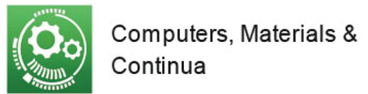

Computers, Materials & Continua

Tech Science Press

https://doi.org/10.32604/cmc.2026.078762

ARTICLE

Integrating FDC and Machine Learning for Enhanced Anomaly Detection in WB Bonding Joint Quality

Chin Ta Wu[1,2], Shing Han Li[3,*] and Ching Shih Tsou[4]

[1]College of Business, National Taipei University of Business, Taipei, Taiwan
[2]Powertech Technology Inc., Hsinchu, Taiwan
[3]Department of Accounting Information, National Taipei University of Business, Taipei, Taiwan
[4]Institute of Information and Decision Sciences, National Taipei University of Business, Taipei, Taiwan

*Corresponding Author: Shing Han Li. Email: shli@ntub.edu.tw

Received: 07 January 2026; Accepted: 14 April 2026; Published: 08 May 2026

ABSTRACT: In semiconductor packaging processes, the wire bonding procedure, which connects chips to substrate lead frames using metal wires, is a crucial step. The quality of the bonding joints significantly affects product performance, including signal integrity and reliability, and is challenging to verify after subsequent processes. To mitigate the risk of defective bonding joints entering the assembly packaging stages of production, this study integrates the concepts of Fault Detection and Classification (FDC) and machine learning into the wire bonding process for enhanced anomaly detection. Production data from the machines were collected and analyzed using statistical methods to filter out normal bonding joint data. After conducting feature engineering, we developed an anomaly detection model specifically for bonding joints. This inference model was subsequently deployed and validated using actual production data. During the validation phase, the proposed anomaly detection system effectively assisted the production line in identifying ball-related anomalies, thereby preventing these defects from advancing to later stages and ensuring overall product quality.

KEYWORDS: Wire bonding; anomaly detection; isolation forest; fault detection and classification

1 Introduction

The semiconductor packaging industry faces unprecedented quality management challenges driven by the rapid growth of automotive electronics and high-performance computing (HPC). In these sectors, the "zero defects" philosophy is the standard, as even minor bonding failures can lead to catastrophic system failures or expensive recalls [1,2].

Concurrently, the application of artificial intelligence (AI) is experiencing swift expansion, particularly with the increasing demand for GPU servers and data centers. The high manufacturing and operational costs associated with these systems [3] underscore the necessity for suppliers to deliver components with near-zero defects to guarantee system stability and performance. In this context, semiconductor wafer packaging serves as a crucial link in the manufacturing process chain, presenting significant challenges in precision control and quality management.

In order to avoid the occurrence of abnormal products, this study proposes an unsupervised anomaly detection framework that integrates FDC data with an ensemble machine learning approach. This paradigm is specifically chosen to address the extreme class imbalance in high-yield semiconductor manufacturing,

Copyright © 2026 The Authors. Published by Tech Science Press. This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

where defect labels are too scarce for effective supervised training. By employing a "three-in-one" ensemble of Isolation Forest, Connectivity-based Outlier Factor (COF), and Multidimensional Scaling (MDS), the system establishes a robust quality barrier to filter risks during the early stages of assembly.

1.1 Research Background

In the semiconductor supply chain, rigorous monitoring of semi-finished products during the manufacturing process is particularly important, as it can identify potential defects early, thereby preventing defective products from progressing through subsequent stages and resulting in greater losses. The preventive quality management strategy not only reduces production costs but also ensures the final product's quality and reliability, aligning with the strict zero-defect requirements of automotive electronics and AI applications.

In recent years, integrated circuit (IC) packaging companies have actively developed smart manufacturing technologies, integrating data and information technology across manufacturing processes, including the Internet of Things (IoT), big data analytics, and AI. Through these data-driven approaches, companies can promptly monitor the status of manufacturing processes, thus maintaining product quality. Key issues related to production equipment include FDC, preventive maintenance, virtual measurement, and anomaly detection. Among these, anomaly detection in process is a crucial research focus. By integrating data from production equipment with machine learning or deep learning models, it is possible to leverage large volumes of normal production data to estimate whether the quality of subsequent products varies. This AI-based anomaly detection system can monitor equipment operation status in real-time, predict potential failures, reduce downtime risks and maintenance costs, and enhance overall production efficiency—all essential for meeting the escalating quality demands in the semiconductor assembly and packaging field [4,5].

In the semiconductor packaging process, wire bonding (WB) is a critical procedure that connects chips on substrates to lead frames or substrates via metal wires. Due to its high stability, mature process, and cost-effectiveness, it is an important process in semiconductor packaging [6]. As the number of stacked layers and the complexity of product leads increase, the quality requirements for WB bonding points have also risen significantly. Traditionally, anomalies in bonding points have relied heavily on manual sampling inspections, X-ray inspections [7], or destructive tests (e.g., pull tests and shear tests), with statistical process control (SPC) managing production parameters within specified limits. These methods are utilized to verify the strength and reliability of bonding points [8].

However, these approaches are more effective at identifying issues such as No Stick on Pad (NSOP) or No Stick on Leadframe (NSOL). Their capability to detect finer quality defects, such as ball offset, oversized or undersized balls, and deformed balls, may be limited. Once anomalies arise that involve multivariable interactions, or if they are not detected during the sampling phase and are discarded, defective products may still enter subsequent processes, impacting quality. Therefore, improving methods for detecting anomalies is crucial for determining product quality.

These references can be useful for substantiating various aspects of the research background and methodologies discussed.

1.2 Research Objectives

The integration of AI in the semiconductor manufacturing industry has become a critical topic, driven by the development of data collectors and AI algorithms [9,10]. Given the rapid generational turnover and high production yield of semiconductor products, anomalous samples are relatively scarce and difficult to obtain [4]. Consequently, traditional supervised learning methods are prone to accuracy issues due to data

imbalance. Therefore, anomaly detection tasks, which identify differences through extensive production data, are more suitable for the current semiconductor manufacturing environment [11].

Recent domain adaptation–based methods, such as the domain feature decoupling network and the joint collaborative adaptation network, improve fault diagnosis robustness under distribution shifts and class imbalance. However, these approaches are primarily designed for vibration or signal-based mechanical systems under varying operating conditions [12,13]. Such models are designed to detect issues early, before defective products proceed to subsequent production stages, thereby creating a protective barrier for quality.

