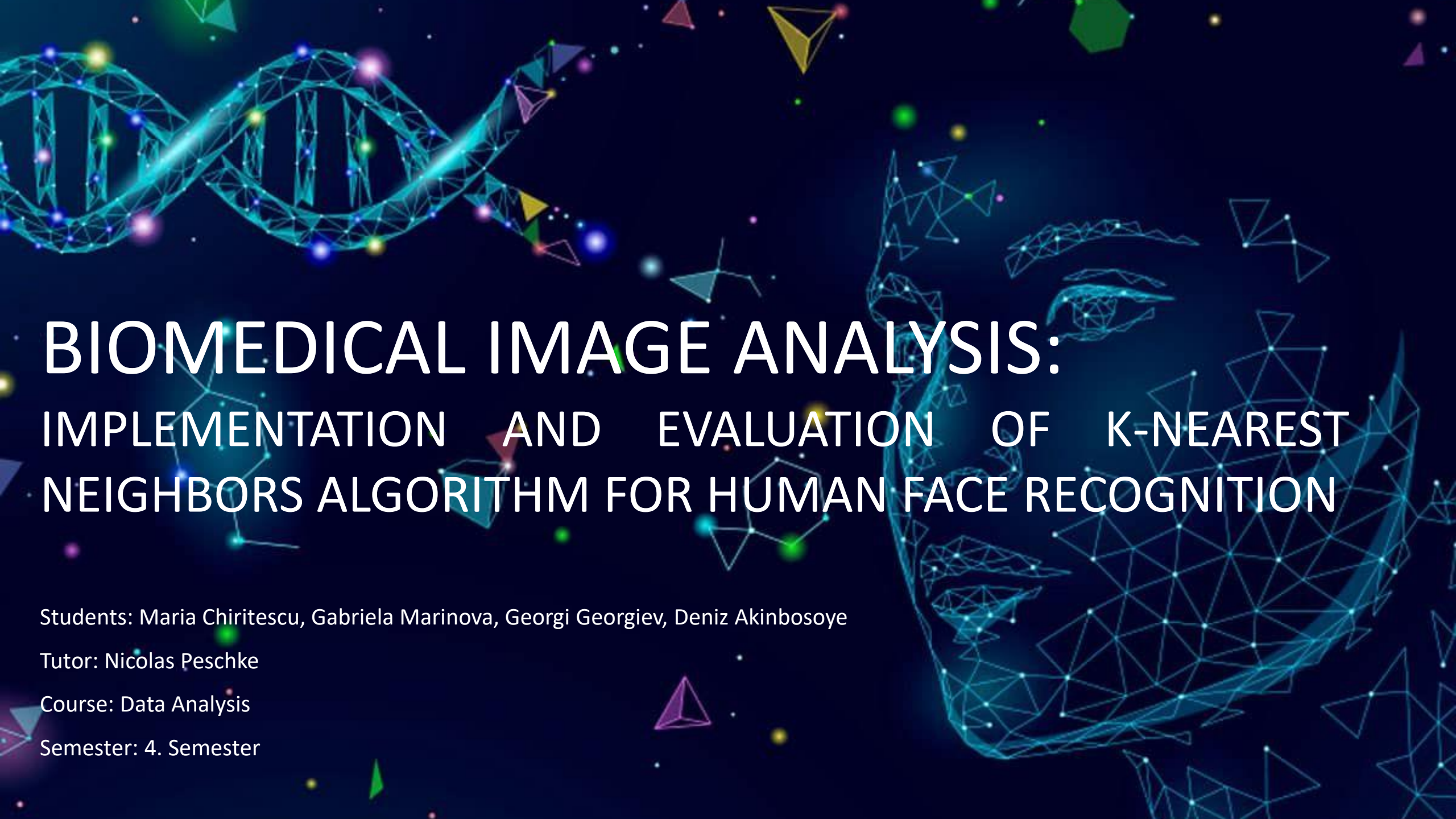


# BIOMEDICAL IMAGE ANALYSIS:

## IMPLEMENTATION AND EVALUATION OF K-NEAREST NEIGHBORS ALGORITHM FOR HUMAN FACE RECOGNITION

Students: Maria Chiritescu, Gabriela Marinova, Georgi Georgiev, Deniz Akinbosoye

Tutor: Nicolas Peschke

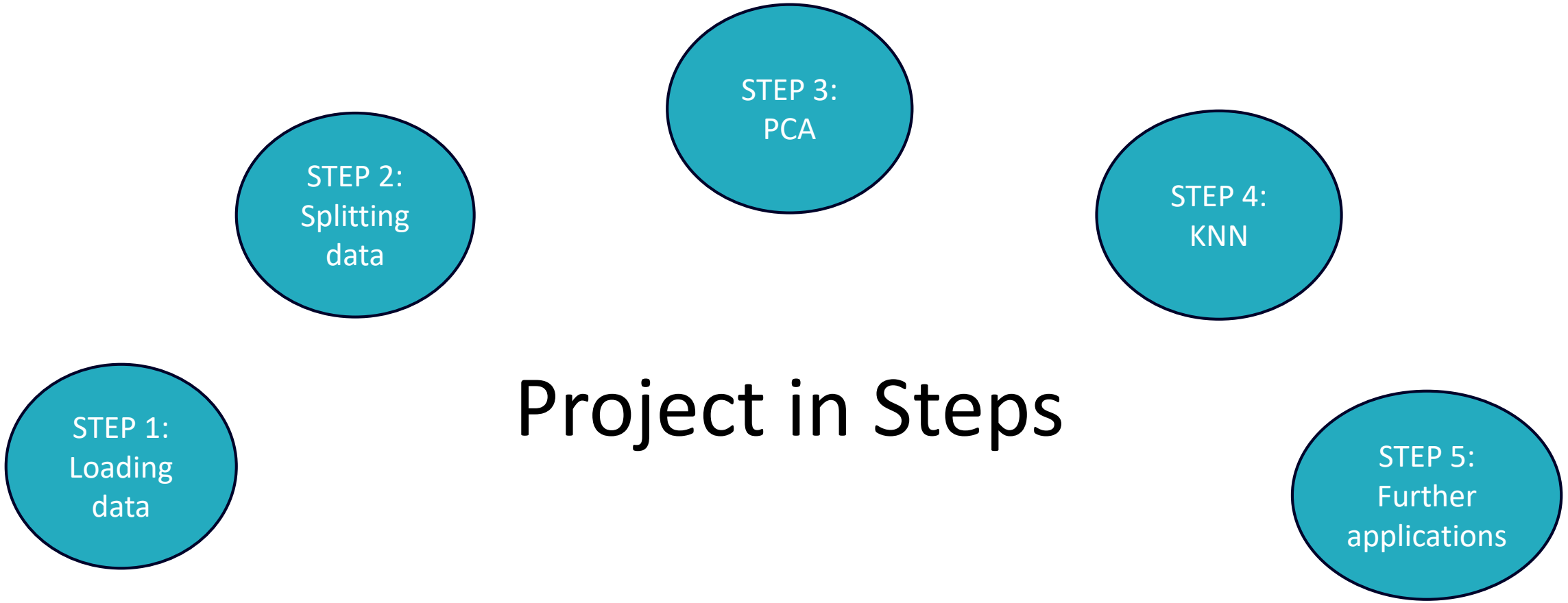
Course: Data Analysis

