EUROPE 2021

november 10-11, 2021

ARSENAL

#BHEU @BlackHatEvents

InOri

Detecting defacement attacks with deep learning

By Nguyen Hoang, Manh Pham & Dong Duong

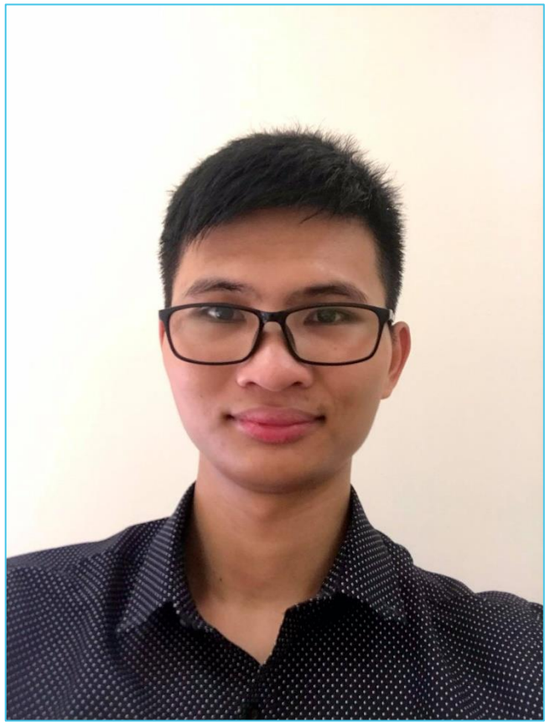
Nguyen Hoang

I'm a final year student at the Academy of Cryptography Techniques, majoring in information security.

I used to work as a System administrator, DevOps. I specialize in building and developing CDN systems, but gradually fell in love with security tinkering.

Currently, working as a security solutions consultant and a penetration tester.

Contact: https://www.linkedin.com/in/hoangtrungnguyen



Manh Pham

A security researcher from Vietnam. With over 4 years dedicated in Information security.

Acknowledged by Microsoft, Google, Facebook, Apple, Offensive Security, etc.

Now, I work for NashTech as a Penetration Tester and also join OWASP as a Contributor

Contact: https://linkedin.com/in/manhnho/

Dong Duong

A student from Phan Dinh Phung highschool.

Software developer, Linux system administrator and security researcher.