1.3 Research Framework

This study is divided into four chapters. The first chapter describes the research background, motivation, objectives, and framework. The second chapter provides a literature review, summarizing relevant literature and technical background knowledge related to the research area. The third chapter details the research methodology and empirical analysis, describing the research methods, empirical data, feature engineering, and empirical results. The fourth chapter summarizes the empirical results and proposes directions for future research.

2 Literature Review

2.1 Wire Bonding

Wire bonding is a well-established and critical process technology in semiconductor packaging. In modern thin or multi-chip packages, after the wafer undergoes front-end processing such as backside grinding and dicing, the die is attached to the IC substrate through a die bonding process. Subsequently, in the WB process stage, the metal pads on the chip are connected to the lead frame on the IC substrate through metal wires, thereby forming an electrical connection. Despite the gradual development of emerging technologies in interconnect processes, such as flip-chip and wafer-level packaging, WB remains the dominant technology in the microelectronics packaging industry due to its mature technological development and flexibility, with over 80% of semiconductor packaging products utilizing this process technology [14,15].

Semiconductor IC packaging encompasses several key processes, including wafer grinding, wafer dicing, die bonding, wire bonding, encapsulation, marking, ball mounting, singulation, and packing, as illustrated in Fig. 1. This study specifically focuses on wire bonding, a critical procedure that involves welding gold or copper wires onto the bonding pads of chips and substrates. These wires serve as essential conduits for electrical connections and signal transmissions between the internal and external circuits of the IC [16].

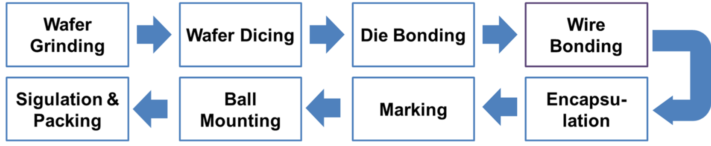


Figure 1: IC packaging key processes.

In the process technology, thermosonic ball bonding is the most widely used method, especially suitable for metal wires such as gold and copper. In this process, a free air ball is generated by an electronic flame-off (EFO) and then pressed onto the bonding pads of the chip and IC substrate with heat, ultrasonic energy, and pressure [14], as illustrated in Fig. 2.

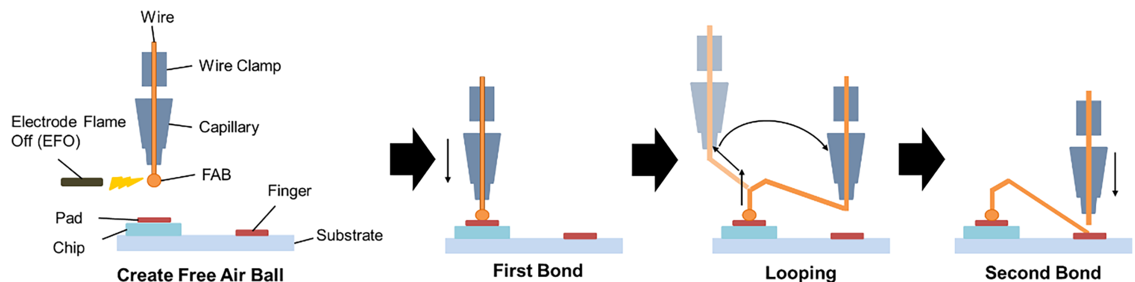


Figure 2: Schematic diagram of the WB process.

In the overall operation flow, the quality of the first bond, the second bond, and the looping is particularly important. Regarding the quality of the bonding points, poor bonding quality, such as oversized or undersized bonding balls or ball position offset, will lead to the functional failure of the packaged product. If defective products are not detected during the inspection phase and problems are exposed to end customers or users, it will cause huge losses to the manufacturing industry. Therefore, companies usually take strict quality control measures, such as ball pull tests, ball shear tests, and automated optical inspection (AOI), to ensure that the quality of each bonding position meets process specifications and to reduce the occurrence of NSOP and NSOL [17,18].

2.2 Fault Detection and Classification Systems

FDC systems involve collecting data from production equipment, such as current values, impedance values, and offsets. By monitoring the changes in this data, the system monitors the equipment and detects anomalies. FDC includes fault detection and fault classification. Fault detection distinguishes between normal and abnormal equipment data, while fault classification categorizes the abnormal data into specific fault modes. By monitoring the production performance of the equipment through FDC, personnel on the production line can take immediate response measures for products and equipment when an abnormality occurs, reducing yield loss caused by the anomaly. FDC systems are crucial for maintaining stable process performance by continuously analyzing operational parameters to identify deviations that could lead to defects or inefficiencies [19,20].

2.3 Anomaly Detection

Anomaly detection refers to the technique of identifying data that deviates from the normal pattern in a dataset. It is widely used in areas such as credit card fraud detection, medical image analysis, equipment failure prediction, and quality monitoring. The use case is that most data follows a certain normal or regular behavior, while a few deviations are considered potential anomalies. These anomalies may represent errors, defects, or other significant events [19]. In semiconductor manufacturing, anomaly detection is crucial for early detection of equipment failures, process deviations, or product defects, helping to improve yield and reduce costs. Common anomaly detection algorithms are summarized in Table 1:

2.4 Isolation Forest

The Isolation Forest algorithm, proposed by Liu et al. in 2008 [21], is a machine learning algorithm used for anomaly detection. According to the literature, anomalies are characterized as being few and different, making them easier to isolate. This algorithm establishes a detection model based on the isolation properties of anomalous data. Its architecture is similar to that of Random Forest, constructing multiple Isolation Trees

to determine anomalous data. Each Isolation Tree randomly selects features and partitions data points based on the differences in these features, creating a tree structure that hierarchically segments the data points. The degree of anomaly for each data point is assessed by calculating the path length of that point in each Isolation Tree. The path length refers to the distance from the root node to the node where the data point is isolated; a shorter path length indicates that the data point was isolated earlier, signifying a greater degree of difference. The average path length across all Isolation Trees is used to determine whether a data point is anomalous.

Table 1: Common anomaly detection algorithms.

