

Thyroid Suite

Evidence Book

Contents

Summary of Evidence	03
Improved characterization for all lexicon descriptors	04
Improved TI-RADS level agreement across all levels	05
AI diagnostics match clinical expertise	06

Summary of Evidence

1. Improved characterization for all lexicon descriptors

- **DeepHealth's AI improves radiologist ability to correctly characterize nodules¹**

On average, between unaided analysis and analysis aided by DeepHealth's AI, characterization improved by 7 percentage points across 5 categories. The AI's stand-alone performance was on par with that of the experts.

2. Improved TI-RADS level agreement across all levels

- **DeepHealth's AI helps radiologist assess levels more accurately¹**

On average, between unaided analysis and analysis aided by DeepHealth's AI, level agreement improved by 9 percentage points across 5 categories, with the highest gains at the most critical TR levels: TR-3, TR-4, and TR-5. The performance of every reader improved.

3. AI diagnostics match clinical expertise

- **Agreement between experts is comparable to agreement with DeepHealth's AI²**

Comparing DeepHealth's standalone AI assessments with that of radiologists, the level of agreement matched or exceeded the agreement reached between radiologists themselves in an MRMC study.

¹ Data on File: Summary of Performance Testing Studies | Thyroid 510(k) clearance

² Data on File: Evaluation of AI-enhanced Thyroid Ultrasound Reporting Software for Improved Diagnostic Consistency

Improved characterization for all lexicon descriptors

Intro

We evaluated 1) the standalone performance of our AI models and 2) that of experienced radiologists with and without the aid of DeepHealth's AI, comparing outputs to ground truth labels.

Methods

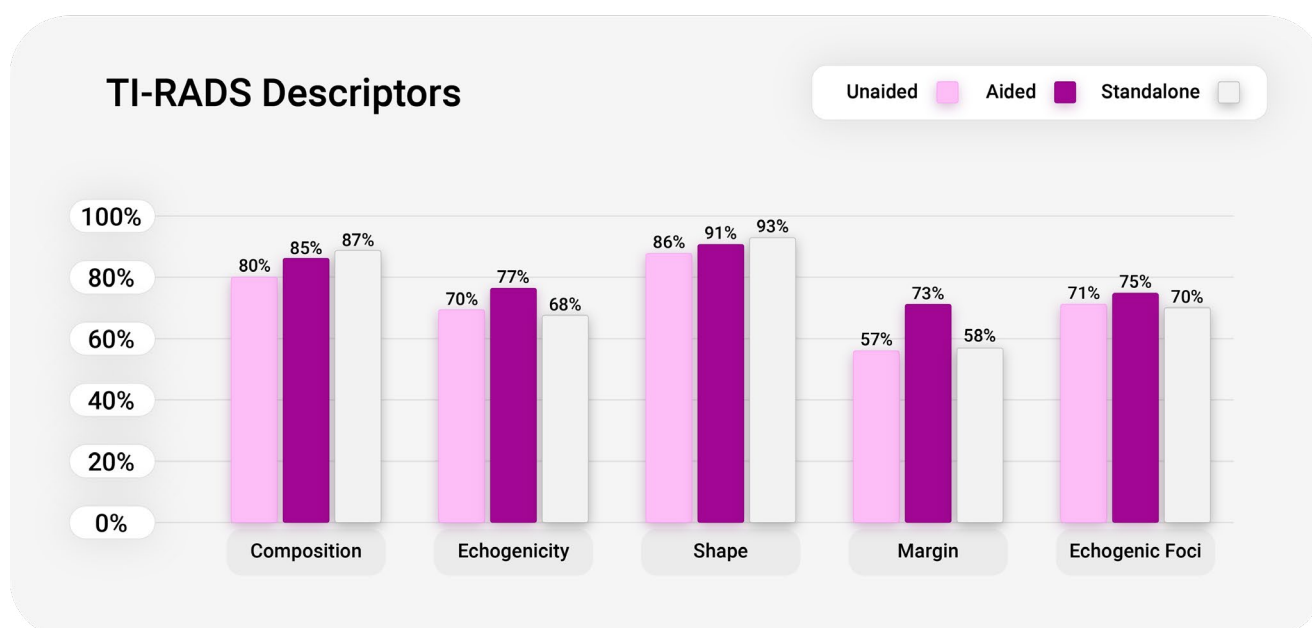
Ground truth labels for ACR TI-RADS lexicon descriptors were based on the labels of two expert US-board certified radiologists and an adjudicator. In a multi-reader multi-case (MRMC) study, 18 experienced radiologists read 600 cases twice, with and without the aid of DeepHealth's AI.

