Gender Detection
Lecture 12 Assignment
Computer Vision

Sagar Ghimire

PROJECT TOPIC: VIOLENT AND NON-VIOLENT SITUATION DETECTION

- This project aims to detect violent or non-violent situations from the images using Computer Vision.
- The datasets are taken from various sources such as Kaggle and Roboflow.com. Some data are images and some are in video form. Different images will be taken from videos for the project.

Resources for dataset:

- 1.Kaggle.com(images and videos)
- 2.Universal roboflow.com(images)

PROJECT

The main goal of this project is to develop a computer vision system that could perform

an object (for example, coins, faces, animals, cars, plants) recognition (classification)

task by using at least 5 local and global features (Haralick Texture, Box Counting,

Chain Code, Area, and Local Binary Pattern)

Project Introduction

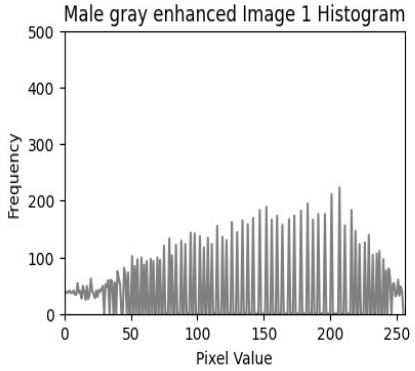
This project is about gender detection where 5 male and 5 female images are used for preprocessing and feature extraction. Finally, the correlation coefficient is used to measure the similarity between males and females respectively.

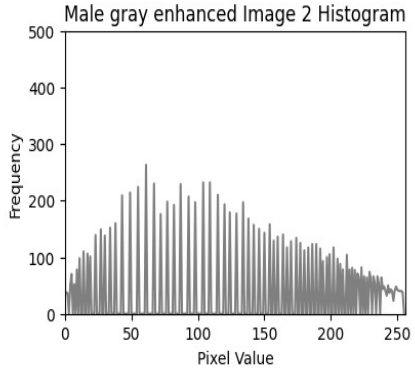
Images taken from Kaggle.com

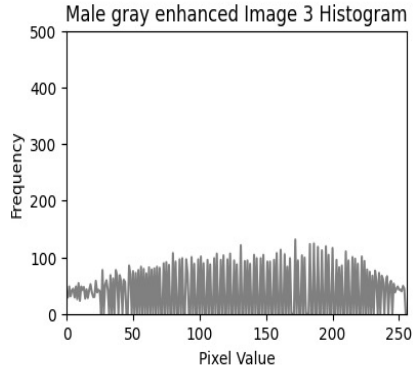
Original color Images

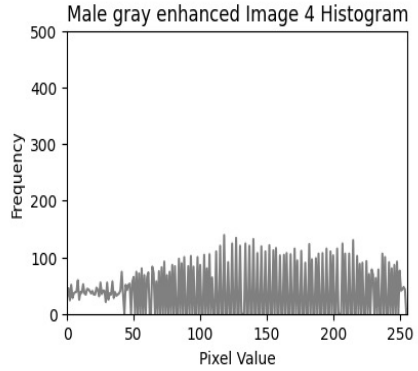
Original gray images

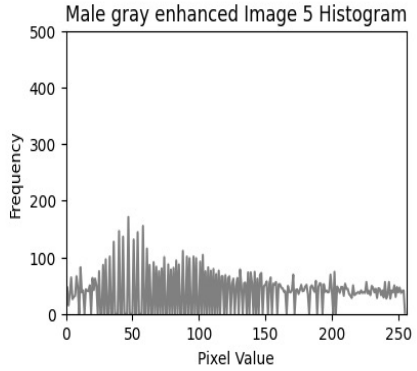
Enhanced Images with Histogram Equalization

