**Toros University**

**Faculty of Engineering**

**Department of Computer and Software Engineering**

**Image Processing – Final Project**

**-MUHAMMED MUSTAFA KAPICI-**

# LIBRARIES

## OpenCV

OpenCV (Open Source Computer Vision Library) is a library of programming functions primarily aimed at real-time computer vision. It is written in C++ and has interfaces for C++, Python, and Java. OpenCV can be used to process images and videos, perform feature detection and extraction, and implement machine learning algorithms for image and video analysis. Some of the most common use cases for OpenCV include object detection and tracking, face recognition, and image restoration.

## SCIKIT-LEARN

Scikit-learn (sklearn) is a Python library for machine learning. It provides a wide range of supervised and unsupervised learning algorithms in Python. Supervised learning algorithms include linear regression, logistic regression, decision trees, and support vector machines, while unsupervised learning algorithms include k-means clustering, hierarchical clustering, and principal component analysis.

Scikit-learn is built on top of NumPy and SciPy, which are Python libraries for numerical computation and scientific computing, respectively. This allows it to efficiently work with large datasets and perform complex computations.

Scikit-learn's API is designed to be simple and consistent, making it easy to use for both beginners and experienced users. It also provides tools for model selection, evaluation, and performance tuning, making it a powerful tool for machine learning tasks.

Scikit-learn is widely used in a variety of applications such as natural language processing, computer vision, bioinformatics, and finance. It's also a popular choice for educational purposes because of its user-friendly interface and a large number of examples and tutorials available.

## MATPLOTLIB

Matplotlib is a plotting library for the Python programming language. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK. Matplotlib is also a popular library for creating static, animated, and interactive visualizations in Python. It supports various types of plots, including line plots, scatter plots, bar plots, histograms, and 3D plots. It also provides a wide range of customization options for creating high-quality visualizations.

## NUMPY

NumPy is a library for the Python programming language, adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays. With NumPy, users can perform mathematical and logical operations on arrays, including basic mathematical operations like addition, subtraction, multiplication, and division. It also supports various forms of indexing and slicing of arrays, which allows users to easily extract and manipulate specific parts of an array. Additionally, NumPy provides support for linear algebra, Fourier transformation, and random number generation. It is a fundamental library for scientific computing with Python and widely used in various data science and machine learning libraries.

## PANDAS

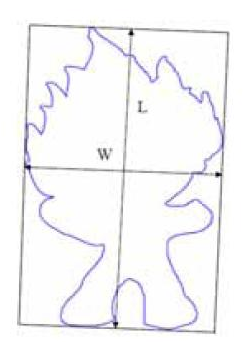
Pandas is a Python library for data manipulation and analysis. It provides data structures and functions for working with structured data, including data frames (similar to tables in a relational database) and series (similar to arrays). Some of the key features of Pandas include powerful indexing and selection capabilities, handling of missing data, and support for merging, grouping, and reshaping data. Pandas is widely used in data science and analytics, and is a fundamental tool for working with data in Python.

# DATASET

# 

# FEATURE EXTRACTION

## Eccentricity

Eccentricity is the measure of aspect ratio.

It’s ratio of length of major axis to minor axis (think ellipse for example)

Calculated by principal axes method or minimum bounding rectangular box

Eccentricity: E = L/W

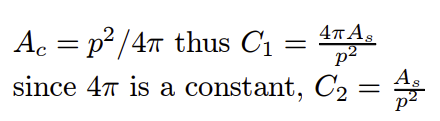
L: length of bounding box

W: width of bounding box

Elongation: Elo = 1 - W/L

## Circularity Ratio

Circularity ratio: How similar to a circle is the shape.

C1 = As /Ac = (Area of a shape)/(Area of circle)

where circle has the same perimeter

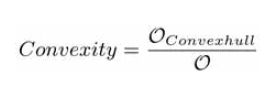
## Ellipse Variance

Ellipse Variance Eva: Mapping error of shape to fit an ellipse with same covariance matrix as shape: Cellipse = Cshape

## Rectangularity

Rectangularity represents how rectangular a shape is how much it fills its minimum bounding rectangle

Where As is the area of a shape, Ar is the area of minimum bounding rectangle.



## Convexity



Convexity is defined as ratio of perimeters of the convex hull Oconvexhull over that of the original contour

# FEATURE EXTRACTION METHODS TO CODE



# FEATURE EXTRACTED DATAFRAME



# PRINCIPAL COMPONENT ANALYSIS (PCA)

Principal Component Analysis (PCA) is a technique used in statistics and machine learning to simplify and reduce the dimensionality of a dataset. It does this by transforming the original set of variables (features) into a new set of uncorrelated variables (principal components) that capture the most important information in the data. The first principal component (PC) explains the largest possible amount of variance in the data, the second PC explains the second largest variance and so on.

# LABEL ENCODER

Label Encoder is a preprocessing method used in machine learning to convert categorical data, represented as text, into numerical values. The process of label encoding maps each unique category to a unique integer value. This allows algorithms that only work with numerical data to operate on the categorical data.

For example, if we have a categorical feature "color" which can take values "red", "green", "blue", the label encoder will assign the values 0, 1, 2 to each of the colors respectively, so "red" will be 0, "green" will be 1, "blue" will be 2.

It's important to note that label encoder assigns the values based on the alphabetic order of the categories, which can lead to bias if the algorithm relies on the ordinal relationship between the labels. To overcome this issue, we can use one-hot encoding, which creates a new binary column for each category

# ADABOOST CLASSIFIER

AdaBoost, short for Adaptive Boosting, is a machine learning algorithm used for classification and regression problems. It is a meta-algorithm that combines a set of weaker "base" classifiers to form a stronger overall classifier. AdaBoost works by weighting the observations in the training set, and training the base classifiers on these weighted observations. The weighting is adjusted at each iteration so that the observations that were misclassified in the previous iteration are given more weight in the current iteration. The final classifier is a weighted combination of the base classifiers, where the weight assigned to each base classifier is determined by its performance on the training set. AdaBoost is a powerful algorithm that is often used in practice due to its ability to improve the performance of the base classifiers, and its relative simplicity.

# GRADIENTBOOST CLASSIFIER

Gradient Boosting is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. It builds the model in a stage-wise fashion like other boosting methods do, and it generalize them by allowing optimization of an arbitrary differentiable loss function.

At each stage a regression tree is fit on the negative gradient of the loss function that is being optimized. The idea is to fit a new model to the new residuals, rather than to the outcome. In this way it corrects the errors of the previous tree, and the final prediction is the sum of the predictions from all the trees.

Gradient Boosting is a powerful and widely used algorithm, particularly in Kaggle competitions and real-world applications, due to its flexibility, efficiency and ability to handle a variety of data types and shapes. However, it is more complex and computationally expensive than AdaBoost

# COMPARSION MACHINE LEARNING CLASSIFIERS



# CODES ON KAGGLE NOTEBOOK:

<https://www.kaggle.com/mustafakapici/image-processing-feature-extraction-classification>

# My interests are Machine Learning , Deep Learning, Computer Vision

<https://www.kaggle.com/learn/certification/mustafakapici/intro-to-machine-learning>

<https://www.kaggle.com/learn/certification/mustafakapici/intermediate-machine-learning>