**HOTARU Model Creation Progress Log**Machine Learning Engineer-in-Charge: Nicholas Dale (falconpunch082)

**About**HOTARU (short for Health Observing Technology for Assessing Risks Unveiled) is a convolutional neural network designed to determine from a picture provided by an end-user whether said picture shows signs of skin cancer (defined in this model as melanoma, basal cell carcinoma and squamous cell carcinoma).

HOTARU is accompanied by SPARK (short for Skin Pattern Analysis and Recognition Kit), a pre-processing Python script that serves to analyse a picture by standardising the picture, isolating the skin-spot, and outputting a pre-processed image of the skin-spot to enable more accurate and efficient determination for HOTARU. While HOTARU can run without using SPARK (granted that there is still a barebones preprocessing script that resizes the image into the size HOTARU accepts, 300x300), SPARK is meant to empower HOTARU to predict accurately, considering the risk of misidentification of skin cancer.

**Data Collection and Preprocessing**Image data was collected from the following websites:

* <https://www.kaggle.com/datasets/ahmedxc4/skin-ds>
* <https://www.kaggle.com/datasets/ismailpromus/skin-diseases-image-dataset>
* <https://www.kaggle.com/datasets/subirbiswas19/skin-disease-dataset>
* <https://api.isic-archive.com/collections/70/> (Challenge 2018, 2019 and 2020)

38663 images containing skin cancer and 14669 images containing non-skin cancer features were gathered.

Through a notebook, training, validation, and test datasets were created with the varying splits based on version. In later versions, this notebook is also where the datasets were pre-processed.

**Model Lifecycle**The table below shows the lifecycle of both HOTARU and SPARK, alongside the notebook hosting both programs.

|  |  |  |
| --- | --- | --- |
| HOTARU Version | SPARK Version | Notebook |
| 1.0 | - | production1.ipynb |
| 1.1 | 1.0 | Production2.ipynb |
| 1.2 | 1.1 | Production3.ipynb |

**HOTARU v1**

Overview

HOTARU v1 was the very first iteration of the CNN model. This model was fed unprocessed images straight from the dataset. Furthermore, SPARK was not introduced at this stage – a preprocessing stage baked into the model was introduced instead.

Model design  
This version featured two sections of the model, which were:

* A Sequential object that encased the CNN, which included:
  + A 16-filter Conv2D input layer with a (3,3) kernel, relu activation and an input\_shape parameter of (300,300,3) followed by a MaxPooling layer,
  + 2 more Conv2D layers (filters were 32 and 16 respectively), each followed by a MaxPooling layer.
  + A Flatten layer for transition from CNN to Dense layers,
  + A Dense layer with 256 neurons and relu activation, and
  + A Dense output layer with one neuron and sigmoid activation (output ranged from 0 to 1, with 0 being non-skin cancer and 1 being skin cancer.), and
* A preprocessing section which included an Input layer of shape (300,300,3) and Resizing layer that converted the input into a size of (300,300).

This model was then compiled and fitted with the following parameters:

* Binary\_crossentropy loss,
* Adam optimiser,
* 20 epochs

Data Preparation  
Training, validation and test datasets were created from the raw dataset in the ratio of 80:10:10 respectively. These photos were reshaped to multi-channel images of size 300x300. No other preprocessing followed.

Training Results  
It took nearly 5 hours and 15 minutes to complete training. Upon testing the model with the test dataset, it was determined that the model boasted 80% accuracy, precision of 71.4% and recall of 100%. This model was saved as ‘prod\_1.h5’.

Discussion and Further Action  
Moving forward, it was hypothesised that a preprocessing script would greatly increase the accuracy of the next version as the model would not consider unimportant features like healthy skin and hair when determining whether one had skin cancer or not. Moreover, the preprocessing section of the model was deemed to be redundant as the data entering the model was already sized to be compatible.

**HOTARU v1.1 and SPARK v1.0**

Overview

HOTARU v1.1 was the second iteration of the model. This model also introduced the very first version of SPARK, the preprocessing script. As such, this model was fed with pre-processed images that mostly featured important skin spot features.

Model design  
This version only featured the Sequential object found in HOTARU v1.0; the preprocessing section was removed entirely. No changes were made to the Sequential object, meaning that the CNN in v1.1 is the same as v1.0

This model was then compiled and fitted with the following parameters:

* Binary\_crossentropy loss,
* Adam optimiser,
* 10 epochs

Data Preparation  
Training, validation and test datasets were created from the raw dataset in the ratio of 90:7.5:2.5 respectively. These photos then underwent preprocessing using SPARK.

SPARK v1.0 featured the following stages:

* Resizing the input image into a 300x300 image to preserve information while still being standardised for the model,
* Normalisation of pixel values,
* Creation of a grey-scaled version of the image for thresholding and masking to segment the skin spot from the rest of the skin, and
* Combining the mask and the original image together to create a processed image displaying only the skin-spot (the background was black).

