

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

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実用化が可能となった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理由のひとつである。

(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が考えられる。
- 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が考えられる。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、機械学習等による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とする。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。

人工知能技術を利用した支援システムのうち、製造販売業者が、市販後学習により生じる性能変化に伴う品質管理を行うことが難しいもの、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検討の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、臨床で使い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によって、承認の軽微な変更にあたるか、一部変更承認を要するものかを検討する必要があるため、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、承認審査においては、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

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、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。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の変化を段階的に行う場合が考えられるが、その場合には、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、使用者はネットワークを介して支援システム等と接続する等、人工知能は製造販売業者が管理した上で、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等製品の品質、有効性または安全性に影響を及ぼす場合、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も考えられる。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についての懸念が議論されている。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、支援システムの保守、設計・仕様上の不具合や故障等のトラブル対応においては、通常の医療機器と同様、製造販売業者が責任を負うことは変わらないが、支援システム等の使用目的、使用方法等を使用者に対して明らかにし、使用者に対するトレーニング等を実施する等、適正使用のために必要な方策を検討し、実施の実効性を担保する必要がある。

(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、それらを規定し、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、技術の進展等により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て必要な項目について内容を明確に示し、また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、市販後に支援システム等の使用に伴い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等)
- 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、その人工知能が使用者の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、次の項目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、以下のガイダンス等を参考に適切な管理を行い個人情報の保護に留意すること
 - 「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なため、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上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。

支援システム等を既存の画像撮影装置に接続して使用する場合、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条件を規定する必要がある、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智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使用した際に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、使用する人工智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に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「人工智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、市販後学習による性能変化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、上記リスクに加え、適応外で使用した際には、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、その時点での性能に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。

製造販売業者は、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され、確定診断、最終判断等、診療行為における最終判断は医師でなければならないことを使用者に明確に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

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ケース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支援システム等の場合、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、臨床上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が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