

平成 29 年度 未来医療を実現する医療機器・システム研究開発事業
(医療機器等に関する開発ガイドライン策定事業)

人工知能分野
開発 WG 報告書

平成 30 年 3 月

国立研究開発法人産業技術総合研究所

平成 29 年度 人工知能分野 開発 WG 委員名簿

(※は座長, 五十音順, 敬称略)

氏名	所属
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
椎名 毅	京都大学大学院 医学研究科人間健康科学系専攻
清水 昭伸	東京農工大学大学院 工学研究院
縄野 繁	国際医療福祉大学 医学部 三田病院 放射線診断センター
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
藤田 広志	岐阜大学 工学部 電気電子・情報工学科
古川 浩	東芝メディカルシステムズ株式会社 経営企画部
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
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
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
横井 英人 (※)	香川大学医学部附属病院 医療情報部

開発 WG 事務局

鎮西 清行 産業技術総合研究所 生命工学領域 健康工学研究部門 副研究部門長

坂無 英徳 産業技術総合研究所 情報・人間工学領域 人工知能研究センター 研究チーム長

人工知能分野開発 WG 委員会 開催日程

第 1 回開発 WG 委員会

開催日 平成 29 年 8 月 3 日 (木)

第 1 回審査・開発 WG 合同委員会

開催日 平成 29 年 9 月 28 日 (木)

第 2 回開発 WG 委員会

開催日 平成 29 年 9 月 28 日 (木)

第 3 回開発 WG 委員会

開催日 平成 29 年 11 月 17 日 (金)

第 2 回審査・開発 WG 合同委員会

開催日 平成 29 年 12 月 11 日 (月)

第 4 回開発 WG 委員会

開催日 平成 29 年 12 月 11 日 (月)

第 5 回開発 WG 委員会

開催日 平成 30 年 1 月 23 日 (火)

第 6 回開発 WG 委員会

開催日時：平成 30 年 2 月 19 日 (月)

目次

1. 当該技術分野の概要および現状.....	1
2. 当該技術分野におけるガイドライン策定の意義.....	2
3. ガイドラインの検討過程.....	3
3.1 ガイドラインの位置付けや作り方について.....	3
3.1.1 人工知能分野 開発 WG 委員会概要.....	3
3.1.2 ガイドライン作成方針の検討.....	10
3.1.3 単体プログラムとしての適用.....	11
3.2 AMED「ICT 技術や人工知能（AI）等による利活用を見据えた、診療画像等データベース 基盤構築に関する研究」について.....	12
3.3 適応対象や使われ方について.....	13
3.4 システムの性能管理について.....	15
3.4.1 検討の観点.....	15
3.4.2 学習方法.....	18
3.4.3 性能評価のタイミング×主体者×方法.....	19
3.4.4 学習によって性能が低下した場合の対応.....	20
3.4.5 学習機能の実行環境とシステムの運用環境とが異なる場合の検証方法.....	21
3.5 医療機器等法に対する考え方について.....	21
3.6 データについて.....	21
3.6.1 データの品質管理について.....	21
3.6.2 データの撮像条件の記録.....	24
3.6.3 データの品質を評価する方法.....	24
3.6.4 学習データ作成機能をシステムが有する場合の対応.....	24
3.6.5 データの保管期限.....	26
3.6.6 データの収集と流通について.....	27
参考資料.....	28
APPENDIX: 用語について.....	29

1. 当該技術分野の概要および現状

深層学習(Deep learning)と呼ばれる多層ニューラルネットワーク、とりわけ、畳み込みニューラルネットワーク(Convolutional Neural Network; CNN)の発展によって、画像自体を直接ニューラルネットワークに入力することで画像のクラス分類を行う手法の開発が急速に発展してきた。また、Fully Convolutional Network (FCN)など、入力、出力ともに画像となるニューラルネットワークも開発され、医用画像処理の分野では数多く利用されている。

畳み込みニューラルネットワーク(CNN)は、その起源を福島らによるNeocognitronとするものであるが、AlexNetなどImageNetデータベースを対象とした認識で、それまでの認識性能を大きく上回るニューラルネットワークシステムが開発されてきた。VGG16、ResNet、DenseNet、YOLO9000など数多くの畳み込みレイヤーを持つニューラルネットワークが開発されてきている。

また、FCN型でのend-to-end(入力データに対する前処理などを行わずに、入出力の関係を単一のモデルで直接的に学習および認識すること)のニューラルネットワークとしては、U-netやV-netのようなニューラルネットワークが開発されてきており、画像のセマンティックセグメンテーションなどに利用されている。

もちろんコンピュータ支援画像診断の分野においても、ニューラルネットワークなどの機械学習を用いた手法は数多く利用されるに至っている。その目的としては、(1)入力される画像を正常、あるいは、異常に分類する画像分類、(2)画像から異常が疑われる部位を検出する異常部位検出、(3)画像において肝臓、心臓、膵臓などといった臓器領域を抽出する画像セグメンテーション、などがあげられる。

これらの目的を達成するためには、従来型の機械学習を用いた方法、あるいは、最近急速に広まっている多層ニューラルネットワークを用いた方法が用いられる。

前者の方法は、比較的古くから行われており、画像から何らかの方法で特徴量を抽出し、その特徴量を3層ニューラルネットワーク(第1層は単なる入力なので、2層ニューラルネットワークと書かれる場合もある)あるいは、サポートベクタマシン(Support Vector Machine)などの分類器によって分類するものである。ここでは、画像全体から特徴量を抽出することもあれば、フィルタ処理などによって得られた異常部位候補領域から得られる平均濃度値、濃度値の分散、面積、体積、直径などの特徴量を抽出する場合もある。例えば、Moriらは超拡大内視鏡画像からLocal Binary Pattern特徴を抽出し、その特徴量をサポートベクタマシンで、大腸ポリープの類型判別を行っている。また、CT画像から抽出された血管領域の各枝から特徴量を算出し、それを条件付き確率場と呼ばれる機械学習の方法で、解剖学的名称を付与する研究などもある。

後者の方法によるものとしては、Itohらは、前出の超拡大内視鏡画像の分類において、CNNを用いた画像分類を導入している。また、腫大リンパ節検出など異常部位の検出においても、CNNが利用されている。医用画像からの臓器セグメンテーションにおいて、FCNを用いているもの、FCNの形態にSkip Connectionを導入した3D U-netを用いるものなどが利用されている。

医用画像処理分野におけるトップカンファレンスである MICCAI (Medical Image Computing and Computer Assisted Intervention) 2017 においては、その約半数が深層学習に関連するものである。また、放射線分野を代表する国際会議 RSNA (Radiological Society of North America) 2017 においては、Subspecialty 分野として Machine Learning が設定されている。このように、医用画像認識とそれを利用する CAD の分野では、急速に機械学習の利用が進んでいる。このような背景から、AI、とりわけ、機械学習を利用した CAD システムに関するガイドラインの策定は急務であると考えられる。

2. 当該技術分野におけるガイドライン策定の意義

既に2回(2012,2015年度版)の発行がなされているコンピュータ診断支援装置(Computer Aided Diagnosis/Detection; CAD)の開発に関するガイドラインに於いて、次の作業として挙げたのが、人工知能(Artificial Intelligence; AI)技術を用いた医療機器の開発に関するガイドライン策定であった。我々は、前述のCADの開発ガイドラインの執筆グループに、AIを用いて開発を行っている研究者や企業を加えて、本ガイドラインの策定を開始した。

初年度、議論の中で最も多くの時間を割いたのは、医療機関が個別にAIを事後学習させ、メーカーの関与がない状態で、これの性能向上もしくは特定の疾患への特化を行うような利用の仕方であった。この利用方法は、これまでの医療機器を提供するメーカーが性能を確定した状態で出荷し、また必要に応じてバージョンアップするなどの作業を通して、その性能をメーカーが管理している状況下で運用する利用方法とは一線を画するものである。

AIは学習の仕方を誤れば、その性能を大きく損なう可能性がある。仮に事後学習を医療機関が独自に行う場合、行う者はAIの特性や学習データの信頼性などに知識を持っている必要があり、その意味では、当該医療機器のユーザーも、本ガイドラインの対象と考える必要がある。

また、AIの計算過程はブラックボックス的であるため、その性能はテストデータにより確認する方法が有力であり、頻回のテストが実施される可能性も考慮しなくてはならない。議論の中では頻回のテストに用いることによってテストデータが学習の対象になってしまう危険性なども指摘され、これを防ぐためには大量のテストデータの存在が必要であるという意見が出された(そのテストデータでよい成績を出した機器でなければ、テストに合格しないことで、当該テストデータ自体が学習結果に反映された形になってしまう)。

