Journal of Reliability
Science and
Engineering

PAPER • OPEN ACCESS

An intelligent diagnosis framework for quality defects in aircraft assembly using weighted text embedding and a stacked sparse autoencoder

To cite this article: Zirong Wang *et al* 2025 *J. Reliab. Sci. Eng.* **1** 035401

View the article online for updates and enhancements.

You may also like

- Adaptive artificial neural network for uncertainty propagation
Yan Shi, Lizhi Niu and Michael Beer
- Time-variant system reliability analysis via stochastic process discretization and most probable point trajectory approximation
Dequan Zhang, Hongfei Zhang, Pengfei Zhou et al.
- Probabilistic calibration of model parameters with approximate Bayesian quadrature and active machine learning
Pengfei Wei, Masaru Kitahara, Matthias G R Faes et al.

This content was downloaded from IP address 118.123.172.215 on 08/09/2025 at 07:29

OPEN ACCESS
IOP Publishing | ISECAEP
Journal of Reliability Science and Engineering
J. Reliab. Sci. Eng. **1** (2025) 035401 (15pp)
https://doi.org/10.1088/3050-2454/adfc76

An intelligent diagnosis framework for quality defects in aircraft assembly using weighted text embedding and a stacked sparse autoencoder

Zirong Wang, Zhen Chen*, Zhaoxiang Chen, Tangbin Xia and Ershun Pan

Department of Industrial Engineering & Management, and the State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, Shanghai 200240, People's Republic of China

E-mail: chenzhendr@sjtu.edu.cn

Received 29 May 2025, revised 26 July 2025
Accepted for publication 18 August 2025
Published 5 September 2025

Abstract

During the aircraft assembly process, quality defects are generally recorded in the form of text and diagnosed manually by engineers. To improve efficiency and make accurate maintenance decisions, a new defect diagnosis approach is needed. However, common diagnosis methods usually involve monitoring digital or image data, and are not suitable for identifying quality defects in text data. Furthermore, current text classification methods are difficult to implement in aircraft assembly. For quality defects involving long Chinese text, this study proposes a novel intelligent diagnosis framework that integrates weighted text embedding and a stacked sparse autoencoder (AE) network. Firstly, the defect text is segmented into individual words using Chinese word segmentation, and the word vectors for keywords are represented by word2vec, a neural embedding model that captures semantic relationships among words. Secondly, an expert knowledge-based semi-supervised attribute annotation algorithm is developed to determine the weights of the keywords, and the text vectors are then calculated through the weighted sum of the word vectors. Thirdly, considering the high dimensionality and sparsity of text vectors, a stacked sparse AE network is constructed to classify the text vectors and identify the types of quality defects. Application to a practical aircraft assembly dataset demonstrates the excellent performance of the proposed framework.

Keywords: quality defect, aircraft assembly, diagnosis framework, Chinese text embedding, stacked sparse autoencoder

* Author to whom any correspondence should be addressed.

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

© 2025 The Author(s). Published by IOP Publishing Ltd on behalf of the Institute of Systems Engineering, China Academy of Engineering Physics

Acronyms and abbreviations

ANN	artificial neural network
BiLSTM	bidirectional long short-term memory
HS	hole spacing
LSTM	long short-term memory
RNN	recurrent neural networks
SSAE	stacked sparse autoencoder
AE	autoencoder
DNN	deep neural network
LLM	large language models
POS	part-of-speech
SAE	sparse autoencoder
SVM	support vector machine

1. Introduction

Quality defects in the aircraft assembly process can arise in many areas, such as hole mis-spacing and surface damage (Zhang *et al* 2024). Hence, management and maintenance engineers need to identify solutions using extensive maintenance manuals (Park and Park 2025). In recent years, with civil aircraft production increasing at a doubling rate, assembly-related quality defects are also increasing, posing huge challenges for maintenance work (Yan and Wang 2022). The problems arising in the maintenance of quality defects in aircraft assembly can be summarized as follows:

- Quality defects records are in the form of unstructured Chinese text and not standardized. The diagnosis of quality defects in the aircraft assembly process is implemented and recorded by assembly workers. There is no standardized record format. Therefore, it is difficult to use numerical methods to analyze the defect data.
- Historical information is not fully utilized. Engineers mainly make repair suggestions based on their own experience and maintenance manuals. Note that a great deal of important information, such as historical solutions to repairing the quality defects, is not effectively considered. This information may provide similar or even the same guidance for handling the current quality defect. Thus, repetitive work (making repair suggestions by referring to maintenance manuals) could be reduced for maintenance engineers .
- The structural complexity of aircraft means that there are various types of quality defects. Correlations and coupling between different subsystems and different parts make it difficult to identify the defect type in the aircraft assembly process. Thus, making disposal opinions based on the corresponding diagnosis results is not effective and accurate.

In order to adapt to the new era of large-scale aircraft manufacturing, it is of great significance to improve the quality defect disposal methods used in aircraft assembly processes, reduce human and material costs, and shorten production times (Shafi *et al* 2023). Enterprises need to make full use of historical data and propose new defect diagnosis methods. Combined with big data technologies, the efficiency of diagnosis and decision-making can be improved, and thereby the contradiction between more and more assembly defects and limited human resources can be solved immediately (Compare and Zio 2025).

Most traditional defect and fault diagnosis models are based on standard knowledge rules and rely on digital health and fault data obtained through condition monitoring (Peng and Yang 2024, Yu *et al* 2024). Liao *et al* (2024) established a hybrid algorithm for aero-engine gas-path fault diagnosis by filtering features from fault data. Yang *et al* (2020) proposed a diagnosis model with a distance metric to transfer diagnostic knowledge from one machine to another based on vibration data. Hui *et al* (2022) presented a data-driven modeling approach to evaluate the assembly quality of the linear axis based on a variable selection method and a hybrid quality evaluation model. Recently, some methods for fault diagnosis have focused on image or short-text data (Yang and Zhang 2022, Deng *et al* 2024, Wang *et al* 2024). Shang *et al* (2023) developed a global prior transformer network for aeroengine blade surface damage detection inspection based on borescope images. Ye *et al* (2023) utilized deep learning technology to establish a text classification model for power equipment defect short-text records. However, in practice, most quality defect information about aircraft assembly exists in the form of text data and long text records. Obviously, it is very difficult to diagnose quality defects based on long-text data using current diagnosis methods.

To compensate for the above shortages, text mining-based methods are necessary (Nan *et al* 2022). Currently, for unstructured long text, the most widely used word representation method is vector representation. Mikolov *et al* (2013) proposed the word2vec algorithm, which maps words into a low-dimensional space in the form of vectors, and reflects the relationship between words and context through the numerical values of vectors. Kim *et al* (2020) proposed a topic modeling method for latent semantic analysis and used word2vec and spherical K-means clustering to better capture and represent the context of the corpus. Xu *et al* (2021) and Zhou *et al* (2021) proposed text-based aircraft fault diagnosis models based on word2vec. In short, vector representation has good applicability in dealing with unstructured data related to quality defects.

