These graphs represent the performance of ParDFROCC on different datasets.

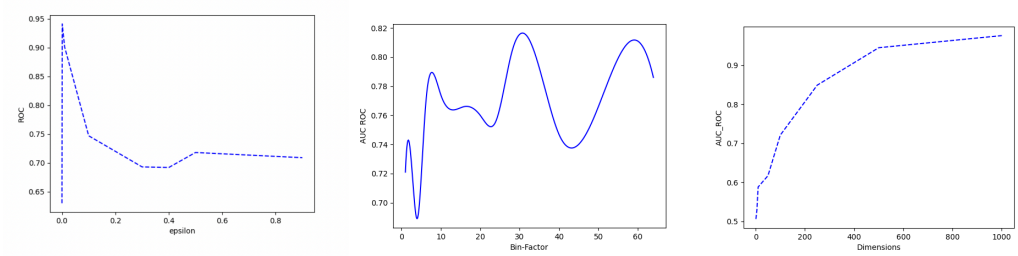
AUC - ROC curve is a performance measurement for classification problems at various threshold settings. ROC is a probability curve and AUC represents the degree or measure of separability. It tells how much the model is capable of distinguishing between classes

These are the default parameters value used while plotting the graphs(unless any change is mentioned)

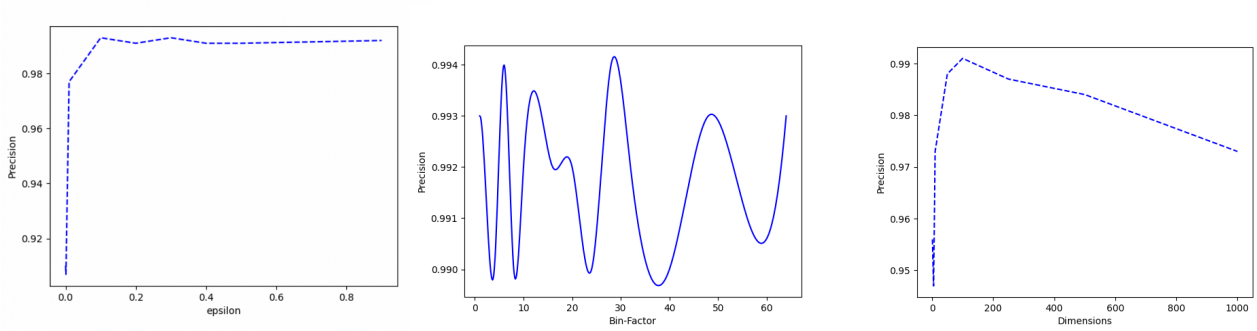
```
epsilon: float = 0.1,
threshold: float = 1,
bin_factor: int = 2,
density: float = 0.01,
kernel: Type[np.dot] = linear_kernel

precision: type = np.float32,

n_jobs: int = 8,
Dimension= 100
```

Mnist 784 data - The MNIST database of handwritten digits with 784 features:

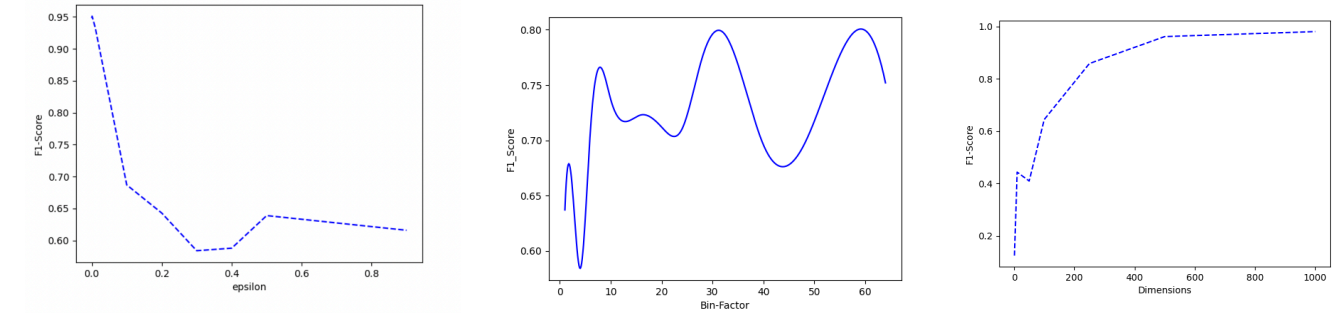
AUC Curve - AUC represents the degree or measure of separability. It tells how much the model is capable of distinguishing between classes. Higher the AUC, the better the model is at predicting 0 classes as 0 and 1 classes as 1



Precision Curve - Precision is one indicator of a machine learning model's performance – the quality of a positive prediction made by the model. Precision refers to the number of true positives divided by the total number of positive predictions (i.e., the number of true positives plus the number of false positives).

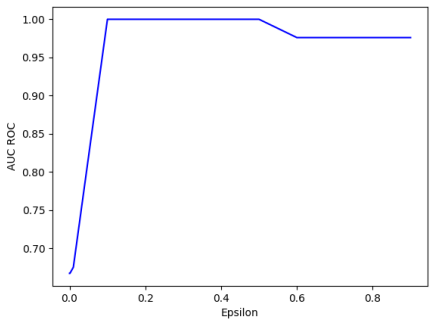


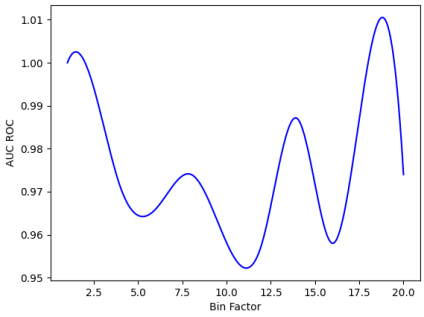
F1-Score-The F1-score **combines the precision and recall of a classifier into a single metric by taking their harmonic mean**. It is primarily used to compare the performance of two classifiers.

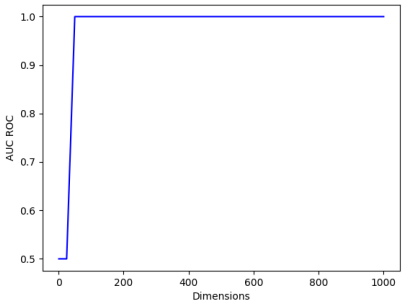


IRIS - The data set contains 3 classes of 50 instances each, where each class refers to a type of iris plant. One class is linearly separable from the other 2; the latter are NOT linearly separable from each ot

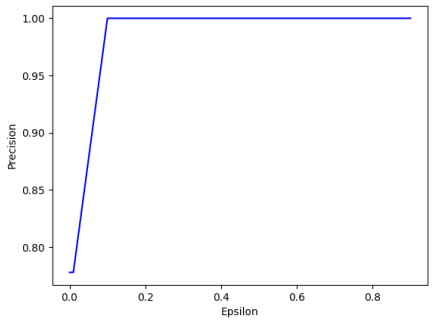
AUC

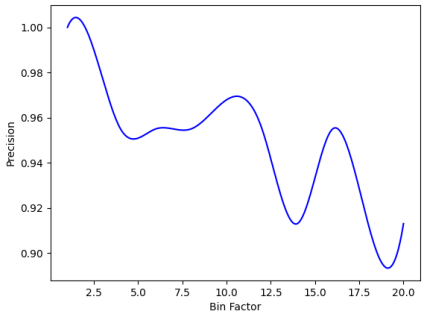


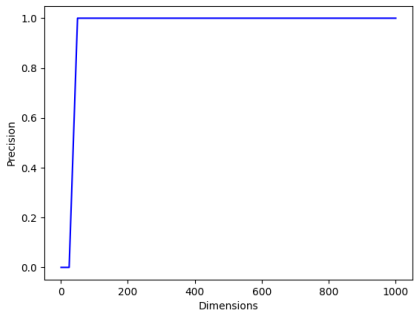




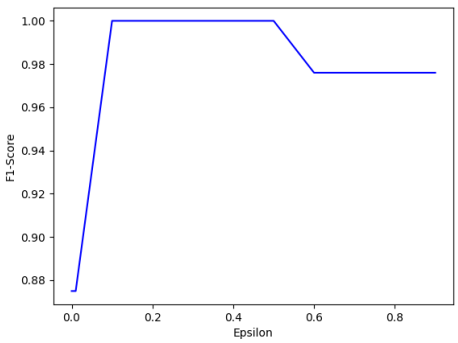
Precision

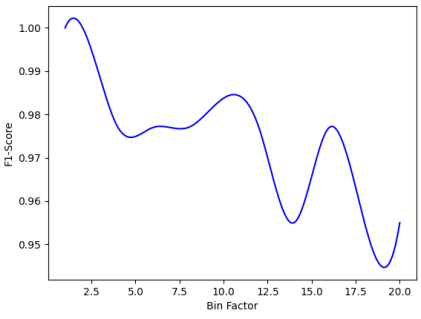


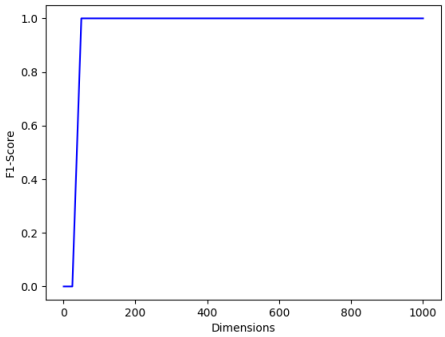




F1-Score

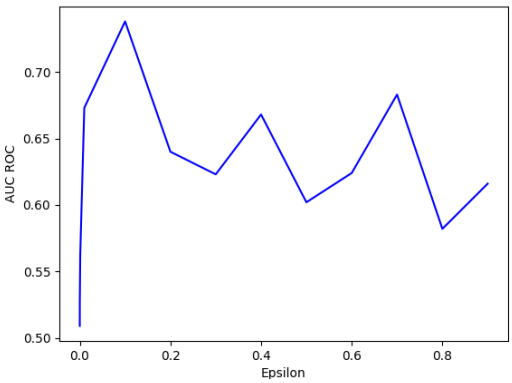


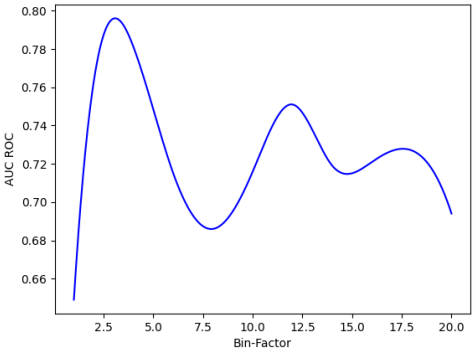


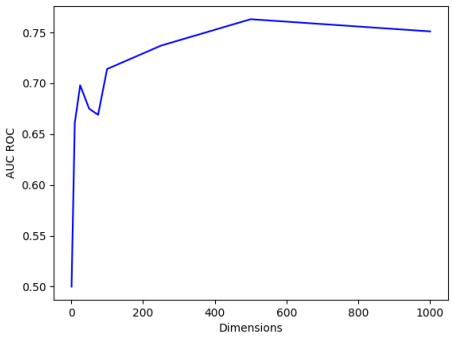


Diabetes -Diabetes tested positive/negative classification based on patient data.

