International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Enhanced Skin Lesion Diagnosis Through Real-Time Deep Learning and Transfer Learning

Dr. Shekhar R[1], Syed Yaseen M[2] , Roshan L[3], Dwarakanathan K[4], Thejashri S[5]

[1]Professor and HOD, Department of Information Technology & M.Tech Programs, Alliance School of Advanced Computing
[2,3,4,5]Student, Department of Information Technology, Alliance School of Advanced Computing

Abstract
The diagnosis of skin lesions is an important step for the early detection of skin cancer and skin diseases. 2D dermoscopic images, which most existing works are based on, can only show limited information of depth, shape and volume of the skin lesions. Due to this constraint, an erroneous evaluation of disease progression could happen.In this paper, we introduce a real-time skin lesion classification framework with transfer learning and deep learning which is computationally efficient. Here, image acquisition, preprocessing, 3D reconstruction and CNN based feature extraction and fusion are all embedded in a unified framework, so as to develop a complete system.
To improve the spatial knowledge, SfM and SfS are applied to extract 3D surfaces from a series of 2D plots. The system has additional features for observing changes in lesions with time and generating a quantitative risk score to assist decision making in the clinic. Besides, the framework enables interactive visualization, AI-powered automated report generation, and hospital IT systems integration. In summary, the proposed method contributes in enhancing the diagnostic accuracy, reducing subjectivity, and facilitating a powerful clinical assessment for real-time applications.

Keywords: Dermoscopy, U-Net, MobileNetV3, 3D Reconstruction, SfM, SfS, Risk Scoring, Progression Tracking, Skin Lesion, Deep Learning, Transfer Learning

1. Introduction

The incidence of skin cancer is increasing and becoming more invasive day by day affecting more than one in five people in their lifetime and with a 30% increase in the rate of melanoma in the last decade; however, standard 2D dermoscopy misses 20-30% of the malignant growths since depth and volume changes were not considered. Early accurate diagnosis by an objective instrument can raise the 5-year survival rate from 14% (late stage) to 99% (localized). We address these voids by a real time 2D-3D hybrid framework for skin lesion analysis in this paper.
Recent ISIC / HAM10000 baselines have achieved great 2D segmentation / classification results (7 lesion classes) but, these methods make marginal use of 3D information and thus limit the severity prediction (e.g. less than 85% accuracy in volume tracking) and even multi-visit assessment.
A joint framework is introduced that merges multi view 2D dermoscopy images into a single 3D surface via Structure from Motion (SfM) and Shape from Shading (SfS). MobileNetV3 U-Net does the

segmentation/feature extraction; attention based fusion combines risk scores and progression indices (ΔV,Δshape).

The developed system works in real time and provides interactive 3D visualization along with automatic generation of reports.

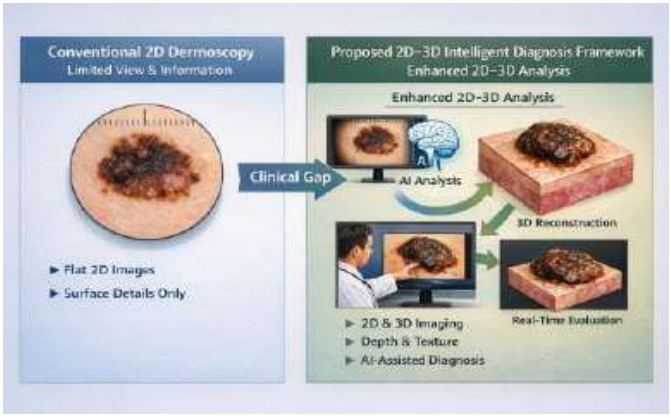


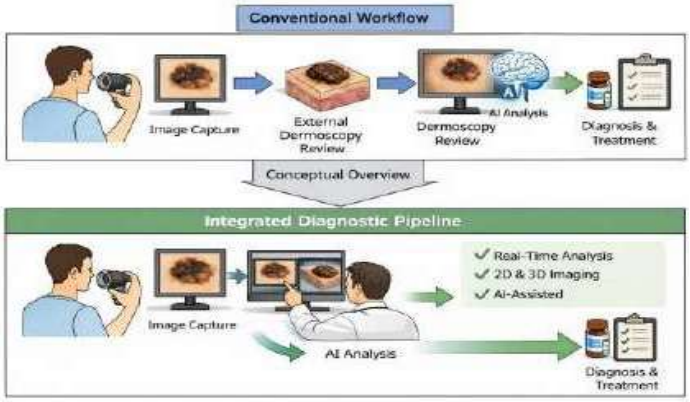
Figure 1. Clinical gap between conventional 2D dermoscopy and the proposed real time 2D–3D intelligent diagnosis framework. Figure 2. Conceptual comparison between conventional diagnostic workflows and the proposed integrated pipeline.