ガイドライン作成事業の初年度に於いて、我々はAIを用いてCADを開発するときの手法について議論した。CADが最も普及している放射線のみならず、動画像を主とする内視鏡、超音波等についても議論を行い、新たな医療機器の開発に際しても、開発者が開発戦略を立てやすいようなガイドラインをまとめようと試みた。このような形で準備したガイドラインが、今後のCAD開発の一助になることを望むところである。

3. ガイドラインの検討過程

3.1 ガイドラインの位置付けや作り方について

近年、各分野で開発が進む人工知能分野について、早期に我が国での医療機器での活用が課題となっている。人工知能分野での装置を開発するために必要なガイドライン策定を検討することを目的に、開発ガイドライン策定ワーキンググループ（以下、開発 WG）委員会を設置し、本年度は、下記に示すとおり 6 回の会合を開催した。人工知能分野は様々な臨床応用も考えられるため、次世代医療機器・再生医療等製品評価指標作成事業 人工知能分野 審査ワーキンググループ（以下、審査 WG）との合同会議も開催し、現状での企業の人工知能分野における開発動向をヒアリングし、CAD 装置として市場でも実績があり、評価指標および開発ガイドラインの作成実績のある画像診断分野で先行し、評価指標およびガイドラインの検討を進めることとなった。

3.1.1 人工知能分野 開発 WG 委員会概要

(1) 第 1 回開発 WG 委員会

- 開催日時：平成 29 年 8 月 3 日(木) 16:00～18:00
- 出席者：
 - 委員：横井英人、片倉由紀子、椎名毅、清水昭伸、縄野繁、野波徹緒、古川浩、宮崎靖、諸岡直樹
 - オブザーバー：
 - 日本医療研究開発機構：浅沼直樹、関口満
 - 経済産業省：中島教順
 - 国立医薬品食品衛生研究所：中岡竜介、加藤玲子
 - 医薬品医療機器総合機構：中里俊章、宮本大誠
 - 日本画像医療システム工業会：田中弘
 - オリンパス株式会社：吉田眞司
 - 事務局：鎮西清行、坂無英徳
- 配布資料：
 - 資料 1: 委員名簿
 - 資料 2: 医療機器開発ガイドラインについて
 - 資料 3: CAD における評価指針・開発ガイドラインの制定経過とその課題
 - 資料 4: 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
 - 資料 5: 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 (手引き)
 - 資料 6: 次世代医療機器・再生医療等製品評価指標作成事業の概要と人工知能分野 審査 WG 活動計画書

- 資料 7: 計算機支援診断の研究・開発動向
- 資料 8: AI 技術を利用した画像診断のこれまでの取り組みや動向について
- 資料 9: AI 関連技術の医療機器への応用例と課題
- 資料 10-1: 「保健医療分野における AI 活用推進懇談会」の関連資料
- 資料 10-2: 「PMDA 科学委員会 AI 専門部会」の関連資料
- 資料 11: 総務省 AI ネットワーク社会推進会議 報告書 2017 の概要
- 資料 12: 想定される論点の案
- 資料 13: 今後の開催日程案およびメーリングリストについて

● 議事概要

1. 開発ガイドライン事業について
2. 話題提供
 - CAD 開発ガイドラインについて
 - 審査 WG について
 - 画像診断支援の研究動向について
 - 医療用 AI に対する産業界の取り組みについて
 - 医療用 AI に関する政府の検討状況について
3. 開発ガイドラインに関する検討の進め方および役割分担
4. その他
 - 今後の開催日程案について
 - メーリングリストについて

(2) 第 1 回審査・開発 WG 合同委員会

- 開催日時：平成 29 年 9 月 28 日(木) 14:00～16:00
- 出席者：

〈審査 WG〉

委員：橋爪誠（座長／九州大学）、佐々木毅（東京大学）、清水昭伸（東京農工大学）、
田中聖人（京都第二赤十字病院）、中田典生（東京慈恵会医科大学）、
藤田広志（岐阜大学）、待鳥詔洋（国立国際医療研究センター国府台病院）、
横井英人（香川大学）

厚生労働省：青柳ゆみ子、他同行者 3 名

内閣官房 健康・医療戦略室：岡本利久、他同行者 2 名

医薬品医療機器総合機構：岡崎譲、加藤健太郎、古森亜矢、谷城博幸、宮崎生子、今
川邦樹、江原輝喜、若尾りか

事務局：齧島由二、中岡竜介、加藤玲子（国立医薬品食品衛生研究所）

〈開発 WG〉

委員：横井英人※、片倉由紀子、椎名毅、清水昭伸※、縄野繁、野波徹緒、
藤田広志※、古川浩、宮崎靖、諸岡直樹、森健策（Skype による遠隔参加）

（※：審査 WG 委員兼任）

オブザーバー：

経済産業省：中島教順

日本医療研究開発機構：浅沼直樹

医薬品医療機器総合機構：宮本大誠

日本画像医療システム工業会：田中弘

オリンパス株式会社：吉田真司

事務局：鎮西清行、坂無英徳

● 配布資料：

資料 1：議事次第（案）

資料 2：出席者名簿（審査 WG）

資料 3：出席者名簿（開発 WG）

資料 4：講演資料 1「個人情報保護法の概要」

資料 5：講演資料 2「医療分野の研究開発に資するための特許加工医療情報に関する法律について」

資料 6：人工知能を利用した画像診断機器に関する評価指標案たたき台

● 議事概要

1. 話題提供

- 医療関連分野ガイダンス等について
- 次世代医療基盤法について

2. 全体討議

- 進捗状況確認及び評価指標案たたき台に関する討議
- 進捗状況、今後の方向性確認及び審査 WG との作業調整討議

(3) 第 2 回開発 WG 委員会

● 開催日時：平成 29 年 9 月 28 日(木) 16:30～17:00

● 出席者：

委員：横井英人、片倉由紀子、椎名毅、清水昭伸、縄野繁、野波徹緒、

藤田広志、古川浩、宮崎靖、諸岡直樹、森健策（Skype による遠隔参加）

オブザーバー

日本医療研究開発機構：浅沼直樹

国立医薬品食品衛生研究所：葩島由二、中岡竜介、加藤玲子

医薬品医療機器総合機構：宮本大誠

日本画像医療システム工業会：田中弘

オリンパス株式会社：吉田眞司

事務局：鎮西清行、坂無英徳

● 配布資料：

資料 1: 平成 29 年度 医療機器開発ガイドライン事業 人工知能分野 開発 WG 第 2 回会議 議事概要（案）

資料 2: 放射線診断の現状とデータベースについて

資料 3: Artificial Intelligence in Medical Image Processing

資料 4: 今年度の活動に関する事務局案

● 議事概要

1. 話題提供「診断支援技術や医療用 AI に対する期待」
2. 話題提供「内視鏡診断支援システムに関する研究紹介」
3. 今後の方向性に関して

(4) 第 3 回開発 WG 委員会

- 開催日時：平成 29 年 11 月 17 日(金) 10:00～12:18

● 出席者：

委員：横井英人、片倉由紀子、椎名毅、野波徹緒、藤田広志、古川浩、
宮崎靖、諸岡直樹

オブザーバー：

日本医療研究開発機構：岩田倫明、関口満

経済産業省：中島教順

国立医薬品食品衛生研究所：中岡竜介、加藤玲子

医薬品医療機器総合機構：宮本大誠、中里俊章

日本画像医療システム工業会：田中弘

オリンパス株式会社：吉田眞司

事務局：鎮西清行、坂無英徳

● 配布資料：

資料 1. 議事次第案

資料 2. 第 2 回 開発 WG 会議 議事概要（案）

資料 3. 第 1 回 合同会議 議事概要（案）

資料 4. 審査 WG からの評価指標たたき台に対するコメント

- 資料 5. 評価指標たたき台に対するコメント
- 資料 6. AI 開発ガイドラインでの検討項目について
- 資料 7. 審査 WG 評価指標案へのコメント
- 資料 8. 開発ガイドラインの纏め方に関して
- 資料 9. 開発ガイドラインにおいて明確化すべきと思われる事項
- 資料 10. 人工知能分野におけるデータの扱い方
- 参考資料 1. 人工知能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案
(10/27 版)
- 参考資料 2. 開発 WG からのコメントに関する評価指標案 (10/27 版) での対応状況
表

- 議事概要

1. 報告書に関する討議
2. 開発ガイドラインに関する討議
3. 今後の開催日程案について

(5) 第 2 回審査・開発 WG 合同委員会

- 開催日時：平成 29 年 12 月 11 日（月）17:13 -18:04
- 出席者：

〈審査 WG〉

委員：橋爪誠（座長/九州大学）、佐々木毅（東京大学）、清水昭伸（東京農工大学）、
田中聖人（京都第二赤十字病院）、中田典生（東京慈恵会医科大学）、
藤田広志（岐阜大学）、待鳥詔洋（国立国際医療研究センター国府台病院）、横
井英人(香川大学)

