DERMOSCOPY AND SKIN IMAGING

STUDY TO ASCERTAIN THE LEVEL OF AGREEMENT FOR THE PRESENCE AND LOCATION OF DERMOSCOPIC STRUCTURES AMONG EXPERTS: A PILOT STUDY

Konstantinos Liopyris [(1)] - Cristian Navarrete-dechent [(1)] - Ayelet Rishpon [(1)] - David Swanson [(2)] - Stephen Dusza [(1)] - Michael Marchetti [(1)] - Ashfaq Marghoob [(1)]

Memorial Sloan Kettering Cancer Center, Dermatology, Nyc, United States [(1)] - Mayo Clinic, Dermatology, Scottsdale, United States [(2)]

Introduction / Importance: Dermoscopic criteria have fair to poor interobserver agreement.

Objective: To determine the level of agreement for dermoscopic features among five readers for seven common diagnoses and to explore the reasons for observed areas of discordance.

Materials and Methods / Design, setting and participants: Twenty lesions within each of 7 diagnostic categories (i.e. angioma; dermatofibroma; solar lentigo/ lichenoid keratosis; seborrheic keratosis; nevus; basal cell carcinoma; melanoma) were selected (n=140) from a database of 32000 dermoscopic lesion images (www.isic-archive.com) by two of the authors in consensus. A novel, web-based, annotation tool was used, and each expert was required to exhaustively annotate each lesion on the super-pixel level.

Results: Five readers annotated 140 dermoscopic images for a total of 1895 features. Each reader annotated an average of 2.7 features per lesion and the number of features annotated increased with the complexity of the lesions (i.e. an average of 1.75 features for angiomas vs 4.54 features for melanomas). All five readers observed the same dermoscopic feature 85 times, 4 readers observed the same feature 76 times, 3 readers observed the same feature 102 times, 2 readers observed the same feature 189 times, while single observation of a feature occurred 482 times. Evaluation of annotations at the super-pixel level revealed that overlapping areas of the lesions were annotated by readers using different terms. Furthermore, for melanocytic lesions, there was a 76% positive agreement for the exemplar dermoscopic feature for which the lesion was selected.

Conclusions: The agreement between dermoscopy users for the presence of specific dermoscopic structures is poor. One possible reason for this discordance appears to be related to differing use and or understanding of specific dermoscopic terms. Agreement observed on the super-pixel level can help to expose the redundant terms and lead to a

more cohesive dermoscopic lexicon.