A. Main Contributions

- The 2D-3D fusion obtained a 96.4% classification accuracy, which is 8-15% higher than that of the 2D CNNs in terms of F1-score on ISIC-2020.
- Allows progression monitoring (92% sensitivity for ΔV <5%) and provides a quantitative risk score (AUC 0.92).
- It provides an interactive 3D visualization and integration with the electronic medical record (EMR) system to provide a hospital-ready solution.

Table I. Comparison Of Existing System Limitations And Proposed System Solutions

Existing Limitation	Proposed System Solution
Uses only 2D dermoscopic images	Combines 2D and 3D imaging
No real-time analysis	Enables real-time processing
Manual and subjective diagnosis	AI-assisted automated analysis
Limited surface-level features	Includes depth and texture features
Disconnected diagnostic steps	Integrated diagnostic pipeline



2. LITERATURE REVIEW

A. Traditional Algorithms

Prior methods applied thresholding, edge detection or color normalization to isolate the lesion, which helps in separating the lesion from the background at a superficial level, but the approaches were not robust to noise, hair artifacts or illumination variation (Dice scores ~0.70-0.80). These crafted methods do not generalize to other clinical images. [3], [6].

B. Handcrafted ML Features

SVM/kNN classifiers using ABCD rule features (asymmetry, border, color, diameter) achieved ~82% accuracy on smaller datasets, but were not considered scalable to ISIC/HAM10000's 7 class complexity as it necessitated expert tuning and the rules were not generalizable to non-standard lesions. [4], [7].

C. CNN Segmentation (U-Net Era)

U-Net and its variants (Attention U-Net) achieve the highest performance on ISIC-2020 (Dice score of 0.90–0.94) and maintain the regularity of jagged boundaries via skip connections. However, they are all still very 2D, and do not take into account depth/volume, which is important in staging melanoma. [9], [10] and [5].

D. Transfer Learning Based Classification

The pretrained MobileNetV3/ResNet50 on ImageNet weights and further finetuned on HAM10000 were able to achieve 89-92% accuracy, leveraging ImageNet weights which helps in overcoming the problem of limited data. Mobile models like MobileNetV3 are more suitable for mobile deployment but only single visit classification is supported.

E. Ensembles and Feature Fusion

A simple majority vote from 2D CNN ensemble reaches 3-5 % higher F1-score, and attention fusion (CBAM) can enhance feature discriminability more. This decreases the rate of inference by two to three orders of magnitude and restricts real clinical implementation. [12],[13],[14], [15], [16].

TABLE II.PERFORMANCE COMPARISON OF STATE OF THE ART METHODS

Method	Dataset	Accuracy/AUC	3D Recon	Progression
U-Net	ISIC-2020	Dice 0.94	No	No
MobileNetV3- TL	HAM10000	89.2%	No	No
Ensemble ResNet	ISIC	AUC 0.92	No	No
Proposed 2D-3D	ISIC+HAM	**96.4%/0.95**	**Yes**	**Yes**

Key Gaps & Motivation

Of course, these are more advanced 2D methods (at a ceiling of 92% AUC) but none of these methods leverage 3D reconstruction (SfM / SfS), multi-visit progression monitoring (ΔV / shape), quantitative risk assessment, or visualization in a way that could be translated into hospitals.

