

平成30年度
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WG報告書

平成31年3月

審査WG座長 橋爪 誠

北九州中央病院

(WG運営：清水 昭伸 東京農工大学)

目 次

I. はしがき.....	1
II. WG 委員等名簿.....	3
III. 人工知能技術を利用したプログラムの医療機器該当性判断要素について(案).....	5
IV. 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（改訂案）... 13	
V. 人工知能技術を利用した医療機器プログラムの承認状況.....	31
VI. 参考資料.....	37
1：韓国 AI ビッグデータガイドライン和訳版（医療機器該当性部分等を抜粋）	
2：中国におけるクラス分類表英訳版（Medical Software 部分を抜粋）	
3：合同検討会報告資料	
4：その他関連 URL リスト	
VII. 平成 30 年度 WG 会議議事概要.....	61

I. はしがき

はしがき

近年、人工知能がディープラーニング（深層学習）の導入により、従来人がなしえなかった判断の支援ツールとしての有用性が多くの分野で実証され、人工知能技術は実用化段階にはいつてきた。人工知能の利用は、現代社会に第3のブームを巻き起こし、産業界だけでなく、医療界への応用も非常に注目されているところである。

人工知能を利用した新しい方法を医師の診療支援として導入することにより、従来の方法では診断・治療が困難であった症例が助かったという事例の報告が散見されるようになってきた。今や世界中で人工知能技術を用いたシステムや医療機器の開発、さらにその製品化に国際競争が激化している。

我が国においてもこのような現状を鑑み、有識者らを構成員としたワーキンググループをたちあげ、平成29年度に人工知能を利用した医療機器、特に「画像診断支援システム等に関する評価指標案」を作成した。平成30年度にはさらに、人工知能を用いた医療機器プログラムが、その特性から市販後に新たなデータを用いて再度プログラムの学習を行い、医療機器承認申請時よりも性能が向上または低下する可能性があることを踏まえ、海外での開発動向を調査すると同時に、プログラムの医療機器該当性判断要素について検討した。

人工知能を利用したプログラムの医療機器該当性を判断する上で考慮すべき要素としては、使用目的や効果・性能が医薬品医療機器法の定義上、「疾病の診断・治療・予防を目的とするもの」か、対象とするプログラムの目的性を考慮する必要がある。また、プログラムにより得られた結果が「疾病の診断・治療・予防に寄与する程度」と、「使用対象者の生命及び健康に与える危害の程度や蓋然性」によって医療機器該当性が判断される。

本WGで作成された評価指標案及びプログラムの医療機器該当性判断要素案は、人工知能技術を利用した医療機器プログラムを評価する上で考慮すべき点等を検討しまとめられたものであり、これらの利用により我が国の人工知能技術とその臨床応用を発展させ、国民が広くその利便性を一早く享受できることを願っている。

人工知能分野審査WG座長
橋爪 誠

Ⅱ. 人工知能分野審査 WG 委員等名簿

人工知能分野審査WG委員等名簿（敬称略）

座長

橋爪 誠 九州大学 名誉教授／北九州中央病院 院長

委員（五十音順）

佐々木 毅 東京大学大学院医学系研究科 次世代病理情報連携学講座 特任教授
（日本病理学会推薦）

清水 昭伸 東京農工大学大学院 工学研究院 教授

田中 聖人 京都第二赤十字病院 医療情報室 室長（日本消化器内視鏡学会推薦）

中田 典生 東京慈恵会医科大学 放射線医学講座 准教授（日本医用画像工学会推薦）

藤田 広志 岐阜大学 工学部 電気電子・情報工学科 情報コース 特任／名誉教授
（医用画像情報学会推薦）

待鳥 詔洋 国立研究開発法人 国立国際医療研究センター 国府台病院 放射線科診療科長
（日本医学放射線学会推薦）

松尾 豊 東京大学大学院 工学系研究科 技術経営戦略学専攻 特任准教授

横井 英人 香川大学医学部附属病院 医療情報部 教授

厚生労働省

中井 清人 医薬・生活衛生局 医療機器審査管理課 課長

田中 大祐 医薬・生活衛生局 医療機器審査管理課 課長補佐

青柳 ゆみ子 医薬・生活衛生局 医療機器審査管理課 医療機器規制国際調整官

藤野 綾太 医薬・生活衛生局 医療機器審査管理課 先進医療機器審査調整官

高村 建人 医薬・生活衛生局 医療機器審査管理課 医療機器係長

小川 雄大 医薬・生活衛生局 監視指導・麻薬対策課 危害情報管理専門官

西澤 洋一 医薬・生活衛生局 監視指導・麻薬対策課 危険ドラッグ監視専門官

宮下 正也 医薬・生活衛生局 監視指導・麻薬対策課

独立行政法人 医薬品医療機器総合機構 (PMDA)

高江 慎一 医療機器審査第一部 部長

岡崎 譲 医療機器審査第一部 審査役

菅原 明彦 医療機器審査第一部 審査役補佐

加藤 健太郎 医療機器審査第一部 審査専門員

谷城 博幸 医療機器審査第二部 審査役

佐々木 拓哉 医療機器審査第二部 審査役補佐

古森 亜矢 医療機器審査第二部 嘱託

小野寺 陽一 医療機器調査・基準部 部長

水上 良明 医療機器調査・基準部 課長
今川 邦樹 医療機器調査・基準部 基準専門員

国立医薬品食品衛生研究所（事務局）

藪島 由二 医療機器部 部長
中岡 竜介 医療機器部 室長
加藤 玲子 医療機器部 主任研究官

産業技術総合研究所（オブザーバー）

鎮西 清行 ヒューマンライフテクノロジー研究部門 副研究部門長
坂無 英徳 人工知能研究センター 人工知能応用研究チーム 主任研究員

Ⅲ. 人工知能技術を利用したプログラムの医療機器 該当性判断要素について（案）

人工知能技術を利用したプログラムの医療機器該当性判断要素について（案）

1. はじめに

近年開発が進んでいる人工知能（AI）技術を利用した場合、患者の詳細な病態変化予測や専門医でも判断困難であった病変の検出が可能になる等、従来のアルゴリズムでは達成困難であった性能を有するプログラムが開発できる可能性を秘めている。但し、AI技術が利用される場合であっても、原則としてプログラムの医療機器該当性は平成26年11月14日付け薬食監麻発1114第5号厚生労働省医薬品食品局監視指導・麻薬対策課長通知「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」（以下「該当性通知」という）で示された基本的な考え方に基づいて判断されることには変わりはない。しかしながら、AI技術を利用したプログラムの医療機器該当性は、学習データのさらなる追加（市販後学習¹）の方法等、その時代の技術レベルに応じた特有のリスクを追加して考慮すべき必要が生じると考えられる。そこで、AI技術を利用して開発され得るプログラムを念頭に、それらの医療機器該当性を判断する上で考慮すべき要素を以下に整理した。また、医療機器該当性判断においては考慮すべき要素を総合的に検討する必要があるため、考え方の参考となり得るプログラムの医療機器該当例及び非該当例を示した。

なお、以下の要素は現時点の知見に基づいて整理しているため、今後の技術革新に伴い変更されることに留意されたい。

2. プログラムの医療機器該当性を判断する上で考慮すべき要素

プログラムの医療機器該当性については該当性通知に示された基本的な考え方に基づいて判断されるが、その判断において重要となる要素は以下のとおりである。

(1) 目的性

医療機器は医薬品医療機器法の定義上、「疾病の診断・治療・予防を目的とするもの」とされているため、まず、使用目的や効果効能から、対象とするプログラムの目的性を考慮する必要がある。

1) 疾病の診断等を目的としない場合の例

- ① 健康管理・患者説明・教育用プログラム
- ② 医療用データの単なる送信・保管を行うためのプログラム

2) 疾病の診断等を目的とする場合の例

- ① 疾病の診断、治療方針の決定や予防のための療法等の決定を支援する情報の提示を行うプログラム

使用者に関係なく、診断・治療・予防の判断を支援するために情報の提示を行うものは、「疾病の診断・治療・予防を目的とした」ものと解される。

¹市販後に再度プログラムの学習を行うことを指す。例えば、承認されたプログラムに新たなデータを用いた学習を追加する場合等が該当する

対象とするプログラムの目的性を判断するにあたって考慮すべき点の例を以下に示す。

- ・ 医師を介さず使用者が直接情報を入力しアドバイス等が表示されるプログラムにおいて、一般的な健康情報が提示されるのか、若しくは疾病名が提示されるのか
- ・ 身体状況に応じた適切な動作等を提示するプログラムにおいて、健康維持のトレーニングを目的とするのか、特定の疾病を予防するためのトレーニングやリハビリテーションを目的とするのか
- ・ 患者が自身の情報を記録するプログラムにおいて、その記録保持のみを目的とするか、あるいはその記録を分析して得られた情報を医師に提示し診断等の判断に利用することを目的とするか

(2) 寄与度・リスク

「薬事法等の一部を改正する法律の施行に伴う関係政令の整備等及び経過措置に関する政令」(平成 26 年政令 269 号)に伴う改正後の「医薬品、医療機器等の品質、有効性及び安全性の確保等に関する法律施行令」(昭和 36 年政令第 11 号)別表第 1 において、プログラム医療機器については、機能の障害等が生じた場合でも人の生命及び健康に影響を与える恐れがほとんどない一般医療機器(クラス I)は、医療機器の範疇から除外されていることを踏まえ、クラス I とクラス II 以上とを分類するために追加で寄与度とリスクを考慮する必要がある。寄与度及びリスクの両者が下記の分類において最も低い場合のみ、クラス I に相当するものとして、医薬品医療機器法の定義上の医療機器に該当しないと判断されると考えられる。

1) 寄与度

寄与度とは、プログラムにより得られた結果が疾病の診断・治療・予防に寄与する程度のことである。診療において参考にする程度の情報であれば寄与度が低いが、その情報を元に診療方針を決定するのであれば寄与度が高いと判断される。これまでの事例から、想定される寄与度の分類例を以下に示す。

寄与度が低い：辞書や公知の診療ガイドライン、過去の統計情報等一般的な情報を提示するプログラム

寄与度が高い：患者に関する情報を分析し、個々の患者に対する診断・治療等の判断を支援する情報を提示するプログラム

2) リスク

リスクとは、プログラムの利用の際、間違った情報が提示される等の不具合が生じた場合に、使用対象者の生命及び健康に与える危害の程度とその危害を生じる蓋然性の組み合わせである。これまでの事例から、想定されるリスクの分類例を以下に示す。

リスクが低い：単純な計算のみを行うプログラム、または医薬品の添付文書や広く使われている診療ガイドライン等公知の情報に従ったもので、使用者がその誤りや不具合を容易に判断可能なプログラム

リスクが高い：独自の分析や判断を行った結果を提示するもので、使用者が一見してその誤りや不具合を判断することが難しいプログラム

寄与度やリスクを判断するにあたって考慮すべき点の例を以下に示す。

- ・ 医師等の医療従事者が使用するものや、医師の指示に基づき患者が使用するものか、あるいは医療従事者以外の使用者が自身の判断で使用するものか
医療従事者を介さずに一般の患者等が直接使用する場合は、プログラムが提示した情報に対して医師等の医学的知識による解釈・判断が行われないため、一般的に寄与度やリスクが高くなると考えられる。
- ・ 患者の情報を分析し診断等を支援するプログラムの場合、分析方法が独自のものか、広く使われている診療ガイドライン等の計算式、表、並びにフローチャート等を改変せずに使用する、あるいは過去の統計情報との比較のみを行うものか
医師や企業が独自に開発した計算式やアルゴリズムに基づく分析は、その分析方法が公知ではないため、出力に誤りが生じても使用者が容易に検知することは難しい上、分析方法自体の医学的な正当性の検証が不十分であるため、リスクが高くなると考えられる。

3. 医療機器該当性判断例

医療機器該当性の判断を行うにあたっては、上記の要素を総合的に考える必要がある。AIを利用したプログラムの医療機器該当性を判断する際の考え方と具体的なプログラムを以下に例示する。

なお、医療機器該当性の判断においては、プログラムそのものの機能や効果だけではなく、機器の標榜（広告や説明）が参考とされるため、標榜も含めて検討する必要がある。プログラムそのものは実態として医療機器に該当しないものでも、実態の効果とは異なる「診断・予防」との標榜により、医療機器に該当するとの判断になる場合もあるため、標榜はプログラムの実態を適切に表した内容にすることに留意すること。

(1) 事例1

医師が使用することを想定しているが、その効能が公知のルールに基づいた単なる計算機の代用等、簡易な補助目的である場合（要素 2/(1)）、そのプログラムは医療機器に該当しない。一方、専門医の専門的知識を代替する目的で研修医や一般医のような専門外の医師が診断時に使用するプログラムの場合、専門医による補助の有無（要素 2/(2)）、提示された情報の種類、根拠、付随情報等（要素 2/(2)/1）に応じてその使用に伴うリスクは高まる。また、その効能によっては、使用に伴う医療安全上のリスク（要素 2/(2)/2）も影響を受けるため、想定される要素を総合的に考慮して、医療機器該当性を判断する必要がある。

(2) 事例 2

使用対象者が体調管理に使用するプログラムにおいて、健康等に関する情報を入力して単に睡眠の質の良し悪しのみを情報として提示する場合(要素 2/(1))には医療機器に該当しない。一方、その他の情報が提示されることにより使用対象者が睡眠時無呼吸症候群等の疾病を診断できる場合には医療機器に該当する。また、何らかの疾病に罹患していることを推測可能な情報が提示される場合(要素 2/(2)/1)、情報内容に応じて寄与度が変化するため、その程度を総合的に考慮して当該プログラムの医療機器該当性を判断する必要がある。

<プログラム例>

1) 医薬品医療機器法において医療機器に該当する事例

- ① 医療画像等を利用してがんを一定以上の確度で検出し平均余命を予測する等、専門医のみが可能な診断を AI が代替するプログラム
- ② 医療従事者以外が入力した体温や血圧等の情報から、罹患している可能性のある疾患名を独自のアルゴリズムに基づいて AI が予測するプログラム(医師による診断行為の代替と判断され、クラス II 以上の医療機器に相当する)
- ③ 疾患が疑われる部位の撮影漏れを医療機器使用中に検査実施者に指摘し、検査を補助する AI を利用したプログラム

2) 医薬品医療機器法において医療機器に該当しない事例

- ① 健康に関するデータを使用者自身が入力して、AI が提示した情報に基づいて体重や健康状態を自己管理するためのプログラム(目的が診断・治療・予防に該当しない)
- ② AI による画像補正等の後、人体画像を 3D で詳細に映し出し、医学生の学習や患者説明用に使用するためのプログラム(目的が診断・治療・予防に該当しない)
- ③ カルテ情報から AI が必要な情報を抽出し、公知の診療ガイドライン等の情報に従って、罹患している可能性のある疾患名をガイドラインの該当箇所及び該当の判断根拠とともに提示するプログラム(医療機器該当例②とは異なり、示された情報は公知の情報のあてはめと考えられるため、医療行為への寄与度が低くなると判断される(要素 2/(2))
- ④ 公知の計算式等を利用して、臨床データから医療行為に必要な情報を提示する AI を利用したプログラム
- ⑤ 医師が診断の際、公知のガイドラインや添付文書の内容を検索する AI を利用したプログラム
- ⑥ 公知のガイドラインに従って医療機器による撮影等を補助する AI を利用したプログラム(内視鏡、エコー、CT 等、一定のルールに従った撮像が必要な医療機器の適正使用等が目的であり、直接医療行為と関連しないプログラム)

注:⑥のプログラムでは再撮影が必要となるケースが想定され、それによって生じる医療安全上のリスク程度に応じて医療機器に該当する可能性がある。また、適正使用の根拠となる機構

が公知のガイドライン以外に基づく場合にも、医療機器に該当する可能性がある。

使用者の健康に関わるプログラムの取り扱いに関する提言（案）

医療機器該当性にかかわらず、使用者の健康に関わるプログラムにおいては、提示した情報を信用して、使用者が医師による診断を受けない間に症状が重篤になる等、悪質なプログラムによる健康被害が生じる可能性がある。また、市販後学習機能を有するプログラムは製造販売承認後にも性能を変化させることが可能であるが、その変化機能に応じてリスクは変化するため、クラス判断にも影響を与える可能性がある。特に、リアルタイムで性能が変化する場合には、製造販売業者は承認申請時に規定した範囲内で性能が変化することを保証しなければならないが、想定外の性能変化を生じる可能性を考慮する必要がある。想定外の性能変化が生じた場合の懸念点を以下に例示する。

1. 性能が規定していた範囲を超え当初は想定しなかった使用方法が可能となる場合

効能・効果・性能を標榜せずにクラスⅠの非医療機器として販売されたプログラムの性能が、販売後の市販後学習によりクラスⅡ以上に相当する判断が可能となる性能を有することも考えられる。この場合、製造販売業者の対応としては、

- (1) 改めて適切なクラスで申請をする
- (2) 非医療機器のままで性能を標榜する

が考えられる。ここで、製造販売業者の対応が(1)の場合であれば問題はないが、(2)の場合には非公式な標榜であっても医薬品医療機器法違反となるため、規制当局による指導等、適切な対応が必要となる。

2. 性能が想定範囲を下回る場合

上記の事例とは逆に、販売されたプログラムの性能が市販後学習によって想定範囲を下回った場合、医療機器として承認されているプログラムであれば、従前通り、製造販売業者がその性能を維持するための対策を取ることで生じうるリスクを回避することができる。一方、クラスⅠと判断された非医療機器プログラムの場合、開発企業の判断によっては、市販後学習により販売直後に達成していた性能を下回る状態となっても当該プログラムの販売が継続される可能性がある。この場合、元々リスクや寄与度が低い非医療機器プログラムであっても、想定されていない健康被害が発生する可能性がある。

これらの事例を全て把握・管理することは、現状では困難であるが、将来、医療機器に該当しないプログラムに対しても、市販後学習に伴う性能変化に対する適切な対応が必要であることを開発企業に啓発する等、健康被害を防ぐための新たな取り組みを検討することが望まれる。

IV. 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（改訂案）

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（改訂案）

1. はじめに

コンピュータ診断支援（Computer-Aided Diagnosis:以下 CAD）とは、X線画像に代表される放射線画像をはじめとする医用画像に対して、コンピュータで定量的に解析された結果を「第2の意見」として利用する「医師による診断の支援」である。

当該診断支援を行う医療機器に関しては、「次世代医療機器評価指標の公表について」（平成23年12月7日付け薬食機発1207第1号厚生労働省医薬食品局審査管理課医療機器審査管理室長通知）別添3「コンピュータ診断支援装置に関する評価指標」にて公表しているところであるが、近年の計算機技術の著しい発展により、学習データのさらなる追加（市販後学習）等により市販後に性能が変化する人工知能技術を利用した診断支援装置については述べられてない。また、人工知能技術の中でも、近年着目されている深層学習では、最終結果を導き出すアルゴリズムがブラックボックス化している点が特徴であると同時に、市販後学習等による性能変化の内容も基本的にはアウトプットでしか評価できないこと等から、その評価や運用に当たって従来にない新たな課題が生じることが指摘されている。当該診断支援装置を用いた際、学習データに含まれていなかったタイプの事例に対する性能をどのように評価するのか（又はそのような事例にどのように対応するのか）、学習データの真正性や偏り等を考慮してどのように学習の品質確保を行うのか、さらに、市販後の性能変化に対する評価方法等に関して、従来とは異なる新しい視点も考慮する必要がある。

このような状況に鑑み、本評価指標は、上記通知も踏まえて、人工知能技術を利用した医用画像診断支援システムの有効性と安全性を評価する際の問題点や留意すべき点について、現時点での考え方をまとめたものである。

2. 用語の定義・説明

本評価指標における用語の定義等は以下のとおりである。

（1）医用画像診断支援システム

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）を行う際に参考となる情報を提供するシステム又はソフトウェアであり、CADeやCADxの機能を持つ。支援の方式としては使用者（医用画像診断支援システム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。

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。

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等の提供等により診断支援を行うものを含む。

注) CAdE と CADx の定義は前出の通知より一部修正の上、引用

(2) 人工知能

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システム又はソフトウェア。本評価指標においては、特に、臨床で使用を開始した後、追加データによる機械学習(例えば、深層学習等)により、その性能が変化する機能(市販後学習機能)を有するものを対象とする。

(3) 機械学習

人工知能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。アルゴリズムはさまざまであり、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 システム開発時に利用されてきている。

(4) 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。ニューラルネットワークには、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。

2006年の Hinton らの研究に始まり、囲碁の世界王者を倒した AlphaGo や、各種コンペティションにおいて上位を占めたソフトウェアで使用され、当時の常識を塗り替えるほど突出した成績を示したことから盛んに研究されている。性能の飛躍的向上の一つの理由は、多層のネットワークを学習可能とする新しい技術の登場である。また、深層ニューラルネットワークの学習のための大量のデータと並列分散計算環境の整備もある。

(5) 転移学習

20年以上も前から認識されている機械学習の一分野であり、「新規タスクの効果的な仮

説を効率的に見つけ出すために、一つ以上の別のタスクで学習された知識を得て、それを適用する問題」とされている。

近年、深層学習でも盛んに転移学習が用いられるようになったが、その具体的な利用例として以下が挙げられる。

- 1) 大規模なデータセットで学習させたニューラルネットワークを特徴抽出器として利用し、得られた特徴を別のタスクへ転用する。例えば、一般の自然画像で学習済みのニューラルネットワークに医用画像を入力し、中間層の出力を特徴量としてその医用画像の認識・診断支援等に利用する場合は考えられる。
- 2) 別の目的で学習済みのニューラルネットワークを用意し、異なるタスクの学習データを用いて再学習させる。例えば、CT 像の認識・診断支援を目的とする場合、1) に例示した自然画像で学習済みのニューラルネットワークに CT 像と教師データを与えて再学習させる場合は考えられる。この場合には、一からニューラルネットワークを学習させる場合と比べて、少数の CT 像と教師データで性能の高いニューラルネットワークが得られることが報告されている。

3. 本評価指標の対象

本評価指標は、医用画像診断支援システムのうち、人工知能技術を利用して様々な画像撮影装置で得られた臨床画像から、

- ・ 疾患名まで特定せず、病変の疑いがある部位のみ検出する（いわゆる CADe）
- ・ 病変の疑いがある部位の検出に加え、疾患名の候補を提示する（いわゆる CADx）
- ・ 疾患名の候補の提示に加え、それらの重み付け（順位づけ）まで行う（同上）

等の機能を有するシステム全て（CADe、CADx 等、以降「支援システム等」と記述する。）を対象とする。

対象とする支援システム等は、人工知能技術の利用の形態や学習アルゴリズムの種類にかかわらず、機械学習等により市販後学習等に伴い診断支援性能が変化することを意図したものである。なお、対象はあくまで診断を行う医師を支援するシステムであり、仮にシステムが想定外の挙動、誤動作をした場合に、使用者側でそれを検知できることが基本要件となる。

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販売業者が行うことが難しい人工知能技術を利用した支援システム等、例えば、使用者が市販後学習させることにより施設ごとに異なる性能変化が生じ得るものに関しては、解決すべき事項が数多く存在することから、本評価指標における対象とすることは困難である。しかしながら、人工知能技術の著しい発展に伴い、これらの支援システム等が近い将来開発される可能性に鑑み、それ

らの品質、安全性及び有効性を確保するための基本的な考え方を別添に取りまとめた。

なお、開発する支援システム等が本評価指標に該当するか判断し難い場合は、必要に応じ厚生労働省医薬・生活衛生局医療機器審査管理課に相談すること。

4. 本評価指標の位置付け

近年、支援システム等の開発では、深層学習のようにデータから自動で学習をする機械学習の利用が主流になっている。特に、大量のデータを継続的に収集する仕組みが整備され、学習に用いるコンピュータの性能が飛躍的に向上したことにより、短期間で繰り返し学習をすることが可能となった。また、臨床で得られたデータや支援システム等の結果を自動的に収集し、市販後学習データとして人工知能に与えて性能を変化させることも技術的には実現可能となった。そのため、臨床で使用しながらその性能が段階的又は連続的に変化する支援システム等の実現が期待されている^{注)}。

本評価指標は、対象とする支援システム等において、現時点で考えられ得る問題点、留意すべき事項を示したものであるが、今後の更なる技術革新や知見の集積等を踏まえ改訂が必要なものであり、承認申請内容等に関して拘束力を有するものではない。

支援システム等の評価に関しては、個別の人工知能構築方法、使用目的及び製品特性を十分に理解した上で、科学的な合理性をもって、柔軟に対応することが必要である。なお、本評価指標以外に前出の「コンピュータ診断支援装置に関する評価指標」及び現存する国内外の関連ガイドライン等、例えば、平成 28 年 3 月 31 日付け厚生労働省医薬・生活衛生局医療機器・再生医療等製品担当参事官室事務連絡「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」別添「医療機器プログラムの承認申請に関するガイダンス」等を参考にすることも考慮すべきである。

注) 深層学習の市販後学習の対象には、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。これらのハイパーパラメータも含めた学習とは、例えば、パラメータの組み合わせごとに教師データとネットワークの出力間の誤差（損失）を最小化する学習を網羅的に行い、それらの中から誤差最小のハイパーパラメータを選択すること等が挙げられる。ただし、変更されたハイパーパラメータの程度によっては、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、事前に独立行政法人医薬品医療機器総合機構と相談することを推奨する。

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

(1) ブラックボックスとしての性質

機械学習、特に深層学習は、その特性上、ニューラルネットワークによる判断の過程がブラックボックスであり、アウトプットの判断根拠を明確に示すことが困難な場合がある。このため、通常の医療機器ではその性能確保のために示すことが求められる動作原理の明示が困難となる。そのため、診断アルゴリズムを詳細に精査するというよりは、一般的な医療機器プログラムの評価と同様、そのインプットに対して所要のアウトプットが得られているかを確認することに重点を置き、その性能を評価することが適当と考えられる。