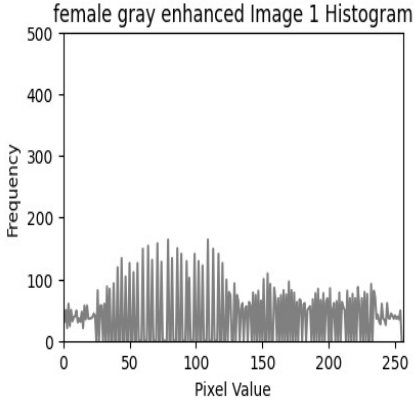


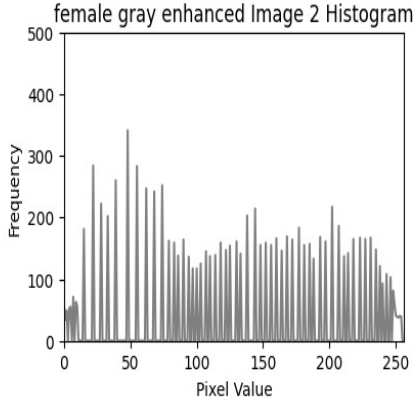


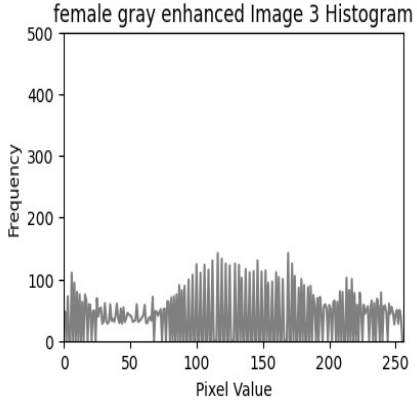


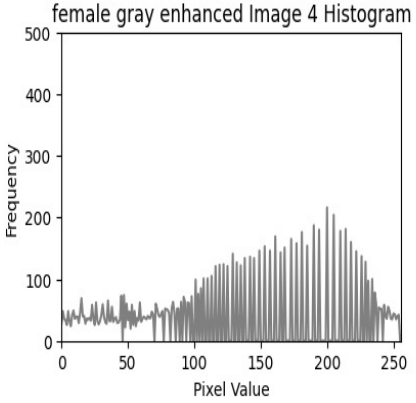


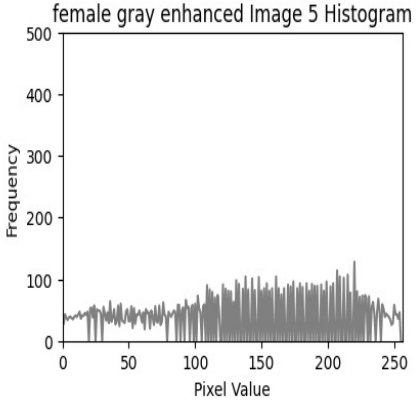






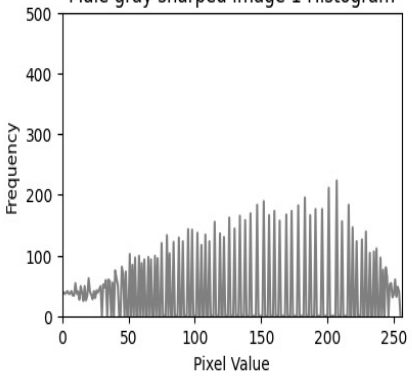


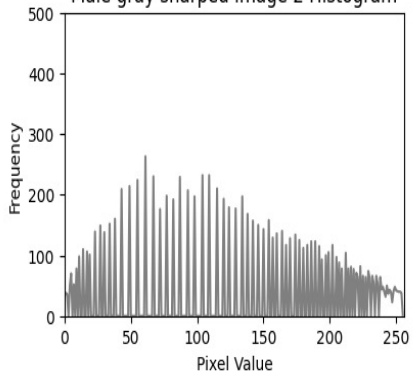


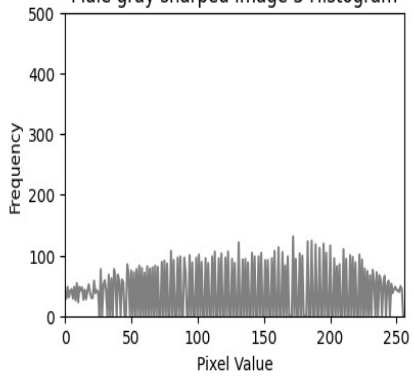


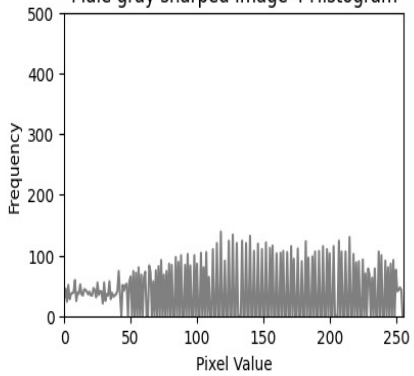
Sharpened Images Using

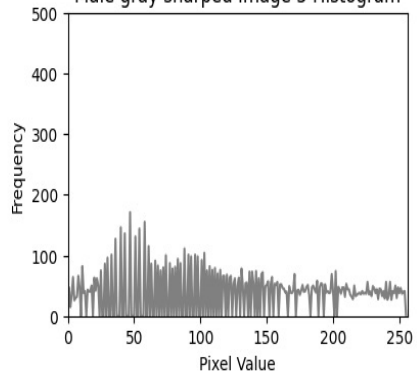
```
kernel = np.array([[-0.5,-.5,-.5],
                   [-.5, 5,-.5],
                   [-.5,-.5,-.5]])
```

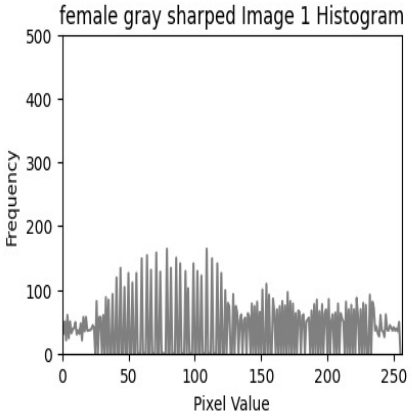


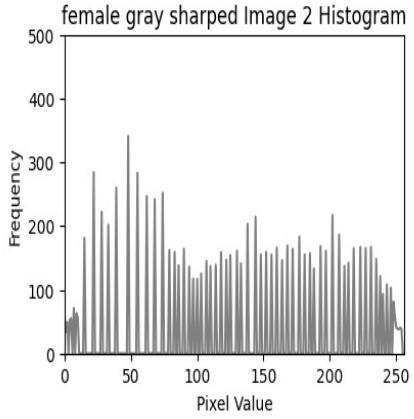


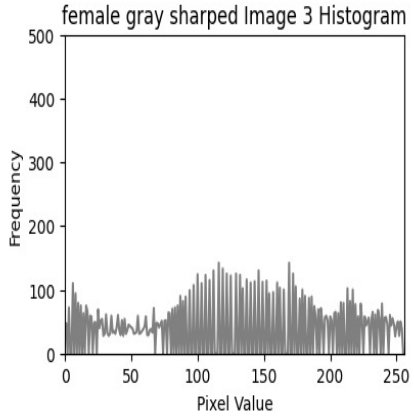


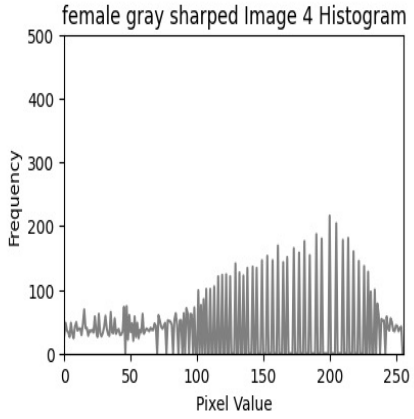


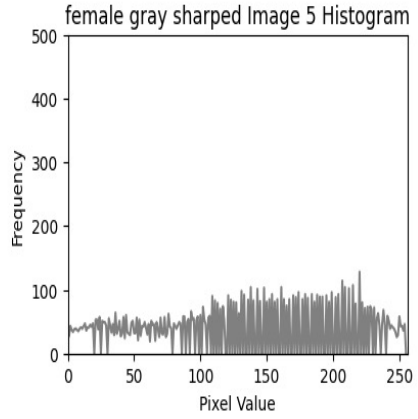












Images Preprocessing and feature generation

1. The images are of different sizes in the list, so resizing them to (100x100) size.
2. Applying Various feature extraction methods to extract features from images.

1.Haralick Texture features

Male Image 1 Haralick values: contrast:
4034.849494949494 homogeneity: 0.1968942557814769
energy: 0.122725693135280 95 correlation:
0.7683814703584944

Male Image 2 Haralick values: contrast:
2312.089191919191 homogeneity: 0.18173083013656305
energy: 0.09992783846718796 correlation:
0.8508849152337469

Male Image 3 Haralick values: contrast:
3983.5212121212126 homogeneity: 0.17383596945223445
energy: 0.10829547736311224 correlation:
0.7590761552708296

Male Image 4 Haralick values: contrast:
2175.6860606060604 homogeneity: 0.16691790537028942
energy: 0.0916323362822626 correlation:
0.8595936550913024

Male Image 5 Haralick values: contrast:
1838.6469696969698 homogeneity: 0.16355958293158473
energy: 0.08306049612007312 correlation:
0.8697979204191079

1

2

3

4

5

Female Image 1 Haralick values: contrast: 1360.080505050505 homogeneity: 0.15385161438812073 energy: 0.051971979217785 21 correlation: 0.8972944219136472

1

Female Image 2 Haralick values: contrast: 1572.3071717171717 homogeneity: 0.13639498081009704 energy: 0.05412770254535551 correlation: 0.8805087357761103

2

Female Image 3 Haralick values: contrast: 1117.5352525252529 homogeneity: 0.147986796652242 energy: 0.052652478008781 88 correlation: 0.9187720432715132

3

Female Image 4 Haralick values: contrast: 2481.2962626262624 homogeneity: 0.143295029754799 02 energy: 0.07394618973019998 correlation: 0.8328574672804369

4

Female Image 5 Haralick values: contrast: 2947.629292929294 homogeneity: 0.1332565737936554 energy: 0.07023587585643595 correlation: 0.8069268116709228

5

2.Box Counting

- Convert to grayscale to deal with binary operations and contour detection
- Setting threshold values to create a binary image from which contour can be detected
- _, thresh = cv2.threshold(gray, 128, 255, cv2.THRESH_BINARY)
- iii. Counter detection in the threshold image which are shapes or boundaries of the object in the image.

