Malware Detection 👀

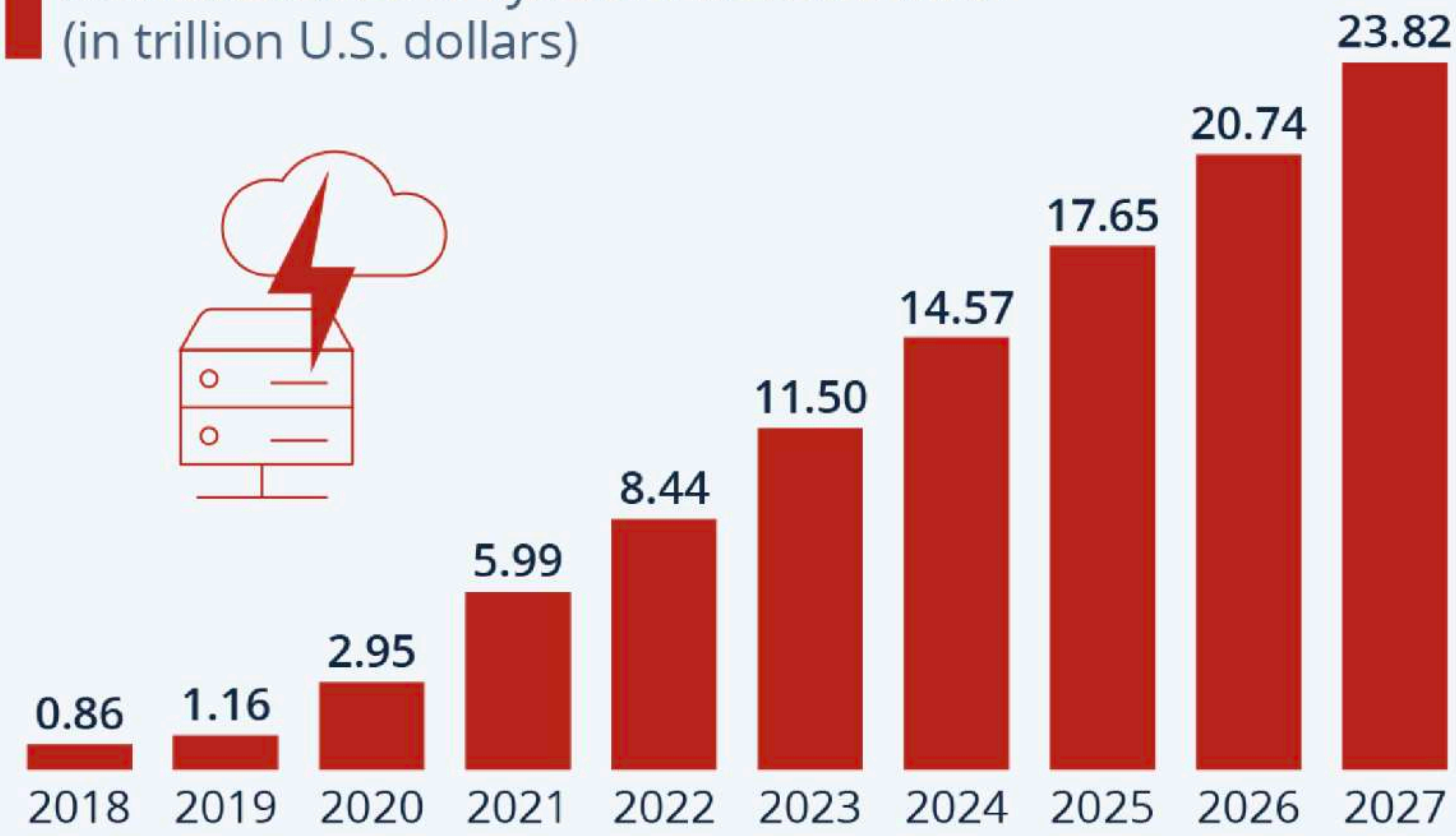
A Springboard Capstone 3 Submission

Greg McKenzie, November 27th 2023

Why Malware?

Cybercrime Expected To Skyrocket in the Coming Years

Estimated cost of cybercrime worldwide
(in trillion U.S. dollars)



As of November 2022. Data shown is using current exchange rates.
Sources: Statista Technology Market Outlook,
National Cyber Security Organizations, FBI, IMF

Who Might Care?

Industries of interest

- Banking and Finance
- Healthcare Industry
- Government and Public Sector
- Cloud Service Providers
- Legal and Compliance Firms
- And many more…

