LBP:

- The limitation of the basic LBP operator is its small 3*3 neighborhood as it cannot capture necessary features with large scale.
- One is the number of points p in a circularly symmetric neighborhood and other one is the radius of the circle r, which allows to account for different scales.
- to be uniform if it has at most two 0-1 or 1-0 transitions
- For 'uniform', the result only includes patterns where all black dots are adjacent and all white dots are adjacent.
- For (8,1) in facial images about 90% are uniform, so it saves memory.
- Rotation of a textured input image cause the LBP patterns to translate in different location and to rotate about their origin.
- no of histogram: uniform > p + 1 for when image rotation are applied value doesn't changes, nri_uniform> p*(p-1)+3 when image rotation affect
- Here, p = 24, r = 8. No of histogram = 26
- As, radius 8 is good to capture necessary feature of large scale. So, 8*24 neighbor possible which increases feature vector size takes more time. So around 15% of that value is used. Value close to this marks give similar result.
- In GBP, center pixels used in LBP are replaced by the mean values of large neighborhood templates, and effects of noises are weakened.
- the resistance to uneven illumination by GBP is worse than that by LBP.

CNN:

- Before the CNN starts, the weights or filter values are randomized.
- The way the computer is able to adjust its filter values (or weights) is through a training process called back propagation.
- Stride controls shifting of conv filter, padding pads the input volume with zeros around the border to fit or drop the part of the image.
- After each conv layer, relu activation layer apply nonlinearity by just changes all the negative activations to 0.
- Pooling layer is down sampling layer.
- Dropout layer "drops out" a random set of activations in that layer by setting them to zero to overcome overfitting problem which means the weights of the network are so tuned to the training examples they are given that the network doesn't perform well when given new examples.
- To increase the stability of a neural network, batch normalization normalizes the output of a previous activation layer by subtracting the batch mean and dividing by the batch standard deviation.
- Alexnet has 5 convolution layer and 3 Fully connected layer

- It is inspired from Alexnet. Extra Convolution Layer is used to increasing the probability of getting better output.

```
model = Sequential()
input_shape = (48,48,1)
model.add(Conv2D(64, (5, 5), input_shape=input_shape,activation='relu', padding='same'))
model.add(Conv2D(64, (5, 5), activation='relu', padding='same'))
model.add(BatchNormalization())
model.add(MaxPooling2D(pool_size=(2, 2)))

model.add(Conv2D(128, (5, 5),activation='relu',padding='same'))
model.add(Conv2D(128, (5, 5),activation='relu',padding='same'))
model.add(BatchNormalization())
model.add(MaxPooling2D(pool_size=(2, 2)))

model.add(Conv2D(256, (3, 3),activation='relu',padding='same'))
model.add(Conv2D(256, (3, 3),activation='relu',padding='same'))
model.add(BatchNormalization())
model.add(MaxPooling2D(pool_size=(2, 2)))

model.add(Flatten())
model.add(Dense(128, name="dense_one"))
model.add(BatchNormalization())
model.add(Activation('relu'))
model.add(Dropout(0.2))
model.add(Dense(7))
model.add(Activation('softmax'))
model.compile(loss='categorical_crossentropy', metrics=['accuracy'],optimizer='adam')
model.summary()
```

SVM:

- The kernel function is that which is applied on each data instance to map the original non-linear observations into a higher-dimensional space in which they become separable.
- C: It is penalty parameter of the error term. It also controls the tradeoff between smooth decision boundary and classifying the training points correctly. (in SVC default 1.0)
- Gamma: It is kernel coefficient for rbf, poly and sigmoid. Higher the value of gamma, SVM will try to exact fit the as per training data set and cause over-fitting problem. (in SVC default 1 / n_features)
- max iter: Number of maximum iteration will be used to find the solution. (in SVC default -1, in LinearSVC default 1000)
- random_state : The seed of the pseudo random number generator used when shuffling the data for probability estimates.
- LinearSVC is actually minimizing squared hinge loss, instead of just hinge loss. (C=100 used)
- Grid Search is done to find the best parameter for performing SVM.
- Data was split in 10 and 2% of data was used for testing while grid searching.
- For SVC max iter=10000, random_state=0, decision function = ovr used.

Classification Report:

- The F1 score is a weighted harmonic mean of precision and recall for which the best score is 1.0 and the worst is 0.0, So, f1 score = (2*precision * recall)/(precision + recall)
- Support is the number of actual occurrences of the class in the specified dataset.

- The precision for a class is the number of true positives (i.e. the number of items correctly labeled as belonging to the positive class) divided by the total number of elements labeled as belonging to the positive class (i.e. the sum of true positives and false positives, which are items incorrectly labeled as belonging to the class). Precision = true positives / (true positives + false positives)
- Recall is defined as the number of true positives divided by the total number of elements that actually belong to the positive class (i.e. the sum of true positives and false negatives, which are items which were not labeled as belonging to the positive class but should have been). Recall = true positives / (true positives + false negatives)
- The other main visualization technique for showing the performance of a classification model is the Receiver Operating Characteristic (ROC) curve. the ROC curve shows how the recall vs precision relationship changes as we vary the threshold for identifying a positive in our model.
- Receiver operating characteristic (ROC) curve: plots the true positive rate (TPR) versus the false positive rate (FPR) as a function of the model's threshold for classifying a positive. An ROC curve plots the true positive rate (recall) on the y-axis versus the false positive rate (probability of a false alarm) on the x-axis. False Positive rate = False positives / (False positives + True Negatives)
- Confusion matrix: shows the actual and predicted labels from a classification problem
- Area under the curve (AUC): metric to calculate the overall performance of a classification model based on area under the ROC curve.