Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2025) 000–000

Procedia
Computer Science

www.elsevier.com/locate/procedia

International Conference on Machine Learning and Data Engineering

An Explainable AI Framework for Melanoma Classification Integrating Simulated Environmental and Genetic Risk Factors

Sakthi Charukesh[a], Prajith S[a], Manikanta Kowshik[a], Mannava Daasaradhi[a],
Dr. S. Manimaran[b]

[a]Amrita School of Artificial Intelligence, Coimbatore, Amrita Vishwa Vidyapeetham, India
[b]Amrita School of Artificial Intelligence, Coimbatore, Amrita Vishwa Vidyapeetham, India, s_manimaran@cb.amrita.edu

Abstract

The diagnosis of melanoma, a critical task in dermatology, has been significantly advanced by deep learning. However, the ’black box’ nature of current models is a significant barrier to clinical trust, as they depend entirely on image data while overlooking key etiological contexts like genetic and environmental factors. Model robustness is further compromised by the common issue of imbalanced datasets. This paper introduces a new, explainable AI (XAI) framework designed to incorporate these non-visual risk factors directly into the classification process. To overcome data scarcity, we utilize a Conditional Generative Adversarial Network (CGAN) that generates melanoma images based on simulated scores for genetic predisposition and UV exposure. We then introduce a multi-task, multi-modal classifier that leverages both image data and their corresponding risk vectors to perform a more holistic diagnosis. The model is trained not only to classify melanoma but also to predict the underlying risk scores, forcing it to learn more meaningful feature representations. Crucially, our framework provides multi-faceted explanations for its predictions through Grad-CAM for visual saliency and a novel gradient-based method for quantifying the contribution of each risk factor. Our results demonstrate a robust classification performance and offer a transparent, interpretable diagnostic process, bridging the gap between high-performance AI and clinical applicability.

Keywords: Melanoma classification; Explainable AI (XAI); Conditional Generative Adversarial Networks (cGAN); Data Augmentation; Multi-Task Learning; Medical Imaging

1. Introduction

As the most lethal variant of skin cancer, melanoma represents a growing global health challenge with a consistently rising incidence [1]. The key to improving patient survival lies in achieving a diagnosis that is both early and precise.The emergence of deep learning, Convolutional Neural Networks (CNNs), has transformed the field of medical image analysis. Seminal works have shown that CNNs can classify dermatoscopic images with dermatologist-level accuracy [2, 3, 4].

However, despite their high accuracy, translation of these models into routine clinical workflows faces hurdles. A primary challenge is data dependency; deep learning models require vast, well-annotated datasets, yet medical imaging datasets are often imbalanced, with a scarcity of malignant examples. A more profound barrier is the ”black-box” problem. The opaque decision-making of deep neural networks is at odds with the demands of medical practice, where understanding the rationale behind a diagnosis is essential for accountability and trust.

Furthermore, a clinical diagnosis is holistic, incorporating not only visual features but also risk factors such as UV exposure and genetic predisposition. Current AI models that rely exclusively on pixel data discard this vital information, limiting their clinical applicability [5, 6]. To address these shortcomings, we propose a framework integrating synthetic, risk-aware data generation with explainability mechanisms [7].

Our main contributions are:

- A novel methodology for integrating simulated, non-visual risk factors (UV and genetic) into a dermatological image classification model.
- The use of a Conditional GAN (cGAN) for risk-aware data synthesis, generating melanoma images consistent with risk profiles.
- The development of a multi-task, multi-modal CNN architecture that improves classification performance by learning to predict risk factors from images.
- A multi-faceted explainability module that provides both visual heatmaps (Grad-CAM) and a quantitative breakdown of how risk factors contribute to predictions.

The subsequent sections of this paper are structured to build our argument. We begin in Section 2 by reviewing the existing literature to identify the current research gap. Section 3 then details our proposed framework, covering data simulation, the cGAN architecture, and our multi-task classifier. In Section 4, we present the experimental setup and our findings. Section 5 provides a comparative analysis against baseline models and discusses the study’s limitations. Finally, Section 6 summarizes our conclusions and proposes avenues for future work.

