ORIGINAL ARTICLE

European Archives of Medical Research

DOI: 10.14744/eamr.2026.03206
Eur Arch Med Res 2026;42(2):188–196

# Machine Learning–Based Ultrasound Radiomics for Differentiating Benign and Malignant Skin Lesions: A Comparative Feature-Set and Model Performance Analysis

Hakan Ayyildiz

Department of Radiology, Basaksehir Cam and Sakura City Hospital, Istanbul, Türkiye

## ABSTRACT

**Objective:** This study aimed to assess the ability of radiomics features extracted from grayscale B-mode dermatologic ultrasound images to distinguish benign from malignant skin lesions and to compare the performance of multiple machine learning (ML) classifiers trained on different radiomics feature domains, including original, Laplacian-of-Gaussian (LoG), and wavelet-transformed images.

**Materials and Methods:** This retrospective study analyzed 190 skin lesions (129 benign, 61 malignant) from a publicly available dermatologic ultrasound dataset. Manual lesion segmentation was performed by an experienced radiologist. Radiomics features were extracted from original grayscale, LoG-filtered, and wavelet-transformed images using PyRadiomics. Low-variance descriptors and highly correlated features were removed, and mutual information–based feature selection was conducted within stratified five-fold cross-validation to avoid information leakage. Logistic regression, support vector machine (RBF), and random forest classifiers were trained using balanced class weights. Performance metrics were computed by pooling predictions from all folds.

**Results:** A total of 851 radiomics features per lesion were initially extracted and reduced to 143 non-redundant descriptors after preprocessing. Across all models and feature domains, the area under the receiver operating characteristic curve (AUC) ranged from 0.64 to 0.71. The highest performance was achieved by the random forest classifier using the combined feature set (AUC: 0.706; accuracy: 0.716). Wavelet-only features also performed well, with random forest achieving an AUC of 0.705 and the highest accuracy (0.742). Logistic regression applied to original features yielded an AUC of 0.697 and comparatively better sensitivity. Specificity was consistently higher than sensitivity across classifiers.

**Conclusion:** Radiomics-based analysis of dermatologic ultrasound provides moderate discriminatory ability for benign–malignant classification. Although limited by dataset size and manual segmentation, the findings support radiomics as a quantitative, operator-independent adjunct for ultrasound-based skin lesion evaluation.

**Keywords:** Machine Learning, Radiomics, Skin Neoplasms, Ultrasonography

**Cite this article as:** Ayyildiz H. Machine Learning–Based Ultrasound Radiomics for Differentiating Benign and Malignant Skin Lesions: A Comparative Feature-Set and Model Performance Analysis. Eur Arch Med Res 2026;42(2):188–196.

**Address for correspondence:** Hakan Ayyildiz. Department of Radiology, Basaksehir Cam and Sakura City Hospital, Istanbul, Türkiye
**E-mail:** hakanayyildiz77@gmail.com **ORCID ID:** 0000-0003-0244-7765

**Submitted:** 27.12.2025 **Revised:** 11.01.2026 **Accepted:** 12.01.2026 **Available Online:** 03.06.2026

European Archives of Medical Research – Available online at www.eurarchmedres.org

## INTRODUCTION

Skin cancer is one of the most rapidly increasing malignancies worldwide, and early diagnosis remains critical for improving patient outcomes.[1] Although dermoscopy and clinical examination are widely used in the initial evaluation of suspicious lesions, imaging modalities such as ultrasound (US) have gained increasing importance due to their accessibility, lack of ionizing radiation, real-time capability, and ability to visualize subcutaneous extension.[2] High-frequency ultrasound, in particular, provides valuable morphological and structural information that assists clinicians in differentiating benign from malignant cutaneous lesions.[3] However, sonographic assessment is inherently operator-dependent, and subtle textural patterns that may be diagnostic are often beyond the threshold of human perception.[3]

