポートベクターマシン（Support Vector Machine: SVM）等が含まれる。

CAD の性能評価は最終製品に対するもの（治験により収集したテストデータによる評価を含む)、開発途上の評価（トレーニングを含む）に大別される。学習データ・バリデーションデータと（最終製品の評価に用いる）テストデータを分離する原則は、最終製品に対する評価で堅持すべきことは当然として、開発途上にあっても理想的には、閾値の設定や手法のトレーニングのために用いられる画像データベースと、その処理手法をテストするための画像データベースが区別されて用意されていることが望ましい。しかしながら、開発段階において、必要な条件を備えた症例数を確保することは困難な状況も想定される。そこで、限られた画像データベースを有効に利用し、かつ処理手法の性能を正確に評価するために、以下に示す四つの評価手法のうち、RB 法以外を推奨する。ただし、RB 法の結果を示しても良い。また、これらの評価手法に限定するものではなく、CAD の特性に応じて適切な評価手法を選択する事が肝要である。

1)　繰り返し代入法（Resubstitution: RB 法）

繰り返し代入法は、最も簡便で単純な方法で、画像データベースに含まれるすべての画像で、処理手法における閾値の設定や識別機構のトレーニングを行い、そして、その処理手法をテストする際にも同じ画像データベースを用いる。テストに用いられる画像がコンピュータのトレーニングに既に用いられた画像であるので、多くの場合に処理手法の性能は過大評価され、画像データベースに含まれる症例数が少ないほど、その傾向は顕著になる。CAD 開発のパイロット研究等で比較的小規模の症例数で行われる場合が多い。パイロット研究の段階である程度の性能が見込めないコンピュータ技術は、臨床的にも有用となる可能性が低く、有用なソフトウェアを開発することが困難なことが予想されることに起因する。

2)　Leave-One-Out 交差検定法（Leave-one-out cross-validation: LOO 法）

交差検定法（cross-validation）は、統計学において標本データを分割し、先ず、その一部を解析して、残る部分を最初の解析の仮説検定に用いる手法である。交差検定法では、最初に解析するデータをトレーニング用データセット、残ったデータをテスト用データセットと呼ぶ。LOO 法はラウンド・ロビン（round-robin）法とも呼ばれる交差検定法の一種で、先ず、画像データベースの中から一症例を取り出して、それをテスト用データセットとし、残りのトレーニング用データセットで学習させた処理手法のテストに用いる。その後、同じ作業をすべての症例について繰り返す。例えば、画像に 100 症例分のデータが含まれる場合、一症例を取り出して、残りの 99 症例分でトレーニングを行い、その取り出した一例をテストする処理を 100 回繰り返す。ここで、同一症例から複数の

標本がデータセットに含まれている場合、上記の一つずつ標本を取り出す段階において、トレーニング用のデータセットに同一症例からの標本が含まれることになるため、そのことがテストにおけるバイアスになる可能性がある。したがって、同一症例からの複数の標本がデータセットに含まれる場合には、一つの標本をテスト用として取り出すのではなく、一つの症例からの標本のすべてをテスト用として取り出して、残りのデータセットでトレーニングを行う Leave-one case-out 法を用いる必要がある。

一般に、LOO 法は繰り返し回数が多いため非常に時間がかかる場合がある。また、LOO 法で評価される処理手法の性能は RB 法に比べて低くなる傾向があり、その差は症例数が大きくなると減少する。LOO 法は、処理手法の性能の評価という点では信頼性が高いが、処理手法における閾値の設定等が、全ての症例についてトレーニングが繰り返される毎に変化するので、臨床応用を考慮する場合には別の手法で閾値を固定させる等の工夫が必要になる（この点は、後述の KCV 法や H 法でも同様）。

3)　K 分割交差検定法（K-fold Cross-Validation: KCV 法）

KCV 法は交差検定法の一つであり、LOO 法が一症例ごとにトレーニング用データセットとテスト用データセットに分割していたのに比べて、画像データベースの全体を K 分割して、そのグループごとにトレーニング用データセットとテスト用データセットを入れ替えて評価を行う。例えば、画像に 100 症例分のデータを 5 分割して 20 症例ずつのグループに分ける場合、一つのグループ（20 症例）を取り出して、残りの四つのグループ（80 症例分）で学習を行う。そして、その取り出した一つのグループをテストする処理を 5 回繰り返す。その後、得られた 5 回の結果を平均して一つの推定をうる。画像データベースに含まれる症例数が比較的多く、LOO 法では時間が必要となる場合に KCV 法は有用な評価法となる。

