Question 1:

Classification Results:

Classifier/ Data Type	Original			Translated			Rotated X | Y | Z		
Metric	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Random Forest	0.3413	0.3472	0.3423	0.3549	0.3593	0.3559	0.3483 0.3407 0.3548	0.3532 0.3480 0.3606	0.3495 0.3417 0.3558
SVM	0.5038	0.5035	0.5039	0.5025	0.5024	0.5026	0.5041 0.5040 0.5039	0.5039 0.5038 0.5036	0.5042 0.5041 0.5040
Decision Tree	0.2645	0.2642	0.2654	0.2803	0.2768	0.2811	0.2640 0.2728 0.2702	0.2588 0.2679 0.2721	0.2646 0.2736 0.2710

Question 2:

For Original Datatype – SVM worked best

For Translated Original Datatype – SVM worked best

For Rotated Original Datatype – SVM worked best

- SVM works well for the given dataset because SVM can handle three-dimensional data, can predict the emotions more accuracy because the dataset doesn't have too many features and also the classes are well-separated.
- We see that Random Forest and Decision Tree gave similar result. This can be due to the fact that Random Forest and Decision Tree work by building a decision tree based on the feature selection.
- We can say that work best means that the model has higher accuracy in the given performance metrices.
- Accuracy is more useful in terms of evaluating the performance of a model. Because, it is the ratio of the number of correct predictions to the total number of predictions.
- Whereas precision gives the false positive predictions and recall gives the ratio of the number of true positive predictions to the total number of actual positive examples.
- Since true positives and false positives don't represent the whole predictions like accuracy does, we can say accuracy is the best metric.

Question 3:

For the SVM Classifier below are the misclassifications according to Datatypes based on average confusion matrix.

- **Original Datatype**: We see that the expression Angry is the mostly misclassified as Disgust.

- **Translated Datatype**: We see that the expression Angry is the mostly misclassified as Disgust.
- **Rotated Datatype**: We see that the expression Angry is the mostly misclassified as Disgust.

This is because of the fact that angry expressions can be very similar to those of disgust expressions. Hence the SVM model might have misclassified Angry as Disgust the most.
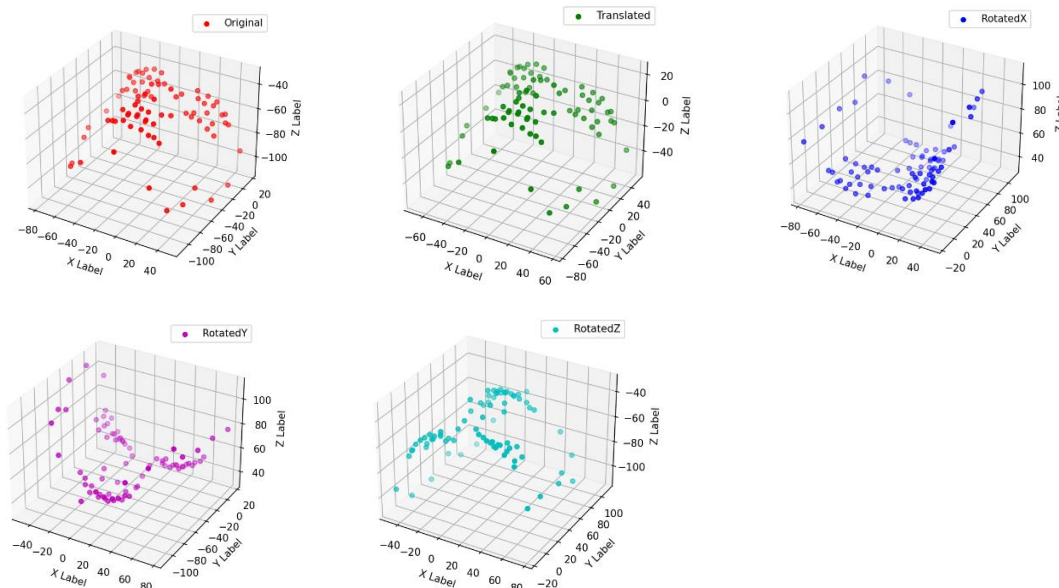
Question 4:

- Comparing the datatypes across the three classifiers, we see that we got similar results.
- Comparing the classifiers in each datatype, we see that SVM got good results.
- Random Forest got the second better results and decision tree being the least.
- We see a huge difference between the results of SVM when compared to Random Forest and Decision Tree. This can be because of the way SVM works. SVM works greatly for high dimensional data which can be linearly separable.
- Random Forest and Decision Tree are similar in a way that Random Forest combines the outputs of multiple decision trees to make a prediction of expression whereas decision trees work on splitting the attributes and builds a tree like model.