Data

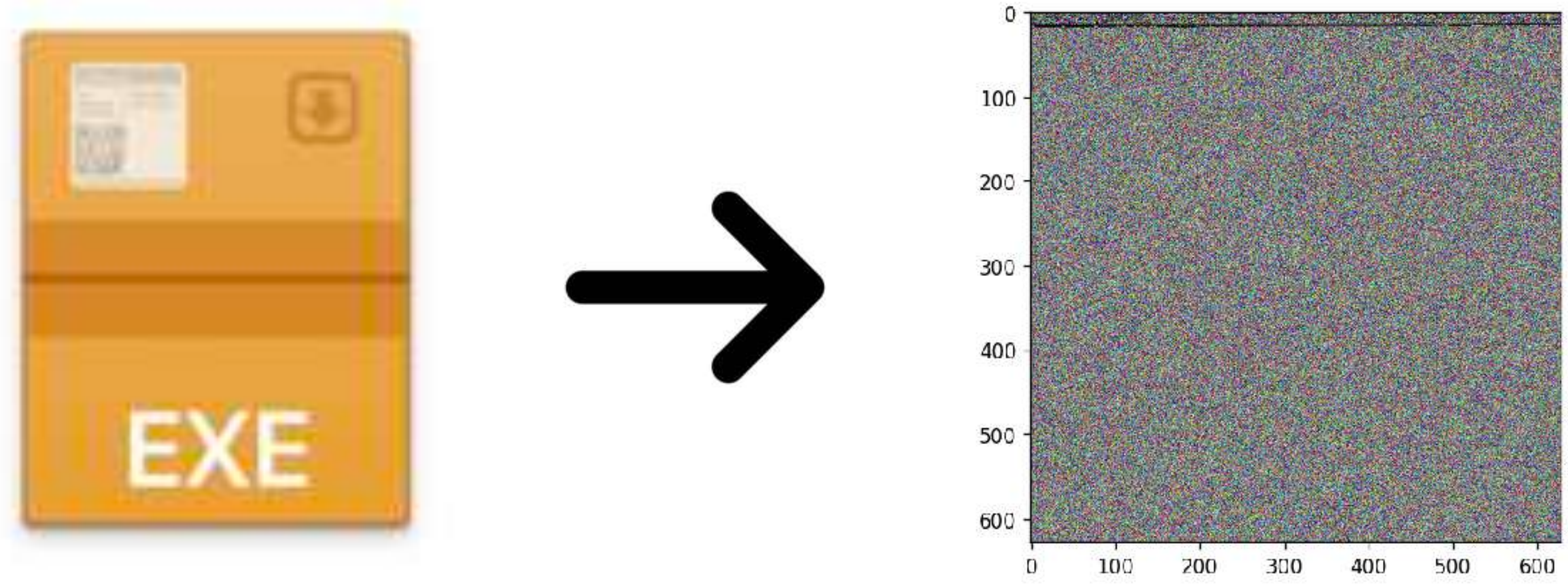
Competition dataset obtained from Kaggle

- 136 unique entries — 75 “Malicious”, 61 “Benign”
- “Malicious” entries were taken from The Zoo
- “Benign” entries were taken from PC Magazine’s The Best Free Software of 2020

Data

How are executable files rendered as images?

- Uses “Mallook” open source software

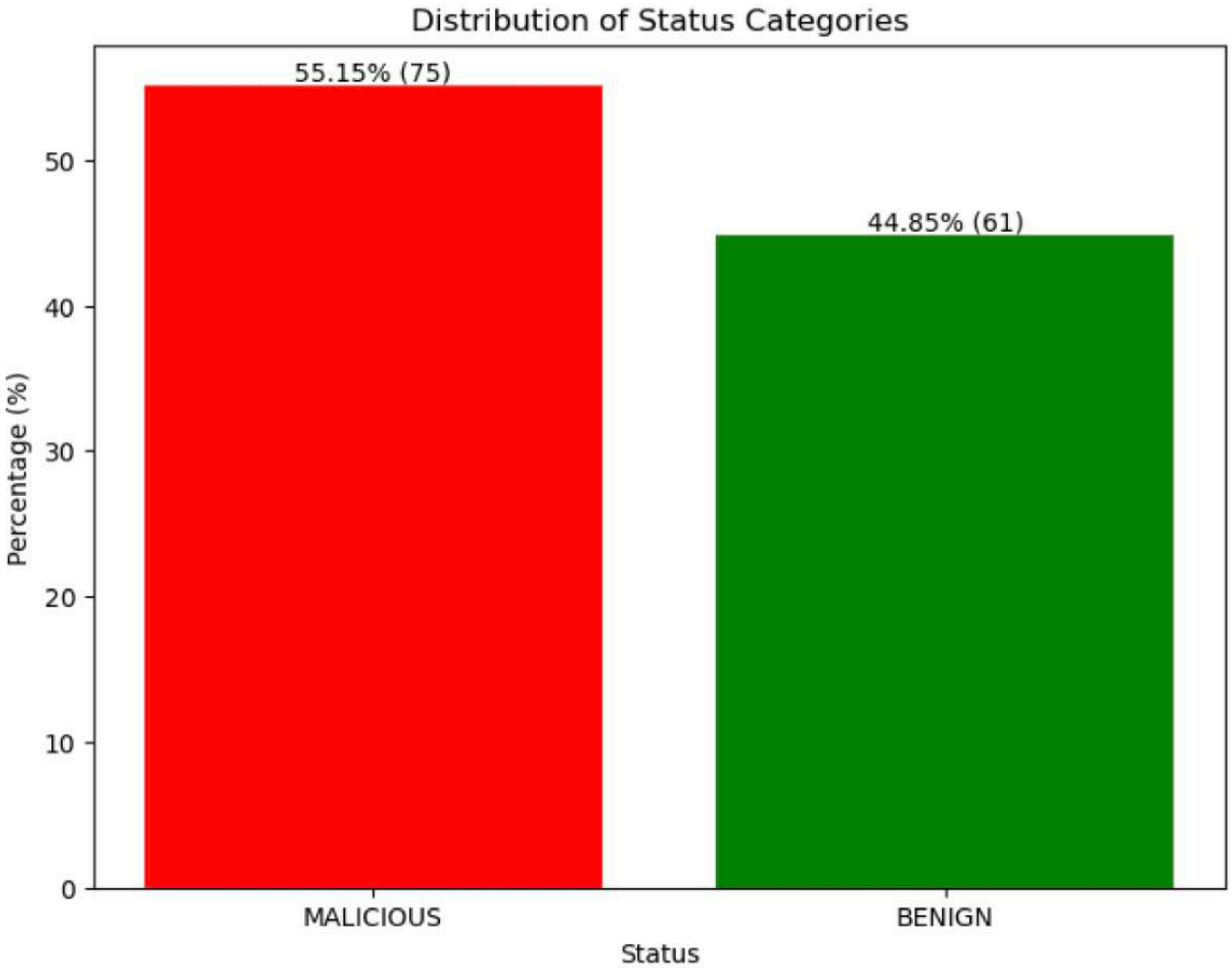


Data

What types of images do we get?

