JOURNAL OF
EMERGING INVESTIGATORS

Article

A novel CNN-based machine learning approach to identify skin cancers

Aditya Rao[1], Lakshmi Kondabagil[1]
[1] Cupertino High School, Cupertino, California

SUMMARY

Skin cancer is the most common type of cancer and includes diagnosis procedures that are costly, time-consuming, inaccurate, and inaccessible to many. The goal of this project was to determine the feasibility of a machine learning algorithm to identify skin cancers and compare the results to the conventional procedures of external visual inspection and biopsies. We used the HAM10000 dataset, a diverse collection of multisource clinical images of cutaneous skin pathologies (based on external appearance) which contained 11,034 unique image files of skin cancers and lesions at the time of this project. We tested and trained a machine learning algorithm with the dataset, and analyzed the accuracy, sensitivity, and runtime of the algorithm for seven skin pathologies, which were either skin cancers or pathologies that could potentially develop into skin cancer. The model was created with AutoKeras to automatically search for and apply the best algorithm. The average accuracy of the model (for each skin pathology type) was 84.05%, which exceeds the accuracy of histopathological diagnoses done by experienced dermatologists by 4.05%. For melanoma, the most fatal form of skin cancer, the model had a 70.63% diagnosis accuracy. Furthermore, the average runtime of the model, 4.9775 seconds, provides a significant advantage when compared to the typical minimum time needed to wait for biopsy results. The increased performance of a machine learning model when compared to conventional methods for identifying skin cancer results makes it a feasible alternative.

INTRODUCTION

Skin cancer, the most common type of cancer, is a life-threatening disease. 20% of Americans will develop skin cancer by the age of 70, leading to over 17,500 annual deaths (1, 2). Unfortunately, the diagnosis of skin cancer is not easy. Early-stage diagnosis is crucial, as diagnosis at later stages comes with decreased survival rates and increased costs. Specifically, in the United States, it costs a patient approximately $1,732 to treat stage 1 skin cancer, while it costs over $56,059 (32 times more) to treat stage 4 cancer (3). These costs, especially at a later stage, can be even greater than the median salary in the US - $54,132 as of the second quarter of 2022 (4). In fact, it was found that almost 100% will survive their skin cancer for at least 5 years after diagnosis if it is at stage 1 (5). However, this survival rate is 80% for stage 2, 70% for stage 3, and rapidly drops down to only 30% for stage 4 (5). The differences in the treatment costs and survival rates at various stages emphasize the importance of early diagnosis.

Although different methods exist to detect skin cancers, each of them comes with its own flaws. Previous research on diagnostic accuracy for skin cancers has been done using histopathological methods (microscopic examination of tissue) to diagnose 128 pathologies that were potential cases of skin cancer (6). The diagnosis accuracy for 2 dermatologists (medical practitioners qualified to diagnose and treat skin disorders) with at least 10 years of experience in dermatology was 80%, while the diagnostic accuracy rates for 2 senior registrars (clinical consultants) each with 3 to 5 years of experience and 6 junior registrars each with 1 to 2 years of experience were 62% and 56%, respectively (6). In this study, variation in visual appearance was a major factor in the diagnosis accuracy of melanoma specifically, with junior registrars recognizing thin and intermediate thickness melanoma pathologies with a third of the accuracy of the dermatologists and senior registrars (6). A German study found that 30% of melanoma cases were incorrectly diagnosed by dermatologists at the first medical visit, indicating a 70% diagnosis accuracy for melanoma (7). Specifically, the similar appearance between melanoma and other conditions such as chronic wounds, warts, and fungal infections, lead to initial misdiagnosis of melanoma and a median delay of melanoma treatment of 9 months (7).

Both aforementioned studies used visual inspection along with a conventional method called excisional biopsy to detect melanoma (whether accurately or not) (6, 7). Additionally, biopsies can be destructive to the skin as the suspected area is removed for microscopic inspection. To perform an excisional biopsy, a cut is made through the skin to remove a suspected area for inspection under a microscope for the presence of a disease, using indicators such as tissue integrity and cell maturity. In fact, a small amount of healthy tissue around the abnormal area may also be removed, resulting in the need for stitches and additional expenses (8). Early identification of skin cancer is even more crucial for people with low income, who are unable to afford the expensive treatments needed for later stages of skin cancer. In addition, they may be more susceptible to such diseases due to underlying conditions like malnutrition (9). Another issue with the current biopsy method for detecting and identifying skin cancer is the length of time it takes until the patient can access test results. It generally takes about 2 to 3 weeks to get the results of a biopsy, which also means that potential treatment is delayed, resulting in a worse prognosis (10). Since biopsies to detect skin cancer are costly, inaccessible, time-consuming, and inaccurate, a software and data-based solution is a potential to remedy these inefficiencies.