Hence we see that SVM works better on 3D data, while Random Forest and Decision Tree gives almost similar results.

Question 5:

Sample plots for each data type – Original, Translated, Rotated on X-axis, Rotated on Y-axis, Rotated on Z-axis respectively.

Question 6:

1) **SVM:**

- SVM is a supervised machine learning algorithm. It works by finding a hyperplane that separates different classes with maximum margin in between.
- Maximizing the margin is essential because this helps classify future datapoints with more confidence.
- The data points are called support vectors. If there are N number of features, the algorithm tries to find a hyperplane in N-Dimensional space.
- To explain in brief, this algorithm can handle both linearly separable and non-linearly separable classes by transforming the data given to a high dimensional space using kernel functions.
- In the higher dimensional space, the classifier finds a hyperplane which can separate the classes with maximum margin.
- There are several advantages for SVM classifier:
- Effective in high dimensional spaces.
- Performs well on data where the number of dimensions are greater than number of samples.

2) **Random Forest:**
- Random forest is a supervised machine learning model which works on the learning methods of ensemble. This learning model will
- combine the predictions of multiple models before obtaining a final prediction. This classifier is a collection of decision trees. Each
- decision tree contains a sample of dataset and a sample of the total features. The sampling is done independently and does not overlap between the models.
- Each decision tree in the model is trained to output the class label or the final prediction.
- The class label or the final prediction which had the majority of votes, i.e., which was labelled output for most of the decision trees is voted as the final output model.
- Advantages of Random Forest:
- Reduces over fitting and provides generalization.
- Provides feature importance scores, which can help us identify the most relevant features of the data.

3) **Decision Tree:**
- Decision Tree is a supervised classification algorithm which works by recursively splitting the input data based on the attributes or features.
- It builds a tree-like model to indicate the possible outcomes of each feature and gives the final prediction.
- For decision tree, there will be nodes and leaf. The attributes are represented as nodes and the data is split according to a selection criteria on the node.
- The final outcome is represented in the leaf. The goal of decision tree is to minimize the differences between each dataset so that one model of decision tree can be used to predict the outcomes with more accuracy.
- The algorithm uses methods such as Information gain, entropy to find out which are the most relevant features and which feature is to be used to build the decision tree.

- Advantages of Decision Trees:
- Ease of use, interpretability
- Ability to handle categorical and numerical data.
-

DETAILED RESULTS:

1. **Random Forest With Original Data**

   Precision: 0.34727069223126467

   Recall: 0.34230708555099587

   Accuracy: 0.34130295262400845

	Angry	Disgust	Fear	Happy	Sad	Surprise
Angry	404.0	222.3	159.0	147.2	333.2	81.6
Disgust	159.0	329.1	165.6	123.3	108.6	73.4
Fear	93.7	125.0	130.4	122.2	146.0	93.2
Happy	65.4	97.7	146.2	387.0	66.8	39.5
Sad	112.2	49.1	95.7	48.6	163.1	56.2
Surprise	178.1	193.9	307.5	169.0	197.5	650.9

2. **Random Forest With Translated Data**

   Precision: 0.3593464332219006

   Recall: 0.35597285990478056

   Accuracy: 0.35497739697476816

	Angry	Disgust	Fear	Happy	Sad	Surprise
Angry	428.2	213.1	159.7	132.8	351.1	81.1
Disgust	151.9	329.4	175.7	117.5	130.7	60.1
Fear	99.5	147.3	158.7	103.4	152.1	95.4
Happy	53.1	86.5	113.8	426.3	59.7	46.1
Sad	106.8	41.3	83.3	51.4	152.5	60.8
Surprise	172.9	199.5	313.1	165.9	168.1	651.3