Algorithm	Characteristics	Source
Isolation Forest	Constructs a tree structure by randomly selecting features to partition data points. Anomalies are identified based on the path length of data points in the tree structure, making it suitable for high-dimensional data.	Liu et al., 2008 [21]
One-Class Support Vector Machine (SVM)	Based on the concept of SVM, this method learns the boundary of normal data and considers points that are far from the boundary as anomalies. It is applicable in situations with only normal samples.	Schölkopf et al., 2001 [22]
Local Outlier Factor	Calculates the Local Outlier Factor (LOF) score based on the density differences between a data point and its neighboring points, identifying points with significantly lower density as outliers. It is suitable for scenarios where normal and abnormal data distributions are imbalanced.	Breunig et al., 2000 [23]
Connectivity-based Outlier Factor	Considers the connection distances between each data point and its neighbors, calculating COF scores based on the chain connection distances to detect outliers according to their scores.	Tang et al., 2002 [24]

The algorithm is an ensemble learning algorithm, which refers to methods that train multiple machine learning sub-models and aggregate their inference results into a single decision. Generally, ensemble learning yields better inference performance than using individual algorithms alone. Unlike density estimation or distance measurement methods for anomaly detection, Isolation Trees are constructed by randomly sampling from the training data without needing to calculate the distances between all samples, making them advantageous for modeling high-dimensional data.

Reviewing the literature [21,25], the main characteristics can be summarized as follows:

1. Isolation Trees use random sampling from subsets of data to determine anomalies based on a tree structure of data differences, making them suitable for high-dimensional datasets.
2. The time and space complexity are low, particularly advantageous for anomaly detection tasks involving large datasets.
3. It employs unsupervised learning, eliminating the need for data labeling, which reduces data preparation costs.

In the semiconductor industry, Isolation Forest has been applied to process quality and defect detection in FDC tasks. Susto et al. (2017) applied Isolation Forest to monitor plasma etching by collecting sensor data to track process quality deviations and anomalies [26]. Puggini and McLoone (2018) applied Isolation Forest to Optical Emission Spectroscopy (OES) data in the etching process after performing dimensionality

reduction during the feature engineering stage; their empirical results showed classification metrics with an Area Under Curve (AUC) ranging from 0.8 to 0.93 [27].

2.5 Connectivity-Based Outlier Factor

The COF algorithm was proposed by Tang et al. in 2002 as a method for identifying anomalies based on data distances [24]. This approach evaluates the degree of outlierness of a data point by assessing the lengths of the connecting paths to its neighboring points. Unlike the LOF, which calculates the density of a data point in relation to its neighbors, COF computes the Set-Based Nearest Path (SBN-path) for each data point relative to several of its nearest neighbors.

The SBN-path begins at the data point and identifies the nearest neighboring points. It then determines the shortest path that connects the starting point to these neighbors based on the distances between them, producing what is known as the Set Based Nearest Trail (SBN-trail). The average chaining distance (AC-dist) is calculated from the distances of the points within the SBN-trail. The AC-dist values for all data points are then converted into COF scores through proportional calculations, allowing for the determination of whether a sample is an outlier based on its COF score. When the COF score approaches 1, this indicates that the average distance between the data point and its neighbors is similar, suggesting it is normal data. Conversely, when the COF score is significantly greater than 1, this indicates a discrepancy in average distance, suggesting that the data point may be an anomalous outlier.

2.6 Multidimensional Scaling

MDS is a dimensionality reduction algorithm that projects the similarities between high-dimensional data into a lower-dimensional space, aiming to preserve the original geometric structure of the data as much as possible during visualization. MDS is widely applied in various fields, including machine learning, psychology, and manufacturing quality monitoring, particularly useful for exploratory data analysis prior to feature engineering [28].

The main characteristics of MDS include:

1. It can handle arbitrary distance definitions, providing flexibility in methodology.
2. The results are visually interpretable, aiding in understanding the distribution and relationships within high-dimensional data.

3 Research Methodology and Empirical Analysis

3.1 Research Methodology

This study focuses on the WB station, collecting machine data during the production of wire bonding machines. Through Feature Engineering, the machine data is transformed into features used to establish a detection model for anomaly detection regarding the quality of the WB bonding joint point. The algorithms employed include Isolation Forest, COF, and MDS. COF and MDS utilize statistical definitions of outliers as criteria for detecting anomalies. A month's worth of production data is collected as training data, and the performance of the detection model is validated using results obtained after the system goes live. Fig. 3 illustrates the overview of the research methodology process.

Figure 3: Overview of the research methodology process.

In this study, we utilize the mathematical complementarity of three distinct algorithms instead of relying a single model to ensure high recall.

1. **Isolation Forest:** Efficiently captures "global sparse anomalies." It isolates samples with extreme feature values through recursive partitioning, making it ideal for detecting major machine malfunctions.
2. **COF:** Focuses on "local density anomalies." It evaluates the Set-Based Nearest Path (SBN-path) to identify process drifts where data points are within univariate specs but exhibit inconsistent topological structures relative to their neighbors.
3. **MDS:** While MDS is often used for visualization, its essence is preserving high-dimensional Euclidean distances in a lower-dimensional projection. Normal WB data forms a stable geometric manifold; deviations captured by the Interquartile Range (IQR) on MDS coordinates represent complex multivariable risks that univariate limits miss.

In the data collection phase, specific recipes are selected as empirical subjects. Based on the operations of the machines under these specified recipes, data is collected on the process parameters of the WB machines over a month of actual production. Parameters include Bond Force, Ultrasonic Impedance, EFO Voltage, and others. The data units pertain to the chips within each packaged IC on the IC substrate, with the data dimensions representing the process parameter values for each bonding joint point. For example, if the research data pertains to a single IC substrate containing 10 packaged ICs, each with 3 chips, and each chip having 100 bonding joint points, this would yield a total of 30 data sets with each data point comprising the process parameter values for 100 joint points during bonding, as illustrated in Fig. 4.

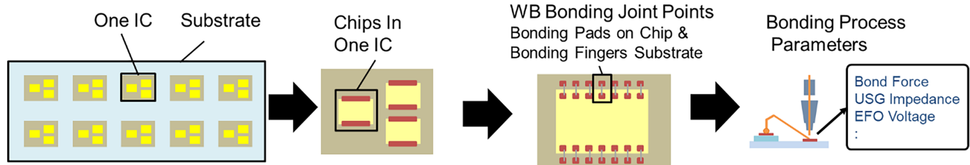


Figure 4: Schematic diagram of WB process parameter data.

