Skin Cancer Image Classification using Convolutional Neural Networks

Nathan Pogue, Tara Chen, David Chen, Ahmed Mohamed, Ajith Rahugnanam

Abstract

Skin cancer is the most common cancer worldwide and has become a major public health issue due to its widespread prevalence and the ambiguity of clinical diagnoses based solely on phenotypic characteristics. Early diagnosis is needed to improve the treatment outcomes. The clinical diagnosis is subject to the discretion of select oncologists. Deep learning has proven to be effective in automatic, high-level feature extraction without prior knowledge. Therefore, we propose the analysis of histopathological images of skin cancer using the Inception V-3 architecture for supervised deep convolutional neural network as a classification tool that can support clinical diagnoses as an auxiliary objective tool. We adapted the Inception V-3 architecture to support multi-class image classification and improved robust classification through geometric image augmentation methods. Our approach reveals that the model has an accuracy of 80%, precision of 85% and recall of 80%. Our experimental results demonstrate that the Inception V-3 architecture is a novel, auxiliary mode that can be applied in the use case of histopathological image classification of skin cancer and poses the potential to further fine tune the model in collaboration with oncologists with a priori knowledge of subsets of characteristic features.

Introduction

Artificial intelligence (AI) can be used for image classification tasks with applications in the medical imaging field. Typically, radiologists visually assess medical images to provide diagnoses and monitor diseases.[1] Convolutional neural networks (CNNs), a subset of machine learning in AI, can recognize patterns in images and provide quantitative assessments of data. CNNs aim to supply more accurate image analyses and reduce workload for radiologists in order to provide more efficient clinical care for patients.[1] Specifically, CNNs can be used to analyze skin lesions and classify skin cancer in patients. This model can be implemented as a web application to provide the public with access to convenient analysis of skin lesions. This report covers key concepts regarding the development of the project including a general background on CNNs, methodology and a discussion of results.

Convolutional Neural Networks

Convolutional neural networks are a class of deep neural networks composed of multiple layers of neuron-like nodes that receive input across adjustable weights. It is a mathematical construct designed to adaptively learn spatial hierarchies of images using multiple filters of increasing complexity. CNNs are comprised of three types of layers: convolution, pooling and fully connected layers; Each layer evaluates input data differently and has specific parameters that can be optimized to increase the accuracy of the model.[2]

Convolution layers extract features from the input image and can perform edge detection, colour detection and sharpening by applying different types of filters or kernels. A kernel is a parameter that is optimized during the training process; The size of the kernel, number of kernels, stride, activation function and padding are all hyperparameters that are set beforehand. A stride is the number of pixels the kernel moves over the input image and a rectified linear unit (ReLU) may be applied to introduce non-linearity to the layer. If the filter does not fit the input image perfectly, the image may be padded with zeros at the edges to avoid losing information at the borders.[2]

Pooling layers perform downsampling of data to reduce the dimensionality of feature maps while retaining important information. Hyperparameters include the specific pooling method, filter size, stride and padding. The two types of pooling are max pooling and average pooling. Max pooling returns the largest element from the feature map, whereas average pooling returns the average of all elements from the feature map.[2]

Output feature maps from the final convolution or pooling layer is flattened into a vector which is a one-dimensional array of numbers. The vector is then attached to fully connected layers where every input is connected to every output through a learnable weight. Finally, extracted features are mapped by a subset of fully connected layers to the final outputs of the network. Softmax is typically applied to the very last layer to convert the output into a probability distribution for the image classification task.[2]

Methodology

This model was based on the Kaggle Kernel “skin lesion Inception V3”. The HAM10000 dataset from Kaggle was used to develop the skin cancer classification model and it consists of 10015 dermatoscopic images organized into several diagnostic categories: actinic keratoses and intraepithelial carcinoma, basal cell carcinoma, benign keratosis, dermatofibroma, melanoma,melanocytic nevi and vascular lesions.[3] The dataset was visualized on graphs based on the age of the patient and amount of images per diagnostic category. This revealed that there

was a significantly greater amount of images classified as melanocytic nevi compared to other minority classes. Additionally, the data frame was split to create the training, validation and test sets.

The data was augmented prior to training the model. The images were randomly rotated 90°, zoomed in by a factor of 0.1, shifted horizontally by 10% and shifted vertically by 10%. Next, the convolutional neural network was built using the Inception V3 model which consists of several symmetric and asymmetric blocks including convolutions, max pooling, average pooling, Dropout, Dense and fully connected layers. BatchNorm was applied throughout the model to activation inputs and loss was computed with Softmax. Once the model was fit, the accuracy and loss of both the training and validation dataset were plotted on a graph to assess the model's performance. The test accuracy was calculated and a confusion matrix was created to summarize the performance of the classification algorithm. Lastly, a classification report revealed the precision, recall and f1-score values for each diagnostic class along with weighted averages.

