Weekly Study Report

Wang Ma

2024-09-20

Electrical, Computer, and Systems Engineering Department
Rensselaer Polytechnic Institute

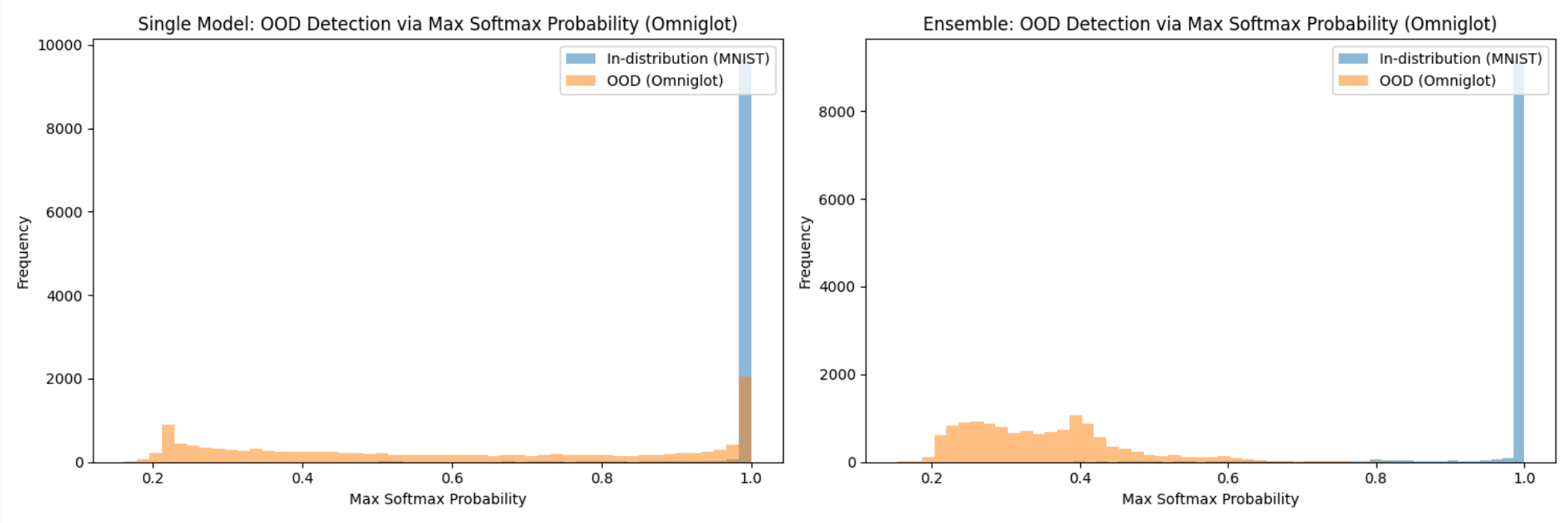
Outline

1. Hands-on: UQ via Deep Ensemble

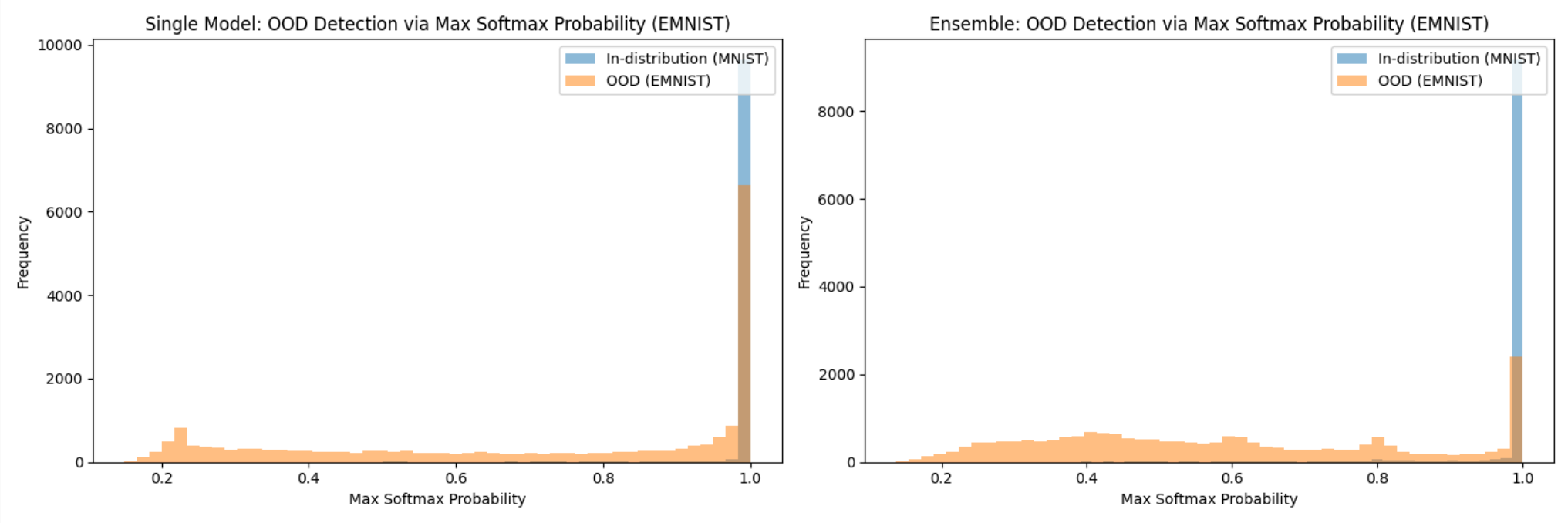
1.1 Problem Encountered / My Farward Progress

- The original paper was published in 2016, so the corresponding code was implemented in Tensorflow. So I **rewrite** a new version of code in Pytorch for classification task on MNIST.
- Trying to be familiar with the server and Linux environment (**configuring Remote Servers**).