また、動作原理及びアルゴリズムの詳細を記述することが困難であるため、性能の確認が行われた事例以外(例えば、開発時の学習データセットに含まれていなかった希少事例)での挙動等が想定しにくいことも考えられる。このようなリスクに鑑み、機械学習、特に深層学習を利用した支援システム等においては、その性能を確保するために、使用方法及び目的に応じて、製造販売業者が対象疾患の検出率、偽陽性率、偽陰性率、検出に要する時間等、性能に係る値を適切に規定し、その性能が常に担保されていることを示すことが必要となる(6(3)項を参照)。

支援システム等が想定外の挙動を示した場合に、使用者に対して当該挙動を通知する仕組みも必要となる。例えば、想定外の挙動が判断可能な使用者のみに使用を許可する等で使用者側に一定の要件を求める、又は臨床での使用状況を逐次フィードバックすることで、製造販売業者側が性能検証可能となる仕組みを構築する等、支援システム等の目的に応じた適切な方策を用意することが望ましい。(6(4)項を参照)。

想定外の挙動、誤判定等を完全に無くすことは極めて難しいが、そのような情報を収集し、対策を講じることができるような運用方法もあらかじめ考慮しておく必要がある。

(2) 性能変化

支援システム等においては、人工知能技術を利用することで市販後学習による診断アルゴリズムの変化を受けてその性能が向上することが期待される反面、学習の内容等によっては意図しない性能変化が生じ、当初規定していた性能を下回る可能性も否定できない。このような利点とリスクとを勘案し、支援システム等には、継続的な性能検証によるほか、学習用データの要件や学習のプロセス等を規定し、学習の品質を確保するための対応策が求められることになる。

1) 継続的な性能の検証方法

支援システム等の品質、安全性及び有効性を確保する上で、性能が変化するたびに検証する必要がある。臨床使用に伴って市販後学習が行われ、性能が連続的又は相当程度高い頻度で変化する場合には、特にその内容、プロセス管理に加え、変化した性能の検

証及びその管理が重要となる。このことを踏まえ、あらかじめ市販後に生じ得る性能変化水準を臨床上許容される範囲かつ統計学的な妥当性に基づいた範囲で目的に応じて規定しておくこと等とともに、その機器に適用可能な検証方法、変化した性能が規定された範囲を逸脱した際の対策等を定めておくことが求められる（6（3）、6（4）項を参照）。また、検証に使用したデータについては、入手元を明らかにした上でその使用根拠や妥当性を示すことが求められる（6（2）2）項を参照）。

2）性能変化に伴う品質確保

人工知能の性能向上には、その人工知能の学習方式に適した市販後学習が必須であることから、その学習方式を明確にした上で、使用した市販後学習データの詳細を示すことが求められる。（6（2）2）項を参照）。

支援システム等においては、まず、一定量蓄積した臨床データを市販後学習データとして使用することで性能が段階的に変化することが考えられることから、基本的にその性能変化のための学習及び性能検証は製造販売業者が実施することとなり、市販前に求められる要件と同様に性能変化後の品質を確保することが求められる（6（3）1）項を参照）。

一方、市販後学習により性能が連続的又は相当程度高い頻度で変化する機器の場合、製造販売業者の意図に反してその性能が低下し、臨床上許容される水準を下回る可能性がある。このような問題を防ぐために、原則、人工知能は製造販売業者が管理した上でネットワークを介して支援システム等と接続する形式とするとともに、当該支援システム等の目的及びリスクを勘案した検証方法や対策を講じることが求められる（6（3）2）②項を参照）。

性能が変化した支援システム等の品質確保に係る作業は製造販売業者が実施することとするが、品質確保の目的で、使用者による市販後学習機能を有さない、誤使用や適応外使用を防止するため使用者を対象としたトレーニングを実施する、性能が下限以下となった場合の対策を使用者側の端末にも施す等、その支援システム等のリスクに応じて必要な追加対策を施すことが望ましい（6（4）項を参照）。

3）薬事上の手続きにおける考え方

一般的には支援システム等の性能を向上させた結果、臨床上の有効性が向上する場合、薬事上の手続きが必要になる場合があるが、その要求範囲は性能変化の内容及びその変化に伴うリスクを勘案した設定が必要となる。

性能向上に係るバージョンアップが新規製品となる、又はその適用に費用が発生する可能性があることから、その性能変化は対象となり得る販売後の製品全てに適用されず、同じ製品であってもバージョンの異なるものが市場に存在することになる可能性が高い。

このような、人工知能の流通体系の特殊性を考慮しながら、バージョンアップに関するフォローアップ体制や薬事上の効率的な手続きについても検討が必要になる。

(3) 責任の所在

人工知能を利用した医療機器においては、その使用に伴う責任の所在が曖昧になることが想定される。特に、支援システム等においては、その目的が「診断支援」であるにもかかわらず、医師が最終診断まで委ねてしまう危険性を含んでいる。よって、対象となる機器の品質、安全性及び有効性の確保、適正使用のために行うべき方策（使用者に対するトレーニング等）、保守、トラブル対応においては製造販売業者が責任を負うこと、その機器の適切な使用及び医療行為に伴う有害事象（支援システム等における設計・仕様上の不具合に起因した有害事象を除く。）等に関しては医師が責任を負うことを明確にする必要がある（6（4）項を参照）。なお、平成30年12月19日付け医政医発1219第1号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」において、当該プログラムを利用して診断、治療を行う主体は医師であり、医師がその最終的な判断の責任を負うことが示されていることも留意されたい。

6. 評価に際して留意すべき事項

(1) 基本的考え方

本評価指標で対象とする支援システム等は、その特性上、通常の医療機器と異なり、原理（実装する検出・診断アルゴリズム等）や設計仕様等のみで性能を確保することが難しいことから、性能に影響する要素、効果を確認できる範囲や限界等を規定する必要がある。当該要求事項を達成するためには、医学的かつ統計学的に妥当な方法で、それらを規定するとともに支援システム等の性能を検証することが求められる。なお、支援システム等自体の品質、安全性及び有効性を確保するために必要となる評価は、原則として前出の「コンピュータ診断支援装置に関する評価指標」を参考に、目的に応じた性能水準を設定の上、実施することになる。

上述した内容を考慮した上で、現時点の科学水準を基に、本評価指標の対象となる支援システム等の評価において特に留意すべき点を示す。

(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

1) 検出・診断の原理（アルゴリズム）

具体的にその記述が可能な検出・診断アルゴリズムにおいては、従前通り、そのアルゴリズムとプログラム概要を示すことが求められる。深層学習のように処理工程がブラックボックス化する原理を用い、承認申請時に検出・診断アルゴリズムを示すことが困難な場

合には、設計開発時の検出・診断用ネットワーク構造とプログラム概要を示す必要がある。なお、本評価指標が公表された後、ブラックボックス化したネットワーク構造等の振る舞いが説明できるようになった場合には、可能な限りその情報を示すことが求められる。

2) 学習

本評価指標が対象とする支援システム等は、適切な学習データを使用した人工知能の学習によりその目的を達成するために必要となる性能を有することが求められる。よって、対象とする支援システム等の機構や規定された性能等に応じて、以下の項目例を参考にし、明確にすべき項目、それらを使用した根拠及び妥当性を示す必要がある。

- ・ 学習アルゴリズムとプログラム概要
(教師あり、半教師あり、強化学習、自己学習、その他等を明確にすること)
- ・ データ (学習データ、バリデーションデータ^{注1)} 及びテストデータ^{注2)} について、以下を参考に必要な項目を明記すること)

注1) バリデーションデータ：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) を決定するためのデータ

注2) テストデータ：システムの性能を評価・検証するためのデータ

- データソース (取得及び管理方法の明確化を含む。特に、支援システム等により取得した画像等を市販後学習データとする場合には、対象となる患者から同意を取得する方法も含む。)
- 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^{注3)} (アノテーション (例：病変等の位置やサイズ (ラベル画像も含む。)、良悪性の鑑別結果) も含む。)
- 臨床画像データ以外のデータ使用の有無^{注4)} (Data Augmentation 等の目的で作成したコンピュータシミュレーションやファントム使用による画像データ)
- 画像データの枚数、サイズ、濃度レベル数等
- 学習前に加工 (例：匿名化) している場合にはその処理方法
- 臨床データの最終判断を行った医師等
- 臨床データ以外のデータを使用した場合は、その最終判断を行った責任者等
- 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

注3) これまでの医用画像診断支援システムは医用画像のみを用い、臨床的信頼度が高いアノテーションを用いて設計することがほとんどであった。しかし、技術の進歩により、近年では、学習データとして自然画像や人工画像、さらには

紐づけされた臨床情報の信頼性が低いデータ（Weak label データ）を用いることが可能となってきたが、このようなデータを利用する場合は、その使用根拠及び妥当性を示す必要がある。また、テストデータに関する各種留意点に関しては、その評価段階に応じて6（3）項を参照すること。

注4）特に転移学習の場合には、自然画像や別の医用画像、さらには人工画像等を学習データとして用いることが想定される。自然画像を用いた場合には画像中の対象部分、種類及び枚数等、別の医用画像を用いた場合には通常の医用画像を使用した場合と同等の事項、人工画像の場合にはその作成方法等について、その詳細を示すこと。

学習において使用する画像等をデータベース等から入手して使用する場合は、その妥当性を確認することが求められる。以下に、データベース等に関して承認申請時に記載が必要と考えられる点を例示する。なお、記載すべき項目は、データベースごとに異なるため、適切に選択、追加すること。

- ・ データベース運営者（学会、認定匿名加工医療情報作成業者等）の概要、組織体制等
- ・ 運営者が所有する事業計画書
- ・ 運営者が外部に委託している業務内容
- ・ データベースに保有するデータの種類
- ・ データベースの概要及び設計書
- ・ データ管理に係る各種手順書とその運用状況 等

上記の学習にかかわる全ての項目については、設計開発時の学習と市販後学習で共通の項目、相違する項目を区別して明記する必要がある。例えば、設計開発時の学習と市販後学習で学習アルゴリズムが異なる場合は、それぞれについて明らかにする必要がある。また、市販後学習により機械学習アルゴリズムのハイパーパラメータも再決定される場合は、変化する範囲をあらかじめ規定したうえで、そのハイパーパラメータの変化が規定範囲内となることを明らかにすることが求められる。さらに、市販後学習後のネットワークの変遷については性能変化ごとに必要な記録を保存し、後日検証可能な状態にすることが必要となる。

3) 使用環境及び情報セキュリティ

人工知能技術の利用において、その技術がコンピュータ上にあるのか又はネットワークを介して接続しているのかを明確にした上で、次の要件を示すことが求められる。

- ・ 画像撮影装置からのデータ転送方法
- ・ 使用に必要なソフトウェアのインストール等による画像撮影装置又はその他のソ

ソフトウェアへの干渉

- ・ 支援システム等がインターネット等外部機器と接続されている場合、平成 27 年 4 月 28 日付け薬食機参発 0428 第 1 号・薬食安発 0428 第 1 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連名通知「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」を参考にサイバーセキュリティ対策に必要な措置を講じていること
- ・ 支援システム等を使用して得られた臨床情報を市販後学習データとする場合は、以下のガイダンスを参考に適切な管理を行い個人情報の保護を行うこと
 - 「医療・介護関係事業者における個人情報の適切な取り扱いのためのガイダンス」平成 29 年 4 月 14 日、個人情報保護委員会、厚生労働省
 - 「医療情報システムの安全管理に関するガイドライン第 5 版」平成 29 年 5 月、厚生労働省

注) 必要に応じて、個人情報保護法及び次世代医療基盤法も参考にすること

4) 併用する画像撮影装置等の条件

支援システム等が所定の機能を達成するにあたっては、あらかじめ併用する画像撮影装置で得られた医用画像データのフォーマットや解像度等が当該システム等の要求を満たす必要がある。また、市販されている特定の画像撮影装置全てに適用することが可能なものであっても、支援システム等の性能は、人工知能の学習に使用する医用画像データの撮像条件や教師データの作成方法等に影響されるため、必要に応じてそれらについて規定する必要がある。

(3) 安全性、品質及び性能評価

1) 承認申請時

支援システム等の評価においては、対象となる機器の目的やリスクに応じた評価が必要のため、品質、安全性及び有効性の評価に必要な規格値（検出率等）を明確にするとともに、その妥当性を示した上で適切な評価を行うことが必要である。

各種評価においては、6（2）項に記載した項目以外に支援システム等に特有の問題として、

- ・ 接続する画像撮影装置の機能（特に既存の画像撮影装置を利用可能なシステムの場合）
- ・ 責任者を含めた性能検証方法の明確化（治験による性能検証の場合は、「正解」となる最終的な確定診断を得るまでのプロトコール、後向き検証の場合は検証用テストデータに関して、その妥当性を示すために必要な情報を示すことが求められる）

- ・ 性能検証に使用したテストデータの種類、入手元及び妥当性（必要に応じて管理方法等も明らかにすること）
- ・ 市販後に生じ得る性能変化の範囲の規定（検出率の下限、偽陽性率及び偽陰性率の上限設定等）及びその規定根拠となる医学的又は統計学的データ等
- ・ 市販後に性能変化が発生しても、有効性及び安全性が確保されていることを確認するための方策（学習させるデータの入手方法、変化後の性能の検証方法等）
- ・ 性能変化に伴って生じ得る不具合

等、その目的に応じて考慮すべき項目を明確にする必要がある。

特に、性能検証用のテストデータセットは、学習データやバリデーションデータとは重複しないことが求められる一方、対象とする母集団の特徴を考慮して、十分な説明が可能となる質と量のデータが必要となる。しかしながら、機械学習の方法、目的及び対象に応じてその必要数は異なり一律に定義することは困難であるため、支援システム等の性能評価に使用したデータセット及び方法論を記載するとともにその妥当性検証スキーム等を個別に記載することが求められる。

2) 市販後

市販後、支援システム等に発生する性能変化に関しては、承認申請時に規定された性能検証方法に従い、市販後学習データにより変化した性能等が臨床学又は統計学的な観点からあらかじめ規定された範囲内を満たし、品質が確保できることを確認の上、その妥当性検証を行うことが求められる。

支援システム等が既存の画像撮影装置に接続して使用可能な場合、画像撮影装置の性能に依存して入手できる画像データの質が多様となることが想定される。この場合には、得られた画像データが学習データとして使用できることを示す必要があり、その条件を満たさないデータは学習データとしないための対策が必要となる。また、性能検証及びそのために用いるテストデータセットに関する考え方は、基本的に前項と変わらない。加えて、市販後学習データ及びテストデータの臨床最終判断を行った医師等を明確にする必要がある点も、前項と変わらない。

なお、人工知能にネットワークを介して接続する支援システム等において、使用者に必要な性能及び動作の詳細を漏れなく把握できるよう、性能変化が生じた際には使用者にその旨を通知するシステムを有することが望ましい。

以下に、性能変化の実現方法ごとに必要となる留意点を示す。

① バージョンアップごとの段階的な性能変化を意図する場合

製造販売業者が臨床現場で入手したデータ等を定期的に収集し、データの偏り等にも配慮しながら適切な形で人工知能に市販後学習させることにより支援システム等の

性能を変化させる場合がこれに該当する。なお、臨床現場で入手したデータ以外を利用することも想定されることから、それらの利用においては承認申請時に明記するとともにその妥当性を示すことが求められる。

この場合、変化した性能の評価や、その変化に伴う安全性及び品質の確保は製造販売業者によってバージョンアップごとに行うこととなるが、それらの評価にあたっては承認申請時と同様に前出の「コンピュータ診断支援装置に関する評価指標」を参考に行うこと。

また、支援システム等の性能変化において必要となる薬事上の手続きについては、システムの特性にも依存するため、想定可能な範囲については独立行政法人医薬品医療機器総合機構と承認申請前に相談して決定しておくことが望ましい。

② 臨床使用に伴った市販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を意図する場合

支援システム等を臨床現場で利用して入手したデータを市販後学習データとして使用し、連続的又は相当程度高い頻度で性能を変化させる場合には、使用する人工知能は製造販売業者が一括して管理し、各支援システム等とネットワークを介して接続していることが想定される。この場合、自動的に追加されたデータを用いた市販後学習により変化した支援システム等の性能や品質があらかじめ規定された範囲に収まることとが特に求められることから、その学習プロセス及びその管理方法を示すことが求められる。加えて、性能が規定範囲に収まることを検証するための機構をあらかじめ付与するとともに、製造販売業者で定期的にその性能検証を実施することが求められる。なお、当該バージョンアップの場合には、想定外の挙動、誤判定等の問題が生じる可能性が高いため、その解決策（リスクマネジメント）を講じておくことが求められる。

(4) リスクマネジメント

1) 基本的考え方

支援システム等におけるリスクとしては、市販後学習の結果性能が変化し「人工知能が規定された変化範囲を逸脱した結果、誤った情報を提示すること」が挙げられる。バージョンアップごとの段階的な性能変化を意図する支援システム等の場合には、その性能変化によって生じた問題は製造販売業者が解決してから上市することが前提となるため、そのリスクは低い。一方、市販後学習データが直ちに実装されることで性能が連続的又は相当程度高い頻度で変化する支援システム等の場合には、上記リスクが生じる可能性があるため、そのリスクを回避するためのシステムが必要となる。

また、適応外使用に伴い、上記リスクに加え、使用者が支援システム等における想定外の挙動に気付かない可能性も増大することが考えられるため、適応外使用を防ぐ方法

を講じることが求められる。

2) 対策について

対策が必要となる支援システム等には、上記リスクを軽減するために、

- ・ 市販後学習データとする画像に関する詳細なログ（使用環境、診断結果、責任者等）を保存するための機能（6（2）を参照）
- ・ 支援システム等の内部データを利用した自己テスト等により、定期的に性能評価を行い、その時点での性能を使用者に示すことで最低限の性能を確保・確認できる機能
- ・ 上記機能により上記リスクが発生していることが確認された場合に、人工知能を市販時の状態に初期化又は変化直前の状態へと復元する機能
- ・ 上記リスク発生時に、必要に応じて、支援システム等を停止させる機能
- ・ 臨床での使用状況を逐次フィードバックして製造販売業者がその性能を適宜検証可能となる機能

等、当該支援システム等の承認申請時に確認された性能を確保するための機能、又は具体的な対策手順をその支援目的に応じて設定しておくことが求められる。

使用者は支援システム等の目的に応じた専門知識を有し、かつ、その使用等に関する教育（トレーニング）を受けた医師を始めとする医療従事者に限定する一方、確定診断、最終判断等、診療行為における最終責任者は医師と規定することが求められる。加えて、使用者の誤使用や支援システム等の目的から逸脱した適応外使用を防ぐため、ユーザトレーニング等の適切な対策を講じておくことが求められる。なお、このユーザトレーニングは関連学会等と連携したものが望ましい。

7. 機器の特性に応じた追加留意事項（案）

(1) 市販後学習による性能変化が診断支援においてハイリスクとなる支援システム等の場合

例えば、病理診断においては、その診断が臨床における最終診断であることから、どのような形式であってもその性能が市販時から変化することは検出率及び正診率低下の可能性を含むこととなり臨床上高いリスクを伴うこととなる。このような支援システム等の場合、その品質、安全性及び有効性を十分に確保する必要があることから、現時点では、市販後学習プロセス、性能変化及びその検証には必ず病理専門医の関与を要求することが望ましい。

(2) 支援システム等の対象画像が動画の場合

消化器内視鏡機器や超音波画像診断装置のように観察像が動画となる場合、人工知能に

対する学習データはその動画をどのような形で利用したかを明確にする必要がある。よって、動画のフォーマットや取得方式を明示するほか、学習データへの変換方法を明らかにする必要がある。

また、当該支援システム等においては、動画の撮影条件及び質が撮像者の技能に左右されることから、学習データとして利用可能となる条件を規定するとともに、その妥当性を明らかにする必要がある。

以上

使用者等が性能変化に関与可能な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基本的考え方（案）

1. 背景

今回、人工知能技術を利用した医用画像診断支援システム（支援システム等）に関する評価指標案を作成するにあたり、近年の計算機技術の著しい発展に伴い、実現可能な様々な支援システム等を全て評価指標案の対象に含めるかについて、WG で議論が行われた。特に、人工知能の中でも、近年着目されている深層学習を利用した場合の可能性を考慮し、コンピュータ技術が発達することにより、現在の大型コンピュータと同レベルのスペックの汎用 PC が使用可能となることで、ネットワークを介さずに PC 中に人工知能を有し、特定の臨床現場で市販後学習を行い性能が変化する場合を対象に含めるかも議論された。

このケースには、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販売業者が行うことが難しい支援システム等、例えば、施設又は使用者ごとの市販後学習による性能変化が可能なシステム、教師なしデータによる学習が可能なシステム等が含まれる。これらの支援システム等は、解決すべき事項が数多く存在することから、本評価指標の対象外とすることとした。しかしながら、人工知能技術の著しい発展に伴いこれらの支援システム等が近い将来開発される可能性に鑑み、それらの品質、安全性及び有効性を確保するための考え方を別添として取りまとめた。

2. 問題点と解決策について

（1）性能変化

例えば、使用者による市販後学習により性能が変化する場合、開発者以外が主体となって市販後学習を行うことになるため、製造販売業者の開発意図と大きく異なる方向にその性能が変化・低下し、臨床許容される水準を下回る可能性が否定できない。このような問題を防ぐためには、適切な質と量のテストデータによる検証に加え、当該支援システム等のリスクを勘案した対策を講じることが求められる。例えば、製造販売業者による品質管理を徹底するための方策を講じる、製造販売業者と同等の品質マネジメントシステムを臨床現場に求める、使用者を対象としたトレーニングを実施することに加え、性能が下限以下となった場合にはそのリスクに応じた追加対策を施すこと等が求められる。

支援システム等の目的は「診断支援」であることから、確定診断時における医師の最終的な判断によりそのリスクを回避することができると考えられるが、可能であれば生じたリスクを漏れなく製造販売業者に連絡する機能、又は手順を必要に応じて設けることが望ましい。

（2）責任の所在

その機器の使用に伴う医療行為、適応外使用、不具合等に対する報告に関しては医師が責任を負うことが原則であるが、その原則を一様に適用できない可能性があることに留意して、製造販売業者及び医師の責任範囲を設定する必要がある。

3. 評価に際して考慮すべき事項（品質、安全性及び有効性の評価）

使用者が施設ごとに学習を行うシステムと、自己学習を行うシステムに関する留意点は以下のとおりである。

（１）使用者による施設ごとの市販後学習が可能な支援システム等

特定の医療機関で得られた臨床データや診断支援の結果を元にした市販後学習を使用者が実施するケースが該当する。この場合、市販後学習データの偏りにより、医療機関ごとに異なる方向に性能が変化し、製造販売業者による品質管理が困難となる。加えて、使用者が市販後学習を実施するため、製造販売業者の開発意図と大きく異なる方向にその性能が変化するリスクがある。そのため、性能評価においては、評価指標に記載した事項に加え、製造販売業者は、

- ・ 市販後学習プロトコル。特に使用者が施設ごとに行う市販後学習データの収集方法や品質管理等も含めたプロトコルの詳細
- ・ 市販後学習後に変化したシステムの性能の検証プロトコルに関する詳細。承認申請時のテストデータに対する性能評価結果のみならず、新しいテストデータを追加する場合には、その収集方法や品質の評価法と結果、及びシステムの性能評価結果に関する詳細

に留意し、製造販売業者が市販後学習を実施した場合と同等の品質マネジメントを臨床現場に求めるための工夫が必要である。

また、市販後学習に加え、変化した性能の評価や、その変化に伴う支援システム等の安全性及び品質の確保を使用者等が行う必要があるため、それらに係る評価は製造販売業者が実施する場合と同等の評価を行う必要がある。そのため、当該支援システム等を有する医療機関内に、以下の要件を充足した責任者の指名と、その責務、性能評価及び品質管理に関する手順書の整備等を求めることになると考えられる。

- ・ 対象となる支援システム等の原理及び学習方法を熟知していること
- ・ 市販後学習データの妥当性を示すことができること

なお、その性能変化の範囲が承認申請時に製造販売業者により規定された範囲内であることが前提となるため、性能評価に関しては製造販売業者が承認申請時に規定した検証方法を用いる必要があり、性能変化が製造販売業者により規定された範囲を超えることが生じないよう、評価指標 6（４）項「リスクマネジメント」に記載した事項に加え、当該支援システム等にはそれを防止するための機構、具体的には、そのリスクを製造販売業者に連絡する機能等を設けることが強く求められる。

（２）「教師なし」データによる自己学習が可能な支援システム等

2017年10月、Googleにより囲碁用人工知能において「教師なし」データを用いた自己学習により連続的な性能変化が可能となったことが Nature にて報告された。将来、同様の人工知能を利用して医師の最終判断や製造販売業者の介入を要することなく、医療現場での使用中、自律的にその性能を変化させていく支援システム等が出現する可能性は否定できない。加えて、その人工知能がネットワークから切り離された状態で搭載され、製造販売業者が関与しない状態で納入先の医療機関ごとに性能が変化する支援システム等の開発も予想される。しかしながら、このような支援システム等がどのように市販後学習を進めるかについては不明な点も多い。

上記のような自律性の高いシステムは、現在の科学技術水準では期待されるレベルの診断支援能力を持つには至っていないが、将来実現した場合には、再学習時に技術者や医師による学習データの作成が不要となり、教師あり学習の場合と比べて支援システム等の自律度は高くなる。また、市販後学習用データの収集や市販後学習のタイミングも全て自ら判断して実行し、技術者や使用者による介入が完全に不要な自律した支援システム等が登場する可能性もある。このような人による介入度が低い、又は不要な支援システム等では特に品質管理が重要となり、承認申請時の性能を担保していることの確認やリスクマネジメントに関する要求は自律度の上昇に伴って高くなる。これらの要求等を充足するためには、前項までに挙げた各種の要求事項をより高度なレベルで満たすことが求められる。充足すべき事項及びそれらのレベルに関しては現時点で評価指標として具体的に記載することは困難であることから、承認申請に先立って独立行政法人医薬品医療機器総合機構と詳細な相談をしておくことが必要になると考えられる。