4)　ホールドアウト検定法（Hold out method）HO 法

HO 法は KCV 法の分割数を 2 にした場合の評価法と基本的には同じであるが、HO 法では、トレーニング用とテスト用のデータセットを入れ替えない点が異なる。

Appendix 6. 観察者特性の評価

図 A1 に示すのは、胸部結節影の検出を目的として処理手法を開発し、その後、診断医の検出能の向上の評価のために ROC 解析が実施された対象モダリティの異なる二つの CAD 研究から得られた ROC 曲線である。一つは対象モダリティが胸部単純 X 線像で、もう一つは CT を対象としている [1,2]。両者の ROC 曲線で、点線で示したのはどちらもコンピュータ単独の性能であり、このコンピュータの手法から得られた出力を、診断医が読影を行う際に提示しなかった場合（without CAD）と提示した場合（with CAD）の差から、開発した処理手法の有用性を検討した。この両者においては、モダリティが違う上に、処理手法の

開発に用いられた画像データベースも違うので、単純には比較することはできない。しかし、コンピュータ単独の性能の評価と、そのコンピュータの出力を診断医に提示した場合の検出能の改善の程度には直接的な関係がないことは明らかである。コンピュータの出力を診断医が利用した場合の臨床における有用性は、観察者実験を行うことによってのみ証明することが可能である。そして、この臨床における有用性が証明されることが、「コンピュータ支援診断」研究にとってはもっとも重要なことである。さらに、図 A1 に示したデータから、コンピュータ単独の性能に対して、診断医がそのコンピュータの出力を利用した場合の検出における性能は、コンピュータ単独の性能を下回る場合（A）もあれば、上回る場合（B）もあることも分かる。

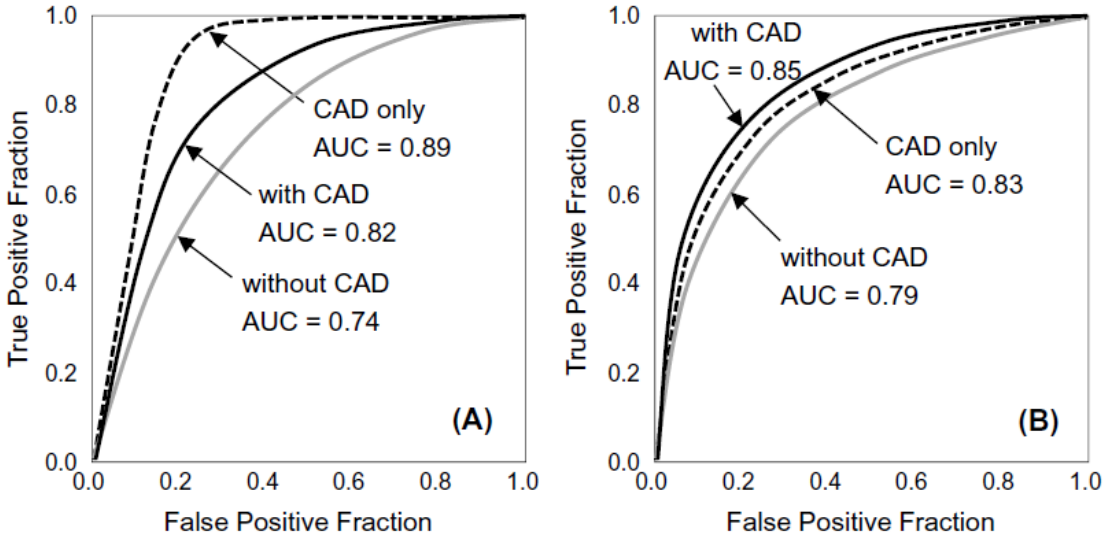

図 A1　2 種類の CAD（(A)胸部単純 X 線像における結節影の良悪性鑑別 [1], (B)胸部 CT における結節影の良悪性鑑別 [2]）を使用した場合の読影医の診断能の向上の ROC 曲線

