Deep Convolution Neural Networks for Automatic Detection of Defects which Impact Hybrid Bonding Yield

Oliver Zhao
Adeia, Inc.
San Jose, CA, USA
Oliver.Zhao@adeia.com

Dominik Suwito
Adeia, Inc.
San Jose, CA, USA
Dominik.Suwito@adeia.com

Bongsub Lee
Adeia, Inc.
San Jose, CA, USA
Bongsub.Lee@adeia.com

Thomas Workman
Adeia, Inc.
San Jose, CA, USA
Thomas.Workman@adeia.com

Laura Mirkarimi
Adeia, Inc.
San Jose, CA, USA
Laura.Mirkarimi@adeia.com

***Abstract*— Fast and accurate defect detection is critical for hybrid bonding because surface defects directly impact yield and processing costs. Optical microscopy provides a low-cost and high-throughput means to assess the quality of hybrid bonded wafers, but manual inspection and defect categorization requires significant time and effort. This paper presents an efficient and accurate optical inspection scheme where an optical microscope with automated stage and customized software collects full wafer images for designs ≥10 µm pad diameter. All the images are fed sequentially into two Convolutional Neural Network (CNN) models: a defect identification model (Model #1) and a defect categorization model (Model #2). With careful tuning of model parameters and category selection, Model #1 and Model #2 achieve 97% and 96% accuracy on the test set, respectively. Model #1 is then further validated by comparing the accuracy and time saved to the performance of a fully manual inspection by a highly trained human operator. The CNN model performs with higher *recall* and saves over 1.5 hours of inspection time per wafer. Finally, limitations of the current models and methods for model refinement are presented with additional discussion on future methods to streamline the inspection process of hybrid bonding.**

Keywords—hybrid bonding, defect detection, convolutional neural network

I. INTRODUCTION

Hybrid bonding is becoming increasingly relevant as the demands for next generation packaging require higher bandwidth architectures. Hybrid bonding enables finer pitch designs for denser interconnect arrays and enhanced electrical and thermal performance compared to incumbent solder-based interconnects [1]. To achieve successful hybrid bonding, both bonding surfaces require stringent surface topography control and cleanliness. The surface topography must be extremely flat without any foreign particles, metal corrosion or missing pads on the bonding surface [2].

Many characterization methods are employed to measure surface topography and ensure the quality of the material before bonding. Atomic force microscopy (AFM) is the current standard to check for Cu pad dishing after chemical mechanical planarization of the bonding surface, but it takes multiple minutes to measure only 9 pads in a 40 µm pitch design [3]. Optical profilometry is being explored as a higher-throughput alternative to AFM, where all the bonding pads on a full 200mm wafer can be inspected [4].

In contrast to the previously mentioned characterization techniques, optical microscopy does provide critical information about the back end of line (BEOL) pattern and surface contamination quality assessment with high throughput. Hybrid bonding is an atomic level bond so particle contamination identification in packaging becomes as important as particle control during wafer processing. Additionally, identification of defects and their root causes that reduce the electrical yield of hybrid bond interconnects are critical for process bring up and high volume manufacturing ramp.

A microscope setup allowing for automated image acquisition can capture the entire surface of a 200 mm wafer within ~1h. The images can then be used for defect inspection and categorization. This optical method captures all bond pads and is therefore magnitudes faster than the previously mentioned techniques of AFM and optical interferometry. In our prototype lab, much of the inspection process prior to bonding is achieved with human operators and can be streamlined with machine learning models that can automatically detect defects.

Many machine learning defect detection models have been developed for various packaging components like interconnect packages [5], printed circuit boards [6], [7], and wafers [8]. Each of these models achieves impressive training and test accuracy ranging from 93% to 99%. They also use different input data streams from camera images [5], thermographic images [7], to low-resolution microscope images [10]. Other models have been developed to examine wire-bonded [9], [10] and flip-chip bonded structures [11] with similar degrees of success, but there has not been as much published on the application of machine learning models to hybrid bonding defects. One such study on hybrid bonding demonstrates training accuracy of 99% and test accuracy of 98%, but uses scanning electron microscopy images as the data input and takes a different approach in the defect categorization [12].

This paper focuses on improving the efficiency and accuracy of the optical microscope inspection process by using CNNs to identify defects important in hybrid bonding. The inspection scheme design, training set classification breakdown, and CNN model architecture are discussed in detail. Results and various metrics highlighting the performance of the CNN models after training are shown. Finally, a validation experiment is conducted which demonstrates the improved efficiency and accuracy of applying CNN models in the hybrid bonding inspection process as opposed to a fully manual approach.