以上

V. 人工知能技術を利用した医療機器プログラムの 承認状況

(公開情報等から確認された製品に限定)

人工知能技術を利用した医療機器プログラムの承認状況 (公開情報等から確認された製品に限定:2019年02月現在)

<米国>

メーカー	製品名	クラス	効能等	アルゴリズム	補足情報	承認日 種類	ソース(Web)
HeartVista	HeartVista	Class II	心臓MRIの撮像効率化を図るプログラムで、位置決めなどに機械学習を使用している。	機械学習	複数台の心臓MRI検査をリモートで操作することにより、放射線技師の人的コストを抑え検査の時間短縮も可能となった。	2014/6/25 510(k)	https://www.accessdata.fda.gov/cdrh_docs/pdf13/K133848.pdf https://static1.squarespace.com/static/f5be92c4a81551253039ac6521t5bc6ac6b42a73da18617ba8/1543269181697/BROCHURERGB.pdf
Riverrain Technologies	ClearReadCT	Class II	無症候性集団を対象としたCT検査時において、放射線医による肺結節の検出を補助する。	機械学習	類似製品としてClearRead Xrayシリーズが販売されており、Detect™は早期の肺がんの可能性がある胸部X線上の領域の特定が可能と標榜されている。	2016/4/26 510(k)	https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmm/pmm.cfm?ID=K161201 https://www.riverraintech.com/ https://www.accessdata.fda.gov/cdrh_docs/pdf16/K163253.pdf
Arterys, Inc.	Caldio DL	Class II	血流解析(シミュレーション)を行う。	不明	HPでは、製品名がCaldio AIと変更されている。	2017/1/5 510(k)	https://www.forbes.com/sites/bernardmarr/2017/01/20/first-fda-approval-for-clinical-cloud-based-deep-learning-in-healthcare/#550ba701161c https://www.prnewswire.com/news-releases/arterys-receives-first-fda-clearance-for-broad-oncology-imaging-suite-with-deep-learning-300599275.html https://www.arterys.com/lung https://www.arterys.com/liver
Quantitative Insights, Inc.	Liver AI Lung AI QuantX	不明 Class II	放射線医による肝臓のMRI及びCT画像、肺のCT画像からの肝がん及び肺がん(結節)検出を補助する。 乳腺MRI画像を基にした放射線医による乳房異常の評価及び特性解析を補助する。	深層学習 機械学習	Webを介して使用するクラウド型プログラムである。メーカーのホームページでは、規制当局の認可を受けているのは米国、ヨーロッパ、カナダのみであること、Lung AIに関しては「研究用途」のみ認められていると紹介されている。 入力した画像のセグメンテーションを行い、医師が指定、あるいは自動検出した特定領域の容量及び形態学的解析を行う。First readerとしての使用は認められておらず、処方目的のみで使用される。	2018/2/15 (press release 日) 不明	https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmm/denovo.cfm?ID=DEN170022 https://www.accessdata.fda.gov/cdrh_docs/pdf17/K170981.pdf
CorTechs Lab	NeuroQuant	Class II	MRI画像からセグメンテーション可能な脳構造及び病変を対象にラベリング、可視化及び体積定量化を自動的に行う。	不明	使用者は放射線医等を含む医療専門従事者のみである。汎用PCにインストールして、DICOMをそのまま入力データとして使用する。	2017/9/17 510(k)	https://www.cortechslabs.com/ https://www.accessdata.fda.gov/cdrh_docs/pdf17/K171816.pdf
Alivecor, Inc.	KardiaBand	Class II	心房細動を検出した場合、その結果をメールで使用者及びかかりつけ医に送信する。	不明	第三世代のApplewatch用にセンサ機能を有するバンドとアプリが一式で承認、販売されている。同様の機能を有するKardia mobileも存在する。	2017/1/16 510(k)	https://www.alivecor.com/press/press_release/fda-clears-first-medical-device-for-apple-watch/

Viz.AI	ContaCT	Class II	CT画像を解析し、患者が脳卒中になる可能性をアラートという形で表示して医療関係者の診断を支援する。	不明	使用が許可されているのは脳血管外科等の専門家のみである。性能評価においては、300のCT画像の後ろ向き試験を実施し、2名の専門家の診断結果との比較を行っている。また、塞栓栓が疑われるケースを検知するための治療も実施されている。	2018/2/13 De novo	https://www.accessdata.fda.gov/cdrh_docs/reviews/DEN170073.pdf https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm596575.htm
Densitas, Inc.	DM-Density	Class II	マンモグラフィ画像から乳腺組織密度を算出するとともに、米国放射線専門医学会が作成した BI-RADS分類を行う。	機械学習	乳房の組織構造を評価するのみで、診断補助を目的とはしないことを標榜している。	2018/2/23 510(k)	https://www.accessdata.fda.gov/cdrh_docs/pdf17/K170540.pdf https://densitas.health/ https://www.accessdata.fda.gov/cdrh_docs/pdf18/K180326.pdf
Icometrix, NV	icobrain	Class II	CT及びMRI画像から脳の容積変化や白質異常を可視化し定量的なレポートを作成することで認知症患者の症状診断を補助する。	機械学習	使用できる画像はMRI画像のみである。国内では株式会社マイクロンが脳画像解析プログラムとして解析プログラム部分のみ認証を受けて販売されている。 (230ALBZX00010A01)	2018/3/8 510(k)	https://icometrix.com/news/icometrix-expands-ai-portfolio-with-new-icobrain-reports-to-aid-the-diagnosis-of-dementia https://micron-kobe.com/archives/works/icobrain
Quantib B.V	Quantib™ Brain 1.3	Class II	脳のMRI画像から、自動でラベリング、可視化及び体積定量化を行う。また、脳の白質、灰白質及び脳脊髄液のセグメンテーション、可視化及び定量化も行う。	深層学習	標榜されている効能は前のバージョンと全く同じだが、アルゴリズムが異なる。また、CT画像解析はこのバージョンから可能となっている。	2018/1/15 510(k)	https://www.accessdata.fda.gov/cdrh_docs/pdf18/K181939.pdf
Edwards Lifesciences	Acumen Hypotension Prediction Index (HPI)	Class II	外科手術中に発生する患者の血圧低下または低血圧症に対処するため、発生前にその可能性を医師に警告する。	機械学習 (200,000例以上のデータを学習)	大脳白質病変及び脳萎縮量の検出が可能 な新製品であるQuantib NDが2019年1月7日に承認された旨がホームページに掲載されている。アルゴリズム等の詳細は不明である。	2018/3/9 510(k)	https://www.accessdata.fda.gov/cdrh_docs/pdf17/K173939.pdf https://www.quantib.com/ https://www.accessdata.fda.gov/cdrh_docs/pdf16/DEN160044.pdf https://www.edwards.com/jp/aboutus/newsroom/pressrelease/20180319-2/
iCAD	PowerLook Density Assessment	Class II	乳房トモシンセシスにより得られた画像を二次元画像として合成し、乳腺密度の濃い部位の検出結果をその上に提示するとともに米国放射線専門医学会が作成した BI-RADS分類を行う。	深層学習	先行品で、同様の仕組みにより異なる領域が強調された画像が表示され、その部分をクリックすると異常候補部位のあるスライス面が表示される機能を有する「Powerlook Tomo Detection」がPMAで2017年3月27日に承認されている。なお、先行品のアルゴリズムも深層学習であることが、適用可能な装置を販売しているGE社のホームページに記載されている。本製品は先行品よりも精度や速度を高めた結果患者の被曝量をより少なくすることができると、GE社だけでなくHologic system社の装置にも適用可能であることが追加で標榜されている。	2018/4/5 510(k)	https://www.accessdata.fda.gov/cdrh_docs/pdf18/K180125.pdf https://www.icadmed.com/newsroom.html#/posts/iCAD-Receiver-FDA-Clearance-of-PowerLook-Density-Assessment-for--Digital-Breast-Tomosynthesis-/109 https://www.fda.gov/downloads/MedicalDevices/ProductsandMedicalPractices/PMAApprovals/UCM551042.pdf (先行品に関する情報)

IDX, LLC	IDx-DR	Class II	網膜カメラ「Topcon NW400」で撮影した画像を解析し、網膜症診断を行う。	不明	過去に網膜症と診断されていた糖尿病患者が使用対象である。画像を読み込み、中程度以上の網膜症症状が検出された場合は眼科専門医の受診を、症状が検出されない場合は12カ月後の再検査を推奨する機能を有している。	2018/4/11 De novo	https://www.accessdata.fda.gov/cdrh_docs/pdf18/DEN180001.pdf https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm604357.htm
Imagen	OsteoDetect	Class II	2方向から取得した2次元X線画像から、成人手首の骨折状態(遠位骨折)の診断を支援する。	機械学習	性能評価においては、1000枚の画像の後ろ向き評価結果を3人の専門医診断結果と比較するとともに、200例を観察した24のproviderとの比較検証を実施し、その有用性を示している。	2018/5/24 De novo	https://www.accessdata.fda.gov/cdrh_docs/pdf18/DEN180005.pdf https://www.fda.gov/news-events/newsroom/pressannouncements/ucm608833.htm
Thirona	LungQ	Class II	胸部CT画像から肺における空気の流入状態等を解析し、肺の亀裂や呼吸障害が疑われる部分を検出し、診断を支援する。	不明	欧州でCEマークを取得済みであり、オーストラリアでも認証済みである。	2018/6/5 510(k)	https://www.accessdata.fda.gov/cdrh_docs/pdf17/K173821.pdf https://thirona.eu/index.php/lungq/
DreaMed Diabetes, Ltd.	DreaMed Advisor Pro	Class II	糖尿病治療支援を目的に、グルコースモニター、インスリンポンプ及び自己モニタリングからのデータを継続的に分析してインスリン投与量を推奨する。	機械学習 (Event-driven)	クラウドベースのプログラムで、学習プロセスを通じて、基礎レート、炭水化物比率、補正係数等の要素をその推奨を判断するために組み入れられる。推奨投与量はモニタリングを担当する臨床医に直接配信され、ボタンを押すのみで患者の糖尿病管理調整が可能である。	2018/6/12 De novo	https://www.accessdata.fda.gov/cdrh_docs/pdf17/DEN170043.pdf https://dreamed-diabetes.com/
Zebra Medical Vision	HealthCCS	Class II	CT画像から自動的に冠動脈石灰化スコアを算出する等により冠動脈疾患診断を補助する。	深層学習	深層学習による画像解析プラットフォームの開発が主目的の企業であり、他の疾患に関する診断補助プログラムも開発中である。	2018/6/13 510(k)	https://www.accessdata.fda.gov/cdrh_docs/pdf17/K172983.pdf https://venturebeat.com/2018/07/12/zebra-medical-gains-fda-approval-for-ai-powered-heart-disease-detection/
Bay Labs	EchoMD AutoEF	Class II	心エコー診断時に適切な撮画像を動画から自動的に選択し左心室駆出率を自動的に計算して診断支援を行う。	深層学習 (9000人の患者から得た400万枚以上の画像データを学習)	使用対象は大人の患者に限られている。	2018/6/14 510(k)	https://www.zebra-med.com/press/ https://www.accessdata.fda.gov/cdrh_docs/pdf17/K173780.pdf https://www.businesswire.com/new-s/home/20180619005552/en/Bay-Labs%E2%80%99-EchoMD-AutoEF-Software-Receives-FDA
Mirada Medical Ltd.	Workflow Box	Class II	DICOMデータからなる各種の医用画像において、臓器等の輪郭を自動的に作成するDLCExpert™や画像データ処理プログラム等から構成され、主に放射線腫瘍治療計画への利用を目的とする。	深層学習	DLCExpert™は、深層学習により、重要な構造物(疾患の可能性がある臓器)の輪郭を自動的に検出する。トレーニングを受けた専門医療従事者のみで使用可能であるが、病変を自動検出することを目的としていない。	2018/7/10 510(k)	https://www.accessdata.fda.gov/cdrh_docs/pdf18/K181572.pdf https://mirada-medical.com/
Aidoc	BriefCase	Class II	医療用画像管理システム(PACS)からデータを受け取って、頭部における急性期の異常(脳内出血など)を検出するための診断支援画像を再構築する。	深層学習	当該製品は米国、欧米で承認を受けている。類似製品として開発されている脊椎疾患を対象としたプログラムは欧州でCEマークを取得済みであり、胸部及び腹腔内疾患を対象としたプログラムは研究目的のみで提供されている。	2018/8/1 510(k)	https://www.accessdata.fda.gov/cdrh_docs/pdf18/K180647.pdf https://www.aidoc.com/

physIQ Inc.	physIQ Heart Rhythm Module (Version 1.0)	Class II	バイオセンサを用いて測定した心電図から心拍数や心拍変動を計算し、心房細動を検出する。	不明	医師及び資格を有する医療従事者のみが使用できる。使用できるバイオセンサはFDAが承認したもののみと限定されている。	2018/8/10 510(k)	https://www.accessdata.fda.gov/cdrh_docs/pdf18/K180234.pdf
Qompium Nv	FibriCheck	Class II	スマートフォンにインストールし、心房細動の罹患率、あるいはその疑いがある健康人が自己診断のために心電計として使用する。	不明	スマートフォンのカメラに指を乗せて心拍を測定し、結果を自動的に登録したかかりつけ医と共有する。	2018/9/28 510(k)	https://www.accessdata.fda.gov/cdrh_docs/pdf17/K173872.pdf
Apple	ECG App	Class II	Applewatchにインストールし、そのセンサーで読み取った心拍等を表示、記録する。	不明	米国では、Applewatchに両方のアプリをインストールして販売されている。	2018/9/11 De novo	http://www.accessdata.fda.gov/cdrh_docs/pdf18/DEN180044.pdf
	Irregular Rhythm Notification Feature	Class II	Applewatchにインストールし、心房細動や心拍異常等の異常を検出した際に使用者に警告を通知する。	不明			http://www.accessdata.fda.gov/cdrh_docs/pdf18/DEN180042.pdf
NinePoint Medical, Inc.	NvisionVLE Imaging System, NvisionVLE Optical Probe	Class II	内視鏡イメージング装置であり、光干渉断層画像計を利用して取得した2次元画像、断面像、深さ方向のリアルタイム画像から、食道等における微細組織構造を評価する。	機械学習	従来のイメージング技術では判別が困難だった潜在的な疾患について、食道表面と表面直下組織を評価することができる。また、最新のアルゴリズムを使用してリアルタイムで食道のセグメンテーションを行い、内視鏡手術中に臨床医による関心領域の識別を支援する。	2018/10/16 510(k)	https://www.accessdata.fda.gov/cdrh_docs/pdf18/K182261.pdf http://www.ninepointmedical.com/
MaxQ-AI, Ltd.	AccipioIx	Class II	取得したCT画像を自動的に解析し頭蓋内出血の疑いを示し速やかな治療を促す。	深層学習	CE社のCT装置に適用可能なプログラムも2018/11/27から提供されるようになったこと、脳内出血による脳卒中や障害の診断補助が可能になったことが標榜されている。	2018/10/26 510(k)	https://www.accessdata.fda.gov/cdrh_docs/pdf18/K182177.pdf http://maxq.ai/2018/11/07/maxq-ai-receives-fda-clearance-for-accipio-ix-intracranial-hemorrhage-platform/
ScreenPoint Medical BV	Transpara	Class II	マンモグラフィのコンカレントリダーとして用い、乳がんが疑われる部位の特定や悪性度評価を補助するために用いる。	深層学習	病変や石灰化が疑われる位置を示すとともに、その疑いの程度をスコア化して示す。	2018/11/21 510(k)	https://www.accessdata.fda.gov/cdrh_docs/pdf18/K181704.pdf http://www.screenpoint-medical.com/transpara
Subtle Medical, Inc	SubtlePET	Class II	放射線医等がフルドオキシジェングルコース、あるいはアミロイドを利用したPETで撮影した画像の移動、保存及びノイズ除去に使用する。	深層学習	画像のノイズ除去及び高画質化処理において深層学習を基にしたアルゴリズムを使用している。	2018/11/30 510(k)	https://www.accessdata.fda.gov/cdrh_docs/pdf18/K182336.pdf https://subtlemedical.com/
iCAD	ProFound AI	Class III	乳房トモシンセシスの各スライス画像で腫瘍性病変や石灰化と判断される候補領域を取り囲んで示す。また、病変単位、あるいは症例単位で検出された異常部位の悪性度を数値化して示す。	深層学習 (がん症例4000以上を含む1万2千症例を学習)		2018/12/7 不明	https://www.icadmed.com/newsroom.html#/posts/ICAD-Announces-FDA-Clearance-of-ProFound-AI-for-Digital-Breast-Tomosynthesis/124

＜韓国＞

メーカー	製品名	クラス	効能等	アルゴリズム	補足情報	承認日	ソース(Web)
VUNO Inc.,	Bone Age	不明	手のX線写真から骨年齢を自動評価し、正常か異常かのみを提示する。	不明	韓国で医療機器として承認済である。異常部位などを示すことは認められていない。骨年齢評価においては、平均で医師の評価と0.9ヶ月の違いに留まったとHPに記載されている。	2018/5/16	https://www.vuno.co/news_detail/48
Lumit	Lumit Insight CXR-Nodule	不明	胸部X線写真を対象にヒートマップあるいは輪郭で異常部位(肺結節)の検出位置を提示する。合わせて、検出確信度(possibility of findings presence)も提示することができる。	深層学習 (7万以上のCT画像を学習)	97%の精度で異常部位を検出し、医師による正診率が20%向上したとHPに記載されている。	2018/8/1	https://lumit.io/product/#INSIGHT-CXR-Nodule

JLK Inspection	JBS-01K	Class III	MRI画像から虚血性脳梗塞の検出及びその重篤度判断を支援する。	不明	不明	2018/9/6	https://www.marketwatch.com/press-release/ai-based-class-iii-medical-device-of-jlk-inspection-first-approved-by-ministry-of-food-and-drug-safety-of-south-korea-2018-09-06
----------------	---------	-----------	---------------------------------	----	----	----------	---

< 欧州、その他 >

メーカー	製品名	クラス	効能等	アルゴリズム	補足情報	承認日	ソース(Web)
qure.ai (インド)	qXR	不明	胸部X線画像から、結核やその他の肺疾患を検出するためのスクリーニングを行う。	深層学習 (150万以上のX線画像を学習)	欧州でCEマークを取得済みだが、米国では研究目的のみで使用可能である。	不明	http://qure.ai
Aidence (オランダ)	Veye Chest	不明	胸部CT画像から異常部位(肺結節)を検出しその位置、大きさ及び状態を測定する。	不明	コンカレントリーダ、あるいはセカンダリーダとして欧州でCEマーク取得済みである。	不明	https://aidence.com/our-solution
contextflow (オーストリア)	製品名不明 (contextflow search engine)	不明	CT画像の特定領域を指定すると、データベース内の類似画像を提示する。	深層学習 (7000のCT画像を学習)	現在欧州でCEマーク取得申請中である。	—	http://contextflow.com/

< 日本 >

メーカー	製品名	クラス	効能等	アルゴリズム	補足情報	承認日	ソース(Web)
サイバネットシステム 株式会社	EndoBRAIN	Class III	超拡大内視鏡画像により、大腸病変の腫瘍/非腫瘍の判別を支援する。	サポートベクターマシン	AMEDの支援を受け、昭和大学、名古屋大学、サイバネットシステム(株)が連携して開発。正診率98%の精度を表現した。	2018/12/6	https://www.amed.go.jp/news/releases_e_20181210.html http://www.innervision.co.jp/products/release/20190115

参考情報

< 中国 >

メーカー	製品名	クラス	効能等	アルゴリズム	補足情報	承認日	ソース(Web)
12 Sigma Technologies	不明	—	CT画像から肺結節を検出する。	深層学習 (10万症例の画像を学習)	中国規制当局からは承認を受けたと報告されている。他に、具体的な箇所は示さないまま、胸部X線画像から15種の病変の有無を検出するプログラム等を開発中。	不明	北米放射線学会(RSNA)2018
IMSIGHT	Lung Sight	—	CT画像から肺結節を検出する。	深層学習 (20万症例の画像を学習)	検出率95%を達成と標榜している。	不明	北米放射線学会(RSNA)2018
	DR Sight	—	トリアージを目的として胸部X線画像から14種類の疾患を検出する。	深層学習 (5万症例の画像を学習)	これらの他に、骨年齢計測、肝臓疾患検出、病理診断及び治療目的のプログラムを開発している。	不明	北米放射線学会(RSNA)2018
Yitu Healthcare	不明	—	CT画像から肺結節を検出し、良悪性等を鑑別する。	深層学習	マンモグラフィや乳房超音波CAD開発のために、150病院から10万症例を収集している。	不明	北米放射線学会(RSNA)2018
United Imaging Intelligent	不明	—	胸部X線画像から、可能性のある疾患の有無と種類を10種類の疾患の中から提示する。	深層学習 (3-10万症例の画像を学習)	胸部CT画像から肺結節、心肥大、肋骨骨折を検出するプログラムも開発している。	不明	北米放射線学会(RSNA)2018
Huiying Medical Technology	不明	—	CT画像から肺結節を検出する。	深層学習 (5万症例の画像を学習)	肺結節の体積を示すことはできない。他に胸部X線を利用したプログラムや、手の骨折診断用プログラムを開発している。	不明	北米放射線学会(RSNA)2018

VI. 参考資料

- 1 : 韓国 AI ビッグデータガイドライン和訳版（医療機器該当性部分等を抜粋）
- 2 : 中国におけるクラス分類表英訳版（Medical Software 部分を抜粋）
- 3 : 合同検討会報告資料
- 4 : その他関連資料 URL

韓国 AI ビッグデータガイドライン和訳版 (医療機器該当性部分等を抜粋)

II 医療機器区分基準および品目分類

1. 医療機器区分基準

a) 概要

現在ビッグデータおよび人工知能技術を活用するソフトウェアが開発されており、今後の技術発展速度に応じて、より多様で複雑な機能を持つ高度な製品が登場することが予想される。これに伴い、現時点で開発・使用されている製品および近い将来登場すると予想される製品に対して、医療機器としての管理必要性を評価する上での医療機器判断基準と管理方を提示することが本節の目的である。

b) ビッグデータおよび人工知能技術が適用された医療分野のソフトウェアに対する規制的アプローチ

ビッグデータおよび人工知能技術が適用された医療分野のソフトウェアは、医療関係者および患者に膨大な情報をリアルタイムで提供することで正確な医療上の意思決定を支援することが目的である。患者診療の質、医療決定の正確度、効率性を高めることで、医療関係者および患者の満足度を高めるのに寄与することができる。

ビッグデータおよび人工知能技術が適用されたソフトウェアに対する規制は、技術発展の速度、頻繁な変更とアップグレード、複雑なアルゴリズムなどを反映することができるように柔軟である必要があり、患者および医療関係者などの使用者との使用環境を考慮しなければならない。したがって、食品医薬品安全処は医療機器の定義を充足することが明確なソフトウェア、または本来の目的通り機能しない場合に患者に危険を及ぼす恐れがあるソフトウェアを医療機器として管理するために、今後の製品開発動向および使用現況などを反映して、医療機器で管理の必要性を持続的に検討する予定である。

c) 医療機器判断基準

医療機器とは、「医療機器法」第2条第1項により人や動物に単独または組み合わせて使用される機構・機械・装置・材料または、これと類似の製品として次の各号のいずれかに該当する製品をいう。

医療機器法第2条(定義) "医療機器"とは、人や動物に単独または組み合わせて使用される機構・機械・装置・材料または、これと類似の製品として次の各号のいずれかに該当する製品をいう。ただし、「薬事法」にともなう医薬品と医薬部外品および「障害者福祉法」第65条にともなう障害者補助器具のうち義肢・補助器は除く。

1. 病気を診断・治療・軽減・処置、または予防する目的で使用される製品

2. 傷害、または障害を診断・治療・軽減、または補正する目的で使用される製品
3. 構造、または機能を検査、置換、変形する目的で使用される製品
4. 妊娠を調節する目的で使用される製品

ビッグデータおよび人工知能技術を活用するソフトウェアの医療機器該当性は使用目的と危害度により判断する。医療機器該当の有無は医療機器法第2条の規定による使用目的と以下の危害度判断要素を考慮して総合的に判断しなければならない。

1) ソフトウェアが意図した通りに作動しない場合、患者に危害を及ぼす可能性があるか

医療用ソフトウェアは正確度が担保されない場合、国民の健康に危害を発生させる恐れがある。例えば、不正確・不適切なデータの入力・学習、アルゴリズムのエラーが原因で病気の有無に対する可能性の程度予測、あるいは異常部位の検出・表示を誤った場合には診断・治療結果に直接的な影響を及ぼす。このような場合、患者は必要な処置を受けることができなかつたり、不必要な検査、手術、医薬品処方を受けたりといった問題が発生する可能性がある。

2) ソフトウェアが医療関係者の臨床的判断を保証できるか

食品医薬品安全処から安全性と性能に対して検証を受けなかった医療用ソフトウェアの場合、医療関係者が自主的に当該ソフトウェアの正確度、信頼度を検証することはできない。このような状況では、当該製品を用いて不正確な診断や治療方法などが決定される可能性があり、その結果、国民の健康に危害を及ぼす恐れがある。したがって、医療関係者が介入する過程の存在を判断することが難しい領域であるかどうかと医療関係者が情報や治療方法勧告の根拠を検討することが可能な合理的機会が与えられているか、検討が必要となる。

医療関係者が患者の臨床的診断や治療方法などを決定する場合、当該製品の推奨だけで主要な判断を下すものではないことを提示しなければならない。また、医療関係者が提供された情報の臨床的根拠を把握できるように、学習データのソース、学習データと当該製品により提供される結果との相関関係などを、十分に説明しなければならない。

具体的な範囲および例示

医療機器の判断基準に基づく、医療機器と非医療機器の範囲は以下の通りである。例として挙げた事例は、基準の理解を助けるためのものであり、実際には製品別の特性と状況、科学的根拠などを考慮して事案ごとに判断する。