Thus, the goal of our study was to build a machine learn-

ing model that outperformed conventional medical methods in detecting skin pathologies that can lead to skin cancers and other medical complications. A convolutional neural network (CNN) was selected due to its superior ability for image classifications, in terms of its relatively high accuracy when compared with other machine learning models such as a multilayer perceptron (MLP) algorithm (11). This is because MLPs take vector inputs while CNNs take tensor inputs, meaning that CNNs can generally detect spatial relations between pixels of images better than MLPs (12). For images of skin cancers and lesions, the spatial relations of pixels are key as certain types of skin lesions appear similar in shape and color but have minor differences that only CNNs would identify and consider to correctly diagnose the disease. Another commonly used machine learning model is a recurrent neural network (RNN), which takes in vector inputs – like MLPs but unlike CNNs (13). In addition to the increased efficiency in accessing data inputs (as previously mentioned), CNNs are inherently faster than RNNs, since the data accessing, augmentation, and processing in CNNs can occur simultaneously, while RNNs need to process image inputs sequentially (since the subsequent steps depend on previous ones) (14). Unlike RNNs, in CNNs its built-in convolutional layer reduces the high dimensionality of images since the same filter is applied to multiple locations of the image at the same time, decreasing the average runtime per image (15).

We hypothesized that the machine learning model would have a total accuracy of 80%, which matches the 80% accuracy rate by experienced dermatologists, and that the model would relay the detected skin cancer result in a matter of seconds as opposed to the 2 to 3 weeks needed for biopsy results (5, 10). Our model analyzed images which included one of 7 skin pathologies and cancers: actinic keratosis, basal cell carcinoma, benign keratosis, dermatofibroma, melanoma, melanocytic nevi, and vascular lesions. After creating and running the CNN algorithm, the final average accuracy of the model across the 7 types of skin pathologies tested was 84.05%, and the average run time for each image was 4.9775 seconds. The high predicted accuracy of the model stems from principles of machine learning and data analysis; with the vast amount of data a machine learning model uses, patterns in the results can be easily detected while smaller errors can be disregarded. On the other hand, even the most experienced doctors can only diagnose based on their observation, leading to the possibility of random error and misclassification. Large datasets, such as those used in machine learning, contain far more images of skin pathologies and cancers than a typical dermatologist has seen. The hypothesis for the quick runtime was made because computers can be fed enormous amounts of data and quickly analyze them for patterns based on prewritten instructions. Computers are generally better suited for executing step-by-step machine learning algorithms (specifically CNN models), as opposed to creative or conditionally complicated tasks. Studies by the IEEE and others confirm that computers can use CNN models to analyze images with several convolutional layers, using algorithms like that in this study, within seconds (16).

RESULTS

Using the HAM10000 database from Kaggle.com (17, 18), the model was trained with training data (90% of all images) and tested with testing data (10%) of all images. This means that the test-train split was 10-90. Trained data is fed one at a time so that the model gets an increasingly better idea of what a certain pathology typically looks like, as it sees both the image and the actual (predetermined in the dataset) pathology type. Using the training data images, which all featured one of the previously mentioned pathologies, the model was run to test previously unseen images and gather data. This ensured that data was not being reused between testing and training phases. Before testing and training of the model however, data augmentation and sampling were used to maintain a balance in the number of images per pathology. Data sampling was done so that the model would not be overfit with excessive data from any certain pathology, and data augmentation was performed so that the model would not have insufficient data to make accurate predictions,

After running 50 iterations (epochs) of the machine learning model on the training dataset, the average accuracy for the 7 types of skin pathology was found to be 84.05%. That is, out of all the approximately 10,000 images in the dataset analyzed by the model, 84.05% of them featured skin pathologies that the model correctly identified and classified as one of the 7 possible pathologies found in the dataset (**Figure 1**). These accuracies were calculated by comparing the predicted results of the model to the testing data. Specifically, the accuracy was calculated for each pathology type by dividing the number of images where the model's prediction matched the truly present pathology by the total number of images of the same pathology encountered by the model. In other words, the equation used to calculate accuracy for each pathology was:

$$\frac{TP+TN}{TP+TN+FP+FN} \cdot 100\%$$

The skin pathology types that had the highest accuracies were dermatofibroma and vascular lesions, with accuracies of

