International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-15 Issue-3, February 2026

Fast Hierarchical and Flat Classification of Dermoscopic Skin Lesions Using Hand-Crafted Features and Classical Machine Learning

Mohamed Amin Belal, Sara. I. Ibrahim, Ben Bella S. Tawfik, Mohamed M. El-Gazzar

Abstract: ***Skin cancer screening using dermoscopic images remains challenging because malignant and benign lesions can share overlapping visual cues, including irregular borders, heterogeneous pigmentation, and complex textures. In addition, real clinical datasets often contain ambiguous samples, acquisition artifacts, and non-uniform illumination, which may degrade training quality and generalization. Although deep learning has achieved strong performance in dermoscopic lesion classification, its dependence on large annotated datasets, limited interpretability, and computational demands can limit adoption in resource-constrained clinical settings. This paper presents a complete classical machine learning framework for four clinically significant dermoscopic categories: basal cell carcinoma (BCC), melanoma, nevus, and pigmented benign keratosis (PBK). A unified hand-crafted feature representation is constructed by combining texture descriptors (Histogram of Oriented Gradients and Local Binary Patterns), region geometry and border regularity measures, Hu moment invariants, lesion intensity statistics, and multi-space colour descriptors (RGB/HSV statistics and HSV histograms). To reduce the effect of label noise and hard-to-separate samples, a margin-based smart cleaning strategy removes low-reliability instances on a per-class basis using a preliminary classifier's confidence margin. Class imbalance is then mitigated through controlled upsampling to a common class count without synthesizing new image content. Feature dimensionality is reduced using minimum-redundancy maximum-relevance (mRMR) ranking, retaining the top 300 features to balance accuracy and runtime. We evaluate both flat multi-class classification and a clinically motivated two-level hierarchical design that first separates Cancer vs non-Cancer, then performs subtype classification within each branch (BCC vs melanoma; nevus vs PBK). Experiments are implemented in MATLAB R2022 using stratified 5-fold cross-validation with strict fold-wise isolation to prevent information leakage. The best flat model, SVM with an RBF kernel, achieves a mean accuracy of 82.87%. The hierarchical system achieves an overall accuracy of 81.80%, with Level-1 accuracy of 86.54%. For Level-1 malignancy detection, ROC AUC = 0.9383 and PR AUC = 0.9367 (all folds combined), indicating strong discrimination. Oracle branch evaluations confirm high intra-branch separability (94.04% for BCC vs melanoma; 94.19% for nevus vs PBK) and show that residual loss is primarily due to Level-1 routing errors.***

Keywords: ***Dermoscopy; Skin Lesion Classification; Hierarchical Classification; Hand-Crafted Features; Support Vector Machine.***

Nomenclature:
HOG: Histogram of Oriented Gradients
LBP: Local Binary Patterns
mRMR: Minimum-Redundancy Maximum-Relevance
SVM: Support Vector Machine
RBF: Radial Basis Function
ROC: Receiver Operating Characteristic
PR: Precision-Recall
AUC: Area Under Curve
BCC: Basal Cell Carcinoma
PBK: Pigmented Benign Keratosis
ECOC: Error-Correcting Output Codes

I. INTRODUCTION

Dermoscopy is widely used for non-invasive inspection of subsurface skin structures and can support earlier identification of suspicious pigmented lesions. Nevertheless, dermoscopic diagnosis remains difficult because multiple lesion categories share visually similar patterns, and clinical datasets may exhibit uncertain boundaries, acquisition artefacts, and labelling variability. Recent studies emphasise robust evaluation and pipeline stability for practical deployment in real workflows; for example, Akram et al. [1] present an optimised fusion strategy for lesion classification and discuss performance under modern evaluation settings.

Public dermoscopy resources also enable standardized benchmarking and reproducible evaluation protocols. The widely used **HAM10000** dataset provides a large multi-source collection of pigmented lesion images that supports comparative studies and method validation, as described by **Tschandl et al. [2]**. More recent multisource datasets expand subclass diversity and acquisition variability, supporting broader generalization analysis, as reported by **Yilmaz et al. [3]**.

This work presents a comprehensive classical framework focusing on: (1) a rich hand-crafted feature space (texture, shape, moments, intensity, and colour); (2) margin-based per-class sample cleaning to reduce noisy or ambiguous training samples; (3) controlled upsampling for class balancing without generating synthetic images; (4) mRMR feature selection to reduce redundancy; and (5)

Manuscript received on 02 February 2026 | Revised Manuscript received on 08 February 2026 | Manuscript Accepted on 15 February 2026 | Manuscript published on 28 February 2026.
*Correspondence Author(s)
Mohamed Amin Belal*, Faculty of Computers and Artificial Intelligence, Modern University for Technology and Information, Egypt. Email ID: abo_amin200@yahoo.com, ORCID ID: 0009-0005-3742-6480
Dr. Sara. I. Ibrahim, Faculty of Computer and Informatics, Suez Canal University, Computer Information Systems Department, Egypt. Email ID: sara_hassan@ci.suez.edu.eg, ORCID ID: 0000-0003-1328-3217
Prof. (Dr.) BenBella S. Tawfik, Faculty of Computer and Informatics, Suez Canal University, Computer Information Systems Department, Egypt. Email ID: benbellat@ci.suez.edu.eg, ORCID ID: 0000-0001-9352-7538
Prof. (Dr.) Mohamed M. El-Gazzar, Faculty of Computers and Artificial Intelligence, Modern University for Technology and Information, Egypt. Email ID: m_elgazzar@cs.mti.edu.eg, ORCID ID: 0000-0003-0144-6891

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open-access article under the CC-BY-NC-ND license https://creativecommons.org/licenses/by-nc-nd/4.0/

comparison between flat 4-class and two-level hierarchical classification, motivated by clinical reasoning, including oracle analysis to isolate routing effects.

