Application of Machine Learning Methods to Predict the Quality of Electric Circuit Boards of a Production Line

Immo Schmidt[1], Lorenz Dingeldein[2], David Hünemohr[3], Henrik Simon[4], and Max Weigert[5]

[1,2,3,4,5] *Technische Universität Darmstadt, Institute of Flight Systems and Automatic Control, Darmstadt, 64287, Germany*
{schmidt, dingeldein, huenemohr, simon, weigert}@fsr.tu-darmstadt.de

Abstract

For the data challenge of the 2022 European PHM conference, data from a production line of electric circuit boards is provided to assess the quality of the produced components. The solution presented in this paper was elaborated to fulfill the data challenge objectives of predicting defects found in an automatic inspection at the end of the production line, predicting the result of a following human inspection and predicting the result of the repair of the defect components. Machine learning methods are used to accomplish the different prediction tasks. In order to build a reliable machine learning model, the steps of data preparation, feature engineering and model selection are performed. Finally, different models are chosen and implemented for the different sub-tasks. The prediction of defects in the automatic inspection is modeled with a multi-layer perceptron neural network, the prediction of human inspection is modeled using a random forest algorithm. For the prediction of human repair, a decision tree is implemented.

1. Introduction

The fourth industrial revolution leads to increasingly automated production and manufacturing. Production machines that are fully connected and fully equipped with sensors generate huge amounts of data enabling new data-driven approaches to assess the quality of the produced parts. Machine learning algorithms are used in an increasing number of applications in production, even if their use is often part of research and not yet widely spread (Mayr et al., 2019; Liukkonen, Havia, & Hiltunen, 2012).

Within the production environment, machine learning provides the opportunity to process the large amounts of data to improve quality, lower costs or increase the flexibility of the process and can contribute to sustainable manufacturing (Mayr et al., 2019; Cioffi, Travaglioni, Piscitelli, Petrillo, & De Felice, 2020). In recent years, deep learning approaches for smart manufacturing have been increasingly studied (Wang, Ma, Zhang, Gao, & Wu, 2018). In contrast to engineering features with expert knowledge of the manufacturing process, deep learning provides an end-to-end machine learning approach, but oftentimes lacks interpretability of the results (Wang et al., 2018).

Within the framework of data challenges in the field of Prognostics and Health Management, several objectives related to various industrial use cases have already been addressed. In this context, the suitability of different algorithms could be demonstrated and compared. As a result, it has been possible to gather new knowledge about problem-specific approaches and insights into general solution strategies (Huang, Di, Jin, & Lee, 2017). Data sets from the manufacturing industry are currently scarce, but very useful for investigating data-based improvements in maintenance and quality control processes (Jourdan, Longard, Biegel, & Metternich, 2021).

Predictive quality is the main focus of the data challenge for the 7th European Conference of the Prognostics and Health Management Society 2022 that is held in cooperation with Bitron Spa. The participants of the challenge receive data from a production line manufacturing electric circuit boards. In the production process, the surface mount technology (SMT) is used, which comprises of several manufacturing and inspection steps. The use of advanced data-driven algorithms for quality management in mass soldering processes dates back to the 1990s (Liukkonen et al., 2012). Since then the technology for the production of electronic components as well as the capabilities of machine learning algorithms evolved. Images of optical inspections of the circuit boards solder prints can be processed with machine vision algorithms to detect defects directly (Zakaria, Amir, Yaakob, & Nazemi, 2020). Indirectly supervised learning approaches enhance the outcome of current automatic optical inspections and measured solder joint dimensions. This can be used in the production line to support the operator in assessing defect calls from the automatic optical inspection and identify false positive classifications (Jabbar et al., 2019). Specifically, Jabbar et al. compared

