2024 25th International Conference on Thermal, Mechanical and Multi-Physics Simulation and Experiments in Microelectronics and Microsystems (EuroSimE) | 979-8-3503-9363-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/EuroSimE60745.2024.10491538

Training Convolutional Neural Networks with Confocal Scanning Acoustic Microscopy Imaging for Power QFN Package Delamination Classification

Henry A. Martin[1,2], Haojia Xu[1,3], Edsger C.P. Smits[1], Willem D. van Driel[2], GuoQi Zhang[2]

[1] *Chip Integration Technology Center (CITC), Transistorweg 5T, 6534 AT Nijmegen, The Netherlands.*
[2] *Delft University of Technology, Mekelweg 4, 2628 CD DELFT, The Netherlands.*
[3] *HAN University of Applied Sciences, Ruitenberglaan 29, 6826 CC Arnhem, The Netherlands.*
henry.martin@citc.org ; h.a.martin@tudelft.nl

Abstract

This study introduces a training protocol utilizing Convolutional Neural Networks (CNNs) and Confocal Scanning Acoustic Microscopy (CSAM) imaging techniques to classify Power Quad Flat No-leads (PQFN) package delamination. The investigation involves empty PQFN packages with varied substrate metallizations subjected to thermal cycling. Four delamination classes were labeled: Die-pad delamination (Class-A), Bond-pad delamination (Class-B), both Die-pad and Bond-pad delamination (Class-C), and No delamination (Class-D). Due to data imbalance, additional randomness was introduced for distribution balancing. Residual Networks (ResNet-18) based CNN model was selected for classification. Five-fold cross-validation assessed overfitting performance concerning input data size, image resolution, and batch size. The ResNet-18 prediction performance was evaluated using precision and recall metrics, with the model achieving average precision and recall scores of 0.86/1 and 0.83/1, respectively. Additionally, a comparison of delamination among different substrate metallizations was presented with Ag and NiPdAu indicating significant delamination compared to bare Cu substrate. This study pioneers the integration of CNNs with CSAM imaging for package defect detection and classification, laying the groundwork for future research to address the complex interplay of multiple failure mechanisms in functional packages.

Keywords

Package delamination, Scanning Acoustic Microscopy, Defect detection and Classification, Deep learning, Residual Networks.

1. Introduction

Power Quad Flat No-leads (or non-leaded) – PQFN packages are widely used surface mount non-hermetically sealed package types in electronic devices, and they are known for compact size and efficient thermal dissipation. While the reliability challenges of PQFN packages are well documented [1, 2], concerns arise on effective methods to identify packaging defects [3–6]. Detection and localization of failure modes on electronic packages are significant to improve the overall package performance. An overview of various failure modes in electronic packaging is provided in [7].

Confocal Scanning Acoustic Microscopy (CSAM) is a powerful non-destructive imaging technique for identifying package defects. An overview of SAM imaging for material evaluation and defect detection is provided in [8–13]. Figure 1 illustrates different failure modes on PQFN packages imaged using CSAM. Five different failure modes were identified on the lifetime-tested sample as compared to a pristine 0-hour sample.

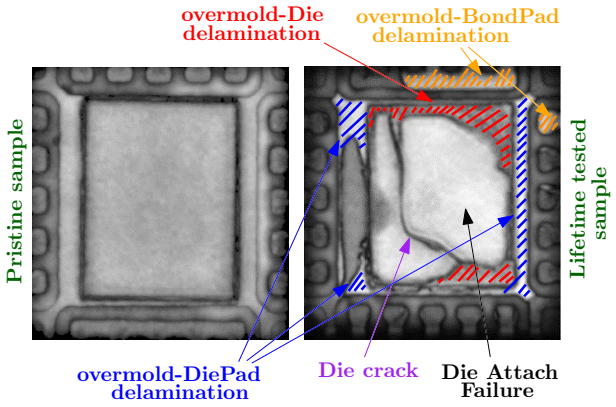


Fig. 1. CSAM image comparison of a pristine sample (left) and a lifetime tested sample (right).

