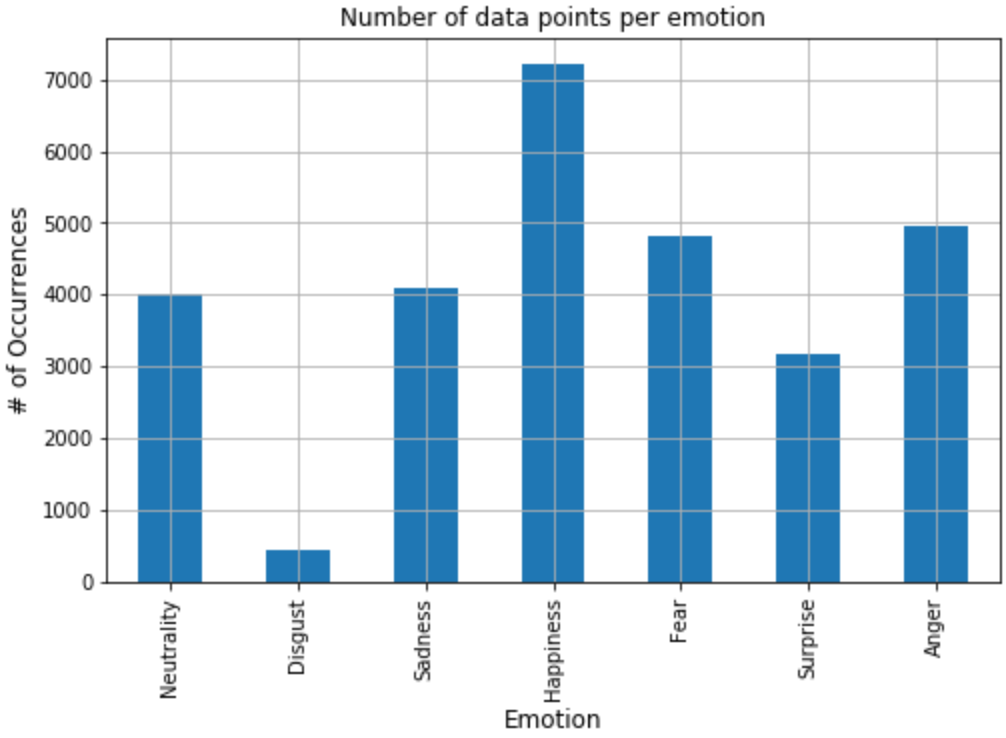
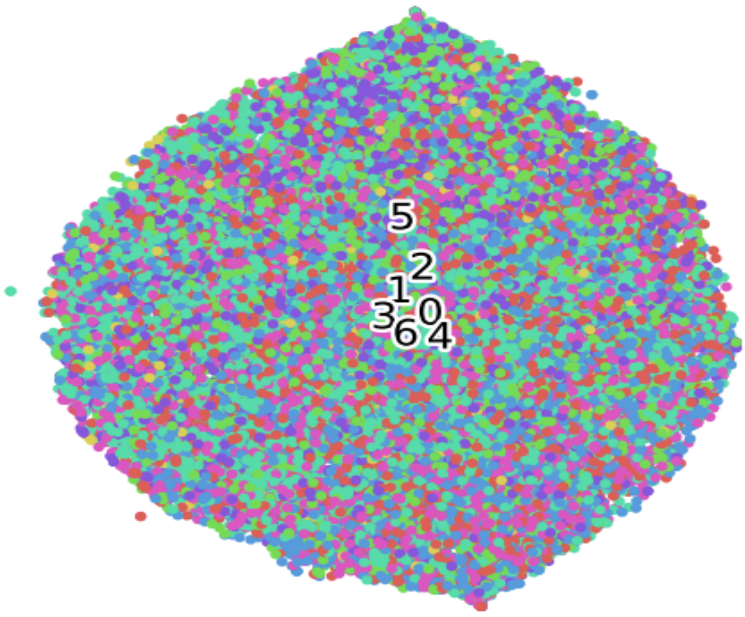
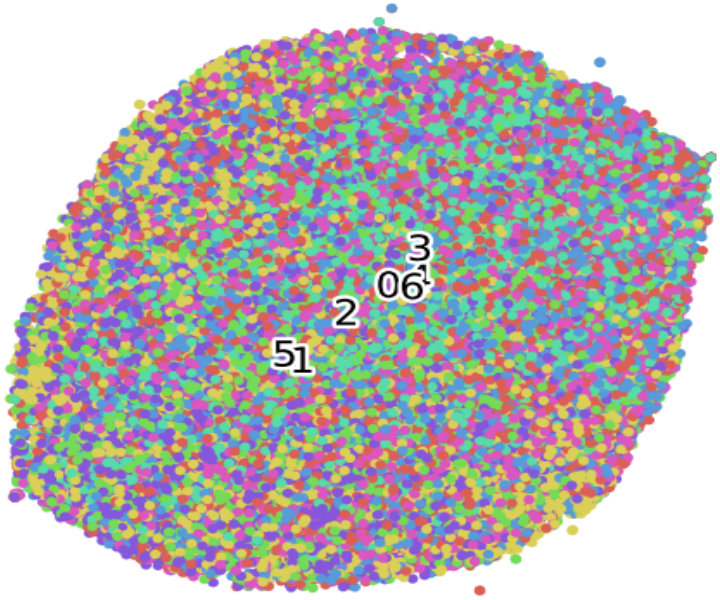
FER Dataset



Data points per class



Training data before Data Augmentation



Training data after Data Augmentation

Useful things to note in FER Dataset:

- FER Dataset has grayscale images.
- Dataset is imbalanced. Class 1 has approx 400 data points.
- So, we augment the data by using width-wise shift,height-wise shift and zoom.
- So, we get the augmented dataset as 42000 with 6000 data points per class.
- Since some data points are not images so we delete those images from training set.
- Since images are good in quality. So, no preprocessing needed on image.
- Since after augmentation, the distribution of data points are not changing from previous one on 2 Dimensional space. So, we can say that augmentation does not change the distribution of original data set.

