Group 17 Skin Cancer classifier

URN:6616059(James Head), 6592956(Andrei Alexandru),6570834(Aliaksandr Bialiatka)

Introduction

For this project we set out to develop an image classifier that could classify an image of a lesion and predict what type of lesions it is from a list of seven common types (Actinic keratoses(akiec), Basal cell carcinoma(bcc), Benign keratosis-like(bkl), Dermatofibroma(df), Melanoma(mel), Melanocytic nevi(nv) , Vascular lesions(vasc)). One of our key aims for this project was to develop a classifier that not only gave the user a prediction but would also output an image showing the user what features of the image the classifier had used to make its prediction. Our aim is that by implementing this classifier we would be able to implement a functioning prototype that could be used in a real-world setting.

Background on skin cancer diagnosis

The NHS is currently in a crisis with record waiting lists (Jefferies, 2023). This is subsequently having a detrimental effect on the wait times for people getting a skin cancer diagnosis (Garcia, 2022). Thus, the NHS is looking to use AI to help speed up patient treatment to help alleviate this crisis (Barclay, 2023). This is where our image classifier could help speed up skin cancer diagnosis, the quicker a patient can get a diagnosis, the quicker any treatment can start. The NHS is already using AI to help speed up treatment (NHS, 2020). We hope that our classifier could be used to help speed up patient treatment during this current crisis.

AI as a Black Box

One of the key aims for our classifier is to ensure that not only does it give a prediction on the inputted image but we also want to provide an outputted image that will clearly indicate what it used to make its prediction. This is because AI systems are traditionally seen as a black box where most users do not understand how or what the system did to provide its output (Yasar, 2023). Our hope is that any user of our system will have a greater degree of certainty in its prediction because they can be sure if our classifier used the correct part of the image to come to its final decision.

Development

Due to the limited number of resources we had available during the development of this project we decided to use a pretrained model as the basis for our application. This is because we believed it would provide us with an effective starting point with which to start our project efficiently. This is opposed to the time that we would had to spent developing a model from scratch that may well not deliver results as good as these pretrained models and take us far longer to develop.

We decided to develop our application using the HAM10000 data set. This data set has 1015 dermatoscopic images which are split between seven different types of skin lesions. This dataset not only has the images and their lesion type but it also includes the method of diagnoses, the age and gender of the patient the lesion was found on. (Harvards Dataverse , 2018)

Knowing that we were going to be using a pretrained model we had to decide which one would be the best for us to use. To decide this, we implemented a basic image classifier and tested Resnet50, VGG16 and Efficientnet pretrained models to see which one would give us the best accuracy after 5 epochs. A table of our results can be seen below in Table A.

Table A

Pre trained Model	5 Epochs(Percentage accuracy)
Resnet50	77.1%
VGG16	66.2%
Efficientnet	63%

We can see from results in table A Resnet50 clearly returned the highest percentage of correct predictions of the three pretrained models that we tested. We them moved to testing all of the different versions of Resnet(Resnet34, Resnet50, Resnet101 and Resnet152). To see which version of Resnet would give us the highest percentage of correct predictions. Again, we tested each version after 5 and 10 epochs and the results can be seen in Table B.

Table B

Resnet Model	Accuracy Percentage
Resnet18	71%
Resnet34	73%
Resnet50	75%
Resnet101	76%
Resnet152	75%

We can see from these results in Table B that Resnet101 returned the highest accuracy of all the available versions of Resnet.

Improving the Dataset

Upon reviewing the data in HAM10000 dataset we saw that there was a large disparity in the distribution of images between the different classes. We saw that Melanocytic nevi had nearly 7000 images and the next nearest class in terms of numbers was Melanoma with around 1000 images. A table showing this distribution can be seen below in figure 1.

When we looked at the accuracy of our models predictions for each class we saw that our classifier was not ever guessing one of our classes correctly and 4 others it was getting below 50% correct. A table of the accuracy percentage for each class and the number of samples in our test set for our model can be seen below in Table C.

