CAPSTONE PROPOSAL

Domain Background

Early detection of skin cancer has the potential to reduce mortality and morbidity. Dermoscopy technique is widely used for image acquisition in diagnosis of skin cancer but artefacts such as air bubbles, ruler markings and hair which are covering skin affect the segmentation process. To improve the accuracy of segmentation, it is necessary to remove these artefacts.
The prevalence of melanoma skin cancer disease is rapidly increasing as recorded death cases of its patients continue to annually escalate. Reliable segmentation of skin lesion is one essential requirement of an efficient noninvasive computer aided diagnosis tool for accelerating the identification process of melanoma.
There has been a number of papers published on classification of skin cancer images using machine learning methods and techniques to process and de-noise dermoscopic images. For machine learning methods, some include;

1. **Deep Neural Networks**
2. **PCA**
3. **Clustering**

Many image processing and de-noising techniques that are being used and considered include color-space transformation, contrast enhancement, artifact removal, saliency-based segmentation.

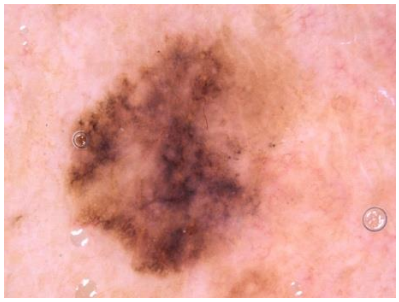
Problem Statement

The scope of the project will be limited to PCA pixel models and given Image data of patients, can a PCA based machine learning model effectively determine whether the patient has malignant skin cancer or not. This leads to sub-questions such as how accurate will this prediction be, how can it remove the different noise types and successfully transform the image, how will the model perform with a feature engineered images compared to a non-feature engineered images, what is the best classifier to use to solve this and more.

Dataset and Inputs

These questions are answered using by data openly available here where there are images of malignant and benign cancer types in zip files and are publicly downloadable. 300 samples each of the malignant and benign image class will be selected for training, validation and testing. Here is a sample of each class:

Benign



Malignant

Each image is 600x400 pixels with different variation of color, luminance, orientation. Also present are hair strands, bubbles, ruler markings etc. The major assumption taken here is the color of a mole or skin spot is a valuable feature for predicting if it is malignant or benign.

Solution Statement

Build a model which can successfully feature engineer and transform the images. It will then be able predict whether the image has skin cancer or not using a trained and tested classical classifier algorithm. This will then be compared against the benchmark non-feature engineered model as well as being compared on the classical classifier types best suited for this.

Benchmark Model

The benchmark model will be the non-feature engineered image model and will be used to be compared against the feature engineered model.

Evaluation Metrics

The metrics to be used includes accuracy, precision, recall, F-1 score and cross validation score.

Project Design

A hybrid technique using PCA model for the classification of the skin images. The proposed hybrid technique consists of three stages, namely, Image processing and feature extraction, dimensionality reduction, and classification.

In the first stage, features related with images using pixel value extraction will be obtained by feature engineering such as greyscaling, image size rescaling and de-noising. In the second stage, the features of skin images will be reduced using principle component analysis to the more essential features. In the classification stage, three classifiers based on supervised machine learning will developed namely k-nearest neighbour, logistic regression and svm which have been selected based on literature.

The classifiers will be used to classify the images as benign or malignant skin cancer images and which will be nominally represented as 1 and 0 respectively.

Finally, the feature engineered model will be compared with non-feature engineered model based on metrics such as accuracy, precision, recall, F-1 score and cross validation score. The best classifier will also be determined from the three proposed.

References

1. Skin cancer reorganization and classification with deep neural network, Hao Chang
2. Automatic skin cancer Images classification, Mahmoud Elgamal
3. Segmentation of Melanoma Skin Lesion Using Perceptual Color Difference Saliency with Morphological Analysis, Oludayo et al
4. Artefact Removal and Contrast Enhancement for Dermoscopic Images Using Image Processing Techniques , Pragati Rajendra Mahajan