Recently acknowledged by Microsoft

Contact: https://cu64.github.io/

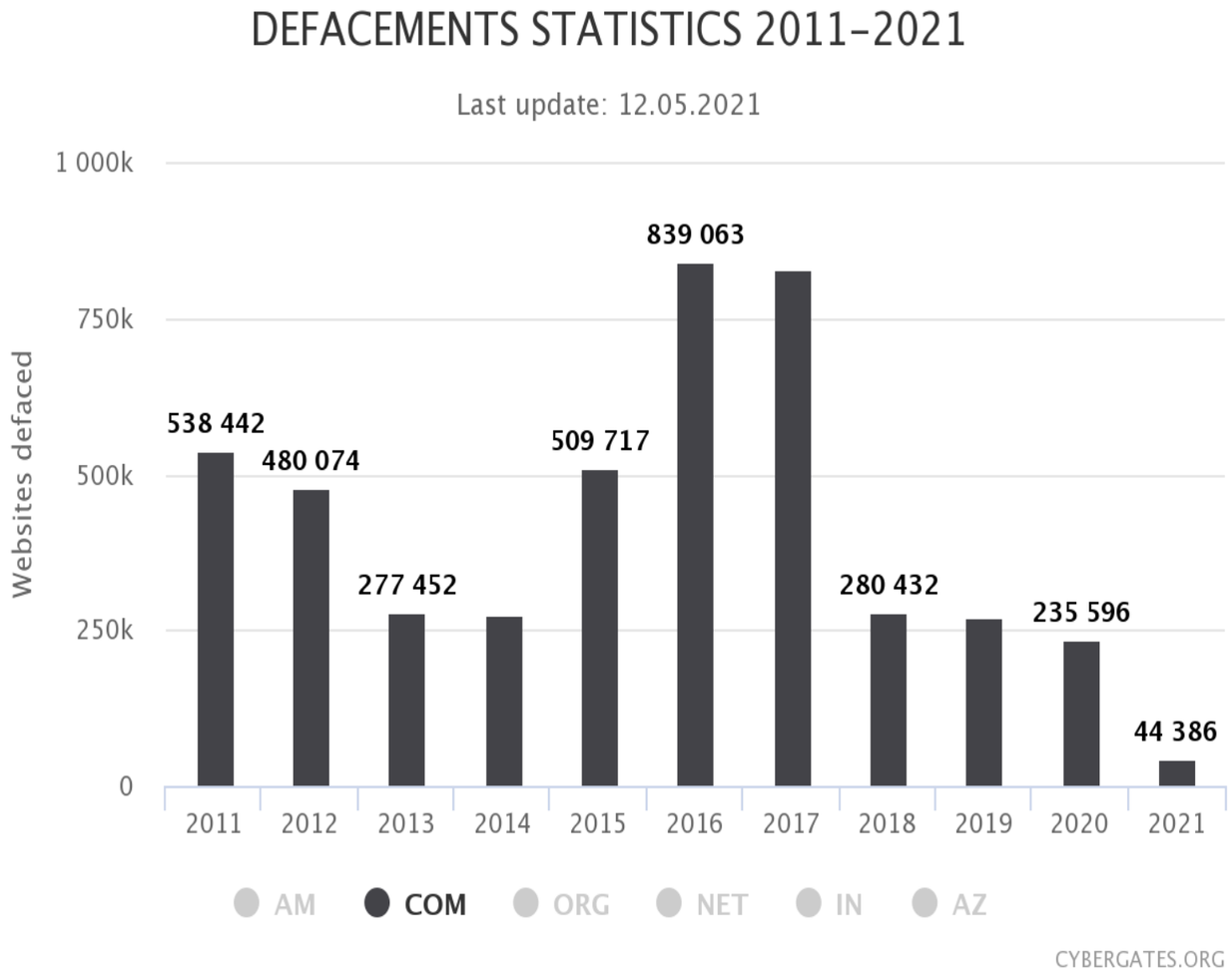
Agenda

1. Motivation
2. Building model Machine Learning
3. Alert system development
4. Experiment and evaluation

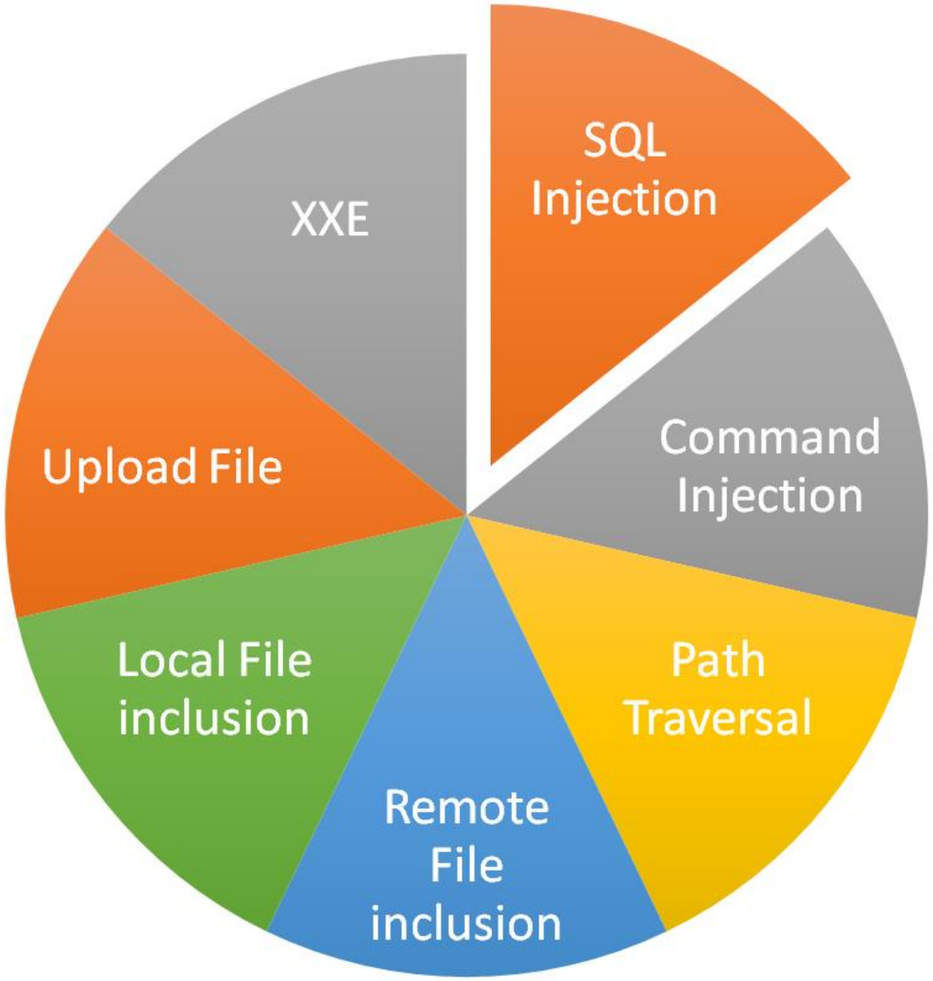
1. Motivation
2. Building model Machine Learning
3. Alert system development
4. Experiment and evaluation