Image 1 

Image 2 

Image 3 

Image 4 

Image 5 

Thresholded 

Thresholded 

Thresholded 

Thresholded 

Thresholded 

Box Counting Values for Male Images:

Male Image 1: 268

Male Image 2: 169

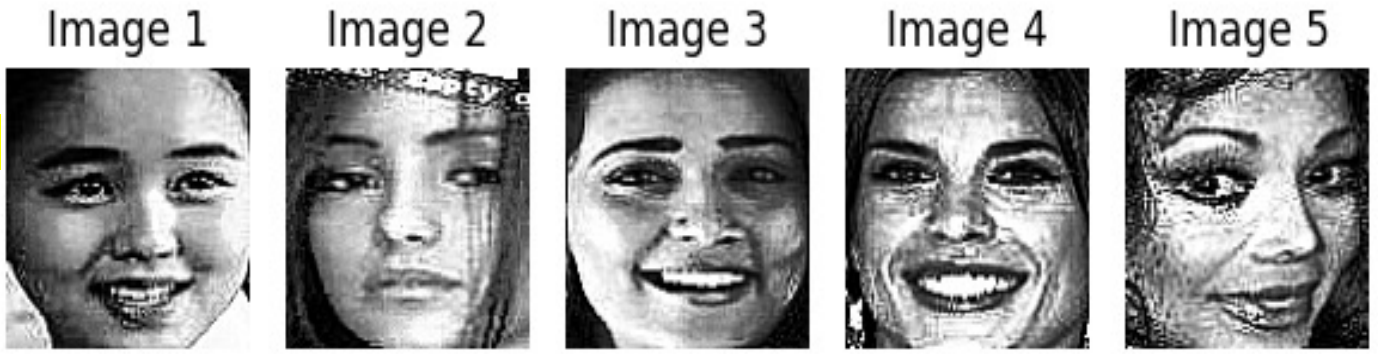
Male Image 3: 268

Male Image 4: 161

Male Image 5: 126

Box counting is a method used in image analysis to determine the complexity or granularity of an image. It involves counting the number of basic shapes (often squares or boxes) required to cover an image at different scales or resolutions.
The process typically involves thresholding the image to convert it into a binary image (black and white), where areas of interest become white (foreground) and the rest become black (background). Then, the number of contours or connected components in the binary image is counted.
.

```
Box Counting Values for Female Images:
 Female Image 1: 113
Female Image 2: 145
Female Image 3: 79
Female Image 4: 169
Female Image 5: 208
```



3.Chain Code

i. Convert to gray for threshold and contour detection
ii. Apply binary threshold and convert to a binary image (0 or 255)
iii. Detects contours in the threshold images

Features extracted

Male Chain Code

Female Chain Code

Male Chain Code

Female Chain Code

Male Chain Code

Female Chain Code

Male Chain Code

Female Chain Code

Male Chain Code

Female Chain Code

Sample values

```
Male Image 1 Chain
Code Values: [[], [],
[], [], [(0, 1)],
[(0, 1)], [(0, 1)],
[(1, 0)], [], [(-1,
1), (0, 1), (-1, 1),
(-1, 0)
```

There are 100's of values for (x,y) for each images

A chain code is a simple and efficient method used in image processing and computer vision for contour representation. It represents a contour by encoding the directions of consecutive boundary points of an object or shape in an image.

4. Extract Area

- Use Numpy to count the number of pixels with values greater than 0. Compute the sum of pixels and calculate the non-zero area regions

```
Male 1:8274
Male 2:8303
Male 3:8118
Male 4:8141
Male 5:8046
```

```
Female 1:8992
Female 2:8303
Female 3:8830
Female 4:8708
Female 5:8678
```

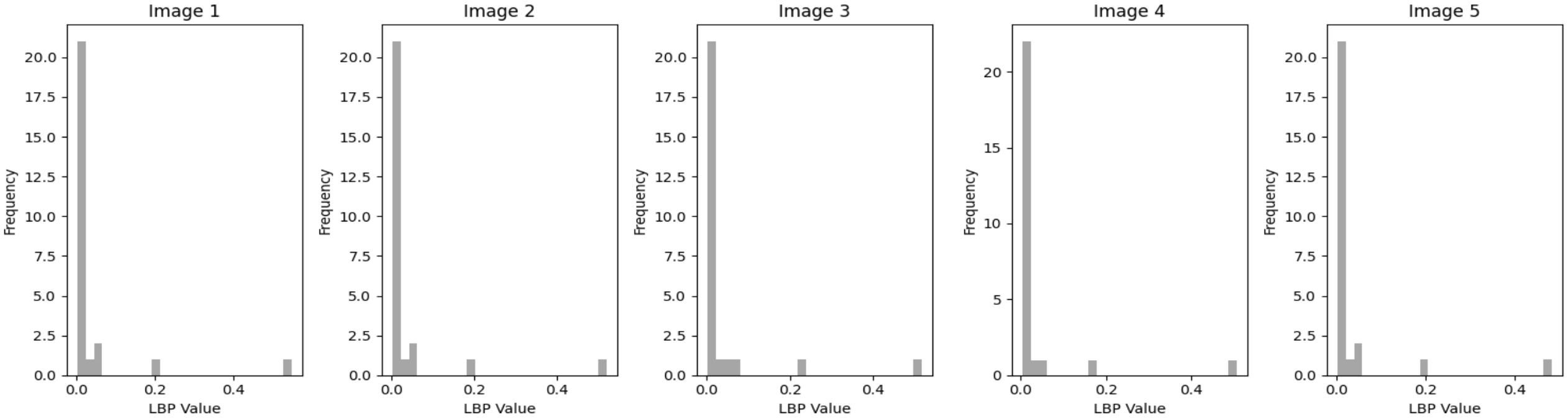
The process of extracting an area from an image involves defining a specific region of interest (ROI) and then cropping the image to retain only that particular area, discarding the rest. This extraction can be based on various criteria such as pixel coordinates, shape, size, or features within the image.

5.Extract lbp(local binary pattern)

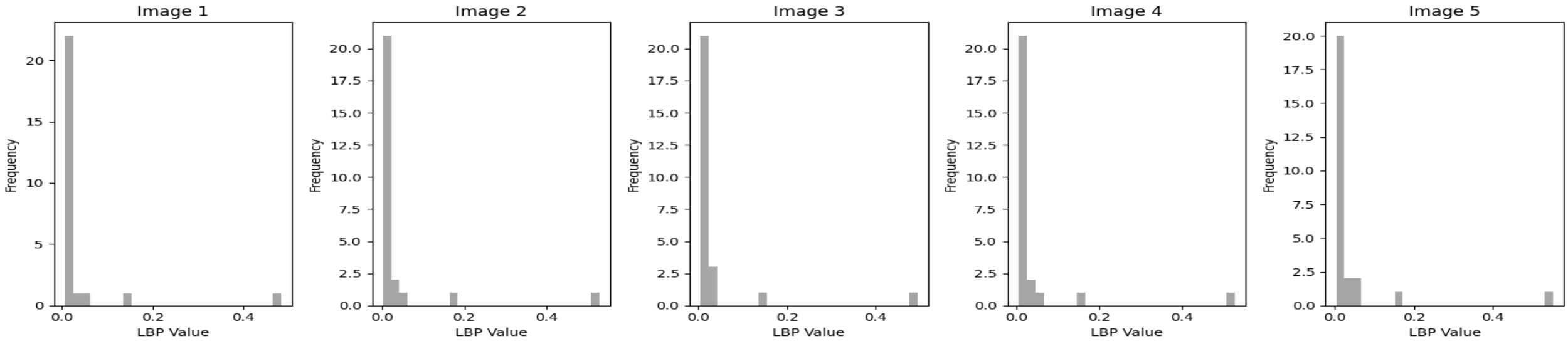
i. Convert to gray
ii. Set the radius and calculate n_points(points in the LBP neighborhood
iii. Computes the LBP image using the parameters
iv. Compute the histogram of the LBP image considering a range of values from 0 to ….

