



Collaborative Mixing of Clinical Concept Sources for Explainable Malignancy Prediction in Dermoscopy

Arianna Bunnell^{1,2}, Samuel Yang^{1,2}, Eujin Cho³, Thomas K. Wolfgruber¹, Kevin Cassel¹, John A. Shepherd¹, and Peter Sadowski²
¹University of Hawai'i Cancer Center, ²University of Hawai'i at Mānoa, ³Research Corporation of the University of Hawai'i

Introduction

Artificial intelligence (AI) assistance has been shown to enhance the diagnostic capabilities of physician users. Human-AI collaboration, where both the AI model and physician contribute to decision making, represents one of the most promising paradigms for clinical adoption of medical AI. One way to facilitate collaboration is with concept bottleneck models (CBM) a type of neural network design in which intermediate representations are aligned with human-understandable concepts [1].

We hypothesize that deterministic, computational definitions of clinical concepts provide complementary information to physician-defined concepts, and that combining the two may improve diagnostic performance. We demonstrate this approach on dermoscopic skin cancer classification using the ABCD criteria: Asymmetry (A), Border (B), Color (C), and Differential structures (D) as defined by the International Society of Dermoscopy [2,3]. The ABCD criteria are clinically used for malignancy determination and the decision to biopsy.

Methods

CoMiC is a CBM [1] developed on publicly-available dermoscopy data with new labels in which each concept in the bottleneck layer is a mixture with a learned weighting. **Fig 1.** illustrates the architecture. CoMiC produces a segmentation mask, expert-predicted ABCD concepts, ABC computational concepts, and a final probability of malignancy. The mixture MLP predicts expert weights per concept/example pair. Computational concept definitions from the literature are used for the concept algorithms [4-6]. The mixed ABC concept scores and optional unexplainable concept are passed to the cancer MLP and a final probability of malignancy is computed. No computational definition exists for D, thus $\lambda_D := 1$.

