Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Effects of objects and image quality on melanoma classification using Deep Neural Networks

Bilge Süheyla Akkoca-Gazioğlu (bilgeakkoca@gmail.com)
Istanbul Teknik Universitesi https://orcid.org/0000-0002-0640-815X
Mustafa Kamasak
Istanbul Teknik Universitesi

Research

Keywords: melanoma classification, deep learning performance, image quality

Posted Date: June 17th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-35907/v1

License: 🅭 🅯 This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Version of Record: A version of this preprint was published at Biomedical Signal Processing and Control on May 1st, 2021. See the published version at https://doi.org/10.1016/j.bspc.2021.102530.

RESEARCH

Effects of objects and image quality on melanoma classification using Deep Neural Networks

Bilge S Akkoca-Gazioğlu[*†] and Mustafa E Kamasak

[*]Correspondence:
bakkoca@itu.edu.tr
Department of Computer Engineering, Istanbul Technical University, Istanbul, Turkey
Full list of author information is available at the end of the article
[†]Equal contributor

Abstract

Background: Melanoma is a type of skin cancer with a higher mortality compared to other types of skin cancers. Early and accurate diagnosis of melanoma has critical importance on its prognosis. Recently, deep neural network based models dominated the CAD systems for classification of the potential melanoma lesions. In clinical settings, capturing impeccable skin images is not always possible. In some cases, an external object such as a ruler is required for determination of lesion size. Sometimes, the skin images can be blurry, noisy or have low contrast. The aim of this work is to investigate the effects of external objects (ruler, hair) and image quality (blur, noise, contrast) on the classification of melanoma using commonly used Convolutional Neural Network(CNN) models.

Results: Performance is analyzed using accuracy, sensitivity, specificity and precision metrics over 6 different test sets. Hair set has 89.22%, ruler set has 86% and none set has 88.81% as the best accuracy with DenseNet121 architecture. Also, DenseNet has the best average accuracy with comparing the other three models in other datasets, which are noise and blur. We find that ResNet is better for contrast dataset. We can infer that DenseNet can be used for melanoma classification with image distortions and degradations.

Conclusion: In this study, we investigate the effect of ruler/hair and image blur, noise and contrast on the melanoma detection performance of four commonly used CNN models: ResNet50, DenseNet121, VGG16 and AlexNet. Melanoma images can be better recognized under contrast changes unlike the benign images, we recommend ResNet model whenever there is contrast issue. Noise significantly degrades the performance on melanoma images and the recognition rates decrease with compared to benign lesions in noisy set. DenseNet121 also works well in this set. Both classes are sensitive to blur changes and best accuracy is obtained with DenseNet model. The images contain ruler has decreased the classification accuracy and ResNet has better performance if there is ruler in an image. Hairy images have the best success rate in our system since it has the maximum number of images in total dataset. DenseNet performs better for both hairy and high quality images.

Keywords: melanoma classification; deep learning performance; image quality

Background

Malignant melanoma is one of the most rapidly increasing cancers in the world [1]. Despite its low prevalence (5%), it has higher mortality rate compared to other types of skin cancers.Just like many other types of cancer, early diagnosis of melanoma is very important for an effective treatment and to avoid poor prognosis. Computer aided diagnosis (CAD) systems can be used to assist physicians and act as a second opinion.

A CAD system for melanoma detection is proposed in many studies. Typically, images are collected using dermatoscopy, lesions are segmented and classified as melanoma or not. The studies differ in their methods to process, segment and classify the lesions.

Lesion segmentation is the primary step for melanoma detection. A typical dermatoscopy image has multiple lesions and suspected regions of interest. Each of these regions has to be processed differently using their characteristics such as border regularity, asymmetry and size of the lesion.

The segmentation of the lesions can be performed via various segmentation algorithms. However with the widespread adoption of deep learning techniques, almost all recent studies employ deep neural network based models for lesion segmentation.

Xie et al. used an algorithm to generate high resolution feature maps which included spatial and discriminative feature of boundaries and obtained segmented lesions with the success rate of 85.7% on PH^2 dataset by using FCN-8 and U-Net CNN models [2]. In another segmentation study by Goyal et al. an ensemble of two different CNN models -Mask RCNN and Deep LabV3- was developed and tested on ISIC datasets. They obtained 93.8% segmentation accuracy [3]. CNN based models were used very commonly for skin lesion segmentation [4, 5]. In a study by Jafari et al., skin images were divided into patches that were sent to a CNN model. CNN model extracted some local and global features from patches and detects the borders of lesion. Their best accuracy was 98.7% [4]. Yuan worked with a few samples of skin images for segmentation and FCN were used in their study. They took the problem as pixel-wise classification instead of segmentation and proposed a novel loss function uses Jacard distance, they achieved the best accuracy with 95.5 % [5].

Besides segmentation deep learning algorithms become very popular in the classification of lesions [6, 7, 8]. Albahar used a novel regularizer, which was used to penalize the weight matrix and to control the complexity of model. This regularizer was embedded to each layer of the network. Average skin lesion classification accuracy was calculated as 97.49% on ISIC dataset [6]. Attention residual learning was proposed for classification by Zhang et al [7]. They obtained attention maps to lower layers from the feature maps in higher layers by using ARL-CNN model on the ISIC dataset. Their average classification performance was 91.7%. Hekler et al. classified the skin images from ISIC dataset by fusing the classification results of 112 dermatologists and a trained CNN. They achieved 82.95% accuracy with fusion method [8]. In addition to these studies, many systematic reviews about lesion classification problem were recently published [9, 10, 11]. In these three review studies, performances of both traditional classifiers (ie. SVM, KNN, Naive Bayes and Decision Trees) and CNN model were investigated for classification of melanoma.

Training of deep learning models require large amounts of labeled data. When the amount of labeled data is low, models that are trained on other images are used as baseline model. These models are then fine tuned for the specific purpose using transfer learning methods. In Harangi ensembles of DNN were used for skin lesion classification [12]. In this study, ISBI 2017 Challenge dataset was used with three classes: nevus, melanoma and seborrheic keratosis. Training images were augmented by rotation, flipping and cropping operations. They recommended to use of fusing four different CNN models: AlexNet, ResNet, VGGNet and GoogLeNet. The fusion of the outputs on the classification layer was handled and the average accuracy was calculated as 0.866 after fusion model with weighting.

In another study by Hosny et al., classification of skin lesion were performed using transfer learning and AlexNet [13]. They used three different datasets, which are MedNode, DermIS- DermQuest and 2017 ISIC Challenge. Their augmentation technique was based on image rotation. They rotated images in two ways. One way was arbitrary rotation and the other was systematic rotation (between 0°-360°with regular intervals of 5°). They used modified AlexNet for transfer learning, and changed the classification layer with softmax layer. Stochastic gradient Descent algorithm was used with a small learning rate to update weights of the CNN model. Their system performance was measured as 0.969, 0.977, 0.959 for Derm, MedNode and ISIC datasets respectively.

Mahbod et al. fused deep features for classification of skin lesion [14]. In this study, ISIC Challenge 2017 dataset was used. First steps in their pipeline were image preprocessing, image normalization, and resizing. After that, four CNN models; AlexNet, VGG16, ResNet18 and ResNet101, were trained for feature extraction. The features were used as the outputs of fully connected (FC) layers of the fine tuning networks. They added two FC layers to the networks as replacement of the last FC and output layer. After extracting features, SVM classifier were trained for three classes. The best performance was obtained with fusion of all fine-tuned networks. The accuracy was 87.26% and 95.52% for melanoma and seborrheic keratosis respectively.

Perez et al. used a CNN model for melanoma classification [15]. In this study, many experiments were conducted to demonstrate the performance of different CNN architectures for melanoma classification. They also used ISIC Challenge 2017 dataset and 9 different CNN models. The models were DenseNet, Dual Path Nets, Inception-V4, Inception-ResNet-v2, MobileNetV2, PNASNet, ResNet, SENet, and Xception. They performed the experiments 3 times on 5 test splits with 9 architectures. They experimented with architectures of the models such as initialization of last layer, augmentation, dropouts, data shuffling. These experiments revealed how they effect the system performance.

