



Molecular Imaging CRO Network

## Micron's ViewPoint

---

医用画像診断支援システムの  
臨床評価：参照基準の重要性  
～医療AIの社会実装に向けて～

---

---



## 目次

|                     |   |
|---------------------|---|
| はじめに                | 3 |
| 医用画像診断支援システムの性能評価指標 | 4 |
| 専門家パネルによる参照基準の決定法   | 6 |
| おわりに                | 8 |
| 参考文献                | 9 |

### 免責事項

本文書に記載された内容は、予告無しに変更される場合があります。  
株式会社マイクロンは、本文書に関していかなる種類の保証（商用性および特定の目的への適合性の黙示の保証を含みますが、これに限定されません）も致しません。  
株式会社マイクロンは、本文書に含まれた誤謬に関しての責任や、本文書の提供、履行および使用に関して偶発的または間接的に起こる損害に対して、責任を負わないものとします。  
執筆者から事前に書面による許可を得ることなく、本文書のいかなる部分も複製、再販売または改変することを禁じます。

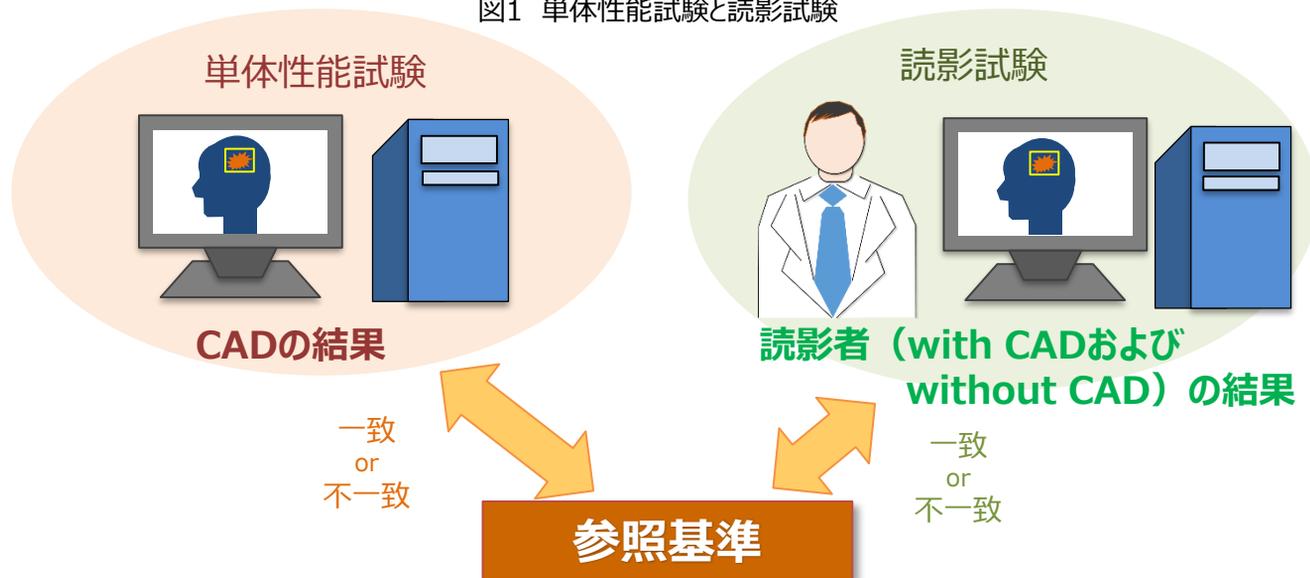
## はじめに

医用画像を定量的にコンピュータで解析し、疾患や病態の有無を検出および診断するコンピュータ検出／診断支援（CAD：Computer Aided Detection/Diagnosis）は、「第2の意見」として医師の診断を支援するコンピュータ診断支援ソフトウェアである。近年、人工知能（AI：Artificial Intelligence）を搭載したCAD（以下「AI搭載CAD」）が臨床現場に実装されつつあり、米国では30品目以上のAI搭載CADがFDAに認証／承認されており、本邦においても昨年にAI搭載CADである内視鏡画像診断支援ソフトウェアが初承認された。このように放射線画像や内視鏡画像、超音波画像などの医用画像を用いて医師の診断を補助するAI搭載CADは、国内外問わず活発に開発され、臨床現場に実装される時代が到来しつつある。