Semester: 4. Semester

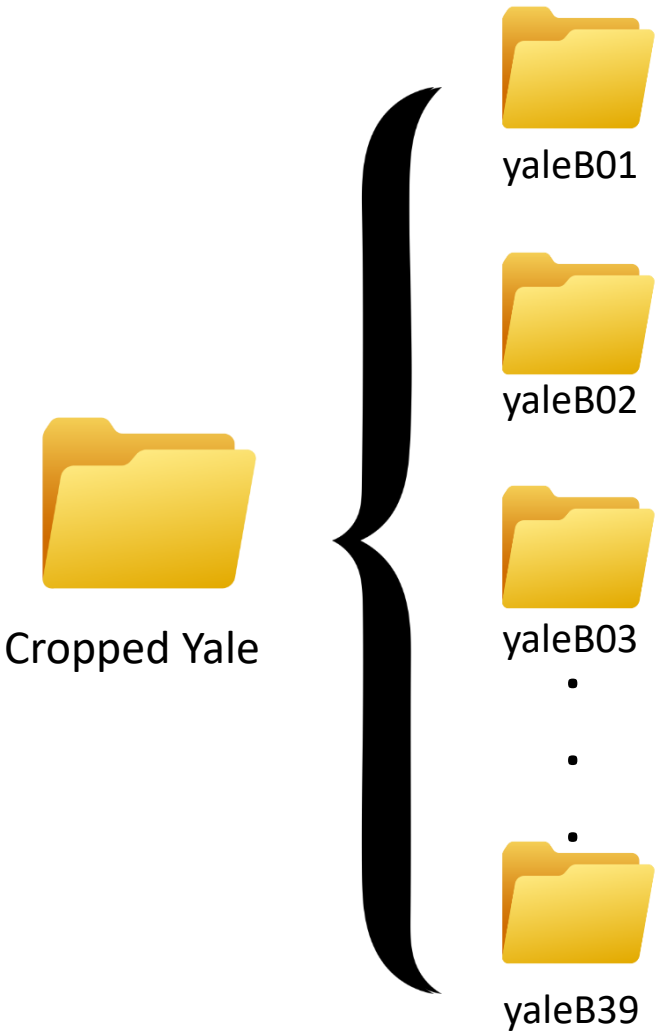
# AIM OF THE PROJECT

Creating a program that recognizes faces in different lighting conditions

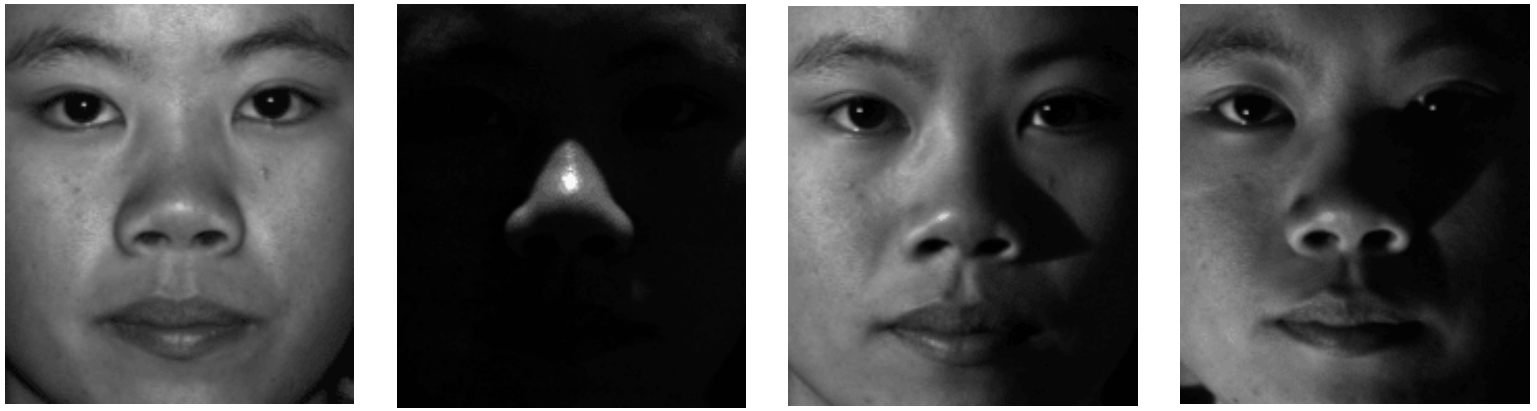
# Project in Steps



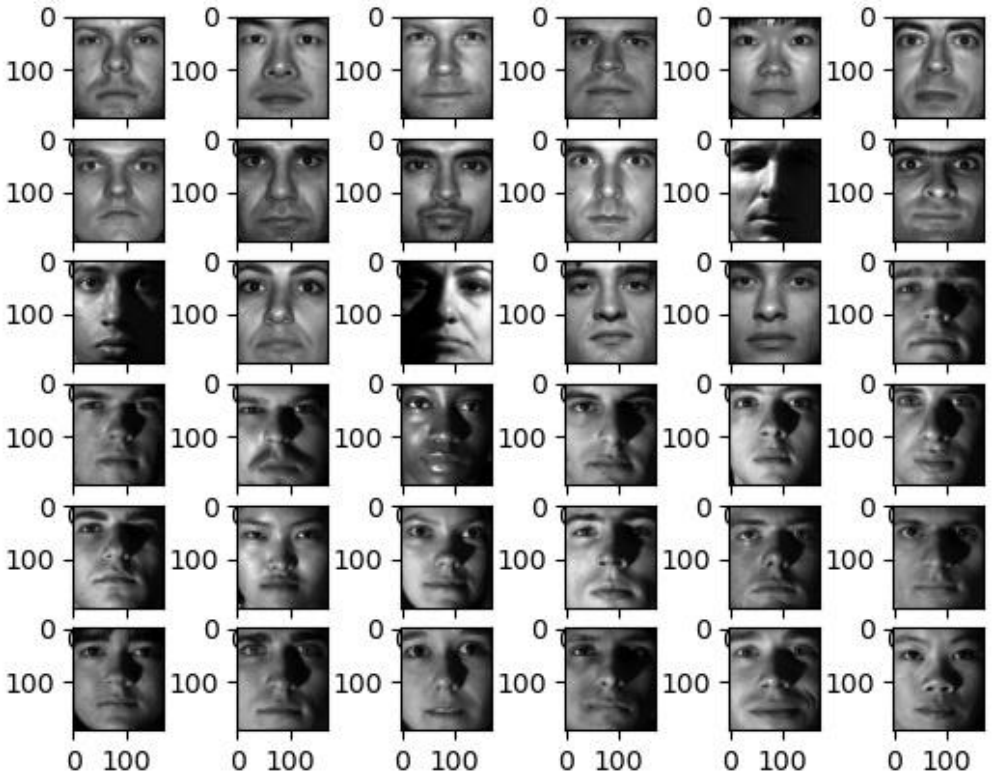
# The Dataset




- Main folder “Cropped Yale” contains 39 subfolders;
- Each subfolder contains pictures of one single person => 39 people in total
- Subfolders are denoted “yaleB01”, “yaleB02”, …, “yaleB39”
- Subfolder “yaleB14” is missing => 38 people in total;
- Each subfolder contains 64 grayscale images of each person that the program analyses;
- The pictures in each subfolder differ in angles and illuminating conditions.



