Original Research Article

SCIENCE PROGRESS

Science Progress
2024, Vol. 107(1) 1–30

DOI: 10.1177/00368504241238612
journals.sagepub.com/home/sci

Production line balance problem identification and improvement based on decision tree: A case study of commercial air conditioner production line

Rui Wang, Tengyuan Xin, Shun Jia, Dawei Ren and Meiyan Li

Department of Industrial Engineering, College of Energy and Mining Engineering, Shangdong University of Science and Technology, Qingdao, China

Abstract

In the production of air conditioners, there are various issues such as complex requirements, redundant stations, excessive man-hours, and low production line balance rate. This paper aims to address these problems by analyzing the historical data of H Company's commercial air conditioner production line. The data is categorized into five aspects: station, working hours, standard working hours, labor capacity, and presence of bottleneck processes. To optimize and improve the second production line, this paper applies the production line balance management method based on data mining. It utilizes the decision tree model in data mining and incorporates lean production knowledge from industrial engineering. The goal is to identify crucial factors that affect the balance of the production line and address the issues caused by these factors. The aim is to reduce and eliminate redundant working hours and enhance the balance rate of the production line. By implementing the approach outlined in this paper, the bottleneck time of the second production line was reduced from 96.67 s to 74.6 s, and the production line balance rate increased from 68% to 85%.

Keywords

Lean production, production line balance, decision tree, bottleneck problem improvement, data mining

Corresponding author:

Shun Jia, Department of Industrial Engineering, College of Energy and Mining Engineering, Shangdong University of Science and Technology, Qingdao, 266590, China.
Email: herojiashun@163.com

Introduction

China's manufacturing industry faces challenges such as long working hours, high operational intensity, and manual labor reliance, resulting in low production efficiency and inconsistent product quality. With the implementation of "Made in China 2025" and a focus on lean production, enterprises are striving to reduce costs, shorten production cycles, and enhance efficiency for flexible production. Achieving a balanced production line is essential for lean production and is a pressing issue for enterprises looking to optimize equipment and labor utilization while eliminating inefficiencies and imbalances.

This research aims to enhance the efficiency of H Company's commercial air conditioner's second production line by integrating a decision tree model and lean production techniques. The primary goal is to pinpoint the factors influencing production line balance, identify critical parameters, and streamline operations to reduce unnecessary working hours and boost the balance rate. In this investigation, we have opted to utilize the decision tree model C4.5.[1] C4.5, an extension of Quinlan's ID3 algorithm, is adept at handling both discrete and continuous attribute types. Data collection pertaining to the commercial air conditioner production line is straightforward, and the C4.5 algorithm does not necessitate a vast amount of data. The C4.5 algorithm is chosen for its interpretability, high accuracy, and its advantage over other commonly used machine learning algorithms. The method framework is illustrated in Figure 1.

The rest of the article is organized as follows: We discuss the related work in Section 2. In Section 3, we introduce the situation of the production line. In Section 4, we introduce the method of constructing the production line balance decision tree model to find the bottleneck process. In Section 5, we describe production line balance improvement plan. Finally, we conclude the paper in Section 6.

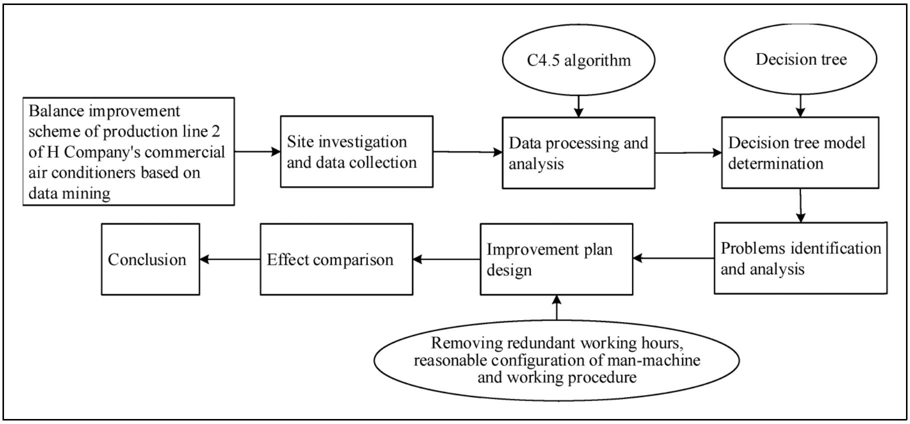


Figure 1. Method framework for problem identification using decision tree model and problem improvement using industrial engineering method.

Related work

Assembly line production has gradually entered the manufacturing industry, bringing about improved production efficiency. However, assembly line production also has its shortcomings. To address these shortcomings, scholars have conducted research on production line balancing and developed the production line balance theory. Production line balance is highly significant in enterprise production as it effectively allocates resources and enhances production efficiency. Extensive studies have been conducted in this area. Salveson[2] established a model on production line balance using the "convergence process method" and solved it analytically to intuitively address the line balancing problem. Helgeson and Birnie[3] employed the rank method to optimize assembly lines. Gutjahr and Nemhauser[4] proposed the network model method, which solves the fewest stations problem by solving the shortest path of the network. Roberts and Villa[5] introduced the concept of the joint priority map to tackle the imbalance problem in mixed production lines, significantly enhancing production efficiency. Bennett and Byrd[6] applied a two-stage heuristic algorithm to further optimize production lines. Dar-El and Rubinovitch[7] proposed a branch-and-bound algorithm, which effectively optimizes production lines. Agrawal[8] introduced the "maximum set principle" for efficient allocation of job elements. Erel[9] utilized 0–1 integer programming model for optimizing mixed production lines, conducting an in-depth analysis of the production line balance problem. Özcan and Toklu[10] proposed the tabu search algorithm and introduced evaluation criteria, namely, the production line balance rate and the smoothing index, as the two main indicators. Huang et al.[11] modified the model by incorporating three types of valid inequalities. Güden and Meral[12] developed the COMSOAL algorithm and ASA algorithm to address the deterministic assembly line balance problem in a multi-objective multi-product model environment. Bortolini et al.[13] proposed a method to appropriately balance the assembly line using different objective functions as features. Krenczyk et al.[14] introduced the Flex Sim simulation method, which utilizes computers to achieve production line balance. Oksuz et al.[15] proposed a U-shaped nonlinear mathematical model and utilized the artificial bee colony algorithm and genetic algorithm to achieve better balance in production lines. Pirogov et al.[16] studied the problem of uncertain process processing time by initially assigning a given set of tasks to modules and subsequently assigning modules to machines to determine the most robust line configuration. Sikora[17] addressed the assembly line balancing problem by employing Benders decomposition and considering uncertain demands. Pinhão et al.[18] proposed models for optimizing indicators such as turnaround time, number of operators, and load balancing in aircraft engine assembly lines. Teshome et al.[19] aimed to balance the garment line of the polo shirt operation using Arena simulation software.

China, being a big manufacturing country with a large number of small and medium-sized enterprises, faces various production line problems. These issues result in low production efficiency, significant resource wastage, and further exacerbate the challenges faced by businesses. Hence, research on production line balancing is urgently required. Li et al.[20] proposed the maximum time priority method to solve the optimal path. Chen et al.[21] introduced an adaptive genetic algorithm based on the eel selection method, which exhibits good periodicity. Song and Han[22] utilized the tabu search algorithm to simultaneously optimize multiple balance problems in the production line. Lu et al.[23] improved the operation of the production line by integrating relevant industrial engineering technologies. Sun et al.[24] proposed the "5S" management method and visual management method, strengthening the production management of enterprises and providing a solid theoretical foundation for production line management. Guo et al.[25] proposed an operation measurement technology to enhance production line balance from the perspective of more effective management and research on operation time. Zhang et al.[26] enhanced the original genetic algorithm for better analysis of the imbalance problem. Wu et al.[27] employed the branch-and-bound algorithm to study work order and station efficiency, proposing the concept of equal-slave production line balance, and summarizing the design process of standard operation. Niu et al.[28] centered their research on lean production and employed methods such as witness simulation and action analysis to address related problems. Li and Meng[29] utilized the MOD method to determine the standard operating time of the process and optimize the layout of the production line, thus enhancing the utilization efficiency. Wang and Li[30] introduced the Benders decomposition algorithm and provided three datasets to test its performance in dealing with uncertainty in mixed-model assembly lines. Zhang et al.[31] investigated the hybrid model U-shaped robot assembly line balance sorting problem, considering energy constraints, and proposed a hybrid multi-objective listening algorithm. Jia et al.[32] presented a novel Therblig-embedded Value Stream Mapping method, which, through the implementation of the Future-State-Map, improves time and energy efficiencies without compromising machining quality. Jia et al.[33] developed a drilling process energy modeling and multi-angle energy visualization analysis method based on the fractional power model, effectively enhancing time and energy efficiency. Huang and Deng[34] conducted a systematic analysis and optimization of the assembly production line of Company A, proposed improvement measures, and evaluated them through simulation software. Chang et al.[35] conducted a study on the balance optimization of the production line in the needle car workshop of J Company using Flexsim simulation and industrial engineering methods.

