**Skin Cancer Prediction Using CNN – Documentation**

**Finished Model – (github link)**

**Step-by-Step Plan**

**Step 1: Load and Define Cancer Data -** Import necessary libraries(pandas, numpy, sklearn, seaborn, tensorflow, matplotlib, imblearn, keras), download the HAM10000 external folder([Skin Cancer MNIST: HAM10000](https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000)).

**Step 1.1: Prepare Cancer Data -** Apply labels: Cancerous (1) vs. Non-cancerous (0) to “hmnist\_28\_28\_RGB.csv file(file with pixel data of each image) based on skin lesion type.

**Step 1.2: Analyze images dataset -** Analyze images dataset and perform oversampling. Normalize data and divide it into training sets(80|20)

**Step 2: CNN –** Construct CNN

**Step 2.1: CNN architecture –** Apply layers: Conv2D, MaxPooling2D, Flatten, Dense

**Step 2.2: Set checkpoints -** Save the best model (by value loss) during training. Implement EarlyStopping that stops training if value accuracy stops improving for 5 epochs.

**Step 3: Train the model**

**Step 4: Evaluate the model**

**Step 4.1: Check for accuracy -**  Check for Final Training Accuracy, Final Validation Accuracy, Test Accuracy

**Step 4.2: Classification report -** Check for precision, recall , f1-score, support

**Step 4.3: Confusion matrix**

**Step 5: Final notes –** Summary of model performance

**Step 6: Example image testing –** Use ready model to test particular image, for example your own skin pigmentation to check whether it is cancer or not.

**Step 6.1: Load image**

**Step 6.2: Process image**

**Step 6.3: Get result –** Put processed image into the model and see result.

**Prepared by:** Alicja Zabłocka

**Project:** Skin Cancer Prediction Using CNN

**Group ID:** 110 - Data Engineering, Semester 4