厚生労働省：青柳ゆみ子

医薬品医療機器総合機構：岡崎譲、菅原明彦、加藤健太郎、桐山瑤子、
古森亜矢、宮崎生子、今川邦樹

事務局：齋島由二、中岡竜介、加藤玲子（国立医薬品食品衛生研究所）

〈開発 WG〉

委員：横井英人※、片倉由紀子、清水昭伸※、野波徹緒、藤田広志※、古川浩、
宮崎靖、森健策（Skype にて遠隔参加）、諸岡直樹（※：審査 WG 委員兼任）

オブザーバー：

日本医療研究開発機構：扇谷悟、浅沼直樹、岩田倫明

経済産業省：尾畑英格

医薬品医療機器総合機構：中里俊章、宮本大誠

日本画像医療システム工業会：田中弘

オリンパス株式会社：吉田眞司

事務局：鎮西清行、坂無英徳

● 配布資料：

資料 1：議事次第（案）

資料 2：出席者名簿（審査 WG）

資料 3：出席者名簿（開発 WG）

資料 4：人工知能技術を利用した画像診断支援システム等に関する評価指標修正案

資料 5：開発 WG における検討状況

● 議事概要

1. 審査 WG 評価指標案に関する確認・討議
2. ガイドライン案の進捗状況に関する報告・討議
3. その他（今後の予定等）

(6) 第 4 回開発 WG 委員会

- 開催日時：平成 29 年 12 月 11 日(月) 18:08～18:45

● 出席者

委員：横井英人、片倉由紀子、清水昭伸、野波徹緒、

藤田広志、古川浩、宮崎靖、諸岡直樹

オブザーバー：

橋爪誠（九州大学、審査 WG 座長）

日本医療研究開発機構：扇谷悟、浅沼直樹、岩田倫明

経済産業省：尾畑英格

国立医薬品食品衛生研究所：齋島由二、中岡竜介、加藤玲子

医薬品医療機器総合機構：中里俊章、宮本大誠

日本画像医療システム工業会：田中弘

オリンパス株式会社：吉田眞司

事務局：鎮西清行、坂無英徳

● 配布資料：

資料 1. 議事次第（案）

資料 2. 平成 29 年度 医療機器開発ガイドライン事業 人工知能分野 開発 WG 第 3 回会議 議事概要（案）

資料 3. 医療機器等に関する開発ガイドライン（手引き）策定事業 平成 29 年度 人工知能分野 開発ワーキンググループ 活動報告書 目次案

資料 4. 開発ガイドライン目次 事務局案

資料 5. CAD 設計開発ガイドライン／CAD 性能評価ガイドライン目次 JIRA 案

● 議事概要

1. 報告書に関する討議
2. 開発ガイドラインに関する討議
3. 今後の開催日程案について

(7) 第 5 回開発 WG 委員会

- 開催日時：平成 30 年 1 月 23 日(火) 16:00～18:20

● 出席者

委員：横井英人、片倉由紀子、清水昭伸、縄野繁、野波徹緒、
藤田広志、古川浩、宮崎靖、諸岡直樹

オブザーバー

日本医療研究開発機構：関口満、浅沼直樹、岩田倫明

経済産業省：中島教順

国立医薬品食品衛生研究所：中岡竜介、加藤玲子

日本画像医療システム工業会：田中弘

オリンパス株式会社：吉田眞司

事務局：鎮西清行、坂無英徳

● 配布資料：

資料 1. 議事次第案

資料 2. 第 4 回 開発 WG 会議 議事概要（案）

資料 3. 活動報告書 骨子案

資料 4. 開発ガイドライン目次案（2017/12/24 版）

参考資料 1. 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
（案）（2018/1/19 版）

● 議事概要

1. 活動報告書に関する討議
2. 開発ガイドラインに関する討議
3. 今後の開催日程案について

(8) 第 6 回開発 WG 委員会

- 開催日時：平成 30 年 2 月 19 日(月) 16:00～18:45

- 出席者

委員：横井英人、片倉由紀子、清水昭伸、縄野繁、野波徹緒、
藤田広志、古川浩、宮崎靖、諸岡直樹、森健策（Skypeにて遠隔参加）

オブザーバー

日本医療研究開発機構：浅沼直樹、岩田倫明

国立医薬品食品衛生研究所：加藤玲子

日本画像医療システム工業会：田中弘

オリンパス株式会社：吉田真司

事務局：鎮西清行、坂無英徳

- 配布資料：

- 資料 1. 議事次第案
- 資料 2. 第 5 回 開発 WG 会議 議事概要（案）
- 資料 3. 活動報告書案
- 資料 4. 開発ガイドライン 事務局案
- 参考資料 1. 審査指標案（2018/2/16 版）

- 議事概要

1. 活動報告書に関する討議
2. 開発ガイドラインに関する討議
3. 来年度の活動内容について

3.1.2 ガイドライン作成方針の検討

CAD においては、厚生労働省から「コンピュータ診断支援装置に関する評価指標（以下、CAD 評価指標）」、経済産業省から「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」と「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」が発行されている。「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」は、CADx（コンピュータ診断支援装置）を対象としている。当初、これに先行し、CADE（コンピュータ検出支援装置）の性能評価開発ガイドラインの発行が企画されていた。「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」の発行後、内容を修正し、CADE の性能評価開発ガイドラインが発行される予定であったが、最終的には発行には至っていない。但し、CAD 評価指標においては、元々CADx、CADE の区別はなく、開発ガイドラインとして、CADx を念頭においた「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」の考え方は、CADE においても適用可能なものである。

開発 WG としては、人工知能分野の開発のあり方について議論を重ねた。WG での議論を最終的にガイドラインとしてまとめるにあたり、ガイドラインの作成方針の検討を行った。画像診断分野の場合、既存の CAD ガイドラインが 2 つ出されているので、AI を導入することにより変わることを、

また近年の新規格の引用などの検討を行った。たとえば、設計および開発プロセスの適用のところで、IEC62304 の適用の考え方を整理し、IEC82304 への言及を追加が必要であるとの意見があった。また、CAD の性能評価ガイドラインの APPENDIX の 4 番において、AI における教師なし／教師ありデータと、CAD のゴールド・スタンダードの関連性を整理する必要があるとの意見があった。また、先日の北米放射線学会（RSNA）などでの顕著な動向として、AI を活用した CAD に関する新興企業の増加が著しい。日本でも同様のスタートアップ企業はふえているので、全部統合した方がわかりやすいのではないかとの意見も多かった。日本医療研究開発機構（AMED）及び経済産業省としても編纂方針については異論なく、委員会で意向に従うとの裁可となり、委員会としては、今年度、既存の CAD 開発ガイドラインをアップデートする形で、人工知能技術を利用した医用画像診断支援システム開発ガイドライン案（以下、AI+CAD 開発ガイドライン案）を作成する方針となった。

CAD に AI が加わることによって変わること

診断支援という大きな使用目的は変わらないが、学習によって性能が変化する特徴から、①性能の定義、②学習データの扱い、③性能変化に対する評価、④市販後の性能管理等における考慮が必要となる。医療機器として考慮すべき評価指標は主に以下の通りとなる。

主な観点	従来 of CAD	AI 導入後の CAD
性能	CAD 評価指標	CAD 評価指標 + AI+CAD 評価指標 ^(注)
学習データ	n/a	AI+CAD 評価指標
市販前の性能評価	CAD 評価指標	同左
市販後の性能変化に対する評価	n/a	AI+CAD 評価指標
市販後の性能変化に対するリスクマネジメント	n/a	AI+CAD 評価指標

(注) AI+CAD 評価指標 : 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、平成 30 年 3 月末日発行

3.1.3 単体プログラムとしての適用

CAD の開発ガイドラインは、単体プログラムが医療機器化された平成 26 年 11 月 25 日改正施行以前に作成されているが、発行時期から、単体プログラムへの適用はある程度考慮されており、平成 26 年 11 月 25 日発行の薬食機参発 1125 第 6 号「医療機器プログラムの製造販売認証申請における取扱いについて」に従って、薬事申請書及び添付資料を対応させることで、基本的には対応可能である。なお、詳細な情報は、「単体プログラムの取扱いに関する厚生労働科学研究班；医療機器に関する単体プログラムの薬事規制のあり方に関する研究（研究代表者：公益財団法人医療機器

センター理事長 菊地眞)」および「AMED 医薬品等規制調和・評価研究事業：医療機器に関する単体プログラムの薬事規制のあり方に関する研究（研究開発代表者：公益財団法人医療機器センター専務理事 中野壮陸）」により検討が行われており、公益財団法人医療機器センターの Web サイトに情報がまとめられているので、参考にされたい（参照 URL: <http://www.jaame.or.jp/mdsi/program.html>）。

