Proceedings of the 14th Asian Control Conference
July 5-8, 2024, Dalian, China

# A Reliable Ensemble Model based on Hierarchical Component Features for Repair Label Prediction of Soldering Defects

1st Longxin Chen
*Department of Automation, University of Science and Technology of China*
Hefei, China
lxchen@mail.ustc.edu.cn

2nd Yunbo Zhao
*Department of Automation and Institute of Advanced Technology, University of Science and Technology of China*
Hefei, China
*Institute of Artificial Intelligence, Hefei Comprehensive National Science Center*
Hefei, China
ybzhao@ustc.edu.cn

3rd Binkun Liu
*Department of Automation, University of Science and Technology of China*
Hefei, China
liubink@mail.ustc.edu.cn

4th Shaojie Dong
*Institute of Advanced Technology, University of Science and Technology of China*
Hefei, China
shaojiedong@mail.ustc.edu.cn

5th Huijuan Zhu
*AHU-IAI AI Joint Laboratory Anhui University*
Hefei, China
wa21301046@stu.ahu.edu.cn

6th Peng Bai
*Department of Automation, University of Science and Technology of China*
Hefei, China
baipeng@lenovo.com

***Abstract*—Using solder paste inspection (SPI) and automated optical inspection (AOI) data, accurate prediction for repair labels of soldering defective printed circuit board (PCB) components can help reduce labor costs. Existing research tries to pick out both the false defect components (actually good) and impossible-to-repair components among defective PCB components, using SPI and AOI data. However, it is inappropriate to pick out the false defect components from screened components using defective information in AOI data. Therefore, the problem setting of existing research is inappropriate, resulting in the algorithm's performance not meeting actual requirements. To address this problem, we only care about the reliable prediction of impossible-to-repair components. We propose a hierarchical component feature extraction method that can comprehensively characterize the degree of component defects from multiple levels, including pin level and component level. Then we apply the ensemble model based on XGBoost and TabNet and adjust the probability threshold of components judged as impossible-to-repair category, achieving the reliable prediction of impossible-to-repair components. Finally, we validated our method on real datasets and achieved better experimental results compared to baseline methods, which can meet actual requirements.**

***Index Terms*—PCB components, Repair label prediction, Hierarchical feature extraction, Ensemble model**

## I. Introduction

In surface mount technology (SMT) process, accurate repair label prediction of defective printed circuit board (PCB) components can reduce labor costs, and the European Prognostics and Health Management Society (PHME) has held a data challenge for it [1]. SMT is a technique used to solder pads of PCB and pins of surface mounted components (SMCs) [2] [3]. During the SMT process, some PCBs are judged to have component soldering defects by automated optical inspection (AOI) machine and an operator [3] [4]. Before considering any repair work, an experienced operator examines these PCB components in more detail and assigns different repair labels depending on whether the component needs to be repaired [5]–[8]. The repair labels classify defective components into three categories: impossible-to-repair, false defect (actually no defects), and others, thus preventing impossible-to-repair and false defect components from entering repair process and saving time. If the repair labels of defective components can be accurately predicted, labour costs will be reduced.

Existing research tries to pick out both the impossible-to-repair and false defect components among defective PCB components, by applying tree model algorithm to predict the repair labels, based on solder paste inspection (SPI) data and AOI data [5]–[7]. To determine whether a defective component can be repaired, the operator needs to evaluate the defects from different perspectives. For example, when there is a significant quantity or severity of defects, the PCB component is often impossible to be repaired or not worth being repaired. AOI data describes the soldering defects of each defective pin (such as misalignment, missing components or solder paste, etc.). And SPI data describes the characteristics of the solder pasted on each component pin (e.g., volume, height, area, offset, etc.). Statistical features such as the standard deviation

This work was supported by the National Natural Science Foundation of China (No. 62173317), and the Key Research and Development Program of Anhui (No. 202104a05020064). Corresponding author: Yunbo Zhao (e-mail: ybzhao@ustc.edu.cn).



