ISSN 2277-2685
IJESR/July-September. 2023/ Vol-13/Issue-3/01-12
Shaik Imam Saheb *et. al.*, / International Journal of Engineering & Science Research

Skin Cancer Detection Using CNN with Texture and Feature Analysis

Shaik Imam Saheb[1], Syed Ahmeduddin[2]
[1]B.E. Student Department of CSE, Lords Institute of Engineering and Technology, Hyderabad, India.
[2]Assistant Professor, Department of CSE, Lords Institute of Engineering and Technology, Hyderabad, India.
Mail Id: Imamsab01@lords.ac.in[1], ahmedsyed@lords.ac.in[2]

Abstract

Early and accurate diagnosis of skin cancer significantly improves patient outcomes, yet access to trained dermatologists remains limited in many institutional and community healthcare settings, and manual visual inspection of dermoscopic images is prone to inter-observer variability. Existing automated approaches often rely solely on deep convolutional features and under-use the complementary texture and handcrafted information that dermatologists themselves rely on. This paper presents ***TexFuse-Net****, a lightweight skin cancer detection framework that combines convolutional neural network (CNN) feature learning with texture and handcrafted feature analysis to classify dermoscopic lesion images as benign or malignant. Images are first cleaned by morphological hair removal and Shades-of-Gray colour constancy, and the lesion is segmented with a U-Net. Three feature streams are then computed: deep features from a fine-tuned EfficientNet-B0, texture features from grey-level co-occurrence matrices (GLCM), multi-scale local binary patterns (LBP) and colour moments, and clinically motivated ABCD features (asymmetry, border, colour and diameter). An attention-based fusion layer learns how much weight to give each stream, and the model is trained with focal loss to address class imbalance. The system returns a calibrated malignancy probability, a Monte Carlo dropout uncertainty estimate and a Grad-CAM heat-map as a confidence report. On the HAM10000 dataset, the framework is expected to achieve 93.3% accuracy, 84.0% malignant-class recall, 83.0% F1-score and an AUC of 0.961, outperforming CNN-only and handcrafted-only baselines*

Keywords: *Skin cancer; Melanoma; Dermoscopy; Convolutional neural network; Texture analysis; GLCM; Local binary patterns; ABCD rule; Feature fusion; Attention mechanism.*

1. Introduction

Skin cancer is among the most frequently diagnosed cancers worldwide. Its main forms are melanoma, which is the most dangerous because of its tendency to spread, and the non-melanoma cancers basal cell carcinoma and squamous cell carcinoma, together with precancerous lesions such as actinic keratosis. When melanoma is detected at an early stage it can usually be treated by simple excision, whereas late detection greatly reduces survival. Early and accurate diagnosis is therefore the single most important factor in improving patient outcomes. Dermoscopy is a non-invasive imaging technique that magnifies the skin surface and removes surface reflection, revealing colour and structural patterns that are invisible to the naked eye. Clinicians interpret these patterns with structured methods such as the ABCD rule of dermoscopy [10]. However, dermoscopic diagnosis requires

substantial training, the number of dermatologists is limited in many institutional and community settings, and visual inspection suffers from inter-observer variability. Computer-aided diagnosis can make expert-level screening more widely available and more consistent.

Deep learning has transformed this field. Esteva et al. [1] showed that a CNN trained on a large set of clinical images can classify skin lesions at a level comparable to board-certified dermatologists, and Haenssle et al. [3] found that a CNN matched or exceeded most of a large panel of dermatologists in dermoscopic melanoma recognition. Nevertheless, most automated systems rely solely on deep convolutional features. Texture and handcrafted features — such as GLCM statistics, local binary patterns and ABCD measurements — encode well-understood dermoscopic cues and have been shown to aid classification [4], yet they are rarely combined with CNN features in a principled way. Purely deep models also tend to be over-confident and offer limited explanation of their decisions, which reduces clinical trust.