In the model training phase, the specified recipes are grouped, and distinct predictive models are established for each group using the algorithms. Considering the impact of the proportion of anomalous data on model accuracy during training for anomaly detection tasks, the training data is cross-referenced with scrap records from the Manufacturing Execution System (MES) and machine alarm events. Data for packaged ICs associated with scrap or alarm records is excluded to ensure that each data point reflects normal production data. For the hyperparameter of Isolation Forest, consider stability of modeling and feature dimensions, tree number set 200, tree sampling number set 256, and feature ratio set 0.8. For the hyperparameter of COF, consider the detection capability of local outlier, neighbor count set to 10. The model deployment process is illustrated in Fig. 5.

Figure 5: Schematic diagram of the system modeling process.

Upon completion of the modeling phase, the detection model is deployed in a live environment to assess its efficacy in identifying anomalies during actual operations. The predictive analysis is conducted on an hourly basis, wherein data is gathered from the machine process parameters while executing the

specified recipes for the designated IC device. The system identifies which bonding joint data for packaged ICs exhibit significant deviations from the training dataset. When the conditions for anomaly detection are met, notifications regarding the anomalous packaged ICs are dispatched to production line supervisors and engineers, who will then instruct operators to conduct a double-check inspection. The inference process is depicted in Fig. 6.

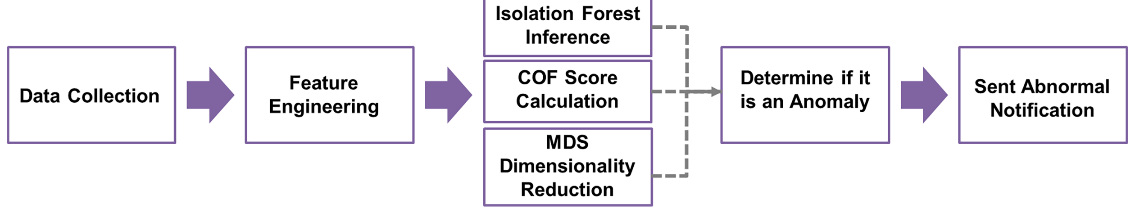


Figure 6: Schematic diagram of the system inference process.

3.2 Empirical Data

This study utilizes internal records of process parameters from WB machines at an Outsourced Semiconductor Assembly and Test (OSAT) facility as empirical data. During the production period, process parameter data for the bonding joints is collected hourly. All parameters are numerical data, and each unit of data can be considered a sequence. This data is combined with domain knowledge from field experts to select relevant process factors associated with bonding joints as modeling features. Additionally, production data is gathered for specific recipes within designated time intervals. Table 2 presents the features selected for modeling in this study.

Table 2: Relevant information of empirical data.

Parameter (Feature) Name	Physical Significance	Potential Failure Mode
Capillary Count	The number of operations performed by the capillary during wire bonding production.	Capillary wear, contamination
Bond Height Delta	The height difference between the first and second solder joints after bonding completion.	Die Attach thickness variation, NSOP
Deformation	The height of the solder ball when bonded to the pad.	Oversized/Undersized ball
Die Height	The height of the chip.	Die Attach thickness variation, NSOP
Die Tilt	Angle of the die relative to the IC substrate.	Ball Misalignment, Deformed Ball
Bond Force	Spindle downward pressure feedback.	Cratering, weak welding
USG Impedance	The ultrasonic impedance value of the machine. Transducer load response during the bond	Capillary wear, contamination
USG Current	The feedback of the ultrasonic current value from the machine.	Capillary wear, contamination

(Continued)

Table 2 (continued)

Parameter (Feature) Name	Physical Significance	Potential Failure Mode
Z at Contact	Spindle height at the moment of initial contact.	Die Attach thickness variation, NSOP
Z at End of Bonding	Spindle height of the main spindle at the completion of the bonding operation.	Die Attach thickness variation, NSOP

In terms of empirical data, this study collects machine parameter records from specified recipes operating under normal conditions to form the training dataset. There are differences in the application products, number of bonds, and data volume of these four formulas. The parameter records collected after the system deployment, along with blind test results, serve as the testing dataset. The training dataset is used for algorithm modeling, while the testing dataset acts as empirical data. The time frame for the training dataset spans from 01 October 2024 to 31 October 2024, while the collection period for the testing dataset ranges from 01 November 2024 to 31 December 2024. Table 3 provides relevant information about the empirical data. In accordance with company confidentiality agreements, the empirical data will be de-identified, and the inference results from the model on the testing dataset will be provided.

Table 3: Number of bonds and data volume in empirical data.

Customer	Application Products	Recipe Name	Number of Bonds	Data Volume
Customer I	Memory	Recipe A	168	1,627,570
Customer II	Logic IC	Recipe B	344	1,010,216
Customer I	Memory	Recipe C	172	2,076,554
Customer III	Memory	Recipe D	172	897,970

3.3 Feature Engineering

Feature engineering refers to the process of transforming raw data into features that enhance the model's visibility into the data. After selecting the features, the raw data undergoes data cleaning processes, including handling missing values, addressing outliers, and normalization. These steps reduce the impact of anomalies and scale differences on model performance. Following data cleaning, data transformations are applied based on the characteristics of each feature. For instance, descriptive statistics are computed for continuous variables, and missing values are imputed using interpolation methods. The processed training data is then used to establish an anomaly detection model. All statistics used for feature scaling and anomaly threshold are calculated solely from the October 2024 training data. Ensuring no future data is used during prediction and better reflecting the real production situation. Based on the nature of the features, the data processing methods selected for this study are presented in Table 4.

Table 4: Data processing methods for selected features.