II. Design of Inspection Process

The inspection process begins with digital dark field images of a wafer at each die site (Fig. 1). The microscope automatically stitches together smaller images taken at 5x magnification into one large image of the full die-site. Before feeding these images into the CNN model, each stitched imaged is divided into 960 smaller images of array size 306 x 240.

After cropping, the images are fed into CNN Model #1 (Defect Identification). CNN Model #1 classifies the images into two categories: "defect" or "non-defect." All images that were labelled as "defect" by CNN Model #1 are then fed into CNN Model #2 (Defect Categorization). CNN Model #2 classifies the "defect" images into one of four categories: "discolored pad", "foreign material", "patterning", and "within pad."

The selection of two separate CNN models was consciously made to allow for introduction of human input, if needed. Defect classification by nature of having greater than two categories is a more complex problem to address than defect identification. Additionally, given that certain defects may be rarer than others, it may be difficult to amass a training set of sufficient size to train Model #2. Humans may be substituted into the process flow to replace CNN Model #2 if there are concerns that model performance is under-spec or if unexpected defects begin to arise. Humans are better at identifying new classes of defects that may occur rarely, whereas the CNN model can only classify defects that it has already been explicitly taught.

CNN Model #1 is simply an Xception model with global average pooling applied to the output of the last convolutional block. The Xception model was chosen because of its demonstrated excellent performance at image classification on ImageNet datasets [13]. Similarly, CNN Model #2 is based on the Xception model but with an additional flatten and dropout layer. The dropout layer was added to avoid overfitting and is discussed in the following sections.

III. CNN Model Training and Validation Experiment

A. Dataset Details

All images in the training, validation, and test sets are dark field images taken at 5x magnification on an Olympus MX61L with motorized stage and customized software for automatic image capture. Dark field imaging was chosen over bright field due to improved contrast for all defects. Additionally, images are from two separate wafer and design layouts. The two designs are 10 μm pad diameter with a 40 μm pitch and a 15 μm pad diameter with a 40 μm pitch.

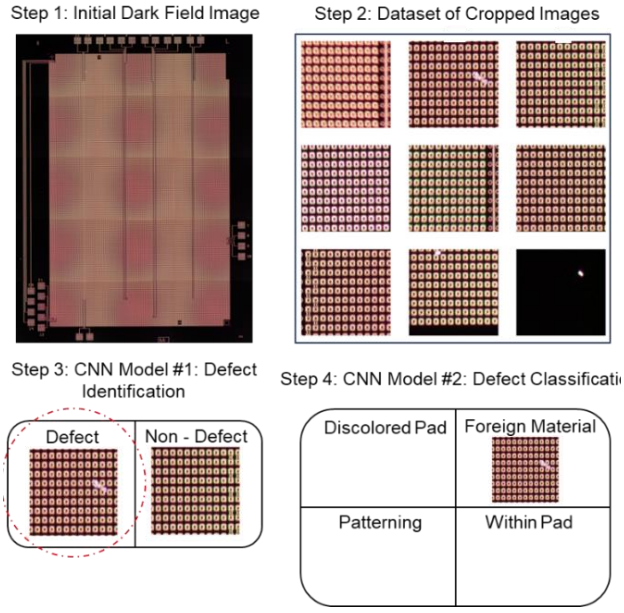


Fig. 1 Inspection process which includes data processing of stitched image into smaller, divided images, which are then fed into Model #1 (defect identification) and Model #2 (defect classification).

B. Training of CNN Model #1: Defect Identification

The breakdown of training, validation, and test set for CNN Model #1 is shown in Table 1. The training set is used to fit the model so all the model parameters can be tuned and optimized. The validation set is used to assess the quality of the training result. The test set is the final dataset, which is separate from training and validation and is as close to real-world data as possible, used to benchmark the model's performance. The binary classification problem of defect identification is by nature a skewed problem. The number of defect images is significantly less than the number of non-defect images in real data (<1% depending on the quality of the material being inspected). Therefore, it is important that the training dataset have enough defect examples to capture all defect types. Thousands of non-defect images were withheld from the training and validation set to avoid the problem of the model overemphasizing the result of a non-defect image during training. Additionally, the training set was augmented by images with varying degrees of saturation, brightness, rotation, zoom, and contrast to account for imaging conditions that may vary depending on the microscope setup.