また、単体プログラムに対しては、近年 IEC82304-1「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」が制定され、JIS 化が進んでいる。この規格では、製品の要求事項及びその検証、製品のバリデーション¹、識別及び付属文書、市販後活動等が定められており、当ガイドラインが発行する頃には、JIS 制定が見込まれている。このため、このガイドラインには、人工知能分野の製品の開発を進めるにあたっては、IEC82304-1 の適用を推奨している。

3.2 AMED「ICT 技術や人工知能（AI）等による利活用を見据えた、診療画像等データベース基盤構築に関する研究」について

CAD の研究開発には、学習および性能確認²、テストのそれぞれの段階で大量のデータが必要である場合が多い。研究者および開発者は、公的ならびに準公的な既存のデータベースなどを使用するか、自ら新たにデータを収集するなどして、これらのデータを準備しなくてはならない。

現在、日本医療研究開発機構（AMED）委託研究として行われているデータベース構築研究に於いて、医療情報にかかる課題についての議論が行われている。

AI 開発を目して、複数の学会により AMED 研究が行われているが、データ収集方法はそれぞれ異なっている。改正個人情報保護法・次世代医療基盤法での運用方法の不確定部分があるため、医療機関からのデータ収集方法や匿名化方法、またその後のデータ保存や利活用方法について、統一的な議論ができないことはやむを得ないところではある。しかし、データ収集に協力する個々の医療機関は、それぞれの方法に応じて個別にシステム接続の対応をせざるを得ず、それらを統合的に管理することが難しいことが問題点として挙げられている。

現在、文字およびコード情報を主とする検体検査結果や病名、処方などの臨床情報は厚生労働省が推奨する標準的なデータ記録方法等により、その収集、運用について徐々に統一的な運用が成立しつつある。しかし、諸外国に比べ、個々の医療機関の規模が大きい日本に於いては、多くの数の医療機関からデータを収集しなくてはならず、ビッグデータの収集方法統一は喫緊の課題である。

CAD の開発および研究に際して、それぞれの画像の特性や研究の方向性に応じたデータの収集

¹ 製造所の構造設備並びに手順、工程その他の製造管理及び品質管理の方法が期待される結果を与えることを検証し、これを文書とすること（「医薬品及び医薬部外品の製造管理及び品質管理規則」（平成 16 年 12 月 24 日、厚生労働省令第 179 号））。

² 学習によってパラメータが変更されたモデルについて、テストデータではなく、validation（確認用）データを用いて性能を推定する処理。英語で Validation と表される処理について、GMP 省令における「バリデーション」との混同を避けるため、本報告書では「性能確認」を用いる。

方法や運用方法があることを否定するところではないが、当分野が早期に、より高度な発展をするために、データ収集や運用方法に統一的な方針を早い段階で見いだす必要がある。いわゆるビッグデータ基盤に関与する関係者は、より「横断的」な運用方針の確立に対する議論を行うべきであろう。また、それぞれの研究者には、個別にデータを収集する場合でも、そのような方針に注意を払い、これへの追従を念頭に置く必要があると考える。本事項は、医療機器の性能に直接関与する内容ではないが、情報処理や管理コストの適正化のためにも考慮すべき事項であろう。

3.3 適応対象や使われ方について

従来のコンピュータ診断支援システムでは、静止画像を主な解析対象とし、放射線画像診断において、病変と想定される部位の抽出（CADE の場合）や、異常部位の良悪性の鑑別（CADx の場合）を行った結果を「第 2 の意見」として提供する事による医師の診断支援を想定していた。現時点では、CAD の結果を一次スクリーニング時の参考にする等の使い方の拡張が検討されている。さらに、画像解析技術の進歩により、内視鏡や超音波などの動画をその場で解析する事が可能になり、患者の診察中にその場で何らかの異常可能性がある部分にその場で情報提示するような使い方も想定されるようになった（元の画像と AI の結果を一緒に見ながら、その場で医師が判断する同時読影（Concurrent）な使い方）。内視鏡や超音波などの動画では画像が操作者のレベルに依存する点も考慮が必要であり、静止画とは違った考え方が必要となる。

病理診断への人工知能の適用も報告されているが、病理診断への人工知能の実用化にはまだ時間がかかると想定される為、今回の開発ガイドラインでは対象外とした。

2018 年 2 月に AI アルゴリズムを利用して画像を解析し、医療関係者へのアラートを表示して医療関係者の意思決定を支援するソフトウェア（Clinical Decision Support Software。コンピュータが医療関係者のトリアージ業務をサポートするデバイス）が FDA に認可された。アラート表示だけの場合には、病変の検出や鑑別を目的としてはいないので、今回評価指標が策定された医用画像診断支援システム（CAD）に該当するとはいえず、今回の開発ガイドラインでは対象外とした。2012 年に発行された FDA の CAD の 510k 申請に関するガイドライン³においても、「CAD のうち、Computer Aided Triage（CAD のうち、医師ほか医療従事者の作業低減または省略を目的として、正常判断されたデータに対しては医師による読影を要さないもの）を目的とするものを除く」との記載がされている。FDA は 2017 年 12 月に Clinical and Patient Decision Support Software に関するドラフトガイダンスを発行しており、今後は、順番を決める等のトリアージ業務など、コンピュータが医師の意思決定をサポート（最終判断は医療関係者）するシステムがトレンドとなっていく事が想定される。今後このようなソフトウェアの扱いも決めていく事が望ましい。

開発ガイドラインの中で CADE と CADx をどう扱うべきかについて議論があった。前述のような

³ US FDA: Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data - Premarket Notification [510(k)] Submissions - Guidance for Industry and Food and Drug Administration Staff, Available at <https://www.fda.gov/MedicalDevices/ucm187249.htm>, Issued on: July 3, 2012.

適用範囲拡大や使われ方の多様化が想定され、CADxに関する検討事項が CADe について検討すべき事項を網羅しているとは現時点では言い難い。CADe と CADx は技術的には検出対象か否かの2値分類と、良性あるいは悪性、その他などの多値分類の違いとなり、分類の内容によっては明確な区別が難しいケースも想定される。また、市販後の性能変化という新たな観点の追加は CADe か CADx かに依らない要件である。これからの議論から、CADe と CADx を区別する必然性もなく、CAD として一律に扱うべきとなった。

事後学習においてユーザーが主体的に性能変化に関与することが可能な診断支援システムを対象にするかについて、性能変化に伴う品質管理をメーカーが行うことが難しいことから、審査 WG において AI+CAD 評価指標の対象外の扱いとなり、別添にて考え方を纏めることになった。並行して本 WG でも多くの議論が行われた。

事後学習により性能向上した場合、医師はより信頼性の高い診断支援を受けられるようになり、その結果患者はより高度な医療を享受できるようになる。しかし、劣性（システムの性能が、事後学習前の性能や、科学的合理性に基づいて承認時に設定した下限を下回ること。§ 3.4.1 を参照）となる可能性があることから性能管理の徹底が条件となる。学習を含む性能管理の主体は、現時点ではメーカーが望ましく、特に深層学習の様にブラックボックスで、システムの入出力関係が高度に非線形なシステムの場合には、学習過程の性能評価と管理が重要となるため、システムの内容を理解している技術者が学習データの内容をよく理解したうえで実施するのが適当だと考えられる。

近年、オープンソースソフトウェア（OSS）やクラウド型計算資源の提供サービスなどの開発環境が整備され、工学系研究者のみならず最終利用者である医師等の医療従事者が独自の人工知能技術を備えたシステムを構築することが想定される。このようなケースではユーザーが主体的に学習や性能管理を実施することになるが、医療従事者や AI 自身が安全に再学習を実施できるようになるための技術的進化が必要と考えられる。

本 WG 会合では、再学習によって変化する性能の管理を徹底するという条件下で事後学習を認めるという意見があった。再学習による性能向上によって、医師はより信頼性の高い診断支援を受けられるようになり、その結果患者は、より高度な医療を享受できる為である。また、ネットワーク構造の本質的ではない点の変更（たとえば、一部のパラメータの最適化）についても、再学習の際に変更を認めるという案である。

人工知能を利用した支援システムにおいては、①どこで誰が学習させるか、②誰がどのような使い方をするかによって、誰がどの範囲に責任⁴を持つかという事も変わる事が想定される。

⁴ ここでは主に道義的責任について述べており、法的な責任については本 WG では議論をしていない。

(例) 性能変化に対する責任の所在

事後学習のさせ方	性能変化に対する責任の所在
メーカーが社内で学習	性能変化はメーカー責任
ユーザーが医療機関で学習	現時点では以下のような議論がなされた ①メーカーは性能変化が承認された規定範囲内である事を検証する手段の提供に責任を負う。 ②ユーザーはメーカーから提供された手段を使って性能を検証し、規定範囲内で利用する事に責任を負う。 詳細については、AI+CAD 評価指標を参照のこと。