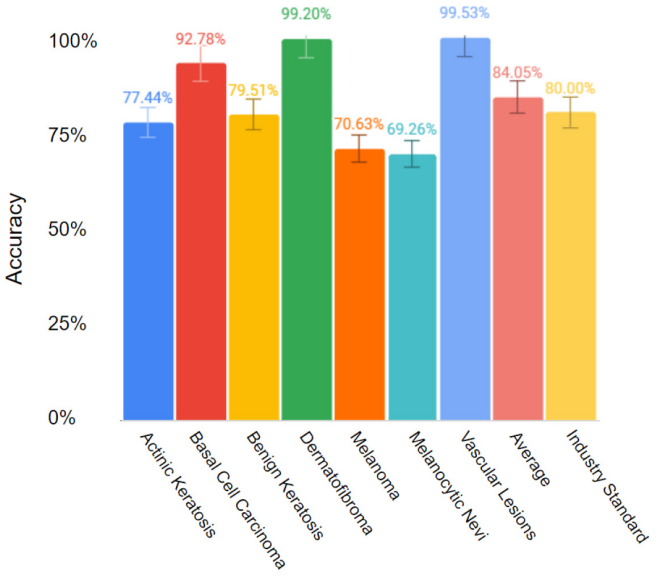

Figure 1: Percent accuracy for each of the seven skin pathology types. The penultimate bar represents the average accuracy across all 7 skin cancer types, which was 84.05%. This exceeds the industry standard of 80% (5) shown in the rightmost bar by 4.05%. Vascular lesions had the highest accuracy (99.53%) and melanoma had the lowest accuracy (70.63%). Data shown as mean ± SD.

99.53% and 99.20%, respectively. The skin pathology types with the lowest accuracies were melanoma and melanocytic nevi, with accuracies of 70.63% and 69.26%, respectively.

The skin pathology type with the highest sensitivity was actinic keratosis, at 0.9268, and the skin pathology type with the lowest sensitivity was benign Keratosis at 0.667. The average sensitivity among all 7 pathology types (also known as the true positive rate, or TPR) was found to be 0.8294 (**Figure 2**). Sensitivity is different from accuracy, which is the ratio of correctly predicted observations to the total observations (19). Accuracy includes true negative values and false positive values, while sensitivity does not. However, accuracy and sensitivity are related. A smaller number of false negative values would increase the sensitivity, or true positive rate. Just as with accuracy, the number of false negatives would decrease with a larger amount of data. Thus, the sensitivity would increase.
The average runtime per image was calculated by the code to be just under 5 seconds, at 4.9775 seconds. This was calculated by dividing the total runtime for the images in all epochs by the number of individual images analyzed.

To evaluate the model for possible overfitting, training and validation loss curves were used to analyze how well it fits the training and new testing data, respectively (20). We noticed that both the training and validation loss curves flattened out and did not diverge from each other at any point (**Figure 3**). In general, we noticed that the accuracy increased logistically (r = 0.732) as the number of original data images (before modification) for each pathology (**Figure 4**).

DISCUSSION

One of the key benefits to a machine learning algorithm is its accuracy. Accuracy is especially important to our CNN model, as it detects potentially life-threatening diseases such as skin cancer. The average accuracy of 84.05% from our machine learning model was 4.05% higher than the accuracy of dermatologists with at least 10 years of experience of 80% (5). Specifically for melanoma, the deadliest form of skin cancer, the algorithm was able to detect the disease with a 70.63% accuracy rate, exceeding the accuracy of the German dermatologists of 70% by 0.63%, although this difference may not be significant (7). Even then, the accuracy of our model was still limited by the data, as inconsistent numbers of images across different skin cancer types meant that sampling and data augmentation techniques had to be used, as opposed to all of the 11,034 images available data in the HAM10000 dataset. That is to say, the data augmentation and sampling in addition to the reduced number of total images used due to these methods, served a net benefit. However, in a more ideal (albeit unrealistic) case, data augmentation nor sampling would need to be used to achieve a balance between the distribution of images across pathologies. This would allow for maximal data inputs, thereby increasing accuracy.

