Comparative analysis of Deep learning models for Surgical Tool Detection

By
Bhavya Gupta

Database	Cholec-tinytools
GPU	NVIDIA Tesla p100-pcie-16gb

INTRODUCTION

While performing robotic surgery or giving training to a junior surgeon through an end-to-end deep learning model, it's very important that our model is able to detect surgical tools accurately so that the chances of error during surgery are less to zero. Therefore to achieve our aim, we will be doing a comparative study of 4 state of the art deep learning models on the basis of various classification parameters. (namely Accuracy, Precision, Recall & F1 score)

Understanding significance of various classification parameters:

1.) Accuracy is how close the predicted value is to the actual value.
Higher accuracy indicates that the chances of making errors by our model are less and we can rely on it for crucial tasks.

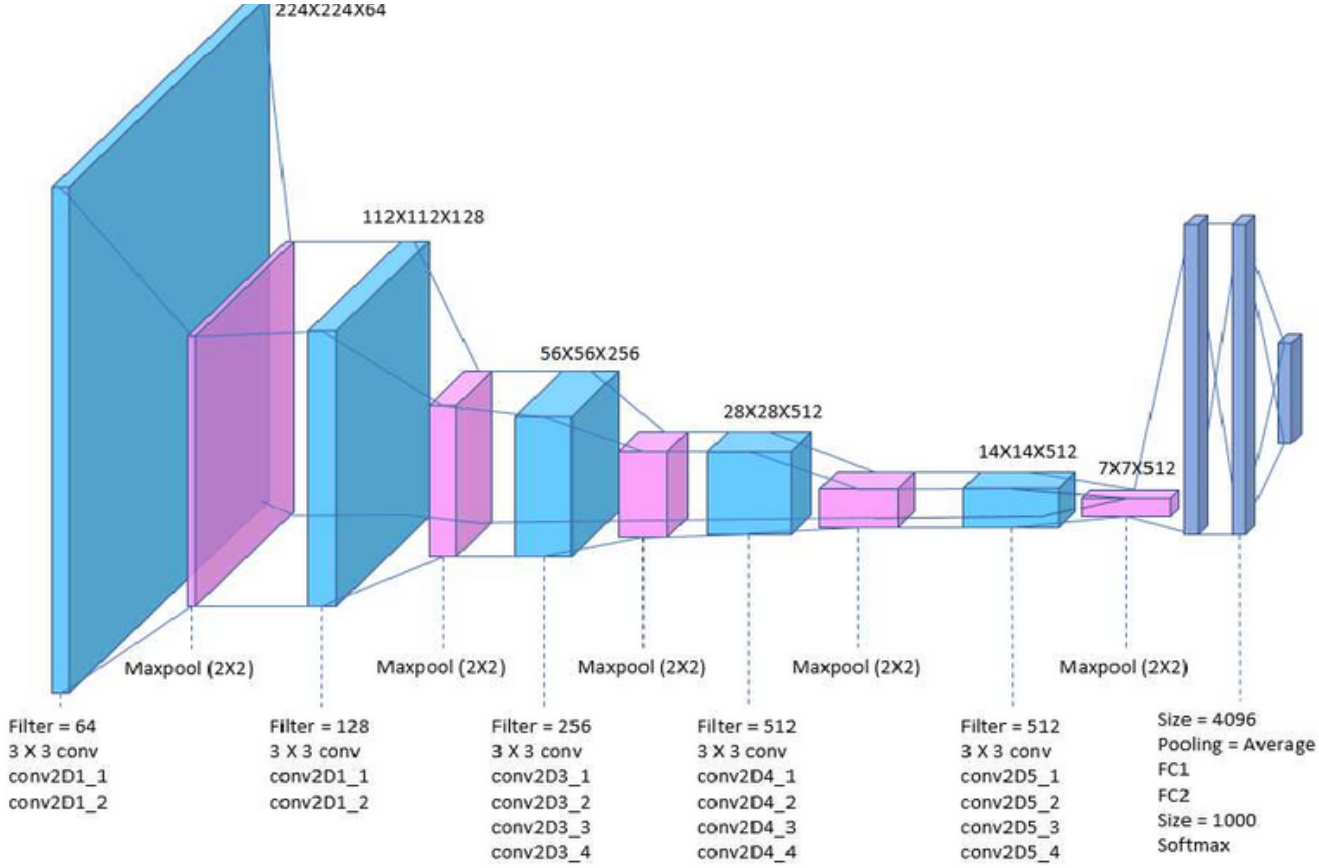
2.) **Higher precision** indicates that predicted values are consistent whereas lower precision indicates that our predicted values vary. However, a highly precise prediction need not always produce an accurate outcome.

