Sensors and Materials, Vol. 37, No. 12 (2025) 5501–5516
MYU Tokyo

S & M 4259

Anomaly Detection in Wafer Grinding Using Accumulated-generating-operation-processed In-process Sensor Data and Ensemble Learning: A Case Study

Chien-Chih Chen,[1] Yao-San Lin,[2] Hung-Yu Chen,[1*]
Wei-Yuan Sun,[1] and Chih-Jung Kuo[1]

[1]Department of Information Management, National Chin-Yi University of Technology, Taiwan, ROC
[2]Department of Industrial Engineering and Management, National Chin-Yi University of Technology, Taiwan, ROC

(Received May 19, 2025; accepted October 8, 2025)

Keywords: wafer grinding, in-process sensors, anomaly detection, time series analysis of sensor data, accumulated generating operation (AGO)

The initial stage of semiconductor packaging, wafer grinding, relies critically on in-process sensors to enable the precise thinning of wafers to a target thickness. Specifically, inner and outer in-process gauges (IPGs) provide real-time height measurements that dictate the vertical positioning of grinding wheels. However, the accuracy of these outer height sensors becomes compromised as wafers become increasingly thin owing to stress-induced lifting on the unground side. This sensor inaccuracy can lead to erroneous feedback, resulting in excessive grinding pressure and a heightened risk of wafer cracking or breakage. The height data acquired from these sensors are presented as variable-length, short-term time series, which are affected by fluctuating production demands. Furthermore, machine recalibration based on individual chuck table heights complicates consistent monitoring using absolute sensor readings. To address these sensor-related challenges in acquiring and interpreting reliable height information, we introduce a novel approach utilizing the accumulated generating operation (AGO) to convert the nonpatterned time series data into discernible patterns, effectively reframing the problem as a pattern recognition task based on sensor data analysis. Addressing the inherent class imbalance due to the infrequency of anomalies (246 abnormal instances within a dataset of 55143), the bootstrap method is employed for data balancing. Subsequently, a bagging ensemble of five back-propagation neural networks is trained for anomaly detection using the transformed sensor data. Experimental results showcase a substantial improvement in the identification of abnormal height patterns in these short-term time series derived from sensor measurements, with the $F1$-score increasing significantly from 0.201 to 0.871. These findings underscore the efficacy of the proposed methodology in enhancing wafer grinding process monitoring and anomaly detection by effectively handling the complexities of in-process sensing.

*Corresponding author: e-mail: chenhy@ncut.edu.tw
https://doi.org/10.18494/SAM5746

ISSN 0914-4935
https://myukk.org/

1. Introduction

The semiconductor packaging industry has grown significantly in strategic importance, driven by the demand for smaller, faster, and more energy-efficient electronics across consumer, automotive, healthcare, and artificial intelligence (AI) sectors. As the critical bridge between silicon dies and functional electronic systems, advanced packaging technologies are essential for maximizing integrated circuit (IC) performance. The wafer grinding process serves as a fundamental initial step that directly determines the thickness, surface quality, and structural integrity of thinned wafers intended for complex packaging. Precise control and monitoring in this stage remain essential for ensuring high production yields and reliable downstream manufacturing.

1.1 Research background

The semiconductor packaging industry relies on precise and efficient manufacturing processes to produce high-quality ICs. The semiconductor packaging processes are illustrated in Fig. 1. Wafer grinding, the initial step in this critical phase, involves the controlled thinning of silicon wafers to achieve specific target thicknesses essential for subsequent packaging operations. This process typically employs grinding wheels with wafers positioned on chuck tables. The precise control of the grinding wheel's height is crucial for achieving the desired wafer thickness and surface finish. This control is facilitated by real-time feedback from in-process gauges (IPGs) strategically positioned to monitor wafer height, as shown in Fig. 2.

During the grinding process, inner and outer IPGs play a crucial role in providing essential height measurements. IPGs are classified into contact and noncontact gauges, as illustrated in

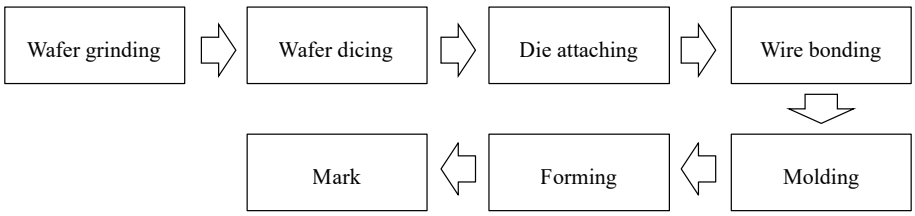

Fig. 1. Subprocesses in the IC packaging process.

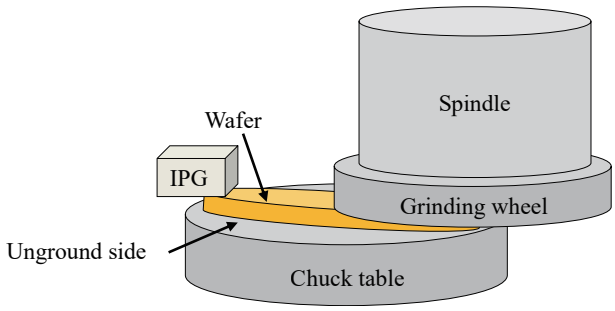

Fig. 2. (Color online) Schematic of the wafer grinding process with IPG.