The research of this paper is based on the actual production line, which is different from most of the theoretical research which mainly focuses on simulation technology. We introduce the decision tree model in data mining into the actual production environment, and use the lean production knowledge in the field of industrial engineering to conduct in-depth research and improvement with the second production line of H Company as a case. Our method is more direct and practical, with the support of real data, a comprehensive review of the production line. It has been proved that our method has achieved remarkable results in solving on-site problems and improving production efficiency, which provides strong support for decision-making in actual production. This paper combines the knowledge of data mining and industrial engineering, and emphasizes the feasibility and effect of practical application. Through the application in the second production line of H Company, we verify the effectiveness of this method in practical scenarios, and provide innovative ideas for production line optimization and problem solving.

Production line situation

The final assembly line of the production line 2 of H Company's commercial air conditioner is divided into three parts: final assembly, pre-assembly, and testing, comprising a

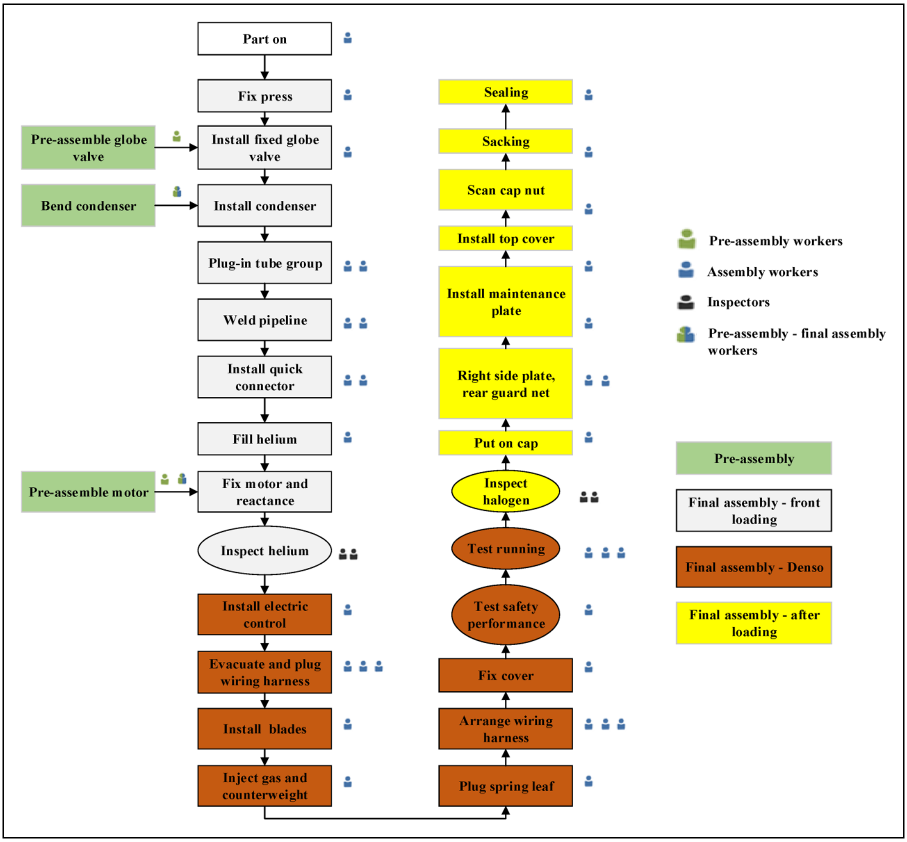


Figure 2. Production flow of the second production line of H Company's commercial air conditioner.

total of 29 stations. The pre-assembly line is located alongside the final assembly line. The production flow is visually represented in Figure 2.

The production line consists of a total of 41 employees, with the distribution as follows: 34 general assembly workers, 4 inspectors, 2 pre-assembly workers, and 2 workers who alternate between the pre-assembly line and the final assembly line. The production operates for a single shift of 10 h. Certain tasks within the production process, such as welding pipelines, electronic control installation, evacuation, gas injection, and safety performance testing, require special qualifications.

The second production line underwent improvements on December 23, 2021, with the production data outlined in Figure 3. The hourly output currently stands at approximately 32 sets, with a production cycle of 112.5 s. However, the production line is plagued by issues such as a high shutdown rate, unnecessary waiting times, suboptimal human-machine coordination, and inconsistent output speeds at "non-shaped" unit stations.

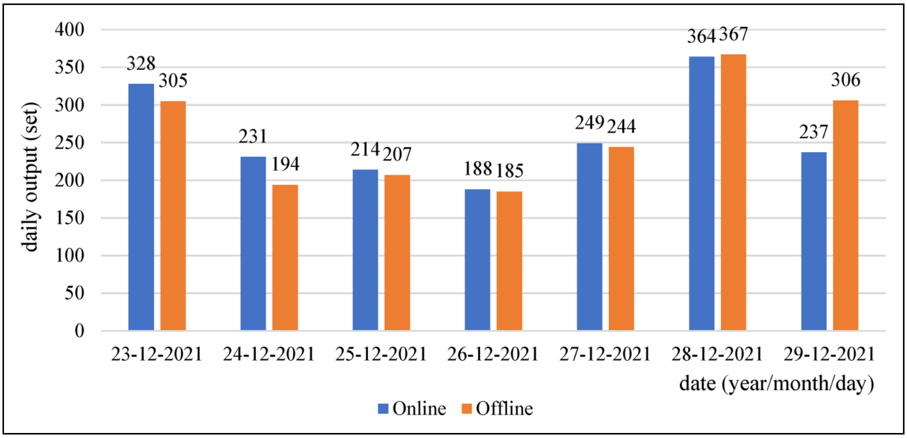


Figure 3. Daily output comparison before and after an improvement.

Problem identification based on decision tree

Method overview

The external commercial air conditioner production line of H Company has a low balance rate, which leads to various issues in the bottleneck process. In order to identify and address the problems within each process of the complex production line, targeted optimization and improvements need to be implemented. This requires the calculation and analysis of historical production line data, as well as in-depth field research to study the production processes of each stage. However, this traditional research method is time-consuming, inefficient, and lacks systematization. To overcome these challenges and enhance the efficiency and systematic approach to production line improvements, this paper aims to leverage machine learning techniques to swiftly identify bottleneck processes and classify existing problems within them.

Before conducting data analysis, it is crucial to collect original data. The historical data should include the name of each station, whether the work is complex, the personnel quota N, and the measurement $T_{x,m}$ representing the working hours measured for the m times at the x station (working hours were measured five times at each station in this study). Complex work refers to stations that require special qualification requirements or man-machine cooperation. In addition to the essential attributes, the training dataset should involve field research to analyze the issues present at each station. The identified problems should be incorporated as target attributes, including issues such as unreasonable working hour allocation and personnel control problem. The data metrics for station load rate l, time difference index t, and variance sq of each station can be calculated as follows:

$$T_x = \frac{\sum_{i=1}^{m} T_{x,i}}{m} \tag{1}$$

$$l = \frac{T_x}{T_{max}} \times 100\% \tag{2}$$

$$sq = \frac{\sum_{i=1}^{m} (T_{x,i} - T_x)^2}{m} \tag{3}$$

$$t_x = \sum_{i=-1}^{1} \max[0, T_x - T_{x+\mathrm{i}}] + \max[0,\ T_x - T_{x-2}] *0.8 + \max[0,\ T_x - T_{x+2}]*0.8 \tag{4}$$

In Equations (1)–(4), $T_{x,i}$ represents the working hours measured for the i -th time at station x. Sum the m measured working hours of station x and calculate the average value T_x, representing the average measurement time at that station. The average working hours for each station are then calculated as the standard working hours of the production line. T_{max} represents the maximum value of the measured working hours among all stations. A station is identified as a bottleneck station if its working hours exceed the standard working hours. The load rate l is calculated and compared to the standard load rate to identify bottleneck processes. Stations with a load rate in the dataset exceeding the standard load rate are considered bottleneck stations that require optimization. The bottleneck processes in the sample data are sorted out, generating four features: personnel quota, whether complex work, time difference index, and variance. Three target attributes are identified: unreasonable working hour allocation, personnel control, and other problems. A decision tree model is established and trained using historical data. The decision tree rules are then applied to classify the stations and identify problems existing at each bottleneck station.

Problem identification model

The primary goal of data analysis is to effectively utilize historical data generated during the production process to establish a model for analyzing production line balance. In this study, the C4.5 algorithm is utilized to develop a comprehensive decision tree model for production line balance. The historical data from H Company's second production line is analyzed using this model to uncover the impact of production process parameters on production line balance. By revealing hidden patterns within the data and providing feedback to the production process, this analysis aims to offer decision support to enterprises in their ongoing efforts to optimize production line balance.

Let S be the training sample set, comprising samples from n categories denoted as $C_1, C_2, \ldots, C_n$. The entropy (expected information) of S can be defined as:

$$E(S) = -\sum_{i=0}^{n} p_i \log_2 p_i \tag{5}$$

In Equation (5), p_i represents the probability of class C_i. Viewing the n types of training samples in S as distinct information categories, the entropy of S signifies the average number of bits needed to encode each type of information. The product of S and $E(S)$ denotes the total bits required to encode S, with $|S|$ representing the sample size. A higher entropy indicates a more balanced probability distribution and increased sample set diversity. Thus, entropy serves as a measure of the training set's impurity. The decision tree's branching principle aims to maximize purity within sample subsets, minimizing entropy to achieve this goal.