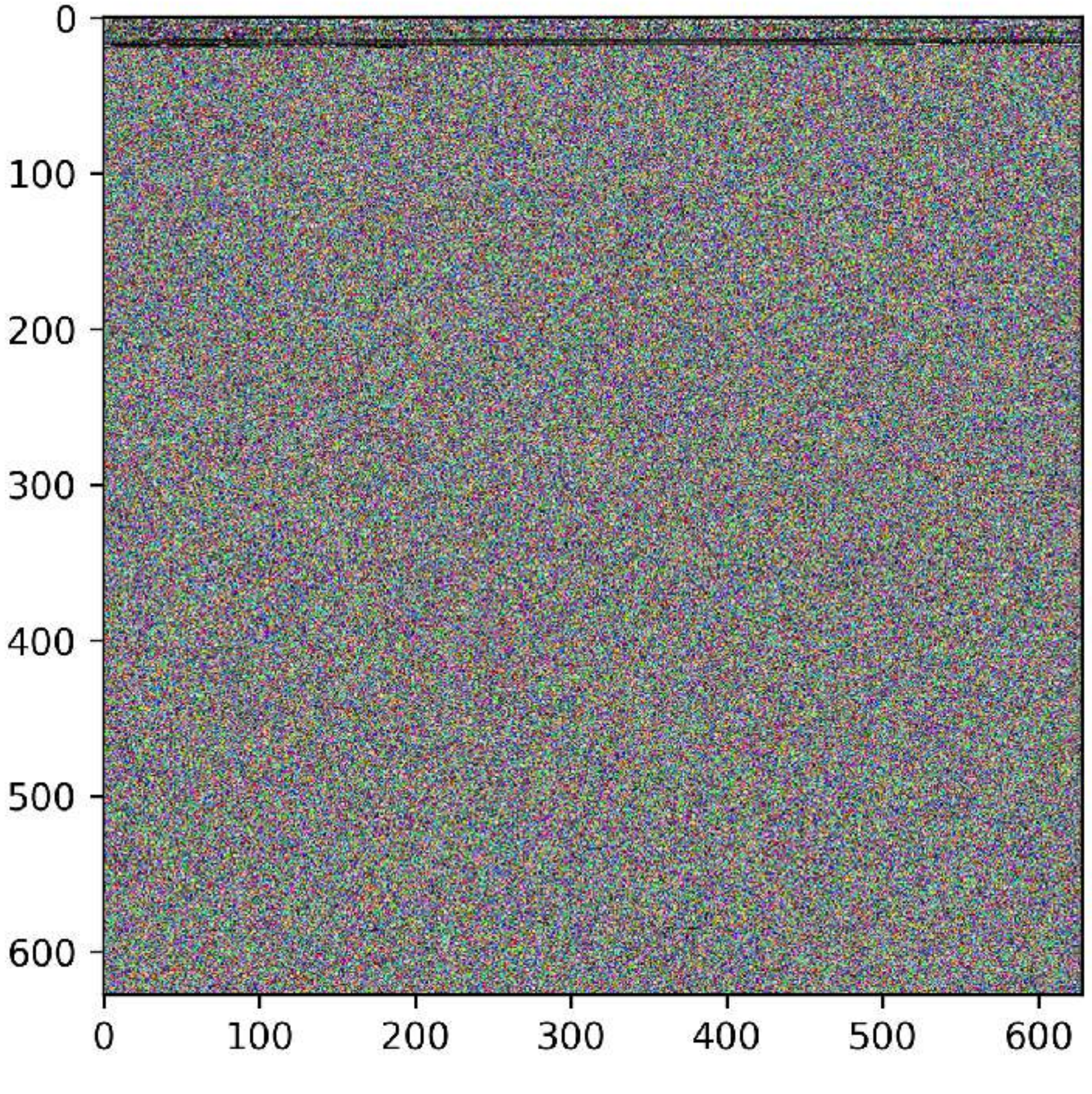
- 8 different images per portable executable file
- DPI: 120, 300, 600, 1200
- Distance: ‘nearest’, ‘lanczos’

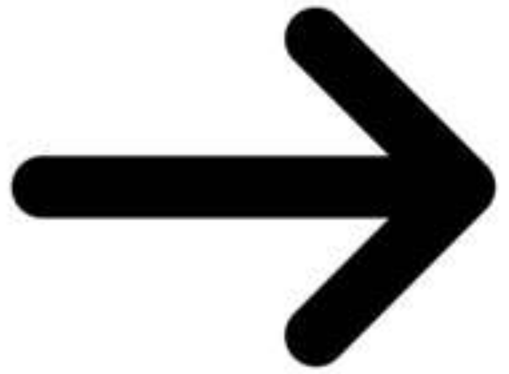
Data


Distribution of Status Categories
55.15% (75)
44.85% (61)
Percentage (%)
0
10
20
30
40
50
MALICIOUS
BENIGN
Status

