Contents lists available at Curevita Journals

CUREVITA INNOVATION OF BIODATA INTELLIGENCE

Cross-Domain Explainable Multimodal Deep Learning for Equitable Skin Cancer Diagnosis Across Diverse Populations

Aarti Chahal[1] and Deepak Dudeja [2]

Article info

Article history: Received 2nd Sept 2025, Revised 18 Nov 2025, Accepted 4th Dec 2025, Published Dec 2025

Keywords: Multimodal Deep Learning, Skin Cancer Diagnosis, Cross-Domain Generalization, Explainable AI (XAI), Fairness and Bias Mitigation.

Corresponding author: Aarti Chahal, PhD Scholar, Email ID: aarti22chahal@gmail.com

[1] PhD Scholar, Maharishi Markandeshwar University, Mullana.

2 Professor, Computer Science and Engineering Department, Maharishi Markandeshwar University, Mullana.

Citation: Chahal Aarti and Dudeja Deepak..2025.Cross-Domain Explainable Multimodal Deep Learning for Equitable Skin Cancer Diagnosis Across Diverse Populations. Curevita Innovation of BioData Intelligence .1,2,132-141.

Publisher: Curevita Research Pvt Ltd

Abstract: Diagnosing skin cancer (melanoma vs. non-melanoma) accurately remains challenging, particularly in differentiating subtypes in clinical settings. Current Deep Learning (DL) models are limited by uni-modal data (often only dermatoscopic images), poor generalizability, and a lack of transparency regarding inherent biases[11]. This paper proposes an Explainable Multimodal Deep Learning (EM-DL) framework for skin cancer subtype prediction. We fuse non-invasive images (e.g., dermatoscopy/clinical images) with tabular clinical data (demographics, lesion history) using a Transformer-based fusion network[22]. Training is conducted on a centralized, augmented multi-center dataset to enhance cross-domain robustness[3]. Finally, we integrate XAI (SHAP and Grad-CAM) [4] to audit model fairness across protected subgroups (e.g., Fitzpatrick skin type, ethnicity, gender) and provide interpretable feature attributions, establishing a new standard for ethical and globally scalable AI diagnostics.

Introduction

Skin cancer, including melanoma and various non-melanoma forms, is a major global health burden. While non-invasive screening using DL is maturing, the next frontier is accurate subtype classification (e.g., differentiating Melanoma from Basal Cell Carcinoma) and ensuring equity in diagnosis. Standard DL models often suffer from unquantified fairness issues, which is especially critical in visually-based diagnoses where darker Fitzpatrick skin types are often underrepresented in training data, leading to potential bias, Jain et al., 2019, Abujaber et al., 2022. This work addresses these limitations by introducing the EM-DL framework, designed to leverage diverse data, improve generalization through data pooling, and deliver transparent, equitable predictions of skin cancer subtypes.

Skin cancer, particularly melanoma, presents a persistent global health challenge due to its clinical complexity, subtle phenotypic variations, and disproportionate outcomes across demographic groups, Esteva, et al., 2017. Traditional diagnostic workflows—largely dependent on clinician expertise and single-modality inputs such as dermoscopic images—often struggle to capture the full biological and sociodemographic variability that defines real-world patient populations, Kedar et al., 2023, Appiahene et al., 2023. according to finfing, disparities in diagnostic accuracy continue to emerge, especially among underrepresented skin types, geographical regions, and age groups.

Prior studies have demonstrated:

- Uni-modal Success: High accuracy in classifying skin lesions using only dermatoscopic images or only electronic health records (EHR) features[6].
- Multimodal Need: Recognition that combining imaging and clinical data (patient history, location) significantly improves complex disease staging[7].
- XAI Integration: The simultaneous demand for XAI to increase clinical trust[8888].

However, few studies have comprehensively integrated multimodal fusion, subtype prediction, and a fairness audit within the context of non-invasive skin cancer diagnostics[9].

Recent advances in Deep Learning (DL) have demonstrated exceptional potential in automating skin cancer detection; however, most existing models remain constrained by uni-domain training, opaque decision pathways, and limited sensitivity to hidden biases embedded in clinical datasets. These limitations hinder both clinical trust and equitable deployment, He et al., 2016,Alwakid et al., 2022.

