**ISIC 2019: Model Building**

**Overview:**

This report summarizes the models created for classifying the 8 classes of cancers for the ISIC 2019 dataset. The metric used for evaluating the performances of the models is “Balanced accuracy”. The same metric has been used to evaluate the models on the ISIC leaderboard. A sample screenshot of the balanced accuracies obtained by the participants of the ISIC 2019 challenge has also been included in this report as a comparison for the performance of the models created by our team. All models are evaluated for a new data or “test data” which is not seen by the model during its training and validation (15% of the initial training data totaling to 3799 images retained during the train test split). This test data is only resized to the input image size of the respective model and is not subjected to any preprocessing or hair removal.

**Balanced Accuracy:** Normalized (or balanced) multi-class accuracy is defined as the accuracies of each category, weighted by the category prevalence. Balanced accuracy is a metric we can use to assess the performance of a classification model. [[1]](#footnote-1) Specifically, it is the arithmetic mean of the (<category>\_true\_positives / <category>\_positives) across each of the diagnostic categories. This metric is semantically equivalent to the average recall score.[[2]](#footnote-2)

Accuracy doesn't make us see the problem with the model. Here, model positives are represented well. So, in a case like this, balanced accuracy is better than accuracy. **If the dataset is well-balanced, Accuracy and Balanced Accuracy tend to converge at the same value.**[[3]](#footnote-3)

**ISIC 2019 Leaderboard scores:[[4]](#footnote-4)**

Graphical user interface

Description automatically generated

Graphical user interface, text, application

Description automatically generated

**Model 1: Research paper**

*Reference File: “model\_researchpaper\_processed.ipynb”*

Train Summary:

Image size: 128 x 128

No. of training images: 72096 (Augmented, balanced hair removal) (8 classes)

No. of validation images: 3799 (Preprocessed - Hair removal) (8 classes)

No. of test images: 3799 (8 classes)

No. of Epochs: 50

We use a total of 72096 images of skin cancer belonging to 8 classes. These images are obtained after balancing the skewed classes by augmentations and appropriate preprocessing. The validation images used has also been subjected to the same preprocessing as the training images.

Model Structure:

Text

Description automatically generated with medium confidence

Training metrics: After 50 epochs

loss: 0.4458 - accuracy: 0.8294 - val\_loss: 1.0900 - val\_accuracy: 0.6570

Confusion matrix and classification report:

Calendar

Description automatically generated

**Overall Model Accuracy / Balanced Accuracy: 0.0526**

Observations: A prediction accuracy of 0.0526, training accuracy of 0.8294 and validation accuracy of 0.6570 is obtained for this model structure.

No overfitting is observed during the training and after the completion of 50 epochs as we could see a gradual decrease in the validation loss and increase in the validation accuracy during training.

Con: Here, we are providing 3799 validation images for a 72096 training images which constitutes to only about 5% of the training data. The training can be further improved by providing more percentage of validation images with respect to the training data.

**Model 2: Research paper (Validation data is given as a 25% split from training data)**

*Reference File: “model\_researchpaper\_processed\_2.ipynb”*

Train Summary:

Image size: 128 x 128

No. of training images: 54072 (Augmented, balanced hair removal)

No. of validation images: 18024 (25% split from training data)

No. of test images: 3799

No. of Epochs: 50

Model Structure: Same as Model 1.

To overcome the drawbacks of model 1, we ran another trail on the same model maintaining all the characteristics of the data same as the model 1, only changing the percentage of validation images provided during the training. Here, we provide about 18024 processed images as validation images to the model for training, given as a 25% split from 72096 training images. As a result, we have 54072 training images and the remaining 18024 used as validation for training.

Text

Description automatically generated with medium confidence

Training metrics: After 50 epochs

loss: 0.3995 - accuracy: 0.8488 - val\_loss: 2.1264 - val\_accuracy: 0.5623

Confusion matrix and classification report:

Calendar

Description automatically generated

**Overall Model Accuracy / Balanced Accuracy: 0.0750**

Observations: This trail is carried out for providing enough number of validation images to evaluate the fit of the model. The training accuracy is found to increase to 0.8488 from 0.8294 but we see a slight decrease in the validation accuracy from 0.6570 (model 1) to 0.5623.

However, the balanced accuracy for the test data increased from 0.0526 to 0.0750 for providing sufficient number of validation images.