Despite the benefits of acoustic imaging in electronic packaging, the manual interpretation of failure modes from acoustic images is subjective, particularly in cases where there is an interplay of multiple failure mechanisms (see figure 1). Though the effect of mold compound delamination is a widely researched topic [14, 15], this paper is an attempt to introduce a training protocol using Convolutional Neural Networks (CNN) in conjunction with Confocal Scanning Acoustic Microscopy (CSAM) imaging technique for overmold delamination detection and classification.

In recent times, machine learning models on image recognition and analysis have gained increasing attention [16–18]. The integration of CNN analysis with CSAM imaging in electronic packaging applications remains largely unexplored, primarily due to its complexity. [19] has explored YOLO (You Only Look Once) series algorithms to target and detect delamination defects based on acoustic images. In this study, a similar approach was implemented on the ResNet-18 framework with a specific

979-8-3503-9363-7/24/$31.00 ©2024 IEEE

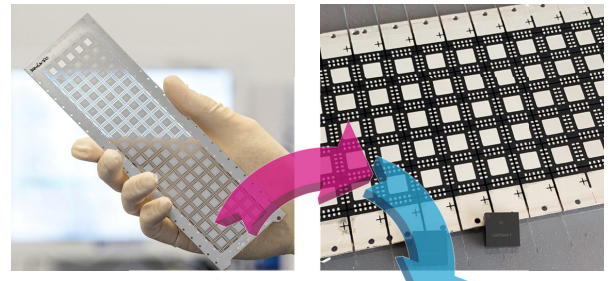
focus on identifying delamination of the epoxy molding compound from the metallized package substrate and classifying it based on the die-pad and the bond-pad areas.

The following section details the sample preparation process, the network architecture, and the workflow. Subsequently, the pre-processing steps involved are documented and the training accuracy vs loss is displayed for varying data size, image resolution, and batch size. The experimental results conclude by highlighting the prediction performance and an overview of failures for different package substrate metallizations.

2. Experimental Methods

A. Sample preparation

PQFN packages with different substrate metallizations, Ag, NiPdAu, and bare Cu (no metallization), were prepared specifically for this study to understand the effect of mold compound delamination from the package substrate. The purpose of metallizing the substrate is to promote adhesion. To quantitatively evaluate the delamination progression of the mold compound from various substrate metallizations, empty packages (without any die) were prepared, thereby the region of focus during CSAM is the overmold-substrate interface. A schematic illustration of the sample preparation process is shown in figure 2.



No.	Substrate type	Quantity
1.	Ag-Metallized	50 PQFN
2.	NiPdAu-Metallized	50 PQFN
3.	Bare Cu	40 PQFN

Fig. 2. Package substrate with different metallizations (Ag, NiPdAu, and bare Cu) was over-molded without any die and singulated. A cross-sectional schematic of an empty PQFN is shown and the quantity of sample for each substrate type is tabulated.

B. Data Acquisition

All assembled samples (140 PQFN packages – 50 Ag-Metallized, 50 NiPdAu-Metallized, and 40 Bare Cu) were scanned using a 50MHz acoustic transducer in echo (C-Scan) mode. Since the pristine samples didn't exhibit significant failures, the samples were subjected to thermal cycling from -55^{o}C to 150^{o}C according to the guidelines outlined in the JESD22-A104 standard.

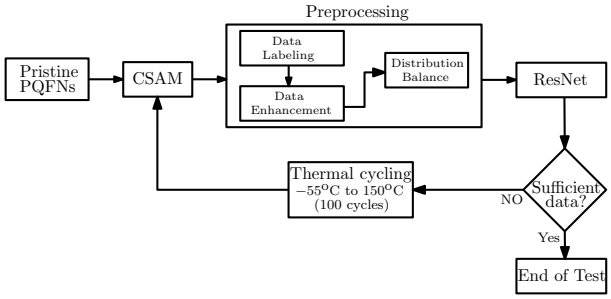


Fig. 3. Data acquisition flowchart. The sufficiency of the data was determined based on the number of images per class.

The samples were scanned intermittently during cycling (i.e., after every one hundred thermal cycles) until a sufficient amount of dataset was collected for training the CNN model. The training and validation accuracy of the model in correlation with the data size is explained later.

