Information Fusion 124 (2025) 103304

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Full Length Article

Fusion of metadata and dermoscopic images for melanoma detection: Deep learning and feature importance analysis

Misbah Ahmad [a,b], Imran Ahmed [c], Abdellah Chehri [d], Gwangill Jeon [e,*]

[a] Centre for Machine Vision, Bristol Robotics Laboratory, University of West of England, Frenchy Campus, Bristol, BS16 1QY, Bristol, UK
[b] Department of Animal and Agriculture, Hartpury University, Hartpury House, Gloucester, GL19 3BE, Gloucestershire, UK
[c] School of Computing and Information Science, Anglia Ruskin University, East Road, Cambridge, CB1 1PT, Cambridgeshire, UK
[d] Department of Mathematics and Computer Science, Royal Military College of Canada (RMC), P.O. Box 17000, Station Forces, Kingston, K7K 7B4, Ontario, Canada
[e] Department of Embedded Systems Engineering, Incheon National University, 119 Academy-ro, Incheon, 22012, Incheon, Republic of Korea

ARTICLE INFO

Keywords:
Melanoma detection
Multi-modal data fusion
Deep learning
Dermoscopic images
Clinical metadata
Smart healthcare

ABSTRACT

In the era of smart healthcare, integrating multimodal data is essential for improving diagnostic accuracy and enabling personalized care. This study presented a deep learning-based multimodal approach for melanoma detection, leveraging both dermoscopic images and clinical metadata to enhance classification performance. The proposed model integrated a multi-layer convolutional neural network (CNN) to extract image features and combined them with structured metadata, including patient age, gender, and lesion location, through feature-level fusion. The fusion process occurred at the final CNN layer, where high-dimensional image feature vectors were concatenated with processed metadata. The metadata was handled separately through a fully connected neural network comprising multiple dense layers. The final fused representation was passed through additional dense layers, culminating in a classification layer that outputted the probability of melanoma presence. The model was trained end-to-end using the SIIM-ISIC dataset, allowing it to learn a joint representation of image and metadata features for optimal classification. Various data augmentation techniques were applied to dermoscopic images to mitigate class imbalance and improve model robustness. Additionally, exploratory data analysis (EDA) and feature importance analysis were conducted to assess the contribution of each metadata feature to the overall classification. Our fusion-based deep learning architecture outperformed single-modality models, boosting classification accuracy. The presented model achieved an accuracy of 94.5% and an overall F1-score of 0.94, validating its effectiveness in melanoma detection. This study aims to highlight the potential of deep learning-based multimodal fusion in enhancing diagnostic precision, offering a scalable and reliable solution for improved melanoma detection in smart healthcare systems.

1. Introduction

In recent years, integrating multimodal data fusion techniques has emerged as a transformative approach in healthcare [1]. Advanced technologies and artificial intelligence-based methods have enhanced diagnostic accuracy and enabled personalized treatment strategies [2,3] for various healthcare applications. In dermatology, accurate detection of melanoma – a type of skin cancer – remains one of the most significant challenges due to the variability in skin lesion appearance and patient demographics. As shown in Fig. 1, melanoma incidence rates vary worldwide, impacting global health statistics. Skin disease detection is a critical concern in public health, as conditions like melanoma have a significant impact on healthcare systems and patient outcomes. Early detection of melanoma and other skin diseases is essential as it improves treatment outcomes, reduces mortality rates, and minimizes healthcare costs. Advanced diagnostic tools, particularly those leveraging artificial intelligence (AI) and multi-sensor data fusion, enhance the accuracy and reliability of detection by integrating clinical and imaging data. This approach plays a key role in addressing the increasing burden of skin diseases, particularly in high-incidence regions. Melanoma detection models typically rely on dermoscopic images, which provide detailed visual representations of skin lesions. However, incorporating additional patient information, such as clinical metadata (age, gender, and lesion location), can further enhance the diagnostic capability of such models.

Existing approaches to melanoma detection have primarily focused on analyzing dermoscopic images using machine learning and

* Corresponding author.
E-mail address: gjeon@inu.ac.kr (G. Jeon).

https://doi.org/10.1016/j.inffus.2025.103304
Received 15 December 2024; Received in revised form 5 March 2025; Accepted 5 May 2025
Available online 6 June 2025
1566-2535/© 2025 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

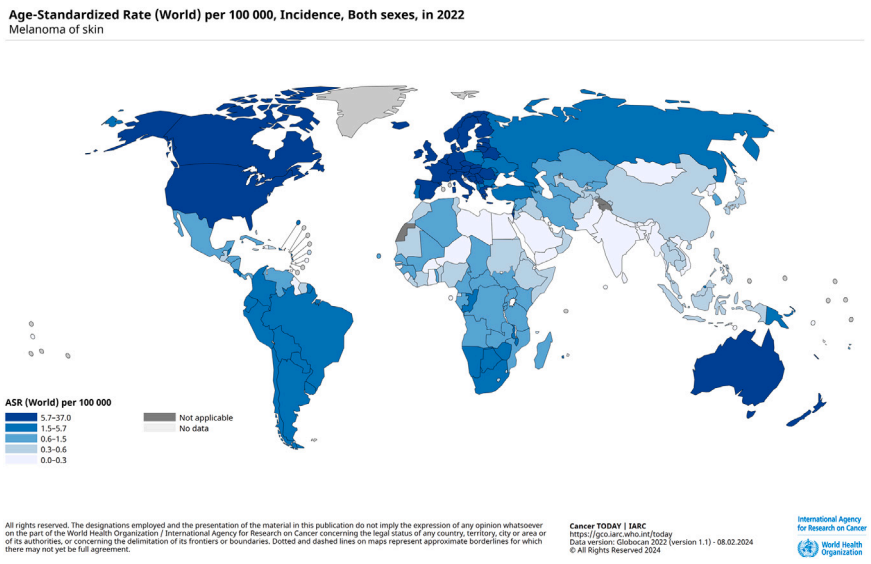


Fig. 1. Data are from GLOBOCAN 2022, Incidences are expressed as the number of cases per 100000 person-years.[1]

deep learning models, particularly Convolutional Neural Networks (CNNs) [4,5]. CNNs have demonstrated high effectiveness in image classification tasks, including medical image analysis. However, despite their success, single-modality models often fail to fully capture the complexities of the disease, particularly when critical factors such as patient demographics or lesion-specific metadata are excluded from the model's decision-making process [6,7]. This diagnostic gap underscores the need for multi-sensor fusion methods that can combine the strengths of both image data and structured clinical information [8]. Multi-sensor data fusion in healthcare integrates diverse data sources to provide a more comprehensive understanding of a patient's condition [9]. By combining visual data (such as dermoscopic images) with clinical metadata, diagnostic models can achieve higher accuracy and reliability [10]. Given the potential of this approach, researchers [11] have extensively explored the fusion of structured metadata and image data in skin disease classification.

Building on this foundation, this study aims to develop a deep learning-based multi-sensor fusion model for melanoma detection using the publicly available SIIM-ISIC Melanoma Classification dataset. This study investigates the fusion of image features extracted via a CNN with patient metadata through feature-level fusion, creating a joint representation that enhances classification performance. Previous research on melanoma detection has highlighted the effectiveness of CNNs for image-based classification but has often overlooked the potential benefits of incorporating patient-specific data. While some attempts have been made to integrate clinical metadata, existing studies are often limited in scope or effectiveness, leaving room for improvement. This raises an important question: Can the fusion of metadata and image data significantly improve melanoma detection accuracy over image-only models?

To answer this, we develop a deep learning architecture that integrates both image and metadata using the SIIM-ISIC 2020 Melanoma Classification Dataset [12]. The model processes dermoscopic images through a multi-layer CNN to extract high-level image features, while clinical metadata is analyzed through a fully connected neural network. These features are then fused at the feature level and passed through additional layers to predict melanoma probability. Data augmentation techniques are applied to address class imbalance, ensuring balanced training. Additionally, exploratory data analysis (EDA) and feature importance analysis are conducted to assess the impact of individual metadata features on classification. Through this comprehensive approach, we aim to demonstrate that multi-sensor data fusion not only enhances classification performance but also provides valuable insights into the role of clinical metadata in diagnostic processes. The primary contributions of this paper are as follows:

- To develop a deep learning-based multi-sensor fusion framework that integrates dermoscopic images and clinical metadata for enhanced melanoma detection.
- To evaluate the diagnostic performance of the fusion model in comparison to single-modality (image-only) models, highlighting improvements in classification accuracy.
- To enhance model training and generalization through data augmentation techniques.
- To conduct exploratory data analysis (EDA) and feature importance analysis to assess the impact of individual clinical metadata features on melanoma detection.
- To demonstrate the advantages of the multi-sensor data fusion framework in outperforming traditional image-only models, particularly in cases with limited training data or different melanoma characteristics.