Feature Name	Data Transformation
Capillary Count	1. For missing values, imputed using the average capillary count from the preceding and following IC substrates. 2. Based on 1.5 IQR rule, removed outliers from the training data.
Bond Height Delta Deformation Bond Force USG Impedance USG Current Z at Contact Z at End of Bonding	1. For missing values, imputed using the median of the same solder joint position on the same IC substrates. 2. Data pre-processing through max-min normalization 3. Based on 1.5 IQR rule, removed outliers from the training data.
Die Height Die Tilt	1. For missing values, imputed using the mode of the same chip. 2. Data pre-processing through max-min normalization

3.4 Anomaly Detection

The empirical method of this study utilizes the inference results from the Isolation Forest, COF scores, and the two-dimensional transformation values from MDS to determine the presence of solder joint anomalies based on machine parameter data. Both COF and MDS require the definition of threshold values to serve as criteria for anomaly detection. Given that the distribution of anomalies typically exhibits skewness and the empirical data have not been labeled, this study adopts the IQR method as the threshold for determining anomalies in COF and MDS. The IQR method, based on Tukey's (1977) box plot analysis [29], identifies outliers without assuming a specific distribution. The calculation of IQR is as follows:

$$\text{IQR} = \text{Q3} - \text{Q1} \tag{1}$$

where Q1 is the first quartile, and Q3 is the third quartile of the data. A data point X is considered an outlier if it satisfies the following condition:

$$x < \text{Q1} - 1.5 \times \text{IQR} \text{ or } x > \text{Q3} + 1.5 \times \text{IQR} \tag{2}$$

Based on the IQR method, Table 5 presents the approaches used by each algorithm to identify outliers in the study. If all three algorithms classify a data point as anomalous, production line staff will be notified to conduct a product confirming inspection.

Table 5: Criteria for anomaly detection in the study method.

Algorithm	Anomaly Detection Criteria
Isolation Forest	Model inference classifies as anomalous
COF	COF > Q3 + 1.5 × IQR.
MDS	MDS < Q1 − 1.5 × IQR or MDS > Q3 + 1.5 × IQR.

3.5 Empirical Results

The objective of this study is to utilize anomaly detection algorithms to assess whether bonding joint anomalies exist in packaged ICs based on machine processing parameters. To evaluate model performance, test dataset results were analyzed. During the system's operational period, approximately 5,612,310 packaged ICs were monitored, and 96 notifications of detected anomalies were sent during the testing phase. After confirmation of material quality by production line staff, it was determined that 49 instances exhibited bonding joint anomalies, while 47 were false alarms. Additionally, among all packaged ICs, there was one anomaly within the observation range that the system failed to detect.

For performance evaluation, a confusion matrix was employed as the assessment criterion for the classification problem. Detected anomalies were categorized as true positives (TP), false alarms as false positives (FP), undetected anomalies as false negatives (FN), and the remaining cases as true negatives (TN).

Based on the confusion matrix, the empirical results can quantify accuracy, recall, and precision, calculated as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{(\text{TP} + \text{NP} + \text{FN} + \text{TN})} \tag{3}$$

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \tag{4}$$

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \tag{5}$$

Accuracy refers to the overall performance of the classification, defined as the ratio of TP and TN among all observed products. Recall is considered a measure of the system's ability to intercept all anomalies, represented by the ratio of detected TP to the total number of anomalies (TP + FN). Precision can be viewed as the hit rate of the system's alerts, calculated as the ratio of true positives (TP) to the total alerts (TP + FP).

According to Formulas (3)–(5), the empirical results indicate that the system achieved an accuracy of 99%, a recall of 98%, and a precision of 51%, as illustrated in Fig. 7.

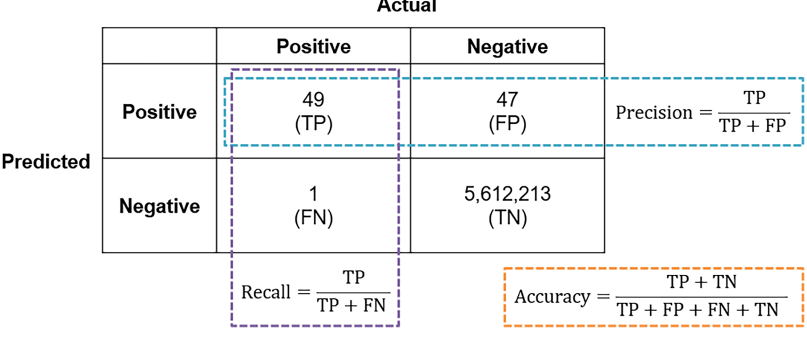


Figure 7: Confusion matrix of empirical results.

In terms of OSAT process quality, the production line must strive to prevent abnormal products from flowing into subsequent processes to avoid increasing impacted in yield loss. The principle for anomaly detection emphasizes "better to overreact than to overlook." From an evaluation perspective, the system's performance prioritizes recall, aiming to minimize the occurrence of FN. Once the system sends an alert, production line personnel must allocate resources to intercept materials and verify whether the notified

packaged products are indeed anomalous and what type of anomaly they represent. If there are too many false alarms, it significantly increases the operational burden on production line staff. Therefore, it is essential to consider the system's precision, ensuring that FN occurrences are avoided while minimizing the frequency of FP.

Based on the inference results during the empirical period, a significant disparity was found between the number of normal and anomalous data points. Despite this data imbalance, the recall rate was 98%, indicating that the system has a certain capability to identify bonding joint anomalies in the WB process. Although precision is 51% for the inference result, this corresponds to only 17 ppm over 5.61 million ICs, which is acceptable for a semiconductor packaging line. When an alert is triggered, the system automatically notifies supervisors via email, and engineers perform a quick double-check within minutes. Compared to the high cost of missed defects, this approach is highly cost-effective. According to the assessments made by inspectors, the system primarily detected anomalies such as ball misalignment and deformed balls, as well as rare anomalies like oversized balls and damaged pads during the empirical period. The detected anomalies are summarized in Table 6.

Table 6: Detected anomalous samples from empirical results.

Anomaly Type	Detected Quantity	Sample Image
Ball Misalignment	37	
Deformed Ball	10	
Oversized Ball	1	
Damaged Pad	1	

In the research methodology, MDS was used to project the empirical data onto a two-dimensional plane, as shown in Fig. 8. The red box indicates the criteria for identifying anomalies using MDS. When performing dimensionality reduction with MDS, the similarity between the original data points is taken into account. The scatter plot reveals that data points associated with ball misalignment are primarily located on the left and right sides of the two-dimensional plane, while deformed balls and oversized balls are positioned on the upper and lower parts. Notably, the data points for damaged pads are the most distant from the others.

Considering the causes of anomalies in the WB process, deformed balls and oversized balls are categorized as shape-related anomalies, differing from the causes of ball misalignment and damaged pads. This phenomenon is also reflected in the distribution of data after the MDS projection.