Cross-Domain Explainable Multimodal Deep Learning offers a transformative paradigm by integrating diverse data sources—dermoscopic images, clinical metadata, histopathology profiles, and population-specific contextual variables—into a unified, transparent diagnostic framework.

By leveraging explainable AI (XAI) mechanisms, the approach not only improves predictive robustness but also reveals the reasoning behind model decisions, enabling clinicians to interrogate and validate outputs. Furthermore, cross-domain alignment ensures that diagnostic performance generalizes effectively across varied skin tones, environmental exposures, and healthcare settings, Lundberg, et al. 2020, Farooq et al., 2025.

This research direction aims to bridge the gap between advanced computational intelligence and equitable medical practice. It advocates for diagnostic systems that are both technically powerful and socially responsible, ensuring that AI-driven skin cancer detection benefits all individuals—irrespective of their background, location, or phenotypic diversity.

Methodology

Data Modalities

The framework uses two primary input streams[10]:

- **Image Data:** Non-invasive images (e.g., dermatoscopic or high-resolution clinical photographs) captured via specialized or standard smartphone cameras[11].
- **Tabular Clinical Data:** Features including age, gender, geographic location, Fitzpatrick skin type, lesion size/location, growth history, and known comorbidities[12].

Centralized Robust Training Architecture

Instead of FL, we utilize a centralized approach focusing on robust data management:

- **Data Aggregation:** Data from multiple clinical centers is anonymized, harmonized, and **pooled** into a single central dataset.
- **Robust Pre-processing:** Techniques are employed to mitigate domain shift (e.g., color augmentation, hardware-specific normalization) introduced by varying acquisition devices across centers.
- **Training:** The model is trained centrally on this comprehensive, large-scale, pooled dataset to maximize exposure to diverse visual and clinical feature variations, enhancing generalizability.

Multimodal Fusion Network

The network utilizes a two-branch architecture[13]:

- **Image Branch:** A pre-trained CNN (e.g., **ResNet-50**) processes the image data[14].
- **Tabular Branch:** A dedicated Multi-Layer Perceptron (MLP) processes the clinical features[15].
- **Fusion Layer:** The feature vectors from both branches are concatenated and passed through a **Transformer encoder block** [16] to capture complex, non-linear cross-modal interactions before the final classification head for subtype prediction (e.g., Melanoma, Basal Cell Carcinoma (BCC), Squamous Cell Carcinoma (SCC)).

Explainability and Fairness Audit

XAI Methods:

- **Grad-CAM (Gradient-weighted Class Activation Mapping):** Applied to the image branch to visualize the regions influencing the prediction (e.g., atypical pigment network or blue-white veil)[17].
- **SHAP (SHapley Additive exPlanations):** Applied to the fused feature space to quantify the individual contribution (positive or negative) of each tabular feature and image feature vector to the final subtype prediction[18181818].

Fairness Audit: We measure **Disparate Impact** (difference in prediction accuracy/sensitivity) across predefined demographic groups (e.g., Fitzpatrick skin type, ethnicity, gender) using a fairness metric such as **Equal Opportunity Difference (EOD)**[19]. The SHAP explanations are then audited to determine if the model relies inappropriately on sensitive attributes (e.g., prioritizing skin type over lesion characteristics) for decision-making[20].

Expected Results and Contribution

- **Superior Subtype Accuracy:** The multimodal fusion is expected to yield significantly higher accuracy ($p < 0.05$) in multi-class skin cancer subtype prediction compared to uni-modal baselines[21].
- **Robust Generalization:** The centralized model trained on pooled, harmonized data is expected to show robust performance when tested on external, independent datasets[22].
- **Actionable Transparency:** SHAP values will reveal the relative clinical importance of features (e.g., growth rate

from history and image-derived boundary features are more critical than age for specific subtype predictions)[23].

- **Bias Identification:** The fairness audit will identify and quantify any prediction disparity, providing the necessary insight for post-processing mitigation or ethical review[24].

This EM-DL framework provides a complete solution for deploying sophisticated, ethical, and transparent AI diagnostics, significantly enhancing diagnostic precision and ensuring equitable healthcare delivery.

