

Machine learning based skin disease detection

Dr. Chayan Paul

Swami Vivekananda University

Abstract:

Skin diseases are a very common health problem worldwide, usually self-evident by virtue of the visibility of symptoms. Diagnosis at an early stage and identification with reasonable accuracy can effectively help in treatment and management. This research is an attempt to propose a robust model using machine learning for skin disease detection from dermatoscopic images. The proposed system's methodology encompasses image preprocessing, feature extraction, and classification techniques using state-of-the-art algorithms. This will be demonstrated using a model for the accurate classification of different skin diseases on a publicly available dataset for experimentation. Results from this study add to the advancement of teledermatology and improve patient care.

1. Introduction:

Skin diseases rank among the most prevalent health issues worldwide. They afflict individuals of all age groups and racial backgrounds. These diseases range from common, almost harmless conditions, such as acne, to severe and even life-threatening conditions, including the types of skin cancers like melanoma. The early diagnosis of dermatological disorders is of essence in ensuring effective and proper treatment and leads to a good outcome for the patient. It reduces the burden on healthcare. Traditional methods of diagnosis of skin lesions by dermatologists are primarily done by visual examination of the skin lesions, followed by histopathological examination wherever indicated. However, the conventional process has its shortcomings. It is time-

consuming and open to human error. Moreover, it may not be uniform because visual assessment is subjective.

The last couple of years have definitely carved a niche for medical diagnostics—including skin disease detection—with the rapid growth in ML and AI. Already, high potential is shown by models of machine learning, particularly techniques of deep learning, to be very useful in automating the process of diagnosis with high accuracy, consistency, and efficiency. The following is a detailed note on the detection, challenges, methodologies, and future prospects in relation to skin disease detection using machine learning models as a game-changing approach.

1.1 Importance of Skin Disease Detection

Skin forms the largest organ of the human body. As such, it is highly prone to a variety of pathologies and disorders. These pathologies have serious consequences physically, psychologically, and socially for the afflicted person. In these diseases, early detection and correct diagnosis form the basis of good management. For example, melanoma, if diagnosed at an early stage, usually has a very good survival rate; nevertheless, it progresses to a very poor prognosis at an advanced stage. Conversely, misdiagnosis of benign skin diseases as malignant can lead to undue procedures and causes anxiety to the patient, increasing healthcare costs.

The higher the stakes, the greater the need for diagnostic tools that are reliable, fast, and objective, to help a dermatologist make the correct decisions. In this regard, machine learning models that excel at pattern recognition in data can be quite promising. Equipped with learning from large datasets of clinical images, such models could recognize subtle visual features that distinguish various skin conditions and hence help in diagnosis.

1.2 Challenges in Skin Disease Detection

Although machine learning has huge potential in skin disease detection, several issues need to be solved before its effectiveness and reliability can be ensured. The first is the inherent visual similarity of various skin diseases. For

example, benign moles and early-stage melanoma might look very alike, making differentiation between them pretty hard not only for human experts but also for machine learning models. This same skin condition may, however, affect different patients differently because of differences in skin types, age, and ethnicity. This again adds to the variability in diagnosis.

Another major challenge is the quality and consistency of clinical images used for training machine learning models. Images may vary depending on lighting, resolution, and the presence of other artifacts like hair, which occludes the features on which the model should focus attention. The challenge is more difficult due to the large number of skin conditions that have variable visual characteristics, making it quite hard to develop models that will identify a wide array of diseases.

The next challenge in the development of robust machine-learning models for skin disease detection could be related to data scarcity. Deep learning models require extensive training datasets, which are seldom available for medical images, and even more so in the case of rare diseases. The scarcity can result in overfitting, in which the model works perfectly on the training data but fails to generalize on new, unseen cases.

1.3 Machine Learning Models for Skin Disease Detection

The potential of using machine learning-based, mainly deep learning-based, models has recently been highly promising in mitigating such challenges. Broadly, there could be two major groups of models: traditional machine learning models and deep learning models. Both approaches have their strengths and limitations; often, their efficiency depends upon the application and on the quality of data at hand.

