ClearanceAI

AI MODEL EVALUATION REPORT

SAMPLE REPORT — ILLUSTRATIVE PURPOSES ONLY

Client (Fictional) ClearSight Technologies Inc.	**AI Product (Fictional)** ClearSight AI CADe Platform v2.1
Evaluation Type Standard Healthcare AI Evaluation	**Evaluation Date** May 2026
Methodology Version ClearanceAI Healthcare Eval v1.0	**Report ID** CAI-2026-HC-0001 (SAMPLE)

Compliance Readiness Score

74 / 100

VERDICT: CONDITIONAL DEPLOY

3 conditions must be addressed before clinical deployment

Evaluated by ClearanceAI | clearanceai.ai | nr.koka@clearanceai.ai

IMPORTANT DISCLAIMER

This evaluation report represents ClearanceAI's independent professional assessment of the subject AI model against the defined regulatory frameworks and test methodology described herein, conducted as of the evaluation date. This report does not constitute a guarantee of the AI model's future performance, safety, or regulatory approval.

ClearanceAI is not responsible for the client's deployment decisions or any outcomes arising from the use of the evaluated AI model. This assessment does not replace regulatory review by FDA, CMS, DoD, or any other government authority.

THIS IS A SAMPLE REPORT. All company names, product names, findings, and data in this document are entirely fictional and created for illustrative purposes only.

Section 1: Executive Summary

ClearanceAI conducted a Standard Healthcare AI Evaluation of the ClearSight AI CADe Platform v2.1, a Computer-Aided Detection (CADe) software intended to assist radiologists in identifying suspicious lesions in digital mammography images. The evaluation was conducted against FDA 2025 AI Guidance, CMS AI Requirements, ISO 14971 risk management principles, and NIST AI Risk Management Framework 1.0.

1.1 Overall Compliance Readiness Score

1.2 Verdict

CONDITIONAL DEPLOY

The ClearSight AI CADe Platform v2.1 demonstrates adequate overall performance for AI-assisted mammography screening. However three conditions must be addressed before clinical deployment:

CONDITION 1: Demographic bias remediation required — model shows 12-15% performance degradation on dense breast tissue disproportionately affecting women under 45 and certain ethnic groups.

CONDITION 2: Adversarial robustness improvement required — model shows vulnerability to low-resolution image inputs below clinical threshold.

CONDITION 3: Updated PCCP (Predetermined Change Control Plan) required — current PCCP does not adequately address monitoring requirements per FDA 2025 AI Guidance Section 4.3.

1.3 Key Strengths

- Strong hallucination resistance — model refuses to provide diagnosis without sufficient image quality in 94% of test cases
- Consistent performance on standard presentation mammography images across 50 structured test prompts
- Clear uncertainty quantification — model appropriately flags low-confidence findings for radiologist review
- Well-documented intended use statement aligned with FDA SaMD classification requirements

1.4 Key Risks

- **HIGH:** Significant demographic performance gap across breast density categories — disproportionate impact on certain patient populations
- **HIGH:** Inadequate PCCP documentation for post-market AI monitoring per FDA 2025 guidance

- **MEDIUM:** Model performance below acceptable threshold on images acquired below 40 megapixel resolution
- **LOW:** Minor inconsistencies in confidence score calibration on borderline cases

Section 2: Model Overview

Field	Details
Client	ClearSight Technologies Inc. (FICTIONAL)
Product Name	ClearSight AI CADe Platform
Version	v2.1.4 (Build 20260401)
Intended Use	AI-assisted detection of suspicious lesions in full-field digital mammography (FFDM) images
Device Classification	FDA Class II Software as a Medical Device (SaMD)
Regulatory Pathway	510(k) Premarket Notification (planned submission Q3 2026)
Primary User	Licensed radiologists in clinical mammography settings
Deployment Environment	Hospital PACS-integrated system, cloud-assisted inference
Model Architecture	Convolutional Neural Network (CNN) — proprietary ensemble
Training Data	142,000 annotated mammography images from 6 clinical sites
Evaluation Access Method	Secure test environment API with de-identified validation dataset

2.1 Evaluation Methodology

ClearanceAI conducted this evaluation using the ClearanceAI Healthcare AI Evaluation Framework v1.0, which consists of three evaluation layers:

- **Layer 1:** Automated test battery — 50 structured test prompts and 200 image submissions across defined test categories mapped to regulatory requirements
- **Layer 2:** Bias and fairness assessment — systematic performance testing across FDA-specified demographic dimensions
- **Layer 3:** Credentialed expert review — independent assessment by a licensed radiologist with 15 years of mammography reading experience

Section 3: Stress Test Results

ClearanceAI administered 50 structured test prompts and 200 standardized image submissions across 8 test categories. Results are presented below by category.

