Table 1. Resumed performances of the models

Models	PR-AUC	F1	PRECISION	RECALL	ACC	DEPTH
Inception_v4	**0.994 +/- 0.002**	**0.988 +/- 0.004**	**0.989 +/- 0.005**	**0.987 +/- 0.005**	**0.987 +/- 0.004**	**75**
ResNet50	0.989 +/- 0.003	0.976 +/- 0.005	0.975 +/- 0.009	0.978 +/- 0.007	0.974 +/- 0.006	53
Darknet53	0.99 +/- 0.003	0.976 +/- 0.006	0.972 +/- 0.009	0.98 +/- 0.007	0.974 +/- 0.007	51
ResNet34	0.988 +/- 0.003	0.975 +/- 0.004	0.968 +/- 0.009	0.976 +/- 0.006	0.969 +/- 0.005	37

Overall, the results were good, because the smallest estimated point is 0.968, which is the Precision of ResNet34. As was observed by [14] "The most straightforward way of improving the performance of deep neural networks is by increasing their size. This includes both increasing the depth – the number of network levels – as well as its width: the number of units at each level", and that was verified at this study. As we can see at Fig. 3, with a pearson correlation coefficient $r = 0.986$, the correlation between the model depth and the ACC is very strong positive, roughly speaking deeper models have tended to have higher ACC.
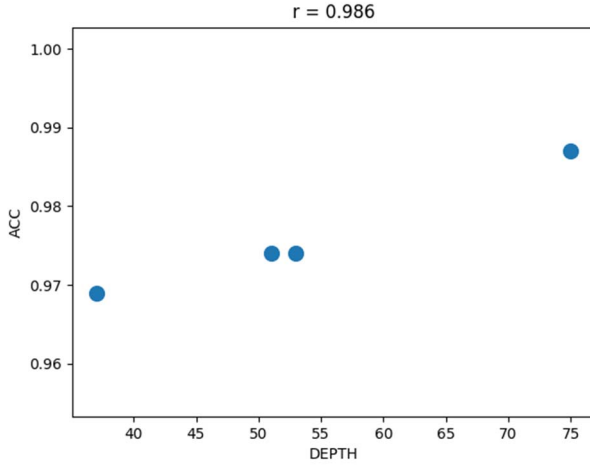

Fig. 3. Very strong positive correlation between the model depth and ACC.

A. Inception_v4

The Table 2 show more details about the performance of Inception_v4, the standard deviation calculated from estimates obtained by running 5-Fold Cross-Validation 6 times is nearly zero, which mean, there is just slightly deference between the obtained estimates or roughly speaking the obtained estimates at each time of test process was nearly equal, which make the model more consistent.

Table 2. The performance of Inception_v4. Margin of Error, E. Sample standard deviation, S. Confidence Interval, **CI.**

Metrics	mean	S	E	CI
PR-AUC	0.994	0.007	0.002	0.994 +/- 0.002
F1	0.988	0.012	0.004	0.988 +/- 0.004
PRECISION	0.989	0.015	0.005	0.989 +/- 0.005
RECALL	0.987	0.014	0.005	0.987 +/- 0.005
ACC	0.987	0.012	0.004	0.987 +/- 0.004

With Precision of 0.989, and Recall of 0.987, the model had on average the absence of false positive (false alarms) and presence of only one false negative. It is important to point out that, the Inception_v4 had less parameters than DarkNet53 and ResNet50. Also, differently from DarkNet53 and ResNet50 Inception_v4 do not use Residual block as building blocks [15], instead it uses Inception block as building blocks [16], which is much deeper, sparse and with very variability in the kernels size, leading to good results. This supports the idea that most of the progress in deep learning is not just on the result of more powerful hardware, larger datasets and bigger models, but mainly a consequence of new ideas, algorithms and improved network architectures [14].

B. ResNet50

The Table 3 show more details about the performance of ResNet50, the standard deviation is nearly zero.