多讀者多病例（MRMC）實驗設計於 AI 輔助肺結節影像判讀成效評估之應用

Evaluating AI-Assisted Pulmonary Nodule Imaging Interpretation Using a Multi-Reader Multi-Case (MRMC) Study Design

Internship Institution：DeepRad.AI Inc.

Student: Yi-Hsin Chang **Advisor**: Charlotte Wang **Internship Supervisor**: Roger Chang

1. Background

- Lung cancer is one of the leading causes of cancer-related deaths worldwide.
- Early detection of pulmonary nodules is critical to improve treatment outcomes and survival.
- Chest CT is the standard tool, but:
  - Large image volumes
  - Variable nodule appearance

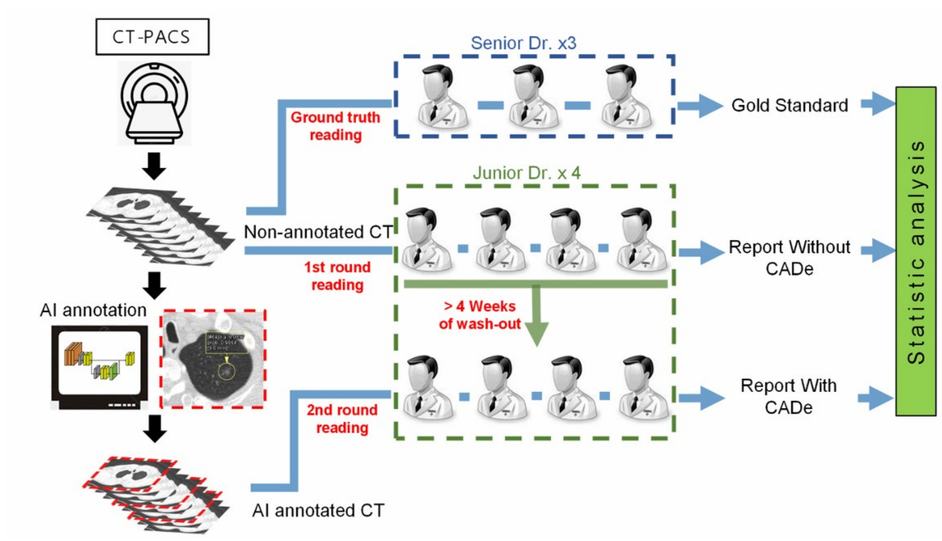
  → High workload and risk of misinterpretation for radiologists.
- AI in medical imaging: shows strong potential, but rigorous evaluation methods are needed to demonstrate real clinical value and meet FDA standards.

2. MRMC Study Design

- FDA guidance recommends Multi-Reader Multi-Case (MRMC) as the standard approach for evaluating CADe systems.
- Core principle:
  - Multiple radiologists (readers)
  - Diverse patient cases

  → Simulates real-world practice and quantifies AI's clinical impact.
- Advantages over small-scale studies:
  - Captures reader variability
  - Captures case heterogeneity

  → More accurate assessment of AI's diagnostic benefit.
- Typical setup:
  - Cases read with and without AI assistance
  - Randomized readers, cases, and order to reduce bias
  - ≥4-week interval for repeated cases to avoid recall bias
- Evaluation metrics: sensitivity, specificity, ROC, FROC, LROC
- FDA recommendation: ROC summary measures as primary metric; sensitivity and specificity as secondary metrics.



[Figure 1] MRMC study design applied to AI-assisted pulmonary nodule interpretation (Chao, H. S. et al., 2023).

3. Applications in Clinical Studies

- Chao et al. (2023) [1]
  - 8 radiologists, 200 chest CT cases (with vs. without AI)
  - Metrics: sensitivity, specificity, ROC
  - Results: Improved overall accuracy, greatest gains in detecting small nodules (4–5 mm) and ground-glass nodules (GGNs).
- Kim et al. (2022) [2]
  - 12 readers, 300 cases; malignancy risk scored on a 100-point scale
  - Results: AI improved accuracy in malignancy risk assessment for indeterminate nodules; enhanced inter-reader consistency in risk stratification and management.
- Hsu et al. (2021) [3]
  - 6 readers, 150 chest CT cases (340 confirmed nodules)
  - Three conditions: unaided, second-read CADe, concurrent CADe
  - Results: Both AI modes improved detection; concurrent CADe also reduced reading time, improving workflow efficiency.

Paper	Chao et al., 2023	Kim et al., 2022	Hsu et al., 2021
Imaging / Task	Chest CT / Pulmonary nodule detection	Chest CT ／ Malignancy risk classification	Chest CT / Pulmonary nodule detection
Readers / Cases	8／200	12／300	6／150
Company	V5med	Optellum	ClearReadCT
Evaluation Metrics	Se, Sp, ROC-AUC, LROC-AUC, FROC	Mean ROC-AUC (OR/DBM), Se, Sp, inter-reader agreement	Se, Sp, ROC-AUC, reading time
Conclusion	Improved overall sensitivity and higher mean AUC	Increased mean AUC and improved risk stratification	Improved AUC and sensitivity, with significantly reduced reading time

[Table 1] Research overview, metrics, and key findings from past studies

4. Conclusion

- Comparison of Chao (2023), Kim (2022), and Hsu (2021) shows broad and practical use of MRMC in AI evaluation.
- MRMC strengths:
  - Handles variability among readers and heterogeneity across cases
  - Provides realistic, representative performance evaluation in clinical settings.
- Practical significance:
  - Aligns with FDA recommendations
  - Frequently adopted by companies to validate AI-assisted systems.
- Takeaway:
  - MRMC is not just a theoretical model—it's a widely trusted, practical tool in both research and industry.
  - Plays an essential role in ensuring objective and reliable AI performance evaluation.

5. References

1. Chao, H. S., Tsai, C. Y., Chou, C. W., Shiao, T. H., Huang, H. C., Chen, K. C., Tsai, H. H., Lin, C. Y., & Chen, Y. M. (2023). Artificial Intelligence Assisted Computational Tomographic Detection of Lung Nodules for Prognostic Cancer Examination: A Large-Scale Clinical Trial. *Biomedicines*, *11*(1), 147.
2. Kim, R. Y., Oke, J. L., Pickup, L. C., Munden, R. F., Dotson, T. L., Bellinger, C. R., Cohen, A., Simoff, M. J., Massion, P. P., Filippini, C., Gleeson, F. V., & Vachani, A. (2022). Artificial Intelligence Tool for Assessment of Indeterminate Pulmonary Nodules Detected with CT. *Radiology*, *304*(3), 683–691.
3. Hsu, H. H., Ko, K. H., Chou, Y. C., Wu, Y. C., Chiu, S. H., Chang, C. K., & Chang, W. C. (2021). Performance and reading time of lung nodule identification on multidetector CT with or without an artificial intelligence-powered computer-aided detection system. *Clinical radiology*, *76*(8), 626.e23–626.e32.