For the training of Model #1, the binary cross-entropy function was implemented as the *loss* function, where N = total number of images, y_i = true label for the image, p_i = predicted probability of class 1 (1). Training was conducted over 200 epochs with the adaptive gradient optimizer and step size of 1e-3.

$$-\frac{1}{N}\sum_{i=1}^{N}[y_i * \log(p_i) + (1 - y_i) * \log(1 - p_i)] \quad (1)$$

C. Training of CNN Model #2: Defect Categorization

Categories for defects are ideally selected based on similarities in visual signature as well as similarities of the defect

TABLE I. MODEL #1: DEFECT IDENTIFICATION DATASET BREAKDOWN

Dataset	Binary Defect	
	Defects	***Non-Defects***
Training/ Validation [a]	508	3000
Test	218	7840

[a.] 70% of images in the Training/Validation split are for training, 30% of the images are for validation.

root cause. For instance, our standardized categories of patterning, foreign material, scratch, and discolored pad each have a distinct visual signature (Fig. 2). Patterning defects are indicated by misplaced or missing metal features and are caused by plating or lithography issues. Large foreign material defects like particles have a characteristic high brightness in dark field mode and a size and shape that is typically not correlated with the pad design. Scratch defects are thin line defects that typically have their origin in the chemical mechanical polishing step. Finally, discolored pad defects appear as hazy and confined features on the metal pad area and are typically a result of oxidation or corrosion of the metal surface.

The high-throughput method of optical inspection has the limitation of low magnification (5x) not being sufficient to resolve smaller defects. Cases arise where the defect signature is ambiguous, and the defect category is not obvious. Specifically, many smaller defects within the pad may appear similar, and additional characterization needs to be completed to verify the correct category. For instance, a defect which initially appears to be a scratch under high magnification optical inspection may actually be a foreign material defect as confirmed by scanning electron microscopy (SEM) and energy dispersive x-ray (EDX) analysis (Fig. 3). Table II shows results of the EDX analysis and how spectra collected in regions with the defect have significantly higher carbon concentration than the region in the copper pad away from the defect.

Given the limitation of the input data, different categories were selected for the training of CNN Model #2. In addition to the previously mentioned categories of discolored pad, foreign material, and patterning, a fourth category of defect "Within Pad" was created. The "Within Pad" category contains defects

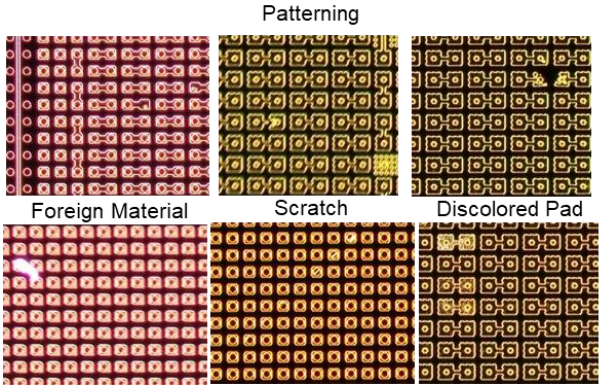


Fig. 2 Examples of clearly identifiable defects in dark field images.

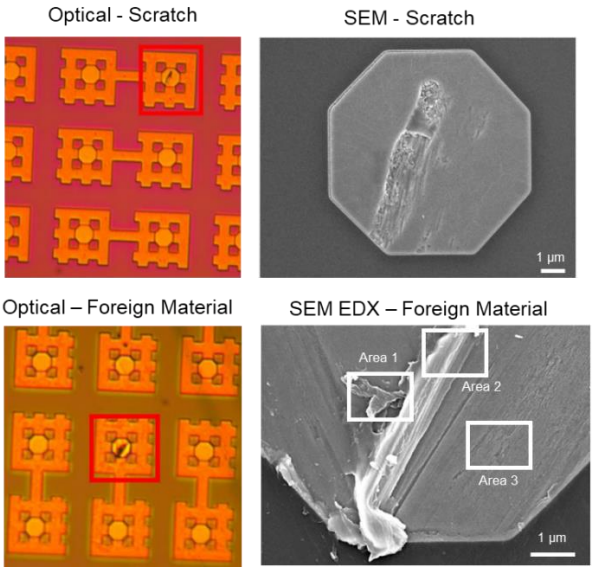


Fig. 3 Examples of defects that are ambiguous under high-magnification optical microscopy which need SEM and EDX analysis to resolve the defect type.