Logistic Regression

With PCA with 30 components (Using Intensity Value)

Standard deviation : 0.0033948406900262008
Mean of accuracy : 0.244047619047619
Accuracy for each validation :[0.24214285714285713, 0.24583333333333332, 0.24166666666666667, 0.24988095238095237, 0.24071428571428571]
Test Accuracy by best model in cross validation : 0.27793257174700475

With PCA with 30 components (Using HOG Value)

Standard deviation : 0.002275522914661031
Mean of accuracy : 0.2585952380952381

Accuracy for each validation : [0.26, 0.256547619047619, 0.256547619047619, 0.2623809523809524, 0.2575]
Test Accuracy by best model in cross validation : 0.36625801058790747

With PCA with 30 components (Using LBP Value)
Standard deviation : 0.0041758402415988174
Mean of accuracy : 0.2061904761904762
Accuracy for each validation : [0.2038095238095238, 0.21333333333333335, 0.2007142857142857, 0.2069047619047619, 0.2061904761904762]
Test Accuracy by best model in cross validation : 0.25828921705210367

With PCA with 30 components (Using HOG + LBP Value)
Standard deviation : 0.002439285423783941
Mean of accuracy : 0.2667380952380952
Accuracy for each validation : [0.2708333333333333, 0.2679761904761905, 0.2646428571428571, 0.26607142857142857, 0.26416666666666666]
Test Accuracy by best model in cross validation : 0.3689049874616885

Using Landmarks (Naive Bayes)
TRAIN: [4008 4009 4010 ... 20033 20034 20035] TEST: [0 1 2 ... 4005 4006 4007]
Accuracy 0.26397205588822353
TRAIN: [0 1 2 ... 20033 20034 20035] TEST: [4008 4009 4010 ... 8012 8013 8014]
Accuracy 0.28375343149488397
TRAIN: [0 1 2 ... 20033 20034 20035] TEST: [8015 8016 8017 ... 12019 12020 12021]
Accuracy 0.2707761417519341
TRAIN: [0 1 2 ... 20033 20034 20035] TEST: [12022 12023 12024 ... 16026 16027 16028]
Accuracy 0.2665335662590467
TRAIN: [0 1 2 ... 16026 16027 16028] TEST: [16029 16030 16031 ... 20033 20034 20035]
Accuracy 0.281756925380584

Using Landmarks (Logistic Regression)
TRAIN: [4008 4009 4010 ... 20033 20034 20035] TEST: [0 1 2 ... 4005 4006 4007]
Accuracy 0.532185628742515
TRAIN: [0 1 2 ... 20033 20034 20035] TEST: [4008 4009 4010 ... 8012 8013 8014]
Accuracy 0.5393062141252808
TRAIN: [0 1 2 ... 20033 20034 20035] TEST: [8015 8016 8017 ... 12019 12020 12021]
Accuracy 0.5208385325680059
TRAIN: [0 1 2 ... 20033 20034 20035] TEST: [12022 12023 12024 ... 16026 16027 16028]
Accuracy 0.5507861242825056
TRAIN: [0 1 2 ... 16026 16027 16028] TEST: [16029 16030 16031 ... 20033 20034 20035]
Accuracy 0.5440479161467432

Logistic Regression:

Without PCA (Using Alexnet features)(Logistic Regression)
Standard deviation : 0.0045072378175290626
Mean of accuracy : 0.3267142857142857
Accuracy for each validation : [0.3317857142857143, 0.3317857142857143, 0.32535714285714284, 0.32011904761904764, 0.32452380952380955]
Test Accuracy by best model in cross validation : 0.36653663973251605

Without PCA (Using VGG16 Features)
Standard deviation : 0.002858729718057837
Mean of accuracy : 0.3246904761904762
Accuracy for each validation : [0.32880952380952383, 0.32595238095238094, 0.32416666666666666, 0.32, 0.32452380952380955]
Test Accuracy by best model in cross validation : 0.35775982167734743

With PCA n_comp=1000 (Using Alexnet features)(Logistic Regression) (without using last 3 layers of sequential layers)
Standard deviation : 0.00480721
Mean of accuracy : 0.39671814277
Accuracy for each validation : [0.3917814283, 0.38857142857143, 0.3953154284, 0.39611904761904764, 0.39452380952380955]
Test Accuracy by best model in cross validation : 0.4305

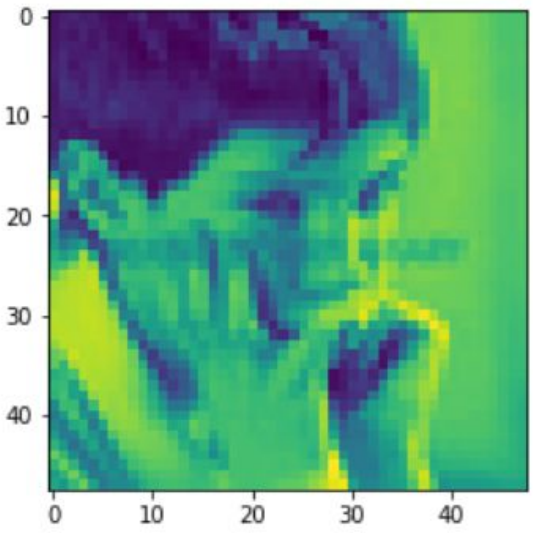
With PCA n_comp=1000 (Using VGG16 Features) (without using last 3 layers of sequential layers)
Standard deviation : 0.007021751110807875
Mean of accuracy : 0.44592857142857134
Accuracy for each validation : [0.44892857142857145, 0.4461904761904762, 0.4517857142857143, 0.43238095238095237, 0.45035714285714284]
Test Accuracy by best model in cross validation : 0.4827249930342714

