**Project: Milestone 1**

<https://docs.google.com/spreadsheets/d/1vA6_vbJ7ecGvYHG0KeizhzZ55FLGuZ0_NUtxdOGEIkE/edit?usp=sharing>

On **7 Dec**, you need to complete **Milestone 1** and give a short presentation (each group 10 minutes) answering the following questions:

1. What is the size of your dataset? Is there a separate batch for train, test, and validation, or do you need to do it manually? What are the different classes in your dataset?
2. What is the distribution of your dataset, i.e., the number of samples per class? Show it using a plot. Is the distribution balanced or unbalanced?
3. Which technique are you going to use to tackle your unbalanced dataset and why? NOTE: you DO NOT need use any such technique for Milestone 1, but you have to use it for the next milestones.
4. Use the logistic regression shown in the last class to classify. Report the following results separately for train and test - accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), the confusion matrix, and the F1 score. How can we interpret these results?
5. Train a five layer (one input layer, one output layer, three hidden layers) neural network using keras with some random number of neurons in the hidden layers on your dataset and use the test dataset to evaluate the model's performance (accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), the confusion matrix, and the F1 score). Use the logistic activation functions for the hidden layers, and softmax for the output layer. Compare the results with that of the Logistic Regression model.