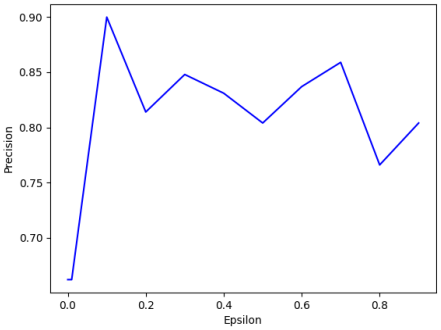
ROC

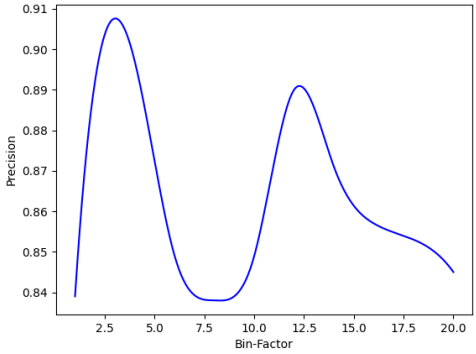


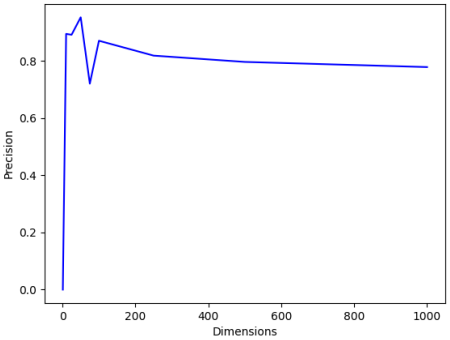




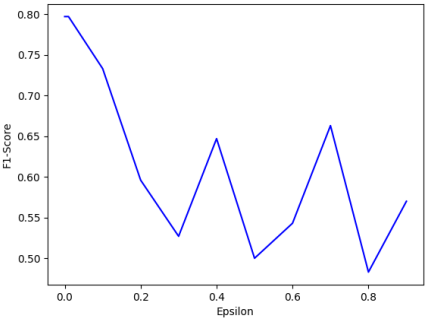
Precision

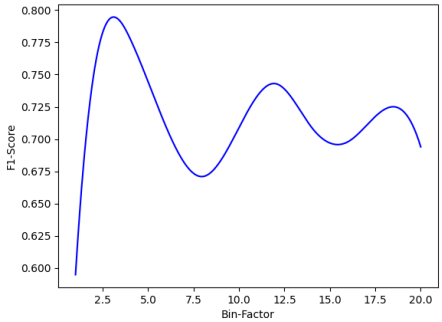


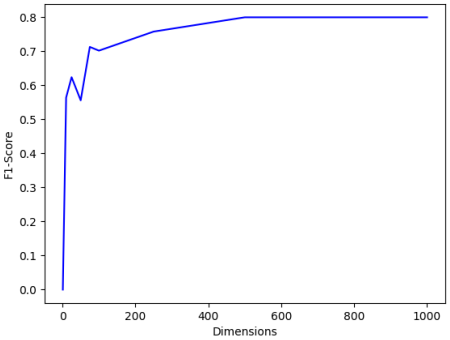




F1-Score

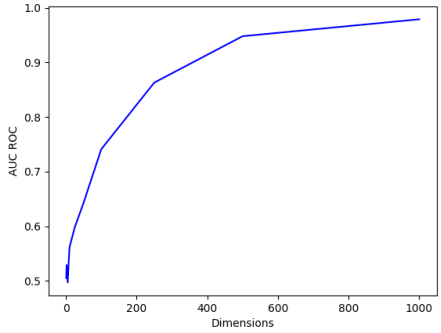


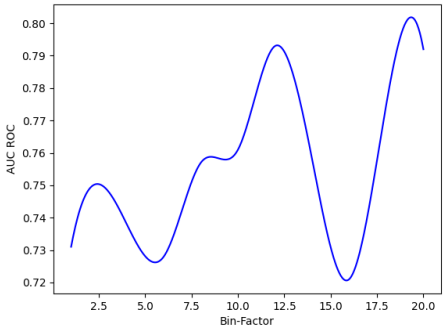


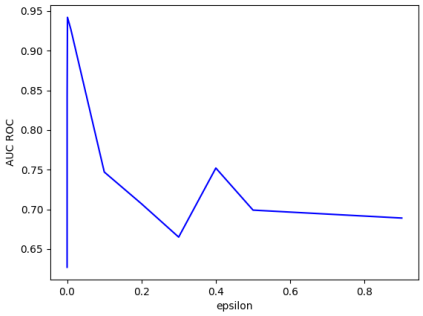


CIFAR-CIFAR-10 is a labelled subset of the 80 million tiny image dataset. It (originally) consists of 32x32 colour images representing 10 classes of objects: **Dimensions set to 1000 for graphs.**

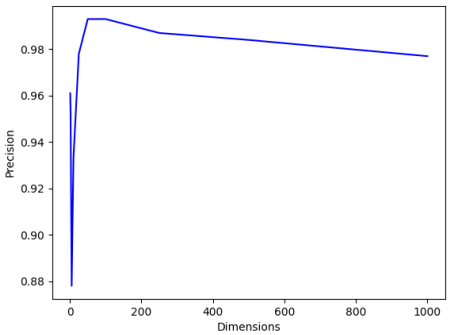
AUC

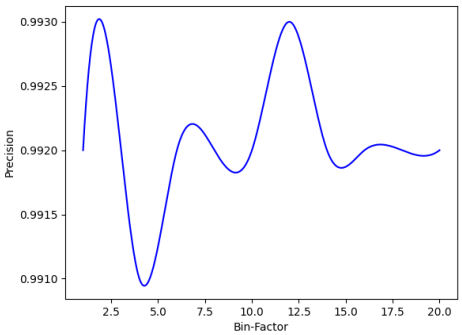


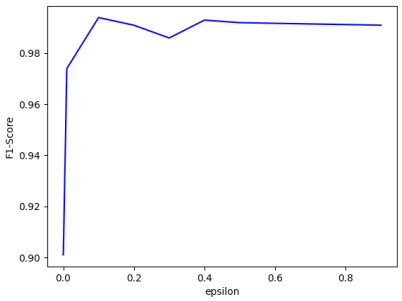




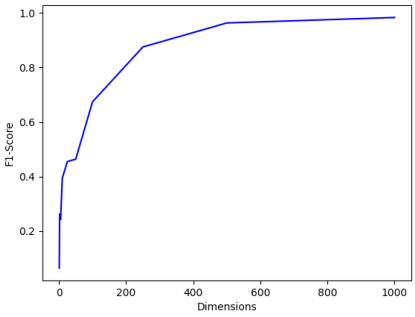
Precision

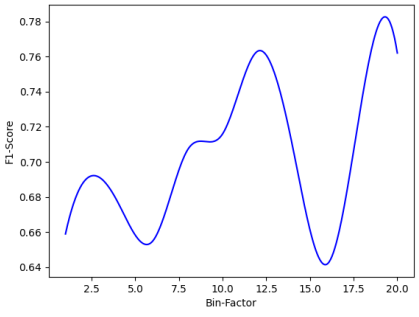


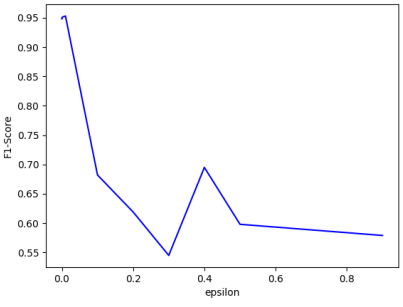




F1-Score

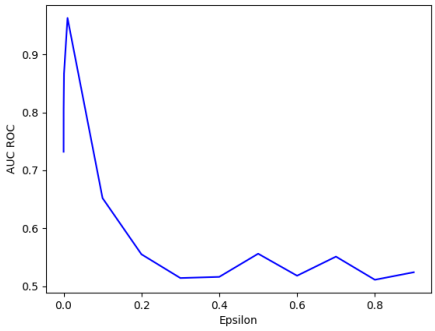


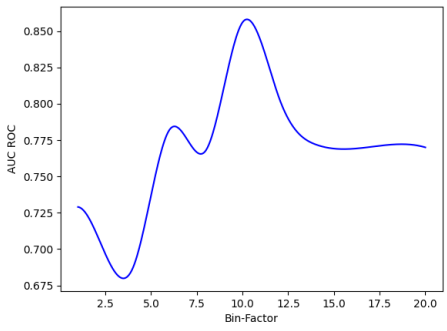


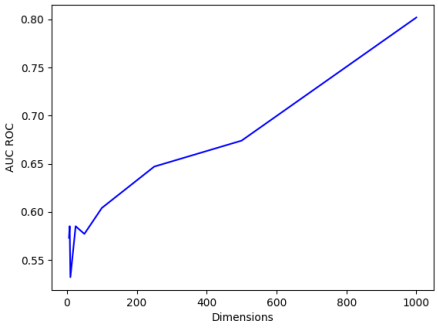


Tic-Tac-Toe - This database encodes the complete set of possible board configurations at the end of tic-tac-toe games, where "x" is assumed to have played first. The target concept is "win for x" (i.e., true when "x" has one of 8 possible ways to create a “three-in-a-row").

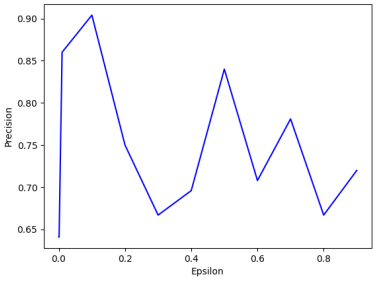
AUC

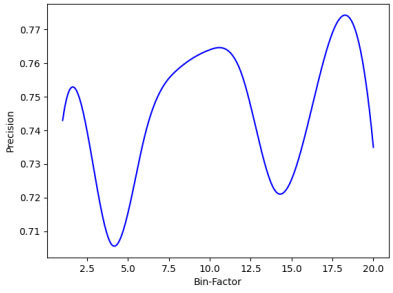


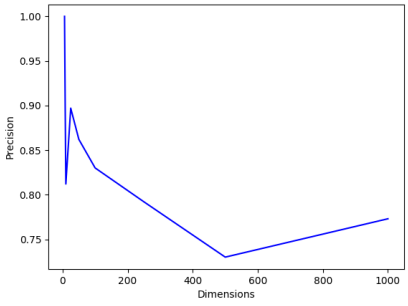




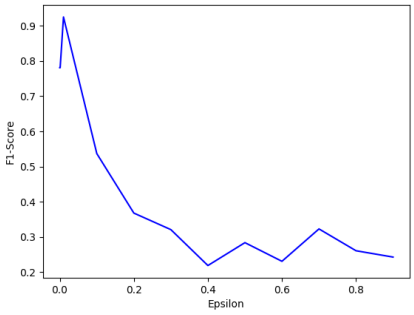
Precision

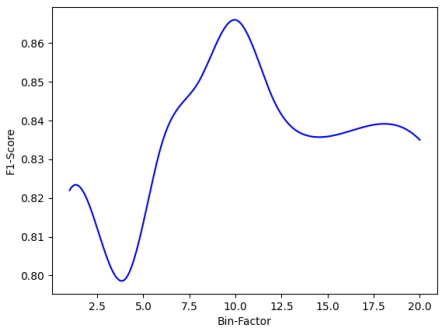


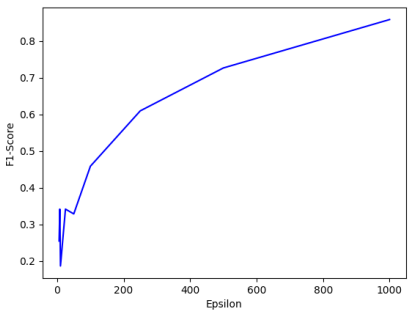




F1-Score



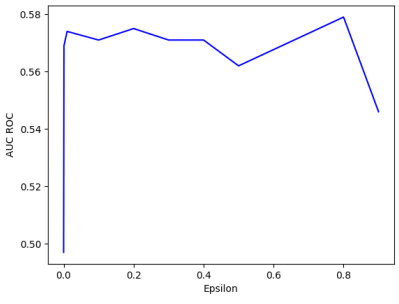


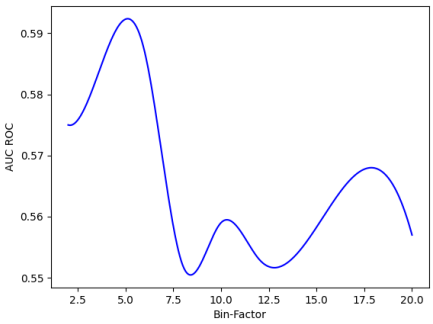


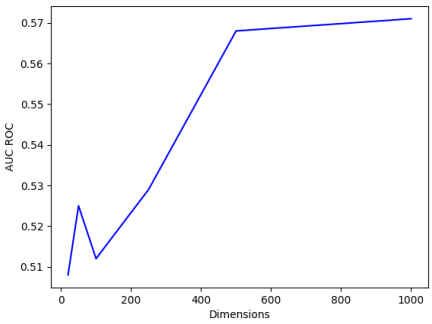
Dimension-1000

Blood-transfusion-service-center - The target attribute is a binary variable representing whether he/she donated blood in March 2007 (2 stands for donating blood; 1 stand for not donating blood).