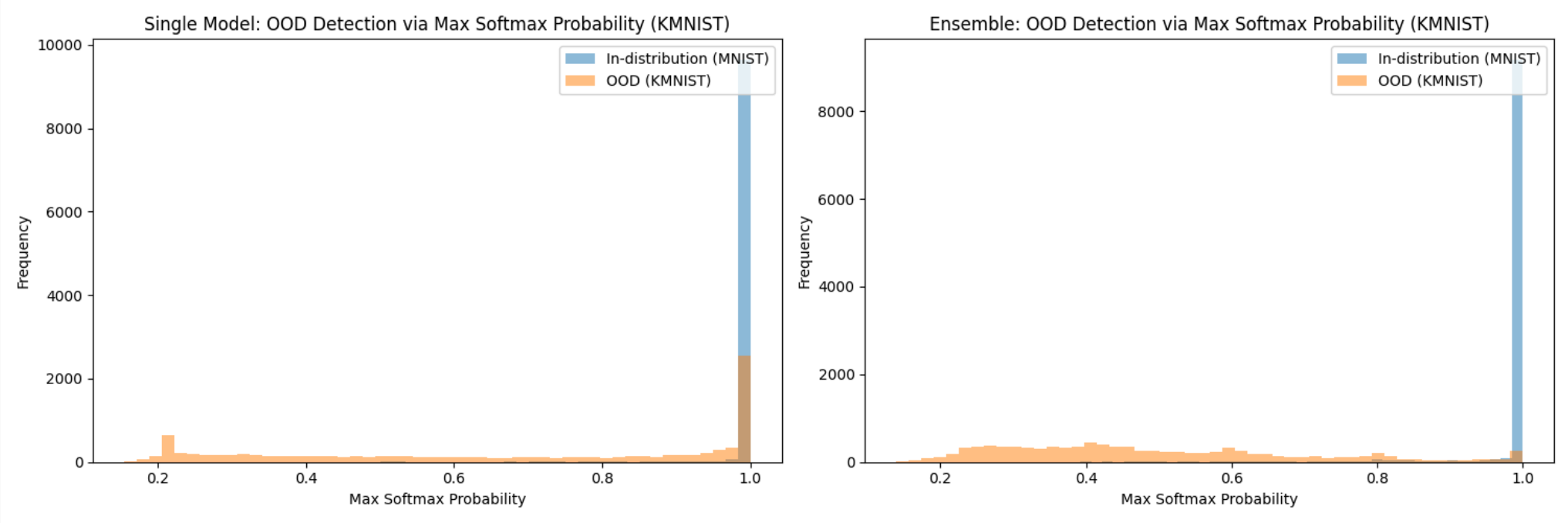
1.2.1 Max Probability on Omniglot



1.2.2 Max Probability on EMNIST



1.2.3 Max Probability on KMNIST



Deep Ensemble Model result

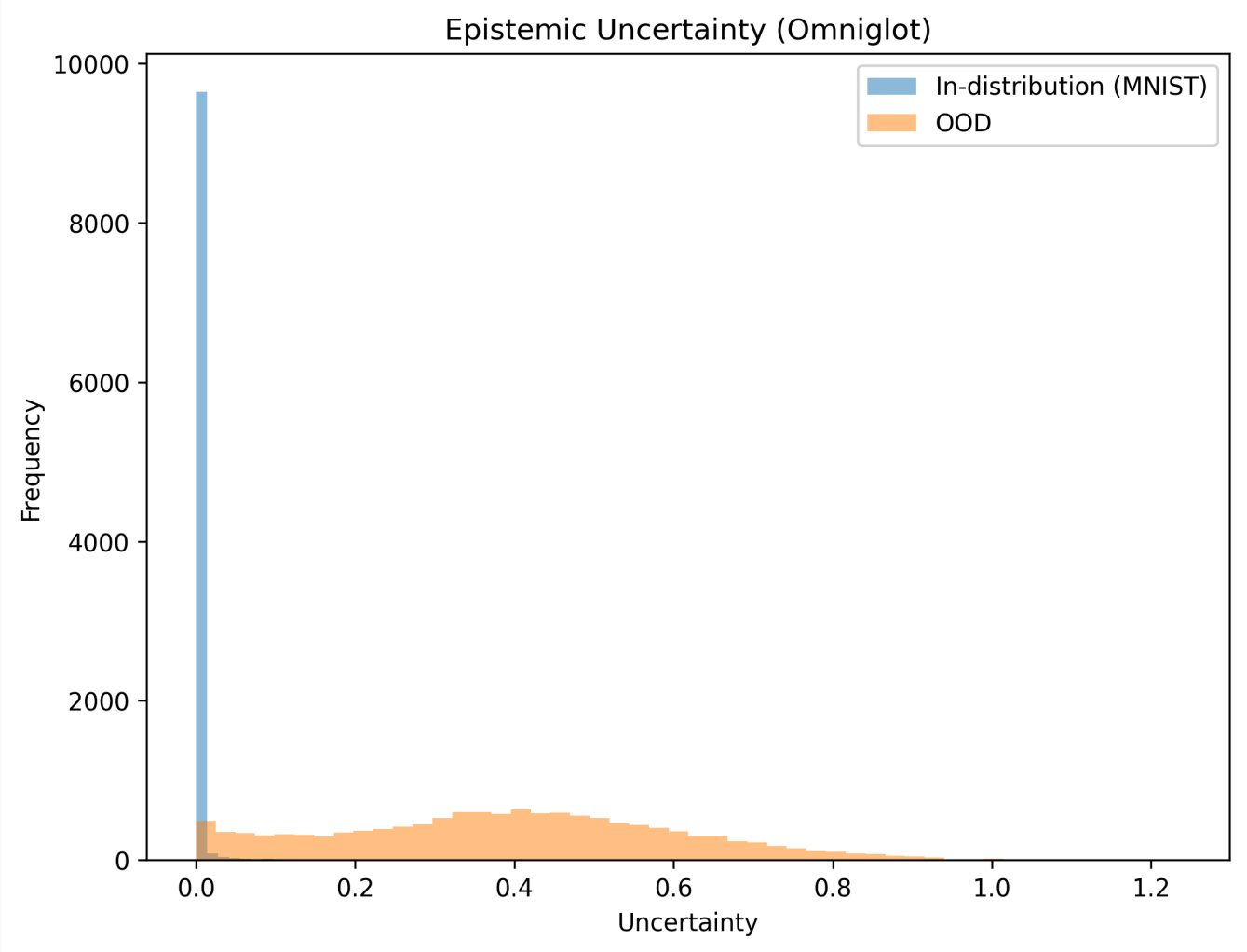
	AUROC	AUPR
Omniglot	0.9931	0.9934
EMNIST	0.9505	0.9744
KMNIST	0.9843	0.9822