Around 80% of all diagnosed skin cancer cases are basal cell carcinoma, making it not only the most common form of skin cancer but also the most common cancer in the world (21). Our model had a high accuracy for this disease, at 92.78%, making it the third highest accuracy of the seven pathologies/cancers detected by the model. In fact, the 92.78% accuracy is well over the 78.5% accuracy for dermatologists, supporting that such a model could have important clinical implications (22). Comparatively, melanoma is a rarer form of skin cancer, accounting for less than 1% of all cases (23). Due to its high mortality rate, a high diagnosis accuracy is important. Although not the highest accuracy compared to that of other pathologies/cancers detected by our model, the accuracy for detecting melanoma was 70.63%, slightly higher than the 70% accuracy of dermatologists (7). The most accurately detected pathology by the model was dermatofibroma at a rate of

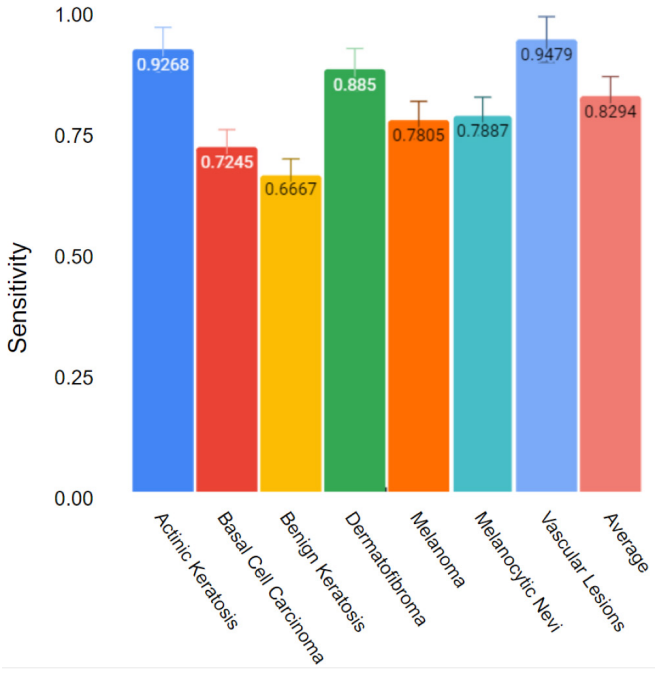

Figure 2: Sensitivity for each of the seven skin pathology types. The rightmost bar represents the average sensitivity across all 7 skin cancer types, which was 0.8294. Vascular lesions had the highest sensitivity (0.9479) and benign keratosis had the lowest sensitivity (0.6667). Data shown as mean ± SD.

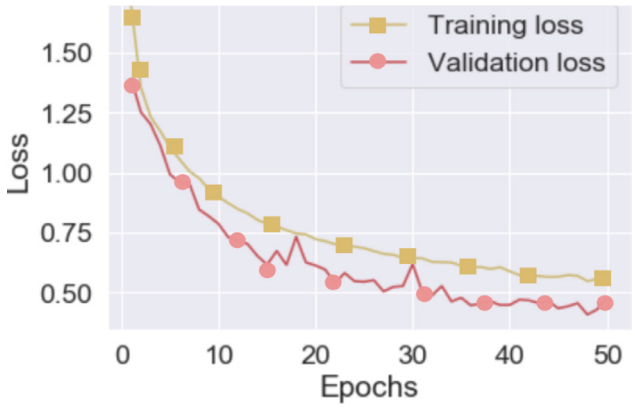

Figure 3: Training and validation loss curves. The training loss curve indicates how well the model has adapted to the training data being fed in, while the validation loss curve evaluates how well the model is curve is generalizing patterns to make future improvements to the model. Since accuracy and loss are inversely proportional, the lower the loss value (y axis), the higher the accuracy and consistency of the model. The number of epochs (x axis) indicates the number of iterations run thus far on the model. Since the curves do not diverge from each other and flatten out as the number of epochs increases, overfitting is at a minimum.

