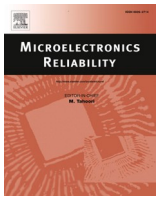
Microelectronics Reliability 150 (2023) 115220

Contents lists available at ScienceDirect

Microelectronics Reliability

journal homepage: www.elsevier.com/locate/microrel



# Synthetic data augmentation to enhance manual and automated defect detection in microelectronics

Adrian Phoulady [a], Yara Suleiman [a], Hongbin Choi [a], Toni Moore [a], Nicholas May [a,b], Sina Shahbazmohamadi [a,*], Pouya Tavousi [a,*]

[a] University of Connecticut, Storrs, CT, USA
[b] Manhattan College, Bronx, NY, USA

ARTICLE INFO

*Keywords:*
Machine learning
Data augmentation
Synthetic dataset
Defect detection

ABSTRACT

Failure analysis and defect detection are crucial processes in industries, governments, and societies to mitigate the risks associated with defective microelectronics. The accurate identification of faulty parts is vital for preventing potential damages. However, traditional manual and automated defect detection approaches face challenges due to the scarcity of ground truth data from defective parts. This limitation hampers the effectiveness of subject matter experts and machine learning models in recognizing and classifying new instances of defects. To address this issue, we propose a synthetic data augmentation workflow that generates virtual defective parts, effectively overcoming the data scarcity problem and enabling the creation of large datasets at a low cost. Our approach enhances defect detection capabilities, empowering industries and governments to improve the quality and reliability of electronic devices.

## 1. Introduction

The semiconductor industry heavily relies on failure analysis and defect detection to ensure high-quality electronic devices [1,2]. Traditional approaches involve capturing 2D and 3D images of components, which are subsequently examined either by experts or through automated processes [3,20–29]. X-ray imaging, particularly X-ray tomography, offers a valuable non-destructive and high-resolution method for failure analysis in the industry.

However, traditional manual inspection methods present challenges due to subjectivity, time consumption, and susceptibility to human error. To address these limitations, there is a growing interest in utilizing machine learning and computer vision techniques to automate and enhance defect detection. Despite advancements in imaging and failure analysis [4], X-ray tomography remains resource-intensive and costly, and the scarcity of defective parts hinders the acquisition of diverse datasets covering potential defects, complicated further by logistical challenges.

Training subject matter specialists to accurately identify and detect defects in microelectronics faces challenges due to limited exposure to real cases and comprehensive datasets. Augmenting training with synthetic datasets offers a potential solution by employing generative models and simulation techniques to create supplementary training examples, enhancing specialists' understanding and detection capabilities.

Automated defect detection is crucial to reduce reliance on human cognition in manual inspection, as it faces challenges in availability of trained personnel, security clearance, and subjective analysis leading to false positives and missed defects. However, the effectiveness of machine learning algorithms in defect assessment heavily relies on large training datasets. Data scarcity is a significant challenge in acquiring annotated datasets, primarily due to the rarity of suitable defective samples and labor-intensive annotation process with large datasets.

To address these challenges, researchers and industry practitioners have explored alternative approaches to dataset creation, such as data augmentation and synthetic data generation techniques. Synthetic datasets can be generated using computer simulations and modeling, overcoming the scarcity of annotated samples. By synthesizing various defect types under different conditions, datasets encompassing a wide range of defect scenarios can be created. Synthetic datasets, when combined with real-world annotated data, offer a more comprehensive and diverse training set for automated defect detection systems.

Previous studies have investigated the application of deep learning for fault detection in electronic and microelectronic parts [5–8]. Some

* Corresponding authors.
*E-mail addresses:* sina.shahbazmohamadi@uconn.edu (S. Shahbazmohamadi), pouya.tavousi@uconn.edu (P. Tavousi).

https://doi.org/10.1016/j.microrel.2023.115220
Received 3 June 2023; Received in revised form 5 August 2023; Accepted 25 August 2023
Available online 1 October 2023

studies relied on existing datasets, some attempted traditional augmentation methods, while others automated the annotation process for specific applications. In response to the data scarcity challenge, researchers have made efforts to create synthetic datasets [9–12]. To generate synthetic datasets, generative adversarial networks [13] have been widely employed, requiring careful training and validation to ensure fidelity to real defects.

