6

Failure Detection in Silicon Package

Saad Al-Baddai and Jan Papadoudis

Infineon Technologies AG, Germany

Abstract

In an ever more connected world, semiconductor devices represent the core of every technically sophisticated system. The desired quality and effectiveness of such a system through assembly and packaging processes is high demanding. In order to achieve an expected quality, the output of each process must be inspected either manually or rule-based. The latter would lead to high over-reject rates which require a lot of additional manual effort. Moreover, such an inspection is sort of handcrafted by engineers, who can only extract shallow features. As a result, either more yield-losses due to an increase in the rejection rate or more products with low quality will be shipped. Therefore, the demand for advanced image inspection techniques is constantly increasing. Recently, machine learning and deep learning algorithms are playing an increasingly critical role to fulfil this demand and therefore have been introduced in multiple applications. In this paper, an overview of the potential use of advanced machine learning techniques is explored by showcasing of image and wirebonding inspection in semiconductor manufacturing. The results are very promising and show that AI models can find failures accurately in a complex environment.

Keywords: anomaly detection, labelling, manufacturing AI solutions, AI integration, transfer learning, scalability.

6.1 Introduction and Background

Semiconductor manufacturing produces the most highly advanced microchips in the world. A manufacturing process of these chips goes through multiple sequences and interacting sub-processes and during that operates in extreme quality-demanding conditions. Thus, it has an increasing complexity and demand on quality requirements, as electronics increasingly become an important part of modern society. In principle semiconductor manufacturing is equipped with lots of sensors to monitor the processes but it lacks a suitable way to make use of this data. However, due to the complexity of the processes and unknown correlation among the collected data, such traditional techniques become quite limited. Here's where AI takes the initiative and offers a promising solution for feature extraction, condition monitoring and fault modelling for anomaly/defect detection using sophisticated algorithms [5]. Therefore, one of the success factors in optimizing the industrial processes is either automatic anomaly detection, supervised learning or both, which leads to prevent production flaws, herewith improving quality, increasing yields and making benefits. The popular way of anomaly detection in many of industrial application is by adjusting digital camera parameters or sensors during collecting images or time series data. This is basically an image or signal anomaly detection problem that is searching later on for patterns that are different from normal data [4]. As a human one can easily manage such task by recognizing of normal patterns, but this is relatively not easy for machines. Unlike other classical approach, image anomaly detection faces some of the following difficult challenges: class imbalance, quality of data, and unknown anomaly [4]. A prevalence of abnormal events are generally exception, whereas normal events account for a significant proportion. Some techniques usually handle the anomaly detection problem as a "one-class" problem. Here models learn by using the normal data as truth ground and afterwards evaluates whether the new data belong to this truth ground or not, by the degree of similarity to the truth ground. In the early applications of surface defect detection, the background is often modeled by designing handmade features on defect-free data. For example, Bennatnoun et al. used blobs technique [3] to characterize the correct texture and to detect deviations through changes in the charter ships of generated blobs. Amet et al. [2] used wavelet filters to extract different scales of defect-free images, then extracted the informative features of different frequency scales of images. However, most of these methods focus can work with homogeneous date with good quality and would require a prior knowledge. Generally, still some

challenges which strongly depend on the field of application. Thus, there is no universal pattern or system, which does not directly allow to use techniques developed for one application to another. Thus, machine/deep learning offers promising solutions in such complex environment. However, the former can be adapted or scaled to other application or use cases. Due to these above-mentioned challenges unsupervised anomaly detection on multi-dimensional data is very highly demanding in machine learning and business applications [6]. Please note, this paper is extended of the published work in [1]. The latter focused on the data preparation, labelling techniques and preliminary results. A new contribution related to quantities, framework and transfer learning and scalability is presented. Therefore, a short description about the data is introduced. Then, labelling approach is shortly discussed. Afterwards, framework is depicted and effective of transfer learning is discussed. Finally, the results are showed and conclusion is drawn.

6.2 Dataset Description

This manuscript showcases dealing with time series data as well as with images at different processes during packaging. The data for the first case is collected in the early phase, at wirebonding process. These data are collected from three different sensors. Namely a current sensor, located at the transducer, a displacement sensor measuring the deformation of the wire respectively the path of the bonding tool and a frequency sensor, also located at the transducer of the wirebonder. Each of these sensors collects roughly 432 features during 143 timestamps. However, the collected data are highly redundant (see Figure 6.1). This is because there is multiple bond connection on one device which share the same process parameters and behave quite similar. However, sometimes, contamination of the device or a misadjusted machine would cause misaligned or deformed bonds, see Figure 6.1. Here, there is a need to develop a ML solution for detecting such deviations. Similarly, the biggest challenge of the outgoing optical inspection (OOI), in the second use case, is the defect detection on the heatsink, see Figure 6.1, which consists of a rough copper surface. It needs to inspected for scratches, metal or mold particles as well as for mechanical damage like imprints. However, this surface shows a very high variety in appearance, as it is oxidized during preceding high temperature testing steps. Hence, the inspection cannot be carried out using rule-based algorithms, as the oxidized areas cannot be distinguished clearly from true defects by a rule-based algorithm. In this