Reference result

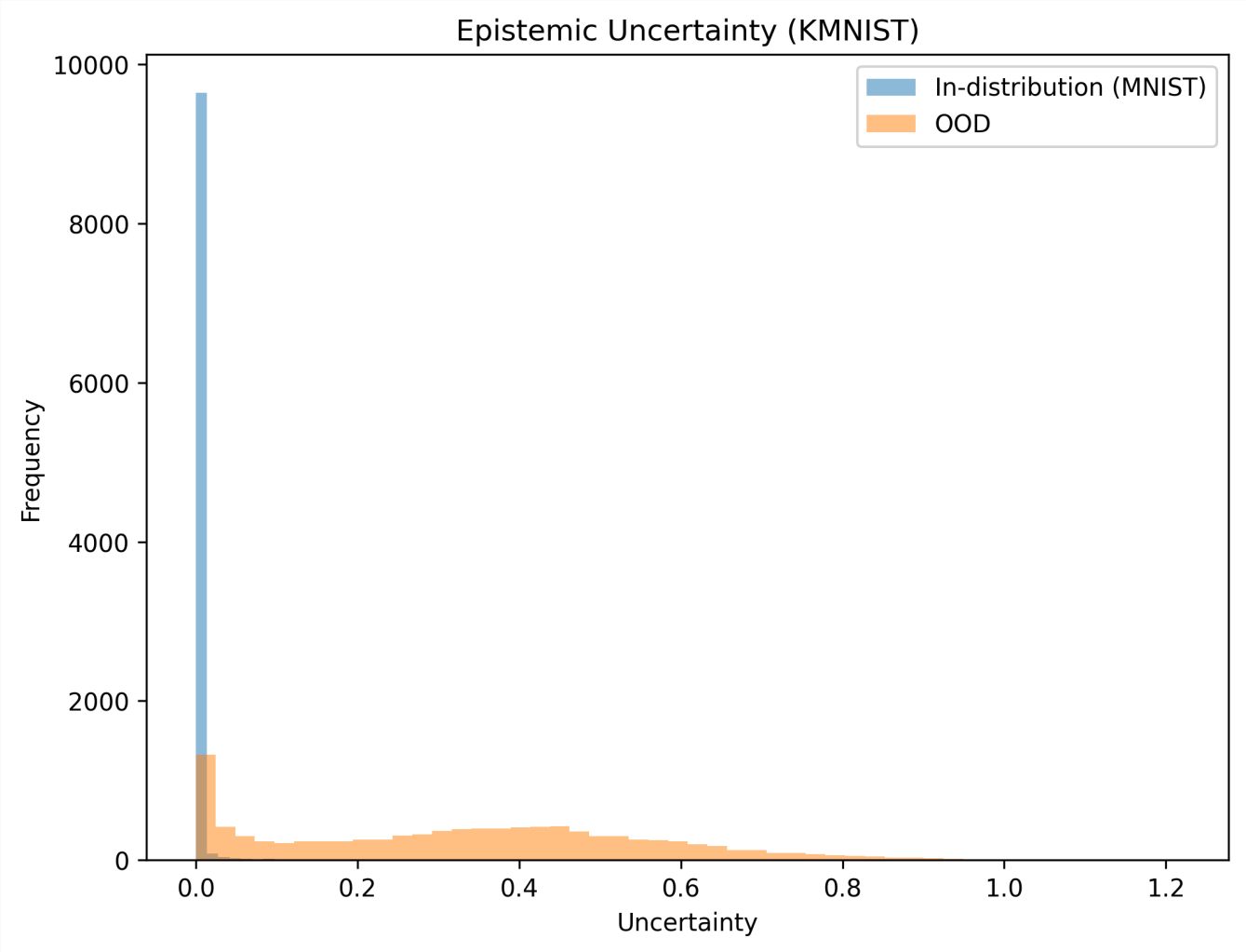
	AUROC	AUPR
Omniglot	0.9792	0.9733
EMNIST	0.9732	0.9610
KMNIST	0.9792	0.9713