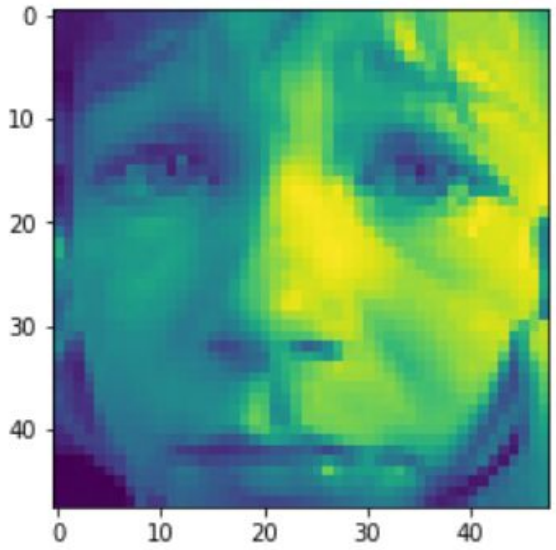
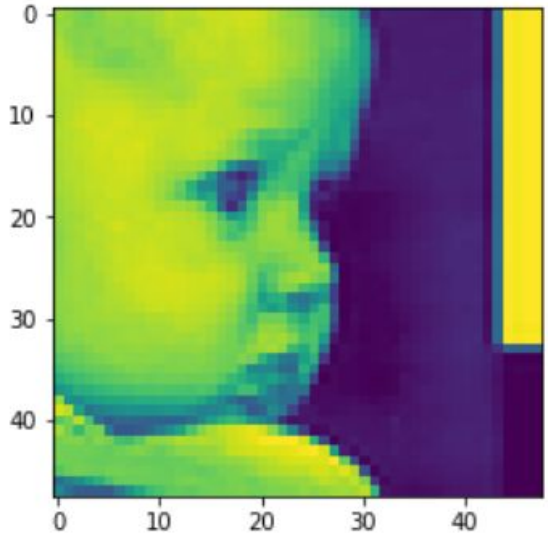
With LDA (Using VGG16 Features) (without using last 3 layers of sequential layers)
Standard deviation : 0.00371000974872046
Mean of accuracy : 0.6148809523809524
Accuracy for each validation : [0.6176190476190476, 0.6088095238095238, 0.6152380952380953, 0.6132142857142857, 0.6195238095238095]
Test Accuracy by best model in cross validation : 0.47757035385901364
Note:** It is overfitting since when we do LDA on train data it tried to reduce dimension and classification as according to train data and when we use that transformation onto test set we got less accuracy than cross validation accuracy.

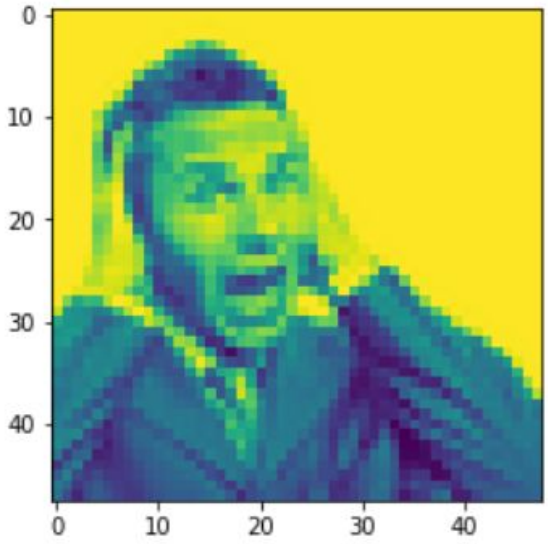
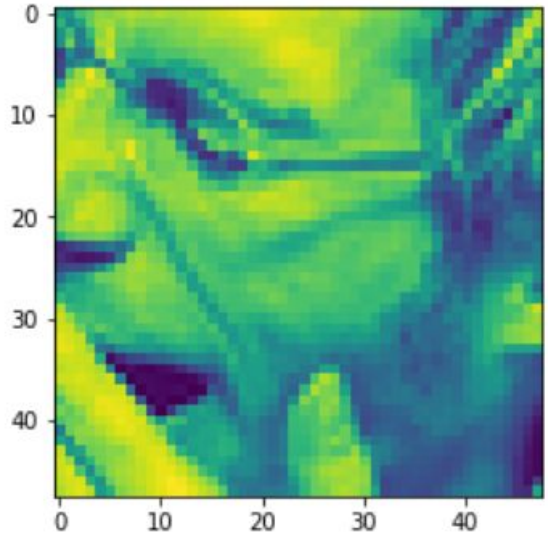
With PCA n_comp=1000 (Using VGG16 + Alexnet Features) (without using last 3 layers of sequential layers)

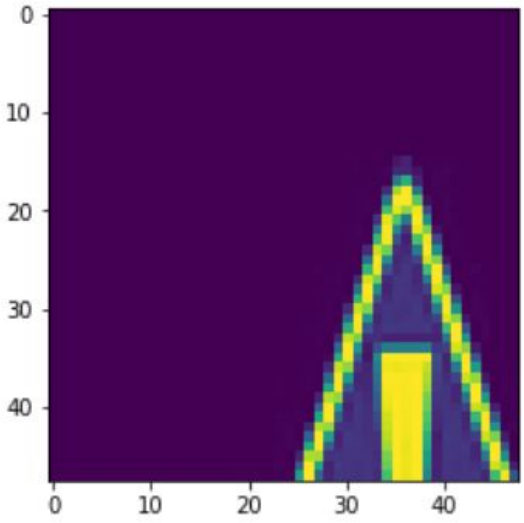
Standard deviation : 0.0038902168124762123
Mean of accuracy : 0.38399999999999995
Accuracy for each validation : [0.38166666666666665, 0.379047619047619, 0.3823809523809524, 0.3871428571428571, 0.38976190476190475]
Test Accuracy by best model in cross validation : 0.4027249930342714

Images in the training set whose landmarks could not be extracted:

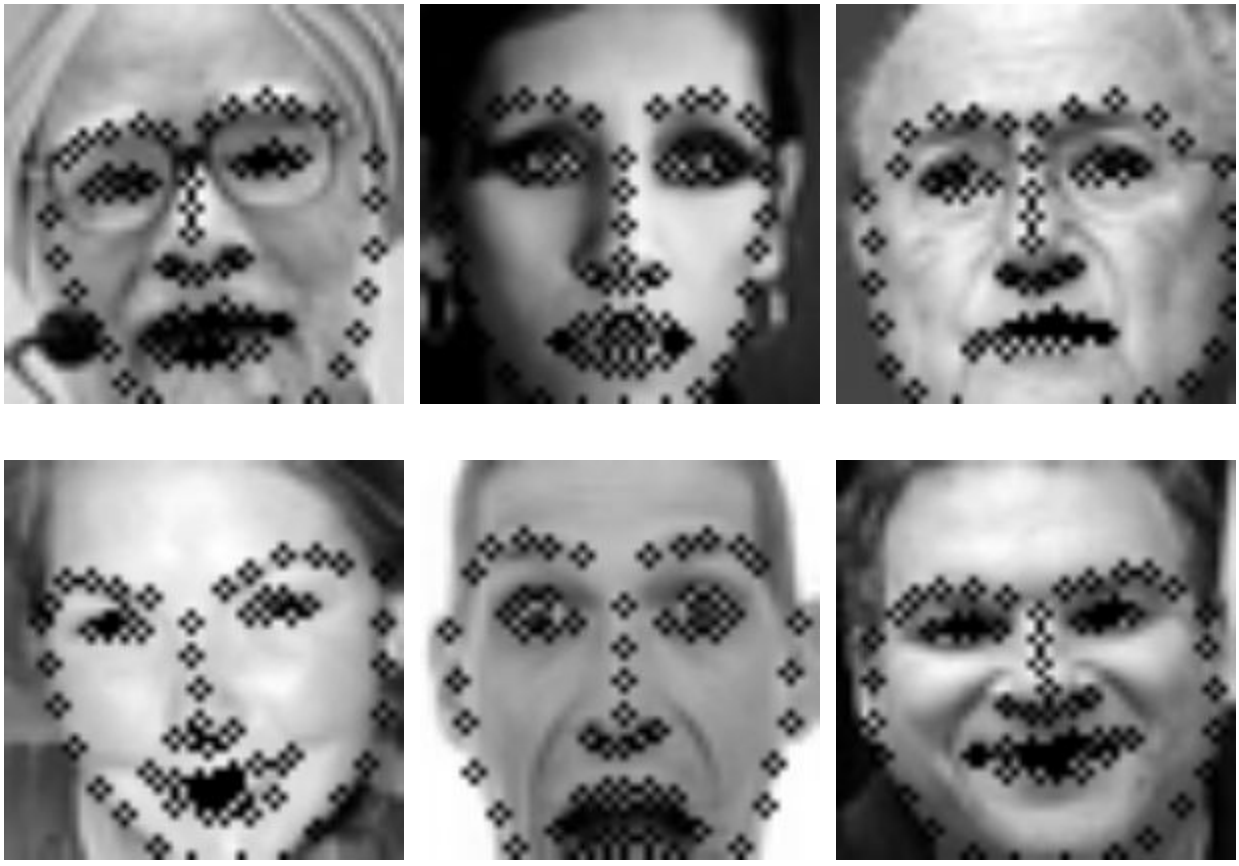
Image #	Image
4	

7	 0 10 20 30 40 0 10 20 30 40
13	 0 10 20 30 40 0 10 20 30 40

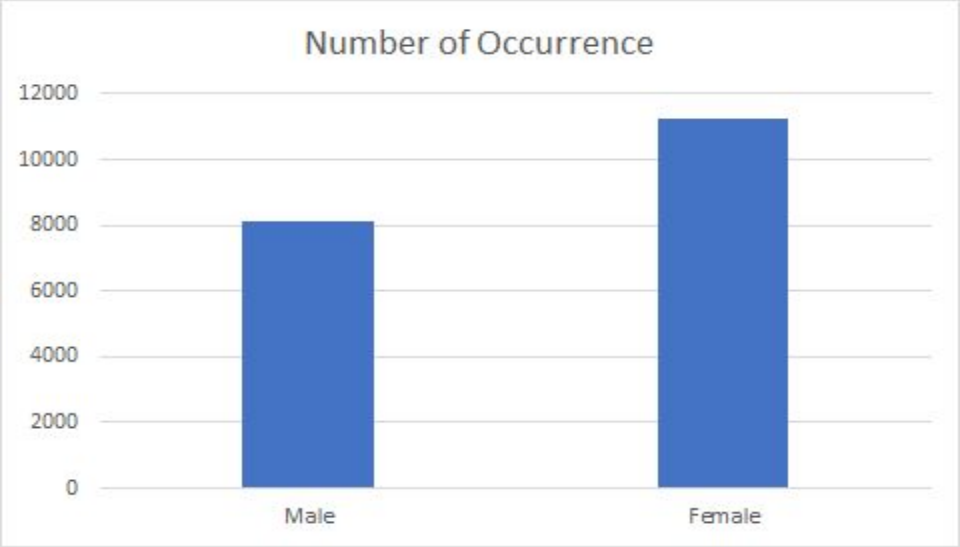
18	
19	

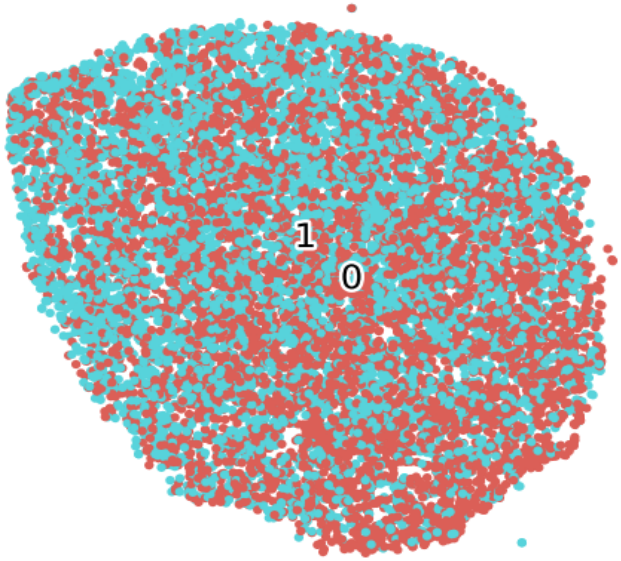
60	

Landmark detection on FER dataset:

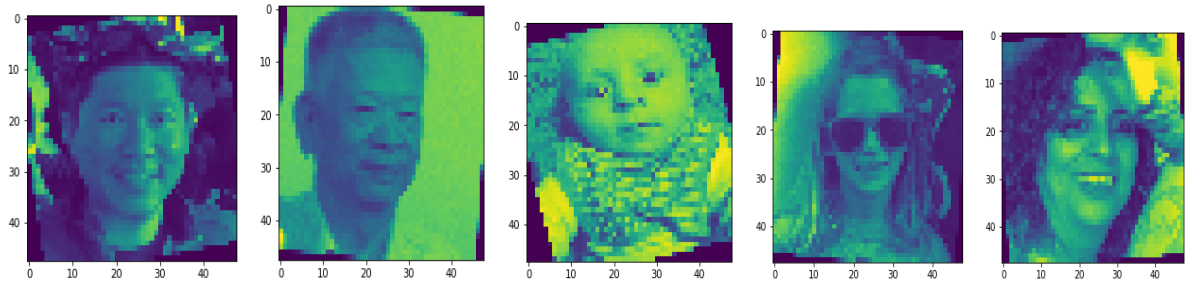


Gender Dataset





Sample data point:



M **M** **M** **F** **F**

Useful things to note in Gender Dataset:

- Gender Dataset has RGB images.
- Dataset is imbalanced. Class 1(Male) has approx 8000 and Class 2(Female) has approx 11000 data points.
- So, we random sample data and take 8000 from each of them.
- So, we get the dataset as 16000 with 8000 data points per class.

Logistic Regression:

With PCA with 30 components (Using HOG Value)

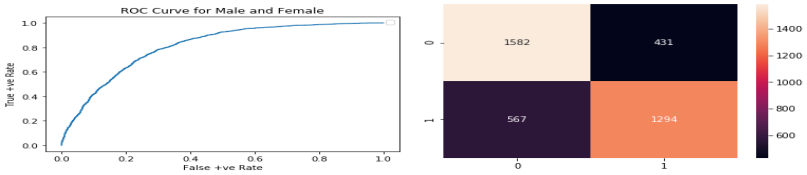
Standard deviation : 0.012408096933824959
Mean of accuracy : 0.7180564802381622
Accuracy for each validation : [0.7148387096774194, 0.7299128751210068, 0.7053888351080994, 0.7050661503710874, 0.7350758309131978]
Test Accuracy by best model in cross validation : 0.7289623128549303
Macro Precision 0.7291307604571524
Micro Precision 0.7289623128549303
Macro Recall 0.7254779777978682
Micro Recall 0.7289623128549303
Macro F1 measure 0.7261584794938971
Micro F1 measure 0.7289623128549304
Precision 0.7289623128549303
Recall 0.7289623128549303
F1 measure 0.7289623128549304

With PCA with 30 components (Using LBP Value)

Standard deviation : 0.005658661501793306
Mean of accuracy : 0.6346799487868096
Accuracy for each validation : [0.6341935483870967, 0.6324620845434011, 0.632784769280413, 0.6453694740238787, 0.6285898676992578]
Test Accuracy by best model in cross validation : 0.6484254001032524
Macro Precision 0.6506403606546523
Micro Precision 0.6484254001032524

Macro Recall 0.6456677893245057
Micro Recall 0.6484254001032524
Macro F1 measure 0.6442841009112561
Micro F1 measure 0.6484254001032524
Precision 0.6484254001032524
Recall 0.6484254001032524
F1 measure 0.6484254001032524

With PCA with 30 + 30 components (Using HOG + LBP Value)
Standard deviation : 0.007945775773508761
Mean of accuracy : 0.7328990829507959
Accuracy for each validation : [0.7287096774193549, 0.7386253630203291, 0.7195869635366248, 0.7357212003872217, 0.7418522103904486]
Test Accuracy by best model in cross validation : 0.7423851316468766
Macro Precision 0.7431506396638814
Micro Precision 0.7423851316468766
Macro Recall 0.7406083989799779
Micro Recall 0.7423851316468766
Macro F1 measure 0.7409534596204524
Micro F1 measure 0.7423851316468766
Precision 0.7423851316468766
Recall 0.7423851316468766
F1 measure 0.7423851316468766



Without PCA (Using VGG16 Features)
Standard deviation : 0.015154601838502141
Mean of accuracy : 0.6693334582435542
Accuracy for each validation : [0.6780645161290323, 0.6453694740238787, 0.6582768635043562, 0.686350435624395, 0.6786060019361084]
Test Accuracy by best model in cross validation : 0.6969540526587507

Without PCA (Using VGG16 Features) w/o last 3 layers
Standard deviation : 0.0062622901140199595
Mean of accuracy : 0.7640683050723960
Accuracy for each validation : [0.7616129032258064, 0.7534688609228783, 0.7647628267182962, 0.7708938367215231, 0.7696030977734754]
Test Accuracy by best model in cross validation : 0.7570986060918947
Macro Precision 0.7570535431655587

Micro Precision 0.7570986060918947
Macro Recall 0.7569743919501105
Micro Recall 0.7570986060918947
Macro F1 measure 0.7570026077281835
Micro F1 measure 0.7570986060918947

With PCA ncomp=1000 (Using Alexnet Features)
Standard deviation : 0.0075391619056626455
Mean of accuracy : 0.7615512183951119
Accuracy for each validation : [0.7638709677419355, 0.749919328815747, 0.7586318167150694, 0.7621813488222007, 0.7731526298806066]
Test Accuracy by best model in cross validation : 0.761
Macro Precision 0.77
Micro Precision 0.77
Macro Recall 0.781
Micro Recall 0.78
Macro F1 measure 0.7923
Micro F1 measure 0.7919

With LDA (Using Alexnet Features)
Standard deviation : 0.0025254973069680055
Mean of accuracy : 0.8734514151287096
Accuracy for each validation : [0.8703225806451613, 0.8712487899322362, 0.8735075830913198, 0.877379799935463, 0.8747983220393676]
Test Accuracy by best model in cross validation : 0.7312854930304594
Macro Precision 0.7321775795846799
Micro Precision 0.7312854930304594
Macro Recall 0.7306918566955869
Micro Recall 0.7312854930304594
Macro F1 measure 0.7306443861955745
Micro F1 measure 0.7312854930304594