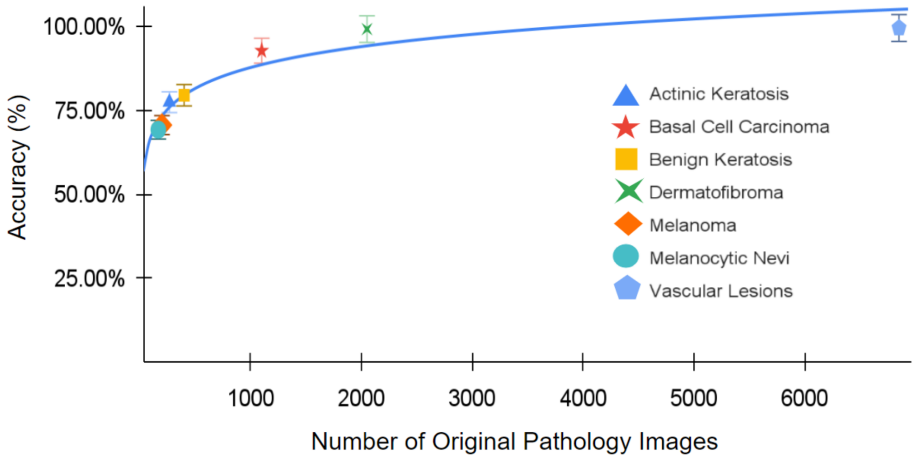

Figure 4: Model accuracy vs original sample size for each of the seven skin pathology types. The blue line is a logarithmic line of best fit; there is a strong, positive, logarithmic relationship (r=0.732) between the two variables. The error bars represent the standard error of the model accuracy for all skin pathologies combined. Data shown as mean ± SD.

99.53%. Dermatofibroma has high real-world applicability given that the pathology is somewhat common. Around 3% of all skin lesions, which are not necessarily dangerous or cancerous lesions, are cases of dermatofibroma (24). Our model's ability to detect dermatofibroma at a high rate is key, given that dermatofibroma patches on women's bodies are a key indicator of risk developing breast cancer (25).

The pathology that was detected with the lowest accuracy by the model was melanocytic nevi (69.26%). The lesion associated with melanocytic nevi is quite common, especially for people with lighter skin, and can lead to a higher risk of melanoma (26). However, its implications are less severe; neurocutaneous melanosis (a broad term for rare neurological disorders) affects 5–10% of people that have a giant congenital melanocytic naevus, the majority of whose cases remain asymptomatic (27). The relatively low accuracy of the prediction of melanocytic nevi can be explained by the smaller number of data images for the pathology, as there were only 167 images that featured melanocytic nevi in the original HAM10000 dataset. Fortunately, the accuracy for all skin cancer/pathology types, including melanoma, can increase with a larger amount of data, especially when such an algorithm is implemented into the real world and more images are gathered through patient diagnosis.

A 0.1 test size was used so that the model would generalize the patterns more accurately as a result of having more data to train with. It would have more images to compare to the patterns in the test images, improving the pattern recognition of the image processing. This smaller test size allowed for an even larger training size (90% of all data), which was especially important given the limited number of images in the dataset.
Overfitting is often a concern with machine learning, which occurs when a function is too closely aligned to a limited set of data points. It causes the model to pick up noise or random fluctuations in the data with higher frequency and learned as patterns instead of variations. Thus, overfitting increases the number of false positives (28). In this case, overfitting could potentially occur for the skin pathology types that had fewer images (melanocytic nevi and melanoma) in the original HAM10000 dataset. This would cause the model to correctly identify the skin pathologies in a small dataset but fail to work for larger datasets. However, by using data augmentation techniques where the data was varied in ways such as changing coloration, stretching, rotating, and blurring, we created a larger and more varied dataset to reduce overfitting (29). Thus, the potential problem of overfitting was limited by creating an artificially larger dataset. Since the training and validation loss curves flattened out as the number of epochs increased, there is no significant overfitting of the model, suggesting its validity (**Figure 2**).

Specifically, this can be explained by the fact that a more varied pool of relevant data diversifies the ways in which data variables (such as pixel density and colors) are learned by the model in the training phase, thus increasing the accuracy. Although data augmentation techniques are not perfect, since they create slightly modified copies of images rather than unique images, giving the model a less diverse pool of images to learn from. However, in general, the use of data augmentation is beneficial for creating a more diverse dataset. While data augmentation is by no means a substitute for real datasets, it is a better alternative to using an insufficient set or subset of data in the absence of a larger one. The model had a lower accuracy when classifying pathologies with an initially small number (melanocytic nevi and melanoma) of data images in the original data set, and thus relied more heavily on augmentation to reach the 2000 images used in the model, had. On the other hand, pathologies that initially had over 2000 images (actinic keratosis, benign keratosis, basal cell carcinoma, vascular lesions, and dermatofibroma) did not have to rely on data augmentation; instead, a sample of the available images was selected for training and testing the model, resulting in a higher accuracy. If data augmentation techniques were not used, the accuracy for pathologies that had relatively fewer images in the original data set (like melanocytic nevi and melanoma) may have been much lower. Indeed, if this project were implemented in the real world, neither overfitting nor the need to create an artificially larger dataset would be a concern; hospitals could add to and access a large, public database and thus increase accuracy.

