Article

Using Customized CNN and AlexNet to Classify Skin dermatoscope of seborrheic keratosis, melanoma, and nevus

Aman Kumar Soni,

1. School of Computer Science and Engineering, Lovely Professional University, Punjab 144411, India; shivsoni377@gmail.com, Github Link

Abstract

Skin cancer, the most painful human condition, is diagnosed primarily by appearance, starting with the first clinical examination and followed by a possible skin analysis, biopsy and histopathological examination. The automatic separation of skin lesions using images is a challenging task due to the well-differentiated appearance of the skin lesions

Deep convolutional neural networks (CNNs) demonstrate the dynamics of normal and highly flexible functions in many fine character categories. Here we show the separation of skin lesions using a single CNN, trained end-to-end exposure to direct images, using only pixels and disease labels as inserts. We train CNN using IHAM's clinical photographic database - two orders of magnitude larger than previous databases - containing a variety of diseases. We examine the efficacy of its clinical images with two critical cases of the use of binary segregation: malignant carcinomas compared with benign seborrheic keratoses; and worse melanomas compared to malignant nevi. The first case represents the most common cancer, the second case identifies the most lethal skin cancer

Background:

Classifiers based on convolutional neural networks (CNNs) shown to differentiate skin images cancer is the same as a dermatologist and can allow a quick and life-saving diagnosis, even outside the hospital with the inclusion of applications on mobile devices. To our knowledge, there is currently no review of current work in this research area.

Objective:

This study provides a systematic review of the first state-of-the-art study on the separation of skin lesions with CNN. We limit our review of skin lesion separators. Methods that use CNN only are categorized or categorized of dermoscopic patterns that are not considered here. Besides, this study discusses why comparisons are presented the processes are very complex and what challenges must be addressed in the future.

Result:

We found 13 papers separating skin lesions using CNN. The methods of separation can be differentiated according to three principles. Methods using CNN already trained with another large database and upgrading its parameters in the separation of skin lesions are the most widely used and show good performance with currently limited data sets are available.

Conclusions:

CNN's show high performance as high-quality materials for filtering skin lesions. Unfortunately, it is difficult to compare differentiated editing methods because some methods use non-public data sets for training and / or testing, making production more difficult. Future publications should make use of publicly available benchmarks and fully reflect the methods used for training to allow comparisons

Keywords: deep learning; transfer learning; ImageNet; K-nearest; medical image processing

Introduction:

For the past 10 years, from 2008 to 2018, the yearly incidence of melanoma cases increased by 53%, as a result of increased UV exposure [1,2]. Although melanoma is one of the most deadly forms of skin cancer, a quick diagnosis can lead to a much higher chance of survival.

The first step in diagnosing a serious ulcer is a dermatologist is a visual examination of a sensitive skin area. Proper diagnosis is important because of the similarity of other types of ulcers; moreover, the diagnostic accuracy is consistent firmly with the professional experience of the physician [3].

Without additional technical assistance, dermatologists have 65% -80% accuracy in melanoma screening [4]. In suspicious cases, visual inspections are supplemented with dermatoscopic images taken with special high-resolution and magnify the camera. During the recording, the light flashes controlled and the filter is used to reduce the appearance of the skin, thus making deeper layers of skin visible. With this technology support, the accuracy of the skin lesion diagnosis can be increased at a continuation of 49% [5]. A combination of visual testing and dermatoscopic images ultimately leads to complete loss accuracy of 75% -84% diagnosis of melanoma by dermatologists [6,7]. For some time, the problem of separating skin lesions has also resurfaced transferred to the focus of the machine learning community.

Automatic wound isolation can support doctors in their daily clinical practice and enables faster and cheaper access to this life-saving diagnosis, even outside the hospital installation of applications on mobile devices [8,9]. Before 2016, the research followed mainly the old workflow of the machine reading: pre-processing, segmentation, feature removal, and separation [9-11]. However, a higher level of System-specific expertise is required, especially in the aspect output, and selection of the most adequate features time-consuming. Also, errors and loss of information in the first steps of processing have a very strong influence on separation quality. For example, the effect of poor segregation often leads to negative consequences for the removal of the feature and, as a result, low accuracy in stages. In 2016, a change occurred with lesion research separation strategies. An indication of this change may be available on the routes installed in the 2016 International Symposium on Biomedical Imaging (ISBI) [12]. 25 the participating groups did not use the traditional machine learning methods; instead, they all use the in-deep reading process: convolutional neural networks (CNNs) [13].

This report provides a systematic review of the initial file of modern research to classify skin lesions using CNN. The methods presented are divided into whether CNN is used only as a feature release even if it is in use learning at the end. The end of this paper discusses why a comparison of strategies is presented so much It is also difficult to face what challenges must be met in the future.

Methods:

Search Strategy:

Google Scholar, PubMed, Medline, ScienceDirect, and Web of Science Web sites are required for systematic review and original research articles published in English. Search names for convolutional neural networks, in-depth study, skin cancer, included ulcers, melanoma, and carcinoma. Only papers which indicated that adequate scientific procedures were included in this review

Study Selection:

We have limited our review of ways to differentiate skin lesions. In particular, methods that use CNN only for wound separation or separation of dermatoscopic patterns as in

Demyanov et al [14] is not considered in this paper. Moreover, only the papers show sufficient scientific evidence continuity is included in this review. This is the last process including understandably introducing methods and adequately discuss the results. It works when the root performance was unthinkable not considered in this case apply, for example, to Carcagnì et al [15] or Dorj et al [16].

