International Journal of Artificial Intelligence, Data Science, and Machine Learning
Grace Horizon Publication | Volume 7, Issue 2, 139-144, 2026
ISSN: 3050-9262 | https://doi.org/10.63282/3050-9262.IJAIDSML-V7I2P121

*Original Article*

# Production Deployment of Computer-Aided Detection Systems in Mammography Screening: Throughput, False Positive Reduction, and Clinical Workflow Integration

Sri Gantikota
Senior Software Engineer, San Diego, California 92101, USA.

***Received On: 12/03/2026*** ***Revised On: 06/04/2026*** ***Accepted On: 13/04/2026*** ***Published On: 24/04/2026***

***Abstract*** *- Computer-aided detection systems for mammography have been part of clinical practice since the late 1990s, with the first generation FDA-cleared in 1998 and rapid adoption across United States screening facilities reaching a substantial majority by the mid 2010s. The first generation systems were associated with concerns about increased false positives and unclear effects on diagnostic accuracy in real-world deployment. The current generation, based on deep learning and rebranded by some vendors as AI-CAD to distinguish it from the classical CAD that preceded it, has produced results in observational studies that suggest meaningfully better diagnostic performance, although the long-term effect on patient outcomes remains the subject of ongoing prospective trials. This paper describes the production deployment of a mammographic prediction system that processed more than five thousand images per week, flagged suspicious anomalies for radiologist review, and contributed to a ten percent reduction in patient recall rates in the deployment setting. The paper focuses on the engineering decisions that determine whether such a system delivers its potential in production: throughput design, integration with the radiology workflow including PACS and reporting, false positive management, alert formatting to support rather than disrupt radiologist judgment, and the ongoing monitoring required to detect performance drift. The paper closes with a discussion of the regulatory environment, the limitations of single-site observational deployment as evidence, and the ongoing prospective trial work that is needed to confirm long-term outcomes.*

***Keywords*** *- Mammography, Computer-Aided Detection, CAD, AI-CAD, Deep Learning, Radiology Workflow, PACS Integration, Recall Rate, False Positive Reduction, Screening, Breast Cancer, Clinical Deployment*.

## 1. Introduction

Breast cancer is the most commonly diagnosed cancer among women worldwide. Mammographic screening is the standard population-level intervention for early detection, and its effectiveness depends on the radiologist\'s ability to identify suspicious findings in screening images. The radiologist\'s workload is high; a busy screening practice may interpret hundreds of mammograms per day, with each mammogram requiring the radiologist to compare multiple views and ideally to compare against prior studies. Computer-aided detection systems have been developed since the 1990s to assist the radiologist by highlighting regions of the image that warrant attention.

The first generation of computer-aided detection systems for mammography received FDA clearance in 1998 and was rapidly adopted across United States screening facilities. By the mid 2010s a substantial majority of US mammography facilities used CAD. The early evidence on the clinical impact of these systems was mixed, with subsequent studies raising concerns about increased false positives, higher recall rates, and the absence of demonstrated improvement in cancer detection when CAD-assisted reads were compared with unassisted reads in real-world settings. These concerns motivated more rigorous evaluation of the next generation of computer-aided detection technology.

The current generation of computer-aided detection systems is based on deep learning rather than on the classical computer vision methods that the first generation used. Some vendors and clinicians refer to this generation as AI-CAD to distinguish it from the classical CAD that preceded it. Observational studies of the current generation have produced more encouraging results, with reports of improved sensitivity, improved specificity, and reduced recall rates compared with both classical CAD and unassisted radiologist reads. A pragmatic randomized controlled trial comparing screening mammography with and without AI assistance is currently underway and will provide stronger evidence on the clinical effect of AI-CAD in real-world screening practice.