# Dividing the Dataset

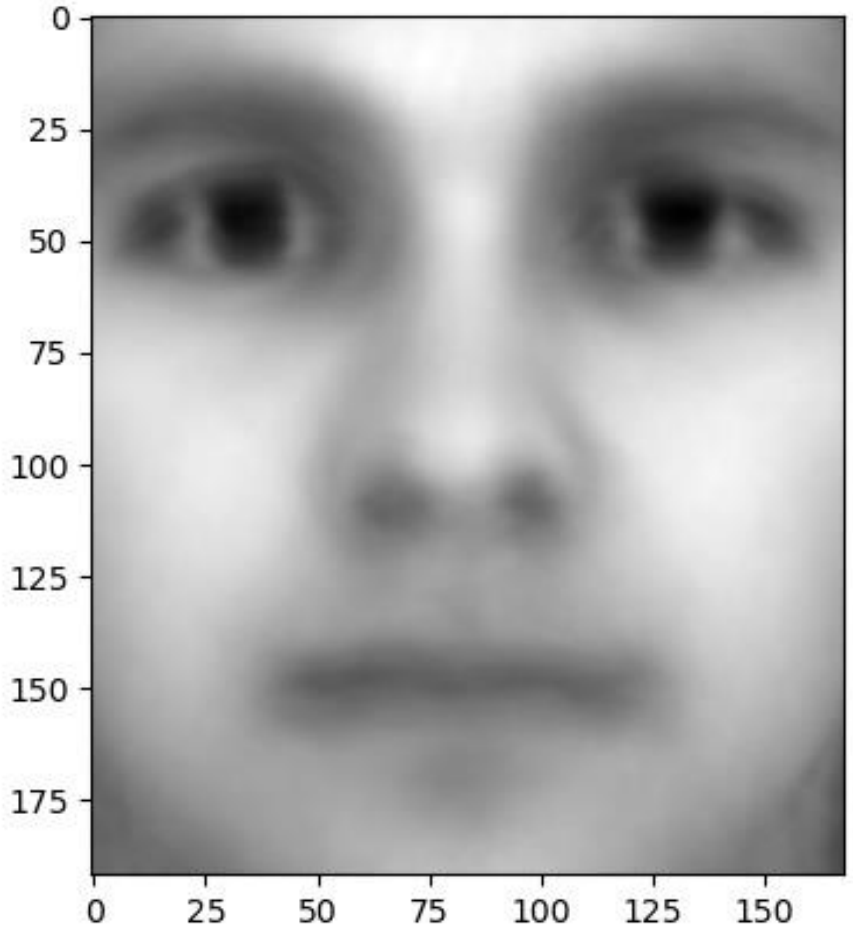


2414 images overall

80 % of the images for training set

20% of the images for the testing set

# Mean Face



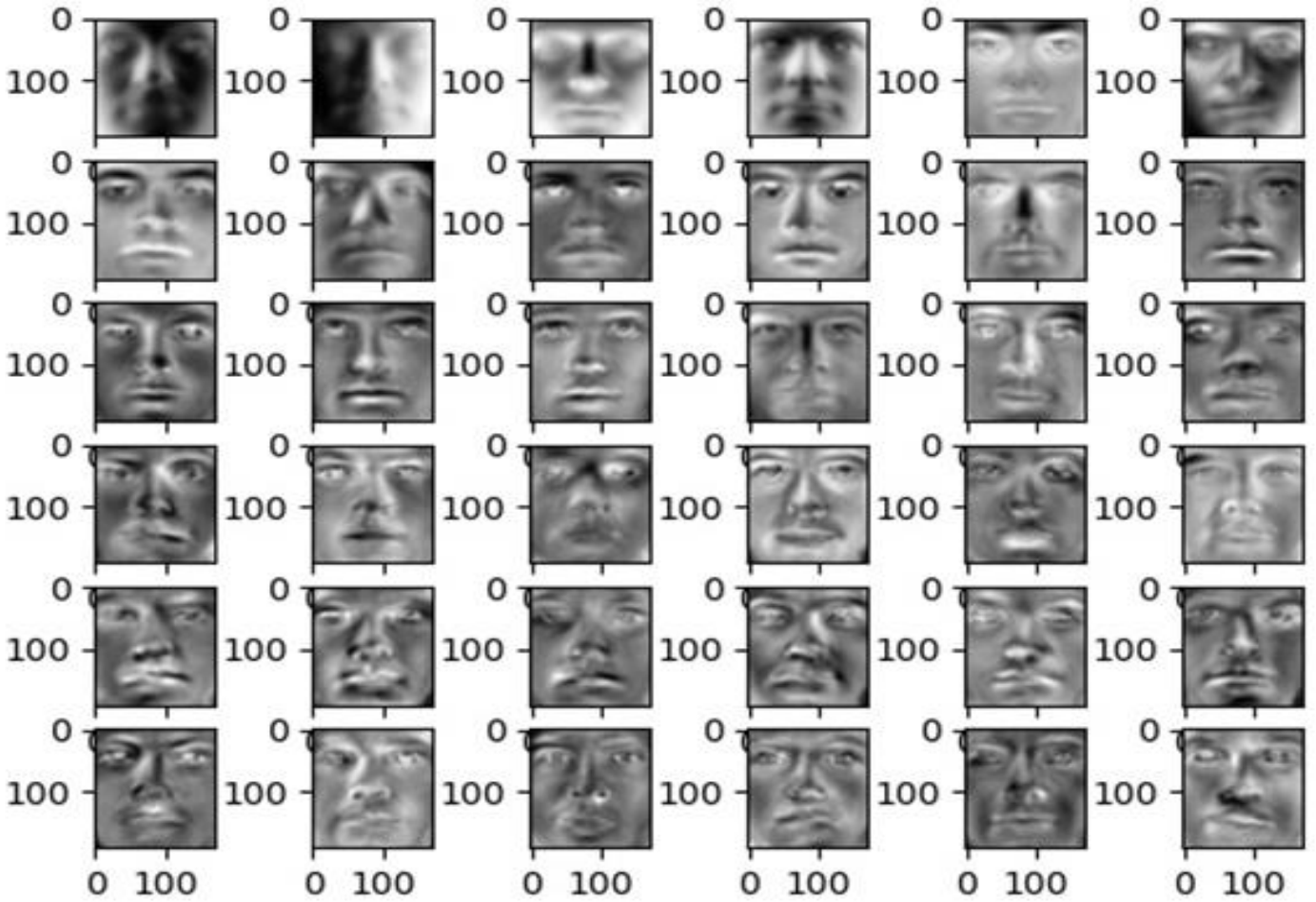

**MEAN FACE EXTRACTION:**

- Image centering;
- Allows the comparison of the images;
- After centering mean becomes zero.