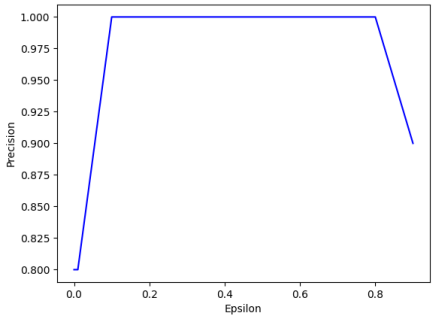
AUC

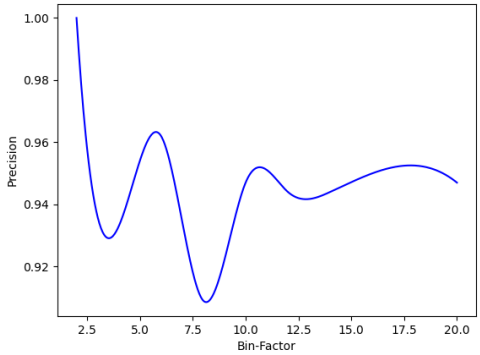


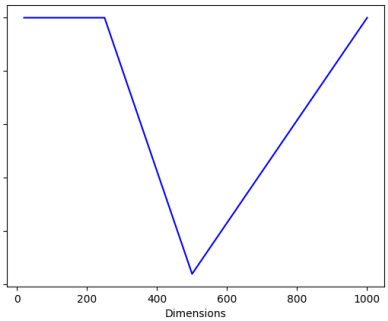




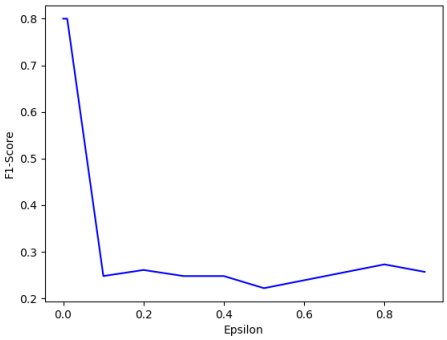
Precision

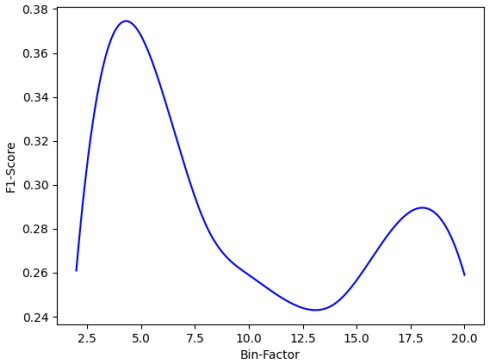


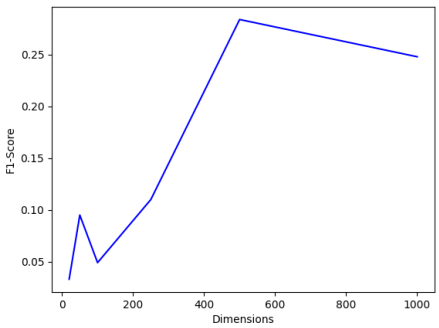




F1-Score

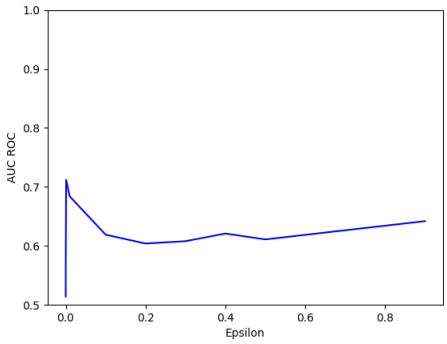


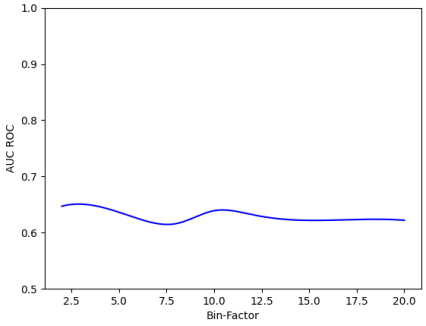


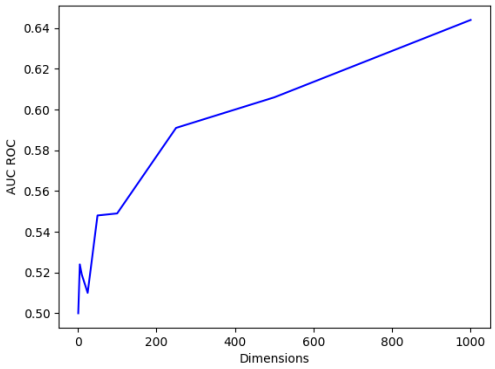


Jannis

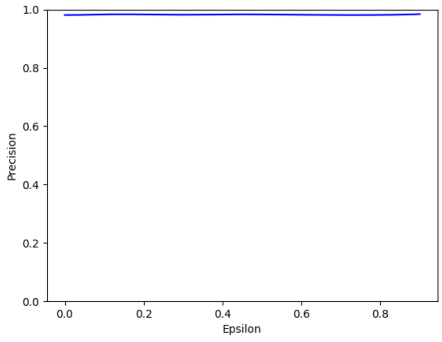
AUC

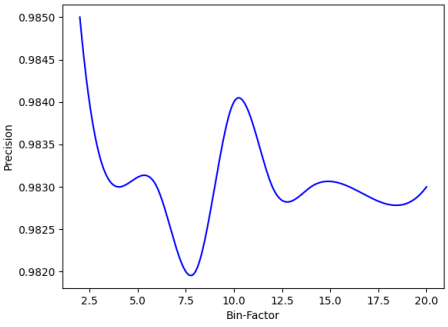


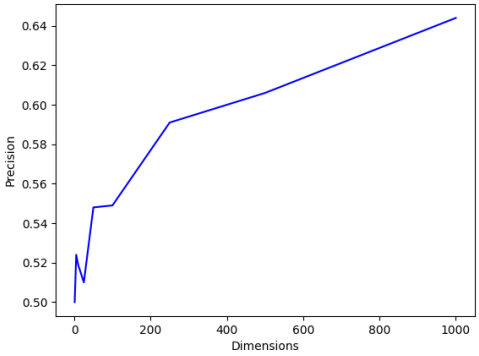




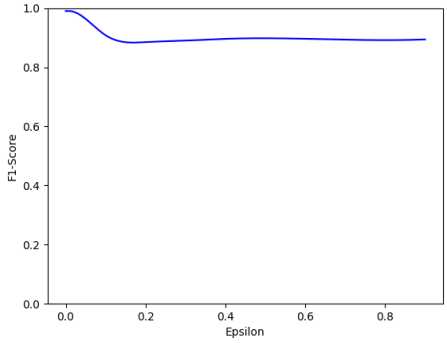
Precision

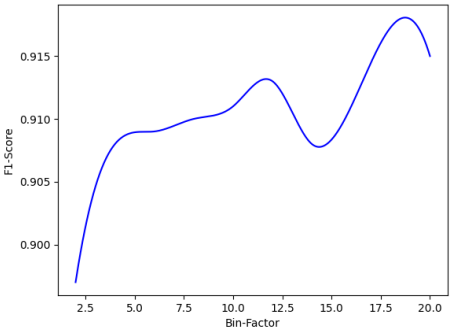


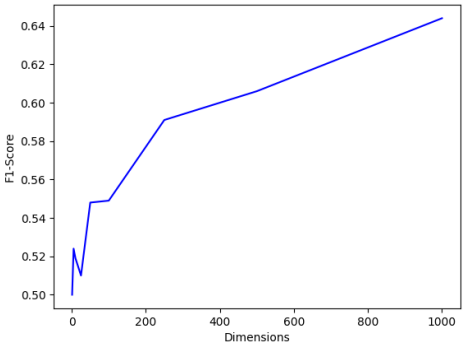




F1-Score



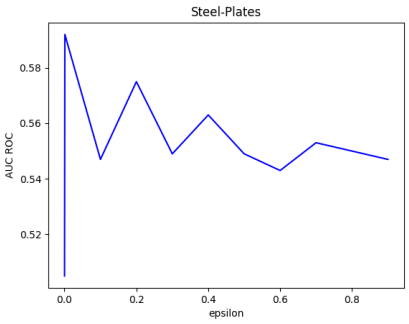


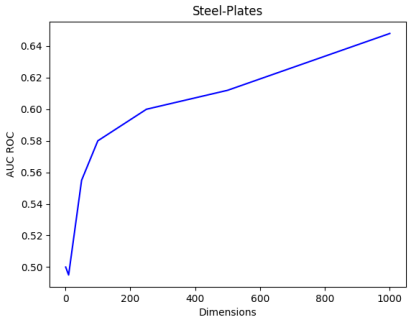


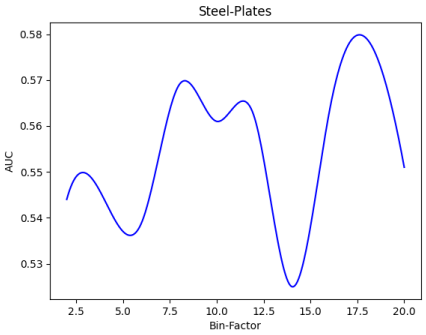
Steel-Plates-Fault

The dataset consists of 27 features describing each fault (location, size, ...) and 7 binary features indicating the type of fault (on of 7: Pastry, Z_Scratch, K_Scatch, Stains, Dirtiness, Bumps, Other_Faults). The latter is used as binary classification target ('common' or 'other' fault.)