Fig. 3. Contact gauges employ physical probes that directly interface with the wafer surface to measure thickness. The measurements represent the total distance from the chuck table to the wafer surface, including the back-grinding (BG) tape thickness, as depicted in Fig. 4(a). In contrast, noncontact gauges utilize optical technologies, such as laser-based systems, to measure wafer thickness without physical contact, as shown in Fig. 3(b). The measurements directly represent the actual wafer thickness, excluding the BG tape, as illustrated in Fig. 4(b). This fundamental difference in measurement methodology makes each gauge type suitable for different stages of the grinding process.

However, as wafer dimensions shrink and target thicknesses decrease, the reliability of outer IPG measurements can be significantly compromised. Wafer warpage, primarily from thermal expansion differences between materials, causes the unground portion of the thin wafer to rise slightly from the chuck table, introducing inaccuracies in the outer height readings, as shown in Fig. 2. This inaccurate feedback can lead to excessive pressure from the grinding wheels, remarkedly increasing the risks of wafer cracking, breakage, and ultimately, reduced production yield and increased costs.

Furthermore, the height data acquired from these IPGs present unique analytical challenges. Production schedules often fluctuate in accordance with customer orders, resulting in variable-length and short-term time series data for each grinding cycle. Adding to this complexity, the measurement baseline for the IPGs is typically reset on the basis of the height of the current chuck table in use. This machine-specific recalibration renders absolute height values less meaningful for consistent, cross-machine monitoring and anomaly detection. Identifying deviations from normal grinding patterns within these short, variable time series, especially when relying on potentially noisy outer IPG data, becomes a significant hurdle. The inherent

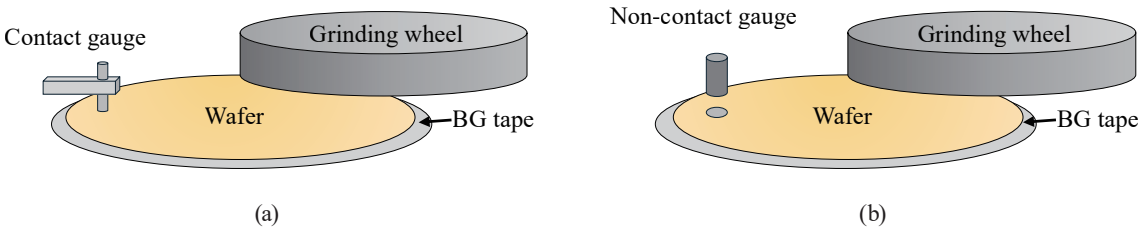

Fig. 3. (Color online) (a) Contact and (b) noncontact gauges.

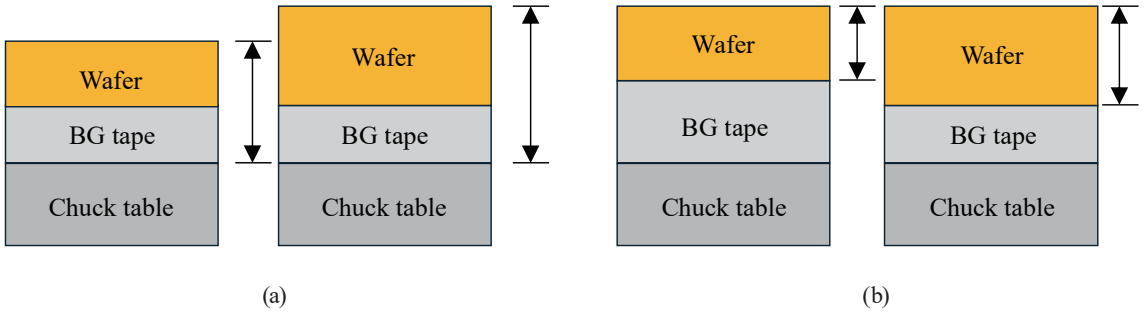

Fig. 4. (Color online) Measured heights of (a) contact and (b) noncontact gauges.

variability in chuck table height during processing creates significant measurement inconsistencies across different lots, challenging traditional data-driven approaches. The characteristics of height measurements are summarized below.

- Extremely short time series data.
- Height measurements lack intrinsic quantitative significance.
- Independence between grinding lots.
- Variable measurement lengths due to various product thicknesses.
- Highly imbalanced class distributions (normal versus abnormal).

1.2 Research motivation

Recent studies have been focused on enhancing the effectiveness of wafer grinding by addressing critical operational aspects. Yao *et al.*[1] investigated how grinding residual height affects the wafer surface shape and developed a model and proposed methods to improve wafer flatness. Tao *et al.*[2] advanced this work by predicting and measuring grinding force in wafer surface rotational grinding (WSRG), creating an improved model to analyze the impacts of grinding forces on precision and surface integrity under various parameters. In further research,[3] material removal and surface generation in WSRG were explored by introducing a surface topography model that optimizes parameters for smoother surfaces while considering grain size effects. These studies provide essential theoretical frameworks and practical guidance for improving wafer flatness, surface integrity, and overall quality through a better understanding and control of residual height, grinding forces, and material removal processes. While previous investigations were focused on optimizing the physical aspects of the grinding process itself, in this study, we take a different approach by focusing on the detection of abnormal patterns directly from in-process height measurements, offering a complementary method for enhanced process monitoring and real-time anomaly detection.