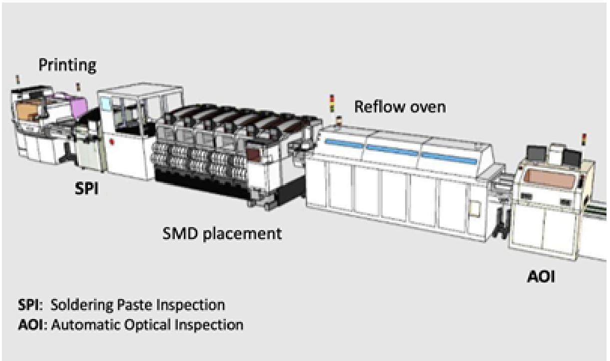

Figure 1. PCB production line (PHM Society, 2022)

tree-based machine learning algorithms for this use case and achieved a good performance (Jabbar et al., 2019). Supervised deep learning has also been applied to real production environment data to enhance the outcome of the optical inspections and reduce labor cost (Chang, Wei, Chen, & Hsieh, 2018).

1.1. Use Case

The production line for the manufacturing of electric circuit boards is depicted in Figure 1. In a first production step, a printing machine places the solder paste on the initially bare printed circuit boards (PCB). Subsequently, the electronic components are mounted on the PCB in the surface mount device (SMD) placement and the soldering process is finished in the reflow oven. Two different datasets are retrieved from the production process. The first one stems from a solder paste inspection (SPI) conducted after the placement of the solder paste and before mounting the electronic components. In this inspection, the quality of the solder paste placing in terms of, among others, volume, height and position is measured for each pin of the PCB. The second set of data is recorded in an automatic optical inspection (AOI) that follows after the soldering. During this inspection, defects on the produced PCBs shall be automatically detected. In case of defect detection, an operator visually inspects the PCB and confirms or rejects the defect found by the AOI. Only in case of a confirmation by the operator, a second operator investigates whether the defect can be repaired.

1.2. Objectives

The objective of the data challenge is the prediction of the automatic optical inspection for each component and, in case of a defect, the prediction of the assessment of the two operators. Consequently, the challenge is divided into three sub-tasks:

1. Prediction of components with a detected defect in AOI based on the SPI data. For each detected defect, a so-called *AOILabel* with information on the defect type is assigned to the component. Prediction of the defect type is not required, the only objective is to predict whether there is an *AOILabel* for a component. The models are evaluated using the F1-score of the defect class.
2. Prediction of human inspection. In case of an AOI defect, predict whether the defect is confirmed or rejected by the human operator, that means predict the binary *OperatorLabel*. For this task, the SPI data and the assigned *AOILabels* can be used as input data. The F1-score of the class of confirmed defects is used for model evaluation.
3. Prediction of Human Repair. For confirmed defects predict whether the component is false scrap or not possible to repair, that means predict the *RepairLabel*. As for the prediction of human inspection, both the SPI data and the *AOILabels* can be used as input data. The models are evaluated using the macro-averaged F1-score of the *RepairLabel*.

2. Approach

In order to solve the given task, the cross-industry standard process for data mining CRISP-DM is followed (Chapman, P. et al., 2000). After having understood the use case defined in the section above, the next important step is to examine the given data. The SPI data set contains information on every pin of the printed circuit boards. Apart from the necessary information to identify the pin, it contains geometry data of the solder paste that is placed for soldering of the pins. This includes the measured volume, area and height of the paste. For volume and area, the data contains also percentages in relation to the target values. Furthermore, it also includes information on the shape and the target sizes of the solder paste deposit, the target position and the percentage offset in x- and y- direction. At last, there is a *SPI Result* indicating several warnings if one of the geometric values of a pin exceeds or falls below certain thresholds.