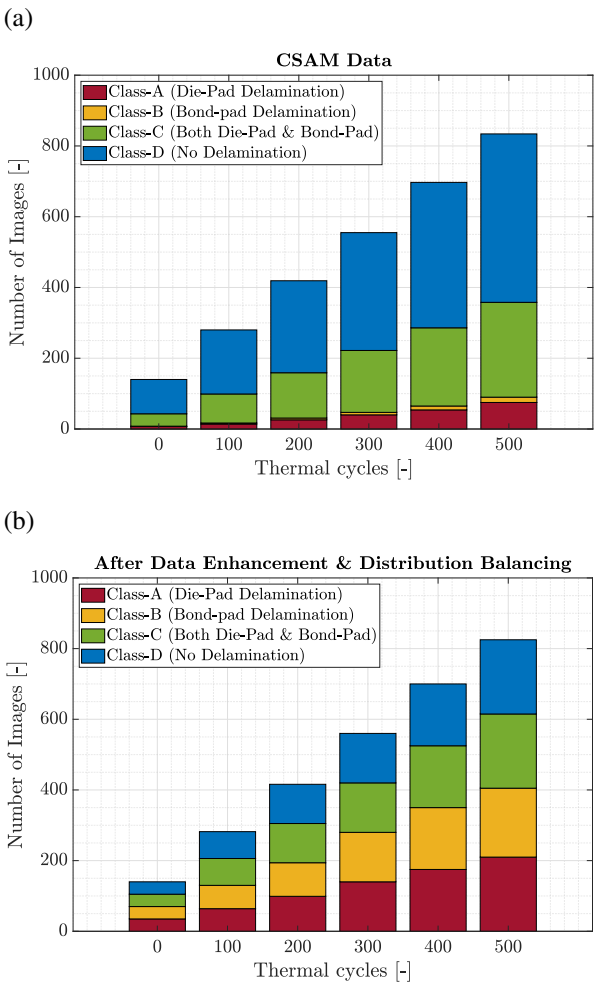


Fig. 4. A comparison of the original CSAM data among all classes (a) and the data after being subjected to preprocessing steps (b).

The data acquisition process flow is charted in figure 3 and the acquired images are labeled into four classes based on the type of failure mode:

- Class A – Delamination of Die-pad area
- Class B – Delamination of Bond-pad area
- Class C – Delamination of both Die-pad and Bond-pad areas
- Class D – No delamination.

Upon labeling, additional randomness (rotation and contrast) was introduced to the original dataset. The data balancing was further performed to ensure all classes have an equal amount of data. A comparison of the labeled data before and after enhancement & distribution balancing is shown in figure 4.

The original CSAM data (see figure 4a) had data imbalance among the four classes, particularly Class-B (Bond-pad delamination) was significantly lower. Upon correction and balancing, an equal amount of datasets were gathered for all four classes as shown in figure 4b.

C. Choosing Network Architecture

Convolutional Neural Networks (CNNs) leverage learnable filters to recognize patterns, such as edges, textures, etc., from an input image. Each filter is convolved across the image dimensions, computing a dot product between the filter and the input, thereby producing a 2-dimensional map of that filter to recognize patterns. However, CNN performance degrades as the network depth increases, leading to vanishing gradients through each layer and overfitting (training accuracy is significantly higher than the validation accuracy).

ResNet (Residual Network) introduced residual learning to address the degradation problem in CNN. A schematic overview of the ResNet architecture is shown in figure 5.

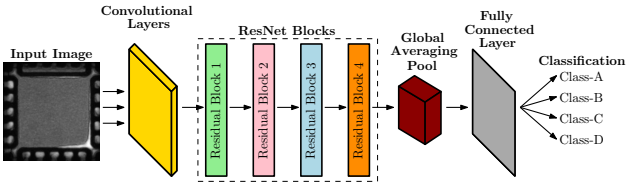


Fig. 5. Schematic overview of Residual Network architecture for image classification

The Residual blocks in ResNet include two main paths. The first path is a sequence of layers (convolution, batch normalization, Rectified Linear Unit, etc.,) that learns residual functions about each layer input. The second path is the skip connection that bypasses the layers where the input and the output of that particular layer are the same. The skip function allows the ResNet to learn the necessary identity information and solves the vanishing gradient problem, thereby allowing deeper networks. The fundamentals of image recognition and classification using deep learning and residual learning are provided in [20, 21]