1) 医療機器に該当する医療用ソフトウェア

- a) 医療用ビッグデータに基づいて、医療情報を分析して得られた臨床情報（例えば、腫瘍病変サイズ・位置など）を利用して、患者の病気の有無、状態などの可能性の程度を自動的に診断・予測、監視、または治療するためのソフトウェア

- b) 医療用ビッグデータに基づいて、医療画像、体外診断機器から出てきたシグナル、信号獲得システム（心電計、脳波計など）から出るパターン、またはシグナルを分析して、診断・治療に必要な臨床情報を提供するソフトウェア

<医療機器の例>

- 肺 CT 画像を分析して、肺がんの有無や肺がんの進行状況（悪性度）を自動的に診断するためのソフトウェア
- 心電図測定結果を利用して、不整脈を診断・予測するソフトウェア
- 組織検査、電子カルテ（EMR）などの医療情報に基づいて、特定の癌の発症確率を計算するためのソフトウェア
- 皮膚病変の映像を分析して皮膚癌の有無を診断するためのソフトウェア
- 血糖データ、食品摂取量、インスリン注射などの情報を分析して低血糖症を事前に予測するソフトウェア
- 緊急治療室で測定・統合した生体信号を分析して、呼吸困難などの緊急事態を予測したり、アラームなどの警告を与えたりするソフトウェア
- CT 画像解析により、異常部位を検出・表示するスクリーニング用ソフトウェア
- 医療画像を分析して血流速度、血管径などの血管特定部位の定量的数値を提供するソフトウェア
- 医療データに基づいて、放射線治療計画を策定するためのソフトウェア

2) 医療機器に該当しない医療用ソフトウェア

- a) 医療機関の行政事務（病室・在庫管理、電子手続きなど）をサポートするソフトウェア

<非医療機器の例>

- 保険金請求データの収集・処理を行うソフトウェア
- 医師の診療スケジュール、病室、患者の投薬時間などを管理するためのソフトウェア
- 患者の医療費請求や診療施設の電子手続きなどをサポートするソフトウェア

b) 運動・レジャーや日常的な健康管理を目的とするソフトウェア

*細部基準は、「医療機器とパーソナルヘルスケア（ウェルネス）製品の判断基準」に従う

<非医療機器の例>

- 健康的な食事、運動、減量、健康的なライフスタイルを奨励・促進するソフトウェア
- 健康データを収集・分析して、高血圧、慢性疾患患者の体重管理、栄養情報を提供するソフトウェア
- 健康データを収集・分析して、有酸素運動プログラムを提供するソフトウェア

c) 教育・研究目的のソフトウェア

<非医療機器の例>

- 大学、研究所などで研究、教育する目的でのみ製造・使用するソフトウェア
- 医療関係者の教育・訓練のための解剖学図や医療画像などを提供するソフトウェア

d) 疾患の治療・診断などとの関係のない医療記録管理を目的とするソフトウェア

<非医療機器の例>

- 電子カルテ（EMR）を保存・管理するためのソフトウェア
- 医薬品処方伝達システム（OCS）
- 臨床研究のための患者の診療、検査、IRB 審議などの記録をサポート・管理するためのソフトウェア

e) 医療関係者に、患者の健康情報や診療情報を整理および追跡するためのツールを提供したり、医療情報に簡単にアクセスできるよう支援したりするソフトウェア

<非医療機器の例>

- 従来の処方された薬や治療法などの情報を交換、変更することなしに、処方・診療に関する文献情報等を検索し、整理するためのツールを医療従事者に提供するソフトウェア
 - 電子カルテシステムから他の患者の処方（治療法？）や、処方された薬のリストを検索して提示するソフトウェア
 - 電子カルテシステムで、患者の医療映像と最も類似した他の患者の医療画像を検索して提示するソフトウェア
 - 患者の健康状態や治療に関する医療情報への簡便なアクセスを支援するソフトウェア
 - 標準治療法、臨床文献などを検索し、その内容を要約して提示するソフトウェア
 - 薬の副作用予防のための薬物間相互作用およびアレルギー反応を検出するソフトウェア
- ※ただし、学習データを取得（優先度、要約の提供などを含む）することを越えて、多様な学習データを再解析し、特定の患者に適した新しい診断や治療方法を提示するものは、医療機器に分類される

非医療機器に該当する場合の管理方案

食品医薬品安全処は、医療機器規制の適用の必要性を判断するために、国内外の資料を調査・分析したり、定期的の実態調査を実施したりして、開発中の製品についての要素（該当要件？）が確認された場合には、医療機器に分類して管理することができる。この場合、ガイドラインや告示の改正手続きにおいて利害関係人の意見を収集するなど、十分な議論の過程を経る予定である。

2. 医療機器品目分類

医療機器の等級分類は、「医療機器法」第3条（等級分類と指定）に応じて、医療機器の使用目的と使用時の人体に及ぼす潜在的な危険性の違いで分類され、品目分類は使用目的、特性などを総合的に考慮して、「医療機器の項目および品目別評価に関する規定」に基づいて分類される。

医療画像を利用するビッグデータと人工知能技術が適用された医療機器は表1の項目に基づいて分類することができる。

表1. 医療画像を利用したビッグデータと人工知能技術が適用された医療機器の品目例

番号	品目（評価）	定義
1	医療画像解析装置 ソフトウェア（2）	医療映像を取得して模擬治療、模擬施術、診断に使用可能となるように分析する装置に使うソフトウェア
2	放射線治療計画 ソフトウェア（2）	獲得された医療用画像を利用して、放射線模擬治療と模擬手術に使用されるソフトウェア
3	医療画像検出補助 ソフトウェア（2）	医療映像内で正常と違った異常部位を検出した後、輪郭線、色または引き出し線等で表示して医療関係者の診断決定を補助するために使用されるソフトウェア
4	医療用画像診断支援 ソフトウェア（3）	医療画像を使用して、病気の有無、病気の重症度や病気の状態などの可能性程度を自動的に表示して医療関係者の診断決定を補助するために使用されるソフトウェア

医療画像以外の医療情報（生体信号、血液や体液などの人体に由来した試料を使用した検査結果）を分析するビッグデータと人工知能技術が適用された医療機器は、分析しようとする情報に基づいて、内蔵の機能検査用機器[A26000]と「血液検査容器[A22000]」内の小分類品目名に従う。

例えば、血圧や酸素飽和度などの患者の生体情報を分析して、測定情報が正常範囲を外れた場合、アラームを提供するビッグデータと人工知能技術が適用された医療機器は、内蔵機能検査用機器[A26000]内に品目分類された患者監視装置【A26090.01】で申請することができる。

ただし、申請する製品の使用目的などを比較して、該当する項目がない場合は、中分類（例えば、内蔵機能検査用機器[A26000]、血液検査用機器[A22000]など）を指定して申請することができ、必要に応じて別の項目を新設することができる。しかし、医療用画像以外の様々な医療情報を利用して、病気を診断・予測する医療機器は、現在の「医療機器の項目および品目別評価に関

する規定」に適切な品目分類がないことから、今後表 2 に示した項目に分類して管理する予定である。

表 2. 医療画像以外の医療情報を利用したビッグデータと人工知能技術が適用された医療機器の品目（案）

番号	品目（評価）	定義
1	生体信号検出補助ソフトウェア（2）	患者の各種生体情報（医療画像を除く）を使用して、通常と異なる異常信号を検出した後、アラームを提供したり、色や引出線などで表示したりして医療関係者の診断決定を補助するために使用されるソフトウェア
2	生体信号の診断補助ソフトウェア（3）	患者の各種生体情報（医療画像を除く）を使用して、病気の有無、病気の重症度や病気の状態などを診断または予測したり、可能性の程度を自動的に表示したりして医療関係者の診断決定を補助するために使用されるソフトウェア
3	人体由来検体検出補助ソフトウェア（2）	人体由来検体を分析して、通常とは異なる特異的な結果を提供することで医療関係者の診断決定を補助するために使用されるソフトウェア
4	人体由来検体の診断補助ソフトウェア（3）	人体由来検体を分析して、病気の有無、病気の重症度や病気の状態などを診断または予測したりする可能性の程度を自動的に表示することで医療関係者の診断決定を補助するために使用されるソフトウェア

許可・審査方法

ビッグデータと人工知能技術が適用された医療機器の承認審査時には、以下のような事項を考慮しなければならない。なお、一般的に医療機器ソフトウェアの許可・審査時に考慮すべき事項は「医療機器ソフトウェアの許可・審査ガイドライン（2015）」を参照すること。

1. 許可審査申請書の「性能」に記載する内容

許可審査申請書に記載する「性能」は、主な機能、クラウドサーバーのオペレーティング環境、クラウドサービスの形態、セキュリティ規格などの技術仕様を記載する。

主な機能は、診断・予測のためにユーザーが入力した情報と出力情報、学習データの更新周期、診断結果精度の事項を記載し、クラウドサーバーが使用されている場合には、クラウドサーバーの動作環境と形態を記載する。また、セキュリティの点では、データの記号化、匿名化政策などのセキュリティ規格を記載する。

<作成例>

1) 主な機能

a) 医療画像の入力と診断結果の出力

- 入力情報：肺 CT 画像
- 出力情報：肺がんの有無、がんの重症度、診断の正確さ、肺がんのいち（CT 画像に表示）

b) 学習データ

- 学習データ更新周期：製造元によるアップデート 1 年

c) 性能

- 感度：90%以上、特異度：90%以上

2) クラウドサーバーのオペレーティング環境

※許可申請製品が製造業者ではなく、第3者（クラウドサービスプロバイダ）が提供する IaaS や PaaS ベースに開発・実装された場合、第3者の相互、クラウドサービスの形態（IaaS または PaaS）、クラウド構成形態（プライベート、共用、ハイブリッド）を記載

※許可申請製品が SaaS 形態で人工知能のサービスを提供する場合、クラウド構成形態（プライベート、公共、ハイブリッド）を記載

3) セキュリティ規格

※データが暗号化・復号化、匿名化政策などに適用されるセキュリティ規格に記載

<クラウドサービスの形態>

1) IaaS (Infrastructure as a Service)

- クラウドを運営するための物理的な資源、すなわち、サーバ、IP アドレス、Network、Storage、電力などのインフラを提供するサービス形態

2) PaaS (Platform as a Service)

- クラウドを利用したアプリケーションを開発することができる安定した環境 (Platform) を提供するサービス形態

3) SaaS (Software as a Service)

- クラウド環境で動作するアプリケーションを提供するサービス形態

2. 性能と臨床的検証項目

ビッグデータと人工知能技術が適用された医療機器の性能および臨床的検証は、感度 (Sensitivity)、特異度 (Specificity)、陽性予測値 (Positive Predictive Value)、陰性予測値 (Negative Predictive Value)、ROC (Receiver Operating Characteristic) Curve、AUC (Area Under the Curve) などの項目を利用して、製品の診断精度を確認することができる。

性能と臨床的検証に使用されるデータは、客観性を維持するために開発データとの相互独立性などを考慮しなければならない。

<性能と臨床的検証項目の例>

1) 感度 (Sensitivity)

- 実際に特定の病気に罹患した人の中から、その病気があると分類して出力する確率

2) 特異度 (Specificity)

- 実際に特定の病気がない人の中で、その病気がないと分類して出力する確率

3) 陽性予測値 (Positive Predictive Value)

- 特定の病気に罹患していると分類された人々の中で、実際にその病気に罹患している人が占める分率

4) 陰性予測値 (Negative Predictive Value)

- 特定の病気に罹患していないと分類された人々の中で、実際にその病気に罹患していない人が占める分率

5) ROC (Receiver Operating Characteristic) Curve

- 診断テストの結果に基づいて、感度と特異度を利用して描かれたグラフであり、陽性と陰性を区別する診断性能を評価することができる

6) AUC (Area Under the Curve)

- ROC Curve 下の面積。診断精度としては 0.5~1.0 の間の値を取り、1 に近いほど理想的な性能と判断することができる。

また、クラウドコンピューティング技術などを適用して、ネットワークを介して医療情報の保存・転送が行われる場合には、医療情報のセキュリティとクラウドの転送に応じて、医療情報の変化または損失が発生するかどうかを検討することが必要となる。

ネットワークの使用に関するセキュリティ要件には、サーバへのアクセス制御、ユーザー認証、医療情報の転送、保存時の暗号化技術の使用、非識別化などがあり、「ヘルスケア医療機器システムの許可の審査ガイドライン（2016）」の「3. システムデータのセキュリティの要件」を参照して、製品の特性に合わせて設定することが求められる。

3. 臨床的有効性の確認

ビッグデータと人工知能技術が適用された医療機器に適用可能な臨床的有効性の確認方法としては、前向き研究、後ろ向き研究、さらにはそれら両者を並行して行う前向き・後ろ向き研究があり、製品特性に応じて適切な臨床試験の方法を設計することができる。

<臨床的有効性の確認方法>

1) 前向き研究（Prospective Study）

- 研究しようとする要因（リスク）を事前に設定した後、一定期間の変化を追跡することで、リスクが引き起こす変化を観察する研究

2) 後ろ向き研究（Retrospective Study）

- 研究対象者と直接接触することなく行われる研究で、被験者のカルテを調査して、特定のデータを収集、統計処理し、結果を算出する研究

※後ろ向き研究では、被験者のカルテ、医療映像、生体信号、病理検査などのデータに、臨床試験の結果などを使用できること

臨床的有効性確認資料は、「医療機器の許可・届出・審査等に関する規定」第 29 条（添付資料の要件）の「12. 臨床試験に関する資料」ロに基づいて、臨床試験の方法、臨床結果、臨床評価の基準が考慮されるべきである。

後ろ向き研究時の臨床試験に使用する被験者のデータは、製品開発プロセス中に使用されたデータとは独立しなければならない。また、被験者の選定基準・除外基準と目標にした被験者数は、偏向（Bias）が発生しないように、被験者のデータの収集に関連する内容（収集方法、収集場所、収集形式、収集事項など）、被験者データの測定時期と対象者数、被験者データの選定基準と除外基準などを考慮しなければならない。

患者の医療映像、生体信号、体外診断テストの結果など、診断・治療に必要な臨床情報を分析して病気を診断・予測する医療機器は、韓国人のデータを使用して電子カルテに記載されている確定された診断結果との比較を通じて臨床的有効性を確認することができる。民族的要因の違いがないことを証明する場合にも、韓国人のデータを含むことが必要である。

(4. 提出の範囲、5. 変更許可・認証対象は省略)

6. バージョン管理

ビッグデータと人工知能技術が適用された医療機器のバージョン管理規則は、製造業者によって製品の構造、設計などを管理する部分と、学習データの追加などを管理する部分に区分することができる。製品の設計変更は、一般的な医療機器のソフトウェアのバージョンと同じように管理することができるが、学習データの追加などの変更によりバージョンが変更される場合は、製造業者および医療機関のデータ管理ポリシーに基づいて、適切なバージョン管理方法を適用して許可審査申請書に記載しなければならない。表3のようにビッグデータと人工知能技術が適用された医療機器のバージョン管理方法は、主な機能の変更、単純な変更、軽微な変更と学習データの変更の管理に分けることができる。学習データの変更によって性能（精度）が変更されている場合は、主な機能変更管理が、承認時に記載した性能（精度）の範囲内で変更される場合は、学習データの変更のバージョン管理規則を設定して、製造業者が自律的に管理するようにして、許可審査申請書には、バージョンの数を具体的に表示せずに「X」と表記することができる。

表3. バージョン管理方法

区分	内容（例）	管理方法
主な機能の変更	作用原理の変更、使用目的の変更、性能変化(学習データの変更による性能変化は許可時に記載された性能(精度)の範囲外になる場合のみ)	変更許可申請
単純変更	ユーザー画面（Graphic User Interface: GUI）の設計変更	変更許可申請
軽微な変更	バグ修正、ユーザー画面（GUI）の色やメニュー位置の変更など	届出、または報告
学習データの変更	許可時に記載された性能（精度）範囲内での学習データの変更	製造業者自主管理

7. 学習データの管理

ビッグデータと人工知能技術が適用された医療機器は、病気の診断・予測のための特徴を抽出するために、電子カルテ（Electronic Medical Record、EMR）、医療文献（臨床論文は、臨床学会のガイドライン等）、医療画像など、さまざまな学習データが必要であり、これらの学習データは、製品の性能と有効性に影響を与える。したがって、製造業者は、学習データの有効性が一貫して維持されるように、データ管理ポリシーを策定しなければならず、学習データの更新時期は製造業者および医療機関との協議により設定することができる。データ管理ポリシーでは、製造業者および医療機関の作業により、データ確保計画と確保されたデータの効果的な運用管理システム計画を定義しなければならない。また、データの管理においては、原則、管理組織、品質管理プ

プロセスのシステムおよび計画を策定しなければならない。特に、データ管理組織は、学習データの品質管理項目と範囲、基準を設定するとともに、定期的または非定期的に追加された学習データによって変化した製品アルゴリズムの品質評価を実施しなければならない。

8. クラウドコンピューティング技術の適用による許可範囲

ビッグデータと人工知能技術が適用された医療機器に使用することができるクラウド構成形態は、医療機関独自のデータセンターを内部に設置して、医療機関でのみ利用できるプライベートクラウド (Private cloud) と外部サービスプロバイダから提供されるクラウドサービスを利用するパブリッククラウド (Public cloud)、プライベートクラウドとパブリッククラウドを組み合わせるハイブリッド・クラウド (Hybrid cloud) に分類することができる。クラウドコンピューティング技術が適用される場合、クラウドサーバーは医療機器の管理対象ではないが、クラウドサービス形態、サーバの運用環境は許可審査申請書に記載することが求められる。医療機器承認・認証後、クラウドサービスの形態 (例えば、IaaS、PaaS、SaaS など) の変更や医療機器の性能に影響を与えるクラウドサーバーの運用環境に変更が発生した場合には、変更許可を受けなければならない。しかし、医療機器の性能に影響を及ぼさない場合には、変更許可・認証なしで製造業者および医療機関で自主的に管理することで問題ない。また、医療機器の製造業者 (輸入業者) は、医療機器ソフトウェアのセキュリティ (アクセス制御、個人情報以外の識別化、データアーム・復号など) に必要な技術的措置を実行し、文書化しなければならない。クラウドサーバーのセキュリティに必要な管理的・物理的・技術的事項は、医療法と個人情報保護法に基づいて管理する。

中国におけるクラス分類表英訳版
(Medical Software 部分を抜粋)

21. Medical Software

1. Scope

includes software that does not depend on the medical sub-directory medical equipment.

2. Frame structure

this sub-directory it does not include the software components, it shows the only independent software. Medical diagnosis and treatment is 3 exemplary software-based, "video", can be divided "data", to, depending on the purpose of processing, they are classified into classes depending on the intended use, "image data". This sub-directory, the product includes examples of 51 people, six categories, will be divided into two product categories 13.

Correspondence between the next 2002 edition of the product categories directory as product categories and sub-directories:

Category correspondence table of the 2002 edition

Category	2002 version of the Product category	Remarks
21-01 treatment planning software	6870-1 program of software function	/
21 -02 image processing software,	6870-2 image processing software diagnostic	/
	6870-4 transfer of the image file, processing system software	
21-03 data processing software	6870-3 diagnostic data processing software	/
21-04 Decision support software	/	Added. Such software provides a determination of the proposal of diagnosis and treatment.
21-05 for in vitro diagnostic software class	/	Sets, respectively. Diagnostic software, some special software can not be classified according to the image processing vitro has both may include the image processing software and data processing software including data processing.
21-06 Other	/	New. Rehabilitation software special properties, can not be categorized in accordance with the image processing software and data processing software.

3. Third, it indicates that the other

- using the "image" or "image", the current existing registered product names that are named. Take the "image processing" inclusive range, the product development in the future to account the "image processing", unified in the "image" of all of the sub-directory, image to be larger than the "image" statements, both of image Although it contains, also, the "image." contains video content
- habits and Chinese translation, and some as the product name of the issues have been

included, "*** system (system) , "to determine whether or not the very ambiguous independent software products. Because of the special nature of the "21 medical software", industry-specific: If there is a term (such as TPS, the consensus in the product name of PACS and other industry), reservation "system software ***" naming, or remove the product name "system" of wordsbasis:.

3. The degree of risk of the diagnostic function software is determined according to the degree of risk, maturity, and publicity of the algorithm used, and not just the object of treatment (eg, images of diseases such as cancer and malignant tumors).
If diagnostic software provides diagnostic advice through algorithms, it has only auxiliary diagnostic function and does not submit diagnostic conclusions directly. Related products in this subcategory are managed according to Class II medical devices. The risk level is relatively high if the diagnostic software automatically identifies the lesion with an algorithm (such as CAD without bone mineral density) and provides a clear diagnostic tip. Related products in this subcategory are managed according to Class III medical devices.
4. Have closely related navigation software and navigation devices no navigation device, navigation software, you will not be able to achieve the intended purpose. currently, registered product "surgical navigation software," of the hardware it contains the most. In principle, in order to determine the surgical navigation products, including hardware specifications of the 01 active surgical instruments, this sub-directory revision process. Any hardware related to "surgical planning software", as a medical software, can not be incorporated into this sub-directory. On the other hand, in order to avoid confusion, to delete the word "navigation" is, for example, to adjust the name.
5. If the medical information management software medical information management software hospital management tools of the attributes defined principles only of, content management, rather than content, in accordance with medical equipment management, is a non-medical diagnostic patient information and / or treatment. If the medical information management software, patient diagnosis, comprises a therapeutic, imaging data is different software object (image data), is processed in accordance with the product specification to "image processing software 21-2" or software, "21- 3 the data processing software.
6. "Different telemedicine consultation system software used to implement the transmission of health information and consultation platform functions between medical institutions. This subdirectory in "21-02-01 Picture Archiving and Communication System Software" or "21-03-02 medical image processing software", including the transmission specifications of the telemedicine software, video data, or contains the medical image data inchfit in the absence controlled medical devices that have not
7. risk to different mobile medical software operating platform, other image processing software, data processing and significant other functions running on a common platform been.This sub-directory, different software processing object (image data), based on the product specifications in the "image processing software 21-2" or software does not reflect "portable medical software," "21-3 data processing software."

8. "21 since -02 image processing software "image processing software name rule, therefore, secondary product category names are attached to the device that are not regulated as in." I am not used CT and stand-alone software after the "image processing software because other software, there is no risk of confusion.
9. Medical image processing software, for single-mode or multi-mode from the medical image processing. If you do not have a diagnosis function of what the image processing software, to unify the X-ray image processing system, "6870 from 2 diagnostic image processing software" down directory to the first category, nuclear medicine imaging, medical magnetic resonance imaging systems and other 2002 the second category of management.
10. Category in the 2002 version of "Human Anatomy 6870 from 5 measurement software." Now a valid certificate of registration information of such products, and because its intended use does not comply with the completely definition of medical devices, this products are not included in this sub-directory.
11. IVD software is, if it contains computer-aided diagnosis function, it should be classified as a decision support software.