In this study, we propose a data augmentation workflow to overcome the shortage of real data in defect identification. Synthetic data is generated by introducing defects into flawless 3D CAD models of microelectronic components, and images are simulated from these defective instances. Each image is labeled with defect information, providing an inherently labeled training dataset crucial for machine learning applications. This method empowers both the workforce and machine learning algorithms for defect identification, finding applications in various industries and governments.

## 2. Materials and methods

### 2.1. Overview of the method

We propose an in-silico method to generate training data for microelectronic defect detection. The process involves: (1) constructing a 3D model of the microelectronic part through imaging-enabled reverse engineering; (2) generating a large number of instances of the part's 3D model; (3) introducing known defects with specific characteristics into the 3D models to obtain a population of defective parts in silico; and (4) obtaining X-ray images of the defective parts using state-of-the-art X-ray CT simulation software. The resulting X-ray images and information about the virtual part's defects are compiled into a dataset for training supervised learning algorithms and workforce training for manual investigations. Fig. 1 illustrates the proposed workflow.

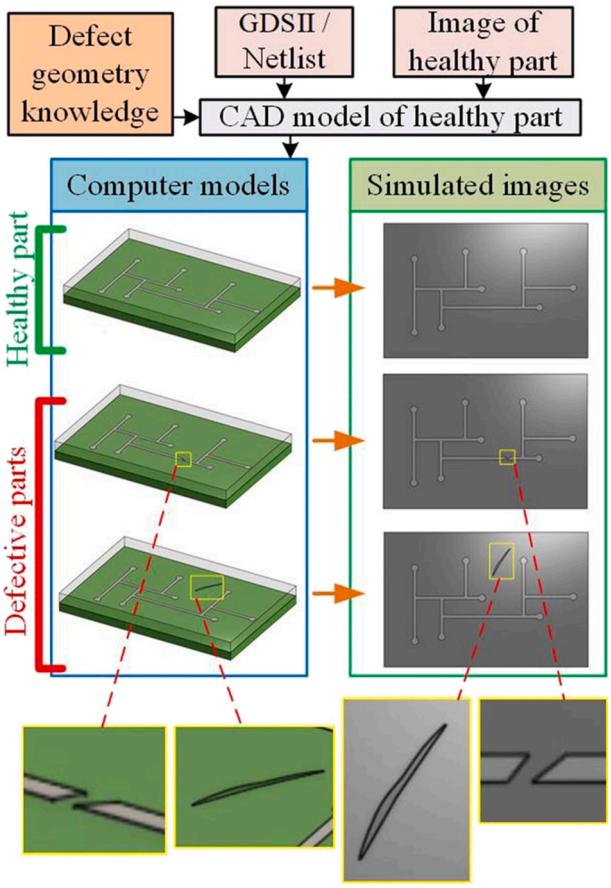



**Fig. 1.** Workflow for synthesizing training data for workforce training and training of machine learning algorithms.

### 2.2. Image acquisition

To generate synthetic training data in two or three dimensions, the reconstructed 3D images of microelectronic chips are required to extract the CAD model. Thus, in our method, we use the last step of acquiring 3D images as the source for generating the synthetic dataset. The CAD model is manually segmented into two categories: electrical parts and packaging.

### 2.3. Reverse engineering and CAD extraction

The generated synthetic dataset can be both in two dimensions and three dimensions, based on the need. However, as the CAD model of the chip is needed for generating the synthetic images, whether 2D or 3D, in our method, the reconstructed 3D images of the chip is acquired. This 3D image is segmented manually using a 3D image analysis tool. For our purpose in this paper, the CAD model is mainly segmented into two categories: the electrical parts and the packaging.

### 2.4. Image simulation

To simulate X-ray imaging, we used gVirtualXRay [14] and the CAD model extracted from real images. We chose appropriate materials for each type of CAD to have the correct attenuation coefficient for X-ray imaging. Then, the image was simulated from the desired spatial angle of the CAD.