Layer (type)	Output Shape	Param #
inception_v3 (Model)	(None, 2048)	21802784
dropout_4 (Dropout)	(None, 2048)	0
dense_4 (Dense)	(None, 128)	262272
dropout_5 (Dropout)	(None, 128)	0
dense_5 (Dense)	(None, 7)	903

Total params: 22,065,959
Trainable params: 22,031,527
Non-trainable params: 34,432

Figure 1. Summary of the model.

Results

The graph of the training and validation loss of the model reveals that the validation dataset experienced less loss compared to the training dataset. The graph of the training and validation accuracy of the model reveals that the validation dataset had a greater accuracy than the training

dataset. The classification table reveals that the model has an accuracy of 80%, precision of 85% and recall of 80%.

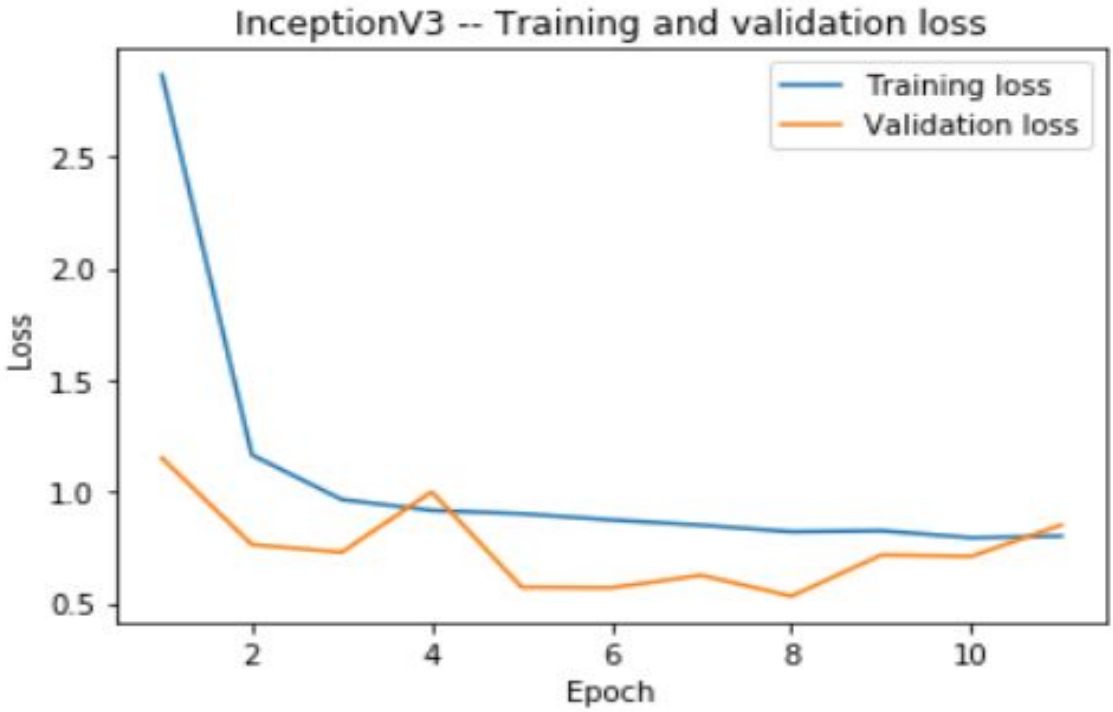

Figure 2. Graph of the training and validation loss of the model.