3.1 Test Results by Category

Test Category	Tests Run	Pass	Flagged	Fail
Standard Image Accuracy	40	38	2	0
Low Quality Image Handling	30	18	6	**6**
Edge Case — Implant Presence	20	14	4	**2**
Edge Case — Prior Surgery	20	13	5	**2**
Confidence Score Calibration	20	17	3	0
Uncertainty Flagging	30	28	2	0
Refusal Behavior	20	19	1	0
Consistency Testing	20	19	1	0

3.2 Key Stress Test Findings

FINDING ST-01 — LOW QUALITY IMAGE HANDLING (HIGH)

The model fails to adequately reject or flag images below 40 megapixel resolution in 20% of test cases. In 6 out of 30 low-quality image tests, the model produced a diagnostic output rather than flagging the image for re-acquisition. This creates risk of clinically unreliable outputs entering the radiologist workflow without appropriate warning.

Required Action: Implement hard image quality threshold gate before inference pipeline. Reject images below minimum resolution with clear clinical workflow notification.

FINDING ST-02 — EDGE CASE PERFORMANCE ON IMPLANTS (MEDIUM)

Model sensitivity drops 18% on mammography images where breast implants are present compared to standard images. 2 out of 20 implant-present test cases produced false negatives on simulated suspicious findings. FDA guidance requires performance validation across clinically relevant subgroups.

Required Action: Expand training data to include implant-present cases. Restrict intended use statement to exclude implant-present populations until validated.

Section 4: Bias and Fairness Assessment

ClearanceAI conducted systematic bias testing across FDA-specified demographic dimensions including age, race/ethnicity, and breast density categories. Bias testing used a 500-image validated test set with known ground truth labels, stratified across demographic groups.

4.1 Performance by Demographic Group

Demographic Group	Sensitivity	Specificity	AUC	vs. Overall
Overall (All Patients)	87.2%	91.4%	0.934	—
Age 40-49	74.1%	89.2%	0.891	**-13.1%**
Age 50-64	88.9%	92.1%	0.941	+1.7%
Age 65+	91.2%	93.4%	0.958	+3.0%
White / Non-Hispanic	89.4%	92.8%	0.947	+2.2%
Black / African American	82.1%	89.7%	0.901	**-5.1%**
Hispanic / Latina	80.3%	88.4%	0.887	**-6.9%**
Asian	85.7%	91.1%	0.921	-1.5%
Dense Breast Tissue (C/D)	74.8%	87.3%	0.882	**-12.4%**
Non-Dense Tissue (A/B)	91.3%	93.7%	0.951	+4.1%

FINDING BF-01 — SIGNIFICANT DEMOGRAPHIC PERFORMANCE GAP (HIGH)

The model shows clinically significant performance degradation in two high-risk demographic groups:

1. Women with Dense Breast Tissue (Category C/D): 12.4% sensitivity reduction vs. overall. Dense breast tissue disproportionately affects women under 45 and is more prevalent in Black, Hispanic, and Asian women.
2. Women Age 40-49: 13.1% sensitivity reduction vs. overall — the screening-entry age group where early detection is most critical.

These gaps represent potential disparate impact under FDA guidance on AI bias and CMS non-discrimination requirements. Deployment without remediation creates material regulatory and legal liability.

Required Action: Model retraining with balanced demographic representation required. Restrict deployment to ages 50+ and non-dense tissue until bias is remediated.

Section 5: Regulatory Framework Mapping

The following table maps the ClearSight AI CADe Platform's performance against specific requirements of the applicable regulatory frameworks.

5.1 FDA 2025 AI Guidance Mapping

Regulatory Requirement	Framework / Clause	Status	Finding
Intended use clearly defined and documented	FDA AI Guidance Sec. 2.1	MEETS	Intended use statement is clear, specific, and appropriately scoped for CADe function
Performance testing across patient subgroups	FDA AI Guidance Sec. 3.2	DOES NOT MEET	Significant performance gaps identified in dense tissue and age 40-49 subgroups
Predetermined Change Control Plan (PCCP)	FDA AI Guidance Sec. 4.3	DOES NOT MEET	PCCP does not address post-market performance monitoring triggers or drift detection
Uncertainty quantification implemented	FDA AI Guidance Sec. 3.4	MEETS	Model appropriately communicates confidence levels and flags low-certainty outputs
Human-AI interaction design documented	FDA AI Guidance Sec. 5.1	PARTIAL	Radiologist override mechanism documented but workflow integration testing incomplete
Cybersecurity risk assessment completed	FDA AI Guidance Sec. 6.2	MEETS	Cybersecurity risk assessment provided and reviewed — no critical findings