Radiomics offers a solution to these limitations by converting medical images into high-dimensional, mineable quantitative data.[4] Through the extraction of texture, intensity, and wavelet-based features, radiomics can reveal microstructural heterogeneity that correlates with the biological characteristics of tumors. Although radiomics has been extensively studied in computed tomography and magnetic resonance imaging, the number of studies applying radiomics analysis to ultrasound for skin lesions remains very limited.[5] The acoustic properties and speckle patterns inherent to ultrasound present a unique opportunity for texture-based characterization but also introduce technical challenges that require careful feature extraction and model optimization.[6]

In recent years, machine learning (ML) algorithms have demonstrated promising performance across various oncologic imaging tasks.[7] However, the optimal feature domains and model types for ultrasound-based skin lesion radiomics remain unclear. Prior studies have typically focused on handcrafted features from original images without systematically exploring the impact of filtered feature sets, such as Laplacian-of-Gaussian (LoG) or wavelet transformations, which may provide additional discriminatory power by emphasizing different textural scales.[8] Moreover, comparative evaluations of multiple machine learning classifiers applied to different radiomics feature domains in skin lesion ultrasound imaging are scarce.[9]

Given these gaps, there is a need for a comprehensive analysis that examines how different radiomics feature sets and machine learning models influence diagnostic performance in distinguishing benign from malignant skin lesions. Such an approach would not only provide insights into the most informative texture domains of ultrasound images but also highlight which ML methods are most robust for this classification task.

Therefore, this study aimed to evaluate and compare the performance of multiple radiomics feature groups, including original, Laplacian-of-Gaussian-filtered, and wavelet-transformed features, combined with different machine learning classifiers for differentiating benign and malignant skin lesions on ultrasound. By systematically assessing these combinations using a standardized cross-validated framework, we sought to identify effective radiomics–machine learning strategies for non-invasive lesion characterization.

## MATERIALS AND METHODS

### Dataset and Study Design

This retrospective radiomics analysis was performed using an openly available dermatologic ultrasound dataset published by Laverde-Saad et al.,[10] and hosted on the Kaggle platform.[11] The dataset consists of grayscale B-mode and color Doppler dermatologic ultrasound images acquired according to DERMUS guidelines. A total of 235 surgically excised or biopsied skin lesions were scanned, and seven diagnostic categories representing the most common benign and malignant cutaneous lesions were included. All images were de-identified before release and contained no patient-identifying information. For the present study, grayscale B-mode images and corresponding lesion masks were used; cases without grayscale B-mode ultrasound images or complete data were excluded, resulting in 190 lesions (129 benign, 61 malignant) for final analysis.

All ultrasound scans in the source dataset were acquired using a single ultrasound platform (MyLab Class C, Esaote, Geneva) with 10–22 MHz or 18 MHz linear probes to maintain image homogeneity, as described in the original publication.[10]

The images were obtained from a publicly available and fully de-identified dataset; therefore, no institutional ethical approval or informed consent was required for this study.

### Lesion Segmentation

All images were segmented manually by a radiologist with 8 years of experience in diagnostic ultrasound. Segmentation was performed in a blinded manner, without any knowledge of the histopathological diagnosis (benign or malignant) of the lesions, to prevent reader bias.

Segmentation was carried out using 3D Slicer software. A freehand region of interest (ROI) was drawn on the grayscale B-mode ultrasound image containing the largest cross-sectional diameter of the lesion. The ROI was carefully adjusted to include the entire lesion while excluding adjacent normal tissues or artifacts. Segmentation masks were exported in NRRD labelmap format and reviewed for accuracy before radiomics feature extraction. The radiomics feature domains used in this study are illustrated in Figure 1, which

