Technical Report

Department of Computer Science
and Engineering
University of Minnesota
4-192 Keller Hall
200 Union Street SE
Minneapolis, MN 55455-0159 USA

TR 18-014

Computer Aided Diagnosis of Skin Lesions from Morphological Features

Nicholas Heller, Erika Bussmann, Aneri Shah, Joshua Dean, Nikolaos Papanikolopoulos

August 24, 2018

Computer Aided Diagnosis of Skin Lesions from Morphological Features

Nicholas Heller, Erika Bussman, Aneri Shah, Joshua Dean,
and Nikolaos Papanikolopoulos

Computer Science and Engineering, University of Minnesota – Twin Cities
{helle246, deanx252, papan001}@umn.edu
{ecbussmann, anerishah00}@gmail.com

Abstract. Skin cancer is the most common cancer, accounting for over 40% of all cancer cases. The morphological features of skin lesions are an integral component of skin cancer detection and diagnosis. With the rapid progress in the field of image classification, increasing attention has been put towards the Computer Aided Diagnosis of skin lesions based on their morphological features. The International Skin Imaging Collaboration (ISIC) Archive is the largest publicly available collection of dermoscopic images of skin lesions, and in 2018 ISIC hosted an image recognition challenge for dermoscopic images. This short paper describes the CAD system that we submitted to this challenge.

1 Background

Skin cancer is the most common form of cancer, accounting for more than 40% of all cancer cases worldwide [2]. Skin cancer is primarily caused by exposure to ultraviolet radiation from the sun [7], but other risk factors include tobacco use, HPV, and artificial UV radiation (e.g. from tanning beds) [15].

Cancerous skin lesions are primarily detected based on their morphological presentation, but a biopsy is usually required to make a diagnosis. Dermoscopy is an imaging technique in which the lesion is submerged in a liquid medium such that images of the lesion can be captured in the absence of skin surface reflections. Research has shown that expert dermatologists make more reliable and more accurate diagnosis based on dermoscopic images than they do with standard photography [12].

The ISIC Archive is the largest public collection of dermoscopic images of skin lesions in the world. In 2018 ISIC hosted an image recognition challenge in which competitors were to train a dermoscopic image classifier using the HAM 10,000 dataset to recognize one of 7 classes: melanoma, melanocytic nevus, basal cell carcinoma, actinic keratosis, benign keratosis, dermatofibroma, and vascular lesion [18] (see figure 1. This challenge is a follow-up of a similar challenge last year, hosted in conjunction with the 2017 International Symposium on Biomedical Imaging [5].

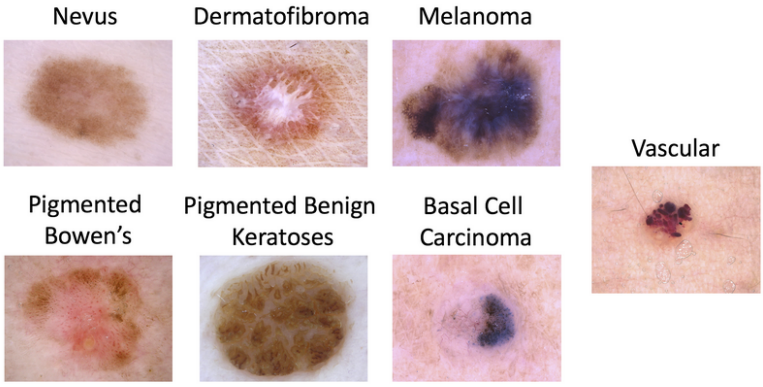

Fig. 1. An example from each of the 7 classes in the HAM10000 dataset [18]

2 Methods

We applied an ensemble of Deep Neural Networks (DNNs) [13] to this task. 10,000 images is a relatively small dataset for training DNNs on a difficult classification problem such as this so we heavily relied on transfer learning [14] and data augmentation [19].

2.1 Models

Our ensemble consisted of five DNNs: ResNet 50 v1 [9], Inception v3 [17], Xception [4], DenseNet 201 [10], and InceptionResNet v2 [16].

Our CAD system is built on top of Keras [3] with a TensorFlow [1] backend. We used the `keras.applications` implementations of each of the DNNs in our ensemble and fine-tuned their available ImageNet [6] initializations. We found the `tf.data` API exceptionally useful for creating a performant data pipeline.