# Dimensionality-Reduction Using PCA

**PCA (Principal Component Analysis):**

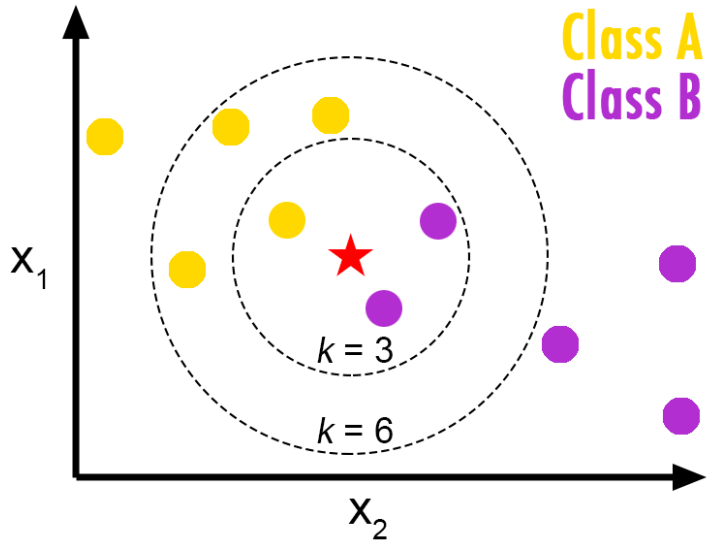
- number of principal components < number of pixels;
- Each variable in the original dataset can be represented in terms of principal components;
- result: dimension reduction.



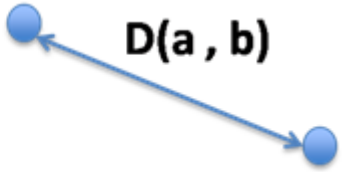
# Classification Using KNN

**KNN (K-Nearest Neighbors):**

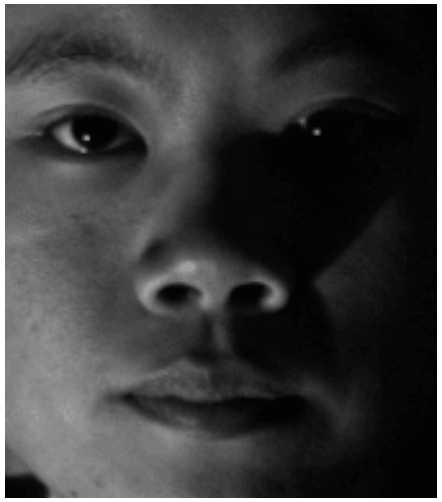
- compares the input image with the ones in the dataset;
- measures the Euclidian distance between the pixels of different pictures;
- assigns the new objects to the category of their most similar K nearest neighbors.