Convolutional Neural Networks:

CNN's neural networks have certain properties it has been shown that there is great power in places like image recognition and segregation [17]. CNN has always been there shown to better identify faces, objects, and road signs there are people so it can be found on robots too self-driving cars. CNN is a targeted learning method and is therefore trained using data written by appropriate classes. CNN learns the relationship between inputs and category labels and contains two elements: layers are hidden there features are extracted and, at the end of processing, I fully integrated layers used for actual separation function. Unlike conventional neural networks, the hidden layers of CNN have a certain structure. In standard neural networks, each layer is made up of a group of neurons and one neuron of a layer is connected to each neuron of the previous layer. The structure of the hidden layers on CNN is slightly different. The linear particles are not connected to all neurons of the previous layer; rather, they are connected only to the smallest number of neurons. This limit on local communication as well as additional integration layers that summarize the effects of local neuron a single value results in flexible translation features. This has consequences with a simple training process and the difficulty of a low model

Dataset:

Skin cancer is the most common type of cancer. Melanoma, in particular, accounts for 75% of skin cancer deaths, although it is the most common skin cancer. The American Cancer Society estimates that there are more than 100,000 new cases of melanoma to be discovered by 2020. It is estimated that at least 7,000 people will die from the disease. As with other cancers, early and accurate diagnosis - which can be aided by data science - can make treatment more effective.

Currently, dermatologists examine all of a patient's warts to identify external lesions or “bad ducks” that may be melanoma. Existing AI methods have never adequately considered this clinical framework. Dermatologists can increase their diagnostic accuracy if the discovery of algorithms focuses on "contextual" images within the same patient to determine which images represent melanoma. If successful, the dividers will be more accurate and can better support the work of the dermatologist.

As a leading healthcare organization for informatics in medical thinking, the aim of the Society's Illustration Informatics in Medicine (SIIM) is to advance medical informational photography through education, research, and innovation in a multicultural society. SIIM is being joined by the International Skin Imaging Collaboration (ISIC), an international effort to improve the diagnosis of melanoma. The ISIC archive contains the largest collection of publicly controlled images of skin insects.

In this competition, you will identify melanoma in pictures of skin lesions. You will use images within the same patient and determine which ones can represent melanoma. Use patient-level content information can assist in the development of image analysis tools, which can better support dermatologists in clinics.

Melanoma is a deadly disease, but if it is caught early, most melanoma can be treated with minimal surgery. Automatic image analysis tools for the diagnosis of melanoma will improve.

In this project I have used HAM1000 ("Human Against Machine with 10000 training images") dataset

The HAM10000 ("Human Against Machine with 10000 training images") dataset which contains 10,015 dermatoscopic images was made publically available by the Harvard database on June 2018 in the hopes to provide training data for automating the process of skin cancer lesion classifications. The motivation behind this act was to provide the public with an abundance and variability of data source for machine learning training purposes such that the results may be compared with that of human experts. If successful, the appplications would bring cost and time saving regimes to hospitals and medical professions alike.

Apart from the 10,015 images, a metadata file with demographic information of each lesion is provided as well. More than 50% of lesions are confirmed through histopathology (histo), the ground truth for the rest of the cases is either follow-up examination (follow_up), expert consensus (consensus), or confirmation by in-vivo confocal microscopy (confocal)

Data source: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T

The 7 classes of skin cancer lesions included in this dataset are:
1. Melanocytic nevi
2. Melanoma
3. Benign keratosis-like lesions
4. Basal cell carcinoma
5. Actinic keratoses
6. Vascular lesions
7. Dermatofibroma

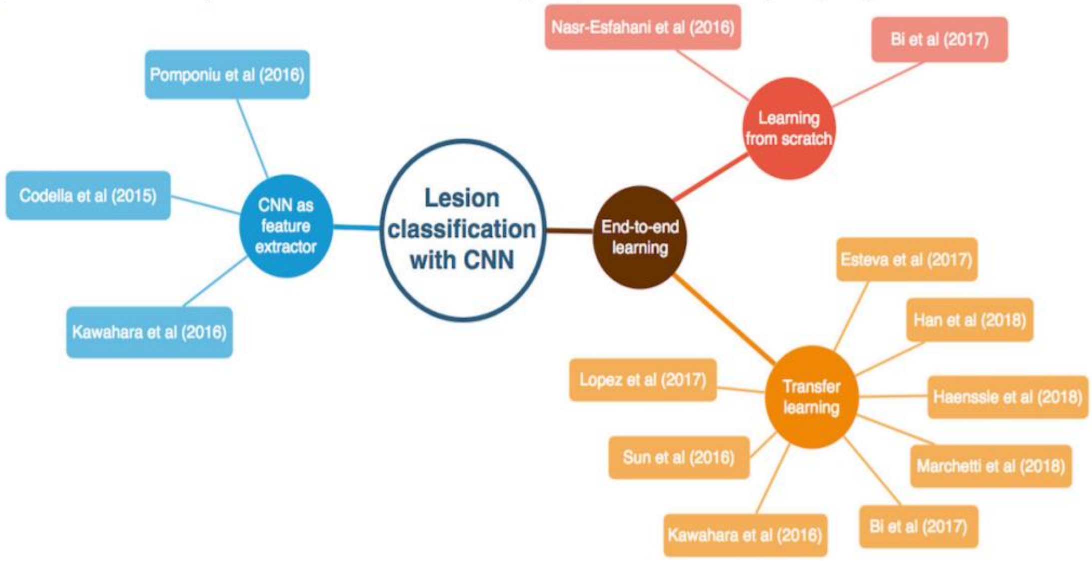
Current Classifiers for Skin Lesions Based on Convolutional Neural Networks:

In this section, the CNN methods are individually used for the classification skin lesions are given. CNN's can be used to separate the skin wounds in two completely different ways. On the other hand, CNN is pre-made in another big database, like ImageNet [18], which can be used as a feature release. In this case, the separation is done by another separator, e.g. k-nearest neighbors, vector support devices, or artificial neural networks. Alternatively, CNN can read the file directly the relationship between raw pixel data and category labels by end-to-end reading. In contrast to the classical workflow commonly used in machine learning, feature exclusion becomes an important part of segregation and not has long been regarded as an independent isolation measure. If CNN is trained in end-of-life learning, research can do it further divided into two different modes: reading model from scratch or transfer learning. Overview of introduced CNN channels shown in Figure 1.

The basic requirement for successful CNN in-depth training models that such sufficient data for training labels and classes are available. Otherwise, there is a risk of neural violations network and, as a result, inadequate performance unknown network input network assets. There is so much a limited amount of information publicly available by division skin lesions. Almost all published methods use data sets that contain less than 1000 training details in each training class. By comparison, CNN's well-known image models in categories, such as AlexNet [18], VGG [19], GoogLeNet [20], or ResNet [21], are trained with a large database of images ImageNet also has over 1000 training images for each training section.

However, through a specific training process called to transfer learning, CNN models have a capacity of several million free parameters that can also be categorized, even if only a small amount of training data is available. In this case, CNN was created before using a very large database, e.g. ImageNet; used for CNN startup for proper work. In particular, the last fully integrated layer of the previously created CNN model is converted according to the number of training classes in the practice of physical separation. Bangu then there are two CNN weight loss options made earlier: in fine-tune all CNN layers or hold the front layers due to problems of overcrowding and fine-tuning only other back layers of the network. The concept of this process is that the front layers of CNN contain the standard features (e.g. edge or color block) that are useful many jobs, but the background layers of CNN are increasingly increasing directly into the details of the actual content classes data. In the following discussion, the statistics should be checked a separate division is introduced. Next, the methods they use CNN was introduced as a feature. The last paragraph provides an overview of the involved methods when using CNN learning at the end.

Figure 1. An overview of the presented convolutional neural networks (CNNs) methods and the corresponding categorization.



Skin Lesion Classifier Using End-to-End Learning:

Convolutional Neural Network Model Training With Transfer Learning:

Because publicly available data sets are limited, it is normal The method of dissection of the skin lesion involves the transfer of learning. Therefore, all such activities make CNN via ImageNet dataset; next, the weighting parameters of the CNN are fine-tuned to the actual classification problem.

Esteva et al [26] presented a landmark book. The first time, the CNN model was trained with a large amount of data, mainly 129,450 photographs, of which 3374 were found dermatoscopic devices and represented by a separate 2032 skin ulcers. Two binary split problems were considered: keratinocyte carcinomas compared with benign seborrheic keratosis as well melanoma is worse than a malignant tumor. The final separation was divided for both clinics at once dermatoscopic images. Authors used the startup of GoogLeNet v3 of the division, first made with ImageNet big data image. CNN model at the time was well designed to separate skin lesions using transfer reading. A special property of this method is the use of the novel the immune system formed by the drug in which a person diseases form the leaves of the tree. A group of internal nodes together individual clinical and clinical diseases the same. CNN does not have a two-sided vector as result instead, it reports on the distribution of opportunities through over 757 training classes. Determining coarser skills lesion class (i.e., the internal node at the highest level in the tree), chances of having children's places for this stage of coarser lesion he summed up together. The authors point out within the examination that CNN-trained good classes are better performance than CNN trained differently classes are interested in the problem. CNN trained tested with fully proven diagnostic data by biopsy as well won ROC AUC for .96 carcinomas, ROC AUC .96 for melanomas, and ROC AUC 94 for melanomas separated only by dermatoscopic images.

Hassle et al [3] presented a similar approach to Esteva et al [26]. The GoogLeNet Inception v3 model is designed for separation of the skin lesion by transfer reading, where the instruments are well aligned in all layers. The analysis was limited in melanoma skin lesions compared to malignant nevi also The AUC ROC achievement for this project was .86 (Esteva et al [26]:.94). The exact number of training details points was not given not all the details confirmed by biopsy. However, the publication includes the largest number of dermatologists to date (n = 58) and was the first to show that clinical supplement The details improve both; sensitivity and specificity of dermatologists.

Han et al [27] are particularly notable for their scientific discoveries open operation as they created their computer algorithm available to the public for external testing. The group presented the classification of 12 different skin diseases based on medical imaging. They made a well-designed ResNet model with 19, 398 training photos. With Asan data available to the public, The CNN model has earned ROC AUCs for basal diagnosis cell carcinoma, squamous cell carcinoma, intraepithelial carcinoma, and melanoma of .96, .83, .82, and .96, respectively. CNN collection of melanomas classification Comparisons of nevi or lentigines are presented by Marchetti et al [13]. Use five methods to mix all the default predictions from the 25 participating teams in ISBI 2016 Challenge one split result. For this purpose, they tested two modes of illiteracy and three machine learning methods. Fusion algorithms were trained with 279 dermatoscopic images from the ISBI 2016 Challenge database and were tested with another 100 leather models from the same database. Depending on the median accuracy, the greedy combination became a 58% sensitivity method with a sensitivity of 58% and 88% specificity.