3.) A **high recall** value means there were very few false negatives and that the classifier is more permissive in the criteria for classifying something as positive which means models need high recall when we need output-sensitive predictions. For example, predicting cancer needs a high recall, in other words, we need to cover false negatives as well. It is ok if a non-cancer tumor is flagged as cancerous but a cancerous tumor should not be labeled non-cancerous.

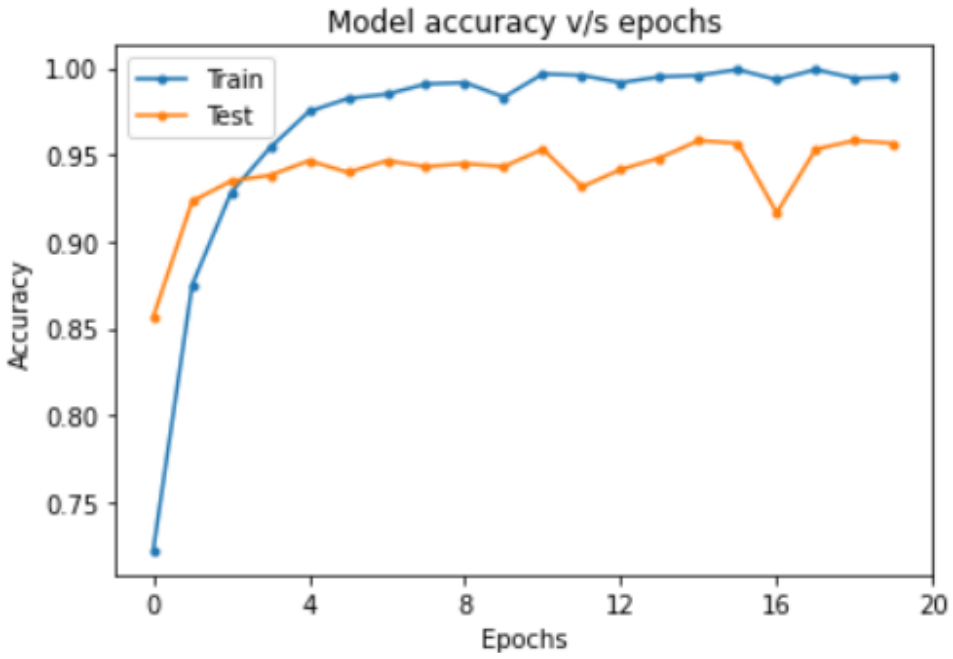
4.) F1 score is defined as the harmonic mean between precision and recall. It is also a measure to rate the performance of the model.

Model Testing :