Density =true normalize the hist to be the probability density

Histograms of LBP Values for Male Images



Histograms of LBP Values for female Images



Male Image LBP Values: Image 1 LBP values: [0.0608 0.0611 0.0229 0.0105 0.0049 0.0041 0.0019 0.0029 0.0026 0.0031 0.0028 0.0028 0.0024 0.0038 0.0025 0.003 0.0019 0.0034 0.0032 0.0032 0.0044 0.0065 0.0089 0.0183 0.211 0.5471] Image 2 LBP values: [0.0556 0.0437 0.0305 0.0146 0.009 0.0071 0.006 0.0055 0.0054 0.0057 0.0076 0.008 0.0085 0.01 0.0064 0.0059 0.0038 0.0037 0.004 0.005 0.0051 0.0083 0.0145 0.0164 0.1898 0.5199] Image 3 LBP values: [0.0652 0.0518 0.0264 0.0132 0.0088 0.0065 0.0029 0.0032 0.0036 0.0046 0.0035 0.0045 0.0046 0.0052 0.0035 0.0039 0.003 0.0032 0.0024 0.0033 0.0032 0.006 0.0111 0.0154 0.2246 0.5164] Image 4 LBP values: [0.0528 0.0428 0.0205 0.0117 0.0063 0.0056 0.0059 0.0071 0.0066 0.0079 0.0124 0.0144 0.0203 0.0181 0.0099 0.0085 0.007 0.006 0.0043 0.0063 0.0066 0.0097 0.013 0.0187 0.1721 0.5055] Image 5 LBP values: [0.0525 0.0425 0.0237 0.0154 0.0116 0.0099 0.0071 0.0085 0.0083 0.0083 0.0137 0.0169 0.0189 0.0168 0.0088 0.0076 0.0057 0.0051 0.0041 0.0028 0.0028 0.0063 0.0085 0.0101 0.2008 0.4833]

Female Image LBP Values: Image 1 LBP values: [0.0542 0.0396 0.0213 0.0123 0.01 0.0096 0.0095 0.0114 0.0139 0.016 0.0183 0.0211 0.023 0.0226 0.0128 0.0103 0.0075 0.0073 0.0063 0.007 0.0075 0.0079 0.0107 0.0166 0.1385 0.4848] Image 2 LBP values: [0.0519 0.0418 0.0266 0.0134 0.0104 0.0078 0.0081 0.0073 0.0073 0.0077 0.0099 0.0116 0.0134 0.0118 0.0081 0.0069 0.0063 0.0046 0.0023 0.0048 0.0057 0.0052 0.0111 0.0166 0.1713 0.5281] Image 3 LBP values: [0.0422 0.0385 0.0213 0.0152 0.0127 0.0082 0.0083 0.0105 0.0102 0.0127 0.0173 0.0234 0.0248 0.0207 0.0123 0.0106 0.0072 0.0068 0.0049 0.0064 0.0057 0.0075 0.0106 0.0122 0.1546 0.4952] Image 4 LBP values: [0.0523 0.0424 0.0256 0.011 0.0087 0.0063 0.0048 0.0059 0.0063 0.0059 0.0101 0.0133 0.0161 0.0159 0.01 0.0083 0.0071 0.0064 0.0043 0.0048 0.0058 0.0105 0.0138 0.0167 0.158 0.5297] Image 5 LBP values: [0.062 0.0507 0.0259 0.0125 0.0065 0.0045 0.0032 0.0032 0.0037 0.0067 0.0048 0.0062 0.0084 0.008 0.0055 0.0055 0.0043 0.004 0.0047 0.0031 0.0053 0.0087 0.0126 0.025 0.1622 0.5528]

Now, combining features

- Null Values check
- **Gender 0**
- **Haralick_Contrast 0**
- **Haralick_Homogeneity 0**
- **Haralick_Energy 0**
- **Haralick_Correlation 0**
- **Box_Counting_Values 0**
- **Area_Values 0**
- **LBP_Values 0**
- **Chain_Code_Values 0**

	Gender	Haralick_Contrast	Haralick_Homogeneity	Haralick_Energy	Haralick_Correlation	Box_Counting_Values	Area_Values	LBP_Values	Chain_Code_Values
0	Female	1360.080505	0.153852	0.051972	0.897294	113	9291	[0.0542, 0.0396, 0.0213, 0.0123, 0.01, 0.0096,...	[[(-1, 1), (0, 1), (-1, 1), (-1, 0), (1, 0), (...
1	Female	1572.307172	0.136395	0.054128	0.880509	145	9276	[0.0519, 0.0418, 0.0266, 0.0134, 0.0104, 0.007...	[[(1, 1), (1, 0), (1, 1), (1, 0), (1, 0), (1, ...
2	Female	1117.535253	0.147987	0.052652	0.918772	79	9382	[0.0422, 0.0385, 0.0213, 0.0152, 0.0127, 0.008...	[[], [], [], [(1, 0), (1, 0), (1, 0), (1, 1), ...
3	Female	2481.296263	0.143295	0.073946	0.832857	169	9055	[0.0523, 0.0424, 0.0256, 0.011, 0.0087, 0.0063...	[[(0, 1), (0, 1), (0, 1), (0, -1), (0, -1)], [...
4	Female	2947.629293	0.133257	0.070236	0.806927	208	8994	[0.062, 0.0507, 0.0259, 0.0125, 0.0065, 0.0045...	[[], [(0, 1), (0, 1), (0, 1), (0, -1), (0, -1)...