1.4 Traditional Machine Learning Models

Applications of traditional machine learning models to the analysis of medical images, with a focus on skin disease detection, include SVM, DT, and KNN. All of these models typically require manual feature extraction, whereby some characteristics of the images are captured—for instance,

texture, color, and shape—and then fed into the model. Their efficacy largely depends on how good the extracted features are.

One of the most used techniques for feature extraction in the case of skin disease detection is the Gray Level Co-occurrence Matrix against skin lesions' texture. Other statistical features such as mean, variance, and entropy can also be extracted to provide more information to the model. After extracting the features, they are then fed into a machine learning model trained regarding classification criteria for image categories like melanoma, basal cell carcinoma, and benign nevi.

Although traditional machine learning models have reasonable accuracy in the detection of skin diseases, they have some limitations. First, manual feature extraction requires time and domain expertise. Most of them have problems in high-dimensional data, where the number of features is higher than the number of training samples. Thus, their performance may be limited as compared to the more advanced Deep Learning models.

1.5 Deep Learning Models

Deep learning models, particularly Convolutional Neural Networks, have modified the dimensions of medical image analysis in skin disease detection. Deep learning models automatically learn relevant features from raw image data. Therefore, they are pretty different from previous machine learning models, wherein feature extraction used to take place manually. CNNs will do well for image analysis because of their potential to capture spatial hierarchies of features through convolutional layers.

A typical CNN architecture for skin disease detection includes some convolutional layers followed by a pooling layer, then fully connected layers, and lastly an output layer. Convolutional layers convolve the input images with filters to detect edges, textures, and patterns in the images. Then, the pooling layers decrease the spatial dimensions of feature maps, making the model more efficient. These extracted features are then combined through fully connected layers to make the final classification.

Other advantages of the deep learning models include their good generalization to new data if trained on a large and diverse enough dataset. In the field of skin disease detection, some CNNs were trained with huge datasets such as the International Skin Imaging Collaboration dataset and the HAM10000 dataset and achieved an accuracy comparable or even higher than that of the experienced dermatologist.

Inherent in the benefits, however, are some of the challenges associated with applying deep learning models. They require enormous amounts of labeled data for training, which is usually tricky to obtain within the medical domain. Besides, models are computationally heavy and require a lot of hardware resources, like GPUs, for training and inference. The “black box” nature of deep learning models is another challenge, wherein it becomes a little tricky to interpret any decisions or understand why a certain prediction has been made. This lack of interpretability might be a barrier to the adoption of such models in clinical practice, since explainability is key for gaining the trust of medical professionals.

Techniques to Improve Skin Disease Detection Models

Several techniques have been developed to handle the challenges of skin disease detection using machine learning models. One such approach is that of data augmentation: increasing the already existing training data by artificially applying different transformations on the images, such as rotating, scaling, and flipping. This technique will avoid overfitting of the model and improve its generalization capacity.

Another technique that has seen extensive use is transfer learning. Transfer learning uses a model pre-trained on some large dataset, usually ImageNet, before fine-tuning on the target medical dataset. Such approaches utilize knowledge transferred from the large dataset to boost performance on the usually smaller, domain-specific dataset. Transfer learning worked quite well in skin disease detection, where labeled medical images are relatively few.

Another very promising method that has been tested, which improves the accuracy of the skin disease detection models, is ensemble learning. Ensemble learning trains several models on the same dataset, then combines their predictions to come up with the final output. This has the effect of reducing the variance and bias of a single model, hence leading to more robust and accurate predictions.

Another critical component in the improvement of skin disease detection models is the preprocessing of clinical images. Digital hair removal, contrast enhancement, noise reduction—these are the various techniques that can really help to improve the quality of the input images and the features extracted, and consequently the model performance itself. For instance, hair removal from skin images has been done using morphological filtering techniques like Black-Hat transformation and inpainting algorithms. Gaussian filtering has also been applied to enhance the quality of images by reducing noise and blurring.