ISIC dataset which is commonly used in this problem has high quality images. However, the images taken by clinicians may not have such good quality due to imaging system, illumination and other issues in clinical environments. As a result, poor quality images with problems like brightness, blur, and noise is acquired. Some of these problems cannot be alleviated using image processing techniques. In image recognition field, there are various studies about effects of image distortions on the output [16, 17, 18, 19, 20, 21]. They analyzed the strengths and weaknesses of CNN models against image degradation. There are also a few studies about effects of image quality in skin lesion classification. One study investigate the CNN performance under different noise distortions in the skin images [22]. In another study, the researchers used low quality images that are degraded with blurring, noise compression and channel errors to investigate the system performance [23].

In this study, we analyzed how the classification performance was affected by degradation and defects in lesion images. We considered contrast, noise and blur distortions and additional external objects such as hair and ruler. For this aim, we obtained new test sets and classify the images by using commonly used deep learning architectures for medical image analysis. We used four different CNN architectures which were commonly used for melanoma classification problem: AlexNet [24, 25, 26], VGG16 [25, 27, 28], ResNet50 [28, 29, 30], DenseNet [31, 32]. We investigated how these CNN models perform under image degradation and defects.

Results

We trained our models until training loss is lower than 10^{-6}. It is the early stopping technique to avoid overfitting. After training, we saved the models and test sets are tested with same models for each CNN architecture. We evaluated the results of the trained models with different metrics which were commonly used in classification problems. These metrics are as follows:

$$
\begin{aligned}
Accuracy &= \frac{t_p + t_n}{t_p + f_p + f_n + t_n} \\
Sensitivity &= \frac{t_p}{t_p + f_n} \\
Specificity &= \frac{t_n}{f_p + t_n} \\
Precision &= \frac{t_p}{t_p + f_p}
\end{aligned}
$$

where t_p and t_n mean true positive/negative and f_p and f_n mean false positive/negative respectively.

We divided our test sets into overlapping 10 folds. All metrics were calculated for each fold and we averaged of ten folds were reported per each test set. Figure 1, 2 and 3 show that the performance metric results that were calculated over the noisy, blurry and contrasted image datasets, respectively.

Figure 1 Graphics of metrics over noisy images.

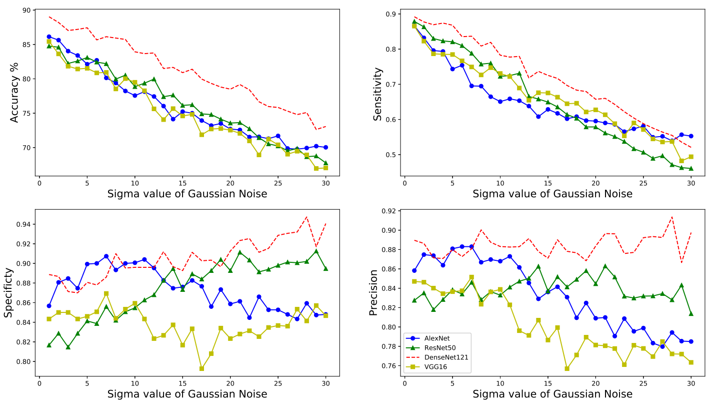


Figure 2 Graphics of metrics over blurry images

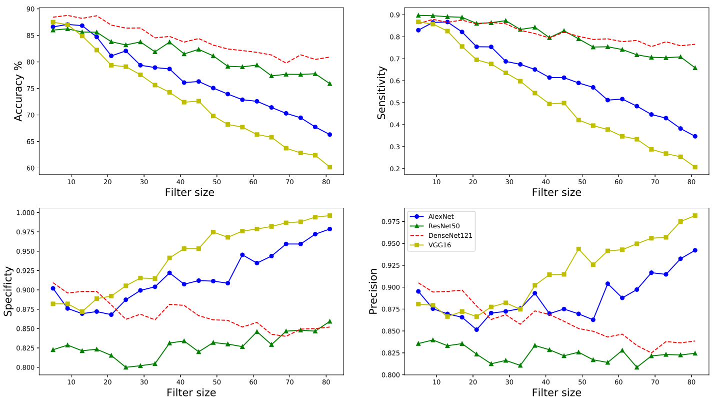
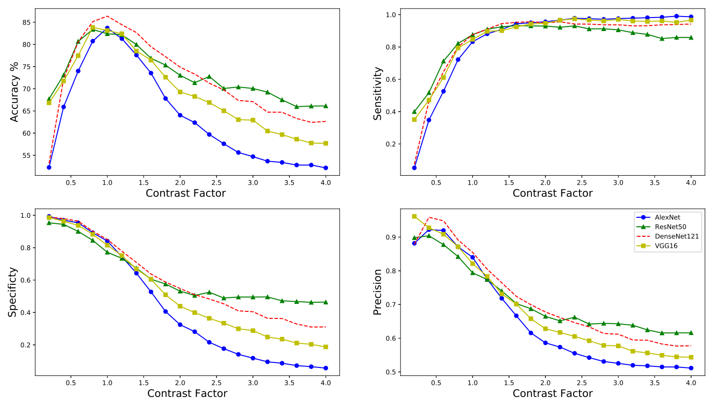


Figure 3 Graphics of metrics over contrasted images



In Table 1, 2 and 3, we calculated the accuracy, sensitivity, specificity, precision and true_predicted-all metrics over 3 test sets: hair, none and ruler respectively by using 4 different CNN models. Accuracies over three test sets are shown in Figure 4.

Table 1 Performance metrics of the models for $Test\_Hair$ dataset

CNN Models	Accuracy	Sensitivity	Specificity	Precision	True-Predicted/All
DenseNet121	89.22	0.8955	0.8883	0.9002	2275/2550
ResNet50	87.96	0.8681	0.8925	0.9008	2243/2550
AlexNet	84.51	0.8688	0.9133	0.8676	2155/2550
VGG16	86.2	0.8177	0.9116	0.9123	2198/2550

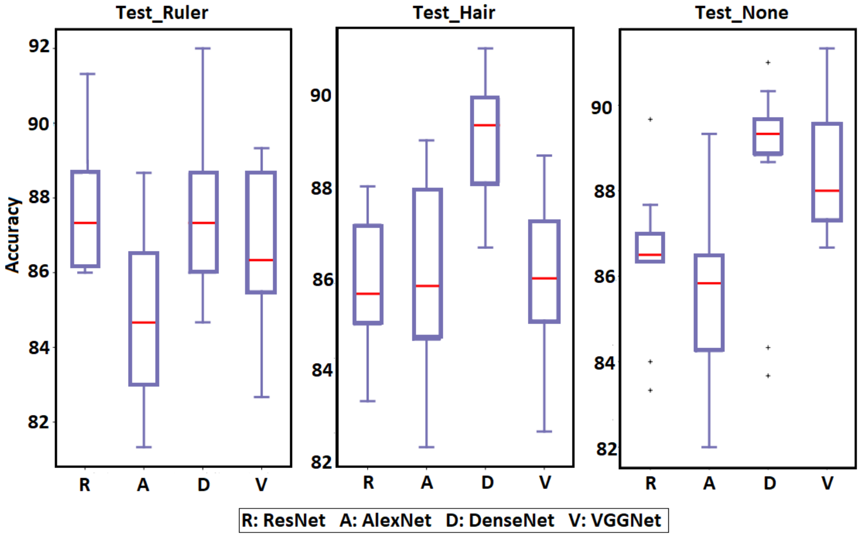
Table 2 Performance metrics of the models for $Test\_None$ dataset

CNN Models	Accuracy	Sensitivity	Specificity	Precision	True-Predicted/All
DenseNet121	88.81	0.9209	0.8552	0.8641	1865/2100
ResNet50	86.95	0.9142	0.8247	0.8391	1826/2100
AlexNet	84.95	0.8657	0.8333	0.8391	1784/2100
VGG16	86.57	0.8161	0.9152	0.9059	1818/2100

Table 3 Performance metrics of the models for $Test\_Ruler$ dataset

CNN Models	Accuracy	Sensitivity	Specificity	Precision	True-Predicted/All
DenseNet121	86	0.9100	0.8171	0.8100	559/650
ResNet50	85.23	0.8933	0.8171	0.8072	554/650
AlexNet	80.92	0.8466	0.7771	0.7650	526/650
VGG16	81.69	0.9433	0.7085	0.7350	531/650

Figure 4 Accuracies of deep neural network models for test sets that are calculated over 3 datasets.



