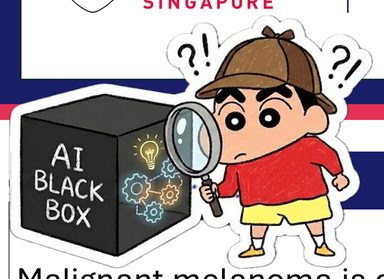
Explainable Melanoma Diagnosis with Contrastive Learning and LLM-based Report Generation

Junwen Zheng[1], Xinran Xu[1], Li rong Wang[1], Chang Cai[1], Lucinda Siyun Tan[2], Wang Ding Yuan [2], Tey Hong Liang [1,2], Xiuyi Fan [1]

1 Nanyang Technological University, Singapore

2 National Skin Centre, Singapore

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE | Centre of AI in Medicine



PURPOSE / OBJECTIVES

Malignant melanoma is clinically high-stakes [1]. While deep models achieve high accuracy, their predictions often lack transparent, criterion-grounded rationales required in dermatology, which limits clinician confidence and real-world uptake [2]. To address this gap, we present CEFM, an explainable framework that anchors predictions in the ABC criteria. CEFM consists of three components:

(i) ABC quantification: quantify asymmetry, border irregularity, and color variation from lesion segmentation.

(ii) Cross-modal alignment: align ABC descriptors with image representations via contrastive learning.

(iii) Report generation: produce a structured diagnostic report for clinician review.

On ISIC2020 datasets, CEFM achieves 92.79% accuracy and 0.961 AUC; dermatologist assessment supports the clinical consistency of the explanations.

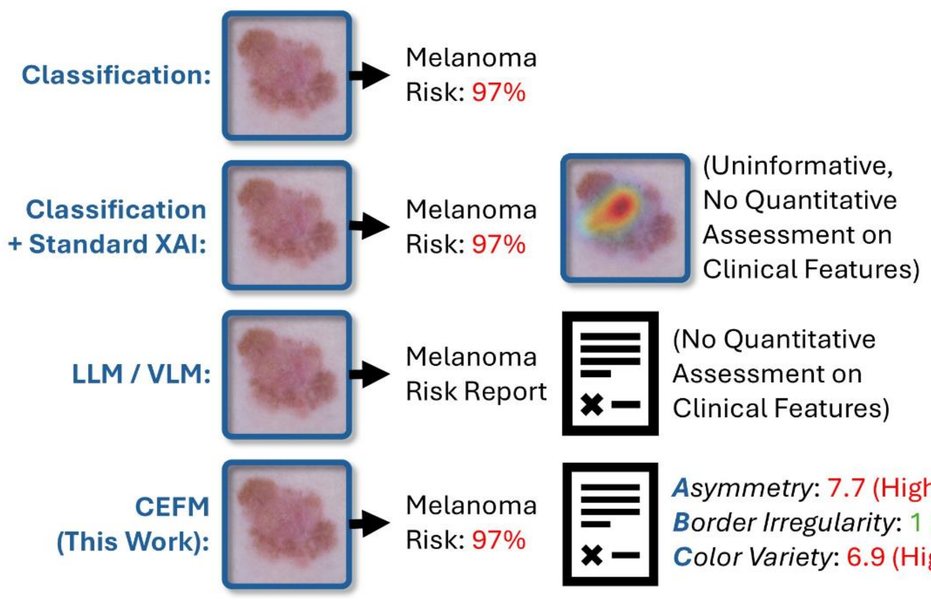
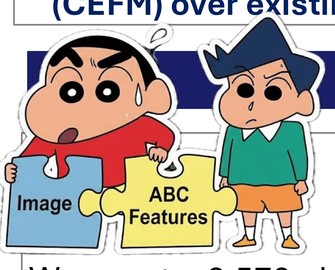


Figure 1: Advantages of the Cross-modal Explainable Framework (CEFM) over existing frameworks.



MATERIAL & METHODS

We curate 3,578 dermoscopic images from ISIC 2018/2020 after quality control. Lesion masks from ISIC 2018 supervise segmentation training. ISIC 2020 is used for classification; lesion masks on ISIC 2020 are obtained by applying the trained segmenter and refined at inference with SAM2. A Vision Transformer (ViT) encodes images. Our contributions are three-fold:

(i) Label-efficient lesion masking that supports ABC evidence extraction on datasets lacking expert masks.

(ii) Cross-modal alignment that maps ViT representations to quantitative ABC descriptors, yielding clinically meaningful embeddings.

(iii) Clinician-readable reporting that translates ABC evidence into structured diagnostic summaries.

Report generation: retrieve concept cues with CLIP and generate structured diagnostic reports using a domain-adapted DeepSeek model.

Fig. 2 details the alignment module; Fig. 3 summarizes the end-to-end pipeline.

