# CSE 555 Project
Group DAGS

Supplementary Images for the project

We trained multiple ML and DL models on the six types of MNIST dataset (Original, Balanced, Asymmetric Balanced, Asymmetric Imbalanced, Symmetric Balanced, Symmetric Imbalanced) for the completion of our project.
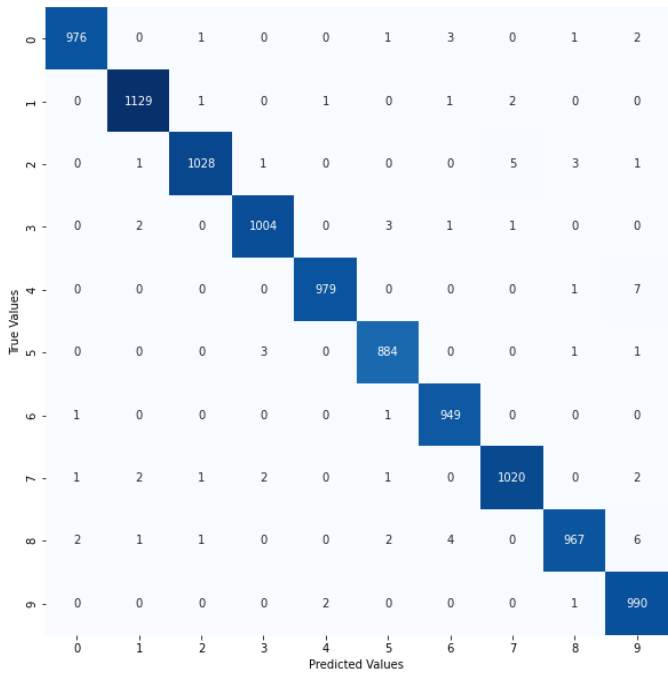
In this file we provide some images showing the performance of the various models. We have shown the confusion matrix along with the Training and Validation Loss graphs for the DL models and Precision vs Recall and ROC curve for the ML models.
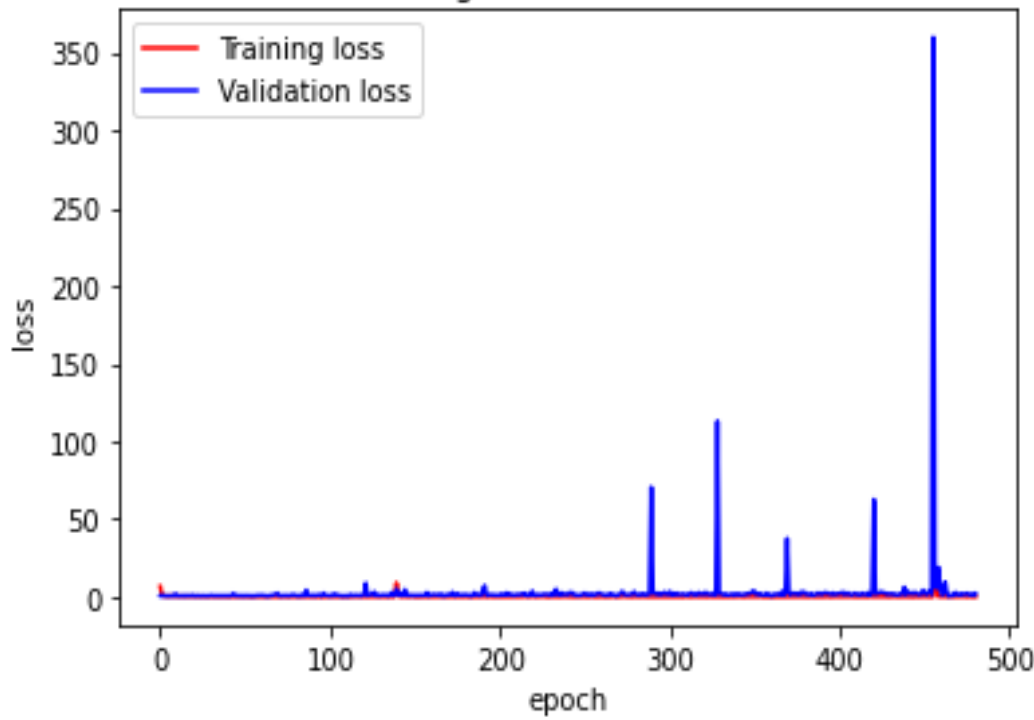
LDAM-DRW
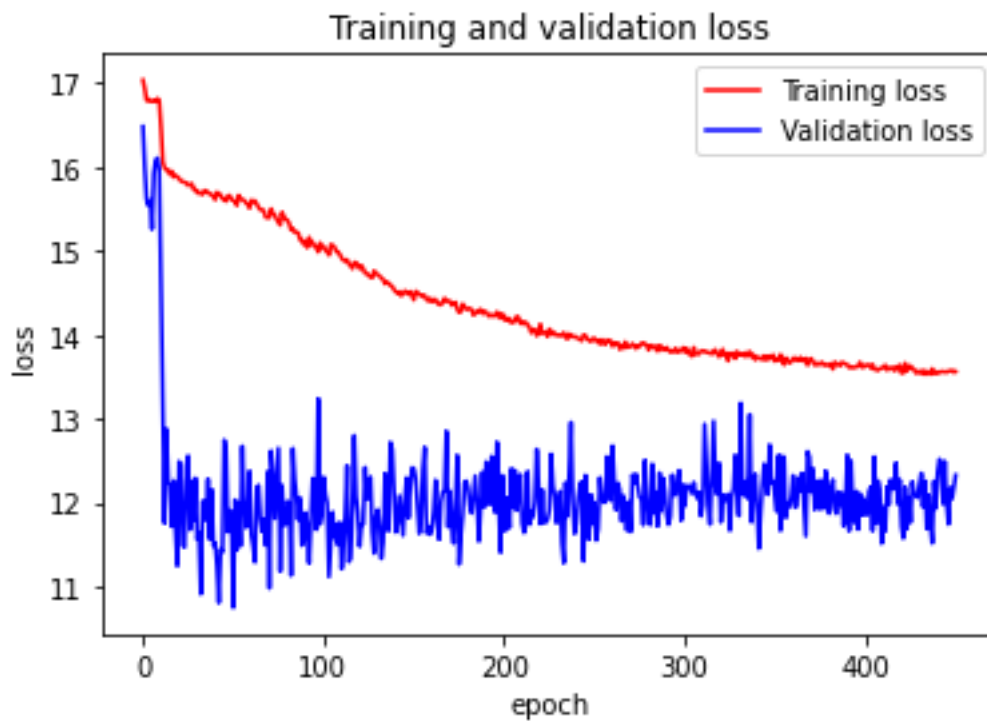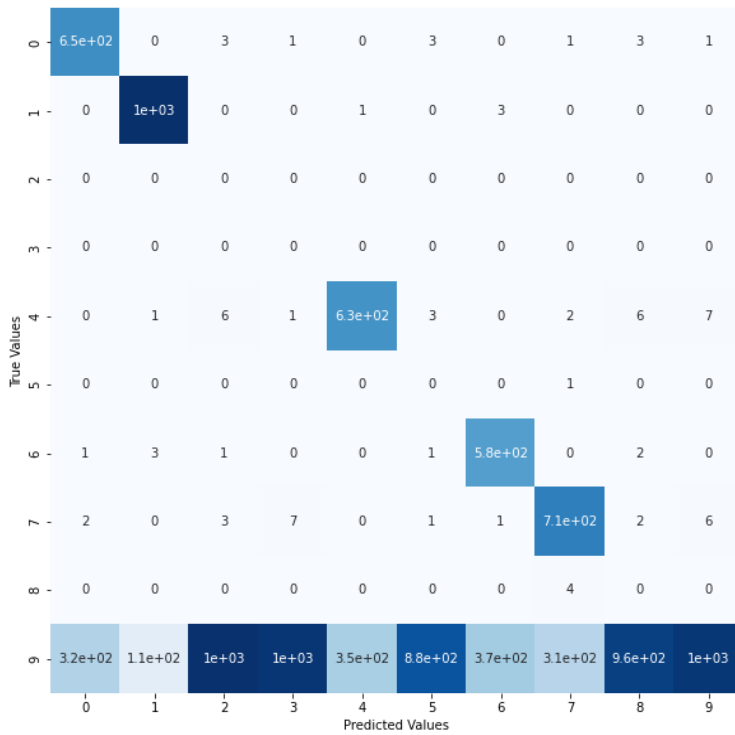
Original Balanced Dataset:
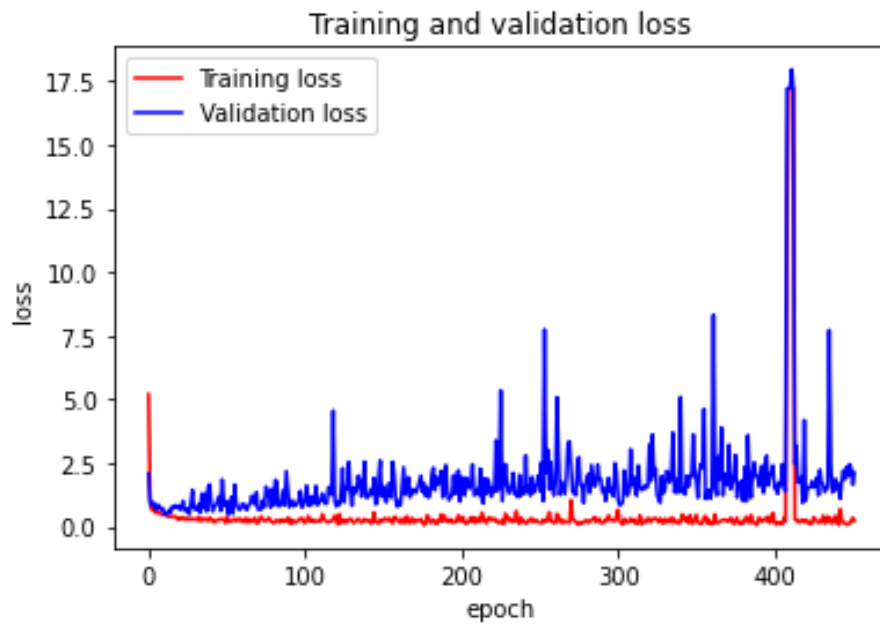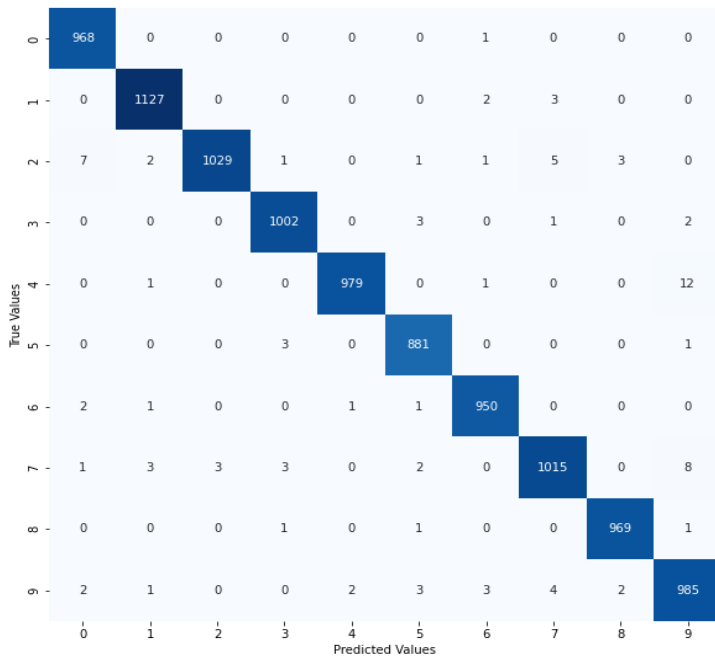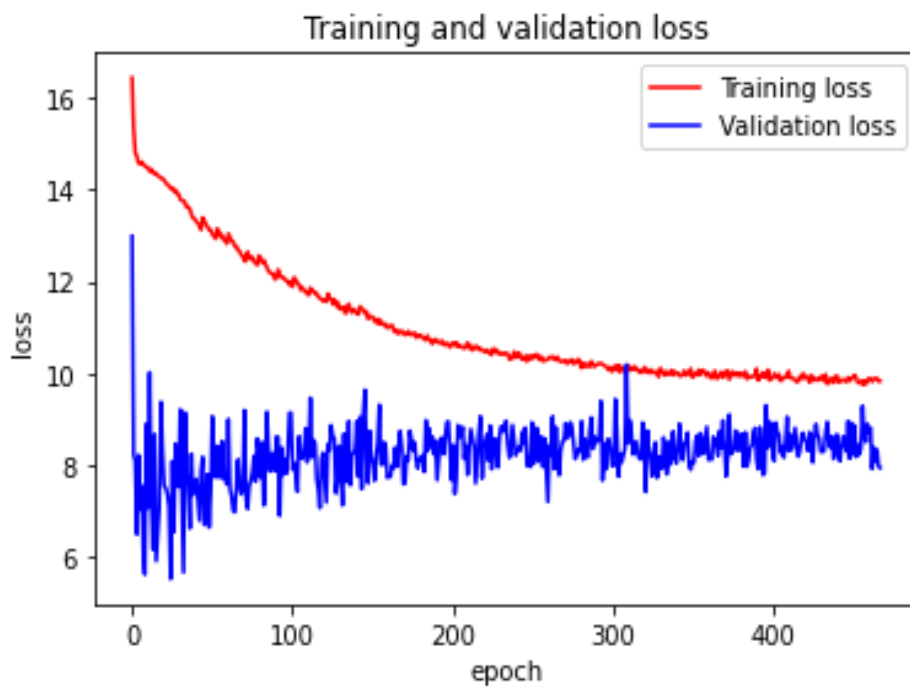
Original Imbalanced Dataset:

Training and validation loss

Balanced Symmetric Dataset:

Training and validation loss

Imbalanced Symmetric Dataset:

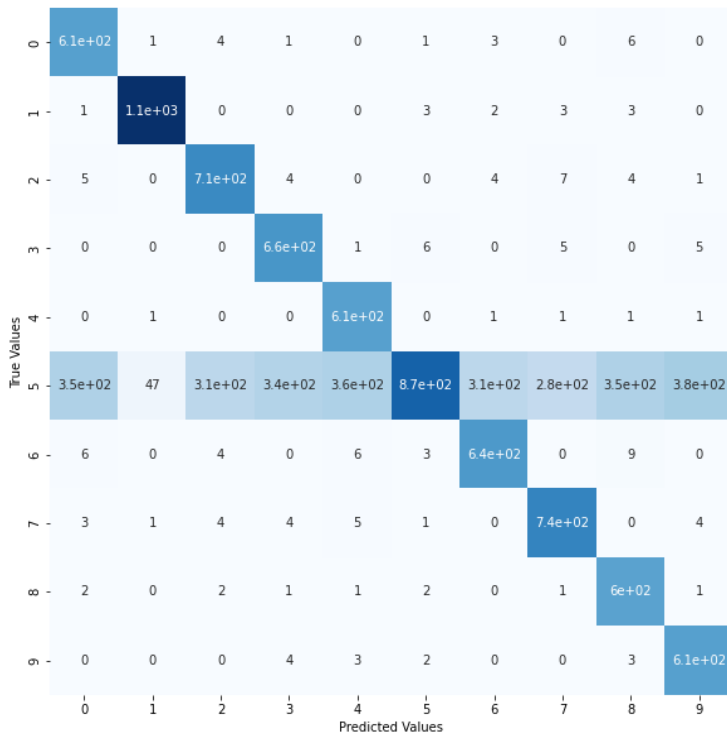|	0	1	2	3	4	5	6	7	8	9
0	6.1e+02	1	4	1	0	1	3	0	6	0
1	1	1.1e+03	0	0	0	3	2	3	3	0
2	5	0	7.1e+02	4	0	0	4	7	4	1
3	0	0	0	6.6e+02	1	6	0	5	0	5
4	0	1	0	0	6.1e+02	0	1	1	1	1
5	3.5e+02	47	3.1e+02	3.4e+02	3.6e+02	8.7e+02	3.1e+02	2.8e+02	3.5e+02	3.8e+02
6	6	0	4	0	6	3	6.4e+02	0	9	0
7	3	1	4	4	5	1	0	7.4e+02	0	4
8	2	0	2	1	1	2	0	1	6e+02	1
9	0	0	0	4	3	2	0	0	3	6.1e+02

True Values / Predicted Values

Training and validation loss

Balanced Asymmetric Dataset:

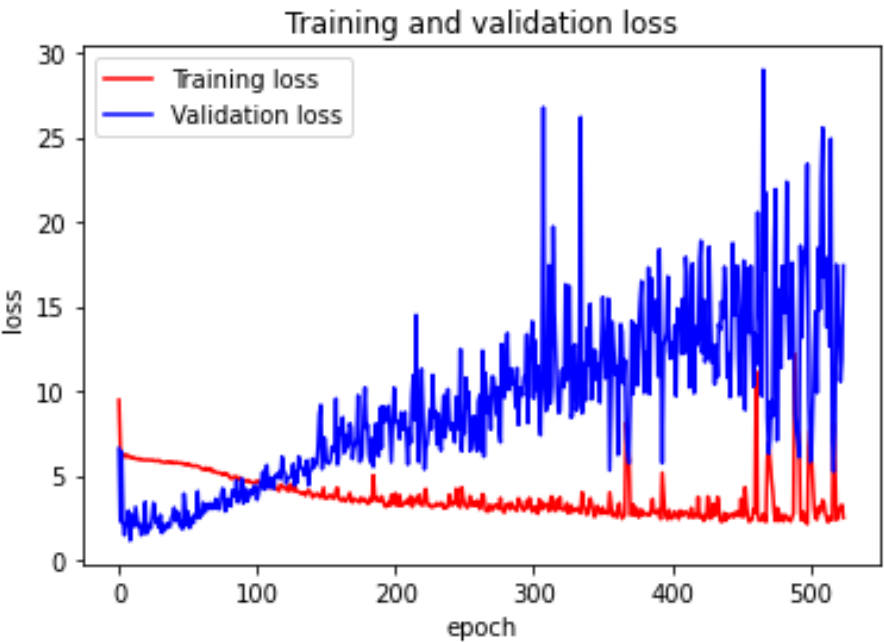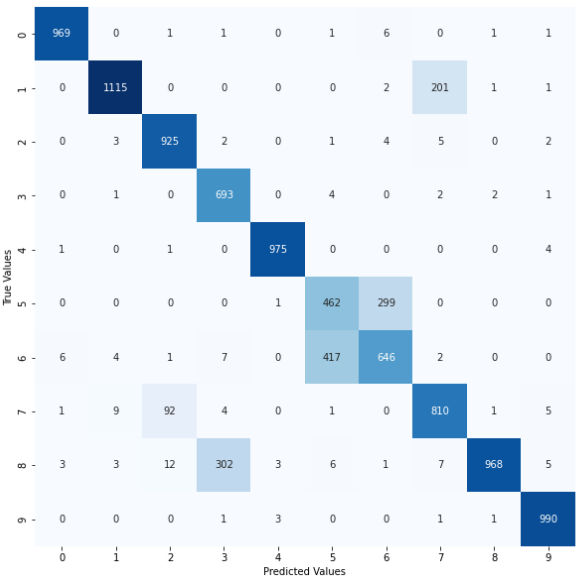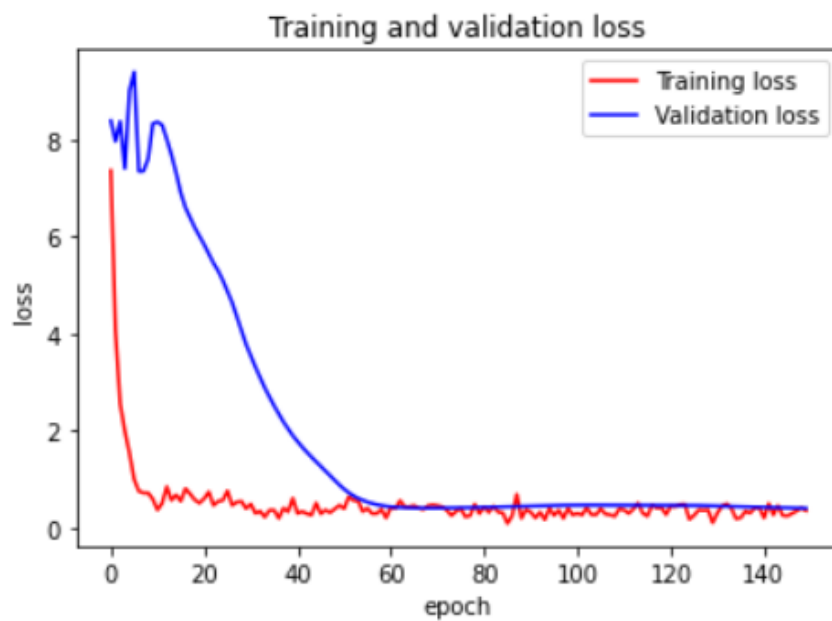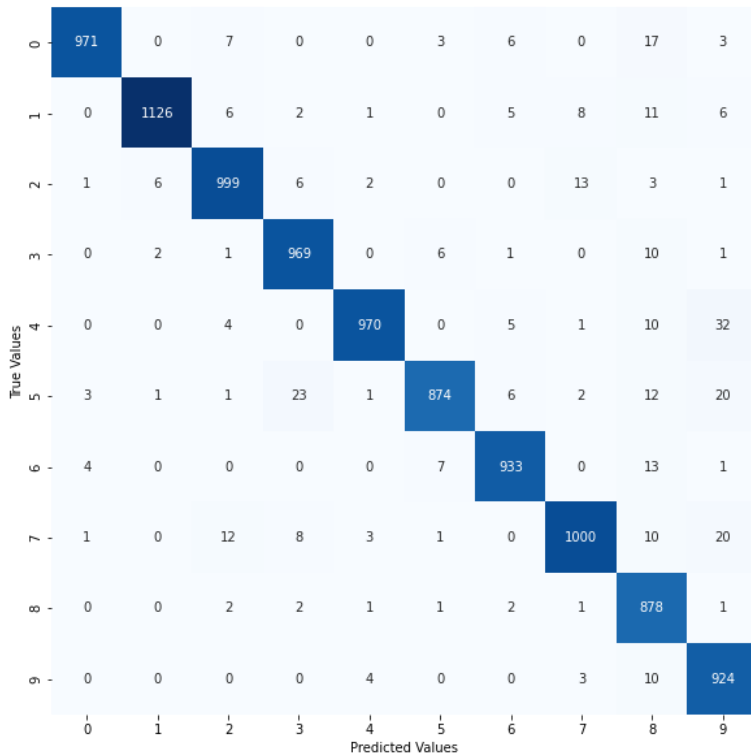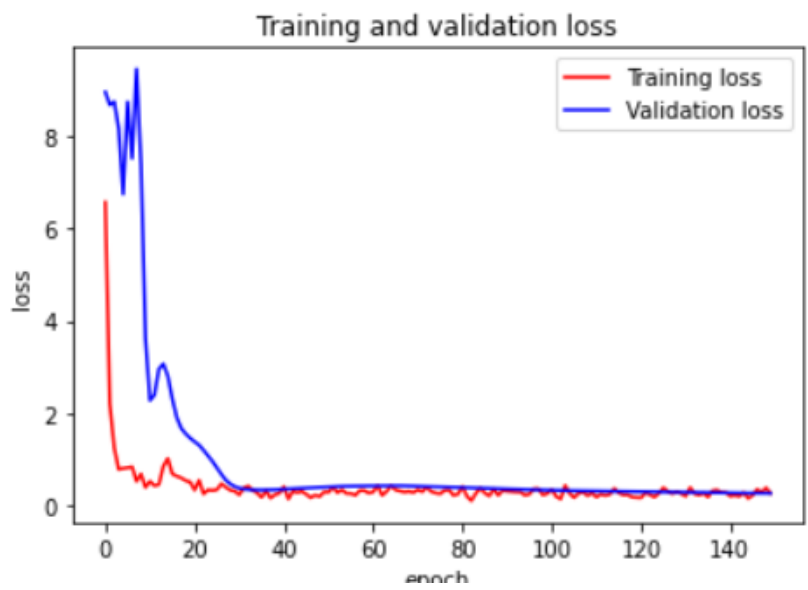Imbalanced Asymmetric Dataset:

Symmetric Learning

Original (Imbalanced) Dataset:

Balanced Dataset:
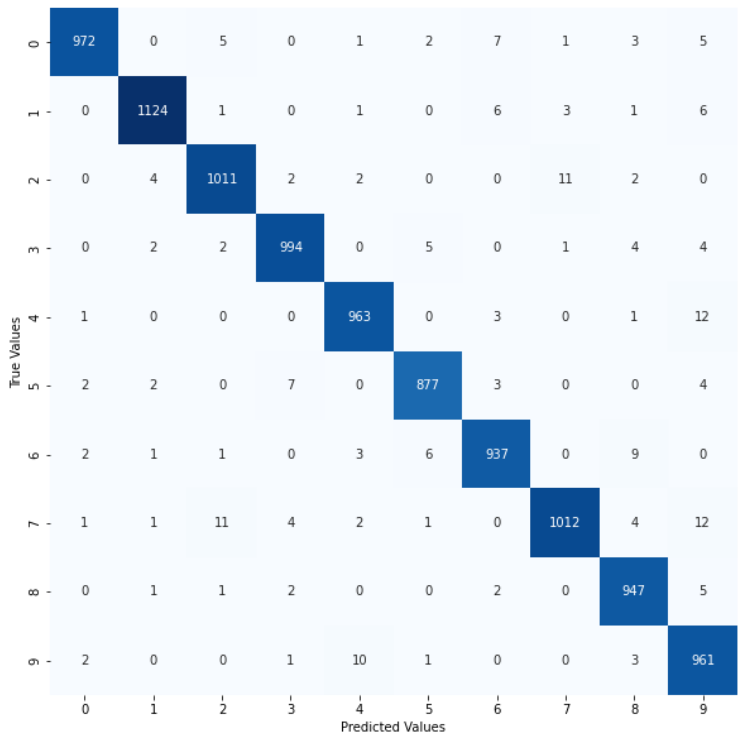
Symmetric Balanced Dataset:
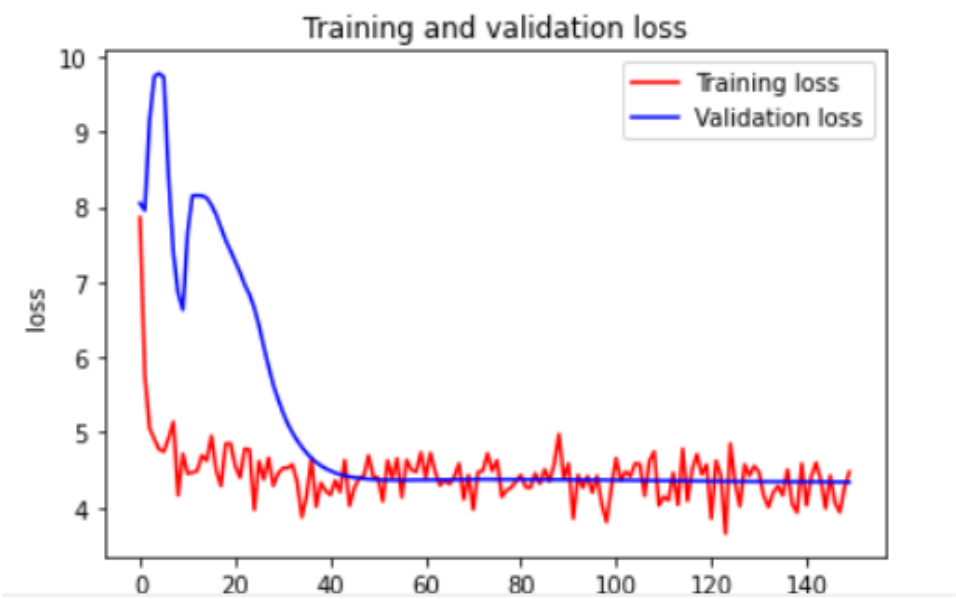
Asymmetric Balanced Dataset:

Symmetric Imbalanced:

Asymmetric Imbalanced Dataset:

True Values \ Predicted Values	0	1	2	3	4	5	6	7	8	9
0	9.7e+02	0	0	1	0	1	3	0	2	2
1	1	1.1e+03	3	0	4	2	3	3.0e+02	1	3
2	0	1	6.7e+02	3	1	0	1	0	4	2
3	0	3	2	6.8e+02	0	6	0	0	1	2
4	1	0	1	0	9.6e+02	2	3	2	3	8
5	1	1	1	0	3	5.7e+02	3.0e+02	0	1	1
6	4	2	1	6	1	3e+02	6.4e+02	0	4	2
7	1	3	3.5e+02	4	2	0	1	7.2e+02	0	3
8	0	2	4	3.2e+02	1	4	4	1	9.6e+02	3
9	0	0	1	0	7	2	0	1	1	9.8e+02

Training and validation loss

— Training loss
— Validation loss
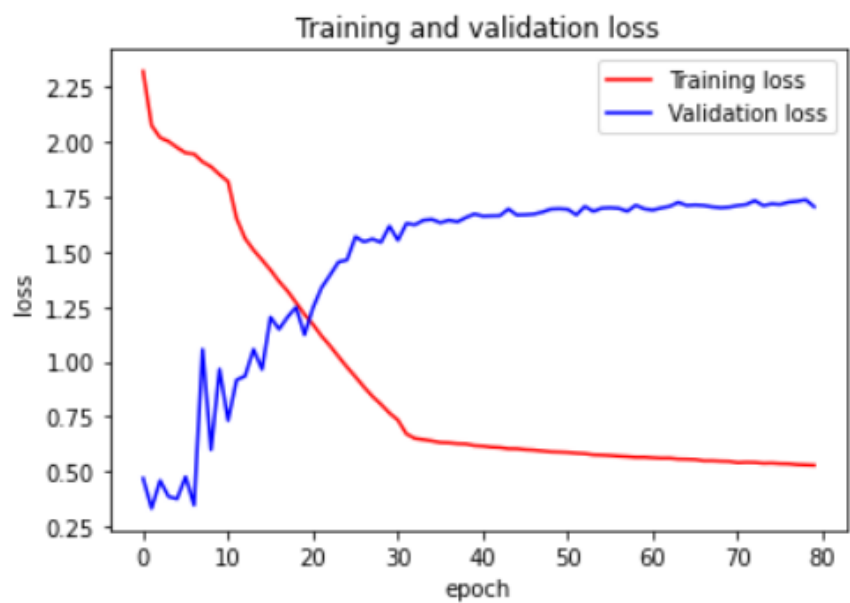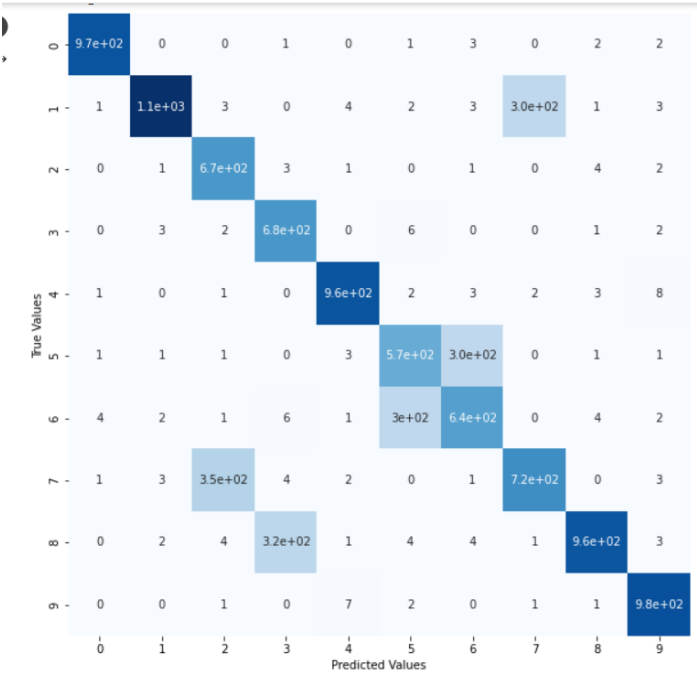
Normalized SL (Proposed Model)