Motivation

- Goals:
  - Cause political conflict
  - For fun
  - Warn the system administrator



Attack Techniques Observed

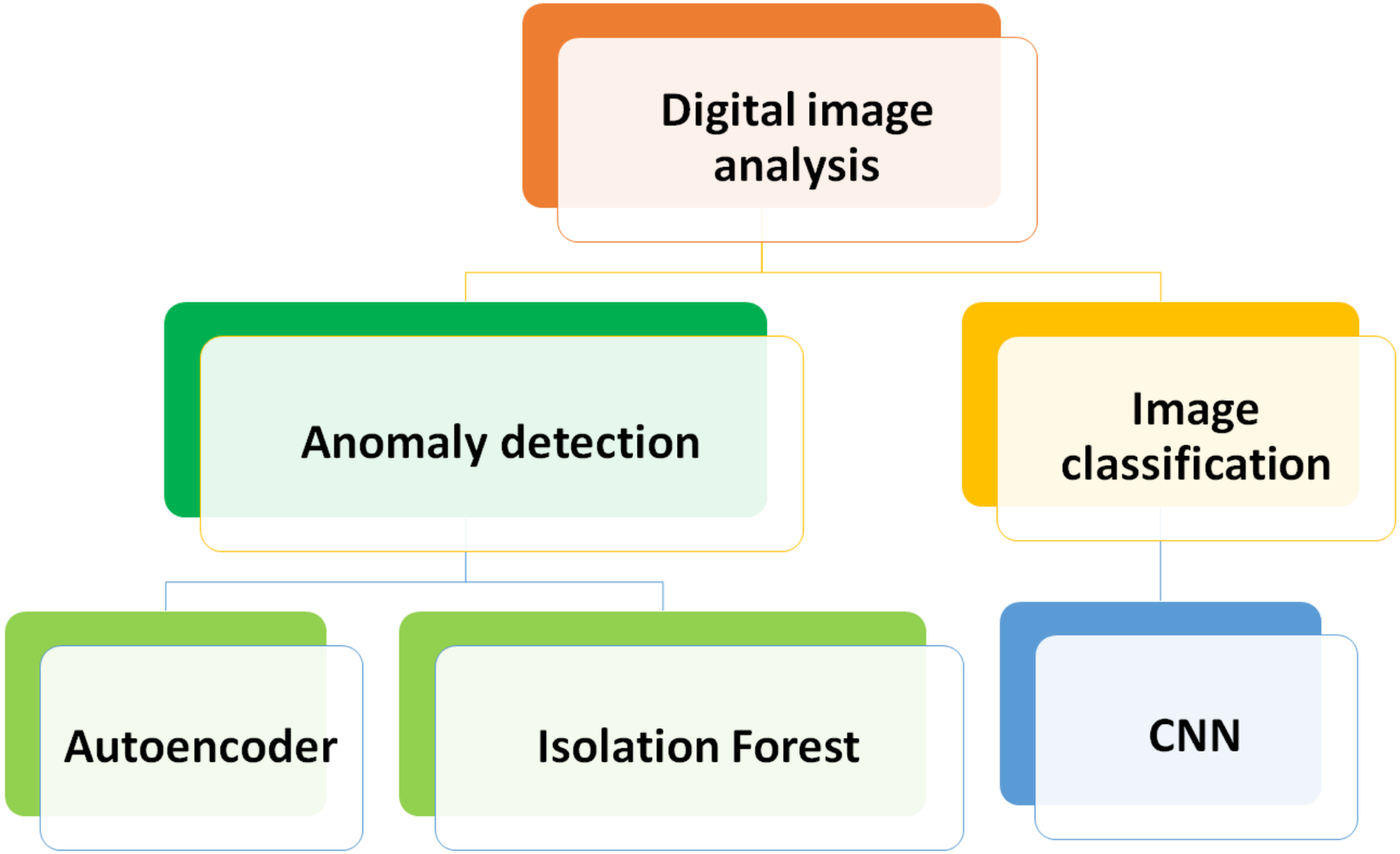


Critical web vulnerabilities

Detection Techniques

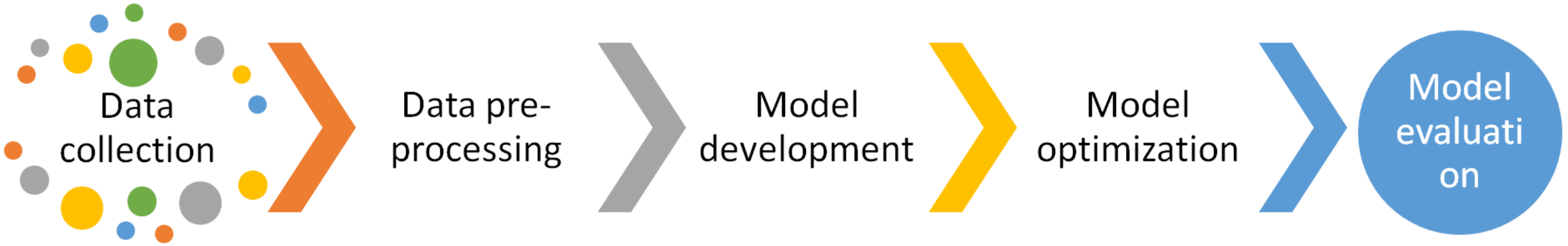
Hash	• Comparing hashes
Signature	• Determine attacker's signature • Determine web page's signature
Compare differences	• Compare changes in source code • Compare DOM Tree
Machine Learning	• Anomaly detection • Image classification

Approach


Digital image analysis
Anomaly detection
Image classification
Autoencoder
Isolation Forest
CNN

1. Motivation
2. Building model Machine Learning
3. Alert system development
4. Experiment and evaluation

Approach


Data collection
Data pre-processing
Model development
Model optimization
Model evaluation