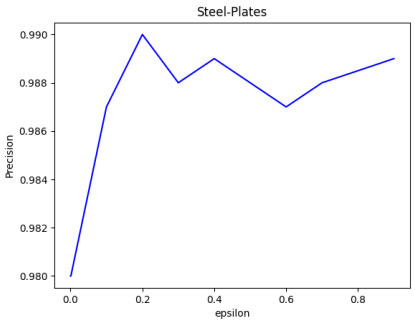
AUC

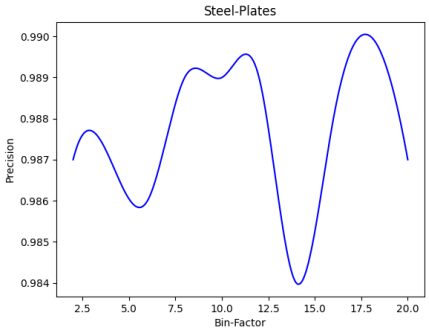


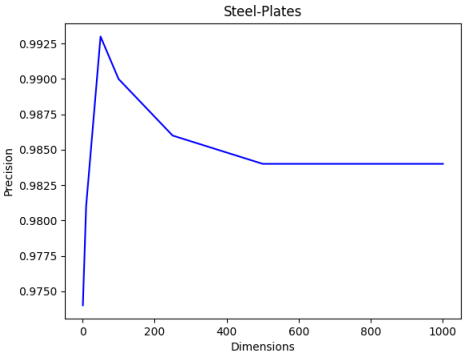




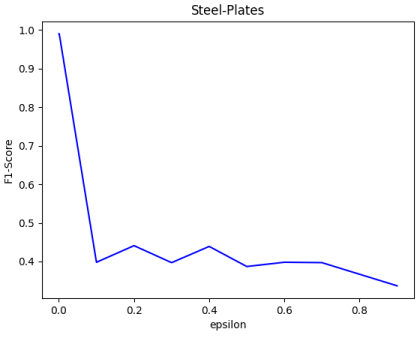
Precision

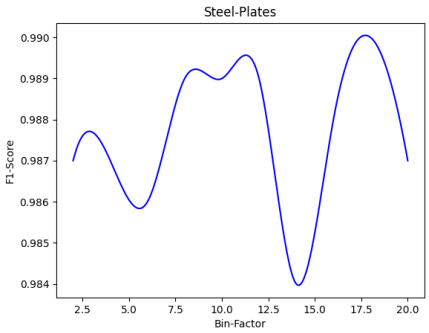


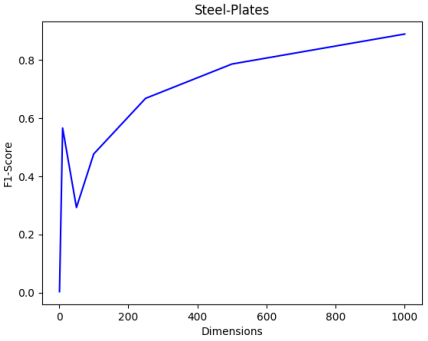




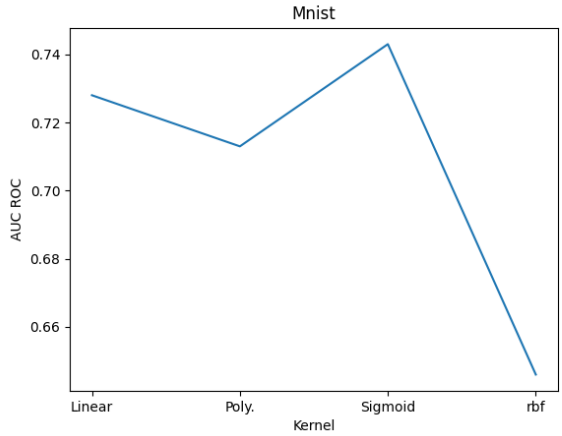
F1-Score

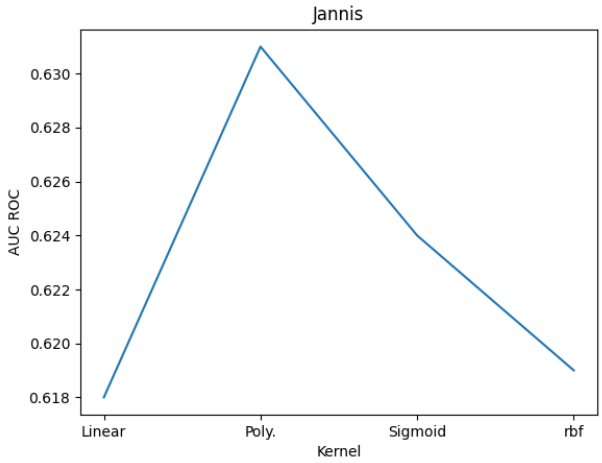


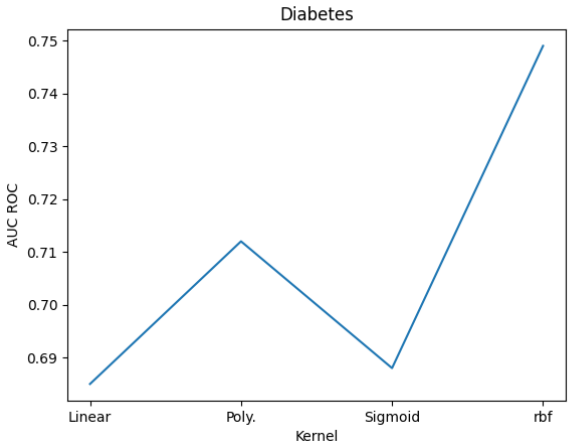


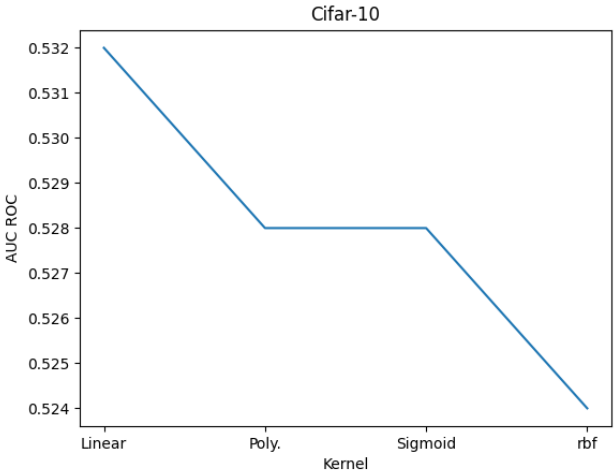


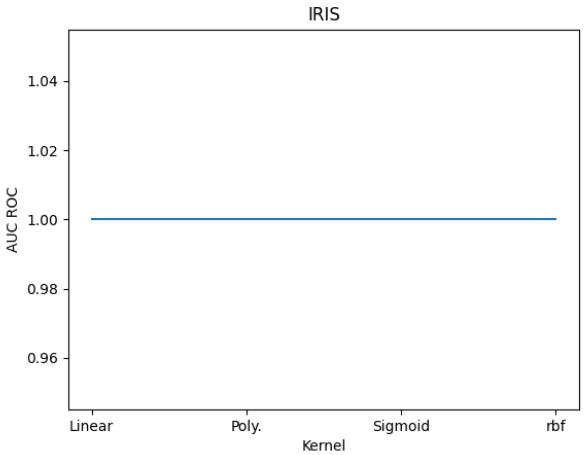
KERNELS

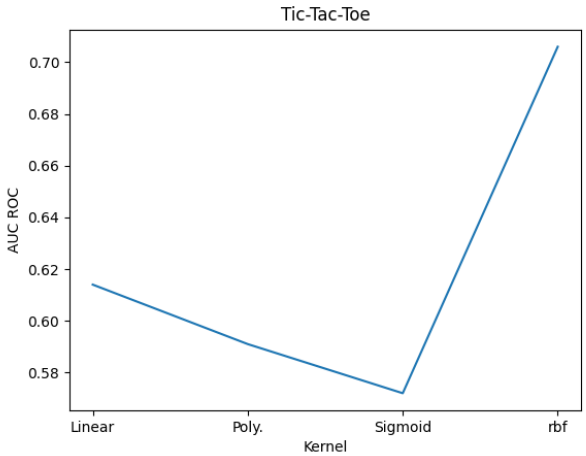

Mnist
AUC ROC
0.74
0.72
0.70
0.68
0.66
Linear
Poly.
Sigmoid
rbf
Kernel

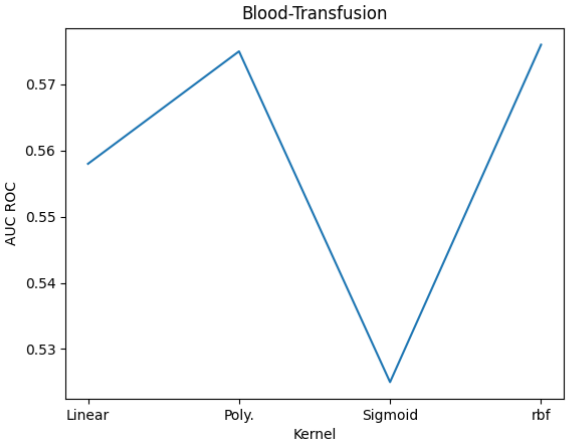

Jannis
AUC ROC
0.630
0.628
0.626
0.624
0.622
0.620
0.618
Linear
Poly.
Sigmoid
rbf
Kernel


Diabetes
AUC ROC
0.75
0.74
0.73
0.72
0.71
0.70
0.69
Linear
Poly.
Sigmoid
rbf
Kernel


Cifar-10
AUC ROC
0.532
0.531
0.530
0.529
0.528
0.527
0.526
0.525
0.524
Linear
Poly.
Sigmoid
rbf
Kernel


IRIS
AUC ROC
1.04
1.02
1.00
0.98
0.96
Linear
Poly.
Sigmoid
rbf
Kernel


Tic-Tac-Toe
AUC ROC
0.70
0.68
0.66
0.64
0.62
0.60
0.58
Linear
Poly.
Sigmoid
rbf
Kernel



Conclusions Drawn:

Epsilon: We observe that when the epsilon is very low the AUC ROC and Precision are low and then as epsilon increases the value increases and then after a certain value of epsilon it doesn't show much change.

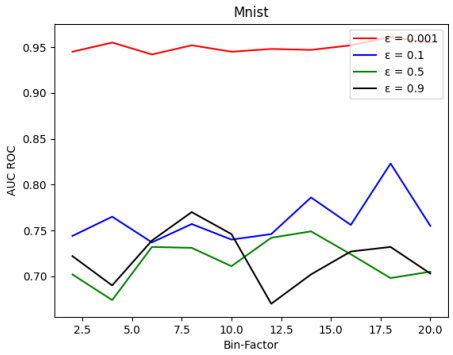
Dimensions - All three measures increase as the value of dimensions is increased.

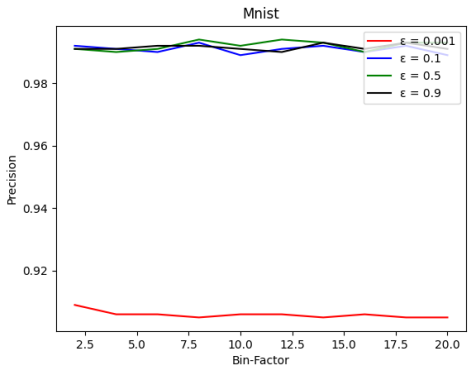
Bin-Factor - There is no correlation between bin-factor and performance as it keeps on increasing and decreasing randomly as the bin-factor is increased.

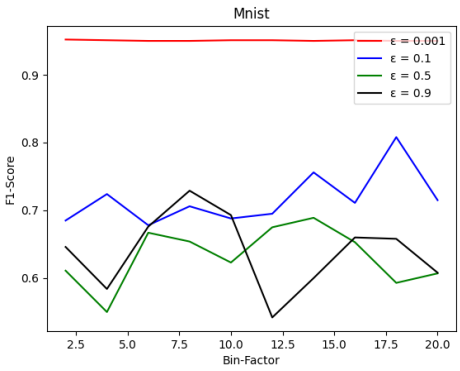
Kernel - Different kernel perform differently depending upon the dataset used.