Original (Imbalanced) Dataset:
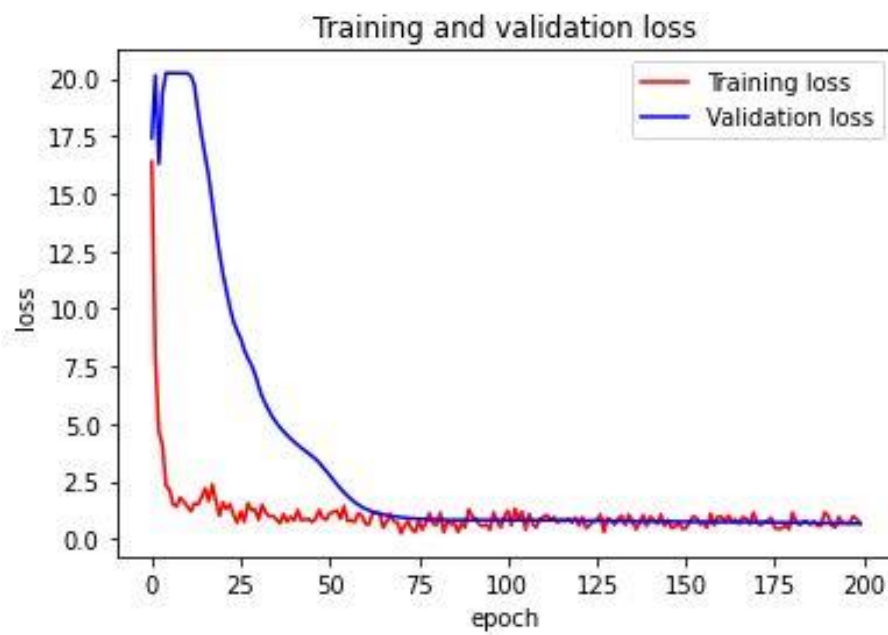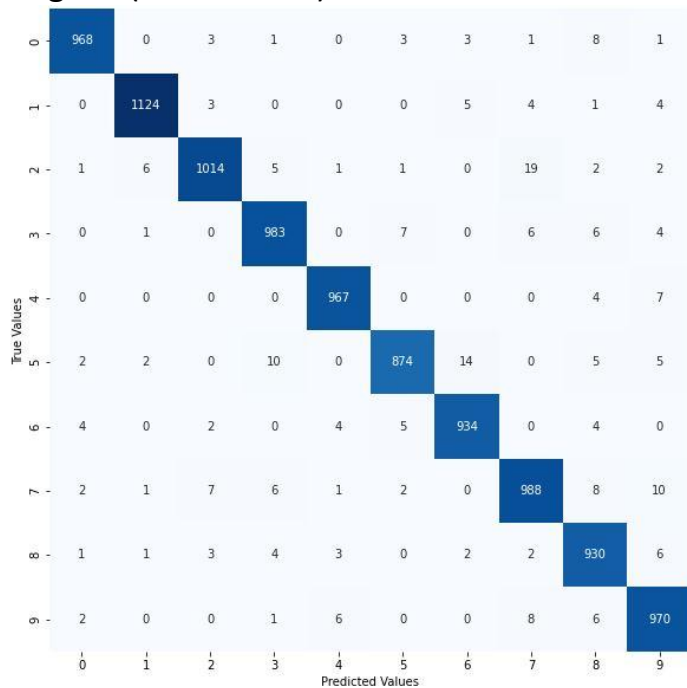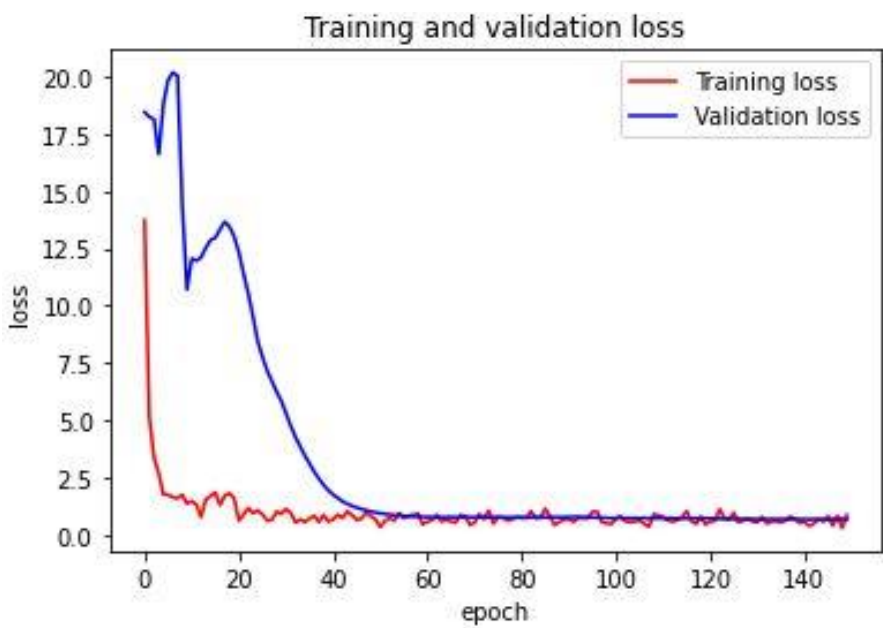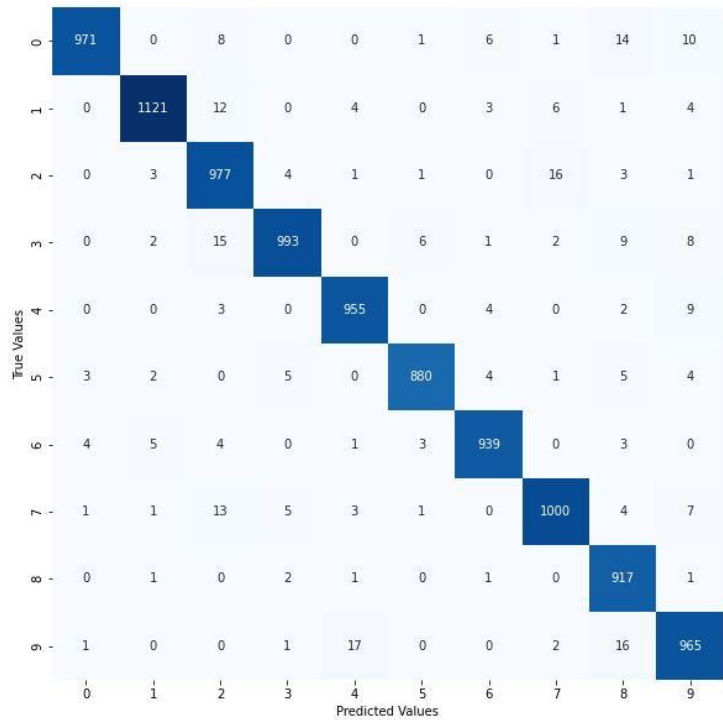
Original Balanced Dataset:
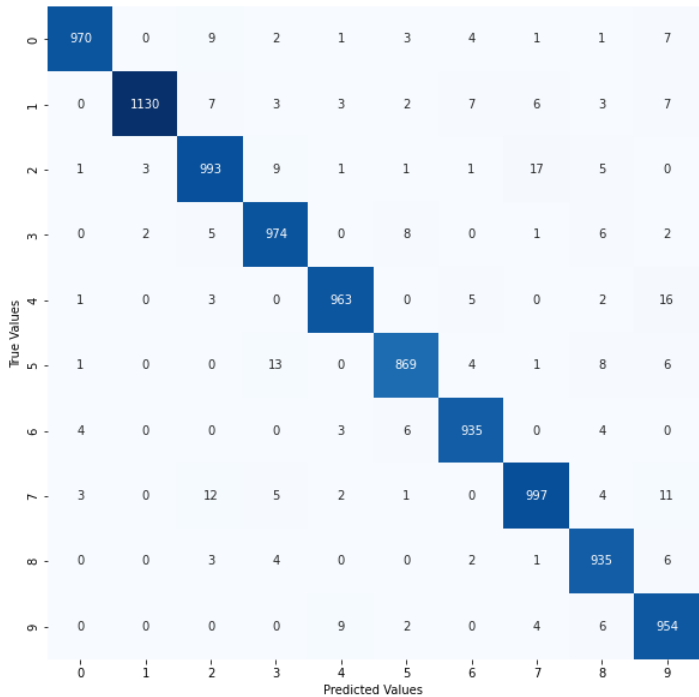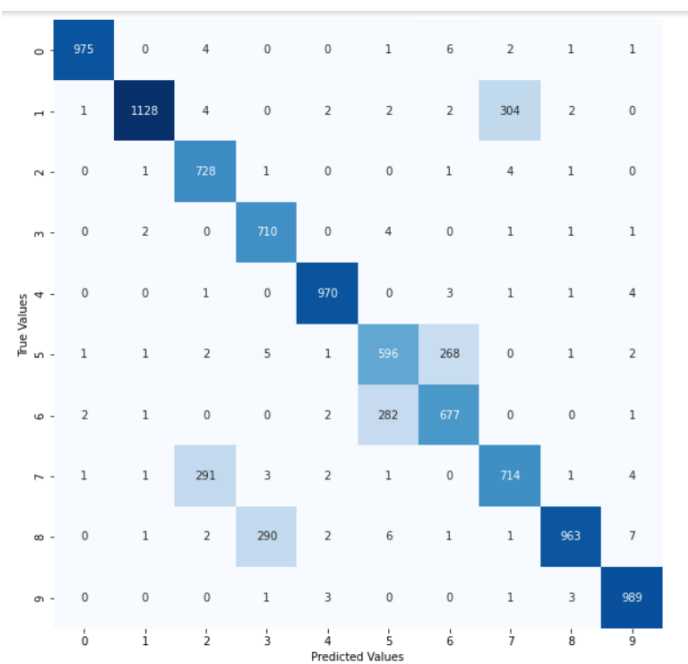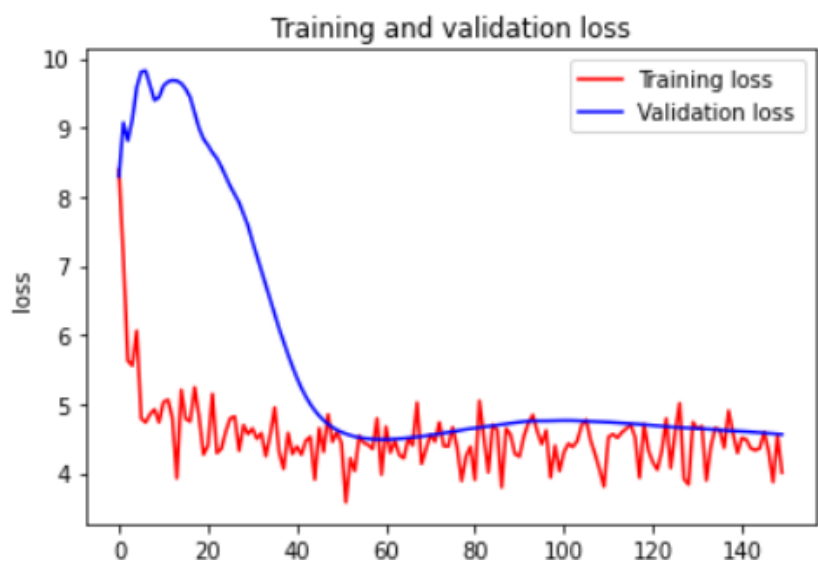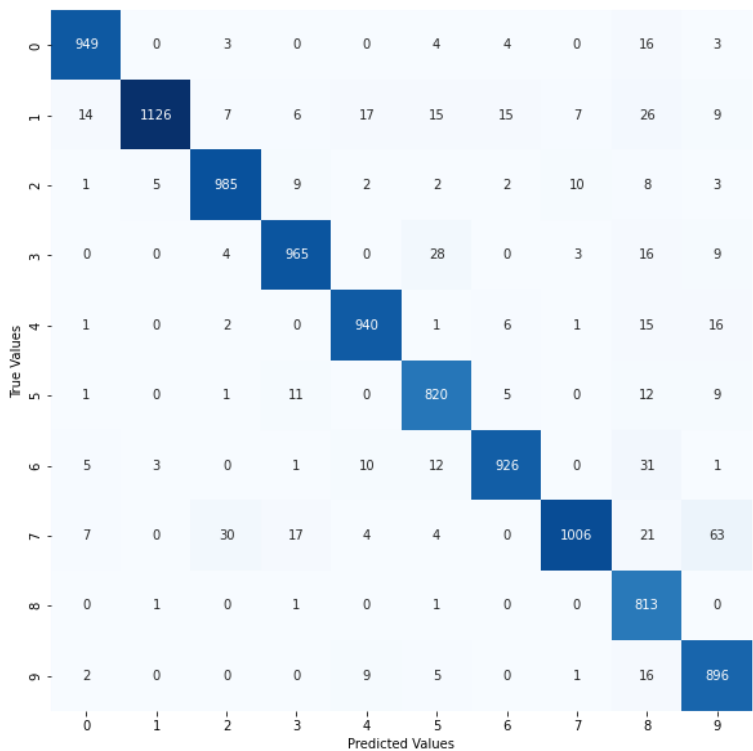
Asymmetric Imbalanced Dataset:
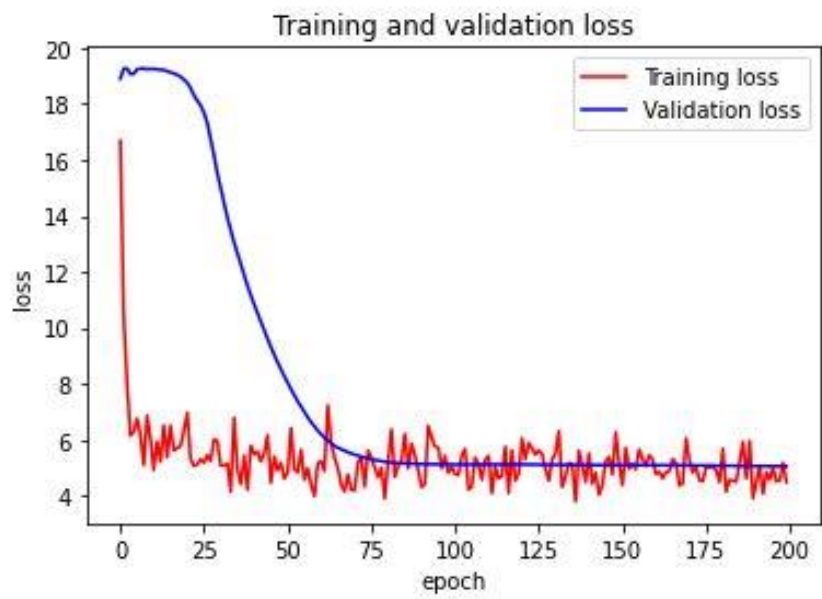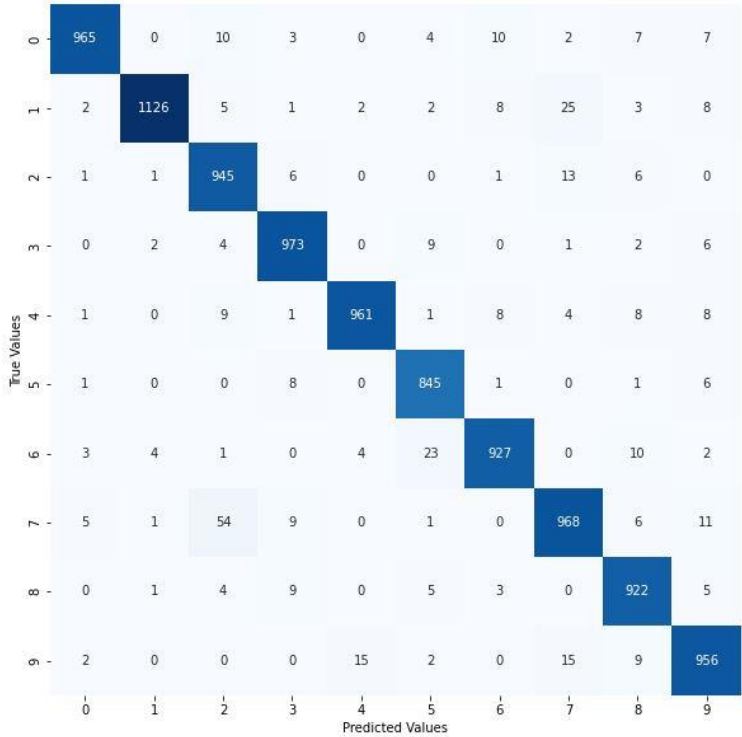
Asymmetric Balanced Dataset:
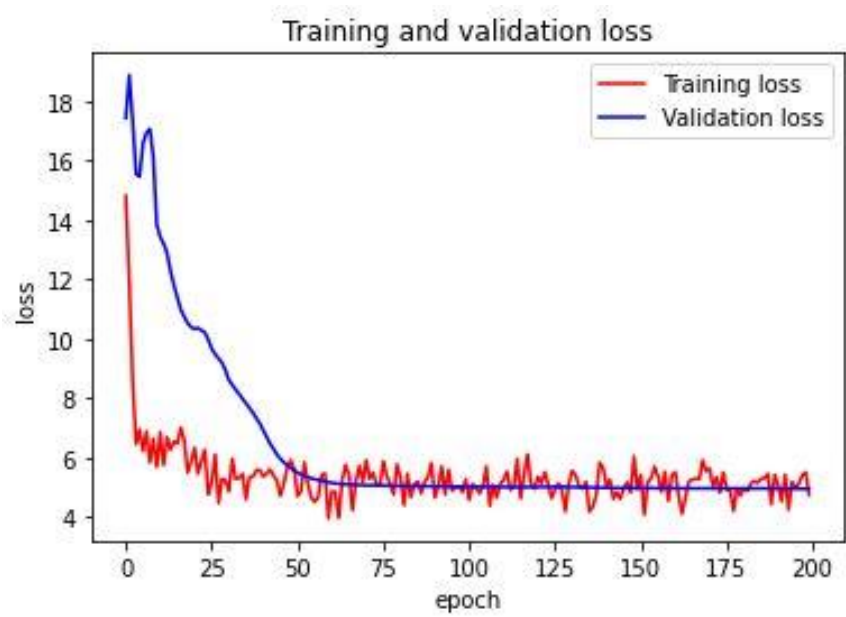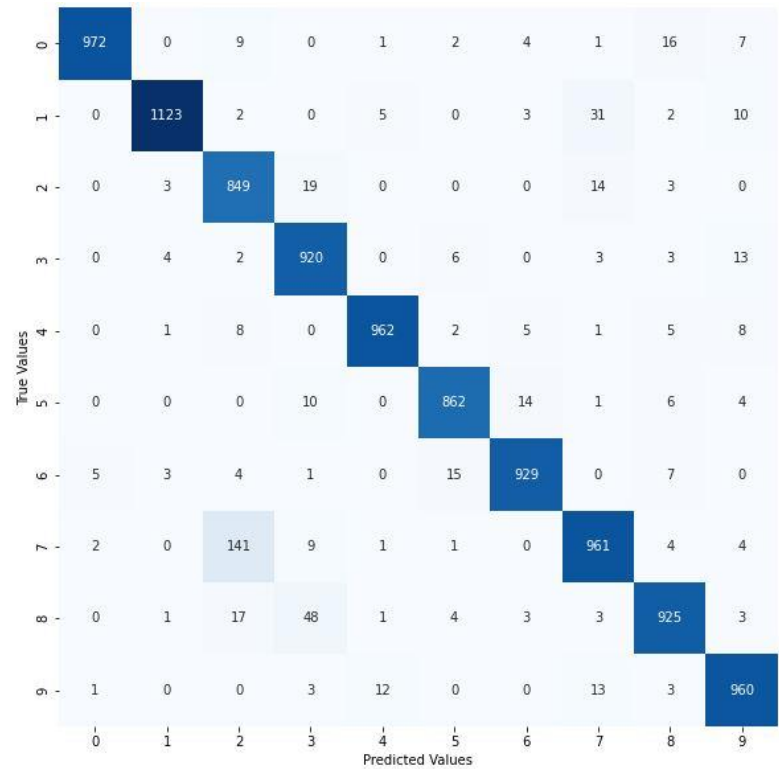
Symmetric Balanced Dataset:
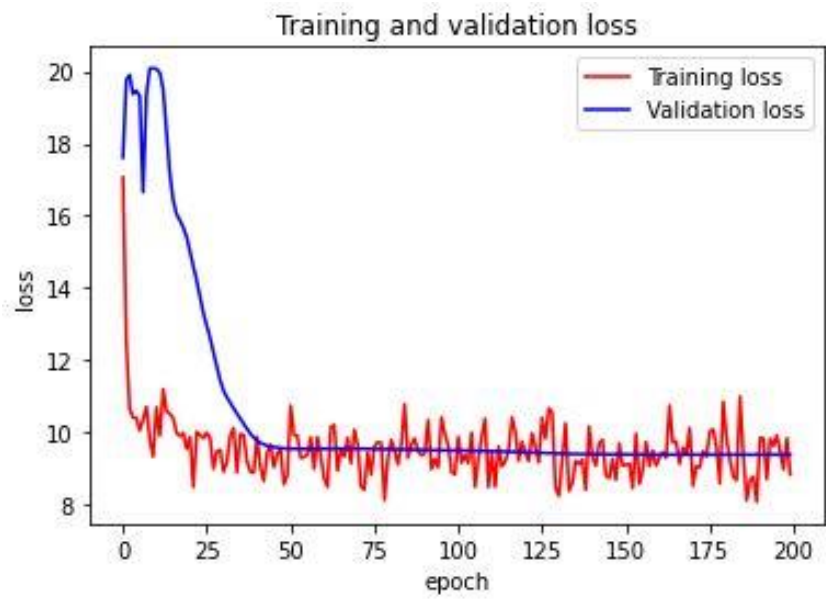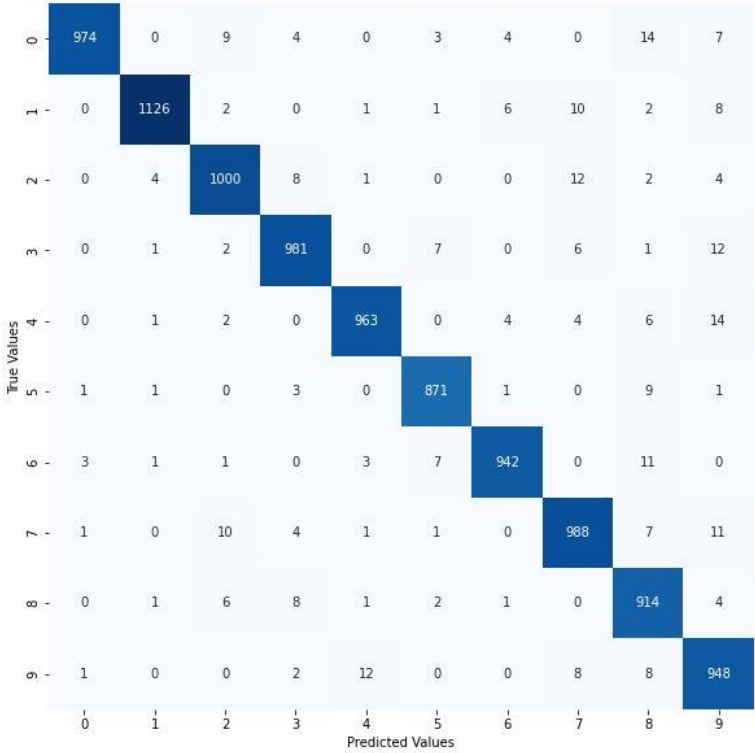
Symmetric Balanced Dataset:
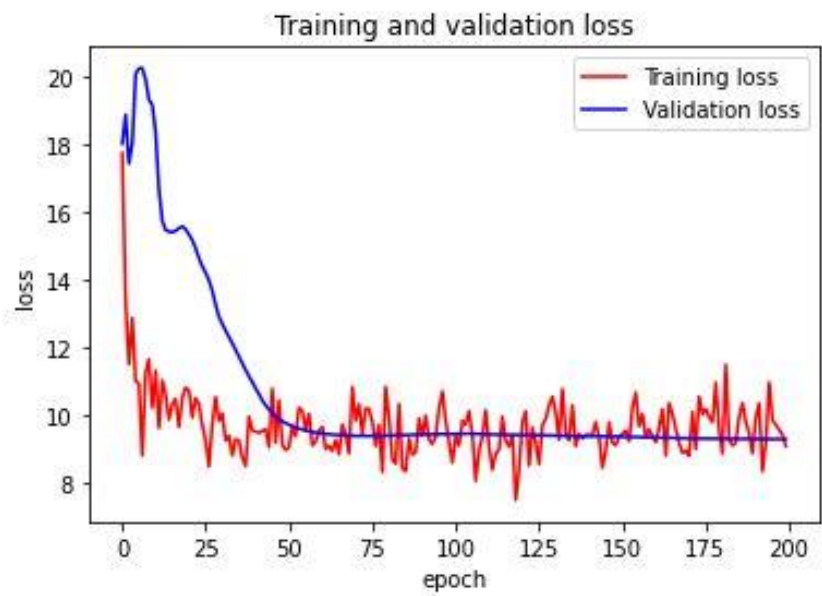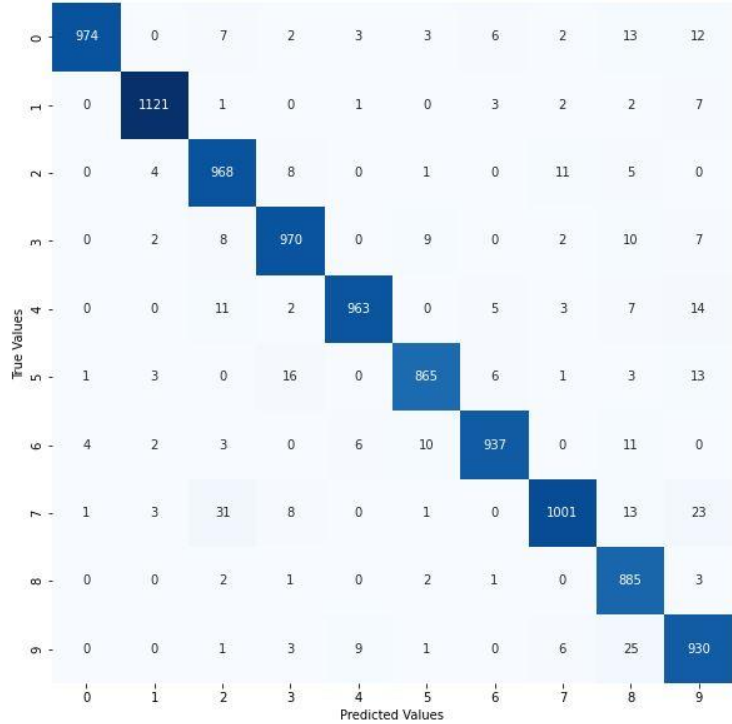
PCA+ KNN (Proposed ML)

