

Impact of Image Degradation on Melanoma Detection by Artificial Intelligence: A Patient Safety Perspective

Marco A. Aguirre L. ^{1,✉}, Alan G. Aguirre-Lam ¹, Mayda Lam-Maldonado ¹, J. F. Franklin Rodríguez-Zapata ¹

¹. División de Estudios de Posgrado e Investigación, Tecnológico Nacional de México/Instituto Tecnológico de Ciudad Madero, Madero, MEX

Received: May 06, 2026 | Review began: June 09, 2026 | Review ended: June 24, 2026 | Published: July 01, 2026

© Copyright 2026

This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Introduction

The integration of artificial intelligence (AI) in teledermatology presents significant diagnostic potential, yet the vulnerability of these models to real-world image degradation remains a major patient safety concern. This study investigates the accuracy discrepancy, where global metrics obscure severe sensitivity drops in melanoma detection under visual stress.

Methods

We evaluated two deep learning architectures, ResNet-50 and MobileNetV2, using the HAM10000 dataset. An experimental pipeline was developed using the Python programming language to simulate three clinical photographic artifacts: Gaussian noise, focal blur, and low luminance. The models were subjected to stress tests to observe shifts in diagnostic classification and evaluate their structural stability.

Results

While baseline global accuracy remained deceptively stable, targeted degradation exposed catastrophic failures. Under moderate degradation, ResNet-50 sensitivity collapsed to near-zero levels, erroneously classifying malignant lesions as benign nevi. MobileNetV2 exhibited similar structural vulnerabilities, proving that standard evaluation metrics fail to capture critical diagnostic risks.

Conclusions

Deep learning models in dermatology exhibit significant vulnerability to common image degradation, posing a direct threat to patient safety. The implementation of automated image quality gates is essential before deploying these algorithms in clinical settings to prevent AI-induced diagnostic errors.

Categories: Safe and trustworthy AI in healthcare, Image Processing and Analysis, Deep Learning

Keywords: teledermatology, deep learning, melanoma detection, image degradation, patient safety, convolutional neural networks, mobile health, diagnostic accuracy

Introduction

Malignant melanoma remains the most aggressive form of skin cancer, responsible for the vast majority of deaths related to cutaneous neoplasms globally [1]. Patient prognosis and survival are intrinsically linked to early detection and timely surgical excision. Historically, initial dermatological screening has relied on in-person clinical evaluation using

How to cite this article:

Aguirre L. M, Aguirre-Lam A, Lam-Maldonado M, et al. (July 01, 2026) Impact of Image Degradation on Melanoma Detection by Artificial Intelligence: A Patient Safety Perspective. Cureus J Comput Sci 3 : es44389-026-00147-2. DOI <https://doi.org/10.7759/s44389-026-00147-2>

dermoscopy; however, geographical barriers and healthcare system saturation have driven the accelerated adoption of teledermatology [2]. In particular, direct-to-consumer mobile applications allow patients to capture images of their own skin lesions using smartphones to receive an initial assessment [3].

In this evolving landscape, the integration of artificial intelligence (AI) and deep learning has revolutionized the state of the art in medical image analysis. Multiple studies have demonstrated that convolutional neural networks (CNNs) can match, and even exceed, the diagnostic accuracy of board-certified dermatologists [4]. These algorithmic milestones have been made possible by the availability of large-scale standardized datasets, such as the HAM10000 (Human Against Machine) repository, which provides 10,015 high-resolution dermatoscopic images with optimal lighting and histopathological validation [5].

Nevertheless, a significant gap exists in the literature regarding the translation of these laboratory models into real-world clinical practice. While algorithms are trained on pristine images captured by specialists, telemedicine applications operate in an environment of high optical variability. Photographs captured by patients (peripheral telephotography) are frequently affected by blur due to kinetic movement, low luminance conditions, or electronic noise derived from lower-tier camera sensors [6]. This phenomenon, known in machine learning as domain shift, threatens to compromise the reliability of diagnostic inferences.

Furthermore, the current state of evaluation metrics in medical informatics tends to prioritize global accuracy. In intrinsically imbalanced datasets where benign lesions overwhelmingly predominate, this metric can mask severe deficiencies in the performance of the minority class (malignancy) [7]. Concurrently, the need to implement these models directly on mobile devices (edge computing) has led to the use of lightweight or compressed architectures, whose robustness to signal degradation compared to dense server-grade models (such as ResNet-50) has not been exhaustively characterized regarding iatrogenic risk [8].

Consequently, the objective of the present analytical study is to quantitatively evaluate the impact of real-world photographic degradation on the diagnostic safety of two paradigmatic architectures (ResNet-50 and MobileNetV2), through the controlled mathematical simulation of mobile capture artifacts. It is important to note that the synthetic degradation protocol described below was applied to dermatoscopic images acquired under standardized clinical conditions, rather than to native smartphone macrophotographs; therefore, the decline in sensitivity reported in this study should be interpreted as a controlled, conservative estimate of algorithmic fragility under visual stress, rather than a direct replication of the unconstrained photographic variability characteristic of direct-to-consumer teledermatology applications.

Materials And Methods

The present study was structured under a computational and quantitative experimental design. In the development of deep learning models for medical use, validating algorithmic stability against altered or out-of-distribution data is an essential methodological requirement prior to their evaluation in real-world clinical settings [9]. In the practice of asynchronous teledermatology, the variability in the quality of photographs captured by patients represents a technical challenge that can compromise diagnostic reliability [10,11]. This variability is frequently influenced by mobile device hardware limitations, user inexperience, and non-standardized environmental conditions.

To quantify the impact of this phenomenon under strictly controlled and reproducible conditions, the methodology was developed in four sequential phases. The first phase involved the selection, standardized partitioning, and preprocessing of a reference dermatological clinical dataset. The second phase comprised the initialization and fine-tuning of two commercial-grade CNN architectures. The third phase focused on the design and systematic mathematical injection of photographic artifacts prevalent in telemedicine, such as blur, luminance reduction, and sensor noise, exclusively into the test cohort. The final phase consisted of the extraction and statistical analysis of the resulting classification metrics, prioritizing diagnostic sensitivity for the detection of malignant lesions.

How to cite this article:

Aguirre L. M, Aguirre-Lam A, Lam-Maldonado M, et al. (July 01, 2026) Impact of Image Degradation on Melanoma Detection by Artificial Intelligence: A Patient Safety Perspective. *Cureus J Comput Sci* 3 : es44389-026-00147-2. DOI <https://doi.org/10.7759/s44389-026-00147-2>

Dataset, ethical considerations, and experimental partitioning

The foundational data corpus for this study was the public HAM10000 repository [5], hosted on the Harvard Dataverse. The total set comprises 10,015 dermatoscopic images corresponding to seven diagnostic categories validated through histopathological confirmation, clinical follow-up, or expert consensus. Given that the images and associated clinical metadata are strictly anonymized and in the public domain, this computational study was exempt from Institutional Review Board review. The complete Python codebase utilized for model training and simulated degradation experiments is publicly available to ensure reproducibility [12].