21 Medical software

#	Product Category	Second classification	Descriptions	Intended Use	Category name	Class
01	Treatment planning Software	Radiation therapy	The typically installed by (from the network or the installer downloads) CD-ROM software composition. Normal circumstances with one or more of the specific algorithms of human organs (necessarily) has been absorbed dose distribution estimation under.	For radiation treatment planning of patients.	Radiation treatment planning software, gamma ray stereotactic radiotherapy planning software, radioactive seeds implanted species on the treatment planning software	III
		Radiotherapy support software	the CD-ROM, often is the component (downloaded from the network or installer) software. Provided, or define the display setting of the data processing apparatus, manually directly from the data, enter or import data from other devices, and record the entire treatment period.	Before starting the radiation therapy machine, compares the current parameter a preset parameter for each processing stage, before starting the radiation therapy planning, and record the data of the actual processing steps.	Radiotherapy record and verify system software	III
		Surgical planning software	usually consists of the installation (or Please download the program from the network) in the software CD. Image information obtained before radiation therapy, and by using the analysis result, and determines a target position or target coordinates.	To support the completion of radiation therapy.	Contours radiation therapy software, radiation therapy simulator software	III
			is, in general, the components of the installation (or downloaded from the network installer) in the software CD. Image acquisition information and the results of the analysis, the use of operational plan or program.	For the operation plan of the previous surgery. ENT except dentistry, the class.	Stereotactic surgical planning software, software simulation surgery, surgical planning software is	III
			typically installed (or a program downloaded from a network) configured by software. Use and results of the	Oral surgery, used to develop a plan in front of the ENT surgery.	Digital implanting design software, was software design denture	II

				analysis of the obtained image information, dental, development of ENT surgical planning and programs.					
		Picture archiving communication system software		typically(download or from a network installer) CD-ROM software is installed by the composition. Medical imaging device, in order to receive the medical imaging departments, the image and / or video signals acquired stored on the hard drive of the computer, output, and / or hospital, a display, a memory, and the output of another process transmission between various sections between.	Medical imaging, such as, transmission, display, storage, to receive the processing output, and is used for clinical diagnosis and treatment.	Picture archiving and communications system software, medical imaging and communication systems management software, Picture Archiving and Communication System Software	II		
02	Image processing software	medical image processing software		typically installed components (or downloaded from a network installer) software CD. Image processing method, we used the three-dimensional reconstruction of medical images as such registration process.	For single-mode or multi-mode from the medical image processing.	Ultrasonic image management software, the endoscope graphics workstation software, digital ultrasound workstation software, image processing software magnetic resonance, nuclear medicine workstation software, CT image processing software, X-rays angiographic image processing software, digitizing X-ray image processing software	II		
03	software processing of the data	Monitoring software physiological signal processing software		Typically, components that are installed (or downloaded from a network installer) software CD. Data communication, to get the data from the monitoring device by a centralized real-time display and alarm. for CD-ROM typically composition (downloaded from the network or installer) software is. Analysis and / or transport of collected EEG, ECG, EMG and other physiological signals.	Monitoring equipment, to get the data from a centralized real-time display and alarm. EEG, ECG, for analysis and / or transport of EMG and other physiological signals.	Central monitoring station software, software management of the central monitoring, the central monitoring center software information and pregnancies software monitoring, Holter analysis software, workstation software ECG, ECG data management software	II		
04	Decision support	01 calculation		usually, the composition via the (from a network or installer download) of the CD-ROM software.	For injecting the medicine clinical advice.	Insulin calculation software	III		

			Pharmacokinetic and / or drug model to calculate signs of physiological parameters and drug injection protocol patients, provide recommendations for clinical drug injection.			
	software, drug	Diagnosis / Analysis software is	Typically installed by the software in the CD-ROM (download from or network installer) composition. Image processing and / or using data processing techniques, the lesion is given on the basis of the treatment with the clinical diagnosis by a computer software for identification, such as the nature of the automatic diseases, and / or recommendations. CD-ROM, from the composition (downloaded from the network or installer) normal software. Video and data analysis of clinical reference values, analysis.	Lesions, automatically as based on the proposal of the nature of the clinical diagnosis and treatment and / or disease, was identified by a particular computer software.	Computer-aided diagnosis software Breast X-ray image, computer-aided diagnosis software colon, lung, computer-aided diagnosis software breast ultrasound diagnostic software	III
			The software in the CD-ROM (download from or network installer)composition, it is usually. With treatment of the theory of TCM, using statistical techniques, provides for achieving the analysis and diagnosis of various symptoms, and / or treatment recommendations.	It provides a recommended treatment for implementing the analysis and / or various incidents of diagnosis.	BMD software-aided detection	II
			typically (setup downloaded from or network) software CD-ROM in the implementation component. Device, the transmission, synthesizing observation, analysis, processing, and reporting from having various access obtained microscopic images from the function.	Microscope imaging equipment obtained as, analysis, will be used for a variety of diagnostic and archive.	Visceral Qjhuang auxiliary processing software,Chinese medical	II
05	Vitro Diagnostics software,software	medical microscopic image analysis software 01	installed by the composition (program downloaded from or network installation) II 02,CD-ROM software typically software ing. Clinical, biochemical analysis of	Disease, screening for the evaluation.	Medical pathological image acquisition software, medical microscopic image analysis software, software, chromosome analysis, the urine sediment analysis software	II
		Screening			Risk analysis software prenatal screening, prenatal screening for Down's syndrome analysis software, neural tube defects in prenatal	II

06	Other	Rehabilitation training	immunological and other measurement data, calculate, diagnosing a disease whereby the evaluation through.		screening analysis software, spreadsheet software trisomy 21, trisomy 18 risk calculation software of the risk calculation software of neural tube defects, blood glucose data analysis software	
			software, usually, the components that are installed (or downloaded from the network installer) in the software CD. Grading, and check the training module of adjuvant therapy, the software system is may be comprised of a single module.	For adjuvant therapy and rehabilitation.	Training software of visual function with low vision, visual function training software children	II

平成30年度次世代医療機器・再生医療等製品評価指標作成事業 **人工知能分野審査WG活動報告**

委員構成(敬称略)

座長 橋爪 誠
(九州大学/北九州中央病院)

- 委員(五十音順)**
 佐々木 毅 (東京大学)
 清水 昭伸 (東京農工大学大学院)
 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 昭洋 (国立国際医療研究センター一府台病院)

松尾 豊 (東京大学)
 横井 英人 (香川大学)

事務局(国立衛研)
 龍島由二, 中岡竜介, 加藤玲子

医学系委員:6名
工学系委員:3名
うち学会推薦委員:5名
(関連学会と連携)

- 日本病理学会推薦
 日本消化器内視鏡学会推薦
 日本医用画像工学会推薦
 医用画像情報学会推薦
 日本医学放射線学会推薦

(独)医薬品医療機器総合機構
 医療機器審査第一部、第二部、
 調査・基準部

当該審査WGの目標

コンピュータ技術の著しい発展により、学習データのさらなる追加(市販後学習)等に伴い市販後の性能変化も可能となった**人工知能技術を利用した医療機器**の開発が活発化している。コンピュータ診断支援を行う医療機器に関しては、既に本事業による次世代医療機器評価指標通知が発出されているが、上述した人工知能技術の適用は当該通知の対象とされていなかった。特に**深層学習**を用いた人工知能技術では、**最終結果を導き出すアルゴリズムがブラックボックス化**している点が特徴であると同時に、事後学習等による性能変化の内容も基本的にはアウトプットでしか評価できないこと等から、その評価や運用において従来にはない**新たな課題**が生じることが懸念されている。また、当該診断支援装置を用いた場合、特に市販後の性能変化に対する評価方法等に対して、**従来とは異なる新しい視点**も考慮する必要が生じた。

このような背景に鑑み、本WGを立ち上げ、まず、**人工知能技術を利用した医用画像診断支援システム**の有効性と安全性を評価する際の問題点や留意すべき点について、現時点での考え方をとりまとめ、評価指標案を作成する。

人工知能応用例(内視鏡画像診断支援)

平成30年12月6日承認: EndoBRAIN(サイバネット株式会社)

内視鏡画像 人工知能

当該プログラム(EndoBRAIN)が、医師の読影の補助を行う

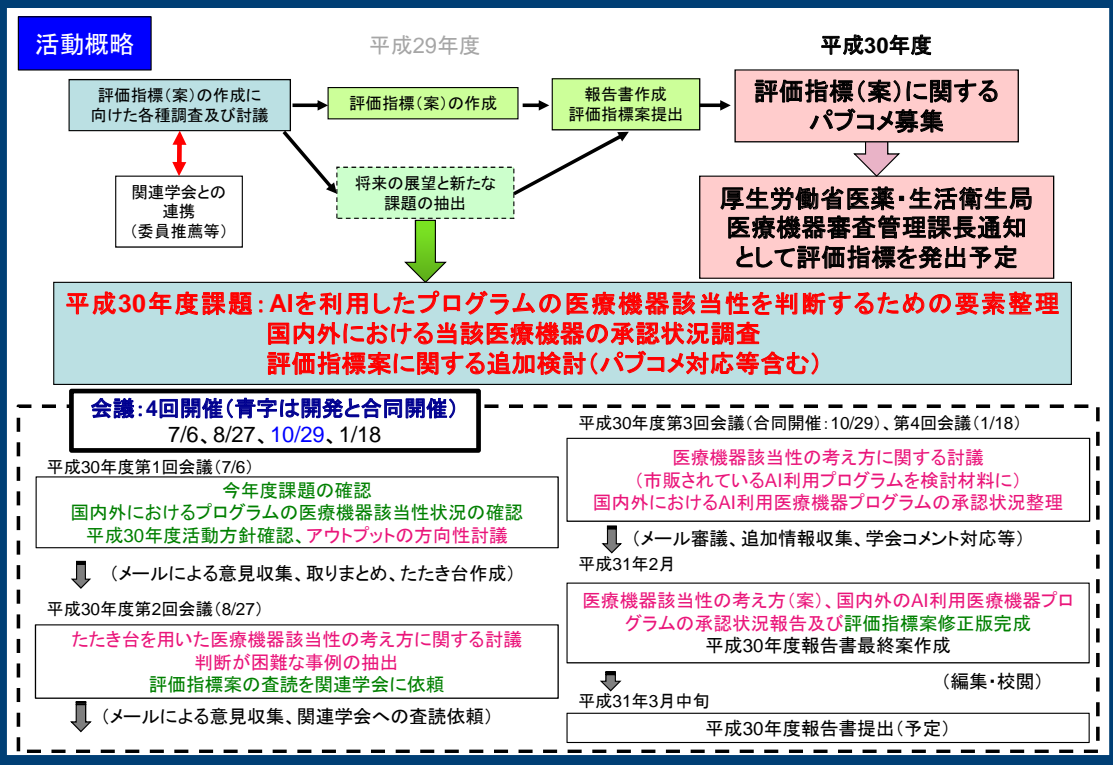
(表示例)

腫瘍の確率
 Neoplastic: 89 %
 Non-neoplastic: 11 %

- サポートベクターマシンを使用し、約6万枚の画像を学習
- 市販後学習機能は有さない

昭和大大学プレスリリースより引用

平成30年度活動内容



平成30年度の課題と対応方針

1. AIを利用したプログラムの医療機器該当性判断に資する要素の整理

医薬品医療機器法施行令別表において医療機器として規制されるものが定められているが、一般医療機器相当のプログラムは規制の対象から外されている

人工知能技術を用いた場合、従来のアルゴリズムでは困難だった性能を有するプログラムが出現する可能性があり、新たな留意点を検討する必要が生じている

海外の動向も参考に現在の考え方を踏襲しつつ、考慮すべき要素を抽出・整理する

2. 国内外における当該医療機器プログラムの承認状況調査

医療機器該当性判断要素の整理にも資する情報であるため、委員から情報を収集・整理し現状把握も兼ねた資料を作成する

3. 評価指標案に関する追加検討(パブコメ対応等含む)

- AIを利用した医療機器プログラムの開発研究を実施中の学会からのコメントを考慮する必要がある→日本眼科学会、日本皮膚科学会、日本超音波学会に査読を依頼
- パブリックコメント募集で寄せられたコメントも必要に応じて確認
- WG内でも再度査読を行い、修正の可否等を検討

課題に対してWGが今年度実施した対応

1. AIを利用したプログラムの医療機器該当性判断に資する要素の整理

- 平成26年11月14日付け薬食監麻発1114第5号通知に示されたプログラムの医療機器該当性に関する基本的な考え方は、AIを利用しても変わらないことを確認
- 医療行為への関与を判断する「目的性」、医療行為の関与程度を判断する「寄与度」、不具合等により生じる「リスク」を該当性判断要素として整理し、実例を含む資料案を作成
- 「寄与度」、「リスク」の双方が低い場合は「一般医療機器」となり非該当となることを明示

2. 国内外における当該医療機器プログラムの承認状況調査

委員から収集した情報を整理し、我が国を含め4ヶ国における当該プログラムの承認状況を取りまとめた→報告書に掲載

3. 評価指標案に関する追加検討(パブコメ対応等含む)

- 日本眼科学会、日本皮膚科学会、日本超音波学会から寄せられたコメントを確認→現状の評価指標案で対応可能と判断し、その根拠を添えて各学会に返答
- 11月に実施されたパブリックコメント募集で寄せられたコメントを確認→一部コメントを受け入れ評価指標案を修正することを確認
- 委員より、「事後学習」が人工知能研究者の理解と異なる趣旨で使用されており誤解が生じる可能性があるとの指摘→当該用語を「市販後学習」に変更して評価指標案を修正

今後

- 今年度で活動は終了するが、技術革新に伴う対応が必要となった場合には活動再開も検討
- 医療機器のAI利用により生じる業事的な課題全般に関しては別事業で対応予定

その他関連 URL リスト

- 「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」 平成 26 年 11 月 14 日 薬食監麻発 1114 第 5 号
<https://www.pmda.go.jp/files/000159993.pdf> (2019.02.27)
- 「「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」 平成 30 年 12 月 28 日 薬生監麻発 1228 第 2 号
<https://www.pmda.go.jp/files/000227451.pdf> (2019.02.27)
- 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
<https://www.pmda.go.jp/files/000227450.pdf> (2019.02.27)
- FDA Guidance “Clinical and Patient Decision Support Software”
<https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm587819.pdf#search=%27clinical+and+patient+decision+support+software%27> (2019.02.27)
- Guidance: Medical device stand-alone software including apps (including IVDMDs) (provided by Medicines and Healthcare products regulatory agency, UK)
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/717865/Software_flow_chart_Ed_1-05.pdf (2019.02.27)
- Medical devices: Guidance document “Guidelines on the qualification and classification of stand alone software used in healthcare within the regulatory framework of medical devices” (published by European Commission, July 2016)
<https://ec.europa.eu/docsroom/documents/17921/attachments/1/translations/en/renditions/native>
(2019.02.27)

Ⅶ. 平成 30 年度 WG 会議議事概要

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 WG
平成 30 年度 第 1 回会議 議事概要

1. 開催日時：平成 30 年 7 月 6 日（金）15:00 – 17:59
2. 開催場所：TKP 品川カンファレンスセンター ANNEX カンファレンスルーム 4
（東京都港区高輪 3-13-1 TAKANAWA COURT 3F）
3. 出席者（委員：座長以下五十音順・敬称略）
座長：橋爪 誠（北九州中央病院）
委員：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：
Skype で出席）
厚生労働省（厚労省）：田中 大祐、青柳 ゆみ子、坂西 義史、宮下 正也
医薬品医療機器総合機構（PMDA）：岡崎 譲、加藤 健太郎、桐山 瑤子、古森 亜矢、
大澤 智子、水上 良明、今川 邦樹
オブザーバー：鎮西 清行（産業技術総合研究所（産総研））、坂無 英徳（産総研）、
浅沼 直樹（日本医療研究開発機構）、近藤 昌夫（大阪大学）
事務局（国立医薬品食品衛生研究所）：葩島 由二、中岡 竜介、加藤 玲子
4. 配布資料
資料 1：第 1 回審査 WG 会議議事次第（案）
資料 2：委員名簿
資料 3：今年度活動計画案
資料 4：米国における承認状況
資料 5：EU における SaMD 規制状況
資料 6：英国における SaMD 規制状況
資料 7：韓国における規制状況（抜粋版：機械和訳）
資料 8：FDA ガイダンス案（和訳付）
資料 9：中国における医療機器ソフトウェア分類状況
資料 10：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）
資料 11：「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」（平成 26 年 11 月 14 日付け薬食監麻発 1114 第 5 号厚生労働省医薬食品局監視指導・麻薬対策課長通知）
5. 議事内容
 - 5-1. 開催挨拶及び資料確認
事務局より、開会挨拶、配布資料の確認及び新規参加者の紹介が行われた。
 - 5-2. 今年度の課題説明
事務局より資料 3 を用いて今年度の課題について説明された。主な内容は以下のとおりである。
 - (1) 評価指標案の追加修正の要否
人工知能分野は技術進歩が激しいことから、昨年度作成した評価指標案、特に別添に

取りまとめた施設毎の性能変化が可能な装置や教師なし学習を行う装置を対象に、追加修正の可否を含めてさらに議論する必要があると思われる。事務局が想定する検討課題案としては下記の事項が挙げられる。

1) 追加すべき項目の有無

<例 1> 新たな原理に基づいた画像診断支援システムに対応するための項目

- ・ 新規対象となる機器に特化した記載の追記は必要か？
- ・ 新たに着目されている人工知能技術や製品はないか？
- ・ 国内外における研究・開発動向調査が必要か？

<例 2> 機器の特性に応じた追加留意事項

- ・ 病理と内視鏡以外に記載が必要となる機器はないか？

<例 3> 臨床試験（治験）の必要性

- ・ 事例毎の判断になると思われるが、評価指標案に特記すべき要件はないか？

2) 別添の更新の可否

「使用者による施設毎の事後学習が可能な製品」及び「教師なしデータによる自己学習が可能な製品」については、医療機器としての品質確保が難しい等、現時点では課題が多いが、そのニーズや将来の発展を考慮し、以下の点を考慮して、追加ブラッシュアップの可否を検討する必要があるか？

- ・ 別添に記載した原理のシステムの実現性
 - ・ 施設毎に異なる性能変化が可能なシステムの問題点
 - ・ 医療機器に該当するための要件整理の必要性
- 等

(2) 人工知能を利用した医療機器の国内外における開発・利用状況調査

本会議に先立ち、委員から直近で開催された学会や見本市で収集した国内外の人工知能を利用した医療機器に関する情報が報告された。また、開発 WG の協力を得て事務局が調査した結果、米国では人工知能を利用した診断支援装置等が少なくとも 5 件承認されていることが報告された。事務局としては、評価指標案の更新作業、並びに次に作成する評価指標案の対象機器を選定する上で、国内外における最新の開発・利用状況調査を継続する必要があると考えている。

(3) 人工知能を利用したヘルスケア製品の医療機器該当性

医療機器プログラムの場合、該当性通知（資料 11）に従い、一般医療機器に相当する製品は規制対象外となっている。医療機器該当性は治療方針等の決定への寄与度や、不具合が生じた際のリスク等を勘案して判断されるが、今後開発が促進される人工知能等を利用した様々な製品については、その境界を判断するための考え方を整理する必要があるため、今年度は該当性通知の記載例を参考として、人工知能等を利用したヘルスケア製品の医療機器該当性判断に係る基礎資料を作成する。

厚労省及び開発 WG の協力を得て、海外における医療機器プログラムの規制状況を資料 3 に取りまとめた。主な内容は以下のとおりである。

<米国>

FDA は、臨床判断支援プログラム（CDS）が医療機器除外となるための 2 要件を関連ガイダンスで明示している。一方、CAD のクラス分類に関しては、本来クラス III と分類される新規医療機器であっても一定の条件（アルゴリズム、学習プロトコール、データセット、性能評価の詳細等の明示、使用時における各種危険性等のラベリング等の記載）を満たせばクラス II に分類することが提案され、現在、パブリックコメントが

実施されている。また、クラスダウン提案を行う一方で、監視体制を強化する動きも見られる。その他、21世紀治療法第3060条において、医療機関の経営支援用ソフトや電子カルテ等に加え、医療機器非該当とする4要件及びその解釈を示したガイダンス案が公表され、パブリックコメントが実施されている。

<英国>

診断支援時に生データを確認できない製品は、単純な支援目的でも医療機器に該当するとの判断が示されている。

<韓国>

CDSの医療機器該当性に関する2要件が明らかとなっている。また、予め申請していた性能の範囲であれば、市販後の学習データ追加に伴う性能向上は届け出も不要としている。

<その他の国>

中国ではクラス分類の方針が示されているのみで、除外区分に関する記載はない。カナダもCDSの取り扱いに関する大まかな方針が示されているのみである。オーストラリア、ブラジル、メキシコ、ロシア、シンガポール、マレーシア、タイ、台湾、インド等は特段の規制を行っていないと考えられる。

事務局の説明後、厚労省よりプログラムの医療機器該当性の考え方について補足説明があった。

- ・ 該当性通知別添には、業界と連携して収集した医療機器プログラムの該当例及び非該当例が列記されている。人工知能を利用した医療機器プログラムの該当性についても、該当性通知に事例を追記することで対応可能と考えている。
- ・ 医療機器プログラムの場合、リスクの極めて低いクラスIに相当する製品は規制対象外とされることが、プログラムの医療機器該当性判断を困難にする一因となっている。本WGでは、特に、人工知能を利用した医療機器におけるクラスIとクラスIIとの違いを判断するための基準を検討して欲しい。
- ・ プログラムの医療機器該当性の考え方は、国際医療機器規制当局フォーラム(IMDRF)で定義された医療機器ソフトウェア(SaMD)のカテゴリー分類の考え方を考慮したものである。IMDRFでは、プログラムが使用される状況(医療現場や病態の現状)は、「危機的」、「深刻」及び「深刻でない」に区分されており、使用方法は、「治療又は診断」、「臨床管理の運用」及び「臨床管理に関する情報提供」に区分されている。
- ・ プログラムの医療機器該当性に関する事例として、一般的な書物等に記載されている情報であれば寄与度は低いですが、治療法の決定に大きく寄与するデータを医師に提示するような場合は、寄与度が高いと考えられる。一方、不具合が発生した場合に人の生命及び健康に影響を及ぼす総合的なリスクの蓋然性の程度は、通常の医療機器クラス分類の考え方を一部踏襲している。例えば、添付文書に記載されている計算式等を使用して医療従事者が医薬品の投与量を決定する行為を代替するプログラムは不具合によって誤った値を提示した場合にも、医療従事者が容易に気付くことができるため、クラスIに相当すると考えられる。本WGでは、人工知能を利用した製品の寄与度及びリスクを考える上で指標になり得る項目抽出及び上記と同様の該当性判断例の作成を検討して欲しい。

(4) 今年度の課題に関する対応方針

上述の報告に基づいて、今年度の課題に対する方針として以下の事項が確認された。

1) 評価指標案の追加修正

連携学会・企業から意見を聴取し、必要に応じて追加修正する。

2) 人工知能を利用した医療機器の国内外における開発・利用状況調査

審査 WG 委員に最新情報の収集を依頼し、事務局側で取りまとめる。必要であれば、タスクフォースを設立して必要な作業を行う。

3) 人工知能を利用したヘルスケア製品の医療機器該当性

海外規制等も参考にした上で、厚労省から提示されている「治療・診断への寄与度」と「リスクの蓋然性」に沿って人工知能を利用した製品に特有の項目を抽出し、我が国における当該製品の医療機器該当性に関する考え方を整理する。また、医療機器該当性を理解するための実例も複数作成する。

5-3. 総合討論

人工知能を利用した製品の開発・利用状況及び医療機器該当性等に関して討議した。主な内容は以下のとおりである。

- ・資料3に、「医師なしで網膜症を診断する」と記載されている IDX 社製 Idx-DR は「ヘルスケアプロバイダー」が関与することが要求されている。専門医の関与は必須ではないと考えられるが、検査技師等、「ヘルスケアプロバイダー」の範囲をどこまで広げていいのか？（委員）→ FDA のプレスリリースによると、医者が使用することが前提となっている。専門医による診断の要否を判断するスクリーニング目的で使用されるため、診断という表現が使用されていると理解している。従来の診断支援装置の場合、症状の最終診断は医師が行うのに対して、当該機器は重症度まで提示するため、「医師なし」で判断すると表現されていると考えられる。ただし、この機器は医師が使用することを前提としていると思われる。（厚労省）
- ・薬剤投薬量を提示するプログラムは医療機器非該当ということだが、TCI ポンプのように投薬積算量を継続的に観察する動的プログラムも非該当か？（委員）→ 動的、静的に関わらず、公知の情報に従って計算するのみのプログラムは非該当である。（厚労省）
- ・保険診療内か否かは全く関係ないということか？（委員）→ 考え方が逆である。医療機器は品質、有効性及び安全性を確認した上で承認され、保険点数はその承認品に付加される。医療機器の場合、手術等にあたり、医師が特定の機器や新材料を使用すると技術料が付加されることもある。臓器モデルを使用した診断が保険適用となることもあるが、使用するモデルには医療機器に該当しない製品を使用している場合もある。（厚労省）
- ・検診分野では様々な萌芽的技術が開発されてきており、人工知能を利用して胃のポリープを検出する製品を非医療機器として上市しようと考えている企業が存在している。その製品が日本に導入された時の規制側の対応が重要になると考えている。本 WG では、可能な限り製品化を認めるための議論を行うのか、ハイレベルな製品を規制するための議論が対象なのかを整理して欲しい。（委員）→ 薬機法の規制対象は製造販売業者である。医療機器該当性の判断は、医師の使用方法に拘わらず、製造販売業者が対象製品を販売する上で標榜する使用目的・方法等が重要になる。（厚労省）→ 例えば商業目的ではなく、学術グループ内で作成されたプログラム

- を複数メンバーが使用する場合は規制対象外か？（委員）→ 厳密に言えば、規制対象となり得る。金銭の授受に関わらず、あるプログラムを他人に譲渡・頒布する場合には薬機法の範囲に入り規制対象となる。ただし、臨床研究の枠組みの範囲内で使用する場合、複数の関係者が使用しても薬事規制は適用されない。（厚労省）
- 薬剤の選択まで行う人工知能プログラムの開発を視野に入れている企業が存在する。このような製品は治療支援まで行うことになるが、昨年度作成した評価指標案には治療支援に関わる項目が不足していると思われる。（委員）→ 評価指標の使用方法に絡む案件と思われる。審査側では、承認申請の内容に応じて、利用可能な事項については評価指標を参照するが、個別対応が必要な部分は、対処法を申請者側と一緒に考えていくことになる。（PMDA）
 - 診断結果と併せて緊急対応が必要な場合に警告するトリアージタイプの CAD も開発されている。現時点では想像できないプログラムが今後開発される可能性があることから、評価指標案はフレキシブルな対応を可能にしておいた方が良い。（委員）→ 従来の CADe や CADx のカテゴリーとは異なる、トリアージタイプを含めた AI-CAD と呼ばれる多様な製品が開発されてくるとと思われる。市販後、事後学習するものはクラス III に相当すると思われるが、今後開発される多様な製品にも対応するために、CAD を再定義する必要があると考える。（委員）
 - AI-CAD の場合、ある特定の病院で豊富な臨床情報を学習して作製された製品が別の病院で転移学習を行い独自に進化していくことが繰り返される可能性もあるので、対応を考えておくべきである。（委員）→ 臨床試験以外で他の医師に未承認の医療機器を譲渡することは、金銭の授受に関係なく薬機法 55 条 2 項違反となるので、現状でも規制可能である。（厚労省）
 - 該当性通知が発出されてから現時点までに様々な承認事例等があると思われるが、その経験を踏まえ、該当性の線引きに関する情報はあるのか？（委員）→ 該当性通知の発出当初、承認事例を用いた Q&A を追加することを想定しているが、現在までの申請では、別添の記載事例と大きく異なった製品がなかったことから、該当性判断のラインを明確化するに至っていない。（厚労省）
 - 人工知能によって新たに実現可能となった製品の一例としてトリアージ CAD が挙げられるが、それ以外に例示できる製品はあるか？（委員）→ FDA ガイダンスに記載されている CDS の概念に包括される様々な製品が該当すると考えている。（厚労省）
 - 治療法のサポート、提案及び投薬する薬剤候補を上げるような製品の該当性については我が国でも議論になっている。アメリカでは医療機器非該当と報道されたこともあるが、FDA は正式な判断を示していない。それらの製品は FDA ガイダンスの記載の範疇であるため、プログラムが回答を導き出す過程が確認できればクラス I 相当と考えられる。しかし、ディープラーニングで学習した結果の回答であれば、その根拠が不明となりリスクが高くなると思われることから、現時点では判断が難しい。（厚労省）
 - 院内業務支援プログラムの中には医師の定型業務を代行するプログラムも開発されている。このような機器への対応も必要になると思われる。（委員）→ 医療機器非該当となる院内業務支援プログラムは該当性通知に例示しているが、そのプログラムはこれらの例示に該当しないのか？（厚労省）→ 開発されている製品は、医師のみが行うことが可能な院内業務を代行するプログラムである。これは該当性通知で示された例のいずれにも該当しないと考えている。必ずしも人工知能技術が必要とするものではないが、今後、対応が必要なプログラムだと思っている。（委員）
 - 診療行為を代替するプログラムは医療機器に該当するのか？（PMDA）→ 従来の問診が紙からタブレットに変更されたのみであればリスクは低く医療機器に該当