TABLE II. EDX CHEMICAL ANALYSIS RESULTS

Area	Carbon (%)	Oxygen (%)	Silicon (%)	Copper (%)
1	22.9	2.6	1.0	73.5
2	21.1	2.4	1.0	75.5
3	7.6	2.3	1.3	88.7

that may be scratches, foreign material which preferentially adheres to the Cu surface, or corrosion of a small portion of the pad.

The breakdown of training, validation, and test set for CNN Model #2 is shown in Table 2. The skewed nature of training/validation set is due to the limited quantity of material that is available. Of the available material, discolored pads and patterning defects are rarer than other defects. As more inspections are completed over time, the distribution of the training set can be adjusted for more robust training of CNN Model #2. Similar to CNN Model #1, data augmentation with images of varying degrees of saturation, brightness, rotation, zoom, and contrast was also applied.

For the training of Model #2, the categorical cross-entropy function, a generalized form of (1), with softmax activation was implemented as the *loss* function, where N = total number of images, y_i = true label for the image, p_i = predicted probability of image i being in a certain class, C (2). Training was conducted over 200 epochs with the adaptive gradient optimizer and step size of 1e-3.

$$-\frac{1}{N}\sum_{i=1}^{N}\log p_i[y_i \in C_{y_i}] \quad (2)$$

TABLE III. MODEL #2: DEFECT CATEGORIZATION DATASET BREAKDOWN

Dataset	Defect Categories			
	Discolored Pad	***Large Foreign Material***	***Patterning***	***Within Pad***
Training/ Validation [b]	42	394	117	279
Test	5	244	18	399

[a.] 70% of images in the Training/Validation split are for training, 30% of the images are for validation

D. *Model Validation Experiment*

After iterations of training, the efficiency and accuracy of the CNN models were assessed in the context of a real inspection process for hybrid bonding. In this experiment, a fully manual inspection approach was compared against a CNN-aided approach. We tested two wafers with different designs, a host wafer and a die wafer. To start, stitched dark field optical images of all sites on both wafers were taken. In the manual approach, a highly trained human operator examined the stitched optical microscope image and created a defect inspection map by categorizing the defect and labelling the severity of the defect. In the CNN-aided approach, CNN Model #1 flagged defective images and then a human created a similar defect inspection map but using only the images that were flagged by CNN Model #1. In this experiment, the inspection map only contained information on the most severe defect per die or bond site as decided by the human operator in case the die or bond site showed more than one defect. The severity of a defect is important to decide whether a defect is likely to impact the performance of the bonded device. After mapping, the total number of defective sites and the inspection time were compared.

IV. RESULTS AND DISCUSSION

A. *CNN Model #1: Defect Idenification*

The accuracy, which is defined in Equation 3 where N = total number of images in the dataset (3), and *loss* (1) (2), a measure of how incorrect the predicted values of a set of images compare to the true labels (defect vs. non-defect), of training, validation, and test sets are shown in Fig. 4. Both training and validation accuracy are 98.8% with test accuracy at 97.2%. Similar performance across all three datasets is also evident when examining the *loss* function: 0.039 for training, 0.043 for validation, and 0.022 for test. Because the accuracy and *loss* of the test set are similar to the other datasets, the model is not overfitting the data and the complexity of the model is correct for the given defect identification problem. As a result, no additional dense layers were added to the output of the Xception model.

A *precision-recall* curve of Model #1 applied to the test data was plotted using scikit-learn [14] by varying the threshold of defect vs. non-defect from 0 to 1. The area under the curve score was measured to be 0.936, which is comparable to other models used for defect classification of packaging materials [7], [15]. For reference, a confusion matrix (Table IV) and definitions of *precision* (4) and recall (5) are shown. For selecting a threshold