It is crucially important to predict diseases correctly. Therefore, we measured True Negative Ratio (TNR) on benign lesions and True Positive Ratio (TPR) on melanoma images separately for four models. It is given in Table 4.

$$TPR = \frac{t_p}{t_p + f_n}$$

$$TNR = \frac{t_n}{f_p + t_n}$$

Table 4 Accuracy of the models for all test sets for both classes.

	Benign TNR				Melanoma TPR			
Test Sets	ResNet50	AlexNet	DenseNet121	VGG16	ResNet50	AlexNet	DenseNet121	VGG16
Ruler Set	**87.73**	77.6	83.33	83.6	87.73	**92**	91.73	89.86
None Set	90.73	84.8	90.86	**91.53**	82.13	**86.2**	**86.2**	85.33
Hair Set	**93.66**	88.13	93	89.53	78.06	83.73	**85.06**	82.46
Blur Set (9x9)	87.6	82.86	**89.6**	88.2	86.53	**89.6**	87.93	85.8
Blur Set (77x77)	97.2	84.66	85	**99.4**	38.26	70.86	**75.93**	25.4
Noise Set ($\sigma = 2$)	88.06	82.86	**88.66**	85	83.2	86.33	**87.73**	82.26
Noise Set ($\sigma = 28$)	85.93	90.2	**94.73**	84.13	53.93	47.13	**55.46**	53.73
Contrast Set (c = 0.4)	97	94.4	**97.93**	96.33	34.8	**51.73**	46.53	47.13
Contrast Set (c = 3.8)	6.53	**46.26**	31	20.33	**99.06**	85.93	93.86	95.2

Discussion and Conclusions

In this study, we analyzed the effects of image degradation and external objects on skin lesion classification performance. We used the datasets from ISIC Challenges, and we split the dataset for training and test purposes. We formed 3 subsets for testing that are ruler (Test_Ruler), hair (Test_Hair) and lesion (Test_None: images with lesion only). Then, we applied Gaussian kernels to obtained blurred images by using 20 different σ values, added zero mean Gaussian noise with 30 different σ values. In addition, we modified the contrast of the images with a contrast factor varies from 0.2 to 4.0. We used four pretrained CNN architectures namely, AlexNet, DenseNet121, ResNet50 and VGG16. We replaced the last FC layer with our classifier. It consisted a FC layer with one input layer with 1000 neurons, one hidden layer with 40 neurons and one output layer with 2 neurons. After the training with four different models, we tested the system with 6 different test sets.

The classification ratio of malignant melanoma increases until the contrast factor get the value of 1, that is, as the contrast value of the image increased, in cases where the value was greater than 1, it continued to perform well and did not deteriorate. On the contrary, classification of benign lesions continued to decrease. Due to the fact that, we obtained high accuracy over the dataset of malignant melanomas with grater values of the contrast factor, we can easily observe that melanoma lesions were not badly affected by contrast. When we looked at overall success for both classes, ResNet50 worked better than other CNN models in changes of contrast factor.

In experiments with different noise levels, it was seen that the melanoma images were more sensitive to noise. Despite this, benign recognition had a much better performance compared to melanoma. We can say that the best of the 4 models was DenseNet121on noise dataset.

When we used the VGG16 and AlexNet models for the classification of skin lesions in blur datasets, we get rapidly decreasing success rate in melanoma lesions however increasing accuracy for the benign lesions. Nevertheless, the performance decreased in melanoma class, while there was an oscillating graphic (means that the accuracy sometimes decreased and sometimes increased) in the benign lesions by using in DenseNet121 and ResNet50 models. In this situation, false negative results were increasing while the false positives are decreasing. We considered that the models evolved towards making benign biased decisions because of the increasing information loss in the corners of the skin lesion images. The accuracies were at acceptable levels for both classes. As a result, the best classification ratio was measured with DenseNet as overall system performance.

The ruler in the images has decreased the accuracy. We can explain that the number of ruler images were less than the other in both training and test sets. So, the learning part of rulers were not sufficient for this set. The classification ratio of malignant melanoma was better than benign. ResNet50 performed better with compared to other models.

In hairy images, the system produced more successful results than other datasets. Because of the more hairy images, the better learning of hairy parts in the images by deep learning models. DenseNet had higher accuracy for this test set.

We expect that the highest accuracy was obtained over hairy image set but we measured better accuracy of Test_Hair than Test_None. Wing to the fact that, there were more hairy images compared to mole images. Moreover, we can interpret the results such that wrinkles or various patterns on the skin were also remarkable for classification. DenseNet had achieved better performance for this test set as well.

Consequently, considering 6 test sets, DenseNet had a better classification accuracy regarding image degradation problems.

In this study, we dealt with some issues and the most important one is overfitting problem in training of the networks. We performed many strategies to eliminate overfitting of data. Our strategies that we used were as follows:

- Loss functions: We tried two different loss functions: Negative Log Likelihood loss function and Cross Entropy Loss Functions.
- Optimizer: ADAM and Stochastic Gradient Descent (SGD) optimizers are handled for the experiments.
- Weight decay: The weight decay parameter is set from 0.1 to 0.9, and we do our experiments with 9 different values of this parameter.
- Early stopping: The training loss is calculated after each epoch and if the loss is lower than a predefined value, the training is cut at that epoch.
- Learning rate: Different values of learning rate parameter are set for the training.

After trying all solutions, we obtained the minimum training loss with early stopping technique. We set the expected loss as $1e^{-}6$. If the calculated loss is lower than $1e^{-}6$, we stopped the training at that point.

Methodology

Dataset

As many other studies, we also used ISBI Challenge datasets that were released in 2017, 2018 and 2019 [33, 34, 35, 36]. In 2018 and 2019, the labels of training part of the dataset were given to the competitors, whereas the labels for the test and validation parts were not released. Hence, training, validation and test sets of 2017 and training sets of 2018 and 2019 are used for the experiments. We combined all images in these three datasets. After that, we split the dataset according to the class labels. There were labels for 9 and 7 classes in 2019 and 2018 datasets respectively. In the 2017 dataset, there were only seborrheic keratosis and melanoma classes.

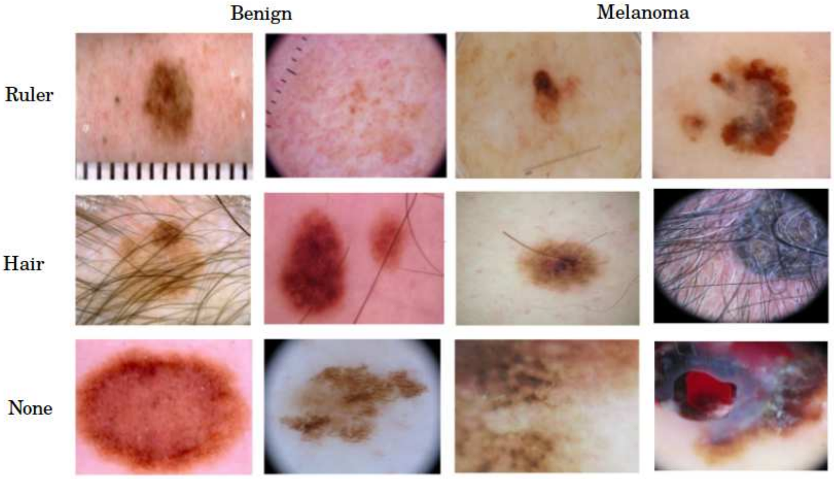
In this study, we dealt with only two classes benign and melanoma. Hence, lesions labeled as melanocytic nevus and benign keratosis and seborreheic keratosis classes were considered as benign. We divided the images in 2 classes and split the dataset into training and test partitions for both classes. We produced our partitions at approximately 80-20% ratio for training and test respectively.