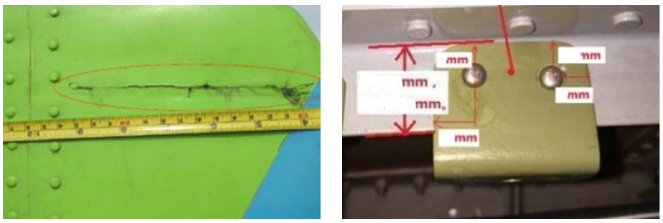
Only when unstructured text data is transformed into structured data can diagnosis algorithms perform quality defect identification. Hence, after obtaining the vectors of word representations, text embedding is another important task (Jing *et al* 2024). That is, using the word vectors in a sentence to constitute a text vector. Considering the temporality and complex semantics of Chinese contexts, deep learning methods are often used for Chinese text embedding (Wang *et al* 2022, Gao *et al* 2024). It should be noted that quality defect diagnosis in this context can be regarded as a Chinese text classification problem. The classification task is accomplished by adding a SoftMax layer as the output of the network. Therefore, most existing studies developed algorithms to deal with text embedding and text classification simultaneously (Hu *et al* 2024, Sheikh *et al* 2024, Jiang *et al* 2025). Zhou *et al* (2016) and Liu and Guo (2019) respectively utilized recurrent neural networks and bidirectional long short-term memory (BiLSTM) to reduce the influence of word segmentation and improve the accuracy of Chinese short text classification.

However, there are two deficiencies in these deep learning-based text classification methods. (1) As the long texts in quality defect records contain a great deal of domain knowledge and weak semantic relationships, it is difficult for deep learning-based text embedding methods to represent defect texts. (2) Although the most common deep learning models, such as RNN, LSTM and BiLSTM, can effectively capture complex temporal characteristics in contexts, they require a lot of labeled data for training and are prone to overfitting. The amount of text data in manufactory industries is far less than that in internet industries. Therefore, there is a significant need to develop new methods for quality defect diagnosis based on long Chinese text data in aircraft assembly.

Recent advances in large language models (LLMs) such as Deekseek, Gemini and GPT-4 have demonstrated impressive zero–shot text classification capabilities (Ma *et al* 2025). However, deploying an LLM–based solution for aircraft–assembly defect logs faces three practical barriers: (i) data–privacy constraints: uploading proprietary defect records to external APIs might be forbidden; (ii) computational and economic overheads: running or fine–tuning a model on–premise entails hardware and energy costs; and (iii) safety–critical interpretability: the opaque reasoning of LLMs is currently incompatible with aviation quality–assurance requirements (Zhang *et al* 2025).

Under these constraints, this paper proposes a lightweight yet interpretable diagnosis framework for quality defects in aircraft assembly, which integrates expert–weighted text embedding with a stacked sparse autoencoder (SSAE). Firstly, the defect records are segmented into individual words and the extracted keywords are then transformed into structured vectors by word2vec. Secondly, an expert knowledge-based attribute annotation method is used to divide the keywords into different categories with different weights. Correspondingly, the text vectors can be obtained by a weighted sum of word vectors. Weighted text embedding can integrate expert knowledge and thereby describe the unique features of aircraft assembly. Thirdly, considering the high dimensionality and sparsity of text vectors, a SSAE network is developed to classify the text vectors so that the quality defect can be diagnosed effectively. The labeled defect texts are used as a training dataset. Compared with other deep learning models, the SSAE requires relatively little labeled training data and has better generalized performance. Finally, the effectiveness and generalization ability of the proposed framework is illustrated using a practical quality defect dataset. To the best of our knowledge, this is the first intelligent diagnosis framework for quality defects in aircraft assembly based on defect records in the form of long Chinese text.

The reminder of this paper is as follows. The research background and purpose are described in section 2. The basic theory of the SSAE is introduced in section 3. A detailed description of the proposed framework is illustrated in section 4. The proposed method is demonstrated using a real dataset in section 5. Section 6 provides the general conclusion and discusses future research aspects.



(a) (b)

Figure 1. Two typical defects in aircraft assembly: (a) cover plate crack, (b) bracket positioning accuracy out-of-tolerance. Note: numeric dimension values are redacted at the collaborating enterprise's request; units are retained for illustrative purposes.

Table 1. An example of a defect record form with its main contents.

Attribute name	Content
Form number	D-G00101-A01-001
Component name	Aircraft panel (飞机壁板)
Aircraft number	101
Defect description	There is one superfluous hole at the butt-joint point of the left panel, and the aperture is l mm. Details is in the attached figure. (在左侧壁板对接处一个多余孔, 孔径为 l mm, 具体描述见附图)
Maintenance action	Repair

Remark: All entries are stored in Chinese in the management system, and only some of the Chinese text is presented.

2. Task description

In this study, we mainly focus on the task of diagnosing quality defects in the process of aircraft assembly. Quality defects occur throughout the whole assembly process. Figure 1 shows two typical quality defects, i.e. cracks and out-of-tolerance positioning accuracy. The defect details are input into the management system in the form of records by field engineers. The defect records are mainly in Chinese. An example of a defect record form is presented in table 1. Only the main contents are given and other information used in the diagnosis framework is omitted. For ease of understanding, English translations are provided and only a part of the raw Chinese text is given.

From table 1, we can see that the defect information is unstructured data, which needs to be transformed into structured data for the next processing step. Generally, the defect description attribute contains the most information about the defect type and is presented in Chinese. In addition, the defect description text is translated into English for better understanding. In the following, we identify the defect type by learning the information in the defect descriptions. Therefore, text mining techniques are needed. Firstly, the text needs to preprocessed to clean invalid and missing data. Secondly, three dictionaries of technical terms for aircraft assembly are established based on expert and field knowledge. The component dictionary includes the names of aircraft subsystems and components. The defect dictionary contains words that describe the details of aircraft quality defects. Note that, a word in Chinese

may consist of several Chinese characters. The stop-word dictionary contains unrelated punctuation symbols and words. With technical and general dictionaries, continuous text can be divided into vocabulary lists based on semantic rules using a Chinese word segmentation algorithm. For words that are not included in the dictionaries, a new word discovery method based on the hidden Markov model is adopted. Thirdly, after obtaining vocabulary lists, word- and sentence-embedding methods are used to construct digital vectors to represent the defect text. Finally, an SSAE network is developed to identify the defect types by training the constructed text vectors.

3. SSAE

The weighted–embedding vectors produced from the defect–log text are high–dimensional and sparse, while the number of available labeled samples is only in the hundreds. The SSAE is attractive in this context because it learns compact low–dimensional representations via unsupervised layer–wise pre–training, reducing the risk of over–fitting under small–sample conditions. It also incorporates a Kullback-Leibler (KL) divergence sparsity constraint that naturally matches the sparsity of the input vectors and is lightweight enough for on–premise deployment on a standard industrial workstation.

Having motivated the choice of the SSAE, we follow the formulation of Xu *et al* (2015) and introduce its constituent parts from the inside out, i.e. from autoencoder (AE) to sparse AE (SAE) to SSAE.

(1) AE: a basic AE is a kind of unsupervised neural network and typically consists of one input layer, one output layer and one hidden layer (Xu and Tan 2019). Figure 2 depicts the single-layer AE structure.

The output values are set to be equal to the input. Deep learning methods always use AEs to obtain appropriate representations from the input during training of hidden layers.

