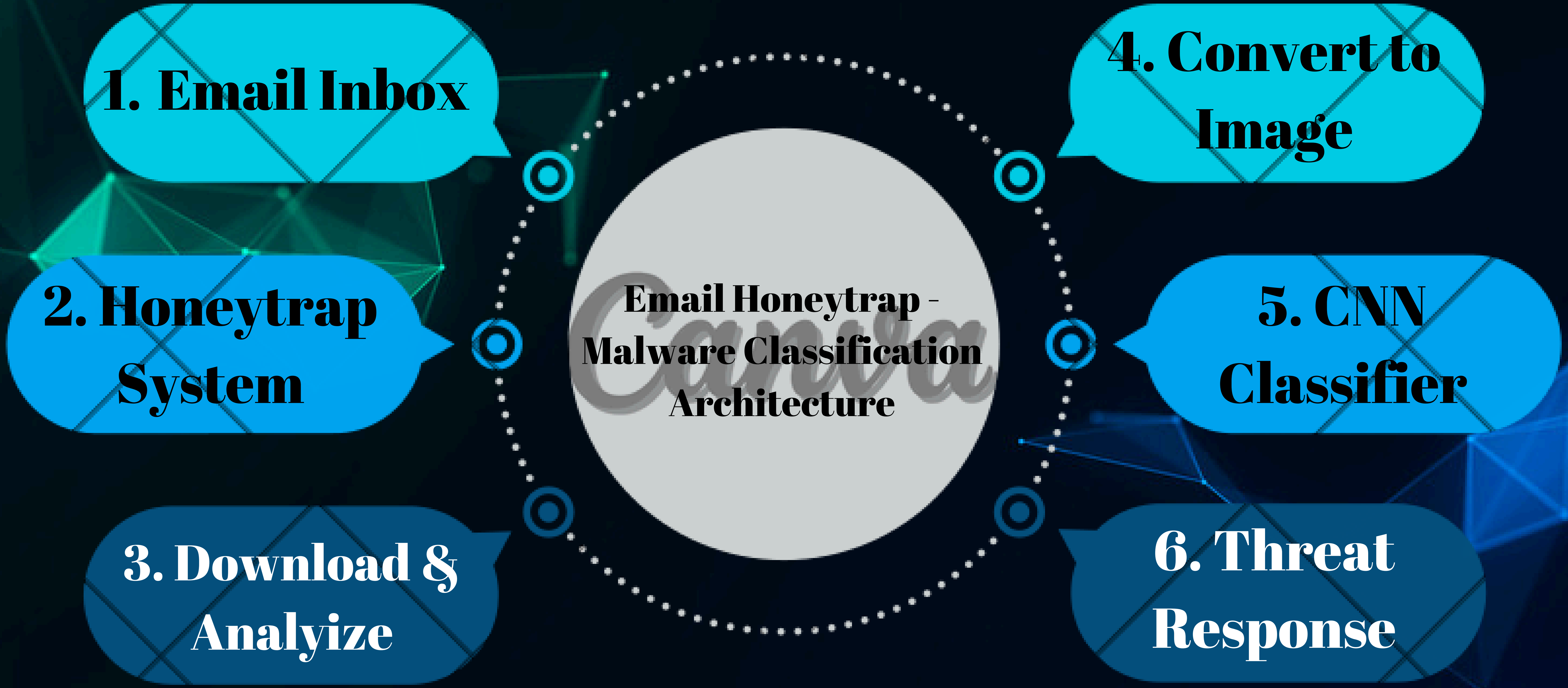
MALWARE DETECTION USING CNN

Neural Networks and Deep Learning - 19CSE456

INRODUCTION

Email remains a key attack vector for malware delivery, with traditional antivirus solutions struggling against zero-day threats. This project proposes a CNN-based approach for malware detection by converting binary executables into grayscale images for classification. An Email Honeytrap captures suspicious attachments, processes them into images, and uses a trained CNN model for real-time threat detection and quarantine. Leveraging Python, TensorFlow, and cloud services, this system enhances malware detection accuracy and response speed, demonstrating the effectiveness of deep learning in cybersecurity