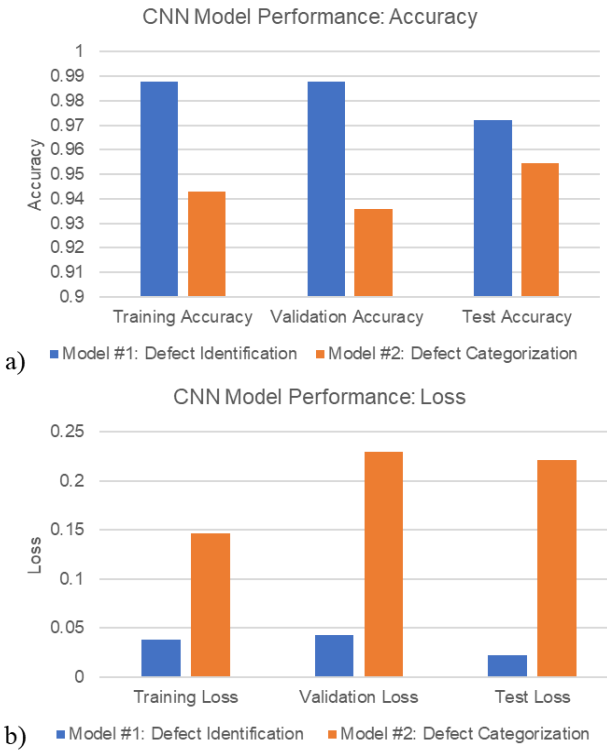


Fig. 4 Performance metrics of a) accuracy and b) *loss* for both CNN Model #1 and Model #2.

of defect versus non-defect, a threshold of 0.11 (p > 0.11 = predicted defect) was chosen which has a computed *precision* of 72% and *recall* of 92%. This threshold was chosen because higher *recall* is more important than a high *precision* in the case of hybrid bonding. The material cost of producing hybrid bonding samples is large so the yield must be high (i.e., FN must be small) even if the inspection process takes a longer time.

$$Accuracy = \frac{\text{TP} + \text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} = \frac{TP+TN}{N} \quad (3)$$

$$Precision = \frac{\text{TP}}{\text{TP}+\text{FP}} \quad (4)$$

$$Recall = \frac{\text{TP}}{\text{TP}+\text{FN}} \quad (5)$$

After determining the optimal threshold, the fully manual inspection approach, where a human operator locates the defects and creates the inspection map, was compared against the CNN-aided approach, where Model #1 was applied to flag

TABLE IV. GENERALIZED CONFUSION MATRIX

		Actual	
		Positive (defect)	Negative (no defect)
Model Predicton	Positive (defect)	True Positive (TP)	False Positive (FP)
	Negative (no defect)	False Negative (FN)	True Negative (TN)

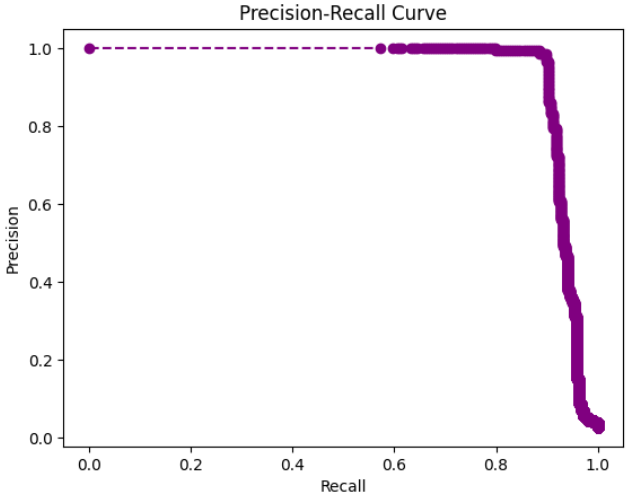


Fig. 5 *Precision-recall* curve of Model #1 applied to test set data.

the defects and then a human operator creates the inspection map. The inspection maps of each approach are shown in Fig. 6. The CNN inspection map has considerably more die sites labelled as defective than the manual approach. If it is assumed that the human operator categorizing the defects only identifies true defects, then the results indicate that the *recall* of the CNN approach is better than the manual approach. All the defects identified by the CNN approach and not by the manual approach were true defects that the human operator failed to identify. Fig. 6 also shows specific die sites labelled with different defect types and severity depending on the inspection scheme, e.g. sites 10, 82, 119. This is explained either by the fact that 1) the images which the CNN marks as defective are cropped and zoomed in which helps the human operator better resolve the defect type or 2) a different and more severe defect was found in a given site by the manual method since only the most severe defect is recorded for each site.

Comparing the total number of defects found by the two inspection methods, the CNN method identified almost twice as many defects as the manual method (Table V). Of the extra defects found, many are within the "Foreign Material" category. Since a large portion of the foreign material defects on this wafer are small points of bright pixels (Fig. 7), they are difficult to identify in the manual approach where the image is a dense, stitched image of tens of thousands of copper pads.