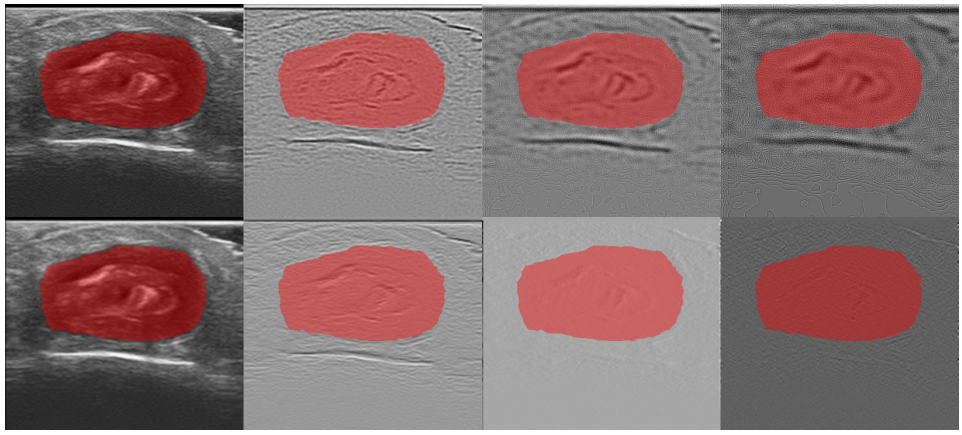

**Figure 1.** Visualization of multi-domain radiomics filters applied to a grayscale ultrasound image of a cutaneous lesion. The first row demonstrates the original image and Laplacian of Gaussian (LoG) filters with σ = 2, 4, and 6. The second row shows single-level Haar wavelet decompositions (LL, LH, HL, HH). The manually-drawn segmentation mask is overlaid on all images to illustrate the correspondence of texture-derived features within the lesion boundaries.

displays the original ultrasound image, LoG filters (σ=2, 4, 6), and Haar wavelet sub-bands (LL, LH, HL, HH) with the lesion segmentation overlaid.

### Radiomics Feature Extraction

Radiomics features were extracted using the PyRadiomics framework implemented in Python. Before feature extraction, images were preprocessed using the default PyRadiomics pipeline: Z-score normalization within the ROI, gray-level discretization using a bin width of 10, B-spline interpolation, and binary mask thresholding to maintain segmentation integrity.

To investigate the impact of different texture domains, radiomics features were calculated from original grayscale images, Laplacian-of-Gaussian (LoG)-filtered images, and wavelet-transformed images (eight directional decompositions).

A total of 851 features per lesion were initially extracted, including first-order statistics, shape features, gray-level co-occurrence matrix (GLCM), gray-level run-length matrix (GLRLM), gray-level size-zone matrix (GLSZM), gray-level dependence matrix (GLDM), and neighboring gray-tone difference matrix (NGTDM) features. Automatic diagnostic metadata generated by PyRadiomics were excluded.

### Feature Processing and Machine Learning Workflow

Following feature extraction, all radiomics variables underwent a structured preprocessing workflow to ensure robustness and reduce dimensionality. Initially, the dataset was examined for non-informative descriptors, and features with zero variance were removed. This step eliminated attributes that did not contribute measurable variability between lesions. Subsequently, redundancy among the remaining features was addressed by constructing a Pearson correlation matrix. When two features demonstrated a correlation coefficient exceeding 0.85, one of them was discarded to prevent multicollinearity and ensure a more stable learning environment. After this unsupervised reduction process, three distinct feature groups were retained for modeling: original grayscale features, wavelet-transformed features, and a combined feature set incorporating both domains.

### Statistical Analysis and Model Evaluation

Machine learning analyses were then conducted using Python-based pipelines. To avoid information leakage, all supervised feature selection procedures were embedded strictly within the training folds of the cross-validation framework. Mutual information (SelectKBest) was employed to identify the most informative features, with a maximum of 30 variables selected in each fold, depending on feature availability after unsupervised filtering. This ensured that each classifier was trained on an optimized subset of radiomics descriptors without exposing the test folds to any stage of the feature selection process.

Model construction was performed using three commonly applied algorithms in radiomics research: logistic regression with L2 regularization, support vector machine with a radial

basis function kernel, and random forest classifier with 500 trees. Because of the class imbalance between benign and malignant lesions, all models were trained using balanced class weights to prevent bias toward the majority class. Model performance was evaluated using five-fold stratified cross-validation, allowing malignant lesions to be proportionally represented across folds. Probabilistic outputs from all folds were aggregated to compute pooled performance metrics, including accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve. By assessing each combination of feature group and classifier, the workflow enabled a comprehensive comparison to identify the most discriminative radiomics-based approach for distinguishing benign from malignant skin lesions. A visual summary of the overall workflow, including dataset curation, segmentation, feature extraction, and model evaluation, is presented in Figure 2.