We aim to demonstrate that the proposed multi-sensor fusion approach enhances melanoma detection by improving classification accuracy, sensitivity, and specificity compared to single-modality models. This work highlights the benefits of integrating multiple data sources to bridge the gap between deep learning and clinical decision-making. To validate its effectiveness, we evaluate the framework on a benchmark melanoma dataset, compare it against traditional models, and analyze the impact of clinical metadata features. By providing a scalable and reliable solution, this study contributes to the broader adoption of multi-sensor data fusion in healthcare. The rest of the paper is organized as follows: Section 2 provides details about related work, Section 3 details the methodology, dataset and proposed deep learning model, and Section 4 presents experimental results. Sections 5 and 6 provide discussion, including the conclusion, study limitations, and future research directions.

2. Related work

In recent years, researchers have extensively explored deep learning and advanced digital technologies for cancer detection and classi-

[1] Source: International Agency for Research on Cancer (IARC). Available at https://gco.iarc.fr/today/en/dataviz/maps-heatmap?mode=population.

fication. Melanoma skin cancer detection has gained significant attention, with various approaches presented to improve classification performance across different datasets.

Early skin lesion analysis relied on conventional image processing techniques, particularly for segmentation, which is important in identifying regions of interest in dermoscopic images. A technique introduced in [13] used machine learning principles to enhance segmentation accuracy and classification performance. Similarly, [14] employed Harris corner detection and region-growing methods for lesion segmentation. Another method in [15] utilized global and local feature extraction with blind deconvolution and color-space transformations to improve segmentation quality. With advancements in machine learning, more sophisticated models emerged, offering automated feature extraction and improved classification accuracy. A framework integrating feature engineering with deep learning concepts was applied in [16], bridging the gap between traditional image processing and modern deep learning. A two-step approach utilizing transfer learning and ensemble methods significantly enhanced binary melanoma classification, as demonstrated in [17] using the SIIM-ISIC 2020 dataset.

The adoption of CNNs marked a significant shift in melanoma classification, enabling automatic feature extraction directly from dermoscopic images. The DenseNet architecture was applied in [18] to identify melanoma-prone regions, outperforming conventional approaches. Further advancements were made in [19], where an ensemble of CNNs achieved an AUC of 0.9411, showcasing the effectiveness of combining multiple models. Another study [20] highlighted the importance of extensive data augmentation techniques in improving model robustness and generalization. To further enhance classification performance, researchers explored ensemble learning and multi-scale approaches. An EfficientNet ensemble model used in [21] demonstrated that combining multiple CNN backbones improves accuracy. A multi-scale deep ensemble learning approach introduced in [22] effectively handled images of varying sizes and captured lesion features across multiple levels. Automated ensemble learning strategies were also explored in [23], which considerably improved malignant melanoma detection. Hybrid approaches, which combine traditional machine learning with deep learning, have also been investigated for melanoma classification. A segmentation and classification model incorporating deep learning with a Markovian approach was presented in [24]. Another study in [25] focused on ensemble learning techniques, outperforming state-of-the-art models on the SIIM-ISIC dataset.

Researchers have also explored modifications to CNN architectures and feature extraction techniques to improve melanoma classification. The impact of varying CNN backbones and input sizes on model sensitivity was demonstrated in [26]. The DenseNet architecture was employed in [27] for skin lesion classification, proving its robustness in distinguishing malignant and benign lesions. NASNet deep features were leveraged in [28], improving feature extraction quality. A modified EfficientNet model proposed in [29] significantly enhanced classification performance, while [30] introduced new baselines for feature representation learning in melanoma classification. The evolution of ensemble learning has led to model blending techniques for improving classification accuracy. A study in [31] employed an ensemble of CNNs with loss balancing and data augmentation to enhance performance. To address class imbalance, [32] integrated knowledge distillation into ensemble models, ensuring better representation of minority classes such as melanoma.

The rise of generative AI has facilitated synthetic data augmentation for melanoma classification. Generative models were used in [33] to create synthetic dermoscopic images, significantly enhancing classifier robustness, particularly in scenarios with limited annotated datasets. Several studies have incorporated clinical metadata into deep learning models to improve classification accuracy. A patient-centric dataset combining dermoscopic images and clinical metadata was introduced in [34], aiding in AI-driven melanoma detection. Another study in [35] explored melanin and erythema indices in CNN models to improve melanoma classification. Multi-stage recognition frameworks using deep residual networks and hyper-parameter optimization were proposed in [36], demonstrating improved decision support for melanoma diagnosis.

Despite significant advancements in melanoma detection using deep learning, several challenges still exists for example, one major limitation is the reliance on single-modality approaches, where most studies focus solely on dermoscopic images, neglecting the potential advantages of integrating clinical metadata for a more comprehensive diagnosis. Additionally, class imbalance remains a critical issue, as many works primarily employ data augmentation rather than advanced resampling techniques like weight loss, which can more effectively address imbalance-related biases in model training. Another challenge lies in generalization, as many models demonstrate high accuracy on specific datasets but struggle when applied to unseen data, underscoring the need for more robust learning strategies. Furthermore, deep learning models, particularly CNNs, often function as black boxes, lacking interpretability, which raises concerns about clinical trust and decision-making transparency. To overcome these challenges, this work proposes a multi-sensor fusion framework that integrates dermoscopic images with clinical metadata while leveraging deep learning and advanced resampling techniques to enhance classification accuracy, robustness, and generalization.

3. Methodology

This section outlines the methodology presented in this work. The overall methodology framework adopted for developing and evaluating the presented deep learning based multi-sensor fusion model for melanoma detection is shown in Fig. 2. The methodology includes dataset acquisition and preprocessing, exploratory data analysis (EDA), model architecture design, training process, and feature importance analysis. This section outlines the methodology presented in this work. The overall methodology framework adopted for developing and evaluating the presented deep learning based multi-sensor fusion model for melanoma detection is shown in Fig. 2. The methodology includes dataset acquisition and preprocessing, exploratory data analysis (EDA), model architecture design, training process, and feature importance analysis.

3.1. Dataset and preprocessing

This work used the publicly available SIIM-ISIC Melanoma Classification dataset [12] https://challenge2020.isic-archive.com/. The dataset is widely recognized as a benchmark resource for melanoma detection research and provides a robust foundation for developing machine learning and deep learning models. It contains high-resolution dermoscopic images, typically around 1024 × 1024 pixels, ensuring detailed lesion visualization. The dataset enables the exploration of multi-modal approaches by incorporating both image and clinical metadata for accurate and reliable skin lesion classification. The dataset comprises 33,126 high-quality dermoscopic images of unique benign and malignant skin lesions collected from over 2,000 patients across six international institutions. Each image is associated with comprehensive clinical metadata, including patient demographics and lesion characteristics. Additionally, a separate test set of 10,982 images is provided specifically for model evaluation. The images are available in DICOM format with embedded metadata and as JPEG files for easier accessibility. The metadata includes essential clinical details such as patient ID, lesion ID, sex, age, and anatomical lesion location. The diagnostic ground truth of the dataset is established through a rigorous validation process: histopathology for malignant lesions, while benign lesions were confirmed through expert consensus, longitudinal follow-up, or histopathology. While the dataset is highly representative and diverse, it is not exhaustive like a census; instead, it serves as a large-scale, well-curated dataset reflecting real-world melanoma cases. A detailed description of the dataset is provided in Table 1.