All the data from the SPI can be used as features for the tasks to solve. However, not all of the features are independent of each other. For example, the volume of the solder paste is the simple product of area and height. These relations are considered later when selecting features for the machine learning applications. Additionally, we define the percentage height in relation to the target height of the solder paste deposit as a new useful feature. It is deducted from the available geometry information as shown in Eq. 1.

$$Height(\%) = \frac{Height(\mu m) Volume(\%) Area(\mu m^2)}{Volume(\mu m^3) Area(\%)} \quad (1)$$

The AOI data set consists of all defects that are found in the automatic inspection. For every defect, the affected component is indicated and the *AOILabel* with the type of defect, *OperatorLabel* and *RepairLabel* are given. Moreover, the affected pin of the component is given if available. This means that some of the defects can be assigned to a specific pin while others can only be assigned on component level. It is also possible that there is more than one defect for a pin or a component. As a consequence, one step for the data preparation must be the assignment between the SPI data that are available for each pin and the AOI data that might not be assigned to a specific pin. With the three sub-tasks being independent of each other, the way of preparing the data and modeling is different and specific for each task. In the following, the chosen approach for each of the sub-tasks will be explained.

For the purpose of validation, the available data are divided into a training and a hold-out validation set. This allows the testing of models on data not used for training and the optimization of hyperparameters. To split the data, 25% of the total PCB panels are randomly chosen and form the validation set while the other 75% form the training set. After validation, the generated models are trained again on the whole data set to use all available data for learning.

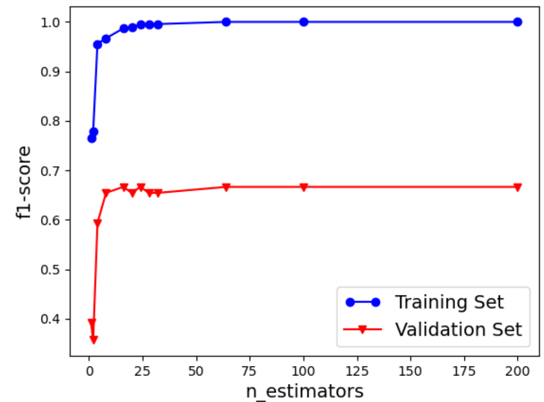

Figure 2. Optimization of the number of estimators for the prediction of human inspection.

2.1. Prediction of AOI defects

A first, simple idea for the prediction of defect components is that high deviations from the target values of the solder paste deposit lead to a defect warning in the AOI. As there is a *SPI Result* in the SPI data indicating exactly these high deviations, a first try is to classify all components with at least one *SPI Result* that is not ”GOOD” as defect. However, this rule-based approach only leads to an F1-score of approximately 7%.

As a consequence, several more complex models are investigated. Besides decision tree and random forest classifiers, a neural network is chosen instead of the simple rule-based approach and leads to best results. The model is implemented using the Scikit-Learn library (Pedregosa, F. et al., 2011). The model is trained on pin level. That means, every pin in the SPI data is labeled depending on the presence of its component in the AOI data. The algorithm then predicts for every pin whether it is faulty. A component is classified as defect when there is at least one faulty pin.

As input features, the geometrical information on the solder paste deposit are taken from the SPI data without any further preprocessing. This includes the features ”Volume(%)”, ”Height(μm)”, ”Area(%)”, ”OffsetX(%)”, ”OffsetY(%)”, ”SizeX”, ”SizeY”, ”Shape(μm)”, ”PosX(mm)” and ”PosY(mm)”. The neural network is a multi layer perceptron trained with 3 hidden layers consisting of 20, 50 and 10 neurons, which are the results of a small grid search. The ReLu activation function is used and alpha is set to 0.00001.

2.2. Prediction of human inspection

The human operator decides for every component with an AOI defect whether the component is ”Good” or ”Bad”. To predict the label given by the operator, we also build a model that works on component level. By consequence, the information on the pins in the SPI data and the AOI defects in the AOI data has to be aggregated for each component.

To use the categorical data from the *AOILabel*, one hot encoding is performed and the encoded labels are summed up for each component. Moreover, the total number of AOI entries is counted for each component and used as an additional feature. For the SPI data, the maximum and minimum values of the percentage area, height, offset in x-and y-direction and the shape are calculated for every component. By consequence, only the pins with the highest deviations from the target values are used for the prediction. The remaining information on the solder paste is not taken into account. Furthermore, the size of the solder paste deposit in x- and y-direction is considered as an additional feature, as percentage deviations might be more or less critical depending on the size of the pin.