### 2.5. Noise addition

Adding the right noise to the synthesized image is critical for effective use of the dataset in training a deep learning model. X-ray and CT imaging systems use X radiation to capture images, which are usually corrupted by noise following a Poisson distribution [15]. Although modern noise reduction algorithms are applied, there are still noises present in the final image. While Monte Carlo simulation of X-ray [16–19] can account for some of these noise effects, it is time-consuming and impractical for generating a large dataset. Therefore, using a ray tracing approach and noise addition is necessary.

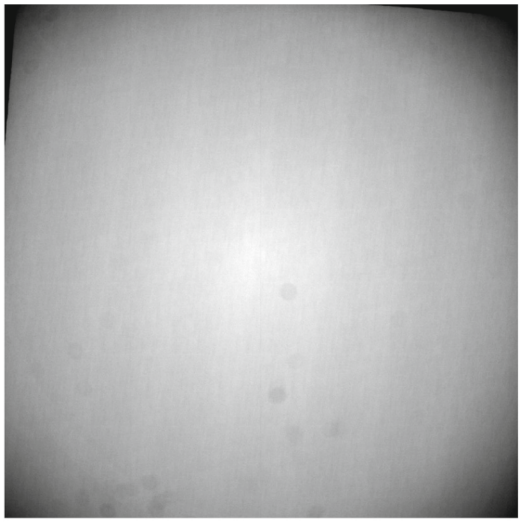
We propose a novel noise addition protocol that involves taking two consecutive images with the X-ray system, from just the air or an arbitrary sample. As these images are of the same thing, they should be identical except for the noise introduced. Subtracting these images and dividing the result by two produces a good approximation of the noise profile of the imaging system, which is then multiplied by a random real number and added to the synthesized images.

In our experiments, we observed that the introduction of artificially added Poisson, white, and salt and pepper noises to the images did not effectively train the deep learning network to perform well on real data. Despite this, we found that this method of simulating noise successfully served our intended purpose.

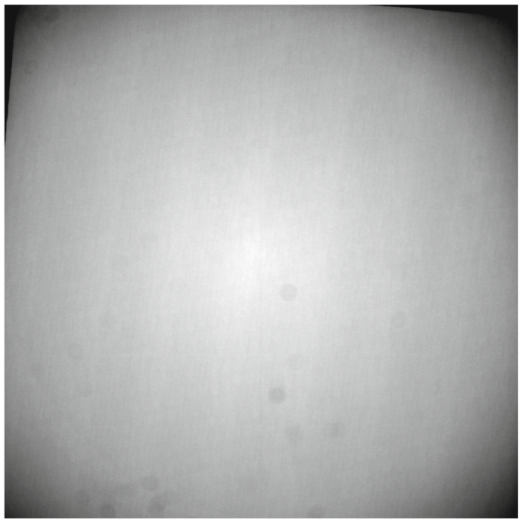
Fig. 2 displays two X-ray images of air, along with the accompanying noise profile obtained from their subtraction. A noticeable effect known as vignetting is apparent in the air images, where the intensity gradually decreases as we move away from the center. In contrast, the noise profile, generated by computing the difference between the two images, demonstrates a consistent texture throughout the entire image. To improve clarity, the histogram of the noise image has been adjusted to enhance visibility and facilitate the observation of the noise profile.

## 3. Results and discussion

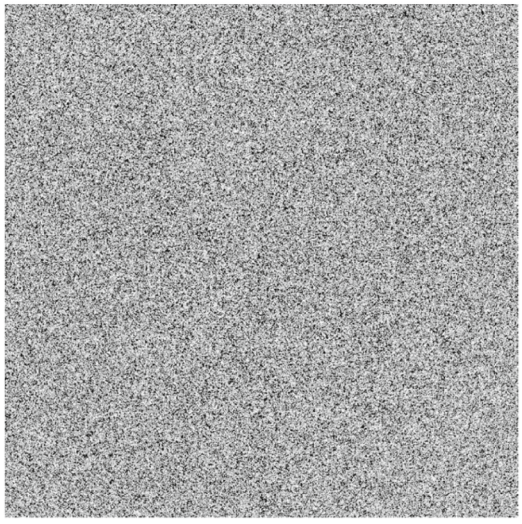
To demonstrate the effectiveness of our proposed workflow, we applied it to an integrated circuit (IC) sample. First, an X-ray CT scan