Compared to our machine learning algorithm's low average runtime per image (4.9775 seconds), the wait time for biopsies to provide test results (2 to 3 weeks) is much slower than the

algorithm to output results (10). It is in the best interest of the patient to be diagnosed quickly to prevent disease progression and defray the cost differences between treating the different stages. The typical wait time for biopsy results is enough time for a skin cancer to escalate to a higher stage of severity. The quicker the skin cancer is identified, the sooner treatment can begin for a patient.

Although no exact statistics on the ancestral or racial distribution of the skin pathology images in the HAM10000 dataset are available, the dataset was put together from a collection of international images, encompassing different skin colors. It can be inferred that cases where the skin color matches the pathology color decreases the accuracy and increases the runtime of the detection. If such a model is implemented into the public healthcare system, as even more data is inputted into the model from various nations and the model's performance improved, a variety of images would potential allow for a high level of performance regardless of background skin color. This would also allow for adaptation of the model for different skin colors and increase the accuracy for each, where different variations of the model could be trained for different skin colors.

One of the primary drawbacks of CNN models may be their strength, depending on the nature of the purpose for their task at hand; CNNs analyze each layer independently at the same time, and the relation between different layers of the image is often disregarded (30). For this project, the issue was not a major one since images of skin lesions are relatively simple, where the lesion is surrounded by skin of a somewhat distinct color.

The results of the algorithm met or exceeded the performance metrics obtained from past research on conventional diagnoses of skin pathologies and cancers and therefore support its feasibility. A novel machine-learning algorithm that diagnoses cutaneous pathologies including, but not limited to, potentially malignant skin cancers can be used to create a more convenient solution to skin pathology identification without the need for traditional medical procedures, which can be inaccurate, time-consuming, and inaccessible. The fast runtime, coupled with the high sensitivity and accuracy of the model, indicates that it may outperform conventional methods. If the software were incorporated into a clinical setting, an increasingly large amount of data would be beneficial to increase the sensitivity and accuracy of the algorithm. Future studies can explore the degree of benefit of a larger amount and variation of data for the model. Further, with the constant addition of data, which a machine learning model uniquely allows for, the model's accuracy and sensitivity metrics could increase at rates potentially faster than healthcare professionals gain expertise. The fact that an automated software solution is cheaper (both for hospitals and patients) and more globally accessible makes it even more favorable for detecting diseases such as skin cancer or identifying skin pathologies prior to developing into more serious complications. The cost for hospitals to use such algorithms would further decrease as the software could be used on the mass market.

MATERIALS AND METHODS

To create the model, the Python programming language was used along with the Jupyter Notebook Integrated Development Environment (IDE). All the 11,034 images and CSV files for the CNN model were taken from the HAM10000 dataset, a publicly available dataset on the website Kaggle.com (17). The images are all of skin pathologies, both benign and malignant. No particular subset was chosen prior to running the model to increase the potential sample size of data; all images were used prior to data augmentation. CSV files were used to read and visualize the distribution of the data for factors like skin localization (which part of the body the skin pathology was located), age, and gender but were not actually used for preparing the data or training and testing the model itself.

Since there were different numbers of image files for each skin pathology, it was necessary to edit the dataset to keep the number of data inputs consistent. The number chosen for this was 2000 images per skin pathology/cancer, as this was approximately the average number of images per pathology. To do this, all data files were separated into classes, randomly resampled, and combined back into a single data file that the code would read. Resampling was done by repeatedly drawing subsets from the training dataset and refitting a particular iteration of the model on each sample. It was key so that the images for all 7 pathology types would be in a random order and easier for the model to quickly access and analyze.

Since the dataset included 7 skin pathology types with fewer than 2000 images, data augmentation was used to artificially generate new images from the originals and thus increase the number of images. The data augmentation was done with the Sequential function of the Keras package. Different augmentations were applied to the image, by individually accessing Keras preprocessing layers and modifying each of them using a random combination of 7 different methods. The 7 used for this model specifically were: RandomFlip, RandomRotation, RandomContrast, RandomCrop, RandomZoom, Resizing, and Rescaling. These processed and modified versions were finally made part of the model using Conv2D function. By modifying the layers of images in this manner, new images were artificially created to add images to any pathology types for which there were less than 2000 images in the original HAM10000 dataset. For pathology types with more than 2000 images, we selected a random selection of 2000 images for analysis.

Since the model would need to train itself while testing its predictions against images that are already known, both training and testing data had to be used. To do so, the test size was 0.1, meaning that 10% of all the data was used to test the model, and the remainder of the data was used to train the model through iteration.