While recurrent neural networks (RNNs), particularly long short-term memory (LSTM)[4] networks, have shown promise in various time series anomaly detection tasks, their direct application to variable-length data in wafer grinding faces significant hurdles. LSTMs typically require input sequences of a fixed length, necessitating either the padding or truncation of the time series, which can lead to information loss or the introduction of artificial signals. Moreover, training robust LSTM models often requires substantial amounts of labeled data, which is particularly scarce for abnormal grinding events that are the focus of anomaly detection. The inherent variability in the length of these short-term time series, coupled with the limited availability of anomaly data, makes it challenging to train effective LSTM models that have good generalizability to unseen patterns of various durations in the wafer grinding process. Over the past two decades, the grey model (GM)[5] and its extensions have shown their effectiveness in handling short-term time series, such as in the fields of tourism,[6] medicine,[7] energy,[8,9] environmental protection,[10] and decision making.[11] However, even a precise prediction may not tackle this issue because abnormalities can occur within the first few height measurements. Additionally, absolute height values are meaningless.

To address these sensing and data analysis challenges inherent in the wafer grinding process, we propose a novel approach utilizing the accumulated generating operation (AGO).[5] This

technique transforms the nonpatterned, short-term time series data obtained from the IPGs into data exhibiting more recognizable patterns, effectively shifting the analytical focus from a complex time series analysis to a more straightforward pattern recognition. There are other ways to aggregate time series data into patterns, such as the fractional-order accumulation operation. However, determining the order to obtain better results can be challenging.[12] In this study, AGO is adopted as an example. The aggregated time series presents monotonically increasing trends, mitigating the impact of uncertainty in the stochastic process. After AGO, the dynamic time series data are transformed into static data that can be learned by classification algorithms, such as back-propagation neural networks (BPNNs). Recognizing the inherent rarity of abnormalities in the wafer grinding process, as evidenced by a highly imbalanced dataset with only 246 abnormal samples out of 55143, in this study, we employ the bootstrap[13] method to mitigate the impact of skewed class distributions during model training. Subsequently, a bootstrap aggregating[14] (hereafter referred to as bagging) ensemble comprising five BPNNs is utilized as the learning framework for robust anomaly detection based on the transformed sensor data.

The experimental results presented in this paper demonstrate that the proposed methodology significantly enhances the accuracy of identifying abnormal height patterns within these challenging short-term time series. Specifically, a substantial increase in $F1$-score from 0.201 to 0.871 is observed, highlighting the effectiveness of the combined AGO-bootstrap-bagging approach. This improvement underscores the potential of the proposed method to overcome the limitations of traditional time series analysis and the challenges posed by sensor inaccuracies in wafer grinding process monitoring, ultimately contributing to improved quality control and reduced manufacturing defects.

The remainder of this paper is structured as follows. In Sect. 2, the proposed method is introduced. In Sect. 3, we explain the experimental settings and results. Further application is discussed in Sect. 4, and the conclusions of this study are given in Sect. 5.

2. Proposed Method

To address the aforementioned challenges in analyzing variable-length, short-term time series data with limited labeled anomalies, in this study, we develop a novel pattern recognition method for effectively detecting abnormal height measurement patterns. As illustrated in Fig. 5, our approach comprises three key stages: sequence transformation using AGO, abnormal data augmentation via the bootstrap method to address class imbalance, and robust anomaly classification through a bagging ensemble of BPNNs. The primary objective is to accurately distinguish subtle abnormality patterns (represented by red lines in Fig. 5) from dominant normality patterns (represented by blue lines) by utilizing the power of ensemble learning on the transformed and balanced data. This pattern recognition framework offers a departure from traditional time series analysis and is specifically designed to overcome the limitations posed by the unique characteristics of the wafer grinding height data.

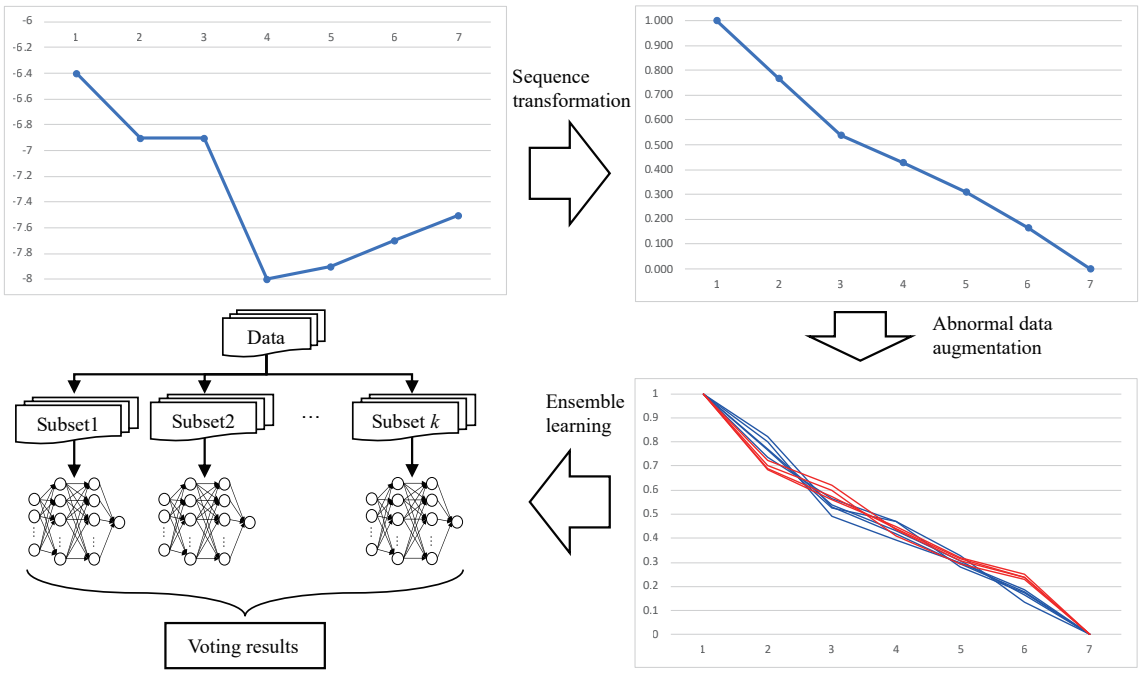

Fig. 5. (Color online) Processes of the proposed method.