医用画像を用いて医師の診断を支援するシステム、すなわち医用画像診断支援システムを実臨床に導入する決定因子のひとつとして、当該システムの客観的な診断性能がある。加えて、この客観的な診断性能は、医療機器として認証／承認されるときに重要なデータにもなる。システム単体の性能を評価する単体性能試験や読影者がどの程度システムから影響を与えられるか評価する読影試験で診断性能を評価するが（図1）、当該試験における診断性能を評価する指標として、感度や特異度、受信者動作特性（ROC：Receiver Operating Characteristic）曲線下面積（AUC：Area Under the Curve）が一般的である。これらの評価指標は、現時点で最も正確度の高い診断法を基準とし算出するが、その基準となる診断法のことを参照基準※（Reference Standard）という。参照基準は、「その症例が間違いなく研究対象の疾患である」または「間違いなく疾患ではない」ということを証明するための証拠のことであり、参照基準の精度が低い場合、読影試験の信頼性が低い結果となってしまう。

本稿では、医用画像診断支援システムの読影試験における参照基準の決定について、その概論を記載する。

図1 単体性能試験と読影試験



※至適基準（Gold Standard）や真の標準（Standard of truth）とも呼ばれることもある。

（参考文献 1、2、3、4、5、6を引用）



## 医用画像診断支援システムの性能評価指標

医用画像診断支援システムの有用性の評価は、単独で医用画像診断支援システムを用いたときの対象疾患検出／診断能力だけでは十分でない場合が多い。すなわち、実臨床において医用画像診断システムがどの程度有用であるか証明することが重要である。

医用画像診断支援システムの臨床における有用性は、医師や医療従事者などの読影者による読影試験を行うことによって証明することが可能である。読影試験は、視覚評価による診断精度や読影時間の短縮能、疲労への影響について評価するが、本稿では視覚評価で得られる結果を中心に述べる（表1）。

| 指標   | 説明   |
|--|--|
| <b>感度</b><br>Sensitivity                                     | 参照基準が陽性であった場合の診断の結果が陽性（真陽性）である割合                             |
| <b>特異度</b><br>Specificity                                    | 参照基準が陰性であった場合の診断の結果が陰性（真陰性）である割合                             |
| <b>陽性的中率</b><br>Positive Predictive Value                    | 診断の結果が陽性であった場合の参照基準が陽性（真陽性）である割合                             |
| <b>陰性的中率</b><br>Negative Predictive Value                    | 診断の結果が陰性であった場合の参照基準が陰性（真陰性）である割合                             |
| <b>陽性尤度比</b><br>Positive likelihood ratio                    | 真陰性よりも真陽性である方が何倍診断検査で陽性になりやすいかを示す比                           |
| <b>陰性尤度比</b><br>Negative likelihood ratio                    | 真陰性よりも真陽性である方が何倍診断検査で陰性になりやすいかを示す比                           |
| <b>AUROC</b><br>Area under Receiver Operating Characteristic | 縦軸を真陽性率（感度）、横軸を偽陽性率（1－特異度）として閾値が取りうるすべての値においてプロットして得られた曲線の面積 |

表1 性能評価指標

（参考文献 5、6を引用）



## 医用画像診断支援システムの性能評価指標

感度や特異度のような診断性能の評価指標の算出のためには、以下のような2×2の四分割表を作成するのが基本である。

|          | 参照基準<br>陽性      | 参照基準<br>陰性       |                    |
|----------|-----------------|------------------|--------------------|
| 検査<br>陽性 | 真陽性<br>a        | 偽陽性<br>b         | 陽性的中率<br>$a/(a+b)$ |
| 検査<br>陰性 | 偽陰性<br>c        | 真陰性<br>d         | 陰性的中率<br>$d/(c+d)$ |
|          | 感度<br>$a/(a+c)$ | 特異度<br>$d/(b+d)$ |                    |

表2 四分割表

四分割表を基に分布図（図2）を作成し、閾値※を変化させ真陽性率と偽陽性率をプロットしたものがROC曲線（図3）である。ROC曲線下の面積であるAUROCは、旧来より読影試験で広く利用され、読影試験で信頼される評価指標と言われている。ROCの読影試験の種類や方法論等について、次稿以降のWhite Paperで詳細に説明する予定である。

四分割表の通り、感度や特異度、評価指標の算出は、参照基準が「真の患者の状態を表している」という大前提のもと議論が進められている。このように精度の高い参照基準は、読影試験の成功を左右するファクターのひとつである。