Email Honeytrap - Malware Classification Architecture

1. Email Inbox

2. Honeytrap System

3. Download & Analyize

4. Convert to Image

5. CNN Classifier

6. Threat Response

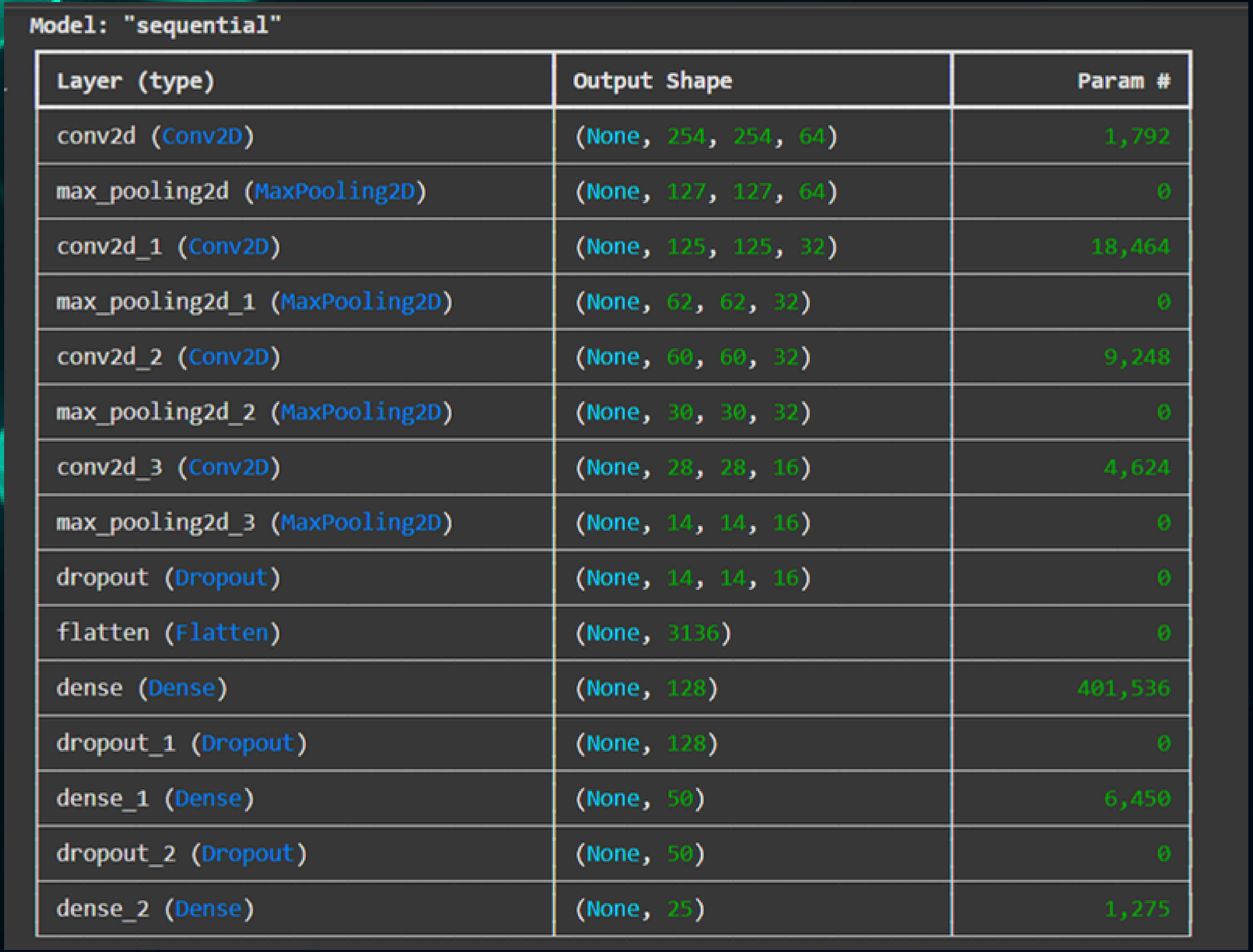
DATASET OVERVIEW

- Training Set: 7,455 images across 25 classes.
- Validation Set: 922 images.
- Test Set: 955 images.
- Preprocessing:
  - Image augmentation (rotation, zoom, flip) for training.
  - Rescaling (1.0/255.0) for validation and test sets.
- Class Distribution: Analyzed using bar plots for train and test sets