With all these features, a random forest is learned on the training data to perform the binary classification. Again, the random forest algorithm is implemented using the Scikit-Learn library (Pedregosa, F. et al., 2011). To tackle the high class-imbalance with only about 1.5% of the components classified as ”Bad”, the class weight is set to balanced. Consequently, the ”Bad” examples are weighted much higher than the ”Good” examples. A hyperparameter optimization is performed using the validation set to find the optimum number of estimators and maximum depth of the trees. The ideal number of estimators is found to be around 16 as higher numbers of estimators do not lead to a higher performance. This is shown in Figure 2. The maximum depth of the trees is set to 10. Figure 3 shows that higher depths only lead to overfitting but do not significantly increase the performance on the hold out validation set.

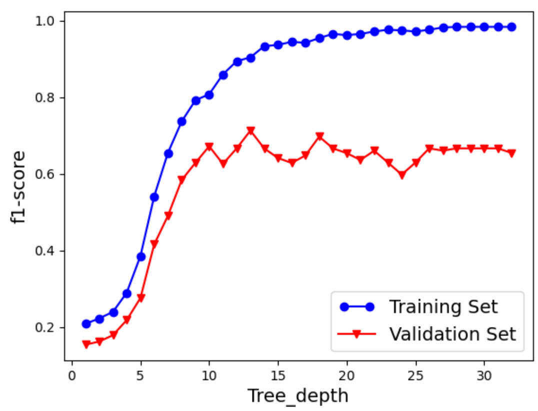

Figure 3. Optimization of the depth of the trees for the prediction of human inspection.

Table 1. Comparison of the number of AOI entries per component.

RepairLabel	Components	AOI entries	Ratio
NotPossibleToRepair	225	995	4.4
FalseScrap	122	139	1.1

2.3. Prediction of human repair

The *RepairLabel* of the second operator is only assigned to components indicated as "Bad" by the first operator. As mentioned, this concerns only about 1.5% of the components with an AOI defect. Thus, the data base used to learn the model for the prediction of human repair is rather small. The data base is even more reduced as some of the components have a *RepairLabel* set to "NotClassifiedYet" and the prediction of not classified components is not part of the task. In total, there are only 347 components left out of initially more than 27,000 components with an AOI defect.

Due to this low number of data, we tried to keep the model as simple as possible in order to reduce the risk of overfitting and guarantee the generalizability to unknown data. As shown in Table 1, components that are not possible to repair have in general much more entries in the AOI data than "FalseScrap" components. The number of AOI entries per component is on average four times higher. As a result of these considerations, we count the number of AOI entries for each component and learn a decision tree on that sole feature.

The resulting model is equivalent to a rule-based approach where all components with only a single defect entry in the AOI data are classified as "FalseScrap" and components with more than one entry in the AOI data are classified as "NotPossibleToRepair".

Table 2. Final scores of the chosen models.

Data set	Task 1 (ANN)	Task 2 (RF)	Task 3 (DT)
Training	0.39	0.81	0.85
Validation	0.39	0.66	0.87
Test	0.41	0.38	0.70

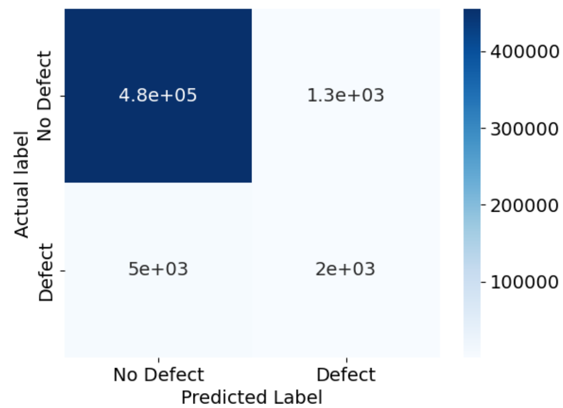

Figure 4. Confusion matrix on validation set for the prediction of AOI defects.