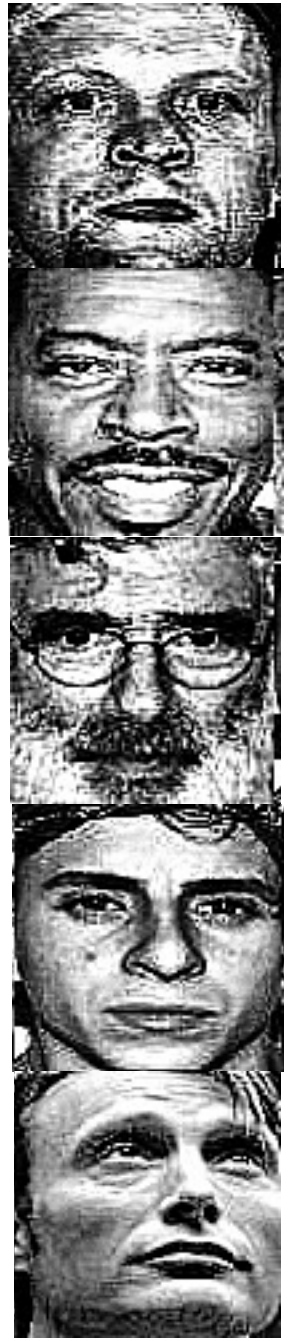
	Gender	Haralick_Contrast	Haralick_Homogeneity	Haralick_Energy	Haralick_Correlation	Box_Counting_Values	Area_Values	LBP_Values	Chain_Code_Values
0	Male	4034.849495	0.196894	0.122726	0.768381	268	8606	[0.0608, 0.0611, 0.0229, 0.0105, 0.0049, 0.004...	[[(1, 0), (1, 0), (-1, 0)], [(0, 1), (1, 1), (...
1	Male	2312.089192	0.181731	0.099928	0.850885	169	8827	[0.0556, 0.0437, 0.0305, 0.0146, 0.009, 0.0071...	[[], [], [], [(1, 0), (1, 0), (-1, 0)], [], [(...
2	Male	3983.521212	0.173836	0.108295	0.759076	268	8605	[0.0652, 0.0518, 0.0264, 0.0132, 0.0088, 0.006...	[[], [], [(0, 1)], [], [(1, 0)], [(0, 1), (0, ...
3	Male	2175.686061	0.166918	0.091632	0.859594	161	8760	[0.0528, 0.0428, 0.0205, 0.0117, 0.0063, 0.005...	[[(0, 1), (0, 1), (0, 1), (0, 1), (0, 1), (0, ...
4	Male	1838.646970	0.163560	0.083060	0.869798	126	8957	[0.0525, 0.0425, 0.0237, 0.0154, 0.0116, 0.009...	[[], [(-1, 1), (0, 1), (1, 1), (0, 1), (0, 1),...

Similarity Check for features using the Correlation Coefficient

Using LBP and Chain Code features complicated the calculations as they have 1000 of (x,y) values in array form, also they reduced the Similarity measures, so dropping these two features from the data frame and also tried distance metrics which doesn't seem to be working as threshold needed to be about 5000.

We are taking row 0 to compare with other rows to check the similarity measures

	Gender	Haralick_Contrast	Haralick_Homogeneity	Haralick_Energy	Haralick_Correlation	Box_Counting_Values	Area_Values	LBP_Values	Chain_Code_Values
0	Male	4034.849495	0.196894	0.122726	0.768381	268	8606	[0.0608, 0.0611, 0.0229, 0.0105, 0.0049, 0.004...	[[(1, 0), (1, 0), (-1, 0)], [(0, 1), (1, 1), (...
1	Male	2312.089192	0.181731	0.099928	0.850885	169	8827	[0.0556, 0.0437, 0.0305, 0.0146, 0.009, 0.0071...	[[], [], [], [(1, 0), (1, 0), (-1, 0)], [], [(...
2	Male	3983.521212	0.173836	0.108295	0.759076	268	8605	[0.0652, 0.0518, 0.0264, 0.0132, 0.0088, 0.006...	[[], [], [(0, 1)], [], [(1, 0)], [(0, 1), (0, ...
3	Male	2175.686061	0.166918	0.091632	0.859594	161	8760	[0.0528, 0.0428, 0.0205, 0.0117, 0.0063, 0.005...	[[(0, 1), (0, 1), (0, 1), (0, 1), (0, 1), (0, ...
4	Male	1838.646970	0.163560	0.083060	0.869798	126	8957	[0.0525, 0.0425, 0.0237, 0.0154, 0.0116, 0.009...	[[], [(-1, 1), (0, 1), (1, 1), (0, 1), (0, 1),...



Male Images features

Row 0 Threshold	Similarity	Dissimilarity
0.91	row1, row2, row3, row4	None
0.95	row1,row2,row3,row4	None
0.99	row2	row1, row3, row4
1	None	row1,row2,row3,row4

	Gender	Haralick_Contrast	Haralick_Homogeneity	Haralick_Energy	Haralick_Correlation	Box_Counting_Values	Area_Values	LBP_Values	Chain_Code_Values
0	Female	1360.080505	0.153852	0.051972	0.897294	113	9291	[0.0542, 0.0396, 0.0213, 0.0123, 0.01, 0.0096,...	[[(-1, 1), (0, 1), (-1, 1), (-1, 0), (1, 0), (...
1	Female	1572.307172	0.136395	0.054128	0.880509	145	9276	[0.0519, 0.0418, 0.0266, 0.0134, 0.0104, 0.007...	[[(1, 1), (1, 0), (1, 1), (1, 0), (1, 0), (1, ...
2	Female	1117.535253	0.147987	0.052652	0.918772	79	9382	[0.0422, 0.0385, 0.0213, 0.0152, 0.0127, 0.008...	[[], [], [], [(1, 0), (1, 0), (1, 0), (1, 1), ...
3	Female	2481.296263	0.143295	0.073946	0.832857	169	9055	[0.0523, 0.0424, 0.0256, 0.011, 0.0087, 0.0063...	[[(0, 1), (0, 1), (0, 1), (0, -1), (0, -1)], [...
4	Female	2947.629293	0.133257	0.070236	0.806927	208	8994	[0.062, 0.0507, 0.0259, 0.0125, 0.0065, 0.0045...	[[], [(0, 1), (0, 1), (0, 1), (0, -1), (0, -1)...

Female Images features

Threshold	Similarity	Dissimilarity
0.91	row1, row2, row3, row4	None
0.95	row1, row2, row3, row4	None
0.99	row1, row2, row3	row4
1	None	row1, row2, row3, row4

Applied chi-square

	Gender	Haralick_Contrast	Haralick_Homogeneity	Haralick_Energy	Haralick_Correlation	Box_Counting_Values	Area_Values	LBP_Values	Chain_Code_Values
0	Male	4034.849495	0.196894	0.122726	0.768381	268	8606	[0.0608, 0.0611, 0.0229, 0.0105, 0.0049, 0.004...	[[(1, 0), (1, 0), (-1, 0)], [(0, 1), (1, 1), (...
1	Male	2312.089192	0.181731	0.099928	0.850885	169	8827	[0.0556, 0.0437, 0.0305, 0.0146, 0.009, 0.0071...	[[], [], [], [(1, 0), (1, 0), (-1, 0)], [], [(...
2	Male	3983.521212	0.173836	0.108295	0.759076	268	8605	[0.0652, 0.0518, 0.0264, 0.0132, 0.0088, 0.006...	[[], [], [(0, 1)], [], [(1, 0)], [(0, 1), (0, ...
3	Male	2175.686061	0.166918	0.091632	0.859594	161	8760	[0.0528, 0.0428, 0.0205, 0.0117, 0.0063, 0.005...	[[(0, 1), (0, 1), (0, 1), (0, 1), (0, 1), (0, ...
4	Male	1838.646970	0.163560	0.083060	0.869798	126	8957	[0.0525, 0.0425, 0.0237, 0.0154, 0.0116, 0.009...	[[], [(-1, 1), (0, 1), (1, 1), (0, 1), (0, 1),...