1.2.4 Uncertainty Vs Dataset

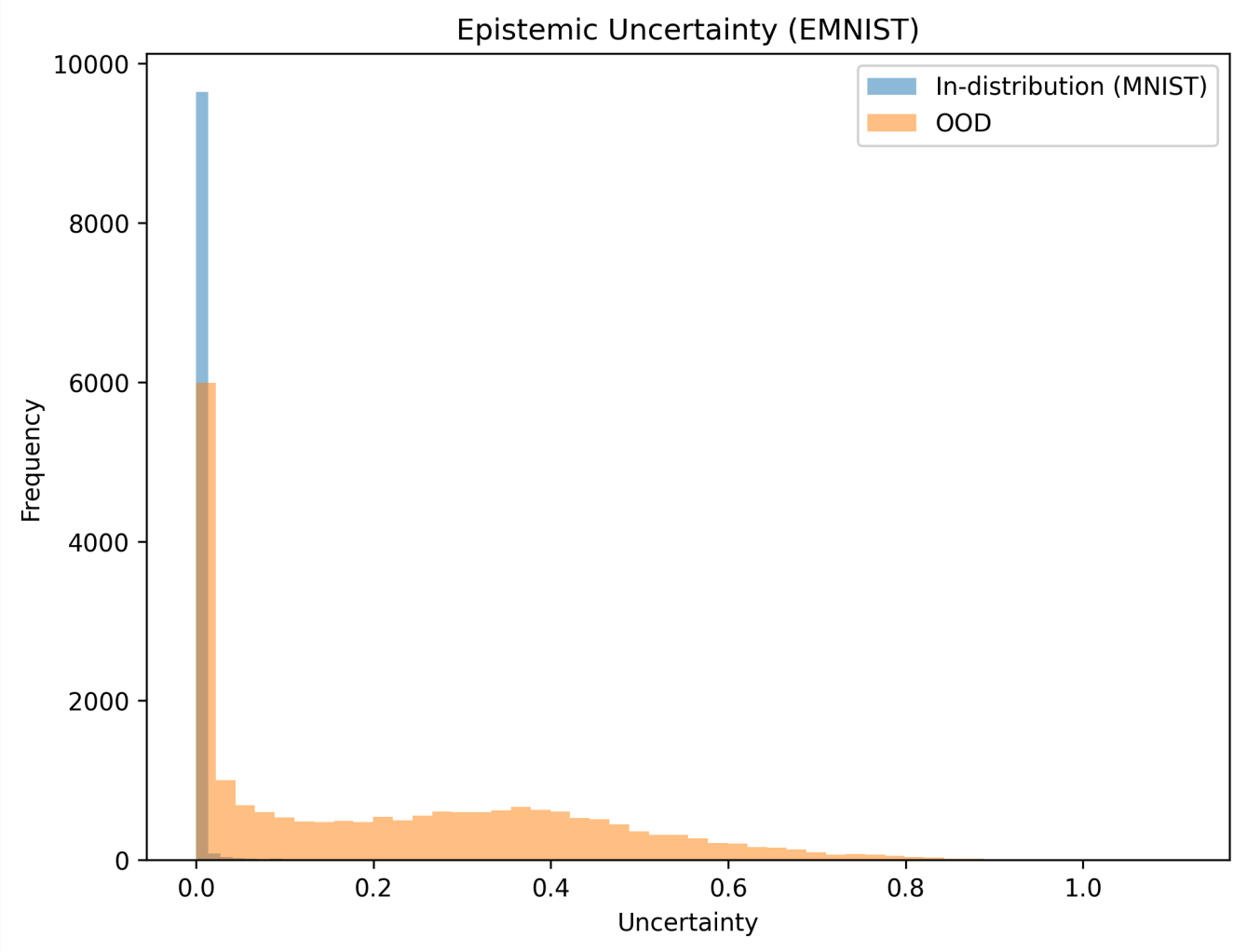
1.2 OOD Test


Epistemic Uncertainty (Omniglot)
In-distribution (MNIST)
OOD
Frequency
Uncertainty
10000
8000
6000
4000
2000
0
0.0
0.2
0.4
0.6
0.8
1.0
1.2

1.2 OOD Test


Epistemic Uncertainty (KMNIST)
In-distribution (MNIST)
OOD
Frequency
Uncertainty
10000
8000
6000
4000
2000
0
0.0
0.2
0.4
0.6
0.8
1.0
1.2