しないと判断されるが、その情報から一定のアルゴリズムを用いて病名や重症度に
応じた診察順を提示する機能があれば医療機器に該当する可能性があり、それを判
断するための要件について検討が必要となる。今回は特に人工知能を使用した場合
に着目すべき点を明らかにして、医療機器該当性を考慮するための項目を明確にし
て欲しい。(厚労省) → 問診時に自然言語解析等を用いて、その文章を何らかのパ
ラメータに落とし込むのは「転送」になるのか？(委員) → 自然言語認識の場合、
言葉の抽出が的確に行われているか否かが問題になるため、今回、論点になると思
われる。(厚労省)

- 多様性に富んだ製品の開発が進む中、医療機器該当性に係る具体例を全て揃えるこ
とは難しいため、基本的考え方をまとめていけば良いのではないかと(委員) → 使
用目的、効能・効果が同じ製品であっても人工知能の使用の有無により該当性が変
わり得るのか？ 従来であれば、技術によって機器のクラスが変わることはなかつ
たが、そこを含めて検討する必要があるのか？(PMDA) → ポイントになる部分
の考え方を列挙し、実例を補足してまとめていく形になると思われる。現状の薬機
法に示されている医療機器の定義は変わらないため、同じ効果・効能を標榜した機
器では、人工知能の使用の有無に関わらず該当性は変わらない。今回は、人工知能
を使用した場合に生じる新たなリスクや寄与度に関する討議が重要になると考え
ている。従来技術では不可能だったが、人工知能を利用して生まれ得る全く新し
い概念の医療機器に関する議論をお願いしたい。なお、最終判断は医師が行う場合
であっても、診断・治療法の決定等に寄与度が大きければ医療機器に該当すると考
えている。(事務局)
- 海外では、画像ベースのトリアージ CAD が既に承認されている。これは人工知能
を利用して、放射線科医の判断無しに、緊急度の高い患者を脳神経外科医等が優先
して診断するように指示する機器であり、誤った判断が出た時には生命に関わるた
めそのリスクは大きいと考えられる。このような CAD でもクラス II に分類され得
ることから、今回の議論対象になると思われる。(委員) → 不具合の発生が人命に
重大な影響を及ぼす可能性がある場合は医療機器になると考えられ、英国のように
生データを確認できるか否かがポイントになると思われる。生データを常に確認で
き、機器が間違った判断を下した場合でも医師が即座に正しい判断を下すことが可
能な場合は医療機器に該当しないと判断して良いというような基準を作成するとい
う認識で良いか？(委員) → その認識で良いと思われる。プログラムの判断に
対して医師の判断がどの程度依存し得るかということが該当性を考える軸となり、
依存度が高ければリスクも高くなると考えられる。(事務局) → 判断基準として適
切な項目の抽出作業を実例も挙げながら行っていくと良いと思われる。(委員)
- 複数のプログラムが組み合わされた場合、リスクと寄与をそれぞれ単独で評価する
のか、あるいは全体として評価するのかを討議する必要があると思われる。一般
的にリスクが高い方を評価すれば良いと思われるが、組み合わせることで影響が
変わる場合も考えられる。(委員) → 国内外の最先端の開発状況も含め、様々な機器を
想定・列挙し、それらの医療機器該当性について議論して欲しい。(事務局) → 英
国で生データがなければ医療機器に該当すると判断された事例についてその背景
を調べることや、開発者側の意見を収集することは可能である。海外規制当局の情
報は厚労省や事務局で調べて欲しい。(委員) → 開発 WG の参加企業から、今回
の対象となり得る開発中の製品に係る医療機器該当性やクラス分類を判断する自社
基準等があれば提示して欲しい。(事務局) → 開発 WG 企業委員は主に日本画像医
療システム工業会 (JIRA) に参加している放射線機器の企業であるが、現存する医
療機器を超える製品の開発は想定していない。(開発 WG 事務局)
- 単体では医療機器に相当する機能を発揮しないが、人工知能プログラムの根幹とな

り得る学習済みの機械学習モデルを第三者に販売譲渡、貸与することは医療機器プログラムに該当し得るのか？ また、IT メーカーが公に提供する学習済みモデルを利用して特定のデータを学習させることで、そのプログラムが診断支援可能となるような組み合わせの場合、その該当性はどうか？ 薬機法上、医療機器プログラムの定義は電子計算機に対する指令と記載されていることから、データ自体は医療機器に該当しないと思われる。学習済みモデルが自由に利用できるようになると、医療機器には該当しない「データ」を学習させるのみで特定の機能が発揮できるようになることも考えられる。現状の解釈では、データが自由に流通して様々な形で医療に用いられることが危惧される。(開発 WG 事務局) → 学習用データを院外に持ち出すことは困難なため、ディープラーニング用システムのみを病院内に設置して、院内で学習済みの機械学習モデルを作成することは想定される。そのモデルを業者が持ち出して自社製品開発に利用する場合、モデル自体は個人情報ではないので持ち出しできる可能性があるが、その規制は検討しなければならない。一方、Google が 30 万症例の眼底写真を利用したシステムを開発したが、そのデータ情報が一般公開された場合、その取り扱いに関する新たな問題が生じ得る。これらは、現時点で我々が想像できない問題である。(委員) → データのみでは医療に利用できない。それを特定のプログラムに学習させて診療に使用するのであれば、原則、違法になると思われる。(委員) → データが機器でないことは明確であるが、プログラムと組み合わせた時の使用目的が診断や治療につながるのであれば医療機器に該当すると考えられ、どのように規制するのか考慮しなければならない。(委員) → 個人情報保護法等の適用を回避するには、個人情報として扱われないように何らかの加工を施すことも一手段になる。ただ、その情報を診断や治療目的に使用するのであれば、その製品は医療機器と判断されることになると思われる。しかし、アメリカでは支援目的であれば、非医療機器になる可能性があるとも認識している。今年度はそのようなデータを使って作製される新しい機器に対する考え方もまとめる必要があると思っている。(委員)

- FDA は医療機器に該当する可能性が高い CDS の監視を強化しているようだが、我が国でも同様の監視活動が必要と思われる。(委員)
- 今後、学会が中心となって教師付きデータを評価用データセットとして集積する事業を展開すると共に、適切な評価機関に関する検討を開始する。データを大量に収集するにあたっては、様々な問題があるため、各領域の有識者等が評価可能なデータベースを作るよう規制当局から指示された方が学会として動きやすい。共通の大きな枠組みは作成できると思われるが、撮像条件や撮影方向まで機器が制御する客観性の高い放射線機器の映像群と、超音波や内視鏡のように、撮像者毎に映像が変化する客観性の低い映像群とでは、評価・判断基準も異なってくる。また、病理診断のように、一枚の画像が持っている情報が複雑化する場合、統一基準によるデータベース化は更に難しくなると思われる。(委員) → 企業が申請時にデータの出典や品質の科学的根拠を明示できれば、医療機器として承認される。Google 等から収集してきたデータの評価については今後の課題になると思われるが、それを見越した上で該当性の考え方を決定していく必要があると思われる。(委員)
- 規制の必要性は認めるが、心電図のミネソタコードを表示する機器のように、誤診があった場合、医師の責任となることを明示する装置については、厳しい規制をかける必要がないと思っている。この範疇を広げるための議論であれば、様々な意見を出すことができると思う。(委員)
- 学会では、医療画像を収集し、評価用教師付データセットの作成を進めていると共に、その運用と当該データを用いて開発された人工知能を利用した医療機器の取り扱いについて検討している。また、治療に直結する結果が得られるコンパニオン診

断薬又はその代替となる遺伝子検査と人工知能を利用した医療機器を組み合わせた技術を倫理的な問題なく開発する手法についても検討しているが、結論が得られていない。(委員) → 本 WG では開発振興に関する議論ではなく、人工知能の利用によって新たに可能となった診断や治療のリスクと寄与度を考慮して、不具合発生が生命等に及ぼす影響の度合いについて検討して欲しい。(厚労省)

- ・ 従前から一定のアルゴリズムに基づいて判定を行う医療機器は存在している。PMDA では、検証データを用いて開発したアルゴリズム、臨床データを使用した感度・特異度等の科学的妥当性が医療機器として適切であるか判断している。人工知能を利用した機器の場合も同様の作業が行われることになるが、人工知能に特化した問題が存在する場合、それが本 WG の検討ポイントになると思われる。(PMDA)
- ・ アメリカでは産業振興を目指し、大規模データを一般公開して教師付データや学習用データとして使用させようとしている。規制は必要であるが、医師が責任を取れる範疇については、多くの製品を世に出すことが可能な規制内容とし、性能やユーザビリティ等、医師の使用感覚に基づいた判断で良い製品が勝ち残るシステムにした方が良いと思われる。(委員)
- ・ 患者自身が使用する医療機器の規制や該当性判断基準に関する議論も必要だと思われる。人工知能が搭載されたチャットボットは問診回答に従った質問過程を繰り返し、最終回答を導く装置である。医師による診察の必要性にまで踏み込んだ回答を出すようになった場合、誤診による患者リスクが高くなると思われる。(PMDA) → 患者自身が使用する医療機器も存在するため、人工知能を使用するそのような機器も対象に医療機器該当性を検討して欲しい。(厚労省)
- ・ 人工知能を利用した皮膚がん診断用アプリケーション等も開発されると思われるが、無料製品でも規制対象となるのか？(委員) → アメリカで規制の網をかいくぐって販売されていたプログラムが、FDA から警告を受けている(資料 3)。我が国で同様の監視体制を構築することも検討する必要がある。(委員)
- ・ 自社開発したテストセットによる評価では高い AUC が得られるが、第三者が作成したテストセットで評価すると悪い結果となる場合、自社評価結果で申請しても良いのか？(委員) → 審査上は一般化した性能が得られるか否か議論することになるが、自社製の評価データのみを根拠にした承認申請に関しては我々も問題視している。この問題は人工知能特有の評価に関する議論になることから、本 WG で検討して欲しい。(PMDA) → 企業側から提出された評価結果の信頼性は、審査側が自ら評価する適切な方法を持っていなければ判断できないと思われる。(委員) → 適切な評価用データセットを準備しておき、PMDA を含む第三機関でも、申請企業と同様の結果が得られることを確認する必要がある。(PMDA) → 自社の評価用データセット以外で高い AUC が得られない機器は自然淘汰されていくと考えている。それを良く理解している企業は、海外で開発した製品も、日本で上市する前に日本のデータで再評価している。(委員) → 有名な病理診断支援ソフト e-Pathologist の場合、日本人の検体を解析させたところ、期待された精度が得られないことが論文で発表されている。(委員) → そのような結果が出た場合、その機器は自然淘汰されると予想される。本 WG では、医療機器として求められる一定の水準を決定するのみで良いと思われる。承認後の課題は市場に委ねれば良い。(委員)
- ・ 人工知能は利用していないが、患者の遺伝子変異パターンに基づいて薬剤効果を判断するコンパニオン診断機器では、データベースを市販後に随時更新することで性能が変化する製品が最近承認された。市販後の定期的な性能報告を義務付けることを承認条件としたが、このケースと同様に、海外承認品を日本で上市する場合、市販後の性能報告を義務付けると共に、性能が維持されていなかった場合の対処法等を含めて評価・承認することも可能と思われる。(PMDA) → 前年度の議論でも、

申請者に性能等に関する許容範囲を明示させ、市販後にその範囲から逸脱する場合は、薬事的に必要な手続きを行うべきであるという結論になっていた。実質的にはそれと同じ考え方だと思われる。(委員)

- IDx-DR は、ほぼ自動診断が可能で、専門医以外の使用が認められている。公表臨床データは検出率 90%、特異度 90%である。一方、検出率 60%台の CAD が過去に認可されているが、審査側はどのような基準で承認の可否を判断しているのか？(委員) → 審査においては、既存機器や治療法を対照として、優勢又は非劣勢であるかを確認していると思われる。(委員)

以上の議論を受けて、先に事務局より提案されていた今年度の課題に関する対応方針(5-2/(4)項)について討議を行った。主な内容は以下のとおりである。

- 5-2/(4)項での提案 1), 2)に関しては画像診断支援装置に特化した作業、3)に関しては、医師以外が使用するヘルスケア製品も含めた該当性を討議するという理解で良いか？(委員) → ご理解のとおりであるが、2)に関しては人工知能を利用した医療機器に相当する製品全般を対象に情報を収集して欲しい。(事務局)
- 評価指標案別添は企業側から反響があったと聞いている。別添の内容に即した新しい製品が海外で開発される可能性が高いことを考慮して、内容を再度精査する必要があると思われる。(委員) → 施設毎に学習する製品に関しては、ベンチャーが開発している可能性があるため、その点も考慮して討議して欲しい。(事務局)
- 評価指標案は実用的な内容となっていることが理想であるため、更新作業では、現実とのギャップや実際の問題点を的確に抽出・分析することが重要となる。企業等からコメントが集まっているのであればそれを活用すると良いのではないか？(PMDA) → 提案に同意する。審査 WG 委員、厚労省、PMDA 及び開発側からも追加コメントが欲しい。(事務局)
- 学会や企業に意見を募集したか？(委員) → 昨年度末、学会推薦委員を介して関連学会に査読を依頼した。期限が短かったこともあり、特段のコメントは届いていないが、再依頼は可能である。企業コメントは、開発 WG 側から収集済みである。(事務局) → 関連学会を中心に再査読を依頼する必要があると思われる。(委員)
- 評価指標案を学会 HP 上で公開し、学会員から意見を求めることは可能であるか？(委員) → 可能である。国立衛研・医療機器部 HP に評価指標案を掲載した報告書が公開されているので、積極的に利用して欲しい。(事務局) → 学会の常任理事会では討議したが、人工知能に詳しい会員の意見も得られるように学会 HP に公開し、コメントを求めたいと思う。(委員)
- 申請が近い製品があると思われる皮膚科と眼科分野の関連学会の協力を得てはどうか？(委員) → 両分野とも既に優れた装置の開発研究が進んでいることから、当該分野の専門家の意見を取り入れることは重要である。必要と考えられる分野のアカデミアを学会推薦委員として追加招聘し意見を聞くことは可能か？(委員) → 新たな委員の増員も可能であるが、まずは重要と考えられる学会に評価指標案の査読を依頼してコメントを得る方法も考えられる。本事業の趣旨を伝えた上で、その学会から参加希望があれば委員追加は可能と思われる。事務局で候補となる学会を選定し、ML を利用して委員に確認する。その他に追加すべき学会があれば推薦をお願いしたい。(事務局)
- 評価指標案に関しては、昨年度に査読が終了し、必要な修正を行っているため、現時点で問題視される点はない。むしろ、開発における評価指標案の実装可能性が気になっている。開発者側から評価指標案の記載事項に対応できるか意見を聴取し、問題点を検討していくのが良いと思われる。(PMDA) → 評価指標案は、可能な限

り開発側の意見を反映させて作成した。最終版に対するコメントは募っていないことから、必要であれば再度依頼する。(事務局) → 意見を募った企業が偏っていることや、企業戦略として我々が気づいていない評価項目を敢えて指摘していない可能性等にも留意すべきである。(委員)

- この分野への参入が急速に増加すると思われるベンチャー企業の意見は反映されているか? (PMDA) → 開発 WG はアカデミア及び JIRA 推薦企業委員から構成されているため、JIRA に参加していないベンチャー企業は含まれていない。ベンチャーからの意見はパブリックコメントにより募集することになると思われる。(事務局)
- 本 WG の発足当時、初年度中に評価指標案を完成させ、次年度(今年度)は速やかな通知化と並行して、必要な討議を行う計画だったと認識していたが、その予定に変更はないか? (事務局) → 今年度から 2 年越しで通知化の作業を進めることになったと認識している。(厚労省)
- ベンチャーからの意見募集は難しいが、良い方法はないか? (事務局) → 通常はパブリックコメントを経て 1 ヶ月程で通知を发出するが、ベンチャーからのコメントが寄せられることを期待したパブリックコメントを先に行い、寄せられた意見への対応を本 WG で行った後に通知化する流れが可能か検討する。(厚労省) → 本 WG のタスクとして、日本発の実現化促進が課せられていると認識している。そのためには、開発者側に現時点での情報を積極的に公表し、収集した意見を評価指標案に反映させる努力を行うことで、より良い効果を生み出せると考えられる。(委員・座長) → その場合、大幅に修正を行った後、通知化を目的とした 2 度目のパブリックコメントを実施することになると思われる。(厚労省)
- 第 3 回会議までに、関連学会、開発 WG、ベンチャーを含むその他の関連企業から意見を収集し、評価指標案の更新、並びに医療機器該当性の考え方に関する基礎資料とし取りまとめる作業を進めて欲しい。(委員) → 評価指標案や該当性に関するコメント及び国内外における開発状況に関する調査結果は、適宜メーリングリストを通じて報告し、関係者全員で共有したい。(事務局)

討議の結果、今年度の課題の作業分担が以下のとおりに決定された。

- 評価指標案に対する追加コメント収集：事務局 (関連学会、ベンチャーを含む開発企業を対象に実施)
- 海外における開発利用及び規制状況調査：主に中田委員、藤田委員及び松尾委員
- 該当性判断の考え方と事例：全委員

5-4. その他

次回以降の審査 WG 会議の全日程が確認された。開発 WG 事務局と協議した結果、10 月 29 日(月)に第 1 回合同検討会を開催する方向で調整することが決定された。

5-5. 今後の会議日程

第 2 回会議：8 月 27 日(金) 15:00 – 18:00

場所：TKP 品川カンファレンスセンター ANNEX (予定)

以上

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 WG
平成 30 年度 第 2 回会議 議事概要

1. 開催日時：平成 30 年 8 月 27 日（月）15:00 – 18:03
2. 開催場所：TKP 品川カンファレンスセンターANNEX カンファレンスルーム 4
(東京都港区高輪 3-13-1 TAKANAWA COURT 3F)
3. 出席者（委員：座長以下五十音順・敬称略）
座長：橋爪 誠（北九州中央病院）
委員：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）
厚生労働省：田中 大祐、藤野 綾太、高村 建人、小川 雄大、宮下 正也
医薬品医療機器総合機構：岡崎 譲、加藤 健太郎、谷城 博幸、桐山 瑤子、
古森 亜矢、今川 邦樹
オブザーバー：鎮西 清行（産業技術総合研究所）、坂無 英徳（産業技術総合研究所）、
浅沼 直樹（日本医療研究開発機構）、岩田 倫明（日本医療研究開発機構）、近藤 昌夫（大阪大学）
事務局（国立医薬品食品衛生研究所）：葩島 由二、中岡 竜介、加藤 玲子
4. 配布資料
資料 1：第 2 回審査 WG 会議議事次第（案）
資料 2：委員名簿
資料 3：審査 WG 第 1 回会議議事概要（案）
資料 4：人工知能技術（AI）を利用したプログラムの医療機器への該当性に関する考
え方について（案）
資料 5：委員コメント集
5. 議事内容
 - 5-1. 開催挨拶及び資料確認
座長の挨拶後、事務局より配布資料の確認及び新規参加者の紹介が行われた。
 - 5-2. 今年度課題の現状説明
事務局より現在の活動状況が説明された。
 1. 「昨年度作成した評価指標案の追加修正について」
委員からの追加コメントはなかったが、前回会議での指摘に基づいて、AMED「臨床研究等 ICT 基盤構築・人工知能実装研究事業」で新たに採択された課題に関与している『日本眼科学会』、『日本皮膚科学会』及び『日本超音波医学会』の 3 学会を対象に、AMED 関係者の協力を得て査読を打診している。学会事務局、あるいは理事長・会長から内諾を得ることができれば、評価指標案を送付し、コメントを収集する予定である。
JIRA に参画していないベンチャー企業を対象としたコメント収集は、人工知能を活用した医療機器・サービスの研究開発に対して支援事業を行っているエムスリー株式会社の協力を得て実施する。
 2. 「人工知能を利用した医療機器の国内外における開発状況調査」

前回会議において、藤田委員、中田委員、松尾委員を中心に情報収集を依頼したが、現時点で新しい情報は届いていない。

3. 「人工知能を利用したヘルスケア製品の医療機器該当性について」

討議資料として、資料4「人工知能技術を利用したプログラムの医療機器への該当性に関する考え方について」と資料5「委員コメント集」を作成した。資料4は本WGにおける成果物の素案ではなく、あくまでも討議を行う上での参考資料である。本WGでは、資料4を参考として、人工知能を利用したヘルスケア製品の医療機器該当性判断基準に関する項目の抽出と実例を列挙する。

5-3. 総合討論

現在の活動状況に関する事務局からの説明を受けて、総合討論を行った。主な内容は以下のとおりである。

- ・ 評価指標案の英訳版があれば FDA 等に意見を求めることも可能である（委員）
→WG で必要性があると判断されれば、委員の協力を得て英訳版を作成することも検討する。（事務局）
- ・ 医療機器該当性に関しては、欠席委員から以下のコメントを受領した。（事務局）
 - 現在、医療情報分野では様々なアラートシステムが提案されているが、いずれも薬事対象にはなりにくい。人工知能（AI）を利用したシステムも開発されているが、多くは保険請求の可否判断等、医療事務を補助する製品である。
 - 放射線読影レポートの未読・既読管理システムに関しては、昨今、検査の見逃し事例が問題になっていることから注目を集めている。
 - 最終的な使用方法で医療への寄与度が変わることから、対象製品の使用目的と性能を明らかにすれば該当性も判断できると考える。
 - 例えば、治療方針や使用薬剤を決定するプログラムの場合、不適切な情報が提示され、それに基づいて医師が不適切な治療を行うことも想定される。しかし、投与後の病態変化により治療変更の必要性を考える「時間的余裕」が生じるケースのみに使用する場合には、患者への影響は小さくリスクが低いと判断できると考える。
- ・ 該当性を判断する上で、使用方法是重要なポイントである。また、時間的余裕の解釈は一樣ではないと思われるが、医療行為において今後患者に生じ得る影響やそれに対する寄与を考慮して解釈すべき事項と考えられる。（委員）
- ・ AI を利用した製品であっても医療機器該当性の判断基準は従来と変わらないと思われる。それらに対する線引きの追加議論を求めているのか？（委員）→医療機器該当性に関する考え方は変わらないが、従来想定できなかった用途を持つ製品が AI を利用して開発されてきているため、該当性判断に必要な新たな留意点を追加抽出する。（事務局）→例えば、一般的な読影レポート管理システムは AI の最大のポイントである性能変化を想定していない。このように性能変化を前提としない製品は、今回の議論では対象外と理解して良いか？（委員）→資料4に記載されているとおり、人の生命及び健康に影響を与える恐れがある製品が医療機器として取り扱われると考えられる。AI を利用して性能が変化する製品を開発したとしても、それ自体が生命及び健康に大きく影響しないと判断されれば医療機器非該当である。しかしながら、AI を利用した読影システムで収集したデータを基に全く新しい読影システムや別の医療機器が開発されることも考えられる。そのようなケースをどのように取り扱うかも議論すべきと考える。（委員）→医療機器プログラムに関しては、クラス I に相当すると判断された製品は医療機器非該当となるため、クラス I と II の線引きに必要な要件の議論も必要である。（事務局）

- ・ 今後開発される可能性がある AI を利用した製品の全体像を明らかにした後、個々の事例毎にリスクを整理することにより、該当性を議論してはどうか？（PMDA）→研究開発中の製品を調査すること自体が難しい。調査にあたっては、委員のほか、PMDA の協力も得たいが、今回、委員から該当性が疑わしい製品が非医療機器として販売されている事例が報告されているため、その製品を対象に議論を進めてはどうか。（事務局）→報告された製品は、今回の議論において非常に重要なポイントを含んでいると思われる。（委員）
- ・ 内科医は、正常ではない組織等を検出する「存在診断」、正常ではない組織が悪性か良性かを判断する「質的診断」、悪性組織の進展度を測定する「進展度診断」の段階を踏んで診断する。診断支援に使用可能な製品の医療機器該当性を考えるためには、その製品の診断レベルや性能が明示されていないとリスクも寄与度も判断できないと考えている。また、医療機器の場合、正診率のみではなく、偽陽性率、偽陰性率の明示も重要となると考えている。（委員）
- ・ 不慣れた医師が検診を行った場合、通常は専門医が二次読影を行い、重要な撮影部分が漏れていた場合に指摘を行う等、その医師の技術向上を目的とした教育を行う。しかし、市場には、二次読影なしで撮影漏れの防止を標榜している製品が存在する。資料には「専門医でも 35%が直近 1 年以内に撮影失敗を経験」、「診断支援ではなく適切な撮影・保存を支援します（非医療機器）」と記載されている。この製品は撮影漏れを防ぐのみで診断に寄与しないと標榜しているように思えるが、開発メーカーは一万枚程度のがん画像を学習データとして使用して、がん検出を目的とした AI を作製しており、この製品にその AI を流用している。アルゴリズムを変更すれば存在診断レベルの検出が可能と考えられる。（委員）
- ・ この製品は「撮影漏れの防止を支援する」と標榜して非医療機器として販売しているが、同じ性能でも標榜の仕方により医療機器に該当すると思われる。（委員）→現在、医療機器該当性は、厚労省監察官の指導に基づいて都道府県が判断している。この製品は、いずれかの都道府県から「診断には当たらない」と確認を取った上で非医療機器として販売されている可能性がある。（PMDA）→単純に撮影漏れの部位の有無を情報として伝えるのみであれば、医療機器非該当と判断することはあり得る。ただし、病変検出も可能と口頭説明した場合、その性能を標榜したことになり医療機器に該当する可能性が生じる。（厚労省）→窓口が異なっても判断が同一になるための基準を本 WG で作成していると理解している。（委員）→医療機器該当性やクラス分類に関する判断が難しい製品については、厚労省へ照会されている。医療機器該当性判断に対して影響が大きいと考えられる照会事項については、国が回答後、他の都道府県に情報提供・共有することにより、判断基準の均一化を図っている。この WG で作成された報告書が公になり、記載された考え方にに基づき該当性を判断するよう国が推奨することができれば、判断基準を一定水準で維持する有効な手段となる。（厚労省）→該当性の判断が困難であった製品の情報を収集し、リスト等に取りまとめる仕組み等が存在すれば非常に有用と思われる。（委員）
- ・ 撮影の不備をリアルタイムで検知する場合、再撮影が必要と判断されれば医療行為が増えることになる。AI の判断が誤っていた場合、医師及び患者双方の負担は増加し、放射線撮影の場合には患者の被曝量も増えるため、医療安全上のリスクは増大する。このような製品でも医療機器非該当となるのか？（委員）→指摘されているケースでは、診断の補助行為に該当する可能性がある。撮影漏れをピントのずれ等以外の要因で検知している場合、該当性判断に影響する可能性があるが、それらのケースにおける該当性を即答することは困難である。（厚労省）→結果的に医療行為自体や患者への影響の有無が、医療機器該当性を判断するための要件の一つになるとと思われる。（委員）

