

## 日本医療研究開発機構 医療機器開発推進研究事業 事後評価報告書

公開

### I 基本情報

研究開発課題名：

(日本語) 画像認識 AI による細菌感染症の菌種分類精度向上と抗菌薬適正使用に関する臨床研究  
(英語) Study to improve the accuracy of image recognition AI for bacterial species classification and proper use of antimicrobial.

研究開発実施期間：令和3年4月1日～令和6年3月31日

研究開発代表者 氏名：(日本語) 山元 佳  
(英語) Kei Yamamoto

研究開発代表者 所属機関・部署・役職：  
(日本語) 国立国際医療研究センター国際感染症センター 医師  
(英語) Disease Control and Prevention Center, National Center for Global Health and Medicine, physician

### II 研究開発の概要

研究開発の成果およびその意義等

和文：2 ページ以上

英文：1 ページ程度

## <研究の概要>

本開発研究は、感染症診療において早期に結果を得ることが可能なグラム染色所見について、感染症診療に長けた医療者が不在の状態でも一定の所見判読成績を示すことができるような画像 AI モデルをカーブジェン株式会社と共同で開発し、その有効性を検証するために実施された。

1) 教師データとなるグラム染色画像を 2 医療機関のグラム染色レジストリより収集し、AI モデルを構築、2) グラム染色で得られた情報を抗菌薬の選定に活かすために地域別あるいは施設別の抗菌薬感受性表（アンチバイオグラム）を参照する辞書機能を開発、3) 開発 AI モデルについてグラム染色の判読に長けた医療者との非劣性試験を行い、その有効性を評価する形で実施した。

## <研究開発の成果>

- 尿検体、血液培養検体を用いた AI モデルを開発し、事前検証において主要評価項目となる菌形態分類予測に際して 95%前後の正確度、95%を超えるマクロ平均感度を示した。
- グラム染色の結果を抗菌薬選定に活かすためのアンチバイオグラム参照システムを開発し、グラム染色の結果に応じた抗菌薬選定に活かす補助機能を実装した。
- 尿検体を用いて、本研究の開発品の単体性能試験を実施し、正確度についてグラム染色判読に長けた医療職との非劣性試験においては非劣性を示した（正確度：感染症専門職 83.0% [95%CI:81.6 to 84.3%] vs CAD 87.9% [83.7 to 91.3%]）。副次的な評価項目となるが、グラム染色を用いた抗菌薬選定により重要な情報を含む詳細な菌種分類においては開発品が医療職と比べて優越性を示す結果となった（正確度：感染症専門職 34.3% [95%CI: 32.6-36.0%] vs CAD 46.1% [40.4-51.8%]）。
- 尿検体について本開発品の薬事申請を行った。
- 血液培養検体を用いて、本研究の開発品の単体性能試験を実施した。また、グラム染色判読に長けた医療職との非劣性試験を行った。
- 血液培養検体について本開発品の薬事申請を検討中である。

### (1) 尿検体および血液検体グラム染色画像収集

### (2) 画像 AI モデルの構築

令和 3 年度から令和 5 年度にかけて国立国際医療研究センター（NCGM）および神戸大学病院（KUH）のグラム染色レジストリより、教師データあるいは検証用データとなるグラム染色の撮像画像ファイルを尿検体で 17,927 枚（1,639 症例）、血液培養検体で 26,350 枚（2,340 症例）を作成し、カーブジェン株式会社に提供した。

iPhone を撮像デバイスとしてバーミー法で染色した尿検体、血液培養検体における画像 AI モデルを構築した。教師データ症例を含まない検証データに対する各モデル菌形態分類に基づいた分類性能は以下の通りであった。