観察者特性の評価において、その結果にバイアスが含まれる可能性のある主要な因子を以下に示す。(1) 試料画像の読影の難易度、(2) 読影を行う試料画像の枚数、(3) 読影を行う試料画像の選択方法、(4) 読影実験の慣熟に用いられる試料画像、(5) 読影実験に使用される評価の方法、(6) 読影順序効果、(7) 読影実験に使用される環境、(8) 読影を行う観察者（読影医等）の数、(9) 読影を行う観察者（読影医）の臨床経験、(10) 実験結果に対して行われた統計的解析の手法である。これらのうち、CAD における観察者の特性の評価用データベースとして、特に考慮すべき項目は、(1)、(2)、(3)の読影試料である。

観察者実験に使用される試料画像の読影が非常に困難な場合、CAD の支援によって異常陰影が検出されていたとしても、読影医がその情報に同意しなければ読影医の診断能は低くなり、CAD の支援による利益は非常に小さくなる [3]。一方、読影の難易度が非常に低い場合も、CAD の支援がなくても読影医は容易に異常陰影を検出することが可能なので、CAD の支援による利益は小さくなる。

読影に用いられる試料画像の枚数は、統計解析のために必要とされる数と、観察者が読影実験によって受ける疲労又は集中力の欠如、観察者実験を分割して実施する場合の弊害等を総合的に考慮して決定する必要がある。一般に、1 回の読影実験の所用時間は観察者の疲労や集中力を考慮すると 1 時間以下が望ましいので、読影枚数が多くて 1 回の読影ですべての試料を観察することが困難な場合は、読影試料を分割して観察者実験を行う。なお、一つの試料に一つの信号（異常陰影）という制限のある ROC や LROC（ROC-type curve for task of detection and localization）解析に比べて、一つの試料に複数の信号が存在することを許容する FROC（free-response receiver operating characteristic）や JAFROC（Jackknife Free-Response ROC）解析では、症例数を増やさずに信号の数を増やすことができるので、画像の数と読影に必要な時間の観点から、効率の良い観察者実験の計画が期待できる。
読影実験に用いる試料の数と同様に、試料の選択方法も実験結果にバイアスを与えないための重要な因子である。試料の選択方法には、Appendix 3「テストデータにおける画像データベースとゴールド・スタンダード」で述べたランダム標本と階層標本がある。原則として、観察者実験に用いられる試料に対する CAD の性能は、処理手法の開発時にテスト用の画像データベースで得られた性能と、ほぼ同等でなければいけない。例えば、処理手法の開発時の感度と特異度が共に 75%であるにも関わらず、観察者実験に用いる試料に対する CAD の性能が 90%の感度と 80%の特異度であったとすれば、その観察者実験で得られる結論は実際の CAD を過大評価している可能性が高い。しかし、将来的に達成可能なレベルの CAD の性能を仮定して観察者実験を行う研究、例えば、CAD の性能について感度がどの程度であれば、臨床的な有用性が認められるかを証明するために、仮想的に CAD の性能を高くする場合がある [4]。

【参考文献】

1) Shiraishi J, Abe H, Engelmann R, Aoyama M, MacMahon H, Doi K: Radiology 227(2), 469-474, 2003.
2) Li F, Aoyama M, Shiraishi J, Abe H, Li Q, Suzuki K, Engelmann R, Sone S, MacMahon H, Doi K: AJR Am J Roentgenol, 183(5), 1209-1215, 2004.
3) 石田隆行, 桂川茂彦, 藤田広志監修：医用画像ハンドブック, 544-558, オーム社, 2010.
4) Shiraishi J, Abe H, Engelmann R, Doi K: Acad Radiol, 10(11), 1302-1311, 2003.

Appendix 7. テストデータの量

テストデータ量の確定は開発する機器の性能を評価する上で不可欠である。「コンピュータ診断支援装置に関する評価指標」（平成 23 年 12 月 7 日薬食機発 1207 第 1 号（別添 3））において、評価試験に必要なテストデータ数は「装置の目的や主要評価項目等を踏まえ、検出率や偽陽性・偽陰性率算出に必要なデータ数とする」こととしている。開発及び製造販売を想定する CAD に対して標榜する性能や検出する特徴、評価方法等が異なることから、テス

トデータの量（データ量の上限値と統計的な分布）を定量的に示すことは困難であるが、当該機器において標榜する性能が統計的に明示されることが推奨される。
テストデータを収集する施設の数は、収集データの客観性を保証するために 2 施設以上で収集することが望ましい。これは、単一の施設では疾患や入力装置に偏りが生じることが懸念されるためである。単一の医療機関でのみ収集する場合は複数施設で収集した場合と実質同等であることを示すべきである。