A. Main Contributions

i. A unified feature vector combining HOG/LBP, region geometry, Hu moments, intensity statistics, and RGB/HSV descriptors.
ii. A per-class margin-based smart cleaning mechanism for removing unreliable samples.
iii. Controlled upsampling to balance classes while preserving the empirical distribution.
iv. mRMR feature selection to retain the most informative subset of features.
v. A fair comparison between flat and hierarchical decision strategies, including oracle analysis to isolate routing effects.

II. RELATED WORK

Early dermoscopic pipelines relied on engineered descriptors (border irregularity, asymmetry, texture, and colour variation) combined with classical learners such as SVM and ensemble trees. Standardized community evaluations accelerated progress by providing benchmarks and shared protocols; for example, **Codella et al. [4]** documented the ISIC challenge setting and evaluation framework that has been widely used for comparative analysis.

Deep learning subsequently dominated many dermoscopy benchmarks with strong discriminative performance, and combined-decision strategies have been investigated to improve classification, as shown by **Imran et al. [5]**. Lightweight attention-based architectures have also been explored; for example, **Yadav et al. [6]** proposed a cross-attention transformer design for skin lesion classification. Transformer-based learning has additionally been used to improve statistical texture representations for skin lesion segmentation, as presented by **Xu et al. [7]**.

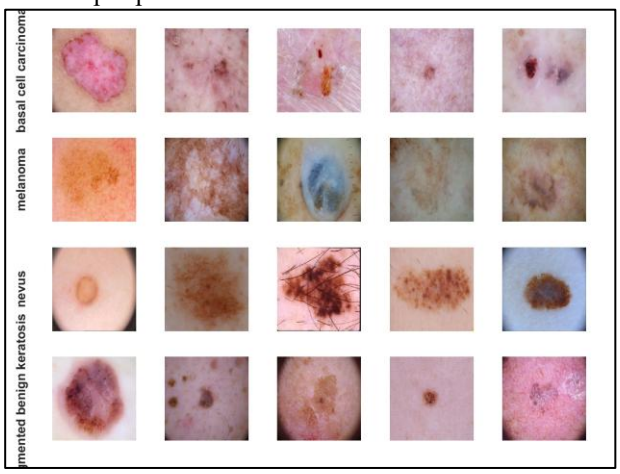
Medical imaging literature increasingly emphasises robustness and practical deployment, including calibration, uncertainty awareness, and generalisation under dataset shift; Meedeniya et al. [8] provide a focused survey on skin cancer identification. Zhang and Metaxas [9] discuss broader perspectives on foundation models and practical deployment constraints in medical image analysis. Data imbalance remains a key concern; augmentation methods, including GAN-based strategies, have been proposed to mitigate multi-class imbalance (Su et al., [10]). Finally, segmentation reliability can directly affect mask-dependent feature pipelines; an efficient, lightweight segmentation approach is described by **Yuan et al. [11]**.

III. DATASET DESCRIPTION

The dataset contains **1445 dermoscopic images** from four diagnostic categories: **BCC (344), melanoma (407), nevus (309), and PBK (385)**. The distribution is moderately imbalanced, which motivates class-aware preprocessing and balancing.

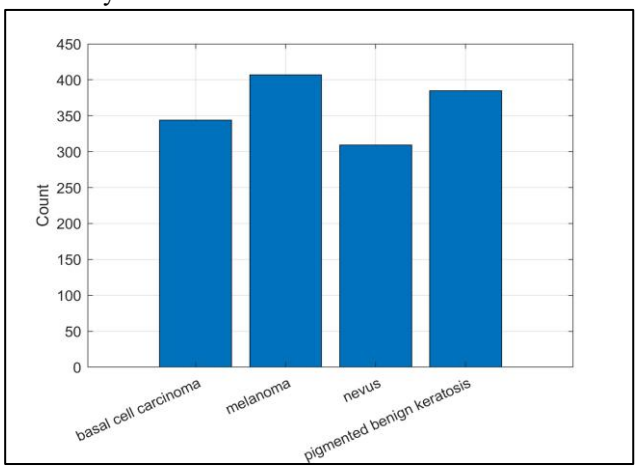
A. Class Descriptions (Clinical/Visual Intuition used by the Pipeline)

i. *BCC:* Common skin cancer with localised structures; texture, local colour distribution, and lesion geometry can be informative.
ii. *Melanoma:* Malignant lesion often showing irregular borders and varied pigmentation; asymmetry and atypical texture cues are important.
iii. *Nevus:* Benign mole with more uniform structure; challenging cases overlap visually with early/ambiguous melanoma.
iv. *PBK:* Benign keratosis with patchy pigmentation and surface texture; HSV distribution descriptors often help separate PBK from nevus.



[Fig.1: Sample Dermoscopic Images from Each Class (Intra-Class Variability and Inter-Class Overlap)]

Fig. 1. Sample dermoscopic images for the four studied classes (basal cell carcinoma, melanoma, nevus, and pigmented benign keratosis). Each panel should include multiple examples per class to show intra-class variability (e.g., border irregularity, pigmentation heterogeneity, texture patterns) and to illustrate inter-class visual overlap, particularly between melanoma and nevus.