3. **Random Forest With Rotated on X-axis Data**

   Precision: 0.3532893321016867

   Recall: 0.34954841150218563

   Accuracy: 0.3483882010904475

	Angry	Disgust	Fear	Happy	Sad	Surprise
Angry	401.5	227.4	173.3	134.9	351.3	70.2
Disgust	167.0	337.4	174.3	142.8	101.1	68.9

Fear	90.9	131.3	139.7	96.6	136.5	92.1
Happy	46.6	74.5	136.5	407.3	66.6	44.1
Sad	127.7	50.8	79.0	48.1	151.9	50.6
Surprise	178.7	195.7	301.6	167.6	179.8	668.9

4. Random Forest With Rotated on Y-axis Data

Precision: 0.34807736377338855

Recall: 0.3417092342344784

Accuracy: 0.34076109529584203

	Angry	Disgust	Fear	Happy	Sad	Surprise
Angry	391.9	225.1	166.1	141.8	338.0	77.5
Disgust	189.0	329.1	171.2	112.0	133.5	69.8
Fear	77.2	113.6	133.8	114.5	136.5	93.6
Happy	69.5	88.6	130.6	401.2	51.7	47.9
Sad	96.3	57.1	93.7	47.4	150.8	52.1
Surprise	188.5	203.6	309.0	180.4	196.7	653.9

5. Random Forest With Rotated on Z-axis Data

Precision: 0.36066443049986624

Recall: 0.3558126879003327

Accuracy: 0.3548233527022132

	Angry	Disgust	Fear	Happy	Sad	Surprise
Angry	416.1	202.6	160.7	113.6	343.6	89.6
Disgust	177.1	359.3	158.2	123.9	119.4	59.7
Fear	89.2	123.3	137.7	131.1	166.4	91.9
Happy	59.0	80.2	137.0	427.0	47.9	43.1
Sad	122.4	51.3	98.8	56.7	153.2	58.4
Surprise	148.6	200.4	312.0	145.0	183.7	652.1

6. Decision Tree With Original Data

Precision: 0.26429532065482186

Recall: 0.2654142912793082

Accuracy: 0.26454660157715787

	Angry	Disgust	Fear	Happy	Sad	Surprise
Angry	275.3	210.9	166.8	162.8	213.4	106.9

Disgust	184.9	192.9	162.1	167.0	122.1	113.4
Fear	136.5	181.8	175.5	134.9	177.5	153.4
Happy	101.6	170.7	148.2	298.9	106.6	68.4
Sad	171.6	113.1	117.7	84.9	188.0	83.9
Surprise	142.5	147.7	234.1	148.8	206.6	468.8

7. Decision Tree With Translated Data

Precision: 0.27681172693783196

Recall: 0.28117603822028947

Accuracy: 0.28032436324784754

	Angry	Disgust	Fear	Happy	Sad	Surprise
Angry	312.8	211.3	143.3	159.1	195.0	101.8
Disgust	145.3	202.5	161.2	140.4	128.4	98.5
Fear	172.7	189.1	193.8	142.8	180.5	156.7
Happy	88.6	166.4	153.0	294.1	103.4	76.3
Sad	174.9	110.6	130.0	70.3	211.6	81.9
Surprise	118.1	137.2	223.1	190.6	195.3	479.6

8. Decision Tree With Rotated on X-axis Data

Precision: 0.2588328933910288

Recall: 0.2646140310262492

Accuracy: 0.2640620794426615

	Angry	Disgust	Fear	Happy	Sad	Surprise
Angry	303.6	224.4	148.1	119.8	199.7	91.5
Disgust	153.4	172.5	147.1	142.1	155.8	123.2
Fear	144.1	204.6	191.3	151.9	168.7	173.2
Happy	109.7	163.1	159.0	292.7	118.5	66.6
Sad	168.4	108.8	140.2	110.2	182.4	86.3
Surprise	113.2	143.7	218.7	180.6	189.1	454.0

9. Decision Tree With Rotated on Y-axis Data

Precision: 0.2679199670542702

Recall: 0.27369530456890034

Accuracy: 0.2728766065930631

	Angry	Disgust	Fear	Happy	Sad	Surprise
Angry	283.7	213.2	141.8	162.9	216.5	109.0
Disgust	165.5	205.4	166.4	150.0	139.2	95.2
Fear	185.3	184.4	197.1	124.6	136.8	149.9
Happy	108.4	171.1	141.1	279.8	125.1	80.7
Sad	127.5	95.6	115.9	86.6	214.1	90.2
Surprise	142.0	147.4	242.1	193.4	182.5	469.8