Assume that attribute A divides S into m parts, denoted by $Q_1, Q_2, \ldots, Q_m$. The characteristics of attribute Q divide it into k parts, respectively denoted by $L_1, L_2, \ldots, L_k$, then the entropy calculation method of Q_i is:

$$E(Q) = -\sum_{i=1}^{k} q_i \log_2 q_i \tag{6}$$

In Equation (6), q_i represents the probability of L_i. The information entropy of attribute A can be calculated as:

$$E(A) = \sum_{i=1}^{m} w_i \, E(Q_i) \tag{7}$$

In Equation (7), w_i represents the probability of Q_i. The information gain is utilized to quantify the expected reduction in entropy. The information gain derived from the partitioning of attribute A into S is:

$$G(S, \; A) = E(S) - E(A) \tag{8}$$

Calculate attribute splitting information metrics:

$$H(A) = -\sum_{i=1}^{m} w_i \log_2 w_i \tag{9}$$

The gain ratio, $GR(S, A)$, can be calculated using Equation (8) and Equation (9):

$$GR(S, \; A) = G(S, \; A) \,/\, H(A) \tag{10}$$

The gain ratio is calculated as the ratio of information gains to segment information. The information gain and gain ratio are computed for each attribute iteratively. The attribute with the highest gain ratio is selected as the root node of the decision tree. Subsequently, the values of each attribute of the root node are expanded recursively to form the decision tree structure.

The decision tree is a supervised classification technique in data mining that infers classification rules based on a set of unordered and unorganized examples. This method is widely used in various classification problems, such as network traffic analysis and quality evaluation, due to its generalizability and comprehensibility. The C4.5 algorithm, known for its high classification accuracy and speed, utilizes the information gain ratio to select attributes. Additionally, it can handle both discrete and continuous attributes, as well as incomplete data.

Accuracy measures the overall correctness of the model's predictions by calculating the ratio of correctly predicted instances to the total number of instances.

Precision quantifies the model's ability to correctly classify positive instances out of the total instances it predicts as positive. It is calculated as the ratio of true positives to the sum of true positives and false positives.

Recall represents the model's ability to correctly identify positive instances from the actual positive instances. It is calculated as the ratio of true positives to the sum of true positives and false negatives.

F1 score is a balanced measure that combines both precision and recall. It is calculated as the weighted harmonic mean of precision and recall, providing a single score that represents the model's overall performance.

Based on the evaluation metrics, we compare the performance of different models as shown in Table 1. The C4.5 decision tree model outperforms the other models in all evaluation metrics. It has the highest accuracy (94%) and precision (96%), while also achieving a high recall (90%) and F1 score (92%). In comparison, the softmax regression model has lower accuracy (72%) and F1 score (69%), while the random forests model has slightly higher accuracy (87%) and F1 score (83%). The support vector machines and K-nearest neighbors models both have accuracy and F1 scores above 80%, and the Naive Bayes model performs well in terms of accuracy (85%) and F1 score (80%), although its recall (76%) is slightly lower. Taking all these metrics into consideration, we choose the C4.5 decision tree model as the preferred choice due to its consistently high performance across all metrics. It has high accuracy and precision, good recall, and a high F1 score. Therefore, the C4.5 Decision Tree model can provide more reliable and accurate predictions.

Problem identification of bottleneck processes

Data collection is conducted for the commercial air conditioner production line of H Company. The sample data of production line 2 is used as the training data set, as shown in Table 2. The data collection focuses on production line 2 and includes various inputs such as station, complex work judgment, man hour measurement, and personnel quota. Additionally, three target attributes are included: unreasonable working hour allocation, personnel control, and other problems. If a station has no problems or does not belong to a bottleneck process, it is indicated with a "\". The collected data reveals that the average working hour for the production line is 78.22 s, with a

Table 1. Model performance comparison.

Evaluation metrics Model	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
C4.5 Decision Tree	94	96	90	92
Softmax Regression [36]	72	83	65	69
Random Forests [37]	87	81	85	83
Support Vector Machines [38]	80	83	78	79
K-Nearest Neighbors [39]	83	85	81	82
Naive Bayes Model [40]	85	82	76	80

Table 2. Sample data collected from the commercial air conditioner production line of H Company.

Station	Working hour 1	Working hour 2	Working hour 3	Working hour 4	Working hour 5	Personnel quota	Whether complex work	Possible problems
Part on	68.48	67.72	74.15	74.64	74.23	1	No	\
Fix press	42.32	33.36	46.12	42.18	45.82	1	No	\
Bend	46.62	45.27	48.84	48.52	48.34	2	No	\
Install gas-liquid separator and globe valve	87.13	90.12	78.72	97.34	100.62	3	No	unreasonable working hour allocation
Fix condenser and stick felt	39.45	43.54	44.48	36.83	39.13	1	No	\
Intubation	91.36	87.62	93.54	89.48	84.57	2	No	unreasonable working hour allocation
Put wet cloth	24.34	33.36	29.75	32.38	41.93	1	No	\
Welding 1	83.32	85.24	81.98	84.75	90.83	1	Yes	personnel control problem
Welding 2	79.34	82.12	85.38	86.85	76.74	1	Yes	personnel control problem
Inject gas	55.53	54.75	57.74	55.82	61.37	1	No	\
Install quick connector	146.38	107.13	112.16	113.59	138.49	1	No	unreasonable working hour allocation
Fill helium	75.27	73.54	83.13	76.02	79.36	1	No	personnel control problem
Fix motor	133.19	142.55	128.14	138.37	140.12	2	No	unreasonable working hour allocation
Inspect helium	48.53	74.36	75.88	49.28	60.92	1	Yes	\
Install electric control	46.25	45.73	52.11	49.42	54.24	1	Yes	\
Evacuate	146.25	145.73	152.11	149.42	144.24	4	No	other problem
Inject gas	55.64	54.71	57.91	55.82	56.36	1	No	personnel control problem
Install blades	44.74	39.82	40.83	54.62	67.17	1	No	\
Plug sensor	57.36	68.15	64.73	61.75	71.23	1	Yes	

(Continued)

Table 2. (continued)

Station	Working hour 1	Working hour 2	Working hour 3	Working hour 4	Working hour 5	Personnel quota	Whether complex work	Possible problems
Part on	68.48	67.72	74.15	74.64	74.23	1	No	\
								personnel control problem
Plug spring leaf	57.36	68.15	64.73	61.75	71.23	1	No	\
Arrange wiring harness 1	83.14	75.75	75.46	94.94	87.37	1	Yes	personnel control problem
Arrange wiring harness 2	94.23	103.23	87.73	89.32	93.37	1	Yes	personnel control problem
Arrange wiring harness 3	90.52	94.14	89.25	85.17	81.45	1	Yes	personnel control problem
Fix cover	80.15	87.81	73.82	77.42	91.41	2	No	other problem
Test safety performance	70.31	58.41	61.51	70.53	81.96	1	No	\
Inspect halogen	67.16	80.134	74.21	77.32	69.41	2	Yes	personnel control problem
Fix right side plate	73.43	81.12	82.21	88.41	82.28	1	No	unreasonable working hour allocation
Test running	88.41	91.23	98.12	94.21	93.51	4	Yes	other problem
Put on cap	92.21	88.64	85.23	81.24	84.32	1	No	personnel control problem
Fix nut, install rear net, install side plate	123.56	102.42	99.12	105.35	117.8	3	No	unreasonable working hour allocation
Install front cover and side panel	73.43	81.12	82.21	88.41	82.28	1	No	other problem
Fix top cover	72.4	73.53	83.23	80.24	78.13	1	No	\
Paste relevant certificate	63.45	68.74	64.12	59.19	62.29	1	No	\
Sacking	58.21	53.31	51.41	61.16	71.51	1	No	\
Sealing	63.45	68.74	64.12	59.19	62.29	1	No	\

minimum of 33.36 s and a maximum of 152.11 s. The median working hour is 74.23 s. The average personnel quota for the production line is 1.59 s. Further analysis of the data identifies 9 workstations with complex work and 13 workstations that may have potential issues. This sample data set will serve as the training data set for the decision tree-based production line balance analysis. By organizing and analyzing this data, H Company aims to optimize the production line's efficiency and identify any bottlenecks or potential issues that may arise.

After performing pre-calculation, we obtain the time difference index, variance, and station load rate for each station. The standard working hours of the production line are calculated to be 72.31 s, allowing us to identify the bottleneck station. Using the traditional improvement method, we identify potential issues within the station, such as unreasonable working hour allocation, personnel control problems, and other problems. These problems are recorded as target attributes. Table 3 displays the station attributes for production line 2.