When comparing the performance of the two inspection schemes for the die wafer, the CNN method again finds more total defective sites (Table VI). However, in this case, the number of scratch defects is significantly fewer in the CNN approach than the manual. Scratch defects are generally not as detrimental to yield as compared to other defects. As long as the surface of a Cu pad is not protruded above the dielectric, the of material in the scratch will likely not create an electrical failure. Given the relatively mild nature of the scratch defect, many sites were labelled with other more severe defects in the CNN approach.

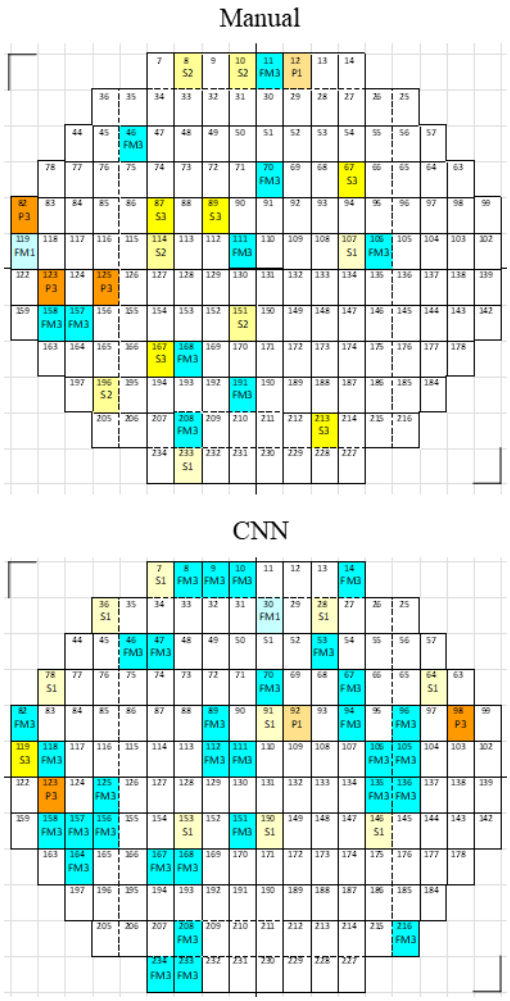


Fig. 6 Inspection map of host wafer generated with a manual approach vs. CNN-aided approach. Defect key: FM = foreign material. S = scratch. P = patterning. 3 = critical severity. 2 = moderate severity. 1 = mild severity.

The CNN approach reduced the required human labor by 105 minutes per host wafer and 95 minutes per die wafer over the manual method. This significant labor saving can be improved further with the development of a CNN categorization model which takes into account defect severity.

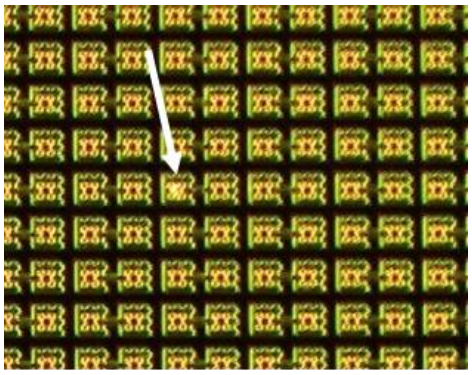
An additional metric of throughput time, which includes time needed for the stitched microscope image to be divided into smaller images and then processed by CNN Model #1, was tracked to compare the two approaches. The throughput time

TABLE V. NUMBER OF DEFECTIVE DIE SITES ON HOST WAFER

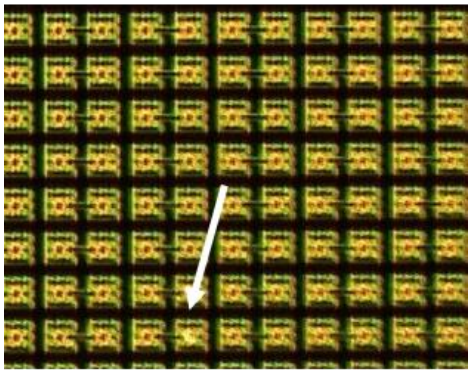
	Discolored Pad	***Foreign Material***	***Patterning***	***Scratch***	***Total***
Manual [a]	0	11	4	12	27
CNN [b]	0	33	3	10	46

a. Completed in 150 minutes of human labor / throughput time.

b. Completed in 45 minutes of human labor time and 150 minutes of throughput time.