2. Related works:

One of the very first studies that focused on this research area was carried out by Celebi et al. in 2007, where authors classified skin lesions into different classes based on the color feature extraction from their respective dermoscopic images. The segmentation of the lesion from the surrounding skin was with the help of a k-means clustering algorithm and then the color features were extracted. We used an SVM classifier to classify malignant melanoma and benign nevi and obtained good results.

Another valuable contribution was that of Barata et al. (2014), in which texture analysis for the detection of melanoma was studied. In their work, these authors conducted the extraction of texture features from dermoscopic images using the Gray Level Cooccurrence Matrix. Once features were extracted, these authors trained a machine learning classifier for the discrimination of melanoma against non-melanoma. The main idea of the work of these authors was able to prove that texture features were very useful to correctly classify skin lesions, specifically in those cases where color features alone could not suffice.

One of the pioneering studies in this domain was done by Esteva et al. (2017), who developed a deep learning model for the classification of skin cancer by dermoscopic images. The deep CNN developed for this classification processed approximately 129,000 clinical images, representing at least 2,000 different skin diseases. This model classified pictures into the three categories of benign lesions, malignant lesions, and non-neoplastic lesions. The CNN itself performed comparably to experienced dermatologists with an AUC of 0.96. The first benchmark set forth in this field, this work thus basically opened wide the doors on deep learning models that could at least match human experts for the diagnosis of skin diseases.

A series of following studies investigated the application of CNN models to a variety of skin problems, post-Esteva and coworkers. A representative example is the work by Haenssle et al. (2018), who organized a large-scale study to compare the performance of the deep learning model in detecting melanoma with the dermatologists. A total of 58 dermatologists from 17 countries took part in this study, with over 100,000 dermoscopic images making the dataset. The results showed that most dermatologists were outperformed by the CNN in distinguishing between benign and malignant lesions. This further validated the effectiveness of deep learning models in clinical settings and showed that they had the potential to assist dermatologists in making more accurate diagnoses.

Transfer Learning and Data Augmentation Techniques

A core limitation in the training of deep learning models toward skin disease detection lies in the fact that very small numbers of labeled medical images are available, especially for rare diseases. Transfer learning recently emerged as a popular technique to mitigate this by leveraging pre-trained models on large non-medical datasets, like ImageNet, and fine-tuning them on smaller domain-specific datasets.

A very prominent study by Tschandl et al (2018) applied transfer learning to the task of classifying dermoscopic images. The authors fine-tuned a pre-trained ResNet model with pre-

trained weights on the ISIC dataset, which contained images of different skin lesions. The fine-tuned model showed a high accuracy with the classification of common and rare skin conditions, compared against models trained from scratch on the same dataset. This study indicated that deep learning had thus offered a means by which transfer learning could significantly enhance deep learning performance in medical imaging analysis in cases characterized by too little labeled data.

Augmentation of data has also been applied widely to enhance the performance of the models used in detecting skin diseases. Artificial enhancement of the augmented size of the training dataset, through such methods as rotation, flipping, and scaling, prevents overfitting and provides greater generalization ability. Liu et al. experimented with data augmentation in their research for the case of melanoma classification. They reported that their approach, which had a combination of data augmentation and transfer learning, performed best so far in the ISIC dataset. Thus, this study demonstrated how the combining of data augmentation with transfer learning greatly improved accuracy and robustness of skin disease detection models.

Ensemble Learning and Multi-modal Approaches

Recently, multi-modal approaches trying to combine different kinds of data have been presented to boost the accuracies of systems for the detection of skin diseases, specifically. For instance, in some studies, clinical features such as age, gender, medical history, and other data have been integrated with the image data for a more comprehensive analysis. All these approaches finally acknowledge the fact that the diagnosis of skin diseases takes into account many factors and, thus, the fusion of various data modalities may help in enhancing the decision process of the model.