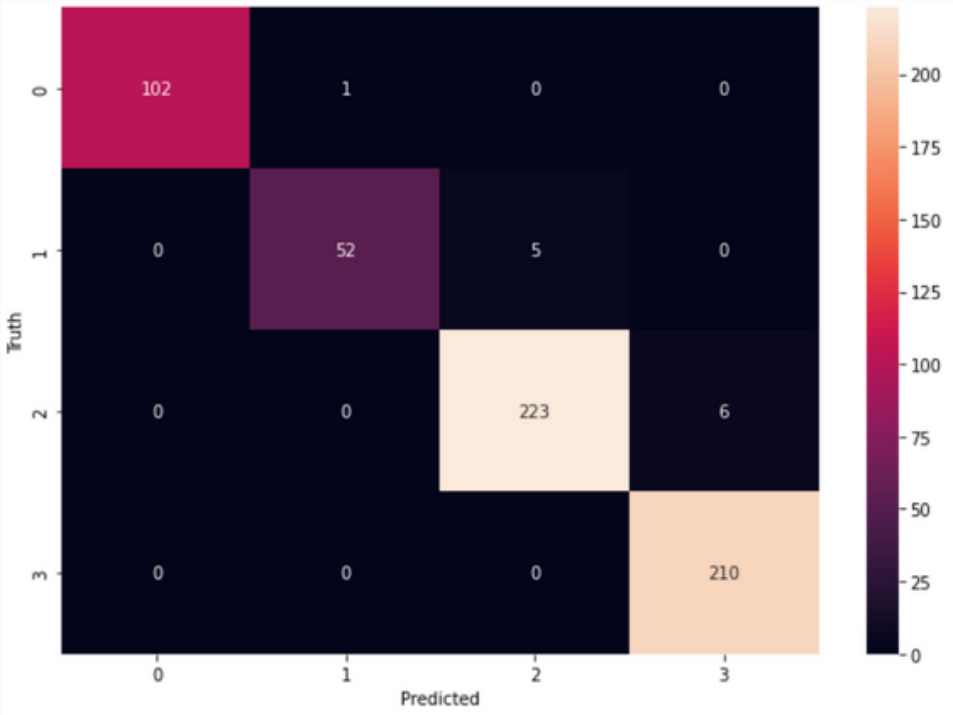
1.] VGG 19



Total number of parameters:	143.7 M
Training time :	4 min
Total number of epochs :	20
Callbacks :	True (Patience =3)
Optimizer :	RMSprop
Learning rate :	2e-5
Loss :	Categorical Crossentropy
Train size :	1200
Validation size :	200
Test size :	599