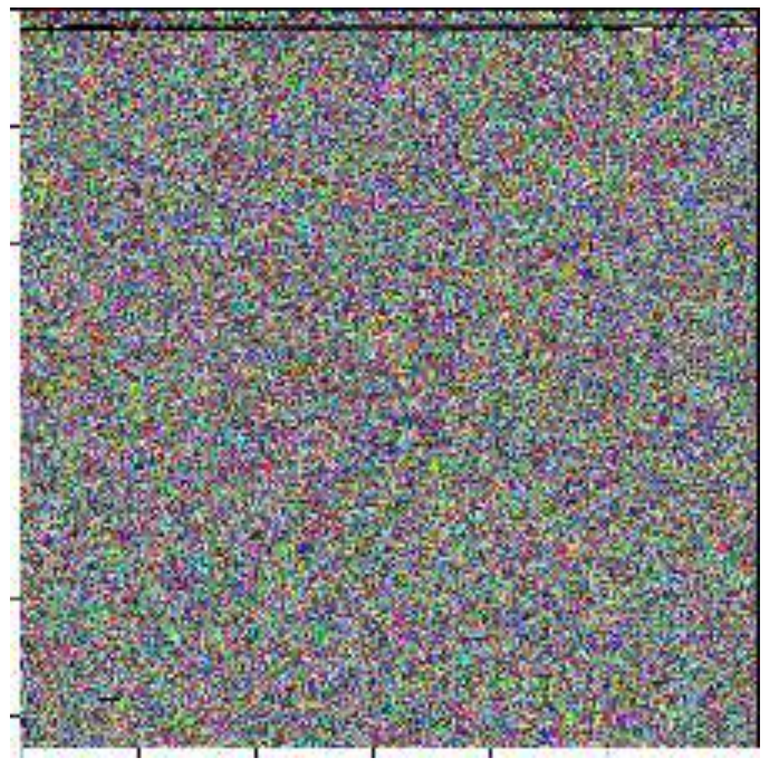
Data

Image processing

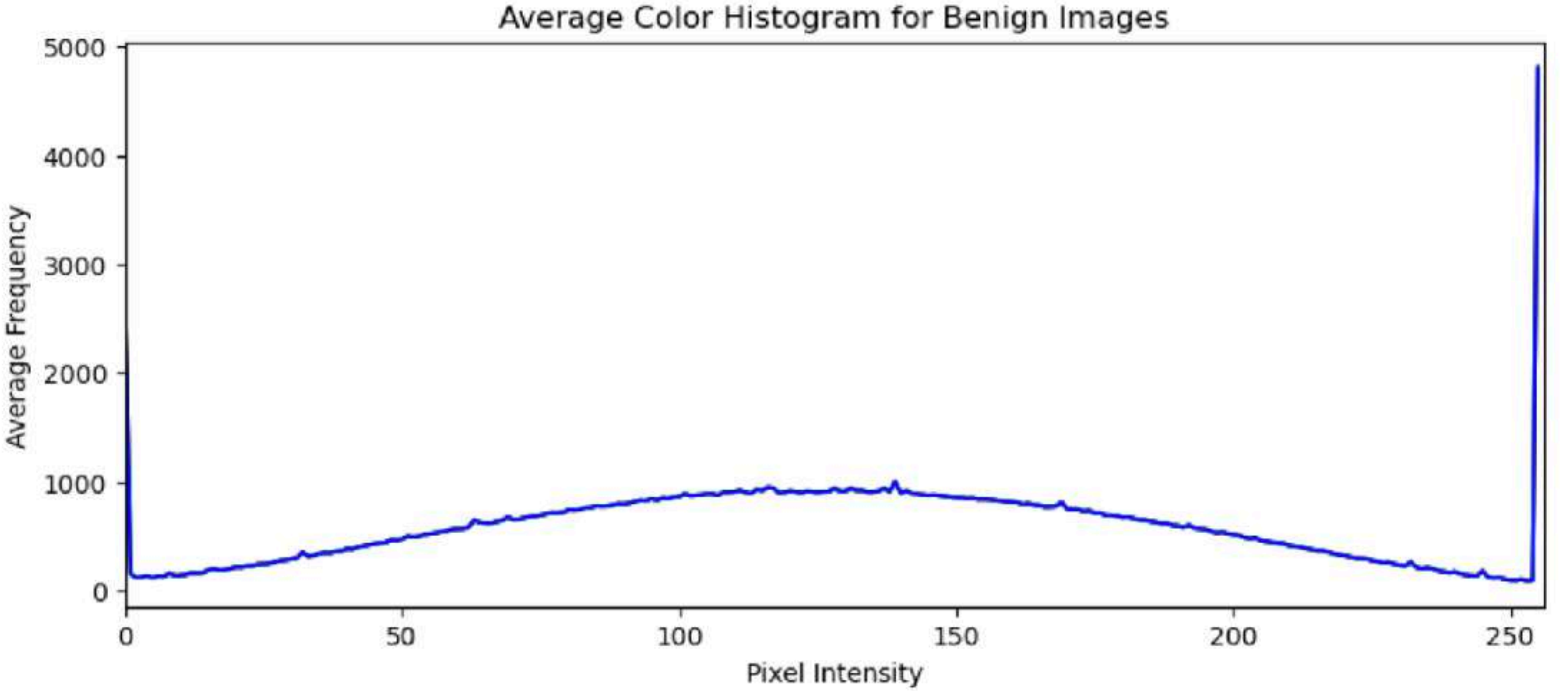


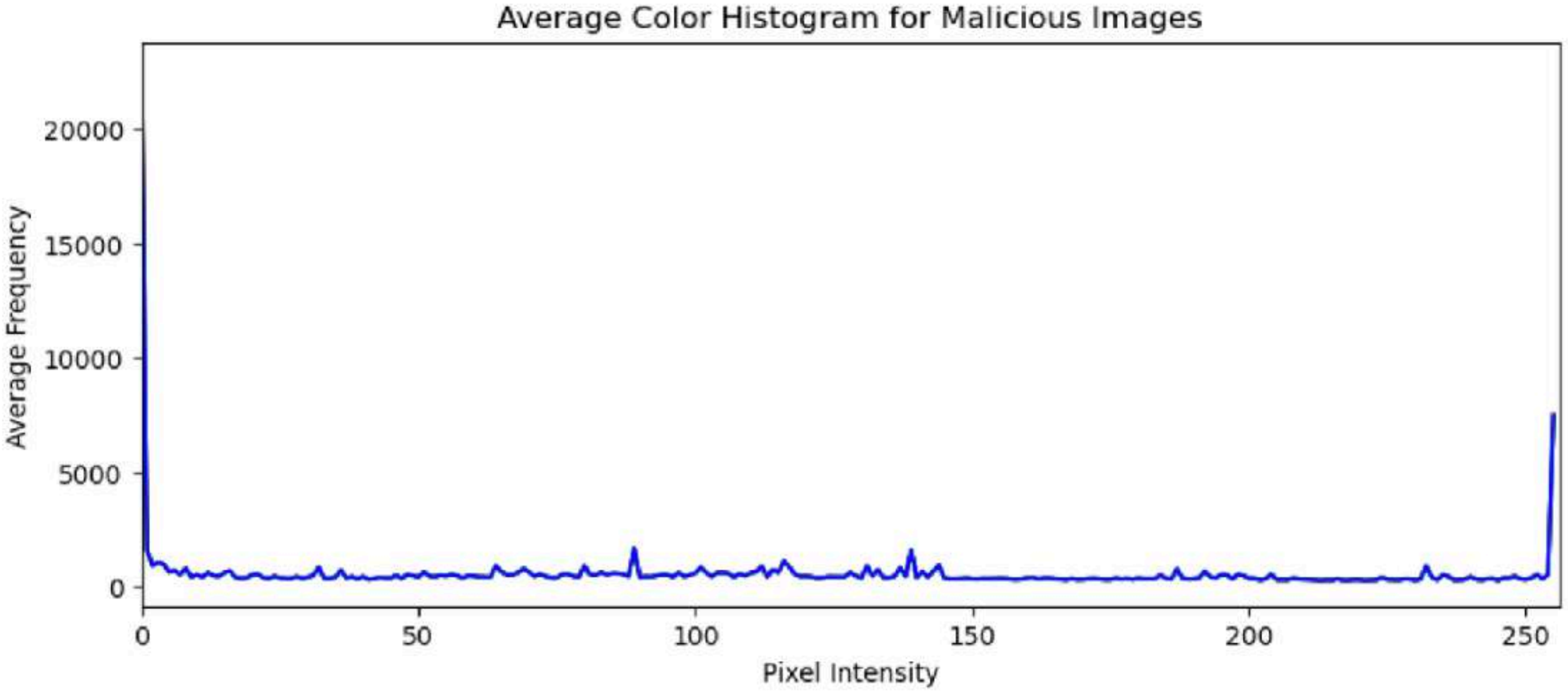


224 X 224