図2 両正規分布図

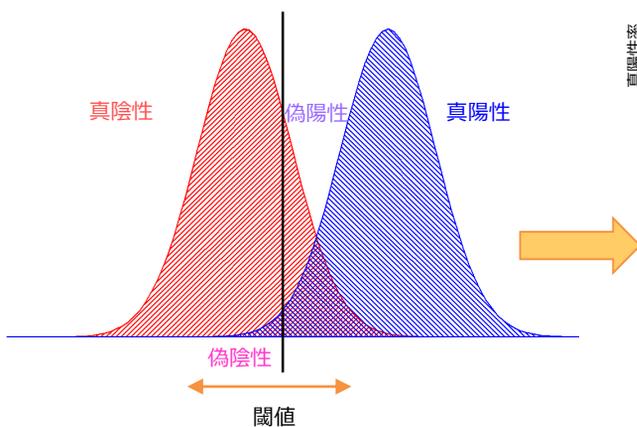
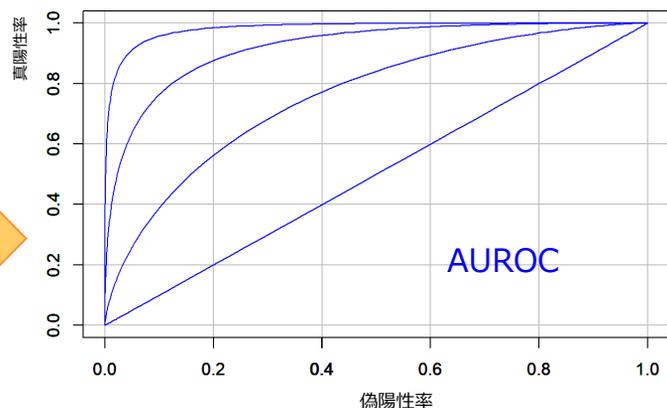


図3 ROC曲線



※ 医用画像診断システムの場合、閾値は確信度で代替される場合が多い。

(参考文献 7を引用)

## 専門家パネルによる参照基準の決定法

参照基準は、実施できる最善で妥当な方法で決定されなければならない。参照基準の決定手法は慎重にデザインしなければならない。参照基準の決定方法として以下の方法があげられる。

- 他の医用画像診断装置の出力結果
- 確立された臨床判断（例；病理所見、臨床検査結果）
- 画像検査によるフォローアップ結果
- 画像診断以外のフォローアップ結果
- 専門家パネル※による解釈

読影試験が後向き試験などで病理所見や臨床検査結果を入手できない場合、または明確な基準で参照基準を決定できる方法がない場合、独立した第三者の専門家パネルによる解釈により、参照基準が決定されることがある。

読影試験の参照基準を決定する専門家パネルは、以下の4つの要素を考慮して構築される。

### ① 委員の構成要素

1人の委員の評価結果より独立した複数名によって導出された評価結果は、より正確な結果の解釈を導く確率が高いため、専門家パネルは複数名により構成されることが多い。また、投票による決裁方式を用いる場合、可否同数を避けるため奇数人数で専門家パネルを構成するのが望ましい。

委員の臨床的背景も重要である。可能な限り対象疾患に関連した全領域の委員で構成することが望ましい。また、対象疾患の熟練度や経験年数も、委員を選定する検討要素の一つである。

※専門家パネル（Expert panel）とは、当該分野の専門家で構成された集団

（参考文献 4、5、8を引用）

## 専門家パネルによる参照基準の決定法

### ② 専門家パネルに提供する情報

専門家パネルに提供する情報は、読影試験のデザインに依存して設定される。どの情報を専門家パネルに提供するか、対象疾患に関連したガイドラインや系統的レビュー、Key paperなどの根拠を明確にする必要がある。情報の具体例として、画像情報に加え、患者の既往歴、合併症、治療歴、身体所見、血液検査結果などがあげられる。

一方、臨床情報によるバイアスを与えないため、盲検化も重要である。盲検化の種類として、完全盲検化や部分的盲検化、段階的盲検化、非盲検化があり、試験特性を考慮して選択する必要がある。

### ③ 決定プロセス

参照基準は、対象疾患の有無や重症度、診断確度のカテゴリー分類、病変の位置情報などで構成されるが、その決定方法は独立した複数名の委員による多数決が基本である。合議は他の委員の評価の影響を受ける可能性があり、基本的に避けた方が好ましい。合議の実施が避けられない状況においても、事前に個々の委員が独立評価し、診断結果を確定する。この委員個々の評価は、専門家パネル全体で評価する必要のないサブグループを決定する場合でも利用することができる。一部症例を専門家パネルによる評価症例から除外することにより、委員の労力を削減することができる。

独立した委員の評価結果が一致しない場合、評価結果の確定方法として、新しい委員の追加や情報の追加提供がある。いずれにせよ、不一致時の評価結果確定手順は試験開始前に決定しておかなければならない。

### ④ 専門家パネルのValidity

明確な決定基準のない参照基準の場合、結果の解釈は主観的であり正確でない可能性があるため、専門家パネルの評価結果は再現性があることが望ましい。委員間の評価のバラツキを回避するためには、可能な限り客観的な評価基準を設定する。必要に応じて、事前に委員のトレーニングを実施する方法もある。また、読影試験の信頼性を担保する手法として再現性の定量化がある。具体的には、カテゴリカルデータの場合には $k$ 係数、連続量データの場合には級内相関係数（ICC：Intraclass Correlation Coefficient）が利用可能である。