Authorized licensed use limited to: University of Science & Technology of China. Downloaded on October 08,2024 at 10:42:37 UTC from IEEE Xplore. Restrictions apply.

of solder paste height on each component can be calculated with SPI data, which can be used to help assess the impact of solder paste quality on component soldering. Therefore, it is possible to achieve the prediction of repair label by using AOI data, supplemented by SPI data. Existing studies use feature extraction methods and only extract features from a single level (component level [1] [6] or pin level [7]). Existing research predicts the repair labels of defective components, and choose the $F1$ score as the final evaluation metric [5]–[7].

However, the inappropriate problem setting in existing research leads to algorithm's performance not meeting actual requirements, because it is inappropriate to pick out the false defect PCB components from the screened defective components using defective information in AOI data. Before determining the repair labels, components identified as defective by the AOI machine have been visually inspected by an operator, who then picks out false defect components from them. Meanwhile, existing research indicates that AOI data is more important in this prediction task than SPI data [5]. But AOI data only contains the defective information of components. So it is inappropriate to pick out the false defect components among the defective components that have already been screened, using the defective information of AOI data. Whereas existing research attempts to pick out false defect components from defective components, so the algorithm's performance cannot meet actual requirements and would increase costs by judging impossible-to-repair components as false defect components falsely and allowing them to enter the subsequent process.

To address the above problem, we are only concerned with the reliable prediction of impossible-to-repair components. We propose a reliable ensemble model based on hierarchical component features. Firstly, we propose a hierarchical component feature extraction method that can comprehensively characterize the degree of component defects from pin level and component level. We extract component-level features by calculating the statistical values of the raw features, and extract the raw features of the first two pins as pin-level features. Then we apply the ensemble model based on XGBoost and TabNet and adjust the probability threshold of components judged as impossible-to-repair category, achieving the reliable prediction of impossible-to-repair components.

Finally, We still choose the $F1$ score of impossible-to-repair components as metrics, and set the $F1$ score under the precision of 85%, 90%, 95% and 100% as metrics, considering the accuracy requirements in actual production. We also use $WorkReduce$ under different precision as metrics, which can reflect the operator's workload reduced by algorithm.

The contributions of this paper are as follows:

1. A hierarchical component feature extraction method, which can comprehensively characterize the degree of component defects from multiple levels, including pin level and component level, is proposed and proven to be effective.

2. An ensemble model based on XGBoost and TabNet is applied for reliable prediction of impossible-to-repair components, and the best results are achieved. And the model's performance can meet actual requirements well.

## II. Related Works

SMT production technology consists of three consecutive steps [3] [4] [9]. The first step is solder paste printing: a solder paste printer is used to leakage print the solder paste onto the pads of the PCB. The second step is component placement: SMCs are mounted to fixed positions on the PCB using a Surface Mounting Device (SMD) machine. The third step is solder paste reflow: use a reflow oven to melt the solder paste so that the pins of the SMC can be firmly affixed to the pads of the PCB. In addition, in order to ensure the quality of surface mounting, the SMT production line is usually equipped with a SPI machine and an AOI machine. The SPI machine is placed after the solder paste printer to check the quality of solder paste printing, while AOI machine is placed after reflow oven to check surface defects of the soldered components on PCB.

SPI and AOI data have been used by researchers for quality assessment. Liu et al. [10] achieved solder paste position offset prediction using SPI data. For the misjudgement problem of AOI machine, Chang et al. [3] applied a deep learning algorithm to predict the quality of welded joints effectively using SPI data. Park et al. [11] proposed a two-layer defect detection method for early defects detection, using SPI data. For the repair label prediction of defective components, Immo Schmidt et al. [5] proposed a decision tree model based on single component-level feature. John Taco et al. [1] and Alexandre Gaffet et al. [6] took different component-level feature extraction approaches, and used XGboost algorithm to predict repair labels. Tang et al. [7] used Random Forest algorithm to predict repair labels based on pin-level features.

## III. METHOD

The overall framework of the method is shown in Fig. 1. Firstly, SPI data and AOI data are collected. Nextly, build the feature engineering on SPI data and AOI data respectively, and then SPI features and AOI features are integrated to get the final features. Finally, XGBoost and TabNet models are constructed separately and the weighted average of their outputs is used to predict the repair labels.