In order to analyze the effects of image external objects, we formed 3 test subsets; $Test_{Ruler}, Test_{Hair}, Test_{None}$. However, training dataset had all images with and without hair and ruler. An artificial ruler was added to some of the images which are then added to $Test_{ruler}$s set. For $Test_{Hair}$ dataset, we selected images that include any hair. It can be a few or more hairs. $Test_{None}$ set included only the lesion without hair or ruler. We discarded the images have both ruler and hair to understand the effects of the separately. The number of images per classes in the sets is shown in Table 5 and some sample images from the test sets are shown in Figure 5.

Table 5 The number of images per classes in the sets.

	Benign	Melanoma
Training	11400	3500
Test_Ruler	350	150
Test_Hair	1200	270
Test _None	1050	350

Figure 5 Sample images from 3 test sets.



Data augmentation

There is imbalance in numbers of images in the classes as seen in Table 5. Therefore, we apply data augmentation processes on melanoma images in all sets to make the image numbers equal in the benign and melanoma classes. Rotation and flipping were used for augmentation. We rotated the images 45°, and flipped the images in horizontal, vertical and both axis.

Test sets with degraded images

Image processing algorithms can have different performances under different conditions such as illumination, noise etc. We considered that Deep learning techniques, Convolutional Neural Networks can also be affected from these conditions. The quality of the images may adversely affect classification. Hence, we wanted to analyze the effects of using synthetically degraded images in melanoma classification. We used different levels of blur, noise and contrast degradation. We used the images in the $Test_{None}$ dataset for both classes.

Blurry image dataset

Blur is one of the main problems in image processing researches. It makes the classification harder, so many studies use deblurring algorithms as preprocessing. In this study, we did the opposite, we applied the Gaussian filters to our images. We change the filter sizes and obtain different blurring levels. For this aim, we changed the value of σ from 1 to 20 and we calculated the filter size width and height according to the formula of $w, h = 2(2\sigma) + 1$. The filter size varied from (5 x 5) to (81 x 81) and we obtained 20 test sets of different levels of blur.

Noisy image dataset

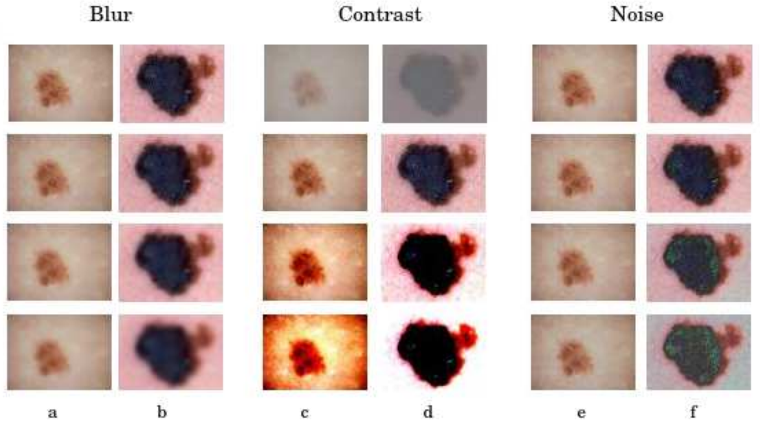
Noise is inherent in all images with different levels. Hence, we add Gaussian distributed noise with zero mean and various scales (standard deviations) σ. We considered noise with scale (σ) varied from 1 to 30. Gaussian noise were added to all channels of skin lesion images separately. As a final step, image pixel values were normalized to 0- 255 range by MinMax Normalization algorithm. We obtained 30 different probe sets to analyze the effects of noise in melanoma classification.

Contrasted image dataset

The images with different contrasts may also have different impact on classification performance. For this aim, we adjusted the contrast in this part of the study using a contrast factor. Contrast factor is a parameter where 0 means no contrast (all gray) and 1 means original contrast level. We obtained 20 different sets for the values of contrast factor from 0.2 to 4 increments of 0.2.

Some samples of artificially degraded images are shown in Figure 6. Also, number of images after data augmentation are given in Table 6.

Figure 6 Sample images from synthetic image test sets.



a,c, e columns are benign; b,d,f columns are melanoma images
For Blur set, the images in the rows are blurred by 5x5, 21x21, 41x41, 81x81 filters, respectively.
For Contrast set, the images in the rows are enhanced contrast by factors of 0.2, 1, 2, 4, respectively.
For Noise set, the images in the rows are added noise with sigma = 1, 10, 20, 30 respectively.

Table 6 The number of images per classes in the sets after data augmentation.

Sets	# of Benign	# of Melanoma
Training_Set	11400	14000
Test_Ruler	350	300
Test_Hair	1200	1350
Test_None	1050	1050
Test_Blurred	1050	1050
Test_Noisy	1050	1050
Test_Contrasted	1050	1050

After specifying training and test sets, we applied pre-processing steps to our images. Before training the networks, we first resized the images (224 x 224) as expected input size of the CNN Models. We also normalized the images with a predefined function in Pytorch library [37]. This function normalizes each channel (RGB) of the image with the following formula

$$x(c,i) = \frac{x(c,i) - \mu_c}{\sigma_c}$$

where $x(c,i)$ is the image pixel i in channel c ($c \in \{R,G,B\}$), μ_c and σ_c are the mean and scale of channel c respectively. μ_c and σ_c are computed using all images in ISIC datasets.

Deep CNN models

ResNet

ResNet model has a different structure from the traditional sequential network architectures and was developed by He et al. [38]. ResNet differs from other architectures with its micro architectural module structure. Changes between some layers ignored and it can be preferred to switch to the lower layer. With this property, ResNet architecture has better performance rate.In addition to this layer structure, we can find the information about the connections between layers. There are various ResNet models: ResNet50 consists of 177 layers with over 23 million trainable parameters.

DenseNet

DenseNet can be summarized as Dense Transform Network that connects each layer to all other layers in the feed forward style. It was developed by Huang et al. [39]. It requires fewer parameters than traditional networks; because it is not necessary to relearn the feature maps. DenseNet layers are very narrow (for example, 12 feature maps per layer), it adds a small set of feature maps to the network's collective information and keeps the remaining feature maps unchanged. The final classifier decides the class of input based on all feature maps on the network. One advantage of DenseNet is advanced information and gradient flows that makes it easier to train across the network. Each layer has direct access to gradients from the loss function and the original input signal, which leads to an implicit deep control. This helps the training of deeper network architectures. Moreover, intensive connections have a regular effect that reduces excessive compliance with tasks with smaller training set sizes.

AlexNet

It is one of the first studies to provide convolutional neural network and deep learning models to become popular in 2012. This network is developed by Alex Krizhevsky, Ilya Sutskever and Geoffrey Hinton [40]. There are basically consecutive layers of convolution and pooling. ReLU (Rectified Linear Unit) is used as activation function, and max-pooling is used in pooling layers. This deep convolutional neural network consists of 25 layers, 5 convolution layers, 3 max-pooling layers, 2 dropout layers, 3 fully connected layers,7 ReLU layers, 2 normalization layers, softmax layer, input and classification (output) layers. Approximately 60 million parameters are calculated during the learning process.