[Fig. 2: Original Class Distribution Before Preprocessing (Total 1445 Images): BCC=344, Melanoma=407, Nevus=309, PBK=385]

Fig. 2. Original class distribution before preprocessing (total 1445 images). Bar chart showing

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

counts: BCC=344, Melanoma=407, Nevus=309, PBK=385. This figure motivates class-aware preprocessing and balancing.

IV. FEATURE EXTRACTION METHODOLOGY

Each image is mapped to a fixed-length feature vector integrating complementary descriptors. Feature extraction captures: (i) local texture and gradient structure; (ii) region geometry and boundary regularity; (iii) global shape invariants via Hu moments; (iv) lesion intensity statistics; and (v) colour distribution in multiple colour spaces. Region-based features are computed from a lesion mask derived using adaptive thresholding and largest-blob selection (fallback to full mask if needed).

A. Texture Descriptors

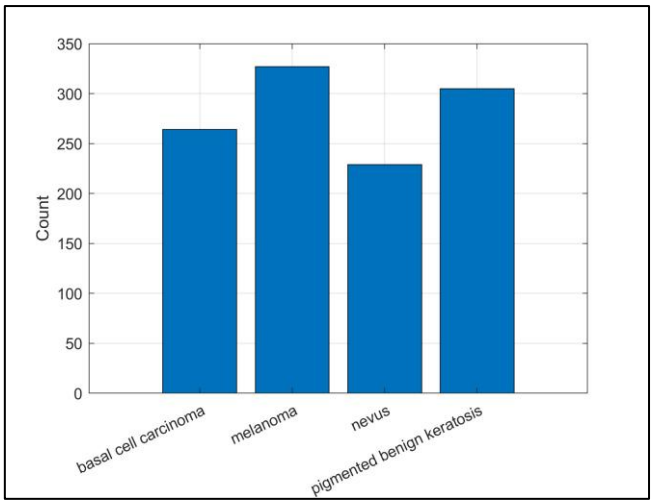
HOG summarizes gradient orientations over spatial cells and captures edge density and directional patterns; it is widely used as an efficient gradient descriptor, as introduced by **Dalal and Triggs [13]**.

LBP encodes micro-texture by thresholding neighbourhood intensities, providing robustness to monotonic illumination changes; texture-driven descriptors are often included in recent dermoscopic pipelines, such as the stacking approach reported by **Balasundaram et al. [14]**.

B. Region Geometry and Border Regularity

Region geometry is computed from the lesion mask using area, perimeter, eccentricity, solidity, extent, and related measures. Circularity is used as a compact measure of boundary regularity:

$$\text{ircuarity} = \frac{4\pi A}{p2+\varepsilon} \quad \ldots \quad (1)$$



[Fig.3: Effect of Preprocessing and Feature Extraction (Pipeline Overview)]

Fig. 3. Preprocessing and feature extraction overview. Diagram showing: image input → mask/segmentation (or provided mask) → feature computation blocks (HOG, LBP, region geometry including circularity, Hu moment invariants, intensity statistics, RGB/HSV statistics and histograms) → concatenation into a fixed-length feature vector → normalization. Include an annotation that feature extraction is parallelized in MATLAB for efficiency.

$$m_{pq} = \frac{\sum_X \sum_Y X^p Y^q}{(X,Y)} \quad \ldots \quad (2)$$

The centroid is computed from first-order moments:

$$\overline{X} = \frac{m_{10}}{m_{00}} \quad , \quad \overline{Y} = \frac{m_{01}}{m_{00}} \quad \ldots \quad (3)$$

Central moments remove translation dependence.

$$\mu_{pq} = \frac{\sum_X \sum_Y (X-\bar{X})^p (Y-\bar{Y})^q}{(X,Y)} \quad \ldots \quad (4)$$

Normalized central moments provide scale invariance:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\left(\frac{p+q}{2}+1\right)}} \quad \ldots \quad (5)$$

Finally, the seven Hu invariants are computed from normalized central moments:

(Scale invariance can be obtained by normalization)

$$\emptyset_1 = \eta_{20} + \eta_{02} \quad \ldots \quad (6)$$

$$\emptyset_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \quad \ldots \quad (7)$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \mu_{03})^2 \quad \ldots \quad (8)$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \mu_{03})^2 \quad \ldots \quad (9)$$

$$\begin{aligned}\phi_5 = {} & (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 \\ & - 3(\eta_{21} + \eta_{03})^2] \\ & + (3\eta_{21} - \eta_{03})(\eta_{21} \\ & + \eta_{03})[3(\eta_{30} + \eta_{12})^2 \\ & - (\eta_{21} + \eta_{03})^2] \quad \ldots \quad (10)\end{aligned}$$

$$\begin{aligned}\phi_6 = {} & (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\ & + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \quad \ldots \quad (11)\end{aligned}$$

$$\begin{aligned}\phi_7 = {} & (3\eta_{21} + \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 \\ & - 3(\eta_{21} + \eta_{03})^2] \\ & - (\eta_{30} - 3\eta_{12})(\eta_{21} \\ & + \eta_{03})[3(\eta_{30} + \eta_{12})^2 \\ & - (\eta_{21} + \eta_{03})^2] \quad \ldots \quad (12)\end{aligned}$$

We extract seven features from each image. These input features helped the classification model to classify the images. However, after extracting these features, we first applied the discrete cosine transformation. It is used in lossy image compression because it exhibits strong energy compaction, i.e., a large amount of information is stored in the signal's low-frequency components. The remaining frequencies have very small amounts of data, which can be stored using a very small number of bits (typically 2 or 3 bits). To perform the DCT Transformation on an image, we first extract the image file information (pixel values in the range 0–255), divide it into 8×8 blocks, and then apply the discrete cosine transform to each block.