Results

On average, between unaided analysis and analysis aided by DeepHealth's AI, characterization of TI-RADS descriptors improved by 7 percentage points across 5 categories (between 4 and 16 percentage points). The AI's standalone performance was on par with that of the experts.

Conclusion

DeepHealth's AI improves radiologist ability to correctly characterize nodules with improved characterization for all lexicon descriptors.



¹ Data on File: Summary of Performance Testing Studies | Thyroid 510(k) clearance

Improved TI-RADS level agreement across all levels

Intro

We evaluated 1) the standalone performance of our AI models and 2) that of experienced radiologists with and without the aid of DeepHealth's AI, comparing outputs to ground truth labels.

Methods

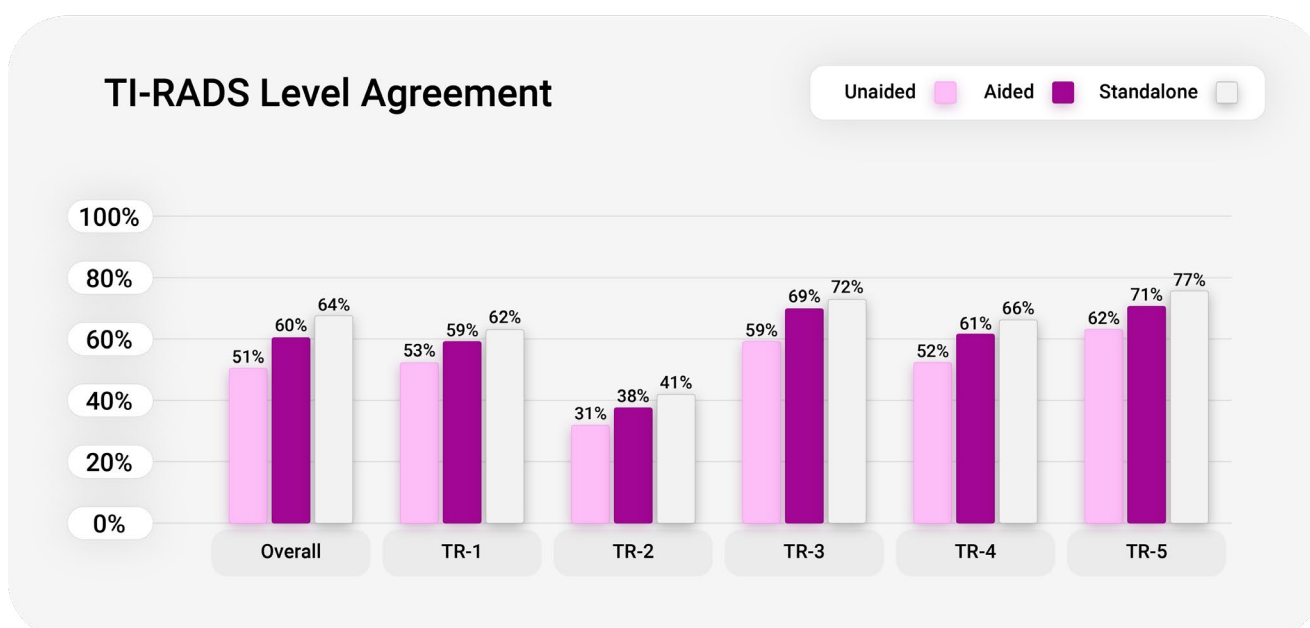
Ground truth labels for ACR TI-RADS levels were based on the labels of two expert US-board certified radiologists and an adjudicator. In a multi-reader multi-case (MRMC) study, 18 experienced radiologists read 600 cases twice, with and without the aid of DeepHealth's AI.