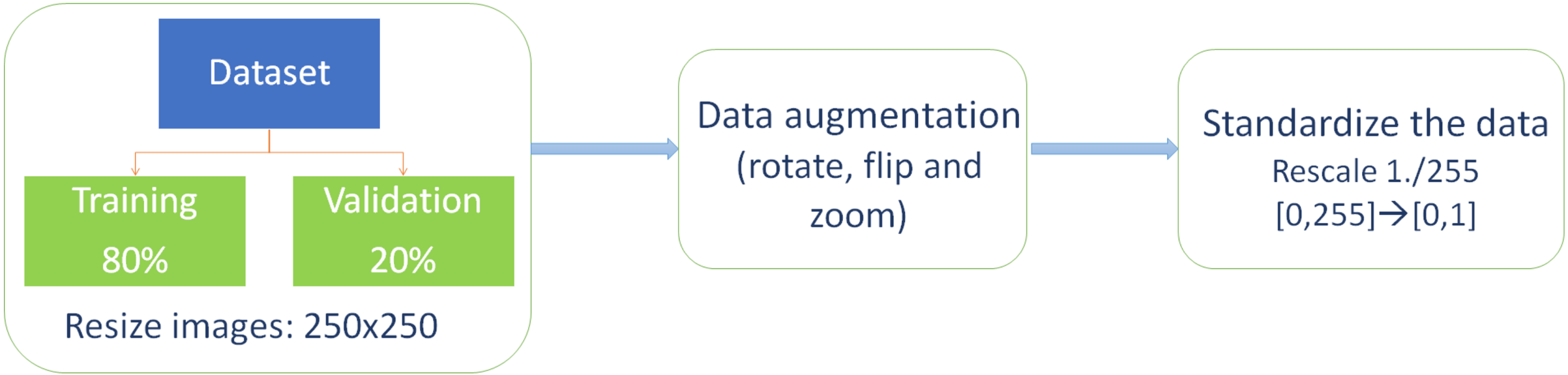
Training data collection

- Normal website data
  - moz.com/top500
  - github.com/GSA/govt-urls

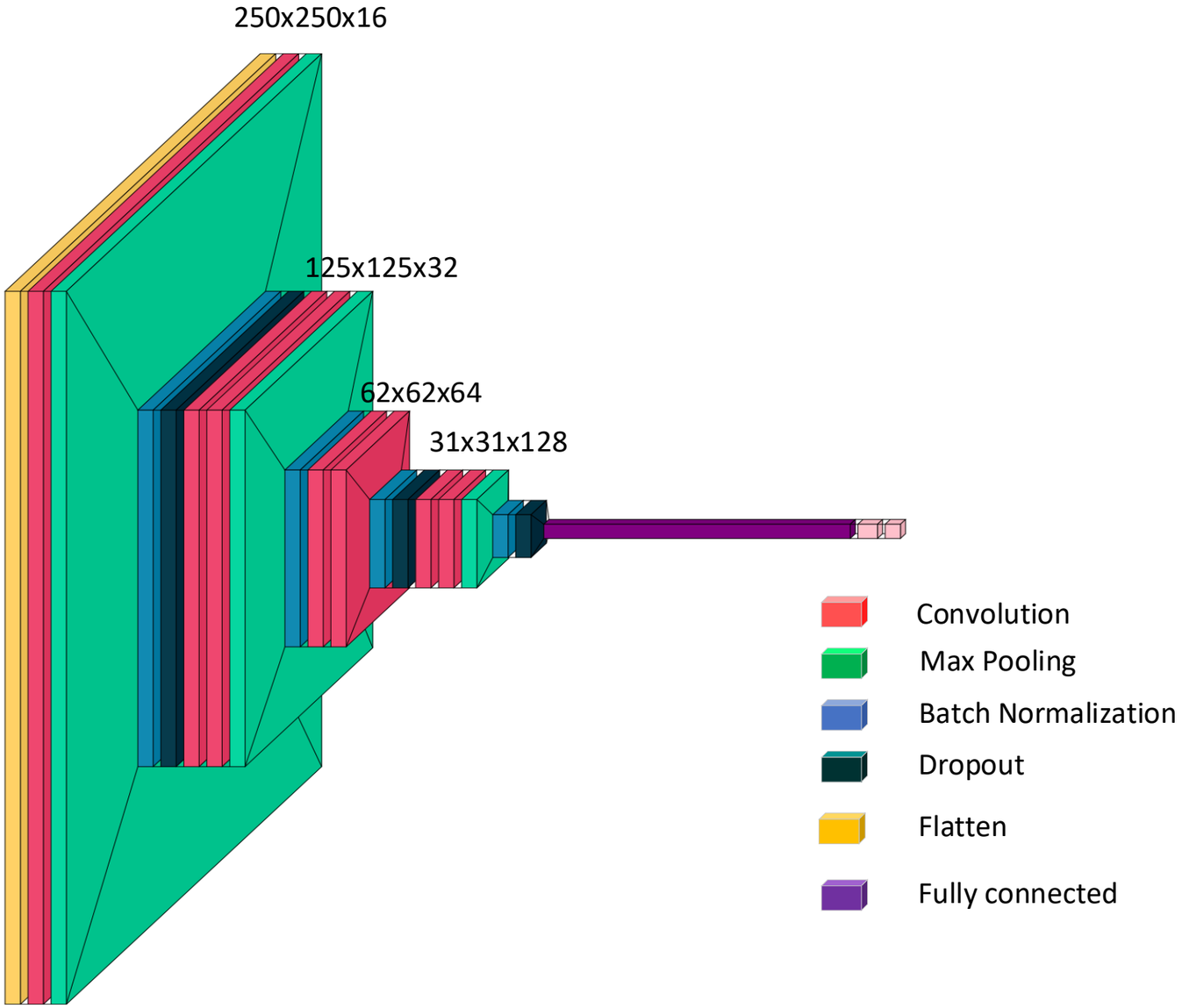
- Defaced website data
  - mirror-h.org
  - zone.kurd-h.org
  - www.zone-h.org
  - www.xatrix.org/defac.php

