

Evaluating AI Performance in Abdominal Aortic Aneurysm Detection: Metrics and Clinical Implications

Rasit Dinc

Rasit Dinc Digital Health & AI Research

Published: September 6, 2025 | AI in Healthcare

DOI: [10.5281/zenodo.17996411](https://doi.org/10.5281/zenodo.17996411)

Abstract

Learn how AI performance metrics like sensitivity, specificity, and AUC evaluate accuracy in detecting Abdominal Aortic Aneurysm for reliable clinical use.

Evaluating AI Performance in Abdominal Aortic Aneurysm Detection: Metrics and Clinical Implications

Artificial Intelligence (AI) has emerged as a transformative tool in medical imaging, offering promising advancements in the detection and diagnosis of life-threatening conditions such as Abdominal Aortic Aneurysm (AAA). AAA, characterized by the abnormal dilation of the abdominal aorta, poses significant risks including rupture and sudden death if left undiagnosed or untreated. Accurate and timely detection is therefore paramount. AI algorithms, particularly those based on deep learning, have demonstrated remarkable potential in automating AAA detection from imaging modalities such as computed tomography angiography (CTA) and ultrasound.

To evaluate the clinical utility and reliability of AI models in AAA detection, it is essential to understand and apply rigorous performance metrics. This article explores key AI evaluation metrics, their clinical significance, current research evidence, practical applications, challenges, and future directions in AAA care.

Understanding AI Performance Metrics for Abdominal Aortic Aneurysm (AAA) Detection

Confusion Matrix: The Foundation of AI Evaluation

The confusion matrix is the cornerstone of classification model evaluation. It compares AI-predicted outcomes with ground truth clinical diagnoses, categorizing results into four groups:

Actual / Predicted	AAA+ (Positive)	AAA- (Negative)	-----	-----
-----	AAA+ (Positive)	True Positive (TP): 95	False Negative	-----

(FN): 5 | | **AAA- (Negative)** | False Positive (FP): 10 | True Negative (TN): 90 |

- **True Positive (TP):** Cases where the AI correctly identifies the presence of AAA.
- **False Negative (FN):** Cases where the AI fails to detect AAA (missed diagnosis).
- **False Positive (FP):** Cases where the AI incorrectly flags non-AAA as AAA.
- **True Negative (TN):** Cases where the AI correctly identifies absence of AAA.

This matrix forms the basis for calculating several critical performance metrics.

Key Performance Metrics and Their Clinical Relevance

1. **Sensitivity (Recall):** Sensitivity measures the AI's ability to correctly detect true AAA cases and is crucial to minimize missed diagnoses that could result in catastrophic outcomes.
$$\text{Sensitivity} = \frac{TP}{TP + FN} = \frac{95}{95 + 5} = 95\%$$
 A sensitivity of 95% indicates that the AI detects 95 out of 100 true AAA cases.
2. **Specificity:** Specificity evaluates the AI's ability to correctly identify patients without AAA, reducing false alarms and unnecessary interventions.
$$\text{Specificity} = \frac{TN}{TN + FP} = \frac{90}{90 + 10} = 90\%$$
 A specificity of 90% means the AI accurately excludes 90 out of 100 non-AAA cases.
3. **Positive Predictive Value (PPV):** PPV reflects the probability that patients flagged positive by AI truly have AAA, informing clinical decision confidence.
$$\text{PPV} = \frac{TP}{TP + FP} = \frac{95}{95 + 10} = 90.5\%$$
4. **Negative Predictive Value (NPV):** NPV indicates the likelihood that patients classified as negative are truly free of AAA.
$$\text{NPV} = \frac{TN}{TN + FN} = \frac{90}{90 + 5} = 94.7\%$$

ROC Curve and Area Under the Curve (AUC): Measuring Discriminative Ability

- **Receiver Operating Characteristic (ROC) Curve:** The ROC curve illustrates the trade-off between sensitivity and (1-specificity) across various classification thresholds. This visualization helps clinicians and developers select optimal thresholds balancing false positives and negatives.
- **Area Under the ROC Curve (AUC):** AUC provides a single scalar metric summarizing overall model discrimination, ranging from 0.5 (random guessing) to 1.0 (perfect accuracy). An AI model for AAA detection with an AUC of 0.96 demonstrates excellent ability to differentiate between aneurysmal and non-aneurysmal aortas.

AUC Value	Interpretation	0.9 - 1.0	Excellent
0.8 - 0.9	Good	0.7 - 0.8	Fair
0.6 - 0.7	Poor	0.5	No better than chance

Clinical Implications of AI Performance in AAA Detection

The integration of AI into clinical workflows hinges on the balance between sensitivity and specificity:

- **High Sensitivity (95%)** minimizes the risk of missed AAAs, which is critical given the asymptomatic nature of aneurysms until rupture. - **High Specificity (90%)** reduces false positives, thereby decreasing unnecessary diagnostic procedures, patient anxiety, and health care costs. - An **AUC of 0.96** confirms the AI system's robustness, supporting its use as a reliable adjunct diagnostic tool.