COMPARISION BETWEEN TWO MODELS

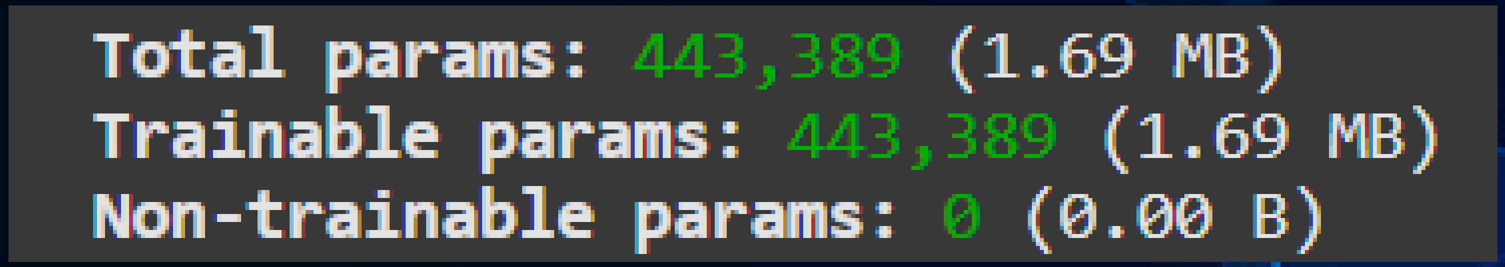
Aspect	Baseline Model	Improved Model
Batch Normalization	Not present	Added after Conv2D and Dense layers
Activation Layers	Implicit (within Conv2D/Dense)	Explicit Activation layers (e.g., ReLU)
Total Parameters	443,389 (1.69 MB)	444,677 (1.70 MB)
Non-trainable Params	0	644 (2.52 KB) due to BatchNormalization
Training Stability	Noisy accuracy/loss curves	Smoother accuracy/loss curves
Convergence	Reaches ~0.95 accuracy, ~0.2 loss (30 epochs)	Reaches ~0.95 accuracy, ~0 loss (50 epochs)
Training Epochs	30	50

BASELINE MODEL



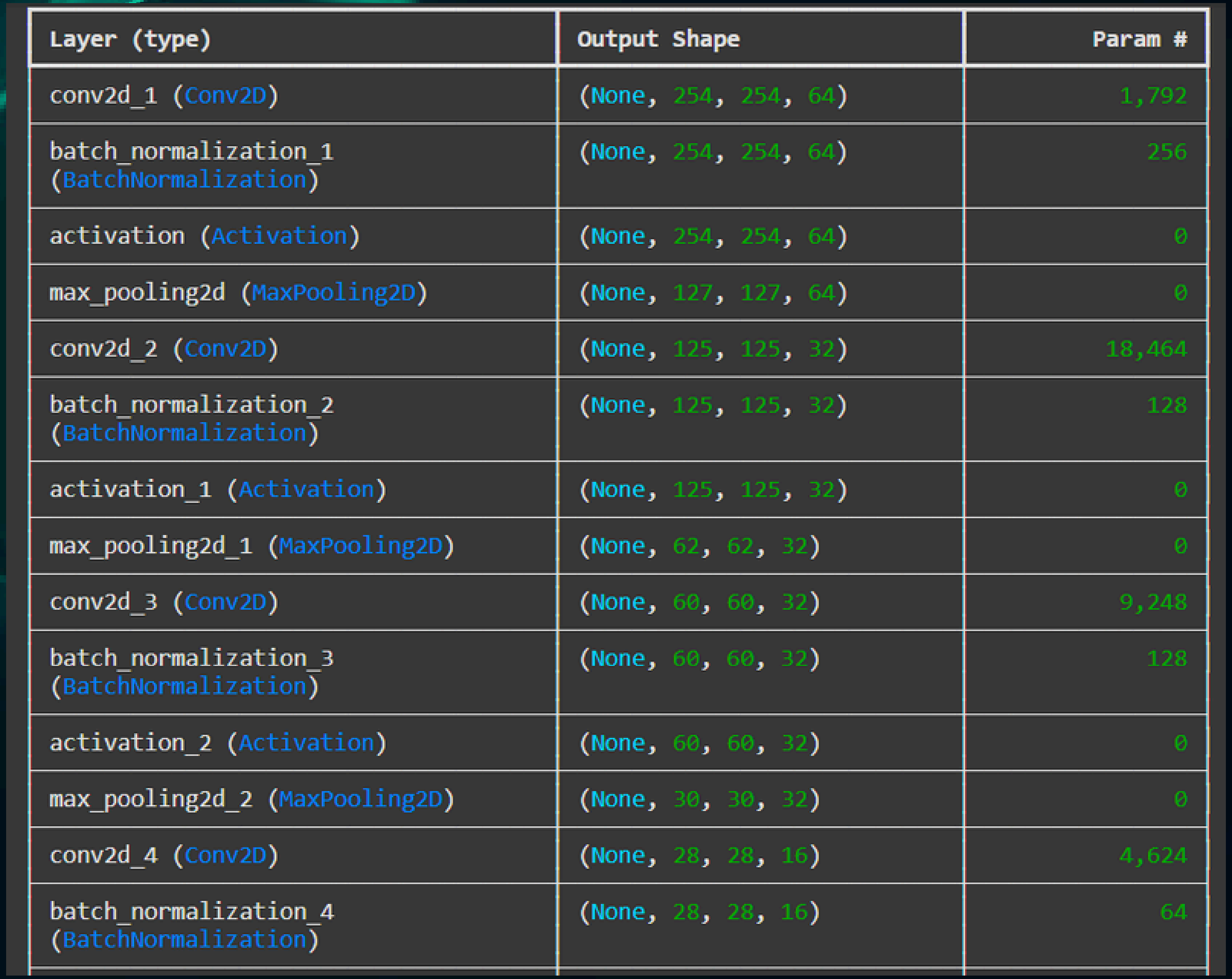
Model: "sequential"