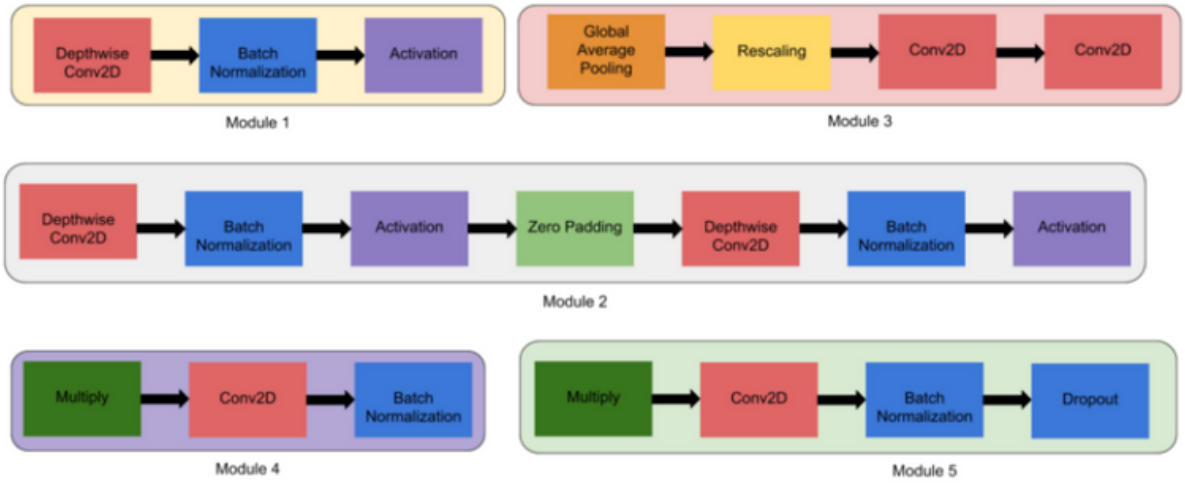


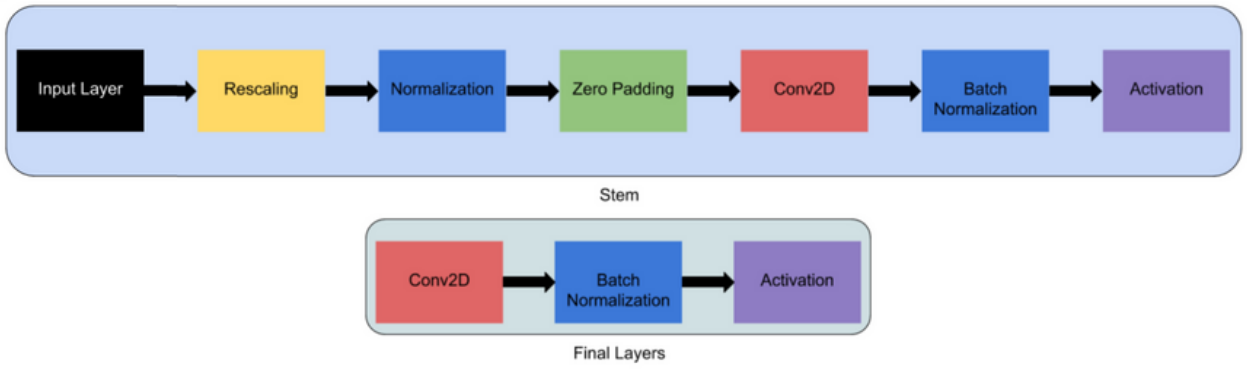
Confusion Matrix

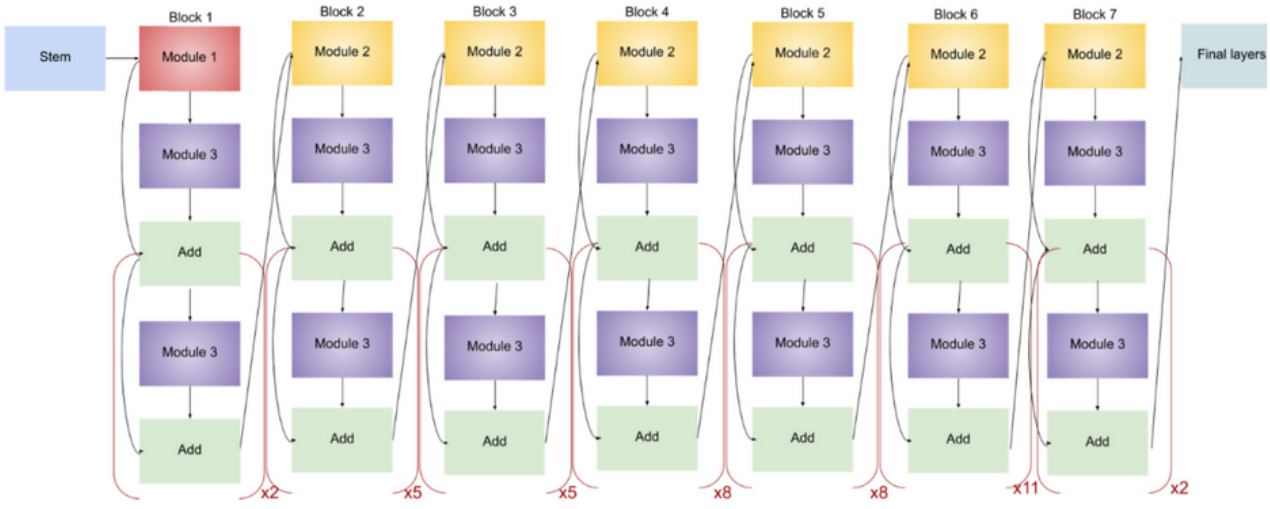


Metric	Measurement
Accuracy	95.66
Precision	98
Recall score	0.956
F1 score	0.956

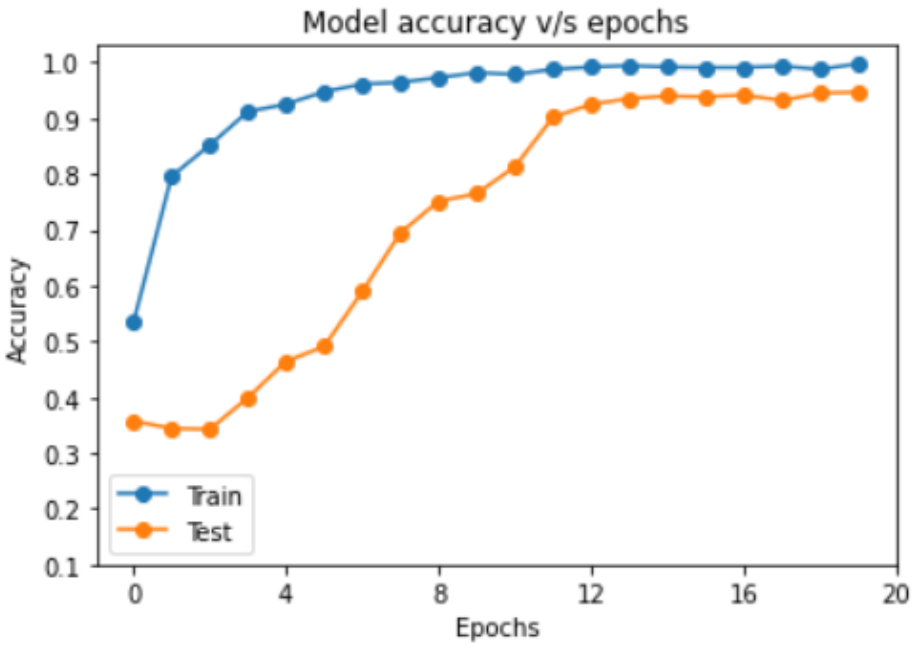
2.) Efficient NET (B7)