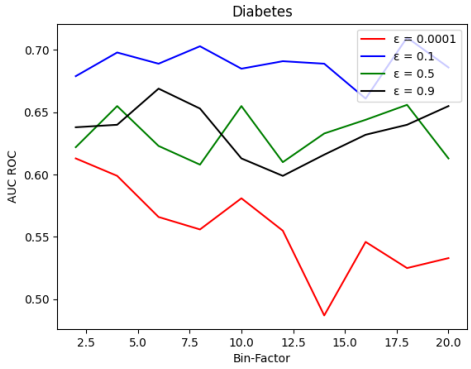
Bin Factor Analysis:

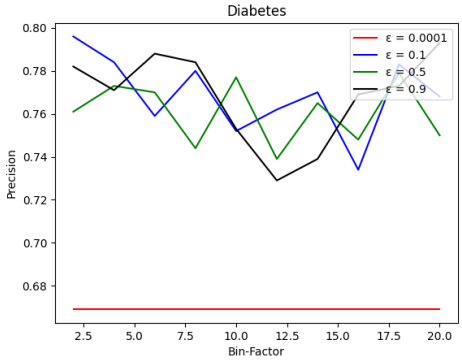
Parameters vs Bin-factor (for different epsilons)

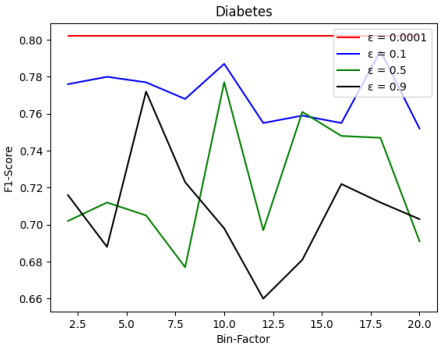


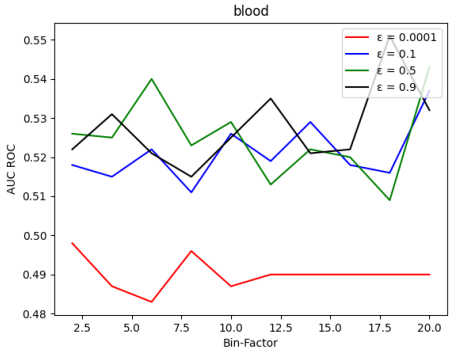


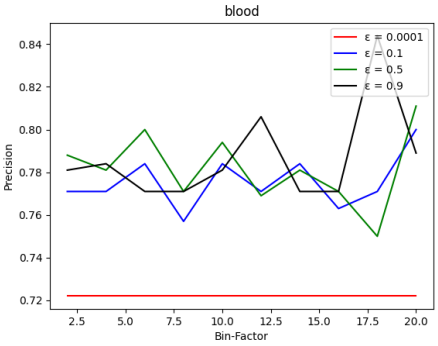


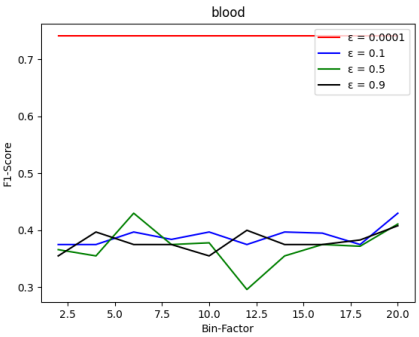


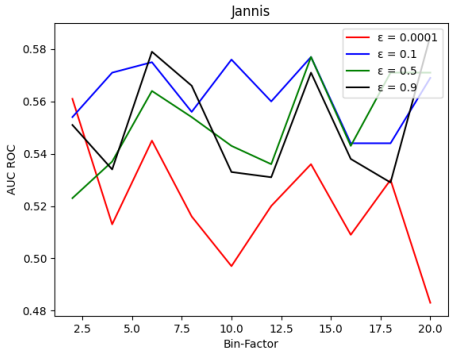


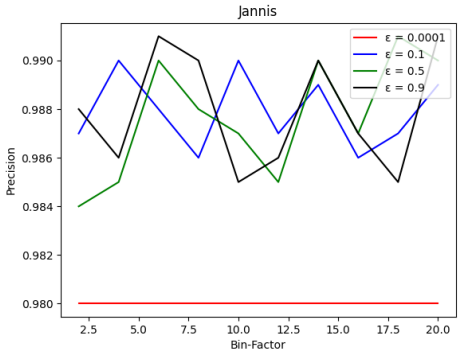


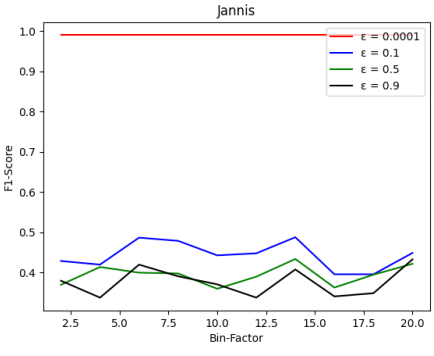


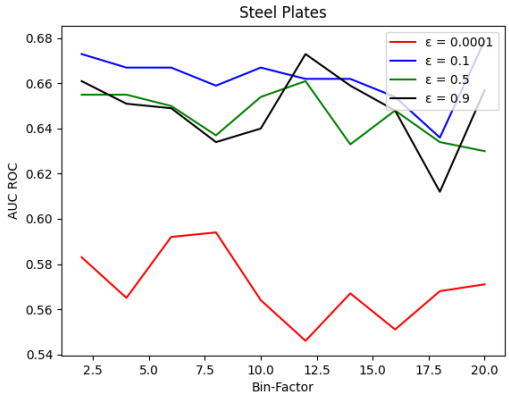


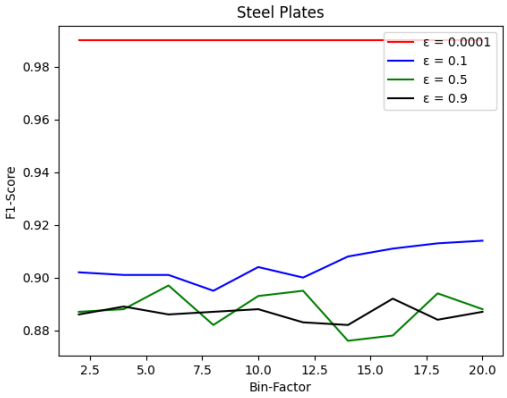








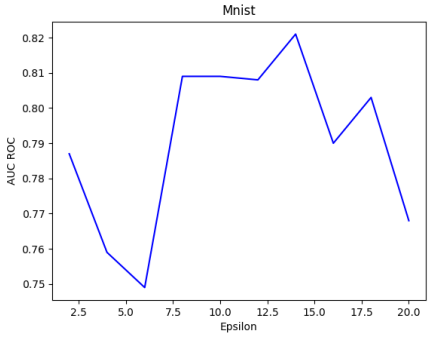


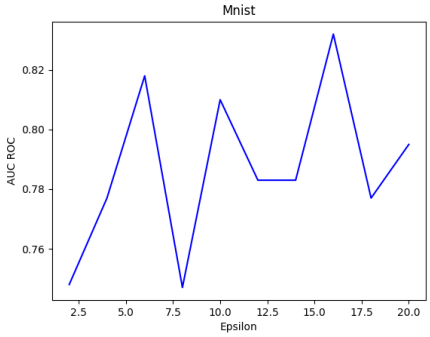


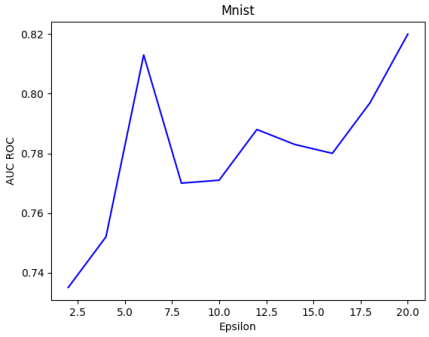
Conclusion:

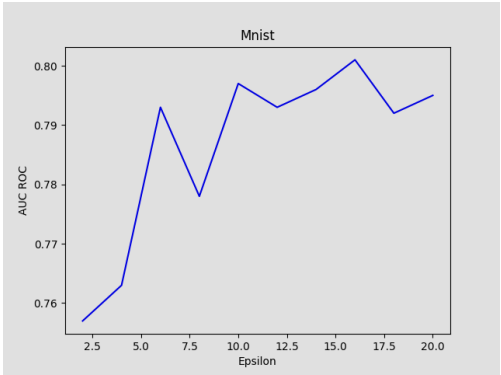
- Higher Epsilons(0.1,0.5,0.9) show maxima at high bin factor values whereas low epsilon (0.0001) show maxima for AUC-ROC at lower bin factor values.