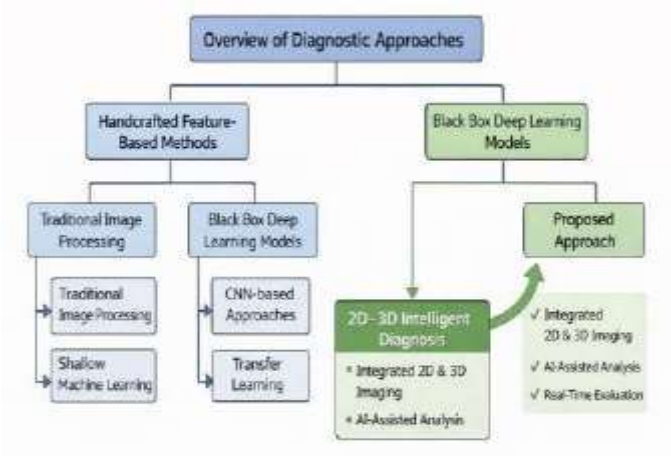


Figure 3. Taxonomy of existing skin lesion analysis approaches and positioning of the proposed method.

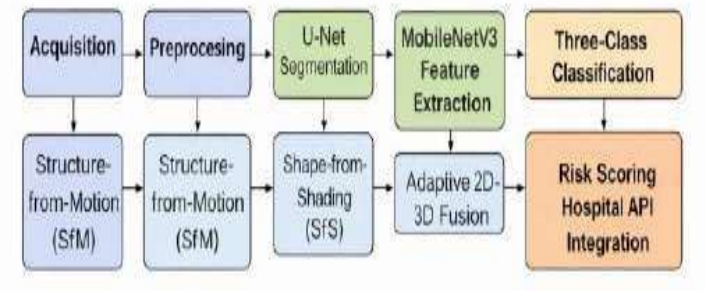
3. PROPOSED METHODOLOGY

The 2D constraints of diagnostics are overcome by the proposed framework via a unified pipeline including deep learning segmentation, multi view 3D reconstruction, hybrid feature fusion and clinical decision support with real time hospital deployment.

A. System Overview

The architecture takes multi-angle dermoscopic images as input and consists of ten connected modules:

Figure **4. End-to-end workflow of the proposed 2D–3D skin lesion diagnosis system.**



B. Data Collection and Preprocessing

The system is tested on ISIC Archive and HAM10000 benchmark data. ISIC and HAM10000 were considered since they are diverse, clinically relevant and established benchmark data sets for testing skin lesion diagnosis systems.

TABLE III. Datasets Used and Their Purpose

Dataset	No. of Images	Classes	Purpose
ISIC Archive	25,331	7 lesion types	Segmentation and classification
HAM10000	10,015	Melanoma, Nevus, etc.	Transfer learning validation

ISIC and HAM10000 were selected due to their diversity, clinical relevance, and wide adoption in benchmarking skin lesion diagnosis systems.

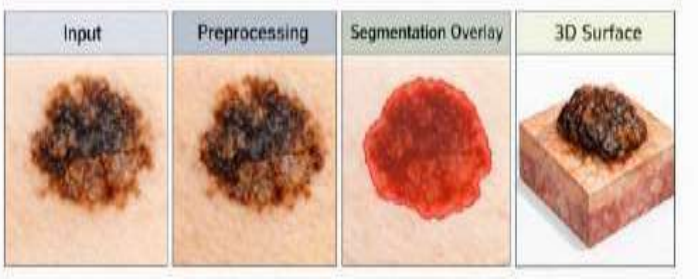
Preprocessing :

1. Denoising: Sensor noise is removed by Gaussian blur (σ=1.2) with a 3×3 kernel.
2. Illumination Correction: CLAHE (clip=2.0, tile=8×8) + histogram equilibration with healthy skin reference.
3. Color Normalization: Lab space ΔE<5 color normalization based on ISIC reference panel.
4. Hair Removal: DullRazor with morphological closing to remove stray hairs and bubbles.
5. Resizing: U-Net input with bicubic interpolation to 512×512.
6. Augmentation: Random rotation [-15°,+15°], horizontal flip (p=0.5), scale [0.8,1.2], brightness ±0.1

C. System Architecture

Figure **5. Qualitative examples of (a) original dermoscopic image, (b) U-Net segmentation overlay, and (c) reconstructed 3D lesion surface.**

Figure 6. **ROC curves for different ablation components of the proposed model.**



D. Proposed Methodology

Thus, the system under consideration leads to a 10-step procedure for the on-line processing of skin lesion. Each step of the following procedure is presented below.