適用範囲や使われ方が変わる場合、ユーザーと CAD の責任範囲も変わる事が想定される。メーカーの責任範囲は、意図する用途やリスクによって変わる為、装置毎に明確にすることが必要である。例えば、検診等で1次読影(First Reader)で利用する場合には、医師は装置の性能を理解した上で CAD 結果を参考にして検診を行い、診断結果に責任を負う等が考えられる。いずれの場合も、CAD の結果を利用した診断に関しては医師の責任となる。“評価指標 (案) 5.支援システム等で想定される主な課題と解決の方向性 (3) 責任の所在”も確認の上明確にしておくこと。

3.4 システムの性能管理について

3.4.1 検討の観点

機械学習における性能管理の概要

「学習」は、設計した学習アルゴリズムを実行する「学習プログラム」で学習用データと教師情報⁵からその特徴を学習し、「学習済みモデル」を得る工程である。学習済みモデル（以下、単にモデルと称す。深層学習を想定した場合、ネットワーク構造とパラメータを含む）は診断アルゴリズムとして医療機器に実装される。ここで言う性能管理は、学習済みモデル（診断アルゴリズム）の規定される診断支援性能（事後学習するシステムでは変化後の性能含む）、および支援システム等としての品質、安全性、有効性を保証・維持するために必要な事項と定義される。具体的には学習工程、性能評価工程、評価後の対応等の下記に例示した各工程に渡る管理が必要となる。

- 学習工程：申請時に性能を規定する処理部（診断アルゴリズム）を開発する工程。そのために必要な学習プログラム、学習データセット、モデル、およびそれらを用いて実施した一連の学習履歴等が管理対象に含まれる。事後学習を行うシステムでは、追加するデータセット、事後学習後のモデルや学習を実施するタイミング、学習履歴等が管理対

⁵ 教師データなしの学習アルゴリズムも存在するが、教師に相当する情報あるいは仮定が全くないものは登場していないと考えられる。

象に含まれる。学習過程では、学習データとは別に用意した性能確認データセットを用いてパラメータ等のチューニングが行われる。学習履歴は、例えば外部導入した学習済みモデルを用いて自己の管理するデータセットで追加学習を実施した等の一連の情報管理を指す。

- 性能評価工程：モデルの性能を評価および検証する工程。評価対象のモデル、評価に用いるテストデータセット、評価結果等が管理対象に含まれる。事後学習を行うシステムでは、事後学習後のモデルとその評価結果が管理対象に加わり、申請時に規定した性能に対する事後学習後モデルの性能変化を評価する。
- 性能評価後の対応：性能評価の結果を受けて必要な対応を実施する工程。学習履歴と紐づけたモデルの性能情報管理、診断アルゴリズムとしてのシステムへの登録、登録履歴等が管理対象に含まれる。事後学習を行うシステムでは、最終的に診断に用いるモデルの登録、登録履歴管理等が対象に含まれる。特に、事後学習後のモデルが劣性となった場合は、診断アルゴリズムとして登録せずにさらに追加学習する、学習以前のモデルに戻す、あるいは事後学習前のモデルを併存して結果を出力する（劣性となっている事実についてユーザーへの強い注意喚起が必要）、などの対応が想定されるため、システムの仕様に応じた管理が必要となる。

ユーザーが性能管理を主体的に実施するシステムにおいては、上記に示した管理を実施するための支援機能（例えば管理ツール）等をメーカーが提供することが必要となる。

原理および機能に関する観点

原理的には、診断アルゴリズムの性能を実現した解析アルゴリズムを明確化する。ネットワーク構造や学習工程の概要説明（アーキテクチャ上の特徴、採用した学習方法や使用したデータなど）、事後学習を実施するシステムでは学習のタイミングなどが必要となる。また、システムを利用する上でのリスク、再事後学習等によって変化するもの（性能、ネットワークパラメータ等）などの観点が必要となる。

機能としては、最終責任者である医師に対して提供する、診断支援の機能仕様を明確にすると共に、システムが介在することによって自動化される処理範囲や機能的な制限事項等を明示する。

性能範囲の規定と維持

事後の性能変化を意図したシステムでは、申請時の性能範囲の規定、および市販後の性能の管理と維持によって、診断支援システムの性能に基づく安全性や品質は担保される。

規定した性能はその性能指標値の妥当性、テストデータを含む性能検証方法の妥当性を示す。

市販後の性能変化は以下の場合があり、いずれの場合もメーカーが品質管理を実施し性能を維持することが望ましい。

- 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 評価指標を参照のこと）
- ② 事後学習により高頻度で性能変化する場合

システムの性能が、科学的合理性に基づいて承認時に設定した下限を下回らないようにすることが重要である。ここで科学的合理性に基づく設定とは、例えば、統計数理的や医学的な視点から設定することを指す。前者の場合は、標本データから求めた性能指標の統計的ばらつきの範囲を推定し、それに基づいて下限を設定する。後者の場合には、医学的観点から意味のある性能変化の範囲を合理的に決定し、それに基づいて下限を設定する。なお、性能は、承認時のテストデータのみならず、さらに多くのデータを用いてテストすることも想定される。初回の承認時と同じテストデータ⁶を用いてテストを行い、承認時に設定した下限を下回らないようにする事が最低限必要である。また、下回った場合には、システムの停止、関係者への速やかな連絡、性能復帰などの対策が必須となる。

一方、性能が有意に向上した場合には、支援システムとユーザーとの間の関係に質的变化が生じる可能性がある。例えば、システムの出力を過度に信頼し、結果として重要なケースの見落としや最終的な診断精度の低下に結び付く可能性がある。そのため、性能が統計的、あるいは、医学的な観点から有意に向上した場合には、ユーザーにその旨を知らせ、不適切な使用法につながらないように注意喚起をすることが重要である。

なお、システムの入出力の特徴や想定される問題点などをユーザーに対して事前に伝えることや、問題発生時の危機管理についてシミュレーションなどを通じてユーザーが事前に体験することは、システムの安全な運用のためには有効な手段である。企業や関連学会が、想定されるユーザーに対して教育やトレーニングを実施することを推奨する。

⁶ 承認時のデータ：承認を受ける際に当該システムの性能を評価したテスト用データ。学習データとは（深層学習の場合には性能検査データとも）区別して管理され、システムの学習や設計にはまったく用いられていないことが必須。何らかの理由でこの条件を満たさなくなった場合には、代替となるデータを用意して評価する方法が考えられるが、承認時のデータを用いた評価と代替データベースによる評価が同等であることを科学的合理性に基づいて示す必要がある。

テストデータは、学習や設計にまったく用いられないことが肝要である。例えば、同一時期に条件（学習データやモデル）を変えて10通りの事後学習を行い、事後学習後の10通りのシステム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システムを選択したり、もっとも性能が良いシステムを事後学習後のシステムとして選択して運用することは絶対に実施してはならない。なぜなら、システム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）ことになるためである。したがって、テストデータの管理と性能評価は、システム開発者とユーザーが担当することは難しく、理想的には第三者機関が実施し、システム開発者やユーザーには基準を満たしたか否かだけが分かるようにし、その結果を性能管理に利用することが望ましい（注：同等レベルのことが可能であれば、実施形態は上記の限りではない。例えばテストデータの暗号化を徹底し、使用履歴も完全に残し、開発者は処理結果の詳細を見られないように工夫できれば、開発者の下で適切に管理することも可能かもしれない）。また、性能評価は、原則として一度に一つのモデルのみを対象とすることが望ましい。

システムの特質に応じた性能管理

システムとしては、ユーザーサイトに設置された装置上で提供される（オンプレミス型⁷）か、ネットワーク等を介して提供される（クラウド型⁸）などの形態が想定される。クラウド型の場合には、集中管理可能であり性能向上を全体で利用できるメリットがあるが、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。オンプレミス型の場合には、サイト毎の管理が必要になり、性能変化の範囲は各サイトに限定されるので、各サイトの利用に合わせた性能変化が可能になる。性能管理も、各サイトの利用における許容範囲で管理ができれば良い。

また、ネットワークから切り離された状態で提供される場合は、メーカーがユーザーサイトへ定期的に訪問するなどの対応の他、性能変化をユーザーが把握するための通知機能などが重要となってくる。特に自律的な学習を意図したシステムでは、データ収集サイクルや学習タイミング等の観点、システムが備えている①学習に用いるデータの妥当性を確認する手段、②性能検証方法、③性能が劣性となった場合の動作、④対応方法の観点が必要となってくる。

一方で、ユーザーが主体となって学習することを意図したシステムも想定される。メーカーは学習機能以外に、性能管理機能も合わせてユーザーに提供する必要がある。具体的には、顧客の要望に対応するカスタマイズ機能などが想定される。