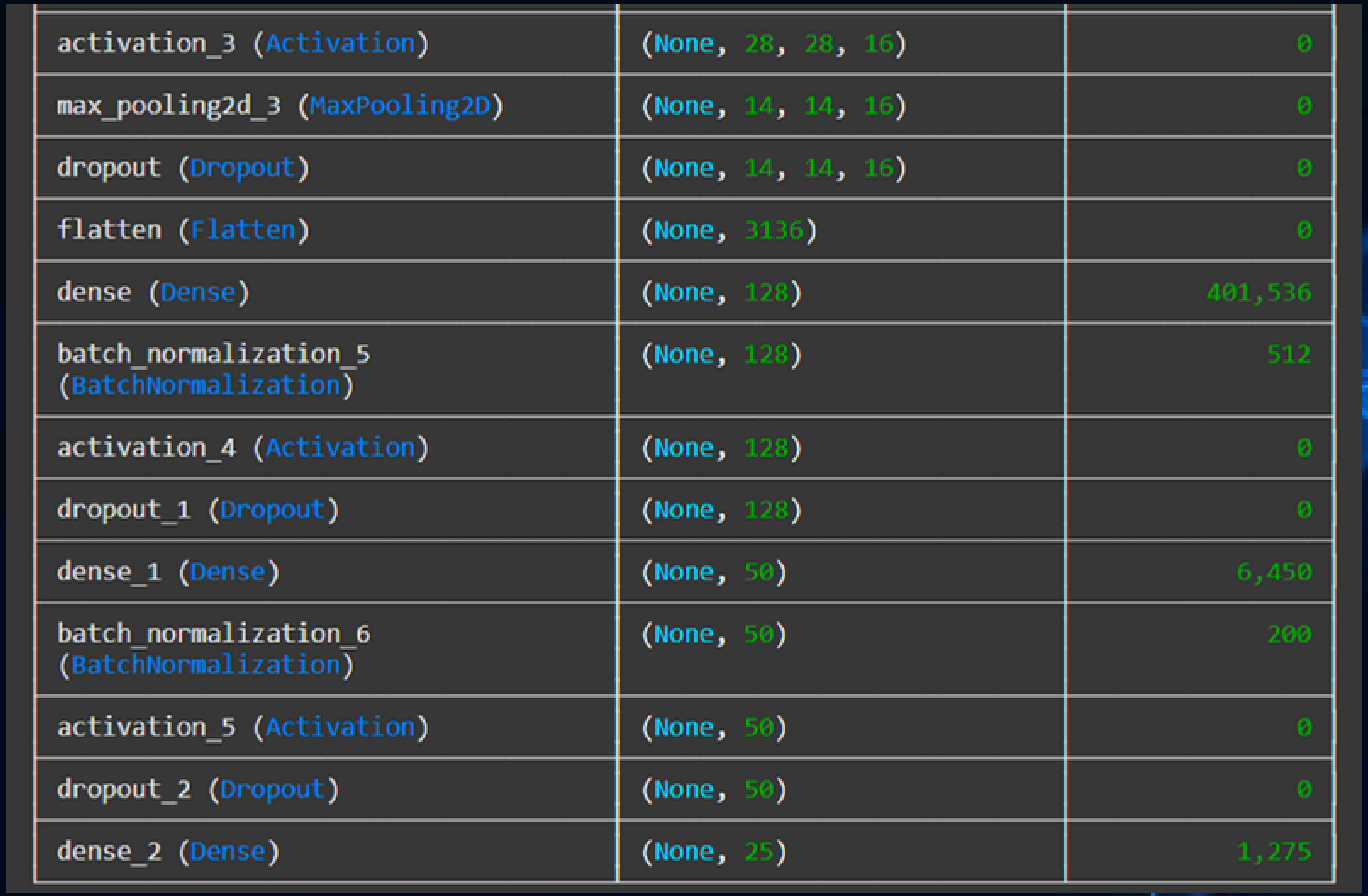
Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 254, 254, 64)	1,792
max_pooling2d (MaxPooling2D)	(None, 127, 127, 64)	0
conv2d_1 (Conv2D)	(None, 125, 125, 32)	18,464
max_pooling2d_1 (MaxPooling2D)	(None, 62, 62, 32)	0
conv2d_2 (Conv2D)	(None, 60, 60, 32)	9,248
max_pooling2d_2 (MaxPooling2D)	(None, 30, 30, 32)	0
conv2d_3 (Conv2D)	(None, 28, 28, 16)	4,624
max_pooling2d_3 (MaxPooling2D)	(None, 14, 14, 16)	0
dropout (Dropout)	(None, 14, 14, 16)	0
flatten (Flatten)	(None, 3136)	0
dense (Dense)	(None, 128)	401,536
dropout_1 (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 50)	6,450
dropout_2 (Dropout)	(None, 50)	0
dense_2 (Dense)	(None, 25)	1,275