Understanding the network architecture allows us to make an informed decision to choose the right network. ResNet architecture contains different variants, such as ResNet-18, ResNet-50, and ResNet-101, each indicating the network depth. The desired depth of the network is a choice depending on the computational resources, and performance requirements. The problem of identifying overmold-substrate delamination from an empty PQFN is a relatively simple task. Hence, ResNet-18 – a shallow variant with 18 layers was chosen, which is computationally faster and less prone to overfitting. Accordingly, the training and validation accuracy of ResNet-18 architecture was analyzed in the subsequent section with varying data size (total number of input data), the effect of input image size (number of pixels), and the effect of batch size (number of images per iteration).

3. Experimental results

A. ResNet-18 cross-validation

The overfitting of the ResNet-18 architecture concerning data size, image pixel size, and batch size was evaluated by computing the delta accuracy and delta loss between the training and validation sets. A five-fold cross-validation (80% data for training and 20% data for validation) was performed by training the ResNet-18 model on each subset, and a cross-entropy loss function was used to evaluate the model's performance. The results on the effect of data size, image size, and batch size are shown in figure 6.

(a) Effect of data size on overfitting:

- To analyze the effect of data size on overfitting, the incremental number of preprocessed images at every thermal cycling step was taken into account (see figure 4b).
- The delta accuracy and delta loss (Training – Validation) were determined and the results are shown in figure 6a. The ResNet-18 tends to have the best overfitting performance with a small dataset of 420 images.
- However, this limits the data collection to 200 thermal cycles (420 images), which contains insufficient information on overmold-substrate delamination. Hence, a data size of 840 images (500-thermal cycles) is preferred. A comparison of the model prediction performance for both 420 and 840 training data sizes is shown in the next subsection.

