Skin Lesion Classification

October 1, 2022

Description

Thanks to the pollution and improper dietary habits, people are suffering with a lot of skin issues. Some of the skin issues are very dangerous to health. While we not being doctors, we would like to identify if any image belongs to one of the 8 skin lesions.

Team Members

- Deepika Maddali (deepikamaddali@ischool.berkeley.edu)
- Peeti Sriwongsanguan (peeti@ischool.berkeley.edu)
- Sudha Ravi Kumar Javvadi (javvadis@ischool.berkeley.edu)

Problem Statement

We attempt to solve a multi-class classification problem containing 9 classes.

- Melanocytic nevus
- Basal cell carcinoma
- Actinic keratosis
- Benign keratosis (solar lentigo / seborrheic keratosis / lichen planus-like keratosis)
- Dermatofibroma
- Vascular lesion
- Squamous cell carcinoma
- None of the above

Objective

In this project, we will train the below mentioned Machine-Learning and Deep-Neural-Network models to make predictions for this skin lesion classification problem, and later provide a summary of how well each model is able to generalize.

Machine Learning Models	Artificial Neural Networks
K-Nearest Neighbor (KNN)	Recurrent Neural Networks (RNN)
Support Vector Machine (SVM)	Long Short-Term Memory (LSTM)
Multi-class AdaBoost	Gated Recurrent Units (GRUs)

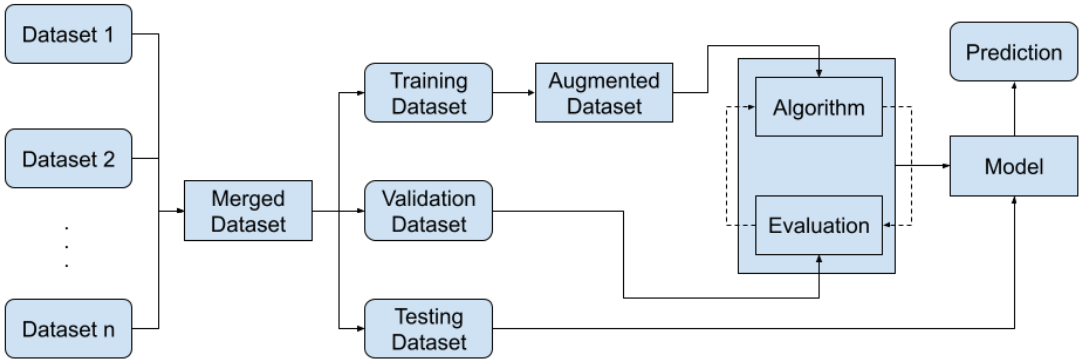
Approach/Methodology

We plan to identify various skin lesion related datasets publicly available on Kaggle. Post dataset-identification, we would merge the data to a unified dataset which is then divided into 60-20-20 split of training, validation, and testing datasets. Now, we do exploratory data analysis (EDA) on our training data and then implement various machine-learning algorithms mentioned in the Objective section. Later, we would train each of our models while figuring out the best paramaters for the multi-class classification problem. Post evaluation of our models and finding the best parameters, we would retrain our models to include the validation dataset. Now, we see how well each of our models is able to generalize on the testing data. All of our observations will be added to a report and summarized in a slide deck.

Following is the anticipated timeline:

Steps	Timeline
Data Collection	1 week
Exploratory Data Analysis	1 week
Implementation	3 weeks
Training & Evaluation	2 weeks
Generalizations & Reporting Results	1 week
Report & Presentation	1 week

Block Diagram



Datasets (Potential)

We would potentially work on the following datasets

- https://www.kaggle.com/datasets/andrewmvd/isic-2019
- https://www.kaggle.com/datasets/wanderdust/skin-lesion-analysis-toward-melanoma-detection
- https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000
- https://www.kaggle.com/c/siim-isic-melanoma-classification/data

Success/Failure Criteria

We would consider an accuracy of 22% and above as a success. Also, an accuracy of 11% or below for a machine learning algorithm would be considered as failure.

Evaluation Parameters (Potential)

- *Recall*: We are working on classifying a skin lesion which could lead to ill health to people. So, our major concern is to be able to correctly predict harmful skin lesions. So, recall would be the most important evaluation parameter for this project.
- *Accuracy*: Another important evalution parameter is accuracy for this project. This is because we do not want to scare people by telling they are having harmful symptoms when they do not. So, we would like to have have as much correct predictions as we could.