(a)



(b)



(c)

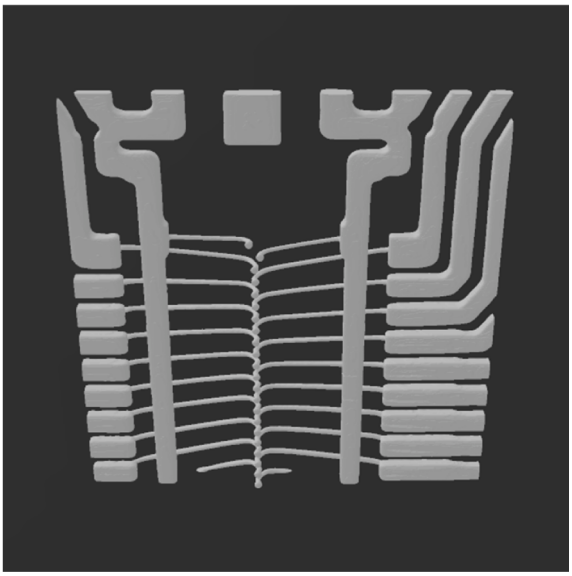
**Fig. 2.** (a, b) Two different X-ray images of air and (c) the normalized resulted noise profile.

was performed on the IC, and the resulting 3D image was used to generate two STL files: one for the electronic parts and one for the packaging (Fig. 3).

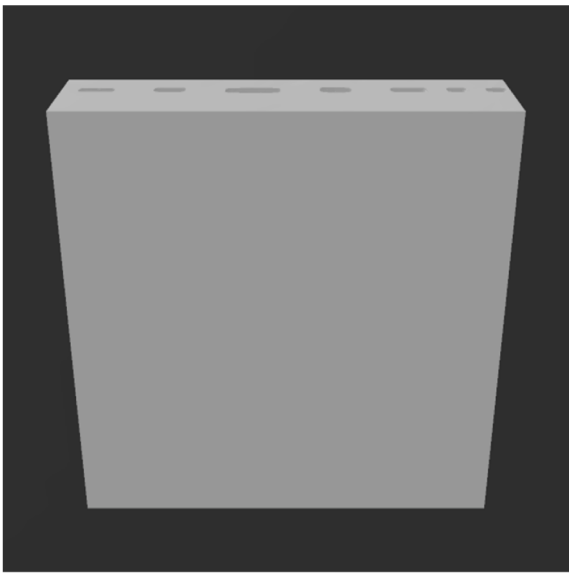
These STL files, along with the assigned materials, were then used in the image simulation algorithm to generate 2D X-ray projections of the sample. Fig. 4 shows the process of generating the synthetic X-ray image using gVirtualXRay. The synthesized image is shown in Fig. 5 after adding noise along with a real 2D X-ray projection of the device. The two images are highly similar, indicating that the simulated images are realistic and suitable for the intended purposes described in this paper.

To quantify the similarity between the synthesized and real X-ray images, the mean absolute error (MAE), mean squared error (MSE), and normalized mean absolute error (NMAE) were employed as performance metrics. If there are total $n$ pixels in the images, $x_i$ is the pixel value of the real image, and $y_i$ is the pixel value of the synthetic image,

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n},$$



(a)



(b)