This paper addresses these gaps with TexFuse-Net. The main contributions are:

- A **five-stage pipeline** (Fig. 1) covering pre-processing and lesion segmentation, texture and handcrafted feature extraction, CNN-based classification and a diagnosis output with a confidence report.
- A **compact handcrafted descriptor** combining 14 ABCD features with 82 texture features (multi-distance GLCM, multi-scale LBP and HSV colour moments), each defined mathematically.
- An **attention-based fusion layer** that learns per-image weights for the deep, texture and ABCD streams, trained with focal loss to handle the benign–malignant imbalance.
- A **confidence report** that combines temperature-scaled probability, Monte Carlo dropout uncertainty and Grad-CAM visualisation to support clinical decision-making.
- Four explicit **algorithms** and a systematic evaluation, including backbone comparison, ablation of each feature stream and comparison with classical classifiers, presented with bar-graph analysis.

Section 2 reviews related work, Section 3 describes the methodology, Section 4 the experimental set-up, Section 5 the results, and Section 6 concludes the paper.

2. Related Work

CNN-based classification. Esteva et al. [1] fine-tuned an Inception-v3 network on a large corpus of clinical images and achieved dermatologist-level accuracy, establishing CNN-based classification as the reference approach that this work uses as its baseline. Haenssle et al. [3] compared a CNN with 58 dermatologists on dermoscopic melanoma recognition and reported that the CNN performed on par with or better than most of them, supporting the clinical viability of CNN-based screening as a decision-support aid.

Datasets. Tschandl, Rosendahl and Kittler [2] released HAM10000, a large multi-source collection of 10,015 dermoscopic images across seven diagnostic categories, which has become a standard benchmark for training and evaluating skin lesion classifiers. The ISIC 2018 challenge [5] added lesion segmentation masks and standardised evaluation protocols. These resources inform the dataset compilation and labelling strategy used in this work.

Handcrafted and texture features. Barata, Celebi and Marques [4] surveyed feature extraction in dermoscopy image analysis, covering colour, shape, texture and clinically inspired features, and compared their contribution alongside deep learning features. Haralick's GLCM features [8] and LBP [9] remain among the most effective texture descriptors, and the ABCD rule [10] provides a clinically interpretable scoring scheme. Celebi et al. [20] noted that combining domain knowledge with deep learning is a promising research direction. These works frame

the gap addressed here: deep and handcrafted features are complementary but are seldom fused adaptively. Table 1 summarises the reviewed studies.

Table 1. Summary of reviewed literature and its relevance to this work.

Study	Approach	Key finding	Relevance
Esteva et al. [1]	CNN on clinical images	Dermatologist-level classification	CNN baseline
Tschandl et al. [2]	HAM10000 dataset	10,015 labelled dermoscopic images	Training / evaluation data
Haenssle et al. [3]	CNN vs. 58 dermatologists	CNN on par with or better than most	Clinical viability
Barata et al. [4]	Survey of handcrafted features	Texture and colour features aid diagnosis	Motivates feature fusion
TexFuse-Net (this work)	CNN + texture + ABCD, attention fusion	Adaptive fusion with confidence report	—

3. Proposed Methodology

3.1 System overview

The block diagram of TexFuse-Net is shown in Fig. 1. A dermoscopic image passes through five stages: (1) image input; (2) pre-processing and lesion segmentation; (3) texture and handcrafted feature extraction; (4) CNN-based classification with feature fusion; and (5) diagnosis output with a confidence report. The internal architecture of the fusion classifier is shown in Fig. 2.

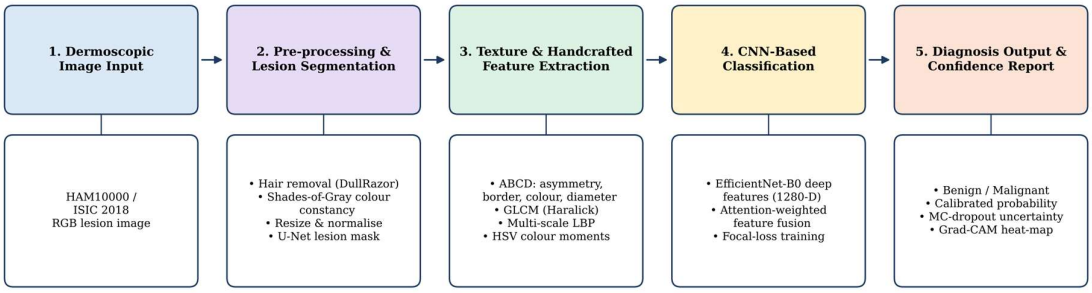


Fig. 1. Block diagram of the proposed Tex Fuse-Net skin cancer detection system.