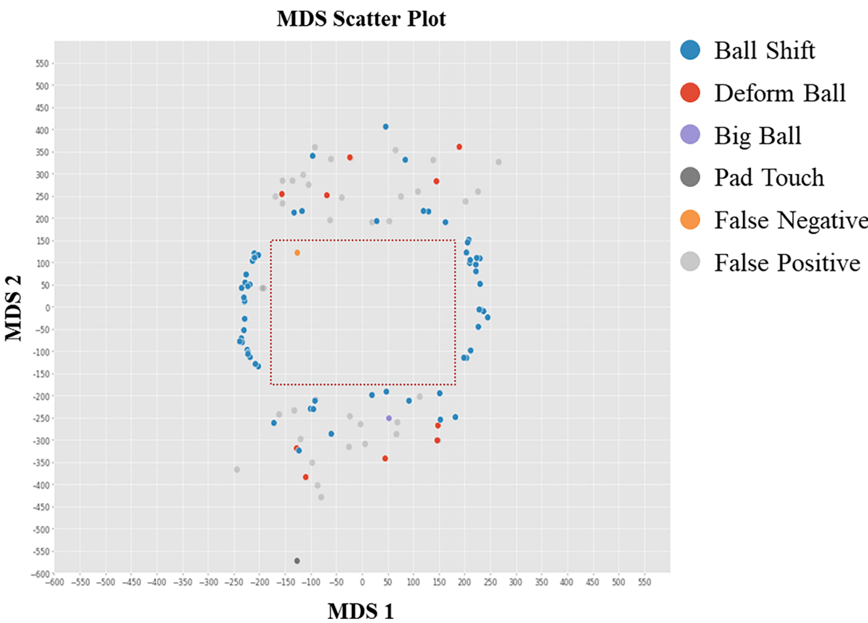
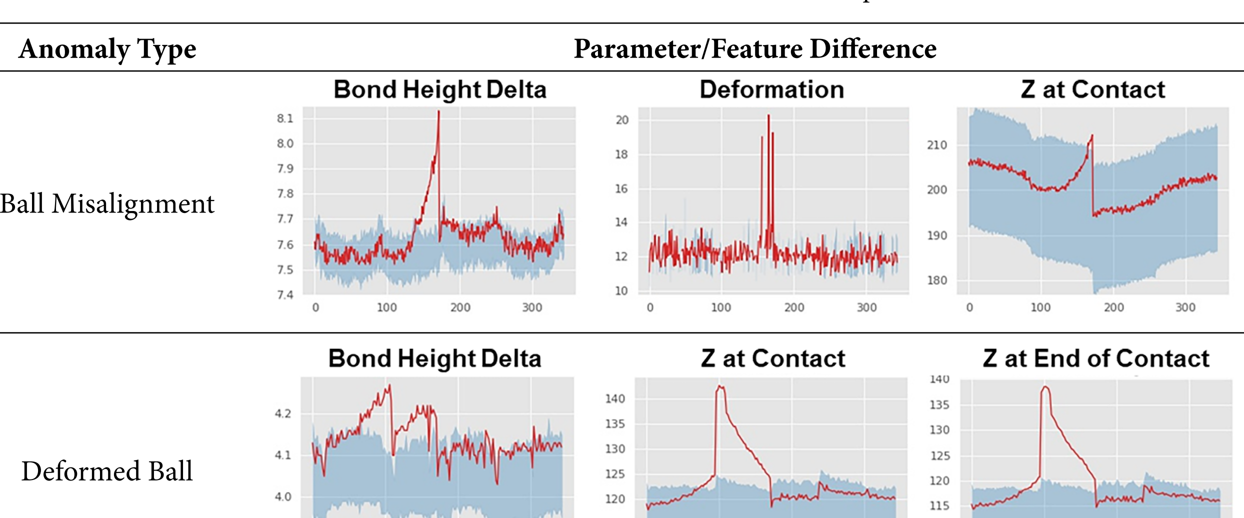


Figure 8: Scatter plot of empirical data projected onto a two-dimensional plane using MDS.

Further analyze the differences between normal data and abnormal data, as shown in Table 7. The red line represents the parameter value of the abnormal packaged integrated circuit, and the blue area represents the parameter value range of the normal packaged integrated circuit on the same substrate. It can be found that the trend of the abnormally packaged integrated circuits is different from those of other packaged integrated circuits on the same substrate.

Table 7: Parameter differences of anomalous samples.



Anomaly Type	Parameter/Feature Difference		
Ball Misalignment	Bond Height Delta	Deformation	Z at Contact
Deformed Ball	Bond Height Delta	Z at Contact	Z at End of Contact

3.6 Performance Evaluation

To further validate the necessity of the proposed three-model ensemble framework, an ablation study was conducted to evaluate the performance of individual models and their combinations. In addition, since the research data is a multivariate time series, in order to benchmark the proposed method against modern approaches, Long Short-Term Memory (LSTM) was adopted as a model baseline for comparison. The method is a classic deep learning algorithm suitable for predicting time series data.

For research methodology, we verified the necessity of the 3 algorithms through an ablation study, as shown in Table 8. The results indicate that single models exhibit limited recall performance. Combining Isolation Forest and COF improves recall but still generates excessive false positives. The proposed three-model ensemble achieves the highest recall while improving precision to 51%. The improvement confirms that the three models capture complementary anomaly characteristics.

Table 8: Ablation study result.

Algorithm	Recall	Precision
Isolation Forest	0.92	0.40
COF	0.88	0.41
MDS & IQR	0.84	0.46
Isolation Forest + COF	0.96	0.44
Proposed Methodology	0.98	0.51

For inference result, in order to validate the generalization ability of the research methodology, we adopt holdout cross-validation to estimate precision and recall. For each iteration, 80% of the data is randomly selected as training data for modeling, and the remaining 20% is used for inference. After 100 iterations, the mean and standard deviation of precision and recall were estimated, and estimating 95% confidence intervals of both criteria. The sampling result is summarized in Table 9, and a small standard deviation indicates that the experimental results are reliable.

Table 9: Holdout cross-validation result.

Criteria	Mean	Standard Deviation	95% Confidence Interval
Recall	0.979	0.060	[0.964, 0.988]
Precision	0.498	0.155	[0.468, 0.528]

Furthermore, we used LSTM as a benchmark model for comparison with the research methods. The following is the model architecture,

1. Input: 1-dimensional array reshape from experimental data.
2. LSTM layers: 3
3. Hidden units: 64
4. Activation function: Rectified Linear Unit (ReLU).
5. Loss function: Mean Squared Error (MSE).
6. Anomaly threshold: 99% reconstruction error on training data.