To evaluate the generalization capacity of the neural architectures and empirically quantify performance vulnerability to photographic noise, the total dataset D was partitioned into two mutually exclusive subsets: the training set D_{train} and the isolated test set D_{test} . The structural relationship of the partition is defined by the following spatial non-intersection condition:

$$D = D_{train} \cup D_{test} \quad \text{subject to} \quad D_{train} \cap D_{test} = \emptyset \tag{1}$$

The standardized split ratio was 80% for the synaptic weight optimization phase and 20% for inference evaluation. In terms of population cardinality, this results in

$$|D_{train}| = 8012, \quad |D_{test}| = 2003 \tag{2}$$

To prevent data leakage and guarantee the strict reproducibility of the experimental protocol across both architectures (ResNet-50 and MobileNetV2), the uniform random sampling $x_i \in D$ was controlled by initializing the pseudorandom number generator with a deterministic computational seed $S = 42$. This methodology ensured that the models were exposed to photographic perturbations using a subset of images rigorously unseen during training (Table 1).

Diagnostic Class	Abbreviation	Training Set	Test Set	Total
Melanocytic nevi	nv	5,366	1,339	6,705
Melanoma	mel	885	228	1,113
Benign keratosis	bkl	897	202	1,099
Basal cell carcinoma	bcc	409	105	514
Actinic keratosis	akiec	252	75	327
Vascular lesions	vasc	115	27	142
Dermatofibroma	df	88	27	115
Grand Total	-	8,012	2,003	10,015

TABLE 1: Population distribution of diagnostic classes in the training and test subsets.

AI architectures

To model the different operational environments in which teledermatology systems are currently deployed, two CNN architectures with distinct structural characteristics and computational demands were selected and evaluated: ResNet-50 and MobileNetV2.

ResNet-50 (Server Standard)

ResNet-50 was established as the control model representative of centralized clinical environments or cloud computing. This deep architecture of 50 layers solves the vanishing gradient problem through the implementation of shortcut connections that skip one or more layers. Mathematically, the residual network building block learns a mapping function defined as

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \{\mathbf{W}_i\}) + \mathbf{x} \quad (3)$$

where $\mathbf{x}$ and $\mathbf{y}$ represent the input and output vectors of the block, and $\mathcal{F}(\mathbf{x}, \{\mathbf{W}_i\})$ denotes the underlying residual mapping to be learned by the network [13]. This formulation allows the training of significantly deeper networks, achieving high theoretical accuracy metrics in the laboratory.

MobileNetV2 (Mobile Device Standard)

In contrast, MobileNetV2 was included as the experimental model representative of edge computing, being the predominant architecture in telemedicine applications executed directly on smartphones. Its design drastically reduces the number of parameters and computational load through the use of depthwise separable convolutions and inverted residual bottlenecks [14]. Its inclusion is essential to evaluate how an inherently compressed model responds to visual signal degradation.

Domain adaptation and loss function: Given that both architectures are preconfigured to classify 1,000 categories, an adaptation of the network topology for the dermatological domain was performed. The final fully connected layer of both models was replaced by a linear classifier with an output dimension equal to 7, corresponding to the diagnostic classes of the HAM10000 set. For the adjustment of synaptic weights during training, the categorical cross-entropy loss function was utilized, which is standard for mutually exclusive multiclass classification problems, calculated using the following equation:

$$\mathcal{L}_{CE} = - \sum_{i=1}^C t_i \log(p_i) \quad (4)$$

where C represents the total number of diagnostic classes $C = 7$, t_i is the binary indicator (1 if class i is the true label, 0 otherwise), and p_i is the probability predicted by the probabilistic output of the model (Softmax) for that class.

Experimental photographic degradation protocol

To systematically evaluate the vulnerability of the architectures to the domain shift characteristic of mobile teledermatology, an algorithmic visual perturbation protocol was designed and implemented. It is important to note that the mathematical transformations were applied exclusively to the test set ($\mathbf{D}_{test}$), maintaining the training set ($\mathbf{D}_{train}$) in its optimal original state. This design guarantees that the evaluation strictly measures the generalization capacity of the model against previously unobserved corrupted data, simulating the deployment of a clinical algorithm in an uncontrolled environment.

The computational perturbations were developed using the Torchvision (Meta Platforms, Inc., Menlo Park, CA) and Pillow (Python Software Foundation, Wilmington, DE) libraries within the PyTorch framework (Meta Platforms, Inc., Menlo Park, CA), modeling three suboptimal scenarios prevalent in image capture by patients.

How to cite this article:

Aguirre L. M, Aguirre-Lam A, Lam-Maldonado M, et al. (July 01, 2026) Impact of Image Degradation on Melanoma Detection by Artificial Intelligence: A Patient Safety Perspective. Cureus J Comput Sci 3 : es44389-026-00147-2. DOI <https://doi.org/10.7759/s44389-026-00147-2>

Spatial blur (simulation of kinetic tremor or focus failure): To emulate the loss of sharpness caused by the movement of the patient's hand (motion blur) or deficiencies in the device's autofocus system, a spatial smoothing filter was applied. This was achieved through the two-dimensional convolution of the image matrix with an isotropic Gaussian kernel. The spatial weighting function $G(x, y)$ with a standard deviation of $\sigma = 2.5$ pixels is defined as

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (5)$$

Clinically, this mathematical transformation acts as a low-pass filter, attenuating high spatial frequencies. This suppresses morphological characteristics essential for dermatological diagnosis, such as the irregularity of the lesion borders and the reticular structures of the pigment network.

Luminance reduction (simulation of poor ambient lighting): At-home photographs frequently lack the standardized lighting of a clinical dermatoscope. To model an underexposed capture, a linear attenuation transformation was applied to the color channels. Let $P_{in}(x, y)$ be the normalized intensity of the original pixel; the resulting pixel $P_{out}(x, y)$ was obtained by multiplying the signal by a scalar retention factor of $\alpha = 0.4$:

$$P_{out}(x, y) = \alpha \cdot P_{in}(x, y) \quad (6)$$

This 60% reduction in the dynamic range of the image severely decreases local contrast, hindering algorithmic differentiation between the pigmentation of the lesion (for example, the blue-gray tones characteristic of melanoma) and the surrounding skin tissue.

Sensor noise injection (simulation of low-fidelity hardware): Optical sensors of entry-level mobile phones, especially when operating at high ISO sensitivities to compensate for the lack of light, introduce random electronic noise into the image. To replicate this photographic artifact, additive white Gaussian noise was injected into the normalized tensor representation of the image. Defining $\mathcal{N}(\mu, \sigma^2)$ as a random variable drawn from a normal distribution with a mean of $\mu = 0$ and a standard deviation of $\sigma = 0.1$, the perturbation is defined as

$$P_{out}(x, y) = \max(\min(P_{in}(x, y) + \mathcal{N}(0, 0.1), 1), 0) \quad (7)$$

The clipping function, represented by the maximum and minimum operations, ensures that the resulting tensor values remain within the valid normalized color space of 0 to 1. The addition of this stochastic variance corrupts the image gradients, introducing artificial textures that can confound the feature extraction of the early convolutional layers.

The visual representation of this algorithmic protocol, detailing the progressive degradation of a standardized clinical sample, is illustrated in Figure 7.