$$D(a,b) = \sqrt{\sum_{i=1}^{n}(b_i - a_i)^2}$$



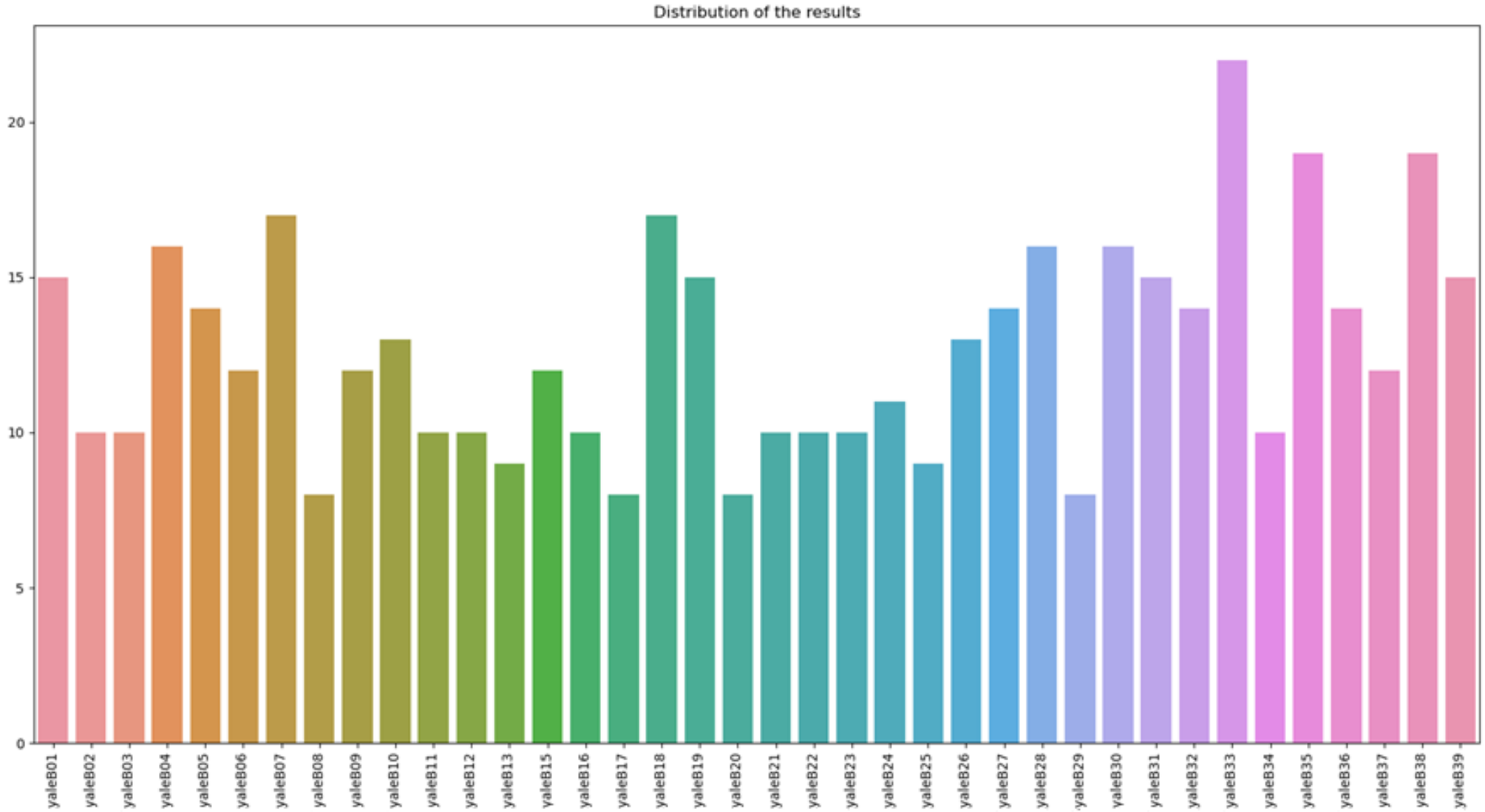

Input



Output

yaleB37

# KNN Classifies Test Pictures Based on the Faces Recognized in Them

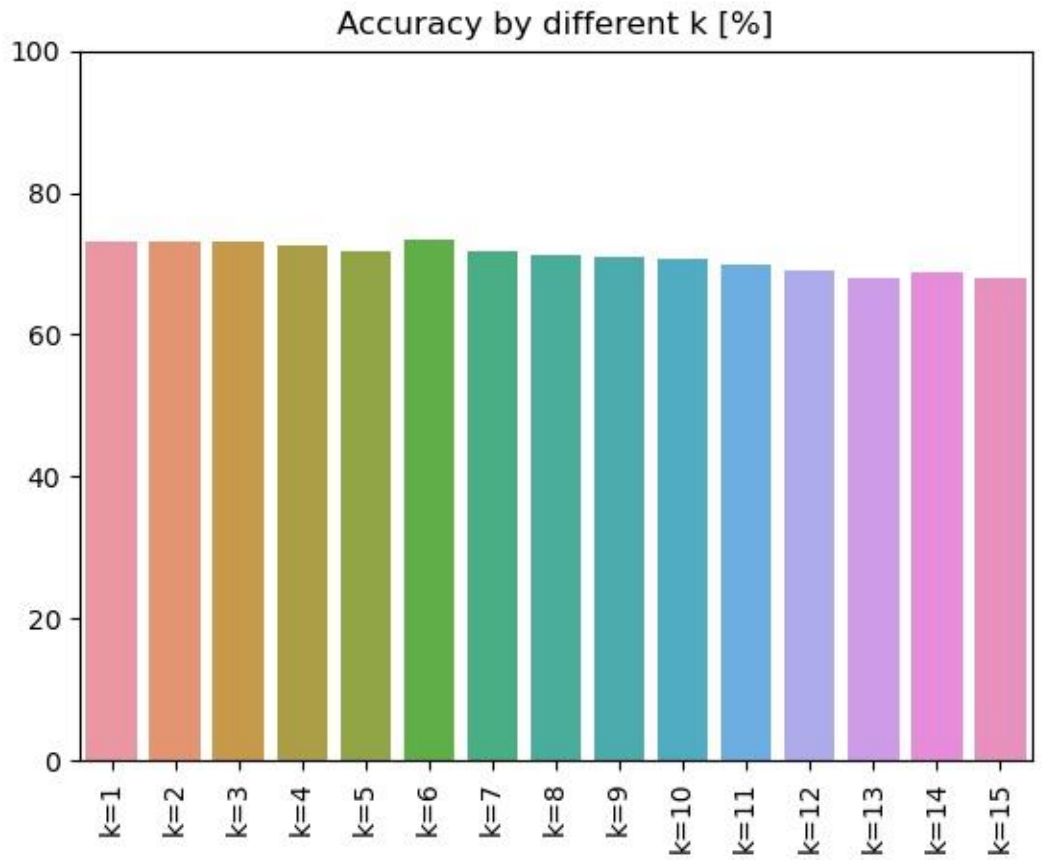


# Evaluating the Results

The performance of the system is measured in terms of accuracy:

$$Accuracy = \frac{Correctly\ Detected\ Test\ Images}{Total\ Number\ of\ Test\ Images} \text{x}100$$


# Accuracy by Different k-Values




- Choosing the proper number of neighbors is vital for the performance of the program;
- There is no perfect k-value for any classification algorithm;
- The best k-value is determined experimentally: it appears that the best accuracy is reached for k the best k-value;
- In our case: the best accuracy appears for k = 6: 73.498964803312 6%.