Male

Threshold =0.05
No similar rows

Threshold = 0.5
Row 2 similar to row 0

Threshold= 1000
Row1, Row2, Row3, Row4 similar to Row

	Gender	Haralick_Contrast	Haralick_Homogeneity	Haralick_Energy	Haralick_Correlation	Box_Counting_Values	Area_Values	LBP_Values	Chain_Code_Values
0	Female	1360.080505	0.153852	0.051972	0.897294	113	9291	[0.0542, 0.0396, 0.0213, 0.0123, 0.01, 0.0096,...	[[(-1, 1), (0, 1), (-1, 1), (-1, 0), (1, 0), (...
1	Female	1572.307172	0.136395	0.054128	0.880509	145	9276	[0.0519, 0.0418, 0.0266, 0.0134, 0.0104, 0.007...	[[(1, 1), (1, 0), (1, 1), (1, 0), (1, 0), (1, ...
2	Female	1117.535253	0.147987	0.052652	0.918772	79	9382	[0.0422, 0.0385, 0.0213, 0.0152, 0.0127, 0.008...	[[], [], [], [(1, 0), (1, 0), (1, 0), (1, 1), ...
3	Female	2481.296263	0.143295	0.073946	0.832857	169	9055	[0.0523, 0.0424, 0.0256, 0.011, 0.0087, 0.0063...	[[(0, 1), (0, 1), (0, 1), (0, -1), (0, -1)], [...
4	Female	2947.629293	0.133257	0.070236	0.806927	208	8994	[0.062, 0.0507, 0.0259, 0.0125, 0.0065, 0.0045...	[[], [(0, 1), (0, 1), (0, 1), (0, -1), (0, -1)...

Female

Threshold =0.05
No similar rows

Threshold = 0.5
No similar rows.

Threshold= 1000
Row1, Row2, Row3, Row4 similar to Row

Let's try a few more features

1.Gabor_filter

A Gabor filter is a linear filter used in signal processing and image processing for tasks like texture analysis, edge detection, and feature extraction. It's particularly useful for analyzing textures within images.

male_df

	Haralick_Contrast	Haralick_Homogeneity	Haralick_Energy	Haralick_Correlation	Box_Counting_Values	Area_Values	BP_Values	Chain_Code_Values	Male_Gabor_Pixel_Values_1	Male_Gabor_Pixel_Values_2	Male_Gabor_Pixel_Values_3	Male_Gabor_Pixel_Values_4	Male_Gabor_Pixel_Values_5
0	4034.849495	0.196894	0.122726	0.768381	268	8606	[0.0608, 0.0611, 0.0229, 0.0105, 0.0049, 0.004...	[[(1, 0), (1, 0), (-1, 0)], [(0, 1), (1, 1), (...	[223, 255, 255, 255, 255, 255, 255, 255, 255, ...	[223, 255, 255, 255, 255, 255, 255, 255, 255, ...	[223, 255, 255, 255, 255, 255, 255, 255, 255, ...	[223, 255, 255, 255, 255, 255, 255, 255, 255, ...	[223, 255, 255, 255, 255, 255, 255, 255, 255, ...
1	2312.089192	0.181731	0.099928	0.850885	169	8827	[0.055... 0.0437, 0.0305, 0.0146, 0.009, 0.0071...	[[], [], [], [(1, 0), (1, 0), (-1, 0)], [], [(...	[70, 76, 108, 139, 166, 236, 255, 255, 255, 25...	[70, 76, 108, 139, 166, 236, 255, 255, 255, 25...	[70, 76, 108, 139, 166, 236, 255, 255, 255, 25...	[70, 76, 108, 139, 166, 236, 255, 255, 255, 25...	[70, 76, 108, 139, 166, 236, 255, 255, 255, 25...
2	3983.521212	0.173836	0.108295	0.759076	268	8605	[0.0652, 0.0518, 0.0264, 0.0132, 0.0088, 0.006...	[[], [], [(0, 1)], [], [(1, 0)], [(0, 1), (0, ...	[255, 255, 255, 255, 255, 255, 248, 138, 162, ...	[255, 255, 255, 255, 255, 255, 248, 138, 162, ...	[255, 255, 255, 255, 255, 255, 248, 138, 162, ...	[255, 255, 255, 255, 255, 255, 248, 138, 162, ...	[255, 255, 255, 255, 255, 255, 248, 138, 162, ...
3	2175.686061	0.166918	0.091632	0.859594	161	8760	[0.0528, 0.0428, 0.0205, 0.0117, 0.0063, 0.005...	[[(0, 1), (0, 1), (0, 1), (0, 1), (0, 1), (0, ...	[58, 58, 90, 142, 176, 194, 167, 131, 110, 128...	[58, 58, 90, 142, 176, 194, 167, 131, 110, 128...	[58, 58, 90, 142, 176, 194, 167, 131, 110, 128...	[58, 58, 90, 142, 176, 194, 167, 131, 110, 128...	[58, 58, 90, 142, 176, 194, 167, 131, 110, 128...
4	1838.646970	0.163560	0.083060	0.869798	126	8957	[0.0525, 0.0425, 0.0237, 0.0154, 0.0116, 0.009...	[[], [(-1, 1), (0, 1), (0, 1), (0, 1), (0, 1), ...	[255, 255, 255, 255, 255, 255, 255, 255, 255, ...	[255, 255, 255, 255, 255, 255, 255, 255, 255, ...	[255, 255, 255, 255, 255, 255, 255, 255, 255, ...	[255, 255, 255, 255, 255, 255, 255, 255, 255, ...	[255, 255, 255, 255, 255, 255, 255, 255, 255, ...