Data pre-processing

- Dataset
  - Clean: 7,277 pics
  - Deface: 4,954 pics

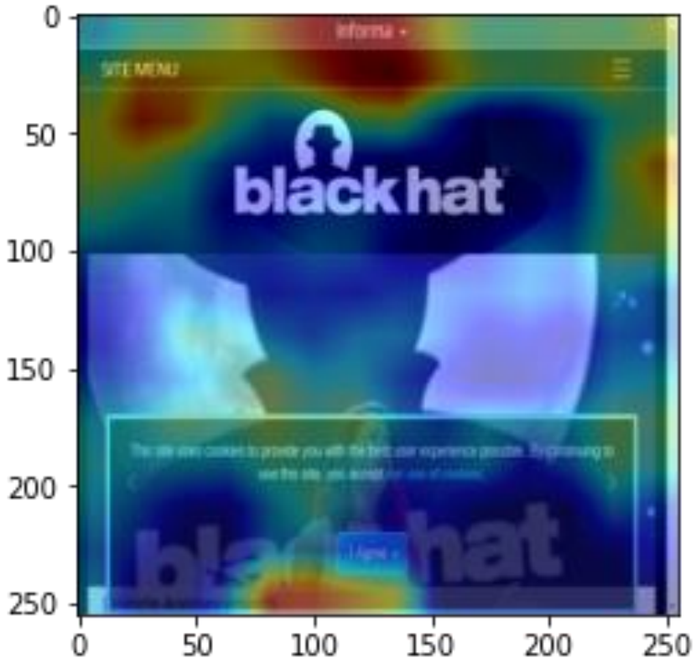


Building detection model

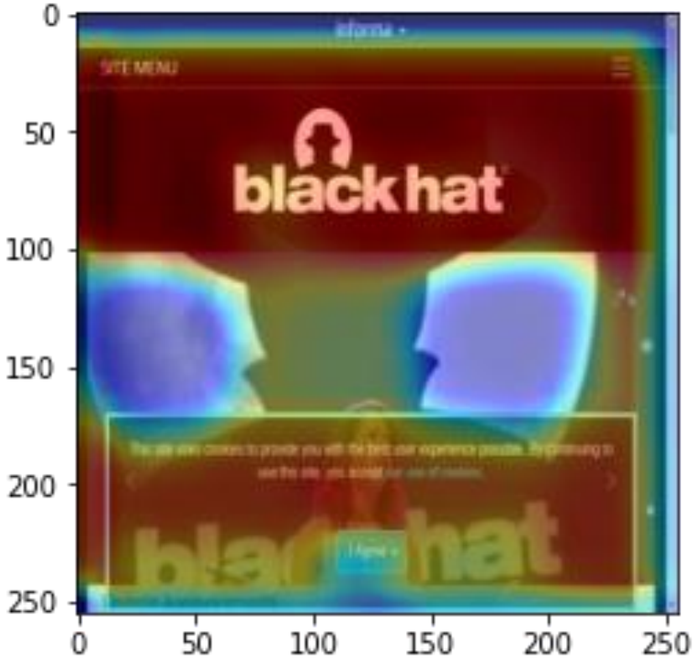


Model of the Convolutional Neural Network

Building detection model



VGG 16



My Model

Model optimization

- Reduce overfitting
  - Add Dropout layers
  - Data augmentation
  - Add Batch Normalization layers