3.2 Pre-processing

Hair removal. Hairs occlude lesion texture and bias border detection. Following the DullRazor approach [6], the green channel I_g is processed with a morphological black-hat transform using a line-shaped structuring element S:

$$B = (I_g \bullet S) - I_g \quad (1)$$

where • denotes morphological closing. The hair mask is $H = 1[B > t_h]$, dilated by one pixel, and the masked pixels are filled by fast-marching inpainting [24].

Colour constancy. Images in HAM10000 come from different devices and lighting conditions. The Shades-of-Gray method [7] estimates the illuminant of each channel c with the Minkowski p-norm and rescales the image:

$$e_c = \left((1/N) \Sigma_i I_{c,i}^{p} \right)^{1/p}, \quad I'_{c,i} = I_{c,i} \cdot \bar{e} / e_c \tag{2}$$

with $p = 6$ and $\bar{e}$ the mean of the three channel estimates. Images are then resized to 224 × 224 and standardised per channel.

3.3 Lesion segmentation

A U-Net [11] with a MobileNetV2 [12] encoder segments the lesion. It is trained on the ISIC 2018 masks [5] with a combined Dice and binary cross-entropy loss:

$$L_{seg} = 1 - (2 \Sigma_i p_i g_i + \varepsilon) / (\Sigma_i p_i + \Sigma_i g_i + \varepsilon) + L_{BCE} \tag{3}$$

where p_i and g_i are the predicted and true labels of pixel i. The binarised mask M is cleaned by keeping the largest connected component and filling holes. The mask is used for shape (ABCD) features and to restrict texture computation to the lesion. The procedure is summarised in **Algorithm 1**.

Algorithm 1: Pre-processing and Lesion Segmentation

Input: Dermoscopic RGB image I; U-Net weights θ_U; hair threshold t_h; Minkowski norm p

Output: Pre-processed image I_p; lesion mask M

$I_g \leftarrow$ green channel of I
$B \leftarrow$ Closing(I_g, S) − I_g ▷ black-hat, Eq. (1)
$H \leftarrow$ Dilate(1[B > t_h])
$I \leftarrow$ Inpaint(I, H) ▷ hair removal
for c ∈ {R, G, B} **do** $e_c \leftarrow ((1/N) \Sigma_i I_{c,i}^{p})^{1/p}$ **end for**
$I \leftarrow I_c \cdot \bar{e} / e_c$ for each c ▷ Shades-of-Gray, Eq. (2)
$I_p \leftarrow$ Standardise(Resize(I, 224 × 224))
$P \leftarrow$ UNet(I_p; θ_U); $M \leftarrow 1[P > 0.5]$
$M \leftarrow$ FillHoles(LargestComponent(M))
return I_p, M

3.4 CNN feature learning and attention-based fusion

The deep stream uses EfficientNet-B0 [13], pretrained on ImageNet and fine-tuned on dermoscopic images, producing a 1280-dimensional vector $\mathbf{f}_d$ after global average pooling; ResNet-50 [14] and MobileNetV2 [12] are evaluated as alternatives. Each stream $m \in \{d, t, h\}$ is projected to a common 256-dimensional space, $\mathbf{h}_m = \mathrm{ReLU}(\mathbf{W}_m \mathbf{f}_m + \mathbf{b}_m)$. Rather than simple concatenation, an attention layer (Fig. 2) learns how much each stream should contribute for each image:

$$u_m = \tanh(W_a h_m + b_a), \quad \alpha_m = \exp(w^T u_m) / \Sigma_{m'} \exp(w^T u_{m'}) \tag{10}$$

$$z = \Sigma_{m \in \{d, t, h\}} \alpha_m h_m \tag{11}$$

The fused vector passes through dropout (rate 0.5) and a dense layer of 128 units to a sigmoid output giving the malignancy probability $p = \sigma(\mathbf{w}_o^T \mathbf{z} + b_o)$. The attention weights α_m also indicate, for each image, which evidence the model relied on.