This paper describes the production deployment of a mammographic prediction system that processed more than five thousand images per week, flagged suspicious anomalies for radiologist review, and contributed to a ten percent reduction in patient recall rates in the deployment setting. The deployment was performed in the mid 2010s as part of healthcare imaging software development. The contribution

of the paper is the engineering perspective: the deployment focused on the practical concerns that determine whether such a system delivers its potential in production, including throughput design, workflow integration, false positive management, alert formatting, and ongoing performance monitoring. These concerns are distinct from the algorithmic concerns that dominate the academic literature on mammography AI, and they are the concerns that determine whether the algorithm\'s reported performance translates to clinical impact.

The rest of the paper is organized as follows. Section 2 covers the background on computer-aided detection in mammography. Section 3 describes the production system architecture. Section 4 covers integration with the radiology workflow. Section 5 covers false positive management and alert formatting. Section 6 reports the deployment results. Section 7 covers monitoring and drift detection. Section 8 discusses the regulatory environment. Section 9 covers limitations. Section 10 concludes.

## 2. Background

### *2.1. The First Generation of CAD*

First-generation computer-aided detection systems for mammography were based on classical computer vision techniques. They identified candidate findings such as microcalcification clusters and masses by applying handcrafted feature extractors to the image, scoring the candidates with a classifier trained on annotated cases, and rendering marks on the image at the locations of high-scoring candidates. The radiologist would interpret the original image, observe the marks, and use them as a secondary check. The systems were positioned as a second reader rather than as a replacement for the radiologist\'s primary read.

Real-world evaluation of first-generation CAD raised concerns. A widely cited body of work reported that CAD-assisted reads were associated with lower diagnostic accuracy, higher false positive rates, and higher costs compared with unassisted reads in the screening setting. The mechanism proposed for the increase in false positives was that radiologists became less willing to dismiss CAD marks as false positives, leading to more recalls of patients for diagnostic follow-up that turned out to be unnecessary. The same body of work suggested that early adoption had proceeded without strong clinical validation.

### *2.2. The Current Generation*

The current generation of computer-aided detection is based on deep learning, typically convolutional neural networks trained on large mammography datasets. The architectural details vary across vendors but the general pattern is end-to-end training on labeled cases, with the model producing both per-image scores and per-region localization of suspicious findings. The model is integrated into the radiology workflow in roughly the same way the first-generation systems were, as a second reader whose findings the radiologist considers alongside their own.

Observational studies of current-generation systems have reported encouraging results. One study published in the European Journal of Radiology Open reported that stand-alone AI-CAD detected an additional 17.9 percent of cancers initially overlooked by radiologists, although the same study noted that the AI-CAD also produced higher recall rates with 89.0 percent of its marks falling on negative mammograms. Another study comparing radiologist performance with and without AI-CAD assistance found that radiologists without AI-CAD showed lower specificity and accuracy and higher recall rates than stand-alone AI-CAD, suggesting that the assistance improved radiologist performance toward the AI-CAD baseline. The evidence base is consistent with a meaningful improvement over the first generation, but prospective randomized trials are required to confirm clinical outcomes.

### *2.3. The Specific Deployment in This Paper*

The deployment described in this paper is a mammographic prediction system that processed more than five thousand images per week in a clinical screening setting. The system flagged suspicious anomalies for radiologist review and contributed to a ten percent reduction in patient recall rates in the deployment setting. The system predates the most recent generation of deep learning approaches but operated in the same workflow position as the modern systems do, and the engineering concerns it addressed are common to deployments across the generations. The paper\'s focus is on those engineering concerns rather than on the specific algorithm.

## 3. Production System Architecture

### *3.1. Throughput Requirements*

Processing more than five thousand images per week corresponds to roughly seven hundred images per day in a typical screening practice, with the actual rate concentrated during the working hours of the imaging center. The system must keep up with the imaging center\'s capture rate so that predictions are available when the radiologist is ready to read the cases. The throughput requirement drives the choice of compute infrastructure, the parallelism of the prediction pipeline, and the queueing model that handles burst arrivals.