2.1 Notation and example definitions

The wafer height measurements are time series data. Here, we suppose the time series data $x = \{x(k) \mid k = 1, 2, \ldots, n\}$, where n is the data size, k denotes the kth stage, and $x(k)$ denotes the kth stage value. To explain our procedure more specifically, an example of height measurements is given as $x = \{-6.4, -6.9, -6.9, -8.0, -7.9, -7.7, -7.5\}$ for the following subsections.

2.2 Sequence transformation

The sequence transformation, as illustrated in Fig. 6, comprises four subprocesses: nonnegative and nonzero transformation, AGO, Min–Max normalization, and data inversion.

2.2.1 Nonnegative and nonzero transformation

This initial transformation step ensures that the input series x becomes nonnegative and nonzero, resulting in a new series x' using the equation

$$x'(k) = x(k) - \min(x) + 1, k = 1, 2, ..., n, \tag{1}$$

where $\min(x)$ represents the minimum value in the original series x. Adding 1 ensures that all values in x' are strictly positive. The primary objective of this transformation is to preserve the underlying pattern of the series data after the subsequent AGO, defined in Eq. (2).

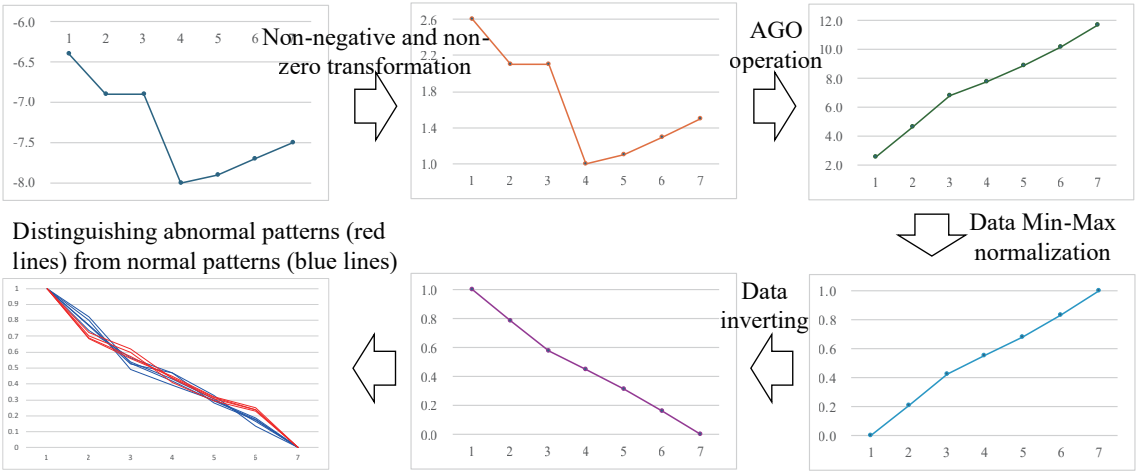

Fig. 6. (Color online) Processes of the proposed sequence transformation.

Consider, for example, the series $x = \{-6.4, -6.9, -6.9, -8, -7.9, -7.7, -7.5\}$. Applying Eq. (1) yields $x' = \{2.60, 2.10, 2.10, 1.00, 1.10, 1.30, 1.50\}$. In contrast, applying only the shift for nonnegativity ($x(k) - \min(x)$) results in $\{1.60, 1.10, 1.10, 0.00, 0.10, 0.30, 0.50\}$.

Applying AGO [Eq. (2)] to these three series, the transformed x', the nonnegative-only series, and the original x, results in $\{2.60, 4.70, 6.80, 7.80, 8.90, 10.20, 11.70\}$, $\{1.60, 2.70, 3.80, 3.80, 3.90, 4.20, 4.70\}$, and $\{6.4, 13.3, 20.2, 28.2, 36.1, 43.8, 51.3\}$, respectively. As illustrated in Fig. 7, the accumulated series derived from x' retains a pattern more similar to the original series x compared with the series generated without adding 1. Directly applying AGO to x also significantly alters the original data pattern. This highlights the importance of the initial transformation in preserving the essential characteristics of the data before AGO.

2.2.2 AGO

The AGO proposed by Deng[5] transforms nonpatterned time series data into data with recognizable patterns that present strictly monotonic increase. The accumulated series $x'' = \{x''(k) | k = 1, 2, \ldots, n\}$ is calculated as

$$x''(k) = \sum_{i=1}^{k} x'(k),\ k = 1, 2, ..., n \text{ and } x''(0) = x'(0). \tag{2}$$

2.2.3 Min–Max normalization

The series x'' is then normalized by the Min–Max method. The new series x''' is obtained as

$$x'''(k) = \frac{x''(k) - \min(x'')}{\max(x'') - \min(x'')},\ k = 1, 2, \ldots, n, \tag{3}$$

where $\max(x'')$ and $\min(x'')$ are the maximum and minimum values of x'', respectively.