3. Results

The results of the chosen models are presented for each subtask separately. An overview of the scores on the different data sets is given in Table 2. The test set was not provided to the participants and only used for evaluation of the data challenge by the organizers.

3.1. Prediction of AOI defects

The neural network for predicting components with AOI defects reaches an F1-score of 0.39 on the training and validation set. A further look into the confusion matrix depicted in Figure 4 shows that the precision is at approximately 0.60 while the recall is 0.28 meaning that the model is rather weak at predicting actual components with AOI defect correctly. However, it is much better at avoiding false positives and predicting healthy components correctly.

This behaviour can be partly explained by the given class imbalance. Since there are much more components without a defect, the model is fitted to those components. Furthermore, an accurate prediction of defect components solely relying on the information on the solder paste as input might not be possible. The cause of an AOI defect can theoretically lie in a later production step like the mounting of the components for which no data are available. As shown in Table 2, the model works well on unknown data and achieves comparable scores on the test set.

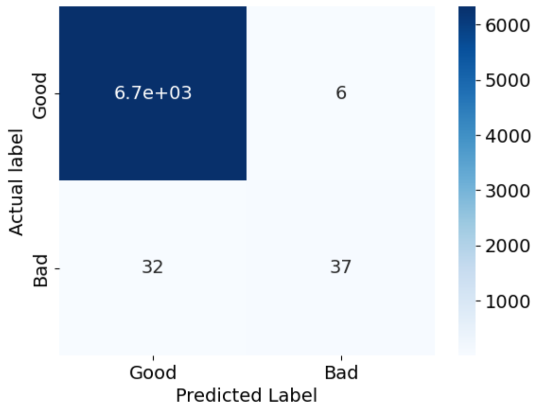

Figure 5. Confusion matrix on validation set for the prediction of human inspection.

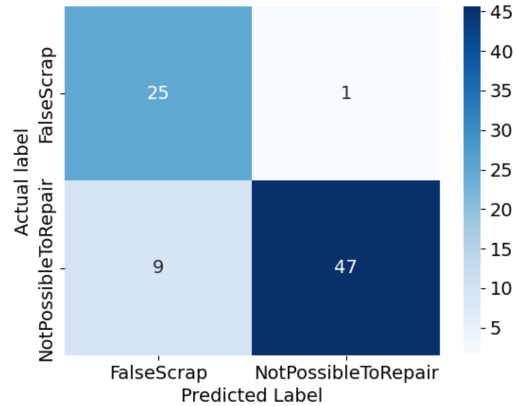

Figure 6. Confusion matrix on validation set for the prediction of human repair.

3.2. Prediction of human inspection

For the prediction of human inspection, the generated random forest classifier reaches an F1-score of 0.81 on the training data and 0.66 on the validation data. Similar to the prediction of AOI defects, the precision of the model is higher than the recall. For the training data, the precision is 0.88 and the recall 0.73

The confusion matrix of the prediction on the validation set is shown in Figure 5. Apparently, the model precision is high even on data not used for training while the recall drops significantly. The model has no problem in predicting good components correctly. Despite balanced weights during training, it has more difficulties in predicting bad components. One reason for that might be the relatively low total number of components with a "Bad" *OperatorLabel*.

The test score only amounts to 0.38 and is thus significantly lower than the score on the validation set. This indicates that the trained and validated model tends to overfit and does not generalize well on the unknown test data.

3.3. Prediction of human repair

The decision tree shows good results on training and validation set with a combined F1-score of about 0.85. Thus, the extremely simple model seems to be a surprisingly good predictor. The confusion matrix on the validation set in Figure 6 shows that almost all of the "FalseScrap" components are correctly predicted as "FalseScrap". For the "NotPossibleToRepair" components, there are a few more falsely predicted components. However, the correct label is assigned to the prevailing majority of components.