These invariants capture global geometric signatures that can differentiate lesions with similar colour and texture but distinct macroscopic geometry. They are especially useful when subtle shape differences drive class boundaries.

© Copyright: All rights reserved.

$$\frac{X - \mu}{\sigma + \varepsilon} \quad \ldots \quad (13)$$

$$m_i = s(i, y_i) - \max_{k \neq y_i} s(i, k) \quad \ldots \quad (14)$$

C. Intensity Statistics

Intensity features computed within the lesion region include mean, standard deviation, skewness, kurtosis, and entropy. Beyond classification accuracy, uncertainty-aware analysis in medical imaging encourages reporting robust summary statistics that reflect heterogeneity, as discussed by **Peeters et al. [15]**.

D. Colour Descriptors and Normalization

Colour features are computed in both RGB and HSV spaces using per-channel mean and standard deviation, plus normalised HSV histograms. Handling imbalance and acquisition variability is important for stable colour statistics; imbalance-aware strategies are discussed in augmentation-focused work by **Su et al. [10]**.

All features are standardized using z-score normalization:

V. SMART CLEANING AND DATA BALANCING

A. Margin-Based Sample Reliability

A preliminary ECOC classifier provides class scores (s(i,k)). The margin for sample (i) is:

$$m_i = s(i, y_i) - \max_{k \neq y_i} s(i, k) \quad \ldots \quad (15)$$

Negative margins indicate misclassification, while small positive margins indicate weak confidence. Samples are ranked within each class, and removal begins with the sample with the smallest margin.

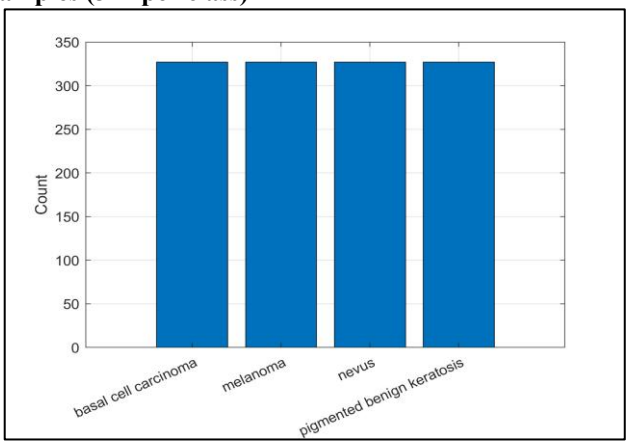
B. Per-Class Cleaning

Cleaning is applied independently to each class, with **Nremove = 80** samples per class. After cleaning, counts become:

BCC (344 → 264), melanoma (407 → 327), nevus (309 → 229), PBK (385 → 305), total (1125).

C. Controlled Upsampling

Each class is upsampled to the maximum cleaned class count (**327**) by controlled replication, producing **1308 samples (327 per class)**.



[Fig.4: Dataset Evolution (Original → Cleaned → Upsampled) with Counts Shown for Each Stage]

Fig. 4. Dataset evolution through preprocessing. Three grouped bars (or three panels): (a) Original counts (1445 total), (b) After smart cleaning with Nremove=80 per class (1125 total; BCC 264, Melanoma 327, Nevus 229, PBK 305), and (c) After controlled upsampling to maxCount=327 (1308 total; 327 per class).

VI. FEATURE SELECTION USING mRMR

The unified feature vector is high-dimensional and can contain redundancy. mRMR ranking is used to select the most informative features by maximising relevance while minimising redundancy; the mutual-information formulation was introduced by **Peng et al. [12]**. A compact objective can be written as:

$$R = \frac{1}{|S|^2} \sum_{f_j,\, f_k \in S} I(f_j ; f_k) \quad \ldots \quad (16)$$

$$\max \Phi = D - R \quad \ldots \quad (17)$$

The top **300** features are retained for all comparisons to balance accuracy and runtime. In high-dimensional settings, feature selection is commonly used to improve generalization and reduce redundancy, as illustrated in an applied classification context by **Gudla et al. [18]**.

VII. CLASSIFICATION MODELS

Four models are evaluated in a flat setting: SVM-RBF, Random Forest, KNN, and a shallow Neural Network baseline.

A. SVM with RBF Kernel

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad \ldots \quad (18)$$

SVM-RBF provides flexible nonlinear separation and remains competitive in dermoscopic classification comparisons, as shown by **Imran et al. [5]**.

B. Random Forest

Random Forest aggregates multiple decision trees and is robust to nonlinear feature interactions. Ensemble-based approaches are widely used in medical prediction settings, as evidenced by systematic reviews by **Wulaningsih et al. [16]**.

C. k-Nearest Neighbors

KNN performs neighbourhood voting in feature space; it can be sensitive to residual noise and high dimensionality, motivating normalisation and feature selection. Multi-site and distributed learning considerations affecting feature geometry are summarised by **Guan et al. [17]**.

D. Shallow Neural Network

A shallow NN is used as a lightweight nonlinear baseline; practical trade-offs between capacity, compute, and reliability are discussed more generally in medical imaging by **Zhang and Metaxas [9]**.

© Copyright: All rights reserved.