**Effect of Preprocessing / Hair removal on the model performance:**

**Model 3: Research paper (Both training and validation datasets are unprocessed)**

*Reference File: “model\_researchpaper\_without HR.ipynb”*

Train Summary:

Image size: 128 x 128

No. of training images: 72096 (Augmented, balanced and **NO** hair removal) (8 classes)

No. of validation images: 3799 (**NO** hair removal) (8 classes)

No. of test images: 3799 (8 classes)

No. of Epochs: 50

Model Structure: Same as Model 1

All characteristics of this model, and the number of training and validation images remain same as model 1. The only difference in the training and validation images provided to this model is that they are not subjected to the hair removal preprocessing as they were in model 1.

Training metrics: After 50 epochs

loss: 0.6791 - accuracy: 0.7498 - val\_loss: 1.3141 - val\_accuracy: 0.5969

Confusion matrix and classification report:

Calendar

Description automatically generated

**Overall Model Accuracy / Balanced Accuracy: 0.1112**

Observations: There are two key points to note from this model:

1. The training accuracy of this model after 50 epochs is 0.7498 and validation accuracy is 0.5969 as opposed to the training and validation accuracies of model 1 which are 0.8294 and 0.6570 respectively. We can observe that for the data without hair removal, as in this model, the training and validation accuracies lowered significantly as compared to model 1 in which the data was processed by hair removal.
2. The balanced accuracy of this model is 0.1112 as opposed to the balanced accuracy of 0.0526 of model 1. Here, we can see an increase in the prediction accuracy of the new data. This increase in accuracy for the new data can be attributed to the fact that the test data is also unprocessed similar to the unprocessed training and validation data used in this model. It could be possible that the model has learnt the hairs as a feature which is helping in better predicting the test data which also has hairs in the images.

**Model 4: Transfer Learning – NasNetMobile (Fine Tuned)[[5]](#footnote-5)**

*Reference File: “model\_NasNetMobile.ipynb”*

Train Summary:

Image size: **224 x 224**

No. of training images: 57680 (Augmented, balanced hair removal)

No. of validation images: 14416 (20% split from training data)

No. of test images: 3799

No. of Epochs: 30

Model Structure:

Next, we try to use transfer learning techniques to classify the 8 classes of skin cancer. We start by developing a smaller architecture, so that the hardware configuration available at hand, is able to train the model. The feature vector of NasNetMobile model trained on ImageNet dataset is imported and fine tuned for our training images. The methodology to use this transfer learning model and fine tuning is available in the documentation of this model provided in the footnote.

**Text

Description automatically generated**

Training metrics: After 30 epochs

loss: 0.2906 - accuracy: 0.0093 - precision: 0.9747 - recall: 0.9713 - auc: 0.9991 - val\_loss: 2.6806 - **val\_accuracy: 0.0133 - val\_precision: 0.6455 - val\_recall: 0.6348 - val\_auc: 0.8606**

Confusion matrix and classification report:

Calendar

Description automatically generated

**Overall Model Accuracy / Balanced Accuracy: 0.1405**

Observations: 30 epochs took about 12 to 13 hours to train. No overfitting was observed as the validation loss decreased after each epoch and the validation accuracy was found to gradually increase. Also, the model is not yet saturated after 30 epochs. It could be configured for a greater number of epochs to improve the accuracy. We obtain a better-balanced accuracy of 0.1405 on the evaluation of test data than other models described in this report.

In order to further improvise the accuracy of predictions, we can use bigger transfer learning architectures like Inceptions, Efficient Nets and Resnets which are more preferred in the classification of medical images. However, training and fine tuning these architectures requires high configuration of RAM and GPU.

**Extras:**

**Model 5: Transfer Learning – EfficientNetV2 (Fine Tuned)**

*Reference File: “model\_efficient\_netV2.ipynb”*

Unable to train due to insufficient hardware requirements:

Text

Description automatically generated

**Model 6: Transfer Learning – InceptionResnetV2 (Fine Tuned)**

*Reference File: “model\_inception\_resnet\_v2.ipynb”*

Unable to train due to insufficient hardware requirements:

Text

Description automatically generated

**Model 7: Autokeras – Searches for the best model for the given data based on all available models**

*Reference File: “Autokeras model.ipynb”*

**Text

Description automatically generated**

1. <https://www.statology.org/balanced-accuracy/#:~:text=Balanced%20accuracy%20is%20a%20metric,model%20is%20able%20to%20detect>. [↑](#footnote-ref-1)
2. https://challenge2019.isic-archive.com/evaluation.html [↑](#footnote-ref-2)
3. https://neptune.ai/blog/balanced-accuracy [↑](#footnote-ref-3)
4. https://challenge.isic-archive.com/leaderboards/2019/ [↑](#footnote-ref-4)
5. https://tfhub.dev/google/imagenet/nasnet\_mobile/feature\_vector/5 [↑](#footnote-ref-5)