The C4.5 algorithm is employed to perform classification calculations on the station attributes of production line 2, as depicted in Table 3. This process results in the generation of a decision tree model. The target attribute of the training set is the potential problems that may arise on the production line. The attribute values consist of working stations, working hours, standard working hours, enterprise complement, and the presence of a bottleneck process. The information entropy of these attributes is as follows:

$$H\ (possible\ problem) = 1.4197367178034823$$

$$E(Station) = 0.0$$

$$E(Time\ difference\ index) = 0.0$$

$$E(Variance) = 0.0$$

$$E(Personnel\ quota) = 1.2202816797100022$$

$$E(Whether\ complex\ work) = 0.6844674843612774$$

Using Equation (8), the information gain of each attribute can be obtained as:

$$G\ (S,\ station) = 0.0$$

$$G(S,\ Time\ difference\ index) = 0.0$$

$$G(S,\ Variance) = 0.0$$

$$G(S,\ Personnel\ quota) = 0.19945503809348009$$

$$G(S,\ Whether\ complex\ work) = 0.735269233442205$$

Equation (10) is used to calculate the gain ratio of each attribute:

$$GR(S,\ Station) = 0.2839473436$$

$$GR(S,\ Time\ difference\ index) = 0.2839473436$$

Table 3. Station attributes of production line 2 obtained by pre-calculation.

Station	Time difference index	Variance	Personnel quota	Whether complex work	Possible problems
Install gas-liquid separator and globe valve	133.61	7.73	1	No	unreasonable working hour allocation
Intubation	108.86	3.08	1	No	unreasonable working hour allocation
Welding 1	78.56	3.03	2	Yes	personnel control problem
Welding 2	64.83	3.73	3	Yes	personnel control problem
Install quick connector 1	179.75	15.77	1	No	unreasonable working hour allocation
Fill helium	56.74	3.41	1	Yes	personnel control problem
Fix motor and reactance	89.32	7.46	2	No	unreasonable working hour allocation
Evacuate	241.27	9.76	4	No	other problem
Inject gas	98.36	3.38	1	Yes	personnel control problem
Plug sensor	39.25	9.23	1	Yes	personnel control problem
Arrange wiring harness 1	45.80	7.36	1	Yes	personnel control problem
Arrange wiring harness 2	48.02	5.4	1	Yes	personnel control problem
Arrange wiring harness 3	25.45	4.39	1	Yes	personnel control problem
Fix cover	46.64	6.54	2	No	other problem
Inspect halogen	5.10	4.82	2	Yes	personnel control problem
Test running	45.86	3.23	4	Yes	other problem
Put on cap	14.06	3.77	1	No	personnel control problem
Fix nut, install rear net and side plate	90.44	9.4	1	No	unreasonable working hour allocation
Install front cover and side panel	18.34	4.78	1	No	other problem
Fix right side panel	98.86	8.53	1	No	unreasonable working hour allocation

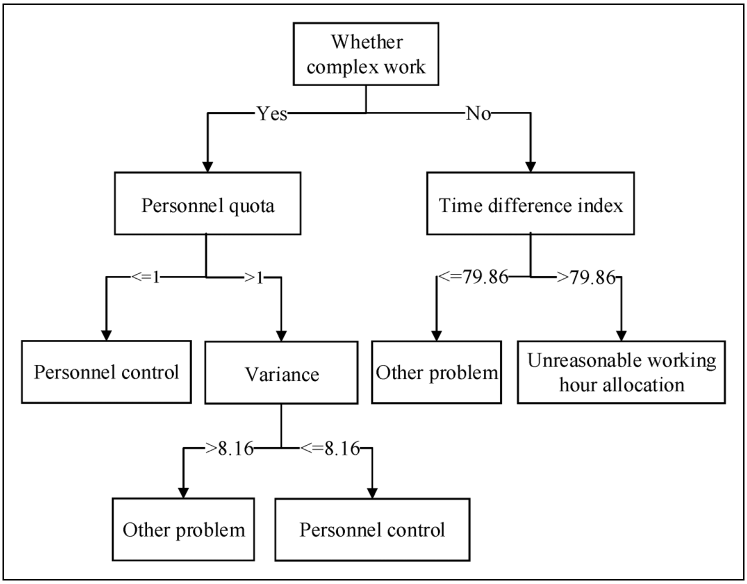


Figure 4. Decision tree for identifying bottleneck processes of production line 2.

Table 4. Extraction rules.

No.	Rules
1	(whether complex work = Yes→ Personnel quota <=1→Personnel control problem)
2	(whether complex work = Yes→ Personnel quota > 1→ Variance<=8.16 →Personnel control problem)
3	(whether complex work = Yes→ Personnel quota >1→ Variance > 8.16 →Other problem)
4	(whether complex work = No→ Time difference index <=79.86 → Personnel unprofessional problem)
5	(whether complex work = No→ Time difference index >79.86 → Station scheduling problem)

$$GR(S, Variance) = \ 0.2839473436$$

$$GR(S, Personnel\ quota) = \ 0.1626725838$$

$$GR(S, Whether\ complex\ work) = 0.8578085211$$

Based on the calculations conducted above, it is evident that the gain ratio for "whether complex work" is notably higher compared to other attributes. The gain ratio for station, time difference index, and variance are equal, while the gain ratio for personnel quota is the lowest. Consequently, "whether complex work" is chosen as the root node for constructing the classification decision tree, as illustrated in Figure 4.

The decision tree (Figure 4) generates a set of rules (Table 4) derived from the analysis.

Based on the established decision tree rules, the bottleneck processes in production line 2 can be identified during other production cycles, allowing for the discovery of the root causes of these bottlenecks. Through pre-calculation, all bottleneck stations on the production line are identified, their characteristic attributes are assessed, and the target attributes for each station are determined based on the decision tree rules, indicating potential issues at each station. Given the complexity of each station's actual conditions, specific problems identified may deviate. Therefore, a comprehensive analysis of the specific problems at each station should be conducted in conjunction with the actual circumstances. For instance, issues at the wire harness sorting station may be attributed to human management and control problems. Additionally, due to the station's unique characteristics, significant variance fluctuations may lead to the identification result being categorized as "other".

Production line balance improvement plan

Improvement of unreasonable working hour allocation

Based on the target attribute identification results outlined in Table 3 for production line 2, it is evident that the stations facing process timing issues include intubation, install quick connector, and fixed motors. Despite the install quick connector station having an "unreasonable working hour allocation", further analysis reveals that the installation of 8 quick connectors in this station averages 110.70 s, with the steps being unsuitable for allocation to other stations. Therefore, this station is categorized as "other" and alternative improvement methods are employed based on the station's specific characteristics. Additionally, while the installation of the gas-liquid separator and globe valve station is initially classified as presenting "other problem", a closer examination reveals that the process can be reallocated to other stations. This flexibility allows for the rearrangement of working hours.

Subsequently, the stations identified with unreasonable working hour allocations are the fix motor and reactance station, intubation station, and install gas-liquid separator and globe valve station. To address these issues, improvements are made based on the station load rate, with a focus on optimizing the process flow and reallocating time to stations with lower working hours before and after, considering the time differences and the overall process rationality. Process flow diagrams for some workstations are drawn to facilitate this optimization process.

Fix motor and reactance station. Analyzing the data in Figure 5(a) reveals that the average time spent on fixing the motor is 136.8 s, while filling helium takes an average of 79.2 s. This results in a time difference of 57.6 s between the two processes. During the motor fixing process, it takes an average of 30.60 s to transfer the partition to the base. To achieve a balanced flow, we currently place the clapboard in the filling helium process and preload it. The improved fixing motor process now takes an average of 106.20 s, while filling helium averages 109.8 s.

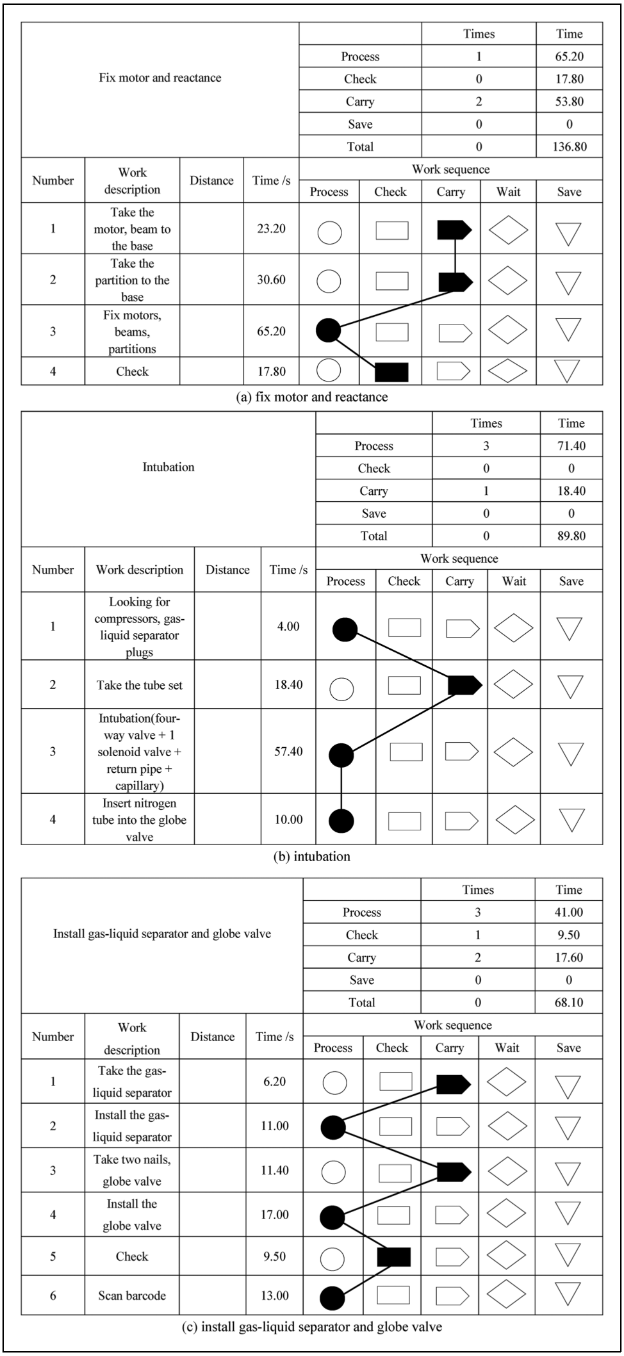


Figure 5. Flow charts for station processes of (a) fix motor and reactance, (b) intubation and (c) install gas-liquid separator and globe valve.