Another type of CNN ensemble was introduced by Bi et al [28]. They looked at the classification of melanomas in comparison to seborrheic keratosis compared to nevi using dermatoscopic images. They did not train many CNNs in the same classes problem instead, three ResNets of various problems were present trained: one of the original problem with three and two binary classes classifiers (i.e., melanoma compared to other stages of the lesion or seborrheic carcinoma compared to other stages of the lesion) approx. well-prepared CNN pre-made. The test used 150 dermatoscopic images and led to ROC AUC of .854 for melanomas, ROC AUC of .976 for seborrheic carcinomas, and the central ROC AUC in all classes of .915. The special design of the CNN collection is presented by Kawahara et al [29]. CNN is made up of many parts where each section looks at the same picture in contrast decision. Next, the last layer covers the results from many decisions into a single layer. CNN points out interaction in various image and weight decisions the parameters are made entirely by end-to-end reading. The algorithm has obtained standard phase accuracy for 79.5% in the public photo library Dermot.

Similar to Esteva et al [26], Sun et al [30] introduced the separator that used 198 well-defined training classes. In total, 6584 images of public archive clinics available to the public DermQuest is used for training and testing as well the performance of CNN models CaffeNet and VGGNet was examined this separation problem. Very good rating 50.27% accuracy overall 198 categories were obtained using The invented VGGNet, optimized for weighted boundaries.

The modified VGGNet was also used by Lopez et al [31], where the melanoma classification is compared to nevi or lentigines it was spoken using leather images. Authors compared to CNN's precision accuracy trained from scratching, CNN pre-made with transfer readings and frozen layers, as well as CNN made before reading the transfer once good order of weighted boundaries. All three configurations tested with 379 photos from the ISBI 2016 Challenge dataset, and the last-mentioned configuration detected a very high accuracy of 81.33%.

Skin Lesion Classifier Using Transfer Learning :

In addition to the huge increase in the number of trained images, the available size of the database is not enough to train a new in-depth model from scratch. To overcome this problem, the concept of learning transfer is applied to AlexNet's previous formats in three different ways. First, the separation layer is replaced by a layer of softmax in two or three phases. Second, the instruments are well prepared and back distribution is done to train new instruments. A low level of learning is used when the weight of the convolutional layer is not significantly altered, while the fully connected layer layers are initiated randomly. The stochastic gradient descent (SGD) algorithm used for network regeneration is based on the data used for skin cancer. Finally, extended data sets to increase the number of images available to train a deeper network. This process resulted in accurate measurements and achieved a good degree of separation from the new layer replaced by softmax.

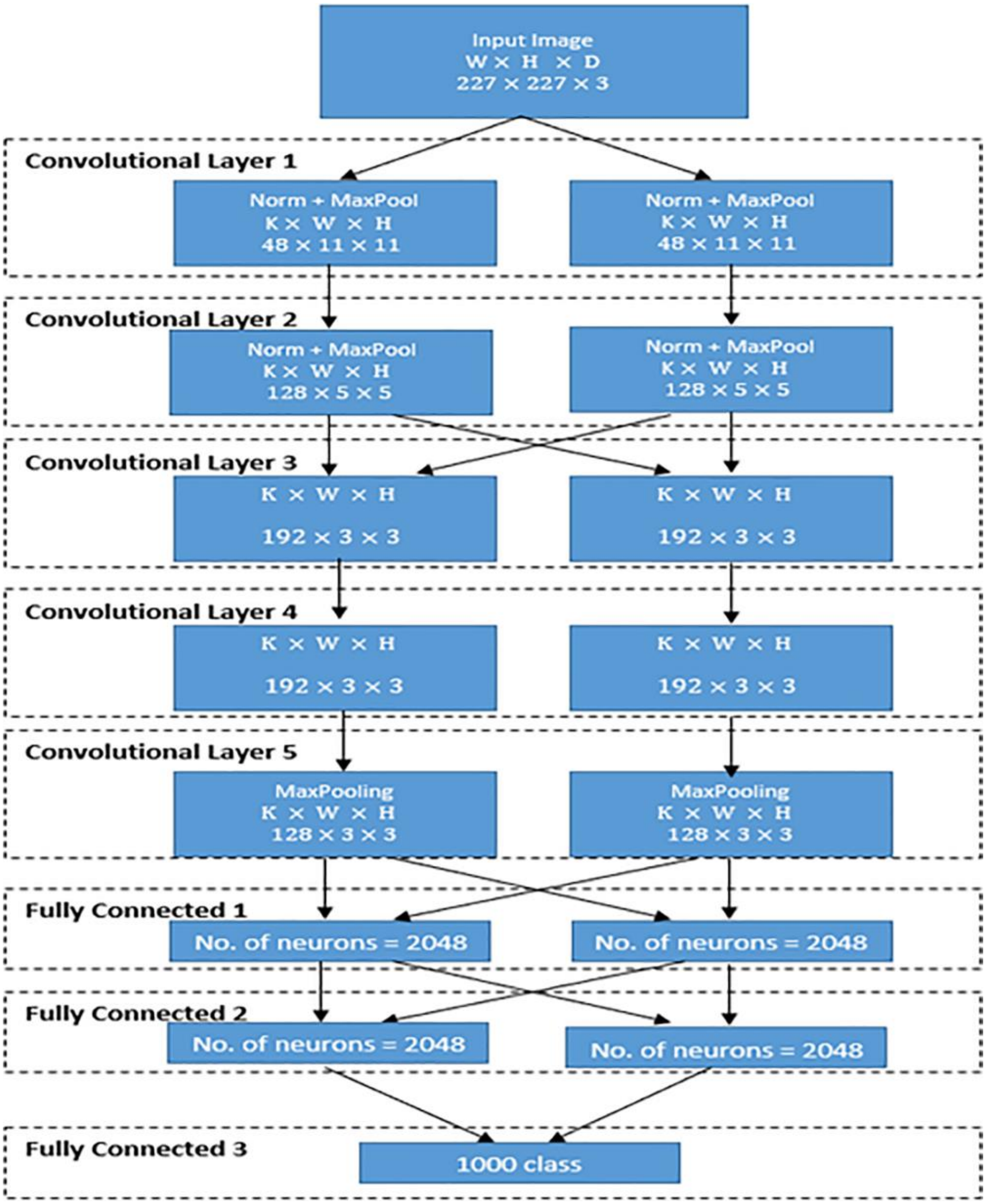
As mentioned above, melanomas have a wide range of intraclass, and there is a high apparent similarity between melanoma and non-melanoma lesions, which greatly influences the function of recognition especially when skin lesion is performed using the first limited doscopy images.