VGGNet

Vgg16 is a simple architecture network. It is developed by Simonyan and Zisserman in 2014[41]. Vgg16 has 13 convolutions 3 fully connected layers. There are 41 layers including Maxpooling, fully connected, ReLu, dropout and softmax layers in total. Approximately 138 million parameters are calculated. It has a structure in traditional sequential network architecture.

We removed the last fully connected layer of all networks and added our custom classifier. This classifier had a fully-connected network, which had 1000 input neurons and a custom hidden layer which has 40 input neurons, also the output neuron number was set to 2 due to the binary nature of classification. We also defined the activation function as ReLU and a dropout was set to 0.2. After that, we defined the loss function and optimizer in our network. We used NLL (Negative log-likelihood) Loss Function and ADAM optimizer for this problem. We also set the learning rate as 0.0001. After all definition, the model was prepared for training.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable

Author's contributions

BSAG: responsible for the design and implementation of research, and wrote the main body of the manuscript. MK: supervised the work and interpreted the results. All authors read and approved the final manuscript.

Acknowledgements

Not applicable

References

1. American cancer society: Survival rates for melanoma skin cancer, 2019
2. Xie, F., Yang, J., Liu, J., Jiang, Z., Zheng, Y., Wang, Y.: Skin lesion segmentation using high-resolution convolutional neural network. Computer Methods and Programs in Biomedicine **186**, 105241 (2020)
3. Goyal, M., Oakley, A., Bansal, P., Dancey, D., Yap, M.H.: Skin lesion segmentation in dermoscopic images with ensemble deep learning methods. IEEE Access (2019)
4. Jafari, M.H., Nasr-Esfahani, E., Karimi, N., Soroushmehr, S.R., Samavi, S., Najarian, K.: Extraction of skin lesions from non-dermoscopic images for surgical excision of melanoma. International journal of computer assisted radiology and surgery **12**(6), 1021–1030 (2017)
5. Yuan, Y., Chao, M., Lo, Y.-C.: Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance. IEEE transactions on medical imaging **36**(9), 1876–1886 (2017)
6. Albahar, M.A.: Skin lesion classification using convolutional neural network with novel regularizer. IEEE Access **7**, 38306–38313 (2019)
7. Zhang, J., Xie, Y., Xia, Y., Shen, C.: Attention residual learning for skin lesion classification. IEEE transactions on medical imaging **38**(9), 2092–2103 (2019)
8. Hekler, A., Utikal, J.S., Enk, A.H., Hauschild, A., Weichenthal, M., Maron, R.C., Berking, C., Haferkamp, S., Klode, J., Schadendorf, D., *et al.*: Superior skin cancer classification by the combination of human and artificial intelligence. European Journal of Cancer **120**, 114–121 (2019)
9. Marka, A., Carter, J.B., Toto, E., Hassanpour, S.: Automated detection of nonmelanoma skin cancer using digital images: a systematic review. BMC medical imaging **19**(1), 21 (2019)
10. Brinker, T.J., Hekler, A., Utikal, J.S., Grabe, N., Schadendorf, D., Klode, J., Berking, C., Steeb, T., Enk, A.H., von Kalle, C.: Skin cancer classification using convolutional neural networks: systematic review. Journal of medical Internet research **20**(10), 11936 (2018)
11. Pathan, S., Prabhu, K.G., Siddalingaswamy, P.: Techniques and algorithms for computer aided diagnosis of pigmented skin lesions—a review. Biomedical Signal Processing and Control **39**, 237–262 (2018)

12. Harangi, B.: Skin lesion classification with ensembles of deep convolutional neural networks. Journal of biomedical informatics **86**, 25–32 (2018)
13. Hosny, K.M., Kassem, M.A., Foaud, M.M.: Classification of skin lesions using transfer learning and augmentation with alex-net. PloS one **14**(5) (2019)
14. Mahbod, A., Schaefer, G., Ellinger, I., Ecker, R., Pitiot, A., Wang, C.: Fusing fine-tuned deep features for skin lesion classification. Computerized Medical Imaging and Graphics **71**, 19–29 (2019)
15. Perez, F., Avila, S., Valle, E.: Solo or ensemble? choosing a cnn architecture for melanoma classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 0–0 (2019)
16. Dodge, S., Karam, L.: Understanding how image quality affects deep neural networks. In: 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX), pp. 1–6 (2016). IEEE
17. Pei, Y., Huang, Y., Zou, Q., Zhang, X., Wang, S.: Effects of image degradation and degradation removal to cnn-based image classification. IEEE transactions on pattern analysis and machine intelligence (2019)
18. Roy, P., Ghosh, S., Bhattacharya, S., Pal, U.: Effects of degradations on deep neural network architectures. arXiv preprint arXiv:1807.10108 (2018)
19. Zhou, Y., Song, S., Cheung, N.-M.: On classification of distorted images with deep convolutional neural networks. In: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1213–1217 (2017). IEEE
20. Ghosh, S., Shet, R., Amon, P., Hutter, A., Kaup, A.: Robustness of deep convolutional neural networks for image degradations. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2916–2920 (2018). IEEE
21. Grm, K., Štruc, V., Artiges, A., Caron, M., Ekenel, H.K.: Strengths and weaknesses of deep learning models for face recognition against image degradations. Iet Biometrics **7**(1), 81–89 (2017)
22. Fan, X., Dai, M., Liu, C., Wu, F., Yan, X., Feng, Y., Feng, Y., Su, B.: Effect of image noise on the classification of skin lesions using deep convolutional neural networks. Tsinghua Science and Technology **25**(3), 425–434 (2019)
23. Elmahdy, M.S., Abdeldayem, S.S., Yassine, I.A.: Low quality dermal image classification using transfer learning. In: 2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), pp. 373–376 (2017). IEEE
24. Kawahara, J., BenTaieb, A., Hamarneh, G.: Deep features to classify skin lesions. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), pp. 1397–1400 (2016). IEEE
25. Mahbod, A., Schaefer, G., Wang, C., Ecker, R., Ellinge, I.: Skin lesion classification using hybrid deep neural networks. In: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1229–1233 (2019). IEEE
26. Codella, N.C., Nguyen, Q.-B., Pankanti, S., Gutman, D.A., Helba, B., Halpern, A.C., Smith, J.R.: Deep learning ensembles for melanoma recognition in dermoscopy images. IBM Journal of Research and Development **61**(4/5), 5–1 (2017)
27. Lopez, A.R., Giro-i-Nieto, X., Burdick, J., Marques, O.: Skin lesion classification from dermoscopic images using deep learning techniques. In: 2017 13th IASTED International Conference on Biomedical Engineering (BioMed), pp. 49–54 (2017). IEEE
28. Harangi, B.: Skin lesion detection based on an ensemble of deep convolutional neural network. arXiv preprint arXiv:1705.03360 (2017)
29. Matsunaga, K., Hamada, A., Minagawa, A., Koga, H.: Image classification of melanoma, nevus and seborrheic keratosis by deep neural network ensemble. arXiv preprint arXiv:1703.03108 (2017)
30. González-Díaz, I.: Incorporating the knowledge of dermatologists to convolutional neural networks for the diagnosis of skin lesions. CoRR (2017)
31. Gessert, N., Sentker, T., Madesta, F., Schmitz, R., Kniep, H., Baltruschat, I., Werner, R., Schlaefer, A.: Skin lesion diagnosis using ensembles, unscaled multi-crop evaluation and loss weighting. arXiv preprint arXiv:1808.01694 (2018)
32. Li, K.M., Li, E.C.: Skin lesion analysis towards melanoma detection via end-to-end deep learning of convolutional neural networks. arXiv preprint arXiv:1807.08332 (2018)
33. Tschandl, P., Rosendahl, C., Kittler, H.: The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific data **5**, 180161 (2018)
34. 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: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pp. 168–172 (2018). IEEE
35. Codella, N., Rotemberg, V., Tschandl, P., Celebi, M.E., Dusza, S., Gutman, D., Helba, B., Kalloo, A., Liopyris, K., Marchetti, M., 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)
36. Combalia, M., Codella, N.C., Rotemberg, V., Helba, B., Vilaplana, V., Reiter, O., Halpern, A.C., Puig, S., Malvehy, J.: Bcn20000: Dermoscopic lesions in the wild. arXiv preprint arXiv:1908.02288 (2019)
37. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: Pytorch: An imperative style, high-performance deep learning library. In: Wallach, H., Larochelle, H., Beygelzimer, A., d' Alché-Buc, F., Fox, E., Garnett, R. (eds.) Advances in Neural Information Processing Systems 32, pp. 8024–8035. Curran Associates, Inc., ??? (2019). http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf
38. 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)
39. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708 (2017)