### *3.2. The Prediction Pipeline*

The prediction pipeline ingests mammographic images from the imaging modality through the standard healthcare imaging protocol, performs the necessary preprocessing including normalization and view detection, runs the prediction model, and stores the results in a form that the downstream workflow can consume. Each stage is implemented as a discrete component so that the stages can scale independently. Preprocessing tends to be CPU-bound; the prediction model is GPU-accelerated in the current generation and was CPU-bound in earlier generations. The discrete-stage architecture isolates the bottleneck to a single stage at a time, which is what allows the throughput to be tuned by adding capacity to the bottleneck rather than scaling the whole pipeline uniformly.

### *3.3. Storage and Persistence*

Predictions are persisted in a database that the workflow components query. The persistence is needed both for the immediate workflow, in which the radiologist consumes the predictions during the read, and for ongoing monitoring, in which the system\'s performance is tracked over time by comparing predictions against the eventual ground truth from the patient\'s clinical follow-up. The persistence layer is designed to retain predictions for the period required by the monitoring use case, which is substantially longer than the immediate workflow use case requires.

### *3.4. The Image Handling Question*

Mammographic images are large; a single screening study consists of multiple views and may total tens of megabytes when the original resolution is preserved. The architecture choices about whether to retain the full-resolution images alongside the predictions, whether to retain downsampled versions for monitoring purposes, and whether to rely on the imaging archive as the source of truth all have implications for storage cost and for the workflow consequences of needing to retrieve an image for review. The deployment in scope here retained the predictions in the system\'s own database and relied on the imaging archive for the images themselves, accessed through the standard imaging protocol when needed.

## 4. Integration with the Radiology Workflow

### *4.1. The PACS Integration*

The Picture Archiving and Communication System, commonly known as PACS, is the central system in a modern radiology workflow. PACS receives images from imaging modalities, stores them, and serves them to viewing workstations when the radiologist is ready to read them. The computer-aided detection system integrates with PACS as both a consumer of the images and a producer of the prediction overlays that the radiologist sees alongside the images. The integration uses the standard imaging protocol so that the prediction overlays appear in the radiologist\'s normal reading workflow rather than in a separate application that the radiologist would have to switch to.

### *4.2. Reporting Integration*

Radiology reports are generated through a reporting system that the radiologist dictates into. The reporting system records the radiologist\'s interpretation and assigns a BI-RADS category that summarizes the level of concern. The computer-aided detection system\'s predictions are available to the radiologist during reporting so that the radiologist can take them into account. The system does not write directly into the report. The decision about what to include in the report is the radiologist\'s, both because the radiologist is the responsible interpreter and because the predictions are an aid to interpretation rather than an interpretation themselves.

### *4.3. The Question of Timing*

The timing of when the prediction is shown to the radiologist matters for the cognitive ergonomics of the workflow. A prediction shown before the radiologist has formed their own impression can anchor the interpretation in a way that the radiologist may not want. A prediction shown after the radiologist has formed their impression can act as a second reader that the radiologist consults to validate their conclusion. Different deployments choose different timings, and the choice has implications for how the radiologist uses the system. The deployment in scope here showed the predictions in the radiologist\'s reading view from the start of the read, with the radiologist free to consult them when they preferred.

### *4.4. Order and Result Messaging*

HL7 v2 messages convey the imaging orders into the system and the results back out. The integration uses the same messaging infrastructure described in companion work on HL7 v2 integration pipelines. The computer-aided detection system reads order messages to know which studies it should process when they arrive and writes result messages to convey the predictions into the downstream systems that consume them. The use of standard messaging is what allows the system to integrate with the rest of the hospital information environment without requiring custom integration for each connected system.

## 5. False Positive Management and Alert Formatting

### *5.1. The Cost of False Positives*

False positives in mammographic screening have real costs. A patient flagged for recall undergoes additional imaging, possibly biopsy, and the time between the recall and the resolution carries substantial anxiety. The reduction of unnecessary recalls is therefore a clinical good in itself, separate from the question of cancer detection. A computer-aided detection system that increases the recall rate by a meaningful amount has to be weighed against any improvement in cancer detection it produces, and the balance is not always favorable.