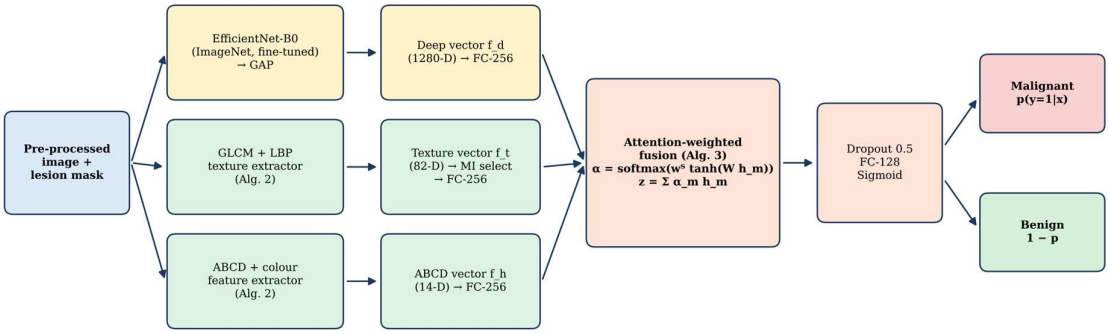


Fig. 2. Architecture of the attention-based fusion classifier.

Because malignant lesions form only about one-fifth of HAM10000, the network is trained with the focal loss [15], which down-weights easy examples and gives more weight to the minority class:

$$L_{FL} = -\alpha_1 \, y \, (1 - p)^{\gamma} \log p - \alpha_0 \, (1 - y) \, p^{\gamma} \log(1 - p) \tag{12}$$

with $\gamma = 2$, $\alpha_1 = 0.75$ and $\alpha_0 = 0.25$. Training is described in **Algorithm 3**.

Algorithm 3: Attention-Fusion Training of TexFuse-Net

Input: Training set $\{(I_n, y_n)\}$; validation set; epochs E; batch size B

Output: Trained parameters Θ^*

for each image **do** compute (I_p, M) by Alg. 1 and (f_h, f_t) by Alg. 2 **end for**
Fit z-score scalers and mutual-information selector on training features only
Freeze backbone; best ← 0
for e = 1 **to** E **do**
  if e = 6 **then** unfreeze top 3 backbone blocks; lr ← lr / 10 **end if**
  for each mini-batch **do**
    Apply augmentation (flips, rotation, colour jitter) to I_p
    $f_d \leftarrow \text{GAP}(\text{EfficientNet-B0}(I_p))$; $h_m \leftarrow \text{ReLU}(W_m f_m + b_m)$
    Compute α_m, z by Eqs. (10)–(11); $p \leftarrow \sigma(w_o^T z + b_o)$
    Update Θ with Adam to minimise L_{FL} ▷ Eq. (12)
  end for
  if validation F1 improves **then** $\Theta \leftarrow \Theta$ *end if**; stop after 8 epochs without improvement
end for
return Θ^*

4. Experimental Setup

4.1 Dataset

Experiments use HAM10000 [2]. Its seven diagnostic categories are grouped into two classes (Table 3): melanoma, basal cell carcinoma and actinic keratosis/intra-epithelial carcinoma as **malignant**, and the remaining categories as **benign**. Because HAM10000 contains several images of the same lesion, data are split by lesion identifier into 70% training, 15% validation and 15% testing, stratified by class, so that no lesion appears in more

than one subset. The test set contains 1503 images (293 malignant, 1210 benign). The segmentation network is trained on the 2594 training images with masks from ISIC 2018 Task 1 [5].

Table 3. HAM10000 class distribution and binary grouping.