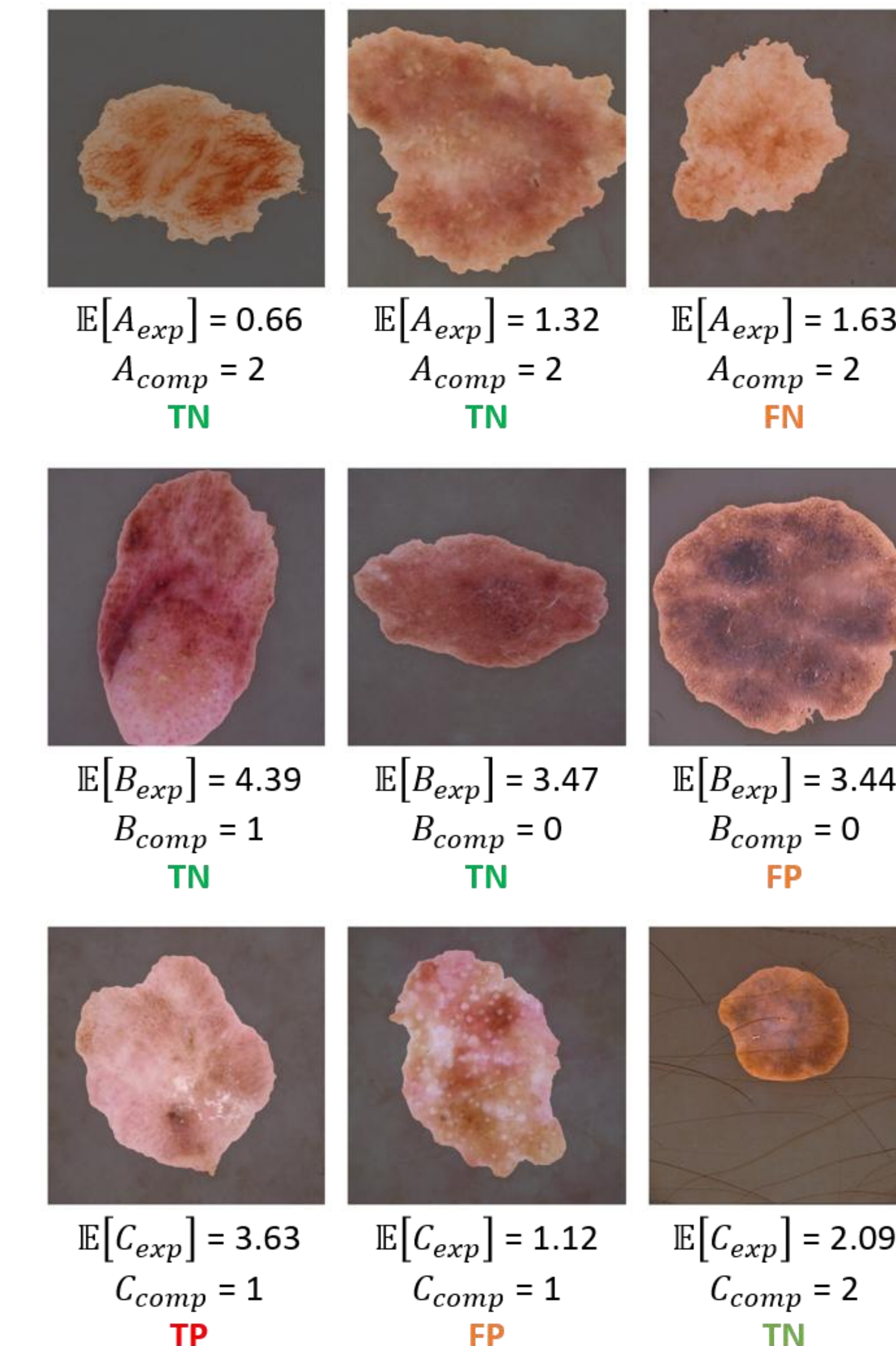
Data

Dermoscopy images and lesion masks were collected from the International Skin Imaging Collaboration (ISIC) challenge images from 2016-2020 [7-12] and the PH2 dataset [13]. ISIC was supplemented with masks from [14]. A random sample of 4,000 ISIC images were annotated with the ABCD criteria by a dermatologist with 17 years of clinical experience. A total of 8,116 images from ISIC and PH2 are included. Of these, 1,920 images have complete mask/ABCD/malignancy labels, 1,952 have ABCD/malignancy labels, 2,471 have mask labels only, and 1,773 have mask/malignancy labels.

Results

Data were randomly split into training (70%), validation (10%), and testing (20%), with no images shared between splits at any training stage. Model hyperparameters were systematically optimized over 40 independent trials. The final model was chosen based on performance on the validation set. CoMiC segmented skin lesions (Stage 1) with Dice Similarity Coefficient 0.89 on the unseen test set. For the expert-labeled concept prediction task (Stage 2), the model performs with AUROC 0.765 (95% CI: 0.746-0.783) for concept A; 0.970 (0.963, 0.977) for concept B; 0.891 (0.881, 0.901) for concept C; and 0.809 (0.797, 0.822) for concept D. When including computational concepts with no unexplainable concept (explainable and collaborative model), CoMiC performed with 0.85 (0.82, 0.87) AUROC in predicting lesion malignancy (Stage 3). When intervention was allowed on concept A to adjust incorrect predictions, CoMiC's diagnostic performance improved to 0.87 (0.84, 0.89) AUROC, matching the unexplainable baseline. See **Figure 2** for example predictions with both high and low expert reliance.

Low λ examples (low expert reliance)



High λ examples (high expert reliance)