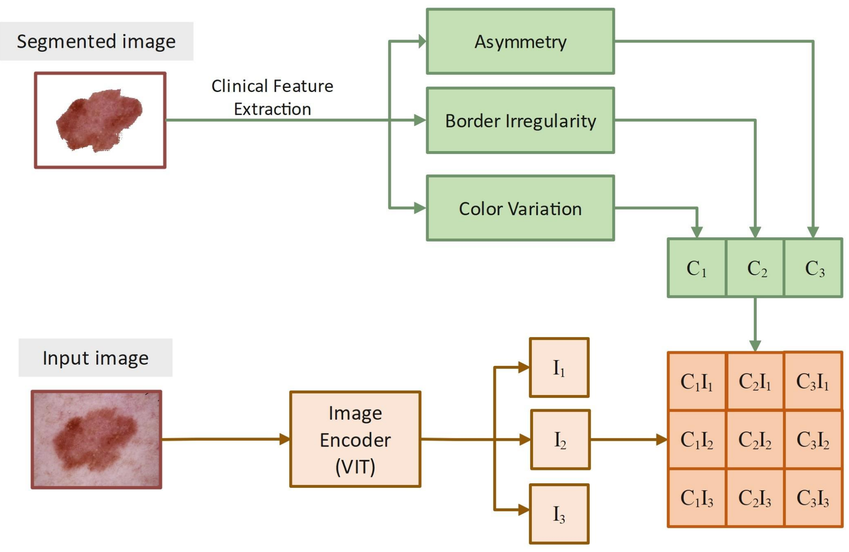


Figure 2: Overview of the cross-modal alignment process.