2.2 Training

Our training procedure was identical for each model. First, we removed the final fully connected layer of each pretrained network and replaced it with a randomly initialized (Glorot Uniform [8]) matrix. We penalized the fully connected layer with both an L2 and an L1 regularizer, each weighted at 0.01. We then optimized these regularization terms along with a categorical softmax cross entropy term weighted at 1.0 with the Adam optimizer [11]. All weights were optimized at all steps of training. Our initial learning rate was 10^{-4} and when the validation accuracy did not improve for 10 steps, we scaled the learning rate by a factor of 0.1. The remaining parameters of the Adam algorithm were left at their default values. We used the largest batch sizes that would fit on a NVIDIA Titan X GPU with each model.

Each of the models we optimized took input tensors with spatial sizes of either 299×299 (Inception, Xception, and InceptionResNet) or 224 (Resnet and DenseNet). For evaluation and prediction, we first scaled the images to either 329×329 or 254×254, cropped the central region of the appropriate size, and fed it to each model's respective preprocessing function.

For training, we first scaled each image to normally distributed random dimensions centered at 30 more than the desired dimensions with a standard deviation of 15. We then randomly rotated the image using bilinear interpolation, randomly cropped a region, flipped vertically and or horizontally, and randomly adjusted the brightness and contrast with a max change of 30% in both cases. Finally we fed this perturbed image to each model's preprocessing function before training. These random transformations were performed at each training step.

Finally, we further augmented our training set with all of the remaining images from the ISIC archive whose "diagnosis" tag appeared to fit under any of our 7 classes.

2.3 Predictions

The final predictions we submitted were the result of averaging the softmax layer from each network for each testing instance.

The code used for preparing the dataset and training our model can be found at https://github.com/neheller/isic18.

3 Results

At the time of writing our performance on the challenge hold out set is unknown, but we report here our results on a held-out 10% of the provided training set. We evaluated our models after each 200 training steps using 20 random batches of class-balanced validation data. Top overall validation accuracy reached by each of the models during training is shown in table 1.

Model	Accuracy %
Inception	87.7
Xception	86.2
Inception-ResNet	87.8
ResNet	85.2
DenseNet	88.2
Ensemble	85.2

Table 1. Peak validation accuracy for each model during training

We used weights from peak validation accuracy in our ensemble. **It should be noted that this procedure will bias our validation results upward**, therefore, we expect performance on the challenge hold-out set to be slightly worse. Reporting an unbiased estimate of performance metrics is impossible without a third "testing" set, and we decided that would be a waste of training data since the model will be evaluated on a true hold out set by the challenge. We report validation precision and recall for each class in tables 2 and 3.

	Melanoma	Melanocytic Nevus	BCC	Actinic keratosis	Benign keratosis	Dermatofibroma	Vascular Lesion
Inception	0.649	0.957	0.809	0.735	0.818	0.850	0.952
Xception	0.591	0.959	0.847	0.672	0.794	0.882	0.952
Inception-ResNet	0.573	0.966	0.921	0.760	0.793	0.933	0.870
ResNet	0.689	0.947	0.873	0.667	0.818	0.882	0.875
DenseNet	0.587	0.967	0.763	0.632	0.754	0.895	0.909
Ensemble	0.709	0.965	0.822	0.725	0.827	0.944	0.909

Table 2. Precision of each model by class

	Melanoma	Melanocytic Nevus	BCC	Actinic keratosis	Benign keratosis	Dermatofibroma	Vascular Lesion
Inception	0.743	0.923	0.923	0.720	0.818	0.944	0.909
Xception	0.760	0.901	0.923	0.780	0.794	0.833	0.909
Inception-ResNet	0.802	0.892	0.897	0.760	0.861	0.778	0.909
ResNet	0.689	0.939	0.885	0.680	0.842	0.833	0.955
DenseNet	0.749	0.891	0.910	0.720	0.800	0.944	0.909
Ensemble	0.772	0.930	0.949	0.740	0.867	0.944	0.909

Table 3. Recall of each model by class

As can be seen, despite its relatively high incidence in the training set, melanoma is one of the most challenging for the model to identify, and it along with Actinic Keratosis are among the lowest in both precision and recall.

The confusion matrix in Fig. 2 shows that outside of the confusion between melanoma and melanocytic nevus, actinic keratosis was often mistaken for basal cell carcinoma and benign keratosis.