AlexNet CNN model:

Krizhevsky et al. [18] established AlexNet to implement the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) [21]. The first layer of AlexNet is used to filter image input. Input image must be wide (W), height (H), and depth (D); $227 \times 227 \times 3$ when $D = 3$ refers to red, green and blue. As mentioned above, the first convolutional layer used to filter an input image with 96 (K) and 11x11-sized filter (F) in addition to 4 pixels called stride (s). The distance between the adjacent field centers of the neighboring nerves on the kernel map is called a step. The mathematical formula, $((W - F + 2P) / S) +1$, is used to calculate the size of the convolution layer when P refers to the number of pixels combined, equal to zero. Using this formula, the volume of the convolution layer output is $(227-11 + 0) / 4) +1 = 55$. The second convolutional layer input will be ($55 \times 55 \times$ none of the filters) and the number of filters in this layer is 256. As the function of this layer is still distributed over 2 GPUs, each load is divided by 2 to 2 GPUs.

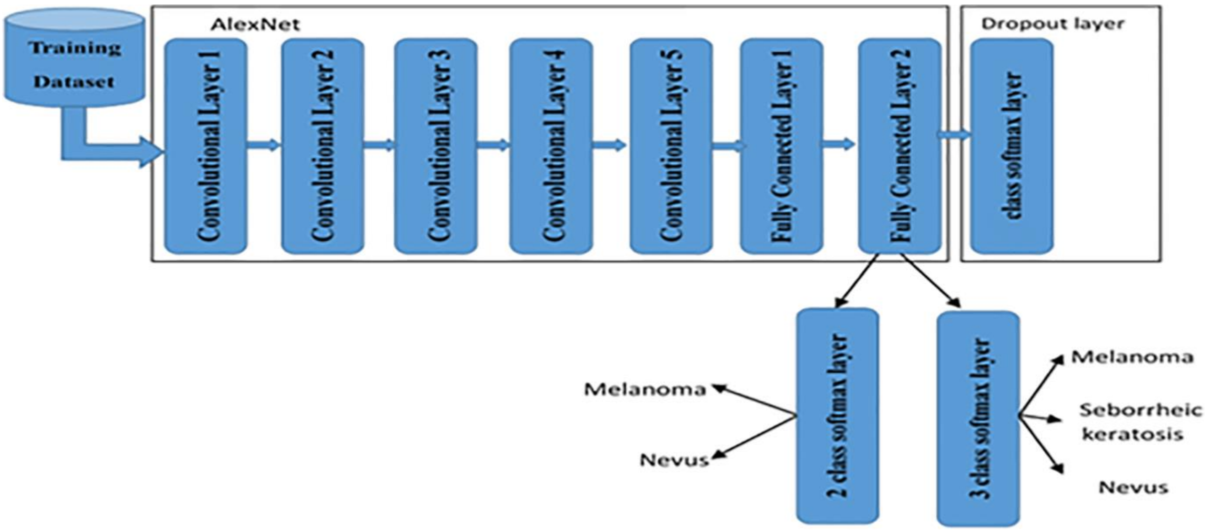
Next is a convolutional layer followed by a consolidation layer. The integration layer tries to reduce the map size of each feature while maintaining the key features, where the integration can be Sum, Max, Rating, etc. AlexNet uses a large integration layer. The insertion of this layer is 256 filters. Each filter is $5 \times 5 \times 256$ in addition by 2 pixels as a step. Using two GPUs, the function will be divided into $55/2 \times 55/2 \times 256 / 2 \approx 27 \times 27 \times 128$ per GPU.

Combined with the standard two-layer connections are connected in the third layer with 384 kernels, each with a size 3×3. There are 384 characters in size 3×3 in the fourth convolutional layer and will be separated by more than 2 GPU so each GPU load will be $3 \times 3 \times 192$. The fifth convolutional layer has 256 characters per letter 3×3 and will be separated by more than 2 GPU so each GPU load will be $3 \times 3 \times 128$. It should be noted that the third, fourth and fifth layers the conversion of the decision was made out of the ordinary and the integration of the layers. The release of these three convolutional layers is transmitted as input to the number of 2 layers that are fully connected when each layer contains 4096 neurons. Figure 1 shows the complete construction of Alex-Net to separate classrooms using imageNet [21] as a training database.