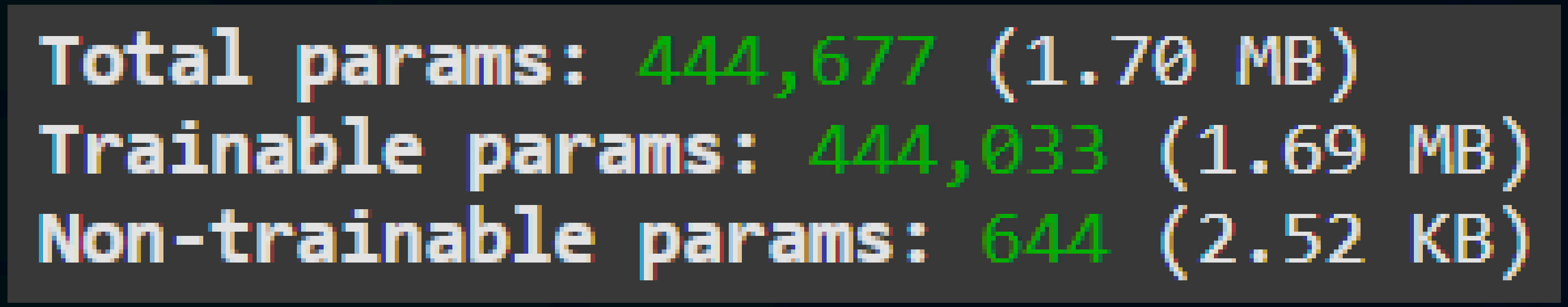
```
Total params: 443,389 (1.69 MB)
Trainable params: 443,389 (1.69 MB)
Non-trainable params: 0 (0.00 B)
```