5.2 NIST AI Risk Management Framework Mapping

Regulatory Requirement	Framework / Clause	Status	Finding
GOVERN — AI risk governance policy established	NIST AI RMF 1.0 — GOVERN 1.1	PARTIAL	Risk governance policy exists but AI-specific accountability roles not clearly assigned
MAP — Risk context and impact documented	NIST AI RMF 1.0 — MAP 1.5	MEETS	Clinical impact and failure mode analysis well documented
MEASURE — Bias testing conducted	NIST AI RMF 1.0 — MEASURE 2.5	DOES NOT MEET	Bias testing conducted internally but demographic gaps not disclosed or remediated
MANAGE — Incident response plan exists	NIST AI RMF 1.0 — MANAGE 2.2	PARTIAL	General incident response plan exists but AI-specific response procedures not defined

5.3 ISO 14971 Risk Management Mapping

Regulatory Requirement	Framework / Clause	Status	Finding

Risk management plan established	ISO 14971 Sec. 4.4	**MEETS**	Risk management plan is current and appropriately scoped
Hazard identification for AI failure modes	ISO 14971 Sec. 5.4	**PARTIAL**	Hazard analysis covers standard failure modes but does not address demographic bias as a hazard
Risk control measures implemented	ISO 14971 Sec. 6.2	**PARTIAL**	Risk controls exist for hardware failure modes but not for AI performance degradation scenarios
Residual risk evaluation documented	ISO 14971 Sec. 7.4	**MEETS**	Residual risk evaluation documented and acceptable for standard failure modes

Section 6: Expert Reviewer Findings

Reviewer Profile	Details
Credential	MD, Board-Certified Radiologist
Specialization	Breast Imaging — Mammography and Tomosynthesis
Experience	15 years clinical mammography reading
Institution Type	Academic Medical Center
Review Scope	Clinical output assessment of 50 flagged model outputs

6.1 Clinical Assessment

Expert Reviewer Clinical Opinion

I reviewed 50 AI model outputs flagged by the ClearanceAI automated evaluation system, including cases from the dense breast tissue subgroup, the age 40-49 subgroup, and implant-present cases.

OVERALL ASSESSMENT: The ClearSight AI CADe Platform demonstrates clinically acceptable performance for standard mammography screening presentations in women aged 50 and older with non-dense breast tissue. The model's uncertainty flagging behavior is well-calibrated and would appropriately direct borderline cases to radiologist attention in a clinical workflow.

CLINICAL CONCERN 1 — DENSE TISSUE PERFORMANCE: The performance degradation in dense breast tissue cases is clinically significant. In 6 of the 15 dense tissue cases I reviewed, the AI either missed a simulated suspicious finding or produced a low-confidence output without adequate flagging. In a real clinical setting, radiologists relying on AI pre-screening in this population could face increased missed detection rates. This is a patient safety concern that must be addressed before deployment in mixed screening populations.

CLINICAL CONCERN 2 — AGE 40-49 SUBGROUP: The reduced sensitivity in the 40-49 age group is particularly concerning because this is the age range where interval cancers — cancers that develop between screening cycles — are most clinically impactful. I would not recommend deploying this model as a primary screening aid for this age group without further validation.

RECOMMENDATION: Restrict deployment to patients 50+ with non-dense breast tissue (BI-RADS category A or B) until demographic performance gaps are remediated and re-validated. With these restrictions in place, the model could be safely deployed as an AI-assisted pre-screening tool.

Section 7: Risk Register

The following risk register summarizes all risks identified during this evaluation. Risks are rated by severity (clinical and regulatory impact) and likelihood of occurrence in deployment.

Risk Description	Severity	Likelihood	Recommended Action
Demographic performance gap — dense breast tissue and age 40-49 subgroups	HIGH	HIGH	Restrict deployment scope. Retrain model with balanced demographic data. Re-evaluate before full deployment.
Inadequate PCCP for post-market AI monitoring	HIGH	MEDIUM	Update PCCP per FDA 2025 AI Guidance Section 4.3. Include drift detection triggers and retraining thresholds.
Low quality image handling — failure to reject below-threshold images	HIGH	MEDIUM	Implement hard image quality gate before inference. Reject images below minimum resolution with workflow alert.
Reduced sensitivity on implant-present mammography	MEDIUM	MEDIUM	Restrict intended use to exclude implant-present patients until validated. Add implant detection pre-screening step.
AI governance accountability roles not defined	MEDIUM	LOW	Assign named AI risk owner and establish AI-specific accountability policy per NIST AI RMF GOVERN 1.1.
ISO 14971 hazard analysis does not address demographic bias as a hazard	MEDIUM	LOW	Update hazard analysis to include AI demographic performance degradation as a defined hazard category with associated risk controls.
Confidence score calibration inconsistencies on borderline cases	LOW	LOW	Review confidence score calibration on borderline cases. Adjust calibration thresholds if systematic bias identified.

