In computer vision, deep learning has reached state-of-the-art performance in many classification tasks. When it comes to skin lesion classification, deep neural networks are able to match dermatologist performance in a single image classification setting. A shortcoming, however, is that these networks only output the class likelihood, which is no substitute for model (un)certainty. To overcome this, we approximate statistical inference using a technique called “Monte Carlo dropout”, which we extend to obtain a metric for output confidence. We apply this to skin lesion classification.

We evaluate our technique on the HAM10000 data set. We hold out a test set, and use the remainder of the data to train a convolutional neural network. For each image in our test set, we pass 200 duplicates through the network, while applying Monte Carlo dropout, to obtain a probability distribution for each class. To evaluate the confusion between the two most likely classes we calculate the Bhattacharyya coefficient between their distributions. The result is a value between 0 and 1, where 0 indicates a high confidence, and 1 stands for a low confidence.

Future work includes further validation of this technique, more focused on the clinical application in dermoscopy. Therefore, we have developed an online evaluation tool (https://lesion-toolkit.intec.ugent.be/) where dermatologists can classify a series of skin lesions. Besides asking for a classification, the toolkit also queries the dermatologist’s classification confidence. We aspire to compare the classifications made by our convolutional neural network, along with its confidence, to these made by dermatologists.