Despite high performance, residual false negatives (5%) and false positives (10%) necessitate that AI outputs be interpreted in conjunction with clinical judgment, imaging findings, and patient history. AI should augment—not replace—clinician expertise, acting as a decision support system that enhances diagnostic accuracy and workflow efficiency.

Research Evidence Supporting AI in AAA Detection

Recent peer-reviewed studies corroborate the utility of AI in AAA detection:

- A 2023 multicenter study involving over 5,000 CT angiograms demonstrated that deep learning algorithms achieved sensitivity and specificity upwards of 93% and 89%, respectively, aligning with the metrics discussed above. - Research published in *Radiology: Artificial Intelligence* reported that AI-assisted screening improved early AAA detection rates by 20%, leading to timely surgical referrals and reduced rupture incidence. - Meta-analyses reveal that AI models consistently outperform traditional rule-based algorithms and manual measurements in detecting subtle aneurysm morphologies and predicting growth trajectories.

Such evidence underscores AI's potential to improve population-level AAA screening programs, especially in resource-constrained settings where specialist radiologists may be scarce.

Practical Applications and Integration in Healthcare

- **Screening Programs:** Automated AAA detection algorithms can be embedded in routine abdominal imaging, flagging high-risk patients for further evaluation. - **Clinical Decision Support:** AI tools provide real-time alerts to radiologists and vascular surgeons, enhancing diagnostic confidence and reducing oversight. - **Surgical Planning:** AI can assist in precise aneurysm sizing and morphology assessment, critical for endovascular repair planning. - **Telemedicine:** AI-powered platforms enable remote AAA screening and monitoring, expanding access to underserved populations.

Challenges and Limitations

Despite promising results, several challenges remain:

- **Data Variability:** AI models trained on specific populations or imaging protocols may underperform when generalized to diverse clinical settings. - **Interpretability:** Black-box AI models lack transparency, complicating

clinician trust and regulatory approval. - **False Positives/Negatives:** Even low rates can lead to clinical consequences; thus, continuous model refinement and validation are essential. - **Integration Barriers:** Workflow disruptions, interoperability issues, and cost constraints can impede adoption.

Ethical considerations around patient data privacy, informed consent, and algorithmic bias must also be addressed as AI becomes more widespread in AAA care.

Future Directions

Future research and development in AI for AAA detection should focus on:

- **Multimodal Data Integration:** Combining imaging with clinical, genetic, and biochemical markers to improve predictive accuracy. - **Explainable AI:** Developing interpretable models that provide rationale behind predictions to enhance clinician acceptance. - **Prospective Clinical Trials:** Validating AI tools in real-world settings to assess impact on patient outcomes, cost-effectiveness, and workflow efficiency. - **Personalized Surveillance:** Leveraging AI to tailor monitoring intervals and intervention thresholds based on individual risk profiles.

Advances in federated learning and privacy-preserving AI may also facilitate collaborative data sharing while maintaining patient confidentiality.

Summary of AAA AI Performance Metrics

Metric	Value	Clinical Meaning
Sensitivity	95%	Detects most AAA cases, minimizing missed diagnoses
Specificity	90%	Accurately excludes non-AAA cases, reducing false alarms
Positive Predictive Value	90.5%	High confidence in positive AI predictions
Negative Predictive Value	94.7%	Reliable exclusion of AAA in negative predictions
AUC	0.96	Excellent overall discrimination ability between AAA and non-AAA

Frequently Asked Questions

Q1: Why are sensitivity and specificity crucial in AAA AI screening? Sensitivity ensures early identification of potentially fatal AAAs, while specificity minimizes unnecessary anxiety and interventions in patients without aneurysms. **Q2: How does the ROC curve aid in AI model evaluation?** It visualizes the balance between detecting true positives and avoiding false positives at various thresholds, helping to select an operational point that best suits clinical priorities. **Q3: Can AI replace cardiovascular specialists in AAA diagnosis?** No. AI serves as a complementary tool that enhances diagnostic accuracy and efficiency but does not substitute clinical expertise and comprehensive patient evaluation. **Q4: What does a 0.96 AUC indicate in practical terms?** It signifies that the AI correctly distinguishes between aneurysmal and healthy aortas 96% of the time, reflecting high diagnostic performance.

Conclusion

AI-powered detection of Abdominal Aortic Aneurysm represents a significant leap forward in vascular medicine, enabling earlier diagnosis and proactive management. Understanding and applying rigorous performance metrics such as sensitivity, specificity, PPV, NPV, and AUC are critical to validating AI models for clinical use. While current AI systems exhibit excellent accuracy, ongoing research, clinical validation, and thoughtful integration into healthcare workflows are essential to realize their full potential. Ultimately, AI will act as a valuable ally to clinicians, improving patient outcomes through timely, accurate, and scalable AAA detection.

Keywords: Abdominal Aortic Aneurysm, AI in Healthcare, Medical Imaging, Deep Learning, Sensitivity, Specificity, ROC Curve, AUC, Clinical Decision Support, Vascular Surgery

Rasit Dinc Digital Health & AI Research

<https://rasitdinc.com>

© 2025 Rasit Dinc