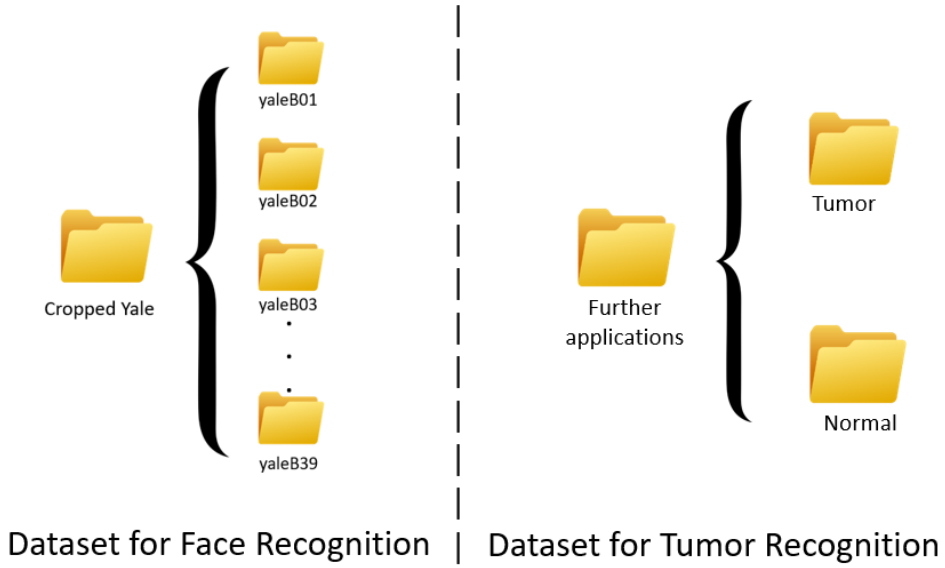
# Further Applications of the Program

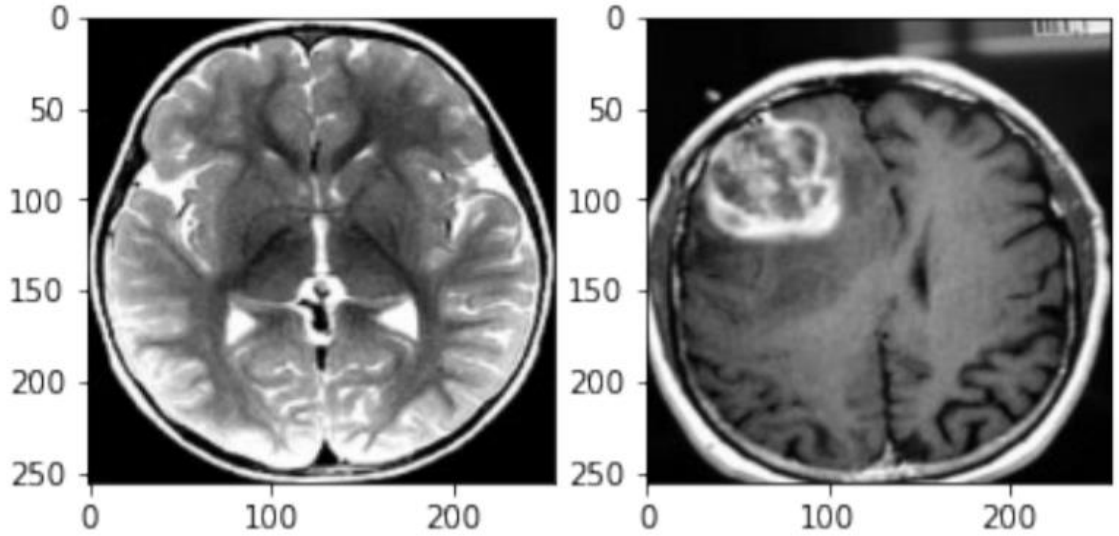

- As seen before, the described algorithm can successfully classify unknown images into a folder based on common features they share with the images present in that folder;
- The algorithm can be successfully used in the context of facial recognition, but its outreach does not stop there;
- Image classification tasks can be executed with this program in a variety of fields.

# Examples of further applications of this program

- Taking the input image from a video or from a webcam;
- Detecting patterns in credit card usage;
- Analyzing register data and detecting suspicious activity.
- Object recognition;
- Diagnosing diseases;