### A. Problem Setting

Given SPI and AOI data of each defective component, we firstly extract features from them, which can be described as follows:

$$X_1^{'}, X_2^{'}, \cdots, X_M^{'} = g(X_1, \cdots, X_T, X_{T+1}, \cdots, X_N) \quad (1)$$

where $X_1, \cdots, X_T$ are raw SPI features of a component, $T$ is the number of SPI features for the component; $X_{T+1}, \cdots, X_N$ are raw AOI features of the same component, $N-T$ is the number of AOI features for the component; $X_1^{'}, X_2^{'}, \cdots, X_M^{'}$ are features extracted from raw data, $M$ is the number of extracted features for the component; $g(\cdot)$ is mapping function constructed by feature extraction method.

Then we construct model to predict the probability that the defective component is of the impossible-to-repair category,


Authorized licensed use limited to: University of Science & Technology of China. Downloaded on October 08,2024 at 10:42:37 UTC from IEEE Xplore. Restrictions apply.




Fig. 1. Overall framework of the methodology

based on features extracted from SPI and AOI data. Our goal is described as follows:

$$P = f(X_1^{'}, X_2^{'}, \cdots, X_M^{'}) \tag{2}$$

where $P$ is the probability that the component is predicted to belong to impossible-to-repair category, $f(\cdot)$ is mapping function constructed by the ensemble model.

When probability exceeds the set threshold, the component is predicted as an impossible-to-repair component. We achieve reliable and accurate prediction by increasing this threshold.

### B. Data Description

PCB panelization technology is usually used in the SMT production line, that is, a PCB panel consists of several separate boards. Fig 2. shows an example of a panel with 4 figures [12], and each figure is a single PCB. All PCBs on the panel have the same characteristics, with the same components to be placed. Each panel is identified by a unique *PanelID*. The PCB is identified by a unique combination of *PanelID* and *FigureID*. Each component of the PCB is identified by a *ComponentID*, and each pin of the component is identified by a *PinNumber*.

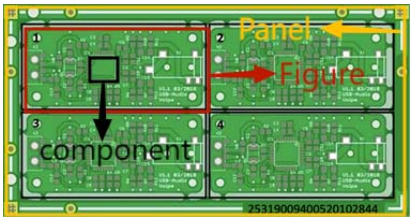


Fig. 2. An example of a panel with 4 figures [12]

SPI Data Description. The SPI data describes the characteristics of the solder paste on each pin. The SPI data contains the characteristics shown in Table I.

AOI Data Description. The AOI data contains defect information only for pins that the AOI machine detects as defective. The AOI data contains characteristics as shown in Table II.

TABLE I
CHARACTERISTICS OF SPI DATA

| Characteristics | Description |
|---|---|
| PanelID | the panel identifier |
| FigureID | the figure identifier |
| ComponentID | the component identifier |
| PinNumber | the number of the pin considered for the component |
| PadID | the unique identified of the Pad that supports the pins |
| Date | the date when the SPI operation was performed |
| Time | the time ins seconds when the SPI operation was performed |
| PosX(mm) | the X coordinate of the pin considering starting from the bottom left cornet |
| PosY(mm) | the Y coordinate of the pin considering starting from the bottom left cornet |
| PadType | the type of Pad |
| Features regarding welding paste characteristics | Volume(%), Height(um), Area(%), OffsetX(%), OffsetY(%), SizeX, SizeY, Volume(um3), Area(um2), Shape(um) |
| Result | the result of the SPI inspection |

TABLE II
CHARACTERISTICS OF AOI DATA

| Characteristics | Description |
|---|---|
| PanelID | the panel identifier |
| FigureID | the figure identifier |
| ComponentID | the component identifier |
| PinNumber | the number of the pin considered for the component |
| MachineID | the identifier of the machine performing the AOI operation |
| AOILabel | the label applied by the AOI machine, i.e., the type of defect |
| OperatorLabel | the label applied by the human operator after visual inspection.<br>Value: Good–the AOI raised a false defect, Bad–the operator confirms the AOI message |

### C. Feature Engineering

Firstly, the three identifiers (*PanelID*, *FigureID*, *ComponentID*) are indexed in SPI data and AOI data respectively to find the data entries of the same component with the same (*PanelID*, *FigureID* and *ComponentID*).