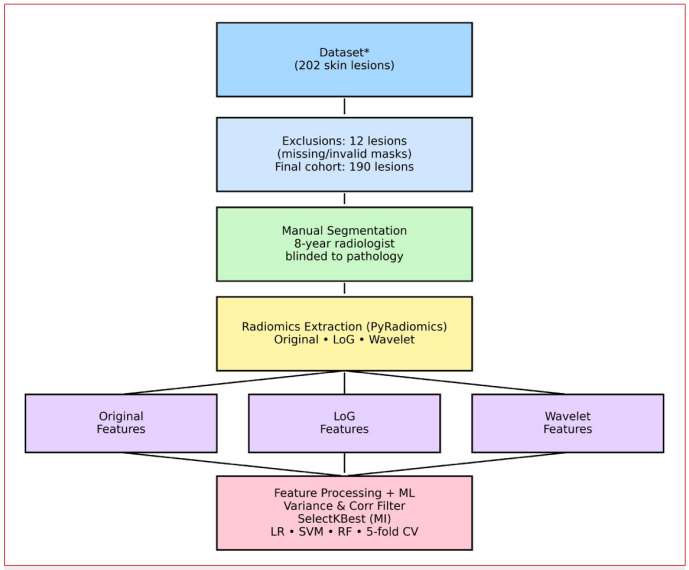


**Figure 2.** Workflow of the radiomics pipeline, including dataset selection, segmentation, radiomics feature extraction (original, LoG, and wavelet), and machine learning model development.

## RESULTS

### Dataset Characteristics

A total of 202 dermatologic ultrasound lesions were initially reviewed. Twelve lesions were excluded due to missing or invalid segmentation masks, yielding a final cohort of 190 analyzable lesions, including 129 benign and 61 malignant cases. All remaining lesions successfully underwent radiomics processing.

### Radiomics Feature Extraction and Reduction

Radiomics extraction from the original, LoG, and wavelet-transformed images produced 851 features per lesion. The dimensionality of this feature set was subsequently reduced through a multistep process. Low-variance filtering removed two near-constant features, followed by Pearson correlation filtering, which eliminated highly collinear variables and reduced the dataset to 143 non-redundant features. Further refinement was performed inside the training folds of a stratified five-fold cross-validation framework using mutual information–based SelectKBest, resulting in the selection of the top 30 discriminative features for each classifier during model training.

### Machine Learning Model Performance

Three classifiers—logistic regression, a support vector machine with an RBF kernel, and random forest—were trained on the original, LoG, wavelet, and combined feature sets using stratified five-fold cross-validation. Model performance varied across feature domains and algorithms. The highest overall performance was observed with the random forest model trained on the combined feature set, which achieved an AUC of 0.706 and an accuracy of 0.716. This model demonstrated high specificity (0.891) for benign lesions but more modest sensitivity (0.344) for malignant lesions.

Wavelet-derived features also showed strong discriminative value. The random forest model trained solely on wavelet features reached an AUC of 0.705 and the highest accuracy among all models at 0.742. Original features yielded intermediate performance, with logistic regression achieving an AUC of approximately 0.697 and offering relatively better malignant sensitivity compared with the other models. SVM-based models consistently showed high specificity but lower sensitivity, indicating stronger performance in identifying benign lesions. Table 1 summarizes model performance across classifiers and feature sets, while Figure 3 presents the corresponding ROC curves for each feature domain and classifier.