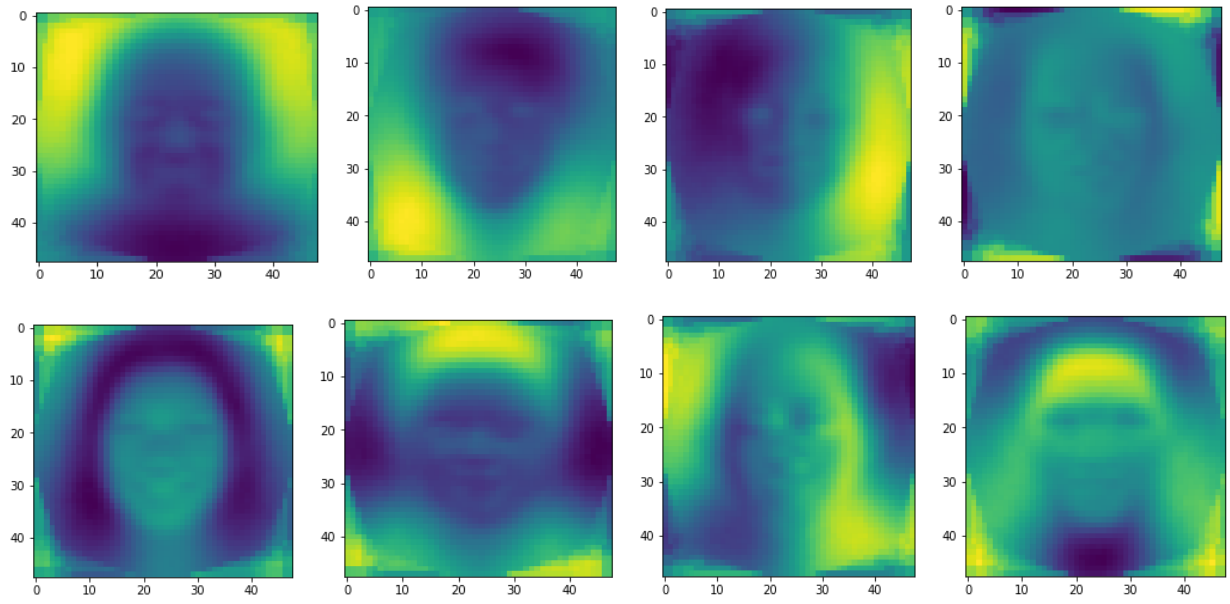
Note:** It is overfitting since when we do LDA on train data it tried to reduce dimension and classification as according to train data and when we use that transformation onto test set we got less accuracy than cross validation accuracy.

With PCA ncomp=1000 (Using Alexnet + VGG16 Features)
Standard deviation : 0.006674107406229556
Mean of accuracy : 0.7566468683966732
Accuracy for each validation : [0.756774193548387, 0.7679896740884156, 0.7525008067118425, 0.7479832203936754, 0.7579864472410455]
Test Accuracy by best model in cross validation : 0.7558079504388229
Macro Precision 0.756190689605011
Micro Precision 0.7558079504388229
Macro Recall 0.7554148111654254

Micro Recall 0.7558079504388229
Macro F1 measure 0.755479422075898
Micro F1 measure 0.7558079504388229

With PCA ncomp=500 (Using Alexnet) + HOG (30)+LBP(30)
Standard deviation : 0.007929958213612588
Mean of accuracy : 0.7567120298951795
Accuracy for each validation : [0.7470967741935484, 0.7560503388189739, 0.7515327525008068, 0.7583091319780574, 0.77057115198451111]
Test Accuracy by best model in cross validation : 0.7506453278265359
Macro Precision 0.7520988339452916
Micro Precision 0.7506453278265359
Macro Recall 0.7499616103191462
Micro Recall 0.7506453278265359
Macro F1 measure 0.7498963438449096
Micro F1 measure 0.750645327826536

Note* : Since after PCA using eigen faces, it is classifying better.
So, it is seen that they show good features of faces and that is why they are good for classifying gender.
Eigen Faces:



HOG + LBP after quantile transform (PCA n=30 each)
Standard deviation : 0.005497186369479125
Mean of accuracy : 0.6895332937784302

Accuracy for each validation : [0.6816129032258065, 0.6873184898354308, 0.6895772829945144, 0.6905453372055502, 0.6986124556308486]
Test Accuracy by best model in cross validation : 0.697986577181208
Macro Precision 0.6977580234177818
Micro Precision 0.697986577181208
Macro Recall 0.6965173784032345
Micro Recall 0.697986577181208
Macro F1 measure 0.6967235886203198
Micro F1 measure 0.697986577181208

HOG + Histogram Equalisation + quantile (PCA n=30 each)
Standard deviation : 0.010196252564463453
Mean of accuracy : 0.6854672995451185
Accuracy for each validation : [0.6841935483870968, 0.691513391416586, 0.6708615682478218, 0.6798967408841562, 0.7008712487899322]
Test Accuracy by best model in cross validation : 0.6768198244708312
Macro Precision 0.6757146928959592
Micro Precision 0.6768198244708312
Macro Recall 0.6752056316262979
Micro Recall 0.6768198244708312
Macro F1 measure 0.6753800469463459
Micro F1 measure 0.6768198244708312

Intuition for accuracy difference in Naive Bayes and Logistic Regression(Using Landmarks):
Since features selected for Landmarks are :
----> Naive Bayes approach assumes features to be independent of each other. However, the features (space coordinates of facial features) obtained through landmark detection are dependent on each other. Due to this dependency, Naive Bayes works poorly for this particular feature set.
----> In case of Logistic Regression, where there are no assumptions about features being independent, it performs relatively good on this particular feature set.

Description about feature set created from Landmark detection:

I. **X-Y coordinates:** These correspond to the position of landmarks expressed as numerical values depicting position in 2D space. The position of the face in the image does not make a difference due to the location invariance of landmarks. However, different ranges of numerical values, due to arbitrary positions of face, are not desired and therefore must be normalized. One approach is to calculate the position of all points relative to each other.
II. **Distance from centroid:** To calculate the relative positioning of points, we first calculate the centroid of the landmarks by finding the mean along both horizontal and vertical axis. Then distance of each landmark value from the centroid is computed and summed together to obtain a constant value.
III. **Rotation angle:** One more thing which needs to be taken into account is of face alignment. Faces in images could be tilted at arbitrary angles and hence need to be fixed for better

detection and classification. The image can be rotated about the nose tip and the rotation angle (arctangent) thereby calculated acts as a feature.