40. 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)
41. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)

Figures

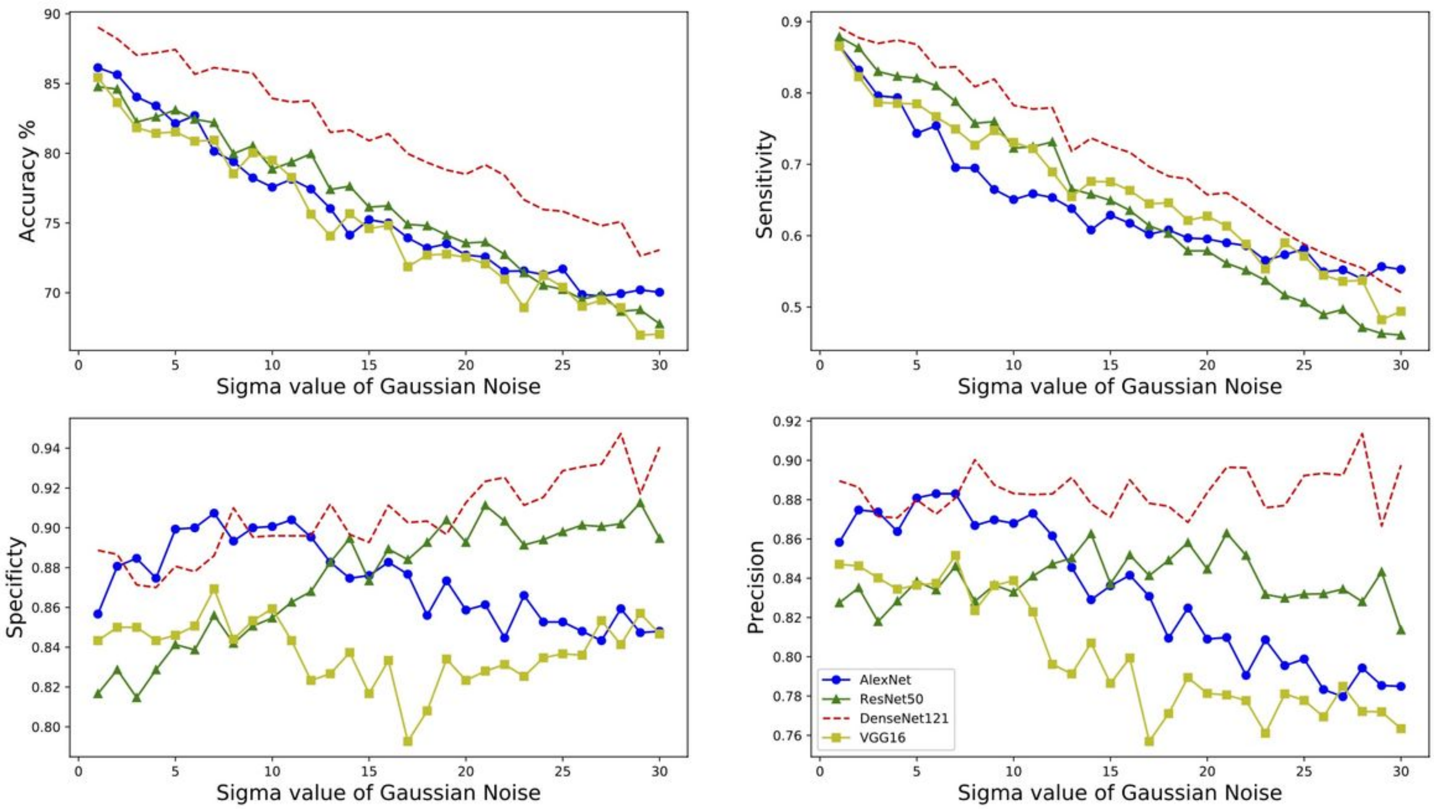


Figure 1

Graphics of metrics over noisy images.