Intubation station. The analysis in Figure 5(b) indicates that the average time required for intubation is 89.80 s, making it a bottleneck process. The preceding put wet cloth process, on the other hand, averages 31.80 s. This significant time difference of 69 s highlights the prolonged waiting time for the put wet cloth process. To address this issue, we have streamlined operations within the intubation process by transferring unnecessary tasks to the put wet cloth process, thereby reducing the time and workload of intubation. Upon examining the intubation workflow, it was found that it takes an average of 18.40 s to take the tube group and 10 s to insert nitrogen tube into the stop valve. As a solution, the worker responsible for putting wet cloth can assist the intubation worker by holding the tube set, preloading the tube, and inserting the nitrogen filling tube into the next intubation process's port before the subsequent intubation process begins. Following these improvements, the average time for intubation has decreased to 61.40 s, while the average time for putting wet cloth is now 60.20 s.

Install gas-liquid separator and globe valve station. Upon reviewing the data in Figure 5(c), it is evident that the installation of the gas-liquid separator and globe valve takes 68.10 s, which includes 6.20 s for turning around and 11.40 s for retrieving parts. In comparison, the fixing press process only takes 41.60 s, a significant 20.80-s difference from the installation process. To address this imbalance, we have decided to eliminate the part retrieval step from the gas-liquid separator installation process and integrate it into the fixing press process. By preloading the upper part of the gas-liquid separator during the fixing press operation, we have successfully reduced the installation time to 56.20 s, while the fixing press now takes 47.8 s.

Improvement of personnel control problem

On-site problem analysis. After utilizing the decision tree for identification, it was determined that stations facing personnel control issues include welding, helium filling, halogen inspection, and sensor plug installation. Upon conducting an on-site investigation and detailed analysis, it was observed that the arrange wiring harness station exhibits a significantly higher variance compared to other stations, potentially resulting in deviation in identification results. Despite this variance, the station should still be classified as having a personnel control problem. We have categorized these stations into three groups based on their characteristics:

1. Specialized stations (requiring specific qualifications or certifications): welding, helium filling, halogen inspection, etc.
2. Error-prone stations (requiring designated personnel for inspection): sensor plug installation.
3. Miscellaneous work station: arrange wiring harness.

Upon analyzing the production capacity in Section 3, it was noted that when the bottleneck process time is considered as the working hours, the calculated daily output exceeds the actual daily output. Further observations on-site revealed significant downtime waste during actual production. Aside from brief downtime due to evacuation and man-machine operations in the test room, most downtime is attributed to personnel issues. Additionally, a notable disparity exists between actual working hours and standard working hours, with many workers lacking the necessary qualifications and effective management. The following specific issues were identified:

1. Operational errors leading to defective products left in subsequent processes, necessitating diagnosis and re-operation, resulting in downtime.
2. Welding and helium filling stations require specialized and qualified personnel for critical processes that cannot be substituted by ordinary workers. Emergencies causing worker unavailability can lead to work stoppages.
3. Operators at stations like halogen inspection fail to timely inspect tested products, causing product accumulation and shutdowns.
4. The plug sensor station is prone to errors due to ineffective management and lack of standardization.
5. The arrange wiring harness station's complexity makes it challenging to accurately predict worker completion times. Fluctuations in working hours at this station significantly impact overall production line productivity.

Improvement plans and suggestions. **For special stations:**

1. Ensure the presence of professionals by increasing the number of technical personnel and conducting examinations to select suitable candidates. Implement salary incentives to attract and retain skilled workers.
2. Utilize a mentorship program where experienced employees provide targeted guidance to new hires, accelerating the learning curve and enhancing the expertise of the new workforce.
3. Maintain continuity by assigning experienced employees to special stations, fostering a culture of dedicated personnel and enhancing overall efficiency. Avoid frequent personnel changes to promote stability and expertise.
4. Develop contingency plans for emergencies at special stations. Identify backup personnel from similar stations or other production lines who can seamlessly step in if regular workers are unavailable. Ensure that production lines can continue operating smoothly without disruptions.

For error-prone stations:

1. Provide regular training for the production line monitor to ensure a clear understanding of their responsibilities. Conduct routine inspections of the production line to identify and address any issues promptly. Assign the squad leader to the bottleneck station of the day, allowing them to assist with the bottleneck process during idle periods. Offer incentives to efficient squads to boost the motivation and management effectiveness of squad leaders.
2. Develop a system for handling problem products at error-prone stations. If workers can quickly identify and rectify the issue, they should make on-site revisions. If identification takes longer or is not possible, the products should be

moved to the defective products area for further inspection and rework, to be handled by the squad leader.
3. Implement daily checks by the squad leader to ensure that personnel are following standardized operating procedures. This will help to establish a consistent workflow for inspection personnel, reducing errors and improving efficiency.

For arrange wiring harness stations:

1. Consider adjusting the personnel allocation, expanding the workbench size, or establishing multiple smaller stations to handle wiring harness assembly. Each station should be equipped to perform the same tasks efficiently. In practice, monitor the progress of workers and adjust processes between stations as needed to maintain consistent cycle times for each production line.
2. Simplify the wiring harness arranging process by breaking it down into smaller, more manageable tasks. Increase the number of stations and allocate resources accordingly. Tasks such as extracting wire ties, arranging bundling lines, cutting off tying wire ends, arranging the wiring harness, marking with a pen, and applying cotton on joints. Assign these relatively simple tasks to each small station to ensure stable working hours for this task.

Improvement results analysis. The key to improving stations with personnel control issues lies in effective people management. It is important to note that the effects of such improvements may not be immediate, but they will become increasingly evident over time. The final measurement of man-hours and the corresponding improvement effects are summarized in Table 5.

Improvement of other problems

Stations with other issues often require a comprehensive approach to improvement, as direct analysis of data may not yield specific reasons for bottleneck processes. These improvements typically involve various methods, such as simplifying and enhancing each process step at the workstation, improving actions, enhancing coordination between humans and machines, and adjusting personnel coordination.

Table 5. Time comparison of improvement effects between different stations.

Station name	Before improvement (s)	After improvement (s)
Welding 1	81.37	74.60
Welding 2	83.61	74.00
Fill helium	79.25	53.12
Halogen inspection	94.21	82.19
Plug sensor	88.94	69.33
Arrange wiring harness	96.22	67.52

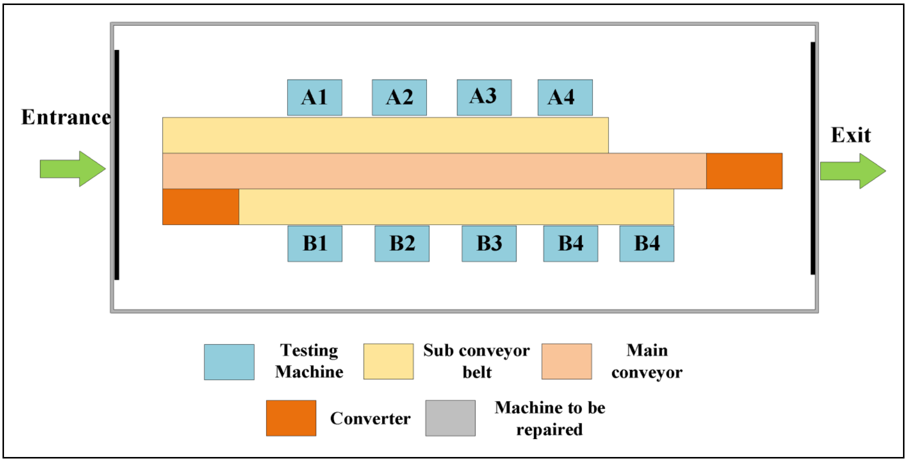


Figure 6. Original plan for the evacuate operation station.

The decision tree identifies several stations with other problems, including wiring harness arrangement, test running, evacuation, enclosure fixing, machine connector installation, and gas-liquid separator and globe valve installation. Among them, we focus on analyzing the gas-liquid separator and globe valve installation station, where we find that the station time could be rearranged, falling under the category of "unreasonable working hour allocation". These stations are then improved based on the station load rate.