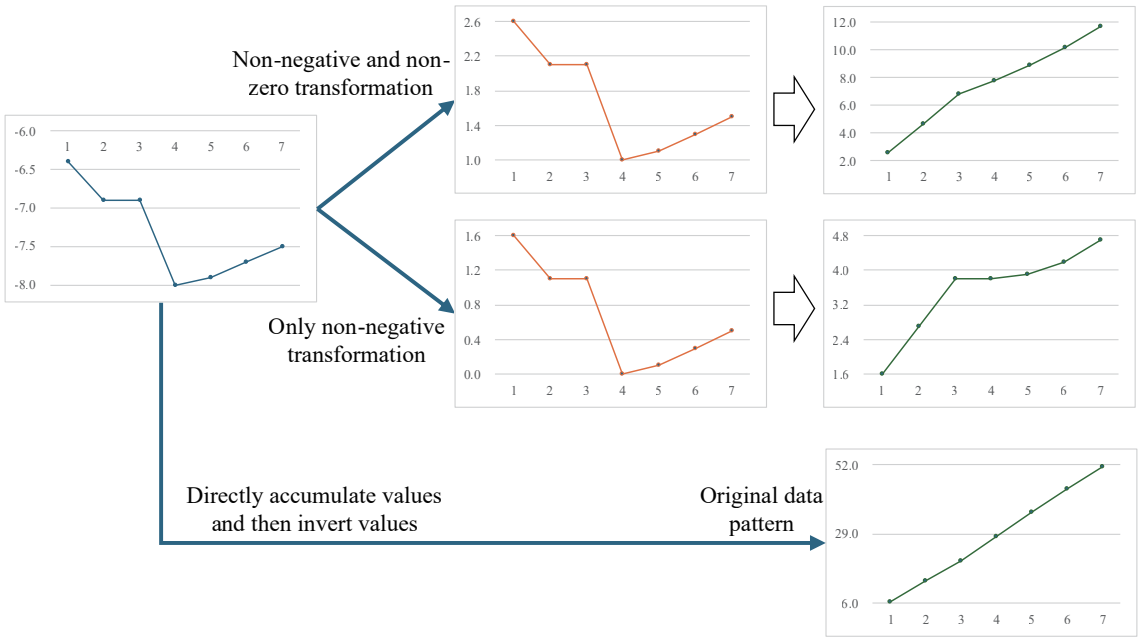

Fig. 7. (Color online) Accumulated series after different initial transformations.

2.2.4 Data inversion

To reduce the computational complexity of the series, the series x''' is further transformed as

$$x''''(k) = 1 - x'''(k), k = 1, 2, \ldots, n. \tag{4}$$

2.2.5 Implementation example

In this subsection, we illustrate the implementation of our sequence transformation on the example series x = {−6.4, −6.9, −6.9, −8.0, −7.9, −7.7, −7.5}, where the minimum value is −8.0. First, applying Eq. (1), the nonnegative and nonzero series x' = {2.6, 2.1, 2.1, 1.0, 1.1, 1.3, 1.5} is obtained. Subsequently, employing AGO using Eq. (2) yields the accumulated series x'' = {2.6, 4.7, 6.8, 7.8, 8.9, 10.2, 11.7}. To ensure consistent scaling for the subsequent model, x″ is then standardized using Min–Max normalization [Eq. (3)], resulting in x''' = {0.000, 0.231, 0.462, 0.571, 0.692, 0.835, 1.000}. Finally, the inverted series x'''' = {1.000, 0.769, 0.538, 0.429, 0.308, 0.165, 0.000} is generated using Eq. (4). The data collected by these steps are summarized in Table 1. It is important to note that each time a new height measurement is recorded, the same procedure will be followed to generate the final series x'''' series as input for our anomaly detection model.

2.3 Data augmentation

In this study, we employ the bootstrap method to address the significant class imbalance caused by the extreme rarity of abnormal height measurements compared with normal ones.

Table 1
Implementation results of the proposed sequence transformation for a given example.

Subsection	Series	Stage						
		1	2	3	4	5	6	7
2.1	x	−6.400	−6.900	−6.900	−8.000	−7.900	−7.700	−7.500
2.2.1	x'	2.600	2.100	2.100	1.000	1.100	1.300	1.500
2.2.2	x''	2.600	4.700	6.800	7.800	8.900	10.200	11.700
2.2.3	x'''	0.000	0.231	0.462	0.571	0.692	0.835	1.000
2.2.4	x''''	1.000	0.769	0.538	0.429	0.308	0.165	0.000

When training data exhibit class imbalance, most algorithms (e.g., C4.5 decision trees) tend to treat minority class instances as acceptable errors to maximize overall accuracy rates, resulting in models unable to effectively identify these critical minority cases.[15] The bootstrap method is a robust resampling technique widely used in statistics and machine learning for estimating population properties, such as the mean, variance, and confidence intervals, based on a single sample. This method proves particularly valuable when the underlying theoretical distributions are complex or unknown, as it avoids relying on distributional assumptions. Instead, the bootstrap method simulates multiple sampling events by repeatedly drawing data points with replacement from our original observed dataset, effectively generating synthetic samples of the minority (abnormality) class to balance the class ratio for improved model training.

2.3 Ensemble learning

In this study, we employ bagging, an ensemble learning technique, to construct an anomaly detection model. Such ensemble learning has been widely applied to various sensor domains.[16,17] Bagging enhances model accuracy and stability by reducing variance through the ensemble averaging of multiple models trained on bootstrapped subsets of the data. It also helps mitigate overfitting during the training process. In this study, BPNNs are used as base learners to capture the nonlinear relationships between time stages. The overall model structure is illustrated in Fig. 5.