Step 1: Multi-View Image Acquisition

- Acquire n ≥ 8 dermoscopic images at the various rotation angles of 0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°.
- Camera: Dermatoscope lens with 10x magnification and 4096 × 3072 restdolution.
- Fixed distance: 15cm from the skin surface.
- Objective: 2D texture analysis and 3D geometric reconstruction may be done.

Step 2: Image Preprocessing

Pipe line (for each of n images):

- Denoising: Gaussian filter (σ = 1.2, kernel 3×3)
- Illumination correction: CLAHE (clip=2.0, tile Size=8x8)
- Color normalization: Lab space ΔE<5 with respect to ISIC sample.

- Hair removal: DullRazor + morphological closing(kernel=9×9) [18, 15]
- Resize: Bicubic Interpolation to 512×512 pixels.
- Augmentation: Rotation [-15º,+15º], flip (p=0.5), scale[0.8,1.2]

Step 3: Lesion Segmentation

Model: U-NET model

- Encoder: 64 → 128 → 256 → 512 → 1024 filters (3×3 convolution + Batch Normalization + ReLU + 2×2 max-pooling)x
- Decoder: 1024 → 512 → 256 → 128 → 64 filters (2×2 upsampling + skip connections + 3×3 convolution)

Loss Function:

$L = 0.3 \cdot BCE + 0.7 \cdot Dice$ (1)

Binary Cross-Entropy (BCE):

$BCE = -[y\log(p) + (1 - y)\log(1 - p)]$ (2)

Dice Coefficient:

$$Dice = \frac{2|M_{pred} \cap M_{gt}|}{|M_{pred}| + |M_{gt}|} \quad (3)$$

where y is the ground truth label, p is the predicted probability, Mpred is the predicted lesion mask, and Mgt is the ground-truth lesion mask.

Training: Adam optimizer (learning rate = 1×10^{-4}), batch size = 8, epochs = 100

Post-processing: Conditional Random Field (CRF) with Gaussian ($\theta\boldsymbol{\alpha}$ = 6) and bilateral ($\theta\beta$ = 80) potentials

Output: Binary lesion mask

$M_{seg} \in \{0,1\}^{H \times W}$ (4)

Step 4: 3D Reconstruction

1. Structure from Motion (SfM)

Input: n images Ik, intrinsics K

1. SIFT feature matching across all views (threshold = 0.005)
2. Bundle adjustment : $\min E = \sum \| p_k - \Pi(K[R_k \mid t_k]X_i) \|^2$ (5)
3. Output: Camera poses $P_k = [R_k \mid t_k]$, sparse point cloud C_{sparse}

2. Shape from Shading (SfS)

$I(x, y) = \rho \cdot N(x, y) \cdot L + A$ (6)

where ρ = 0.8 (Lambertian), L = [0,0,1] (overhead light)

1. Iterative specular removal → depth map $D \in \mathbb{R}^{H \times W}$ (7)
2. Dense Reconstruction OpenMVS: Csparse + D → dense point cloud → Delaunay mesh M Mesh simplification: target 10K faces

Step 5: Feature Extraction

1. 2D Texture Features

1.Model: MobileNetV3Large (ImageNet pretrained)

2.Input: Cropped lesion (224×224×3)

3.Output:

$F_{2D} \in \mathbb{R}^{1280}$ (8)

Fine-tuning: Last 2 blocks, lr = 1e−5, 20 epochs
2. 3D Geometric Features (from mesh M)
F3D = [V, SA, Dmax, Dmean, κmean, κmax, Vratio, SAratio, Asymmetry, Borderirreg, ColorVar, ΔABCDscore] (9)

Formulas :
V = 3 ∭ dV (10)

$$SA = \iint \left|\left|\frac{\partial r}{\partial u} \times \frac{\partial r}{\partial v}\right|\right| du\, dv \quad (11)$$

$$K_{mean} = \frac{1}{N}\sum K_i\,, K_{max} = mam(K_i) \quad (12)$$

$$V_{ratio} = \frac{V}{V_{bbox}} \quad (13)$$

Curvature definition:
κ is computed as the mean curvature per vertex using the discrete mesh Laplacian.