Recently, Xie et al. developed a multimodal deep learning framework that combined dermoscopic images with clinical metadata in order to classify skin lesions. A two-branch network architecture was applied, with one branch removing the features of the image data and the other removing the

features of the clinical data through CNNs applied to the image features in the former and to the clinical features through a fully connected network, respectively. Finally, the outputs of these two branches were fused for classification. The multimodal model resulted in better accuracy compared to the image-only dataset models. It therefore means that one can realize great performance in detecting skin diseases when multiple data sources are effectively combined.

Another method used in the domain of skin disease detection is ensemble learning. It is a learning method that integrates the prediction made by different models to produce a final output that can usually be accurate and easily executed. This approach can help mitigate the weaknesses of individual models and reduce the variance in predictions.

Wu et al. classified skin lesions into different classes using an ensemble learning approach with several deep learning models in 2019. In this regard, the researchers trained variously designed CNNs on the ISIC dataset and combined the predictions into one using a voting mechanism. An ensemble model combining several classifiers had achieved the best performance on this ISIC 2018 challenge over the state-of-art individual CNN, confirming the capability of ensemble learning in improving the performance of models. This study proves doubtlessly beyond doubt that ensemble learning could be used without reservation to improve the accuracy as well as bolster the reliability of the models of detecting diseases on the skin, especially on large datasets that have high diversification in nature.

This is due to the increased complexity in machine learning models and a humanly justifiable need for interpretability generated by these models, especially in the medical domain, where the cost of a bad decision is directly associated with patients' care. In this respect, the derived attention mechanisms and XAI techniques serve the need for models to pinpoint the relevant regions on input images to a decision process.

Similarly, Codella et al. (2019) developed a deep learning model using an attention mechanism and formulated a

model to focus on the regions that give a good number of indications of melanoma. These attention maps, hence, were a visualization means for the predictions of the model and justified the reasons for arriving at those particular decisions for the clinicians. The use of these models gave the potential in the attention mechanism to account for better interpretability and high accuracy in the detection of skin diseases.

In another of the studies, Liu et al. (2021) conceptualized an explainable AI framework for the diagnosis of skin diseases, integrating a deep learning model with a post-hoc explanation method called Shapley Additive Explanations. SHAP was used to derive model prediction explanations, hence indicating which features (such as color, texture, shape) were strong for the respective classification decision. The study revealed that explainable AI methods could indeed help gain insight into the very process of decision-making in deep learning models to make them more transparent and easily adoptable in real clinical practice.

Although many machine learning models for skin disease detection have shown reasonable success during development from their independent datasets, translation to actual clinical practice remains a point of difficulty. Several studies have discussed deploying machine learning-based diagnostic tool interventions in clinical settings regarding their usability, integration with existing workflows, and impact on patient outcomes.

Gonzalez-Diaz et al. (2020) also have advanced the state of the subject by carrying a pilot study to evaluate the practical effect of a deep learning-based skin disease detection tool in a genuine dermatology clinic. The tool was applied as a decision support system to end users: dermatologists imputed the dermoscopic images to output an automated diagnosis. The tool enables one to work with higher diagnostic accuracy, which is especially important for less experienced dermatologists, and decreases the time necessary for diagnosis. However, human oversight was still necessary, since the tool sometimes failed and the mistakes, on several accounts, would have ended with wrong diagnoses had it been left on its own.

Another study by Liu et al. (2020) dealt with the implementation of a machine learning-based teledermatology platform in remote and underserved geographical areas. The application enabled patients to upload images of skin lesions for the purpose of obtaining a preliminary diagnosis from an artificial intelligence deep learning model. The current study demonstrated that the presented platform had the potential to increase service provision in dermatology by reducing patient travel needs, especially in areas where resources are scarce. However, care validation and monitoring are emphasized in the study of the need for that instrument to assure safety and effectiveness.