Man-machine operation of evacuation stations. The evacuation operation layout is illustrated in Figure 6, with 4 unit operating tables located on the left side of the conveyor belt and another 4 operating unit tables on the right side. Products enter the right sub-conveyor belt and make an initial turn. The average manual operation time for a single person in the evacuation pipe operation is 176.75 s. Due to the complexity of the wiring harness process, there is a significant variance in the operation time among workers. Through on-site observation, the following issues have been identified.

1. While plugging the wiring harness, a worker picks up a wire from the ground between the operating platforms of units 2 and 3, resulting in unnecessary walking and inefficiency.
2. During the machine evacuation process, there is a lack of clarity in the indicator lights for machine progress and completion. Typically, the yellow light is off, and the green light is on to indicate completion, as depicted in Figure 7(a). There are two machines whose yellow light is off to indicate that the evacuate is completed, as shown in Figure 7(b). Additionally, after the evacuation of the machine is completed, if a single product fails to meet the evacuation requirements, the red light will turn on, as shown in Figure 7(c). Workers often need to walk around to assess the completion of wiring harness plugging.

(a) Yellow light off, green on (b) Two yellow lights off (c) Red light on

Figure 7. Three states (a), (b) and (c) of evacuate complete instrument.

(a) Evacuation work console (b) Indicator light

Figure 8. Improvement plans of (a) evacuation work console and (b) indicator light for taking materials at distance and indicator light.

3. Two workers are stationed on the left side of the production line, with their working areas spanning all four unit operation stations from A1 to A4. While this setup allows for mutual assistance, it also leads to inefficiencies due to excessive walking back and forth.
4. On the right side of the production line, there are four unit consoles and only one worker, resulting in significant idle time for the machines.

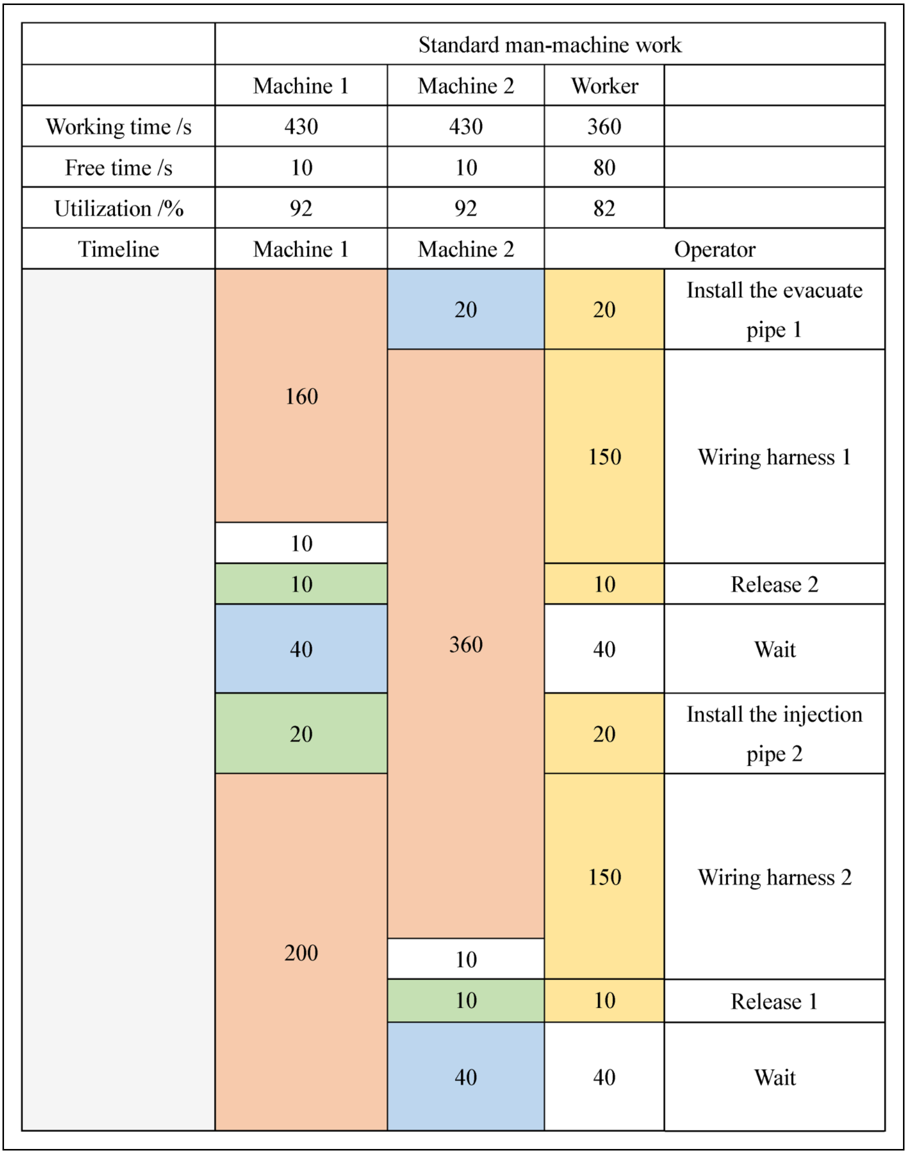


Figure 9. Man-machine operation diagram for analyzing the man-machine improvement.

Improvement plan:

1. To address the issue of workers having to fetch wires from a distance, a box can be installed in the designated area shown in Figure 8(a) to store multiple wires. This will significantly reduce the time required to fetch the wires to approximately 2 s.

2. To enhance the clarity of the indicator lights, it is recommended to improve them based on the indicator lights used in production line 2, as depicted in Figure 8(b). This will ensure consistent and clear indications of machine progress and completion.
3. To address the issues (3) and (4), a man-machine operation analysis is conducted, revealing that units A1, A2, A3, and A4 process an external machine with cycle times of 31 s, 36 s, 47 s, and 56 s, respectively. The machine evacuation time remains fixed at 360 s. Average(/seconds) and relax(10%/ seconds) for the other processes of online scan, patch harness (front), take line, patch harness (rear) and extubation release are 18.8&21, 65.2&72, 2&2, 20.6&23, 10&11, and the total Average(/seconds) and relax(10%/ seconds) are 119.59&131.55.

In Figure 9, the job diagram illustrates the man-machine operation. One worker oversees two machines, ensuring 100% machine utilization and 18% worker idle time. The four stations on the left side of the production line are assigned to two workers, A and B. Worker A handles A1 and A2, while worker B manages A3 and A4. Due to the extended entry and exit times of external machines A3 and A4, worker B retrieves wires from the four unit consoles and places them in the wire box proposed in (2). On the right side of the production line, there is only one worker in charge, resulting in the closure of station B4 and leaving only three unit operating stations active.

Quick connector installation process. The installation of quick connectors is identified as a bottleneck process, taking up a significant amount of time at 149.10 s. The analysis reveals that the problem for this station is “unreasonable working hour allocation”. However, during the improvement process, it is challenging to directly reallocate the man-hours for the installation of the quick connector station to other stations. Therefore, it is divided into “other problem”. The work steps involved in the installation of the quick connector station include three rotations and walking. These steps include taking the quick connector, placing the copper cap, and positioning the damp cloth. The average time required to install eight quick connectors is particularly high, reaching 110.70 s.

1. The ECRS principle is being utilized to merge and rearrange the three turn-and-walk movements. The operator now places the unscrewed copper cap next to the base, and once the quick connector is installed, both the copper cap and damp cloth are placed on the rear workbench. The next set of quick connectors is then retrieved from the workbench. This improvement reduces the number of turn-and-walk movements from three to one.
2. Based on on-site observations, it has been noted that some operators perform the task of installing the quick connector using only one hand, which proves to be less efficient compared to using both hands. Therefore, operators are now encouraged to use both hands during the installation of the quick connector. Observations show that the average time for one-handed operation in installing the quick connector is 159.80 s, while two-handed operation averages at 61.60 s. The time difference between the two operation methods is 98.20 s, allowing operators to significantly enhance operational efficiency by using both hands.

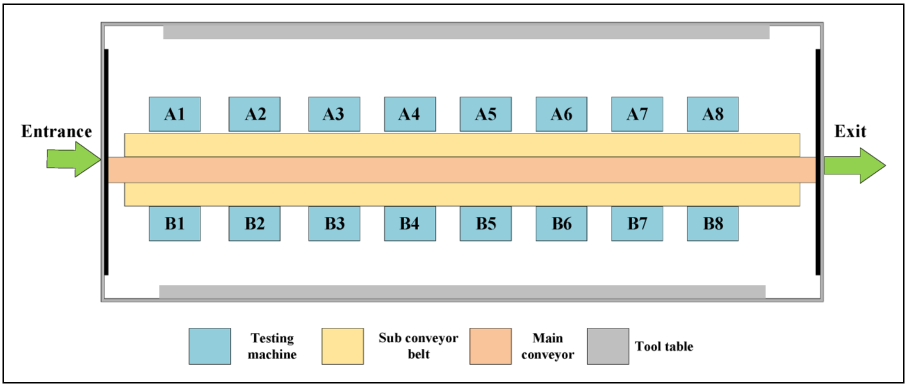


Figure 10. Original plan for test room layout.