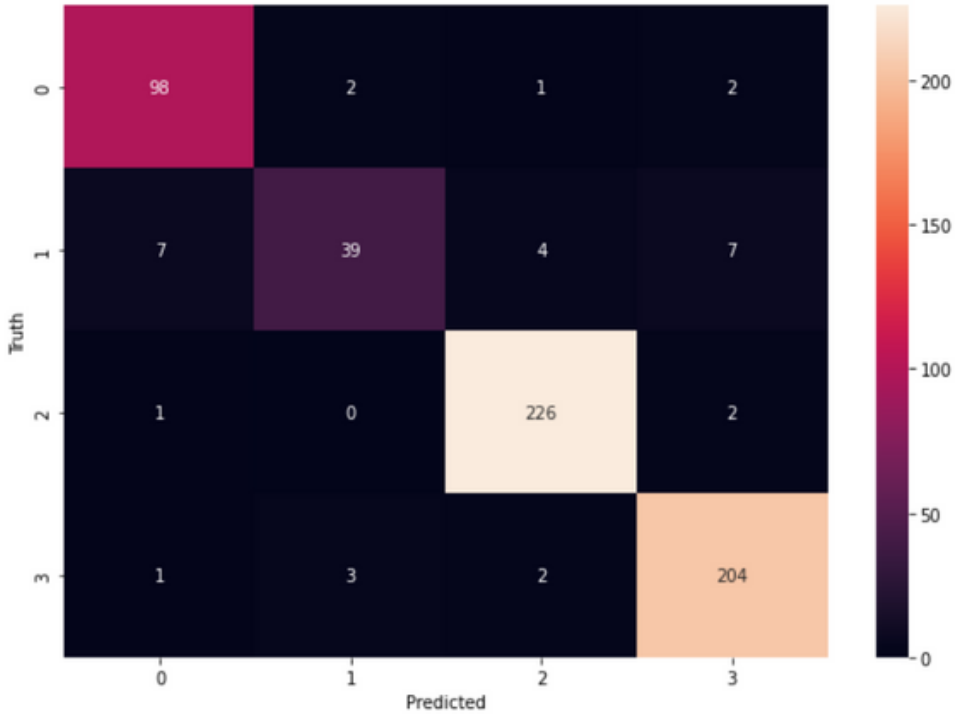




Total number of parameters :	66.7M
Training time :	8 min
Total number of Epochs :	20
Callbacks :	True (Patience =3)
Optimizer :	RMSprop
Learning rate :	2e-5
Loss :	Categorical Crossentropy
Train size :	1200
Validation size :	200
Test size :	599