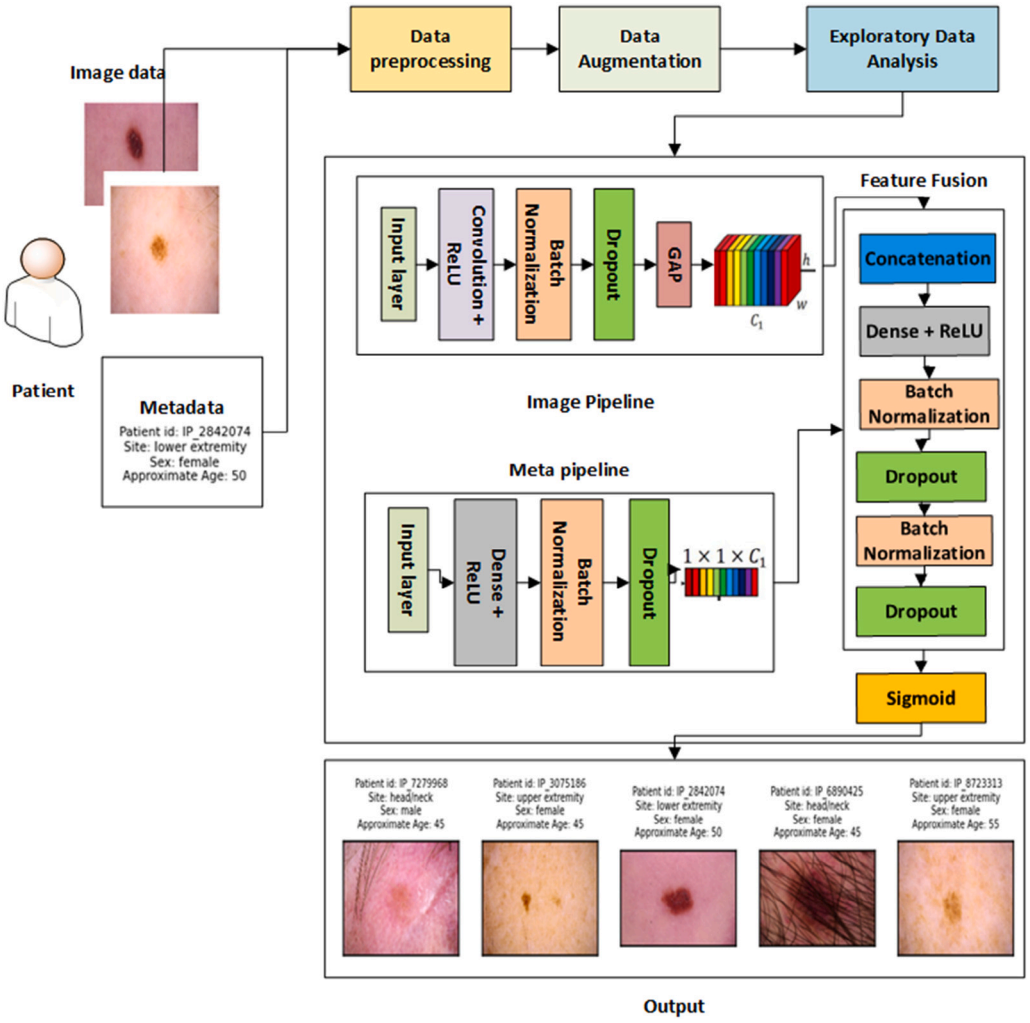


Fig. 2. Overall framework of the presented methodology. The model integrates dermoscopic images and clinical metadata, processed through image and meta pipelines. Extracted features are fused using a feature fusion module, enhancing melanoma detection accuracy through multi-sensor data fusion.

Table 1
Description of the SIIM-ISIC melanoma classification dataset.

Attribute	Description
Number of Images	33,126 high-quality dermoscopic images
Test Set	10,982 images for model evaluation
Image Formats	DICOM format with embedded metadata, JPEG files
Patient Demographics	Patient ID, Lesion ID, Sex, Age
Lesion Characteristics	General anatomical site
Diagnostic Validation	Histopathology (malignant lesions), expert consensus, longitudinal follow-up, or histopathology (benign lesions)

3.2. Exploratory data analysis (EDA)

As discussed in the objectives of this work, we provided the EDA in order to understand insights into the metadata. The following figures show some basic EDA analysis of the data set. The distribution of gender in the dataset is first to analyzed. In Fig. 3, the bar chart shows the distribution of genders (male, female, and unknown) in the training dataset. It can be seen that the distribution is approximately balanced, with a slight increase in the number of male samples. This analysis is very helpful in metadata and image fusion, it ensures that demographic attributes, such as gender, are properly incorporated into the model, allowing it to learn meaningful correlations between patient characteristics and melanoma risk. By integrating gender information, the model can better understand potential variations in disease presentation across different groups, improving classification accuracy and fairness. It is important to know that the combined information from both sources is consistent, accurate, and meaningful, as metadata provides contextual details. In this case, gender enhances the interpretation of image features. This analysis allows for identifying biases or missing data, providing the fusion process leads to more robust and reliable results for diagnosis or pattern recognition applications. The second main analysis is to understand the distribution of benign and malignant cases by gender, as shown in Fig. 4. It can be seen that males have a slightly higher number of cases, with both benign and malignant cases being predominantly represented.

Fig. 5 illustrated the percentage distribution of the anatomical sites where images were taken. This analysis highlights the anatomical site

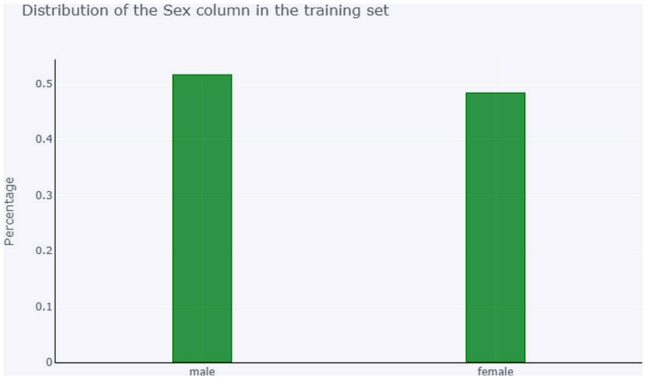


Fig. 3. Distribution of the gender (Sex) column in the training set.

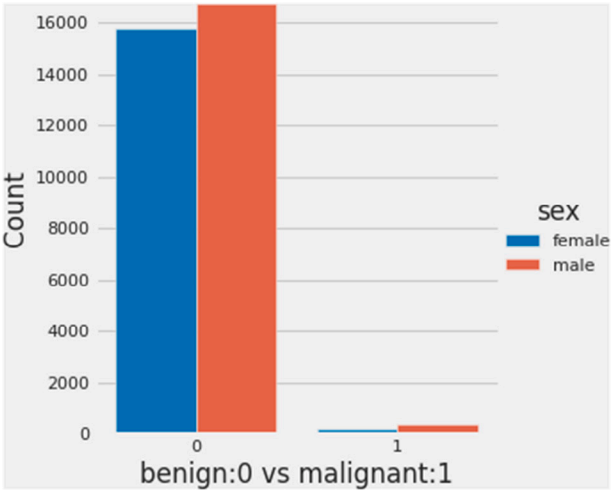


Fig. 4. Distribution of benign (0) vs malignant (1) cases by gender.

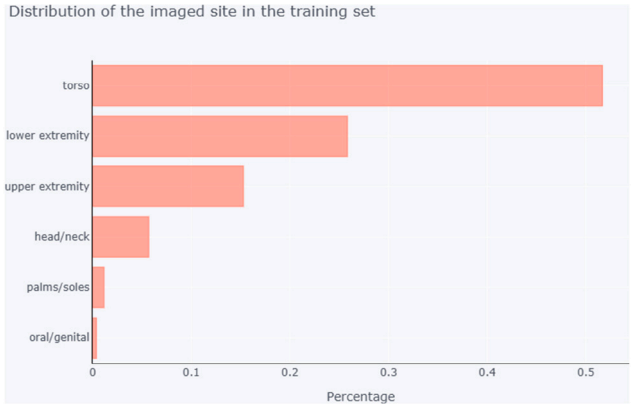


Fig. 5. Distribution of the imaged site in the training set.

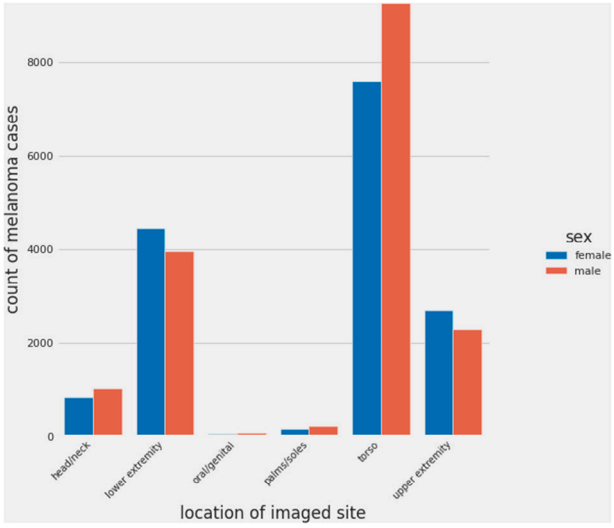


Fig. 6. Count of melanoma cases by the location of the imaged site and gender.

distribution, indicating that the torso is the most frequently imaged area, which could suggest a higher prevalence of conditions or lesions in that region. Understanding this distribution is important for identifying anatomical patterns of occurrence, which can guide clinical focus and resource allocation for diagnosis or treatment. We also explored the relationship between the imaged site, malignancy, and gender as depicted in Fig. 6. This analysis is important because it reveals how melanoma occurrence changes across anatomical sites and between genders. It indicates potential gender-specific risk factors or exposure patterns (e.g., lifestyle or biological differences). It can be seen that the torso is the most common site for melanoma, with notable differences in male and female representation. Such insights are valuable for tailoring prevention strategies, improving diagnostic accuracy, and enhancing understanding of gender-specific disease manifestations.

The age distribution in the dataset is shown in Fig. 7. This information provides insights into the age groups most affected by the condition, allowing to determine the population at higher risk. Understanding the age distribution is important for designing targeted screening programs, developing age-specific prevention strategies, and tailoring medical interventions to the requirements of the most impacted demographic. The distribution in Figure follows a near-normal pattern, with most cases focused around 40-60 years of age.