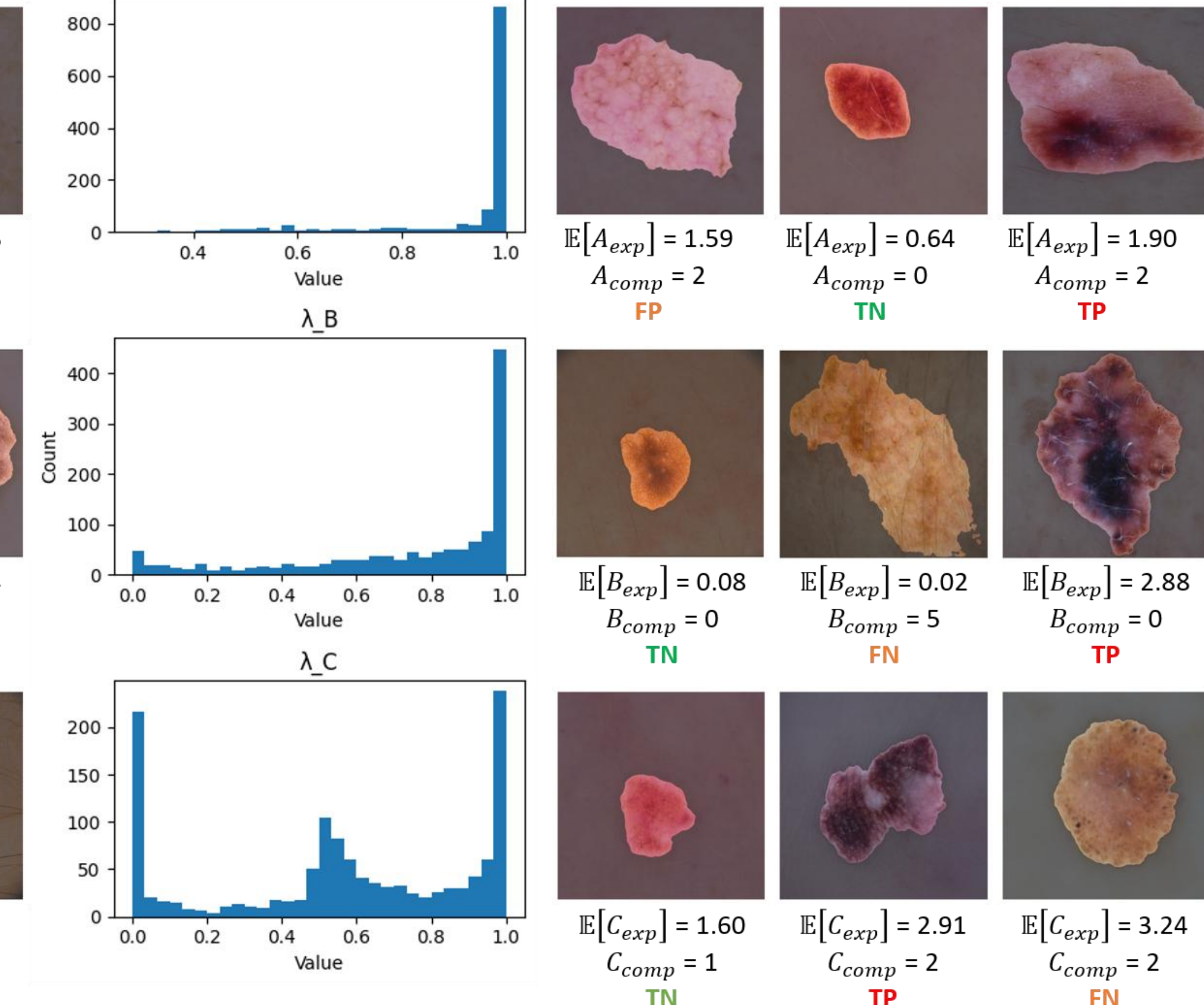


Figure 2. Exemplary images from the testing set with predicted masks overlaid. Their computational concepts, predicted expert concepts, final classification, and ground truth status are displayed.

Conclusion

We propose a new AI model that provides interpretable predictions in terms of clinical concepts with minimal loss of performance. In experiments, we demonstrate its use on a skin cancer prediction task. Furthermore, we demonstrate for the first time that the combination of expert-defined and computationally-derived concepts may increase diagnostic task performance in concept bottleneck models from clinical lexica.

Funding Information

This study was funded through UH CCSG SMART study (P30 2P30CA071789-18) and a UHCC SEED grant.

References

[1] Koh, P.W., et al., Concept Bottleneck Models. (2020). [2] Malvey, J., et al., Dermoscopy report: proposal for standardization: results of a consensus meeting of the International Dermoscopy Society. *Journal of the American Academy of Dermatology*. **57**(1): p. 84-95 (2007). [3] Argenziano, G., et al., Dermoscopy of pigmented skin lesions: results of a consensus meeting via the Internet. *Journal of the American Academy of Dermatology*. **48**(5): p. 679-693 (2003). [4] Stoecker, W.V., et al., Automatic detection of asymmetry in skin tumors. *Computerized Medical Imaging and Graphics*. **16**(3): p. 191-197 (1992). [5] Kasmi, R. and K. Mokrani, Classification of malignant melanoma and benign skin lesions: implementation of automatic ABCD rule. *IET Image Processing*. **10**(6): p. 448-455 (2016). [6] Majumder, S. and M.A. Ullah, Feature extraction from dermoscopy images for melanoma diagnosis. *SN Applied Sciences*. **1**(7): p. 753 (2019). [7] Codella, N., et al., Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). *arXiv preprint arXiv:1902.03368*, (2019). [8] Codella, N.C., et al. Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018). 2018. IEEE. [9] Gutman, D., et al., Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC). *arXiv preprint arXiv:1605.01397*, (2016). [10] Tschandl, P., et al., The HAM10000 dataset, a large collection of multi-source dermoscopic images of common pigmented skin lesions. *Scientific data*. **5**(1): p. 1-9 (2018). [11] Hernández-Pérez, C., et al., Bcn20000: Dermoscopic lesions in the wild. *Scientific Data*. **11**(1): p. 641 (2024). [12] Rotemberg, V., et al., A patient-centric dataset of images and metadata for identifying melanomas using clinical context. *Scientific data*. **8**(1): p. 34 (2021). [13] Mendonça, T., et al., Ph2: A public database for the analysis of dermoscopic images. *Dermoscopy image analysis*. **2** (2015). [14] Tschandl, P., et al., Human-computer collaboration for skin cancer recognition. *Nature Medicine*. **26**(8): p. 1229-1234 (2020).