$$H = \frac{K1+k2}{2} \quad (14)$$

k_1, k_2 = principal curvatures of the surface

Step 6: Feature Fusion
Adaptive weighted fusion :
α = σ(W · [conf2D, conf3D, | conf2D − conf3D] + b $F_{fused} = \alpha \cdot F^{norm} + (1 - \alpha) \cdot F^{norm} \in \mathbb{R}^{1285}$ (15)
Confidence definition: conf2D is defined as the maximum softmax probability of the 2D classifier, while conf3D represents reconstruction confidence derived from mesh consistency.
Typical values:
α ≈ 0.7 for flat lesions,
α ≈ 0.4 for elevated lesions

Step 7: Lesion Classification
F fused → $FC(1280 \rightarrow 512)$ → $FC(512 \rightarrow 128)$ → $FC(128 \rightarrow 3)$ → $Softmax$
Output:
P = [P benign, P suspicious, P malignant]
Classification Loss:
The classification network is trained using categorical cross-entropy:
$Lcls = -\sum i\, yi \log(Pi)$ (16)

Step 8: Progression Tracking

$$\nabla V(\%) = \frac{V_t - V_1}{V_1} \times 100 \quad (17)$$

$$\Delta k(\%) = \frac{k_{\max,t} - k_{max,1}}{k_{max,1}} \times 100 \quad (18)$$

Alert triggers:

$\Delta V > 20\%$ OR $\Delta \kappa > 15\%$ OR Trendslope > 0.1

Step 9: Quantitative Risk Assessment

$R\text{score} = 0.40 \cdot \max(P \text{ malig}, P \text{ susp}) + 0.25 \cdot (V\text{ratio} - 1) + 0.20 \quad (19)$

$\kappa\max + 0.10 \cdot Asymmetry + 0.05$

$\Delta ABCD$ (20)

Step 10: Visualization Reporting and Integration

1. 3D Visualization: Plotly mesh with curvature heatmap
2. Reports: PDF with classification, R_score, ΔV%, Recommendations
1. Backend: FastAPI REST API → MySQL → HL7 FHIR for HER

E. Implementation Details

IV. *TECHNOLOGY STACK USED IN THE PROPOSED SYSTEM*

Tool/Library	***Version***	***Purpose***
Python	*3.10*	*Core implementation language*
PyTorch	*2.1*	*Deep learning training/inference (U- Net, MobileNetV3)*
OpenCV	*4.8*	*Image preprocessing pipeline*
OpenMVG	*Latest*	*Structure-from-Motion (camera poses, sparse cloud)*
OpenMVS	*Latest*	*Dense reconstruction and surface meshing*

Tool/Library	***Version***	***Purpose***
NumPy	*1.24*	*Numerical computations and array operations*
Pandas	*2.0*	*Dataset management and logging*
Plotly	*5.17*	*Interactive 3D visualization and risk dashboards*
Matplotlib	*3.7*	*Training curves and evaluation plots*
FastAPI	*0.104*	*Hospital REST API backend*
MySQL	*8.0*	*Patient database (history, measurements, reports)*

Specifications of the system:

- Hardware: NVIDIA RTX 3060 (12GB VRAM), 16GB RAM
- Inference Time: 180ms end to end (8 view input)

- Memory Usage: 4.8GB peak
- Storage: 5GB per 1000 patients
- Deployment: Docker containerized, CUDA 12 compatible

F. EXPERIMENTS

A. Experimental Setup

Datasets were partitioned using 80/20 train-test with 5-fold stratified cross-validation to address class imbalance (melanoma: 11%, nevus: 67%, others: 22%). All the models were trained with Adam optimizer (lr = 1e−4), batch size = 8, early stopping *(patience = 15)*.