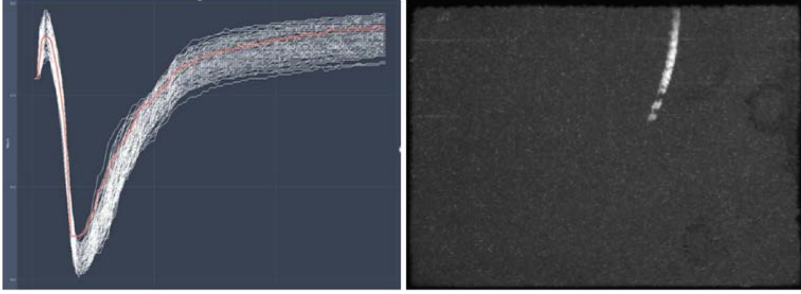

Figure 6.1 *Left*: Curve with abnormal minimum position (red) in comparison to normal ones (white) of recorded sensor data during wirebonding process. *Right*: shows an example of abnormal OOI image with shown crack on the surface.

context, trained personnel took care of the heatsink inspection and was used to label the image data, roughly 300 images, for supervised learning.

6.2.1 Data Collection and Labelling

Data labelling is an essential step in a machine learning area. Here, the common phrase "Garbage in - Garbage out" is used very commonly in the ML community, that means the quality of the model strongly depends on the quality of the (labelled) training data. In this work, two approached are considered:

- X $\rightarrow$ Y

Indeed, data labelling is a task that requires a lot of manual work. In this approach, labelling data(images) is done based on human experience. Luckily, only few percent of data had to reviewed after applying the tool introduced in [1] for reducing the effort. This process is done by review the sorted data of historical images and recognize on the defects by looking closely at heat sink surface. Thus, there is no need for prior knowledge about the status of Y machine to sort out X data. Afterwards, simply, the data can be categorized into two categories as either healthy(good) or unhealthy(fail). These data, then, can be used for training the AI model. This approach is used for labelling the first case OOI.

- Y $\rightarrow$ X

Contrary to the first approach, in this approach human's experience unfortunately is not fully helpful for labelling data, as the data is very complex. Hence, the design of experiment (DOE) is set by checking the machine status while collecting data. Therefore, a predefined mis-adjustment in Y wire bond should be known to get deviation on X data.

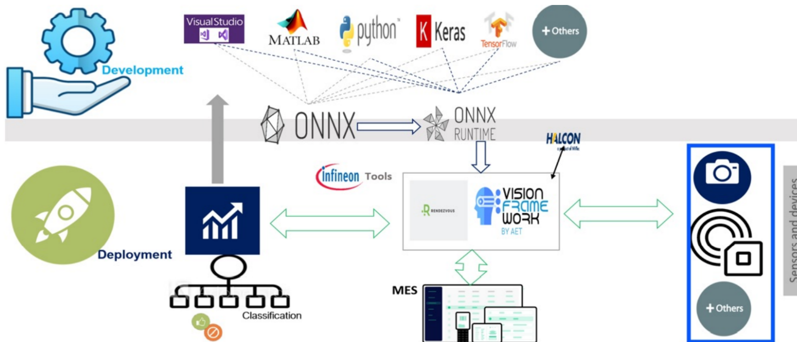

Figure 6.2 Flow chart of development and deployment life cycle for AI solution at IFX. In development phase data scientists could use different programming language as the final model can be converted to ONNX. In deployment phase, the vision frame can simply access to ONNX and run during inference time.

6.3 Development and Deployment

In order to satisfy the robustness requirements of AI model, we propose the AI framework to be adapted to the best practices with the following characteristics

- Short adaption cycles.
- Testing in every stage and automatically integration and deployment.
- Reproducible processes and reliable software releases.

Figure 6.2 shows a typical DevOps process which is the basis for continuous integration and delivery. Thus, the following feedback loops are added to the process in order to integrate central ML lifecycle steps:

- Define and build a suitable model and improve it based on demo feedback through experiments using any suitable programming language.
- Converting the optimal model, based on observed model performance, into ONNX (or other suitable format) and integrating it to the target AI platform.
- Retrain, when it is needed, an operational model based on new real-life data and report the performance.
- Adapt result of the whole process based on the performance of the models on productive data.