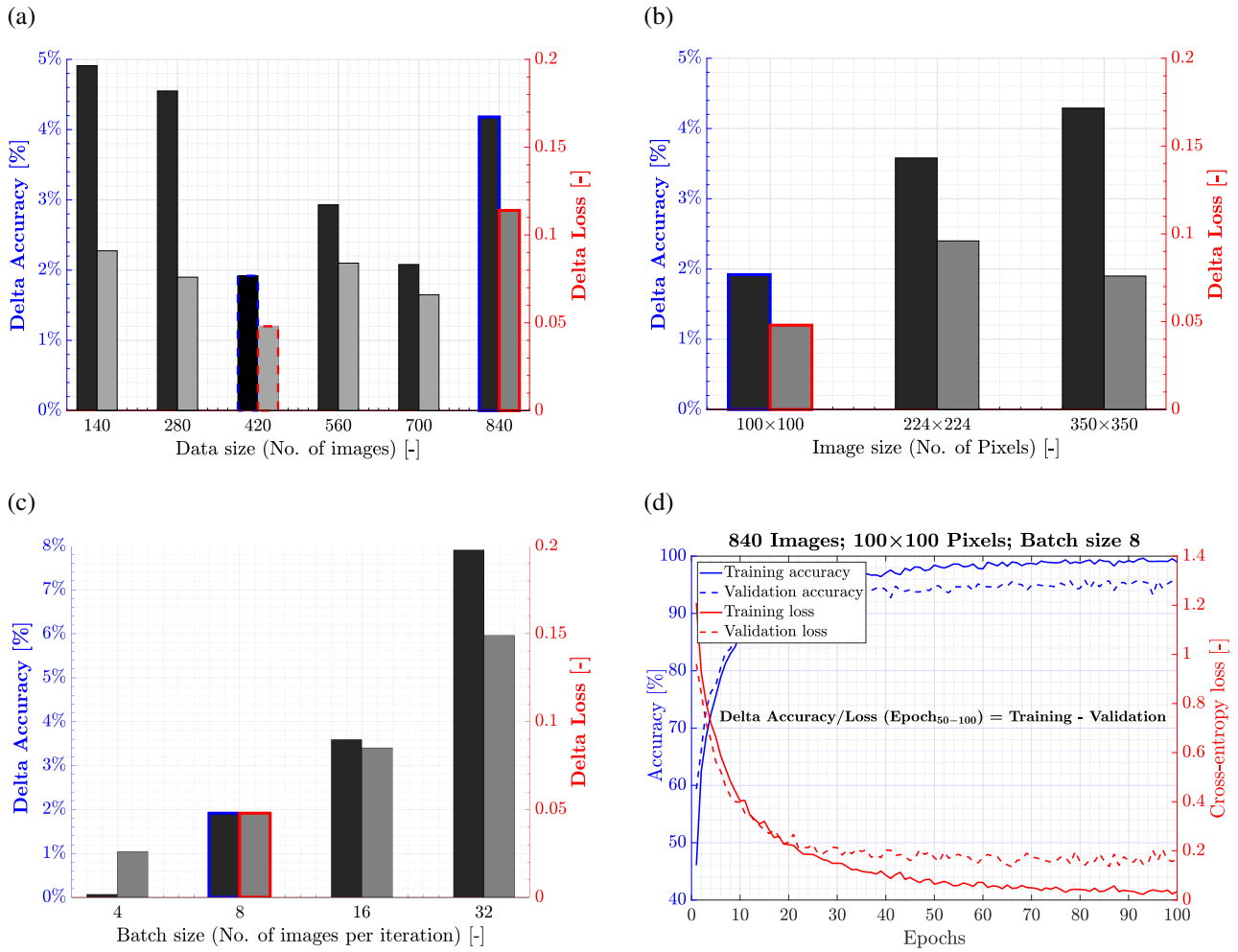


Fig. 6. Five-fold cross-validation of the ResNet-18 architecture by varying the data size (a), image size (b), and batch size (c) is shown, and the chosen parameters are highlighted. (d) indicates the overfitting performance (delta accuracy and delta loss) of the ResNet-18 model for the chosen parameters.

- It is important to be aware that the delta accuracy and delta loss shown in figure 6 includes all four classifications.

(b) Effect of image resolution on overfitting:

- The effect of image pixel size on overfitting was analyzed by varying the number of pixels, and the results are shown in figure 6b.
- The resolution of a CSAM image is 20μm per pixel, which translates around 400×400 pixels.
- Having an exact number of pixels as the original image would be computationally slower and further leads to poor overfitting performance. Hence, we limited our analysis to a maximum of 350×350 pixels.
- The best overfitting performance was achieved for 100×100 pixels, which provides an overall good balance between computational speed and capturing the necessary information.

(c) Effect of batch size on overfitting:

- The batch size, or the number of images per iteration during training affects the training dynamics and the model performance.
- Though larger batch size improves the computational efficiency for training, it leads to overfitting and requires more memory while smaller batch size may result in inherent noise.
- The effect of batch size was evaluated and the results are graphically shown in figure 6c. An optimal batch size of 8 indicates good overfitting performance, faster convergence during training, and requires less memory.

Figure 6d indicates the overfitting performance (training accuracy & loss in comparison to the validation accuracy & loss) for the ResNet-18 model with 840 input images, 100×100 pixels, and a batch size of 8. The trained ResNet-18 model was further evaluated with a testing dataset and the results are shown next.

B. ResNet-18 test results

The prediction performance of the trained ResNet-18 model was assessed using a unique testing dataset of 88 CSAM images from all four classifications without preprocessing, i.e., the testing dataset has a distribution imbalance with a sample size of 22 images for Class-A; 23 images for Class-B; 18-images for Class-C; and 23 images for Class-D. The most common metrics to evaluate the prediction performance of a trained model are as follows:

- **Accuracy** - It provides an understanding of how often the model is true but fails to evaluate the true performance of the prediction compared to the actual.

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

where, *TP* - True Positive; *TN* - True Negative; *FP* - False Positive; *FN* - False Negative.

(a) Dataset – 420 images

(b) Dataset – 840 images

Fig. 7. Confusion matrix for all four classes (Actual vs Predicted) is plotted for training dataset of 420 images (a) and 840 images (b). The average precision and average recall values for 420 datasets are 0.68/1 and 0.59/1, which is significantly lower. The average precision and average recall for 840 datasets are 0.86/1 and 0.83/1, which indicates a relatively better prediction performance.