Appendix 8. 市販後学習による性能変化の課題

市販後学習による性能変化を生じる場合、特に施設ごとに異なる性能変化を生じる場合、技術的、法的課題が未解決である。本 Appendix では製造業者が品質管理を実施する CAD について技術的課題を生じうるケースにつき述べる。なお、法的課題も同時に生じる場合があることに留意する。

1) 性能範囲の設定と維持

市販後学習による性能変化に対しては、製造業者が品質管理を実施し、仕様として事前に設定した許容される性能範囲内に維持する必要がある。
市販後の性能変化には以下の場合がある。

① バージョンアップ毎の段階的に性能変化する場合（バージョンアップについては AI-CAD 評価指標の 6.(3)2)「市販後」に記載がある）
② 市販後学習により高頻度で性能変化する場合

特に後者の場合に製造業者による品質管理維持が難しくなる。

2) CAD の提供形態に応じた性能管理

CAD の提供形態としては、医療機関・施設等に設置された装置上で実行される形態（オンプレミス型[11]）か、公衆ネットワーク等を介して実行される形態（クラウド型[12]）が想定される。
クラウド型は、集中管理可能である一方、公衆ネットワーク障害の影響を受ける。また情報セキュリティ上の十分な対策が必要となる。性能変化する場合は製造業者による品質、バージョン管理が可能である。性能向上を全体で利用できる一方、性能低下も全体に影響するので全体展開を考慮した性能管理が必要となる。
オンプレミス型は、公衆ネットワーク障害の影響が小さい。特に公衆ネットワーク対応しない医療データ管理を行なっている施設にとって導入が容易である。反面、施設毎の管理が必要になる。オンプレミス型の CAD をネットワークで接続して学習データを「共用」することも技術的には可能であるが、オンプレミス型であることのメリットを減じる。市販後学習

[11] 自身が管理する設備内に情報システムを設置・運用する形態（総務省「クラウドサービス提供における情報セキュリティ対策ガイドライン」）
[12] 従来は使用者が手元のコンピュータで利用していたデータやソフトウェアを、ネットワーク経由で、サービスとして使用者に提供する形態（http://www.soumu.go.jp/main_sosiki/joho_tsusin/security/basic/service/13.html）

データの偏りに起因して、施設毎に性能の異なる CAD が存在することは製造業者による品質管理、安全対策を難しくする。
オンプレミス型の CAD がネットワークから切り離された状態で運用される場合（公衆ネットワークの障害が起きた場合も含む）は、製造業者が公衆ネットワーク経由でメンテナンスが行えない。
使用者が主体となって学習させることを意図した CAD も考えられるが、本ガイドラインでは扱わない。

3) 市販後学習による性能変化のその他の課題

市販後学習による性能変化に関しては以下のような課題も予想される。

(1) 臨床で得られるデータには教師ラベルがない、又は簡略的、間接的な教師ラベルしか得られないことが一般的である。得られた場合も、専門医による判定を経た教師ラベルつき学習データと同等に扱えない可能性が高い。そのようなデータを用いて市販後学習することの効果、効率、データクレンジングを行う場合はそのコスト等、市販後学習を行うことについてコストパフォーマンスの観点からも慎重な検討が必要である。
(2) 特定の病態を多く含む偏りのあるデータで市販後学習を繰り返すと一見性能が向上したように見えても、平均的な病態に対しては診断精度が低下する可能性がある。

Appendix 9. 個人情報保護法概要と対応

個人情報の保護に関する法律（平成 29 年改正施行、以下：個人情報保護法）は、その組織が、民間や私立大学の場合に適用され、独立行政機関の場合は、独立行政法人等の保有する個人情報の保護に関する法律が、国立の研究機関の場合には、行政機関の保有する個人情報の保護に関する法律が、また、公立の研究機関の場合には、個人情報保護条例が適用されるため留意すること。なお、適用される法律は異なるが、基本的には同等である。

（医療機関から提供を受ける診療情報の扱い）

個人情報保護法において、診療情報は要配慮個人情報とされ、取扱いに注意することが求められているとともに、その情報の医療機関以外への取扱いに関しては、患者の同意を得て扱うことが必要となる。AI 技術を用いた医療機器の学習に用いること、学習結果を製品に適用すること等、使用目的を想定した同意を得ておくことも必要になると考えられるため、注意が必要である。