Experiments and results

Experiments are performed using an HP computer equipped with a core i5-6300HQ processor, 8 GB DDRAM and a NVIDIA GeForce 950M graphic card. The MATLAB 2017 x64-bit is used to execute the coded program. Three datasets, ISIC, MED-NODE, and DermIS—DermQuest, of RGB colored skin images are used in these experiments. The first dataset consists of three labeled data/classes, melanoma, seborrheic keratosis, and nevus. The second and the third datasets consist of only two labeled data/classes, melanoma and nevus. The code is converted from MATLAB 2017 to CUDA to be run over GPU. Using GPU enables us to use a huge number of training data with low error rate of models. In many works, like that with DCNN, the classification layer may be dropped out and replaced with other classification methods like multi-class SVM. In this work, the classification layer called softmax is replaced with a new softmax layer to be appropriate for skin lesion where three classes are used. Fig 2 illustrates the modified pre-trained AlexNet with the new softmax layer.



There are two kinds of performed experiments with the three datasets. The first one is to evaluate the proposed method using original datasets without image augmentation. The second one is to evaluate the proposed method with augmented datasets. All experiments are performed with fixed values of the batch size, 10, the number of training epochs, 32, and the initial learning rate, 0.001. All color images are segmented using the segmentation methodology [41].

$$\text{Accuracy} = \frac{t_p + t_n}{t_p + f_p + f_n + t_n}$$

$$\text{Sensitivity(TPR)} = \frac{t_p}{t_p + f_n}$$

$$\text{Specificity(TNR)} = \frac{t_n}{f_p + t_n}$$

$$\text{Precision(PPV)} = \frac{t_p}{t_p + f_p}$$

Four evaluation measures are used to evaluate the performance of proposed method. These measures are accuracy, sensitivity, specificity, and precision [42]. These measures are computed using the following equations:

Where t_p, f_p, f_n, *and* t_n refer to true positive, false positive, false negative, and true negative respectively. The acronyms, TPR, TNR, and PPV refer to true positive rate, true negative rate, and positive prediction value. The rates of true negative and false positive should be large and small, which makes most of the points fall in the left part of the receiver operating characteristic (ROC) curve [43].

All experiments start by loading the color images form the data source, then by passing it to the segmentation step. According to the pre-trained AlexNet, the size of the input image cannot exceed 227×227, and the depth-limit of the image is 3. Therefore, after segmentation, a check step is performed to ensure the suitability of the image's size. If the size of the image exceeds the size limit, a resizing process to 227×227×3 for width, height, and depth is imperative.

In the first type of experiment, the 10-fold cross validation have been used to divided MEDNODE and DermIs–DermQuest dataset into groups for training and testing without any augmentation. Each group have been used at least once as training and once as testing but not in the same run. Then the modified AlexNet after applying transfer learning theory have been used. This process was repeated 10 time and the average accuracy for the 10 runs times was computed to be the overall accuracy of the proposed model. In the first type of experiment, runs are performed using the original datasets of color images without any augmentation. The is applied where the is pre-trained. The classification layer, softmax, is modified to work with 2 classes instead of the ImageNet classes.

The first run was done with the DermIS—DermQuest dataset which contains low quality images of two classes, melanoma and nevus. To evaluate the performance of the proposed method, the values of the four measures are computed for each class separately. Therefore, the average value for these measures are computed. The average of the computed measures is 88.24%, 86.79%, 86.79%, and 89.01% for accuracy, sensitivity, specificity, and precision respectively. The confusion matrix of this experiment is shown in Fig 3.

Confusion matrix for original derma

Actual \ Prediction	Nev	Mel
Nev	30	4
Mel	4	30

Confusion matrix for augmented derma

Actual \ Prediction	Nev	Mel
Nev	1822	59
Mel	59	1822

The second run is performed using the same conditions and architecture with the second dataset, MED-NODE. The MED-NODE dataset consists of high quality dermoscopy images. This dataset is divided into two classes, melanoma and nevus. The pre-trained AlexNet that has transferred learning with modified softmax layer to be appropriate with two classes of skin lesions. The average values of measures are computed where these values are 91.18% for average accuracy of the two classes, 91.43% for average accuracy of sensitivity and specificity, and 90.70% for average precision. The confusion matrix of this experiment is shown in Fig 4.

Confusion matrix for original MedNode

Actual \ Prediction	Nev	Mel
Nev	31	3
Mel	3	31

Confusion matrix for augmented MedNode

Actual \ Prediction	Nev	Mel
Nev	1827	43
Mel	43	1827

The ISIC dataset is used in the third run where this dataset consists of three classes, melanoma, seborrheic keratosis, and nevus. The ISIC dataset is relatively big and originally divided into training and test groups; so, we ignore 10-fold cross validation. Similarly, the transfer learning is applied to the pre-trained AlexNet where the softmax layer is modified to be worked with three classes. The values of batch size, the number of training epochs and initial learning rate were fixed for all runs as 10, 32, and 0.001 respectively. The average computed for all measures, accuracy sensitivity, specificity, and precision were 87.31%, 62.02%, 79.07%, and 73.07% respectively. The confusion matrix of this experiment is shown in Fig 5

Confusion matrix for original ISIC

Actual \ Prediction	Nev	Mel	Seb. Ker.
Nev	245	14	30
Mel	38	235	16
Seb. Ker.	9	3	265

Confusion matrix for augmented ISIC

Actual \ Prediction	Nev	Mel	Seb. Ker.
Nev	15113	142	640
Mel	710	14937	248
Seb. Ker.	124	88	15683

The second kind of experiments is performed with the same datasets, DermIS- DermQuest, MED-NODE, and ISIC. DermIS _dermQuest, and MED-MODE dataset have been splitting into training and testing groups using 10-fold cross validation. These groups in addition to the ISIC training and testing dataset groups are augmented by rotating each image with 55 different rotation angles ranging from 0^0 to 355^0 with a constant step 5^0.