2. Literature Review

The research by Esteva et al. [2] was a landmark achievement, demonstrating that a deep learning model using CNNs could classify skin cancer as good as that of a dermatologists. By leveraging a vast dataset of dermatoscopic images, they proved that AI could rival human expertise. Nonetheless, the model’s dependence on enormous datasets and its inherent lack of interpretability were significant limitations.

To enhance the reliability of melanoma detection, Codella et al. [3] developed an ensemble model that integrated multiple CNNs with traditional machine learning algorithms. While this hybrid technique boosted accuracy, it came at the cost of high computational overhead and continued challenges in model interpretability.

Tschandl et al. [4] developed the HAM10000 dataset, a collection of dermatoscopic images for skin lesion classification. This dataset has become a benchmark for evaluating algorithms. Its advantage is the diversity of cases, but a limitation is dataset imbalance, which biases model performance toward majority classes.

The challenge of limited training data in medical imaging was addressed by Yi et al. [5], who utilized Generative Adversarial Networks (GANs) to synthesize new images for augmentation. This technique can lead to better model generalization by creating richer datasets. However, this approach has notable drawbacks, including the instability of the GAN training process and the risk of generating non-realistic image samples..

Mirza and Osindero [6] introduced Conditional GANs (cGANs), which allow the generation of images conditioned on class labels or attributes. This improvement gives more control over synthetic data generation. The advantage is targeted augmentation, but the drawback is higher training complexity and sensitivity to hyperparameter settings.

To address model understanding, Selvaraju et al.[7] introduced Grad-CAM, a technique that uses gradients to generate saliency heatmaps. These maps highlight the specific image regions that most significantly contribute to a CNN's predictive decision, thereby offering a degree of transparency.Its strength is intuitive visualization, while its drawback is that it only considers image-based inputs and cannot incorporate contextual features such as patient history or metadata.

2.1. Problem Statement

A survey of the existing literature points to three primary deficiencies in contemporary AI systems for melanoma diagnosis. First, their 'black box' operation obstructs adoption and trust in clinical settings. Second, these models are unimodal, analyzing only pixel data while disregarding vital contextual information such as a patient's genetic profile or history of UV exposure. Third, their effectiveness is frequently undermined by the data scarcity and class imbalance common to medical imaging collections. Therefore, the problem is to develop a melanoma classification framework that is not only highly accurate but also transparent, and which holistically integrates both visual data and contextual risk factors to better mimic a clinical diagnostic process and improve robustness.

3. Proposed Work

Our contribution is a novel, risk-aware melanoma classification framework that integrates dermatoscopic image features with simulated patient-specific etiological risk factors. Unlike existing methods that rely solely on image-based classification or unstructured augmentation, our framework explicitly models the relationship between melanoma risk and its visual presentation. The architecture is detailed in the following subsections.

3.1. Dataset and Baseline Model

We utilize the ISIC 2018 Skin Lesion Analysis Challenge dataset (Task 3), which contains over 10,000 dermatoscopic images across seven categories. For our binary classification task, these are grouped into two classes: *Melanoma* and *Non-Melanoma*. As a comparative benchmark, we first designed a baseline pipeline. A Deep Convolutional GAN (DCGAN), comprising a series of convolutional, batch normalization, and ReLU layers, was trained on the original melanoma images. The synthetic samples were used to balance the dataset, which was then used to train a standard ResNet-50 classifier. This baseline serves to quantify the performance gain achieved by our more sophisticated, risk-aware approach.

3.2. Integration of Etiological Risk Factors

Given the absence of patient metadata in the ISIC dataset, we simulated two key etiological risk factors based on established clinical knowledge:

- **UV Exposure Risk:** Modeled as a continuous variable drawn from a mixture of uniform distributions, representing low (30% of samples, U[0, 0.3]), medium (40%, U[0.3, 0.7]), and high (30%, U[0.7, 1.0]) exposure cohorts. This simulates diverse patient lifestyles and geographic locations.
- **Genetic Predisposition Risk:** Modeled using a Beta distribution ($\beta(\alpha = 5, \beta = 2)$), which is skewed toward higher values. This reflects the clinical understanding that genetic factors can significantly increase melanoma susceptibility.