How to cite this article:

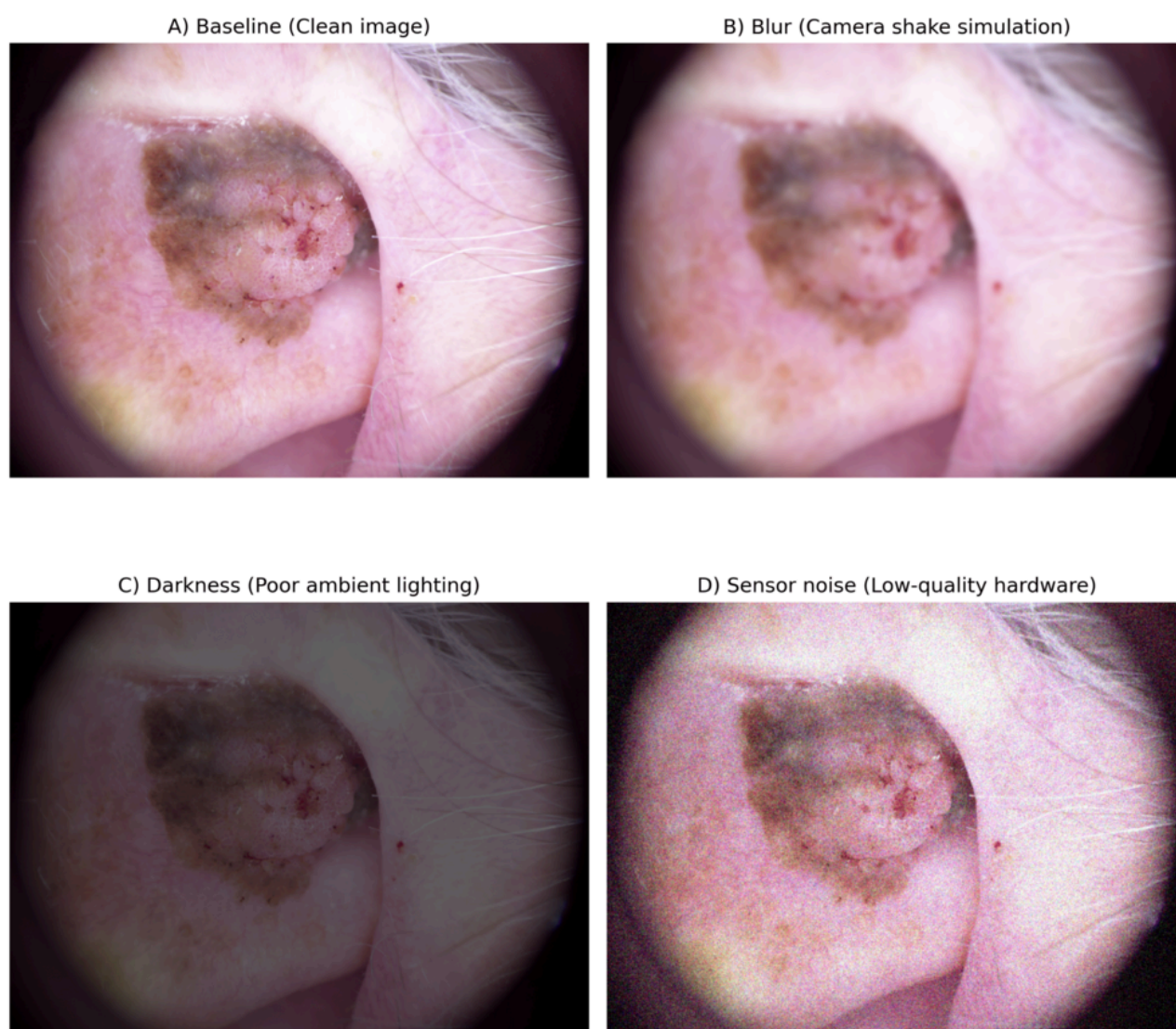


FIGURE 1: Simulation of photographic artifacts in teledermatology.

Simulation of photographic artifacts in teledermatology. All panels are original figures generated by the authors by applying the degradation pipeline developed in this study to a representative dermoscopic sample from the HAM10000 public dataset [5]. A) Baseline (clean image): original high-resolution dermoscopic sample. B) Blur (camera shake simulation): application of a Gaussian filter ($\sigma = 2.5$). C) Darkness (poor ambient lighting): luminance reduction to 40% of the original spectrum ($\alpha = 0.4$). D) Sensor noise (low-quality hardware): injection of additive white Gaussian noise ($\mu = 0, \sigma = 0.1$).

Statistical analysis and diagnostic evaluation metrics

The quantitative evaluation of the models' performance was conducted using the scikit-learn scientific computing library (scikit-learn developers). Given that the HAM10000 dataset presents a significant class imbalance (where melanocytic nevi represent approximately 67% of the total population), exclusive reliance on global accuracy was considered methodologically insufficient, as it tends to mask severe failures in minority but clinically urgent classes.

For a comprehensive analysis, metrics derived from the confusion matrix were calculated based on the counts of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The equations governing the evaluation of the models are as follows:

How to cite this article:

Aguirre L. M, Aguirre-Lam A, Lam-Maldonado M, et al. (July 01, 2026) Impact of Image Degradation on Melanoma Detection by Artificial Intelligence: A Patient Safety Perspective. Cureus J Comput Sci 3 : es44389-026-00147-2. DOI <https://doi.org/10.7759/s44389-026-00147-2>

Global accuracy: This measures the total proportion of correct predictions against the total number of evaluated cases. Although reported for comparative purposes, its clinical value is limited in the context of imbalanced oncological screening.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

Sensitivity (recall) as the primary endpoint: This metric constituted the most vital safety parameter of the study, specifically for melanoma detection. It represents the model's capacity to correctly identify subjects who actually present with the malignant pathology, minimizing the false-negative rate.

$$\text{Sensitivity (Recall)} = \frac{TP}{TP + FN} \quad (9)$$

Positive predictive value (precision): This calculates the probability that a lesion identified as positive by the model is genuinely an instance of that disease, reflecting diagnostic specificity against false alarms.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (10)$$

F1-score: This was used as a measure of the harmonic balance between precision and sensitivity, providing a comprehensive view of the model's stability in classes with high variability, such as basal cell carcinoma and actinic keratosis.

$$\text{F1-Score} = 2 \times \left(\frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \right) \quad (11)$$

The statistical analysis was systematically executed for the two architectures across the four experimental conditions. For each scenario, a detailed classification report and a confusion matrix were generated to visualize the redistribution of diagnostic errors under the effect of photographic artifacts [15,16].

Results And Discussion

All experiments were conducted within the PyTorch 2.1.0 framework with CUDA-enabled hardware acceleration. Both architectures - ResNet-50 and MobileNetV2 - were initialized with ImageNet-1K pretrained weights and fine-tuned for 5 epochs using the Adam optimizer ($\text{lr} = 1 \times 10^{-4}$) and categorical cross-entropy loss, with a batch size of 32. The HAM10000 dataset (10,015 images across 7 diagnostic classes) was strictly partitioned using a deterministic random seed ($S=42$) into a training subset (8,012 images; 80%) and an isolated test cohort (2,003 images; 20%), the latter remaining completely unseen during training. This test cohort was then subjected to three mathematical degradation pipelines - spatial blur ($\sigma=2.5$), luminance reduction ($\alpha=0.4$), and additive Gaussian noise ($\mu=0$, $\sigma=0.1$) - to systematically evaluate the structural stability of both architectures under identical conditions. Performance was quantified using Global Accuracy and Melanoma-Specific Sensitivity (Recall) as primary endpoints, computed via the scikit-learn 1.3.0 classification report.