なお、学術研究の場合は、同法 76 条の適用除外条項により、個人の同意を得なくても個人情報を研究に用いることができるが、製品開発等の場合には同条は適用されないこと、一方で診療情報の学術研究での利用に関しては、臨床研究法の定める臨床研究又は「人を対象とする医学系研究に関する倫理指針」の定める「研究」に該当する可能性が高いこと、これらは研究対象者の同意取得を原則的に求めていることに注意する。

このため、企業が製品開発等のための画像情報等を収集するためには、個人情報保護法に従

い、医療機関等において匿名化の処理を行い、個人情報に該当しない情報として提供を受け用いるか、患者から利用に関する同意を得た要配慮個人情報として提供を受ける必要がある。提供を受けた企業等では、個人情報に該当しない場合であっても情報の適切な管理体制の下で用いる必要がある。医薬品医療機器等法に基づく臨床試験（治験）に基づくデータ収集・利用・保管は GCP 省令（平成 17 年 厚生労働省令第 36 号）に従う必要がり、個人情報保護法第 16 条 1 項、2 項の適用を受けないが、医療・介護関係事業者における個人情報の適切な取扱いのためのガイダンス等に従い、情報の取り扱いには十分な配慮が必要である。

（海外へのデータ送信）

CAD の運用において要配慮個人情報に該当する画像データ等を海外のクラウド等へ送信して処理する場合は、データの管理・処理方法の水準につき国内と同等以上であるかを確認するとともに、海外への送信及び処理に対して患者同意を得る必要があると考えられる。

（市販後学習を行う場合）

市販後学習を行う場合については、開発時と同等の患者同意が必要となる。

（同意の撤回があった場合）

なお、同意が後から撤回された場合、適切に匿名化したのちに取得した当該患者のデータを遡って削除する必要はない。次のように判断できる。

1) 「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」（平成 29 年 5 月 29 日一部改定）第 12 の 8「同意の撤回」では、『3「当該撤回又は拒否の内容に従った措置」とは、例えば、既に取得した試料・情報の使用停止・廃棄、他機関への試料・情報の提供の差し止め等が想定される。』との記載がある。
2) 一方、同意の撤回されたデータについては、その受領者がこれを特定することが不可能であり、よってこれだけを削除することができない。同ガイダンスの当該記載は匿名化されたデータまでを意図したものでないと考えられる。
3) 個人情報保護法は、そのようなデータが含まれているデータセットの破棄やデータセットに基づく研究開発の成果物の破棄や差し止めまでを法的に求めているものではない。

（データを用いた成果物の扱い）

深層学習において、学習済みモデルは、学習の結果得られたニューラルネットの重み係数とネットワーク構造であるが、重み係数は統計処理した結果であるので個人情報には該当しない。ネットワーク構造も同様に個人情報には該当しない。患者の同意が学術研究目的のみであったデータを用いて得られた学術研究の成果物を、製品に利用することが可能であるかについて明確化が図られることが期待される。

学習・評価に用いたデータについて同意撤回があった場合も、そのデータの成果物を遡って修正・削除する、又はその学習モデルを用いた製品又は匿名化されたデータからなるデータベース等の使用停止等の措置は原則的には必要ないと考えられる。

（参考資料）

- 医療・介護関係事業者における個人情報の適切な取扱いのためのガイダンス（平成 29 年 4 月 14 日通知、個人情報保護委員会事務局・厚生労働省）

・ 「医療・介護関係事業者における個人情報の適切な取扱いのためのガイダンス」に関するＱ＆Ａ（事例集）（平成 29 年 5 月 30 日適用、個人情報保護委員会事務局・厚生労働省）

なお、医療ビッグデータの利活用の重要性が認識されるなか、企業等第三者が診療情報を入手する手続きが厳格になったこと、医療機関が匿名加工を行うことが現実的でないこと、診療情報の匿名加工はその情報の性質によって様々な形態が考えられその判断には高度な専門性を要すると考えられること等から、認定匿名加工業者による匿名加工等を導入した次世代医療基盤法が制定され、平成 30 年 5 月に施行された。今後 AI の開発・性能改善に活用できるデータの提供が望まれる。