However, for deployment, it gets more complex, because of additional types of IFX infrastructure must be considered. Here, Figure 6.3 shows the process which is extended by the new development into the existing IFX

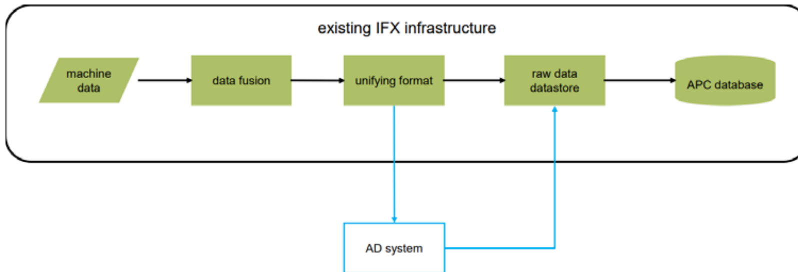

Figure 6.3 Process flow integration of the developed AD solution into an existing IFX infrastructure.

infrastructure. From the perspective of a classic ML lifecycle, the role setting of Business Analysts together with Data Scientists and Data Engineers is sufficient for conducting a working ML solution which proves to deliver all required benefits.

6.4 Transfer Learning and Scalability

Transfer learning is simply fine-tuning previously trained neural networks. In this context we transfer the trained model on OOI data into other processes of packaging, see Figure 6.4 Thus, instead of creating an AI model from scratch, only a few images of the new process are enough for fine tuning the pre-trained model of OOI images. Interestingly, not only the collected images from new process are similar to the OOI images but the defect types as well. As a result, the model reports a high accuracy as is shown in Table **??**. The anomaly detection for the wire bonding process has a wide range of application, as there are multiple Infineon sites and multiple machines of the same type. The training of an anomaly detection model can benefit from unlabelled data under the assumption that the majority of the data is good. Given the general high yield this assumption is valid. Given multiple similar machines there are two approaches to scale one model to multiple machines.

- Using data from multiple machines for the training. Thus, the model implicitly learns differences between the machines and the same model can be used for multiple machines.
- Using an anomaly detection model, which was trained on a prior defined machine and setting up all other machines to behave most similar to the selected machine. Thus, all other machines generate raw data of the same input space as the selected machine.

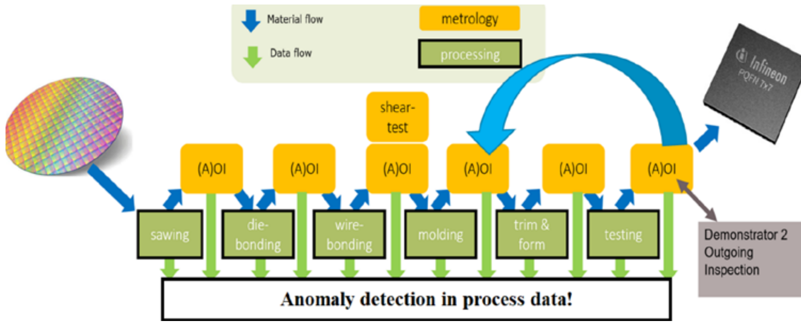

Figure 6.4 show the flow processes during silicon package, the backside blue arrow shows the position of transfer learning from OOI backwards to taken images after molding process, see Figure 6.5

Figure 6.5 shows an example of the OOI image on left side (This image is taken before shopping and after electrical test) and example of image after molding process on right side.

With this procedure it was possible to scale one model to s complete production line with more than 30 machines.

6.5 Result and Discussion

For wire bonding use case, two different approaches to validate the system were made. The first one was to simply calculate the percentage of devices which showed an anomaly in the dataset and compare this to the process yield. If these percentages align this is a good indicator that the anomaly detection represents the product quality. Additionally, a statistical significantly correlation between high anomaly values and bad electrical test results is considered. For the second approach, we gathered multiple devices which showed a high anomaly value and examined them thoroughly. In all of the cases different influences could be found on the device, like a contaminated device, reduced shear value or input material which was out of specifications. But not all findings, even though varying from the normal, will lead to a malfunctioning device. However, an important aspect