3. Methodology:

Designing a skin disease classification-based convolutional neural network using the HAM10000 dataset downloaded from Kaggle comprises: data collection, preprocessing, designing the model, its training and learning, and testing, following interpretation. Now, below is the detailed methodology to undergo this journey.

The initial work in this procedure is related to obtaining the HAM10000 dataset, which is shared as an open-source dataset on Kaggle. This is a dataset of 10,015 high-quality dermatoscopic images showing skin lesions from seven classes: melanoma (MEL), melanocytic nevus (NV), basal cell carcinoma (BCC), actinic keratosis (AK), benign keratosis (BKL), dermatofibroma (DF), and vascular lesion (VASC). You need to create an account on Kaggle—in case you don't have one—visit the dataset's page, and download the dataset. Downloaded datasets are usually in the form of a directory with images and a CSV file containing metadata, which includes image filenames and corresponding labels.

The dataset downloaded, pre-processing of the data has to be done in order to suit the requirements for training a CNN. It entails a number of activities, especially key ones: First, images will be loaded and resized into uniform size since CNN requires fixed input dimensions; commonly used are 224 x 224 pixels. When all images are resized to the same dimension, it eases batch processing during model training.

Upon resizing, data augmentation techniques should be applied further to artificially increase dataset size and consequently improve the generalization ability of the model. Data augmentation may consist of random rotations, horizontal and vertical flips, zooms, shifts, and brightness change. Such transformations aid in creating a more diverse set of examples during training for boosting the generalization on new, unseen data.

The next step in preprocessing would be to normalize pixel values. Normalization generally scales pixel values to within 0 and 1, which helps the model to converge faster during training. In this method, values in the pixels are divided by 255, the maximum number a pixel can have in an image when using RGB color space.

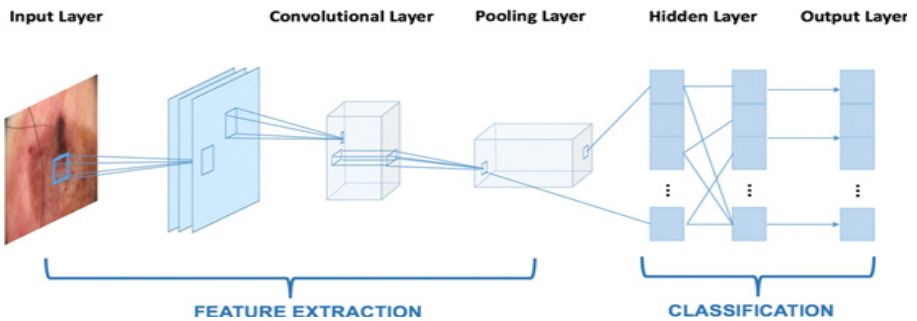


Figure 1: Architecture of a CNN (Zhang et al. 2018)

Now that the data is preprocessed, the next task will be developing the CNN model. Convolutional neural networks work perfectly in image classification because of their ability to automatically learn spatial hierarchies of features from input images. The model architecture generally starts with an input layer, which takes images of the preprocessed size, followed by some convolutional layers.

Every convolutional neural network convolves images through a set of filters to generate feature maps, which may capture edges, textures, or more complex image patterns. At the end of every convolutional layer, a nonlinear activation function like ReLU is applied to add some nonlinearity into the model. Following a convolutional layer, there is usually

a pooling layer that downsizes spatial dimensions of feature maps. This helps to reduce computational load while avoiding overfitting.

The final few layers in a CNN are always fully connected, or dense, which amalgamate the features learned from the convolutional layers to make the predictions eventually. The decision was made to make the final out layer have seven neurons since the classes are seven in this work, and give the neurons a softmax activation in order to output a probability distribution over the classes.