VIII. HIERARCHICAL CLASSIFICATION FRAMEWORK

We use a clinically motivated two-level hierarchical decomposition:

- **Level-1:** Cancer vs non-Cancer
- **Level-2a:** BCC vs melanoma
- **Level-2b:** nevus vs PBK

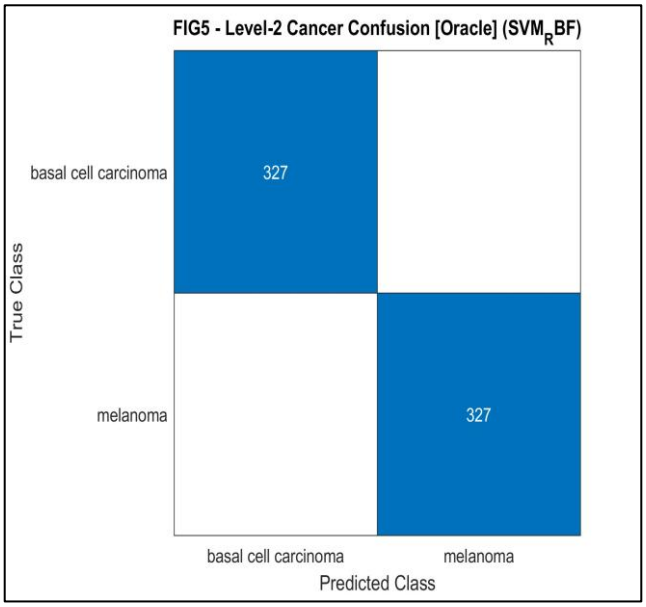
Level-1 decision:

$$\hat{y}_1 = \arg_{C \in \{\text{Cancer},\ \text{Non-Cancer}\}} \max f_1(X) \quad \ldots \quad (19)$$

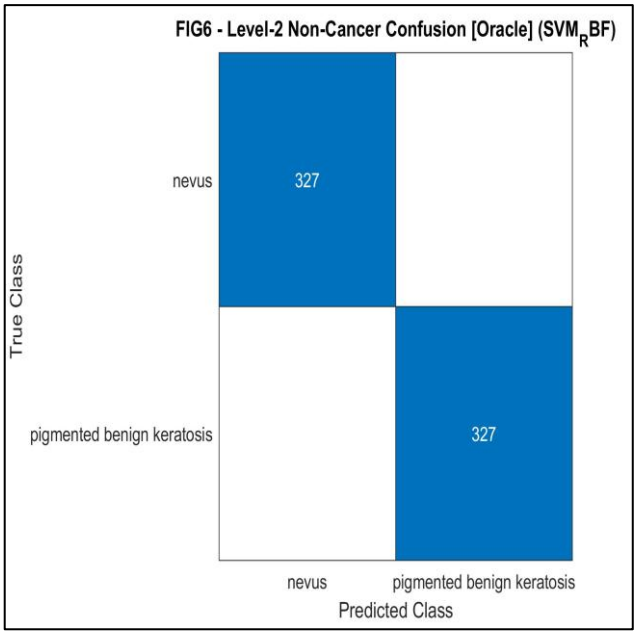
Level-2 decision (branch-dependent):

$$\hat{y} = f_2^{(b)}(x),\ b \in \{\text{Cancer},\ \text{Non} - \text{Cancer}\} \quad \ldots \quad (20)$$

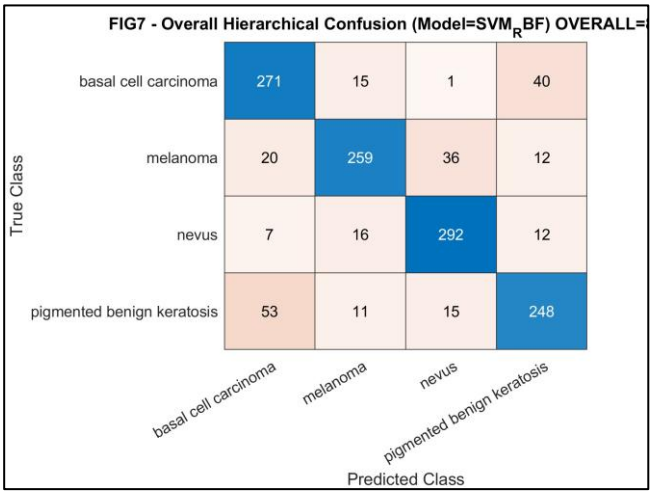
Oracle evaluation supplies true Level-1 labels when testing Level-2 to isolate subtype separability without routing errors.



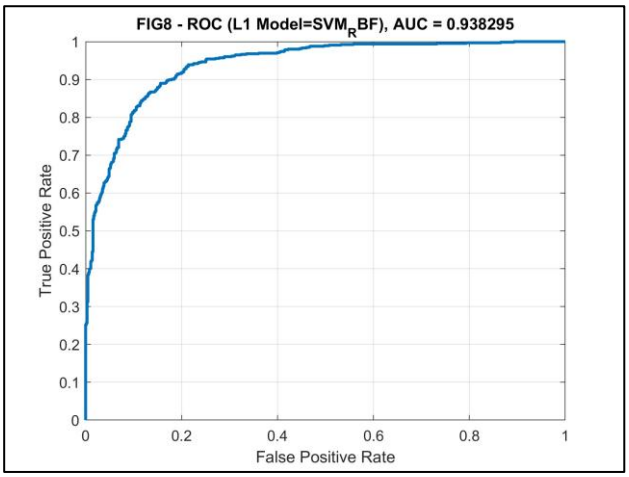
[Fig.5: Level-1 Confusion Matrix (Cancer vs Non-Cancer) — Level-1 Accuracy = 86.54%]



[Fig.6: Level-2 Cancer Oracle Confusion (BCC vs Melanoma) — 94.04%]



[Fig.7: Level-2 Non-Cancer Oracle Confusion (Nevus vs PBK) — 94.19%]



[Fig.8: Overall Hierarchical Confusion (4 classes) — Overall Accuracy = 81.80%]

IX. EXPERIMENTAL SETUP

A. Cross-Validation Protocol

All experiments use stratified **5-fold cross-validation**. In each fold, the pipeline is applied strictly within the training split: normalisation parameters are fit on the training data only, smart cleaning is performed using a preliminary model trained on the training data only, upsampling is applied to the training data only, and feature selection is fit on the training data only. Leakage-avoidance principles are consistent with challenge-based evaluation practices described by **Codella et al. [4]**.