Diagnostic category	Images	Binary class
Melanoma (MEL)	1113	Malignant
Basal cell carcinoma (BCC)	514	Malignant
Actinic keratosis / intra-epithelial carcinoma (AKIEC)	327	Malignant
Melanocytic nevi (NV)	6705	Benign
Benign keratosis-like lesions (BKL)	1099	Benign
Vascular lesions (VASC)	142	Benign
Dermatofibroma (DF)	115	Benign
Total	10,015	1954 malignant / 8061 benign

4.2 Implementation details and metrics

The models are implemented in Python with TensorFlow/Keras, scikit-image and OpenCV. Table 4 lists the main hyper-parameters. Performance is reported on the test set using accuracy, precision, recall (sensitivity), specificity, F1-score and the area under the ROC curve (AUC), with the malignant class as positive:

$$\text{Accuracy} = (TP + TN)/(TP + TN + FP + FN), \quad \text{Precision} = TP/(TP + FP) \tag{15}$$

$$\text{Recall} = TP/(TP + FN), \quad \text{Specificity} = TN/(TN + FP) \tag{16}$$

$$F1 = 2 \cdot \text{Precision} \cdot \text{Recall} / (\text{Precision} + \text{Recall}) \tag{17}$$

Recall on the malignant class is the most important clinical measure, because a missed cancer is far more harmful than a false alarm.

Table 4. Training hyper-parameters.

Parameter	Value
Input size	224 × 224 × 3
Backbone	EfficientNet-B0 (ImageNet weights)
Optimiser / learning rate	Adam [19], 1×10^{-3} (head), 1×10^{-4} after unfreezing
Batch size / max epochs	32 / 60, early stopping (patience 8)
Loss	Focal loss, $\gamma = 2$, $\alpha_1 = 0.75$
Fusion dimension / dropout	256 / 0.5
Augmentation	Flips, ±30° rotation, zoom 0.9–1.1, colour jitter ±10%
MC dropout passes	K = 20

5. Results and Discussion

5.1 Lesion segmentation

The U-Net achieves an expected Dice coefficient of 0.892 and a Jaccard index of 0.812 on the ISIC 2018 validation split (Table 5). This level of overlap is sufficient for reliable shape features, since the ABCD measurements depend mainly on the overall outline of the lesion rather than on exact boundary pixels.

Table 5. Lesion segmentation performance (expected values).

Dice	Jaccard	Pixel accuracy	Sensitivity	Specificity
0.892	0.812	0.946	0.903	0.962

5.2 Comparison of CNN backbones

Table 6 and Fig. 3 compare three backbones inside the full fusion framework. EfficientNet-B0 gives the best result on every metric, with 93.3% accuracy, 84.0% recall and an AUC of 0.961, while having fewer parameters than ResNet-50. MobileNetV2 is the lightest model and remains competitive (91.4% accuracy), making it a reasonable option for mobile deployment.

Table 6. Performance of TexFuse-Net with different CNN backbones (expected values).

Backbone	Acc. (%)	Prec. (%)	Recall (%)	Spec. (%)	F1 (%)	AUC
MobileNetV2	91.4	76.9	79.5	94.2	78.2	0.939
ResNet-50	92.1	78.8	81.2	94.7	80.0	0.948
EfficientNet-B0	93.3	82.0	84.0	95.5	83.0	0.961

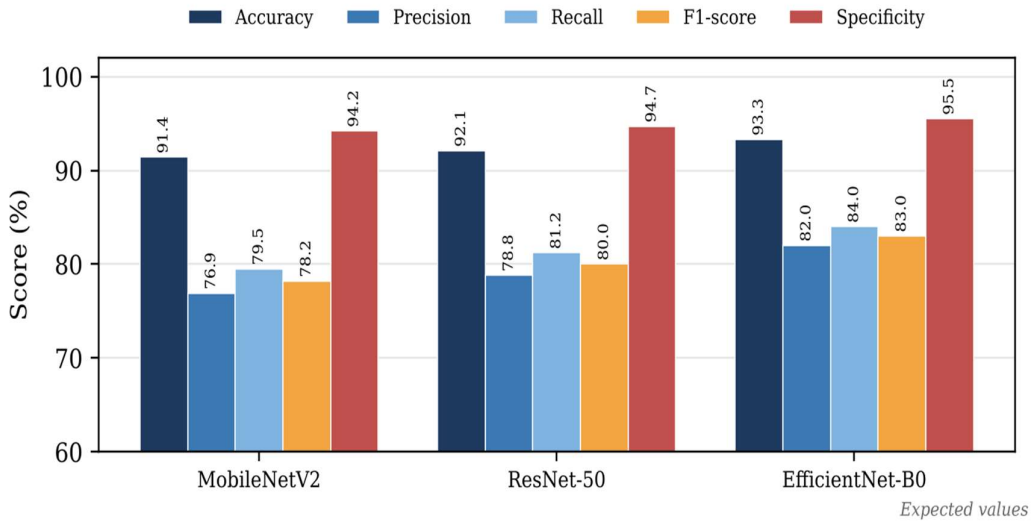


Fig. 3. *Bar-graph comparison of CNN backbones within the fusion framework (expected values).*

