

---

# Dermatologist-Level Classification Of Skin Cancer With Deep Neural Networks

## Introduction

The research paper 'Dermatologist-level classification of skin cancer with deep neural networks' proposes that mobile devices combined with artificial intelligence have the potential of providing "low-cost universal access to vital diagnostic care." This means that there is a rise in technology to enable medical diagnosis in an effective and inexpensive way. To support their claim, the authors utilized a single convolutional neural network (CNN) to predict skin cancer diagnosis by classifying skin lesion images into different skin diseases. The CNN model is meant to mimic human brains by training it with large datasets so that it learns from patterns in the data. The authors' goal was to show readers a real-life and applicable study that proves that artificial intelligence could reach a higher objective. Furthermore, the higher purpose of this algorithm is to track skin lesion and detect skin cancer at an earlier stage, giving patients a longer survival rate than those who are diagnosed at later stages. This study introduces CNNs to a function that could potentially save lives. Finally, more research should be considered in order to fine tune the algorithm and scale it to several medical entities.

## Background

In this study the authors decided to focus on skin cancer detection because of its high death rate. Every year, 5.4 million new cases of skin cancer are diagnosed in the United States. Melanomas, which is the strongest type of skin cancer, represents 5% of all skin cancers, but "they account for approximately 75% of all skin-cancer-related deaths, and are responsible for over 10,000 deaths annually in the United States alone." Patients diagnosed at an early stage with melanoma are estimated a 5-year survival rate, when diagnosed at a later stage this rate can drop to about 14%; therefore, detecting skin cancer at an early stage is critical for patients. The data driven method utilized by the authors allows for medical practitioners and patients to proactively track skin lesions and as a result detect skin cancer earlier.

For many years the process for detecting skin cancer has been done through visual diagnosis, followed by a dermoscopic analysis, biopsy, and a histopathological examination. Visual diagnosis has many problems attached to it, not only are some skin lesions difficult to identify as cancerous through the naked eye, but also it requires a certain level of expertise. Dermoscopy analysis and histopathological examination have proven to be far better than visual inspection due to standardized imagery; however, they still require an expert to differentiate skin lesions. The technique proposed by the authors of this research paper has shown advances in other areas involving visual analysis such as playing Atari games or strategic board games and has even performed better than humans. This method applied to skin cancer diagnosis has the ability to be scalable and overcome variability in images, which make classification more challenging, by training and validating 1.41 million images that create a robust classification algorithm.

## Research Design and Methods

---

In general, to obtain a high accuracy when testing a machine learning model, one must first train the model to find patterns in the data then apply the model to never seen data, called test data. In this study, the data consisted of 129,450 images, of those 127,463 were labelled by dermatologists and were selected to train and validate the model. The remaining 1,942 images were labelled by biopsy results and were destined to test the model. The images labelled by dermatologists were not necessarily confirmed by a biopsy. Therefore, the accuracy of the training and validation is not reliable since dermatologists could have mistakenly categorized a skin lesion as cancerous when it was not, yielding to false negatives. On the other hand, it is safe to say that the biopsy-labelled images were correctly labelled, hence why they were used to test the CNN model. Although the dermatologist labelled data set is somewhat unreliable, it serves a good purpose to find patterns in the data.

It demonstrates the high-level deep CNN process for classifying skin images. Essentially, in each layer the model is detecting similarities by performing regression and learning the correlations between the labels previously stated in the data. The skin image is then classified into 757 fine-grained training classes that yield a probability distribution over these classes. This means that the model outputs the weight of each training class obtained from the skin image. Afterwards, the algorithm allots images to coarser inference classes such as malignant melanocytic lesion, which could be conformed of amelanotic melanoma and acrolentiginous melanoma. The inference classes are accompanied by a probability metric which is calculated by adding the probabilities of the training classes corresponding to the inference class.

This process not only breaks down the images into more digestible classes, but also outputs a tangible metric giving the model the ability to categorize an image with a level of importance and a corresponding accuracy. This is especially important because medical practitioners, when diagnosing, can express the likelihood that a skin lesion is malignant, benign, or non-neoplastic and further assess its composition. Suddenly the original black and white process has more layers. Here is it being showcased the intricacy of the model and its ability to measure results.