3. Empirical Evaluation

In this section, we describe our experimental datasets, environment, and results. All experiments were conducted in Python, following the workflow illustrated in Fig. 8, which consists of five key stages: sequence transformation, abnormal data augmentation, stratified k-fold cross-validation, model training, and performance evaluation.

For implementation, we utilized several functions from the scikit-learn (sklearn) package. The bootstrap method was implemented using sklearn.utils.resample for addressing class imbalance. We employed sklearn.model_selection.StratifiedKFold to generate representative training and testing sets, and sklearn.neural_network.MLPClassifier to construct five BPNNs forming our ensemble learning architecture. Performance evaluation relied on classification accuracy and $F1$-score metrics, calculated using sklearn.metrics.accuracy_score and sklearn.metrics.f1_score, respectively.

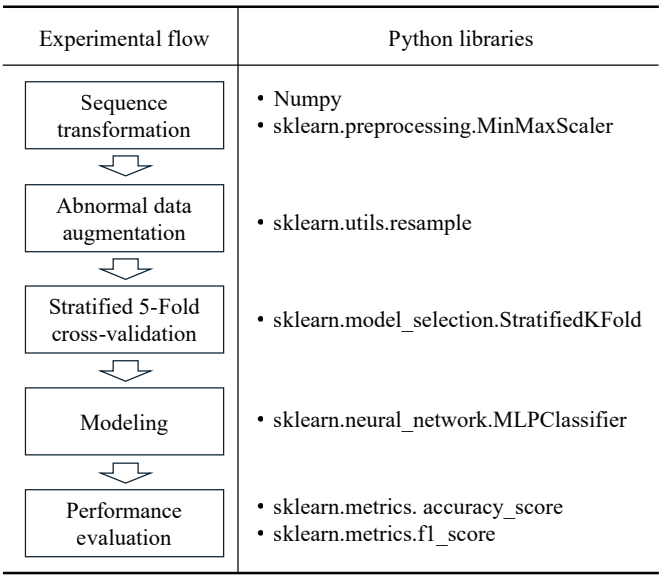

Fig. 8. Experimental workflow: five-stage process and corresponding Python implementations.

3.1 Data preparation

We collected 55389 instances spanning the period from February 1, 2021 to February 28, 2023. The dataset initially comprised 55143 normal instances and 246 abnormal instances, with various data lengths depending on the product types. After applying the sequence transformation described in Sect. 2.2 and removing duplicate instances where values were identical up to three decimal places, the transformed dataset contained 30830 normal instances and 218 abnormal instances (normal-to-abnormal ratio of 141.4:1). Table 2 shows the class distribution based on data length, where "Data length" refers to the number of consecutive outer height measurements recorded before an abnormality occurred. For example, a data length of 3 indicates that an abnormality was detected after recording three outer height measurements.

For model evaluation, we employed the stratified 5-fold cross-validation instead of the traditional *k*-fold cross-validation to address this severe class imbalance. This approach ensures that both training and testing sets contain representative proportions of minority class instances (abnormalities), as the dataset split summarized in Table 3. To further address the imbalance, we applied the bootstrap method to balance class ratios within the training sets only. Table 4 presents the final dataset distribution across all five folds (hereafter referred to as subsets), showing both the original class distribution and the bootstrap-balanced training sets.

3.2 Evaluation metrics

For a binary classification dataset, the model's performance is evaluated using a confusion matrix, as presented in Table 5. On the basis of the four fundamental outcomes, true positive (*TP*), true negative (*TN*), false positive (*FP*), and false negative (*FN*), we compute key

Table 2
Dataset after sequence transform and data cleaning.

Data length	Normality	Abnormality
3	275	42
4	2708	40
5	4492	32
6	7987	50
7	4999	22
8	3012	9
9	2340	6
10	1776	1
11	1467	9
12	1243	6
13	531	1
Total	30830	218

Table 3
Distributions of normal and abnormal instances across stratified 5-fold cross-validation sets.

Fold	Training sets		Testing sets		Total
	Normal	Abnormal	Normal	Abnormal	
1	24664	174	6166	44	31048
2	24664	174	6166	44	31048
3	24664	174	6166	44	31048
4	24664	175	6166	43	31048
5	24664	175	6166	43	31048

Table 4
Class distributions after bootstrap balancing in training sets.

Subset	Training sets		Testing sets		Total
	Normal	Abnormal	Normal	Abnormal	
1	24664	24664	6166	44	55538
2	24664	24664	6166	44	55538
3	24664	24664	6166	44	55538
4	24664	24664	6166	43	55537
5	24664	24664	6166	43	55537

Table 5
Confusion matrix.

		Actual values	
		Positive	Negative
Predicted values	Positive	*TP*	*FP*
	Negative	*FN*	*TN*

performance metrics, including accuracy and $F1$-score, using Eqs. (5) and (6), respectively. Consistent with the problem domain, we designate the abnormality class as the positive class and the normality class as the negative class.

$$Accuracy\ rate = \frac{TP + TN}{TP + FP + FN + TN} \tag{5}$$

$$F1\text{-}score = \frac{2TP}{2TP + FP + FN} \tag{6}$$

3.3 Determining optimal BPNN configuration

In this study, we employed a genetic algorithm (GA) to determine the optimal hyperparameters for BPNNs using the entire dataset presented in Table 2 as the training set.[18] To mitigate class imbalance during hyperparameter tuning, the number of abnormal instances was augmented from 218 to 30830 using the bootstrap method. The MLPClassifier function in scikit-learn contains 23 configurable parameters, such as activation functions and solver types. Notably, certain parameters are solver-dependent; for instance, the 'learning_rate' parameter is only activated when the solver is set to 'sgd', as specified in the scikit-learn documentation.