B. Metrics

Accuracy is reported as the primary metric, and confusion matrices provide class-wise insight. For Level-1 malignancy detection, ROC and PR curves are computed, and AUC values are reported.

C. Implementation Details

Experiments are implemented in MATLAB R2022. Feature extraction is parallelized to reduce runtime. Flat and hierarchical models are compared using the same best-performing model configuration for fairness.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

X. RESULTS AND ANALYSIS

A. Dataset Processing Summary

Original dataset: **1445** images (BCC 344, melanoma 407, nevus 309, PBK 385).

After smart cleaning (Nremove=80 per class): **1125** total (BCC 264, melanoma 327, nevus 229, PBK 305).

After controlled upsampling: **1308** total (**327 per class**).

Feature selection retained the **top 300** features.

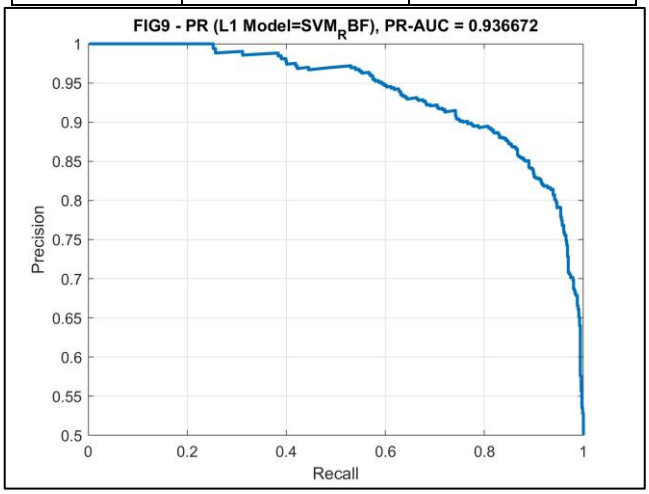
B. Flat 4-Class Model Comparison

Best model: **SVM-RBF,** mean accuracy **82.87%**.

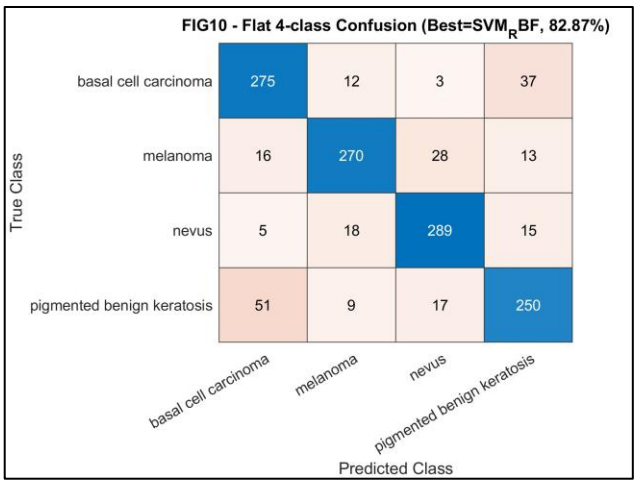
Other models: **RF 81.65%**, **NN 80.97%**, **KNN 60.24%**.

Table 1: Flat 4-Class Accuracy Comparison Across Models (5-fold Mean Accuracy)

Model	Mean Accuracy (%)	Evaluation
SVM-RBF	82.87	Stratified 5-fold CV
Random Forest	81.65	Stratified 5-fold CV
Shallow NN	80.97	Stratified 5-fold CV
KNN	60.24	Stratified 5-fold CV



[Fig.9: Confusion Matrix of Best Flat Model (SVM-RBF)]



[Fig.10: Fold-wise Accuracy for SVM-RBF Across 5 Folds]

C. Hierarchical Results and Oracle Analysis

Using SVM-RBF for both levels:

i. Level-1 accuracy = **86.54%**

ii. L2 Cancer oracle accuracy = **94.04%**

iii. L2 Non-Cancer oracle accuracy = **94.19%**

iv. Overall hierarchical accuracy = **81.80%**

The gap between Oracle Level-2 performance and final hierarchical accuracy indicates that the dominant source of degradation is Level-1 misrouting.