**Fig. 3.** CADs of (a) the electronic parts and (b) packaging.

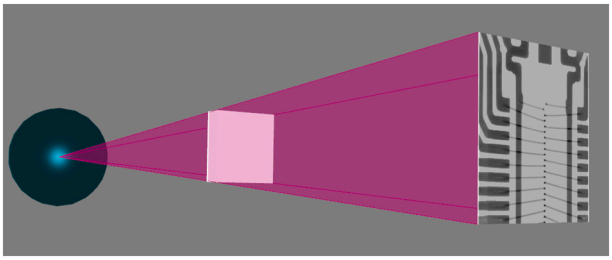
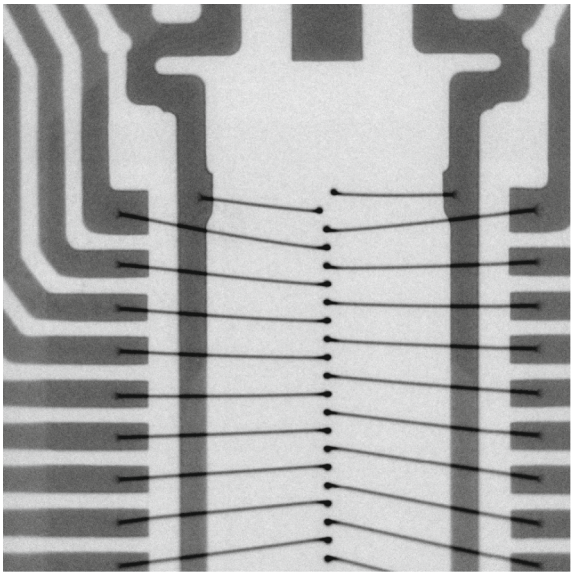
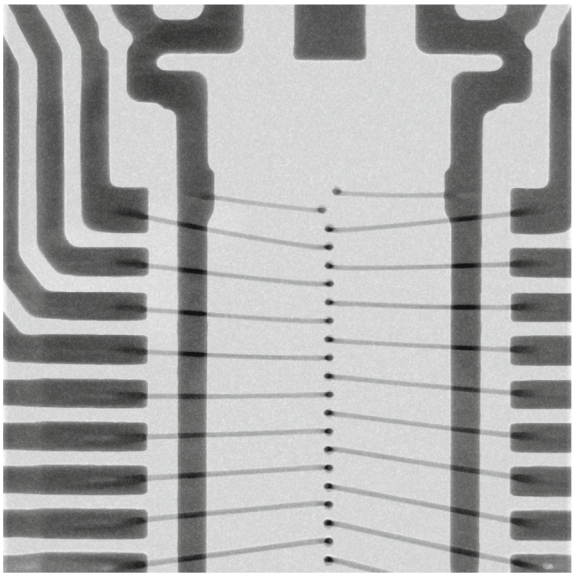


**Fig. 4.** The X-ray simulation from the CADs using gVirtualXRay.



(a)



(b)

**Fig. 5.** (a) A real X-ray image and (b) synthetic image.

$$MSE = \frac{\sum_{i=1}^{n}(y_i - x_i)^2}{n},$$

$$NMAE = \frac{\sum_{i=1}^{n}|y_i - x_i|}{\sum_{i=1}^{n}x_i}.$$

The MAE measures the average absolute difference between corresponding pixels in the two images, MSE measures the average of the squares of the difference between corresponding pixels, and NMAE normalizes MAE by the mean intensity of the real image. The normalized MAE accounts for intensity variations and allows for a more meaningful dissimilarity quantification.

In our analysis, an MAE value of 0.046, MSE value of 0.004, and NMAE value of 0.074 were obtained, indicating a relatively low error between the synthesized and real images. Although it should be noted that all these metrics are sensitive to translation, meaning that even slight shifts in the position of the synthesized image can affect the calculated values. Despite not aligning and registering the two images, the obtained metric values were remarkably low.

Another IC was subjected to X-ray CT imaging, and its CAD model was intentionally modified to produce a missing bond wire and dented packaging. Fig. 6 displays the CADs and a synthetic 2D X-ray projection having these defects. The defects can range from flaws in the microelectronic parts, such as broken or overlapping bond wires, to flaws in the packaging, such as voids or dents.

In the presented case study, the defects were manually applied. However, it is important to highlight that the proposed approach has the potential to be extended and automated. By annotating CAD parts and devising rules, it is possible to streamline the process of applying defects in an automated fashion. This would eliminate the need for manual manipulations and enable a more efficient and scalable approach to generate synthetic images of defects.

## 4. Limitations

While our approach demonstrates the capability to create synthetic X-ray images with rare defects, it is important to acknowledge certain limitations that should be considered. These limitations pertain to the requirement of having or extracting CAD models of the part and the need for expertise in understanding the types of defects.