To compare the age distribution between benign and malignant cases, we analyzed as shown in Fig. 8. It can be seen that malignant cases are more general in older age groups, indicating that age might be a significant risk factor for malignancy. This information is

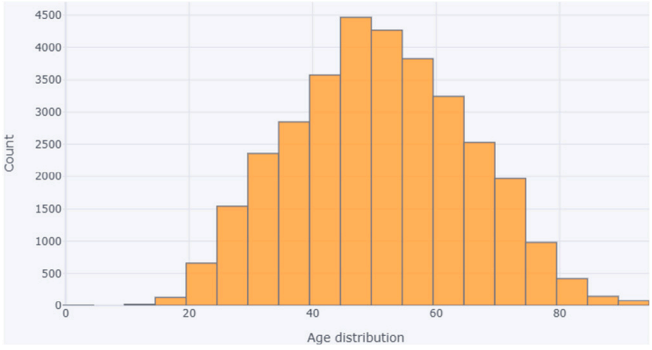


Fig. 7. Age distribution in the dataset.

important, as previously stated, for developing age-specific diagnostic protocols, early detection strategies, and understanding the condition's progression with age.

In Fig. 9, we aim to understand the age distribution trends, which are generally consistent across genders; there are slight differences in the representation of certain age groups. These variations could point to gender-specific factors influencing the condition's onset or progression, which can inform more personalized approaches to diagnosis, prevention, and treatment.

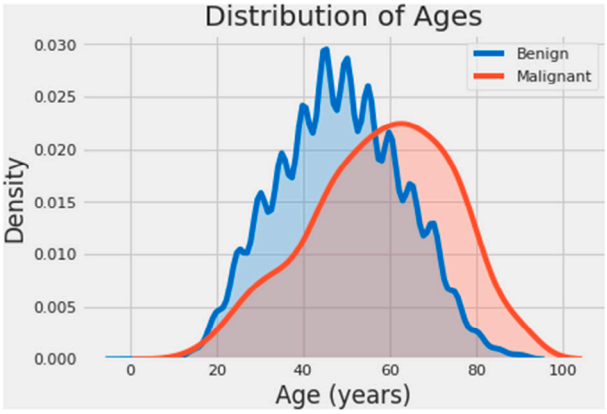


Fig. 8. Distribution of ages for benign and malignant cases.

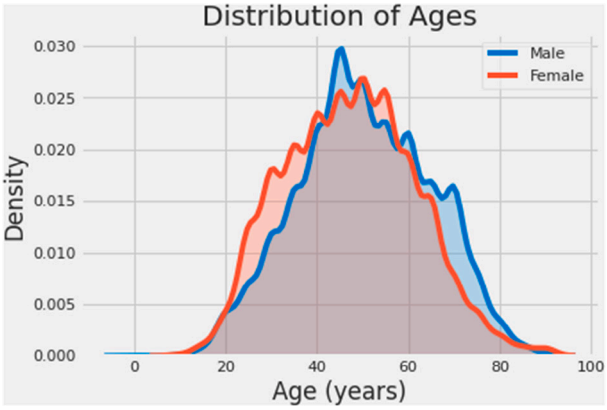


Fig. 9. Age distribution by gender.

Figs. 10 and 11 analyze the intensity values of the red, green, and blue channels in benign and malignant images. This analysis demonstrates that malignant images have distinct color intensity patterns compared to benign ones, likely reflecting differences in pigmentation or structural characteristics of the lesions. These findings are important for developing image-based diagnostic tools as they highlight features that can be used to distinguish between benign and malignant conditions effectively.

3.3. Data augmentation techniques

Data augmentation is crucial for training deep learning models in melanoma detection, addressing the following challenges:

- Enhancing the diversity of the training data to reduce overfitting.
- Simulating real-world scenarios by introducing variations in lighting, orientation, and noise.
- Complementing class imbalance handling techniques such as weighted loss, ensuring a well-balanced training distribution.
- Simulating real-world scenarios by introducing variations in lighting, orientation, and noise.

The images presented in Fig. 12 illustrate various image transformations applied to dermoscopic images for preprocessing and enhancement. (a) B&W Transformation converts images into gray scale, removing color information while preserving texture and contrast, aiding in feature extraction. (b) Ben Graham: grays cale without Gaussian Blur applies a color removal technique but retains the high-frequency noise, which may affect analysis. (c) Ben Graham: greyscale with Gaussian Blur smooths the images, reducing noise and enhancing the visibility of key patterns in dermoscopic features. (d) HSB Adjustments modify Hue, Saturation, and Brightness to simulate variations in lighting conditions and skin tones, improving model robustness against real-world variations. Lastly, (e) LUV color space transformations shift the image representation into a perceptually uniform color space, optimizing color contrast while maintaining human visual perception fidelity. These transformations contribute to improving image-based analysis, particularly in medical imaging tasks such as skin lesion classification.

Beyond these transformations, advanced augmentation techniques are implemented to further diversify the dataset. Center cropping ensures that lesions remain the focal point, while color jittering introduces variations in brightness, contrast, and saturation to account for different imaging conditions. Random gray scale conversion simulates variability in image color properties, and random vertical flipping enhances spatial invariance by exposing the model to different orientations of lesions. These augmentation strategies, illustrated in Fig. 13, strengthen the model's ability to generalize across diverse real-world scenarios.

This section comprehensively examines the dataset's key characteristics, including demographic distribution, malignancy trends, anatomical site distribution, and age-related insights. The analysis begins with the gender distribution, highlighting a nearly balanced dataset with a slight male predominance, ensuring demographic attributes are properly incorporated into the model. Further, the malignancy distribution

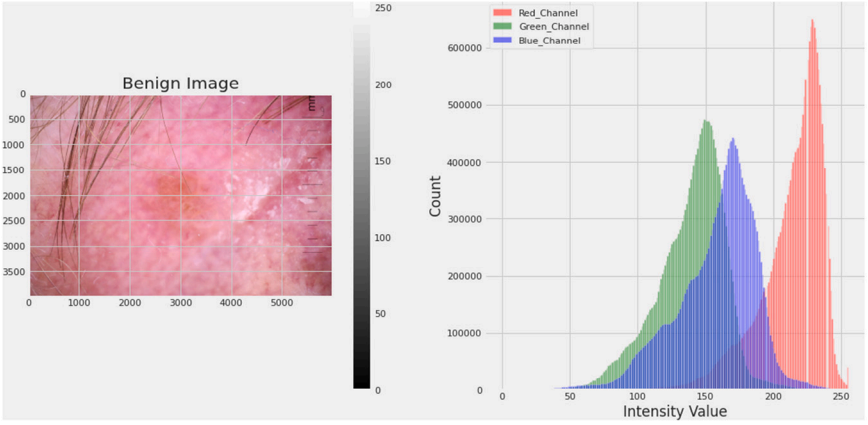


Fig. 10. Color intensity distribution for benign images.

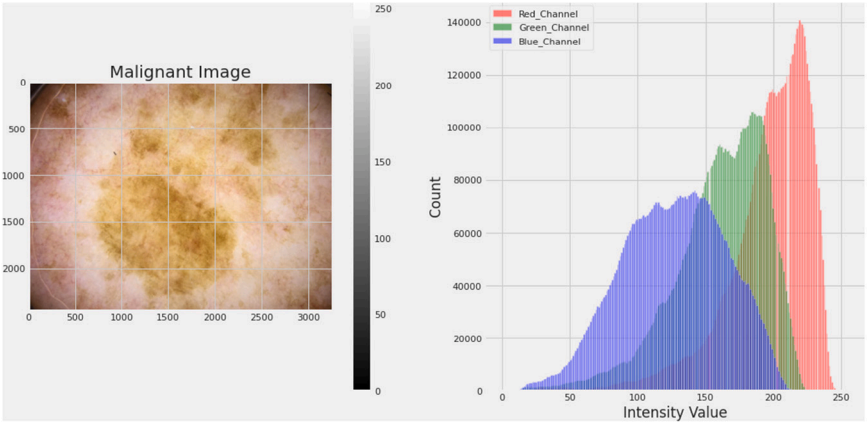
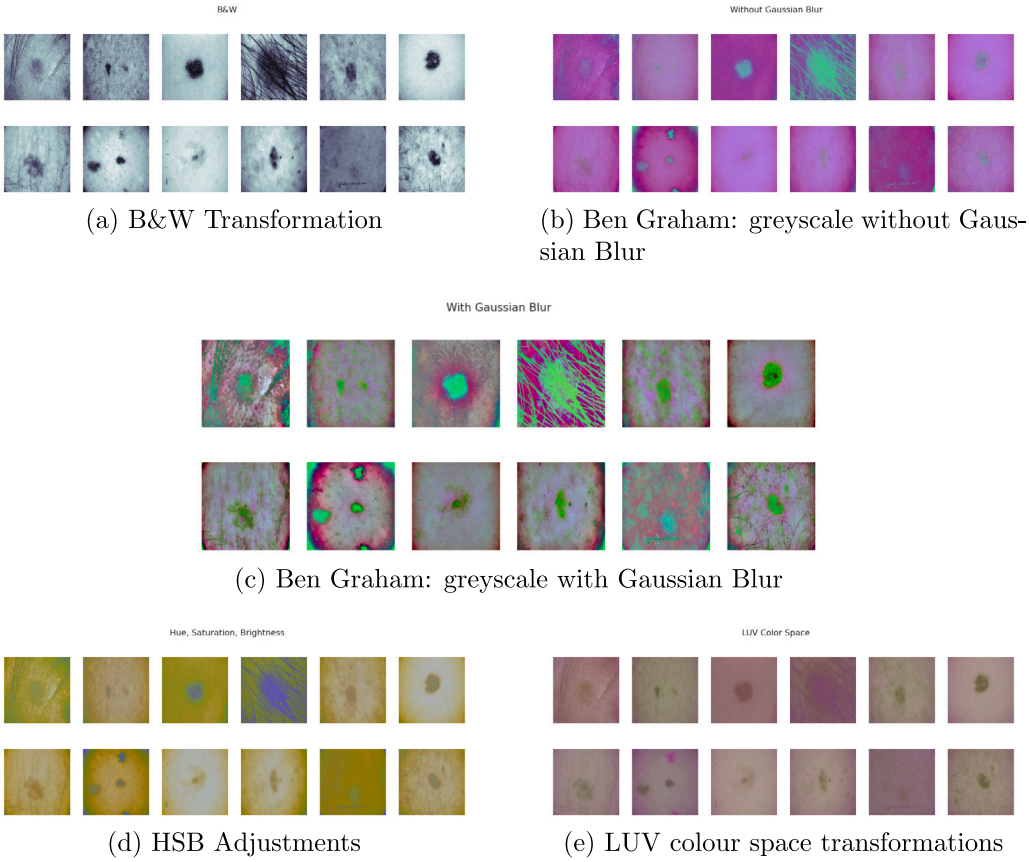