The evaluation of the diagnostic performance of the ResNet-50 and MobileNetV2 architectures yielded significant discrepancies between global accuracy metrics and class-specific sensitivity when subjected to photographic perturbations. The quantitative results detailed below are based on the computational inference performed on the strictly isolated test set, guaranteeing that none of the evaluated images were previously processed during the model optimization phase.

To structure the clinical interpretation of the findings, the analysis has been divided into two fundamental analytical axes. First, the baseline performance of both models is established under control conditions (high-fidelity clinical images). Subsequently, the impact of the domain shift is quantified by applying the three simulated degradation artifacts: spatial

How to cite this article:

blur, luminance reduction, and sensor noise. The primary focus of the metric report lies in the behavior of false negatives and the consequent collapse of diagnostic sensitivity in melanoma detection.

Diagnostic performance under baseline conditions

Before the systematic introduction of photographic artifacts, the baseline performance of both architectures was determined using the original test set D_{test} , characterized by high-fidelity dermatoscopic images with standardized lighting and focus. The results obtained in this phase represent the upper limit of diagnostic performance achievable by the models under laboratory conditions.

The ResNet-50 architecture demonstrated superior diagnostic stability, achieving an overall accuracy of 95.8%. In terms of clinical relevance, this model achieved a sensitivity (recall) for melanoma detection of 88.6%, suggesting an advanced morphological discrimination capacity under optimal conditions. Conversely, the MobileNetV2 architecture, optimized for processing on mobile devices, exhibited significantly lower performance. Although its overall accuracy stood at 84.7%, its specific sensitivity for melanoma was considerably reduced, reaching only 64.0%.

This initial contrast is important: MobileNetV2, as the architecture designed for mobile environments, fundamentally presents an oncological false-negative rate exceeding 35%, even with professional-quality images. These findings suggest that, in this specific architecture, structural compression (via depthwise separable convolutions) might sacrifice the capacity to extract subtle features necessary for the identification of malignancy. The complete comparative data are presented in Table 2, and the visual disparity between accuracy and sensitivity is illustrated in Figure 2.

Architectural Model	Evaluation Scenario	Global Accuracy	Melanoma Sensitivity (Recall)
ResNet-50	Baseline	95.8%	88.6%
ResNet-50	Moderate blur	68.5%	1.3%
ResNet-50	Luminance reduction	67.0%	0.9%
ResNet-50	Sensor noise	51.8%	6.1%
MobileNetV2	Baseline	84.7%	64.0%
MobileNetV2	Moderate blur	67.3%	0.0%
MobileNetV2	Luminance reduction	67.6%	8.3%
MobileNetV2	Sensor noise	66.8%	0.0%

TABLE 2: Comparative analysis of diagnostic performance and oncological sensitivity under photographic degradation conditions.

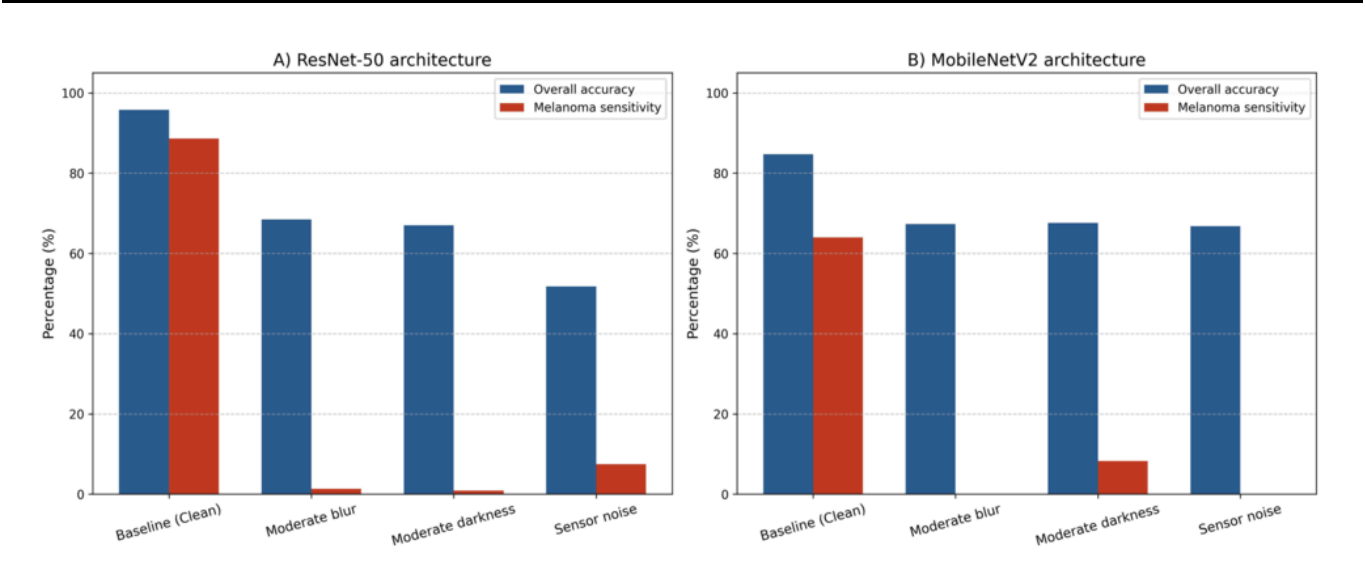


FIGURE 2: Accuracy discrepancy between overall diagnostic accuracy and class-specific sensitivity for melanoma detection.

(A) Performance metrics for ResNet-50. (B) Performance metrics for MobileNetV2. While overall accuracy remains artificially stable across most conditions, melanoma recall drops to near-zero levels under image degradation.

Impact of degradation on diagnostic sensitivity

Upon subjecting the test set to controlled photographic artifacts, the analysis of the confusion matrices exposed a severe vulnerability to domain shift. A profound discrepancy was identified between the apparent stability of the global accuracy and the actual clinical performance of the models, which suffered an algorithmic collapse characterized by a systematic bias towards the majority class (melanocytic nevi, nv) [17].

Effects of Spatial Blur and Luminance Reduction

Under optimal control conditions, the ResNet-50 architecture successfully extracted the morphological features of melanoma, identifying 202 of the 228 confirmed cases, representing a baseline sensitivity of 88.6%. However, the application of moderate blur ($\sigma = 2.5$) acted as a low-pass filter that suppressed the high spatial frequencies necessary for detecting irregular borders. As a result, of the 228 evaluated melanomas, 221 cases were erroneously classified as benign nevi. This massive redistribution reduced the class sensitivity to 1.3% (Table 3), even though the overall accuracy metric of the model remained deceptively stable at 68.5% [18].