IMPROVED MODEL



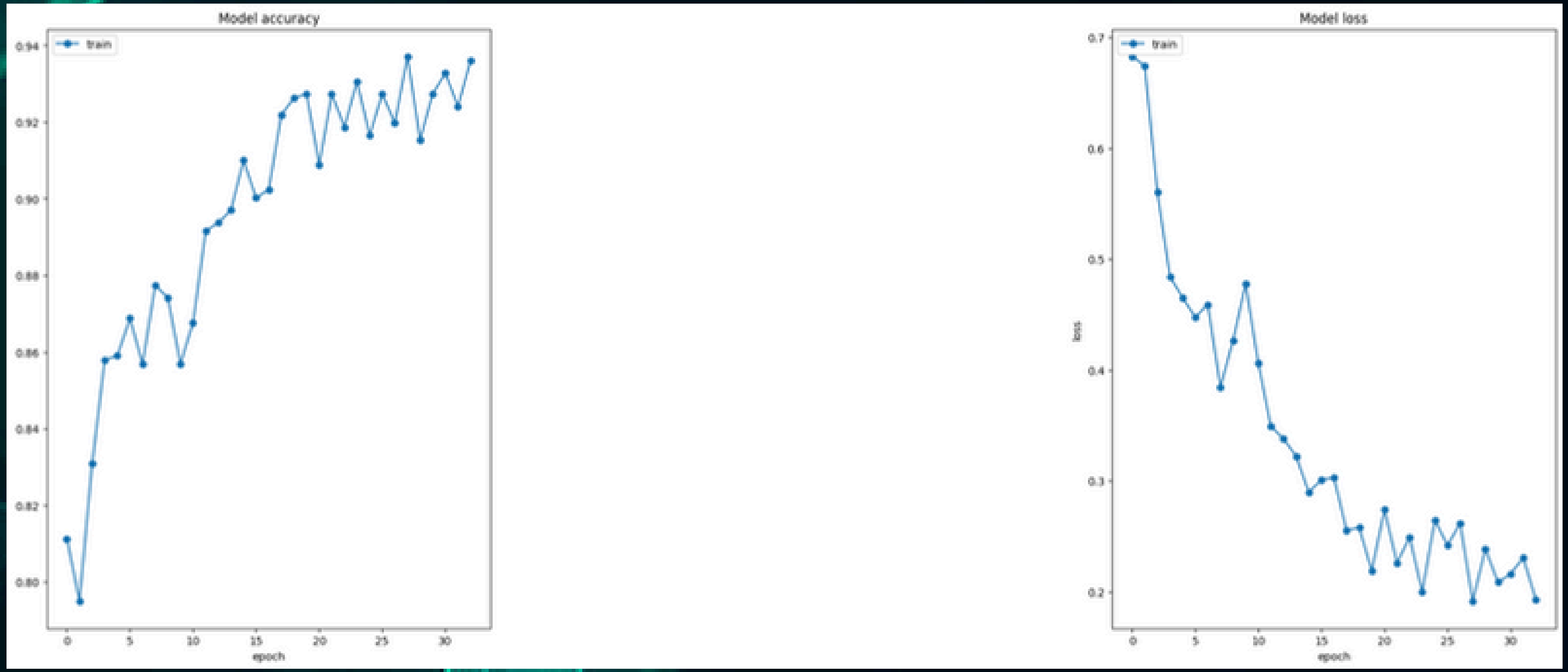


Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 254, 254, 64)	1,792
batch_normalization_1 (BatchNormalization)	(None, 254, 254, 64)	256
activation (Activation)	(None, 254, 254, 64)	0
max_pooling2d (MaxPooling2D)	(None, 127, 127, 64)	0
conv2d_2 (Conv2D)	(None, 125, 125, 32)	18,464
batch_normalization_2 (BatchNormalization)	(None, 125, 125, 32)	128
activation_1 (Activation)	(None, 125, 125, 32)	0
max_pooling2d_1 (MaxPooling2D)	(None, 62, 62, 32)	0
conv2d_3 (Conv2D)	(None, 60, 60, 32)	9,248
batch_normalization_3 (BatchNormalization)	(None, 60, 60, 32)	128
activation_2 (Activation)	(None, 60, 60, 32)	0
max_pooling2d_2 (MaxPooling2D)	(None, 30, 30, 32)	0
conv2d_4 (Conv2D)	(None, 28, 28, 16)	4,624
batch_normalization_4 (BatchNormalization)	(None, 28, 28, 16)	64
activation_3 (Activation)	(None, 28, 28, 16)	0
max_pooling2d_3 (MaxPooling2D)	(None, 14, 14, 16)	0
dropout (Dropout)	(None, 14, 14, 16)	0
flatten (Flatten)	(None, 3136)	0
dense (Dense)	(None, 128)	401,536
batch_normalization_5 (BatchNormalization)	(None, 128)	512
activation_4 (Activation)	(None, 128)	0
dropout_1 (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 50)	6,450
batch_normalization_6 (BatchNormalization)	(None, 50)	200
activation_5 (Activation)	(None, 50)	0
dropout_2 (Dropout)	(None, 50)	0
dense_2 (Dense)	(None, 25)	1,275