Fig. 11. Color intensity distribution for malignant images.



(a) B&W Transformation

(b) Ben Graham: greyscale without Gaussian Blur

(c) Ben Graham: greyscale with Gaussian Blur

(d) HSB Adjustments

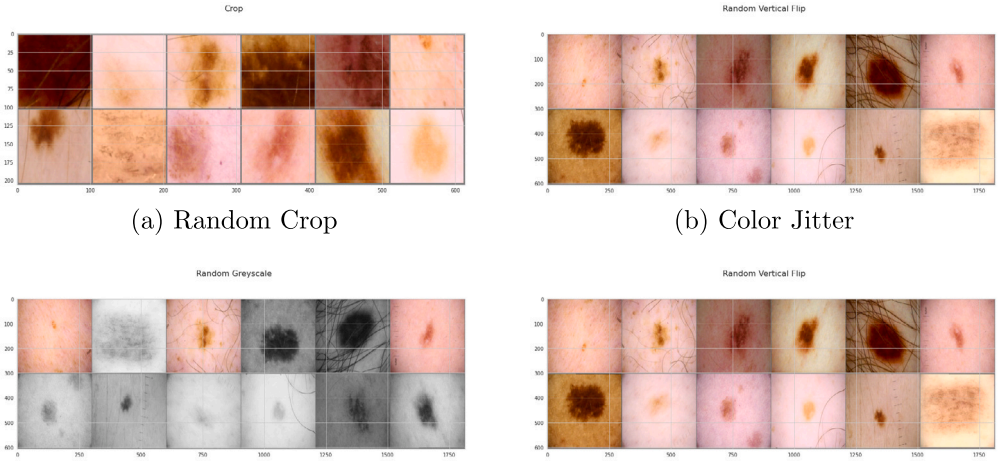
(e) LUV colour space transformations

Fig. 12. Examples of image transformations used in this study.

by gender indicates a higher prevalence of cases among males, reinforcing the importance of demographic-based risk assessment. The imaged site distribution reveals that the torso is the most frequently examined region, emphasizing potential anatomical risk patterns. Additionally, the age distribution follows a near-normal pattern, with most cases concentrated between 40 and 60 years, underlining age as a significant risk factor. Further, comparing age distribution across benign and malignant cases shows that malignant cases are more common in older age groups. This highlights the importance of age-specific screening strategies. Finally, the color intensity analysis of benign and malignant images indicates distinct pigmentation patterns, reinforcing the significance of image-based diagnostic features. The data augmentation techniques showcased, including gray scale transformations, HSB adjustments, and advanced transformations like random cropping, flipping, and color jittering, enhance model generalization by simulating real-world variations. These insights collectively strengthen the model's ability to recognize meaningful patterns and improve classification accuracy for melanoma detection.

3.4. Proposed model architecture

The presented deep learning model used to classify skin diseases consists of two parallel pipelines: one for processing image data and another for processing metadata. These pipelines are fused at the feature level, as shown in Fig. 2, to form a unified representation. The model is implemented and trained on a computer system equipped with an Intel Core i7 processor and an NVIDIA GeForce RTX GPU



(a) Random Crop (b) Color Jitter

(c) Random Gray Scale (d) Vertical Flip

Fig. 13. Examples of augmentations applied (a) Random crop, (b) Color jitter, (c) gray scale, (d) Horizontal flip, . These transformations showcase common preprocessing techniques used for deep learning model training.

(Nitro 5 series), ensuring efficient computation and accelerated deep learning processing. The implementation is done in Python, utilizing libraries such as TensorFlow, PyTorch, OpenCV, and scikit-learn for model development, image processing, and evaluation. The detailed architecture is described as follows:

3.4.1. Image processing pipeline

The image data is processed through Convolutional Neural Network (CNN) layers, which are designed to extract high-level features. The pipeline includes the following components:

1. Input layer. The input image is represented as:

$$\mathbf{X} \in \mathbb{R}^{H \times W \times C} \tag{1}$$

In Eq. (1):

- H and W are the height and width of the image.
- C is the number of channels (for RGB images, $C = 3$).

2. Convolutional layers. After the input layer, we have a series of convolutional layers, which are used to extract the hierarchical features of the image. These layers perform convolution operation, which is defined in Eq. (2) as follows:

$$\mathbf{Y}_k(i,j) = \sigma\left(\sum_{c=1}^{C}\sum_{m=1}^{M}\sum_{n=1}^{N}\mathbf{W}_k(c,m,n)\cdot\mathbf{X}(i+m,j+n,c)+b_k\right) \tag{2}$$

This Equation describes how convolutional layers transform input images into feature maps by applying learnable filters (kernels). Each parameter plays a crucial role in this transformation and is defined as:

- $\mathbf{X}(i+m, j+n, c)$ represents the input feature map at spatial position $(i+m, j+n)$ and channel c.
- $\mathbf{W}_k(c,m,n)$ denotes the convolutional kernel (filter) for the k-th output channel, with dimensions $M \times N$. Each filter slides over the input feature map, extracting spatial patterns such as edges, textures, and high-level features.
- C represents the number of input channels, where $C = 3$ for RGB images and $C = 1$ for gray scale images.
- $M \times N$ defines the size of the convolutional filter. Common values include 3×3, 5×5, or 7×7, with smaller filters capturing fine-grained features and larger filters capturing broader spatial information.
- b_k is the bias term associated with the k-th output channel, ensuring the activation function can shift the learned features.
- $\sigma(x)$ is the activation function, typically ReLU, given in Eq. (3):

$$\sigma(x) = \max(0, x) \tag{3}$$

  ReLU introduces non-linearity, preventing vanishing gradient issues and accelerating convergence.
- $\mathbf{Y}_k(i,j)$ is the output feature map corresponding to the k-th filter, obtained after applying convolution and activation.

This convolution operation allows the network to learn spatially invariant features by detecting meaningful patterns at different levels. As the network deepens, successive convolutional layers capture increasingly complex representations, enabling effective classification of skin lesions.

3. Batch normalization. As shown in Fig. 2, Batch normalization is used to stabilize the training by normalizing the activations after each convolution. This is given as;

$$\hat{x}_i = \frac{x_i - \mu}{\sqrt{\sigma^2 + \epsilon}}, \quad y_i = \gamma \hat{x}_i + \beta \tag{4}$$

In Eq. (4) μ and σ^2 are the batch mean and variance, ϵ is a small constant for numerical stability, and γ, β are learnable scaling parameters. These parameters control the scaling and shifting of the normalized activations. They allow the model to preserve representation flexibility, improving training stability and convergence speed.

4. Dropout. Dropout is used to reduce overfitting by randomly setting a fraction p of the activations to zero during training and is calculated in Eq. (5);

$$y_i = \begin{cases} 0 & \text{with probability } p \\ x_i/(1-p) & \text{with probability } 1-p \end{cases} \tag{5}$$

5. Global average pooling (GAP). To reduce the spatial dimensions of the feature maps, we used GAP by computing the average across each channel. It is determined using the following;

$$y_k = \frac{1}{H' \cdot W'}\sum_{i=1}^{H'}\sum_{j=1}^{W'}\mathbf{Y}_k(i,j) \tag{6}$$

In Eq. (6),

- H' and W' represent the height and width of the feature map produced by the last convolutional layer before GAP.
- $\mathbf{Y}_k(i,j)$ denotes the activation value at position (i,j) in the k-th feature map.
- y_k is the final output for the k-th feature map after applying GAP.

Table 2
Presented Fused CNN architecture combining image and metadata pipelines.