female_df

	Haralick_Contrast	Haralick_Homogeneity	Haralick_Energy	Haralick_Correlation	Box_Counting_Values	Area_Values	BP_Values	Chain_Code_Values	Female_Gabor_Pixel_Values_1	Female_Gabor_Pixel_Values_2	Female_Gabor_Pixel_Values_3	Female_Gabor_Pixel_Values_4	Female_Gabor_Pixel_Values_5
0	1360.080505	0.153852	0.051972	0.897294	113	9291	[0.0542, 0.0396, 0.0213, 0.0123, 0.01, 0.0056,...	[[(-1, 1), (0, 1), (-1, 1), (-1, 0), (1, 0), (...	[25, 26, 29, 34, 37, 35, 37, 33, 35, 43, 55, 7...	[25, 26, 29, 34, 37, 35, 37, 33, 35, 43, 55, 7...	[25, 26, 29, 34, 37, 35, 37, 33, 35, 43, 55, 7...	[25, 26, 29, 34, 37, 35, 37, 33, 35, 43, 55, 7...	[25, 26, 29, 34, 37, 35, 37, 33, 35, 43, 55, 7...
1	1572.307172	0.136395	0.054128	0.880509	145	9276	[0.0519, 0.0418, 0.0266, 0.0134, 0.0104, 0.007...	[[(1, 1), (1, 0), (1, 1), (1, 0), (1, 0), (1, ...	[255, 255, 255, 255, 255, 255, 255, 255, 255, ...	[255, 255, 255, 255, 255, 255, 255, 255, 255, ...	[255, 255, 255, 255, 255, 255, 255, 255, 255, ...	[255, 255, 255, 255, 255, 255, 255, 255, 255, ...	[255, 255, 255, 255, 255, 255, 255, 255, 255, ...
2	1117.535253	0.147987	0.052652	0.918772	79	9382	[0.0422, 0.0385, 0.0213, 0.0152, 0.0127, 0.008...	[[], [], [], [(1, 0), (1, 0), (-1, 0), (1, 1), ...	[225, 222, 226, 225, 247, 242, 227, 206, 210, ...	[225, 222, 226, 225, 247, 242, 227, 206, 210, ...	[225, 222, 226, 225, 247, 242, 227, 206, 210, ...	[225, 222, 226, 225, 247, 242, 227, 206, 210, ...	[225, 222, 226, 225, 247, 242, 227, 206, 210, ...
3	2481.296263	0.143295	0.073946	0.832857	169	9055	[0.0523, 0.0424, 0.0256, 0.011, 0.0087, 0.0063...	[[(0, 1), (0, 1), (0, 1), (0, -1), (0, 1)], [...	[255, 255, 255, 255, 255, 255, 255, 255, 255, ...	[255, 255, 255, 255, 255, 255, 255, 255, 255, ...	[255, 255, 255, 255, 255, 255, 255, 255, 255, ...	[255, 255, 255, 255, 255, 255, 255, 255, 255, ...	[255, 255, 255, 255, 255, 255, 255, 255, 255, ...
4	2947.629293	0.133257	0.070236	0.806927	208	8994	[0.062, 0.0507, 0.0259, 0.0125, 0.0065, 0.0045...	[[], [(0, 1), (0, 1), (0, 1), (0, -1), (0, -1)...	[255, 255, 244, 188, 149, 127, 136, 157, 186, ...	[255, 255, 244, 188, 149, 127, 136, 157, 186, ...	[255, 255, 244, 188, 149, 127, 136, 157, 186, ...	[255, 255, 244, 188, 149, 127, 136, 157, 186, ...	[255, 255, 244, 188, 149, 127, 136, 157, 186, ...

PCA

```
Transformed Images for Males:
Image 1 - Transformed Values: [-3.41003199e+03 -3.84031191e+03 -4.89872177e+03 -1.66046846e+03 2.44999530e-12]
Image 2 - Transformed Values: [-2.01860698e+03 6.12583677e+02 6.07860477e+02 6.28077781e+03 2.44999530e-12]
Image 3 - Transformed Values: [ 8.19002341e+03 -3.64009948e+02 -1.22320292e+03 5.27395010e+01 2.44999530e-12]
Image 4 - Transformed Values: [-1.00270469e+03 -3.38774879e+03 5.60569700e+03 -2.16166229e+03 2.44999530e-12]
Image 5 - Transformed Values: [-1.75867975e+03 6.97948696e+03 -9.16327783e+01 -2.51138657e+03 2.44999530e-12]
```

```
Transformed Images for Females:
Image 1 - Transformed Values: [7.63262172e+03 3.34708167e+03 1.05123081e+02 3.83509980e+02 2.11030588e-12]
Image 2 - Transformed Values: [ 1.29391428e+03 -4.45219382e+03 -1.51947379e+03 -4.03004876e+03 2.11030588e-12]
Image 3 - Transformed Values: [-5.64016238e+03 4.49466163e+03 -9.29661705e+02 -2.05250355e+03 2.11030588e-12]
Image 4 - Transformed Values: [-1.61902870e+03 -1.80431681e+03 -3.54873939e+03 4.41191379e+03 2.11030588e-12]
Image 5 - Transformed Values: [-1.66734492e+03 -1.58523267e+03 5.89275180e+03 1.28712855e+03 2.11030588e-12]
```

A combined data frame with all the features

Excluding the crossed parts they were affecting the similarity measures.

	Haralick_Contrast	Haralick_Homogeneity	Haralick_Energy	Haralick_Correlation	Box_Counting_Values	Area_Values	Male_Gabor_Pixel_Values_1	Male_Gabor_Pixel_Values_2	Male_Gabor_Pixel_Values_3	Male_Gabor_Pixel_Values_4	Male_Gabor_Pixel_Values_5	Male_PCA_1	Male_PCA_2	Male_PCA_3	Male_PCA_4	Male_PCA_5
0	4034.849495	0.196894	0.122726	0.768381	268	8606	[223, 255, 255, 255, 255, 255, 255, 255, 255, ...	[223, 255, 255, 255, 255, 255, 255, 255, 255, ...	[223, 255, 255, 255, 255, 255, 255, 255, 255, ...	[223, 255, 255, 255, 255, 255, 255, 255, 255, ...	[223, 255, 255, 255, 255, 255, 255, 255, 255, ...	-3410.031991	-3840.311905	-4898.721772	-1660.468456	2.449995e-12
1	2312.089192	0.181731	0.099928	0.850885	169	8827	[70, 76, 108, 139, 166, 236, 255, 255, 255, 25...	[70, 76, 108, 139, 166, 236, 255, 255, 255, 25...	[70, 76, 108, 139, 166, 236, 255, 255, 255, 25...	[70, 76, 108, 139, 166, 236, 255, 255, 255, 25...	[70, 76, 108, 139, 166, 236, 255, 255, 255, 25...	-2018.606979	612.583677	607.860477	6280.777813	2.449995e-12
2	3983.521212	0.173836	0.108295	0.759076	268	8605	[255, 255, 255, 255, 255, 255, 248, 138, 162, ...	[255, 255, 255, 255, 255, 255, 248, 138, 162, ...	[255, 255, 255, 255, 255, 255, 248, 138, 162, ...	[255, 255, 255, 255, 255, 255, 248, 138, 162, ...	[255, 255, 255, 255, 255, 255, 248, 138, 162, ...	8190.023413	-364.009948	-1223.202923	52.739501	2.449995e-12
3	2175.686061	0.166918	0.091632	0.859594	161	8760	[58, 58, 90, 142, 176, 194, 167, 131, 110, 128...	[58, 58, 90, 142, 176, 194, 167, 131, 110, 128...	[58, 58, 90, 142, 176, 194, 167, 131, 110, 128...	[58, 58, 90, 142, 176, 194, 167, 131, 110, 128...	[58, 58, 90, 142, 176, 194, 167, 131, 110, 128...	-1002.704694	-3387.748787	5605.696996	-2161.662287	2.449995e-12
4	1838.646970	0.163560	0.083060	0.869798	126	8957	[255, 255, 255, 255, 255, 255, 255, 255, 255, ...	[255, 255, 255, 255, 255, 255, 255, 255, 255, ...	[255, 255, 255, 255, 255, 255, 255, 255, 255, ...	[255, 255, 255, 255, 255, 255, 255, 255, 255, ...	[255, 255, 255, 255, 255, 255, 255, 255, 255, ...	-1758.679749	6979.486963	-91.632778	-2511.386572	2.449995e-12

```
combined_female_with_pca1= combined_female_with_pca.drop(["LBP_Values", "Chain_Code_Values"], axis=1)
combined_female_with_pca1
```