Model Evaluation

❖ Evaluation criteria:

Accuracy

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = 98.01$$

False positive rate

$$FPR = \frac{FP}{FP + TN} = 3.73$$

Positive predictive value

$$PPV = \frac{TP}{TP + FP} = 94.00$$

True positive rate

$$TPR = \frac{TP}{TP + FN} = 98.12$$

False negative rate

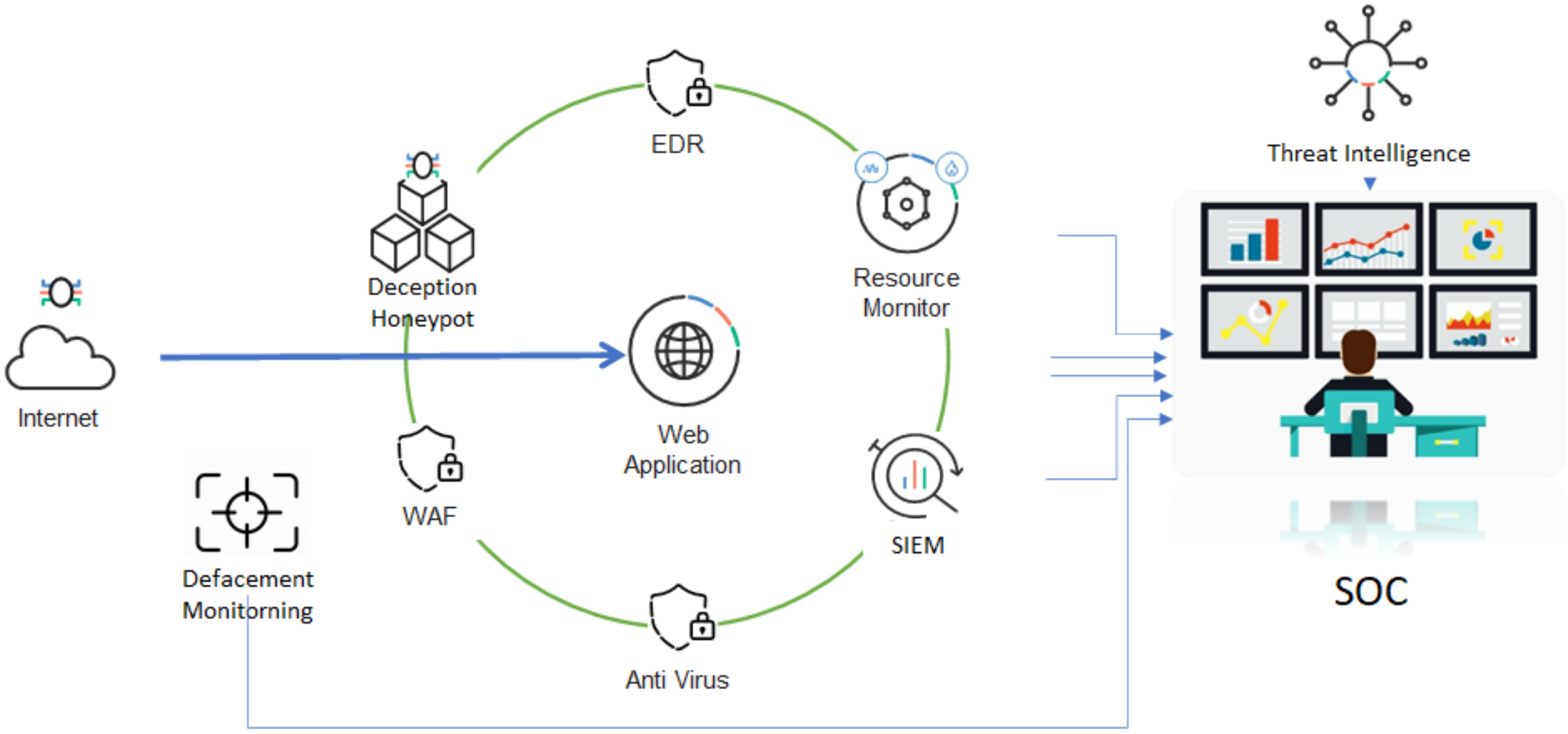
$$FNR = \frac{FN}{FN + TP} = 1 - TPR = 1.88$$