For the GA-BPNN experiment, parameters were categorized into controllable and fixed types, as summarized in Tables 6 and 7, respectively. GA optimizes the values of controllable parameters on the basis of training accuracy as the fitness function, while fixed parameters remain set to predefined constants or default values. The "hidden_layer_sizes" parameter follows a specific scaling mechanism: when the multiplier value is 1, the three hidden layers contain 12, 24, and 12 nodes, respectively, which correspond to the 12 input features in our

Table 6
Controllable parameters and their search ranges during hyperparameter optimization.

Parameter	Value type	Possible value
hidden_layer_sizes	Integer	[1, 10]
activation	Categories	{'identity', 'logistic', 'tanh', 'relu'}
solvers	Categories	{'sgd', 'adam'}
learning_rate*	Categories	{'constant', 'invscaling', 'adaptive'}
learning_rate_init	Float	[0.001, 0.005]
random_state	Integer	[0, 50]
momentum*	Float	[0.50, 0.99]

*Only used when solver = 'sgd'.

Table 7
Fixed parameters and their values during hyperparameter optimization.

Parameter	Value	Parameter	Value
alpha	0.0001	nesterovs_momentum	True
batch_size	200	early_stopping	True
power_t	0.5	validation_fraction	0.1
max_iter	200	beta_1	0.9
shuffle	True	beta_2	0.999
tol	0.0004	epsilon	1×10^{-8}
warm_start	True	n_iter_no_change	10

BPNN architecture. This scaling approach maintains architectural consistency while allowing GA to explore different network capacities. The network architecture maintains a fixed three-hidden-layer structure throughout all experiments. The experimental results of the GA-BPNN is summarized in Table 8.

3.4 Experimental results

The individual testing results of the five BPNNs, obtained using the five pairs of training and testing sets (detailed in Table 4) and the two optimized BPNN hyperparameter settings (specified in Table 8), are summarized in Table 9. Notably, BPNN #1 refers to the BPNN trained on the first training subset with the second hyperparameter configuration and so forth. While the individual BPNNs exhibited high accuracy, ranging from 98.183 to 98.985%, their corresponding $F1$-scores were considerably lower, falling within the range of [0.175, 0.222]. To illustrate the impact of class imbalance, a naive classifier that predicted all testing instances as belonging to the majority class (normality) would achieve an accuracy of 99.291 or 99.307% (depending on the specific fold's class distribution) but would yield an $F1$-score of zero, highlighting its inability to identify abnormal instances. These initial results underscore the critical role of the bootstrap method in addressing the inherent class imbalance within the data. Furthermore, the application of the ensemble learning technique resulted in substantial improvements in both overall accuracy and $F1$-score, reaching 99.804% and 0.871, respectively. These experimental findings demonstrate the robustness and effectiveness of the proposed method for learning from variable-length short-term time series data in the presence of significant class imbalance.

4. Further Discussion

To further evaluate the generalization capability and robustness of the proposed method to newly encountered abnormal instances, we collected an additional 73 abnormal samples spanning the period from March 1, 2023 to June 30, 2023, to constitute an independent testing set. The initial classification accuracy of the pretrained model on these previously unseen abnormal instances was 75.343%, a performance level deemed insufficient for practical application. However, upon applying a continual learning strategy to adapt the model to this new data, the classification accuracy improved significantly to 97.260%. The detailed results of this evaluation on the new testing set are summarized in Table 10.

Table 8
Final optimized BPNN hyperparameter settings.

Parameter	Value
hidden_layer_sizes	(120, 240, 120)
activation	'tanh'
solvers	sgd
learning_rate	'adaptive'
learning_rate_init	0.001
random_state	42
momentum	0.9

Table 9
Testing results in ensemble learning.

Model	Subset No.	Accuracy (%)	*F*1-score
BPNN #1	1	98.183	0.175
BPNN #2	2	98.729	0.218
BPNN #3	3	98.583	0.200
BPNN #4	4	98.776	0.191
BPNN #5	5	98.985	0.222
Ensemble result		99.804	0.871

Table 10
Evaluation results of new testing data.

Data length	Abnormality	Number of correct classifications	Number of correct classifications after continual learning
3	16	16	16
4	13	13	13
5	3	3	3
6	22	17	22
7	6	4	6
8	1	1	1
9	0	0	0
10	3	1	3
11	1	0	1
12	8	0	6
13	0	0	0
Total	73	55 (75.343%)	71 (97.260%)

5. Conclusions

In this study, we addressed the anomaly detection challenge in wafer grinding processes based on height measurements obtained from IPGs. These height measurements manifest as variable-length, short-term time series data. To tackle this issue, we developed a novel sequence transformation method employing AGO. This transformation converts the time series data into a more pattern-recognizable format, facilitating the application of ensemble learning for effective anomaly detection. Experimental results demonstrated the superiority of the proposed method over the direct analysis of the raw time series data, achieving a significant improvement in *F*1-score from 0.201 to 0.871. This outcome highlights the efficacy of the AGO transformation in revealing underlying patterns indicative of abnormal grinding behavior. Furthermore, we established the importance of addressing class imbalance, effectively achieved through the bootstrap method, and the benefits of employing an ensemble of BPNNs to enhance robustness and generalization. The analysis of solver and activation function combinations revealed that tanh activation, paired with both sgd and adam solvers, yielded optimal performance for the BPNN architecture. The method's ability to maintain high performance in the presence of noisy sensor data and its adaptability to new abnormal instances through continual learning further underscore its potential for practical implementation in real-world wafer grinding operations. Future research should explore the real-time implementation and optimization of this method, as well as its applicability to other manufacturing processes characterized by similar data challenges.