Secondly, construct the component-level statistical features


Authorized licensed use limited to: University of Science & Technology of China. Downloaded on October 08,2024 at 10:42:37 UTC from IEEE Xplore. Restrictions apply.

on SPI data and AOI data respectively. For the same component, we calculate the mean, variance, standard deviation, minimum, maximum, and median of different SPI pin features as SPI component-level features, and calculate the median or count of different AOI pin features as AOI component-level features.

Thirdly, considering that most of the components have two pins [7], we extract the original SPI and AOI features of pin 1 and pin 2 as pin-level features. For the components with only 1 pin, the features of pin 2 are set to 0.

The features extracted from SPI data are shown in Table III, with a total of 163 SPI features. In Table III, SPI pin features *1-12* are original features in SPI data, and SPI pin features *13-23* are new features computed from SPI pin features *1-12*. For extracted features, *pin1* and *pin2* are pin-level features, while the others are component-level features.

The features extracted from AOI data are shown in Table IV, with a total of 35 AOI features. *ComponentType_** denotes the type of the component, and *pin1_** and *pin2_** denote the defective condition of pin1 and pin2, respectively. For extracted features, *pin1_** and *pin2_** are pin-level features, while the others are component-level features.

Finally, SPI features and AOI features with the same (*PanelID*, *FigureID*, *ComponentID*) are combined as final features for the component, which includes component-level and pin-level features.

The feature extraction method here can comprehensively characterize the degree of component defects from multiple levels, including pin level and component level, which addresses the shortcomings of existing methods. Some component-level statistical features are referenced in [13].

### D. Model Construction

This paper constructs XGBoost and TabNet models separately and finally uses the ensemble of XGBoost and Tabnet models for prediction. In this problem, there is no temporal and spatial relationship between the final extracted features, which is a typical tabular data. For classification or regression problems of tabular data, tree integration algorithms (e.g., XGBoost) are the preferred algorithms. Considering that the XGBoost algorithm is currently the best performing model on repair label prediction of defective components, we choose the XGBoost algorithm. In addition, it has been pointed out that the ensemble of TabNet and XGBoost performs better than XGBoost alone on some datasets [14] [15], so we finally choose the ensemble of XGBoost and TabNet models for repair label prediction. TabNet is a deep learning model designed by Google Cloud AI for tabular data [16] [17].

The outputs of both the XGBoost and the TabNet are the probability that the component belongs to the impossible-to-repair category. The model solve the following optimization problems:

$$min[-plog\hat{p} - (1-p)log(1-\hat{p})] \tag{3}$$

where $p \in \{0,1\}$, $\hat{p} \in [0,1]$. $p$ is the probability that the defective component belongs to the impossible-to-repair category in fact: if defective component is impossible-to-repair, $p = 1$, otherwise $p = 0$. $\hat{p}$ is the probability output of model.

After constructing the XGBoost model and TabNet model respectively, we use the weighted average of the two models's outputs as the final prediction result:

$$p(y|x) = \omega_{XGBoost} p_{XGBoost}(y|x, \theta_{XGBoost}) + \omega_{TabNet} p_{TabNet}(y|x, \theta_{TabNet}) \tag{4}$$

where $\omega_{XGBoost} + \omega_{TabNet} = 1,\ \omega_{XGBoost} \geq 0,\ \omega_{TabNet} \geq 0$. The values of $\omega_{XGBoost}$ and $\omega_{TabNet}$ are ultimately determined by the actual results of each model.

## IV. EXPERIMENT

### A. Dataset

The dataset is obtained from PHM Europe 2022 data challenge [12]. The dataset is obtained from a real industrial production line. In this production line, PCBs detected with defects by the AOI machine are visually inspected by operator 1. If operator 1 confirms the presence of defects, then operator 2 evaluates the repair plan. There are three types of repair labels assigned by operator 2: NotPossibleToRepair(impossible to repair), FalseScrap(the component does not need any repair operation) and NotYetClassified(no information about the repair is available).