For an input vector $\boldsymbol{x} \in \mathbb{R}^{\mathrm{m}}$, the hidden representation is

$$h(\boldsymbol{x}) = \sigma(\boldsymbol{W}_1 \boldsymbol{x} + \boldsymbol{b}_1), \tag{1}$$

where $\sigma(\cdot)$ is a nonlinear activation function, $\boldsymbol{W}_1$ is a weight matrix and $\boldsymbol{b}_1$ is a bias vector. The calculation process from the input layer to the hidden layer is called an encoder. The output of the AE maps h back to a reconstruction $\hat{\boldsymbol{x}} \in \mathbb{R}^{\mathrm{m}}$, i.e.

$$\hat{\boldsymbol{x}} = \sigma(\boldsymbol{W}_2 h(\boldsymbol{x}) + \boldsymbol{b}_2), \tag{2}$$

where $\boldsymbol{W}_2$ is a weight matrix and $\boldsymbol{b}_2$ is a bias vector. This process is called a decoder. Given a training dataset χ, the parameters $\theta = \{\boldsymbol{W}_1, \boldsymbol{W}_2, \boldsymbol{b}_1, \boldsymbol{b}_2\}$ are optimized by minimizing the reconstruction error, and thereby the objective function of an AE can be given by

$$J(\theta) = \sum_{x \in \chi} \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2^2. \tag{3}$$

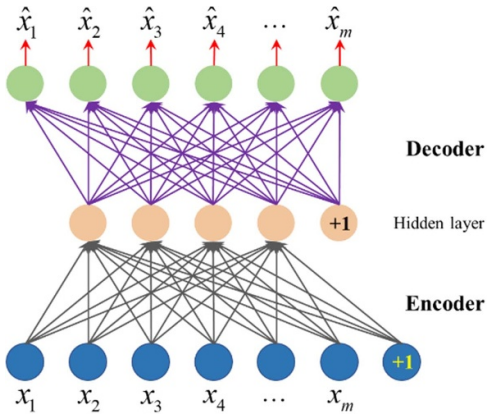


Figure 2. Architecture of an AE.

Remark: in fact, both the encoder and decoder networks of an AE can have multiple layers; irrespective of depth, they share the same objective function and can be optimized with stochastic gradient descent.

(2) SAE: to avoid simply copying the input to the hidden layer, the SAE, which captures the sparse representation by penalizing the average activation of hidden neurons (Wen *et al* 2017), was developed. The objective function of an SAE can be represented in the following formulation

$$\sum_{\boldsymbol{x} \in \chi} \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2^2 + \beta \sum_{j=1}^{r} KL\left(\rho \| \hat{\rho}_j\right) + \frac{\gamma}{2}\left(\|\boldsymbol{W}_1\|_F^2 + \|\boldsymbol{W}_2\|_F^2\right), \tag{4}$$

where $KL\left(\rho \| \hat{\rho}_j\right) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1-\rho) \log \frac{1-\rho}{1-\hat{\rho}_j}$ is the sparse penalty term, $\hat{\rho}_j = \frac{1}{s} \sum_{i=1}^{s} h_j(x_i)$ is the average activation of a hidden neuron j, ρ is a sparsity parameter, r is the number of hidden neurons and the third term penalizes the weights to prevent overfitting. The sparsity penalty coefficient β and the weight decay coefficient γ are determined through cross-validation experiments. In our implementations, $\beta = 3$ balances reconstruction accuracy and the sparsity constraint, while $\gamma = 0.001$ provides an appropriate regularization strength. When the dimension of the hidden layer is smaller than that of the input, then the SAE can learn a compressed and sparse representation of the input data. The SAE can also be trained using gradient descent algorithms.

(3) SSAE: multiple basic SAEs are stacked to form the SSAE (Tao *et al* 2015). The structure of an SSAE is shown in figure 3. More specifically, the outputs of each hidden layer are used as the inputs of each successive layer, which is also named pre-training. The pre-trained weights are connected and finetuned to obtain the final weights of the SSAE. For simplicity, the decoder parts of each basic SAE are not presented in figure 3.

An SSAE is usually trained with a layer-by-layer greedy algorithm (Yu *et al* 2021). Generally, an SSAE can fully represent original features of input datasets better than a basic SAE (Xu *et al* 2015). However, the SSAE is an unsupervised model used only for feature extraction and

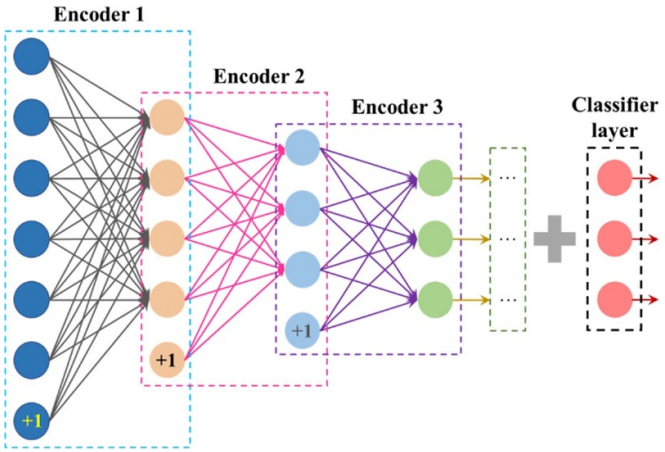


Figure 3. Architecture of an SSAE.

data reconstruction. If we want to use an SSAE for defect diagnosis, a classifier layer should be added after the last encoder layer.

SoftMax logistic regression is a commonly used classifier for the construction of SSAE-based classification models. The output of the SoftMax-based classifier layer is given by

$$\delta(h) = \frac{\left[\exp\left(-\boldsymbol{W}_{c,1}{}^{\mathrm{T}}h\right), \cdots, \exp\left(-\boldsymbol{W}_{c,k}{}^{\mathrm{T}}h\right)\right]}{\sum_k \exp\left(-\boldsymbol{W}_{c,k}{}^{\mathrm{T}}h\right)}, \qquad (5)$$

where $\delta(\cdot)$ is the SoftMax function with weight matrix $\boldsymbol{W}_{c,k}{}^{\mathrm{T}}$ and h is the output of a high-level representation of encoders learned by the SSAE. Note that the classifier is a supervised model. Hence, $\boldsymbol{W}_c$ can be determined by minimizing the cost function with a labeled training dataset.

4. The intelligent diagnosis framework

To achieve the quality defect diagnosis with raw defect records, a general intelligent framework consisting of data acquisition, text mining, weighted text embedding and defect diagnosis is proposed herein. Figure 4 shows the implementation flowchart of the proposed framework. The input data of this framework include raw defect records of aircraft assembly and a thesaurus of technical terms. Then, the Chinese word segmentation method is used to segment the defect records into individual words. Only the keywords for each record are preserved by using the TextRank algorithm. In the third part, as for the extracted keywords, we adopt word representation and attribute annotation based on expert knowledge to obtain the structured text vectors. Finally, the labeled text vectors from different defect types are used as a training dataset. Based on the SSAE and gradient descent algorithms, the diagnosis model can be learned using the training dataset. Substituting a new defect record into the trained framework can output the corresponding defect type. Further, the related techniques of the proposed framework are detailed in the following subsections.



Figure 4. Flowchart of the proposed intelligent diagnosis framework for quality defects in aircraft assembly.

4.1. Text mining of defect records

In order to extract information from the defect descriptions in historical defect records, we need to preprocess the form data first. Incomplete or unrecognized information is deleted. In the text of defect descriptions, Chinese characters are retained and other contents including numbers and punctuation are cleaned. Then, a Chinese word segmentation method is needed to segment continuous text into vocabulary lists according to certain semantic rules. Three dictionaries of technical terms are constructed as the basis of segmentation by making full use of expert knowledge. Parts of these three dictionaries are displayed in table 2.

We use the Jieba method, which was developed based on the reverse directional maximum matching algorithm, to conduct word omni-segmentation for defect description text in the Python programming language. The segmentation results are thereby achieved, and the segmentation results of the examples in table 1 are shown in table 3. It can be seen that the segmented words are consistent with Chinese vocabularies. This indicates that long texts were segmented into individual words accurately and explicitly.

Considering that the importance of each word in defect descriptions is different, the keywords of texts can be extracted to represent the defects in the form of weight sums more

Table 2. Parts of the three constructed dictionaries.

Component dictionary	Defect dictionary	Stop-word dictionary
Upper panel (上壁板)	Hole dislocation (孔错位)	Action (处理)
Elevator (升降舵)	Out of tolerance (超差)	Find (发现)
Power module (电源模块)	Interference (干涉)	Inspect (检查)
Flaps (襟翼)	Surface damage (表面损伤)	Engineering (工程)
Tube (导管)	Warping (翘曲)	Requirements (要求)
…	…	…

Table 3. Segmentation results of the defect description by Chinese text segmentation.