- **Precision** - It evaluates the accuracy of the positive predictions made by the model, i.e., how many predicted positives are true positve? It is the ratio of true positive predictions to the total positive predictions. A precision value of 1 would indicate 100% accuracy for positive predictions.

$$\text{Precision} = \frac{TP}{TP + FP}$$

- **Recall** - It evaluates the accuracy of the predictions (both positive and negative) made by the model, i.e., how many predicted positives and negatives are true? It is the ratio of true positive predictions to the total actual positives. A recall value of 1 would indicate 100% accuracy for both positive and negative predictions.

$$\text{Recall} = \frac{TP}{TP + FN}$$

In this study, precision and recall are used as metrics to evaluate the ResNet-18 model performance. The results of the model performance are plotted as a confusion matrix (see figure 7) that summarizes the performance of the classification model (predicted vs actual). The confusion matrix is plotted for both data sizes; 420 images (see figure 7a) and 840 images (see figure 7b).

The following conclusions can be drawn from the confusion matrix shown in figure 7:

- From figure 7a, a training dataset of 420 images leads to poor prediction, particularly for Class-A – Delamination of Die-pad and Class-B – Delamination of Bond-pad area. Class-B has the least original CSAM images (see figure 4a), which has led to insufficient learning due to data enhancement and biased overfitting performance from other classes.
- Comparatively, a dataset of 840 images resulted in better prediction performance (see figure 7b). While the majority of the classes indicate relatively good predictions, the model still struggles to detect Class-B failure mode.
- For a dataset of 420 images, the average precision score is 0.68/1, and the average recall score is 0.59/1.
- Whereas, the dataset of 840 images indicates an average precision score of 0.86/1 and an average recall score of 0.83/1, which is far superior to the trained model with 420 images.
- The provided confusion matrix and the metric scores highlight the goodness of the ResNet-18 model for PQFN package delamination classification.

Furthermore, it was also realized that the epoxy mold compound delamination was predominantly from the NiPdAu metallized die-pad substrates. Some of the Ag metallized substrates also indicated delamination of overmold. The bare Cu substrate indicated no sign of die-pad delamination, but some packages indicated minimal delaminations in the bond-pad region. A comparative overview of all three substrate types, imaged at 0-hour and after 500-thermal cycles, is shown in figure 8. It is important to be aware that the delamination of overmold–substrate also depends on the type of epoxy molding compound and the filler size.

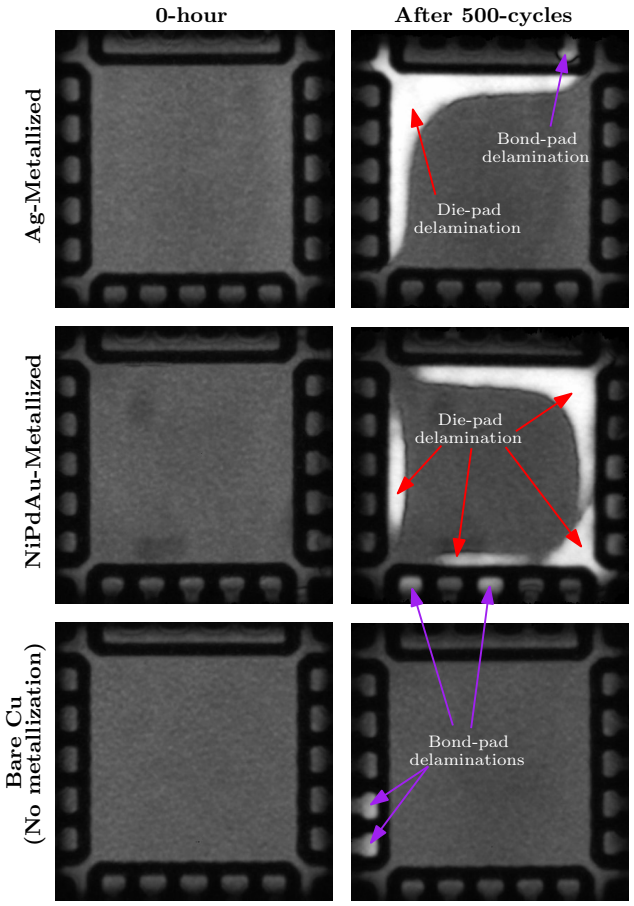


Fig. 8. A comparative overview of overmold delamination from different substrate metallizations. Ag and NiPdAu metallized substrates indicate significant delamination from the die-pad area and some marginal delaminations from the bond-pad area after 500 thermal cycles. Bare Cu substrate (without any metallization) indicates minimal delaminations only from the bond pads.