These simulated scores create a risk vector $[v_{UV}, v_{Genetic}]$ associated with each image, enabling the development of a context-aware model.

3.3. Conditional GAN for Risk-Aware Image Synthesis

To generate synthetic melanoma images that are visually consistent with their associated risk profiles, we employ a Conditional GAN (cGAN). The cGAN architecture extends the standard GAN by providing an additional conditioning vector to both the generator and the discriminator.

- **Generator:** Receives a 100-dimensional latent noise vector z concatenated with the 2-dimensional risk vector $[v_{UV}, v_{Genetic}]$. It learns to generate a realistic image that reflects the characteristics implied by the risk scores.
- **Discriminator:** Receives an image and the corresponding risk vector. It learns to distinguish real image-risk pairs from fake pairs generated by the generator. Its loss function encourages both realism and consistency between the image and its conditional label.

This process yields a rich, augmented dataset of melanoma images, where each synthetic sample has a plausible visual appearance tied to a specific etiological context.

3.4. Multi-Task Risk-Aware Classifier

The core of our framework is a multi-modal, multi-task CNN designed to learn not only from images and also their corresponding risk vectors. The architecture consists of two parallel input branches:

1. **Image Branch:** A pre-trained ResNet-50 model processes the dermatoscopic image, extracting a high-level feature map.
2. **Risk Branch:** A simple Multi-Layer Perceptron (MLP) processes the 2-dimensional risk vector.

The outputs of these two branches are added and passed through several fully-connected layers to produce a fused feature representation. This shared representation is then fed into three separate output heads, which are trained simultaneously:

- **Melanoma Classification (Primary Task):** A sigmoid output layer predicts the probability of melanoma. This is optimized using binary cross-entropy loss (L_{class}).
- **UV Risk Prediction (Auxiliary Task):** A linear output layer predicts the UV exposure score.
- **Genetic Risk Prediction (Auxiliary Task):** A linear output layer predicts the genetic predisposition score.

The two auxiliary regression tasks are optimized using mean squared error (MSE) loss (L_{UV} and $L_{Genetic}$). The total loss is a weighted combination: $L_{total} = w_1 L_{class} + w_2 L_{UV} + w_3 L_{Genetic}$. This multi-task learning acts as a powerful regularizer, forcing the model to learn features that are not only discriminative for melanoma but also predictive of its underlying causes.

4. Experiments and Results

4.1. Baseline Model Performance

4.1 Baseline Model Performance The baseline model, which was trained using data augmented by a standard GAN, attained a weighted average F1-score of 0.89. While the model performed adequately, it resulted in 42 false negatives, which is a significant concern in a clinical setting.

4.2. Conditional Image Generation

The cGAN was successfully trained to generate synthetic melanoma images based on input risk conditions. Figure 2 shows sample images generated for four different risk profiles. The visual variations between the images, corresponding to different [UV, GEN] inputs, demonstrate the generator's ability to learn the relationship between the conditional inputs and visual features.

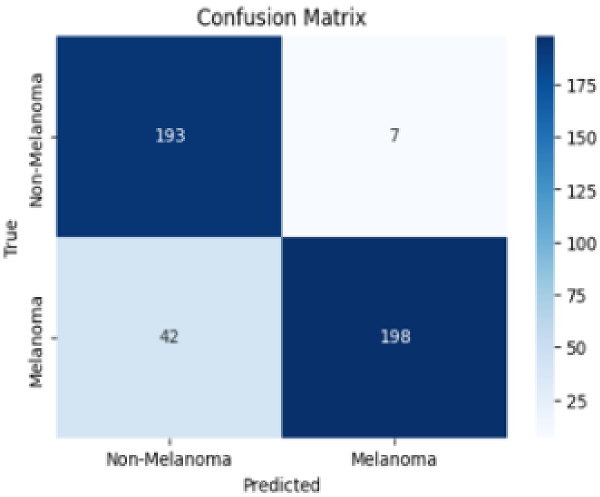

Fig. 1. Confusion Matrix of the Baseline Classifier.