Training results  
Unlike v1, v1.1 was created after an hour of training. Furthermore, upon testing the model, it boasted an accuracy of 97.2%, with precision of 97.2% and recall of 98.9%. This model was saved as ‘prod\_2.h5’ and ‘prod\_2.pkl’.

Discussion and Further Action  
Upon reviewing the dataset, it was hypothesised that if the dataset was to be cleaned up so that it only contained pictures of skin (and not body parts like genitals), accuracy would improve as SPARK, and in turn HOTARU, would be more efficient in fulfilling their purpose. In addition, further inspection of the processed photos shows that SPARK v1.0 tended to include body hair into the mask, which resulted in the addition of an unnecessary feature. Moreover, considering that a lot of time was shaved off from training, it is now plausible to add in more layers to HOTARU to make it more accurate without consuming too much time and compute power. Therefore, changes were made to address these hypotheses.

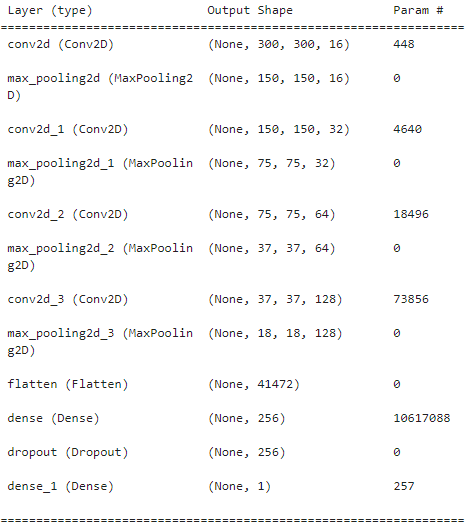
**HOTARU v1.2 and SPARK v1.1**

Overview

HOTARU v1.2 was the third iteration of the model. Improvements were made in terms of data preparation and complexity of the model to achieve better real-life accuracy, in exchange for a slower processing time and the increased likelihood of false positives.

Model design  
This version featured improvements to the CNN. It introduced the parameter padding, which was set to ‘same’, to all Conv2D layers. In addition, both the number of Conv2D layers and filters in each Conv2D layer were altered, and another Dense layer alongside a Dropout layer were added at the end of the Sequential object. The output layer remained the same.

The following is a screenshot of the structure of the CNN.



This model was then compiled and fitted with the same parameters as in HOTARU v1.1.

Data Preparation  
Prior to partitioning the raw dataset, the raw dataset was reviewed, and unsatisfactory photos were removed, resulting in the removal of around 11 thousand photos. While the previous versions featured a cancer:non-cancer picture ratio of 2.6:1, this version would learn from a dataset which featured a ratio of 11.7:1 (meaning for every 12 photos with skin cancer featured, there also existed a photo not featuring skin cancer). As such, the disparity between the two classes was exacerbated; the skin cancer class was being oversampled more than before. While this will inevitably result in the appearance of false positives, considering that the use case of this model, it is better to get something checked and find out it is not cancer, than being given a false negative, only to suffer from misidentification later on.

Training, validation and test datasets were created from the improved raw dataset in the ratio of 90:7.5:2.5 respectively. These photos then underwent preprocessing using a new version of SPARK.

SPARK v1.1 featured the same stages as v1.0, except that the masking process was enhanced with a DullRazor hair removal algorithm. In v1.0, the masking process would include body hair, adding unnecessary noise when training the model. This is fixed in v1.1.

Training results  
During the creation of HOTARU v1.2, F1-score was added as an additional metric. In addition, the evaluation now included a confusion matrix and a classification report to further scrutinise v1.2’s performance.

Training took an hour and 15 minutes to finish. By the end of training, the model provided the following statistics: accuracy of 90-92.1%, with precision of 93.2%, recall of 98.78% and F1-score of 95.9%. When compared to HOTARU v1.1 using the same test dataset (which yielded an accuracy of 90-92.65%, precision of 93.4%, recall of 98.9% and F1-score of 96.1%), it is evident that v1.2 performed slightly worse than v1.1. Furthermore, the confusion matrixes of both versions were nearly identical.

Discussion and Further Action  
The addition of more layers, the cleanup of the raw dataset and the implementation of a new version of SPARK did little to improve the performance of HOTARU, despite all of those actions initially being seen as capable of improving performance. Upon looking at the pre-processed photos, it was observed that some photos were not processed as expected (e.g., the feature was completely left out, leaving a halo of healthy skin). Thus, an attempt was made to improve on SPARK again to hopefully improve HOTARU’s performance.

**HOTARU v1.25 and SPARK v1.2**

Overview

HOTARU v1.

Model design

Data Preparation

Training results

Discussion and Further Action