PROJECT PROPOSAL

- **Team Members**

NAME	CWID
Kalebu Patan	A20446212
Robert Shyroian	A20484362
Rutika Dhoka	A20501969

- **Project Topic:** Skin Cancer Classification with Machine Learning

- **Objective:** To develop and compare multiple models for accurate classification of skin lesions as benign or malignant using the Skin Cancer MNIST: HAM10000 dataset, aiming to improve diagnostic accuracy and reduce unnecessary biopsies.

  Some potential sub-objectives to achieve this overarching goal could include:
  1. Preprocessing the dataset to remove duplicates and correct any class distribution imbalances.
  2. Exploring and visualizing the dataset to gain insights into the features that distinguish benign from malignant lesions.
  3. Implementing and comparing the performance of different machine learning models, such as logistic regression, decision trees, random forests, support vector machines (SVMs), and neural networks, in classifying skin lesions.
  4. Tuning the hyper-parameters of the chosen models to improve their performance.
  5. Evaluating the performance of the models using metrics such as accuracy, precision, recall, and F1 score, and comparing the results to determine the most effective approach.
  6. Interpreting the results of the models to gain insights into the most important features for classification and potential areas for further research.

  Overall, the goal of the project would be to develop a reliable and accurate model for skin cancer classification that could potentially be used to improve clinical diagnosis and reduce the number of unnecessary biopsies.

- **Abstract:**

  Skin Cancer is one of the most common types of disease in the United States. Over the year, up to 4 Million cases have been reported dead due to skin cancer in the United States. Skin cancer is an uncontrolled development of abnormal skin cells potentially due to excessive exposure to sun, history of sunburns, less melanin, Precancerous skin lesions, moles, etc. This occurs when unprepared DNA damages the cells of the skin. It is one of the diseases that are viewed on its quick evolution and the most common type of cancer that endangers life. In this project, we are going to classify different types of skin cancers using Skin Cancer MNIST: HAM10000 dataset. We will implement different ML algorithms for classification and compare their performance.

- **References:**

[1] https://openaccess.thecvf.com/content_cvpr_2018/html/Yang_Clinical_Skin_Lesion_CVPR_2018_paper.html

[2] https://www.sciencedirect.com/science/article/pii/S2352914819302047

[3] https://www.frontiersin.org/articles/10.3389/fonc.2022.893972/full

[4] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6231861/

[5] https://www.ejcancer.com/article/S0959-8049(21)00444-5/fulltext