専門家パネルで留意すべき点として、多大な時間と労力を要してしまうことである。読影試験開始前に効率的で最適なデザインを設定することが期間の短縮を左右する。また、読影試験、特に後向き試験において、他の臨床試験と同様にバイアスのリスクがある。バイアスの混入により読影試験の結果が歪められると、誤った結果の解釈となり、適切な評価結果が得られない可能性がある。バイアスを排除した試験デザインの構築が重要である。読影試験におけるバイアスの種類およびその対応策について、次項以降のWhite Paperで詳細に説明する。

(参考文献 5、7、8、9を引用)



## おわりに

個々の患者に最適な治療を提供するPrecision Medicineの実現のために、より精度の高い診断検査が重要視されています。特にAIを利活用した画像診断は、深層学習を中心とした第三次AIブームのなかで特に期待されている重点分野のひとつです。

医用画像診断支援システムの診断精度を適切な方法で評価するためには、適切な臨床性能試験の確立が重要であり、それが市販承認申請データとして利用されるならば、より一層の合理的で妥当な試験デザインの構築が望まれます。

マイクロンは、国内はもとよりアジア圏のイメージングCROのリーディングカンパニーとして、画像を用いた臨床試験を多くサポートしています。イメージング臨床試験の支援機関として、試験依頼者のパートナーとなり、最適な臨床試験を提案しています。

AI搭載CADや医用画像診断支援システムの認証・承認に関するコンサルテーションおよびご相談に関するお問い合わせは、下記連絡先までお問い合わせください。

### 会社概要

東京本社 : 〒103-0027 東京都中央区日本橋一丁目5番3号 日本橋西川ビル4階  
TEL.03-6262-2830 FAX.03-6262-2831

大阪支社 : 〒532-0003 大阪市淀川区宮原四丁目5-36 セントラル新大阪ビル6階  
TEL.06-6399-0007 FAX.06-6399-0008

名古屋事業所 : 〒474-8511 愛知県大府市森岡町七丁目430番地  
国立研究開発法人 国立長寿医療研究センター 南病棟2階  
TEL.0562-46-2105 FAX.0562-46-2106

事業内容

- ①イメージング技法の活用を機軸とした医薬品、診断薬、バイオマーカーの開発支援
- ②臨床開発支援  
(治験モニタリング、品質管理、画像解析、画像ハンドリング、読影支援等)
- ③PET標識化合物、PET動物評価、PET薬剤製造支援
- ④薬事コンサルティング (信頼性保証体制の構築、PET治験薬 GMP、c-GMP等)
- ⑤医薬品開発に係るコンサルティング

ホームページ : <https://micron-kobe.com>

お問合せEmail : [imagingbiomarker@micron-kobe.com](mailto:imagingbiomarker@micron-kobe.com)

## 参考文献

1. 「次世代医療機器評価指標の公表について 別紙4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」(令和元年5月23日付け薬生機審発0523第2号)
2. 「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月国立医薬品食品衛生研究所)
3. [https://www.amed.go.jp/news/release\\_20181210.html](https://www.amed.go.jp/news/release_20181210.html)
4. FDA. Clinical Performance Assessment: Considerations for Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data - Premarket Approval (PMA) and Premarket Notification [510(k)] Submissions  
<https://www.fda.gov/media/77642/download>
5. 「コンピュータ診断支援装置の性能評価開発ガイドライン2015（手引き）」(平成27年12月経済産業省／国立研究開発法人日本医療研究開発機構)
6. 杣取恵太・坂本次郎・時田棕子・鈴木彩夏・国里愛彦. 診断精度研究の系統的レビューとメタアナリシス. 専修人間科学論集 心理学篇 Vol. 6, No. 1, pp. 41～58, 2016
7. 「診断用放射性医薬品の臨床評価方法に関するガイドライン」(平成24年6月11日付け薬食審査発0611第1号)
8. Bertens LC, Broekhuizen BD, Naaktgeboren CA, Rutten FH, Hoes AW, van Mourik Y, Moons KG, Reitsma JB. Use of expert panels to define the reference standard in diagnostic research: a systematic review of published methods and reporting. PLoS Med. 2013 Oct;10(10):e1001531.
9. Klaus Gottlieb, Fez Hussain. Voting for Image Scoring and Assessment (VISA) - theory and application of a 2 + 1 reader algorithm to improve accuracy of imaging endpoints in clinical trials. BMC Med Imaging 2015 Feb 19;15:6.