3. Through on-site observations, it has been identified that there are three types of quick connectors, which are currently stored either in the locker on the rear side or in the two storage boxes on the back workbench. When operators retrieve the quick connectors, they need to identify the model themselves, resulting in time wastage. To address this issue, it is recommended that the feeder differentiates the placement of the three types of quick connectors during replenishment. Additionally, three storage boxes can be placed on the workbench, each storing a different type of quick connector, to facilitate easy access for operators.

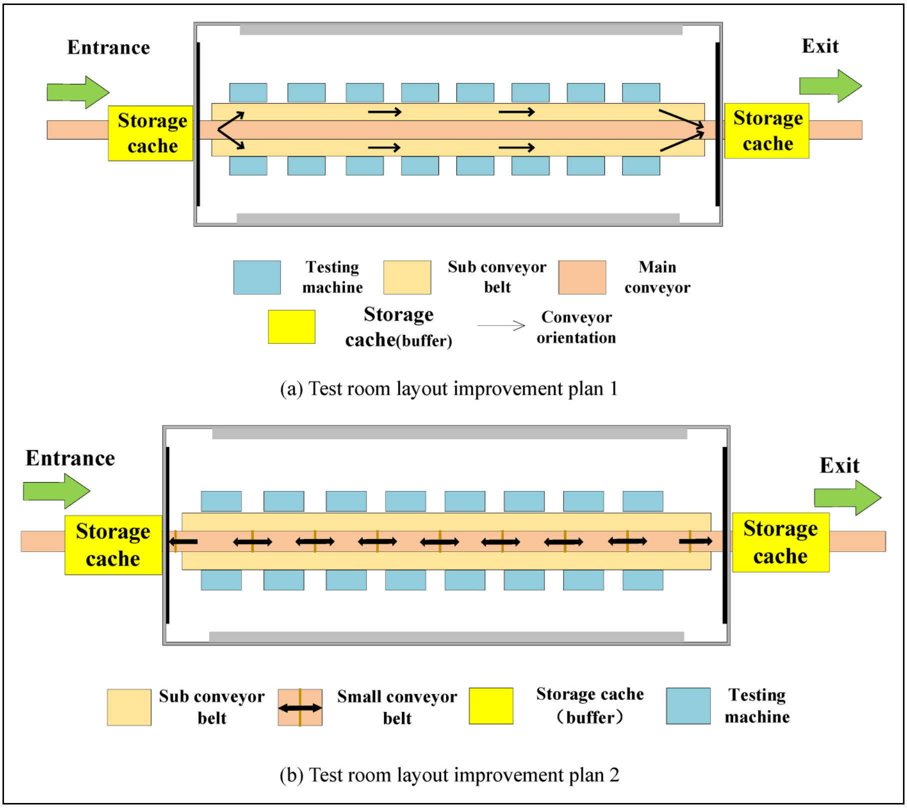
Test room. There is stagnation within the test room, with the main reasons being:

1. Imbalance in the output of the pre-process results in fluctuating start beats in the test room, with the amplitude of fluctuation being related to the pre-process beat.
2. Unclear division of labor among testers leads to overlapping operations, causing incomplete grasp of test information for each external machine and potential work inefficiencies.
3. Bottlenecks or significant fluctuations in the post-process cause machines that have completed testing to stall in the test room. The stagnant external machines may obstruct the filling of subsequent external machines.

Based on the current layout and distribution method (4 workers on both sides and 2 workers on one side), the upper and lower areas of the test room are divided into 4 small areas, with each worker responsible for one area, namely [A1-A4], [A5-A8], [B1-B4], [B5-B8]. At the entrance, the conveyor belt allocates external machines awaiting testing to the area with the most available space. The layout of the test room is shown in Figure 10. For each zone, an external machine awaiting testing is connected every 80×4 s.

Improvement plan:

Assign two testers on each side with different tasks:



(a) Test room layout improvement plan 1

(b) Test room layout improvement plan 2

Figure 11. Improvement plans of (a) and (b) for test room layout.

Primary mission for Worker A: Wiring, pipe installation, power plug insertion, first test, and monitoring the indicator light until it changes from yellow to green.

Primary mission for Worker B: Power plug removal, tube replacement, second test, tube retrieval, and wire withdrawal.

Secondary mission: Mutual assistance and support between the two workers.

Improvement Plan 1: Create a buffer zone after the crowded sections in the test room, specifically in the first two and last two sections. The layout for this improvement plan is depicted in Figure 11(a).

Improvement Plan 2: Transfer the crowded sections in the front of the test room to the main conveyor belt. The main conveyor belt consists of eight smaller conveyor belts, each capable of moving left and right. To allow a single external machine to move freely on the main conveyor belt, an external machine rotating device is installed on each small conveyor belt. This device determines the areas with available vacancies in real-time. The layout for this improvement plan is shown in Figure 11(b). This solution helps save space but requires enhancements to the conveyor belt system.

Fix enclosure process. The fix enclosure process, which requires 84.95 s, is identified as a bottleneck process with various operational issues categorized as “other problem”. Upon analyzing the workstation, it is determined that the layout of the operation site is not optimized. Specific issues include:

1. Workers have to turn around and walk 2–3 meters to retrieve screws.
2. The placement of the screw gun is uncertain, leading to unnecessary movements and inconvenience for workers.

Improvement plan:

1. Fix the screw box beside the conveyor belt or provide a designated space for workers to access screws easily.
2. Allocate multiple locations for screw gun placement next to the conveyor belt.

Improvement impact:

After implementing the improvement plan, the man-hours for the fix enclosure station decrease from 84.95 s to 69.32 s, resulting in a reduction of 15.63 s. Additionally, the stability of working hours improves significantly after conducting multiple measurements of station performance.

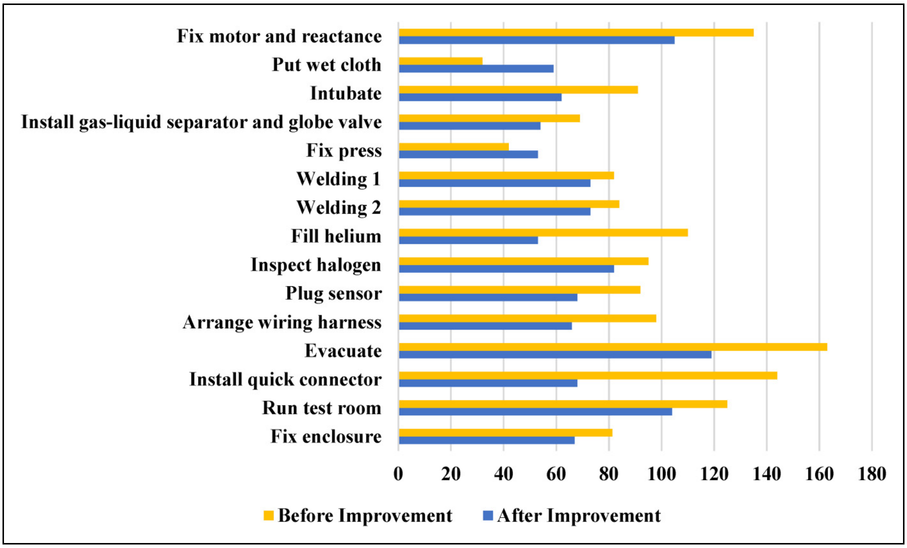


Figure 12. Time comparison between before and after improvement.

Evaluation and discussion of expected effects

Through the calculation of the optimized stations, the predicted station time for these improvements was determined. The analysis indicates that the bottleneck time can be reduced from 170.80 s to 119.59 s, as illustrated in Figure 12. This improvement is based on the following formula:

$$Balance\ rate = \frac{sum\ of\ each\ process\ time}{number\ of\ stations\ *\ bottleneck\ process\ time} \tag{11}$$

The production line balance rate has been enhanced from the initial 51% to 66%.

Production line balancing is a well-established research area in China's manufacturing industry. However, utilizing a decision tree optimization scheme enhances the accuracy and efficiency of identifying workstation issues, making the optimization process more streamlined and intelligent. By leveraging machine learning algorithms, production line problems can be systematically identified and classified, ultimately leading to improved production line efficiency. This paper introduces a problem identification system based on the bottleneck process, effectively addressing the balance issues in H Company's commercial air conditioner production line. Through a comprehensive analysis of the production line improvement process, a total of 15 bottleneck stations were identified, with 12 stations correctly pinpointed, resulting in an 80% accuracy rate. Subsequent improvements were successfully implemented in 13 stations, achieving an impressive application rate of 86.67%.

Despite the progress made in this study, there are several areas that require further research and improvement:

1. Utilizing the existing infrastructure, enhancing the production line's digitization by implementing "Kanban" systems can elevate the level of informatization and management efficiency, consequently boosting overall production efficiency.
2. Strengthening personnel management is essential for the success of lean production initiatives. Future efforts should focus on developing a workforce with diverse skill sets through human factors engineering principles and establishing a sustainable evaluation mechanism to facilitate mutual growth of both the company and individuals.
3. Consider incorporating additional measurement data to expand the scope of attributes identified by the decision tree method. This can lead to more precise and effective identification results, enhancing overall performance and decision-making processes.