As for the model itself, AutoKeras, an open-source software library for automated machine learning, was used to find and implement the most optimal machine learning model out of 25 possible neural network architectures (31). This was done in tandem with compiling the model with the Adam optimizer algorithm, a popular optimization technique for gradient descent, and the categorical cross-entropy loss function (popular for multi-class classification tasks) to finalize the implementation of the model (32, 33). The specific model used, AutoModel, combines a HyperModel and a Tuner to tune to HyperModel (34). To further increase the efficiency of the model, the MaxPool2D layer and Dropout technique were used to down sample the input by taking the maximum value over the input windows and dropping out some nodes of the network, respectively (35). Next, the SoftMax function, which converts

a vector of numbers into a vector of probabilities, was used to allow for the easy selection of the vector with the 7 highest probabilities for the model (36). The last step was to train the model. The training used a batch size of 200 images for each of the 50 epochs, using the vast majority of the allocated training data. Training and validation loss curves were graphed to check for potential overfitting as well.

After the model was created and run, a confusion matrix was generated using the heatmap function of the Seaborn package, which was ultimately used to calculate the average sensitivity of the model for each of the seven pathology types. This confusion matrix was used to indicate the number of predictions that were classified correctly. True positive and false negative values were used to analyze the performance of the model. A true positive (TP) outcome is one where the model's prediction matches the actual skin pathology present. A false negative (FN) outcome is one where the model's prediction says a truly present skin pathology is not featured in the image. The sensitivity or true positive rate (TPR) was calculated with the following equation:

$$TPR = \frac{TP}{TP+FN}$$

In general, sensitivity is a measure of the proportion of actual positive cases that got predicted as positive (or true positive). Thus, it is also known as the true positive rate, or TPR (19). Specifically, it implies that there would be another group of cases where the model encounters an image that contains a certain disease but believes that the actual disease is not featured.

ACKNOWLEDGEMENTS

We would like to thank the Journal of Emerging Investigator's scientific reviewers and editors for their thoughtful and helpful feedback throughout the revision process.

Received: February 26, 2022
Accepted: July 11, 2022
Published: November 18, 2022

REFERENCES

1. "Skin Cancer Facts & Statistics." *Skin Cancer Foundation.* www.skincancer.org/skin-cancer-information/skin-cancer-facts. Accessed 02 Jan. 2022.
2. "Immunotherapy: The Next Frontier of Skin Cancer Treatment." *Focus*. focus.masseyeandear.org/immunotherapy-the-next-frontier-of-skin-cancer-treatment/ Accessed 21 Dec. 2021.
3. "Skin Cancer Treatment Costs." *Molescope.* molescope.com/blog/cost-skin-cancer. Accessed 20 Dec. 2021.
4. "Median Salary in the U.S." *The Balance.* www.thebalancemoney.com/average-salary-information-for-us-workers-2060808. Accessed 20 Dec. 2021.
5. "Survival." *Cancer Research UK.* www.cancerresearchuk.org/about-cancer/melanoma/survival. Accessed 17 Feb. 2022.
6. Morton, C A, and R M Mackie. "Clinical accuracy of the diagnosis of cutaneous malignant melanoma." *The British journal of dermatology* vol. 138,2 (1998): 283-7. doi:10.1046/j.1365-2133.1998.02075.x
7. Sondermann, Wiebke et al. "Initial misdiagnosis of melanoma located on the foot is associated with poorer prognosis." *Medicine* vol. 95,29 (2016): e4332. doi:10.1097/MD.0000000000004332
8. "Excisional biopsy, N." National Cancer Institute, www.cancer.gov/publications/dictionaries/cancer-terms/def/excisional-biopsy. Accessed 29 Dec. 2021.
9. Walson, Judd L, and James A Berkley. "The impact of malnutrition on childhood infections." *Current opinion in infectious diseases* vol. 31,3 (2018): 231-236. doi:10.1097/QCO.0000000000000448
10. "Tests to Diagnose Skin Cancer." *Cancer Research UK.* www.cancerresearchuk.org/about-cancer/skin-cancer/getting-diagnosed/tests-diagnose. Accessed 12 Jan. 2022.
11. "CNN vs MLP for Image Classification." *Medium*. medium.com/analytics-vidhya/cnn-convolutional-neural-network-8d0a292b4498 Accessed 15 Aug. 2022.
12. "MLP vs CNN vs RNN Deep Learning, Machine Learning Model." *LinkedIn.* www.linkedin.com/pulse/mlp-vs-cnn-rnn-deep-learning-machine-model-momen-negm. Accessed 15 Aug 2022.
13. "Recurrent Neural Networks (RNNs)." *Carnegie Mellon University.* www.cs.cmu.edu/~mgormley/courses/10601-s18/slides/lecture17-rnn.pdf. Accessed 18 Aug 2022.
14. "Deep Learning Tutorial for Beginners." *Simplilearn.* www.simplilearn.com/tutorials/deep-learning-tutorial. Accessed 22 Sept. 2022.
15. "Using Convolutional Neural Network for Image Classification." *Towards Data Science*. towardsdatascience.com/using-convolutional-neural-network-for-image-classification-5997bfd0ede4. Accessed 27 Jan. 2022.
16. Yamashita, Rikiya et al. "Convolutional neural networks: an overview and application in radiology." *Insights into imaging* vol. 9,4 (2018): 611-629. doi:10.1007/s13244-018-0639-9
17. "Skin Cancer MNIST: HAM10000." *Kaggle.* www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000. Accessed 02 Mar. 2021.
18. "HAM10000." *Papers With Code.* www.paperswithcode.com/dataset/ham10000-1. Accessed 02 Mar. 2021.
19. "Sensitivity vs Specificity." Technology Networks. www.technologynetworks.com/analysis/articles/sensitivity-vs-specificity-318222. Accessed 16 July. 2022.
20. "How to use Learning Curves to Diagnose Machine Learning Model Performance." *Machine Learning Mastery*. machinelearningmastery.com/learning-curves-for-diagnosing-machine-learning-model-performance. Accessed 23 May 2021.
21. "Basal Cell Carcinoma (BCC)." *Yale Medicine.* www.yalemedicine.org/conditions/basal-cell-carcinoma. Accessed 09 July 2021.
22. Mohammad, Ebrahimzadeh-Ardakani et al. "Assessment of clinical diagnostic accuracy compared with pathological diagnosis of basal cell carcinoma." *Indian dermatology online journal* vol. 6,4 (2015): 258-62. doi:10.4103/2229-5178.160257
23. "Key Statistics for Melanoma Skin Cancer." *American Cancer Society.* www.cancer.org/cancer/melanoma-skin-cancer/about/key-statistics.html. Accessed 18 Apr. 2021
24. "Cellular Dermatofibroma." *Cleveland Clinic.* my.clevelandclinic.org/health/diseases/22668-cellular-dermatofibroma. Accessed 09 May 2021.
25. Dantzig, Paul I. "Breast cancer, dermatofibromas and ar-