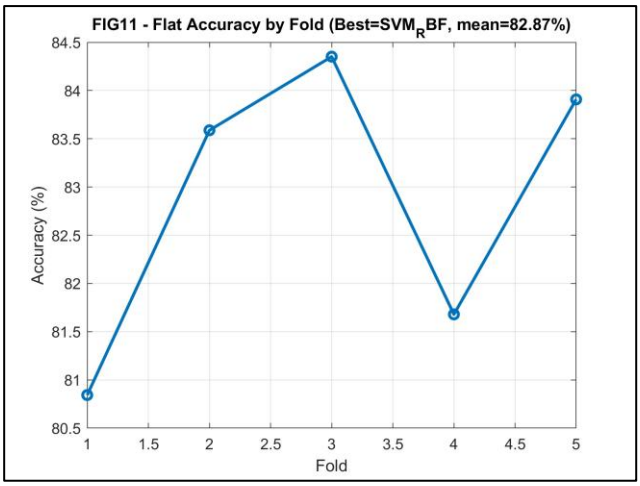
D. ROC and PR Analysis for Level-1 Malignancy Detection

i. ROC AUC (overall) = **0.938295**

ii. PR AUC (overall) = **0.936672**

iii. ROC AUC (mean ± std across folds) = **0.938533 ± 0.014707**

iv. PR AUC (mean ± std across folds) = **0.930521 ± 0.014528**



[Fig. 11: ROC Curve for Level-1 Malignancy Detection Showing AUC]

XI. DISCUSSION

The results show that a carefully engineered classical pipeline can perform strongly on dermoscopic lesion classification when supported by robust preprocessing and feature selection. The best flat model achieves **82.87%** accuracy, while the hierarchical approach achieves **81.80%** and provides a clinically interpretable decision structure.

The hierarchical model is primarily constrained by error propagation, since Level-1 errors cannot be corrected at Level-2. Oracle results (≈approximately 94% in both branches) confirm strong intra-branch separability and indicate that improving routing reliability is the most promising path forward. Practical directions include confidence-aware routing thresholds, calibration, and uncertainty handling, as emphasized in skin cancer surveys by **Meedeniya et al. [8]** and broader medical imaging perspectives by **Zhang and Metaxas [9]**.

Recent trends also include optimised feature-fusion systems (e.g., Akram et al. [1]) and lightweight attention-based transformers (e.g., Yadav et al. [6]), suggesting that classical and deep approaches can be complementary depending on deployment constraints. Finally, because mask quality affects shape and moment descriptors, efficient segmentation improvements—such as the lightweight approach by **Yuan et al. [11]**—can further stabilize region-based features.

© Copyright: All rights reserved.

XII. CONCLUSION AND FUTURE WORK

This paper presents a classical machine learning framework for four-class dermoscopic lesion classification that employs a unified handcrafted feature set, margin-based per-class smart cleaning, controlled upsampling, and mRMR feature selection. Under stratified 5-fold cross-validation, the best flat model (SVM-RBF) achieved **an accuracy of 82.87%**. The hierarchical system achieved **an overall accuracy of 81.80%, with Level-1 accuracy of 86.54%. Oracle** Level-2 accuracies above **94%** confirmed strong subtype separability and identified Level-1 routing errors as the primary bottleneck. Level-1 malignancy detection showed strong discrimination with **ROC AUC = 0.9383** and **PR AUC = 0.9367**.

Future work will focus on improving Level-1 routing reliability through calibration and thresholding, confidence-aware deferral, and cost-sensitive learning. Extending to more lesion categories and validating on additional multi-source datasets will further assess generalization.

DECLARATION STATEMENT

After aggregating input from all authors, I must verify the accuracy of the following information as the article's author.

- **Conflicts of Interest/ Competing Interests:** Based on my understanding, this article has no conflicts of interest.
- **Funding Support:** This article has not been funded by any organizations or agencies. This independence ensures that the research is conducted objectively and free from external influence.
- **Ethical Approval and Consent to Participate:** The content of this article does not necessitate ethical approval or consent to participate with supporting documentation.
- **Data Access Statement and Material Availability:** The adequate resources of this article are publicly accessible.
- **Author's Contributions:** The authorship of this article is contributed equally to all participating individuals.