Defect description	Word omni-segmentation
在左侧壁板对接处一个多余孔, 孔径为1 mm, 具体描述见附图.	在/左侧/壁板/对接/处/一个/一个多/多余/孔, 孔径/为/1 mm, 具体/描述/见/附图. At/left/panel/butt-join/point/one/one superfluous/superfluous/hole, hole diameter/is/l mm. Details/is/in/the attached figure.

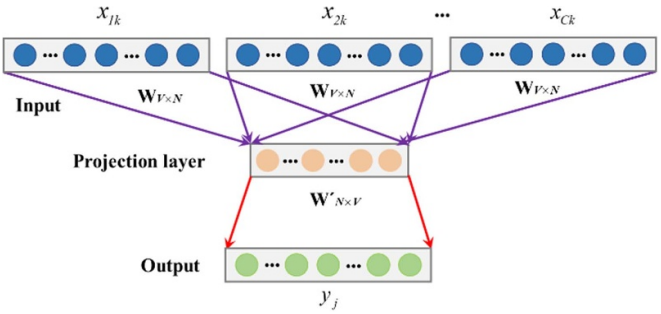


Figure 5. Illustration of word representation by the word2vec model.

concisely. Mining effectiveness is thereby improved. There are two common word frequency-based keyword extraction methods for text mining, i.e. term frequency-inverse document frequency (TF-IDF) and TextRank. Compared to TF-IDF, the TextRank algorithm is more generalized and requires no prior information. Therefore, the latter is employed here.

The TextRank algorithm adopts the relationship between adjacent words to construct a word network, which takes a word in a vocabulary sequence and its adjacent words as nodes with graph adjacent relationships. The contributions of different word-pairs are set as the edge lengths of undirected graphs. The iterative calculation formula is given by

$$\mathrm{WS}(U_i) = (1-\varsigma) + \varsigma \times \sum_{j\in \mathrm{In}(\mathrm{U}_i)} \frac{1}{\sum_{\mathrm{U_j}\in \mathrm{Out}(\mathrm{U_j})} \sigma_{jk}} WS(U_\mathrm{j}), \quad (6)$$

where $\mathrm{WS}(U_i)$, $In(U_i)$ and $\mathrm{Out}(U_i)$ denote the rank value, the precursor node set and the successor node set of nodes U_i, respectively. ς is used for smoothing. The length σ_{jk} represents the importance of two-word links. According to (6), the rank value of each node can be calculated iteratively until convergence. Finally, the edge length is used to measure the importance of keywords. When a word's importance value is larger than the predetermined threshold, it is defined as a keyword. Only the extracted keywords are used for text embedding; other words are omitted for the sake of effectiveness.

4.2. Weighted text embedding

Before text embedding, we still need to achieve the vector representations of keywords. A basic vector representation approach for vocabulary is one-hot encoding, which can transform a vocabulary into a sparse vector. However, in the process of aircraft assembly, there is a high dependency between the keywords of quality defects. Using this method will lose some associated semantic information, which leads to high complexity and dimension disaster. However, distributed word vector representation methods, which compress words into a lower dimensional space and describe the correlation levels by using the elements of vectors, can effectively address these problems.

Word2vec is a commonly used model for distributed word vector representation. It was developed based on a continuous bag of words network and the SoftMax classifier. The architecture of word2vec is shown in figure 5. This model takes a word and its adjacent words as the input and takes itself as the output. The objective function of model training is to maximize the average logarithm probability. The SoftMax classifier is used for prediction tasks. As the training process converges, the weight matrix of the projection layer can be optimized. Then, substituting words into the optimal weight matrix yields vectors that can represent the Chinese keywords. All these vectors are m-dimensional. Throughout this article, we fix $m = 50$ as a trade-off between representational power and efficiency.

When obtaining the word vectors, the text vectors for representing the defect descriptions can be constructed in two ways. One is by calculating the sum or weight sum of the word vectors. The other is by fusing the word vectors into a text vector using machine learning methods. The latter mainly considers the word order and context for text representation, but ignores the importance of technical words whose distinctive attributes are critical to defect diagnosis. Therefore, here we develop a weighted text embedding.

Table 4. Attribute annotation rules based on expert knowledge.

Importance	Category name	Annotation method
High weight, λ_1	1. Component	Manual tagging
	2. Location	Manual tagging
	3. Action	Manual tagging
Medium weight, λ_2	4. Phenomenon	Manual tagging
	5. Material	Retrieval from forms
	6. Correlation	Manual tagging
Low weight, λ_3	7. Measurement	Retrieval from forms
	8. Attachment	Retrieval from forms
	9. Others	—

Remark: $\lambda_1 > \lambda_2 > \lambda_3$, and $\lambda_1 + \lambda_2 + \lambda_3 = 1$.

To obtain the structured data that can represent the whole text accurately, we need to not only consider the word frequency and context, but also consider the importance of different kinds of words. Generally, a word with high frequency is defined as a keyword and endowed with a higher weight in the text embedding. However, some technical terms for aircraft assembly, which describe the defect sites and phenomena, have a relative smaller weight due to lower frequency. This is unreasonable. As technical terms can represent the exclusive features of quality defects better than other common words, higher weight sets for those technical terms would improve the diagnosis accuracy significantly. Hence, we divide these words into several categories according to the expert knowledge. Words from different categories with different attributes should have different importance values. Those critical for a specific defect are endowed with higher importance. Table 4 shows the attribute annotation rules.

Common part-of-speech (POS) tagging is used to distinguish nouns, verbs, adjectives and adverbs. However, in aircraft assembly defect texts, there are many professional terms that cannot be distinguished by POS. Furthermore, these professional words characterize the quality defects exclusively. Thus, there is a need to define a proprietary attribute annotation rule to classify the words segmented from defect texts and add additional semantic information to vocabulary lists. From table 4, all these words are divided into nine categories. As the components, locations and actions are the most important basis for identifying a defect, the words of these three categories should be endowed with the highest weight in the text embedding. During this phenomenon, the material and correlation of defects are endowed with a medium weight. Other categories that are not that critical to defect diagnosis are set with the lowest weight. The weight values are predetermined by expert experience. In this article, we assume $\lambda_1 = 0.5$, $\lambda_2 = 0.3$ and $\lambda_3 = 0.2$. Additionally, the annotation methods for different attribute categories are also different. Manual tagging refers to determining word categories based on expert knowledge, whereas retrieval from forms involves identifying word categories by searching the defect record datasets.

Algorithm 1. The semi-supervised annotation algorithm for keywords.