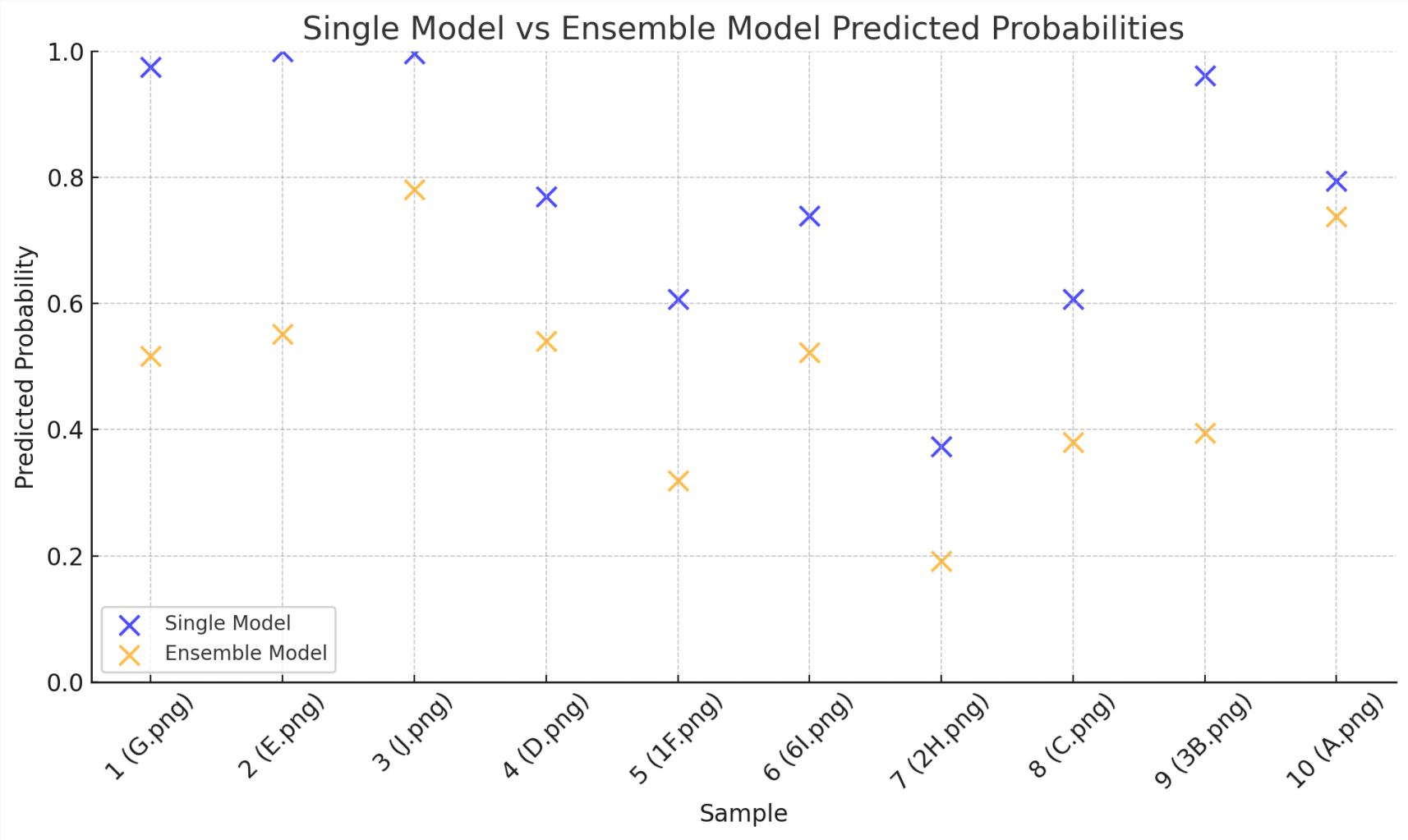
1.2 OOD Test


Epistemic Uncertainty (EMNIST)
In-distribution (MNIST)
OOD
Frequency
Uncertainty
10000
8000
6000
4000
2000
0
0.0
0.2
0.4
0.6
0.8
1.0

1.3 Single Image Test

Sample	Single Model Prediction	Single Model Probability	Ensemble Model Prediction	Ensemble Model Probability
G.png	3	0.9749	3	0.5169
E.png	5	1.0000	8	0.5515
J.png	1	0.9964	1	0.7802
D.png	4	0.7699	4	0.5406
1F.png	4	0.6075	6	0.3192
6I.png	1	0.7395	1	0.5227
2H.png	9	0.3729	8	0.1919
C.png	4	0.6074	1	0.3803
3B.png	4	0.9609	4	0.3950
A.png	4	0.7945	4	0.7379