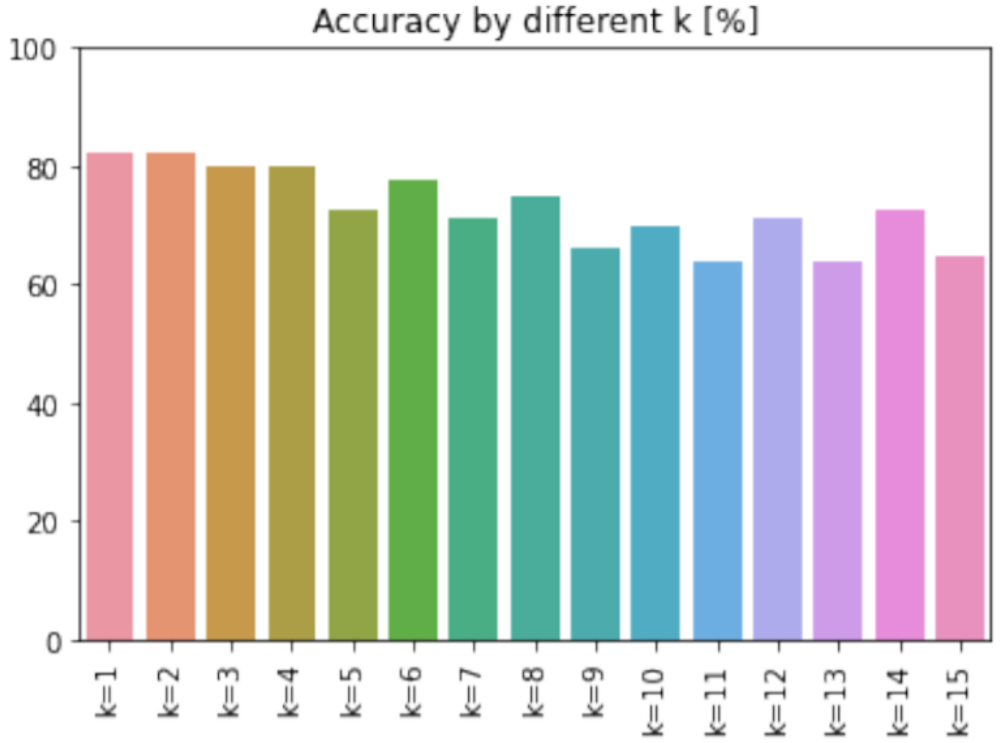
# Further Application of the Program: Tumor Recognition

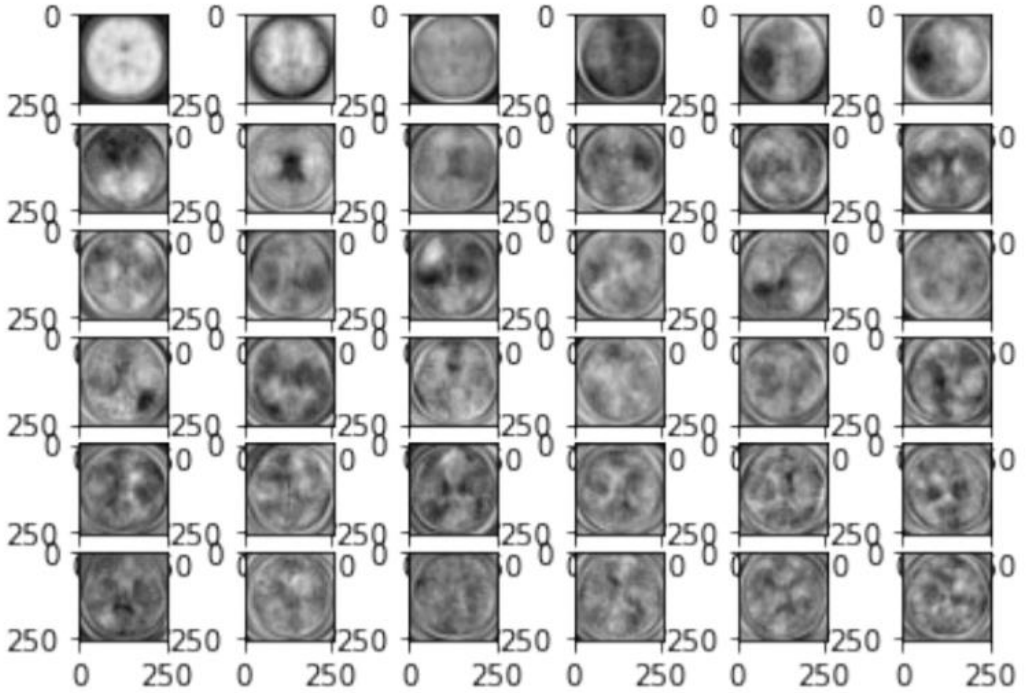



- The program used for face recognition can also be used for tumor recognition;
- The folder “Further applications” contains two sub-folders: “Tumor” and “Normal”;
- The sub-folder “Tumor” contains magnetic resonance images (MRIs) of malignant brains;
- The sub-folder “Normal” contains magnetic resonance images (MRIs) of healthy brains;
- Purpose: given a new brain MRI, decide whether the brain in it is healthy or cancerous.

# Results of Tumor Recognition



- Illustrating the accuracy of the results shows that working with a small k-value deliver more reliable results;
- The maximum accuracy obtained by performing KNN for tumor recognition is higher than the one obtained when running the code for face recognition => The reliability of the program is slightly influenced by the used dataset.


Thank you for your attention!

# Sources


1. Belhumeur, P.N., Hespanha, J.P. and Kriegman, D. (1997). Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. IEEE Transactions on Pattern Analysis and Machine Intelligence (1997) 19, 711-720.
2. Gerbrands, J.J. (1981). On the relationships between SVD, KLT and PCA. Pattern Recognition 14, 1-6, 375-381.
3. Gareth, J. et al. (2013). An introduction to statistical learning. Springer New York 4, 4
4. Sasankar, P. and Kosarkar, U. (2021). A study for Face Recognition Using Techniques PCA and KNN. Research Review Journals.
5. Wirdiani, N. et al. (2019). Face Identification Based on K-Nearest Neighbor. Scientific Journal of Informatics.