Note:**
Since (x,y) coordinate are dependent on each other as feature so using naive bayes gave not good result as compared to model like Logistic Regression.

Why use VGG16?

VGG16 is a deep neural network with 16 hidden layers. The architecture allows even moderately powerful systems to run such a network. Due to this advantage, we preferred the use of this model for extracting features from the training images. This was achieved by taking fc7 layer output which has dimension of 4096.

Note: Histogram Readings in Proposed Architecture:

Classifier	Accuracy	
	Intensity	Histogram
NaiveBayes	23.11	20.31
LogisticRegression	35.83	25.43
SVM (RBF Kernel)	30.58	22.32
Adaboost	22.15	16.15
Bagging	36.48	14.82

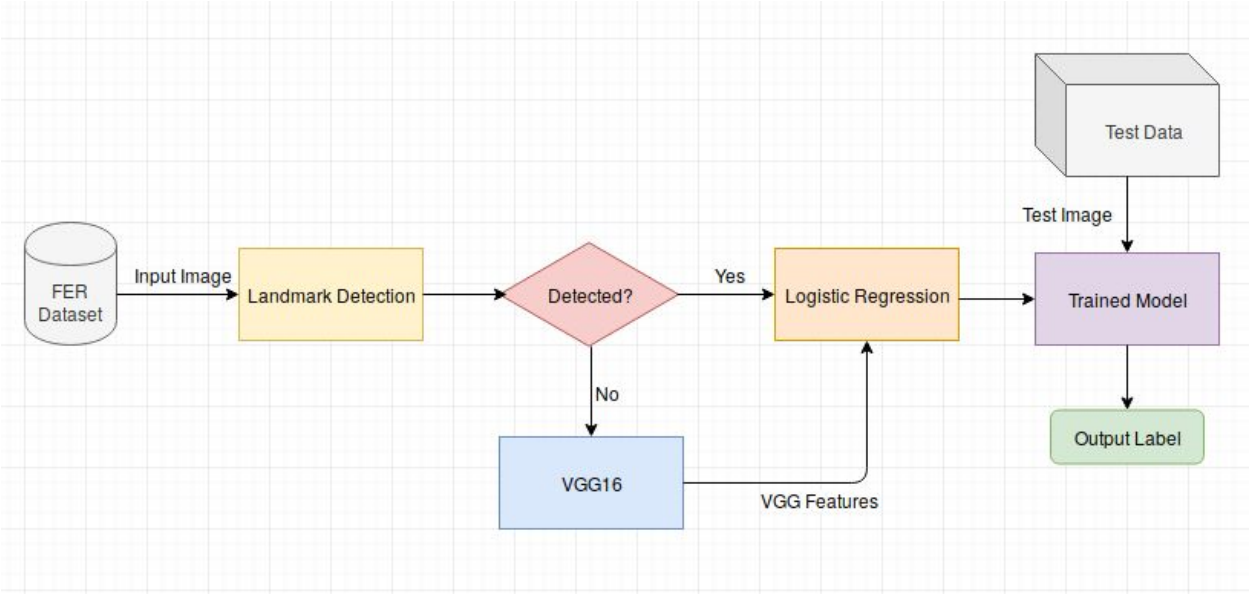
Table 1: Test Accuracy for FER dataset

Classifier	Accuracy	
	Intensity	Histogram
NaiveBayes	59.11	56.22
LogisticRegression	62.04	55.463
SVM (RBF Kernel)	56.49	46.21
Adaboost	42.65	43.76
Bagging	61.86	33.56

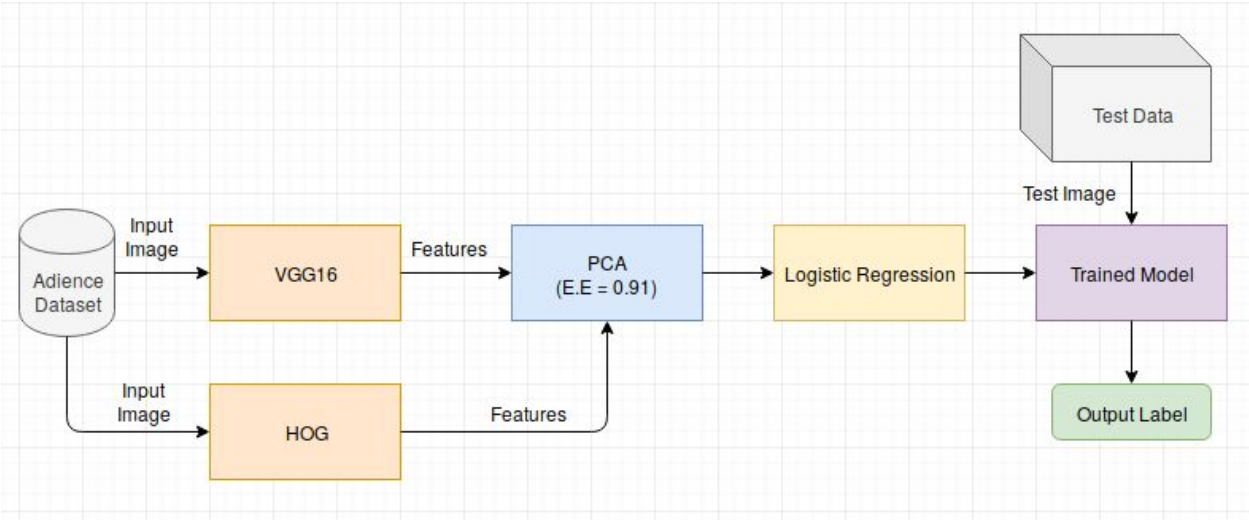
Table 2: Test Accuracy for Adience dataset

Proposed Algorithm:

For FER dataset:

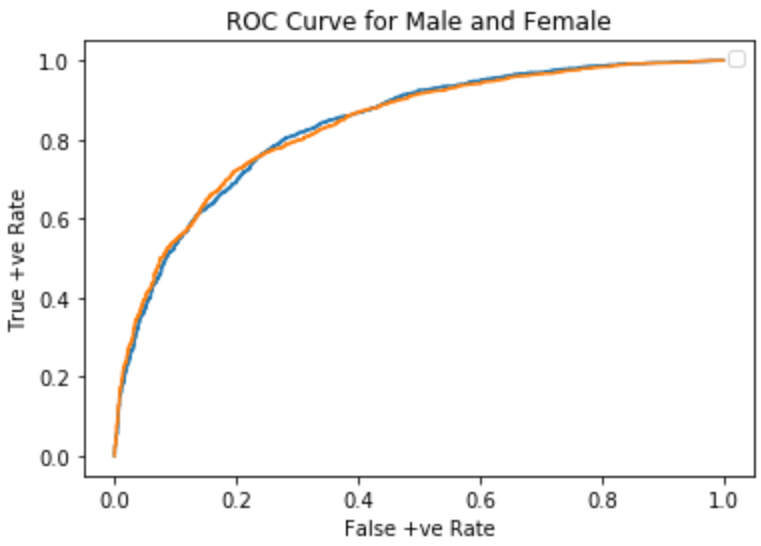


For Adience Data:

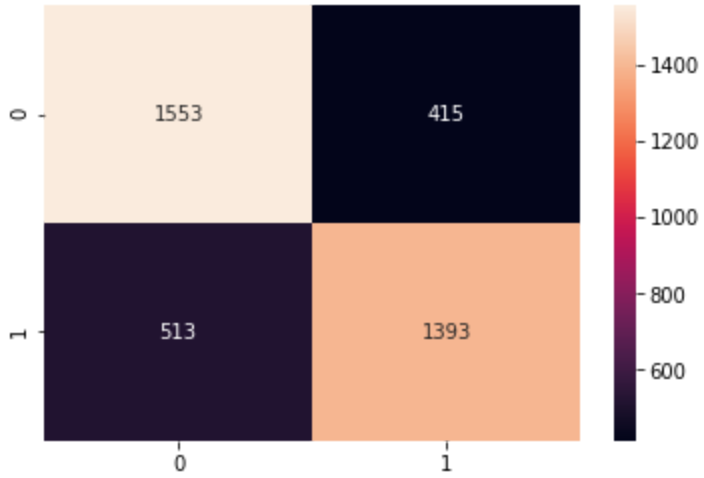


Adience Dataset:

ROC:



Adience Confusion Matrix:



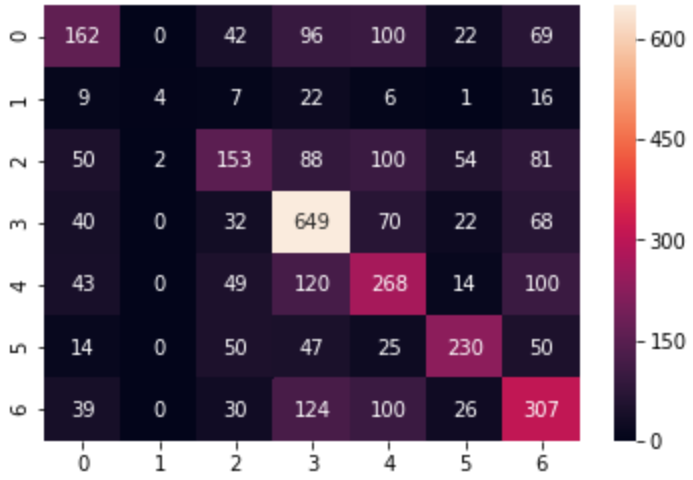
Inferences:

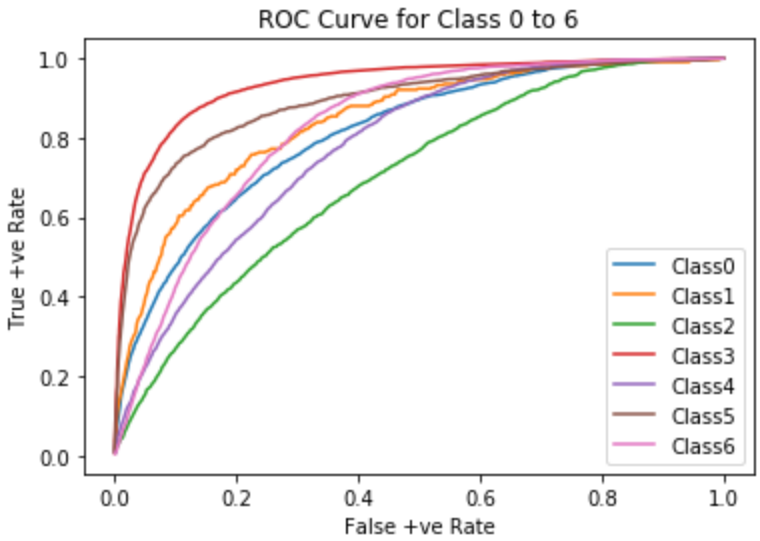
- **PCA worked better as compared to LDA -** It is because LDA caused overfitting on the training data as it tries to reduce dimensions and perform classification at the same time which did not seem to work well with the test data. This means classification performed at a lower dimension worked poorly for both datasets.
- **Facial Landmarks are useful features for capturing the emotional aspects of face as they boosted the classification accuracy considerably.**

- **Naive Bayes performed poorly on landmarks - this is due to the naive assumption of Naive Bayes which assumes features to be independent which is not the case in case of space coordinates.**

Conclusion:

We implemented two separate models for FER dataset and Adience dataset. Since images contained in FER dataset were decent frontal face portraits, facial landmarks could be extracted which helps perform better classification. However, Adience dataset, which contains 'images in the wild', landmarks extraction was not possible. Instead, we used VGG18 to extract image features combined with HOG features on which PCA was applied separately. On trying various models, we found Logistic Regression to be the best working model for both the datasets.




ROC Curve for Class 0 to 6
True +ve Rate
False +ve Rate
0.0
0.2
0.4
0.6
0.8
1.0
Class0
Class1
Class2
Class3
Class4
Class5
Class6