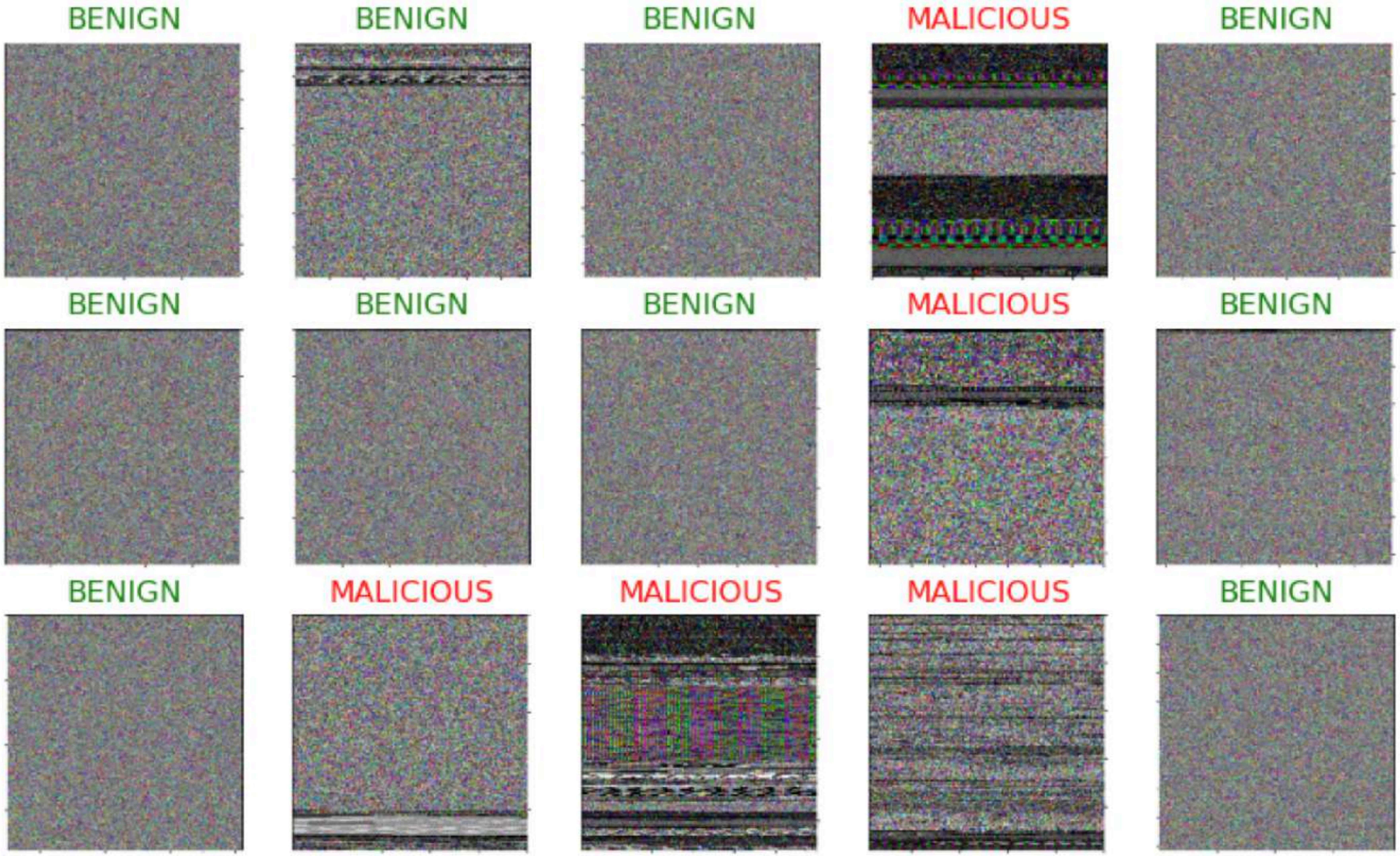


Data


Average Color Histogram for Benign Images
Average Frequency
0
1000
2000
3000
4000
5000
Pixel Intensity
0
50
100
150
200
250