Component	Layer type	Output shape	Details
Image Pipeline	Input	$H \times W \times C$	Input image with dimensions H, W, C
	Convolution + ReLU	$H' \times W' \times C_1$	Kernel size: 3×3, Stride: 1
	Batch Normalization	$H' \times W' \times C_1$	Normalize activations
	Dropout	$H' \times W' \times C_1$	Dropout rate: p_1
	Global Average Pooling (GAP)	C_1	Reduce spatial dimensions
Metadata Pipeline	Input	d	Metadata feature vector size d
	Dense + ReLU	d_1	Number of neurons: d_1
	Batch Normalization	d_1	Normalize activations
	Dropout	d_1	Dropout rate: p_2
Feature Fusion	Concatenation	$C_1 + d_1$	Combine image and metadata features
	Dense + ReLU	d_2	Number of neurons: d_2
	Batch Normalization	d_2	Normalize activations
	Dropout	d_2	Dropout rate: p_3
	Dense + ReLU	d_3	Number of neurons: d_3
	Batch Normalization	d_3	Normalize activations
	Dropout	d_3	Dropout rate: p_4
Output Layer	Sigmoid	1	Probability of melanoma

The results in the above Equation are in a fixed-length feature vector with a size equal to the number of channels.

3.4.2. Metadata processing pipeline

After image processing, the second pipeline is for processing the metadata. This pipeline processes clinical information using a fully connected neural network. The input is a vector $\mathbf{Z} \in \mathbb{R}^d$, where d is the number of metadata features.

1. Input layer. The metadata is passed through a dense layer. It is calculated as;

$$\mathbf{h}^{(1)} = \sigma(\mathbf{W}^{(1)}\mathbf{Z} + \mathbf{b}^{(1)}) \tag{7}$$

In Eq. (7)

- $\mathbf{W}^{(1)}$ and $\mathbf{b}^{(1)}$ are the weight matrix and bias vector.
- σ is the activation function (ReLU).

2. Hidden layers. Additional dense layers are used to refine the metadata representation:

$$\mathbf{h}^{(l+1)} = \sigma(\mathbf{W}^{(l)}\mathbf{h}^{(l)} + \mathbf{b}^{(l)}) \tag{8}$$

3. Batch normalization and dropout. This pipeline also include batch normalization as above. It normalizes the outputs of each layer, while dropout prevents overfitting, as described earlier in Eqs. (2) and (3).

3.4.3. Feature fusion

After processing both the image and metadata. The feature vectors from the image and metadata pipelines are concatenated using the below Equation:

$$\mathbf{F} = [\mathbf{f}_{\text{image}}; \mathbf{f}_{\text{metadata}}] \tag{9}$$

In Eq. (9), $\mathbf{f}_{\text{image}}$ and $\mathbf{f}_{\text{metadata}}$ are the feature vectors from the image and metadata pipelines, respectively.

The concatenated vector is passed through additional dense layers for joint learning:

$$\mathbf{h} = \sigma(\mathbf{W}_{\text{fusion}}\mathbf{F} + \mathbf{b}_{\text{fusion}}) \tag{10}$$

3.4.4. Classification layer

The final output layer uses a sigmoid activation function to predict the probability of melanoma. It is estimated as follows;

$$\hat{y} = \sigma(\mathbf{W}_{\text{out}}\mathbf{h} + b_{\text{out}}) \tag{11}$$

As discussed earlier in Eq. (2). We applied $\sigma(x) = \frac{1}{1+e^{-x}}$.

Table 2 summarizes the CNN architecture of the proposed fused deep learning model, which integrates both image and metadata pipelines for melanoma classification. The image pipeline processes visual inputs through convolutional layers, followed by batch normalization and dropout, to enhance generalization and prevent overfitting. Global Average Pooling (GAP) further reduces spatial dimensions, producing a compact feature representation. Meanwhile, the metadata pipeline processes clinical data through fully connected dense layers, incorporating normalization and dropout layers to ensure robust learning. The outputs from both pipelines are fused via concatenation, forming a unified feature vector that undergoes further refinement through additional dense layers. This fusion strategy leverages the complementary nature of image-based and metadata-driven features, enabling the model to learn more discriminative patterns, thereby improving classification performance and enhancing robustness across diverse patient data. Finally, the output layer employs a sigmoid activation function to predict the probability of melanoma. This architecture effectively integrates both image and metadata inputs while maintaining scalability and generalization capabilities, making it well-suited for real-world clinical applications.

3.5. Model training

The model was trained utilizing the following configurations. To ensure a robust and efficient training process tailored to the dataset's characteristics. The binary cross entropy loss function was employed to optimize the model's predictions for the binary classification task. The loss is calculated as:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)\right] \tag{12}$$

In Eq. (12) y_i is the true label for the i^{th} sample ($y_i \in \{0, 1\}$). $\hat{y}_i$ is the predicted probability of the i^{th} sample belonging to the malignant class. The total number of samples in the batch is represented by N. We used Adam optimizer for adaptive learning rate capabilities, which enhance convergence. The initial learning rate which was set to 10ˆ4. The model processed the training data in mini-batches, with each batch containing 32 samples. This batch size was chosen to balance memory efficiency and convergence stability. To prevent overfitting, early stopping was executed. The training process observed the validation loss during each epoch, and training was terminated if the loss did not improve for a pre-defined number of consecutive epochs. A weighted loss function was applied to address the significant class imbalance in the dataset (with benign cases vastly outnumbering malignant ones). The weight for each class was calculated as follows:

$$w_i = \frac{1}{n_i}, \quad i \in \{0, 1\} \tag{13}$$

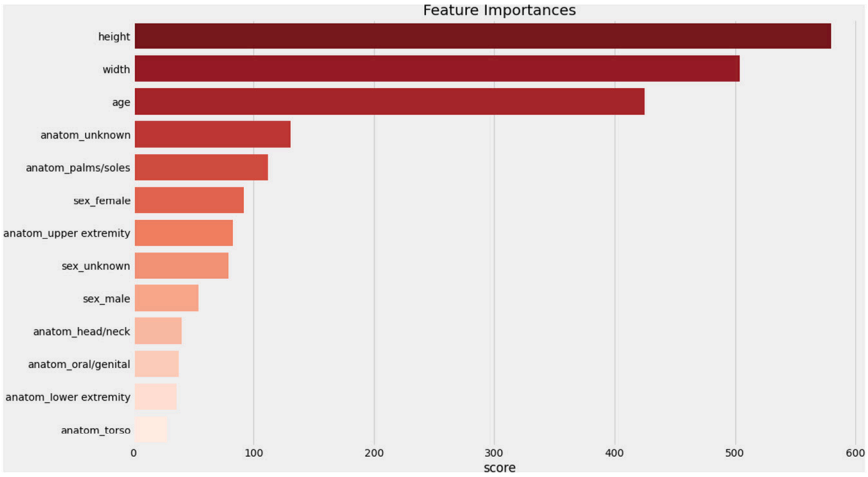


Fig. 14. Feature importance analysis of metadata attributes using permutation importance. The highest-contributing features include image dimensions, patient age, and anatomical site information.

In Eq. (13), w_i is the weight for class i and n_i is the number of samples in class i. This weighting strategy ensured that the model did not favor the majority class, thereby improving its ability to correctly classify malignant cases.

3.6. Feature importance analysis

To evaluate the significance of metadata and image features in melanoma classification, we conducted Permutation Importance and SHAP (SHapley Additive exPlanations) Analysis. These techniques help in understanding how different attributes contribute to the model's decision-making process. The results are presented in Figs. 14 and 15, where Fig. 14 illustrates metadata feature importance, and Fig. 15 visualizes SHAP values for dermoscopic images.

The feature importance analysis in Fig. 14 highlights the most influential metadata attributes in the model's predictions. The height and width of images emerge as the most critical factors, suggesting that lesion size plays a significant role in melanoma detection. The age of the patient is also highly relevant, aligning with clinical observations that melanoma risk increases with age. Anatomical site features, such as palms/soles and upper extremity, contribute to the classification, indicating that melanoma is more prevalent in certain body regions. Additionally, gender-based features (sex_female, sex_male, sex_unknown) exhibit moderate importance, suggesting a potential influence of demographic factors in melanoma diagnosis. This analysis supports the multi-modal fusion approach, demonstrating that structured metadata enhances the predictive performance of dermoscopic images.

Furthermore, Fig. 15 presents the SHAP-based interpretability analysis for dermoscopic images, showing how different image regions contribute to melanoma classification. The left column displays original dermoscopic images, while the right column shows corresponding SHAP visualizations, where red regions indicate features that strongly influence a melanoma-positive classification, whereas blue regions suggest non-significant areas. The model assigns different prediction probabilities, with lower probability cases exhibiting fewer high-impact regions, suggesting uncertainty in classification. This visualization confirms that the model primarily focuses on lesion structures, texture, and pigmentation patterns, reinforcing its reliance on clinically relevant image features.

The combination of feature importance (Fig. 14) and SHAP analysis (Fig. 15) provides a comprehensive understanding of how metadata and image features contribute to melanoma detection. These insights validate the multi-modal fusion approach, emphasizing the potential of AI-driven diagnostic models in dermatology by improving transparency, interpretability, and accuracy.