senic." *Indian journal of dermatology* vol. 54,1 (2009): 23-5. doi:10.4103/0019-5154.48981
26. "What Is Melanoma Skin Cancer?." *American Cancer Society.* www.cancer.org/cancer/melanoma-skin-cancer/about/what-is-melanoma.html. Accessed 16 Mar. 2021
27. "Congenital Melanocytic Naevus." *DermNet.* dermnetnz.org/topics/congenital-melanocytic-naevi. Accessed 09 May 2021.
28. "Overfitting and Underfitting With Machine Learning Algorithms." *Machine Learning Mastery.* machinelearningmastery.com/overfitting-and-underfitting-with-machine-learning-algorithms. Accessed 23 June 2021.
29. Perez, Luis, and Jason Wang. "The effectiveness of data augmentation in image classification using deep learning." *arXiv preprint arXiv:1712.04621* (2017).
30. "Introduction to Convolutional Neural Networks." *Aigents.* aigents.co/data-science-blog/publication/introduction-to-convolutional-neural-networks-cnns. Accessed 06 Aug. 2021.
31. "Data Augmentation Techniques." *OpenGenus IQ.* iq.opengenus.org/data-augmentation
32. "Gentle Introduction to the Adam Optimization Algorithm for Deep Learning." *Machine Learning Mastery.* machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning. Accessed 15 Sept. 2021.
33. "Categorical Crossentropy." *Peltarion.* peltarion.com/knowledge-center/modeling-view/build-an-ai-model/loss-functions/categorical-crossentropy. Accessed 15 Sept. 2021.
34. "Auto Model." *Auto Keras.* autokeras.com/auto_model. Accessed 12 Oct. 2021.
35. "Everything You Should Know About Dropouts And Batch-Normalization in CNN." *Analytics India Magazine.* analyticsindiamag.com/everything-you-should-know-about-dropouts-and-batchnormalization-in-cnn. Accessed 12 Oct. 2021.
36. "Softmax Function." *DeepAI.* deepai.org/machine-learning-glossary-and-terms/softmax-layer. Accessed 12 Oct. 2021.