尿検体：正確度 94.5%、マクロ平均感度 95.4%、F1 値 95.07

血液培養検体（好気ボトル）：正確度 96.5%、マクロ平均感度 97.6%、F1 値 97.1

血液培養検体（嫌気ボトル）：正確度 96.3%、マクロ平均感度 96.1%、F1 値 96.1

(3) アンチバイオグラム連動抗菌薬選  
定システム開発

アンチバイオグラムを用いた抗菌薬選  
定のための補助情報参照システムの開  
発・テストは令和 4 年度には完了し、  
国立国際医療研究センター、神戸大学  
以外の複数医療機関での評価や機能の  
チェックを完了し、右の図 1 のような  
画面遷移で情報を表示することができ  
ることを確認した。



図 1. 本開発品の使用方法と画面遷移

(4) 有用性評価試験

<尿検体>2023 年 4 月から感染症専門医療職、CAD による判読  
試験を開始し、2023 年 5 月に解析報告を行った。単体性能試験  
における全体の正確度は 菌形態で 83.1%(95%CI: 82.4 to  
83.9%) (図 2) で、詳細な菌種分類で 65.5%(95%CI: 64.5 to  
66.4%)であった。

また予測の正確性は確信度が上昇すると増す傾向にあること  
も明らかになった (AUC 0.828)。

感染症専門職との非劣性試験において、菌形態の予測、詳細  
な菌種分類の予測とも開発した CAD は非劣性を示した(正確度:  
感染症専門職 83.0% [95%CI:81.6 to 84.3%] vs CAD 87.9%  
[83.7 to 91.3%]、図 3)。マクロ平均感度は 70.5 vs 79.1%で  
あり各分類の感度(再現率)、陽性的中率(適合率)は図 4 の  
ようになった。

|    |        | 正解  |      |      |     |        |     |
|----|--------|-----|------|------|-----|--------|-----|
|    |        | GNC | GNR  | GPC  | GPR | Others | 真菌  |
| 予測 | GNC    | 39  | 1    | 0    | 0   | 0      | 0   |
|    | GNR    | 63  | 1705 | 25   | 5   | 474    | 0   |
|    | GPC    | 2   | 0    | 1060 | 13  | 200    | 0   |
|    | GPR    | 0   | 1    | 49   | 207 | 247    | 1   |
|    | Others | 148 | 198  | 219  | 44  | 5283   | 7   |
|    | 真菌     | 0   | 1    | 3    | 1   | 36     | 272 |

図 2. 単体性能試験における菌形態分類における  
混同行列

|    |      | 正解  |      |     |     |     |      |     |
|----|------|-----|------|-----|-----|-----|------|-----|
|    |      | GNC | GNR  | GPC | GPR | 複数菌 | 菌体なし | 真菌  |
| 予測 | GNC  | 63  | 6    | 26  | 4   | 5   | 0    | 0   |
|    | GNR  | 52  | 1207 | 25  | 15  | 41  | 3    | 0   |
|    | GPC  | 26  | 8    | 770 | 35  | 33  | 0    | 10  |
|    | GPR  | 0   | 11   | 8   | 81  | 7   | 0    | 0   |
|    | 複数菌  | 0   | 17   | 27  | 10  | 90  | 0    | 0   |
|    | 菌体なし | 39  | 10   | 37  | 34  | 2   | 174  | 16  |
|    | 真菌   | 0   | 1    | 7   | 1   | 2   | 3    | 154 |

|    |      | 正解  |     |     |     |     |      |    |
|----|------|-----|-----|-----|-----|-----|------|----|
|    |      | GNC | GNR | GPC | GPR | 複数菌 | 菌体なし | 真菌 |
| 予測 | GNC  | 11  | 0   | 0   | 0   | 0   | 0    | 0  |
|    | GNR  | 2   | 121 | 0   | 0   | 4   | 0    | 0  |
|    | GPC  | 0   | 0   | 82  | 3   | 3   | 0    | 0  |
|    | GPR  | 0   | 0   | 1   | 12  | 3   | 0    | 0  |
|    | 複数菌  | 0   | 5   | 5   | 0   | 8   | 0    | 0  |
|    | 菌体なし | 5   | 0   | 2   | 3   | 0   | 18   | 1  |
|    | 真菌   | 0   | 0   | 0   | 0   | 0   | 0    | 17 |

