The final report

ID	NAME
20221441990	حازم محمد احمد بكر
20221468416	إبراهيم محمد إبراهيم النقيب
20221450304	احمد سمير عبدالفتاح امين
20221041599	خالد عبدالحميد حمدي محمد
20221373780	احمد سمير عبدالعظيم
20221375760	احمد السيد عجمي احمد
20221150099	نورالدين أنور محمد
20221449224	احمد ناصر عبدالفضيل
20221442040	مروان اشرف محمد عبدالباقي
20221453750	مصطفى صالح مصطفى

Image processing

1. **Data Loading and Preprocessing:**
   a. The script loads bird images from three different directories: training, testing, and validation sets.
   b. Images are read using OpenCV, resized to (224, 224), and normalized to values between 0 and 1.
   c. The script converts the image data and corresponding labels into NumPy arrays.

2. **One-Hot Encoding:**
   a. The script uses **LabelBinarizer** from scikit-learn to perform one-hot encoding on the class labels.

3. **CNN Model Architecture:**
   a. The CNN model is defined using Keras Sequential API.
   b. It consists of several convolutional layers with max-pooling, followed by a flatten layer and dense layers.
   c. The last dense layer has 20 neurons (assuming 20 bird classes) with softmax activation for multi-class classification.

4. **Model Compilation:**
   a. The model is compiled with the categorical cross-entropy loss function and the Adam optimizer.

5. **Model Training:**
   a. The model is trained on the training set for 20 epochs with a batch size of 128. Validation data is provided during training.

6. **Model Evaluation:**
   a. The script evaluates the trained model on the test set and prints the test accuracy and loss.

7. **Model Prediction and Visualization:**
   a. The script predicts bird species for some test images and compares the predictions with the ground truth.

8. **Training History Visualization:**
   a. Training history (accuracy and loss) is plotted using matplotlib.

9. **Classification Report:**
   a. A classification report is generated using scikit-learn's **classification_report** function, providing metrics like precision, recall, and F1-score.

10. **Random Image Visualization:**
    a. The script displays 10 random images from the test set with their original and predicted labels.

11. **Web Scraping:**
    a. The script performs web scraping to retrieve information about a bird species using its name and eBird link.

12. **Loading a Saved Model:**
    a. The script attempts to load a model from a file named 'your_model.h5' using Keras's **load_model** function. Note that you should replace 'your_model.h5' with the actual filename if you intend to use this part.

Lung Cancer

1. **libraries:**
   a. pandas ,scipy.io to read and manipulate the data
   b. numpy for arrays
   c. sklearn for feature selection , building models ,clustering , encoding
   d. matplotlib for plotting the data

2. **data quality :**
   a. check that the data has no duplicates values and there is no missing values.

3. **encode:**
   a. we have 5 different types , so we encode the column y to be in integer format to use as a target column.

4. **Splitting:**
   a. First , split the feature and the target columns in variables X and y
   b. Split the X and y to train and test sets using Train_test_split in sklearn.model_selection.

5. **Feature selection:**
   a. We compine Select K Best and f_classif from sklearn.feature_selection as a filter technique
   b. Train a RandomForest model to choose the best number of features to be used

6. **ML models :**
   a. Build some Machine Learning models to identify what is the best model to fit this data.

7. **Classification Report:**
   a. A classification report is generated using scikit-learn's **classification_report** function, providing metrics like precision, recall, and F1-score.

8. **Feature extraction:**
   a. Used PCA to get the data only in 2 dimension to make it easy to visualize in cluster

9. **Clustering:**
   a. Using HC and k-means to see how the different types separated from each other.