After holdout cross-validation, Table 10 shows inference result between LSTM and proposed method, based on detection performance, the proposed method demonstrates superior stability and higher recall.

Table 10: Inference result between LSTM and proposed methodology.

Model	Recall	Precision
LSTM	0.894 ± 0.156	0.428 ± 0.228
Proposed Methodology	0.979 ± 0.060	0.498 ± 0.155

3.7 Feature Importance

Feature Importance is a metric that measures the influence of each feature on the target variable in machine learning models. It is used to justify feature selection and assess the contribution of features to model inference. This study employs Permutation Importance to evaluate the contribution of selected features to the model's detection capability. This method involves randomly shuffling the data indices of the features in the dataset and measuring the decrease in evaluation criteria to assess the impact of each feature on model performance.

The calculation of Permutation Importance does not rely on the internal structure of the model; rather, it evaluates the influence of each feature based on output results. This approach is applicable to various machine learning models [30]. The calculation process is as follows:

1. **Establish Baseline Performance:** Calculate the evaluation criterion for the original dataset after model inference, serving as the baseline performance, such as MSE, Coefficient of Determination (R^2), accuracy, recall, etc.
2. **Shuffle Feature Indices:** Randomly shuffle the index order of the specified feature and infer the results using the modified dataset. Calculate the evaluation criterion for this new inference.
3. **Record Performance Decrease:** Measure the drop in the evaluation criterion as the Permutation Importance score for that feature.
4. **Repeat for Other Features:** Select other features and repeat steps 2 and 3.

In this study, a recall rate of 0.98 from the empirical results is used as the baseline performance criterion. Based on the detection performance from the empirical results, the Permutation Importance of each feature is further calculated to assess the impact of each variable on the anomaly detection model, as summarized in Table 11.

Table 11: Permutation importance of each feature.

Feature Name	Recall Rate of Original Data	Recall Rate After Shuffling	Permutation Importance
Die Tilt	0.98	0.92	0.06
USG Impedance		0.88	0.1
Capillary Count		0.84	0.14
Die Height		0.84	0.14
Bond Force		0.74	0.24
USG Current		0.87	0.11
Deformation		0.54	0.44
Z at End of Bonding		0.56	0.42
Bond Height Delta		0.72	0.26
Z at Contact		0.61	0.37

Based on the results of the importance scores, the selected features in this study all contribute to the model's detection performance. Notably, the drop in recall rates for ball deformation amount, spindle bonding height, spindle operation completion height, bonding joint height difference, and bonding force is significantly greater than that of the other variables (ranging from 0.24 to 0.44). This indicates that these features have a more substantial influence on the detection method.

In contrast, the decline in recall rates for the other variables falls between 0.06 and 0.14, suggesting that they also have some impact on the model's detection capability, albeit to a lesser extent.

4 Conclusion

4.1 Research Conclusions

The objective of this study was to estimate the bonding joint quality in the WB process through machine production data. Utilizing the operational concepts of FDC, algorithms such as Isolation Forest, COF, and MDS were employed to establish an anomaly detection system for bonding joint quality. Empirically, approximately one month of normal production data was collected to estimate the production conditions over the subsequent two months, with actual inspection feedback from production line staff serving as the empirical results.

From the empirical findings, the proposed method demonstrated an interception capability for WB bonding joint anomalies of about 98%, indicating that the research method possesses a significant detection ability for bonding joint anomalies in the empirical product.

After confirming the feasibility of the research method, the characteristics of the WB bonding joint anomaly detection system proposed in this study can be summarized in two points:

1. **Enhanced Quality Assurance for the WB Process:** As shown in the empirical results, existing detection methods rely on the machine's built-in inspection functions and quality sampling. The proposed system offers a comparative analysis based on machine parameters in a high-dimensional space through anomaly detection algorithms, adding an additional layer of quality protection for WB bonding joints.
2. **Understanding Parameter Differences between Normal and Anomalous Products:** As shown in Table 7, we can find the data difference between the detected abnormal unit and the normal unit. In addition to notifying the machines of any anomalies, this information is also provided to the production engineering team for reference and analysis of possible causes of the phenomenon.

4.2 Future Work

This study demonstrates that the bonding joint anomaly detection system, developed utilizing WB machine parameters, can effectively aid production lines in detecting wire bonding joint anomalies. Due to the challenges related to data collection and machine variability, only four program sets were implemented during the empirical phase. However, in the context of practical OSAT mass production, the number of recipes utilized in the WB process far exceeds four.

Therefore, the subsequent aim of this research is to broaden the detection system to incorporate additional mass production recipes. By harnessing this anomaly detection framework, we seek to mitigate the risk of bonding joint anomalies bypassing the WB process.

As semiconductor manufacturing processes evolve towards finer pitch and increased complexity, assembly packaging processes may introduce subtle defects that current inspection and testing tools struggle to identify. Consequently, methodologies akin to FDC may provide viable solutions for enhancing future process quality control.