- ・「撮影漏れ部位」を検知する製品の例は現時点の情報のみではグレーゾーンに入ると思われることから、該当性を議論する良い事例になると思われる。プログラムの該当性判断を難しくしている要因として、クラス I 相当は医療機器非該当とするルールが存在していることが考えられる。本 WG では対象となる AI を利用した医療機器の範囲は既に整理されていると思うが、それ以外の AI を利用した製品についての該当性を海外の状況も踏まえて検討して欲しい。米国は、AI 利用の有無を問わず、ユーザーが自力で同じ結論を導くことができる製品を医療機器非該当と判断する傾向があるが、これらの製品を例として、AI による判定リスク、その結果の医療への寄与度の考え方について検討すれば良いと思われる。(厚労省) →当該製品は、AI を利用していることを標榜しているが、リスクと寄与度を検討するために必要な情報(AI の方式や使用した学習データの詳細等)も明記する必要があると思われる。(委員)
- ・医療機器該当性の判断上、リスクと寄与度以外で新たに考慮すべきポイントは現時点で思いつかない。具体的な製品例としては、医療機器非該当の製品に AI を利用して新たな機能が付加されることで医療機器に該当する可能性を考慮した。該当性通知において、「医療機器で取得したデータを診療記録として用いるために転送、保管、表示を行うプログラム」は医療機器非該当とされているが、例えば AI が判断して追加されたタグを利用して検索を行う機能が付加された場合、診断に近い意味を持つことも想定されることから、医療機器に該当する可能性がある。(委員)
- ・深層学習を用いた AI は、結論を導いた根拠がブラックボックス化していると共に、事後学習が可能となる特徴を有する。これらの特徴に起因するリスクを想定して、適切な対策を講じればリスクは小さくなり、逆に対応しなければリスクが大きくなることを記載した方がよいと思われる。(委員)
- ・診断等の根拠を提示する際、公知のガイドライン等を並列で示す機能がある場合はリスクが小さいと判断して良いと思われる。生データを同時に示す場合の取り扱いに関しては、追加議論が必要と思われる。一方、診断の根拠なしに結論のみを提示する製品のリスクは高いと判断するのが妥当と思われる。深層学習の場合、真に未知の基準に基づく判断か断定できない。現在利用している診断基準より優れた基準が生まれてくる可能性も留意して議論すべきである。(委員)
- ・事後学習に起因したリスク変化は、医療機器該当性を議論する際に考慮すべき事項である。事後学習の頻度、変化する性能を担保する手段を整理して、リスクをコントロールする具体的な手法の説明もあると良い。(委員) →資料 4 には事後学習について記載されていないが、ブラックボックス性も含め、これらの指摘事項を考慮してリスクを最小化するための対策を明示することが重要であると理解した。(委員)
- ・FDA は CADe に相当する製品を認めてきている。検出目的の機器であるため問題は生じないと思われるが、CADx の場合は、診断基準が相対的であることが多いため、診断結果の確度の判断が困難であると考えられる。一方、異なった基準で計算した結果を並列で同時に示す、優先順位を付して示す、あるいはそれらの結果と併せて統合的な結果も示すこと等で、最終的な判断を人間に委ねる形にすればリスクコントロールが可能になると考える。一方、AI が新たな診断基準を作成することは現時点で想定されていない。(委員) →未知の基準による情報の取捨選択の有無はリスク要因になるが、あまり重要ではないと考える。保証すべき性能や必要なデータ量に関するコンセンサスが得られるまで、診断等の根拠を画像や自然言語で提示する手法はリスクを下げるための重要な要因になると思われる。(委員) →FDA は、判定の根拠に対する信頼性が向上したと判断して CAD を承認している可能性がないか？(委員) →その点は不明である。(委員)

- ・医療への寄与度を考慮する際、公知情報との紐付けが確認された場合はリスクが低いと判断されることが資料4に記載されている。これは、医療機器該当性を直接判断する基準にならないと理解しているが、間違いないか？（委員）→該当性通知において、クラス I相当のプログラムは医療機器非該当として扱われることが示されているため、単純にリスクレベルのみで該当性を判断する物差しにはならない。医療機器非該当と判断されている製品に AI を搭載した場合、新たに発生するリスクの種類と程度を考慮して該当性を議論して欲しい。（事務局）
- ・「疾病の可能性の程度や疾病の候補名を提示するプログラム」の場合、導かれる結論の根拠は診療ガイドラインであるため、提示される結果はガイドラインを逆引きしていると想定していた。この場合でも、目的を考慮すればリスクが高くなるため医療機器に該当すると思われる。ただし、その根拠を自然言語や画像で説明するようであれば、単純に結果のみを提示する場合よりリスクは低いと考えられる。（委員）→これらのプログラムは診断への寄与度が高いと思われるため、事務局では医療機器に該当すると考えている。（事務局）
- ・特に患者が使用する「自己診断プログラム」、「疾病管理プログラム」の医療機器該当性判断が難しいと考えている。（事務局）→患者使用の有無にかかわらず、性能の観点では医療機器と同等の製品になる。その場合、医師が現場で使用できる可能性も考慮する必要はないか？（委員）→これらの製品を医師が使用する可能性は低く、医師がそれらの該当性を論じることは難しい。（委員）
- ・医療機器該当性を議論する上で有用な国内開発製品はないか？（委員）→CAD 関係の研究開発は学会等で発表されている。ワトソンのように治療方針を支援する製品、検査データを利用して心筋梗塞やてんかんの発症を予測して受診勧奨や生活指導を行う製品も存在する。（PMDA）→検査データに基づいて、数年後の疾患発症率を予測するサービスを行う企業のプレスリリースもある。（PMDA）
- ・データに基づいて疾病等を予測する製品は、AI を利用した医療機器として実現可能性が高いが、患者が使う場合にどのように該当性を判断すべきか議論すべきである。（委員）→受診勧奨を判断する健常人向け機器の開発を考えている企業も存在する。（PMDA）→そのような機器において間違った結果が得られた場合、重篤患者にとってはリスクが高くなる。一方、適切な結果が出た場合、医療への寄与度は高いと判断される。（委員）→リスクの大きさは間違った判断により生じた影響の大小にも左右される。通常、リスクが生じる頻度も考慮する必要があるが、FDA ガイダンスでは、プログラムのリスクは生じた影響の大小のみで判断するように示されている。しかし、ブラックボックス性に起因した予見できないリスクを考慮する価値はあると思われる。（PMDA）
- ・本来は医師による使用が前提となっている機器を患者や一般人が使用できるプログラムが実用化された場合、そのリスクはどのように考えるのが良いか？（厚労省）→そのような機器が開発された場合、一般人は通院しなくなる可能性があり、間違った判断がなされた場合のリスクはより高くなると思われる。そのリスクを何らかの形で申告しておくべきである。（委員）→医師の立場からみれば、患者が使用するプログラムでも医師が使用するプログラムと同等の性能が必要であると思われる。（委員）→数年前から iPhone のカメラを利用して心電波形を検出するアプリが存在し、確度は不明だが不整脈を検出できる機能もある。今後、類似のアプリが開発されてくるため、使用者に応じた規制は今後難しくなる。（委員）→現在、予防医療が重要になってきており、将来、患者が自ら診断を行い、その結果に応じて病院を受診するようになると思われる。患者の自己診断用製品に関する規制・基準について考えておく必要がある。（委員）→便潜血キットが薬局でも購入可能となった。しかし、陽性結果が出ても、直ちに大病院で診察を受けることはできず、必

ず開業医から紹介状を得る必要があり、患者にとっては却って手間がかかる。患者が使用可能な AI を用いた診断支援機器が上市されたとしても同様の問題が生じることが予想され、それらの製品が効率的に使用されるためには社会インフラの整備も必要になる。(委員)

- AI を利用した製品の定義や基準を定めることも必要になると思われる。(委員) → ウェアラブルデバイスから医療関連データを取得することが可能となっているが、基本的に医療機器として取り扱われていない。当該デバイスにより大量の医療データを取得することができるため、何らかの規制は必要になると思われる。(委員) → AI の利用に加えて、データの取り扱いに関する基準の作成が必要である。現在は、データ収集に対する明確な規制が存在しないため、企業がデータ収集に邁進している印象がある。(委員)
- AI を利用した製品は、専門医と同等レベルの医療支援が可能な機能を付加しないと開発する意味がない。業者が目的・用途を曖昧にして開発品の医療機器該当性判断を依頼してきた場合、十分な人員と作業時間が必要となるため、都道府県レベルでの確かな判断を行うことは難しいと思われる。(委員)
- AI という言葉が一人歩きし、どのレベルの AI を議論しているか不明瞭になることがある。(委員) → ここでは、ブラックボックス性と事後学習の特徴を備えている AI を対象にして議論すべきである。(委員) → 本 WG では、深層学習を中心として議論していると理解している。(委員) → 利用した AI の種類を明示することが重要になる。(委員)
- ブラックボックス性、事後学習以外にリスクを考慮する上で重要となる項目が他にあるか？(委員) → 医療機器該当性がリスクと医療への寄与度で判断される場合、深層学習の使用の有無等はそれらの 2 軸の大小を左右する要因となるのみで、該当性を判断する根拠になり得ないと思われる。(委員)
- 学習データセット数が少なくても、統計学的な有意差をもって良好な結果が得られることを根拠に医療機器としての有効性を主張するケースが出てくると思われるが、専門的には過学習による結果であることが疑われる。過学習の防止に関する項目は、昨年度作成した評価指標案に設けられていたか？(委員) → 深層学習の場合は、適切な性能評価方法が存在する。これについては昨年度議論し、評価指標案に反映させていると理解している。(委員) → 医療機器非該当の製品は、そのような詳細な評価を行わずに販売されることが問題である。該当性に関する議論は重要であり、医療への寄与度が非常に低い製品のみを非医療機器と判断するような枠組みにする必要があると思われる。(委員)
- AI 利用により生じる影響は、患者への寄与度よりも医師の判断支援に対する寄与度の方が大きいと考えられる。しかし、該当性判断の原則が変わるわけではなく、現時点では、経験を有した医師でも判断が困難であった疾病について、短時間で判定可能となる有用性が生じるのみである。(委員) → AI を利用することで、医療機器該当性に係る判断基準を変える必要が生じるか整理する必要がある。診断が迅速かつ正確になるのであれば問題ないが、従来では難しかった「疾病の初期段階」まで検出できようになった場合は、新たな機能を有する製品となるので、従来とは別の製品として該当性を判断することになる。このような製品を対象として該当性を議論すべきである。(PMDA) → 単純にアルゴリズム化した大量のデータやガイドラインを検索して結果を導く従来のプログラムと異なり、深層学習のように判断過程が不明なアルゴリズムを用いて出力した結果が医療に大きな影響を与える性能を持つ AI が対象となる。該当性については、その影響の種類や度合いを整理した上で議論して欲しい。(厚労省) → 医療機器該当性は、最終的な出力結果の種類で判断することになる。AI の性能と出力結果の変化に伴い該当性判断が変わる事例が存

在する可能性があれば、議論して欲しい（厚労省）→従来困難だった診断が可能となるように性能が向上する場合は、使用方法が変わることになるので、該当性通知に基づいて判断すべきである。しかし、クラス分類は使用されている技術の特性やリスクを踏まえて考える必要がある。最先端技術を利用した場合、従来確認されていないリスクが生じる可能性があることは、いつの時代も変わらないと思われる。

（委員）

- ブラックボックス性が大きなリスク要因にならない背景を説明して欲しい。（委員）
→ブラックボックス性が、プログラム管理の難しさや出力結果の解釈を困難にする可能性があることは理解する。しかし、プログラムの複雑性を理由としてリスクが上がるのではなく、誤った出力結果が患者に与える影響に応じてリスクを考えることになる。（PMDA）
- 該当性、クラス分類を問わず、評価に必要な軸を明らかにすることは必要である。その軸に対する個別事例の線引きは次の作業になると考える。（厚労省）→当初、該当性判断のために必要なリスク要因として「事後学習」による性能変化を想定していたが、それ自体は該当性に影響しないと判断し、今回は検討資料を用意しなかった。（事務局）→リスクの内訳と軸自体を検討することは別事項である。（委員）
→リスクと医療への貢献度からクラスⅠとクラスⅡとの線引きを行うことを考えていたが、今回の議論から判断することは難しい。今後の議論の方向性は、引き続き厚労省と相談する。（事務局）

討議の結果、今後の作業が以下のとおりに決定された。

- 現在市販されている AI を利用した製品（アプリを含む）をリスト化し、該当性通知及び海外の規制状況を参考として、医療機器該当性を判断する作業を依頼する。事務局側で判断理由や意見を整理し、次回会議において該当性判断のポイントとなる項目について議論する。

5-4. 次回の会議日程

期日：10月29日（月）（開発WGと合同開催）

会場：A P 東京八重洲通り13階A室

東京都中央区京橋1丁目10番7号KPP八重洲ビル

<https://www.tc-forum.co.jp/kanto-area/ap-yaesu/>

時間：審査WG会議 15:00 – 16:00

合同会議 16:00 – 17:00

開発WG会議 17:00 – 18:00

以上

次世代医療機器・再生医療等製品評価指標作成事業
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30年度 人工知能分野 審査・開発WG第1回合同会議
（平成30年度 審査WG第3回会議）議事概要

1. 開催日時：平成30年10月29日（月）15:00 – 17:18

2. 開催場所：AP 東京八重洲通り13階A室
東京都中央区京橋1-10-7 KPP 八重洲ビル

3. 出席者（委員：座長以下五十音順・敬称略）

<審査WG>

座長：橋爪 誠（北九州中央病院）

委員：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）、松尾 豊（東京大学）

厚生労働省：田中 大祐、青柳 ゆみ子、藤野 綾太、高村 建人、小川 雄大

医薬品医療機器総合機構：岡崎 譲、谷城 博幸、桐山 瑤子、古森 亜矢、今川 邦樹
事務局（国立医薬品食品衛生研究所）：薮島 由二、中岡 竜介、加藤 玲子

<開発WG>

座長：横井 英人（香川大学/審査WG委員兼任）

委員：片倉 由紀子（富士フィルム株式会社）、清水 昭伸（東京農工大学/審査WG委員兼任）、藤田 広志（岐阜大学/審査WG委員兼任）、古川 浩（東芝メディカルシステムズ株式会社）、宮崎 靖（株式会社日立製作所）、森 健策（名古屋大学/Skype参加）、森 孝夫（オリンパス株式会社）、諸岡 直樹（株式会社島津製作所）

事務局（国立研究開発法人産業技術総合研究所）：鎮西 清行、坂無 英徳

経済産業省：門川 員浩

オブザーバー（日本医療研究開発機構）：浅沼 直樹、岩田 倫明

4. 配布資料

<合同会議配布資料>

資料1：審査・開発WG第1回合同会議議事次第（案）

資料2：委員名簿（審査WG、開発WG）

資料3-1：既存の医療機器プログラム及びヘルスケアアプリについて

資料3-2：国内で承認されている診断用医療機器プログラム

資料3-3：国内で認証されている疾病診断用医療機器プログラム

資料3-4：国内で承認されている疾病治療用医療機器プログラム

資料3-5：Apple Store, Google Play等から入手可能なヘルスケアアプリ

資料4：医療機器の該当性の判断について（審査WG委員案）

資料5：開発ガイドライン案

資料6：開発WG今年度検討事項案

参考資料1：Apple watch4用医療機器アプリに関するFDA letter 1

参考資料2：Apple watch4用医療機器アプリに関するFDA letter 2

参考資料3：心電計アプリ「Embrace」に関するFDA 510(k) summary

参考資料4：心電計アプリ及びウェアラブルデバイス「Kardia Band System」に関する FDA 510(k) summary

参考資料5：Apple heart study に関する Apple press release homepage

参考資料6：Apple heart study に関する FAQs homepage（スタンフォード大学）

参考資料7：プログラムの医療機器該当性について

<審査 WG 会議配布資料>

資料5:第2回審査 WG 会議議事概要（案）

5. 議事内容

5-1. 開催挨拶及び資料確認

審査 WG 事務局より、今回の討議内容に秘匿性がないことから、審査 WG 会議は別途行わずに合同会議を2時間開催することとした。各 WG 座長の挨拶後、両事務局から、橋爪審査 WG 座長を合同会議の議長とすることが提案され、全会一致で承認された。審査 WG 事務局が進行役を務めることも承認され、配布資料の確認及び新規参加者の紹介が行われた。

5-2. 審査 WG 進捗状況報告

当日に追加配布された資料を用いて、審査 WG 事務局より説明が行われた。

(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について

人工知能技術関連の AMED 研究課題に採択されている下記3学会に、昨年度作成した評価指標案の査読を依頼した。これまでの対応は以下のとおりである。

- ・日本眼科学会：研究開発代表者 大鹿 哲郎 先生（筑波大）
理事会決議を得て、学会内に創設されている AI 関連事業担当専門委員会に諮問後、寄せられた答申を取りまとめて本事務局へ返送する。
- ・日本皮膚科学会：研究開発代表者 天谷 雅行 先生（慶應大）
皮膚科学会内に設置された AI 関連 WG 委員が査読した後、コメントを収集・取りまとめ、10月末を目処に事務局へ返送する。
- ・日本超音波医学会：研究開発代表者 工藤 正俊 先生（近畿大）
AMED 研究課題の研究開発代表者としての立場で回答を得た。主な内容は以下のとおりである。

<コメント>

➤ 5/(1)項：ブラックボックスとしての性質

AI による診断は、医師が理解可能な根拠が示されるべきであり、診断名に加えて診断を補助する所見もアノテーションとして付帯させるべきである。

➤ 6/(3)項：安全性、品質及び性能評価

テストデータセットは、画像データの品質とアノテーションの正確性が担保されるべきであり、例えば各学会主導で構築されている全国規模のデータベースを用いた品質及び性能評価を行うことが望まれる。

➤ 6/(4)項：リスクマネジメント

使用者はトレーニングを受けた医師、医療従事者とし、診断や医療行為の最終責任者は医師とすべきである。また、使用者を行政が把握するためのシステム（届出等）、が必要と考えられる。超音波機器の場合、携帯型の機器が存在するため、AI ビルトイン型が上市された場合の施設管理者や使用者を把握し、問題点が発覚した場合に速やかに通知、あるいは報告するシステムが必要と考えられる。

ベンチャー企業を対象にした査読については、ベンチャー支援企業にコメント収集を依頼しており、現在、回答待ちの状況である。

寄せられたコメントについては、第4回審査WG会議までの間に、評価指標案の修正要否も含めメールベースで討議したい。

上記報告後、質疑応答が行われた。主な内容は以下のとおりである。

- ・眼科学会及び皮膚科学会にはAI関連WG等が設置されているようだが、他に同様のWGを設置している学会もあるのか？（開発WG委員）→全てを把握できていないが、病理学会、消化器内視鏡学会及び医学放射線学会にはWG等が設置されており、学会推薦委員を介して当該学会の意見は評価指標案に反映されている。（事務局）
- ・学会等に評価指標案の査読を依頼した経緯を確認したい。（審査WG委員）→本審査WGの設立にあたり、AIに関連したAMED研究課題を実施していた病理学会、内視鏡学会及び放射線学会から委員推薦を受けた経緯に鑑み、新規AMED研究課題を推進している関連学会にも査読を依頼した。また、ベンチャー企業に関しては、業界団体に所属していない場合があることから、ベンチャー支援企業にコメント収集を依頼した。（事務局）
- ・海外では眼科及び皮膚科疾患を対象としたAI診断機器が既に存在することが過去の議論において報告されたことから、対象臓器固有の評価項目の設定要否等に関する情報を収集するために関連学会へ査読を依頼した経緯がある。収集されたコメントに応じて、特段の注意が必要であれば評価項目を追加する。（審査WG委員）
- ・超音波学会から指摘された学習データセットの取り扱いに関する問題は、他の分野にも共通しており、データベース整備が進めば今後追加検討が必要となる可能性がある。（審査WG委員）

(2)人工知能を利用した医療機器の国内外における開発状況調査

AIを利用した自動診断装置としてFDAが承認したIDx-DRに関する講演情報が共有された。主な内容は以下のとおりである。

- ・IDx-DRはクラスIIで承認されており、眼科医以外の医療従事者も診断に使用できる機器である。診断には従来のアルゴリズムが使用されており、ディープラーニングは使用されていない。現在、当該機器はトプコム製眼底カメラシステムに使用が限定されているが、今後の展開は不明である。（審査WG委員）→米国では、一般内科医の検診により糖尿病性網膜症の疑いがある患者に眼科の受診勧告を行うが、その正診率は3-4割と言われている。これに対してIDx-DRで当該疾患の疑いが検出された患者の最終的な正診率は9割弱であったことから、有効性が認められFDAで承認された。一方、日本では、糖尿病内科専門医が患者の状態を判断して眼科専門医の受診を推奨することが一般的で、米国に比べてその正診率が高いことから、IDx社及びトプコム社はIDx-DRが日本で普及するか懸念している。（PMDA）→日本では健診医のほとんどが内科医であることから、健診時のファーストスクリーニングとして使用される可能性があると思われる。（審査WG委員）
- ・現時点では米国でも明確に規制されていないようだが、そのような状況でもIDx-DRが自動診断に使用されても良いのか？（審査WG委員）→講演スライドには「autonomous diagnosis」と表現されていたが、最終的には医師による診断が求められるシステムである。（審査WG委員）
- ・IDx-DRで得られるデータは教師付きデータにならない。事後学習機能の付加につ

いては、別途臨床研究を行う必要があり、コストがかかりすぎる。また、FDA は事後学習機能付き医療機器の承認に対して消極的であること等から、現在、IDx-DR は事後学習機能を有していない。(開発 WG 事務局)

5-3. 総合討議

(1) 「人工知能技術を利用したヘルスケア関連プログラムの医療機器該当性」について資料 3-5 に基づいて討議された。主な内容は以下のとおりである。

- 資料 3-5 に列挙した「一般人が入手可能な代表的なヘルスケア関連プログラム」の医療機器該当性について、事前に寄せられたコメントを精査した。
- 公知のルールや医療計算に基づき電卓等で行っていた計算を代替するのみのプログラムは医療機器非該当となる判断は、全てのコメントで一致していた(事務局)。→プログラムに病名予測機能等が付属している場合、その機能に応じて医療機器該当性を検討する必要がある。(審査 WG 委員) →プログラムが異常値を算出すると警告を発する機能を有する場合は、該当性判断に影響を与えないか？(PMDA) →警告を発する判断の根拠に依存して当該プログラムの医療機器該当性は変化すると思われる。例えば、ガイドライン等で特定マーカの検査値が上限を超えた際に対象患者に施す治療法が定められている場合、測定値とともにその情報を単純に表示するプログラムは、辞書同等と判断可能であり、医療機器に該当しない。一方、治療の是非を固有のアルゴリズムにより判断するプログラムの場合は、単なる辞書とはならず、医療機器該当性を検討しなければならない。(厚労省)
- 文献データベースや検索機能を有するプログラムは、電子辞書相当として医療機器非該当と考えられる。患者向けの説明や医師のトレーニングに用いる場合も医療機器非該当になるが、その使用に伴う治療成績の向上や入院期間の短縮が標榜されている場合は医療機器に該当する可能性がある。(審査 WG 委員) →標榜される効能も医療機器該当性を判断する上で重要である。診療に関連した効能を標榜する場合は、医療機器に該当すると思われる。(審査 WG 委員)
- 病院の検索・予約機能に加え、症状を診断し受診すべき病院・診療科を表示する AI を利用したプログラムが存在するが、疾患名を推測する機能や遠隔診療機能も有するため、医療機器に該当する可能性が高い。(審査 WG 委員) →当該プログラムは明確に「診断」と標榜していることから、医療機器に該当すると思われる。(審査 WG 委員) →遠隔診療であっても初診は対面が原則ではないか？(PMDA) →原則は指摘のとおりと思われる。しかし、画像解像度が向上しているため、将来、その原則は再考する必要が生じるとと思われる。(審査 WG 委員)
- 学会が作成したガイドラインや診断基準に従い、設定された設問への該当数に応じて疾病の可能性を提示するアンケート式プログラムは医療機器に該当しないのではないか？(審査 WG 委員) →フローチャート化した診療ガイドラインに従って「Yes」、「No」で判断した結果を示すプログラムは医療機器に該当しないと思われる。一方、ガイドラインの内容を解読して作成した新たなフローチャートを基に開発されたプログラムの場合は該当性を慎重に判断する必要がある。また、医師はガイドラインの内容を把握しており、提示された情報が誤っていても直ちに指摘できるが、医療従事者以外が使用した場合は、その誤りに気づかないリスクが高まる。医療従事者が使用する場合にはクラス I であっても、医療従事者以外が使用場合はクラスが上がることも考えられる。(厚労省) →診断フローチャートがある程度確立している場合でも、AI を用いるとそのフローチャートでは得られない結果が提示される可能性があるため、診断根拠の精査が必要である。(審査 WG 委員)
- 受診すべき病院・診療科を表示するプログラムでは、「AI が診断」という標榜自体