Average Color Histogram for Malicious Images
Average Frequency
0
5000
10000
15000
20000
Pixel Intensity
0
50
100
150
200
250

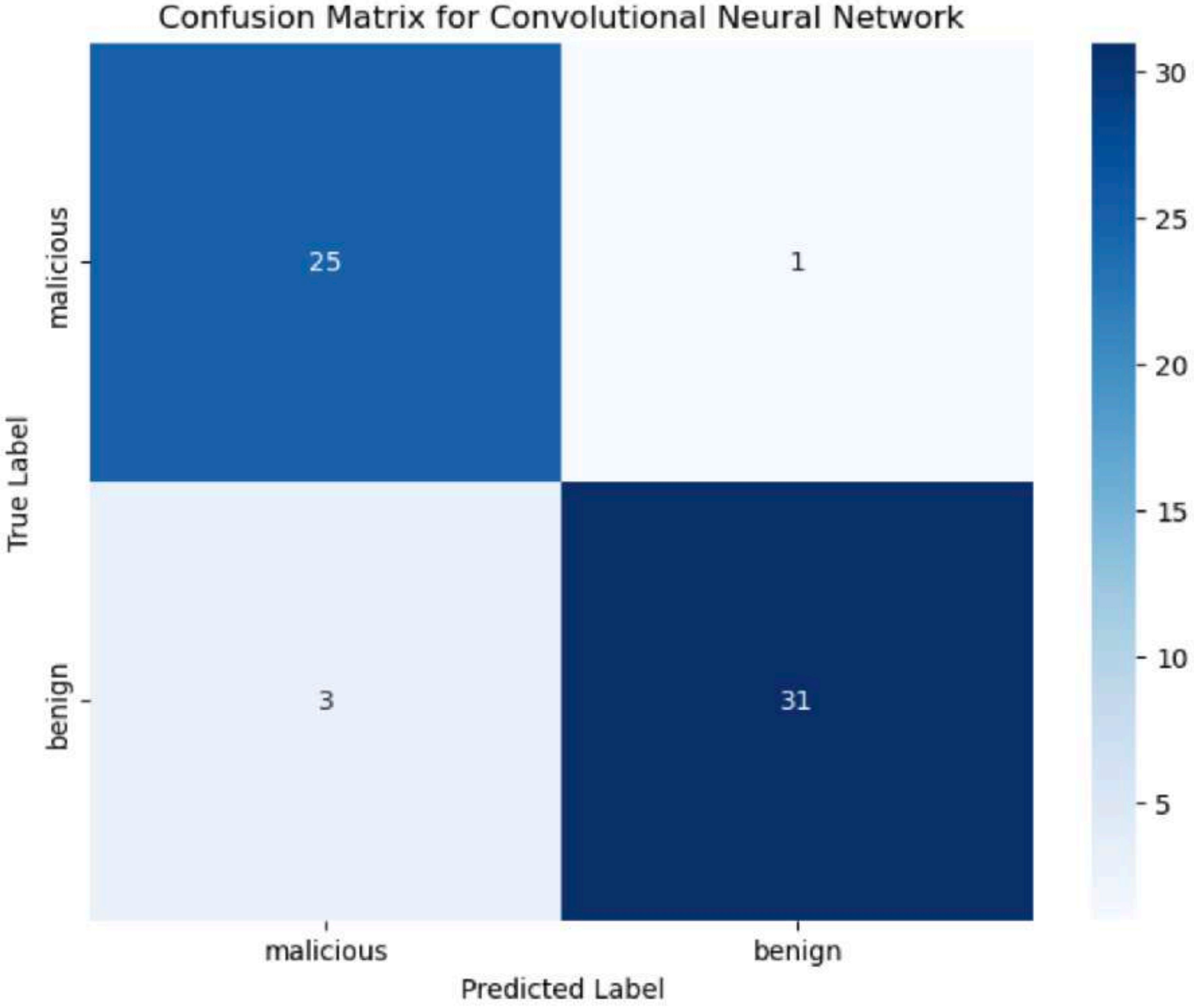
Data


BENIGN
BENIGN
BENIGN
MALICIOUS
BENIGN
BENIGN
BENIGN
BENIGN
MALICIOUS
BENIGN
BENIGN
MALICIOUS
MALICIOUS
MALICIOUS
BENIGN

Modelling

Convolutional Neural Network

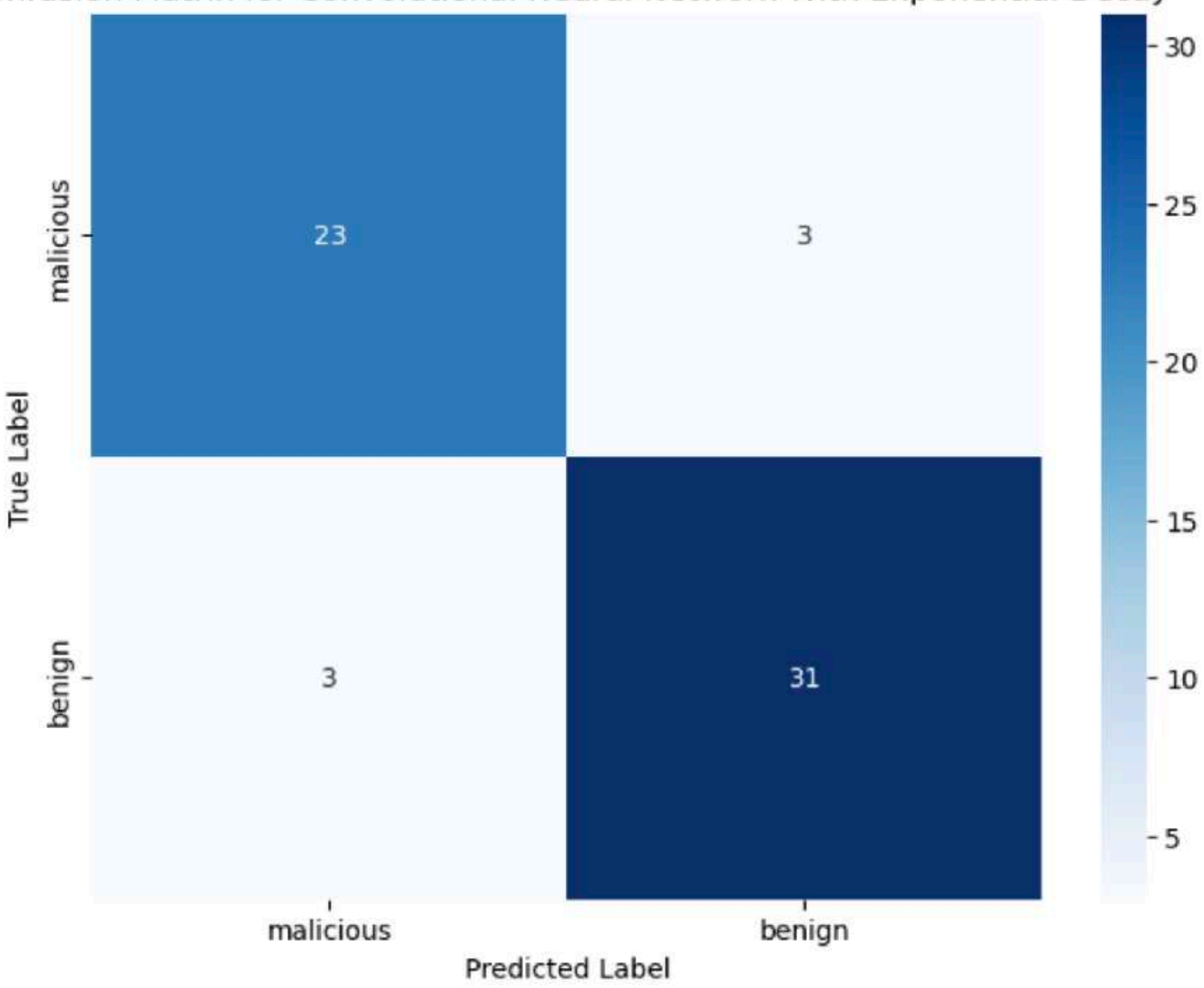
- Accuracy: 0.93