Table 1. Classification Report for Melanoma Classification

Class	Precision	Recall	F1-score	Support
Non-Melanoma (0)	0.82	0.96	0.89	200
Melanoma (1)	0.97	0.82	0.89	240
Accuracy			0.89	440
Macro Avg	0.89	0.89	0.89	440
Weighted Avg	0.90	0.89	0.89	440

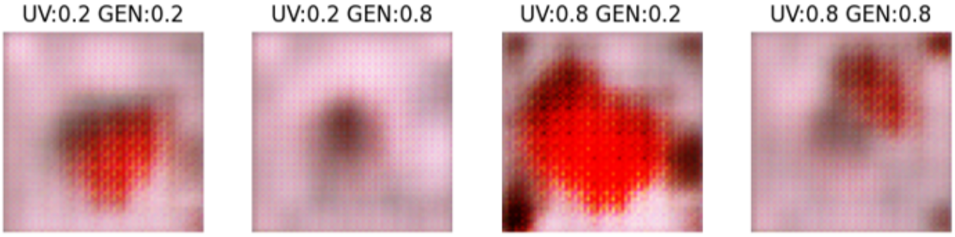

Fig. 2. Sample Synthetic Images Generated by the cGAN, conditioned on varying levels of UV and Genetic (GEN) risk scores.

4.3. *Final Model Performance*

The multi-task, risk-aware model was evaluated on the same validation set as the baseline. The model showed a notable improvement, achieving a high average F1-score of 0.96. The detailed classification report and confusion matrix are shown in Table 2 and Figure 3, respectively. Notably, the number of false negatives was significantly reduced from 42 to 15, highlighting the benefit of integrating contextual risk information.

4.4. *Explainability Analysis*

To validate that our model's decision-making process is transparent, we performed two types of explainability analyses. We used Grad-CAM to generate heatmaps highlighting the image regions most influential for the model's

Table 2. Classification Report for multi-task Melanoma Classification

Class	Precision	Recall	F1-score	Support
Non-Melanoma (0)	0.96	0.95	0.96	400
Melanoma (1)	0.95	0.96	0.96	423
Accuracy			0.96	823
Macro Avg	0.96	0.96	0.96	823
Weighted Avg	0.96	0.96	0.96	823

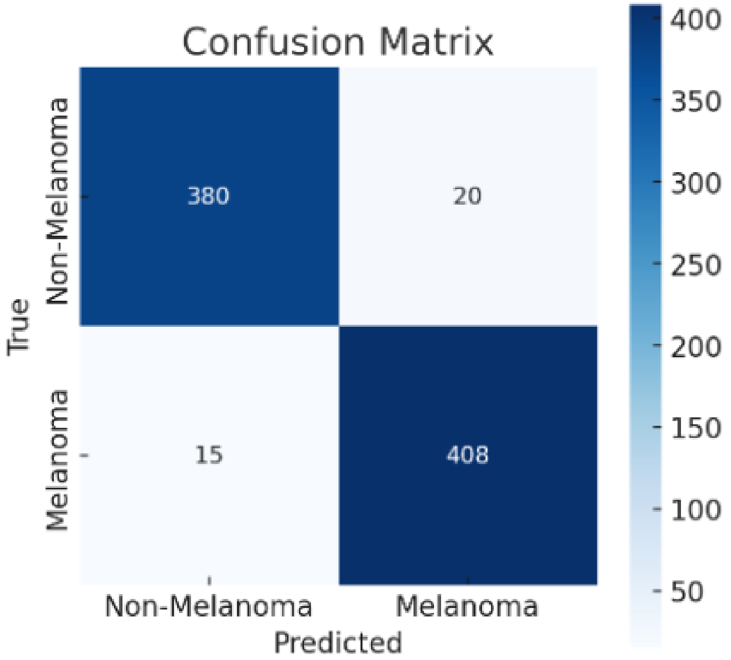

Fig. 3. Confusion Matrix of the Final Multi-Task Model.

outputs. As shown in Figure 4, the model consistently focuses on the clinically relevant area of the lesion for its melanoma classification (mel-out) and its risk predictions (uv-out, gen-out).