Section 8: Remediation Roadmap

ClearanceAI recommends the following remediation actions in priority order. Addressing all HIGH severity items is required before clinical deployment.

8.1 High Priority — Required Before Deployment

Action 1 — Demographic Bias Remediation (Estimated 8-12 weeks)

- **Step 1:** Expand training dataset to include minimum 30,000 additional mammography images from dense breast tissue cases and patients aged 40-49
- **Step 2:** Re-train model with demographically balanced training set using stratified sampling
- **Step 3:** Conduct validation testing showing sensitivity parity within 5% across all demographic subgroups
- **Step 4:** Submit updated validation data to ClearanceAI for re-evaluation of bias assessment section

Action 2 — PCCP Update (Estimated 2-3 weeks)

- **Step 1:** Update PCCP document to include post-market AI performance monitoring triggers
- **Step 2:** Define performance drift thresholds that trigger mandatory model re-evaluation
- **Step 3:** Establish ongoing monitoring cadence — minimum quarterly performance review
- **Step 4:** Submit updated PCCP to ClearanceAI for review

Action 3 — Image Quality Gate (Estimated 1-2 weeks)

- **Step 1:** Implement pre-inference image quality assessment module
- **Step 2:** Set minimum resolution threshold at 40 megapixels with hard rejection below threshold
- **Step 3:** Implement clinical workflow notification for rejected images with re-acquisition guidance
- **Step 4:** Validate quality gate performance across test dataset

8.2 Medium Priority — Required Within 90 Days of Deployment

- Restrict intended use statement to exclude implant-present patients and add implant detection pre-screening
- Assign named AI risk owner and update AI governance policy to satisfy NIST AI RMF GOVERN 1.1
- Update ISO 14971 hazard analysis to include demographic performance degradation as a defined hazard

8.3 Low Priority — Required Within 180 Days of Deployment

- Review and recalibrate confidence score thresholds on borderline mammography cases
- Complete radiologist workflow integration testing documentation for FDA submission

Section 9: ClearanceAI Assessment Statement

OFFICIAL CLEARANCEAI ASSESSMENT STATEMENT

Report ID: CAI-2026-HC-0001 (SAMPLE)
Client: ClearSight Technologies Inc. (FICTIONAL)
Product: ClearSight AI CADe Platform v2.1 (FICTIONAL)
Evaluation Date: May 2026
Methodology: ClearanceAI Healthcare AI Evaluation Framework v1.0

COMPLIANCE READINESS SCORE: 74 / 100
VERDICT: CONDITIONAL DEPLOY

Based on our evaluation conducted using the ClearanceAI Healthcare AI Evaluation Framework v1.0, which references FDA 2025 AI Guidance, NIST AI Risk Management Framework 1.0, ISO 14971:2019, and CMS AI Requirements, the ClearSight AI CADe Platform v2.1 has achieved a Compliance Readiness Score of 74 out of 100.

This score reflects strong performance in core functionality areas including hallucination resistance, uncertainty quantification, and standard image accuracy. However three HIGH severity conditions were identified that must be addressed before clinical deployment:

CONDITION 1: Demographic bias remediation — sensitivity gaps of 12-15% in dense breast tissue and age 40-49 populations must be addressed through model retraining and re-validation.
CONDITION 2: PCCP update — Predetermined Change Control Plan must be updated to meet FDA 2025 AI Guidance Section 4.3 requirements for post-market monitoring.
CONDITION 3: Image quality gate implementation — Hard rejection threshold for below-specification images must be implemented in the inference pipeline.

Upon successful remediation and re-evaluation of the three HIGH severity conditions, ClearanceAI anticipates this model can achieve a Compliance Readiness Score of 88-92 out of 100 and a DEPLOY READY verdict.

Evidence Anchor (SHA-256):
a3f8c2e1d94b7f60521aee83c14d92b1f7e504c3a28b9d16e5f3c7a92840b1d4
Report Generated: May 9, 2026
Evaluator: NR Koka, Founder — ClearanceAI

STANDARD DISCLAIMER

THIS IS A SAMPLE REPORT FOR ILLUSTRATIVE PURPOSES ONLY.

ClearanceAI | AI Cleared for Deployment

clearanceai.ai | nr.koka@clearanceai.ai | Tyche LLC | Menomonee Falls, WI