The neural network follows a nine-fold cross-validation, this means that the validation dataset is randomly pooled nine times. The idea behind cross-validation is to test the trained model and assess its accuracy before it is tested with unseen data. The CNN in this study was validated in two ways, first, it was partitioned into three categories: benign, malignant, and non-neoplastic lesions, reference first-level nodes. Here the CNN model reached an accuracy of 72.1 +/- 0.9% (mean +/- s.d.), against two dermatologist who acquired an accuracy of 65.56% and 66.00% on a subset of the data. Second, it was classified into nine disease options which followed similar medical treatment plans, reference second-level nodes. In this task, the CNN yields an accuracy of 55.4 +/- 7% against the same two dermatologist who achieved an accuracy of 53.3% and 55.0%. This means that the CNN performs better when trained on finer disease partition and overall better than dermatologists. Though, it is important to note that the main idea behind this analysis is to prove that the taxonomy of skin disease is effective, rather than to compare metrics. In fact, more than showing performance, the accuracy percentages indicate "that the CNN is learning relevant information." With this, the authors prove the value of training models as a vital step to continuing the path of artificial intelligence in medical diagnosis.

Finally, the CNN model was measured utilizing a subset of the test data to have more specific results. The researchers focused on keratinocyte carcinoma and melanoma classification. They utilized two metrics to measure success, sensitivity which refers to the true positive rate (TPR)

---

and specificity which is the true negative rate (TNR). The results were plotted into a receiver operating characteristic curve, known as ROC curve, which captures the trade-off between the TPR and TNR as the discrimination threshold varies. The blue line is the ROC curve, the red dots represent a single dermatologist sensitivity and specificity. Those dots under the curve underperform the CNN, which are the majority. Lastly, the green dot is the average of all dermatologists. This is a pictorial representation that the algorithm studied performs better than experts with never seen data. This test starts conversations about the future success of this model if more data is included and it is scaled to other medical areas.

## Conclusions

The authors were proposing a deep learning model for classifying skin lesions into different classes that would revolutionize the skin cancer diagnosis and potentially expand to other medical areas. They based their proposal on similar acting models and data that had performed better than humans. Such models were applied to image recognition in games. Because of the success in other areas they saw opportunity to apply this technique to the medical world. By utilizing a large data set of images, they created a robust classifying model that has the potential to scale and expand to different medical areas. Throughout the research paper, the authors set up a strong case to diagnose skin cancer using CNN by showing that the more reliable data in conjunction with a well-trained model, has the ability to achieve same or better results than experts.

Moreover, this research paper showed that this technique has the potential to diagnose skin cancer at an early stage which could be the difference between a 5-year survival rate versus 8.5 months. The study proved that the process could be as good or better than the current skin cancer diagnosis by comparing metrics achieved by the model against dermatologists' performance. Furthermore, it also showed that the CNN model has the ability to strengthen and scale to other settings using the appropriate set of data.

## Discussion

This study overall showed a strong case for the expanded use of deep neural networks in dermatology. However, there are some future steps to take in order to better apply this method. Since the testing section was performed using a subset of the data which referred to keratinocyte carcinoma and melanoma classification, the authors acknowledged the need for further research in a real-world setting. This would enable them to confirm that this procedure is scalable to all skin lesions and potentially other clinical areas. It is also important to note that the authors recognize that skin cancer diagnosis is not solely based on skin lesion identification. Though, the fact that this model performed equally as good as dermatologist experts shows the potential for a data driven diagnosis approach to grow medical care access. In addition, the main constraint for this technique is the amount of data obtain, as proven above. Therefore, there is still a continuous need to gather more data. Additionally, if this model was to expand to other medical areas research should be considered to find reliable data.

Furthermore, it is lightly mentioned that this model could be easily included in mobile devices, however, there was no prove or test performed in mobile devices. Another consideration for this research would be to include the use of mobile devices and compare its performance through this technology. As this could potentially impact the results in a real-world setting if the

---

experience is not user-friendly.

## References

1. Esteva, Andre and Kuprel, Brett and Novoa, Roberto A. and Ko, Justin and Swetter, Susan M. and Blau, Helen M. and Thrun, Sebastian. Dermatologist-level classification of skin cancer with deep neural networks. Nature, vol. 542, nature21056 (2017).
2. Masood, A. & Al-Jumaily, A. A. Computer aided diagnostic support system for skin cancer: a review of techniques and algorithms. Int. J. Biomed. Imaging 2013, 323268 (2013).

edubirdie.com