10. Decision Tree With Rotated on Z-axis Data

Precision: 0.27216516935973945

Recall: 0.2710682104417925

Accuracy: 0.27023469711602727

	Angry	Disgust	Fear	Happy	Sad	Surprise
Angry	284.1	173.3	174.7	136.3	263.0	114.8
Disgust	147.0	214.5	171.4	169.0	117.6	89.6
Fear	168.1	194.1	169.1	128.3	175.0	169.0
Happy	87.2	118.0	118.8	326.0	102.6	74.0
Sad	181.7	109.2	131.8	65.4	180.6	88.7
Surprise	144.3	208.0	238.6	172.3	175.4	458.7

*** Best classifier is SVM for all Original, Translated, Rotatedx, Rotatedy, Rotatedz ***

11. SVM With Original Data

Precision: 0.5035975061226166

Recall: 0.5039422106872873

Accuracy: 0.5038479441257369

	Angry	Disgust	Fear	Happy	Sad	Surprise
Angry	546.0	100.2	48.9	37.6	55.7	32.5
Disgust	44.7	450.5	84.74	25.5	18.7	18.0
Fear	26.0	75.1	353.0	44.3	23.1	33.0
Happy	20.8	38.3	76.0	534.9	23.9	30.4
Sad	55.9	22.7	53.7	28.5	570.5	40.8
Surprise	19.0	30.3	88.4	23.5	22.3	540.1

12. SVM With Translated Data

Precision: 0.5024550623080682

Recall: 0.5026978444571537

Accuracy: 0.5025896139076137

	Angry	Disgust	Fear	Happy	Sad	Surprise
Angry	543.8	101.2	47.8	37.7	56.7	31.9
Disgust	43.4	447.8	800.6	26.2	18.6	18.4
Fear	25.6	72.9	353.9	44.2	26.8	32.8
Happy	18.7	38.8	79.0	534.9	20.9	29.9
Sad	60.4	25.7	57.0	25.8	567.3	42.1
Surprise	20.5	30.7	86.1	28.5	23.9	569.7

13. SVM With Rotated on X-axis Data

Precision: 0.5039213531611917

Recall: 0.5042790080174459

Accuracy: 0.5041790452876645

	Angry	Disgust	Fear	Happy	Sad	Surprise
Angry	545.5	100.0	49.1	38.0	55.3	31.5
Disgust	44.2	451.3	84.9	25.4	18.5	17.7
Fear	26.1	74.7	353.0	44.7	23.5	32.5
Happy	20.8	38.2	75.9	534.9	23.6	30.3
Sad	56.6	23.7	53.4	27.7	570.8	41.3
Surprise	19.2	29.2	88.1	26.6	22.5	541.5

14. SVM With Rotated on Y-axis Data

Precision: 0.5038632022002009

Recall: 0.5041978930103725

Accuracy: 0.504096266571365

	Angry	Disgust	Fear	Happy	Sad	Surprise
Angry	545.8	10.1	49.9	37.6	55.2	32.0
Disgust	44.0	451.3	84.9	25.9	18.6	17.9
Fear	25.4	74.6	353.1	44.6	24.1	32.5
Happy	21.3	37.4	76.1	535.2	24.0	30.1
Sad	56.4	23.8	52.4	27.7	569.9	41.1
Surprise	19.5	29.9	88.0	26.3	22.4	541.2

15. SVM With Rotated on Z-axis Data

Precision: 0.503665110776906

Recall: 0.5040085466917882

Accuracy: 0.5039141747726078

	Angry	Disgust	Fear	Happy	Sad	Surprise
Angry	545.7	100.4	49.3	37.6	55.2	32.2
Disgust	43.9	450.3	84.5	25.5	18.7	17.9
Fear	25.8	75.4	353.3	44.5	23.1	33.4
Happy	21.2	37..7	75.9	534.9	23.6	30.5
Sad	56.8	23.4	53.6	27.9	571.0	40.6
Surprise	19.0	29.9	87.8	26.9	22.6	540.2