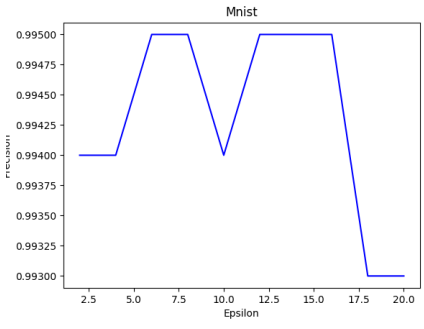
Mnist-784

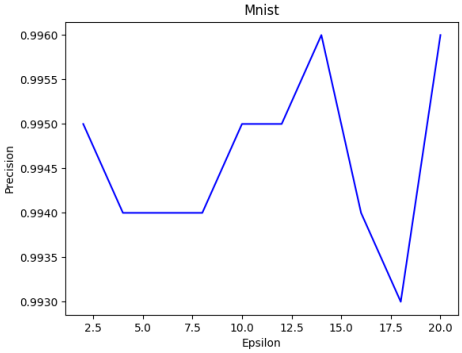

Mnist
AUC ROC
Epsilon

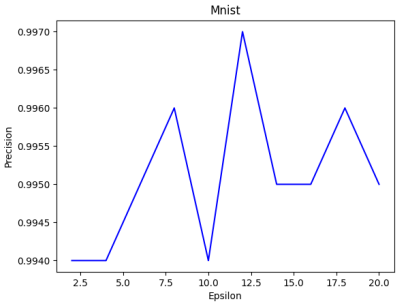

Mnist
AUC ROC
Epsilon

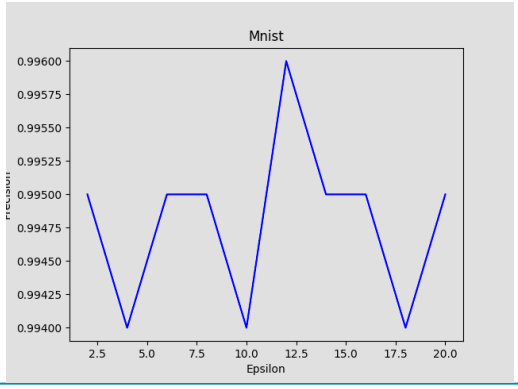

Mnist
AUC ROC
Epsilon

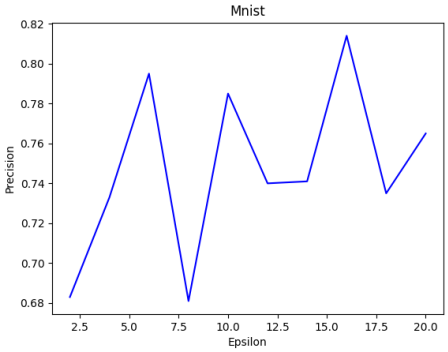

Mnist
AUC ROC
Epsilon

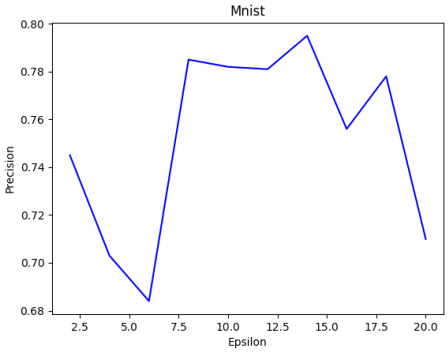

Mnist
Epsilon

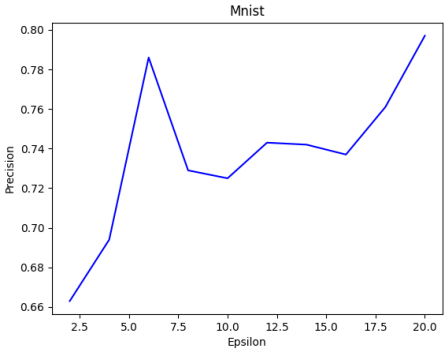

Mnist
Precision
Epsilon

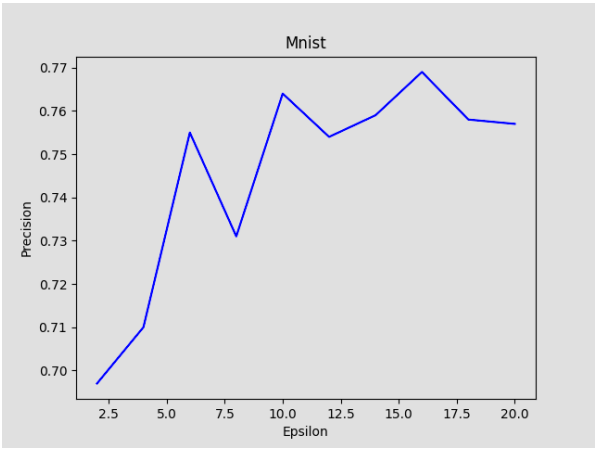

Mnist
Precision
Epsilon


Mnist
Epsilon

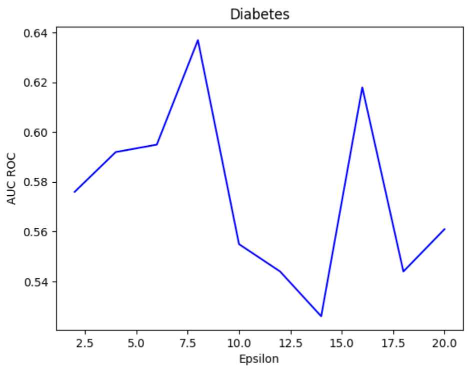

Mnist
Precision
Epsilon
0.82
0.80
0.78
0.76
0.74
0.72
0.70
0.68
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

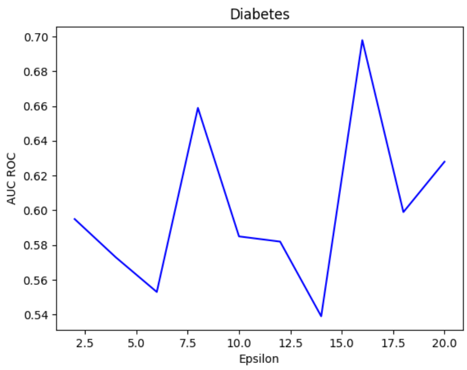

Mnist
Precision
Epsilon
0.80
0.78
0.76
0.74
0.72
0.70
0.68
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

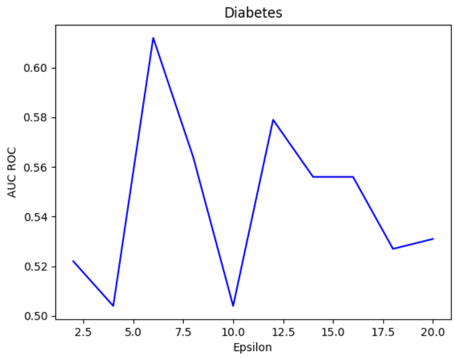

Mnist
Precision
Epsilon
0.80
0.78
0.76
0.74
0.72
0.70
0.68
0.66
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

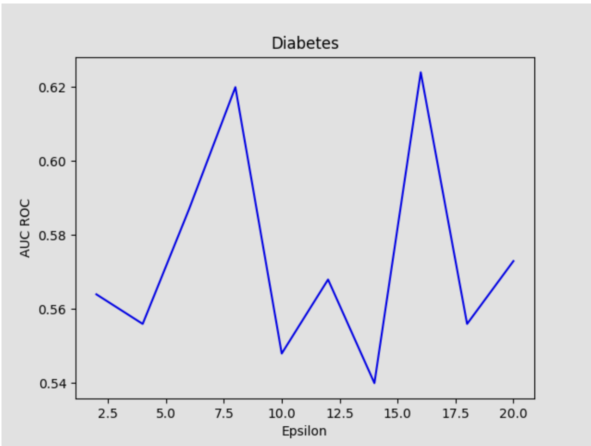

Mnist
Precision
Epsilon
0.77
0.76
0.75
0.74
0.73
0.72
0.71
0.70
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

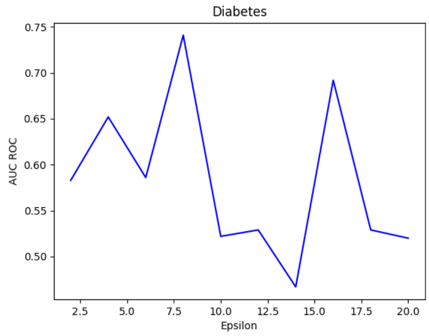
Diabetes

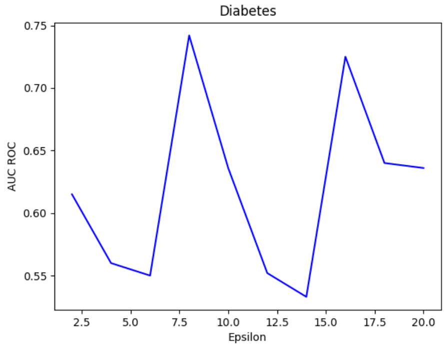

Diabetes
AUC ROC
0.64
0.62
0.60
0.58
0.56
0.54
2.5 5.0 7.5 10.0 12.5 15.0 17.5 20.0
Epsilon

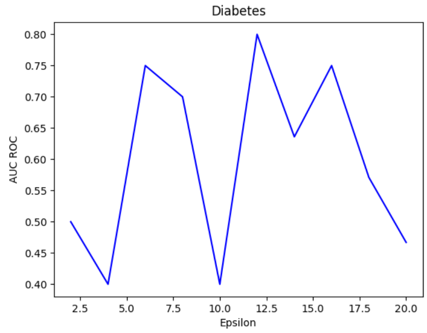

Diabetes
AUC ROC
0.70
0.68
0.66
0.64
0.62
0.60
0.58
0.56
0.54
2.5 5.0 7.5 10.0 12.5 15.0 17.5 20.0
Epsilon

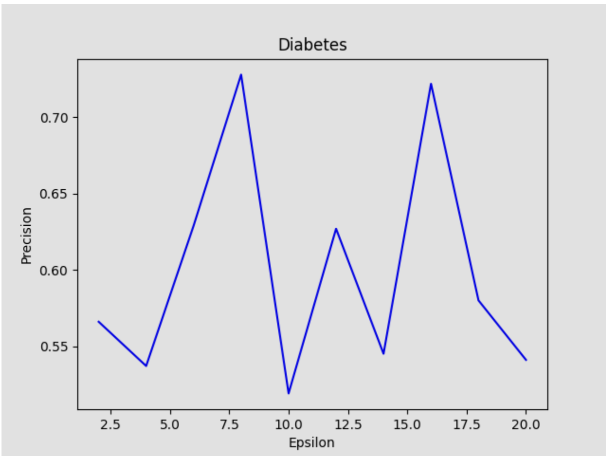

Diabetes
AUC ROC
0.60
0.58
0.56
0.54
0.52
0.50
2.5 5.0 7.5 10.0 12.5 15.0 17.5 20.0
Epsilon

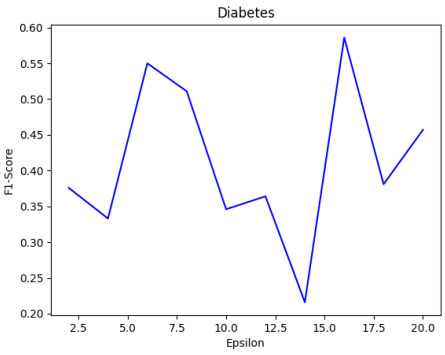

Diabetes
AUC ROC
0.62
0.60
0.58
0.56
0.54
2.5 5.0 7.5 10.0 12.5 15.0 17.5 20.0
Epsilon

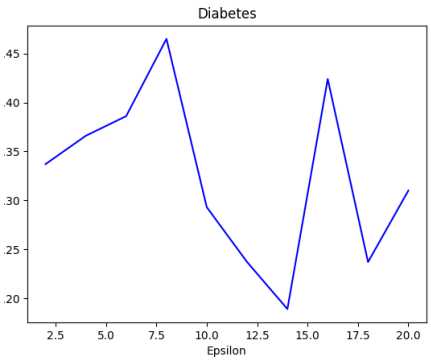

Diabetes
AUC ROC
0.75
0.70
0.65
0.60
0.55
0.50
2.5 5.0 7.5 10.0 12.5 15.0 17.5 20.0
Epsilon

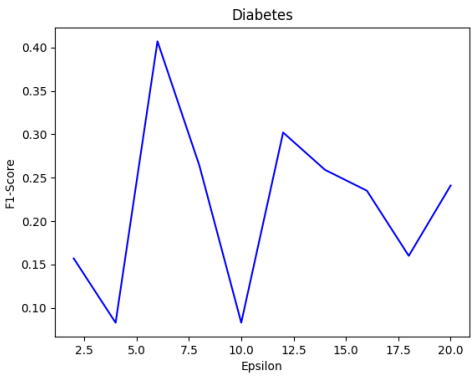

Diabetes
AUC ROC
0.75
0.70
0.65
0.60
0.55
2.5 5.0 7.5 10.0 12.5 15.0 17.5 20.0
Epsilon

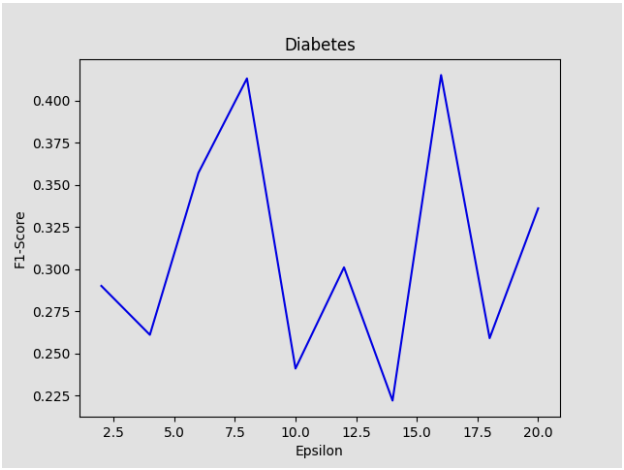

Diabetes
AUC ROC
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.45
0.40
2.5 5.0 7.5 10.0 12.5 15.0 17.5 20.0
Epsilon


Diabetes
Precision
0.70
0.65
0.60
0.55
2.5 5.0 7.5 10.0 12.5 15.0 17.5 20.0
Epsilon