1.3 Single Image Test



2. Hands-on: UQ via MC-Dropout

2.1 Key Points & Progress

- Drop-out during Inference

MC-DropOut Model result

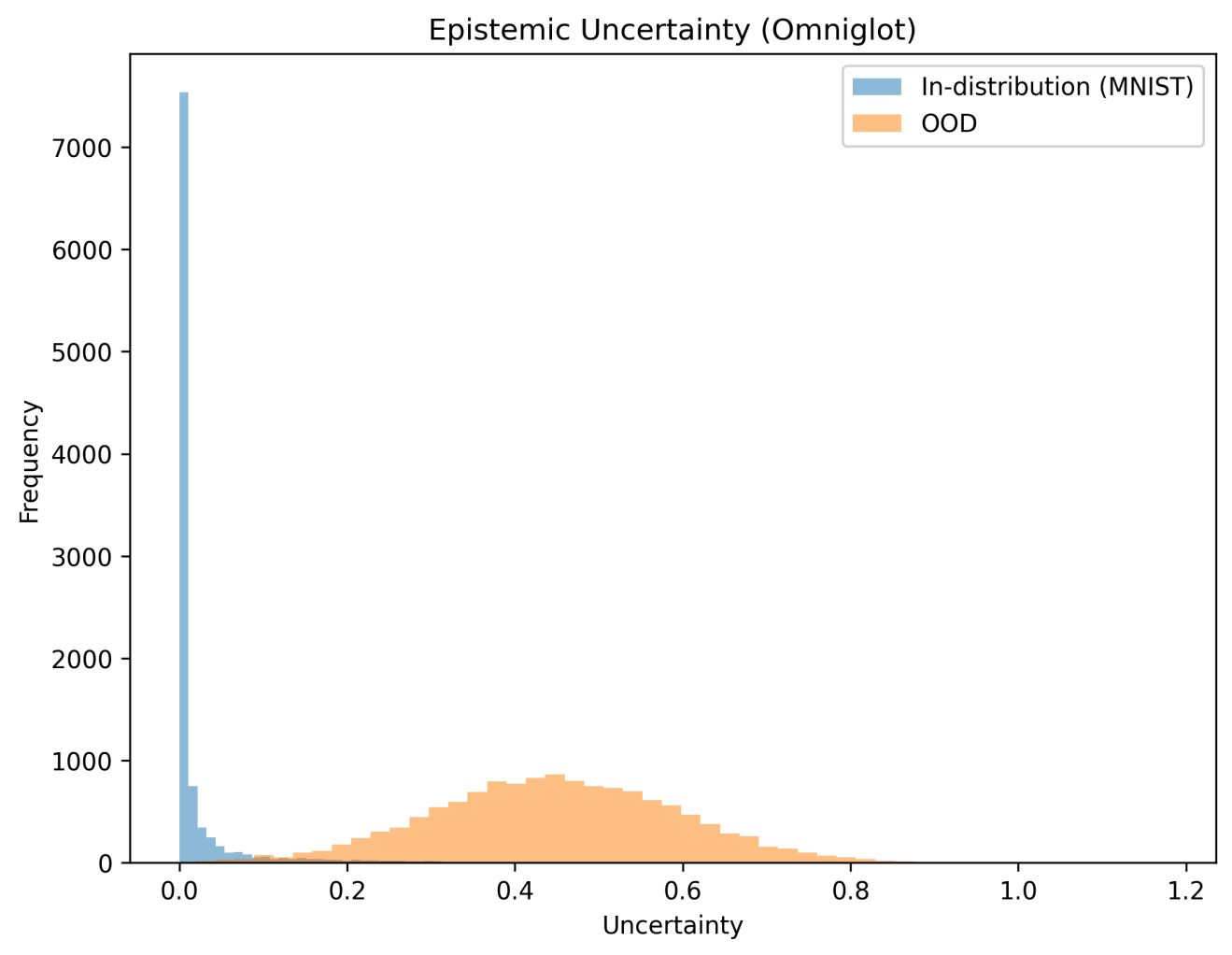
	AUROC	AUPR
Omniglot	0.9949	0.9956
EMNIST	0.9415	0.9703
KMNIST	0.9784	0.9776

Reference result

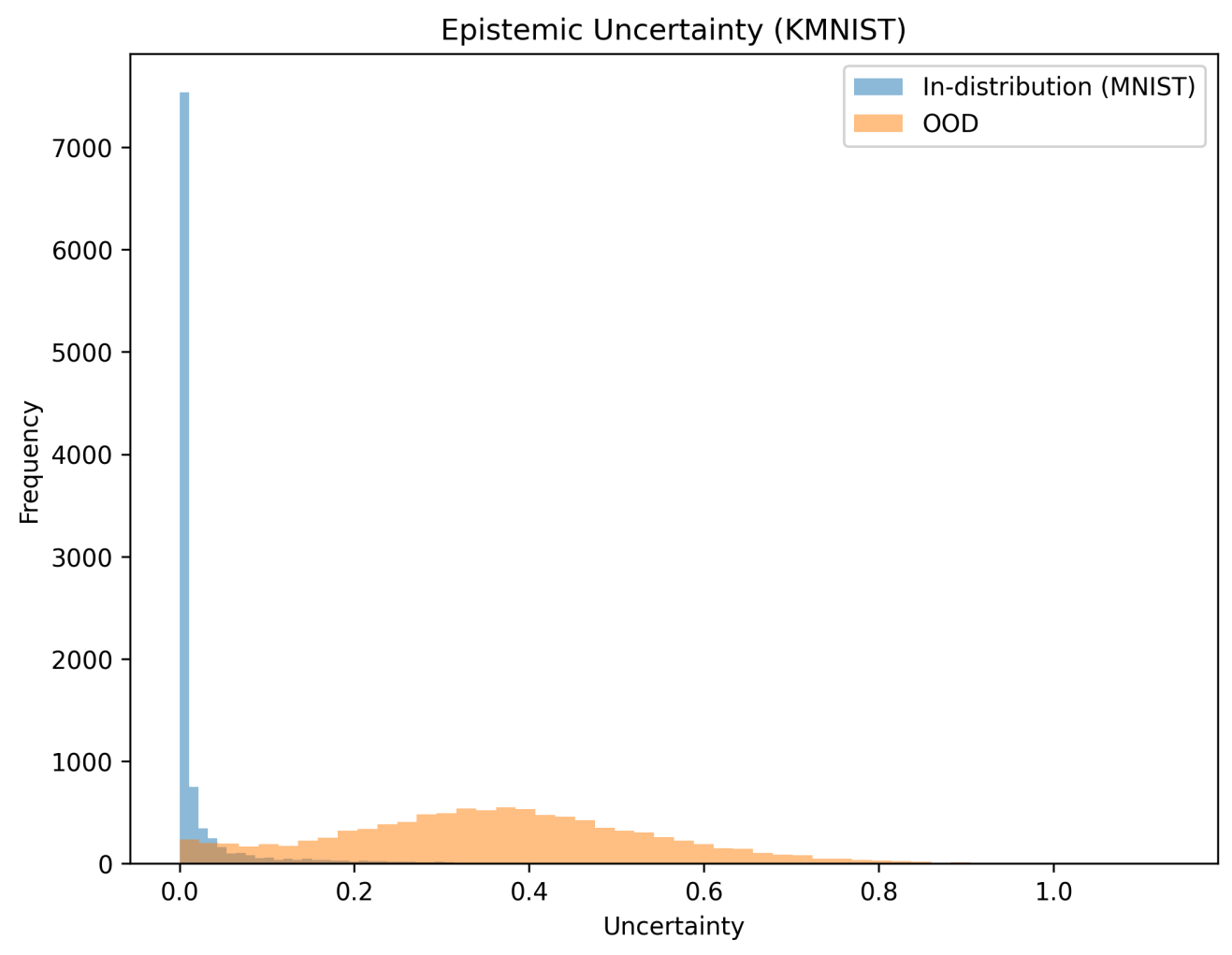
	AUROC	AUPR
Omniglot	0.9792	0.9733
EMNIST	0.9732	0.9610
KMNIST	0.9792	0.9713