Site 9 (FM3)



Site 14 (FM3)

Fig. 7 Examples of defects missed by the fully human approach but caught and identified by the CNN-aided approach.

for the CNN-based approach adds an additional 105 minutes to the human labor time: 5 minutes for image processing and 100 minutes for CNN Model #1 to process the images. The CNN approach did not affect throughput time per host wafer and increased throughput time by 10 minutes per die wafer (Table V and VI). With minimal increases to throughput time, the CNN-based inspection process will not affect queue times between processing steps. It should be noted that all image processing was completed on a Lenovo Tiny Thinkstation P3 with 12th Gen Intel i9-12900T 1.40 GHz processor and no installed GPU. The throughput time of the CNN-approach can be significantly shortened with improved hardware

B. *CNN Model #2: Defect Categorization*

CNN Model #2 achieves acceptable accuracies of 94.3%, 93.6%, and 95.5% for the training, validation, and test sets respectively (Fig. 4). However, when compared to Model #1, both *loss* and accuracy are worse. The decrease in performance is attributed to the limited number of training examples of the discolored pad and patterning defect categories, which could not be overcome even with data augmentation. As more inspections are completed, a larger dataset in those two categories will improve the model performance.

TABLE VI. NUMBER OF DEFECTIVE DIE SITES ON DIE WAFER

	Discolored Pad	***Foreign Material***	***Patterning***	***Scratch***	***Total***
Manual [a]	29	53	48	41	171
CNN [b]	46	70	63	17	196

a. Completed in 250 minutes of human labor / throughput time.

b. Completed in 155 minutes of human labor time and 260 minutes of throughput time.

Model #2 demonstrates similar performance across training, validation, and test sets, indicating that the model complexity is correctly tuned to the classification problem (Fig. 4). Initially, only the Xception model with no additional layers was implemented but resulted in a large difference of > 0.3 in *loss* between test and training sets. To address this, a dropout layer was added before the final output layer, and the dropout factor was varied in increments of 0.25 (Fig. 8). Each dropout factor was trained with 200 epochs and the model with the lowest combined training, validation, and test set *loss* was used to plot Fig. 8. Increasing the dropout factor has the expected result of increasing the training *loss* due to reduced model complexity. But at a dropout factor 0.5, there is a local minimum of both validation and test *loss*, suggesting that a dropout factor of 0.5 is the ideal choice.

V. CONCLUSION

In this paper, an inspection scheme was presented which has two independent CNNs both using the Xception model. CNN Model #1 is for defect identification and CNN Model #2 is for defect categorization. The advantage of a two-model approach is the ability to introduce a human operator into the inspection scheme in the case of model performance drift and the possible presence of unknown defects. Both models demonstrated good performance on all datasets, with Model #1 performing slightly better than Model #2.

The two-model inspection scheme was validated with a simple experiment in the context of a real inspection process for hybrid bonding. The CNN approach approximately halved the operator time necessary for inspection and resulted in more total defects found (higher *recall*) than a fully manual approach. Therefore, we confidently rely on the CNN approach which frees up the time of our technical staff which would otherwise be spent on inspection.

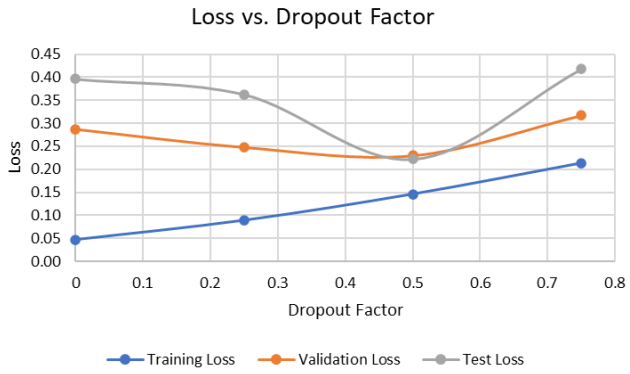


Fig. 8 Effect of changing the dropout factor in the dropout layer on the training, validation, and test *loss*.

These promising results suggest several avenues for future work including further refinement of Model #2. Expanding the dataset to include more examples of the rarer defects such as discolored pad and patterning defects will improve the accuracy and performance of Model #2. Additionally, more relevant defect categories can be used in Model #2 if we can modify the architecture to incorporate an additional CNN. The new CNN would take topography data from other characterization techniques (AFM or optical profilometry) as input and would then be stacked on top of the original CNN such that the final output can make the classification decision based on both optical microscopy and topography images. Smaller "within pad" defects can then be properly binned into the "scratch" or "foreign material" categories if the defect is characterized as a valley or protrusion, respectively. Specifically, high throughput optical profilometry and AFM could be used as options for topography characterization methods. The use of chemical methods such as EDX as shown in Fig. 3 is likely infeasible because of its extremely limited throughput. Finally, the development of a model which can classify defect severity is important because the ability to differentiate between defects that negatively affect bonding yield versus defects that are benign would improve the effective post-inspection yield and reduce material waste. These improvements will further streamline the inspection process, thereby enabling high-yield hybrid bonding with reduced inspection process variability and improved efficiency.