### Overall Classification Trends

Across all model–feature combinations, AUC values ranged from 0.64 to 0.71, and accuracy ranged from 0.63 to 0.74. A consistent pattern was observed in which specificity exceeded sensitivity across classifiers, reflecting better performance in correctly identifying benign lesions than malignant ones. Despite this imbalance, the stable results across feature groups and algorithms support the conclusion that quantitative radiomics features derived from dermatologic ultrasound images contain diagnostically relevant information for benign/malignant differentiation (Table 1).

## DISCUSSION

In this study, we demonstrated the feasibility of using radiomics analysis of grayscale B-mode ultrasound images to

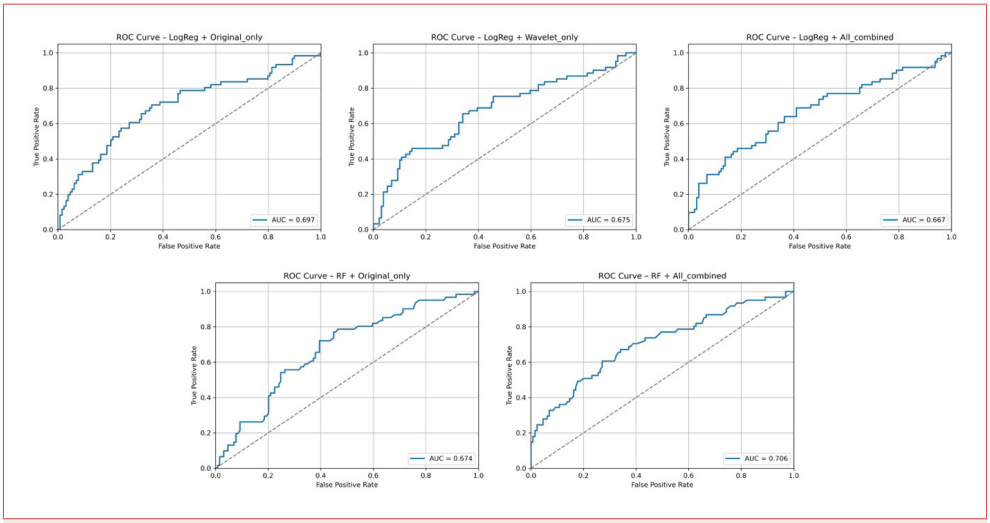


**Figure 2.** Receiver operating characteristic (ROC) curves for logistic regression and random forest models across different radiomics feature domains. The top row displays ROC performance for logistic regression using original features, wavelet features, and all combined features. The bottom row shows corresponding ROC curves for random forest using original and all combined feature sets. The area under the curve (AUC) is reported for each model.

classify skin lesions as benign or malignant. Using a dataset of approximately 190 ultrasound images, we extracted a high-dimensional set of quantitative imaging features and built machine learning models to distinguish malignant lesions. Our accuracy is comparable to the ~77% test accuracy reported for a deep learning classifier on a similar image set and, notably, is in the same range as expert dermatologist performance.[12] Their deep learning model achieved 77.1% accuracy in differentiating benign from malignant lesions, which was not significantly different from the performance of an experienced dermatologist (74.1%) on the same cases. Thus, our radiomics approach performs on par with advanced deep learning and expert interpretation, underscoring that ultrasound radiomics can capture diagnostically relevant information in skin lesions. This comparison is encouraging because it indicates that the diagnostic signal present in skin ultrasound images can be extracted through different methodological approaches. In fact, our radiomics model slightly exceeded the dermatologist-level accuracy reported by Laverde-Saad et al.,[10] albeit within a similar range. Given that the radiomics and CNN models were tested on comparable images, our study provides evidence that machine learning based on engineered radiomics features is a feasible alternative to deep learning for this task.