に問題があるが、その使用目的は診療前に適切な病院や診療科の選択を補助することにある。医療行為に入る前の使用を前提としたプログラムの医療機器該当性はどのように判断すべきか？（開発 WG 委員）→「診断」と標榜していることから、医療機器に該当すると判断するが、痛み等を感じる部位に応じて診療科を単純に提示する程度の効能であれば、適切な標榜により医療機器非該当と判断される。（厚労省）→当該プログラムに関するホームページを確認したところ、その効能は曖昧な表現で標榜されている。当該プログラムが、そのような状況で販売されていることも問題である。（審査 WG 委員）→医療機器として扱うべきプログラムについては、規制当局が市場動向を把握できるようなシステムを構築することも重要と思われる。（審査 WG 委員）

- ・ 但し書きとして「医師診察の代行となる診断を行うものではない」と提示されるプログラムもあるが、「症状から病名を予測」等、診断と同等の効能を標榜するプログラムは医療機器に該当することが厚労省より示されている。（事務局）→曖昧な表現で効能を標榜していても、特定の病名・病態を提案していると思われる場合は、医療行為に相当すると判断して良いか？ 症状に応じた受診勧告を行うプログラムは医療機器非該当との議論があったが、この勧奨も何らかの病態を想定して行われていると考えられるため、切り分けが不明である（審査 WG 委員）→基本的に「目が痛いので眼科を受診する」という判断は特別な知識等が必要ではなく、公知の事実として考えることができる。一方、様々な症状を基に想定される疾患と受診すべき診療科を提示する場合は、その判断手順・内容を踏まえて該当性を判断する必要がある。（厚労省）
- ・ 遠隔診断プログラムの場合、通信機能以上の効能を標榜すると医療機器に該当する可能性があるが、単なるテレビ電話機能や検査データを加工せずに通信することを目的とするプログラムは医療機器非該当と思われる。（審査 WG 委員）→医療機器に該当するプログラムと情報を連携する場合、該当性をどう判断すれば良いか？（審査 WG 委員）→万歩計データ、人間ドック診断結果、睡眠データ等を入力して自己管理することを目的としたプログラムの場合、連携形態が単に当該データの通信であれば、検査データを未加工で伝送するプログラムと同等であり、医療機器非該当と考える。一方、連携の結果、単なるデータ通信に留まらずに診断に対して影響を与え得る機能を有すると判断されるプログラムの場合、医療への寄与度を考慮して医療機器該当性を判断しなければならない。（厚労省）
- ・ AI を利用していても、健康管理、メンタルケア及びダイエットサポート等を目的とするプログラムは、主に健常人が使用することから医療機器非該当であると思われる。一方、AI を利用することで初めて可能となる支援内容を示すプログラムは、医師もその情報を積極的に使用する可能性があるため、健常人が使用する場合でも診断支援に該当し、医療機器に該当する懸念がある。後者に該当する具体例はあるか？（審査 WG 委員）→現時点で具体例の想定は困難であるが、その可能性は留意すべきである。（審査 WG 委員）
- ・ 特定の臓器に障害・疾病を有する患者が同様のプログラムを使用する場合、医療機器該当性はどのように判断すれば良いか？（PMDA）→学会ガイドライン等、公知の情報に基づいて判断され、その根拠を患者が判読可能な場合は医療機器に該当しないと考えられる。一方、それ以外の専門的判断を加味して助言等を行う場合には医療機器に該当する可能性がある。（厚労省）
- ・ 病院向けの AI を利用した問診プログラムは、医療機器として取り扱われているのか？（審査 WG 委員）。→紙の問診票を単にプログラム化した場合は該当しないと思われるが、回答に応じて問診内容が変化するプログラムであれば医療機器に該当する可能性がある。（厚労省）

- ・年齢、生活習慣及び家族の既往歴等から AI を利用してがんのリスクを単純に算出するプログラムは医療機器非該当と思われるが、がんリスクの提示形式に応じて医療機器該当性が異なる可能性がある。(審査 WG 委員) → 遺伝子診断によるがんリスクに関しては文献によって判断が異なっており、確立した科学的根拠を示すデータは存在しない。通常の遺伝子と変異遺伝子のコホートに関する調査研究におけるリスク比やオッズ比に基づいてがんリスクを提示しているケースが多い。医事課は、統計解析結果が正しく説明されている場合、診断に該当しないと判断している。当該プログラムも生活習慣や既往歴等から判断した疫学的な統計値を示すと考えられるので、診断行為に該当しない。(厚労省)
- ・診断精度が悪くても恩恵を受ける患者が多い、あるいは医療費コストの削減に繋がる等の政治的判断が医療機器該当性判断に影響することはあり得るか？(審査 WG 委員) → 有り得ない。但し、医療機器の承認はリスク・ベネフィットバランスを勘案するため、リスクが低ければベネフィットが少なく性能が低い機器であっても承認拒否要件に合致せず承認されることはあり得る。(厚労省)
- ・WHO が骨折リスクを計算するプログラム「FRAX」を公開しており、日本人に合わせたデータも使用可能である。このような公開プログラムを製品開発、あるいは製品自体に利用することも考えられることから、その対応についても検討すべきである。(開発 WG 委員) → プログラムの流通方法として、インターネットを介してウェブからダウンロードできる形態も存在する。例えば、海外のウェブページからダウンロード可能なプログラムが国内規制上は医療機器に該当する可能性も想定されるが、その取り扱いはどのように考えるべきか？(PMDA) → 基本的に、海外のウェブサイトには日本の国内法規が及ばないため検討対象とはならないが、日本向けに海外から提供されている場合は議論する必要がある。有体物の医療機器が未承認の状態でも日本向けに販売されている場合は医薬品医療機器法違反になるため、対象 Web ページの削除をレジストラ等に要請している。しかしながら、WHO で公開されているプログラムの場合、慎重な検討が必要だと考えている。なお、FRAX は連携大学が開発したものであり、WHO はその開発等に一切関与していないことが公表されている。よって、製品開発に利用する場合には、FRAX 自体も審査対象となることに留意されたい。(厚労省) → AI を利用したヘルスケア関連プログラムに対しては、その流通経路等を規制当局が把握できるよう申請等、何らかの作業を義務付ける必要があると思われる。(審査 WG 委員)
- ・診断時に撮影漏れの有無をチェックするプログラムにおいて、リアルタイムで漏れを知らせる機能を有する場合は医療機器に該当する可能性がある。(審査 WG 委員) → リアルタイムで撮影漏れが分かる場合、再撮影を要する等、診療行為に大きな影響を与えるため、医療機器に該当すると考えられる。(厚労省)

(2) 医療機器該当性を判断する上で必要となる要素について

審査 WG 委員から、資料 4 の内容が以下のとおりに概説された。

- ・医療機器該当性の評価軸は「医療への寄与度」と「リスク」であり、原則は変わらない。
- ・該当性の判断要素として、効能の標榜、対象疾患の性質、判断機能の有無、事後学習機能の有無、使用者のほか、誤った提示結果を使用者が指摘できる可能性を想定した。
- ・各要素を組み合わせ、医療機器該当/非該当と考えられる事例をそれぞれ具体的に記載した。
- ・一般人(健常人)が使用するプログラム中、効能を標榜しないと共に、医師又は医

療機関と非連携下に自己完結する製品は医療機器に該当しないと想定した。

- ・ 事後学習機能を有する場合、販売後の事後学習により当初標榜していた性能を担保できなくなるリスクがクラス分類の判断に影響を及ぼすことを想定し、考え得る事例を記載した。特に、非医療機器として販売されたプログラムが事後学習により医療機器相当の性能を有する可能性は否定できない。そのプログラムが医療上、重要な判断に使用される懸念があるため、新たな監視体制等が必要になると考えられる。

説明後、医療機器該当性を判断する上で必要となる要素等について全体で討議した。主な内容は以下のとおりである。

- ・ ヘルスケア関連プログラムの医療機器該当性は、支援等の対象となる「観察」、「検出」、「判断」等、人間の知的活動に基づいた整理が必要と思われる。（開発 WG 委員）
- ・ 医師が新たな使用法を見出した結果、医療機器プログラムが適応外使用される可能性も考えられる。適応外使用に伴いリスクが増大する可能性もあるが、新たな使用法を効能として追加するために薬事対応を行うか、適応外使用を責任範囲外とするかは製造販売業者の判断になる。一方、適応外使用が医師により発信され公知となった場合、その使用が患者にとって有用である可能性があるため、規制当局が適応拡大を啓発するような仕組みの構築を要望する。（審査 WG 委員）
- ・ 開発を促進する上で、例えば、AI を利用した医療機器に対して条件付きで承認を与えた後、一定期間継続的、あるいは一定間隔でモニタリングを行い、適応や性能が変化する毎に製造販売業者が申告し、最終的に達成された性能で承認を与える等、これまでにはない柔軟な規制システムを構築すべきと考える。（審査 WG 委員）

5-4. その他

次回会議までに、以下の事項についてメールベースで討議することとなった。

- ・ ヘルスケア関連プログラムの医療機器該当性について考慮すべき要件（評価軸）の洗い出し。
- ・ 医療機器該当/非該当となるプログラムの具体例の作成。

開発ガイドライン案については、審査 WG 委員も内容を確認し、必要に応じて審査 WG 事務局を介して開発 WG 事務局にコメントを提出することとなった。提出されたコメントは、ガイドライン改訂版作成時に開発 WG で精査し、取り扱いを決定する旨が開発 WG 事務局より説明され、参加者から了承された。

5-5. 次回の会議日程

第 4 回会議：1 月 18 日（金）15:00 – 18:00

会場：TKP 品川カンファレンスセンター ANNEX

以上

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 WG
平成 30 年度 第 4 回会議 議事概要

1. 開催日時：平成 31 年 1 月 18 日（金）15:00 – 18:06
2. 開催場所：TKP 品川カンファレンスセンターANNEX カンファレンスルーム 6
東京都港区高輪 3-13-1 TAKANAWA COURT 3F
3. 出席者（委員：座長以下五十音順・敬称略）
座長：橋爪 誠（北九州中央病院）
委員：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）、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）、松尾 豊（東京大学）、横井 英人（香川大学）
厚生労働省：田中 大祐、青柳 ゆみ子、藤野 綾太、小川 雄大
医薬品医療機器総合機構：岡崎 譲、谷城 博幸、加藤 健太郎、桐山 瑤子、古森 亜矢、今川 邦樹
オブザーバー：坂無 英徳（産業技術総合研究所）、近藤 昌夫（大阪大学）
事務局（国立医薬品食品衛生研究所）：齋島 由二、中岡 竜介、加藤 玲子
4. 配布資料
資料 1：第 4 回審査 WG 会議議事次第（案）（本紙）
資料 2：委員名簿
資料 3：審査 WG 第 3 回会議（合同会議）議事概要（案）
資料 4：人工知能技術を利用した医療機器プログラムの該当性判断要素について（案）
資料 5：人工知能技術を利用した医療機器の国内外における承認状況
資料 6：パブリックコメント時に評価指標案に寄せられたコメント及び厚労省対応案
資料 7：関連学会からのコメント及び審査 WG 対応案
資料 8：平成 30 年 12 月 19 日付け医政医発 1219 第 1 号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」
資料 9：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改訂案
資料 10：報告書目次（案）

5. 議事内容

5-1. 開会にあたり

座長挨拶後、事務局より配布資料の確認が行われた。事前送付が遅れた前回会議議事概要案については、内容を確認の上、意見等があれば 1 月末を目処に事務局へ連絡することとした。

5-2. 総合討議

資料 4「人工知能技術を利用したヘルスケア関連プログラムの医療機器該当性」について、事務局より説明された。主な内容は以下のとおりである。

- ・ 疾病の診断・治療・予防がプログラムの目的である場合は医療機器に該当するため、該当性判断における最初の要素として「目的性」を設けた。

- ・ 「目的性」を判断する事例として、次の3件を記載した。
 - プログラムが使用者に提示する情報の内容が一般的な健康情報であれば非該当であるが、具体的な個別疾病名を提示する場合は医療機器に該当する。
 - 健康維持トレーニングを目的として使用者に特定の身体動作を促すプログラムは非該当であるが、リハビリテーションや疾病予防を目的とする場合は医療機器に該当する。
 - 病態を単に記録するプログラムであれば非該当であるが、その記録を更に分析して医師の診断を支援することを目的としている場合は医療機器に該当する。
- ・ 次に考慮すべき要素として「寄与度」と「リスク」があるが、密接に関連することから一つの項に取りまとめ、両者を評価して医療機器該当性を判断する必要性を記載した。
- ・ 提示される情報の根拠に応じて「寄与度」が変化すると考えられる。公知のガイドライン等に基づいた結果を示す場合は寄与度が低いと考えられるが、独自の分析を行って診療等に重要な結果を示す際は寄与度が高くなることを記載した。
- ・ 「リスク」の大きさは健康に影響を及ぼす程度に左右されることを、具体的な事例と共に示した。
- ・ 「目的性」の観点から医療機器に該当するプログラムであっても、「寄与度」と「リスク」が低い場合には、クラスⅠ相当と判断され非該当となることを示した。また、クラス分類判断に必要と考えられる要素を記載した。
- ・ 3項には医療機器該当性の判断事例を記載した。
- ・ 提言の原案では、事後学習により「想定しなかった使用方法が可能となる場合」の企業対応として想定されるケースを3種類挙げて、それぞれの課題等を記載していた。しかし、薬機法違反となるケースが重複記載されていたことから、一方を削除した。
- ・ また、事後学習により「性能が想定範囲を下回る」場合、医療機器非該当のプログラムであっても健康被害が生じる可能性への懸念が示されていた。しかし、当該プログラムであってもリスクはゼロではないことから、記載不要との意見が寄せられている。この記載の是非については議論を依頼する。

事務局による説明と並行して討議が行われた。主な内容は以下のとおりである。

- ・ 「目的性」の判断事例で、健康管理・患者説明・教育用データの単なる送信・保管を目的とするプログラムは医療機器非該当とある。この場合、データの種類もしくは送信・保管行為のどちらが医療機器非該当と判断される「目的性」になるのか？→非該当と判断される指標は健康管理・患者説明・教育用プログラムである。なお、医療用データであっても電子カルテのように単なる送信・保管プログラムであれば、医療機器非該当となることを補足した。(厚労省)
- ・ クラスⅠとⅡとの分類を判断する追加留意点を想定して別項で取りまとめたが、結果として「寄与度及びリスク」の項に記載した内容と重複した。この項はタイトルも含めて追記・修正が必要と思われるので検討を依頼する。(事務局)→原案は、医療機器該当性判断の流れが分かり難いため、文章全体の構成を再考する必要がある。(委員)→「目的性」で該当性を判断した後、医療機器に該当するプログラムについてはクラス分類を判断するため、「目的性」、「寄与度」、「リスク」をそれぞれ判断するポイントを記載するのが良いと思われる。クラス分類判断に必要な項目は削除して、「寄与度」及び「リスク」に包含させる形に修正するのが望ましい。(厚労省)→「寄与度及びリスク」の項も「目的性」と同じ構造に揃えれば、理解し易くなると思われる。(厚労省)

- ・使用方法・効能は「目的性」に記載されていることから、最初に考慮すべき要素だと思われる。(委員)→「目的性」、「使用方法」、「効能効果」、「寄与度とリスク」の順に判断するのであれば、整理し直すべきである。(委員)→使用方法・効能は「寄与度」と「リスク」にも関連するため、それぞれに留意点として記載すべきだと思われる。(委員)→使用方法・効能のみで「目的性」が判断されることは避けたい。(厚労省)→「寄与度及びリスク」の項に「寄与度」、「リスク」の中項目を設け、それぞれの説明と判断要素を記載するのが良いと思われる。使用方法・効能を含めたクラス分類判断における留意点は、それぞれの要素を考慮するための留意点として整理して記載すれば、位置付けが明確になるとと思われる。(PMDA)→使用方法・効能は「寄与度」と「リスク」の両方に対する留意点として記載されていることに注意した修正が必要である。(委員)
- ・「寄与度」及び「リスク」については、詳細な説明を追記することで理解が深まると考えられる。(委員)→通知として発出されている「プログラムの医療機器への該当性に関する基本的考え方について」に記載されている定義を引用すれば良いと考える。(厚労省)→そのまま引用するのみではなく、「目的性」と同様、定義の後に具体的な事例及び判断時の留意点を記載する形式にするのが良いと思われる。(事務局)
- ・医療機器該当性を判断する上で考慮すべき要素は「寄与度」と「リスク」であることは明確にすべきである。(厚労省)→両要素が共に一番低いと判断される場合はクラスⅠ相当に分類されることを明示したい。(厚労省)
- ・「寄与度」と「リスク」が相関しない場合は特殊事例と考えて良いか？(委員)→その理解で差し支えない。(厚労省)→診断機器等は寄与度が大きいリスクは少ない例と思われる。(委員)
- ・プログラムの医療機器該当性に関する資料は、プログラム開発研究を行っているアカデミアや企業のために作成を依頼している。(厚労省)→判断に必要な要素のみならず、具体例も記載して欲しい。例えば、判断に必要な要素として「医師等の医療従事者が使用するものや、医師の指示の元に患者が使用するものか」と記載されているが、どのような場合に寄与度やリスクが高くなるのか？(委員)→医療従事者以外が使用する場合にはリスクが高くなる。その点は、資料を修正すれば適切な箇所に事例として記載できると思われる。使用者を医師に限定すればリスクは低くなるが、そのみでクラスⅠと判断することはできないことに留意すべきである。(厚労省)
- ・IMDRFにおいて、プログラムの「寄与度」は自明の情報、医療の方向性を示唆する情報、もしくは治療・診断情報を提示するかの3段階が定義されている。(厚労省)→クラスⅠとⅡとの分類が課題となるため、寄与度に関しては1段階目とそれ以外に該当する事例2つを記載すれば良いと思われる。(事務局)
- ・「リスク」に関しても同様の修正を行えば良いか？(事務局)→「リスク」の大小に関する判断は、規制当局が適宜行う。この文書では、「リスク」は健康に対する影響の程度に応じて分類する方向性を示すのみで良いと考える。(座長)
- ・性能の「標榜」は使用方法・効能とは別の問題であるため、削除すべきではないか？(委員)→後述されている事例に補足すれば良いと思われる。(厚労省)→事例に先立って、医療機器該当性判断においては性能の「標榜」が参考になると補足説明がある。医療機器相当の性能を有さないにもかかわらず過度の「標榜」により該当すると判断される、もしくはその使用実態が医療支援相当にもかかわらず医療機器としての性能を「標榜」していないことから非該当と判断される場合の2種類が問題になると想定される。その点を考慮した事例の記載が必要と思われる。(委員)→「標榜」に関する留意点を詳述してはどうか。(厚労省)

- ・資料のタイトルを適切に修正する必要がある。(事務局)
- ・資料中の「医療機器」という用語が、「プログラム医療機器」、あるいは「一般医療機器」を指しているのか区別し難い。「医療機器」を可能な限り使わない形に修正すべきである。(委員)
- ・記載した医療機器該当性判断事例は、該当例と非該当例を対にして示すべきか。(厚労省) →開発者にとっては、非常に有益な情報になるので、その構成での記載を要望する。(委員)
- ・プログラムの目的が「睡眠の質」を判定するのみであれば医療機器非該当と判断されるが、「睡眠時無呼吸症候群」等の具体的な疾患名を提示する、あるいは診断行為に寄与可能な別情報も提示する場合は医療機器に該当することを事例として記載する選択肢もある。(厚労省) →判断上、特定疾患の提示が重要になるのか？(委員) →一般的な総称の場合、判断は難しい。医療従事者以外の大部分の一般人がその総称を疾患と認識する場合は、医療機器に該当すると思われる。(厚労省)
- ・非該当例として記載されている健康管理プログラムにおいて使用するデータ例は、一般的な記述とした方が誤解を生じないと思われる。(厚労省)
- ・「提言」の趣旨を明確にする必要がある。(厚労省) →医療機器に該当しないと判断されたクラスⅠ相当のプログラムにおいて、事後学習により生じうる健康被害を監視する体制整備の要求が「提言」の趣旨である。(事務局) →非該当の場合、薬事上の規制は不可能であり、別の法律で対応することになる。健康被害を被った使用者が企業に対して補償を要求することが実効性のある対応と考えられ、それに代わる具体的な方策を提示しないまま「提言」を提出しても現実性に劣る。(厚労省) →事後学習による性能変化の責任主体が企業であることを明確化することが、この提言の目的の一つであると考えている。(委員) →企業責任を明確にする仕組みの必要性を記載するのは良いが、表現は再考すべきである。(厚労省)
- ・資料4及び評価指標案で使用されている「事後学習」は、人工知能分野では一般的用語ではなく、特定技術における一部分を示すこともあり、定義が曖昧である。今回議論されている「事後学習」は市販後に新たに行う学習を意味していることから、誤解が生じないように定義を明確にする必要がある。(委員)

討議の結果、資料4のタイトルを「人工知能技術を利用したプログラムの医療機器該当性判断要素について(案)」に変更することとなった。また、2項以降については、以下のとおりに修正することとした。

- ・「2. プログラムの医療機器該当性を判断する上で考慮すべき要素」の構成を変更する。
 - 中項目を「(1)目的性」、「(2)寄与度とリスク」とし、記載形式を統一する。また、「寄与度」と「リスク」を個々に項立てし、それぞれに定義及び具体的なプログラム例を示す形式に修正する。
 - 「標榜」や提供される情報の種類について重複する表現が項内に散在していたため、適切な箇所に集約して記載する。
- ・「3. 医療機器該当性判断例」における事例を適切な記載に修正する。
 - 誤った事例を削除し、代替として「睡眠時無呼吸症候群」や睡眠状態に関連する情報を提示するプログラムを例示する。
 - 「プログラム例」の2)/①項に記載された医療機器非該当プログラムで取り扱うデータを「健康に関するデータ」に変更する。
- ・「使用者の健康に関わるプログラムの取り扱いに関する提言(案)」の内容を、現状に即した表現に変更する。
 - 「1. 性能が規定していた範囲を超え、当初想定していなかった使用方法

が可能となる場合」における企業対応のうち、未承認医療機器の販売に相当する2例を整理統合する。

- 医療機器非該当プログラムであっても事後学習による性能変化の結果、使用者に健康被害を生じる危険性への対応については、表現を再考した上で開発企業の責任に言及する内容へ変更する。

5-3. 人工知能を利用した医療機器の国内外における開発状況

藤田委員より、スライドを用いて北米医学放射線学会（2019.11.25～11.30, シカゴ）の概要が以下のとおりに報告された。

- ・多くの中国及び韓国企業が開発中のCADプログラムを紹介しており、国を挙げた産業化が活発に進められている。また、複数の製品が当該国で承認されていることが広報されていた。
- ・米国、欧州及び日本企業も自社製品（既承認及び開発中を含む）を紹介していた。

資料5に基づいて、事務局より説明された。主な内容は以下のとおりである。

- ・米国、韓国及び日本において、人工知能技術を利用した医療機器として承認されているプログラムを対象にクラス分類、効能、アルゴリズム等の情報を取りまとめた。
- ・2017年以降、米国FDAで承認されているほとんどの製品は510k、もしくはDe Novoで認可・承認されていた。調査した17製品中、iCAD社のProFound AIがクラスⅢである以外は、すべてクラスⅡに分類されていた。
- ・韓国で人工知能を利用した医療機器として承認された製品は3種類あった。MRI画像から虚血性脳梗塞の検出及びその重篤度判断を支援する製品はクラスⅢ医療機器として承認されていることも判明した。
- ・日本では、AMEDの支援を受けて昭和大学、名古屋大学及びサイバネットシステム（株）が連携して開発した大腸病変の腫瘍/非腫瘍の判別を支援するEndoBRAINがクラスⅢ医療機器として12月に承認された。使用されているアルゴリズムは機械学習の一つであるサポートベクターマシンであり、承認後の性能向上を目的とした事後学習機能は有していない。

5-4. 昨年度作成した評価指標案に寄せられたコメントと対応

資料6及び資料7に基づいて、事務局より報告・説明された。主な内容は以下のとおりである。

- ・評価指標案に対するパブリックコメント及び厚労省対応案の概要を資料6に示した。審査WGで議論を要するコメントはなかったため、この内容に則って厚労省が対応する予定である。修正又は検討が必要と考えられる場合は、1月末を目処に事務局に連絡して欲しい。
- ・参考までに、厚労省対応案に従い加筆・修正した評価指標案を資料9として配布した。
- ・資料7は、日本眼科学会、日本皮膚科学会及び日本超音波学会（理事長）から寄せられたコメントとメール審議を経て作成した審査WG対応案である。事前回覧ではコメントが寄せられなかったが、特に問題なければ審査WG回答として各学会へ送付したい。→討議の結果、異議がなかったことから、各学会の理事長及び事務局宛てに送付することとなった。

5-5. その他

今後の作業内容とスケジュールについて事務局より説明された。

- 本日の討議内容に基づいて「人工知能技術を利用したプログラムの医療機器該当性判断要素について（案）」を更新し、メール審議を経て最終版を作成する。
- 本WGの平成30年度報告書目次案を資料10に示した。参考資料については、報告書に全て掲載する形式としているが、Web経由で簡単に入手可能な資料をURLリストに差し替える案もある。→討議の結果、後者の形式とすることとした。
- 今後の作業は、資料10に示したタイムスケジュールに従って進めることとなった。

<今後のスケジュール>

- 1月末：該当性判断要素に対する追加コメント、研究開発動向に関する情報提供が切
- 2月中旬：該当性判断要素最終版、研究開発動向資料最終版及び第4回議事概要（案）の確認依頼
- 2月後半：報告書案確認依頼
- 2月末～3月上旬：報告書印刷発注
- 3月中旬：報告書納品、関係者に配布

以上