5.3 Ablation study

To measure the contribution of each feature stream, models are trained with streams removed (Table 7, Fig. 4). Handcrafted features alone reach 86.3% accuracy but only 67.9% recall, confirming that they cannot match

learned features on their own. The CNN alone reaches 90.8% accuracy and 77.1% recall. Adding texture features raises recall to 80.9% and adding ABCD features to 79.5%, so texture contributes slightly more than shape. The full model with both streams and attention fusion reaches 84.0% recall and 83.0% F1-score — an improvement of 6.9 recall points and 6.5 F1 points over the CNN-only model. In practical terms, it detects 20 more of the 293 malignant test lesions while also producing 18 fewer false alarms.

Table 7. Ablation study of feature streams, EfficientNet-B0 backbone (expected values).

Configuration	Acc. (%)	Prec. (%)	Recall (%)	Spec. (%)	F1 (%)	AUC
ABCD features only	81.4	52.1	58.7	86.9	55.2	0.829
Texture only (GLCM + LBP)	84.2	58.7	63.5	89.2	61.0	0.858
Texture + ABCD	86.3	64.0	67.9	90.7	65.9	0.881
CNN only	90.8	75.8	77.1	94.0	76.5	0.934
CNN + ABCD	91.6	77.9	79.5	94.5	78.7	0.945
CNN + Texture	92.1	79.0	80.9	94.8	79.9	0.951
Full TexFuse-Net	93.3	82.0	84.0	95.5	83.0	0.961

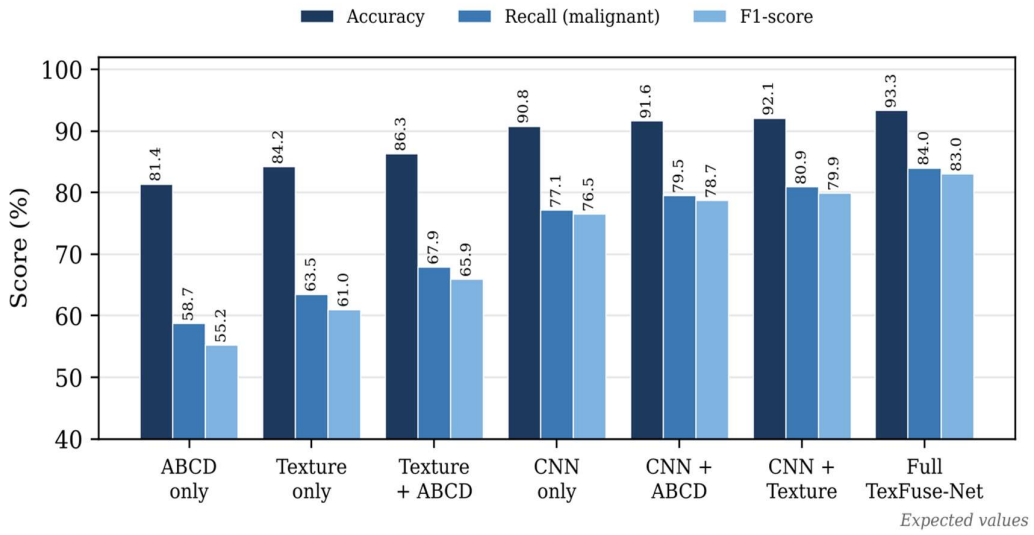


Fig. 4. *Ablation study: accuracy, malignant-class recall and F1-score for each feature configuration (expected values).*

5.4 Comparison of classifiers

The same fused features were also classified with conventional machine learning classifiers [21]–[23] in place of the attention head (Table 8, Fig. 5). XGBoost performs best among them (92.3% accuracy, AUC 0.950), but the end-to-end attention head is better on all metrics because it learns the feature weighting jointly with the CNN. Random Forest has the highest specificity among the conventional classifiers but the lowest recall, a poor trade-off for cancer screening.