How to cite this article:

Experimental Condition	Correct Hits (mel)	False Negatives (nv)	False Negatives (bkl)	Other Errors	Sensitivity
Baseline (Control)	202	21	5	0	88.6%
Moderate blur	3	221	0	4	1.3%
Luminance reduction	2	226	0	0	0.9%
Sensor noise	14	68	146	0	6.1%

TABLE 3: Distribution of predictions and major diagnostic errors for the melanoma class under various experimental conditions (ResNet-50).

An analogous diagnostic degradation was documented under luminance reduction ($\alpha = 0.4$). The loss of chromatic contrast caused 226 of the 228 malignant lesions to be diagnosed as benign tissue, achieving a mere 0.9% sensitivity (two correct predictions). The literature suggests that upon the suppression of variance in the color channels, CNNs abandon active feature extraction and revert to the prior probability distribution of the training dataset [19].

Impact of Sensor Noise Injection

The injection of additive white Gaussian noise severely perturbed the local gradients of the visual tensors. Under this scenario, ResNet-50 retained only a 6.1% sensitivity for melanoma (14 detected cases). Clinically, the error pattern differed significantly from the previous scenarios: 68 melanomas were confused with benign nevi, while 146 malignant cases were erroneously categorized as benign keratoses (bkl). This specific shift suggests that the stochastic noise may have introduced artificial granular textures that the algorithm interpreted as consistent with the typical morphological features of keratotic lesions [18].

Vulnerability Observed in a Compressed Mobile-Oriented Architecture (MobileNetV2)

The performance of the MobileNetV2 architecture exhibited a clinical deterioration of absolute magnitude. Despite maintaining global accuracies above 66% across all degradation scenarios, the model registered a sensitivity of 0.0% under blur and sensor noise (Table 2). The total inability to identify a single case of malignancy under mild perturbations suggests that, in this specific architecture, parametric compression via depthwise separable convolutions may disproportionately sacrifice the stability of fine feature maps [20]. As only one lightweight architecture was evaluated in this study, this finding should not be extrapolated to mobile-oriented architectures as a category; it remains an open question whether this vulnerability reflects a general property of compressed networks or a characteristic specific to MobileNetV2's particular topology.

The unidirectional diagnostic shift and the consequent annulment of patient safety triage are conclusively illustrated in the actual confusion matrices of the study (Figure 3).

How to cite this article:

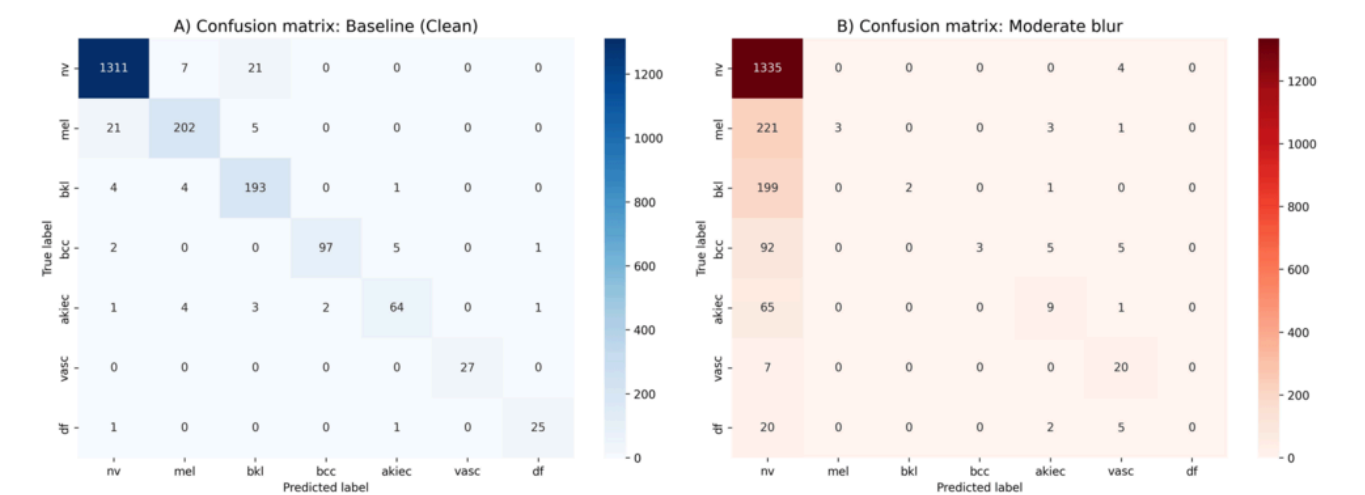


FIGURE 3: Diagnostic shift under image degradation.

Comparative confusion matrices for the ResNet-50 architecture. (A) Under baseline (clean) conditions, the model correctly identifies 202 out of 228 melanomas. (B) Under moderate spatial blur, the feature maps collapse, causing a systematic diagnostic shift where 221 malignant melanomas are severely misclassified as benign melanocytic nevi (nv).

Discussion

The integration of deep learning architectures in teledermatology promises to democratize access to early oncological screening, reducing geographical barriers and optimizing clinical referral times. However, as evidenced by the preceding performance metrics, the high diagnostic accuracy reported by these algorithms in controlled laboratory environments does not fundamentally translate into operational safety under real-world photographic variability conditions.

Through the mathematical simulation of visual artifacts prevalent in image capture by mobile devices, we have documented a severe algorithmic collapse in which the studied CNNs abandon the extraction of morphological features and revert to predictions based on the prior probability of the majority class.

To frame the impact of this phenomenon within current medical practice, our analysis of these findings has been structured into four fundamental axes. First, a technical interpretation of the data is offered, defining the accuracy discrepancy and analyzing the severe structural vulnerability discovered in MobileNetV2. Subsequently, the clinical implications of these algorithmic biases are examined, evaluating the ethical risk and morbidity associated with the systematic generation of false negatives in melanoma screening. Based on this analysis, we propose a framework of software engineering recommendations to mitigate these risks in commercial telemedicine applications. Finally, the scope of our conclusions is delineated through the analysis of the methodological limitations of the adopted experimental design.

Structural Vulnerability of MobileNetV2

The analysis of the evaluation metrics under visual degradation conditions exposes a statistical misconception fundamental to the isolated use of global accuracy in the training of artificial intelligence with imbalanced clinical datasets. The HAM10000 corpus presents a natural and significant population bias, where melanocytic nevi represent approximately 67% of the samples. When photographic perturbations (such as spatial blur or darkness) suppress gradients and chromatic contrast, neural networks lose the capacity to discriminate subtle morphological patterns, such as atypical pigment networks or the blue-white veils characteristic of melanoma.

How to cite this article:

Faced with this informational uncertainty, the redistribution of predictions observed across confusion matrices is consistent with a state of feature collapse, in which the model would abandon the extraction of discriminative visual features and revert toward the prior statistical probability of the training set, systematically predicting the majority class. This interpretation aligns with prior literature on texture bias and class-imbalance-driven prediction shift in convolutional neural networks [7,19]; however, it was not directly validated in the present study through explainability techniques (e.g., Grad-CAM, saliency mapping, or analysis of the softmax output distribution), and should therefore be regarded as a plausible mechanistic explanation rather than a confirmed finding. This phenomenon constitutes an accuracy discrepancy: the model maintains a deceptively stable and theoretically acceptable global performance (ranging between 67% and 68%) simply by diagnosing all lesions as benign nevi, while its actual clinical utility (sensitivity to melanoma) plummets to levels near absolute zero.