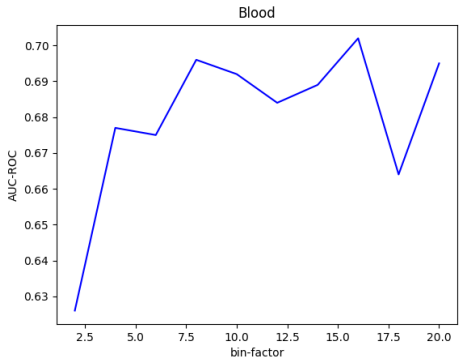

Diabetes
F1-Score
Epsilon

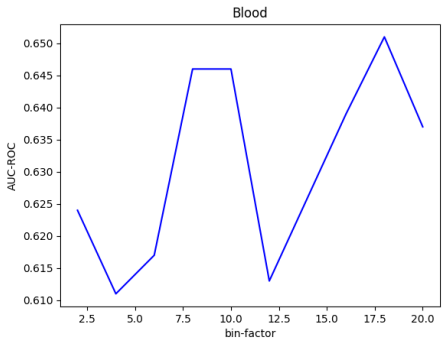

Diabetes
Epsilon

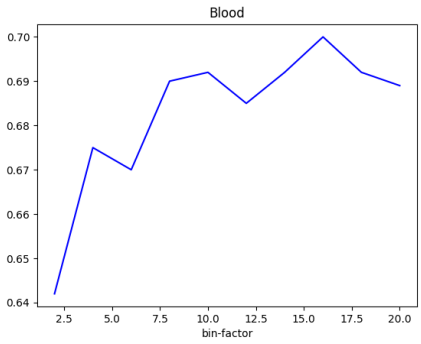

Diabetes
F1-Score
Epsilon

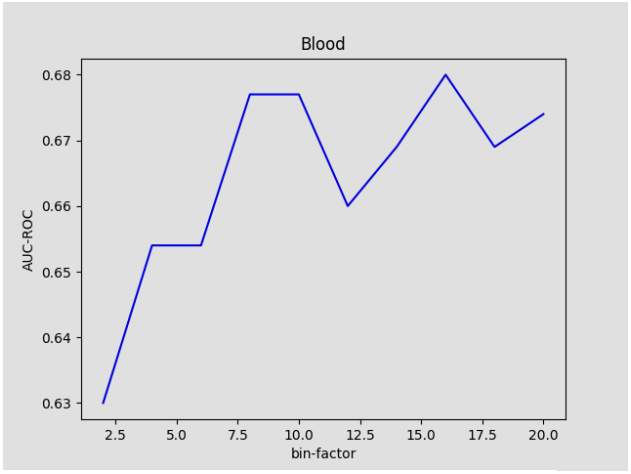

Diabetes
F1-Score
Epsilon

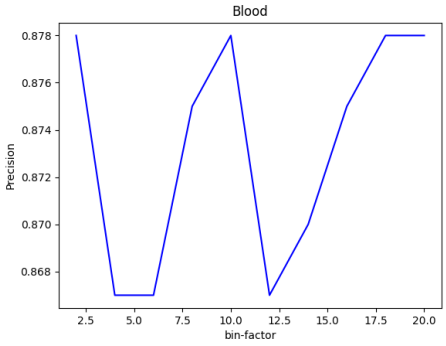
Blood-transfusion-service-center

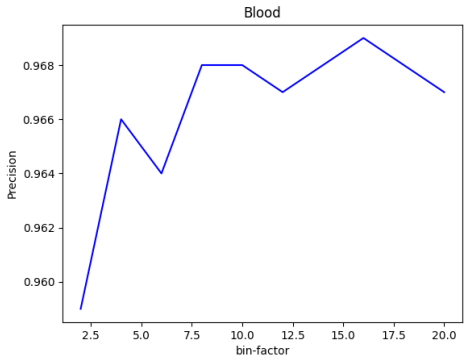


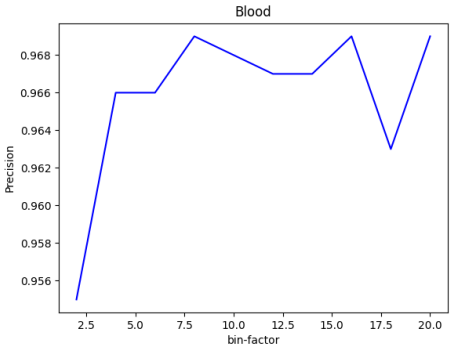


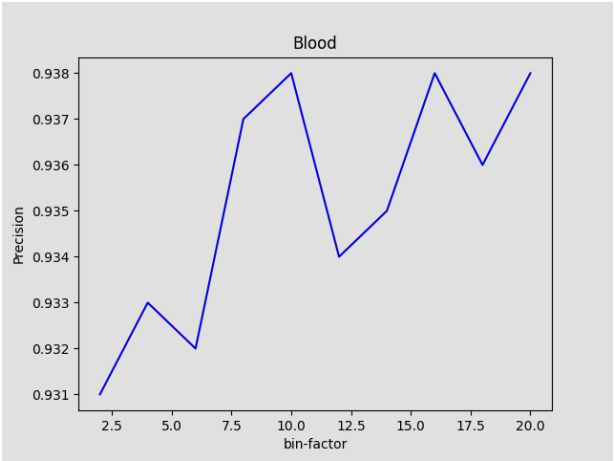


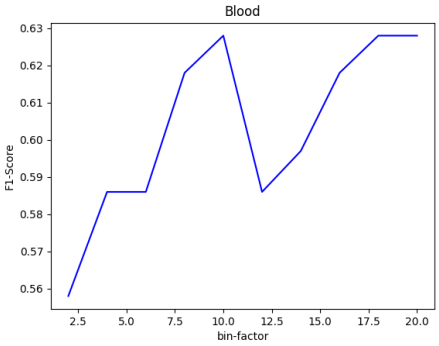


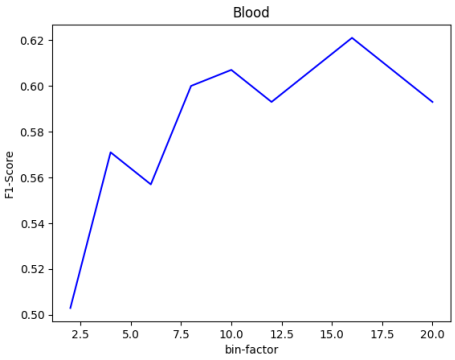


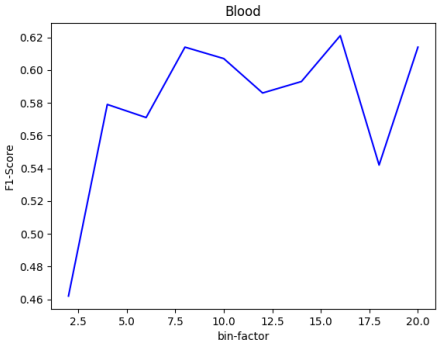


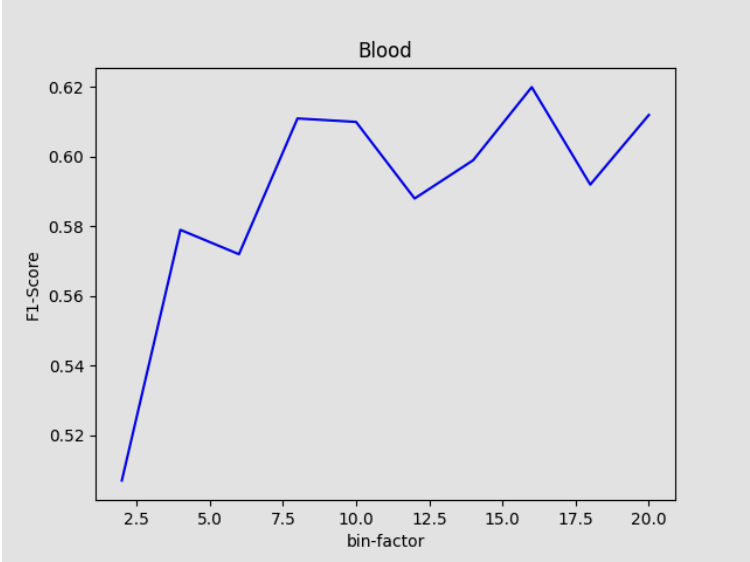




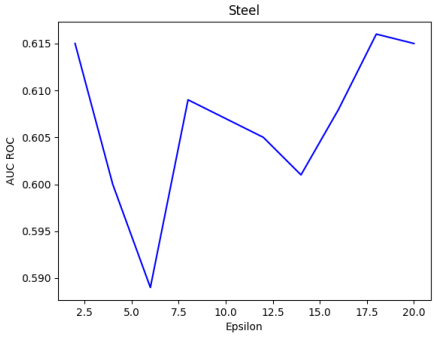

Blood
F1-Score
bin-factor

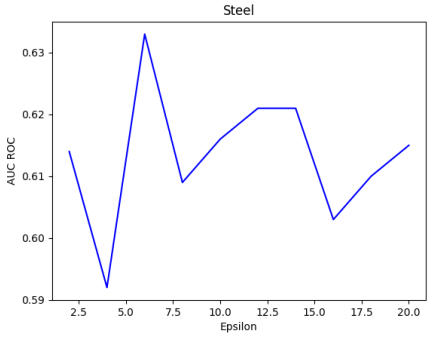

Blood
F1-Score
bin-factor

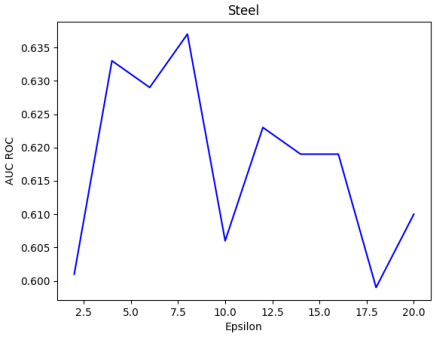

Blood
F1-Score
bin-factor

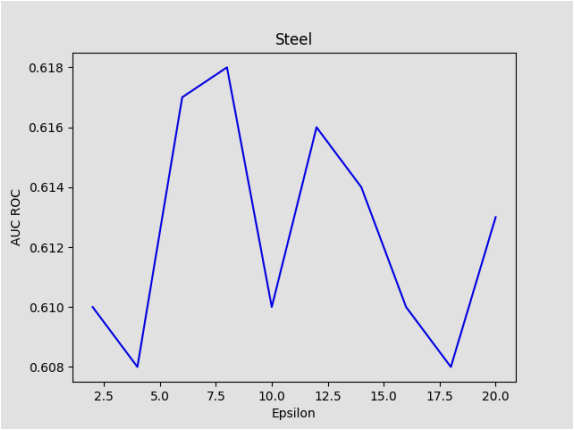

Blood
F1-Score
bin-factor

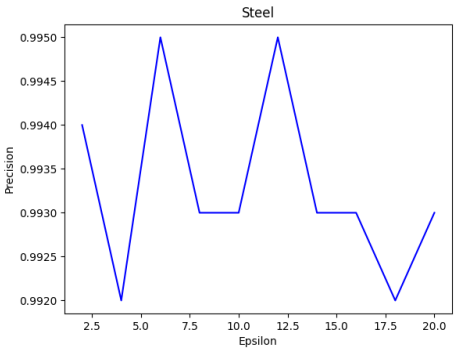
Steel-Plates

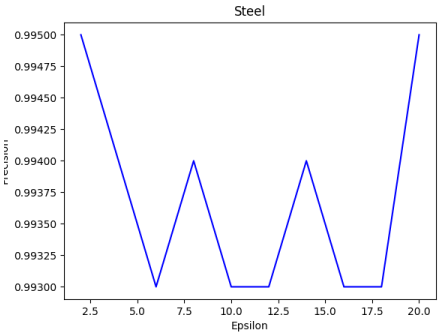


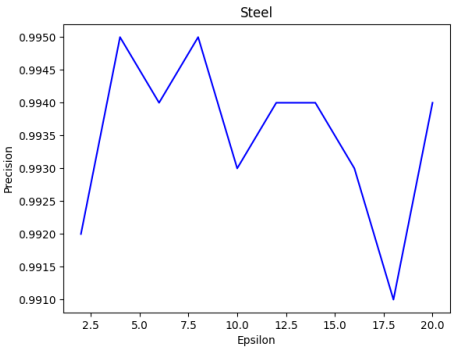


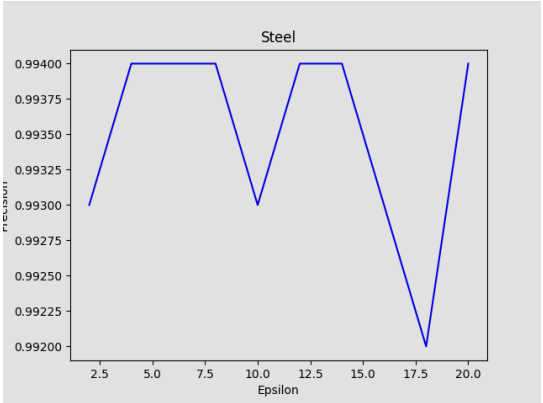


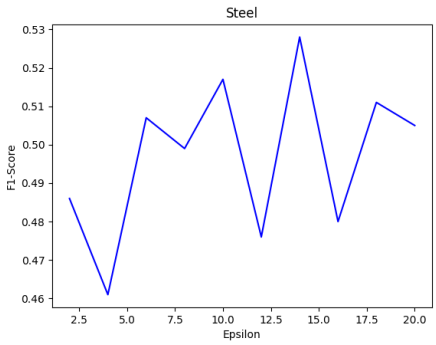