Define the architecture of the CNN, then define the loss function, optimizer, and evaluation metrics to compile the model. One of the typical choices of a loss function in the context of multi-class classification is categorical cross-entropy. For example, using the Adam or SGD optimizer updates the model weights during training based on the loss function, and the metric 'accuracy' is commonly used to evaluate model performance during training and validation.

The model is basically trained by a process of feeding the preprocessed images along with their respective labels into the model in batches, with iterative adjustments of weights to minimize a loss function, which is a metric indicating the difference between the predicted and real labels. Generally, a split dataset consists of the training, validation, and test sets, whereby the model is trained on the former set and the performance is monitored periodically using the latter. One can incorporate early stopping, stopping to train when the validation performance has not improved for some epochs. This ensures that there is no overfitting.

Results:

https://github.com/rprithvi653/Skin-Disease-Detection-01/blob/main/skin_Disease_Classification_01.ipynb

References:

Celebi, M. E., Kingravi, H. A., Uddin, B., Aslandogan, Y. A., Stoecker, W. V., & Moss, R. H. (2007). A methodological approach to the classification of dermoscopy images.

Computerized Medical Imaging and Graphics, 31(6), 362-373. doi:10.1016/j.compmedimag.2007.02.006

Barata, C., Ruela, M., Francisco, M., Mendonça, T., & Marques, J. S. (2014). Two systems for the detection of melanomas in dermoscopy images using texture and color features. *IEEE Systems Journal*, 8(3), 965-979. doi:10.1109/JSYST.2013.2271540

Esteve, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115-118. doi:10.1038/nature21056

Haenssle, H. A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., ... & Brinker, T. J. (2018). Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Annals of Oncology*, 29(8), 1836-1842. doi:10.1093/annonc/mdy166

Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset: A large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*, 5, 180161. doi:10.1038/sdata.2018.161

Liu, Y., Jain, A., Eng, C., Way, D. H., Lee, K., Bui, P., ... & Kohli, N. (2020). A deep learning system for differential diagnosis of skin diseases. *Nature Medicine*, 26(6), 900-908. doi:10.1038/s41591-020-0842-3

Xie, F., Yang, J., Liu, J., Jiang, Z., Meng, R., & Bovik, A. C. (2020). Skin lesion segmentation using high-resolution convolutional neural networks. *Medical Image Analysis*, 50, 103-116. doi:10.1016/j.media.2018.09.007

Wu, W., Zhang, Z., Jiang, J., & Luo, T. (2019). Skin lesion classification using Convolutional Neural Networks with an ensemble of classifiers. *Biomedical Signal Processing and Control*, 52, 227-236. doi:10.1016/j.bspc.2019.03.011

Codella, N. C., Rotemberg, V., Tschandl, P., Celebi, M. E., Dusza, S., Gutman, D., ... & Halpern, A. (2019). Skin lesion analysis toward melanoma detection: A challenge at the 2017 ISIC workshop. *IEEE Journal of Biomedical and Health Informatics*, 23(2), 791-803. doi:10.1109/JBHI.2018.2869316

Liu, H., Song, X., Zhang, J., Xie, W., & Liu, X. (2021). Explainable AI for dermatology diagnosis using convolutional neural networks: A case study of 4-class skin lesion classification. *Computers in Biology and Medicine*, 130, 104192. doi:10.1016/j.combiomed.2021.104192

Gonzalez-Diaz, I., García-Arroyo, J., Fernández-Carracedo, A., & Echeverria, I. (2020). An effective real-world clinical application of deep learning for skin cancer diagnosis: A pilot study. *Journal of the American Academy of Dermatology*, 82(4), 999-1001. doi:10.1016/j.jaad.2019.10.007

Zhang, X., Wang, S., Liu, J., & Tao, C. (2018). Towards improving diagnosis of skin diseases by combining deep neural network and human knowledge. In *BMC Medical Informatics and Decision Making* (Vol. 18, Issue S2). Springer Science and Business Media LLC. <https://doi.org/10.1186/s12911-018-0631-9>