Conditional dependence tests reveal the usage of ABCD rule features and bias variables in automatic skin lesion classification – supplementary material –

Christian Reimers[1,2], Niklas Penzel[1], Paul Bodesheim[1], Jakob Runge[2,3], Joachim Denzler[1,2]
[1] Computer Vision Group, Friedrich Schiller University Jena, Jena, Germany
[2] Institute of Data Science, German Aerospace Center, Jena, Germany
[3] Technische Universität Berlin, Berlin, Germany
{christian.reimers, niklas.penzel, paul.bodesheim, joachim.denzler}@uni-jena.de,
jakob.runge@dlr.de

Appendices

A. Ranges for the colors of the color score

To automatically determine the color features described in Section 4.3, we need to assign color values to each of the six colors. Since the recording process introduces noise and the visual perception changes depending on the environment, it is not optimal to assign an exact value. Hence, we assign ranges to every color. We base these color ranges on the colors provided in the dermoscopedia[1]. Since this feature is designed to be visually evaluated by a dermatologist, we define them in the HSV color-space as it is closer to the human perception than the RGB color space. Table 7 lists the exact ranges.

B. Skin lesion classification performance

This section describes the performances of the two models we analyze in this work.

Test-time augmentation models [28] We train the models of Perez *et al*. [28] with different backbone architectures as the authors did on the ISIC 2017 dataset. The performance scores in Table 4 are evaluated on the corresponding test set. We report similar performance scores as Perez *et al*. on the melanoma classification problem. The models trained to classify seborrheic keratosis perform even better.

EfficientNet ensemble [12] The ensemble consisting of five EfficientNet (B0) classifiers tries to differentiate between eight classes. It achieves a training accuracy of 0.541. The AUC scores for the different classes are listed in Table 5.

[1] https://dermoscopedia.org/ABCD_rule

C. MNIST classifier

To calculate the feature named "MNIST class" we use a classifier for hand-written digits trained on the MNIST dataset [20]. The architecture of this classifier can be found in Table 6. The classifier was trained for 50 epochs using stochastic gradient descent with a learning rate of 0.01, Polyak-momentum [29] of 0.9, and a weight decay of 0.0005. It reaches an accuracy of 0.9939 on the MNIST test set.

D. Skin color feature

For the skin color feature we extract the top left ten by ten area of the image. We perform a PCA, such that every example $\mathbf{x}$ can be expressed as

$$x = \sum_{i=1}^{300} a_i \mathbf{e}_i \tag{5}$$

where $\mathbf{e}_i$ is the i-th principle component. We use a_1 as the score for this feature. Some examples of the extracted pixel areas can be found in Figure 2.

E. Colorful patch examples

Figure 3 displays some examples of the first 10K images we downloaded from the ISIC archive. In Section 5, we analyze if classification models rely on the occurrence of colorful patches.

Table 4: Ranges corresponding to the individual colors named in the Color score of the ABCD-rule.

Color	Hue	Saturation	Value
White	0° - 360°	0% - 10%	90% - 100%
Red	350° - 25°	70% - 100%	80% - 100%
Light Brown	0° - 40°	50% - 100%	70% - 90%
Dark Brown	11° - 47°	60% - 100%	40% - 60%
Gray-Blue	216° - 252°	30% - 100%	60% - 100%
Black	0° - 360°	0% - 100%	0% - 15%

Figure 2: Examples of top left ten by ten pixel areas ordered by the weight of the first principal component.

Table 5: Performance of the Perez *et al*. models [28] on the test set of the ISIC 2017 challenge [8]. We report the Area Under the Receiver Operating Characteristic curve (AUC) and the accuracy (ACC).

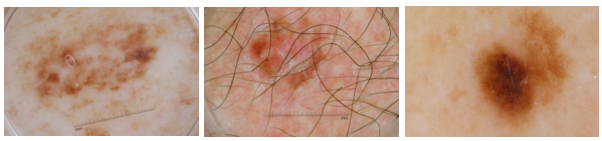
	Melanoma		Seborr. Keratosis	
Backbone	AUC	ACC	AUC	ACC
ResNet152 [15]	0.886	0.868	0.938	0.918
Inception-v4 [42]	0.851	0.835	0.924	0.887
DenseNet161 [16]	0.877	0.852	0.944	0.913

Table 6: Area Under the Receiver Operating Characteristic curve (AUC) from an ensemble of five EfficientNets (B0) for different classes.

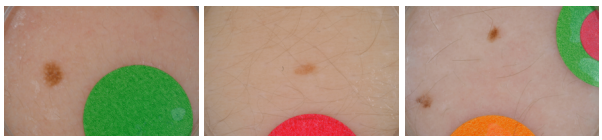
Class	AUC score
Melanoma	0.802
Melanocytic Nevus	0.884
Basal Cell Carcinoma	0.895
Actinic Keratosis	0.889
Benign Keratosis	0.819
Dermatofibroma	0.844
Vascular Lesion	0.938
Squamous Cell Carcinoma	0.882

Table 7: The architecture of the MNIST classifier used to calculate a feature in Section 4.2.

Layertype	Filter Size	Filters	Padding
Convolutional	5×5	32	2×2
Convolutional	3×3	32	1×1
Max Pooling	2×2	-	2×2
Convolutional	3×3	64	1×1
Convolutional	3×3	64	1×1
Convolutional	3×3	128	1×1
Convolutional	3×3	10	1×1
Global Average Pooling	-	-	-



(a) melanoma



(b) benign nevi

Figure 3: Some images from the first 10K images of the ISIC archive. All images containing colorful patches similar to the examples in (b) are benign nevi in children from the SONIC dataset [37].