Original (Imbalanced) Dataset:

```
accuracy: 0.949

[[ 967    0    1    1    1    2    6    2    0    0]
 [   0 1125    3    0    1    1    4    0    1    0]
 [  12    1  973    9    1    1    4    8   21    2]
 [   0    2    7  953    0   12    0   11   19    6]
 [   1    1    4    1  928    0    8    4    2   33]
 [   6    0    2   21    4  835   10    3    5    6]
 [   8    4    2    0    3    8  932    0    1    0]
 [   0   13   12    1    4    1    0  950    2   45]
 [   8    0    7   11    6   27    6    7  894    8]
 [   6    6    3   12   25    3    0   13    8  933]]
```


precision vs. recall curve

ROC curve
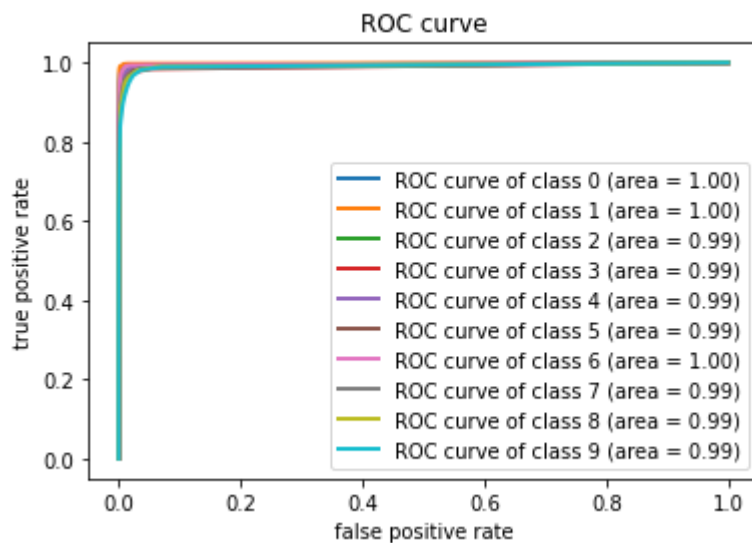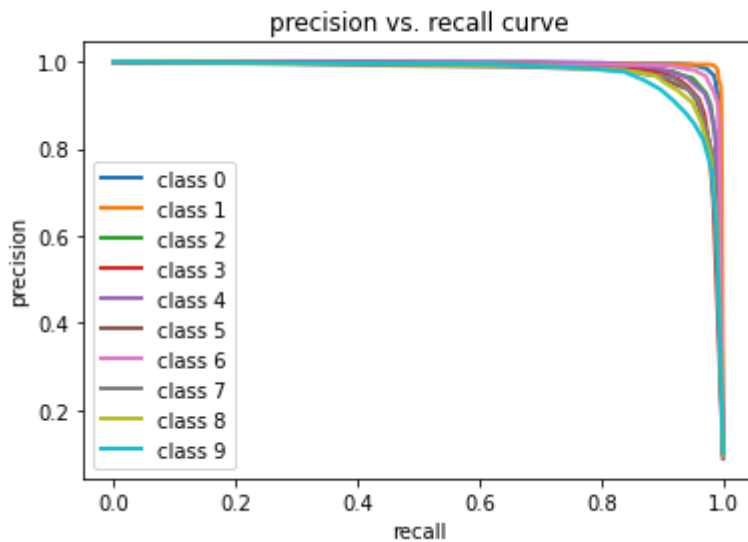
Balanced Dataset:

```
accuracy: 0.9504

[[ 967    0    0    1    1    2    7    2    0    0]
 [   0 1125    4    0    1    1    3    0    1    0]
 [   9    0  981    9    1    1    4    6   19    2]
 [   0    1    7  955    1    9    2   10   18    7]
 [   1    1    4    1  931    0    8    4    1   31]
 [   4    0    1   20    5  835   12    2    4    9]
 [   9    3    2    0    3    6  932    0    3    0]
 [   0   14   11    2    6    1    0  947    3   44]
 [   4    1    3   14    8   26    5    5  901    7]
 [   7    5    3   10   31    3    0   13    7  930]]
```


precision vs. recall curve

ROC curve
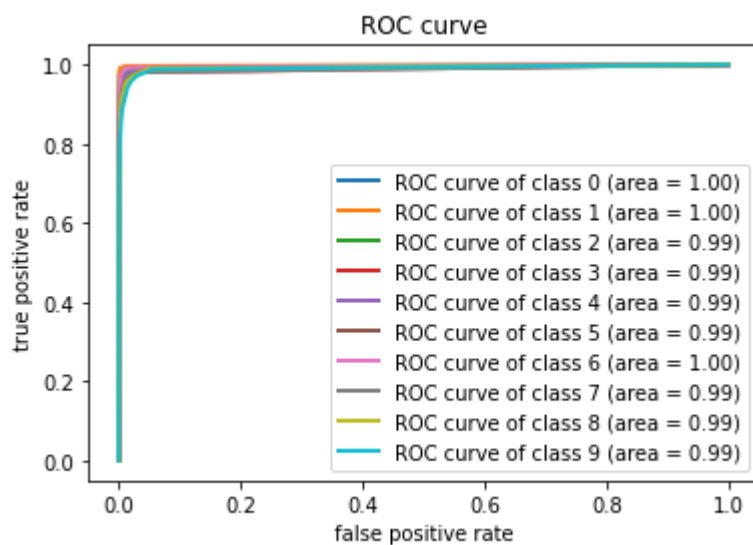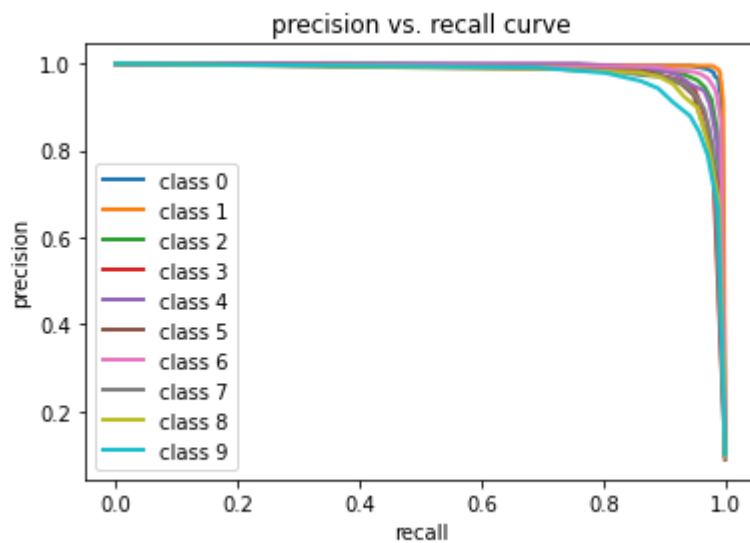
Imbalanced Asymmetric Dataset:

```
accuracy: 0.9264

[[ 960    1    3    1    1    2    5    4    0    3]
 [   0 1125    4    0    1    1    3    0    1    0]
 [  19    4  958   11    3    0    6   10   19    2]
 [   4    6   13  929    2   15    4   15   16    6]
 [   4    9   12    2  909    0    8    8    2   28]
 [  10    9    7   41   16  776   11    4    8   10]
 [  12    6    5    0    4   10  917    1    3    0]
 [   2   18   14    7   13    1    0  933    2   38]
 [  15    6    8   20    1   26    9   12  869    8]
 [  15    8    7   17   35    8    2   22    7  888]]
```