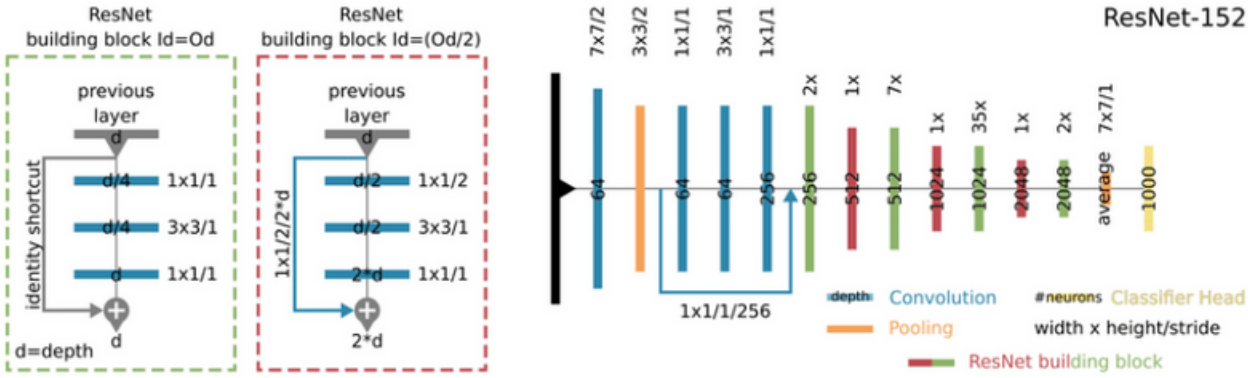


Confusion Matrix

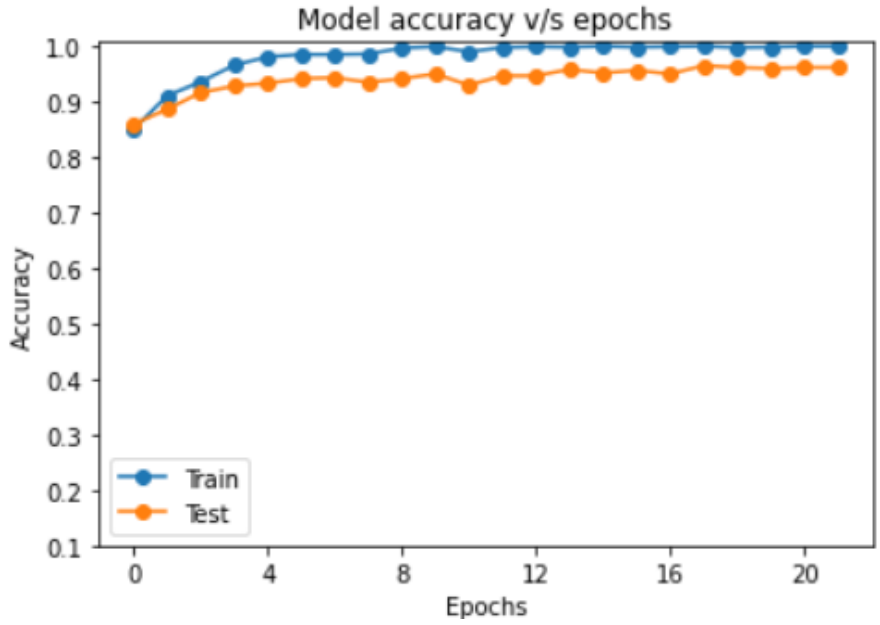


Metric	Measurement
Accuracy	94.66
Precision	94.53
Recall score	0.944
F1 score	0.946

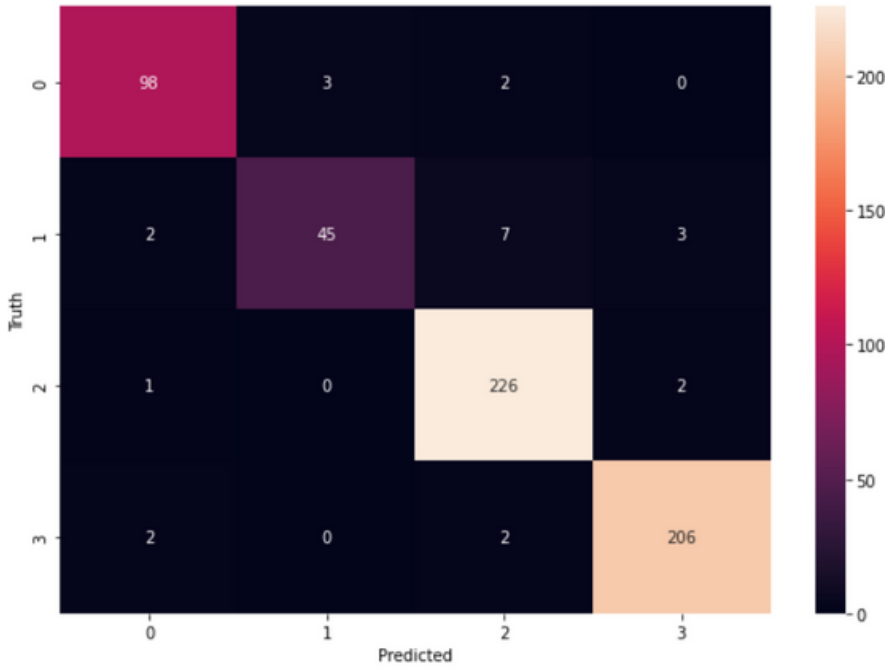
3.] ResNet152_v2



Total number of parameters :	60.4M
Training time :	3 min
Total number of Epochs :	22
Callbacks :	True (Patience =3)
Optimizer :	RMSprop
Learning rate :	2e-5
Loss :	Categorical Crossentropy
Train size :	1200
Validation size :	200
Test size :	599