2.2.1 Uncertainty Vs Dataset

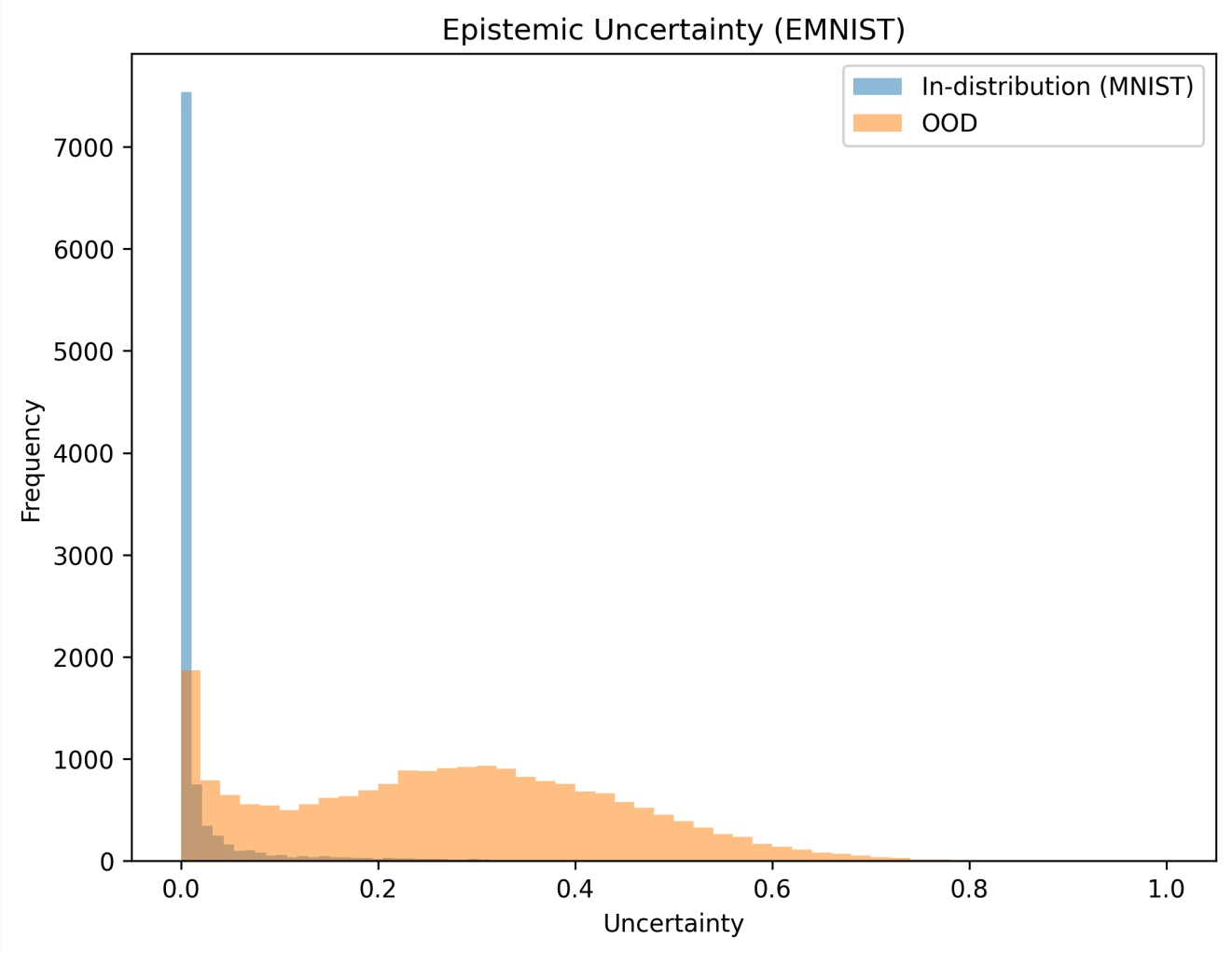
2.2 OOD Test


Epistemic Uncertainty (Omniglot)
In-distribution (MNIST)
OOD
Frequency
Uncertainty
0
1000
2000
3000
4000
5000
6000
7000
0.0
0.2
0.4
0.6
0.8
1.0
1.2