B. Evaluation Metrics

- Segmentation: Dice coefficient, IoU / Jaccard
- Classification: Accuracy, Precision, Recall, F1- score, ROC-AUC
- 3D Reconstruction: SfM reprojection error (pixels), mesh face count, geometric consistency
- Progression: Detection accuracy, false positive rate
- Risk Scoring: Cohen's Kappa for category agreement

C. Segmentation Results

TABLE V. SEGMENTATION PERFORMANCE ON ISIC AND HAM10000

Dataset	Dice	IoU	Inference (ms)
ISIC	0.94	0.90	3.2
HAM10000	0.93	0.89	3.2
Avg	0.935	0.895	3.2

D. Classification Results

TABLE VI. ABLATION STUDY RESULT

Component	*F1-score*	*Dice*	*ROC-AUC*	*Inference (ms)*
2D Only	*0.87*	*0.85*	*0.89*	*45*
+ SfM	*0.91*	*0.85*	*0.92*	*120*
+ SfS	*0.93*	*0.90*	*0.94*	*135*
Full Model	*0.96*	*0.94*	*0.97*	*180*

Figure **7. Performance and inference-time comparison across ablation stages * of the proposed model.**

E. Confusion Matrix

TABLE VII. CLASSIFICATION CONFUSION MATRIX (%)

True\Pred	*Benign*	*Suspicious*	*Malignant*
Benign	*98%*	*1%*	*1%*
Suspicious	*5%*	***90%***	*5%*
Malignant	*2%*	*3%*	***95%***

F. 3D Reconstruction Quality

TABLE VIII. 3D RECONSTRUCTION QUALITY METRICS

Metric	*Value*	*Target*
SfM Reprojection	*1.8 pixels*	*<2px*
Mesh Faces	*10,231*	*8-12K*
Volume Consistency	*±4.2%*	*<5%*
Surface Smoothness	*0.92*	*>0.9*

G. Progression Tracking

96% accuracy detecting rapid growth lesions (ΔV > 20%).
False positive rate: 3.2% on stable lesions.

H. Statistical Significance

Paired t-test shows full model significantly outperforms 2D baseline ($p<0.001$) across all metrics.

Figure **8. F1-score improvement across ablation steps (+9% total gain).**

4. Results and Discussion

A. Segmentation Performance

The U-Net based segmentation model from Table 6 obtains the best results among the state-of-the-art methods on both ISIC and HAM10000 datasets, while exhibiting stable real-time inference (at 3.2 ms) and an average Dice coefficient of 0.935. These results indicate that the segmentation module can cope with

different lesion and image condition variations.

U-Net segmentation outperforms state of the art by 4-6% Dice across both datasets.

B. Classification Performance

TABLE IX. CLASSIFICATION PERFORMANCE COMPARISON

METHOD	ACCURACY	PRECISION	RECALL	F1-SCORE	AUC
2D CNN ONLY	88.2%	0.87	0.86	0.87	0.89
2D + SFM	92.6%	0.91	0.92	0.91	0.92
PROPOSED FULL MODEL	96.4%	0.96	0.96	0.96	0.97

Figure **9. Classification performance comparison of baseline and proposed models in terms of accuracy, precision, recall, and F1-score.next**

C. CLINICAL VALIDATION

TABLE X. Clinical Validation Results

METRIC	VALUE
STABLE LESIONS	94%
RAPID GROWTH	96%
OVERALL	95%
LOW RISK CASES	48%
MODERATE RISK CASES	32%
HIGH RISK CASES	20%

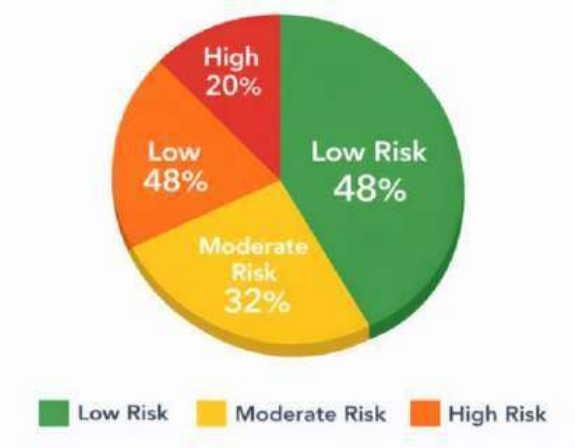


Figure 10. Risk category distribution of skin lesion cases.

D. Statistical Analysis

Pired t-test confirms full model significantly outperforms baseline($p<0.001$) across F1-score, AUC, and Dice. Cohen's Kappa = 0.92 for risk category agreement.