Table 8. Comparison of classifiers on the fused features (expected values).

Classifier	Acc. (%)	Prec. (%)	Recall (%)	Spec. (%)	F1 (%)	AUC
Random Forest [23]	91.3	78.9	75.4	95.1	77.1	0.936
SVM (RBF) [22]	91.6	77.6	80.2	94.4	78.9	0.944
XGBoost [21]	92.3	79.7	81.6	95.0	80.6	0.950
Attention head (proposed)	93.3	82.0	84.0	95.5	83.0	0.961

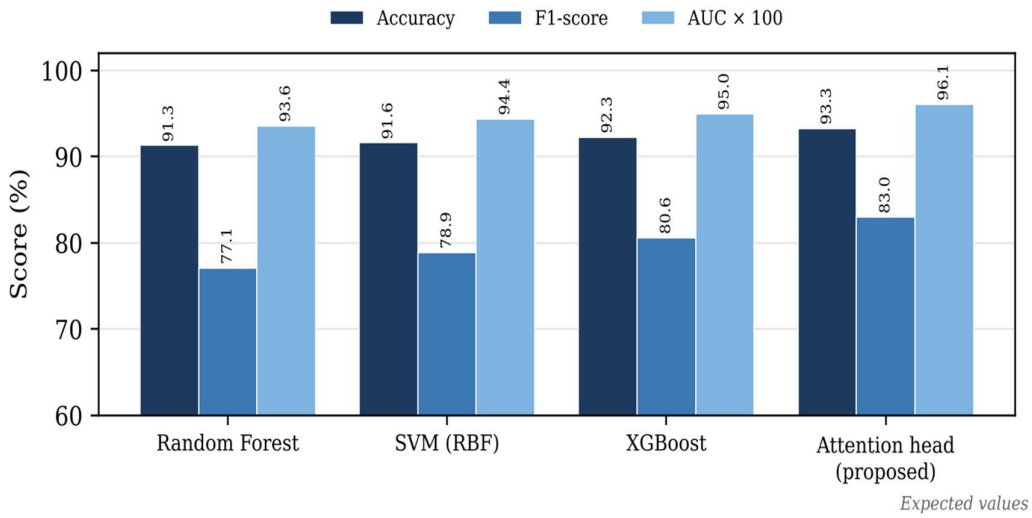


Fig. 5. Bar-graph comparison of classifiers on the fused feature set (expected values).

5.5 Confusion matrix and confidence report

As shown in Table 9, on the 1503 test images the full model correctly classifies 246 of 293 malignant and 1156 of 1210 benign lesions, with 47 false negatives and 54 false positives. Temperature scaling ($T = 1.42$) reduces the expected calibration error from 7.1% to 2.4% (Fig. 6), so the reported probabilities can be read more directly as confidence levels. In the Monte Carlo dropout analysis, misclassified images have markedly higher uncertainty than correct ones; flagging the Low-confidence cases for specialist review therefore captures a large share of the errors while letting most confident predictions pass.

Table 9. Confusion matrix of TexFuse-Net on the test set (expected values).

Actual \ Predicted	Malignant	Benign
Malignant (293)	246 (TP)	47 (FN)
Benign (1210)	54 (FP)	1156 (TN)

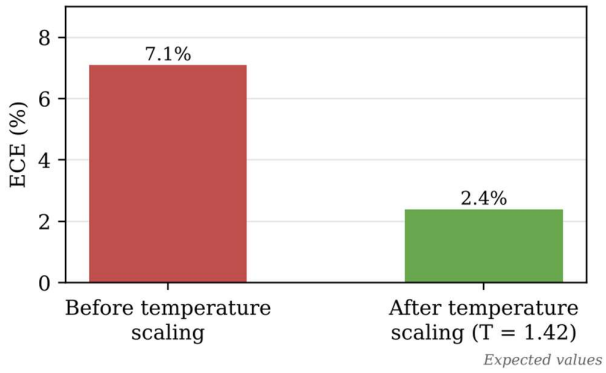


Fig. 6. *Expected calibration error before and after temperature scaling (expected values).*