After data processing, the SPI and AOI data of 412 components in the training set are finally left for the task of this paper, i.e., the number of components whose *OperatorLabel* are 'Bad' is 412. And the testing set has 265 components left for the task of this paper. In the training set, 122 components belong to FalseScrap, 225 components belong to NotPossibleToRepair, and 65 components belong to NotYetClassified. And in the test set, 58 components belong to FalseScrap, 182 components belong to NotPossibleToRepair and 25 components belong to NotYetClassified.

In the training set, the ratio of the number of NotPossibleToRepair components to the number of components in the other categories is approximately 1.2:1. This is not a significant difference in number, so the category imbalance is not considered.

### B. Experimental Settings

*1) Metrics:* We use $F1$ score of impossible-to-repair components and $WorkReduce$ score as metrics. The $F1$ score is a trade-off between recall and precision, which can be calculated as follow:

$$F1 = 2\frac{precision \cdot recall}{precision + recall} \tag{5}$$

$WorkReduce$ can reflect the operators' workload reduced by our method, which can be calculated as follow:

$$WorkReduce = \frac{num_{impossible-to-repair}}{num_{all}} \cdot 100\% \tag{6}$$

where $num_{impossible-to-repair}$ is the number of components predicted as impossible-to-repair, $num_{all}$ is the number of all defective components.


Authorized licensed use limited to: University of Science & Technology of China. Downloaded on October 08,2024 at 10:42:37 UTC from IEEE Xplore. Restrictions apply.

TABLE III
FEATURES EXTRACTED FROM SPI DATA

| SPI Pin Features | | | Extracted Features (Including pin-level and component-level features) | |
|---|---|---|---|---|
| No. | Variable Name | operation required | No. | Feature Name |
| 1 | Volume(%) | | 01-08 | mean, std, var, min, max, median, pin1, pin2 |
| 2 | Height(um) | | 09-16 | mean, std, var, min, max, median, pin1, pin2 |
| 3 | Area(%) | | 17-24 | mean, std, var, min, max, median, pin1, pin2 |
| 4 | OffsetX(%) | | 25-32 | mean, std, var, min, max, median, pin1, pin2 |
| 5 | OffsetY(%) | | 33-40 | mean, std, var, min, max, median, pin1, pin2 |
| 6 | SizeX | | 41-48 | mean, std, var, min, max, median, pin1, pin2 |
| 7 | SizeY | | 49-56 | mean, std, var, min, max, median, pin1, pin2 |
| 8 | Volume(um3) | | 57-64 | mean, std, var, min, max, median, pin1, pin2 |
| 9 | Area(um2) | | 65-72 | mean, std, var, min, max, median, pin1, pin2 |
| 10 | Shape(um) | | 73-80 | mean, std, var, min, max, median, pin1, pin2 |
| 11 | PosX(mm) | | 81-88 | mean, std, var, min, max, median, pin1, pin2 |
| 12 | PosY(mm) | | 89-96 | mean, std, var, min, max, median, pin1, pin2 |
| 13 | OffsetX and Y | $= \sqrt{([4]^2 + [5]^2)}$ | 97-102 | mean, std, var, min, max, median |
| 14 | SizeX and Y | $= \sqrt{([6]^2 + [7]^2)}$ | 103-108 | mean, std, var, min, max, median |
| 15 | Volume/Area | $= [8]/[9]$ | 109-114 | mean, std, var, min, max, median |
| 16 | Volume/Volume | $= [9]/[1]$ | 115-120 | mean, std, var, min, max, median |
| 17 | PosX anf Y | $= \sqrt{([11]^2 + [12]^2)}$ | 121-126 | mean, std, var, min, max, median |
| 18 | OffsetX/OffsetY | $= [4]/[5]$ | 127-132 | mean, std, var, min, max, median |
| 19 | OffsetX/SizeX | $= [4]/[6]$ | 133-138 | mean, std, var, min, max, median |
| 20 | OffsetY/SizeY | $= [5]/[7]$ | 139-144 | mean, std, var, min, max, median |
| 21 | Volume/OffsetX | $= [1]/[4]$ | 145-150 | mean, std, var, min, max, median |
| 22 | SizeY/PosY | $= [7]/[12]$ | 151-156 | mean, std, var, min, max, median |
| 23 | PosX/PosY | $= [11]/[12]$ | 157-162 | mean, std, var, min, max, median |
| | | | 163 | pin_count(number of pins) |