Results

On average, between unaided analysis and analysis aided by DeepHealth's AI, level agreement improved by 9 percentage points across 5 categories, with the highest gains at the most critical TR levels: TR-3, TR-4, and TR-5. The performance of every experienced radiologist in the study also improved, indicating that sonographers would likely experience an even higher level of accuracy when aided by DeepHealth's AI.

Conclusion

DeepHealth's AI helps radiologist assess levels more accurately with improved TI-RADS level agreement across all levels.



AI diagnostics match clinical expertise

Intro

We compared the level of agreement between our AI models and radiologists with the general level of agreement between radiologists themselves, representing multiple healthcare institutions and radiology practices from the United States and Canada.

Methods

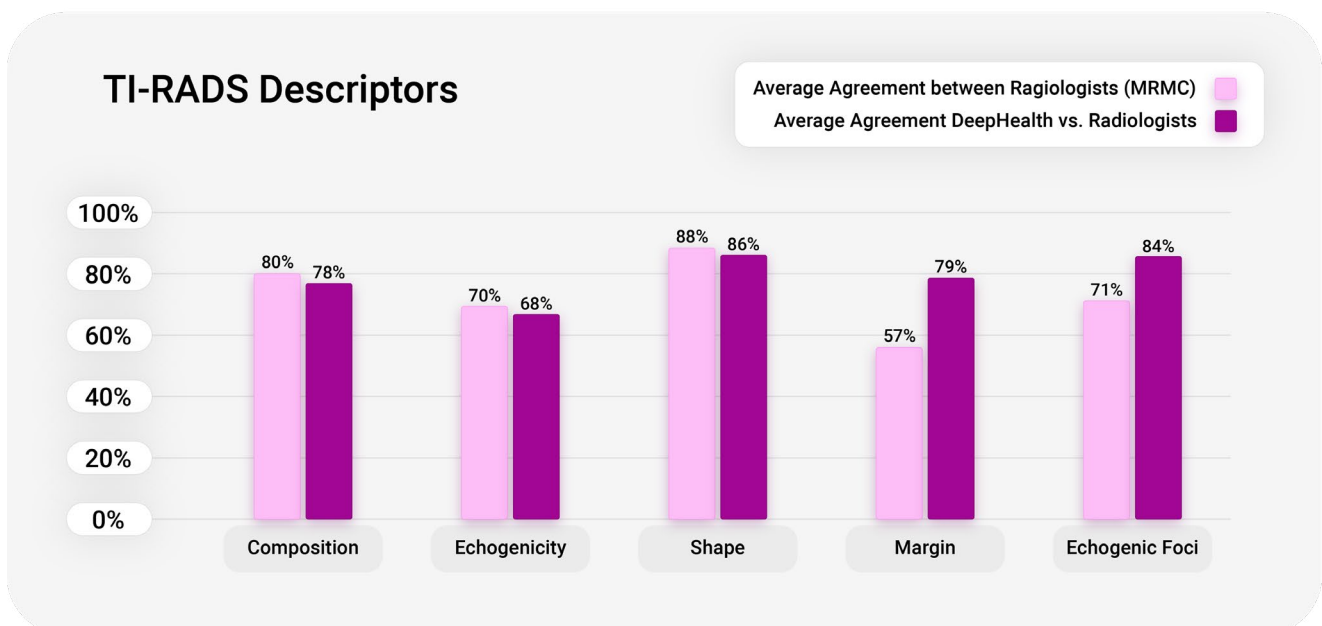
A retrospective analysis was performed on 449 thyroid nodules across 173 ultrasound cases, with participation from 6 institutions in the United States and Canada, including both major health systems and private radiology practices. DeepHealth's AI classified nodules according to ACR TI-RADS descriptors, with ground truth established by the radiologist's historical assessments. The agreement rates between the AI and radiologists were calculated and also compared to results from a recent multi-reader multi-case study (MRMC).

Results

It was observed that agreement between the software and radiologist interpretations was consistently high showing the robustness of the AI models on unseen clinical data.

Conclusion

DeepHealth's performance matches clinical expertise.



³ Data on File: Evaluation of AI-enhanced Thyroid Ultrasound Reporting Software for Improved Diagnostic Consistency

For more information, visit www.deephealth.com

deephealth