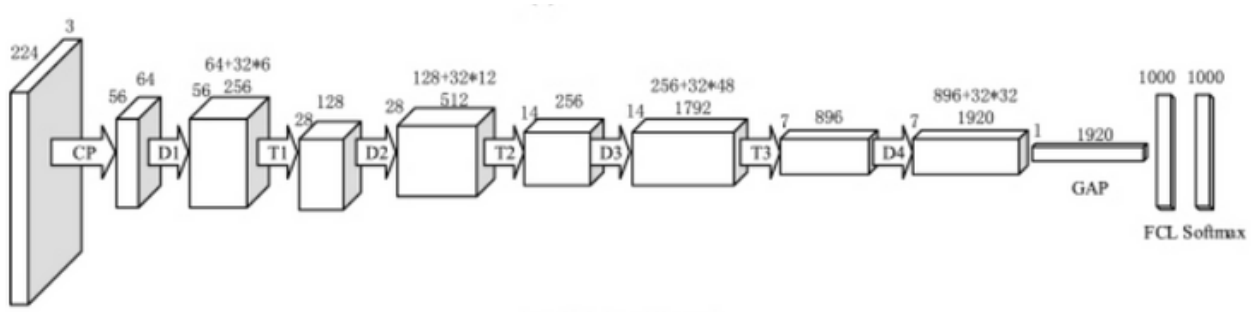


Confusion Matrix

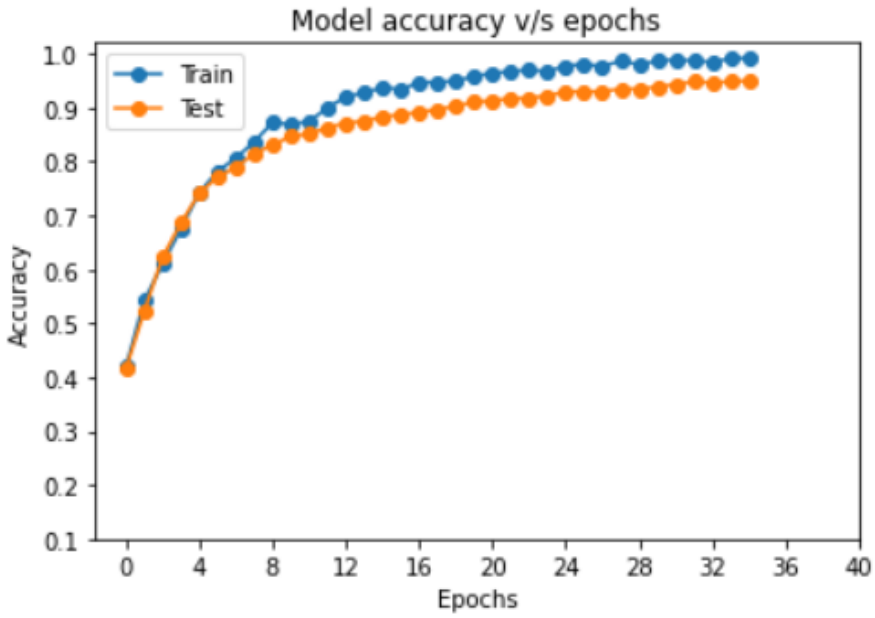


Metric	Measurement
Accuracy	96.01
Precision	95.97
Recall score	0.960
F1 score	0.960