The DermIS- DermQuest is used in the first run of this experiment. The original colored images of this datasets are segmented to reduce the size and remove unwanted complicated background. The segmentation step is performed before the augmentation process, which is applied to all segmented color images. As a result of augmentation, the number of color images becomes 4620 and 4785 for melanoma and nevus respectively. A size-check constraint, 227×227×3, is applied to all input images as done in the first experiment. The same values of batch size,10, the number of training epochs, 32, and initial learning rate, 0.001 are applied. The classification layer softmax is modified to two classes, melanoma and nevus. The same four measures are used to evaluate the performance of the modified pre-trained AlexNet. The computed average values of these measures are 96.86% for accuracy, 96.90% for sensitivity and specificity and 96.92% for precision. The confusion matrix of this experiment is shown in Fig 3.

The second run is performed with the MED-NODE dataset. Like previous runs, the original color images are segmented and then the augmentation process is performed where the number of images becomes 3850 and 5500 for melanoma and nevus respectively. The size of all images is determined where images with exceeded size are resized to 227×227×3 for width, height, and depth. The performance measures are computed where the average values of these measures are 97.70%, 97.34%, 97.34%, and 97.93% for accuracy, sensitivity, specificity, and precision respectively. The confusion matrix of this experiment is shown in Fig 4.

The third experiment is performed with the ISIC dataset. This dataset consists of three classes, melanoma, seborrheic keratosis, and nevus. Therefore, the modified softmax layer will be modified to work with these three classes. After segmentation process, the dataset is augmented where the number of images becomes 20570, 13970 and 75460 images for melanoma, seborrheic keratosis, and nevus respectively. The size of all input images must not exceed the size 227×227×3. The average values of the performance measures are 95.91%, 88.47%, 93.00%, and 92.34% for accuracy, sensitivity, specificity, precision, and negative predication value respectively. The confusion matrix of this experiment is shown in Fig 5.

Table 1 gives an overview of the obtained results for the performed experiments. It is clear that the augmentation processes significantly improve the classification rates. The proposed method achieved a very high classification rates with different datasets.

Dataset	Original dataset				Augmented dataset			
	Average accuracy	Average sensitivity	Average specificity	Average Precision	Average accuracy	Average sensitivity	Average specificity	Average Precision
DermIS-DermQuest	88.24	86.79	86.79	89.01	96.86	96.90	96.90	96.92
MED-NODE	91.18	91.43	91.43	90.70	97.70	97.34	97.34	97.93
ISIC	87.31	62.02	79.07	73.07	95.91	88.47	93.00	92.34

https://doi.org/10.1371/journal.pone.0217293.t001

Conclusion:

Principal Findings:

One issue in comparison is the separation of the skin lesion ways that problematic approaches to the individual tasks vary, sometimes slightly. This is happening not only the training classes considered and the data used, but also the figures presented. Also, some works using non-public archives of skin clinics in addition to publicly available data archives [3,26]. This makes it even. It is very difficult to produce results. As of 2016, ISIC The Melanoma Project has tried to improve this feature by to establish a publicly accessible dermatoscopic skin archive images of wounds as a measure of learning and research [12]. Also, they announced an annual challenge when a clearly defined problem must be fixed. It would be desirable if more work can compare to this sign in to achieve a better level of processes in the province of research.

Another important challenge in this area of research is the development of large public photo archives with images such as representing as much land as possible [33]. The existing photo archives mainly contain skin lesions from people of a fair complexion. Images on the ISIC website, for example, come mainly from the United States, Europe, and Australia. Achieving accurate separation of dark skin and people too, CNN must learn to pull off the skin color. However, this can only happen if it looks good enough photos of black people during training.

The improvement in the quality of the separation can be achieved to add clinical data (e.g. age, gender, race, skin type, and anatomic position) as the inclusion of classifiers. This is additionally the information that is beneficial for decision-making dermatologists, as Haenssle et al [3] point out. Future work should take care of these factors.

Conclusion:

Unfortunately, it is difficult and often impossible to compare the effectiveness of published segmentation results because many authors use non-public data sets for training and/or examination. Upcoming books should be used publicly measurements and full disclosure methods used for training to allow comparisons.

Project Conclusion:

Findings: Accuracy is higher if model is trained on more samples of lower resolution than small samples of high resolutions.

Improvements:

- loss is still increasing in every model (optimizer is not steering the model towards the right direction)
- low resolution
- accuracy is the highest but unstable
- loss: between 1 to 2.5
- adding dropout layers did not help with reducing overfitting, but in turn reduced accuracy scores
- overfit on training data
- everything was done on unscaled, imbalanced class
- high resolution
- accuracy score is lower than the low resolution model
- more unstable and noiser than the low resolution model
- loss is between 2 to 6, much higher than low resolution
- only had 800 pictuers left in total after fixing class imbalance issue

Going forward, can continue to refine the model to achieve a stable decrease in loss function with every epoch, build an interface such that given an image of a skin lesion within the two classes, the output will give a % probability of which of the seven classes it belongs to.