Step 1: Set $s = 0$;

Step 2: Given an un-annotated word vector $\boldsymbol{y}_j$, calculate the similarities between $\boldsymbol{y}_j$ and each center vector, which can be expressed by using the cosine similarity measure as $\cos\left\langle \boldsymbol{y}_{\mathrm{j}}, \bar{\boldsymbol{y}}_{\mathrm{l}}^{(s)} \right\rangle = \frac{\boldsymbol{y}_{\mathrm{j}} \times \bar{\boldsymbol{y}}_{\mathrm{l}}^{(s)}}{\left|\boldsymbol{y}_{\mathrm{j}}\right| \times \left|\bar{\boldsymbol{y}}_{\mathrm{l}}^{(s)}\right|}$;

Step 3: The category label for $\boldsymbol{y}_j$ is annotated as $\theta_j = a_l$, $l = \arg\max_{\mathrm{l}} \cos\left\langle \boldsymbol{y}_{\mathrm{j}}, \bar{\boldsymbol{y}}_{\mathrm{l}}^{(s)} \right\rangle$;

Step 4: Update the center word vector of the lth category, $\bar{\boldsymbol{y}}_{\mathrm{l}}^{(s+1)}$;

Step 5: Compute the maximum cosine-distance change of all centroid vectors, $\Delta = \max_{\mathrm{l}} \left(1 - \cos\left\langle \bar{\boldsymbol{y}}_{\mathrm{l}}^{(s+1)}, \bar{\boldsymbol{y}}_{\mathrm{l}}^{(s)} \right\rangle\right)$

Step 6: If $\Delta < 10^{-4}$ or $s = 20$, terminate; otherwise set $s = s + 1$, and repeat step 2–6

Note that the number of segmented words is so large that we cannot annotate the categories of all words manually within a feasible time. Therefore, here we employ a semi-supervised annotation method. That is, a small portion of words is annotated manually according to expert knowledge. Let $\bar{\boldsymbol{y}}_{\mathrm{l}}^{(0)} \in \mathbb{R}^{\mathrm{m}\times 1}$ and a_l, $\mathrm{l} = 1, 2, \cdots, 9$, denote the initial center word vector and the label of the lth category, respectively. The attribute annotation procedure for un-annotated word vectors is summarized in algorithm 1.

The per-iteration computational cost of algorithm 1 is $\mathcal{O}(NLm)$, where N is the number of unlabeled words, L the number of attribute classes and m the embedding dimension. In our implementation, the annotation process converges typically in approximate 10 iterations and the time cost shows near-linear scalability with vocabulary size.

After completing the attribute annotation, the weighted text vector can be obtained by

$$\boldsymbol{V}_{\text{text}} = \frac{\sum_{i=1}^{n} \lambda_i \boldsymbol{v}_i}{\sum_{i=1}^{n} \lambda_i}, \tag{7}$$

where $\boldsymbol{v}_i$ is the word vector, n is the word number of a text and λ_i is the attribution weight of $\boldsymbol{v}_i$. The flowchart of weighted text embedding is shown in figure 6. From figure 6, the weighted text representation, which combines word frequency, relationship between words and expert knowledge, can characterize the quality defect more specifically and exactly.

4.3. Defect diagnosis based on the SSAE

A defect diagnosis method for aircraft assembly with weighted text embedding and an SSAE network is proposed in this section. To further improve the performance of the SSAE-based classifier, we consider the sparsity of the output class

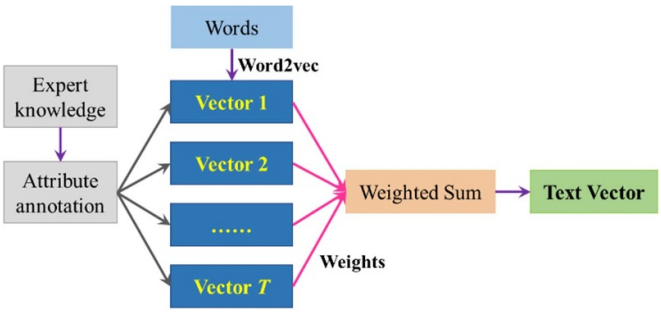


Figure 6. Text representation by attribute annotation and the word2vec algorithm.

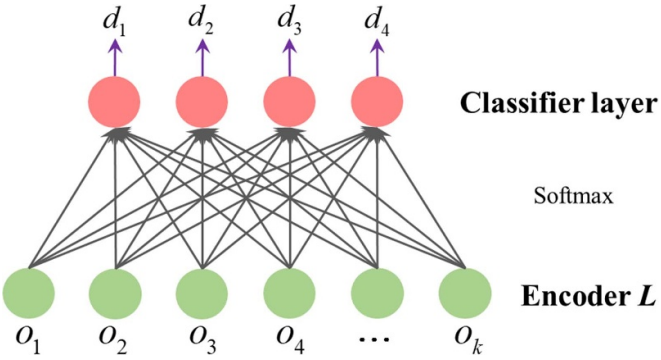


Figure 7. Detailed structure of the classification module in the SSAE.

label. Note that this diagnosis task is a multi-class classification problem. Thus, a SoftMax classifier is added to transform the output of the last encoder into the class label, as shown in figure 3. The labeled data is needed in the model training for this supervised learning.

The weighted text embedding based on the vectors of keywords is used to represent the defect description texts. Let $\boldsymbol{V}_{\text{text}} = [V_1, \cdots, V_m]$ be an m-dim text vector, and input $\boldsymbol{V}_{\text{text}}$ into the SSAE network. Let $\boldsymbol{O} = [o_1, o_2, \cdots, o_{\text{k}}]$ be a vector output of the last encoder of the SSAE, where k is the number of neurons in the encoder. Then, according to the data mapping and transformation procedure of the SSAE, the output $\boldsymbol{O}$ can be calculated as follows

$$o_k = \sigma\left(\sigma\left(\boldsymbol{W}_2 h_1\left(\boldsymbol{V}_{\text{text}}\right) + b_1\right) + \cdots\right),\ h_j(x) = \sigma\left(\boldsymbol{W}_j x + b_j\right), \tag{8}$$

where j = 1, 2, $\cdots$, L and L is the number of hidden layers. The vector $\boldsymbol{O}$ can be regarded as a dimension-reduced realization of text representation by the multi-layer encoders. The detailed structure of the classification module in the SSAE is shown in figure 7, which assumes four types of quality defects as an example.

During the classifier process, the output $\boldsymbol{O}$ of the encoders is mapped to be a scalar d_{i}, $i = 1, \cdots, 4$. These scalars form the class labels. Based on the coding sparsity, the output vector $[d_1, d_2, d_3, d_4]$ for defect text vector V_{text} is represented by a binary number. That is, we can use $[1,0,0,0]$, $[0,1,0,0]$, $[0,0,1,0]$ and $[0,0,0,1]$ to denote the first type (label $(\boldsymbol{V}_{\text{text}}) = 1$), the second type (label $(\boldsymbol{V}_{\text{text}}) = 2$), the third type (label $(\boldsymbol{V}_{\text{text}}) = 3$) and the fourth type (label $(\boldsymbol{V}_{\text{text}}) = 4$) of quality defect, respectively. Through the classification module, the actual output $\tilde{d}_i$ is expressed by using the SoftMax function and weight $\boldsymbol{W}_{c,i}$ as

$$\tilde{d}_i = \frac{\exp\left(-\boldsymbol{W}_{c,i}{}^{\text{T}}\boldsymbol{O}\right)}{\sum_{i=1}^{4}\exp\left(-\boldsymbol{W}_{c,i}{}^{\text{T}}\boldsymbol{O}\right)},\ i = 1, \cdots, 4. \tag{9}$$

When defect samples are equally sampled from the four types and there are $N/4$ in each class, the ideal output for all samples in the output node d_i can be shown as

$$\begin{aligned}
\boldsymbol{d}_1 &= \left[\underbrace{1,\cdots,1}_{N/4},\underbrace{0,\cdots,0}_{N/4},\underbrace{0,\cdots,0}_{N/4},\underbrace{0,\cdots,0}_{N/4}\right]\\
\boldsymbol{d}_2 &= \left[\underbrace{0,\cdots,0}_{N/4},\underbrace{1,\cdots,1}_{N/4},\underbrace{0,\cdots,0}_{N/4},\underbrace{0,\cdots,0}_{N/4}\right]\\
\boldsymbol{d}_3 &= \left[\underbrace{0,\cdots,0}_{N/4},\underbrace{0,\cdots,0}_{N/4},\underbrace{1,\cdots,1}_{N/4},\underbrace{0,\cdots,0}_{N/4}\right]\\
\boldsymbol{d}_4 &= \left[\underbrace{0,\cdots,0}_{N/4},\underbrace{0,\cdots,0}_{N/4},\underbrace{0,\cdots,0}_{N/4},\underbrace{1,\cdots,1}_{N/4}\right]
\end{aligned}. \tag{10}$$

Equation (10) reflects the sparsity in the output of the classification module.