- Precision for ‘malicious’: 0.89
- Recall for ‘malicious’: 0.96
- F1 Score for ‘malicious’: 0.93
- Prediction Wall Time: 647 ms
- 286.9 MB



Modelling

Convolutional Neural Network with Exponential Decay

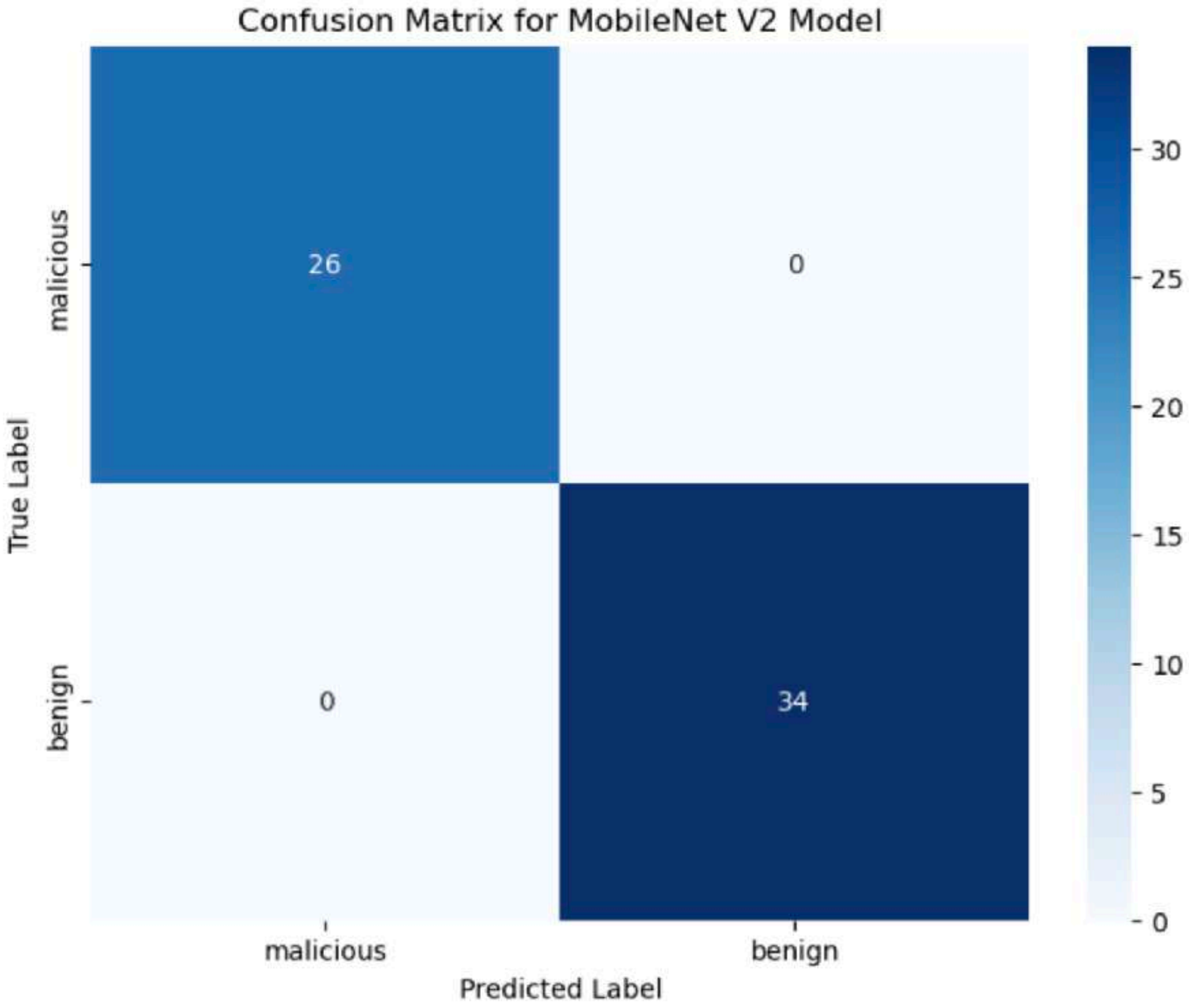
- Accuracy: 0.90
- Precision for ‘malicious’: 0.88
- Recall for ‘malicious’: 0.88
- F1 Score for ‘malicious’: 0.88
- Prediction Wall Time: 438 ms
- 286.9 MB