+ Code   + Markdown

precision vs. recall curve

ROC curve
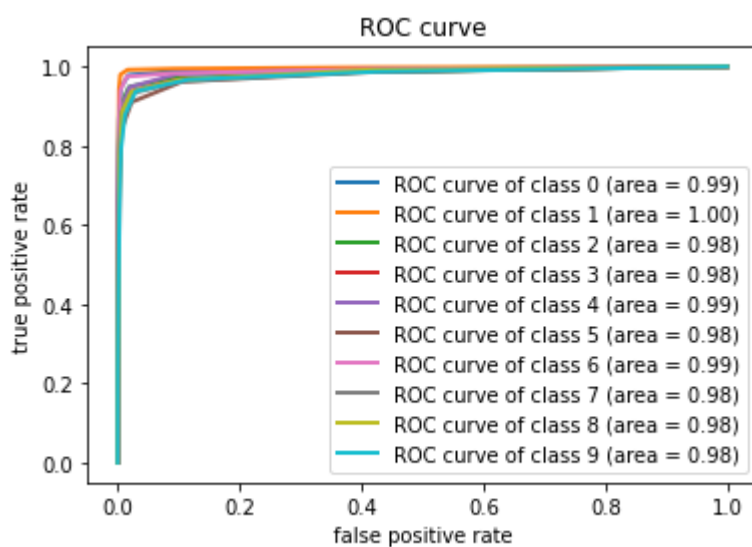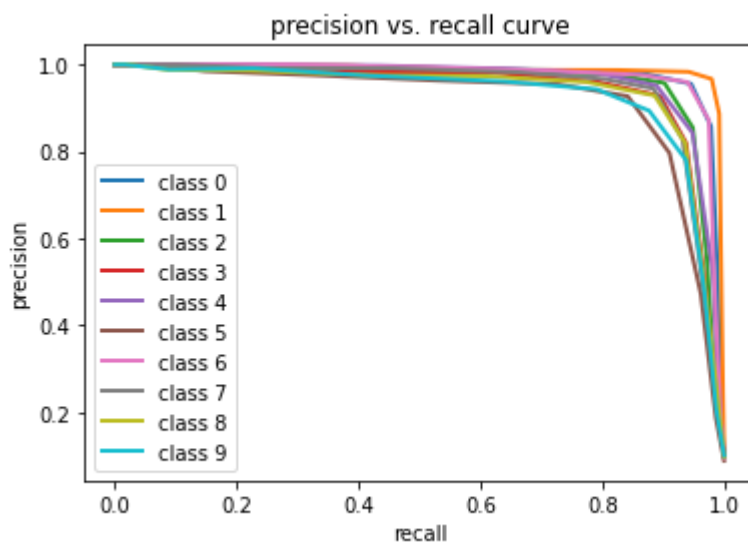
Balanced Symmetric Dataset:

```
accuracy: 0.8859

[[ 940    2    7    1    4    5   11    3    5    2]
 [   7 1089    8    5    6    4    4    5    6    1]
 [  22   19  924   13    7    4    7    8   19    9]
 [  25    6   27  892    1   13    8    9   20    9]
 [  12   13   18   14  870    9   12   10    5   19]
 [  20    8   11   26   14  781   10    4   10    8]
 [  20    9   15    9   18   22  860    2    2    1]
 [   7   27   24   11   23   11    6  877    6   36]
 [  23   12   21   24   19   32   14   13  807    9]
 [  29   12   22   22   47   11   12   22   13  819]]
```


precision vs. recall curve

ROC curve
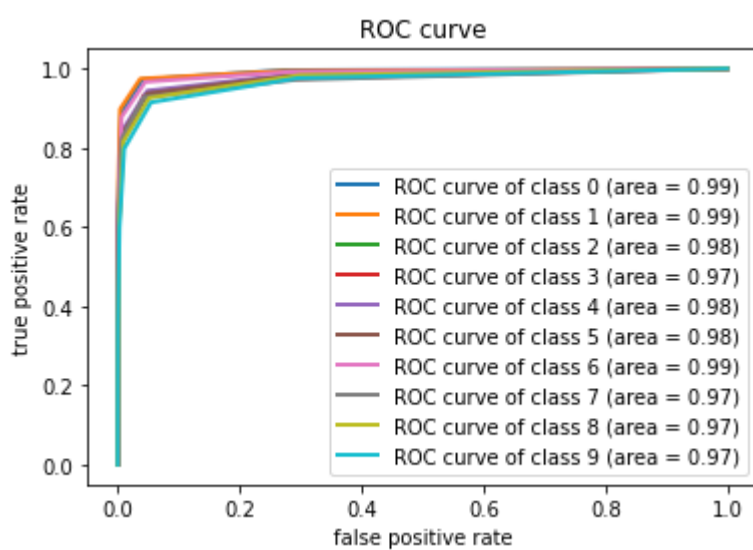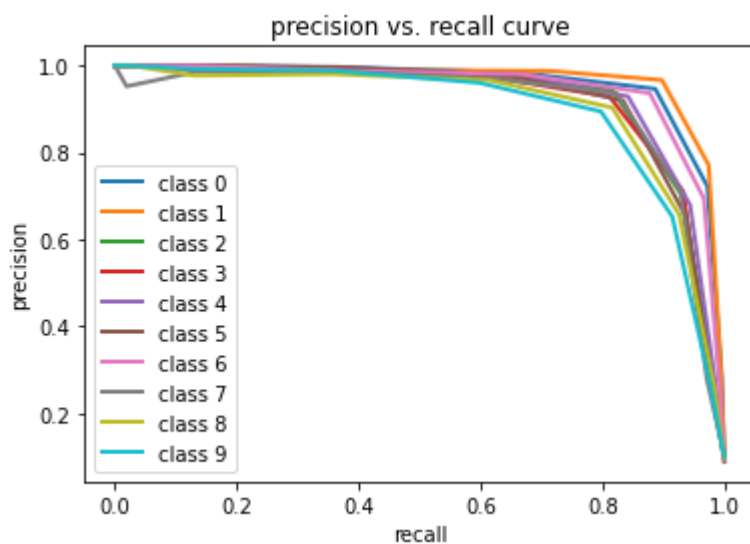
Balanced Asymmetric Dataset:

```
accuracy: 0.8176

[[ 971    1    0    0    1    4    1    1    1    0]
 [   0 1127    1    0    0    2    3    2    0    0]
 [  14    5  794    4    1    3    3  187   19    2]
 [   0    8    3  770    1   13    4    3  202    6]
 [   1    4    4    2  942    2    6    2    2   17]
 [   5    3    3   18    5  699  147    0    7    5]
 [  11    4    1    0    5  411  522    1    3    0]
 [   1  455    8    1    8    0    0  511    2   42]
 [   5    4    5    6    8   18    4    2  915    7]
 [   7    9    4    9   31    3    1   11    9  925]]
```

precision vs. recall curve

ROC curve
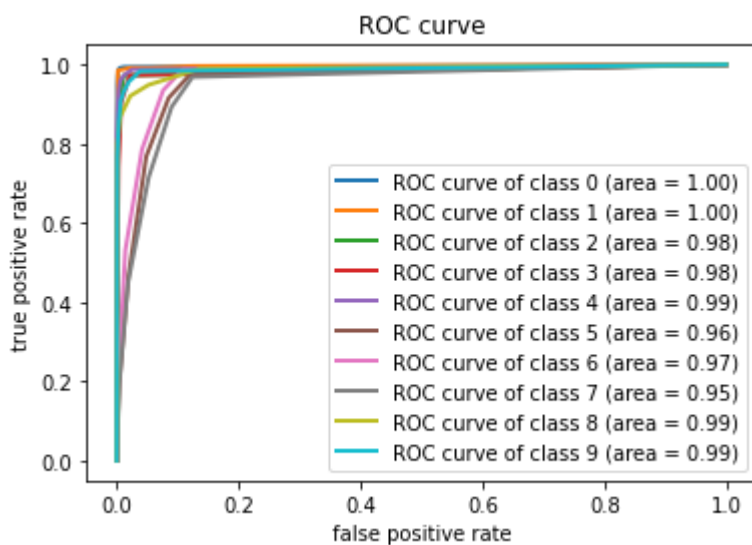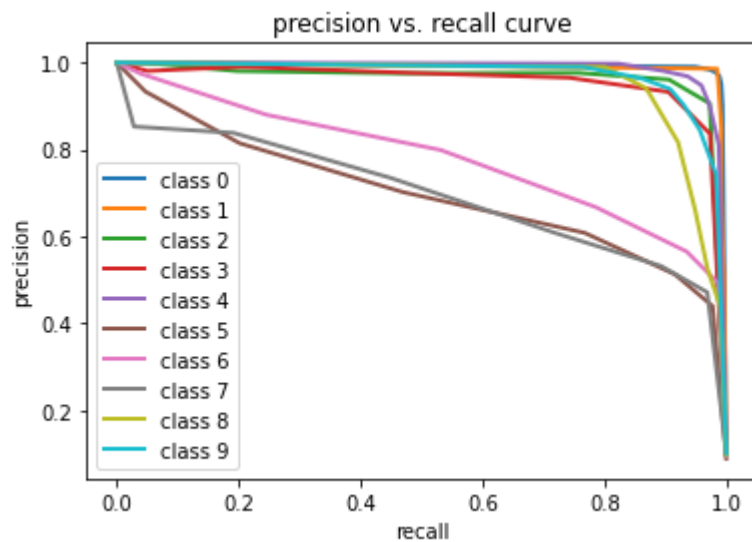
Imbalanced Asymmetric Dataset:

```
accuracy: 0.819

[[ 968    0    0    1    1    2    6    2    0    0]
 [   0 1125    0    0    1    1    4    3    1    0]
 [  12    2  719    6    2    1    1  261   26    2]
 [   0    9    3  631    0    9    5    9  335    9]
 [   1    1    4    1  931    3    4    3    2   32]
 [   8    0    2   12    6  629  215    2    9    9]
 [   8    5    1    0    4  272  665    1    2    0]
 [   0  292   12    1    5    0    0  666    3   49]
 [   7    3    3    2    9   15   10    3  914    8]
 [   6    9    3    4   22    2    1    6   14  942]]
```

precision vs. recall curve

ROC curve
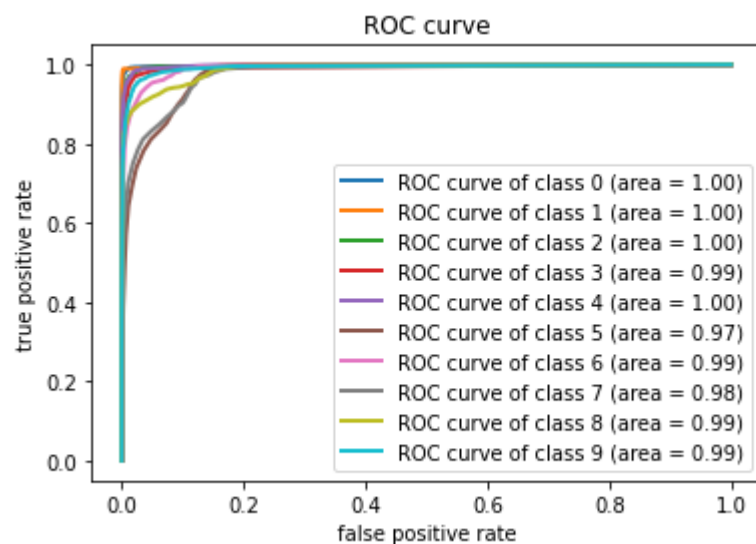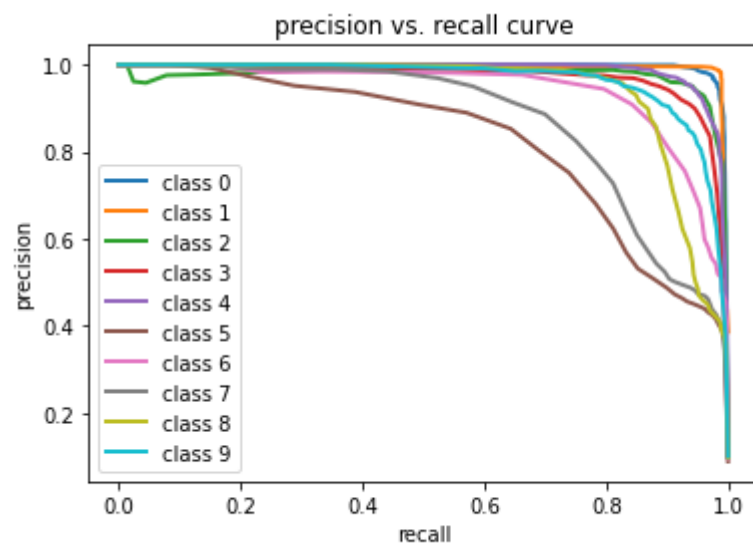
Logistic Regression