ACKNOWLEDGEMENTS

The authors would like to thank Marita Adsuara for her work completing the manual inspection and inspection mapping. We would also like to thank Gill Fountain for his support and providing the wafers necessary to complete this project.

REFERENCES

[1] L. Mirkarimi, "3D Chiplet Integration with Hybrid Bonding," *Chip Scale Review*, vol. 27, pp. 7–11, Apr. 2023.

[2] E. Paul, "3D Technology Platform – Advanced Direct Bond Technology," *3D Integration for VLSI Systems*, p. 175, 2011.

[3] B. Lee *et al.*, "Nanoscale Topography Characterization for Direct Bond Interconnect," in *2019 IEEE 69th Electronic Components and Technology Conference (ECTC)*, Las Vegas, NV, USA: IEEE, May 2019, pp. 1041–1046. doi: 10.1109/ECTC.2019.00163.

[4] B. Lee *et al.*, "High-Throughput Characterization of Nanoscale Topography for Hybrid Bonding by Optical Interferometry," presented at the 2024 IEEE 74th Electronic Components and Technology Conference,

[5] K. M. Malinski and K. Okarma, "Application of CNN-Based Method for Automatic Detection and Classification of the IC Packages," in *2020 16th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, Shenzhen, China: IEEE, Dec. 2020, pp. 944–950. doi: 10.1109/ICARCV50220.2020.9305493.

[6] Y.-S. Deng, A.-C. Luo, and M.-J. Dai, "Building an Automatic Defect Verification System Using Deep Neural Network for PCB Defect Classification," in *2018 4th International Conference on Frontiers of Signal Processing (ICFSP)*, Poitiers: IEEE, Sep. 2018, pp. 145–149. doi: 10.1109/ICFSP.2018.8552045.

[7] M. Jeon, S. Yoo, and S.-W. Kim, "A Contactless PCBA Defect Detection Method: Convolutional Neural Networks With Thermographic Images," *IEEE Trans. Compon., Packag. Manufact. Technol.*, vol. 12, no. 3, pp. 489–501, Mar. 2022, doi: 10.1109/TCPMT.2022.3147319.

[8] J.-C. Chien, M.-T. Wu, and J.-D. Lee, "Inspection and Classification of Semiconductor Wafer Surface Defects Using CNN Deep Learning Networks," *Applied Sciences*, vol. 10, no. 15, p. 5340, Aug. 2020, doi: 10.3390/app10155340.

[9] J. Chen, Z. Zhang, and F. Wu, "A data-driven method for enhancing the image-based automatic inspection of IC wire bonding defects," *International Journal of Production Research*, vol. 59, no. 16, pp. 4779–4793, 2020, doi: 10.1080/00207543.2020.1821928.

[10] W. Pan, T. Tang, M. Chen, and F. Mo, "Automatic detection of wire bonding defects in microwave components using multi-stage hybrid methods based on deep learning," *Measurement Science and Technology*, vol. 34, no. 11, doi: 10.1088/1361-6501/ace926.

[11] Z. Long, X. Zhou, and X. Wu, "Cascaded Approach to Defect Location and Classification in Microelectronic Bonded Joints: Improved Level Set and Random Forest," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4403–4412, Jul. 2020, doi: 10.1109/TII.2019.2950496.

[12] R. Komatireddit, S. Dangayack, P. Lianto, R. Cherikkallil, and S. Rupa, "Defect Classification using Deep Learning for Hybrid Bonding Application," presented at the 2022 IEEE 24th Electronics Packaging Technology Conference (EPTC),

[13] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions." arXiv, Apr. 04, 2017. Accessed: Jan. 08, 2024. [Online]. Available: http://arxiv.org/abs/1610.02357

[14] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *MACHINE LEARNING IN PYTHON*.

[15] A. Bhattacharya and S. G. Cloutier, "End-to-end deep learning framework for printed circuit board manufacturing defect classification," *Sci Rep*, vol. 12, no. 1, p. 12559, Jul. 2022, doi: 10.1038/s41598-022-16302-3.