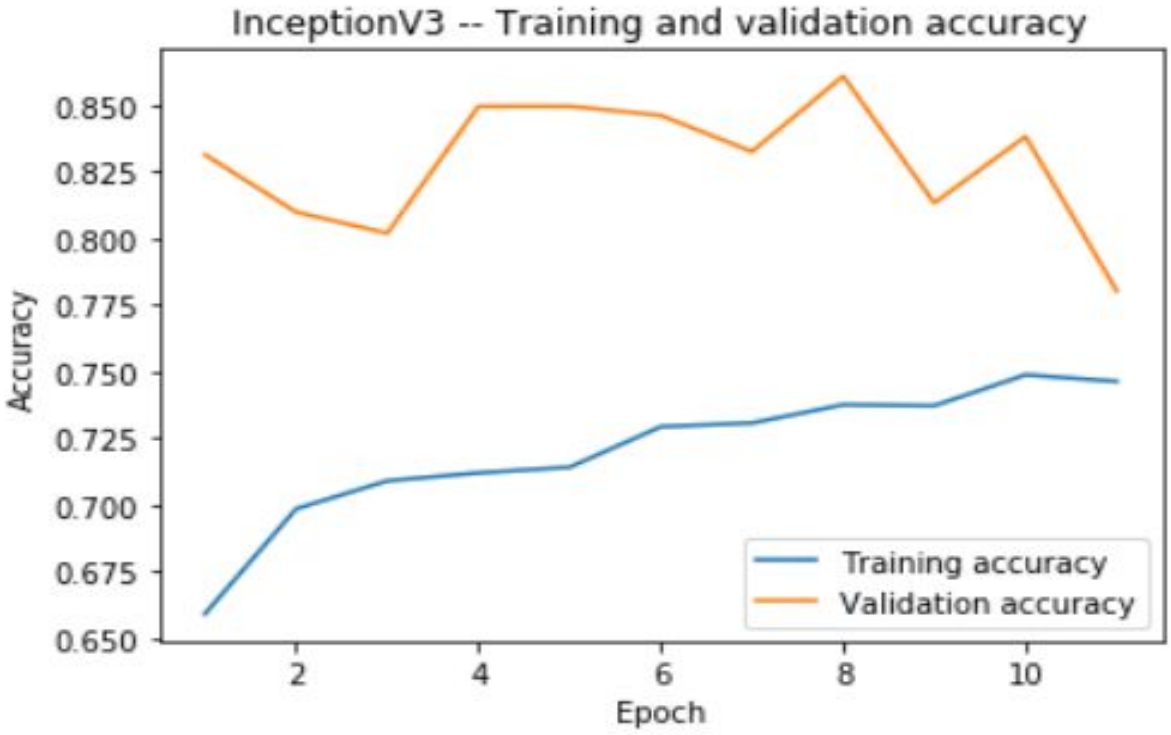

Figure 3. Graph of the training and validation accuracy of the model.

	precision	recall	f1-score	support
bkl	0.32	0.86	0.46	87
nv	0.96	0.90	0.93	886
df	0.00	0.00	0.00	8
mel	0.29	0.22	0.25	36
vasc	0.00	0.00	0.00	14
bcc	0.62	0.12	0.21	40
akiec	0.50	0.06	0.11	32
accuracy			0.80	1103
macro avg	0.38	0.31	0.28	1103
weighted avg	0.85	0.80	0.80	1103

Figure 4. Classification table of the model.

Discussion

In this study, skin cancer histopathology image classification through an augmented CNN model is proposed. We show that our deep learning approach in silico can be used in conjunction with clinical diagnoses as an auxiliary mode of diagnostic confirmation, which can prove to be useful in ambiguous situations that require further investigation. Our front-end user interface maintains a computationally lightweight architecture so that the user experience is streamlined with a pre-built model for ultra-fast and accurate image classification.

We note that the issue of class imbalance in the training sets of supervised artificial intelligence classification models can lead to lack of generalizability to real world data due to the failure in predicting minority classes. This issue is especially apparent in models that fail to consider weighted error measures relative to the class size of majority and minority testing sets. We intend to use a hybrid method of undersampling the majority class and oversampling the minority class with 10-fold cross validation to achieve more representative model predictions in future iterations of the model development.

Furthermore, we suggest that the variation in classification accuracy and the related metrics of precision and recall can be due to: 1) the inherent variation of cross validation splitting of testing and training sets, 2) the variation due to the lack of standardization of image quality included in the dataset, and 3) the partly random nature of phenotypic tumour development that can lead to outliers of one class being predicted as another class. In particular, the evolution of cancer across progressive clinical stages suggests the inherent variation within individual classes. We note that

the dataset did not include metadata that labelled cancer stage within cancer type, which poses the potential for future unsupervised machine learning approaches to predict within-class subclusters of tumour stage.

In the future, we would involve experienced oncologists to guide our deep learning model design. The deterministic nature of model selection and parameterization can make it difficult for the model to accurately capture the nuanced feature sets of skin cancer in real world scenarios. By including expert decisions in model architecture setup, we aim to highlight key feature sets of this image classification problem that can predict the complex, synergistic interplays that lead to the phenotypic effects displayed by skin cancer. We intend to compare the difference between model parameters and clinical evidence a priori to bridge the gap between oncology and deep learning in an increasingly computationally-driven world.

Conclusion

Automatic classification of skin cancer images using convolutional neural networks is proposed. The methodology is based on convolution of an input image with a linear filter across multiple subregions, applying a multi-layer neural network model architecture based on non-linear activation functions with biased weights, then outputting a prediction of the class that the input image belongs to. The accuracy of the CNN model was reported to be 80%. The precision of the CNN model was reported to be 85%. The recall of the CNN model was reported to be 80%. Results suggest deep learning models can be used as an auxiliary mode for clinical diagnostic confirmation based on input images and poses the potential for more accurate models in collaboration with oncologists with nuanced understandings of the capricious nature of cancer.

References

1. Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., & Aerts, H. J. W. L. (2018). Artificial intelligence in radiology. *Nature,* 18, 500-510. doi: 10.1038/s41568-018-0016-5.
2. Yamashita, R., Nishio, M., Do, R.K.G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in biology. *Insights into Imaging,* 9, 611-629. doi: 10.1007/s13244-018-0639-9.
3. Tschandl, P., Rosendahl, C. & Kittler, H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci. Data 5, 180161 (2018). doi: 10.1038/sdata.2018.161.