Haralick_Contrast	Haralick_Homogeneity	Haralick_Energy	Haralick_Correlation	Box_Counting_Values	Area_Values	Female_Gabor_Pixel_Values_1	Female_Gabor_Pixel_Values_2	Female_Gabor_Pixel_Values_3	Female_Gabor_Pixel_Values_4	Female_Gabor_Pixel_Values_5	Female_PCA_1	Female_PCA_2	Female_PCA_3	Female_PCA_4	Female_PCA_5
1360.080505	0.153852	0.051972	0.897294	113	9291	[25, 26, 29, 34, 37, 35, 37, 33, 35, 43, 55, 7...	[25, 26, 29, 34, 37, 35, 37, 33, 35, 43, 55, 7...	[25, 26, 29, 34, 37, 35, 37, 33, 35, 43, 55, 7...	[25, 26, 29, 34, 37, 35, 37, 33, 35, 43, 55, 7...	[25, 26, 29, 34, 37, 35, 37, 33, 35, 43, 55, 7...	7632.621719	3347.081673	105.123081	383.509980	2.110306e-12
1572.307172	0.136395	0.054128	0.880509	145	9276	[255, 255, 255, 255, 255, 255, 255, 255, 255, ...	[255, 255, 255, 255, 255, 255, 255, 255, 255, ...	[255, 255, 255, 255, 255, 255, 255, 255, 255, ...	[255, 255, 255, 255, 255, 255, 255, 255, 255, ...	[255, 255, 255, 255, 255, 255, 255, 255, 255, ...	1293.914279	-4452.193821	-1519.473788	-4030.048762	2.110306e-12
1117.535253	0.147987	0.052652	0.918772	79	9382	[225, 222, 226, 225, 247, 242, 227, 206, 210, ...	[225, 222, 226, 225, 247, 242, 227, 206, 210, ...	[225, 222, 226, 225, 247, 242, 227, 206, 210, ...	[225, 222, 226, 225, 247, 242, 227, 206, 210, ...	[225, 222, 226, 225, 247, 242, 227, 206, 210, ...	-5640.162380	4494.661634	-929.661705	-2052.503553	2.110306e-12
2481.296263	0.143295	0.073946	0.832857	169	9055	[255, 255, 255, 255, 255, 255, 255, 255, 255, ...	[255, 255, 255, 255, 255, 255, 255, 255, 255, ...	[255, 255, 255, 255, 255, 255, 255, 255, 255, ...	[255, 255, 255, 255, 255, 255, 255, 255, 255, ...	[255, 255, 255, 255, 255, 255, 255, 255, 255, ...	-1619.028699	-1804.316812	-3548.739389	4411.913786	2.110306e-12
2947.629293	0.133257	0.070236	0.806927	208	8994	[255, 255, 244, 188, 149, 127, 136, 157, 186, ...	[255, 255, 244, 188, 149, 127, 136, 157, 186, ...	[255, 255, 244, 188, 149, 127, 136, 157, 186, ...	[255, 255, 244, 188, 149, 127, 136, 157, 186, ...	[255, 255, 244, 188, 149, 127, 136, 157, 186, ...	-1667.344919	-1585.232674	5892.751801	1287.128549	2.110306e-12

Taking row 0 as the main row to compare with others

Male

Threshold	Similarity	Dissimilarity
0.30	row1, row2, row3, row4	None
0.45	row1, row2, row3, row4	None

Female

Threshold	Similarity	Dissimilarity
0.30	row1, row2, row3, row4	None
0.45	row1	row2, row3, row4

Tried removing and adding other features but got no better results than above slides

Conclusion:

Having many features doesn't mean that the model will train or work better, we need to check for the best features that are highly correlated and important in the dataset.
Also, we saw correlation coefficient worked better for the first 3 features than Ch--Square Haralick, Box counting and Area but Ch—Square worked very weakly.

Best Worked Measure

	Gender	Haralick_Contrast	Haralick_Homogeneity	Haralick_Energy	Haralick_Correlation	Box_Counting_Values	Area_Values	LBP_Values	Chain_Code_Values
0	Male	4034.849495	0.196894	0.122726	0.768381	268	8606	[0.0608, 0.0611, 0.0229, 0.0105, 0.0049, 0.004...	[[(1, 0), (1, 0), (-1, 0)], [(0, 1), (1, 1), (...
1	Male	2312.089192	0.181731	0.099928	0.850885	169	8827	[0.0556, 0.0437, 0.0305, 0.0146, 0.009, 0.0071...	[[], [], [], [(1, 0), (1, 0), (-1, 0)], [], [(...
2	Male	3983.521212	0.173836	0.108295	0.759076	268	8605	[0.0652, 0.0518, 0.0264, 0.0132, 0.0088, 0.006...	[[], [], [(0, 1)], [], [(1, 0)], [(0, 1), (0, ...
3	Male	2175.686061	0.166918	0.091632	0.859594	161	8760	[0.0528, 0.0428, 0.0205, 0.0117, 0.0063, 0.005...	[[(0, 1), (0, 1), (0, 1), (0, 1), (0, 1), (0, ...
4	Male	1838.646970	0.163560	0.083060	0.869798	126	8957	[0.0525, 0.0425, 0.0237, 0.0154, 0.0116, 0.009...	[[], [(-1, 1), (0, 1), (1, 1), (0, 1), (0, 1),...

Threshold	Similarity	DIssimilarity
0.91	row1, row2, row3, row4	None
0.95	row1,row2,row3,row4	None
0.99	row2	row1, row3, row4
1	None	row1,row2,row3,row4

	Gender	Haralick_Contrast	Haralick_Homogeneity	Haralick_Energy	Haralick_Correlation	Box_Counting_Values	Area_Values	LBP_Values	Chain_Code_Values
0	Female	1360.080505	0.153852	0.051972	0.897294	113	9291	[0.0542, 0.0396, 0.0213, 0.0123, 0.01, 0.0096,...	[[(-1, 1), (0, 1), (-1, 1), (-1, 0), (1, 0), (...
1	Female	1572.307172	0.136395	0.054128	0.880509	145	9276	[0.0519, 0.0418, 0.0266, 0.0134, 0.0104, 0.007...	[[(1, 1), (1, 0), (1, 1), (1, 0), (1, 0), (1, ...
2	Female	1117.535253	0.147987	0.052652	0.918772	79	9382	[0.0422, 0.0385, 0.0213, 0.0152, 0.0127, 0.008...	[[], [], [], [(1, 0), (1, 0), (1, 0), (1, 1), ...
3	Female	2481.296263	0.143295	0.073946	0.832857	169	9055	[0.0523, 0.0424, 0.0256, 0.011, 0.0087, 0.0063...	[[(0, 1), (0, 1), (0, 1), (0, -1), (0, -1)], [...
4	Female	2947.629293	0.133257	0.070236	0.806927	208	8994	[0.062, 0.0507, 0.0259, 0.0125, 0.0065, 0.0045...	[[], [(0, 1), (0, 1), (0, 1), (0, -1), (0, -1)...

Threshold	**Similarity**	**Dissimilarity**
0.91	row1, row2, row3, row4	None
0.95	row1, row2, row3, row4	None
0.99	row1, row2, row3	row4
1	None	row1, row2, row3, row4

##########################THE END##########################