TABLE IV
FEATURES EXTRACTED FROM AOI DATA

| AOI Pin Features | | | Extracted Features (Including pin-level and component-level features) | |
|---|---|---|---|---|
| No. | Variable Name | operation required | No. | Feature Name |
| 1 | PinNumber | | 01 | median |
| 2 | AOILabel | | 02 | count |
| 3 | PadType_* | acquire from SPI data | 03-04 | count{* = 0, 1} |
| 4 | PinNumber_* | acquire from PinNumber | 05-07 | count{* = 1, 2, others} |
| 5 | AOILabel_* | acquire from AOILabel | 08-13 | count{* = Translated, Coplanarity, Size, Soldered, UnSoldered, Others} |
| | | | 14-25 | ComponentType_*{* = C, L, R, D, U, TR, BC, RA, IC, CN, DZ, TRB} (value: 0 or 1, acquire from ComponentID) |
| | | | 26-30 | pin1_* {* = Translated, Coplanarity, Soldered, UnSoldered, Others} (value: 0 or 1, acquire from AOILabel when PinNumber = 1) |
| | | | 31-35 | pin2_* {* = Translated, Coplanarity, Soldered, UnSoldered, Others} (value: 0 or 1, acquire from AOILabel when PinNumber = 2) |

We firstly concerns the precision of the NotPossibleToRepair components, which can ensure that the prediction precision meets actual requirements. A precision of 100% means no errors in prediction, which is the most conservative strategy. However, in practice, there may be different requirements for the precision of the NotPossibleToRepair components prediction, considering the cost trade-off. So we set the $F1$ score and $WorkReduce$ score under the precision of 85%, 90%, 95% and 100% as metrics. After completing the model construction, we iterate through probability thresholds with a step size of 0.01, starting from 0.51 until we reach the highest probability value of the model output, to find the best score at the corresponding accuracy.

*2) Comparative Experiments:* To prove the superiority of the proposed method, we use the method of [1], [6] and [7] for the task of this paper as baseline methods respectively. We conduct the following experiments for comparison:

*Ensemble Model+Hierarchical Features*: Method in this paper. The parameters of XGBoost are referenced as follow: the learning rate is set to 0.1, the maximum number of trees generated is set to 100, the depth of the tree is set to 3, the loss function is chosen to be "binary:logistic", and the number of training rounds for the model is set to 66. The parameters of TabNet are referenced as follow: the learning rate of the TabNet model is set to 0.001, batch_size is set to 32, the maximum number of iterations is set to 500, the optimiser is chosen to be Adam, and momentum is set to 0.01. For models obtained from a total of 500 iteration process, the model with the highest precision is saved.

*XGBoost+Features_1*: Apply the component-level statistical features *Features_1* extracted from [1] and XGBoost model from [1] for the task of this paper.

*RF+Features_2*: Apply the pin-level features *Features_2* extracted from [7] and random forests model from [7] for the task of this paper.

*XGBoost+Features_3*: Apply the component-level statistical features *Features_3* extracted from [6] and XGBoost model from [6] for the task of this paper.

*3) Ablation Experiments:* In order to prove the effectiveness of the feature extraction method and the ensemble model used in this paper, we conduct the following ablation experiments respectively:

*Ablation experiment on feature extraction method*: This experiment is conducted to prove effectiveness of the feature extraction method. We make prediction using only component-level statistical features and pin-level features separately, and compare their results with the result of hierarchical features that include both component-level statistical features and pin-level features. Ensemble model is applied here.

*Ablation experiment on ensemble model*: This experiment is conducted to prove effectiveness of the ensemble model. We make prediction using only XGBoost model and TabNet model separately, and compare their results with the result of the ensemble model. Hierarchical features is used here.


Authorized licensed use limited to: University of Science & Technology of China. Downloaded on October 08,2024 at 10:42:37 UTC from IEEE Xplore. Restrictions apply.