Note that, if a sample belongs to the ith defect type, the output $\tilde{d}_i$ should be equal to 1. Otherwise, $\tilde{d}_i$ is equal to 0. However, in real diagnosis scenarios, the predicted class label is not exactly equal to 1 or 0 due to data noise. Hence, for a new defect sample with text vector $\widetilde{\boldsymbol{V}}_{\text{text}}$, the identified type is designed as

$$\text{label}\left(\widetilde{\boldsymbol{V}}_{\text{text}}\right) = \arg\max_i\left(\tilde{d}_i\right),\ i = 1, \cdots, 4. \tag{11}$$

Before implementing the defect diagnosis using the SSAE, a model training and a classification process is needed. The historical quality defect records are organized to constitute the training dataset $\boldsymbol{T}$, which is in the form of Chinese text. Then, $\boldsymbol{T}$ and the corresponding class labels are used to train an SSAE network. In the test process, the type of new defect data is identified by equation (11). The detailed procedure for the intelligent diagnosis framework is summarized below.

Input	Training dataset $\boldsymbol{T}$, class label, new defect data $\widetilde{\mathrm{t}}$, dictionaries of technical terms.
Step 1:	Conducting Chinese word segmentation for the defect descriptions $\boldsymbol{T}$ based on the Jieba method and dictionaries.
Step 2:	Extracting the keywords from the segmented words by using the TextRank algorithm. Calculating the representation vectors of the keywords with the word2vec algorithm.
Step 3:	Annotating the attributes of keywords using algorithm 1 and assigning weights for each attribute with expert knowledge. Obtaining the text vectors based on the weighted sum of word vectors $\boldsymbol{V}=\left\{\boldsymbol{V}_{\text{text}}{}^{(\text{n})}\right\}$. The training dataset $\boldsymbol{T}$ in the form of Chinese text is transformed into vectors $\boldsymbol{V}$.
Step 4:	Training the SSAE-based classifier with the sample set $\boldsymbol{V}$ and their class labels. First, the initial parameters of the encoders and decoders are estimated by pre-training. Second, the final weights of encoders and classifier are obtained based on gradient descent algorithms.
Step 5:	For a new defect data $\widetilde{\mathrm{t}}$, transforming it into vector $\widetilde{\boldsymbol{V}}_{\text{text}}$ based on Steps 1–3. Input $\widetilde{\boldsymbol{V}}_{\text{text}}$ into the trained SSAE to generate an output as $\left[\tilde{d}_1,\tilde{d}_2,\tilde{d}_3,\tilde{d}_4\right]$.
Output:	Defect type label of $\widetilde{\mathrm{t}}$, i.e. label $\left(\widetilde{\mathrm{t}}\right)=\arg\max_i\left(\tilde{d}_i\right)$, $i=1,\cdots,4$.

Compared with traditional natural language processing, the proposed intelligent diagnosis framework integrates field and expert knowledge and deep learning algorithms, which can make the text embedding more accurate and the output more accurate and reliable diagnosis results. With these diagnosis results, quality engineers can make maintenance actions and quality improvements effectively and quickly.

5. Experiment verification

5.1. Dataset

In order to verify the effectiveness and accuracy of the proposed intelligent diagnosis framework, practical aircraft assembly-related data is acquired from the assembly line of an aircraft manufacturer. There are five types of quality defects in the dataset. Each type has 500 defect records and thereby the total size is 2500. For each defect type, we take 400 records as the training dataset and the other 100 as the testing dataset. Moreover, the class label for each defect type is represented by sparse coding. Details of the dataset are shown in table 5.

Table 5. Details of the quality defect dataset.

Defect type	Abbreviation	Number of records	Class label
Oversize (尺寸超差)	OV	500	[1, 0, 0, 0, 0]
Assembly clearance (装配间隙)	AC	500	[0, 1, 0, 0, 0]
Hole mis-spacing (孔排列错误)	HS	500	[0, 0, 1, 0, 0]
Surface damage (表面损伤)	SD	500	[0, 0, 0, 1, 0]
Crack (裂纹)	CR	500	[0, 0, 0, 0, 1]

5.2. Methods for comparison

To demonstrate the effectiveness of the SSAE classifier, commonly used diagnosis algorithms, including SVM (support vector machine), ANN (artificial neural network), DNN (deep neural network), LSTM, BiLSTM, and SAE, are used for defect diagnosis in aircraft assembly. The classification results obtained by these six algorithms are compared with the diagnosis results from the SSAE.

In general natural language processing or text mining research, deep learning methods have become more and more popular recently. Considering the temporal characteristics of speech and text, LSTM and BiLSTM, which are RNN, are two widely used deep learning methods for processing serialized information and speech recognition. Without losing generality, the traditional classification method, the SVM, is also used for comparison here. The kernel function of the SVM is set to a radial basis function. The nodes of the input layer of these network-based methods are 50, and the number of nodes in the output layer are 5. In an ANN, only one hidden layer is considered, while two hidden layers are considered in DNNs. The BiLSTM consists of two LSTM networks in bi-direction. A softmax layer is taken as the output of these networks. A dropout layer is added before the softmax layer for the LSTM, BiLSTM and DNN to avoid overfitting. All batch sizes and epochs are set as 128 and 100, respectively. The SAE is used to show the development necessity from the SAE to the SSAE.

5.3. Results

Since all historical quality defect records are unstructured data and presented in Chinese text, we need to transform the raw data into structured data. First, based on the defect, general and stop-word dictionaries, the defect records in continuous text are segmented into vocabulary lists with the Chinese word segmentation algorithm. Then, the word2vec model is employed for word vector representation. After attribute annotation, the obtained word vectors are integrated into text vectors by weighted text embedding. Table 6 shows several examples of the processed data. All the vectors are 50-dimensional, but some elements are omitted for better presentation. Then, a vector and its corresponding class label form a pair of training data, in which the former is the input and the latter is the output.

After obtaining the text vectors as in table 6, the SSAE-based classifier can be determined by supervised learning. A total of 2000 data samples (400 for each defect type) are used to train the SSAE network and 500 data samples (100 for

Table 6. Examples of the processed data format for the defect records by weighted text embedding.

No.	Class label	Vector
1	[1, 0, 0, 0, 0]	[0.6572463 , 0.10706615, 0.05572755, −0.754483, 0.5369991, 1.0088301, ……, 0.6806897, −0.00469302, 1.2699736, −0.13783777, 0.11444575, 0.24955863]
2	[0, 1, 0, 0, 0]	[0.7542916 , −0.01749352, −0.36677638, −0.4563792, −0.4995486, −0.281815, ……, 0.55768025, 0.21790005, 1.2711376, 0.06936119, 0.06145719, 0.44860455]
3	[0, 0, 1, 0, 0]	[0.75164133, 0.1171543, −0.24626656, −0.56207055, −0.7033588, −0.145031, ……, 0.62114483, 0.03707135, 1.4960561, 0.11143543, 0.08778881, 0.23797192]
4	[0, 0, 0, 1, 0]	[0.7957793, −0.04374009, −0.1335192, −0.48980123, −0.80299264, −0.25000, ……, 0.43382773, 0.07181711, 1.3889766, 0.21212052, 0.01927389, 0.29392081]
5	[0, 0, 0, 0, 1]	[0.6246822, 0.11401705, −0.15698974, −0.791429, −0.44586647, −0.17210977, …… 0.33895493, 0.17126107, 1.2494333, −0.02494708, 0.03535834, 0.2714129]