Original (Imbalanced) Dataset:

```
accuracy: 0.9201

[[ 960    0    1    2    0    5    6    3    1    2]
 [   0 1112    3    1    0    1    5    1   12    0]
 [   8    8  920   20    9    5   10   11   37    4]
 [   4    0   17  919    2   22    4   12   21    9]
 [   1    2    5    3  914    0   10    2    7   38]
 [  10    2    0   42   10  769   17    7   28    7]
 [   9    3    7    2    6   20  907    1    3    0]
 [   2    7   22    5    8    1    1  950    5   27]
 [  10   14    5   21   14   27    7   11  853   12]
 [   8    8    2   13   31   14    0   24   12  897]]
```


Micro-averaged over all classes

AP = 0.95

Some extension of Receiver operating characteristic to multiclass

- micro-average ROC curve (area = 0.98873)
- macro-average ROC curve (area = 0.98628)
- ROC curve of class 0 (area = 0.99841)
- ROC curve of class 1 (area = 0.99858)
- ROC curve of class 2 (area = 0.97956)
- ROC curve of class 3 (area = 0.98539)
- ROC curve of class 4 (area = 0.99332)
- ROC curve of class 5 (area = 0.97615)
- ROC curve of class 6 (area = 0.99608)
- ROC curve of class 7 (area = 0.99132)
- ROC curve of class 8 (area = 0.96666)
- ROC curve of class 9 (area = 0.97677)

Balanced Dataset:

```
accuracy: 0.9193

[[ 960    0    0    2    0    5    6    4    1    2]
 [   0 1112    3    2    0    1    5    1   11    0]
 [   7    8  917   18   11    7   11   10   40    3]
 [   4    0   19  915    2   23    4   12   22    9]
 [   1    2    3    3  915    0   11    3    6   38]
 [  10    2    0   33    9  776   17    9   29    7]
 [  10    3    7    2    6   20  906    1    3    0]
 [   2    8   23    5    7    2    1  948    5   27]
 [  10   15    7   21   13   30    8    9  849   12]
 [   9    8    1   15   34   12    0   24   11  895]]
```
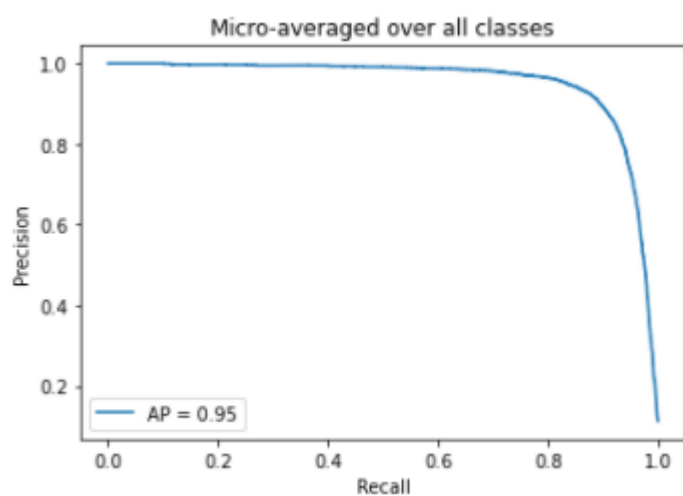

Micro-averaged over all classes

AP = 0.95

Some extension of Receiver operating characteristic to multiclass

micro-average ROC curve (area = 0.98863)
macro-average ROC curve (area = 0.98614)
ROC curve of class 0 (area = 0.99834)
ROC curve of class 1 (area = 0.99859)
ROC curve of class 2 (area = 0.97961)
ROC curve of class 3 (area = 0.98493)
ROC curve of class 4 (area = 0.99317)
ROC curve of class 5 (area = 0.97642)
ROC curve of class 6 (area = 0.99603)
ROC curve of class 7 (area = 0.99121)
ROC curve of class 8 (area = 0.96574)
ROC curve of class 9 (area = 0.97681)
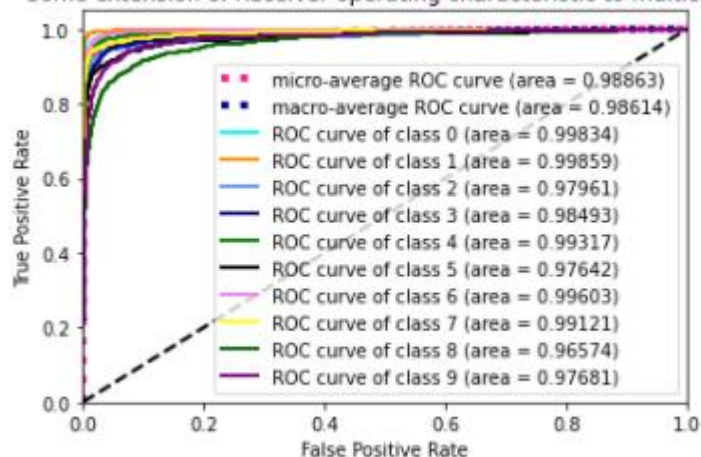
Imbalanced Symmetric Dataset:

```
accuracy: 0.8748

[[ 936    1    3    4    3    6   13    4    6    4]
 [   0 1109    3    4    1    1    5    1   11    0]
 [  15   49  834   25    7    2   25   29   37    9]
 [   6    8   28  882    2   22    8   24   18   12]
 [   2   12    5    2  899    1   11    1    8   41]
 [  14   17    3   61   16  692   15   25   37   12]
 [  13    8    9    2   20   22  872    2    9    1]
 [   6   29   15    5   15    1    2  906    3   46]
 [  12   45    8   33   23   36   13   13  771   20]
 [  14   12    6   13   62    7    3   36    9  847]]
```


Micro-averaged over all classes

Some extension of Receiver operating characteristic to multiclass
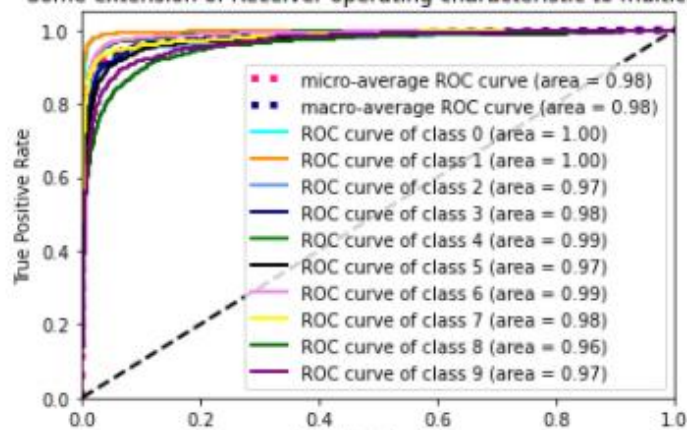
Imbalanced Asymmetric Dataset:

```
accuracy: 0.8123

[[ 966    0    0    1    0    3    3    4    3    0]
 [   0 1098    2    1    1    1    5    1   26    0]
 [  13   21  677   15   15    7   15  160  102    7]
 [   5   17    9  697    2   19    7   22  224    8]
 [   1    8    0    1  913    5    4    4   10   36]
 [  11   19    0   21   18  591   76   20  122   14]
 [  13    9    2    0   10  113  779   18   14    0]
 [   4  266    6    2   14    3    1  627   10   95]
 [   9   22    5    5   13   16   17    5  872   10]
 [   8   22    0    1   32    3    2   11   27  903]]
```
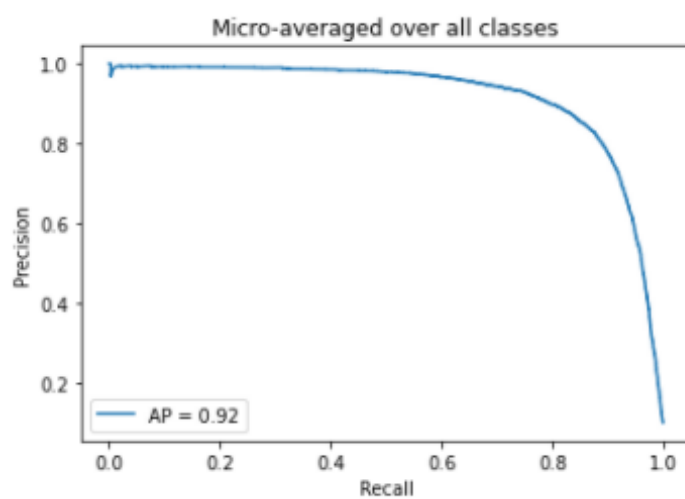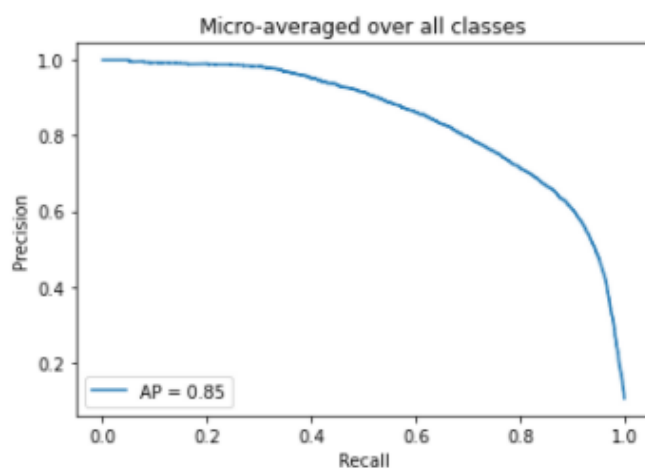

Micro-averaged over all classes

AP = 0.85

Some extension of Receiver operating characteristic to multiclass

- micro-average ROC curve (area = 0.97)
- macro-average ROC curve (area = 0.97)
- ROC curve of class 0 (area = 1.00)
- ROC curve of class 1 (area = 0.99)
- ROC curve of class 2 (area = 0.98)
- ROC curve of class 3 (area = 0.98)
- ROC curve of class 4 (area = 0.99)
- ROC curve of class 5 (area = 0.92)
- ROC curve of class 6 (area = 0.96)
- ROC curve of class 7 (area = 0.95)
- ROC curve of class 8 (area = 0.94)
- ROC curve of class 9 (area = 0.98)
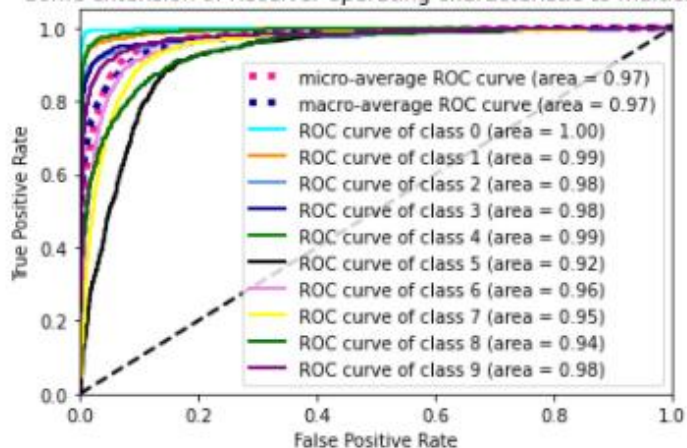
Balanced Symmetric Dataset:

```
accuracy: 0.8807

[[ 946    1    1    1    0    8   13    2    6    2]
 [   0 1097    3    3    1    2    5    0   23    1]
 [  18   35  841   21   21    7   16   22   42    9]
 [   8    8   19  878    5   33    6   19   22   12]
 [   0    9    8    3  895    2    8    3   12   42]
 [  12   10    4   34   18  743   19    8   30   14]
 [  13   10    9    1   21   31  866    1    6    0]
 [   4   30   21    6   15    3    3  893    2   51]
 [  12   27    9   17   25   36   14   16  795   23]
 [   8   11    2   16   60   10    1   38   10  853]]
```