There have been several attempts to improve the model by adding more input features. Since these attempts did not lead to a significantly better training and validation score, it was decided to stay with the simple decision tree based on the number of entries in the AOI data.

The average F1-score on the test data set is 0.70 and therefore a bit lower than the training score. It seems that the relation between the number of AOI entries and the *RepairLabel* is less evident on the test data, but the decision tree classifier still provides satisfactory results.

4. Conclusion

The selected models for predicting the quality of manufactured electric circuit boards show overall satisfactory results and our team was able to reach 4th place at the data challenge. However, there is still room for improvement. The chosen model for the prediction of human inspection clearly shows signs of overfitting and should be adjusted to better classify the components. Results of the prediction of human repair show that in some applications simple rule-based approaches can provide very good results comparable to those of complex machine learning models. A good understanding of the data based on an explorative data analysis is key to identify fundamental relationships in the data.

References

Chang, Y.-M., Wei, C.-C., Chen, J., & Hsieh, P. (2018). Classification of solder joints via automatic mistake reduction system for improvement of aoi inspection. In *2018 13th international microsystems, packaging, assembly and circuits technology conference (impact)* (pp. 150–153).

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). Crisp-dm 1.0: Step-by-step data mining guide..

Cioffi, R., Travaglioni, M., Piscitelli, G., Petrillo, A., &

De Felice, F. (2020). Artificial intelligence and machine learning applications in smart production: Progress, trends, and directions. *Sustainability*, *12*(2). doi: 10.3390/su12020492

Huang, B., Di, Y., Jin, C., & Lee, J. (2017). Review of data-driven prognostics and health management techniques: Lessions learned from phm data challenge competitions..

Jabbar, E., Besse, P., Loubes, J.-M., Roa, N. B., Merle, C., & Dettai, R. (2019). Supervised learning approach for surface-mount device production. In G. Nicosia, P. Pardalos, G. Giuffrida, R. Umeton, & V. Sciacca (Eds.), *Machine learning, optimization, and data science* (Vol. 11331, pp. 254–263). Springer International Publishing. doi: 10.1007/978-3-030-13709-0_21

Jourdan, N., Longard, L., Biegel, T., & Metternich, J. (2021). *Machine learning for intelligent maintenance and quality control: A review of existing datasets and corresponding use cases.* Institutionelles Repositorium der Leibniz Universität Hannover. doi: 10.15488/11280

Liukkonen, M., Havia, E., & Hiltunen, Y. (2012). Computational intelligence in mass soldering of electronics – a survey. *Expert Systems with Applications*, *39*(10), 9928-9937. doi: https://doi.org/10.1016/j.eswa.2012.02.100

Mayr, A., Kißkalt, D., Meiners, M., Lutz, B., Schäfer, F., Seidel, R., … Franke, J. (2019). Machine learning in production – potentials, challenges and exemplary applications. *Procedia CIRP*, *86*, 49–54. doi: 10.1016/j.procir.2020.01.035

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, E. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*(85), 2825–2830.

PHM Society. (2022). *Data challenge: 7th european conference of the prognostics and health management society 2022.* Retrieved 18.06.2022, from `https://phm-europe.org/data-challenge`

Wang, J., Ma, Y., Zhang, L., Gao, R. X., & Wu, D. (2018). Deep learning for smart manufacturing: Methods and applications. *Journal of Manufacturing Systems*, *48*, 144-156. (Special Issue on Smart Manufacturing) doi: https://doi.org/10.1016/j.jmsy.2018.01.003

Zakaria, S., Amir, A., Yaakob, N., & Nazemi, S. (2020). Automated detection of printed circuit boards (pcb) defects by using machine learning in electronic manufacturing: Current approaches. In *Iop conference series: Materials science and engineering* (Vol. 767, p. 012064).