Conclusion

In this paper, a machine learning algorithm is leveraged to uncover the underlying rules governing production line issues, enabling rapid identification and classification of such problems. Through the application of this method, we successfully address the production line balance challenges encountered in H Company's commercial air conditioner production. By utilizing the bottleneck process identification approach, we identify a total of 15 bottleneck stations with an accuracy rate of 80%. Moreover, we effectively enhance 13 of these stations, achieving an impressive application rate of 86.67%. Moving forward, our focus is on disseminating the decision tree-based production line issue identification method and extending its application to diverse production fields to unlock new opportunities for lean production optimization techniques.

Acknowledgements

The work is supported by Project of Shandong Province Higher Educational "Youth Innovation Science and Technology Plan" Team (No.2021KJ060), Humanities and Social Sciences Youth Foundation, Ministry of Education (No.23YJCZH221), Natural Science Foundation of Shandong Province (No.ZR2023QE030) and Qingdao Postdoctoral Funding Project (No.QDBSH20220202063). The authors are very grateful to the financial contribution and convey their appreciation for supporting this basic research.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Qingdao Postdoctoral Funding Project, Natural Science Foundation of Shandong Province, Humanities and Social Sciences Youth Foundation, Ministry of Education, Project of Shandong Province Higher Educational "Youth Innovation Science and Technology Plan" Team, (grant number QDBSH20220202063, ZR2023QE030, 23YJCZH221, 2021KJ060).

ORCID iD

Rui Wang https://orcid.org/0000-0003-2634-4656

References

1. Quinlan JR, 2014, C4. 5: programs for machine learning, Elsevier.
2. Salveson ME. The assembly line balancing problem. *J Ind Eng* 1955; 77: 18–25.
3. Helgeson W and Birnie DP. Assembly line balancing using the ranked positional weight technique. *J Ind Eng* 1961; 12: 394–398.
4. Gutjahr AL and Nemhauser GL. An algorithm for the line balancing problem. *Manage Sci* 1964; 11: 308–315.
5. Roberts SD and Villa CD. On a multiproduct assembly line-balancing problem. *AIIE Trans* 1970; 2: 361–364.
6. Bennett GB and Byrd Jr J. A trainable heuristic procedure for the assembly line balancing problem. *AIIE Trans* 1976; 8: 195–201.
7. Dar-El E and Rubinovitch Y. Must-a multiple solutions technique for balancing single model assembly lines. *Manage Sci* 1979; 25: 1105–1114.
8. Agrawal P. The related activity concept in assembly line balancing. *Int J Prod Res* 1985; 23: 403–421.

9. Erel O. Automated measurement of serum ferroxidase activity. *Clin Chem* 1998; 44: 2313–2319.
10. Ozcan U and Toklu B. Multiple-criteria decision-making in two-sided assembly line balancing: a goal programming and a fuzzy goal programming models. *Comput Oper Res* 2009; 36: 1955–1965.
11. Huang D, Mao Z, Fang K, et al. Combinatorial benders decomposition for mixed-model two-sided assembly line balancing problem. *Int J Prod Res* 2022; 60: 2598–2624.
12. Guden H and Meral S. An adaptive simulated annealing algorithm-based approach for assembly line balancing and a real-life case study. *Int J Adv Manuf Technol* 2016; 84: 1539–1559.
13. Bortolini M, Faccio M, Gamberi M, et al. Including material exposure and part attributes in the manual assembly line balancing problem. *IFAC PapersOnLine* 2016; 49: 926–931.
14. Krenczyk D, Skolud B and Herok A, A heuristic and simulation hybrid approach for mixed and multi model assembly line balancing. In: *International conference on intelligent systems in production engineering and maintenance*, Springer, 2017, pp. 99–108.
15. Oksuz MK, Buyukozkan K and Satoglu SI. U-shaped assembly line worker assignment and balancing problem: a mathematical model and two meta-heuristics. *Comput Ind Eng* 2017; 112: 246–263.
16. Pirogov A, Gurevsky E, Rossi A, et al. Robust balancing of transfer lines with blocks of uncertain parallel tasks under fixed cycle time and space restrictions. *Eur J Oper Res* 2021; 290: 946–955.
17. Sikora CGS. Benders' decomposition for the balancing of assembly lines with stochastic demand. *Eur J Oper Res* 2021; 292: 108–124.
18. Pinhão JMOB, Ignacio AAV and Coelho O. An integer programming mathematical model with line balancing and scheduling for standard work optimization: a realistic application to aircraft engines assembly lines. *Comput Ind Eng* 2022; 173: 108652.
19. Teshome MM, Meles TY and Yang CL. Productivity improvement through assembly line balancing by using simulation modeling in case of Abay garment industry Gondar. *Heliyon*. 2023;10: e23585.
20. Li J, Chen L and Chang Z. Alternative paths of it for process reengineering and a case study. *Ind Eng Manag* 1999; 4: 24–27.
21. Chen X, Xiao T, Hao X, et al. Assembly line balancing using genetic algorithms. *Comput Eng Applic* 2001; 37: 81–83.
22. Song H and Han Y. An optimization approach for multiple-objective assembly line balancing. *Oper Res Manag Sci* 2002; 11: 55–62.
23. Lu J, Lan X, Chen Y, et al. An optimized design in production assembly systems. *Zhejiang Eng Inst* 2003; 31: 290–292.
24. Sun J, Gao G and Jiang Z. Research on the methods of streamline balance. *Group Technology Production Modernization* 2004; 21: 34–36.
25. Guo F, Zhang G and Wen J. Application of work study to the throughput balance of car assembly line. *Ind Eng Manag* 2006; 11: 119–122.
26. Zhang R, Chen D and Yang Q. Solution of assembly line balancing problem based on improved genetic algorithms. *Comput Eng Des*. 2006; 27: 3731–3733.
27. Wu E, Jin Y, Hu X, et al. A branch and bound algorithm for bilateral assembly line balancing. *Machine Made* 2008; 46: 4–8.
28. Niu Z, Wu X and Yue L. Optimization of assembly shop of clutch enterprise based on lean production. *Ind Eng Manag* 2015; 20: 1–6.
29. Li J and Meng C. Application of work study to production line redesign: a case study. *Ind Eng J* 2009; 12: 121–125.
30. Wang J and Li D. Task scheduling based on a hybrid heuristic algorithm for smart production line with fog computing. *Sensors* 2019; 19: 1023.

31. Zhang B, Xu L and Zhang J. Balancing and sequencing problem of mixed-model u-shaped robotic assembly line: mathematical model and dragonfly algorithm based approach. *Appl Soft Comput* 2021; 98: 106739.
32. Jia S, Yuan Q, Lv J, et al. Therblig-embedded value stream mapping method for lean energy machining. *Energy* 2017; 138: 1081–1098.
33. Jia S, Cai W, Liu C, et al. Energy modeling and visualization analysis method of drilling processes in the manufacturing industry. *Energy* 2021; 228: 120567.
34. Huang P and Deng Z. Analysis and improvement of production line balance rate based on witness. *Manufacturing Automation* 2021; 09: 50–55.
35. Chang Y, Wan P and Zhou S. Research on production line simulation and optimization of shoemaking needle cart workshop based on Flexsim. *Logistics Eng. Manag* 2023; 12: 28–34.
36. Jiang M, Liang Y, Feng X, et al. Text classification based on deep belief network and softmax regression. *Neural Comput Appl* 2018; 29: 61–70.
37. Breiman L. Random forests. *Mach Learn* 2001; 45: 5–32.
38. Hearst MA, Dumais ST, Osuna E, et al. Support vector machines. *IEEE Intell. Syst. Appl.* 1998; 13: 18–28.
39. Peterson LE. K-nearest neighbor. *Scholarpedia* 2009; 4: 1883.
40. Murphy KP. *Naive Bayes classifiers*. University of British Columbia, 2006; 18, pp. 1–8.

Author biographies

Rui Wang is currently a lecturer with the College of Energy and Mining Engineering, Shandong University of Science and Technology, Qingdao, China. She received the PhD degree in computer software and theory from Wuhan University, Wuhan, China, in 2019. Her research interests include industrial engineering, data mining, and artificial intelligence.

Tengyuan Xin is currently a master's student in the Department of Industrial Engineering, Shandong University of Science and Technology, Qingdao, China. His current research interests include industrial engineering, data mining, and artificial intelligence.

Shun Jia is currently a professor and doctoral supervisor with the Department of Industrial Engineering, Shandong University of Science and Technology, Qingdao, China. He received the PhD degree in industrial engineering from Zhejiang University, Hangzhou, China, in 2014. His research interests include sustainable design and manufacturing, green manufacturing, and lean production.

Dawei Ren is currently an associate professor with the Department of Industrial Engineering, Shandong University of Science and Technology, Qingdao, China. He received a PhD degree in mining information engineering from Shandong University of Science and Technology, Qingdao, China, in 2010. His current research interests include human factors, safety management, and lean production.

Meiyan Li is currently a professor with the Department of Industrial Engineering, Shandong University of Science and Technology, Qingdao, China. She received the PhD degree in business management from Shanghai Jiao Tong University, Shanghai, China, in 2007. Her research interests include big data analysis and logistics supply chain management optimization.