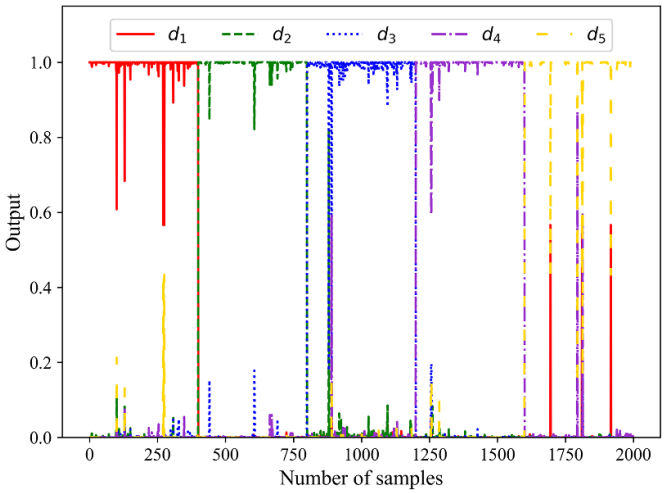


Figure 8. Output of the SSAE for the training dataset.

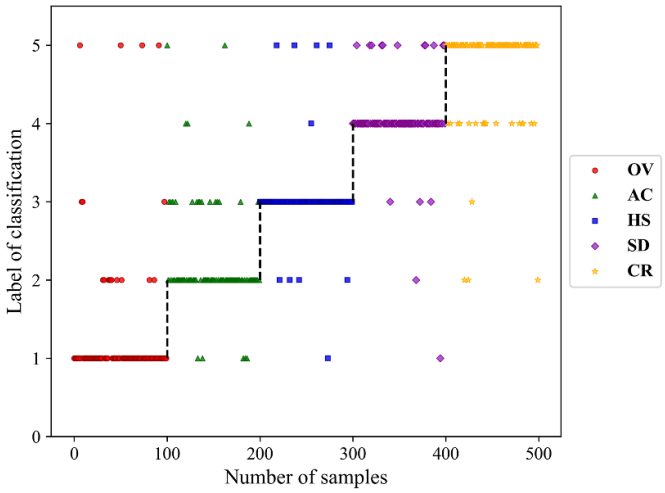


Figure 9. Classification result of quality defects by the SSAE.

each defect type) are taken as the testing samples to verify the effectiveness of the diagnosis framework. To design the SSAE, each 50-dimensional vector sample is input into the basis module of the SSAE, and thus the input layer has 50 nodes. There are five nodes in the output layer corresponding to five defect types, which are respectively denoted by d_1, d_2, d_3, d_4 and d_5. During the training process, the number of basis modules in the SSAE is set to three, the epoch size is 100 and the batch size is 128. Hidden sizes $256 \rightarrow 96 \rightarrow 16$ are selected after a two-stage process combining heuristic scaling and a cross-validated grid search. The text vectors and corresponding class labels are used to train the SSAE network. Figure 8 shows the output of the training dataset using the SSAE. It is found that the sparsity of the class labels is clear, and most of the outputs of the five nodes can indicate the defect type of training samples accurately.

To identify defect types for the 500 testing samples, the text vectors of these samples are input into the trained defect diagnosis framework and the corresponding output can be calculated based on the classification process shown in table 5. The classification results of these testing samples are displayed in figure 9, whose horizontal axis and vertical axis represent the true class labels of the testing samples and the estimated class labels, respectively.

Most samples are correctly classified. For example, for the 1st-100th samples, their true defect type is OV, and obviously most samples in 1st-100th are identified as defect type OV. Furthermore, the confusion matrix of the diagnosis results by the SSAE is shown in figure 10.

For example, the (5, 1)*th* element in the matrix means 83 samples belonging to class OV are classified correctly and the (5, 2)*th* element means that 10 samples belonging to class OV are misclassified as class AC. The total diagnosis accuracy of the SSAE-based classifier is 82.2%, which means effectiveness in practical applications with only a small amount of data. For defect type HS (hole spacing), 90% of these samples can be identified correctly. Meanwhile, the misclassified rates between classes SD and CR are more than other pairs of defects. In addition, 10% samples of class OV and 15% samples of class AC are misclassified as classes AC and

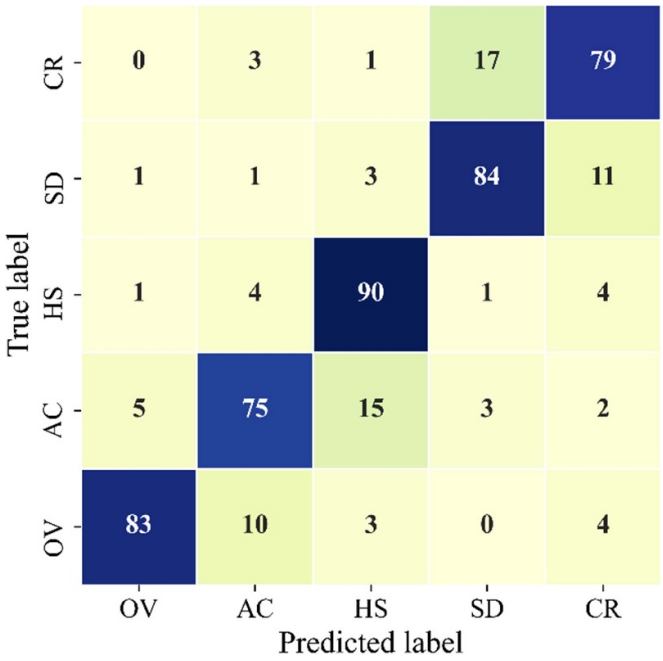


Figure 10. Confusion matrix of the diagnosis results for the five defect types by the SSAE.

HS, respectively. This is because similarity between the segmented words of these defect types is higher, or the attributes of some vocabularies is closer. To further illustrate the effectiveness of the SSAE classifier, the confusion matrix results for comparative methods are shown in figure 11.

From the confusion matrices of the diagnosis results by different diagnosis methods, we can see that the diagnosis performances of the LSTM and BiLSTM are not good. This is contrary to the consensus that LSTM-based algorithms are more suitable to deal with temporal data, especially in speech recognition. Although the network structures of the LSTM and BiLSTM are more complex than those of other methods, they are prone to over-fitting because of insufficient quality defect data. Hence, for many engineering applications in which it is hard to collect big data, an appropriate non-complex model would be preferred. The DNN with two hidden layers here is the simplest deep learning method, as it achieves better results than the other methods in figure 11. This indicates that a relatively simple model can avoid overfitting effectively and thereby lead to better diagnosis effect. The diagnosis accuracy of the SVM is at an intermediate level among these six methods. This suggests that traditional methods can still be used in practical situations due to their simple structure and low computational cost.

Detailed information about the misclassified samples for all five quality defects is presented in table 7. The total classification accuracies of the SVM, ANN, DNN, LSTM, BiLSTM and SAE are 78.6%, 80.6%, 81.6%, 70.0%, 74.6% and 81.2% respectively. Obviously, the accuracy generated by these six methods is lowet than that of the SSAE, which indicates that the SSAE with accuracy of 82.2% performs better than the other diagnosis methods. The comparative results further validate the effectiveness and superiority of the SSAE for quality defect diagnosis.

5.4. Discussion

Having established the SSAE's superior aggregate performance, we now analyze its behavior in detail. Some example comparisons of word segmentation results between different defect types are shown in figures 12–14.

