

## Data Quality Report - PCA Based Face Recognition

Data Source	Data Quality Issue	Severity	Resolution Plan
/content/train, /content/test	Class imbalance – some people have fewer face images than others	Moderate	Apply data augmentation (flipping, rotation) or assign class weights during training.
/content/train, /content/test	Low-resolution images (e.g., 64x64) may not capture enough features for PCA	High	Increase resolution (e.g., 128x128), or use eigenfaces at higher dimensions (n_components = 100–200).
/content/train, /content/test	Noisy or mislabeled face images in folders	Moderate	Manually verify, or use a script to detect and remove wrong samples.
All classes (Train/Test)	Image format inconsistency (JPG, PNG, grayscale vs RGB)	High	Normalize all images to grayscale (cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)) and resize to fixed dimensions.
All classes	Overfitting due to small dataset size	Low	Use regularization (dropout in ANN), data augmentation, or early stopping.
All classes	Uneven distribution between train/test splits	Moderate	Use stratified split to maintain class balance in both sets.