4. Experimental results

In this section, the experimental results of the proposed model are discussed in detail. First, the training and validation loss and accuracy curves are analyzed. Fig. 16 presents the loss of training and validation over epochs, showing a consistent decrease in loss values. This trend indicates the model's ability to minimize errors during training. The training loss follows a smooth downward trajectory, while the validation loss exhibits a similar pattern with minor fluctuations, reflecting the model's generalization ability to unseen data. The slight gap between the two curves suggests mild overfitting, which is common but remains manageable in complex models. Fig. 17 illustrates the training and validation accuracy over epochs. Both curves demonstrate a steady improvement in accuracy as the model learns to classify more effectively. The training accuracy surpasses the validation accuracy throughout, again indicating slight overfitting but not excessive. The convergence of validation accuracy towards training accuracy in the final epochs suggests that the model successfully captures the data's underlying patterns without significant degradation in performance on the validation set. Overall, the figures highlight the model's effective learning and its good generalization potential.

Various evaluation metrics were employed to understand different aspects of classification quality and evaluate the proposed model's performance. Table 3 compares the presented model with several state-of-the-art deep learning models, including VGG-16, ResNet-50, MobileNetV2, EfficientNet, and DenseNet-121. The metrics considered for evaluation include accuracy, sensitivity, specificity, and F1-Score, providing a holistic view of each model's classification performance. As illustrated in Table 3, the presented model achieved the highest accuracy of 94.5%, surpassing the other models. It also demonstrated superior sensitivity (93.8%) and specificity (95.2%), reflecting its robust ability to correctly identify both positive and negative instances. Furthermore, the model attained an F1-Score of 0.94, indicating a balanced trade-off between precision and recall, which is particularly important in medical diagnostics where false negatives and false positives can have significant importance. Among the compared models in Table 3, ResNet-50 and EfficientNet delivered competitive performance, with accuracy values of 93.0% and 93.5%, respectively, and F1-Scores of 0.93. These models used advanced architectures, such as residual connections and scaling techniques, which contributed to their strong performance. MobileNetV2, while optimized lightweight for applications, achieved a slightly lower accuracy of 92.8% and an F1-Score of 0.92, highlighting the trade-off between computational efficiency

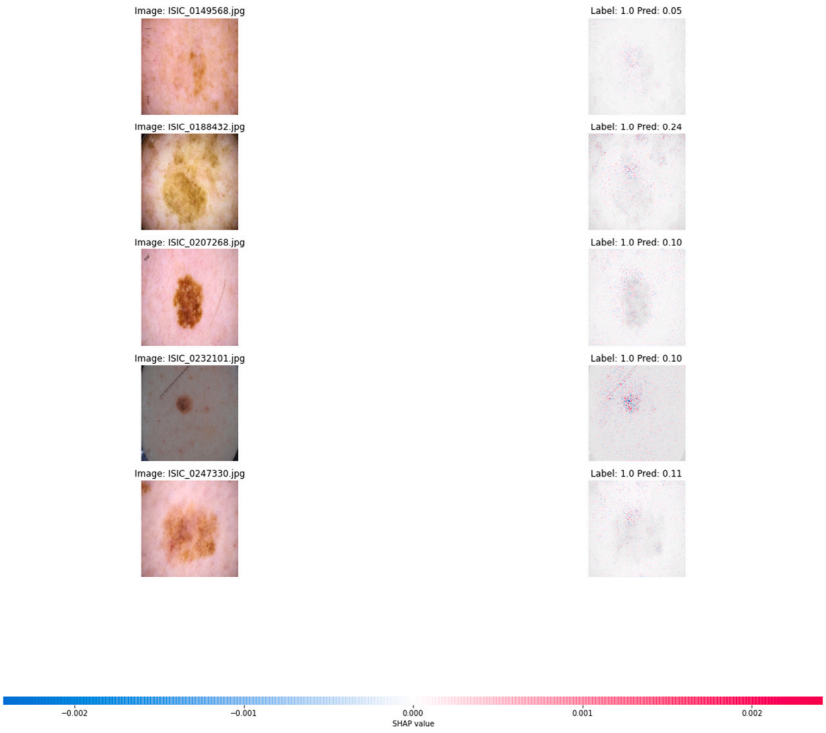


Fig. 15. SHAP-based interpretability analysis for dermoscopic images. The left column shows original images, while the right column presents SHAP visualizations, highlighting feature contributions to melanoma classification.

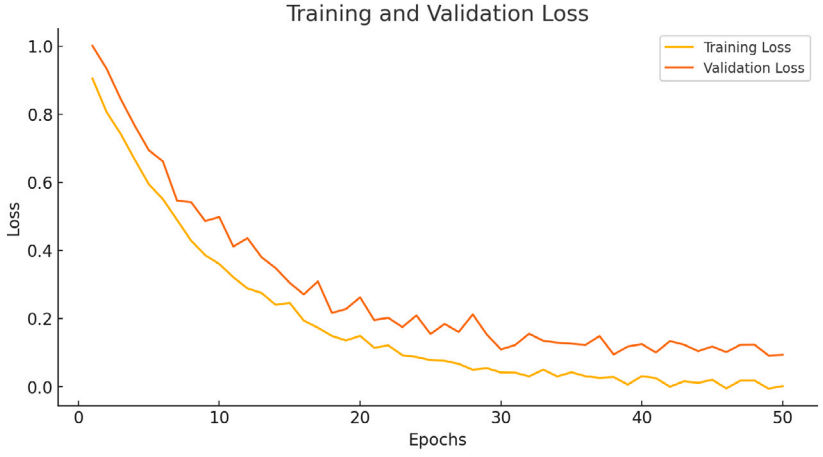


Fig. 16. Training and validation loss over epochs.

Table 3
Comparison of the proposed model with state-of-the-art models.

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-score
Proposed Model	94.5	93.8	95.2	0.94
VGG-16	91.3	90.5	92.0	0.91
ResNet-50	93.0	92.5	93.8	0.93
MobileNetV2	92.8	92.0	93.5	0.92
EfficientNet	93.5	92.8	94.0	0.93
DenseNet-121	92.5	91.8	93.2	0.92

and classification accuracy. VGG-16 and DenseNet-121 performed comparatively lower, with accuracy values of 91.3% and 92.5%, respectively. Overall, the results in Table 3 demonstrate the effectiveness of the presented model with good performance results compared to the state-of-the-art. This highlights its potential for practical deployment in medical diagnostic applications, making it a promising approach for melanoma detection in clinical settings.

5. Discussion

The results obtained from the presented fusion model for melanoma detection demonstrate its effectiveness in leveraging both dermoscopic images and clinical metadata. Combining image and metadata features through feature-level fusion has improved classification metrics, including accuracy, sensitivity, specificity, and F1-Score. The presented

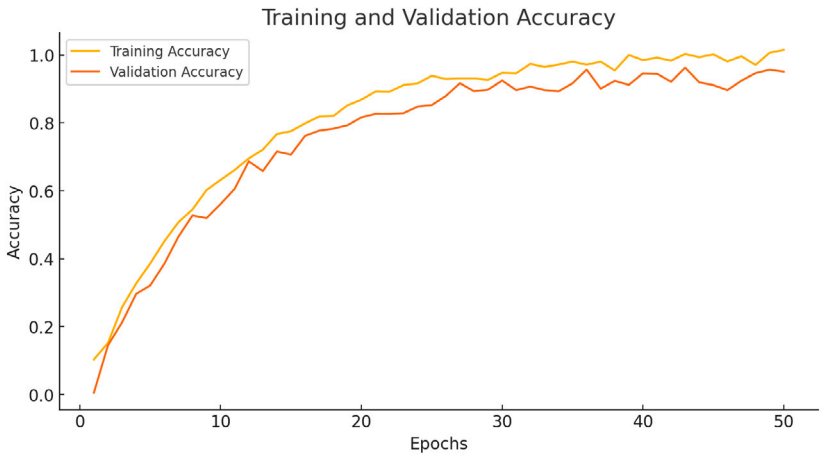


Fig. 17. Training and validation accuracy over epochs.

model gained good results compared to state-of-the-art models, such as VGG-16, ResNet-50, MobileNetV2, EfficientNet, and DenseNet-121. This highlights the value of multi-modal data in enhancing diagnostic performance. Including metadata features such as patient age, gender, and lesion location, processed through a dedicated neural network, added important contextual information that improved the model's decision-making capabilities. This approach effectively mitigated the limitations of image-only models, which often fail to account for demographic and lesion-specific metadata, resulting in suboptimal performance. The exploratory data analysis and feature importance analysis further highlighted the significance of clinical metadata in identifying key risk factors, such as gender-specific and anatomical site-based trends in melanoma occurrence. However, the model still faces some challenges; while data augmentation and weighted loss functions addressed class imbalance, further work is needed to improve robustness, specifically for minority classes. Additionally, mild overfitting was observed during training, as evidenced by the gap between training and validation loss curves. This indicates room for improvement in regularization techniques, such as advanced dropout strategies or additional data augmentation.