Acknowledgement: Not applicable.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Chin Ta Wu and Shing Han Li; methodology, Chin Ta Wu and Shing Han Li; software, Chin Ta Wu; validation, Chin Ta Wu, Shing Han Li and Ching Shih Tsou; formal analysis, Chin Ta Wu; investigation, Chin Ta Wu; resources, Chin Ta Wu; data curation, Chin Ta Wu; writing—original draft preparation, Chin Ta Wu; writing—review and editing, Shing Han Li; visualization, Chin Ta Wu; supervision, Shing Han Li; project administration, Shing Han Li; funding acquisition, Shing Han Li. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The datasets generated and/or analyzed during the current study are not publicly available due to company confidentiality agreements but are available from the corresponding author on reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. von Tils V. Zero defects—reliability for automotive electronics. In: Proceedings of the 2008 International Interconnect Technology Conference; 2008 Jun 1–4; Burlingame, CA, USA. doi:10.1109/iitc.2008.4546908.
2. Raina R. Achieving zero-defects for automotive applications. In: Proceedings of the 2008 IEEE International Test Conference; 2008 Oct 28–30; Santa Clara, CA, USA. doi:10.1109/TEST.2008.5483611.
3. Leberruyer N, Bruch J, Ahlskog M, Afshar S. Toward zero defect manufacturing with the support of artificial intelligence—insights from an industrial application. Comput Ind. 2023;147:103877. doi:10.1016/j.compind.2023.103877.
4. Song S, Baek JG. New anomaly detection in semiconductor manufacturing process using oversampling method. In: Proceedings of the 12th International Conference on Agents and Artificial Intelligence; 2020 Feb 22–24; Valletta, Malta. doi:10.5220/0009170709260932.
5. Raghuwanshi P. Revolutionizing semiconductor design and manufacturing with AI. J Knowl Learn Sci Technol. 2024;3(3):272–7. doi:10.60087/jklst.vol3.n3.p.272-277.
6. Qin I, Shah A, Xu H, Chylak B, Wong N. Advances in wire bonding technology for different bonding wire material. Int Symp Microelectron. 2015;2015(1):406–12. doi:10.4071/isom-2015-wp33.
7. Zhan D, Huang R, Yi K, Yang X, Shi Z, Lin R, et al. Convolutional neural network defect detection algorithm for wire bonding X-ray images. Micromachines. 2023;14(9):1737. doi:10.3390/mi14091737.
8. Rocha LS, Fernandes IJ, Peter C, Da S Pereira PR. Wire bonding failure investigation in semiconductor packaging process. In: Proceedings of the 2024 IEEE URUCON; 2024 Nov 18–20; Montevideo, Uruguay. doi:10.1109/urucon63440.2024.10850051.
9. De Luca C, Lippmann B, Schober W, Al-Baddai S, Pelz G, Rojko A, et al. AI in semiconductor industry. In: Artificial intelligence for digitising industry—applications. Abingdon, UK: Taylor Francis Group; 2022. p. 105–12.
10. Niveditha PKP, Patel PA, Sahu PR. AI-powered predictive analytics for semiconductor manufacturing: case studies and industry applications. Int J Innov Res Sci Eng Technol. 2024;13(6):1–14. doi:10.15680/ijirset.2024.1306280.
11. Gao T, Yang J, Wang W, Fan X. A domain feature decoupling network for rotating machinery fault diagnosis under unseen operating conditions. Reliab Eng Syst Saf. 2024;252(3):110449. doi:10.1016/j.ress.2024.110449.
12. Li Y, Yang J, Wang W, Gao T. A joint collaborative adaptation network for fault diagnosis of rolling bearing under class imbalance and variable operating conditions. Adv Eng Inform. 2026;69(9):103931. doi:10.1016/j.aei.2025.103931.
13. Johnson G. Anomaly detection in semiconductor process validation using unsupervised learning and generative models. Int J Eng Technol Res Dev. 2022;3(2):5–10. doi:10.55248/gengpi.07.0426.20827.

14. Zhou H, Zhang Y, Cao J, Su C, Li C, Chang A, et al. Research progress on bonding wire for microelectronic packaging. Micromachines. 2023;14(2):432. doi:10.3390/mi14020432.
15. Gomes J, Mayer M, Lin B. Development of a fast method for optimization of Au ball bond process. Microelectron Reliab. 2015;55(3–4):602–7. doi:10.1016/j.microrel.2014.12.013.
16. Yu CM, Lai KK, Chen KS, Chang TC. Process-quality evaluation for wire bonding with multiple gold wires. IEEE Access. 2020;8:106075–82. doi:10.1109/ACCESS.2020.2998463.
17. Zhou H, Chang A, Fan J, Cao J, An B, Xia J, et al. Copper wire bonding: a review. Micromachines. 2023;14(8):1612.
18. Sethu RS. Reducing non-stick on pad for wire bond: a review. Aust J Mech Eng. 2012;9(2):147–59. doi:10.7158/m11-771.2012.9.2.
19. Jiang JA, Chuang CL, Wang YC, Hung CH, Wang JY, Lee CH, et al. A hybrid framework for fault detection, classification, and location—part I: concept, structure, and methodology. IEEE Trans Power Deliv. 2011;26(3):1988–98. doi:10.1109/TPWRD.2011.2141157.
20. Chandola V, Banerjee A, Kumar V. Anomaly detection: a survey. ACM Comput Surv. 2009;41(3):1–58. doi:10.1145/1541880.1541882.
21. Liu FT, Ting KM, Zhou ZH. Isolation forest. In: Proceedings of the 2008 Eighth IEEE International Conference on Data Mining; 2008 Dec 15–19; Pisa, Italy. doi:10.1109/ICDM.2008.17.
22. Schölkopf B, Platt JC, Shawe-Taylor J, Smola AJ, Williamson RC. Estimating the support of a high-dimensional distribution. Neural Comput. 2001;13(7):1443–71. doi:10.1162/089976601750264965.
23. Breunig MM, Kriegel HP, Ng RT, Sander J. LOF: identifying density-based local outliers. In: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data; 2000 May 16–18; Dallas, TX, USA.
24. Tang J, Chen Z, Fu AW, Cheung DW. Enhancing effectiveness of outlier detections for low density patterns. In: Advances in knowledge discovery and data mining. Berlin/Heidelberg, Germany: Springer; 2002. p. 535–48. doi:10.1007/3-540-47887-6_53.
25. Hariri S, Kind MC, Brunner RJ. Extended isolation forest. IEEE Trans Knowl Data Eng. 2021;33(4):1479–89. doi:10.1109/tkde.2019.2947676.
26. Susto GA, Beghi A, McLoone S. Anomaly detection through on-line isolation forest: an application to plasma etching. In: Proceedings of the 2017 28th Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC); 2017 May 15–18; Saratoga Springs, NY, USA. doi:10.1109/ASMC.2017.7969205.
27. Puggini L, McLoone S. An enhanced variable selection and Isolation Forest based methodology for anomaly detection with OES data. Eng Appl Artif Intell. 2018;67(3R):126–35. doi:10.1016/j.engappai.2017.09.021.
28. Borg I, Groenen PJF. Modern multidimensional scaling: theory and applications. New York, NY, USA: Springer New York; 2005.
29. Tukey JW. Exploratory data analysis (Vol. 2). Reading, MA, USA: Addison-Wesley; 1977. p. 131–60.
30. Altmann A, Toloşi L, Sander O, Lengauer T. Permutation importance: a corrected feature importance measure. Bioinformatics. 2010;26(10):1340–7. doi:10.1093/bioinformatics/btq134.