データ利活用の観点

学習データ、性能確認データ、テストデータ、それぞれ利用目的に応じた管理が必要である。特に、深層学習ではアルゴリズムがブラックボックス化されてしまう為、学習工程の説明だけで妥当性を示すのは困難であり、性能検証により妥当性を証明することになる。したがって、学習データ以上に性能検証用であるテストデータは質の担保が重要となる。さらに学習データとテストデータのコンタミネーションへも配慮が必要である。

また、転移学習や強化学習など深層学習の領域で研究が進められている学習方法を採用した医療機器の登場も今後想定される。診断アルゴリズムの中に様々なデータを用いた学習結果が混在する可能性もあり、学習結果の統合プロセスを明らかにすることが必要である。

3.4.2 学習方法

深層学習など内部動作説明が難しいアルゴリズムの場合は診断用ネットワーク構造や学習工程の概要説明が求められる。ネットワーク構造は考慮している次元等の設計意図を示す（たとえば、2次元／3次元畳み込みニューラルネット、時間軸他）。学習工程は転移学習などの学習手法や導入

⁷ 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」）

⁸ クラウドサービス：従来は利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、サービスとして利用者に提供するもの。

(http://www.soumu.go.jp/main_sosiki/joho_tsusin/security/basic/service/13.html)

したモデルなど、性能を実現したポイントを示す（たとえば、転移学習を用いた場合、どのような出自の学習済みモデル適応させたか、など）。学習と性能確認に利用したデータの説明（質、量、加工方法、データオーギュメンテーション方法など）も求められる。

学習するタイミングは、学習の主体者であるメーカーが設計意図に基づき予め設定するか、再学習のためのデータの収集状況等、設定された条件のもと判断されると想定される。自律的に事後学習するプログラムを利用したシステムの場合も、基本的にはシステムで設定された再学習条件に基づいた学習および評価タイミングとなる。また、自律的事後学習の場合は性能変化に対応するための性能監視機能等が診断支援システムに組み込まれていることが基本となる。性能監視機能等により非劣性と判定された場合のみ学習結果がシステムに反映されるなどの対応が必要となる。

ユーザーが主体的に事後学習させるシステムの場合、学習のタイミングはユーザーの任意となるが、システムの特성에応じた推奨条件等はメーカー側が提供することが望ましい。

3.4.3 性能評価のタイミング×主体者×方法

以下のマトリクスに示される関係性が考えられる。

性能評価実施の主体者	性能変化のタイミング	性能評価のタイミング
メーカー	バージョンアップごとの段階的な性能変化（バージョンアップについては AI+CAD 評価指標を参照のこと）	メーカーでの学習後、リリース前に評価
	臨床使用に伴った事後学習により連続的又は相当程度高い頻度でのバージョンアップ	学習後かつ更新前
ユーザー（注） （メーカーは性能評価用のデータおよび手段（プログラム）を提供する形で性能評価実施の役割も担う）	バージョンアップごとの段階的な性能変化	バージョンアップ前
	臨床使用に伴った事後学習により連続的又は相当程度高い頻度でのバージョンアップ	学習後かつ更新前

（注）ユーザーが性能変化に関与可能なシステムは AI+CAD 評価指標の範囲外となった。この場合に性能評価において解決すべき課題としては以下のような項目が挙げられた。

- メーカーによる性能評価手段の提供。
- メーカーによる性能監視の方法とタイミング（リモートメンテナンスによる監視などが想定。）
- 性能評価における責任範囲。メーカーは意図した範囲に性能変化が限定され

る事に責任を負うべきという議論があった。メーカーが用意した評価手段をどのよう実施するかはユーザーである医療機関側の責任となり、ユーザー側での品質管理体制の構築が必要であるという議論もあった。

3.4.4 学習によって性能が低下した場合の対応

事後学習によりユーザーが性能変化に関与可能なシステムは、AI+CAD 評価指標の範囲外となったため、現時点で想定される性能変化はメーカーによる学習の場合だけとなる。①バージョンアップ毎の段階的な性能変化を意図するシステムの場合、メーカーはバージョンアップ前に性能低下が無い事を確認して上市する事が前提となるため、あきらかな性能変化が起きる可能性は低い。一方、②臨床使用に伴った事後学習により連続的又は相当高い頻度でのバージョンアップを意図する場合は、データが直ちに実装される事で性能が連続的または相当高い頻度で変化することに対する対策を設定しておく事が求められる(評価指標 6.評価に際して留意すべき事項(4)リスクマネジメント 2) 対策について"を参照の事)。

市販時の状態に初期化または変化直前の状態へと復元する機能や、手順などが想定されている。

これらの手段は実現するためのソフトウェアを提供してもよいし、サービスエンジニアが現地で確認して再インストールなどを行う方法も考えられる。

事後学習データが直ちに実装される事で性能が連続的または相当高い頻度で変化する支援システムの場合は、性能変化におけるリスクを軽減する為に、定期的に性能評価を行い、その時点での性能をエンドユーザーに示す、等の対応を行うこと。

性能低下に関しては以下のような議論があった。

- (1) メーカーが用意したテストデータでは性能範囲内であった場合でも、病院の該当症例で性能の低下とを感じる可能性もある。(該当病院で難しい症例が集まるような場合など)。このような場合は原因に応じた対応を個別に相談することが望ましい。
- (2) 診断支援システムにおいて不具合(性能低下)の定義は非常に重要である。最終的な確定診断では違う症例となった場合でも、画像診断上では診断支援システムの出力した診断結果となる可能性があったのであれば、必ずしも不具合(性能低下)とはいえない。また、医師が見ても見落とすようなケースの場合には、CADの見落としを必ずしも不具合とは言えない。という議論があった。
- (3) 適用外使用(トレーニングを受けていないユーザーによる使用等)により、ユーザーが想定外の挙動に気付かない可能性が増大する事が考えられる為、適用外使用を防ぐ方法も講じること(AI+CAD 評価指標 12 ページ(4)リスクマネジメント参照)。
- (4) ユーザーが性能変化に関与可能なシステムは AI+CAD 評価指標の範囲外となったが、自己の医療施設で得た画像データで事後学習を行う場合には、特に性能維持に関する配慮が必要。特定の病態(まれな症例など)で偏りのあるデータで事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。メーカーが性能評価の手段を提供し、所定の性能が得られる範囲を正常とする等の方法を予め確認しておく事。（“3.4.3 性能評価のタイミング×主体者×方法”を参照の事）

3.4.5 学習機能の実行環境とシステムの運用環境とが異なる場合の検証方法

大量のデータを用いた学習過程は大規模な演算量となるため、学習演算および性能評価はクラウドで提供され、学習済みモデルの実行（運用）環境はオンプレミスで提供される等が想定される。性能検証は実運用環境で実施するのが望ましいが、例示のケースのように異なる環境（クラウド）で性能検証を実施する場合は妥当性（同等性）の説明が必要になる。

3.5 医療機器等法に対する考え方について

- (1) AI を利用したシステムの場合も医薬品医療機器法に基づく製造販売承認事項一部変更申請（一変申請）、軽微変更届の範囲に関する基本的な考え方は他の医療機器と共通である。学習による性能変化があった場合も影響が軽微であり製品の同一性が損なわれないことが明らかな場合には、軽微変更手続きが想定される。但し、変更による影響は製品の性質にも依存するので、製品の承認審査時点で事前に審査機関と確認すること。逆に製品の同一性が明らかで無い場合、例えばメーカーが標榜する使用範囲（対象部位など）や性能を変更する場合には、一変申請または新規申請が必要となる。一変手続きでリリース可能か、新規申請が必要かは、対象となる診断支援装置の用途やリスクによって異なるので事前に審査機関に相談すること。
- (2) 事後学習の対象には、深層学習の場合は、ネットワークの結線（結合）の重み以外に、ネットワークのハイパーパラメータ（層数、特徴マップ数、畳み込みのフィルタサイズ、ユニット数等）も考えられる。ハイパーパラメータ変更はあらかじめ想定され、製品の同一性を損なわない場合もあるが、変更の程度によっては、軽微な変更ではなく支援システム自身の変更と判断される可能性があるため、当該変更の実施については、事前に医薬品医療機器総合機構と相談することを推奨する。

3.6 データについて

3.6.1 データの品質管理について

CAD のシステム開発及び性能検証の際には、妥当な性質ならびに数量の臨床画像、シミュレーションやファントム使用による画像データを使用する。これらの画像に対し、適当な臨床データ（病変の存在、位置やサイズを示す情報、診断結果等）を紐付けし、開発、検証用のデータとして使用する。CAD の開発（学習、性能確認（ハイパーパラメータ調整等）など）に用いるデータと性能検証に用いるテストデータとは厳密に区別する。両者を混在させてはならない。