Micro-averaged over all classes

AP = 0.92

Some extension of Receiver operating characteristic to multiclass

- micro-average ROC curve (area = 0.98)
- macro-average ROC curve (area = 0.98)
- ROC curve of class 0 (area = 1.00)
- ROC curve of class 1 (area = 1.00)
- ROC curve of class 2 (area = 0.97)
- ROC curve of class 3 (area = 0.98)
- ROC curve of class 4 (area = 0.99)
- ROC curve of class 5 (area = 0.97)
- ROC curve of class 6 (area = 0.99)
- ROC curve of class 7 (area = 0.98)
- ROC curve of class 8 (area = 0.95)
- ROC curve of class 9 (area = 0.96)

Balanced Asymmetric Dataset:

```
accuracy: 0.8129

[[ 962    0    0    0    0    6    5    4    3    0]
 [   0 1098    2    1    1    1    6    1   25    0]
 [  12   25  670    9   16   14   15  156  111    4]
 [   6   15    9  694    3   19   11   21  224    8]
 [   1    6    3    0  916    1    6    3    8   38]
 [  10   15    0   18   17  597   82   21  120   12]
 [  13   10    1    0   10  113  776   17   18    0]
 [   4  233    8    3   16    4    2  649    8  101]
 [   9   22    2    6   11   26   13    9  866   10]
 [   8   19    0    3   33    5    1   15   24  901]]
```

Micro-averaged over all classes
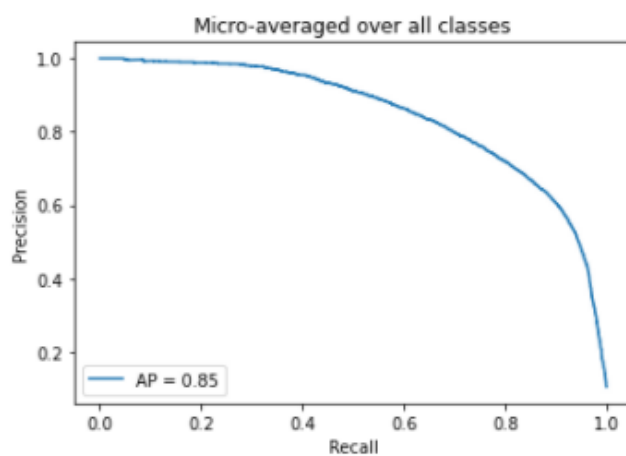
Some extension of Receiver operating characteristic to multiclass
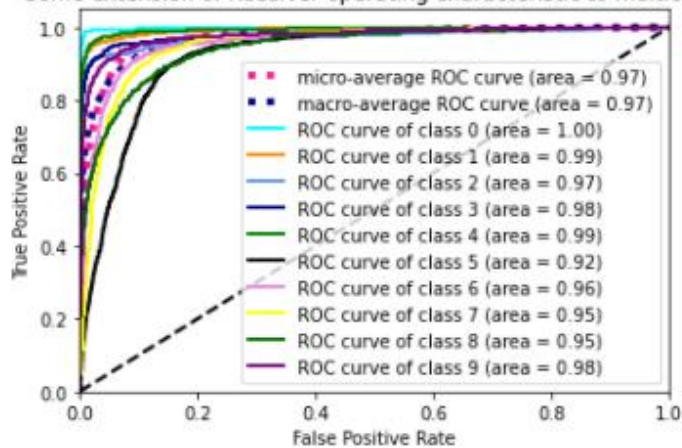
SVM

Original (Imbalanced) Dataset:

```
accuracy: 0.9239

[[ 965    0    1    1    0    2    8    1    2    0]
 [   0 1114    3    2    0    2    4    1    9    0]
 [  10    5  929   13   10    1   13   17   27    7]
 [   4    1   22  922    2   17    5   12   18    7]
 [   1    7    5    0  909    1   11    1    6   41]
 [  13    3    1   24   10  787   20    6   21    7]
 [  10    3    3    1    7   11  920    2    1    0]
 [   5   14   23    3    9    0    2  946    0   26]
 [  11    9    9   16   16   23   12   12  856   10]
 [  13    7    1   11   44   12    2   22    6  891]]
```
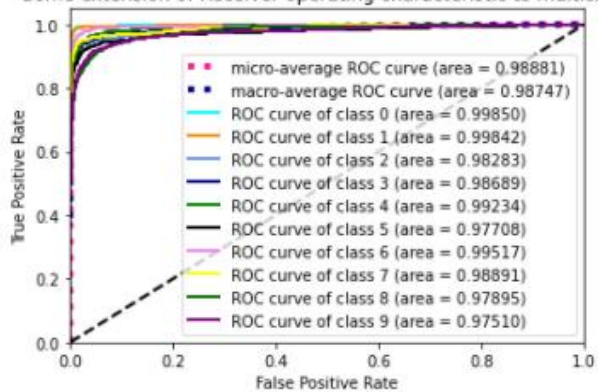
Balanced Dataset:

```
accuracy: 0.9257

[[ 965    0    1    1    0    2    8    1    2    0]
 [   0 1112    3    2    0    2    5    1   10    0]
 [   9    4  932   11   12    1   12   16   28    7]
 [   4    1   21  924    2   17    5   12   16    8]
 [   1    5    5    0  911    1   11    1    7   40]
 [  11    3    1   20    8  796   19    6   21    7]
 [  10    3    3    1    8   14  916    2    1    0]
 [   4   13   23    3    9    0    2  947    1   26]
 [  10    8    8   16   15   23   12   12  861    9]
 [  12    7    1   12   44   12    1   21    6  893]]
```

Micro-averaged over all classes

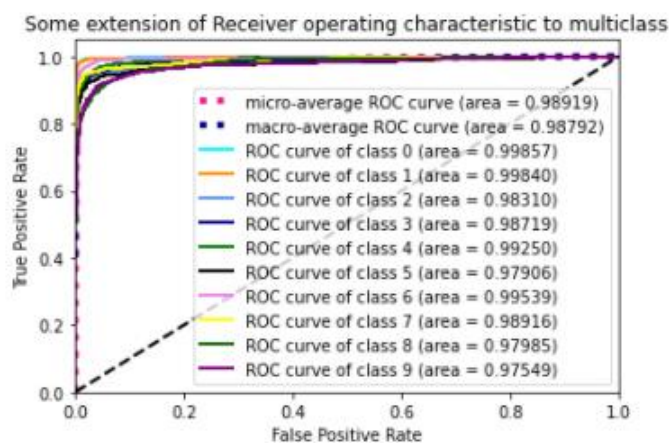

Some extension of Receiver operating characteristic to multiclass
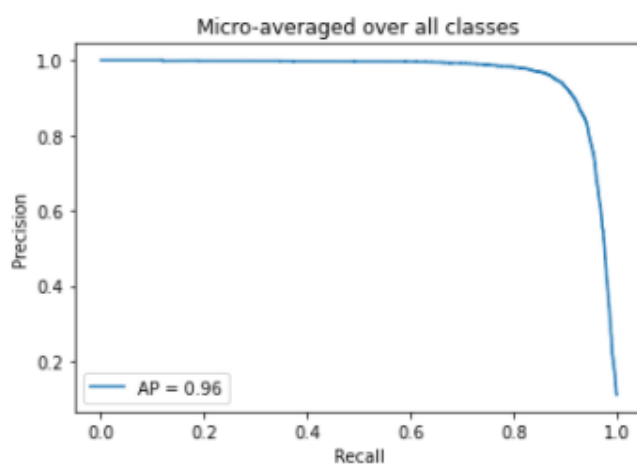
Imbalanced Asymmetric Dataset:

```
accuracy: 0.8637

[[ 967    0    0    0    0    7    3    1    2    0]
 [   0 1123    1    1    1    2    3    1    3    0]
 [   9    0  805    0    3    2    5  174   34    0]
 [   0    1    3  656    0    7    1   17  323    2]
 [   0    1    2    0  953    6    3    2    3   12]
 [   5    1    0    2    1  771   76    1   33    2]
 [   5    3    0    0    6  244  694    2    4    0]
 [   0  228    3    0    3    0    0  774    5   15]
 [   1    0    0    0    1    6    4    4  954    4]
 [   4    8    0    0   16    5    0   10   26  940]]
```


Micro-averaged over all classes

Some extension of Receiver operating characteristic to multiclass
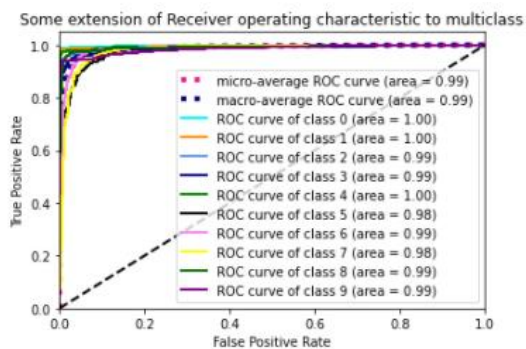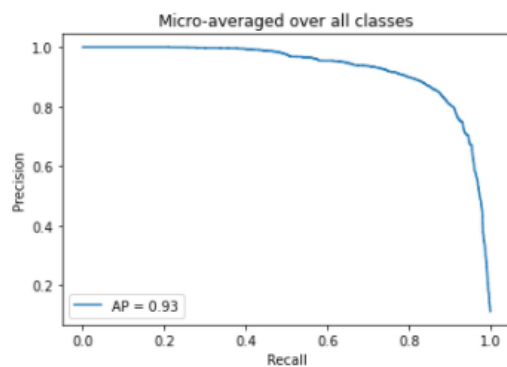
Balanced Symmetric Dataset:

```
accuracy: 0.9588

[[ 966    0    2    0    0    3    6    1    2    0]
 [   0 1123    3    2    1    1    3    1    1    0]
 [   7    1  996    6    0    0    2    9   11    0]
 [   1    0   15  963    0   10    0    9    9    3]
 [   0    0    6    0  947    0   10    2    2   15]
 [   5    2    8   13    1  851    4    1    6    1]
 [   8    3    4    0    4    6  932    0    1    0]
 [   1   13   22    4    8    1    0  968    2    9]
 [   5    0    9   12    5   11    4    7  916    5]
 [   9    6    6   11   26    8    1   12    4  926]]
```


Micro-averaged over all classes

Some extension of Receiver operating characteristic to multiclass
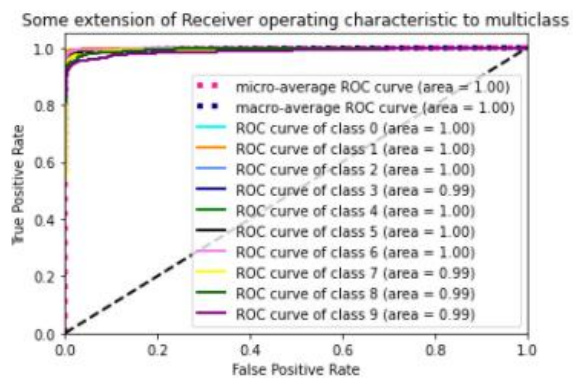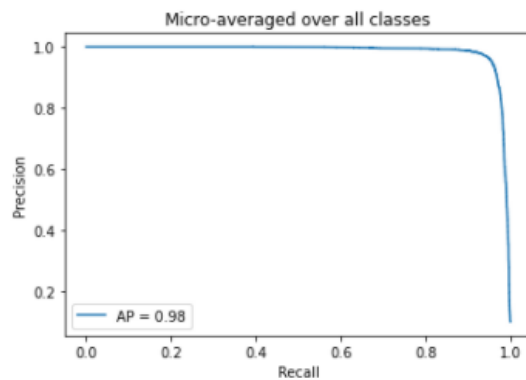
Imbalanced Symmetric Dataset:

accuracy: 0.9567

```
[[ 969    0    3    0    0    3    4    1    0    0]
 [   0 1125    3    2    1    1    3    0    0    0]
 [  11    0  988    7    2    0    2   11   11    0]
 [   1    0   12  970    0    5    1    8    9    4]
 [   2    1    6    0  948    1    6    1    2   15]
 [   5    2    9   25    3  830    9    2    4    3]
 [  14    3    2    0    7    2  930    0    0    0]
 [   2   18   24    2    4    0    0  966    2   10]
 [   5    2   10   17    8    7    1    8  914    2]
 [  11    7    8    8   28    1    1   12    6  927]]
```


Micro-averaged over all classes

AP = 0.98

Some extension of Receiver operating characteristic to multiclass

micro-average ROC curve (area = 1.00)
macro-average ROC curve (area = 1.00)
ROC curve of class 0 (area = 1.00)
ROC curve of class 1 (area = 1.00)
ROC curve of class 2 (area = 1.00)
ROC curve of class 3 (area = 0.99)
ROC curve of class 4 (area = 1.00)
ROC curve of class 5 (area = 1.00)
ROC curve of class 6 (area = 1.00)
ROC curve of class 7 (area = 0.99)
ROC curve of class 8 (area = 1.00)
ROC curve of class 9 (area = 0.99)