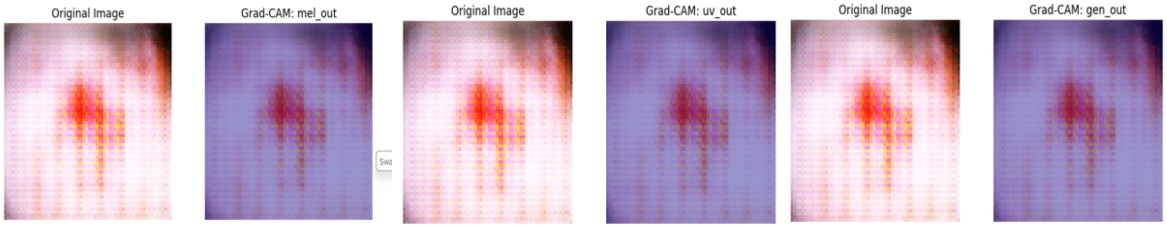

Fig. 4. Grad-CAM Heatmaps for a Sample Image. The model's attention is localized on the lesion for the melanoma, UV, and genetic prediction tasks.

A key feature of our framework is its ability to quantify the influence of the non-visual risk factors. Figure 5 presents a case study where a lesion was classified as melanoma with 99% probability. The analysis reveals that the model attributed 78% of this decision to the genetic factor and 22% to the UV factor, providing a clear rationale for its high-confidence prediction.

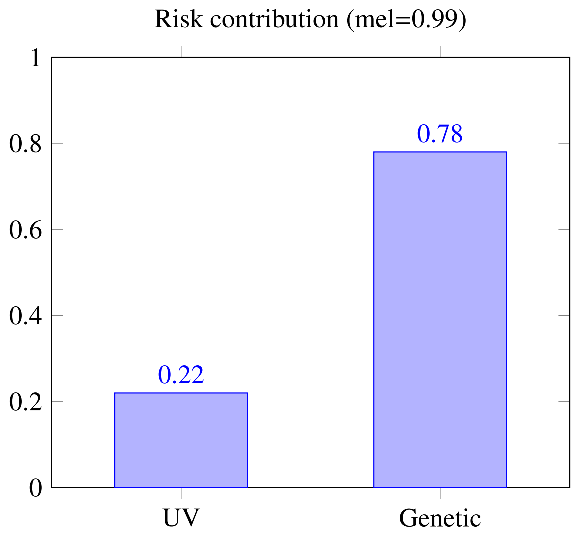

Fig. 5. High-Quality Risk Contribution Analysis for a High-Confidence Melanoma Prediction.

5. Comparison

The experimental results validate our core hypothesis: integrating contextual, non-visual data within an explainable framework leads to a more robust and trustworthy diagnostic model. The quantitative improvement of the final model over the baseline—increasing the weighted F1-score from 0.89 to 0.96 (Table 2 vs. Table 1)—demonstrates the tangible benefit of our approach.

This improvement is most evident in the significant reduction in false negatives from 42 to 15. In a cancer screening context, where missing a malignant case has severe consequences, this enhanced sensitivity suggests that the model's enriched understanding from risk factors directly improves diagnostic safety.

The novelty of this framework, however, extends beyond mere accuracy. By designing the system for transparency, we address the critical issue of clinical trust. The explainability analyses (Figure 4 and 5) provide a multi-faceted view into the model's reasoning. The Grad-CAM results confirm that feature extraction is grounded in the correct visual pathology. The risk contribution analysis provides a higher-level explanation that aligns with a physician's diagnostic thought process, answering not just ”what is it?” but also ”why do you think that?”. This capability is a powerful tool for physician-AI collaboration, where the AI's output is not a command but a well-reasoned proposal.