主に学習により機能構築を行う CAD においては、その性能は学習に使用するデータと学習方法に依るところが大きく、アルゴリズムに相当する部分がブラックボックス化し、学習後のネットワーク構造などの解析は現時点では困難である（特に深層学習ではこの傾向は顕著となる）。このため、テストとそれに用いるデータの品質管理を厳密に行う必要がある。

CAD の開発及び性能評価に用いるデータについては、その出所と質を明示する。データの品質管理項目としては「人工知能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」6.（2）2）学習 の項の記載を参考に、必要な項目を確認し、記載する。

シミュレーションやファントム使用による画像データ用いる際は、その使用目的に相応するレベルで、画像が具備する性質が臨床画像と同質であることを科学的ならびに論理的に説明する必要がある。

事後学習、あるいは、常時学習による更新が技術的に可能となってきている。これらの機能を採用するシステムの場合、学習データの妥当性確認と、適切なテストデータに基づく性能確認を、オンラインまたはクラウド等で行えるようにする必要がある。

ワーキンググループからはデータに対する要件は、以下に例示するように、意図する使い方、目的により異なるとの意見があったが結論に至っていない。

使い方（例）	テストデータ（例）	学習データ
2次読影 (Second Reader) 従来型 CAD	複数読影医の合意や病理診断により確定したデータを、ゴールド・スタンダードとして利用。	以下のようなデータが想定される。従来型 CAD の場合は教師付データを学習データとして用いる場合が多いと思われるが、性能は評価用データで担保する為、学習用データは教師付きデータに限定するものではない。
1次読影 (First Reader) (検診等での利用を想定。)	必ずしも確定診断済みである事は必須では無い。WG の議論では、陽性の見逃し防止に重点化する場合、擬陽性のデータを付加した人工画像等を用いる事も想定されるとの意見もあった。利用の妥当性については別途説明が必要。	① 複数読影医や病理診断判定から確定診断したゴールド・スタンダード ② 一人の読影医による診断結果（※1） ③ 機械的に異常可能性を検出したデータ。 ④ 人工画像
コンカレント (動画での注意喚起等)	同上	⑤ 転移学習用自然画像 など

学習データ

学習データに対しては、必要な品質管理項目を確認し、当該学習データが、開発しようとする

CAD 機器の要求事項、目的、学習方法等の技術的手段に対し妥当なものであることを示す。学習データとしては、臨床情報の信頼性の低いデータ（Weak label データ）を活用する方法、対象の臨床画像以外の、人工生成画像（シミュレーションやファントム使用による画像データ）を用いる方法、転移学習等一般画像を用いる方法などが考案されている。さらに、教師データ無しで学習を行う方法も検討研究されており、それらの CAD 用学習方法としての適否を単純に決定付けることはできない。採用する学習方法と学習データが想定する診断支援の目的に対し妥当であることを示す必要がある。

事後学習に用いられる画像については、システム機能、あるいは、運用により、学習データとしての妥当性を確保することが求められる。

ワーキンググループにおける議論としては、CAD システムの性能はテストデータにより担保されることから、学習データに関する品質管理項目の記載、妥当性の提示については緩くても良いとの意見があった。一方、学習データと学習プロセスについて、一定の正当性を示せるレベルの記載は必要との意見があった。

性能確認データ

性能確認データは、識別器のハイパーパラメータの調整に用いる。学習用データと同様の品質管理対象であるが、テストデータ同等の信頼性、学習データに対する独立性が確保されていることが望ましい。

テストデータ

テストデータは申請時に規定する性能のエビデンス、およびその後の性能変化の判定基準となるものである。開発しようとする CAD 機器の要求事項、目的に対し妥当なテストデータセットであることを示す。

開発過程においては、原則として学習データや性能確認データとは厳密に区別して管理し、独立性を確保する。テスト結果に基づく設計結果の選択等、開発設計への反映が行われぬよう配慮する必要がある。薬事承認を前提とした場合、テストデータは治験に相当する信頼性をもつことが求められる。

シミュレーションやファントム使用による画像データをテストデータとして用いる場合は、学習データとして用いる場合より厳密に臨床画像との同等性を示さねばならない。

事後学習を採用する場合は、テストデータは性能変化が申請範囲内であることの評価用としても用いる。適切な質と量のデータを用いて性能変化の検証を行えるようにすることが特に重要である。

ワーキンググループでは、深層学習の様にブラックボックスで、かつ、入出力の関係が高度に非線形なシステムの場合、当面は治験並みの検証が適切である。ただし、条件付きで後ろ向きの試験（たとえば、過去のデータを用いるが、読影は前向きに実施する）も可とするなど、試験の信頼性

が確保できる範囲内で、可能な限り速やかに認証審査が実施できるようにするための工夫は積極的に導入すべき、との意見があった。

3.6.2 データの撮像条件の記録

診断アルゴリズムの適応範囲は基本的に学習範囲に依存すると考えられるため、開発に用いた学習データと性能確認データの網羅範囲の記録が必要である。また、性能が担保される範囲はテストデータの網羅範囲に依存するため、記録するとともに明示が必要である。

適応範囲、性能の規定にかかわる項目

- 撮像パラメータ等：撮像機器の方式、機種、撮影／撮像プロトコル、画質条件、画像処理のパラメータ等
- 画質条件、画像処理のパラメータ等

また、使用時の性能に影響を与える条件や、学習データとしての利用に不適な因子に関してもリスクに応じた対応が必要である。

- 制約条件、制限事項：パラメータの組合せ等で使用上の制約となる条件等
- 人為的な使い方、設定等に依存する事項：解析対象の視野はみ出し等
- 学習に不適な性能阻害因子：ノイズ、アーチファクトなど、性能劣化を誘引する事項

内視鏡等は、撮像機器の性能、光源の特性、観察モード、画像処理のパラメータ、入力データの規格等による性能への影響が予想される。開発する CAD システムの接続対象とする撮像機器、観察モード等を示し、目的とする CAD 性能に影響のあるパラメータ等について必要のあるものを明示する。

加えて内視鏡等では、視野における対象の捉え方、距離、大きさ、拡大観察、染色法の適用、動画に対する CAD 適用場面等、画像取得条件や検査中の機器操作による性能への影響も想定される。適切な使用場面、方法を規定する必要がある場合は明記する。

3.6.3 データの品質を評価する方法

画像特性として、画像の S/N に関連する撮像条件、空間(x, y, z)及び濃度、時間方向の分解能等、学習データの条件を満たしていることを確認する。

臨床特性として、部位、視野、対象病変（あるいは正常例）としての妥当性確認、病態や病変の描出状態（造影および非造影含む）やアーチファクトの有無を確認する。

3.6.4 学習データ作成機能をシステムが有する場合の対応

CAD の開発や性能強化で使用するデータには、学習データや性能確認データ、製品申請や性能評価のためのテストデータ（以下テストデータ）に分けられる。テストデータを学習データや性能確認データに転用した場合はその価値を失うので注意が必要であり、新たに取得する必要が生じる

(AI+CAD 評価指標案ではテストデータを学習のプロセスから完全に切り離して管理するための方策をとるように求めている)。ただし、当 WG 委員からは、あまり厳しくしなくとも良いのではないかという意見があった。

学習データ作成機能をシステムが有する場合とは、検査に用いられたデータを引き続き開発や性能強化に使用する事を指し、メーカーがクラウドシステム上で追加する場合とオンプレミスで追加する場合が想定される。どちらの場合にも、その検査後データを上記のどのデータとして取り扱うかを明確にする必要がある。

新たに追加するデータについては、以下の観点で品質の評価を行うことを推奨する。

- (1) データに含まれる疾患範囲が意図する性能に対して適切であること。
- (2) 収集方法が適切である事 (収集プロトコルの明示)。
- (3) 当初の開発にあたって使用したデータと同じ性質の画像を用いること。たとえば、単独の医師が確定したゴールド・スタンダード、複数読影医の合議の上で得られたゴールド・スタンダード、病理診断判定から確定診断したゴールド・スタンダードなどがあるが、追加データでもその基準を変えないこと。
- (4) テストデータには不適切なバイアスが無いこと。
- (5) テストデータとしては、文書による患者同意のあるデータを使用するか入手元及び妥当性のあるデータを使用すること。次世代医療基盤法のもとに集められたデータもテストデータとして使用出来ると想定されるが、確認が必要と思われる。
- (6) オンプレミスで登録する場合には、登録する画像の質や撮影条件などチェックする項目を明示し、医師が妥当性をチェックするシステムが必要であり、登録医師と症例登録者は記録される必要がある。
- (7) 非典型的な画像や非常にまれな所見を呈する画像を新たに登録するかどうかは、その CAD システムの趣旨にあっているかどうかを吟味すること。

AI を利用した CAD が臨床で使用されると、実際に診断で使用された画像が AI の学習のためのデータとして利用されると予想される。この場合、CAD を開発した企業が各病院からデータを収集して使用する場合や自施設で取り込んで活用する場合が想定される。