4. Summary and Conclusion

The primary objective of this study focussed on demonstrating a training protocol using Convolutional Neural Networks in conjunction with Confocal Scanning Acoustic Microscopy imaging for classifying Power QFN package delamination classification. Empty packages (without die) were prepared with different substrate metallizations; Ag-Metallized – 50PQFNs, NiPdAu Metallized – 50 PQFNs, and Bare Cu – 40PQFNs. All empty packages were scanned intermittently during the thermal cycling process. The accelerated cycling led to overmold-substrate delaminations over time. Four classes of delaminations were labeled; Class A - Die-pad delamination, Class B - Bond-pad delamination, Class C - Delamination of both die-pad and bond-pad, and Class D - No delaminations.

Due to the imbalance of delamination data among all four classes, data preprocessing was required to introduce additional randomness (image rotation and contrast) for distribution balancing. The choice of network architecture is critical. Residual Networks (ResNet) based CNN models with 18 layers of network depth were chosen in this study for delamination classification. A five-fold cross-validation was conducted to mitigate the overfitting performance introduced by various data sizes, image resolutions, and batch sizes. The prediction performance of the trained ResNet-18 model was further evaluated based on precision and recall metrics. The results of the model were displayed by plotting a confusion matrix between actual and predicted information. The model demonstrated an average precision of 0.86/1 and an average recall of 0.83/1 for all four classes. An overview of the overmold–substrate delamination for all three different substrate metallizations was further provided as a comparison between pristine and aged.

This study has been the first effort to integrate Convolutional Neural Networks with CSAM imaging in electronic packaging for defect detection and classification. Though the current paper focussed on classifying a single interface failure (overmold–substrate), future research will be focused on introducing a protocol for classifying complex multiple failure mechanisms on functional packages with multiple interfaces.

Acknowledgements

The authors extend their sincere gratitude and recognition to all CITC colleagues involved in this test work, providing their valuable support and input. We also would like to acknowledge the support of SENCIO BV in Nijmegen (NL), providing their transfer molding knowledge and capabilities to encapsulate the empty PQFN packages.

References

[1] R. Kong, C. Tulkoff, C. Hillman, and Rkong, "The reliability challenges of qfn packaging," 2010. [Online]. Available: https://api.semanticscholar.org/CorpusID:26888744.

[2] M. Bazu and T. T.-M. Bajenescu, "Failure analysis: A practical guide for manufacturers of electronic components and systems," *Failure Analysis: A Practical Guide for Manufacturers of Electronic Components and Systems*, Apr. 2011. DOI: 10.1002/9781119990093.

[3] P. Aryan, S. Sampath, and H. Sohn, "An overview of non-destructive testing methods for integrated circuit packaging inspection," *Sensors*, vol. 18, no. 7, 2018, ISSN: 1424-8220. DOI: 10.3390/s18071981. [Online]. Available: https://www.mdpi.com/1424-8220/18/7/1981.

[4] Z. Niu, "Design of defect detection system for semiconductor plastic packaging based on machine vision," *Journal of Physics: Conference Series*, vol. 2006, no. 1, p. 012 008, 2021. DOI: 10.1088/1742-6596/2006/1/012008. [Online]. Available: https://dx.doi.org/10.1088/1742-6596/2006/1/012008.

[5] H. A. Martin, E. C. P. Smits, R. H. Poelma, W. D. van Driel, and G. Zhang, "Online condition monitoring methodology for power electronics package reliability assessment," *IEEE Transactions on Power Electronics*, vol. 39, no. 4, pp. 4725–4734, 2024. DOI: 10.1109/TPEL.2024.3352747.

[6] H. A. Martin, R. Sattari, E. C. P. Smits, H. W. van Zeijl, W. D. van Driel, and G. Q. Zhang, "In-situ reliability monitoring of power packages using a thermal test chip," in *2022 23rd International Conference on Thermal, Mechanical and Multi-Physics Simulation and Experiments in Microelectronics and Microsystems (EuroSimE)*, 2022, pp. 1–10. DOI: 10.1109/EuroSimE54907.2022.9758913.