### *5.2. Threshold Selection*

The prediction model produces a score for each suspicious region. The decision to display a region as a finding depends on a threshold applied to the score. A lower threshold catches more cancers and produces more false positives. A higher threshold catches fewer cancers and produces fewer false positives. The choice of threshold is a clinical decision that the system must make explicit so that the deployment site can adjust it to the population it serves and to the workflow capacity of its radiologists.

### *5.3. Alert Formatting*

Alert formatting matters because the radiologist will rapidly tune out alerts that do not match their judgment. Alerts that are too aggressive or too poorly formatted erode trust in the system over time. The pattern in the deployment described here was conservative: alerts were rendered as subtle overlays that the radiologist could choose to inspect, rather than as prominent visual elements that would dominate the reading view. The conservative formatting preserved the radiologist\'s primary read as the primary act, with the system in a supporting role.

### *5.4. The Significance Question*

Not every alert from the system represents a clinically significant finding. Some alerts correspond to features that are visible to the radiologist but that the radiologist would correctly dismiss as benign. Other alerts correspond to features that warrant a closer look. The system does not distinguish these categories on its own; the distinction is part of the interpretation that the radiologist performs. The deployment treats this as a feature rather than a defect of the system, because the alternative would be a system that quietly suppresses some alerts based on its own judgment of significance, which would obscure information from the radiologist.

## 6. Deployment Results

### *6.1. Throughput Achieved*

The system sustained processing of more than five thousand images per week through the deployment period. The throughput was within the design target and was not a binding constraint on the imaging center\'s capture rate. Queueing of incoming studies was minimal under normal conditions, with brief burst-period queueing during the busiest hours that the radiologists did not experience as a delay because the predictions were ready by the time they reached the relevant studies in the reading queue.

### *6.2. Recall Rate Reduction*

Recall rates in the deployment setting were approximately ten percent lower than the baseline observed before the deployment. The interpretation of the reduction is that the system helped the radiologists triage their attention more effectively, with fewer studies flagged for recall that subsequently turned out to be negative. The reduction is the headline outcome that motivated the publication of the work and is the most direct measure of the patient-experienced benefit of the deployment.

### *6.3. Radiologist Acceptance*

Radiologist acceptance of the system was monitored through both formal feedback and observed usage patterns. The conservative alert formatting described in Section 5.3 contributed to acceptance because it did not interfere with the radiologists\' established reading rhythm. The radiologists who initially expressed skepticism about the system reported over time that the predictions had become a useful second opinion that they consulted regularly without depending on.

## 7. Monitoring and Drift Detection

### *7.1. Performance Drift*

A deployed prediction system can drift from its design performance over time. Drift can come from changes in the imaging equipment, changes in the patient population, changes in the way studies are acquired, or accumulated software changes in any of the components that handle the images. The monitoring infrastructure must detect drift early enough that the team can investigate before drift becomes harmful.

### *7.2. Ground Truth Acquisition*

Detecting drift requires comparing the system\'s predictions against eventual ground truth. The ground truth for mammographic screening is the outcome of subsequent clinical follow-up, including biopsy results where biopsy is performed and absence of cancer at the next round of screening where it is not. The acquisition of ground truth is itself a substantial pipeline because the outcomes accumulate over a time scale of months to years, and the linkage from the predictions to the eventual outcomes has to be maintained across that time.

### *7.3. Monitoring Metrics*

The metrics monitored include the cancer detection rate, the recall rate, the positive predictive value of the system\'s alerts, the proportion of cancers detected by the system but missed by the radiologist on the initial read, and the proportion missed by the system but detected by the radiologist. Each of these metrics is computed against the ground truth on a rolling time window so that recent performance can be compared against historical baselines.

### *7.4. Action on Drift*

When drift is detected, the response depends on the suspected cause. Equipment changes may require retraining of the system on data from the new equipment. Population changes may require recalibration of the threshold described in Section 5.2. Software changes may require investigation of which change introduced the drift. The system\'s monitoring infrastructure is therefore not only a passive observer but is also an input to operational decisions about whether to continue using the system, to retune it, or to suspend it pending investigation.