It is noteworthy that the random forest classifier performed best when using wavelet-transformed features alone. Wavelet radiomics features likely captured multiscale textural patterns and vascular detail in the ultrasound images, which may be particularly indicative of malignancy. This finding aligns with radiomics studies in other domains showing that multifrequency texture features can enhance lesion discrimination.[13]

**Table 1.** Performance of Radiomics-Based Machine Learning Models for Classifying Benign and Malignant Skin Lesions on Ultrasound (SVM: Support Vector Machine RBF:Radial Basis Function)

| Radiomics Feature Set | Classifier | AUC | Accuracy | Sensitivity | Specificity |
|---|---|---|---|---|---|
| All features | Random Forest | 0.706 | 71.6% | 34.4% | 89.1% |
| Wavelet features | Random Forest | 0.705 | 74.2% | 40.9% | 89.9% |
| Original features | Logistic Regression | 0.697 | 66.8% | 62.3% | 68.9% |
| All features | SVM (RBF Kernel) | 0.687 | 73.2% | 29.5% | 93.8% |
| Original features | Random Forest | 0.674 | 66.3% | 26.2% | 85.3% |

Our radiomics-based ultrasound classification of skin lesions demonstrates competitive diagnostic performance in line with recent literature. For example, a 2025 study by Qin et al.[8] extracted grayscale ultrasound radiomics features and used LASSO-selected inputs for machine learning, achieving a high AUC of ~0.96 in distinguishing benign from malignant skin tumors. The higher AUC in studies such as that by Qin et al.[8] may reflect differences in dataset homogeneity, lesion types, or optimization strategies, whereas our real-world dataset presents a broader and more challenging diagnostic spectrum. This is comparable to state-of-the-art deep learning models: a 2025 CNN approach combining B-mode and Doppler high-frequency ultrasound images reached approximately 95% accuracy (AUC ~0.98) in binary lesion classification.[14] In a broader prospective study, a deep multimodal network that fused clinical photographs with ultrasound slightly trailed those figures (binary AUC ~0.88), but its performance nearly matched that of expert dermatologists.[15] It should be noted that these performance differences likely reflect variations in model complexity, dataset size, and study design; therefore, direct comparison of AUC values across radiomics- and deep learning–based studies should be interpreted with caution. Accordingly, the contribution of this study lies not in outperforming deep learning models but in providing a systematic and transparent comparison of radiomics feature domains and machine learning classifiers for ultrasound-based lesion characterization.

Overall, both radiomics-based models and modern CNNs have achieved strong accuracy and AUC in dermatologic ultrasound, often markedly outperforming diagnoses based only on clinical features.[8] Key differences lie in feature extraction and modeling. Our radiomics pipeline relies on manual ROI delineation and predefined features, which are statistically filtered and fed into classifiers.[8] This approach can be advantageous for small datasets and offers some interpretability, but it may miss subtle imaging patterns. In contrast, deep learning methods automatically learn abstract features from raw ultrasound images; recent studies have trained CNNs on single-modality ultrasound or dual inputs to capture more information.[14] Others have integrated ultrasound with clinical photographs in a fusion model to improve the classification of skin diseases.[15] Hybrid strategies are also emerging; one 2024 study combined clinical data, radiomics features, and CNN-extracted features using an SVM, yielding an AUC of ~0.95 in external test sets.[16] In comparison, our radiomics and ML method remains robust and relatively data-efficient, but it may be outperformed by deep learning in multiclass or highly varied lesion scenarios. As a trade-off, radiomics models have proven their value by significantly improving diagnostic accuracy over clinical assessment alone, whereas deep learning approaches offer higher automation and the ability to leverage richer image information at the cost of greater complexity.[8] This approach allows direct inspection of feature categories and their relative contributions, offering greater transparency and interpretability compared with deep learning models, which often function as "black boxes." In addition, radiomics-based methods can be more data-efficient and robust in smaller datasets, where training complex deep learning architectures may be impractical or prone to overfitting. Each approach has complementary strengths, and the handcrafted feature model used in this study can be viewed as a solid, interpretable baseline against which newer deep learning or fusion techniques achieve incremental gains in diagnostic performance. Accordingly, the proposed models should be viewed as decision-support tools rather than standalone screening systems, and their outputs must be interpreted in conjunction with clinical and imaging findings.