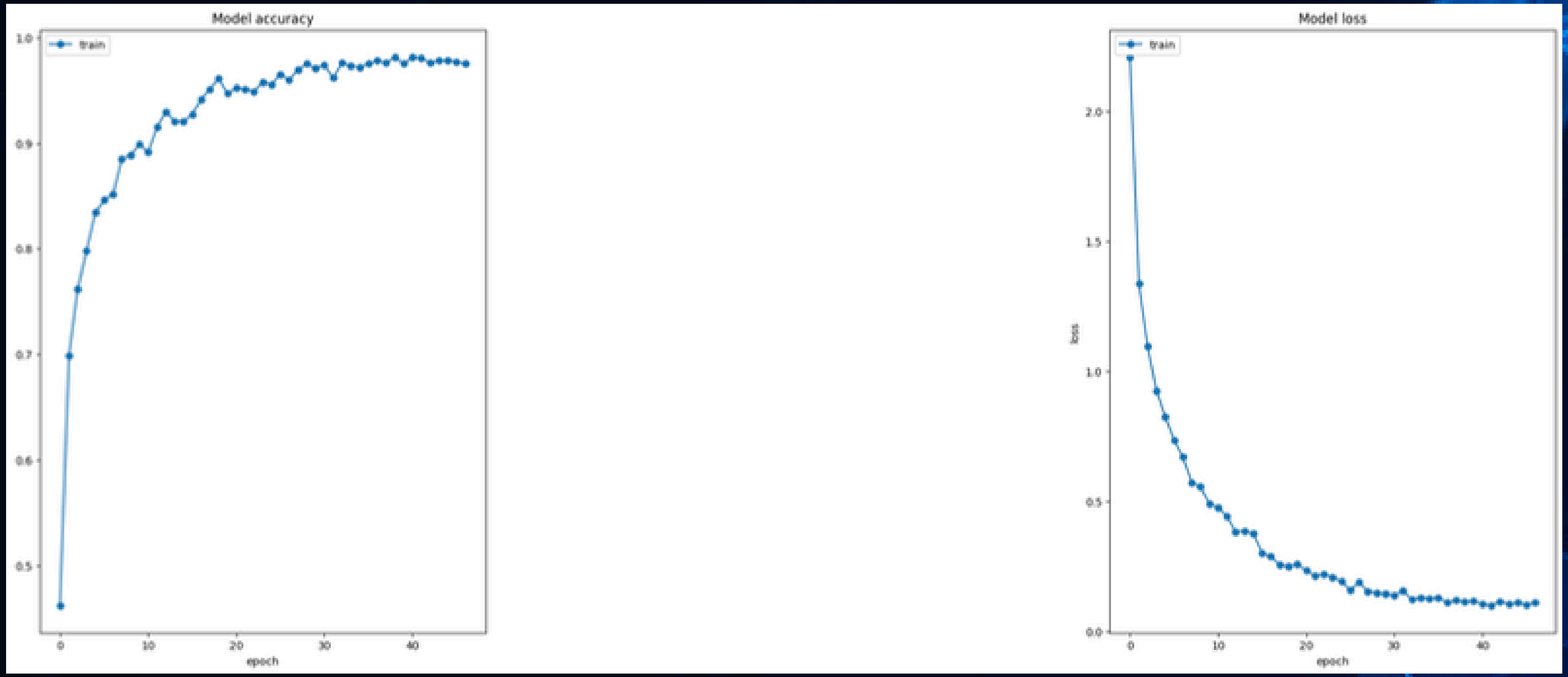


```
Total params: 444,677 (1.70 MB)
Trainable params: 444,033 (1.69 MB)
Non-trainable params: 644 (2.52 KB)
```