Beyond the general impact of photographic artifacts, the results reveal an extreme and disproportionate vulnerability in MobileNetV2 as an architecture optimized for mobile devices. While ResNet-50, a parametrically dense network, demonstrated superior baseline performance and retained a marginal detection capacity under additive noise (6.1%), MobileNetV2 experienced a total diagnostic annulment (0.0% sensitivity) when faced with blur and sensor noise.

The technical explanation for this susceptibility originates from the network topology itself. MobileNetV2 achieves its high computational efficiency and low memory consumption by replacing standard convolutions with depthwise separable convolutions, drastically reducing floating-point operations [8]. However, this severe parameter compression sacrifices redundancy in feature representation [21]. Lacking the synaptic redundancy and massive residual connections of server-standard architectures like ResNet-50, the feature maps extracted by MobileNetV2 are fundamentally unstable. When the input signal is corrupted in an edge computing environment (such as a patient's mobile phone), these compressed representations rapidly degrade to the point of inoperability, turning the mobile application into a potential risk for clinical screening.

Clinical Risks and False Reassurance

Translating the previously described algorithmic vulnerabilities into a real clinical environment poses an unacceptable risk scenario for patient safety. In the current teledermatology ecosystem, particularly in DTC applications, photographic quality is fundamentally variable, user-dependent, and uncontrollable [22].

When a patient captures a suboptimal image of an incipient melanoma, whether due to kinetic tremor, poor ambient lighting, or sensor noise, the diagnostic collapse of artificial intelligence towards the majority class of benign nevi generates an evaluation of false reassurance. Clinically, the issuance of a false negative represents the most severe systemic error that an oncological screening system can commit. Unlike a false positive, which routinely leads to an in-person dermoscopy or the excision of a benign lesion with limited anatomical and psychological morbidity, a false negative completely interrupts the medical care pathway, dissuading the patient from seeking specialized evaluation.

In the tumor biology of melanoma, the most determining prognostic factor for survival is the tumor thickness at the time of excision, known as the Breslow depth. Melanoma typically evolves from a radial growth phase within the epidermis to a vertical growth phase with dermal invasion. A diagnostic delay of just a few months, motivated by the false security provided by a mobile application algorithm, provides the necessary time window for neoplastic cells to reach the dermal lymphatic and vascular plexuses. The oncological evidence is conclusive: the overall five-year survival drastically decreases from over 98% in localized melanoma in situ to less than 30% in the presence of distant metastasis [23].

Consequently, the commercial deployment of AI models with vulnerable architectures that lack stability against visual degradation does not democratize early diagnosis; on the contrary, it induces algorithmic iatrogenesis. Automation bias, which is the psychological tendency of humans to blindly trust the results of computational systems, exacerbates this situation. In this context of silent failure, the technology ceases to be an adjunct to the physician and becomes a major barrier that delays timely biopsy, directly impacting patient morbidity and mortality.

Quality Gate Implementation

How to cite this article:

Aguirre L. M, Aguirre-Lam A, Lam-Maldonado M, et al. (July 01, 2026) Impact of Image Degradation on Melanoma Detection by Artificial Intelligence: A Patient Safety Perspective. Cureus J Comput Sci 3 : es44389-026-00147-2. DOI <https://doi.org/10.7759/s44389-026-00147-2>

The quantitative demonstration of the extreme vulnerability of AI models, specifically compressed mobile architectures like MobileNetV2, to mild visual perturbations emphasizes the imperative need to shift from a model-centric vision toward a data-centric AI approach in teledermatology engineering [24]. Given the severity of the clinical implications of the documented oncological false negatives, mobile-device-based AI screening systems must not operate independently of the input signal quality.

Consequently, we propose a fundamental change in the software architecture of direct-to-consumer telemedicine: the mandatory and standardized implementation of a Quality Gate that executes locally on the edge computing device before activating any diagnostic inference (Figure 4).

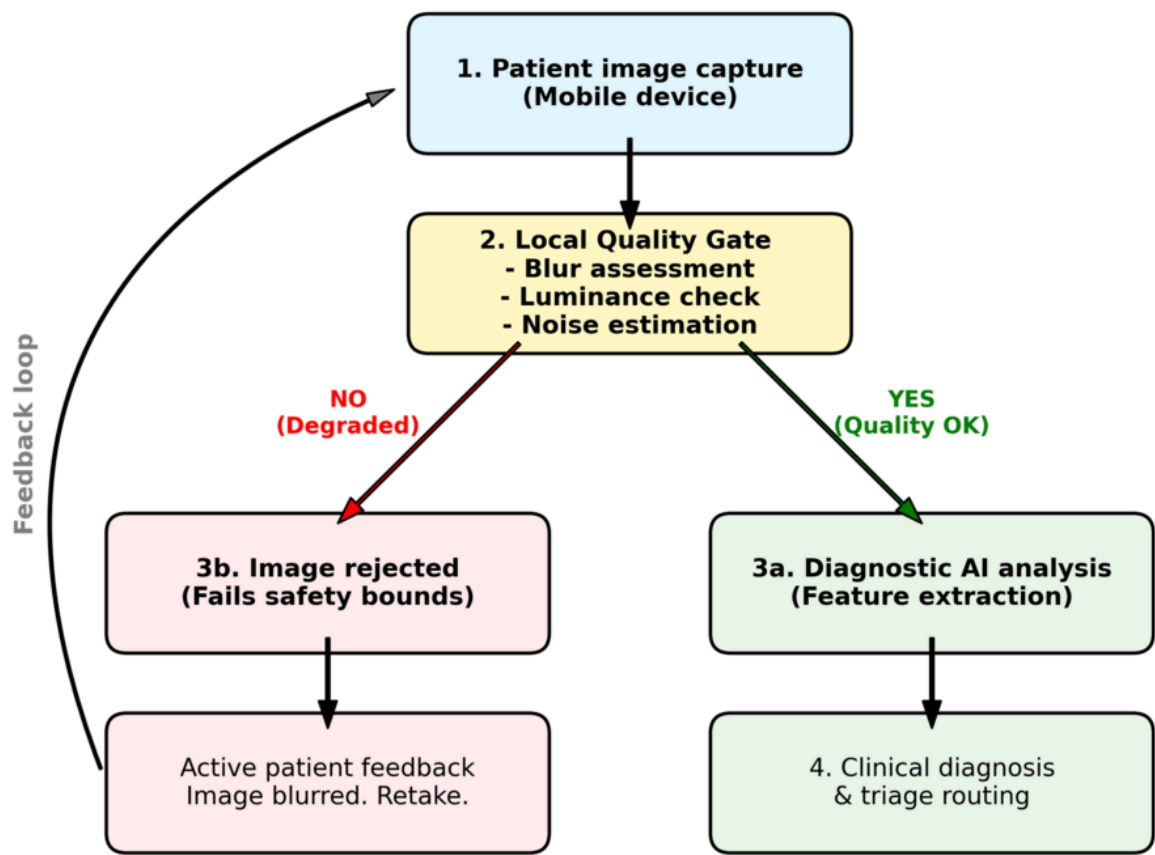


FIGURE 4: Conceptual framework for a patient-safe mobile teledermatology application architecture.

