HOST 2022 Microelectronics Security Challenge: Supply Chain Security Track

Navid Asadi, Brian Knight, Yousef Iskander, and Ujjwal Guin
Emails:nasadi@ufl.edu, brian.knight@microsoft.com, yousefi@microsoft.com, and ujjwal.guin@auburn.edu

Overview: The lack of traceability in the globalized electronics supply chain results in the infiltration of various counterfeit electronic parts, including recycled, remarked, overproduced, cloned, out-of-spec/defective, forged documentation, and tampered types and pose a severe threat to the security of our critical infrastructures. Among them, recycled, remarked, and cloned parts constitute most counterfeit incidents. Over the years, a class of solutions has been proposed to mitigate the widespread infiltration of these fake parts. Physical Inspection methods have gained a lot of attention due to their one-size-fits-all nature as the same methods can be applied to all types of parts (analog, digital, memory, FPGAs. etc.). Among various modalities, including optical, X-ray, thermal, electron beam microscopy, etc., optical imaging is one of the fastest and most affordable modalities.

This challenge requires the competitors to develop a highly-accurate automated method to identify counterfeit ICs from their labeled optical images. The minimum accuracy requirement is at least 60%. Detailed requirements are given in the following table.

Accuracy Range	**Competition Ranking Categories**
90%-100%	Gold
More than or equal to 80%, but less than 90%	Silver
More than or equal to 70%, but less than 80%	Bronze
More than or equal to 60%, but less than 70%	Qualified, but no specified ranking
Less than 60%	Disqualified

Dataset Overview

The image data presented here is collected using the FICS lab facilities with two different image acquisition modalities:

a) DSLR system with color normalization,
b) Zeiss Stemi 508 Stereo Microscope.

Training Data:

The challenge package includes a folder called **'train',** which contains two subfolders named **'DSLR'** and **'STEMI_508'.** Inside the **'DSLR'** folder, there are **40** high-resolution images (both **authentic** and **counterfeit**) of the front and the back surface of **3** different types of ICs. On the other hand, the **'STEMI_508'** folder includes **60** high-resolution images (both **authentic** and **counterfeit**) of the front and the back surface of the same **3** types of ICs mentioned for DSLR. The **'train'** folder also contains an annotation file called **'train.csv',** which has **8** columns as mentioned below:

- **id** - Unique identifier for each sample image. For example, the id **A-M-16DIP-00F-D** has 5 portions which contain 5 information of the sample:

        - **A –** Authentic
        - **M –** Mouser Electronics (Vendor Acronym)
        - **16DIP –** 16 pins dual in-line package (Package type)
        - **00F –** 00 refers to the first image, i.e., serial number, and F refers to the Front side image
        - **D –** DSLR (image acquisition modality)
    - **sample_name –** Manufacturer Product Number. For example, **STM32F105R8T7**
    - **manufacturer -** Manufacturing company's name. For example, **STMicroelectronics**
    - **vendor –** Name of the entity that supplies the product (IC). For example, **Digi-Key Electronics**
    - **product_type –** Type of the IC. For example, **DAC** (Digital-to-Analog converter) or **Embedded – Microcontrollers**
    - **package_type –** Type of the IC package. For example, **16DIP** (16 pins dual in-line package)
    - **modality –** Type of image acquisition modality. For example, **DSLR** or **Stereo Microscope**
    - **label – 0** (authentic) or **1** (counterfeit)

Test Data:

The challenge package includes another folder called **'test',** which contains two subfolders named **'DSLR'** and **'STEMI_508'**. Inside the 'DSLR' folder, there are **10** high-resolution images (both **authentic** and **counterfeit**) of the front and the back surface of a single type (different from training data) of ICs. On the other hand, the **'STEMI_508'** folder includes **10** high-resolution images (both **authentic** and **counterfeit**) of the front and the back surface of the same type of IC mentioned for DSLR. The **'test'** folder also contains an annotation file called **'test.csv'**, with the same column information mentioned in the training data.

Submission Criteria:

- The participants are required to submit a zip file containing codes, sample submission, demo, and presentation. The file name should be "team_name_HOST_2022_SCS.zip using OneDrive.
- **Code:** Can be in any language, i.e., python (recommended) /R/Java/C++ etc.
    - A GitHub repository containing all necessary codes/libraries/helper functions to train, test, and visualize with a complete and comprehensive README file to implement on the hold-out test data.
    - If the README file and Demo video (see below) don't work/cannot help in successful implementation as claimed in the presentation (see below) and demo (mentioned later), the competitor will receive a penalty.
- **Sample submission:** a 'sample_submission.csv' file containing two columns –
    - id - Unique identifier for each sample test image as in 'test.csv'
    - predicted_label – 0 (if the prediction is 'authentic' for a certain sample test image), 1 (if the prediction is 'counterfeit' for a certain sample test image)
- **Demo:** a demo video link (upload in OneDrive) to demonstrate how to run the developed system (train, test, and visualize).
    - Time limit: minimum - 5 mins, maximum - 30 mins, recommended – 15 mins
- **Presentation:** A PowerPoint presentation of a **minimum 5 to maximum 15 slides** in format to show the outcome (must include the method, results, and discussion).

Evaluation Criteria***:

Total: 100 pts

- **Training Strategy (e.g., Cross-validation, Data augmentation, Pre-processing, etc.) - 20 pts**

- **Model Complexity – 10 pts**
- **Model Performance (Accuracy*, Confusion Matrix, Precision, Recall, F-Score, ROC AUC Score,** The Matthews Correlation Coefficient (MCC) **) – 30 pts**
    - Highly accurate on **only DSLR** images – **10 pts**
    - Highly accurate on **only Stereo Microscope** images – **10 pts**
    - Highly accurate on **both DSLR and Stereo Microscope** images – **30 pts**
- **Model's Generalizability/Robustness - 10 pts**
  (To clarify, it will be examined how consistent the model's performance is over different dataset distribution and/or adversarial examples)
- **Inference time/Computational Cost – 20 pts**
- **Model's scalability – 10 pts**
- **Model's explainability** - 10 pts (Bonus)**

*** Mandatory**
****Optional**
*****Tentative**

Some external resources for more clarification:

- **More about counterfeit IC detection:**
    - https://link.springer.com/chapter/10.1007/978-3-030-62609-9_2
- **Model's Performance Metrics:** https://neptune.ai/blog/evaluation-metrics-binary-classification
- **Model's Generalizability/Robustness :**
    - https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7780854&tag=1
    - https://proceedings.neurips.cc/paper/2020/hash/61d77652c97ef636343742fc3dcf3ba9-Abstract.html
- **Model's scalability:**
    - https://www.codementor.io/blog/scaling-ml-6ruo1wykxf
    - https://www.codementor.io/blog/scalable-ml-models-6rvtbf8dsd
    - https://neptune.ai/blog/how-to-scale-ml-projects
- **Model's explainability:**
    - https://neptune.ai/blog/explainability-auditability-ml-definitions-techniques-tools