One notable limitation is the reliance on the availability of CAD models for the target part. The CAD models serve as the foundation for generating the synthetic X-ray images and simulating the defects. Therefore, if CAD models are not readily accessible or if they are incomplete or inaccurate, it may pose challenges in applying our approach effectively.

An additional limitation of the method is its dependence on the specific noise profile of the X-ray system used. Since the noise profile is inherently tied to the characteristics of a particular system, including the source, optics, and detector components, the generated dataset may be less applicable for use with other X-ray systems. As a result, the effectiveness of our approach with different X-ray systems remains unexplored and requires further investigation.

## 5. Conclusion

The detection of defective microelectronics is crucial for various industries, governments, and societies to prevent potential damages. This paper presented a novel workflow to address the challenge of limited ground truth data for detecting defects through imaging. The proposed data augmentation workflow can generate large datasets of virtual defective parts inexpensively. The effectiveness of the method was demonstrated through multiple examples. This workflow can be used to enhance the accuracy of defect detection models and ultimately reduce the risk of faulty microelectronics.

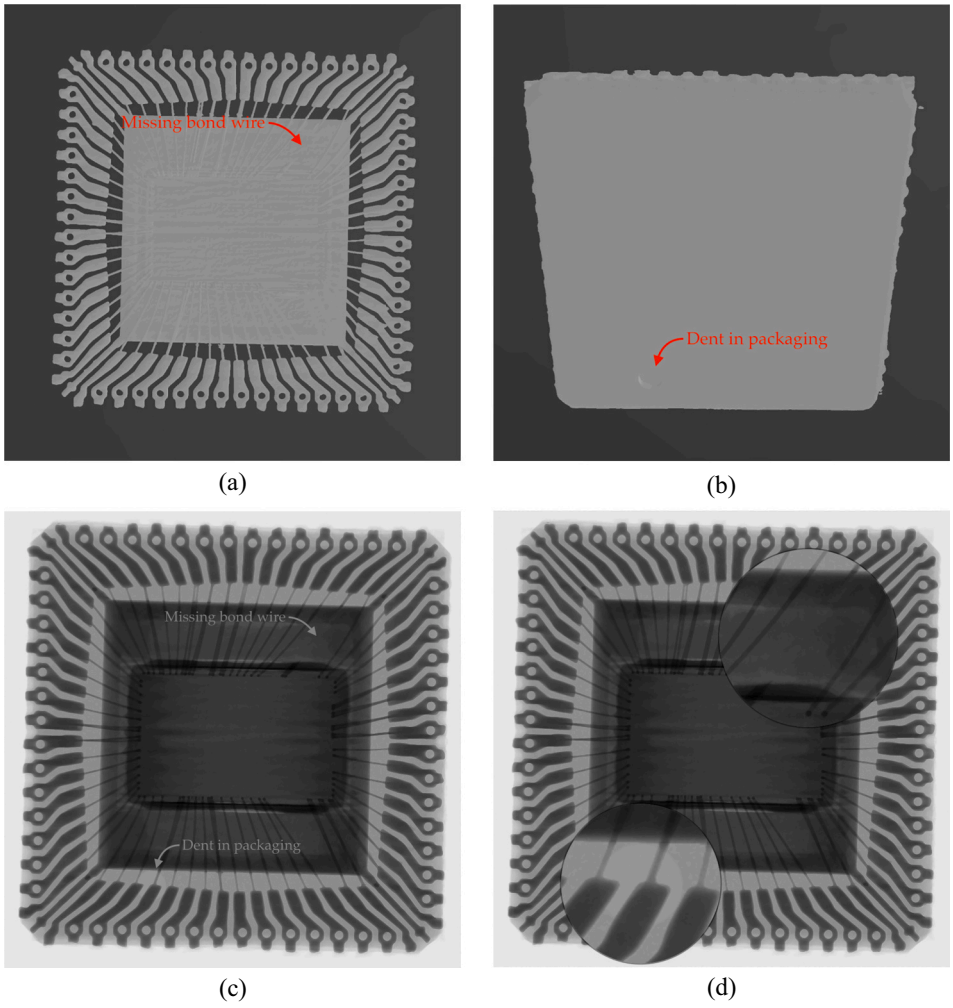



**Fig. 6.** CADs of (a) missing bond wire and (b) dented packaging, (c) the synthetic image with these defects, and (d) the zoomed-in version of the defects.