Reference:

1. Nami N, Giannini E, Burroni M, Fimiani M, Rubegni P. Teledermatology: State-of-the-art and future perspectives. Expert Rev Dermatol 2014 Jan 10;7(1):1-3. [doi: 10.1586/edm.11.79]

2. Fabbrocini G, Triassi M, Mauriello MC, Torre G, Annunziata MC, De Vita V, et al. Epidemiology of skin cancer: Role of some environmental factors. Cancers (Basel) 2010 Nov 24;2(4):1980-1989 [FREE Full text] [doi: 10.3390/cancers2041980][Medline: 24281212]

3. Haenssle H, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, Reader Study Level-I and Level-II Groups. Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition

in comparison to 58 dermatologists. Ann Oncol 2018 Aug 01;29(8):1836-1842. [doi: 10.1093/annonc/mdy166] [Medline: 29846502]

4. Argenziano G, Soyer HP. Dermoscopy of pigmented skin lesions: A valuable tool for early diagnosis of melanoma. Lancet Oncol 2001 Jul;2(7):443-449. [Medline: 11905739]

5. Kittler H, Pehamberger H, Wolff K, Binder M. Diagnostic accuracy of dermoscopy. Lancet Oncol 2002 Mar;3(3):159-165.[Medline: 11902502]

6. Ali ARA, Deserno TM. A systematic review of automated melanoma detection in dermatoscopic images and its ground truth data. Proc SPIE Int Soc Opt Eng 2012 Feb 28;8318:1-6. [doi: 10.1117/12.912389]

7. Fabbrocini G, De Vita V, Pastore F, D'Arco V, Mazzella C, Annunziata MC, et al. Teledermatology: From prevention to diagnosis of nonmelanoma and melanoma skin cancer. Int J Telemed Appl 2011 Sep 01;2011(17):125762 [FREE Full text] [doi: 10.1155/2011/125762] [Medline: 21776252]

8. Foraker R, Kite B, Kelley MM, Lai AM, Roth C, Lopetegui MA, et al. EHR-based visualization tool: Adoption rates, satisfaction, and patient outcomes. EGEMS (Wash DC) 2015;3(2):1159 [FREE Full text] [doi: 10.13063/2327-9214.1159] [Medline: 26290891]

9. Fabbrocini G, Betta G, Di Leo G, Liguori C, Paolillo A, Pietrosanto A, et al. Epiluminescence image processing for melanocytic skin lesion diagnosis based on 7-point check-list: A preliminary discussion on three parameters. Open Dermatol J 2010 Jan 01;4(1):110-115. [doi: 10.2174/1874372201004010110]

10. Hart PE, Stork DG, Duda RO. Pattern Classification. 2nd edition. Hoboken, NJ: John Wiley & Sons; 2000

11. Oliveira RB, Papa JP, Pereira AS, Tavares JMRS. Computational methods for pigmented skin lesion classification in images: Review and future trends. Neural Comput Appl 2016 Jul 15;29(3):613-636. [doi: 10.1007/s00521-016-2482-6]

12. Gutman D, Codella NCF, Celebi E, Helba B, Marchetti M, Mishra N, et al. arXiv. 2016 May 04. Skin lesion analysis toward melanoma detection: A challenge at the International Symposium on Biomedical Imaging (ISBI) 2016, hosted by the International Skin Imaging Collaboration (ISIC) URL:https://arxiv.org/pdf/1605.01397 [accessed 2018-10-06] [WebCite Cache ID 72yoIScsz]

13. Marchetti MA, Codella NCF, Dusza SW, Gutman DA, Helba B, Kalloo A, International Skin Imaging Collaboration. Results of the 2016 International Skin Imaging Collaboration International Symposium on Biomedical Imaging challenge: Comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images. J Am Acad Dermatol 2018 Dec;78(2):270-277.e1. [doi: 10.1016/j.jaad.2017.08.016] [Medline: 28969863]

14. Demyanov S, Chakravorty R, Abedini M, Halpern A, Garnavi R. Classification of dermoscopy patterns using deep convolutional neural networks. In: Proceedings of the 2016 IEEE 13th International Symposium on Biomedical Imaging

(ISBI).: IEEE; 2016 Presented at: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI); April 13-16, 2016; Prague, Czech Republic p. 2-12.

15. Carcagni P, Cuna A, Distante C. arXiv. 2018 Jul 26. A dense CNN approach for skin lesion classification URL:http://arxiv.org/pdf/1807.06416 [accessed 2018-10-06] [WebCite Cache ID 72ypBqU6n]

16. Dorj U, Lee K, Choi J, Lee M. The skin cancer classification using deep convolutional neural network. Multimed Tools Appl 2018 Feb 22;77(8):9909-9924. [doi: 10.1007/s11042-018-5714-1]

17. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015 May 28;521(7553):436-444. [doi: 10.1038/nature14539] [Medline: 26017442]

18. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 2012 Jan:1-23. [doi: 10.1145/3065386]

19. Simonyan K, Zisserman A. arXiv. 2014 Apr 10. Very deep convolutional networks for large-scale image recognition URL:https://arxiv.org/pdf/1409.1556 [accessed 2018-10-06] [WebCite Cache ID 72yq41Xea]

20. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. arXiv. 2014 Sep 17. Going deeper with convolutions URL:https://arxiv.org/pdf/1409.4842 [accessed 2018-10-06] [WebCite Cache ID 72yqGoKqD]

21**.** Deng J., Dong W., Socher R., Li LJ, Li K., and Fei-Fei L., "ImageNet: A large-scale hierarchical image database," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 248–255, 2009. https://doi.org/10.1109/CVPR.2009.5206848