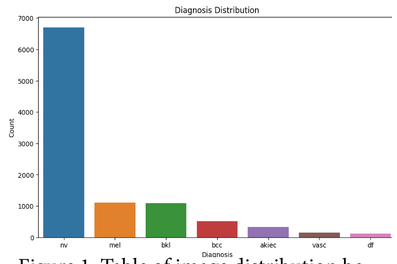

Figure 1. Table of image distribution between classes in HAM10000

Table C

Lesion type	Accuracy percentage	Number of test samples
Actinic keratoses(akiec)	3	61
Basal cell carcinoma(bcc)	35	107
Benign keratosis-like(bkl)	41	224
Dermatofibroma(df)	0	19
Melanoma(mel)	20	211
Melanocytic nevi(nv)	97	1355
Vascular lesions(vasc)	85	26

Seeing that some of our classes had such low numbers of images and that these low numbers were directly affecting the classes prediction, we decided to experiment by transforming images. We did this to a greater and lesser extent depending on the class to create more images for our classifier to train on.

We created three transformation groups. The first group would have small transformations done to them which the Melanocytic nevi class of images would go in. The second group would have all their images transformed once so that there were twice as many images. The Melanoma and Benign keratosis-like classed images would go into this group. The remaining classes (Actinic keratoses, Basal cell carcinoma, Dermatofibroma and Vascular lesions) would go into the last group where their images were copied and transformed 4 times. A graph showing our updated image distribution can be seen below in figure 2.

We then reran the training and testing of our classifier with the same training and test split of our initial dataset (so that the same number of images were included in our test set). The accuracy for each class on the improved data set can be seen in Table D.

Five of our classes saw an improvement in their accuracy, Vascular lesions remained the same and Melanocytic nevi saw a 6% drop. Some of our classes with initially smaller amounts of training data saw a huge jump in their class specific accuracy (Dermatofibroma saw a 63% increase) other classe's saw smaller but still significant increases in their accuracy. Moreover, this led to an overall increase of our model's accuracy by 4% from 76% to 80% for 10 epochs. Compared to the increases we saw in individual classes, 4% is a small increase in the model's overall performance.

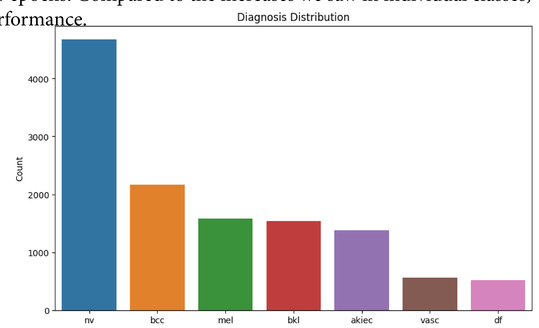

Figure 2. Updated dataset distribution graph

Table D

Lesion type	Accuracy percentage	Percentage improvement from original dataset.
Actinic keratoses(akiec)	44	41
Basal cell carcinoma(bcc)	78	43
Benign keratosis-like(bkl)	50	9
Dermatofibroma(df)	63	63
Melanoma(mel)	49	29
Melanocytic nevi(nv)	91	-6
Vascular lesions(vasc)	85	0

We believe that this was down to the fall in the accuracy for Melanocytic nevi. As this is the class that made up nearly three quarters of the initial dataset, any fall in its accuracy would have a disproportionately large effect on our classifier's overall performance.

These results show that by increasing the images for the classes with an initial smaller number of samples, we can have a substantial improvement on some of the classes specific accuracy. However, this increased accuracy does have a substantial impact on the time taken to train the model. The original data set would take 2 hours to run for 10 epochs. Compared to the expanded data set which took over 6 hours to run for the same number of epochs. Subsequently, this increase had a considerable effect on the resources required to develop our classifier further.

Experiment with the test and train split

We experimented with the distribution of the test and train data that we used to train the model. Our initial split was 80% to train the model and 20% to test it. We changed the split to 70% train and 30% test and added more transformations so that the images for the classes excluding Melanocytic nevi were increased further. This led to only one of our classes seeing an increase in accuracy, all the other classes saw a decrease which meant the overall accuracy for our model fell by 20%. Subsequently, we did not use this test/train split or the increased transformations for the final version of our classifier.