図 3. 非劣性試験における菌形態分類における混同行列 (左: 感染症専門職、右: CAD)

|      | CAD    |        |       | 感染症専門医療職 |       |       |
|------|--------|--------|-------|----------|-------|-------|
|      | 適合率    | 再現率    | F1 値  | 適合率      | 再現率   | F1 値  |
| 菌体なし | 62.1%  | 100.0% | 76.6% | 55.8%    | 96.7% | 70.7% |
| 複数菌  | 44.4%  | 44.4%  | 44.4% | 62.5%    | 50.0% | 55.6% |
| GPC  | 93.2%  | 91.1%  | 92.1% | 87.3%    | 85.6% | 86.4% |
| GPR  | 75.0%  | 66.7%  | 70.6% | 75.7%    | 45.0% | 56.4% |
| GNR  | 95.3%  | 96.0%  | 95.7% | 89.9%    | 95.8% | 92.7% |
| GNC  | 100.0% | 61.1%  | 75.9% | 60.6%    | 35.0% | 44.4% |
| 真菌   | 100.0% | 94.4%  | 97.1% | 91.7%    | 85.6% | 88.5% |

図 4. 非劣性試験における菌形態分類毎の適合率、再現率、F1 値（調和平均）

なお、抗菌薬選定により重要な予測となる詳細な菌種分類の予測については優越性を示す結果となった（正確度：感染症専門職 34.3% [95%CI: 32.6-36.0%] vs CAD 46.1% [40.4-51.8%]）。

<血液培養検体>2024 年 2 月に本試験を実施し、2023 年 3 月に解析報告を行った。

#### (5) 医療機器プログラム新規承認に向けた活動

尿検体については、有用性評価試験の結果を基に 2023 年 5 月に医療機器プログラム新規承認申請を行った。

血液培養検体については、承認申請を検討中である。

### <Outline of Research>

This study was conducted in collaboration with CarbGeM Inc. to develop and validate an computer-aided diagnosis (CAD) with imaging AI model for Gram stain findings, which can be used to obtain to select empiric antimicrobials for bacterial infection, and to demonstrate the effectiveness of CAD even in the absence of medical personnel skilled in the treatment of infectious diseases.

CAD was developed by 1) collecting Gram-stained images as training data from Gram-stained registries of two medical institutions, 2) developing a dictionary function to refer to antibiograms by region or facility to utilize the information obtained from Gram staining in selecting antimicrobial agents, and 3) developing CAD that can be used for Gram stain decipherment in the absence of a medical practitioner. The CAD was evaluated in a non-inferiority trial against medical professionals skilled in Gram stain decipherment.

### <Result of Research>

- An CAD using urine and blood culture specimens was developed, and preliminary validation showed an accuracy of around 95% and a macro-average sensitivity of over 95% in predicting bacterial morphology.
- An antibiogram reference system was developed to utilize the results of Gram staining in the selection of antimicrobial agents, and an auxiliary function for selecting antimicrobial agents according to the results of Gram staining was implemented.
- A stand-alone performance test of the developed product in this study was conducted using urine specimens, it showed non-inferiority in a non-inferiority test with medical professionals skilled in Gram stain deciphering (accuracy: 83.0% [95%CI: 81.6 to 84.3%] for medical professionals vs. 87.9% [83.7 to 91.3%] for CAD). As a secondary endpoint, the developed products showed superiority over the medical profession in the detailed bacterial species classification, which includes important information due to the selection of empiric antimicrobial agents using Gram staining (accuracy: infectious disease professionals 34.3% [95%CI: 32.6-36.0%] vs CAD 46.1% [40.4-51.8%]).
- An application was filed for regulatory approval for CAD for urine specimens.
- A stand-alone performance test of the developed product was conducted on blood culture specimens. Non-inferiority tests with medical professionals skilled in Gram stain decipherment was also conducted.
- Regulatory application is under consideration for this development product for blood culture specimens.

ここまでを総括報告として AMED のホームページに掲載