Confusion Matrix for Convolutional Neural Network With Exponential Decay

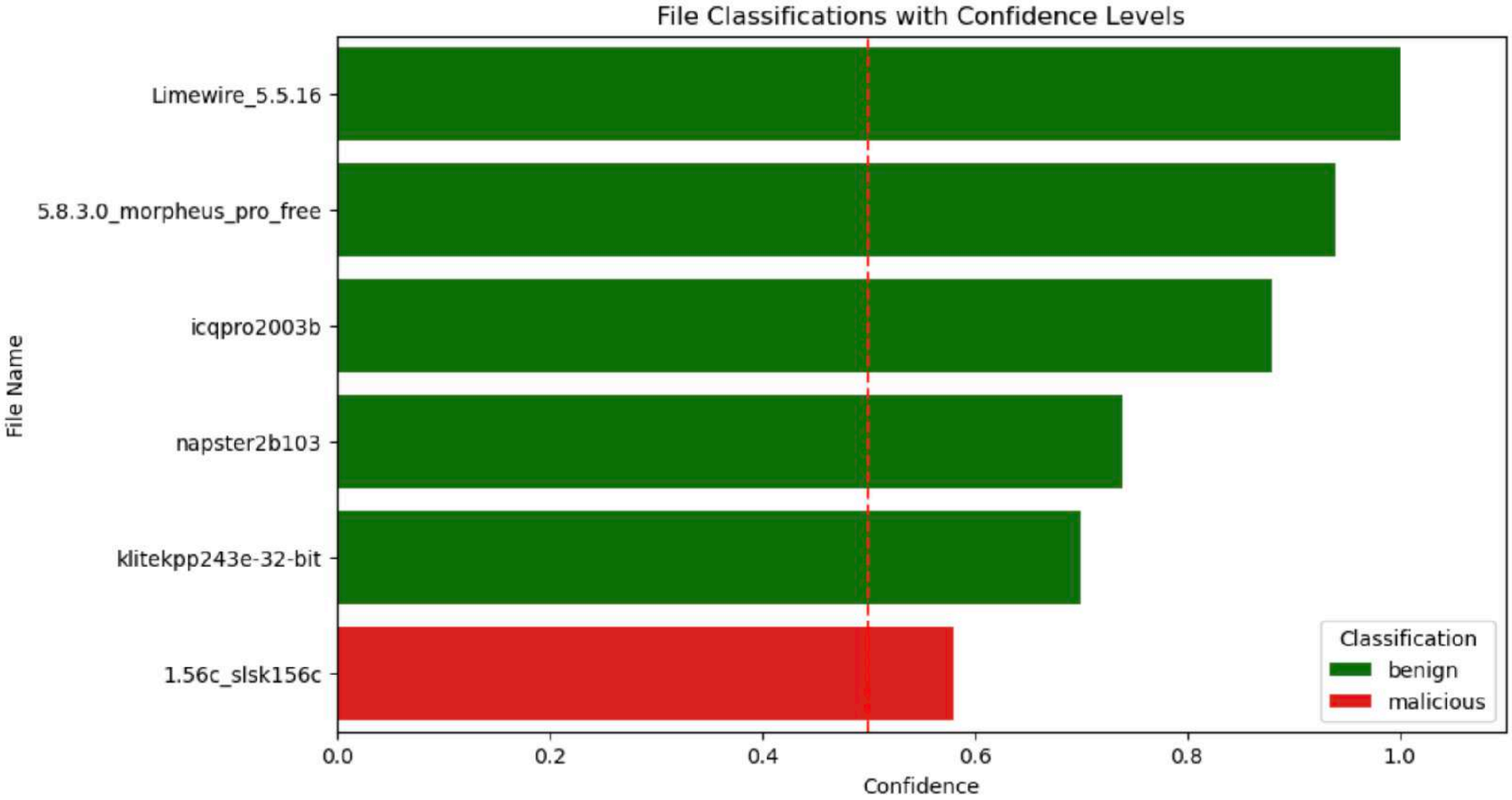
Modelling

MobileNet V2 (Transfer Learning)

- Accuracy: 1
- Precision for ‘malicious’: 1
- Recall for ‘malicious’: 1
- F1 Score for ‘malicious’: 1
- Prediction Wall Time: 1.27 s
- 22 MB



Predictions



Improvements Going Forward

What's next?

- Generate more data
- Create API and front end
- Make it open source?

Malware Detection 👀

A Springboard Capstone 3 Submission

Greg McKenzie, November 27th 2023