To prevent automated false reassurance, an essential "Image Quality Validator" must execute locally on the mobile device. This active sensor layer forms a "Quality Gate" that assesses real-time sharpness, luminance, and noise levels. Diagnostic AI analysis (A) is only initiated if the image meets pre-defined safety bounds (B); otherwise, active, specific guidance is provided to the patient to re-acquire the image (C). This ensures model reliability by rejecting degraded inputs.

The quality validation mechanism: This Quality Gate consists of an algorithmic image quality assessment module that acts as a binary safety filter. This module must stochastically evaluate the same degradation parameters studied in this research.

How to cite this article:

Aguirre L. M, Aguirre-Lam A, Lam-Maldonado M, et al. (July 01, 2026) Impact of Image Degradation on Melanoma Detection by Artificial Intelligence: A Patient Safety Perspective. Cureus J Comput Sci 3 : es44389-026-00147-2. DOI <https://doi.org/10.7759/s44389-026-00147-2>

Sharpness validation: The implementation of traditional computer vision metrics, such as the variance of the Laplacian operator, is required to quantify spatial blur. If the sharpness score falls below a safe operational threshold (conceptually represented by our blur radius of 2.5), the image must be rejected immediately.

Luminance validation: A chromatic histogram analysis must be performed to ensure that pixel intensity and local contrast remain within acceptable diagnostic ranges, avoiding the diagnostic collapses observed under the 60% attenuation of ambient luminosity.

Sensor noise management: Although more complex to quantify without a reference, the software must use local noise variance estimators to identify captures made with low-fidelity hardware under adverse ISO conditions.

Patient feedback loop: The user interface must not present an AI error or a generic diagnosis to the patient when faced with a degraded image. On the contrary, it must provide active and specific guidance to rectify the failure in real time. As conceptualized in Figure 4, if an image is rejected due to blur, the system should instruct the patient: "Image not accepted. The photo is too blurry due to movement. Please hold the phone steady and try the capture again".

This architecture will not only shield patient safety by avoiding automated oncological false reassurance but will also serve to educate the end user on the proper capture of dermatological images, improving the overall data quality in the digital health ecosystem.

Study Limitations and Future Directions

Although the outcomes of this research provide a detailed quantification of the instability of neural networks in the face of visual domain shift, it is essential to interpret the results in the context of their methodological limitations.

The primary limitation lies in the computational nature of the experimental design. The evaluated visual perturbations (blur, luminance attenuation, and sensor noise) were synthetically injected through isolated mathematical transformations. Although this algorithmic approach guarantees strict control over independent variables and allows for exact reproducibility, the generated degraded images do not capture the entirety of real-world optical complexity. In the daily clinical practice of mobile teledermatology, photographic artifacts rarely present in isolation; they are typically compound and synergistic phenomena (for example, motion blur combined with poor peripheral lighting, specular reflections from the flash, and hair occlusion) [25]. It is highly probable that the diagnostic collapse reported in this study underestimates the actual failure rate that these models would experience when faced with the non-linear variability of photographs taken directly by patients.

Secondly, the baseline performance was established using the HAM10000 dataset. This corpus is composed exclusively of dermoscopic images, which were originally captured by specialists using optical equipment with standard magnification, cross-polarized lighting, and immersion lenses. Although applying degradation algorithms to these high-fidelity images served to demonstrate feature collapse, a more representative evaluation of the risk in direct-to-consumer applications would require the use of native clinical teledermatology datasets or macrophotographs taken directly with smartphones [26].

Thirdly, the architectural scope of the evaluation was limited to a single lightweight, mobile-oriented architecture (MobileNetV2). Consequently, it is not possible to determine whether the observed vulnerability to image degradation generalizes to the broader category of edge-computing architectures (e.g., MobileNetV3, EfficientNet-Lite, ShuffleNet) or is instead a characteristic specific to MobileNetV2's depthwise separable convolution and inverted residual bottleneck design. Future studies should replicate this protocol across multiple lightweight architectures to determine whether a common structural pattern of vulnerability exists.

Fourthly, the mechanistic explanations proposed in the Discussion regarding the internal behavior of the evaluated networks under degradation, namely, feature collapse and reversion toward the prior probability of the majority class, were inferred from the redistribution patterns observed in the confusion matrices, and are consistent with prior literature

How to cite this article:

Aguirre L. M, Aguirre-Lam A, Lam-Maldonado M, et al. (July 01, 2026) Impact of Image Degradation on Melanoma Detection by Artificial Intelligence: A Patient Safety Perspective. Cureus J Comput Sci 3 : es44389-026-00147-2. DOI <https://doi.org/10.7759/s44389-026-00147-2>

on texture bias in CNNs [19]. However, these explanations were not directly validated in this study through explainability techniques such as Grad-CAM, saliency mapping, or quantitative analysis of the softmax output distribution. Future research should incorporate such analyses to confirm the underlying mechanism driving the observed diagnostic shift.

Finally, being a retrospective study based on a static dataset, it does not evaluate the impact of the proposed feedback loop in real-world scenarios. Future research should focus on prospective clinical trials that evaluate the efficacy of quality gates integrated into mobile applications. It will be vital to determine not only whether these algorithmic filters successfully reject suboptimal images and mitigate false negatives, but also to evaluate patient adherence and behavioral response to automated warnings of poor photographic quality. Furthermore, future research should investigate whether aggressive data augmentation techniques, specifically training models with synthetically blurred or noisy images, can improve architectural stability, although this does not negate the necessity of a quality gate.

Conclusions

The integration of artificial intelligence in teledermatology represents a promising advancement for early skin cancer detection; however, this study quantitatively demonstrates that high diagnostic accuracy on pristine datasets does not guarantee operational clinical safety. Through the controlled simulation of three photographic artifacts - spatial blur ($\sigma=2.5$), luminance reduction ($\alpha=0.4$), and sensor noise ($\sigma=0.1$) - on the HAM10000 benchmark, we identified and quantified the accuracy discrepancy: both ResNet-50 and MobileNetV2 maintained deceptively stable global accuracy ($\geq 51.8\%$) while melanoma-specific sensitivity collapsed from baseline values of 88.6% and 64.0% to near-zero levels (1.3%-6.1% and 0.0%-8.3%, respectively). This systematic false-negative generation constitutes a direct patient safety threat, delaying critical oncological care and enabling neoplastic progression. As the principal contribution of this work, we propose a data-centric framework: the mandatory integration of automated Local Quality Gates on mobile devices that algorithmically evaluate image sharpness, luminance, and noise prior to diagnostic inference - ensuring artificial intelligence operates as a safe and reliable screening tool rather than a point of silent clinical failure.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Marco A. Aguirre L., Alan G. Aguirre-Lam, Mayda Lam-Maldonado, J. F. Franklin Rodríguez-Zapata

Acquisition, analysis, or interpretation of data: Marco A. Aguirre L., Alan G. Aguirre-Lam

Drafting of the manuscript: Marco A. Aguirre L., Alan G. Aguirre-Lam

Critical review of the manuscript for important intellectual content: Marco A. Aguirre L., Alan G. Aguirre-Lam, Mayda Lam-Maldonado, J. F. Franklin Rodríguez-Zapata

Supervision: Mayda Lam-Maldonado, J. F. Franklin Rodríguez-Zapata

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

How to cite this article:

Aguirre L. M, Aguirre-Lam A, Lam-Maldonado M, et al. (July 01, 2026) Impact of Image Degradation on Melanoma Detection by Artificial Intelligence: A Patient Safety Perspective. *Cureus J Comput Sci* 3 : es44389-026-00147-2. DOI <https://doi.org/10.7759/s44389-026-00147-2>

Data Availability Statements

The data (and/or code) supporting this study are openly available in GitHub at <https://github.com/experiments-cloud/cureus>, reference HAM10000.