Several limitations likely contributed to the moderate performance and constrain the generalizability of our findings. First, the dataset was relatively small, which limits lesion diversity and may predispose the models to overfitting or dataset-specific bias. This restricts the generalizability of the results, particularly across different institutions, ultrasound devices, and acquisition protocols. No external validation was performed; therefore, it remains unknown how well the radiomics model would perform on independent data from other institutions or ultrasound devices. This lack of external testing is a critical weakness, as models that perform adequately on internal data can falter when faced with variations in imaging settings or patient populations.

Additionally, our workflow relied on manual lesion segmentation performed by a single radiologist with 8 years of experience, an inherently time-consuming process that can introduce interoperator variability. Interobserver variability was not formally assessed, which represents a limitation of this study. In fact, the requirement for precise ROI segmentation is a known bottleneck in radiomics; it is cumbersome and highly dependent on the operator's experience.[17] This dependency not only affects the reproducibility of the features but also contrasts with deep learning models, which can learn directly from image data without a separate segmentation step.

Taken together, the limited sample size, single-center data, lack of external validation, and reliance on manual segmentation indicate that our reported AUC of ~0.71 should be interpreted with caution. The dataset also showed class imbalance (129 benign vs. 61 malignant lesions), which likely contributed to the relatively low sensitivity observed across models. SMOTE oversampling was evaluated within the training folds and improved sensitivity but resulted in reduced specificity and a slight decrease in AUC; therefore, class-weighted non-SMOTE

models were retained as the primary analysis, with SMOTE results reported in the supplementary material (Suppl 1). The model's generalizability is likely moderate at best, and further work is needed before it can be considered for clinical use.

Despite these limitations, our study highlights that radiomics analysis of high-frequency skin ultrasound is a viable approach for lesion characterization. The performance, although not superior to deep learning benchmarks, confirms that handcrafted ultrasound features contain relevant diagnostic information. With improvements, such as larger multicenter datasets, automated or standardized segmentation, and prospective validation, the accuracy of radiomics-based models may be enhanced. Future investigations might also explore hybrid approaches, such as combining radiomics features with deep learning or clinical data, to improve performance.

## CONCLUSION

Radiomics proved moderately effective in classifying dermatologic ultrasound lesions as benign or malignant, providing a foundation for quantitative image analysis in this emerging field. Achieving the higher accuracy levels reported with deep learning (AUC ~0.77) will require further research and validation, given the current model's limitations in data scope and generalizability. Our findings encourage ongoing efforts to refine radiomics and integrate it with advanced AI techniques so that non-invasive ultrasound can be used more reliably as a diagnostic tool in dermatologic oncology.

### DECLARATIONS

**Ethics Committee Approval:** The images were obtained from a publicly available dataset in which all identifying information has been completely anonymized; therefore, no institutional ethics committee approval was required for this study.

**Informed Consent:** The images were obtained from a publicly available dataset in which all identifying information has been completely anonymized; therefore, no informed consent was required for this study.

**Conflict of Interest:** The author declare that there is no conflict of interest.

**Funding:** The author received no financial support for the research and/or authorship of this article.

**Use of AI for Writing Assistance:** Not declared.

**Authorship Contributions:** Concept – HA; Design – HA; Supervision – HA; Fundings – HA; Materials – HA; Data collection &/or processing – HA; Analysis and/or interpretation – HA; Literature search – HA; Writing – HA; Critical review – HA.