## 8. Regulatory Environment

Computer-aided detection systems for mammography are medical devices under the regulatory frameworks of the major jurisdictions. In the United States the FDA clears such devices through the premarket pathway. In the European Union the equivalent process is the Medical Device Regulation. The regulatory frameworks require evidence of safety and effectiveness, postmarket surveillance, and adherence to quality management systems in the development and manufacture of the device. The deployment described here operated within these frameworks, with the engineering decisions documented to support both initial clearance and ongoing surveillance.

The most recent thinking in the regulatory community recognizes that AI-based medical devices change in ways that traditional medical devices do not, because retraining and recalibration may be part of the normal operation of the device. Predetermined change control plans, which the FDA has developed guidance on, allow some kinds of changes to be made without requiring a new clearance for each change. The engineering practices that produce the monitoring data described in Section 7 are also the practices that support the operation of such a change control plan.

## 9. Limitations

### *9.1. Single-Site Observational Evidence*

The deployment described here is single-site observational evidence. The ten percent recall rate reduction observed is consistent with reductions reported in observational studies of other systems, but observational evidence has known limitations including the absence of a controlled comparison group. Prospective randomized trials are required to establish causal claims about the effect of computer-aided detection on patient outcomes. The trial referenced in Section 2.2 is one of the studies underway to provide this evidence base.

### *9.2. Generation of System*

The deployment described here predates the current generation of deep learning systems. The engineering concerns covered in this paper apply to the current generation as well, but the algorithmic performance of the current generation differs from the system described here in ways that the published comparisons cover.

### *9.3. The Workflow Question*

The workflow integration choices described in Section 4 are specific to the deployment context. Other deployments may make different choices about timing, alert formatting, and reporting integration that produce different radiologist experiences and different clinical outcomes. The engineering principles transfer but the specific choices need to be made in light of the local workflow and the local radiologists\' preferences.

## 10. Conclusion

Computer-aided detection systems for mammography have evolved substantially since the first generation reached the clinic in the late 1990s. The current generation, based on deep learning, has produced evidence in observational studies of meaningfully better diagnostic performance than the first generation, with prospective randomized trials underway to confirm long-term clinical outcomes. The engineering decisions that determine whether such a system delivers its potential in production are distinct from the algorithmic decisions that dominate the academic literature: throughput design, integration with the PACS and reporting workflow, false positive management, alert formatting that supports rather than disrupts radiologist judgment, and the ongoing monitoring required to detect performance drift. The deployment described in this paper, which processed more than five thousand mammographic images per week and contributed to a ten percent reduction in patient recall rates, illustrates the engineering perspective on these systems and provides a reference for similar deployments. The engineering practices that support such a deployment are also the practices that support the regulatory expectations for medical devices that may change during their operational life.

***Acknowledgments***

This work was performed in the context of healthcare imaging software development at Merge Healthcare. The author thanks the imaging engineering team and the clinical partners whose collaboration made the deployment possible.

**Conflicts of Interest**

The author declares that there is no conflict of interest concerning the publishing of this paper.

## References

[1] Lehman, C. D., Wellman, R. D., Buist, D. S. M., et al. Diagnostic Accuracy of Digital Screening Mammography With and Without Computer-Aided Detection. JAMA Internal Medicine, 175(11), 1828 to 1837, 2015. https://scholar.google.com/scholar?hl=en&q=D., Wellman, R

[2] Fenton, J. J., Taplin, S. H., Carney, P. A., et al. Influence of Computer-Aided Detection on Performance of Screening Mammography. New England Journal of Medicine, 356(14), 1399 to 1409, 2007. https://scholar.google.com/scholar?hl=en&q=J., Taplin, S