F1 Score

$$F1 = 2 \times \frac{PPV \times TPR}{PPV + TPR} = 96.02$$

1. Motivation
2. Building model Machine Learning
3. Alert system development
4. Experiment and evaluation

Alert system development



Some security solutions for a web application

Alert system development

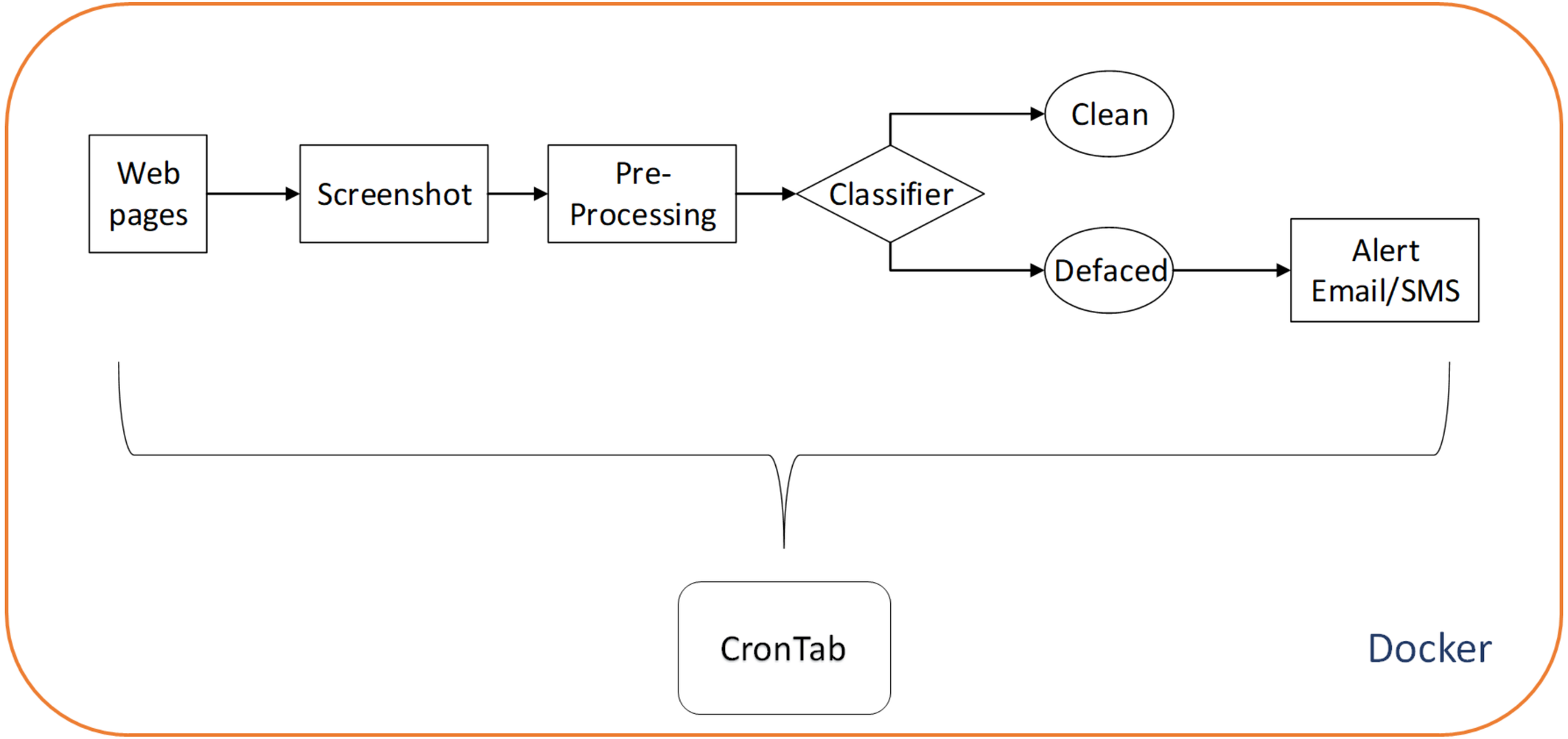
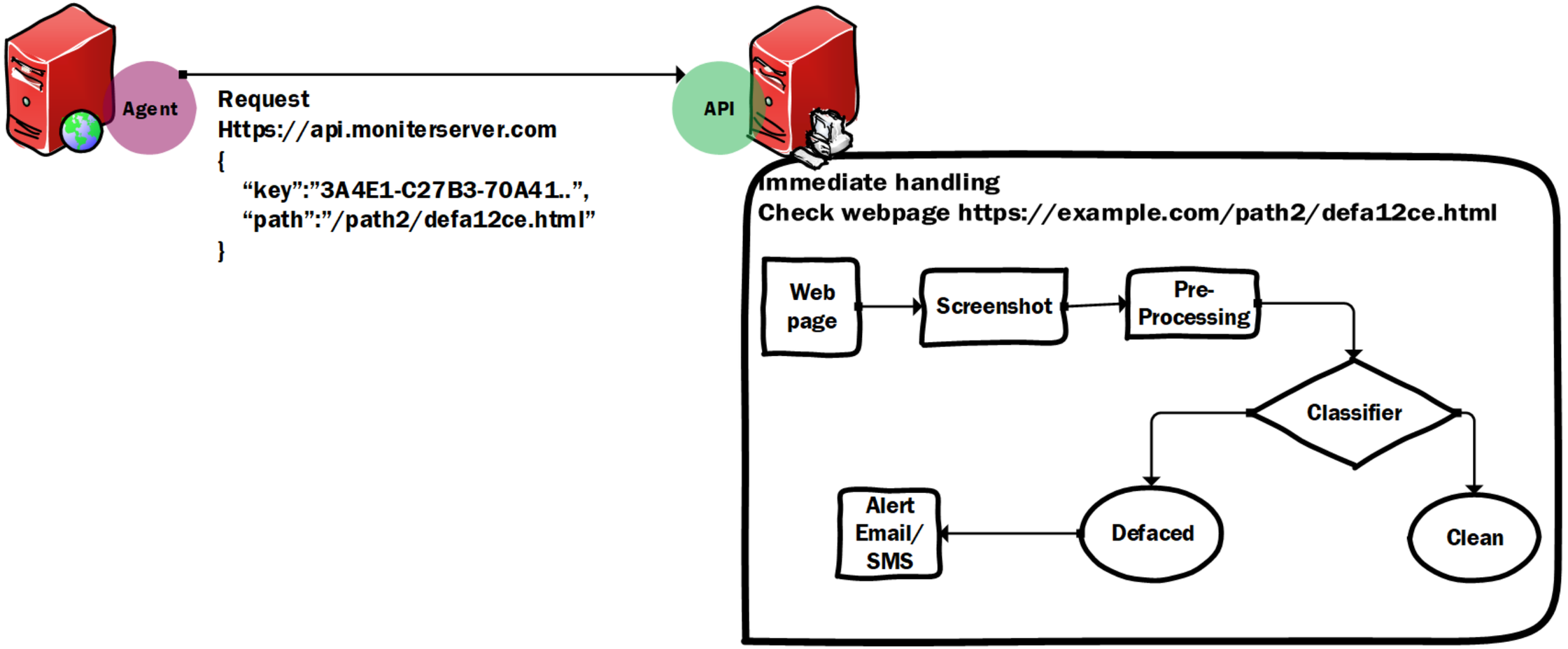


Diagram of the detection and warning system

URLs reported on Zone-h

	R		[redacted].th/wh.html	Linux
	R		[redacted]on.go.th/silence.html	Linux
	R		[redacted]en-yvelines.fr...	Linux
	R		[redacted]rpl.txt	Linux
	R		[redacted]d/haha.html	Linux
	R		[redacted].id	Linux
	R		[redacted]al/kurd.txt	MacOSX
M			[redacted]/qq.html	Win 2012
M	R		[redacted].org/qq.html	Win 2012

Agent deployment



1. Motivation
2. Building model Machine Learning
3. Alert system development
4. Experiment and evaluation