## CRediT authorship contribution statement

All authors contributed to conceptualization, data curation, formal analysis, investigation, methodology, visualization, and writing. Additionally, corresponding authors contributed to project administration and supervision.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## References

[1] U. Guin, K. Huang, D. DiMase, J.M. Carulli, M. Tehranipoor, Y. Makris, Counterfeit integrated circuits: a rising threat in the global semiconductor supply chain, Proc. IEEE 102 (8) (2014) 1207–1228.
[2] B. Ahmadi, P. Tavousi, J. Favata, P. Shahbeigi-Roodposhti, R. Pelapur, S. Shahbazmohamadi, A novel crowdsourcing platform for microelectronics counterfeit defect detection, Microelectron. Reliab. 88 (2018) 48–53.
[3] S. Shahbazmohamadi, D. Forte, M. Tehranipoor, Advanced physical inspection methods for counterfeit IC detection, in: 40th International Symposium for Testing and Failure Analysis, 2014, November, pp. 55–64.
[4] A. Gu, A. Andreyev, M. Terada, B. Zee, S. Mohammad-Zulkifli, Y. Yang, Accelerate your 3D X-ray failure analysis by deep learning high resolution reconstruction, in: ISTFA 2021, ASM International, 2021, October, pp. 291–295.
[5] Z. Chen, S. Deng, X. Chen, C. Li, R.V. Sanchez, H. Qin, Deep neural networks-based rolling bearing fault diagnosis, Microelectron. Reliab. 75 (2017) 327–333.
[6] N. Dimitriou, L. Leontaris, T. Vafeiadis, D. Ioannidis, T. Wotherspoon, G. Tinker, D. Tzovaras, Fault diagnosis in microelectronics attachment via deep learning analysis of 3-D laser scans, IEEE Trans. Ind. Electron. 67 (7) (2019) 5748–5757.
[7] N. Dimitriou, L. Leontaris, T. Vafeiadis, D. Ioannidis, T. Wotherspoon, G. Tinker, D. Tzovaras, A deep learning framework for simulation and defect prediction applied in microelectronics, Simul. Model. Pract. Theory 100 (2020), 102063.
[8] A. Evangelidis, N. Dimitriou, L. Leontaris, D. Ioannidis, G. Tinker, D. Tzovaras, A deep regression framework toward laboratory accuracy in the shop floor of microelectronics, IEEE Trans. Ind. Inform. 19 (3) (2022) 2652–2661.
[9] M. Frid-Adar, E. Klang, M. Amitai, J. Goldberger, H. Greenspan, Synthetic data augmentation using GAN for improved liver lesion classification, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE, 2018, April, pp. 289–293.
[10] H.I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, P.A. Muller, Data augmentation using synthetic data for time series classification with deep residual networks, arXiv preprint, arXiv:1808.02455, 2018.
[11] N. Jaipuria, X. Zhang, R. Bhasin, M. Arafa, P. Chakravarty, S. Shrivastava, V. N. Murali, Deflating dataset bias using synthetic data augmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 772–773.
[12] S. Jain, G. Seth, A. Paruthi, U. Soni, G. Kumar, Synthetic data augmentation for surface defect detection and classification using deep learning, J. Intell. Manuf. (2022) 1–14.
[13] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, Y. Bengio, Generative adversarial networks, Commun. ACM 63 (11) (2020) 139–144.
[14] A. Sujar, A. Meuleman, P.F. Villard, M. Garcia, F.P. Vidal, gVirtualXRay: virtual X-ray imaging library on GPU, in: Computer Graphics and Visual Computing, The Eurographics Association, 2017, September, pp. 61–68.