### C. Experimental Results

All the experimental results are shown in Table V, Table VI and Table VII. It should be noted the ratio of $WorkReduce$ score of XGBoost and TabNet under the 85% precision is about 2:1, so we take $\omega_{XGBoost} = 0.66$ and $\omega_{TabNet} = 0.34$, so that the ratio of $\omega_{XGBoost}$ and $\omega_{TabNet}$ is set to 2:1.

*1) Comparative Experiment:* The results are shown in Table V. For the prediction of NotPossibleToRepair components, compared to baseline method, the proposed method in this paper performs optimally under the requirements of different precision. We can also see that the superiority of our method is more obvious under the precision of 95% and 100%. According to the $WorkReduce$ score under the precision of 100%, our method can reduce the operators' workload by 16.60% without any errors. So the experiment results prove the superiority of the proposed method in this paper.

TABLE V
THE RESULT OF COMPARATIVE EXPERIMENT

| metrics | method | precision | | | |
|---|---|---|---|---|---|
| | | 85% | 90% | 95% | 100% |
| F1 | Ensemble Model+Hierarchical Features | **0.81** | **0.75** | **0.61** | **0.39** |
| | XGBoost+Features_1 | 0.78 | 0.71 | 0.27 | 0 |
| | RF+Features_2 | 0.56 | 0.52 | 0.39 | 0.23 |
| | XGBoost+Features_3 | 0.50 | 0.38 | 0.38 | 0 |
| WorkReduce | Ensemble Model+Hierarchical Features | **61.89%** | **49.43%** | **32.45%** | **16.60%** |
| | XGBoost+Features_1 | 57.74% | 42.64% | 10.94% | 0 |
| | RF+Features_2 | 33.21% | 27.55% | 17.36% | 9.06% |
| | XGBoost+Features_3 | 27.17% | 17.74% | 16.98% | 0 |

*2) Ablation experiment on feature extraction method:* The results are shown in Table VI. The hierarchical features that include both component-level statistical features and pin-level features achieve the best score. So the hierarchical feature extraction method proposed in this paper is effective.

TABLE VI
THE RESULT OF ABLATION EXPERIMENT ON FEATURE EXTRACTION METHOD

| metrics | method | precision | | | |
|---|---|---|---|---|---|
| | | 85% | 90% | 95% | 100% |
| F1 | hierarchical features | **0.81** | **0.75** | **0.61** | **0.39** |
| | only component-level features | 0.76 | 0.72 | 0.50 | 0.15 |
| | only pin-level features | 0.67 | 0.46 | 0.27 | 0.20 |
| WorkReduce | hierarchical features | **61.89%** | **49.43%** | **32.45%** | **16.60%** |
| | only component-level features | 52.83% | 45.28% | 24.53% | 5.66% |
| | only pin-level features | 46.79% | 23.40% | 11.32% | 7.55% |

*3) Ablation experiment on ensemble model:* The results are shown in Table VII. It can be observed through Table VII that the TabNet model can not reach the precision of 95% and 100%. The score of the XGBoost model is better than the TabNet model. And the ensemble model is better than the XGBoost model. This is because XGBoost model and TabNet model learn the features from different perspectives and form complementarity. So the ensemble model is effective.

TABLE VII
THE RESULT OF ABLATION EXPERIMENT ON ENSEMBLE MODEL

| metrics | method | precision | | | |
|---|---|---|---|---|---|
| | | 85% | 90% | 95% | 100% |
| F1 | Ensemble model | **0.81** | **0.75** | **0.61** | **0.39** |
| | XGBoost model | 0.76 | 0.72 | 0.50 | 0.29 |
| | TabNet model | 0.48 | 0.39 | 0 | 0 |
| WorkReduce | Ensemble model | **61.89%** | **49.43%** | **32.45%** | **16.60%** |
| | XGBoost model | 53.21% | 46.04% | 24.15% | 11.70% |
| | TabNet model | 26.79% | 19.25% | 0 | 0 |

## CONCLUSION

In this paper, we propose a hierarchical component feature extraction method that can comprehensively characterize the degree of component defects from multiple levels, including pin level and component level. Then we apply the ensemble model based on XGBoost and TabNet and adjust the probability threshold of components judged as impossible-to-repair category, achieving the reliable prediction of impossible-to-repair components. By conducting comparative and ablation experiments, we prove the superiority and effectiveness of our method, whose performance can meet actual requirements.