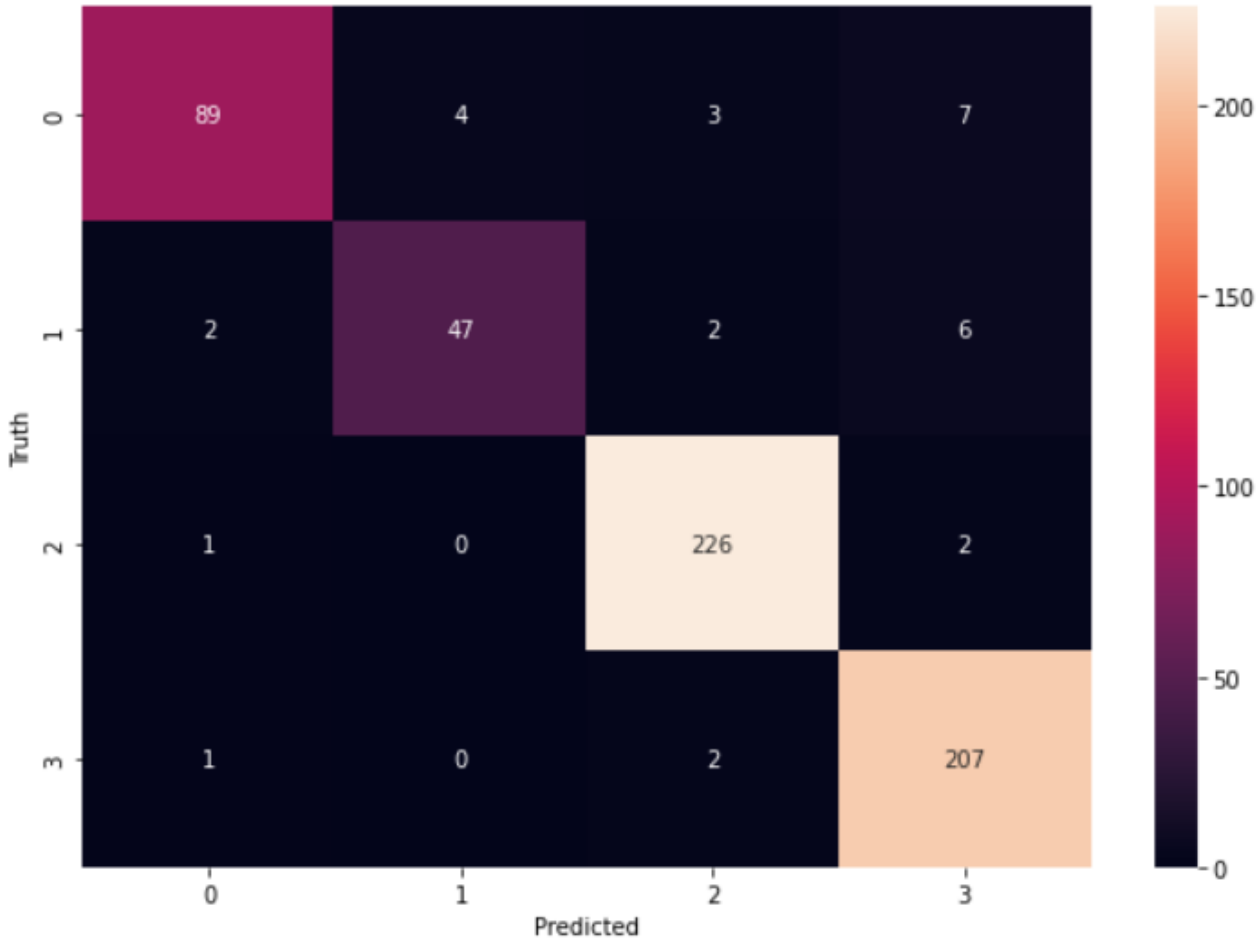
4.]DenseNet_201



Total number of parameters :	20.2 M
Training time :	7 min
Total number of Epochs :	35
Callbacks :	True (Patience =3)
Optimizer :	RMSprop
Learning rate :	2e-5
Loss :	Categorical Crossentropy
Train size :	1200
Validation size :	200
Test size :	599



Confusion Matrix



Metric	Measurement
Accuracy	95.01
Precision	95.00
Recall score	0.950
F1 score	0.949

Comparative analysis :

Model	Accuracy	Precision	Recall score	F1 score	Training time (In min)
VGG_19	95.66	98.00	0.956	0.956	4
EfficientNet[b7]	94.66	94.53	0.944	0.946	8
ResNet152_v2	96.01	95.97	0.960	0.960	3
DenseNet_201	95.01	95.00	0.950	0.949	7

Conclusion :

ResNet152_v2 performs better than the rest of the models in 3 out of 4 parameters Therefore we can safely conclude that out of the four state-of-the-art classification models ResNet152_v2 is the best choice for Surgical tools detection.