Appendix 10. 次世代医療基盤法・臨床研究法の概要と対応

次世代医療基盤法の概要と対応

医療分野の研究開発に資するための匿名加工医療情報に関する法律（平成 30 年施行、以下、次世代医療基盤法）は、「医療分野の研究開発に資するための匿名加工医療情報に関し、国の責務、基本方針の策定、匿名加工医療情報作成事業を行う者の認定、医療情報等及び匿名加工医療情報の取扱いに関する規制等について定めることにより、健康・医療に関する先端的研究開発及び新産業創出を促進し、もって健康長寿社会の形成に資することを目的とする」ものである。

このため、次世代医療基盤法に基づき匿名加工処理されたデータは、個人情報には該当しないものではあるが、取扱いに配慮して扱うべきデータであり、情報入手者は、その情報の提供者が誰であり、いつ入手したかについて、適切に記録を作成し保管する必要がある。

次世代医療基盤法では、基本的には、オプトアウトの申し出がない場合においては、匿名加工業者（認定された匿名加工医療情報作成事業を行う者）により個人情報が除かれた情報であり、研究開発や製品開発に使用することが可能である。ある時点でオプトアウトの申し出が行われた場合においても、それ以前のデータについては、そのまま利用することが可能である。

臨床研究法の概要と対応

臨床研究法の目的は、「臨床研究の実施の手続、認定臨床研究審査委員会による審査意見業務の適切な実施のための措置、臨床研究に関する資金等の提供に関する情報の公表の制度等を定めることにより、臨床研究の対象者をはじめとする国民の臨床研究に対する信頼の確保を図ることを通じてその実施を推進し、もって保健衛生の向上に寄与することを目的とする。」とされており、研究責任医師（研究を行う医師等の代表者）は臨床研究実施基準に従い計画書の妥当性を認定臨床研究審査委員会へ図り、承認を得て、計画書を厚生労働大臣へ提出して行うこととなる。

このため、特定臨床研究（医療機器等の性能・安全性を明らかにすることを目的とする研究

で、未承認品、適用外の機能を用いるもの又は製造販売業者等からの資金提供を受けて行う医療機器（既承認品を含む）を用いるもの）は基準への遵守義務が、またその他の臨床研究の場合には基準への適用の努力義務が求められている。
医薬品医療機器等法に基づく臨床試験（治験）は臨床研究法の適用を受けない。また、通常の診療行為により得られた情報のみを用いた観察研究については同法の適用を受けない。このため、通常の診療において撮影された画像等を用いた研究は同法の適用外となり、この場合は「人を対象とする医学系研究に関する倫理指針（平成 29 年文部科学省・厚生労働省告示第 1 号にて改正）」に従い研究を行う必要がある。
なお、AI 技術の研究や学習のため、通常診療にない制御を医療行為や患者の行動に加える場合は、観察研究に該当しない（臨床研究法の対象となる）可能性があり、注意が必要である。また、該当性の判断については、
厚生労働省の HP（https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000163417.html）で公開されている Q&A 及び事例集を参照して判断すること。判断が難しい場合は認定委員会の意見を聴くことができる。

人工知能分野 開発 WG 委員

座長	横井 英人	香川大学医学部附属病院 医療情報部 部長 教授
	片倉 由紀子	富士フイルム株式会社 メディカルシステム事業部 品証薬事部 スペシャリスト
	椎名 毅	京都大学大学院 医学研究科人間健康科学系専攻 教授
	清水 昭伸	東京農工大学 大学院 工学府　教授
	縄野 繁	国際医療福祉大学 医学部 教授 三田病院 放射線診断センター 画像診断部長
	野波 徹緒	オリンパス株式会社 医療第 3 開発本部 医療イメージング技術開発 1 部 部長
	藤田 広志	岐阜大学 工学部 電気電子・情報工学科 特任教授
	古川 浩	キヤノンメディカルシステムズ株式会社 経営企画部 参与
	宮崎 靖	株式会社日立製作所 ヘルスケア BU 診断システム事業部 ソリューションビジネス本部 本部長付
	森 健策	名古屋大学大学院 情報学研究科 知能システム学専攻 教授
	森 孝夫	オリンパス株式会社 医療第 3 開発本部 医療イメージング技術開発 1 部 リードディレクター
	諸岡 直樹	株式会社島津製作所 医用機器事業部 品質保証部 シニアマネージャー