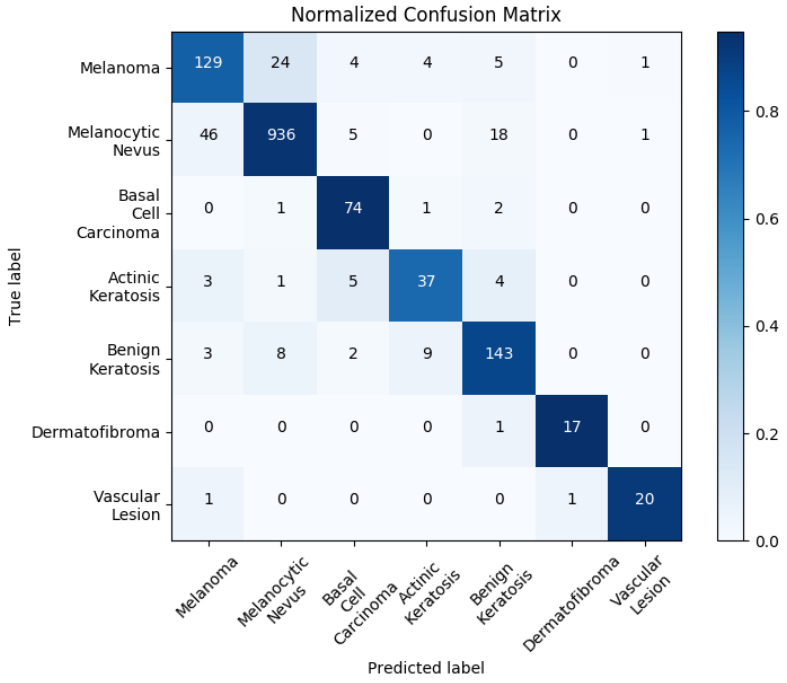

Fig. 2. Normalized (color intensity) confusion matrix of the ensemble on the validation set [18]

Acknowledgements

Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under Award Number R01CA225435. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X.: TensorFlow: Large-scale machine learning on heterogeneous systems (2015), https://www.tensorflow.org/, software available from tensorflow.org

2. Cakir, B.Ö., Adamson, P., Cingi, C.: Epidemiology and economic burden of non-melanoma skin cancer. Facial plastic surgery clinics of North America **20**(4), 419–422 (2012)
3. Chollet, F., et al.: Keras. https://keras.io (2015)
4. Chollet, F.: Xception: Deep learning with depthwise separable convolutions. arXiv preprint pp. 1610–02357 (2017)
5. Codella, N.C., Gutman, D., Celebi, M.E., Helba, B., Marchetti, M.A., Dusza, S.W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H., 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). In: Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on. pp. 168–172. IEEE (2018)
6. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. pp. 248–255. Ieee (2009)
7. Gallagher, R.P., Lee, T.K., Bajdik, C.D., Borugian, M.: Ultraviolet radiation. Chronic Diseases and Injuries in Canada **29** (2010)
8. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics. pp. 249–256 (2010)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
10. Iandola, F., Moskewicz, M., Karayev, S., Girshick, R., Darrell, T., Keutzer, K.: Densenet: Implementing efficient convnet descriptor pyramids. arXiv preprint arXiv:1404.1869 (2014)
11. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
12. Kittler, H., Pehamberger, H., Wolff, K., Binder, M.: Diagnostic accuracy of dermoscopy. The lancet oncology **3**(3), 159–165 (2002)
13. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. pp. 1097–1105 (2012)
14. Pan, S.J., Yang, Q., et al.: A survey on transfer learning. IEEE Transactions on knowledge and data engineering **22**(10), 1345–1359 (2010)
15. Saladi, R.N., Persaud, A.N.: The causes of skin cancer: a comprehensive review. Drugs of Today **41**(1), 37–54 (2005)
16. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A.: Inception-v4, inception-resnet and the impact of residual connections on learning. In: AAAI. vol. 4, p. 12 (2017)
17. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2818–2826 (2016)
18. Tschandl, P., Rosendahl, C., Kittler, H.: The ham10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions. arXiv preprint arXiv:1803.10417 (2018)
19. Van Dyk, D.A., Meng, X.L.: The art of data augmentation. Journal of Computational and Graphical Statistics **10**(1), 1–50 (2001)