**Experimental results**

We use accuracy, precision, recall, fscore and AUC.

We use cross-validation for evaluation on 6 splits. For precision, recall, fscore and AUC we perform a mean over the classes, since they have equal number of images in the dataset.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | From scratch Softmax classifier | | ANN | | CNN | |
| Mean | CI interval | mean | CI interval | mean | CI interval |
| Accuracy | 0.41 | 0.0031 | 0.48 | 0.04 | 0.79 | 0.005 |
| Precision | 0.40 | 0.002 | 0.43 | 0.08 | 0.79 | 0.005 |
| Recall | 0.41 | 0.003 | 0.48 | 0.04 | 0.79 | 0.005 |
| Fscore | 0.40 | 0.002 | 0.43 | 0.08 | 0.79 | 0.005 |
| AUC | 0.90 | 0.0002 | 0.91 | 0.004 | 0.95 | 0.0006 |

In what follows we give some examples of ROC curves and confusion matrixes.

A picture containing graphical user interface

Description automatically generatedGraphical user interface, diagram, application

Description automatically generated A picture containing shape

Description automatically generated

Softmax

CNN

ANN

A picture containing treemap chart

Description automatically generatedTreemap chart

Description automatically generated with low confidence

CNN

ANN

A picture containing text, different

Description automatically generated

Softmax