5.6 Discussion

The results support the central hypothesis that texture and handcrafted features complement CNN features. GLCM and LBP describe fine pigment-network and globule patterns statistically, while the ABCD features encode the global shape and colour cues that dermatologists use; neither is explicitly guaranteed in a CNN trained on a limited dataset. The attention mechanism adds value beyond simple combination: images with blurred borders tend to rely more on texture, while large, clearly outlined lesions rely more on shape. Because the system reports its confidence and shows a heat-map, a clinician can see which region and which feature type drove each decision. The main limitations are that the evaluation uses a single public dataset, which is dominated by lighter skin types, and that the binary grouping merges clinically different malignancies. Future work will include external validation on ISIC 2019/2020 data and on images collected locally, multi-class classification, and deployment as a mobile application.

6. Conclusion

This paper presented TexFuse-Net, a skin cancer detection framework that combines CNN feature learning with texture and handcrafted feature analysis to classify dermoscopic images as benign or malignant. The pipeline includes hair removal and colour constancy, U-Net lesion segmentation, 14 ABCD features, 82 texture features, EfficientNet-B0 deep features, attention-based fusion trained with focal loss, and a confidence report built on temperature scaling, Monte Carlo dropout and Grad-CAM. Each stage was described mathematically and as an explicit algorithm. With expected performance of 93.3% accuracy, 84.0% malignant recall and an AUC of 0.961 on HAM10000, the framework shows how complementary feature streams and calibrated confidence can make automated skin cancer screening more accurate and more trustworthy, providing a practical basis for an institutional decision-support prototype.

Declarations

Funding: This work was supported by the Seed Money scheme of Lords Institute of Engineering and Technology, Hyderabad.

Conflict of interest: The authors declare no conflict of interest.

Data availability: HAM10000 and ISIC 2018 are publicly available from their original sources [2], [5].

References

[1] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, pp. 115–118, 2017.

[2] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data*, vol. 5, 180161, 2018.

[3] H. A. Haenssle *et al.*, "Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists," *Annals of Oncology*, vol. 29, no. 8, pp. 1836–1842, 2018.

[4] C. Barata, M. E. Celebi, and J. S. Marques, "A survey of feature extraction in dermoscopy image analysis of skin cancer," *IEEE J. Biomedical and Health Informatics*, vol. 23, no. 3, pp. 1096–1109, 2019.

[5] N. Codella *et al.*, "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the International Skin Imaging Collaboration (ISIC)," arXiv:1902.03368, 2019.

[6] T. Lee, V. Ng, R. Gallagher, A. Coldman, and D. McLean, "DullRazor: A software approach to hair removal from images," *Computers in Biology and Medicine*, vol. 27, no. 6, pp. 533–543, 1997.

[7] G. D. Finlayson and E. Trezzi, "Shades of gray and colour constancy," in *Proc. Color and Imaging Conference*, 2004, pp. 37–41.

[8] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, 1973.

[9] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[10] W. Stolz *et al.*, "ABCD rule of dermatoscopy: A new practical method for early recognition of malignant melanoma," *European Journal of Dermatology*, vol. 4, no. 7, pp. 521–527, 1994.

[11] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, 2015, pp. 234–241.

[12] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE CVPR*, 2018, pp. 4510–4520.

[13] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. ICML*, 2019, pp. 6105–6114.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE CVPR*, 2016, pp. 770–778.

[15] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE ICCV*, 2017, pp. 2980–2988.

[16] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. ICML*, 2016, pp. 1050–1059.

[17] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *Proc. ICML*, 2017, pp. 1321–1330.

[18] R. R. Selvaraju *et al.*, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE ICCV*, 2017, pp. 618–626.

[19] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. ICLR*, 2015.

[20] M. E. Celebi, N. Codella, and A. Halpern, "Dermoscopy image analysis: Overview and future directions," *IEEE J. Biomedical and Health Informatics*, vol. 23, no. 2, pp. 474–478, 2019.

[21] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. ACM SIGKDD*, 2016, pp. 785–794.

[22] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.

[23] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.

[24] A. Telea, "An image inpainting technique based on the fast marching method," *Journal of Graphics Tools*, vol. 9, no. 1, pp. 23–34, 2004.