References

1 W. Yao, R. Kang, X. Guo, and X. Zhu: J. Mater. Process. Technol. **299** (2022) 117390. https://doi.org/10.1016/j.jmatprotec.2021.117390
2 H. Tao, Y. Liu, D. Zhao, and X. Lu: Int. J. Mech. Sci. **258** (2023) 108530. https://doi.org/10.1016/j.ijmecsci.2023.108530
3 H. Tao, Y. Liu, D. Zhao, and X. Lu: Int. J. Mech. Sci. **222** (2022) 107240. https://doi.org/10.1016/j.ijmecsci.2022.107240
4 S. Hochreiter and J. Schmidhuber: Neural Comput. **9** (1997) 1735. https://doi.org/10.1162/neco.1997.9.8.1735
5 J.-L. Deng: Syst. Control Lett. **1** (1982) 288. https://doi.org/10.1016/S0167-6911(82)80025-X
6 Y.-C. Hu: Grey Syst. Theory Appl. **13** (2023) 808. https://doi.org/10.1108/GS-04-2023-0037
7 Y. Ren, Y. Wang, L. Xia, W. Liu, and R. Tao: Grey Syst. Theory Appl. **14** (2024) 671. https://doi.org/10.1108/GS-01-2024-0005
8 Y. Li, C. Wang, and J. Liu: Grey Syst. Theory Appl. **14** (2024) 621. https://doi.org/10.1108/GS-12-2023-0117
9 F. E. Sapnken, B. S. Diboma, A. Khalili Tazehkandgheshlagh, M. Hamaidi, P. G. Noumo, Y. Wang, and J. G. Tamba: Grey Syst. Theory Appl. **14** (2024) 708. https://doi.org/10.1108/GS-01-2024-0011
10 X. Xie and L. Liu: Energy Sources Part A **45** (2023) 9894. https://doi.org/10.1080/15567036.2023.2240264
11 C.-C. Chen and C. M. Tsai: Grey Syst. Theory Appl. **15** (2025) 602. https://doi.org/10.1108/GS-11-2024-0135
12 Y. Chen, W. Lifeng, L. Lianyi, and Z. Kai: Chaos, Solitons Fractals **138** (2020) 109915. https://doi.org/10.1016/j.chaos.2020.109915
13 B. Efron and R. J. Tibshirani: An Introduction to the Bootstrap (New York, Chapmen & Hall, 1994) 1st ed., Chap. 1. https://doi.org/10.1201/9780429246593
14 L. Breiman: Mach. Learn. **24** (1996) 123. https://doi.org/10.1007/BF00058655
15 C. C. Chen, Y. S. Lin, and H. Y. Chen: Sens. Mater. **36** (2024) 4835. https://doi.org/https://doi.org/10.18494/SAM5224
16 P. Chanchotisatien and C. Vong: Sens. Mater. **33** (2021) 4245. https://doi.org/https://doi.org/10.18494/SAM.2021.3481
17 M. Saito and K. Fujinami: Sens. Mater. **32** (2020) 27. https://doi.org/10.18494/SAM.2020.2585
18 Y. Wu, D. Wu, M. Fei, H. Sørensen, Y. Ren, and J. Mou: Eng. Appl. Artif. **119** (2023) 105738. https://doi.org/10.1016/j.engappai.2022.105738

About the Authors

Chien-Chih Chen is an assistant professor in the Department of Information Management of National Chin-Yi University of Technology, Taiwan. His current interests are focused on machine learning with small data sets. His articles have appeared in Decision Support Systems, Omega, Automation in Construction, Computers and Industrial Engineering, International Journal of Production Research, Neurocomputing, and other publications. (frick@ncut.edu.tw)

Yao-San Lin is an associate professor in the Department of Industrial Engineering and Management of National Chin-Yi University of Technology, Taiwan, and has taught statistical analysis, information management and machine learning for many years. In the past ten years of research, he has twice won the IEEE Best Paper Award (2019, 2013), and his research results have been published in the following international academic journals: Applied Soft Computing, Decision Support Systems, Neurocomputing, European Journal of Operational Research, International Journal of Production Research, Computers & Operations Research, and Expert Systems with Applications. His current research interests are mainly in the field of manufacturing and natural language processing. (yslin@ncut.edu.tw)

Hung-Yu Chen is an assistant professor in the Department of Information Management of National Chin-Yi University of Technology, Taiwan. His current interests are concentrated on human behavior with misinformation. His articles have been published in Decision Support Systems, Neurocomputing, International Journal of Advanced Manufacturing Technology, and other publications. (chenhy@ncut.edu.tw)

Wei-Yuan Sun is a graduate student in the Department of Information Management at National Chin-Yi University of Technology, Taiwan. His research interests encompass machine learning and artificial intelligence, with particular emphasis on their applications in industrial process optimization and data analytics. (4B332005@gm.student.ncut.edu.tw)

Chih-Jung Kuo is an undergraduate student in the Department of Information Management at National Chin-Yi University of Technology, Taiwan. Her research interests encompass big data analytics and heuristic optimization algorithms, with a particular focus on their applications in industrial process optimization. (kitty930211@gmail.com)