**Peer-review:** Externally peer-reviewed.

## REFERENCES

1. Rigel DS, Friedman RJ, Kopf AW. The incidence of malignant melanoma in the United States: issues as we approach the 21st century. J Am Acad Dermatol 1996;34:839–47.
2. Raza S, Ali F, Al-Niaimi F. Ultrasonography in diagnostic dermatology: a primer for clinicians. Arch Dermatol Res 2023;315:1–6.
3. Boostani M, Wortsman X, Pellacani G, Cantisani C, Suppa M, Mohos A, et al. Dermoscopy-guided high-frequency ultrasound for preoperative assessment of basal cell carcinoma lateral margins: a pilot study. Br J Dermatol 2025;193:572–4.
4. Rogers W, Thulasi Seetha S, Refaee TAG, Lieverse RIY, Granzier RWY, Ibrahim A, et al. Radiomics: from qualitative to quantitative imaging. Br J Radiol 2020;93:20190948.
5. Falcone R, Verkhovskaia S, Di Pietro FR, Scianni C, Poti G, Morelli MF, et al. Application of radiomics in melanoma: a systematic review and meta-analysis. Cancers (Basel) 2025;17:3130.
6. Thomson H, Yang S, Cochran S. Machine learning-enabled quantitative ultrasound techniques for tissue differentiation. J Med Ultrason (2001) 2022;49:517–28.
7. Zhang YP, Zhang XY, Cheng YT, Li B, Teng XZ, Zhang J, et al. Artificial intelligence-driven radiomics study in cancer: the role of feature engineering and modeling. Mil Med Res 2023;10:22.
8. Qin Y, Zhang Z, Qu X, Liu W, Yan Y, Huang Y. A machine learning model based on high-frequency ultrasound for differentiating benign and malignant skin tumors. Med Ultrason 2025;27:284–93.
9. Faita F, Oranges T, Di Lascio N, Ciompi F, Vitali S, Aringhieri G, et al. Ultra-high-frequency ultrasound and machine learning approaches for the differential diagnosis of melanocytic lesions. Exp Dermatol 2022;31:94–8.
10. Laverde-Saad A, Jfri A, García R, Salgüero I, Martínez C, Cembrero H, et al. Discriminative deep learning based benignity/malignancy diagnosis of dermatologic ultrasound skin lesions with pretrained artificial intelligence architecture. Skin Res Technol 2022;28:35–9.
11. Dermatologic Ultrasound Images for Classification Dataset. Available from: https://www.kaggle.com/datasets/alfageme/dermatologic-ultrasound-images. Accessed Nov 24, 2025.
12. Borhani-Haghighi A, Emami M, Vasaksi AS, Shariat A, Banihashemi MA, Nikseresht A, et al. Large-vessel stenosis in the patients with ischemic stroke in Iran: Prevalence, pattern, and risk factors. J Vasc Interv Neurol 2015;8:11–6.
13. Lee J, Lee J, Song BI. A machine learning-based radiomics model for the differential diagnosis of benign and

malignant thyroid nodules in F-18 FDG PET/CT: external validation in the different scanner. Cancers (Basel) 2025;17:331.

14. da Silva IRV, Jardim AG, de Souza Faés GR, Alva TAP, Becker CDL, Botelho VR. Deep learning-based skin lesion classification: a CNN approach on high-frequency ultrasound imaging. J Ultrasound Med 2026;45:925–36.
15. Zhu AQ, Wang Q, Shi YL, Ren WW, Cao X, Ren TT, et al. A deep learning fusion network trained with clinical and high-frequency ultrasound images in the multi-classification of skin diseases in comparison with dermatologists: a prospective and multicenter study. EClinicalMedicine 2024;67:102391.
16. Dai X, Lu H, Wang X, Zhao B, Liu Z, Sun T, et al. Development of ultrasound-based clinical, radiomics and deep learning fusion models for the diagnosis of benign and malignant soft tissue tumors. Front Oncol 2024;14:1443029.
17. Zhang X, Zhang Y, Zhang G, Qiu X, Tan W, Yin X, Liao L. Deep learning with radiomics for disease diagnosis and treatment: challenges and potential. Front Oncol 2022;12:773840.

**Appendix 1.** Effect of SMOTE on Random Forest model performance across key radiomics feature sets

| Feature Set | Variant | AUC | Accuracy (%) | Sensitivity (%) | Specificity (%) |
|---|---|---|---|---|---|
| All combined | No SMOTE | 0.706 | 71.6 | 34.4 | 89.1 |
| All combined | SMOTE | 0.688 | 71.1 | 47.5 | 82.2 |
| Wavelet only | No SMOTE | 0.705 | 74.2 | 40.9 | 89.9 |
| Wavelet only | SMOTE | 0.709 | 71.6 | 50.8 | 81.4 |