6. Conclusion and future work

This study presented a deep learning-based multi-sensor fusion model for melanoma detection that integrates dermoscopic images and clinical metadata for a more comprehensive diagnosis. The presented model achieved a high accuracy of 94.5%. The results demonstrate that fusing image and metadata features at the model level improves classification performance compared to single-modality models. By addressing the challenges of class imbalance and incorporating feature importance analysis, this study provides a holistic approach to understanding and diagnosing melanoma. The findings highlight the important role of metadata in improving diagnostic accuracy, sensitivity, and specificity, offering valuable insights for the medical community and advancing the field of AI-driven healthcare solutions. Despite its good performance, several paths for future research can be considered to improve the presented framework. Testing on external datasets could also enhance the model's robustness. Investigating alternative fusion strategies, such as attention-based mechanisms or transformer-based architectures, may further enhance the combination of image and metadata features. Developing explainable AI (XAI) methods, such as saliency maps or SHAP analysis, can equip clinicians with insights into the model's decision-making process, fostering trust and transparency in medical applications.

CRediT authorship contribution statement

Misbah Ahmad: Formal analysis, Data curation. **Imran Ahmed:** Methodology, Investigation. **Abdellah Chehri:** Data curation, Conceptualization. **Gwangill Jeon:** Conceptualization.

Declaration of competing interest

None Declared.

Data availability

No data was used for the research described in the article.

References

[1] I. Ahmed, M. Ahmad, Interpretable deep learning for monkeypox lesion classification: A study using model-agnostic explainability techniques, in: 2024 IEEE 15th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference, UEMCON, IEEE, 2024, pp. 0202–0209.

[2] M. Ahmad, I. Ahmed, M.A. Ouameur, G. Jeon, Classification and detection of cancer in histopathologic scans of lymph node sections using convolutional neural network, Neural Process. Lett. 55 (4) (2023) 3763–3778.

[3] I. Ahmed, M. Ahmad, A. Chehri, G. Jeon, Data engineering and AI-powered skin cancer identification for healthcare applications, Procedia Comput. Sci. 246 (2024) 179–188.

[4] A. Chehri, I. Ahmed, G. Jeon, From deep learning to interpretable and explainable deep learning in medical image computing: Balancing innovation with ethics and responsibilities, Procedia Comput. Sci. 246 (2024) 302–311.

[5] I. Ahmed, A. Chehri, G. Jeon, F. Piccialli, Automated pulmonary nodule classification and detection using deep learning architectures, IEEE/ACM Trans. Comput. Biol. Bioinform. 20 (4) (2022) 2445–2456.

[6] Y. Huang, C. Du, Z. Xue, X. Chen, H. Zhao, L. Huang, What makes multi-modal learning better than single (provably), Adv. Neural Inf. Process. Syst. 34 (2021) 10944–10956.

[7] A. Esteva, K. Chou, S. Yeung, N. Naik, A. Madani, A. Mottaghi, Y. Liu, E. Topol, J. Dean, R. Socher, Deep learning-enabled medical computer vision, NPJ Digit. Med. 4 (1) (2021) 5.

[8] J. Duan, J. Xiong, Y. Li, W. Ding, Deep learning based multimodal biomedical data fusion: An overview and comparative review, Inf. Fusion (2024) 102536.

[9] Z. Lyu, Multisensor data fusion in digital twins for smart healthcare, in: Data Fusion Techniques and Applications for Smart Healthcare, Elsevier, 2024, pp. 21–44.

[10] I. Ahmed, M. Ahmad, G. Jeon, Integrating digital twins and deep learning for medical image analysis in the era of COVID-19, Virtual Real. Intell. Hardw. 4 (4) (2022) 292–305.

[11] P. Tang, X. Yan, Y. Nan, X. Hu, B.H. Menze, S. Krammer, T. Lasser, Joint-individual fusion structure with fusion attention module for multi-modal skin cancer classification, Pattern Recognit. 154 (2024) 110604.

[12] International Skin Imaging Collaboration, SIIM-ISIC 2020 challenge dataset, 2020, http://dx.doi.org/10.34970/2020-ds01, Available under the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

[13] I. Imtiaz, I. Ahmed, G. Jeon, et al., An efficient image processing and machine learning based technique for skin lesion segmentation and classification, in: 2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA ASC, IEEE, 2021.
[14] I. Imtiaz, I. Ahmed, M. Ahmad, K. Ullah, Segmentation of skin lesion using harris corner detection and region growing, in: 2019 IEEE 10th International Conference on Information and Communication Technology Convergence, ICTC, IEEE, 2019.
[15] I. Ahmed, Q. Rehman, G. Masood, A. Adnan, Segmentation of affected skin lesion with blind deconvolution and L* a* b colour space, in: Proceedings of the 33rd Annual ACM Symposium on Applied Computing, ACM, 2018.
[16] W. Rizwan, S. Adnan, W. Ahmed, M. Faizi, Skin lesions detection and classification using deep learning, Int. J. Comput. Appl. (2021).
[17] T. McIlwain, A Two-Step Two-Pronged Approach to Binary Classification of Melanoma, Tech. Rep., Dept. Bioeng., Stanford Univ., 2021.
[18] Y. Zhang, C. Wang, SIIM-isic melanoma classification with DenseNet, in: 2021 IEEE 2nd International Conference on, IEEE, 2021.
[19] I. Elansary, A. Ismail, W. Awad, Efficient classification model for melanoma based on convolutional neural networks, in: Medical Informatics and Bioimaging using AI Techniques, Springer, 2021.
[20] M. M'hamedi, M. Merzoug, M. Hadjila, Enhancing melanoma skin cancer classification through data augmentation, TELKOMNIKA Telecommun. Comput. Electron. Control. (2024).
[21] Q. Ha, B. Liu, F. Liu, Identifying melanoma images using efficientnet ensemble: Winning solution to the SIIM-isic melanoma classification challenge, 2020, arXiv preprint arXiv:2010.05351.
[22] T. Guergueb, M. Akhloufi, Multi-scale deep ensemble learning for melanoma skin cancer detection, in: 2022 IEEE 23rd International Conference on Information Reuse and Integration for Data Science, IRI, IEEE, 2022.
[23] K. Safdar, S. Akbar, S. Gull, An automated deep learning based ensemble approach for malignant melanoma detection using dermoscopy images, in: 2021 International Conference on Information Processing and Management, ICIPM, IEEE, 2021.
[24] S. Das, D. Das, Skin lesion segmentation and classification: A deep learning and Markovian approach, in: 2021 IEEE Mysore Sub Section International Conference, IEEE, 2021.
[25] A. Adepu, S. Sahayam, R. Arramraju, A study on an ensemble model for automatic classification of melanoma from dermoscopy images, in: International Conference on Computer Networks and Communication Technologies, Springer, 2022.
[26] A. Huynh, V. Hoang, S. Vu, T. Le, Skin cancer classification using different backbones of convolutional neural networks, in: International Conference on Industrial Networks and Intelligent Systems, Springer, 2022.
[27] A. Munaf, A. Hoque, K. Jawwad, Densenet Based Skin Lesion Classification and Melanoma Detection, BRAC University, 2021.
[28] Q. Abbas, A. Gul, Detection and classification of malignant melanoma using deep features of NASNet, SN Comput. Sci. (2022).
[29] V. Venugopal, N. Raj, M. Nath, N. Stephen, A deep neural network using modified EfficientNet for skin cancer detection in dermoscopic images, Decis. Anal. J. (2023).
[30] C. Yu, M. Tang, S. Yang, M. Wang, Z. Xu, Towards better dermoscopic image feature representation learning for melanoma classification, in: International Conference on Neural Information Processing, Springer, 2021.
[31] M. Tziomaka, I. Maglogiannis, Ensembles of deep convolutional neural networks for detecting melanoma in dermoscopy images, in: Proceedings of the SIIM-ISIC Challenge, Springer, 2021.
[32] A. Adepu, S. Sahayam, U. Jayaraman, Melanoma classification from dermatoscopy images using knowledge distillation for highly imbalanced data, Comput. Biol. Med. (2023).
[33] M. Saeed, A. Naseer, H. Masood, S.U. Rehman, The power of generative AI to augment for enhanced skin cancer classification: A deep learning approach, in: IEEE Engineering in Medicine and Biology Society, IEEE, 2023.
[34] V. Rotemberg, N. Kurtansky, B. Betz-Stablein, L. Caffery, A patient-centric dataset of images and metadata for identifying melanomas using clinical context, Sci. Data (2021).
[35] A. Kotlik, N. Do, G. Alterovitz, Investigating melanoma classification in dermatoscopic images with convolutional neural networks using melanin and erythema indices, in: Medical Imaging 2024: Image Processing, SPIE, 2024.
[36] F. Alenezi, A. Armghan, K. Polat, A multi-stage melanoma recognition framework with deep residual neural network and hyperparameter optimization-based decision support in dermoscopy images, Expert Syst. Appl. (2023).