2.2 OOD Test


Epistemic Uncertainty (KMNIST)
In-distribution (MNIST)
OOD
Frequency
0
1000
2000
3000
4000
5000
6000
7000
0.0
0.2
0.4
0.6
0.8
1.0
Uncertainty

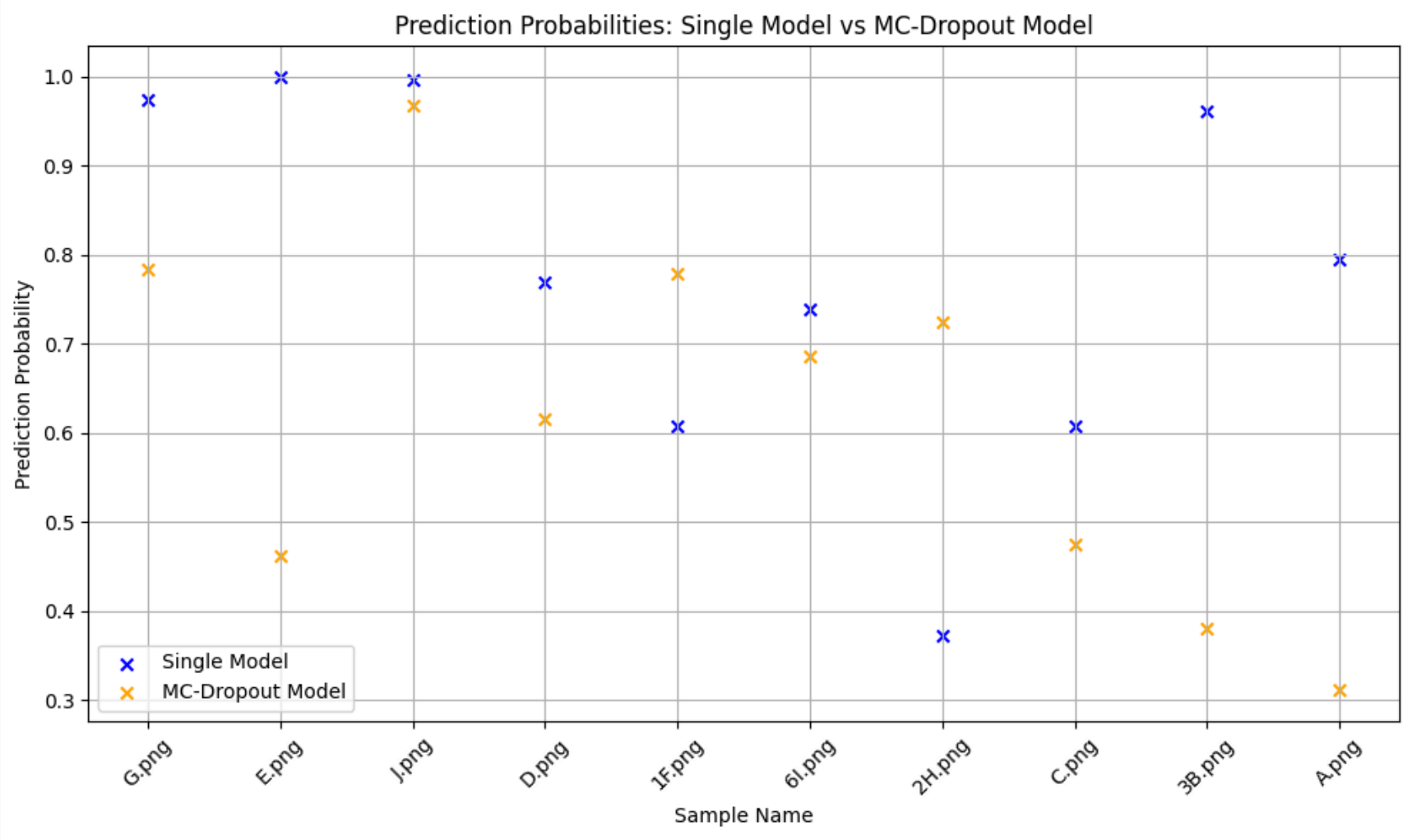
2.2 OOD Test


Epistemic Uncertainty (EMNIST)
In-distribution (MNIST)
OOD
Frequency
0
1000
2000
3000
4000
5000
6000
7000
0.0
0.2
0.4
0.6
0.8
1.0
Uncertainty

2.3 Single Image Test

Sample	Single Model Prediction	Single Model Probability	Ensemble Model Prediction	Ensemble Model Probability
G.png	3	0.9749	3	0.7833
E.png	5	1.0000	8	0.4629
J.png	1	0.9964	1	0.9670
D.png	4	0.7699	4	0.6162
1F.png	4	0.6075	6	0.7789
6I.png	1	0.7395	1	0.6868
2H.png	9	0.3729	8	0.7240
C.png	4	0.6074	1	0.4749
3B.png	4	0.9609	4	0.3802
A.png	4	0.7945	4	0.3116

2.3 Single Image Test



3. Plan for Next Week

3. Plan for Next Week

1. Train larger model on larger dataset (resnet for CIFAR10) and do uncertainty quantification/evaluation.
2. Learn detailedly the metrics to measure uncertainty (dive into reading), such as AUROC, AUPR, ECE, etc..
3. Read/implement gradient-based uncertainty attribution.