[15] D. Thanh, P. Surya, A review on CT and X-ray images denoising methods, Informatica 43 (2) (2019).
[16] J. Giersch, A. Weidemann, G. Anton, ROSI—an object-oriented and parallel-computing Monte Carlo simulation for X-ray imaging, Nucl. Instrum. 509 (1–3) (2003) 151–156.
[17] M.R. Ay, M. Shahriari, S. Sarkar, M. Adib, H. Zaidi, Monte Carlo simulation of x-ray spectra in diagnostic radiology and mammography using MCNP4C, Phys. Med. Biol. 49 (21) (2004) 4897.
[18] R. Gauvin, E. Lifshin, H. Demers, P. Horny, H. Campbell, Win X-ray: a new Monte Carlo program that computes X-ray spectra obtained with a scanning electron microscope, Microsc. Microanal. 12 (1) (2006) 49–64.
[19] E. Bergbäck Knudsen, A. Prodi, J. Baltser, M. Thomsen, P. Kjær Willendrup, M. Sanchez del Rio, K. Lefmann, McXtrace: a Monte Carlo software package for simulating X-ray optics, beamlines and experiments, J. Appl. Crystallogr. 46 (3) (2013) 679–696.
[20] A. Phoulady, N. May, H. Choi, S. Shahbazmohamadi, P. Tavousi, A novel material detection method using femtosecond laser, confocal imaging and image processing enabling endpointing in fast inspection of microelectronics, Microelectron. Reliab. 126 (2021), 114287.
[21] M. Konnik, B. Ahmadi, N. May, J. Favata, Z. Shahbazi, S. Shahbazmohamadi, P. Tavousi, Training AI-based feature extraction algorithms, for micro CT images, using synthesized data, J. Nondestruct. Eval. 40 (2021 Mar) 1–3.
[22] A. Phoulady, N. May, H. Choi, Y. Suleiman, S. Shahbazmohamadi, P. Tavousi, Rapid high-resolution volumetric imaging via laser ablation delayering and confocal imaging, Sci. Rep. 12 (1) (2022 Jul 19) 12277.
[23] H. Choi, N. May, A. Phoulady, Y. Suleiman, D. DiMase, S. Shahbazmohamadi, P. Tavousi, Rapid three-dimensional reconstruction of printed circuit board using femtosecond laser delayering and digital microscopy, Microelectron. Reliab. 138 (2022 Nov 1), 114659.
[24] N. May, H. Choi, A. Phoulady, Y. Suleiman, D. DiMase, P. Tavousi, S. Shahbazmohamadi, Correlative multimodal imaging and targeted lasering for automated high-precision IC decapsulation, Microelectron. Reliab. 138 (2022 Nov 1), 114660.
[25] N. May, H. Choi, A. Phoulady, S. Amini, P. Tavousi, S. Shahbazmohamadi, Gas-assisted femtosecond pulsed laser machining: a high-throughput alternative to focused ion beam for creating large, high-resolution cross sections, PLoS One 18 (5) (2023 May 3), e0285158.
[26] P. Hoveida, A. Phoulady, H. Choi, N. May, S. Shahbazmohamadi, P. Tavousi, Terahertz-readable laser engraved marks as a novel solution for product traceability, Sci. Rep. 13 (1) (2023 Aug 1) 12474.
[27] N. May, A. Phoulady, H. Choi, P. Tavousi, S. Shahbazmohamadi, Single image composite tomography utilizing large scale femtosecond laser cross-sectioning and scanning electron microscopy, Microsc. Microanal. 28 (S1) (2022 Aug) 876–878.
[28] N. May, H. Choi, A. Phoulady, P. Tavousi, S. Shahbazmohamadi, Three-dimensional reconstruction of printed circuit boards: comparative study between 3D femtosecond laser serial sectioning and optical imaging versus 3D X-ray computed tomography, Microsc. Microanal. 28 (S1) (2022 Aug) 284–286.
[29] N. May, J. Favata, B. Ahmadi, P. Tavousi, S. Shahbazmohamadi, Correlative microscopy workflow for precise targeted failure analysis of multi-layer ceramic capacitors, Microelectron. Reliab. 114 (2020 Nov 1), 113858.