Classifier Output

One of our key aims for this project was to provide a prediction of a lesion type along with an output image that would give the user a clear indication as to what it used to make its prediction. We were able to implement this using heat maps with an imported library from grad-cam. An example of an original image from our dataset and the same image with a heat map applied showing where the model generated its prediction can be seen below in figure 2.

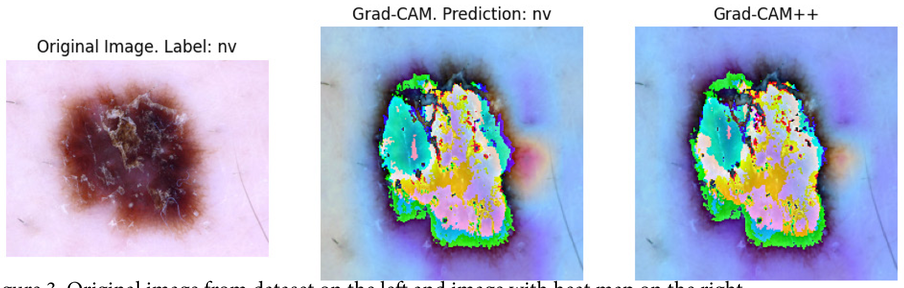

Figure 3. Original image from dataset on the left and image with heat map on the right.

When studying these images side by side we can see that the heat mapped image has the brightest colours directly on the lesion itself and some less prominent shading on the right side of it. Thus, the heat map image has shown that our model is using the lesion to generate its prediction and not another part of the image. This means that any user of our system should have confidence in the model's prediction because they can see that the model is using the correct part of the image to generate its prediction.

Using the metadata

We set about trying to incorporate the metadata that came with each of the images into our classifier. Our logic being that if we could give our model more contextual information it could yield higher levels of accuracy.

We were able to develop a model that would take in the metadata for each sample, study this data and make a prediction using only the metadata. This model was able to give a prediction with an accuracy of 95%. Upon achieving this we tried to integrate it with our image classifier to see if it could increase its accuracy further.

However, when we went about combining the models to work together we found that they both processed data in different orders. Our image classifier would process its dataset in a random order whereas our metadata model would process its data sequentially. Subsequently we were not able to combine the two models to work together and were not able to make use of the metadata that was included in our data set.

Evaluation

If we were to conduct this project again there are several objectives we would look to achieve. The first is that we would want to develop a model that could take an image as well as text data about the patient and where the lesion was found as its input. We believe that this extra contextual information could lead to better predictions as our classifier would have a greater holistic understanding of the patient and lesion rather than just an image. The second objective that we would look to achieve would be to develop a dataset that has a more even distribution of data between classes. Our reason being that if we had a dataset that had a better distribution we would be able to achieve a higher accuracy for the classes that had less images in the dataset.

In conclusion we have been able to develop an image classifier that has an overall accuracy of 80% on the seven types of lesions in our dataset.

References

Barclay, S., 2023. gov.co.uk. [Online]
Available at: https://www.gov.uk/government/news/thousands-of-patients-to-benefit-from-quicker-diagnosis-more-accurate-tests-from-ground-breaking-ai-research
[Accessed 2 5 2023].
Garcia, C. A., 2022. theguardian.co.uk. [Online]
Available at: https://www.theguardian.com/society/2022/oct/02/skin-and-breast-cancer-patients-in-england-waiting-time-increase-nhs
[Accessed 12 5 2023].
Harvards Dataverse , 2018. dataverse.harvard.edu. [Online]
Available at: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T
[Accessed 12 5 2023].
Jefferies, D., 2023. kingsfund.co.uk. [Online]
Available at: https://www.kingsfund.org.uk/blog/2023/01/nhs-crisis-patients-now-waiting-longer-every-type-emergency-care
[Accessed 12 5 2023].
NHS, 2020. https://www.cuh.nhs.uk/news. [Online]
Available at: https://www.cuh.nhs.uk/news/ai-speeds-cancer-treatment/
[Accessed 12 5 2023].
Yasar, K., 2023. techtarget.com. [Online]
Available at: https://www.techtarget.com/whatis/definition/black-box-AI#:~:text=Black%20box%20AI%20is%20any,to%20how%20they%20were%20reached.
[Accessed 12 5 2023].