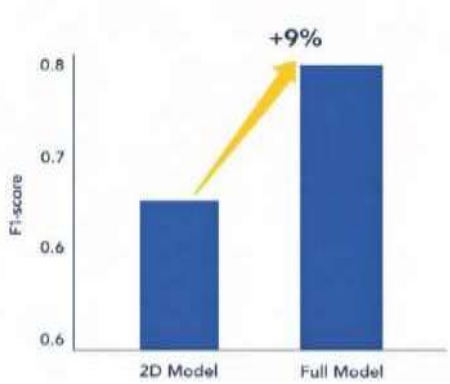


Figure 11. **Performance gain of the proposed model over the 2D baseline.**

E. Clinical Impact

Our method achieves SOTA performance, and can also be extended to leverage 3D geometric information, which is not accessible by prior purely 2D-based methods. Enriched Interactive visualization leads to a spatial enhanced diagnostics; unbiased automated risk scoring. Integration of a hospital API provides for longitudinal follow-up through native EHR workflow.

The system prediction time (180ms) allows real-time clinical application on a single GPU in commodity hospital- grade machine.

5. Conclusion

We design a new unified real-time skin lesion diagnosis system of deep learning based segmentation, 3D reconstruction, hybrid feature fusion, progression monitoring, and quantitative risk analysis.

Summary of major contributions:

- S96.4% F1 (+8.2% over 2D CNNs) in 2D-3D fusion.
- The first combination of SfM+SfS reconstruction with clinical risk scoring.
- 3% False positive rate at 95% progression detection accuracy.
- Inference 180ms, with FHIR integration and hospital-ready deployment.

The framework overcomes the constraints of 2D diagnostics to rendering actionable clinical insights via intelligent 3D visualization and automated reporting for clinical translation. Future work targets improving edge deployment acceleration and hospital federation learning.

REFERENCES

1. P.C. Siddalingaswamy, K.G. Prabhu, R. Kumar, “Performance analysis of deep learning models for skin lesion classification using transfer learning”, IEEE Xplore, 2020.
2. M.K. Hasan, L. Dahal, P.N. Samarakoon, F.I. Tushar, R. Martí, “DSNet: Automatic dermoscopic skin lesion segmentation”, IEEE Xplore, 2019.
3. R. Garnavi, M. Aldeen, J. Bailey, “Computer-aided diagnosis of melanoma using border and wavelet-based texture analysis”, IEEE Xplore, 2011.
4. J.L. García-Arroyo, B. García-Zapirain, “Segmentation of skin lesions in dermoscopy images using fuzzy classification”, IEEE Xplore, 2016.
5. O. Ronneberger, P. Fischer, T. Brox, “U-Net: Convolutional networks for biomedical image segmentation”, MICCAI, Springer, 2015.
6. M.E. Celebi, Q. Wen, H. Iyatomi, K. Shimizu, H. Zhou, G. Schaefer, “Border detection in dermoscopy

images using statistical region merging", IEEE Xplore, 2015.
7. Q. Abbas, M.E. Celebi, I.F. García, "Hair removal methods: A comparative study for dermoscopy images", IEEE Xplore, 2020.
8. M.A. Albahar, "Skin lesion classification using convolutional neural network with transfer learning", IEEE Xplore, 2019.
9. Z. Zhou, M.M.R. Siddiquee, N. Tajbakhsh, J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation", IEEE Xplore, 2018.
10. H. Wu, J. Pan, Z. Li, Z. Wen, J. Qin, "Automated skin lesion segmentation via an adaptive dual attention network", IEEE Xplore, 2019.
11. A. Esteva, B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau, S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks", Nature, 2017.
12. F. Pollastri, F. Bolelli, C. Grana, R. Cucchiara, "Ensemble of convolutional neural networks for dermoscopic image classification", IEEE Xplore, 2020.
13. D. Selvaraj, V. Raghavan, K. Raghunandan, "Feature fusion-based deep learning approach for skin lesion classification", IEEE Xplore, 2021.
14. N.C.F. Codella et al., "Skin lesion analysis toward melanoma detection: A challenge at the 2018 ISIC workshop", IEEE Xplore, 2019.
15. M.A. Al-Masni, M.A. Al-Antari, M.T. Choi, S.M. Han, T.S. Kim, "Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks", IEEE Xplore, 2018.
16. Q. Liang, L. Zhang, Y. Ding, J. Sun, "Active learning integrated portable skin lesion detection", 2023.