REFERENCES

1. T. Akram, et al., "Dermo-Optimiser: Skin lesion classification using information-theoretic deep feature fusion and entropy-controlled binary bat optimisation," *International Journal of Imaging Systems and Technology*, 2024. DOI: https://doi.org/10.1002/ima.23172
2. T. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset: a large collection of multi-sources dermatoscopic images of common pigmented skin lesions," *Scientific Data*, vol. 5, 180161, 2018. DOI: https://doi.org/10.1038/sdata.2018.161
3. A. Yilmaz, et al., "DERM12345: A Large, Multisource Dermatoscopic Skin Lesion Dataset with 40 Subclasses," *Scientific Data*, vol. 11, art. 1302, 2024. DOI: https://doi.org/10.1038/s41597-024-04104-3
4. N. C. F. Codella, et al., "Skin lesion analysis toward melanoma detection: A challenge at the 2017 ISBI, hosted by the ISIC," in *IEEE ISBI*, 2018, pp. 168–172. DOI: https://doi.org/10.1109/ISBI.2018.8363547
5. A. Imran, A. Nasir, M. Bilal, G. Sun, A. Alzahrani, and A. Almuhaimeed, "Skin Cancer Detection Using Combined Decision of Deep Learners," *IEEE Access*, vol. 10, pp. 118198–118212, 2022. DOI: https://doi.org/10.1109/ACCESS.2022.3220329
6. D. P. Yadav, B. Sharma, S. Chauhan, J. L. Webber, and A. Mehbodniya, "Dual-scale lightweight cross-attention transformer for skin lesion classification," *PLOS ONE*, vol. 19, no. 12, e0312598, 2024. DOI: https://doi.org/10.1371/journal.pone.0312598
7. R. Xu, C. Wang, J. Zhang, S. Xu, W. Meng, and X. Zhang, "SkinFormer: Learning Statistical Texture Representation With Transformer for Skin Lesion Segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 10, pp. 6008–6018, 2024. DOI: https://doi.org/10.1109/JBHI.2024.3417247
8. D. Meedeniya, S. De Silva, L. Gamage, and U. Isuranga, "Skin cancer identification utilizing deep learning: A survey," *IET Image Processing*, vol. 18, no. 13, pp. 3731–3749, 2024. DOI: https://doi.org/10.1049/ipr2.13219
9. S. Zhang and D. Metaxas, "On the challenges and perspectives of foundation models for medical image analysis," *Medical Image Analysis*, vol. 91, art. 102996, 2024. DOI: https://doi.org/10.1016/j.media.2023.102996
10. Q. Su, H. N. A. Hamed, M. A. Isa, X. Hao, and X. Dai, "A GAN-Based Data Augmentation Method for Imbalanced Multi-Class Skin Lesion Classification," *IEEE Access*, vol. 12, pp. 16498–16513, 2024. DOI: https://doi.org/10.1109/ACCESS.2024.3360215
11. C. Yuan, D. Zhao, and S. S. Agaian, "UCM-Net: A lightweight and efficient solution for skin lesion segmentation using MLP and CNN," *Biomedical Signal Processing and Control*, vol. 96, pt. B, art. 106573, 2024. DOI: https://doi.org/10.1016/j.bspc.2024.106573
12. H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005. DOI: https://doi.org/10.1109/TPAMI.2005.159
13. N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *CVPR*, 2005, pp. 886–893. DOI: https://doi.org/10.1109/CVPR.2005.177
14. A. Balasundaram, A. Shaik, B. R. Alroy, A. Singh, and S. J. Shivaprakash, "Genetic Algorithm Optimized Stacking Approach to Skin Disease Detection," *IEEE Access*, vol. 12, pp. 88950–88962, 2024. DOI: https://doi.org/10.1109/ACCESS.2024.3412791
15. D. Peeters et al., "Enhancing a deep learning model for pulmonary nodule malignancy risk estimation in chest CT with uncertainty estimation," *European Radiology*, vol. 34, pp. 6639–6651, 2024. DOI: https://doi.org/10.1007/s00330-024-10714-7
16. W. Wulaningsih, et al., "Deep Learning Models for Predicting Malignancy Risk in CT-Detected Pulmonary Nodules: A Systematic Review and Meta-analysis," *Lung*, vol. 202, pp. 625–636, 2024. DOI: https://doi.org/10.1007/s00408-024-00706-1
17. H. Guan, P.-T. Yap, A. Bozoki, and M. Liu, "Federated learning for medical image analysis: A survey," *Pattern Recognition*, vol. 151, art. 110424, 2024. DOI: https://doi.org/10.1016/j.patcog.2024.110424
18. R. Gudla, S. Vollala, S. K. G., and R. Amin, "A novel approach for classification of Tor and non-Tor traffic using efficient feature selection methods," *Expert Systems with Applications*, vol. 249, art. 123544, 2024. DOI: https://doi.org/10.1016/j.eswa.2024.123544

AUTHOR'S PROFILE

Mohamed Amin Belal is a faculty member in the Computer Information Systems Department, Faculty of Computers and Artificial Intelligence, at the Modern University for Technology and Information. He received B.Sc. and M.Sc. degrees from Military Technical College in 1987 and 1994, respectively. He is a Teaching Assistant at MTI University. His interests are image processing, cybersecurity, computer vision, data mining, artificial intelligence, image segmentation and image classification.

Dr. Sara Ibraim Ibrahim is a faculty member in the Computer Information Systems Department, Faculty of Computers and Informatics, Suez Canal University. She earned her B.Sc., M.Sc., and PhD from Suez Canal University in 2011, 2017, and 2023, respectively. She is a Lecturer at Suez Canal University. Her interests include image fusion, image processing, cybersecurity, computer vision, data mining, artificial intelligence, bioinformatics, image segmentation, image classification, and image super-resolution.

© Copyright: All rights reserved.

Prof. Dr. Ben Bella Sayed Tawfik is a faculty member in the Computer Information Systems Department, Faculty of Computers and Informatics, Suez Canal University, and a member of the Planning Committee of the Computer and Information Sector at Egypt's Ministry of Higher Education. He earned his PhD in Radar Image Analysis using the Wavelet Transform from Colorado State University in 1998, with a focus on environmental observation and melting-layer detection. He later completed a Post-Doctoral Certificate in Computer Engineering (2006), including analysis of Tropical Rainfall Measuring Mission satellite data. With over two decades of academic experience, his research spans artificial intelligence and data-driven computing, including machine learning, deep learning, data science, optimisation, forecasting, and wireless sensor networks, with particular contributions to AI applications in graphical and montage design. He has taught in engineering schools in Egypt since 1998 and has held sustained academic leadership responsibilities, including eight years as department chief.

Prof. Dr. Mohamed Mohamed Elgazzar, received the B.Sc in engineering degree from MTC, Cairo, Egypt, in 1984 and M.Sc degree in Computer Networks from Faculty of Engineering, Cairo University, Egypt, in 1991 and the Ph.D degree from Faculty of Engineering, Cairo University, Egypt, and He is a faculty member in the Computer Science Department, Faculty of Computers and Artificial Intelligence, Modern University for Technology and Information, at Egypt's Ministry of Higher Education. He has 36 years of experience, including both academic and scientific research. He is currently a Full Professor in Computer Science. His research interests include Computer Networks, Network Security, Machine Learning, and Cybersecurity. He has published more than 30 research papers across more than 4 research areas.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of the Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)/ journal and/or the editor(s). The Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP) and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

© Copyright: All rights reserved.