**Skin Lesion Analysis Towards Skin Cancer Detection**



**INFO 7390 - Advances in Data Sciences and Architecture**

Group 18: “Wonder Women”

Chinmayi Shaligram (001550682)

Dona Darji (001044833)

Sushmitha Jogula (001546751)

**1. Introduction**

* Cancer is one of the most prevalent and difficult diseases that humanity is still fighting today, and skin cancer is one such type of cancer.
* Skin cancer occurs as a result of aberrant skin cell proliferation, mainly on places exposed to sunlight.
* In the United States, more than 9,500 individuals are determined to have skin malignant growth consistently, with the annual cost of care exceeding $8 billion.
* Skin cancer can be cured if found and treated early; the 5-year survival percentage of the most lethal form, melanoma, can be as high as 99 percent if detected early; but, if diagnosed late, the survival rate drops to only 23 percent.
* This is where Skin Lesion analysis comes into picture which can be employed for early cancer detection in patients and boost up the survival rate.
* Skin Lesions are areas of skin that look different from the surrounding area. They are often bumps or patches, and many issues can cause them. These are changes in skin that are not cancer but could become cancer over time.
* Understanding and detecting these skin diseases early can help millions of people prevent skin mutations.

**2. Problem Statement**

* Skin cancer is the most common cancer in the world that affects one in every five people by the age of 70.
* Early cancer diagnosis saves lives. Knowing what to look for on your own skin might help you spot cancer early, when it's easiest to treat, before it becomes harmful, disfiguring, or fatal.
* Early and precise discovery of tumors could make the treatment more viable. Early detection can prevent it from spreading with minor surgeries.
* Thus, it becomes highly critical to learn to spot the early warning signs.
* Good historical records and knowledge of skincare combined can help detect the type and severity of skin disease.
* Skin Lesion image analysis with various analysis tools helps detect and improves diagnostic accuracy.
* High power lenses and microscopes can help capture lesion pictures to better understand the situation.

**3.** **Goals**

* Analyze various attributes of the patient data along with the images of their skin lesions and predict the disease classification within the dermoscopic images.
* Distinguish the possibilities by analyzing whether skin lesion images addresses any of the possible skin disease categories:
  + Melanoma
  + Melanocytic nevus
  + Basal cell carcinoma
  + Actinic keratosis / Bowen’s disease (intraepithelial carcinoma)
  + Benign keratosis (solar lentigo / seborrheic keratosis / lichen planus-like keratosis)
  + Dermatofibroma
  + Vascular lesion
* Automating the classification of skin lesion detection for detecting skin diseases using several machine learning and deep learning algorithms and assess their performance.
* Identifying feature set for better classifying the images based on patient records.

**4. Importance**

* The good news is that if discovered and treated early enough, 99 percent of all skin cancer cases are curable. However, in order to prevent skin cancer, we must detect it early.
* The accuracy of skin disease diagnoses will be improved by skin lesion image analysis techniques that automate the diagnosis of skin cancers.
* Historical records of patients with skin lesion images can help the model predict and classify whether the patient is diagnosed with a possible skin disease or not.
* Early detection of skin cancer and its types leads to early treatment and a significant boost in the survival rate, thus positively impact millions of people.

**5. Dataset**

* The small size and lack of diversity of the available datasets of dermatoscopic images make it difficult to train neural networks for automated diagnosis of pigmented skin lesions.
* The HAM10000 Dataset is used in our project. This consists of 10015 dermatoscopic lesion images along with patient data (such as age, sex, lesion localization), acquired from different populations and stored using different modalities.
* The diagnostic lesion categories included in our analysis are:
  + Actinic keratoses and intraepithelial carcinoma / Bowen's disease (akiec)
  + Basal cell carcinoma (bcc)
  + Benign keratosis-like lesions (solar lentigines / seborrheic keratoses and lichen-planus like keratoses, bkl)
  + Dermatofibroma (df)
  + Melanoma (mel)
  + Melanocytic nevi (nv)
  + Vascular lesions (angiomas, angiokeratomas, pyogenic granulomas and hemorrhage, vasc)
* More than 50% of lesions are confirmed through histopathology (histo), the ground truth for the rest of the cases is either follow-up examination (followup), expert consensus (consensus), or confirmation by in-vivo confocal microscopy (confocal).
* The HAM10000 dataset can be found on Kaggle here: <https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000?select=hmnist_28_28_L.csv>

**6. Algorithms used**

* Traditional Machine Learning Models:
  + Softmax Regression
  + Support Vector Machine
* Deep Learning:
  + Multi Layer Perceptron (MLP) and Convolutional Neural Network (CNN)
* Ensemble Learning (Boosting)
  + XGBoosting

**7. Models Performance and Comparison**

* At the end of our experiments we compare all model results in the following way:
  + We take the best model over validation set for each classifier and then we measure all of them over the test set.
* The CNN model achieved the best macro-f1 over the test set (same over validation set) and SVM model achieved the best BMCA score.
* The final results confirms that CNNs perform the best in image classification tasks but also the good performances of boosting techniques (XGBoosting).
* The final results also prove that our classification models perform better than our baseline model i.e. random guess classifier.

**8. Conclusion**

* This study's main aim and purpose is to identify the disease classification based on the dermatoscopic images of skin lesions, while taking into consideration several patient attributes and data such as age, sex, skin lesion localization, and skin lesion confirmation/detection method.
* We experimented with various machine learning and deep learning models and methods. The Convolutional Neural Network (CNN) model is the best performing model we deduced of all the models we have implemented in terms of macro F1-score; and the Support Vector Machine (SVM) model is the best performing model in terms of BMCA score.
* This is a fascinating opportunity for healthcare industry to assess these factors related to patients and their skin lesions, and this analysis is highly important for both healthcare providers and patients.
* With this knowledge, groundbreaking work can be performed towards early detection of skin cancer in patients, which leads to early treatment procedures and eventual increase in the survival rate of the patients.

**9. Google Colab Notebook Link:** [**ADS\_FinalProject.ipynb - Colaboratory (google.com)**](https://colab.research.google.com/drive/1mw92bIZB2TNZhSbWCdWbMeAvpcTVbTIv#scrollTo=Wi-_1_iGRKPG)