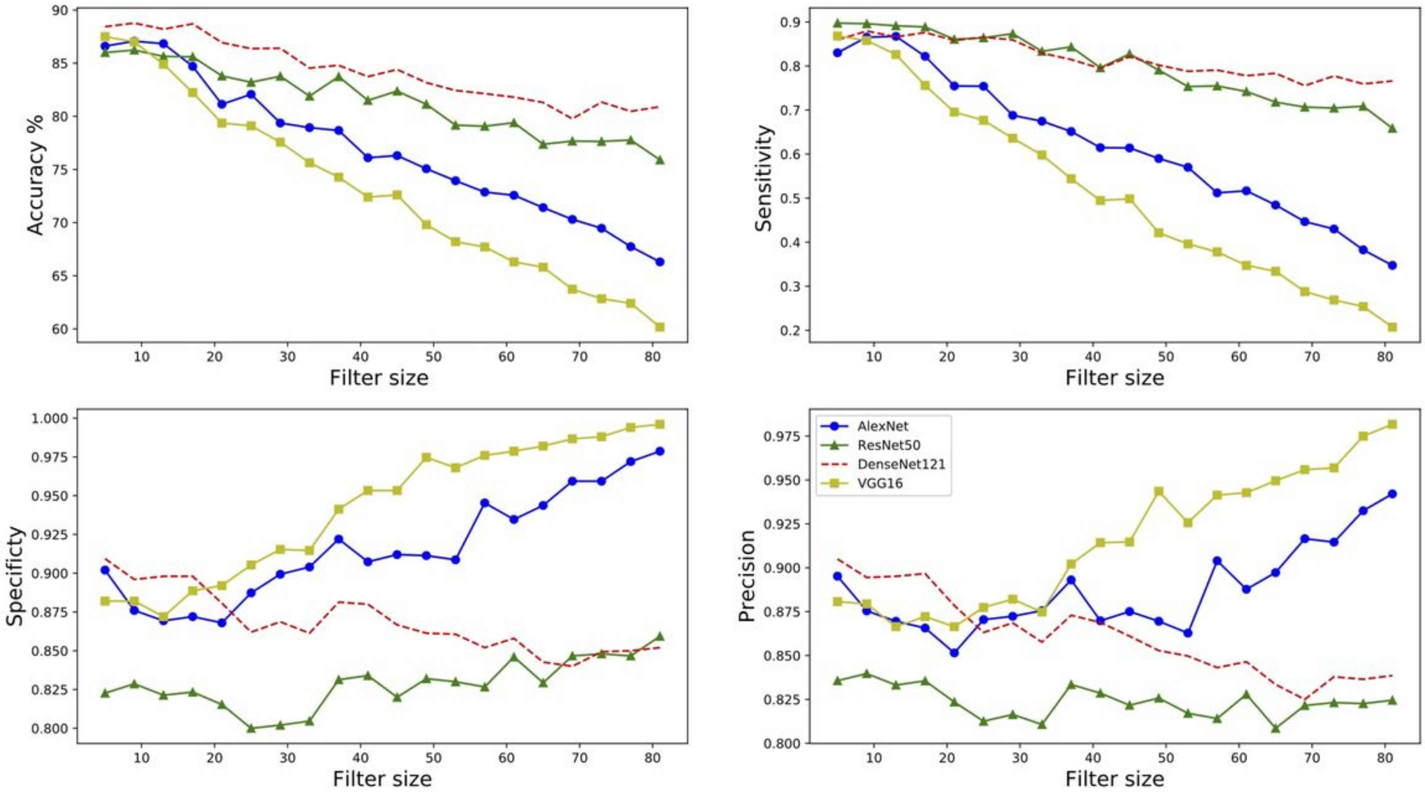


Figure 2

Graphics of metrics over blurry images

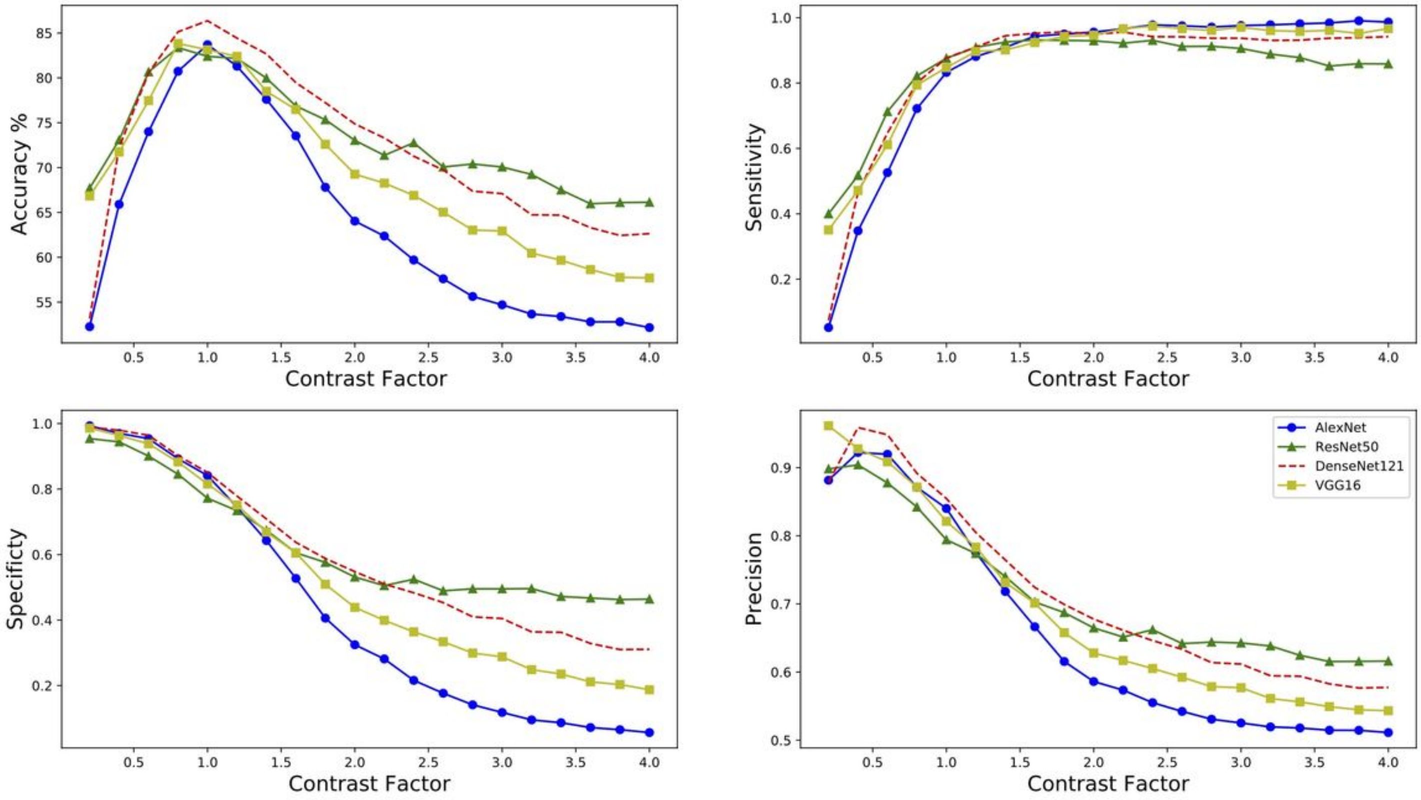


Figure 3

Graphics of metrics over contrasted images

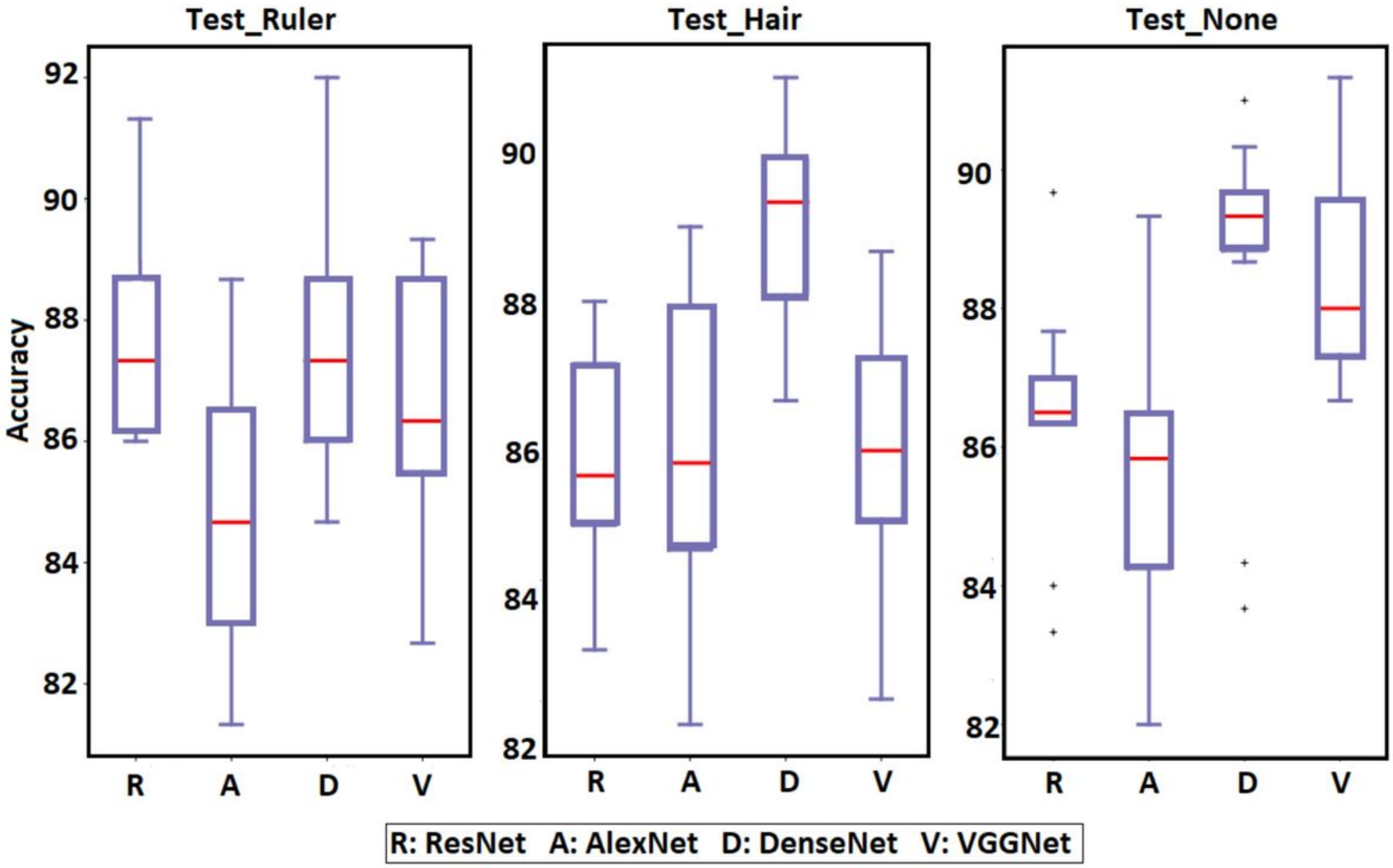


Figure 4

Accuracies of deep neural network models for test sets that are calculated over 3 datasets.