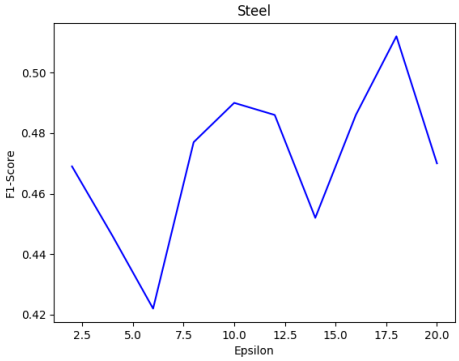


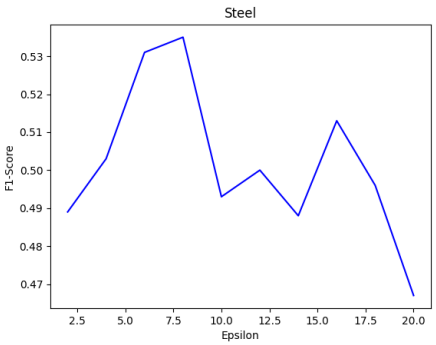


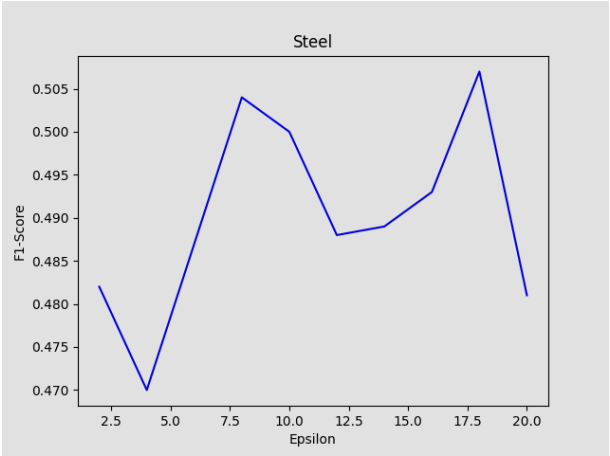




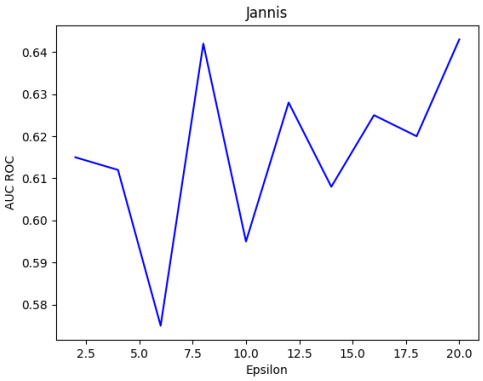

Steel
F1-Score
Epsilon

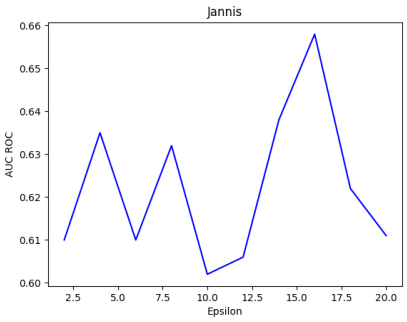

Steel
F1-Score
Epsilon

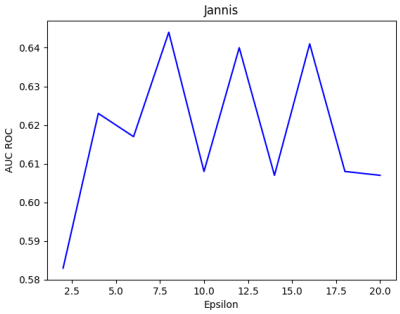

Steel
F1-Score
Epsilon

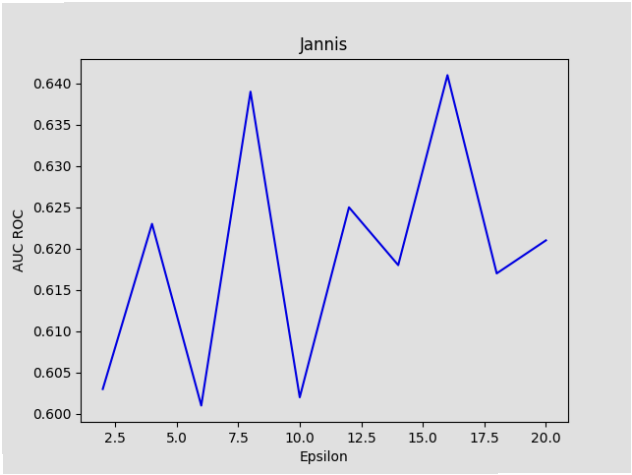

Steel
F1-Score
Epsilon

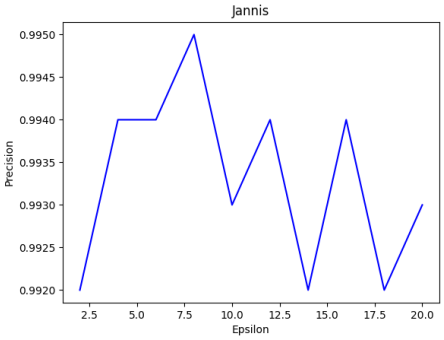
JANNIS

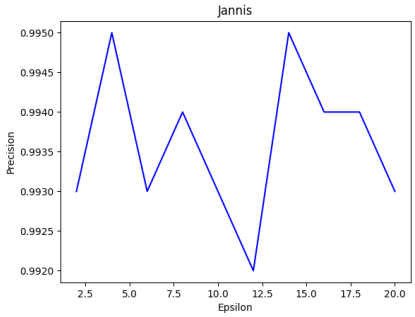

Jannis
AUC ROC
Epsilon

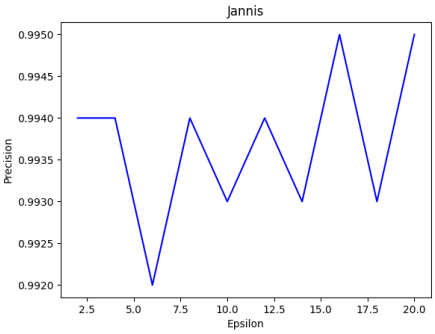

Jannis
AUC ROC
Epsilon

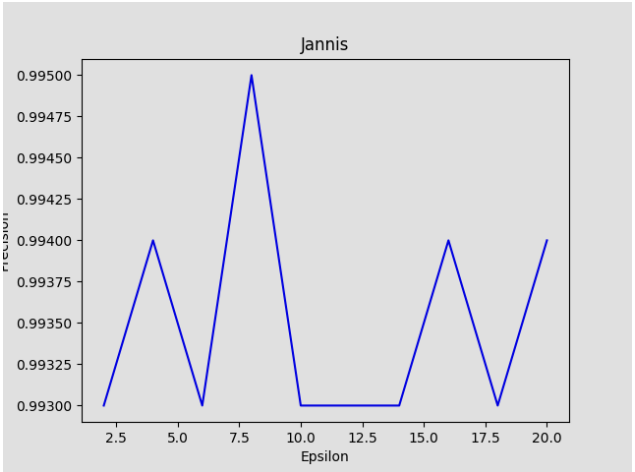

Jannis
AUC ROC
Epsilon

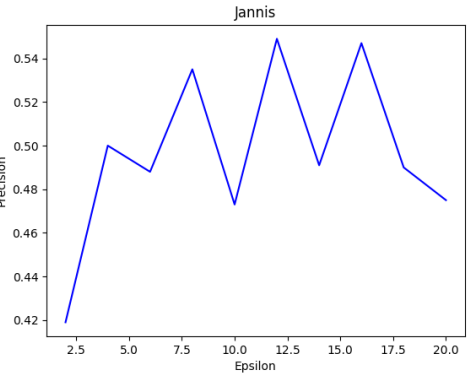

Jannis
AUC ROC
Epsilon

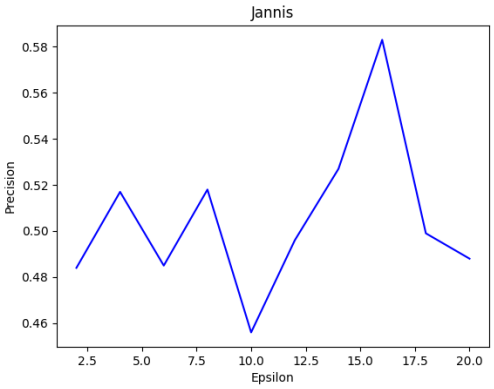

Jannis
Precision
Epsilon

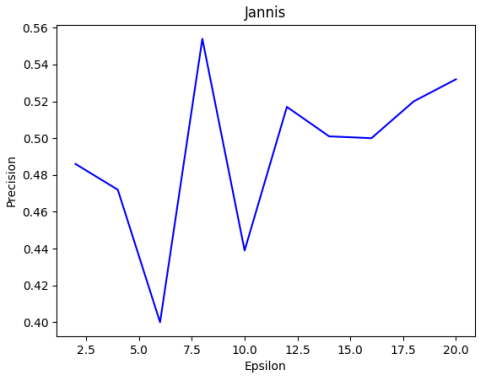

Jannis
Precision
Epsilon

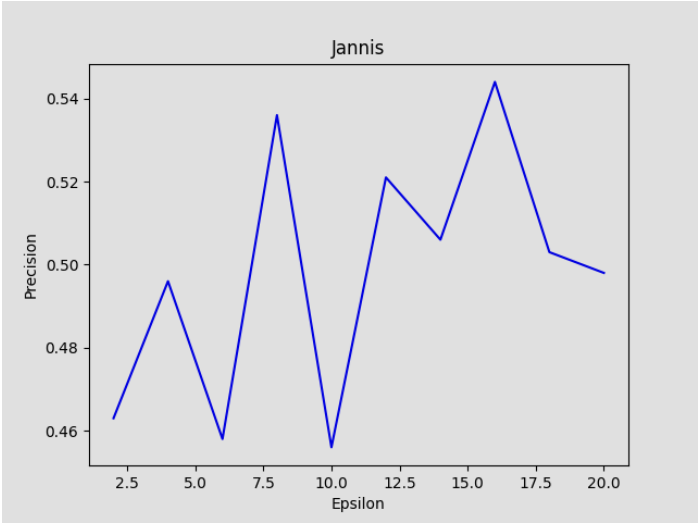

Jannis
Precision
Epsilon


Jannis
Epsilon


Jannis
Precision
Epsilon


Jannis
Precision
Epsilon


Jannis
Precision
Epsilon


Jannis
Precision
Epsilon

Conclusions:

- The three parameters have local minima and maxima at similar values of Bin-Factor generally