of the used anomaly detection was that the result is an anomaly score, indicating how different the raw data from normal is not a Boolean indication anomaly / no anomaly. Thus, it is necessary to find an optimal threshold on which the difference in the raw data influences the quality of the product. An important impact of the work was also the adaptation of the approach to a performant data management infrastructure; i. e. the development of automatable methods for the detection of conspicuous parameter behaviour and its marking and storage. The evaluation was based on sample data and statistical analysis of standard deviations considering Nelson's rules. The work carried out covers both the familiarization with the various technologies and their variants, the adaptation of the methods to the subject area, and the prototypical implementation and testing of the algorithms by embedding them in automated analysis pipelines. Currently the anomaly detection for wirebonding is running on over 40 machines on 3 different IFX sites. During a runtime of 4 months, several misadjusted bonders were detected, random errors and contaminated devices. However, currently a big focus is set to fully integrate the model not only in the infrastructure but also in the day to day workflow of the operators, this also includes a clear definition of action plans for found deviations and trainings of operators. For OOI use case, after collecting images, the labeled images are pre-processed first by cropping the region of interest and normalization the intensity values between 0 and 1. These images are sent to CNN for training purpose. The CNN consist of 100 layers. The latter consisting of different blocks. Each block contains the convolutional, pooling and ReLU layer. Also, before the last layer, fully connected layer, a strict regularization factor is added in order to avoid over-fitting issue by adding dropout layer with value 0.6. The data was splited into 80% training and 20% validation data. The model reported with accuracy higher than 99%. Afterwards, the model is tested on productive data with roughly 25k images. Table 6.1 shows the confusion matrix with the important measures, sensitivity, specificity and accuracy. As, one can see that model to follow zero defect philosophy, as sensitivity value is 100%. The accuracy also is less than 1%. Hence, only the latter have to be reviewed by an expert. Moreover, the performance model after scaling to anew process is still very robust. As one can see in the Table 6.2, which shows the reported results by a model when run on productive data of the new process. Although, one can see there is one escapee in bottom surface (BOT), but the accuracy is still higher than 99%.

Table 6.1 Show the confusion matrix and metrics of the CNN model on productive data for BOT and TOP of OOI images.

BOT	Defect	Good
Defect	250	379
Good	0	39921
Acc.:99,07%	Sen.:100%	Spe.:99,06%

TOP	Defect	Good
Defect	130	220
Good	0	25000
Acc.:99,13%	Sen.:100%	Spe.:99,13%

Table 6.2 Show the confusion matrix and metrics of the CNN model on productive data for BOT and TOP of the new process.

BOT	Defect	Good
Defect	227	198
Good	0	26063
Acc.:99,25%	Sen.:100%	Spe.:99,25%

TOP	Defect	Good
Defect	751	60
Good	1	9353
Acc.:99,40%	Sen.:99,87%	Spe.:99,36%

6.6 Conclusion and Outlooks

In this paper, two use cases show the potential benefits of using AI models in detecting abnormalities in industrial packages. Moreover, the methodology shows the possibility of scaling such solutions to new similar use cases or machines with minimum effort. As a result, not only the manual effort would significantly be reduced, but also costs and the quality of the products would be improved. Additionally, the long-term goal is not only to find the deviation but to detect exactly the root cause behind it. However, there is still a lot of work left, unrealized potentials benefit of AI solutions, but IFX has already taken a step forward in the right direction. Thus, semiconductor community is investing more with AI to harvest its benefits in the short and, most importantly, long term. Generally, the results are promising and would be a good alternative to classical approaches. The next steps are monitoring, optimization and more validation for both solutions in a productive environment.

Acknowledgements

AI4DI receives funding within the Electronic Components and Systems for European Leadership Joint Undertaking (ECSEL JU) in collaboration with the European Union's Horizon2020 Framework Programme and National Authorities, under grant agreement n° 826060.

References

[1] S. Al-Baddai, M. Juhrisch, J. Papadoudis, A. Renner, L. Bernhard, C. Luca, F. Haas, and W. Schober. Automated Anomaly Detection through Assembly and Packaging Process, pages 161–176. 09 2021.

[2] A. Amet, A. Ertuzun, and A. Ercil. Texture defect detection using subband domain co-occurrence matrices. pages 205 – 210, 05 1998.

[3] A. Bodnarova, M. Bennamoun, and K. Kubik. Automatic visual inspection and flaw detection in textile materials: A review. pages 194–197, 01 2001.

[4] T. Ehret, A. Davy, J. M. Morel, and M. Delbracio. Image anomalies: a review and synthesis of detection methods. 08 2018.

[5] G. A. Susto, M. Terzi, and A. Beghi. Anomaly detection approaches for semiconductor manufacturing. Procedia Manufacturing, 11:2018–2024, 12 2017.

[6] B. Zong, Q. Song, M. R. Min, W. Cheng, C. Lumezanu, D. Cho, and H. Chen. Deep autoencoding gaussian mixture model for unsupervised anomaly detection, 2018.