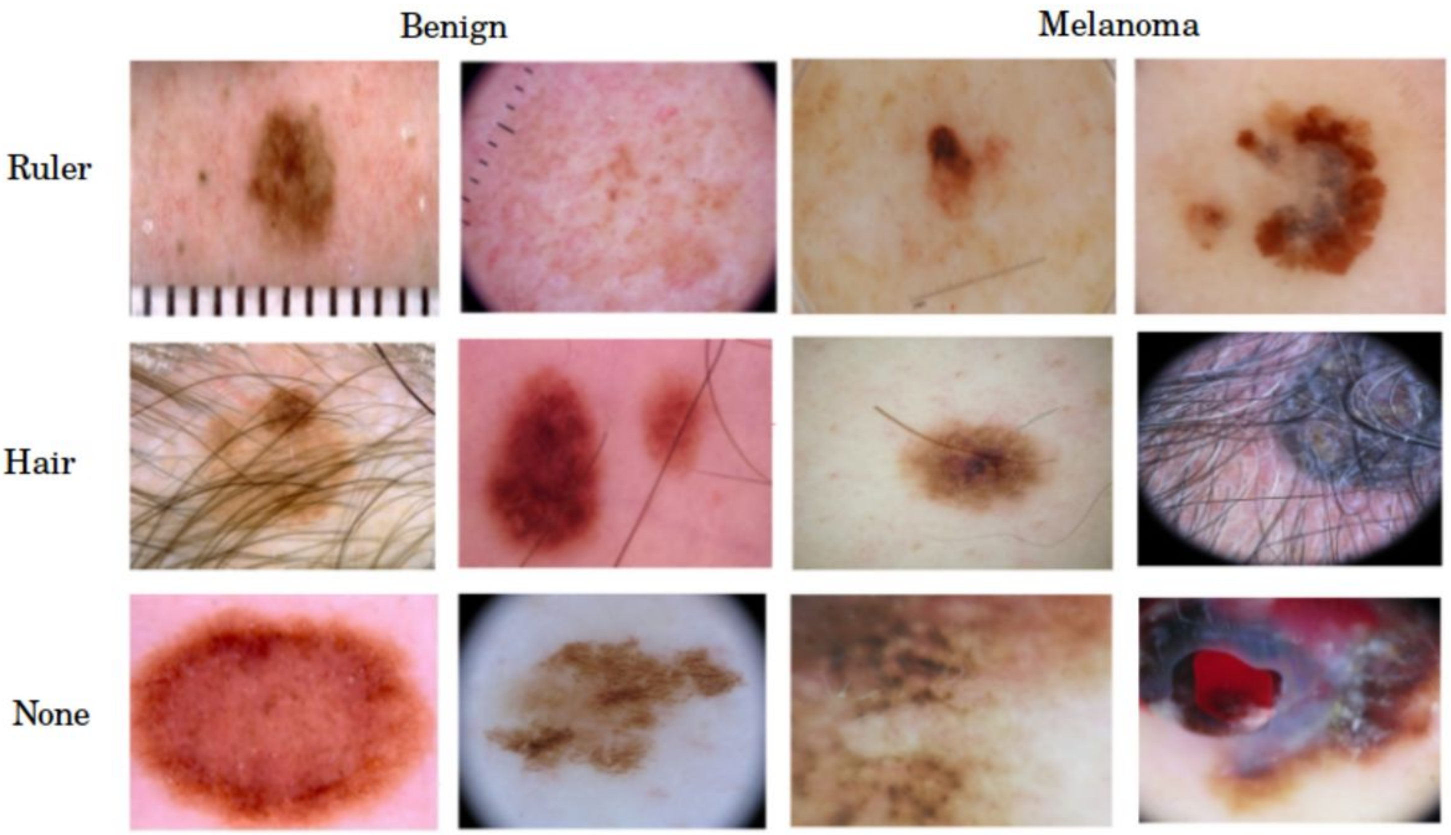
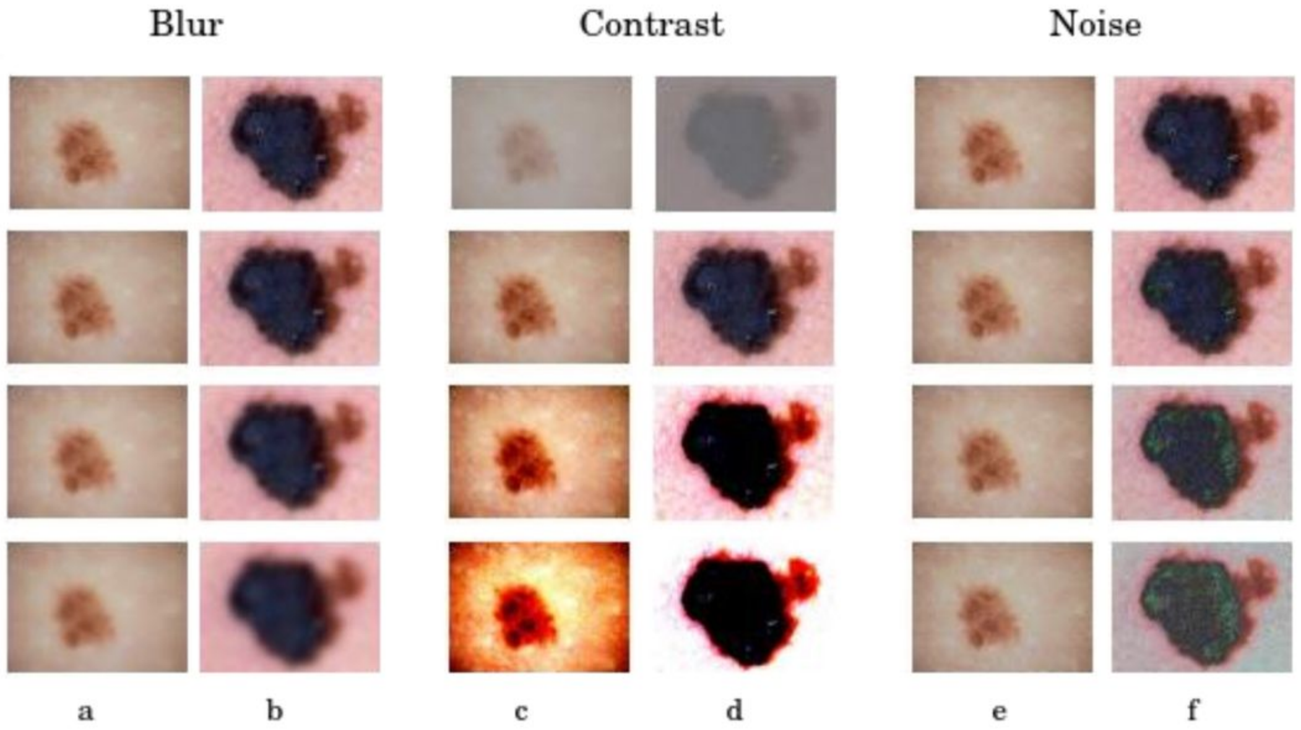


Figure 5

Sample images from 3 test sets.



a,c, e columns are benign; b,d,f columns are melanoma images
For Blur set, the images in the rows are blurred by 5x5, 21x21, 41x41, 81x81 filters, respectively.
For Contrast set, the images in the rows are enhanced contrast by factors of 0.2, 1, 2, 4, respectively.
For Noise set, the images in the rows are added noise with sigma = 1, 10, 20, 30 respectively.

Figure 6

Sample images from 3 test sets.