[7] P. Viswanadham and P. Singh, "Failure modes and mechanisms," in *Failure Modes and Mechanisms in Electronic Packages*. Boston, MA: Springer US, 1998, pp. 136–282, ISBN: 978-1-4615-6029-6. DOI: 10.1007/978-1-4615-6029-6_6. [Online]. Available: https://doi.org/10.1007/978-1-4615-6029-6_6.

[8] H. Yu, "Scanning acoustic microscopy for material evaluation," *Applied Microscopy*, vol. 50, Dec. 2020. DOI: 10.1186/s42649-020-00045-4.

[9] F. Bertocci, A. Grandoni, and T. Djuric-Rissner, "Scanning acoustic microscopy (sam): A robust method for defect detection during the manufacturing process of ultrasound probes for medical imaging," *Sensors*, vol. 19, no. 22, 2019, ISSN: 1424-8220. [Online]. Available: https://www.mdpi.com/1424-8220/19/22/4868.

[10] S. Levikari, T. J. Kärkkäinen, C. Andersson, J. Tammminen, and P. Silventoinen, "Acoustic detection of cracks and delamination in multilayer ceramic capacitors," *IEEE Transactions on Industry Applications*, vol. 55, no. 2, pp. 1787–1794, 2019. DOI: 10.1109/TIA.2018.2873989.

[11] G.-M. Zhang, D. M. Harvey, and D. R. Braden, "Microelectronic package characterisation using scanning acoustic microscopy," *NDT E International*, vol. 40, no. 8, pp. 609–617, 2007, ISSN: 0963-8695. DOI: https://doi.org/10.1016/j.ndteint.2007.05.002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S096386950700062X.

[12] F. Liu, L. Su, M. Fan, J. Yin, Z. He, and X. Lu, "Using scanning acoustic microscopy and lm-bp algorithm for defect inspection of micro solder bumps," *Microelectronics Reliability*, vol. 79, pp. 166–174, 2017, ISSN: 0026-2714. DOI: https://doi.org/10.1016/j.microrel.2017.10.029. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S002627141730505X.

[13] P. Aryan, S. Sampath, and H. Sohn, "An overview of non-destructive testing methods for integrated circuit packaging inspection," *Sensors*, vol. 18, no. 7, 2018, ISSN: 1424-8220. [Online]. Available: https://www.mdpi.com/1424-8220/18/7/1981.

[14] Y. Kim, C. Raleigh, and S. Saiyed, "Design study to prevent mold delamination for overmolded lead frame package," in *2019 18th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)*, 2019, pp. 485–489. DOI: 10.1109/ITHERM.2019.8757448.

[15] H. Zhang *et al.*, "Power qfn down bond lift and delamination study," in *2014 IEEE 16th Electronics Packaging Technology Conference (EPTC)*, 2014, pp. 570–573. DOI: 10.1109/EPTC.2014.7028424.

[16] Y.-J. Huang, C.-L. Pan, S.-C. Lin, and M.-H. Guo, "Machine-learning approach in detection and classification for defects in tsv-based 3-d ic," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. PP, pp. 1–8, Jan. 2018. DOI: 10.1109/TCPMT.2017.2788896.

[17] E. Weiss, "Revealing hidden defects in electronic components with an ai-based inspection method: A corrosion case study," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 13, no. 7, pp. 1078–1080, 2023. DOI: 10.1109/TCPMT.2023.3293005.

[18] V. Adibhatla, J.-S. Shieh, M. Abbod, H.-C. Chih, C. Hsu, and J. Cheng, "Detecting defects in pcb using deep learning via convolution neural networks," Oct. 2018, pp. 202–205. DOI: 10.1109/IMPACT.2018.8625828.

[19] Y. Zhao, M. Xiao, H. Lv, J. Luo, X. Wang, and D. Luo, "Research on scanning acoustic image defects detection of integrated circuits based on yolox," in *2022 23rd International Conference on Electronic Packaging Technology (ICEPT)*, 2022, pp. 1–4. DOI: 10.1109/ICEPT56209.2022.9872652.

[20] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," *Neural Information Processing Systems*, vol. 25, Jan. 2012. DOI: 10.1145/3065386.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.