臨床では治療経過や手術、死後の病理解剖で診断が訂正されることが有るが、匿名加工済みのデータにおける該当事例の特定は困難である。そこで、メーカーが事後学習で性能変化を行うシステムの場合は、診断の訂正の可能性が少ないデータや、データの妥当性が医師により確認済みのデータを用いることで、事後学習による性能確保を達成する事が望ましい。この場合、医師が CAD 用データベースとしてふさわしくないと判断した画像を事前に除外するか又は利用前に妥当性をチェックできるようにすべきである。一方、ユーザーが性能変化に関与可能なシステムの場合は、登録する画像の質や撮影条件などチェックする項目を明示し、医師が妥当性をチェックするシステムが必要であり、症例登録者は記録されるべきである。

ユーザーである医師がデータの妥当性をチェックする例を下記に示す。

1. 病理診断で悪性が示された画像のみを異常としたデータベースの場合、病理診断が無く、医師が悪性と判断した画像は加えることができないようにする。
2. データベースの異常症例が典型症例のみで構成されている場合は、まれな所見を示した症例は登録できないようにする。
3. 珍しい症例がティーチングファイルの代用として登録され、性能に悪影響を及ぼさないか。
4. 読影結果の正当性に裏付けのないデータが登録されようとしていないか。

各施設で収集した画像を AI の学習データとして使用する場合、事前に被験者の同意を書面で得ておく必要があるかどうか、包括同意で良いのかは審査側と協議する必要がある（詳細については、「人を対象とする医学系研究に関する倫理指針」や「臨床研究法」、「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（通称、次世代医療基盤法）」などを参照のこと）。

3.6.5 データの保管期限

(1) テストデータ：

少なくとも他の評価エビデンスと同等期間の保管が必要⁹。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の間は保管すること。

(2) 事後学習用の学習および性能確認データ：

A) メーカーが性能を変化させるための事後学習を行う場合：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必ず残す事が必要。但し、性能はテストデータで担保されるので、下記の理由により、学習および性能確認データ全てを学習時の状態で維持する事は必須では無い。

① ビッグデータの利用等を想定すると学習および性能確認データを学習時の状況で長期保管することは非現実的。

② 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を想定すると、医療機関や海外で学習したデータの提供を受け、長期保管することは難しい。

B) 医療機関が追加学習を行う場合：学習用データについては、医療機関側の責任で必要に応じて保管すること（学習データが残っていない場合は、メーカー側での解析が出来ない）。学習に関する履歴の記録をログとして残す等の対応が検討課題となる。

初期出荷時の学習データについては、今回の人工知能の定義を考えると性能変化の為のデータでは無く、初期の性能を実現する為の材料と考えられる。この性能は初回申請時のテストデータに

⁹ 文書、記録／資料の保管期間については、QMS 省令、GLP 省令、GCP 省令、認証基準告示（平成 17 年厚生労働省告示第 112 号）等、の医療機器に要求される法令のうち、対象の法令に従って保管する事。

よって担保される。

3.6.6 データの収集と流通について

個人情報の保護に関する法律（施行 H29 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。

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、取り扱いに注意することが求められているとともに、その情報の診療以外への取り扱いに対し、オプトアウトで扱うことはできず、患者の同意を得て扱うことが必要となる。医療機関における学術研究の場合は、個人情報保護法の 76 条の適用除外条項により、個人の同意を得なくても診療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に注意する。

このため、製品開発等のための画像情報等を収集するためには、医療機関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。なお、同意が後から撤回された場合、適切に匿名化されていれば当該患者のデータを遡って削除する必要はない。

なお、承認申請に用いる評価に用いる診療情報については、研究と開発のみではなく、承認申請に用いることを含めた患者同意を得ておく必要があることに留意すること。学術研究に用いる情報であっても、製品開発に使用する可能性もある場合においては、あらかじめ、研究のほか、匿名化した情報として製品開発や承認申請等にも用いることの患者の同意を得ておくことが望ましい。治験に相当する試験が必要な場合は、試験の為のデータ収集は GCP 省令に従うこと。

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処理のアウトプットであり、患者情報は含まないのでこちらも個人情報には該当しない。また、研究時点で、患者の同意が学術研究目的のみであった場合も、学術研究の結果として実現された成果物であるコンポーネントに患者データ（画像）が含まれない場合はコンポーネントの製品利用自体は可能である。学術研究の成果は個人情報には該当しないものであり、特段の配慮等は不要であるが、学術研究のために用いた要配慮個人情報は、他への流用はできないことに留意すべきである。このため、前項に述べたように、医療機関において匿名化された情報（個人情報に該当しないもの）として提供する必要がある。

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、企業など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などから、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 月には施行の予定である。

製品における留意点

1) 動作環境について

- ア) 医療機関内にて処理をする場合、要配慮個人情報外部へ持ち出すことがないため、特段の対応は不要である。
- イ) 医療機関外で処理を行う場合は、要配慮個人情報外部へ送付されるため注意が必要である。医療行為の一環として、処理を委託しているものとしての扱うことが可能であるが、処理を行う施設（外部のクラウド等）においては、医療機関の監督下の元で適切に管理できる体制が必要となる。なお、処理施設が国外となる場合については、個人情報保護法にて認定された国に該当しない場合は（現時点で認定された国はないが）、その個別施設ごとに適切に情報が保護されていることを確認することが必要となるため注意すること。この場合は、処理を外部（他の国）で行う旨の患者同意を得ることを検討することが望ましい。

2) データの保管等

処理した画像を用いて、学習のため利用する場合について、性能確認等のために、その情報を保管する場合については注意が必要である。扱いが当該医療機関内に限定される場合は良いが、医療機関外となる場合においては、開発に用いる情報と同様に、医療機関において匿名化された情報とするか、患者の同意を得る必要がある。

参考資料

- IEC 82304-1:2016, Health software -- Part 1: General requirements for product safety
- JIS C 6950-1:2012 「情報技術機器—安全性—第 1 部：一般要求事項」
- JIS T 0601-1:2014 「医用電気機器—第 1 部：基礎安全及び基本性能に関する一般要求事項」
- JIS T 2304:2012 「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」
- JIS Q 13485:2005 「医療機器—品質マネジメントシステム—規制目的のための要求事項」
- JIS T 14971:2012 「医療機器—リスクマネジメントの医療機器への適用」
- IEC 62304, 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)
- IEC/TR80002-1 (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—第 1 部：医療機器ソフトウェアへの ISO14971 の適用の手引き)

- 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」（平成 16 年 9 月 22 日厚生労働省令第 135 号）
- 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）
- 画像診断ワークステーションのウイルス対策ソフトに関するガイドライン（技術資料 No.JESRA TR-0035-2010 制定 2010 年 6 月 25 日）

APPENDIX: 用語について

AI 技術に関してはいくつかの似た用語が用いられている。本報告書では以下の意味で用いられている。

- 事後学習：臨床で使用を開始した後、追加データにより学習させること。
- オンサイト学習：医師等が医療機関で得られたデータにより学習させること。

よって、下記のようなケースが想定される。

- 1) 事後学習かつオンサイト学習
- 2) 事後学習ではないが、オンサイト学習
- 3) 事後学習だがオンサイト学習ではない

事後学習における「性能変化」は下記のように分けられる。

- A) 施設ごとに別々の性能変化が可能なもの、
- B) 施設ごとに別々の性能変化をしないもの（すなわち、全体で共通の性能変化をするもの）。

事後学習における「性能変化」には下記のタイプが含まれる

- ① その性能変化について都度一部変更申請を行い承認されたもの（段階的変化）
- ② それ以外（逐次変化）

また、本ガイドラインでは以下の用語も登場する。

- オンプレミス型：従来のな（AI 技術を用いない）CAD の多くが該当する。
- クラウド型：画像データをクラウド上に置くプロセスが加わる。

この報告書は、平成 29 年度に国立研究開発法人産業技術総合研究所が、国立研究開発法人日本医療研究開発機構からの委託を受けて実施した成果を取りまとめたものです。

— 禁無断転載 —

平成 29 年度 未来医療を実現する医療機器・システム研究開発事業
(医療機器等に関する開発ガイドライン策定事業)

人工知能分野
開発 WG 報告書

連絡先

〒100-0004
東京都千代田区大手町 1-7-1 読売新聞ビル 23 階
国立研究開発法人日本医療研究開発機構
産学連携部 医療機器研究課
TEL : 03-6870-2213
FAX : 03-6870-2242
URL : <http://www.amed.go.jp/>

発行

〒305-8564
茨城県つくば市並木 1-2-1
国立研究開発法人 産業技術総合研究所 健康工学研究部門
医療機器開発ガイドライン事業実務委員会
TEL/FAX : 029-861-7840
E-Mail : md-guidelines@aist.go.jp