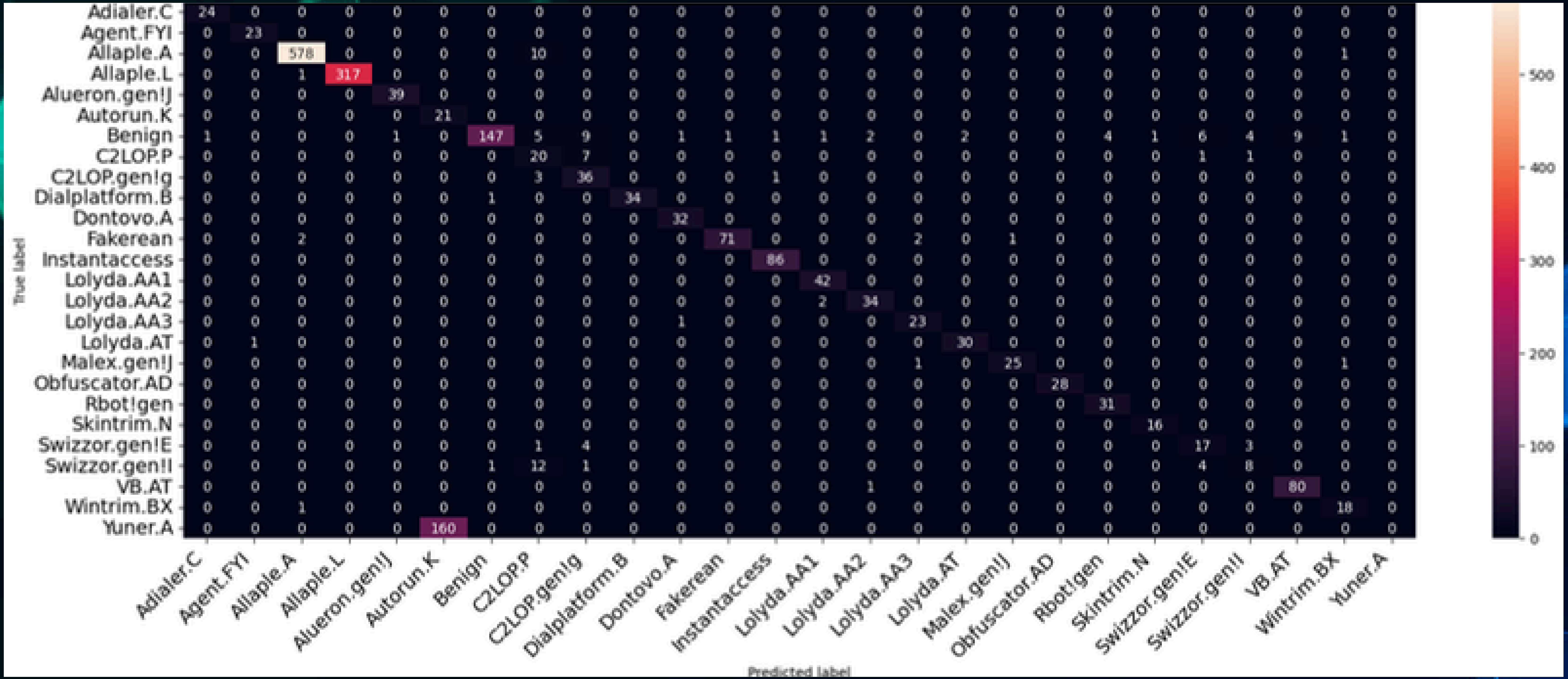
BASELINE MODEL



IMPROVED MODEL



EVALUATION METRICS



	Adialer.C	Agent.FYI	Allaple.A	Allaple.L	Alueron.gen!J	Autorun.K	Benign	C2LOP.P	C2LOP.gen!g	Dialplatform.B	Dontovo.A	Fakerean	Instantaccess	Lolyda.AA1	Lolyda.AA2	Lolyda.AA3	Lolyda.AT	Malex.gen!J	Obfuscator.AD	Rbot!gen	Skintrim.N	Swizzor.gen!E	Swizzor.gen!I	VB.AT	Wintrim.BX	Yuner.A
Adialer.C	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Agent.FYI	0	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Allaple.A	0	0	578	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Allaple.L	0	0	1	317	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Alueron.gen!J	0	0	0	0	39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Autorun.K	0	0	0	0	0	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Benign	1	0	0	0	1	0	147	5	9	0	1	1	1	1	2	0	2	0	0	4	1	6	4	9	1	0
C2LOP.P	0	0	0	0	0	0	0	20	7	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0
C2LOP.gen!g	0	0	0	0	0	0	0	3	36	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Dialplatform.B	0	0	0	0	0	0	1	0	0	34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Dontovo.A	0	0	0	0	0	0	0	0	0	0	32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Fakerean	0	0	2	0	0	0	0	0	0	0	0	71	0	0	0	2	0	1	0	0	0	0	0	0	0	0
Instantaccess	0	0	0	0	0	0	0	0	0	0	0	0	86	0	0	0	0	0	0	0	0	0	0	0	0	0
Lolyda.AA1	0	0	0	0	0	0	0	0	0	0	0	0	0	42	0	0	0	0	0	0	0	0	0	0	0	0
Lolyda.AA2	0	0	0	0	0	0	0	0	0	0	0	0	0	2	34	0	0	0	0	0	0	0	0	0	0	0
Lolyda.AA3	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	23	0	0	0	0	0	0	0	0	0	0
Lolyda.AT	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0
Malex.gen!J	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	25	0	0	0	0	0	0	1	0
Obfuscator.AD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28	0	0	0	0	0	0	0
Rbot!gen	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	31	0	0	0	0	0	0
Skintrim.N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	0	0	0	0	0
Swizzor.gen!E	0	0	0	0	0	0	0	1	4	0	0	0	0	0	0	0	0	0	0	0	0	17	3	0	0	0
Swizzor.gen!I	0	0	0	0	0	0	1	12	1	0	0	0	0	0	0	0	0	0	0	0	0	4	8	0	0	0
VB.AT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	80	0	0
Wintrim.BX	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	18	0
Yuner.A	0	0	0	0	0	160	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

TEAM MEMBERS

Roll Number	Name
CB.EN.U4CSE22015	Guhanesh T
CB.EN.U4CSE22031	Nalan Krishna V
CB.EN.U4CSE22036	Prashanna R
CB.EN.U4CSE22124	Sajeev K