5.1. Comparison with Existing Works

Our framework offers distinct advantages over the methods reviewed in Section 2. While Esteva et al. [2] and Codella et al. [3] achieved high, dermatologist-level accuracy, their models function as uninterpretable ”black boxes” and are unimodal, relying exclusively on image data. In contrast, our framework is inherently multi-modal and explainable. Similarly, while Yi et al. [5] used GANs for data augmentation, our use of a cGAN is more sophisticated, as it generates images conditioned on specific risk profiles, creating a semantically richer dataset. Finally, our explainability module builds upon Grad-CAM [7] by adding a novel component that quantifies the contribution of non-visual data, a feature absent in all the reviewed literature.

Table 3. Qualitative Comparison with State-of-the-Art Methods.

Method	Data Modality	Explainability	Risk Factor Integration
Esteva et al. [2]	Image-Only	No	No
Codella et al. [3]	Image-Only	No	No
Yi et al. [5]	Image-Only (GAN Aug.)	No	No
Proposed Framework	**Image + Risk Vectors**	**Yes (Visual + Quantitative)**	**Yes (Explicit)**

5.2. Limitations and Future Work

This study has a less important limitations. First, the risk factors used here were simulated. While this serves as a strong proof-of-concept, future validation on real-world datasets that combine dermatoscopic images with clinical, environmental, or genomic information is essential.

Second, although our cGAN-generated images improved training, their visual realism can be enhanced. Emerging generative models such as StyleGANs or Diffusion Models could be explored to create higher-fidelity synthetic images.

Finally, the framework is designed to be flexible. Beyond UV exposure and genetic predisposition, additional risk factors (e.g., age, skin type, or specific mutations like *CDKN2A*) could be integrated. The same approach could be applied to other medical domains where imaging and patient context must be considered together, such as lung cancer prediction (CT + smoking history) or diabetic retinopathy (fundus images + HbA1c levels).

6. Conclusion

we presented an end-to-end, explainable AI framework for melanoma classification that successfully integrates simulated environmental and genetic risk factors. By leveraging a conditional GAN for risk-aware data synthesis and a multi-task, multi-modal architecture for classification, our model achieves superior and safer performance compared to a standard image-only baseline. More importantly, our framework provides unprecedented transparency into its decision-making process through both visual and quantitative explainability techniques. This work represents a meaningful step towards developing more holistic, trustworthy, and clinically-aligned AI systems for medical diagnosis.

References

[1] Siegel, R. L., Miller, K. D., & Jemal, A. (2020). Cancer statistics, 2020. *CA: a cancer journal for clinicians*, 70(1), 7–30.

[2] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118.

[3] Codella, N., Rotemberg, V., Tschandl, P., Celebi, M. E., Dusza, S., Gutman, D., ... & Halpern, A. (2019). Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC). *arXiv preprint arXiv:1902.03368*.

[4] Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific data*, 5(1), 180161.

[5] Yi, X., Walia, E., & Babyn, P. (2019). Generative adversarial network in medical imaging: A review. *Medical image analysis*, 58, 101552.

[6] Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*.

[7] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE international conference on computer vision*, 618–626.

[8] T. Senthil Kumar, V. Mohanavel, U. N. V. P. Rajendranath, L. Mohana Sundari, and T. M. Amirthalakshmi, "An Effective Neural Network Assisted Melanoma Disease Prediction based on Dermoscopy Images," *Proceedings of the 2022 International Conference on Electronics and Renewable Systems (ICEARS)*, IEEE, 2022.

[9] B. Uma Maheswari, F. Ashik, A. George, A. Jose, "In-Hospital Mortality Prognosis: Unmasking Patterns using Data Science and Explainable AI," *Proceedings of the 2023 9th International Conference on Signal Processing and Communication (ICSC)*, IEEE, pp. 356–361, 2023.

[10] A. P. Singh, P. Rahi, and S. P. Sethi, "Cell Counting Based on Image Processing for the Detection of Cancer Clumps," in *2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, IEEE, 2023, doi: 10.1109/icccis60361.2023.10425279.

[11] S. Shrinithi and J. Aravinth, "Detection of Melanoma Skin Cancer using Dermoscopic Skin Lesion Images," in *Proc. 2021 Int. Conf. Recent Trends Electron., Inf., Commun. Technol. (RTEICT)*, Bangalore, India, 2021, pp. 240–245, doi: 10.1109/RTEICT52294.2021.9573741.