Conclusion

This research successfully designed and proposed the **Explainable Multimodal Deep Learning (EM-DL) framework**, a novel solution addressing the critical challenges of accuracy, generalization, and transparency in non-invasive skin cancer diagnostics[25]. By architecturally fusing non-invasive images with comprehensive clinical records via a Transformer-based network, the model moves beyond binary classification to provide **context-aware predictions of cancer subtypes**[26]. Training on a large, pooled dataset enhances the model's robustness and cross-domain applicability. Crucially, the mandatory integration of **XAI techniques (Grad-CAM and SHAP)** alongside a rigorous **Fairness Audit** directly tackles the clinical trust deficit and ethical concerns that have long hampered the deployment of high-stakes AI systems[27]. The EM-DL framework establishes a robust, ethical, and scalable blueprint for the next generation of AI-driven diagnostic tools, significantly enhancing

diagnostic precision and ensuring equitable healthcare delivery worldwide[28].

Future Scope

The successful validation and deployment of the EM-DL framework open several compelling and high-impact avenues for future research:

Dynamic Fairness Interventions and Mitigation Strategies

Future work must move beyond simply auditing bias to implementing and evaluating **dynamic mitigation techniques**[29]. This involves integrating fairness-aware loss functions (e.g., using adversarial debiasing or re-weighting schemes) directly into the centralized training process. Research should focus on quantifying the trade-off between subtype prediction accuracy and group fairness (Equal Opportunity Difference) and establishing optimal operating points for clinical deployment.

Temporal Prediction and Risk Forecasting

The current model provides a cross-sectional diagnosis. A significant advancement would be to adapt the multimodal architecture (leveraging the sequential nature of EHR data) to a Recurrent Neural Network (RNN) or pure Transformer model for **temporal prediction**. This would allow the system to forecast a patient's risk of developing malignancy (e.g., Melanoma) within a defined period, transforming the system from a diagnostic tool into a **preventative risk stratification engine**.

Optimization for Resource-Constrained Environments

While the model is centrally trained, deployment often occurs on edge devices (smartphones, handheld

dermatoscopes). Future studies should focus on **model compression techniques** (e.g., quantization, pruning) and optimizing the multimodal inference pipeline to ensure fast, energy-efficient performance on resource-limited hardware, critical for use in remote clinics.

Integration of Novel Non-Invasive Biomarkers

The current framework relies on clinical images. Future research can expand the input modalities to include novel, non-invasive biomarkers, such as **spectroscopic data** (e.g., reflectance confocal microscopy) or **thermographic analysis**, further enriching the multimodal feature space and potentially enhancing sensitivity for early-stage diagnosis.

Clinical Utility and User Experience Validation

The ultimate measure of success is clinical adoption. Prospective studies are required to validate the EM-DL framework's **clinical utility** by measuring its impact on physician workflow, diagnostic agreement (human vs. AI), and patient outcomes. Furthermore, extensive research into the optimal visualization and presentation of XAI explanations for clinicians is necessary to maximize trust and minimize cognitive overload during real-time decision-making.

References

Abujaber, M., et al. (2022). Artificial Intelligence Models for Predicting Skin Lesion Malignancy based on Accessible Clinical Data. *Journal of Intelligent Systems and Engineering Management*, 101090.

Jain, R., et al. (2019). Non-invasive diagnosis of Skin Lesions using deep learning on Clinical Images. *IET Computer Vision*, 13(4), pp. 321-329.

Esteva, A., et al. (2017). Dermatologist-level classification of skin cancer with deep

neural networks. *Nature*, 542(7639), pp. 115-118.

KedarD, P., et al. (2023). Non-Invasive Detection of Skin Lesions Using Deep Learning on Clinical Images. *IEEE Conference*.

Appiahene, P., et al. (2023). Detection of Melanoma by medical images: a comparative study of machine learning algorithms. *BioData Mining*, 16(2).

He, K., et al. (2016). Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778. (Essential DL Architecture)

Alwakid G, Gouda W, Humayun M, Sama NU.Melanoma Detection and Subtype Estimation from Clinical Images Using Deep Learning: A Scalable Solution. (2022). *Translational Vision Science & Technology*, 14(1):20.

Lundberg, S. M., et al. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2(1), pp. 56-65. (Core SHAP).