[3] Yoon, J. H., Han, K., Suh, H. J., et al. Artificial intelligence-based computer-assisted detection/diagnosis (AI-CAD) for screening mammography: Outcomes of AI-CAD in the mammographic interpretation workflow. European Journal of Radiology Open, Vol. 11, 100509, December 2023. https://scholar.google.com/scholar?hl=en&q=H., Han, K., Suh, H

[4] Yoon, J. H., Strand, F., Baltzer, P. A. T., et al. Standalone AI for Breast Cancer Detection at Screening Digital Mammography and Digital Breast Tomosynthesis. Radiology, 307(5), 2023. https://scholar.google.com/scholar?hl=en&q=H., Strand, F., Baltzer, P

[5] Rodriguez-Ruiz, A., Lang, K., Gubern-Merida, A., et al. Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography. Journal of the National Cancer Institute, 111(9), 916 to 922, 2019. https://scholar.google.com/scholar?hl=en&q=Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography

[6] McKinney, S. M., Sieniek, M., Godbole, V., et al. International evaluation of an AI system for breast cancer screening. Nature, 577(7788), 89 to 94, 2020. https://scholar.google.com/scholar?hl=en&q=M., Sieniek, M., Godbole, V., et al

[7] Lang, K., Josefsson, V., Larsson, A. M., et al. Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI). Lancet Oncology, 24(8), 936 to 944, 2023. https://scholar.google.com/scholar?hl=en&q=M., et al

[8] American College of Radiology. Breast Imaging Reporting and Data System (BI-RADS), Fifth Edition. https://scholar.google.com/scholar?hl=en&q=Breast Imaging Reporting and Data System (BI-RADS), Fifth Edition

[9] Digital Imaging and Communications in Medicine Standards Committee. DICOM Standard, NEMA PS3.

https://scholar.google.com/scholar?hl=en&q=DICOM Standard, NEMA PS3

[10] Integrating the Healthcare Enterprise. IHE Radiology Technical Framework. https://scholar.google.com/scholar?hl=en&q=IHE Radiology Technical Framework

[11] United States Food and Drug Administration. Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence/Machine Learning-Enabled Device Software Functions, Guidance for Industry and FDA Staff, 2024. https://scholar.google.com/scholar?hl=en&q=Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence/Machine Learning-Enabled Device Software Functions, Guidance for Industry and FDA Staff, 2024

[12] United States Food and Drug Administration. Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data: Premarket Notification (510(k)) Submissions, Guidance for Industry and FDA Staff. https://scholar.google.com/scholar?hl=en&q=Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data: Premarket Notification (510(k)) Submissions, Guidance for Industry and FDA Staff

[13] European Union. Medical Device Regulation, Regulation (EU) 2017/745. https://scholar.google.com/scholar?hl=en&q=Medical Device Regulation, Regulation (EU) 2017/745

[14] International Organization for Standardization. ISO 13485 Medical devices: Quality management systems. https://scholar.google.com/scholar?hl=en&q=ISO 13485 Medical devices: Quality management systems

[15] International Organization for Standardization. ISO 14971 Medical devices: Application of risk management to medical devices. https://scholar.google.com/scholar?hl=en&q=ISO 14971 Medical devices: Application of risk management to medical devices

[16] World Health Organization. WHO position paper on mammography screening, 2014. https://scholar.google.com/scholar?hl=en&q=WHO position paper on mammography screening, 2014

[17] US Preventive Services Task Force. Final Recommendation Statement: Breast Cancer Screening, 2024. https://scholar.google.com/scholar?hl=en&q=Final Recommendation Statement: Breast Cancer Screening, 2024

[18] Litjens, G., Kooi, T., Bejnordi, B. E., et al. A Survey on Deep Learning in Medical Image Analysis. Medical Image Analysis, 42, 60 to 88, 2017. https://scholar.google.com/scholar?hl=en&q=E., et al

[19] LeCun, Y., Bengio, Y., and Hinton, G. Deep learning. Nature, 521(7553), 436 to 444, 2015. https://scholar.google.com/scholar?hl=en&q=Deep learning