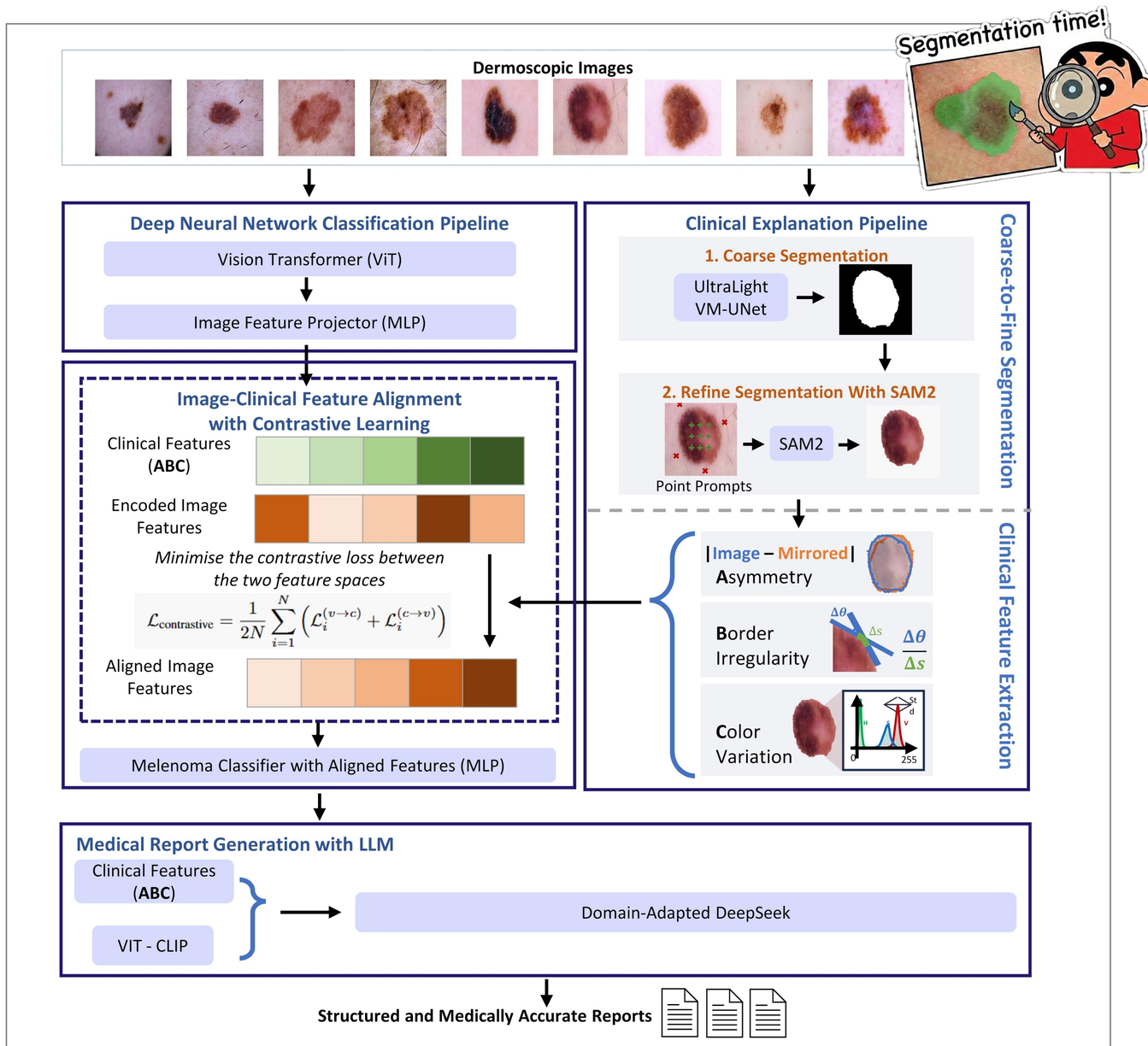


Figure 3. CEFM overview: a ViT encodes dermoscopic images; a coarse-to-fine segmenter (UltraLight VM-UNet + SAM2) yields lesion masks for ABC descriptors; contrastive learning aligns image and ABC embeddings; CLIP concept cues plus a domain-adapted LLM (DeepSeek) generate structured diagnostic reports.

RESULTS

CEFM achieves 92.79% accuracy and 0.961 AUC on ISIC2020. Three board-certified dermatologists evaluated the generated reports, rating interpretability 4.60/5 and consistency with clinical judgment 4.18/5 (Table 1). Experts particularly valued the ABC-based quantitative analysis and the clarity/readability of the structured reports.

Evaluation Dimension	Expert 1	Expert 2	Expert 3	Average Score
1. Consistency with clinical judgment	4.55	4.18	3.81	4.18
2. Usefulness of ABC feature analysis	5.00	4.20	4.00	4.40
3. Report clarity and readability	4.50	4.50	4.50	4.50
4. Interpretability of AI decision-making	5.00	4.40	4.40	4.60
5. Clinical applicability and decision support	4.50	3.50	4.00	4.00

Table 1: Quantitative Expert Evaluation (Likert scale: 1 = strongly disagree; 5 = strongly agree)

Fig. 4 evidences effective contrastive alignment, with positives concentrated at high similarity and negatives near zero.

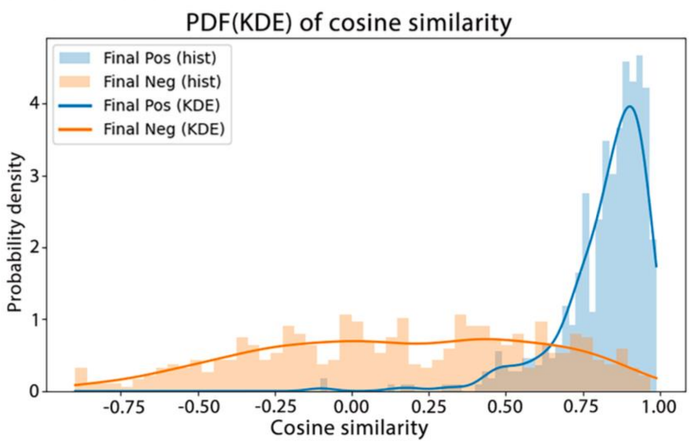


Fig. 4. Aligned cosine similarity

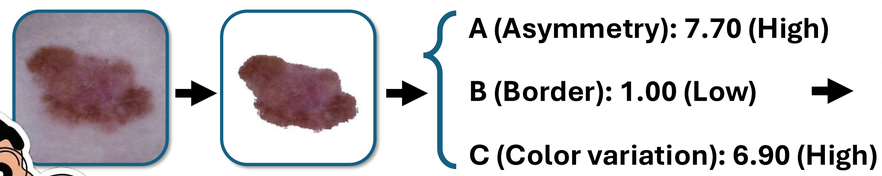


Figure 5. Example output.

SUMMARY/CONCLUSION

We present CEFM, which grounds melanoma predictions in quantitative ABC descriptors by contrastively aligning ABC features with ViT embeddings and translating the aligned evidence into structured diagnostic reports. CEFM achieves 92.79% accuracy and 0.961 AUC, and dermatologist evaluation reports 4.60/5 interpretability and 4.18/5 consistency with clinical judgment. Future work will focus on prospective, multi-center validation and longitudinal lesion tracking.

REFERENCES

[1] Mallardo, D., Basile, D., & Vitale, M. G. (2025). Advances in Melanoma and Skin Cancers. International Journal of Molecular Sciences, 26(5), 1849.

[2] Hossain, M. I., Zamzmi, G., Mouton, P. R., Salekin, M. S., Sun, Y., & Goldgof, D. (2025). Explainable AI for medical data: Current methods, limitations, and future directions. ACM Computing Surveys, 57(6), 1-46.

CONTACT INFORMATION

Corresponding Author: Xiuyi Fan, email: xyfan@ntu.edu.sg