Acknowledgements

The authors would like to thank the Tecnológico Nacional de México (TECNM) for the facilities provided to carry out this research, as well as the Laboratorio Nacional de Tecnologías de la Información (LaNTI) for the technical support and infrastructure provided for the development of the project.

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A: [Cancer statistics, 2022](#). CA: A Cancer Journal for Clinicians. 2022, 72:7-33. [10.3322/caac.21708](#)
2. Coates SJ, Kvedar J, Granstein RD: [Teledermatology: from historical perspective to emerging techniques of the modern era: part I: History, rationale, and current practice](#). Journal of the American Academy of Dermatology. 2015, 72:563-574. [10.1016/j.jaad.2014.07.061](#)
3. Chuchu N, Takwoingi Y, Dinnes J, et al.: [Smartphone applications for triaging adults with skin lesions that are suspicious for melanoma](#). Cochrane Database of Systematic Reviews. 2018, 12:CD013192. [10.1002/14651858.cd013192](#)
4. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S: [Dermatologist-level classification of skin cancer with deep neural networks](#). Nature. 2017, 542:115-118. [10.1038/nature21056](#)
5. Tschandl P, Rosendahl C, Kittler H: [The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions](#). Scientific Data. 2018, 5:180161. [10.1038/sdata.2018.161](#)
6. Navarrete-Dechent C, Dusza SW, Liopyris K, Marghoob AA, Halpern AC, Marchetti MA: [Automated dermatological diagnosis: hype or reality?](#). Journal of Investigative Dermatology. 2018, 138:2277-2279. [10.1016/j.jid.2018.04.040](#)
7. Buda M, Maki A, Mazurowski MA: [A systematic study of the class imbalance problem in convolutional neural networks](#). Neural Networks. 2018, 106:249-259. [10.1016/j.neunet.2018.07.011](#)
8. Howard AG, Zhu M, Chen B, et al.: [MobileNets: efficient convolutional neural networks for mobile vision applications](#). arXiv. 2017, [10.48550/arXiv.1704.04861](#)
9. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D: [Key challenges for delivering clinical impact with artificial intelligence](#). BMC Medicine. 2019, 17:195. [10.1186/s12916-019-1426-2](#)
10. Giansanti D: [The artificial intelligence in teledermatology: a narrative review on opportunities, perspectives, and bottlenecks](#). International Journal of Environmental Research and Public Health. 2023, 20:5810. [10.3390/ijerph20105810](#)
11. Marchetti MA, Codella NCF, Dusza SW, et al.: [Results of the 2016 International Skin Imaging Collaboration International Symposium on Biomedical Imaging challenge: Comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images](#). Journal of the American Academy of Dermatology. 2018, 78:270-277. [10.1016/j.jaad.2017.08.016](#)
12. GitHub: [Source code and experimental data for Impact of image degradation on melanoma detection](#). (2026). Accessed: June 7, 2026: <https://github.com/experiments-cloud/cureus>.
13. He K, Zhang X, Ren S, Sun J: [Deep residual learning for image recognition](#). 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA. 2016, 770-778. [10.1109/CVPR.2016.90](#)
14. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC: [MobileNetV2: inverted residuals and linear bottlenecks](#). 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA. 2018, 4510-4520. [10.1109/CVPR.2018.00474](#)
15. Hicks SA, Strümke I, Thambawita V, et al.: [On evaluation metrics for medical applications of artificial intelligence](#). Scientific Reports. 2022, 12:5979. [10.1038/s41598-022-09954-8](#)
16. Chicco D, Jurman G: [The advantages of the Matthews correlation coefficient \(MCC\) over F1 score and accuracy in binary classification evaluation](#). BMC Genomics. 2020, 21:6. [10.1186/s12864-019-6413-7](#)

How to cite this article:

Aguirre L. M, Aguirre-Lam A, Lam-Maldonado M, et al. (July 01, 2026) Impact of Image Degradation on Melanoma Detection by Artificial Intelligence: A Patient Safety Perspective. Cureus J Comput Sci 3 : es44389-026-00147-2. DOI <https://doi.org/10.7759/s44389-026-00147-2>

17. Finlayson SG, Bowers JD, Ito J, Zittrain JL, Beam AL, Kohane IS: [Adversarial attacks on medical machine learning](#). Science. 2019, 363:1287-1289. [10.1126/science.aaw4399](#)
18. Dodge S, Karam L: [Understanding how image quality affects deep neural networks](#). 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX), Lisbon, Portugal. 2016, 1-6. [10.1109/QoMEX.2016.7498955](#)
19. Geirhos R, Rubisch P, Michaelis C, Bethge M, Wichmann FA, Brendel W: [ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness](#). arXiv. 2022, [10.48550/arXiv.1811.12231](#)
20. Hooker S, Courville A, Clark G, Dauphin Y, Frome A: [What do compressed deep neural networks forget?](#). arXiv. 2019, [10.48550/arXiv.1911.05248](#)
21. Kornblith S, Shlens J, Le QV: [Do better ImageNet models transfer better?](#). 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA. 2019, 2656-2666. [10.1109/CVPR.2019.00277](#)
22. Jalaboi R, Winther O, Galimzianova A: [Explainable image quality assessments in teledermatological photography](#). Telemedicine and e-Health. 2023, 29:1342-1348. [10.1089/tmj.2022.0405](#)
23. Balch CM, Gershenwald JE, Soong SJ, et al.: [Final version of 2009 AJCC melanoma staging and classification](#). Journal of Clinical Oncology. 2009, 27:6199-6206. [10.1200/jco.2009.23.4799](#)
24. Sambasivan N, Kapania S, Highfill H, et al.: ["Everyone wants to do the model work, not the data work": data cascades in high-stakes AI](#). CHI '21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021, 1-15. [10.1145/3411764.3445518](#)
25. Pérez E, Reyes O, Ventura S: [Convolutional neural networks for the automatic diagnosis of melanoma: An extensive experimental study](#). Medical Image Analysis. 2021, 67:101858. [10.1016/j.media.2020.101858](#)
26. Pacheco AGC, Lima GR, Salomão AS, et al.: [PAD-UFES-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones](#). Data in Brief. 2020, 32:106221. [10.1016/j.dib.2020.106221](#)

How to cite this article:

Aguirre L. M, Aguirre-Lam A, Lam-Maldonado M, et al. (July 01, 2026) Impact of Image Degradation on Melanoma Detection by Artificial Intelligence: A Patient Safety Perspective. Cureus J Comput Sci 3 : es44389-026-00147-2. DOI <https://doi.org/10.7759/s44389-026-00147-2>