## REFERENCES


[1] J. Taco, P. Gore, T. Minami, P. Kundu, A. Suer, and J. Lee, "A novel methodology for health assessment in printed circuit boards," in *PHM Society European Conference*, 2022, pp. 556–562.

[2] H. Shamkhalichenar, C. J. Bueche, and J.-W. Choi, "Printed circuit board (pcb) technology for electrochemical sensors and sensing platforms," *Biosensors*, vol. 10, no. 11, pp. 159–174, 2020.

[3] Y.-M. Chang, C.-C. Wei, J. Chen, and P. Hsieh, "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)*, 2018, pp. 150–153.

[4] Y.-M. Chang, J. Chen, and P. Hsieh, "Prediction of solder joint quality using a data mining methodology for surface mounted technology process," in *Proceedings of the 1st International Conference on Big Data Research*, 2017, pp. 52–56.

[5] I. Schmidt, L. Dingeldein, D. Hünemohr, H. Simon, and M. Weigert, "Application of machine learning methods to predict the quality of electric circuit boards of a production line," in *PHM Society European Conference*, 2022, pp. 550–555.

[6] A. Gaffet, N. B. Roa, P. Ribot, E. Chanthery, and C. Merle, "A hierarchical xgboost early detection method for quality and productivity improvement of electronics manufacturing systems," in *PHM Society European Conference*, 2022, pp. 541–549.

[7] H. Tang, Y. Tian, J. Dai, Y. Wang, J. Cong, Q. Liu, X. Zhao, and Y. Fu, "Prediction of production line status for printed circuit boards," in *PHM Society European Conference*, 2022, pp. 563–570.

[8] B. Liu, Y. Zhao, Y. Cao, Y. Kang, and Z. Xu, "A robustness benchmark for prognostics and health management," in *2022 41st Chinese Control Conference (CCC)*, 2022, pp. 7338–7342.

[9] S. Härter, T. Klinger, J. Franke, and D. Beer, "Comprehensive correlation of inline inspection data for the evaluation of defects in heterogeneous electronic assemblies," in *2016 Pan Pacific Microelectronics Symposium (Pan Pacific)*, 2016, pp. 1–6.

[10] B. Liu, Y. Zhao, Y. Kang, Y. Cao, P. Bai, and Z. Xu, "A feature engineering-based method for pcb solder paste position offset prediction," in *2023 6th International Symposium on Autonomous Systems (ISAS)*, 2023, pp. 1–6.

[11] J.-M. Park, Y.-H. Yoo, U.-H. Kim, D. Lee, and J.-H. Kim, "D 3 pointnet: Dual-level defect detection pointnet for solder paste printer in surface mount technology," *IEEE Access*, vol. 8, pp. 140 310–140 322, 2020.

[12] D. Giordano and M. Trevisan, "Phme data challenge." European conference of the prognostics and health management society, 2022.

[13] Y.-M. Chang, C.-C. Wei, J. Chen, and P. Hsieh, "An implementation of health prediction in smt solder joint via machine learning," in *2019 ieee international conference on big data and smart computing (bigcomp)*, 2019, pp. 1–4.

[14] R. Shwartz-Ziv and A. Armon, "Tabular data: Deep learning is not all you need," *Information Fusion*, vol. 81, pp. 84–90, 2022.

[15] A. Lewandowska, "Xgboost meets tabnet in predicting the costs of forwarding contracts," in *2022 17th Conference on Computer Science and Intelligence Systems (FedCSIS)*, 2022, pp. 417–420.

[16] S. Ö. Arik and T. Pfister, "Tabnet: Attentive interpretable tabular learning," in *Proceedings of the AAAI conference on artificial intelligence*, 2021, pp. 6679–6687.

[17] C. Shah, Q. Du, and Y. Xu, "Enhanced tabnet: Attentive interpretable tabular learning for hyperspectral image classification," *Remote Sensing*, vol. 14, no. 3, pp. 716–736, 2022.




Authorized licensed use limited to: University of Science & Technology of China. Downloaded on October 08,2024 at 10:42:37 UTC from IEEE Xplore. Restrictions apply.