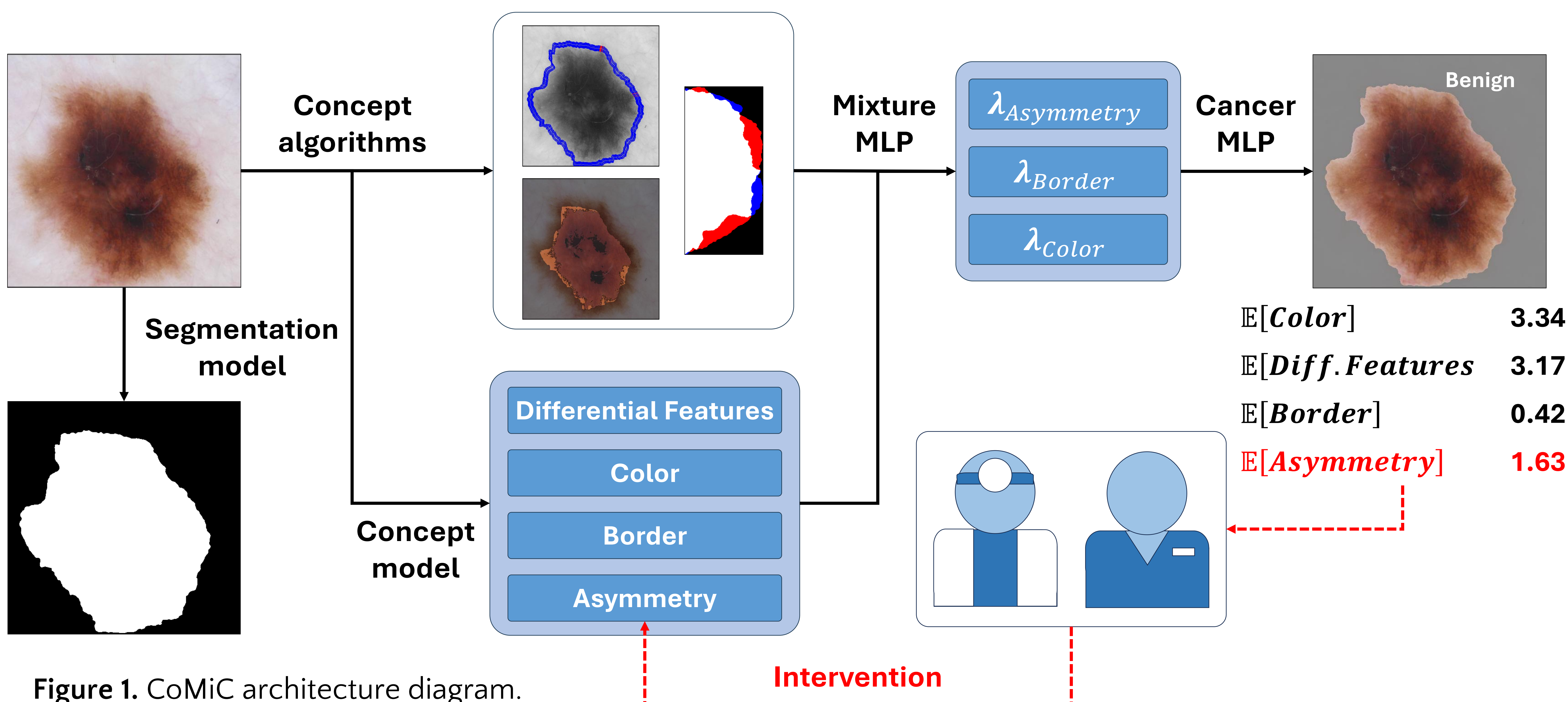


Figure 1. CoMiC architecture diagram.