The words in red are the same, the words in purple are all attributed with the high weight λ_1. The left part of figure 12 represents that sample 1 belonging to class OV is misclassified into class AC, while the right part presents that sample 2 belonging to class AC is identified correctly. From the comparison in figure 13, we can see that some words are the same in these two samples, and some other keywords with high weights have similar locations in the sentence text. Hence, sample 1, which is close to the characteristics of class AC but deviates from the main characteristics of class OV, is easily misclassified. From the comparison between sample 3 and sample 4, the similar conclusion can also be drawn. In addition, it is interesting to find that the number of misclassified samples from AC to OV, and from HS to AC, are much less. This is because the features of most samples in class AC are different from the main characteristics of class OV, so are HS and AC. According to figure 14, both misclassification rates between class SD and class CR are higher. Comparing sample 5 with sample 6 in figure 14, we can find that the defect descriptions are similar, especially the sentence structures. Hence, these two defect types are easy to confuse.

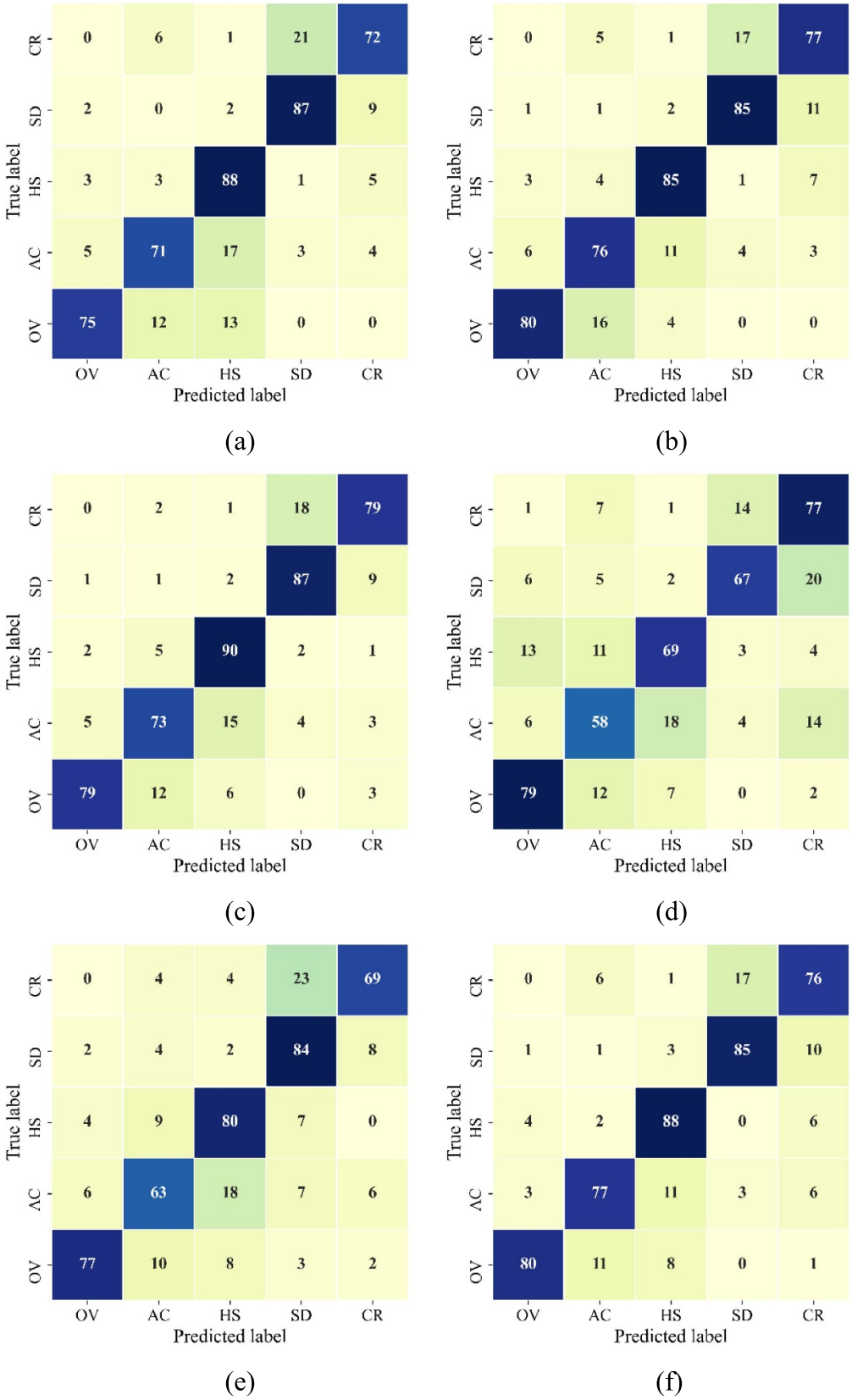


Figure 11. Confusion matrix of the classification results obtained by different methods for quality defects: (a) SVM, (b) ANN, (c) DNN, (d) LSTM, (e) BiLSTM, (f) SAE.

Table 7. Number of misclassified samples for the five quality defects and the total diagnostic accuracy of seven different methods.

Defect types	Diagnosis method						
	SVM	ANN	DNN	LSTM	BiLSTM	SAE	SSAE
OV	25	20	21	21	23	20	17
AC	29	24	27	32	37	23	25
HS	12	15	10	31	20	12	10
SD	13	15	13	33	16	15	16
CR	28	23	21	23	31	24	21
Accuracy	78.6%	80.6%	81.6%	70.0%	74.6%	81.2%	**82.2%**

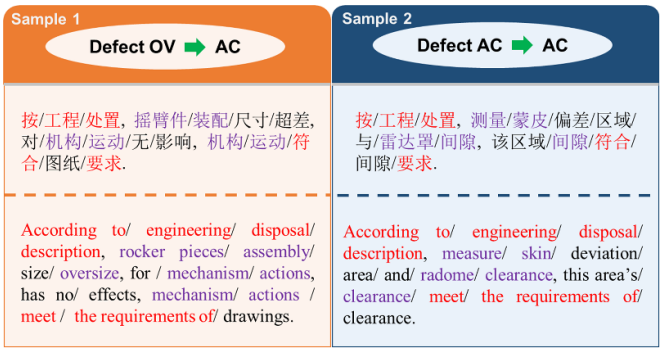


Figure 12. Comparison of the word segmentation result of defects OV and AC.

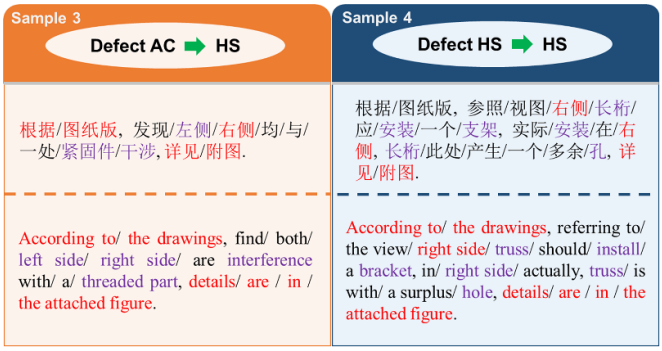


Figure 13. Comparison of the word segmentation result of defects AC and HS.

Nevertheless, the diagnosis performance by the SSAE network is satisfactory overall. Practically, we still need to improve diagnostic accuracy by extending the framework to handle more categories, rarer defects and fresher domain-specific terminology. To achieve this, three lightweight enhancements should be conducted: (i) the SoftMax layer can be freely resized to any number of categories, while focal-loss training and synonym-based data augmentation will compensate for class imbalance; (ii) the semi-supervised annotation step only needs a handful of seed words to initialize a new class centroid, so incorporating additional defect types incurs minimal manual effort; and (iii) continuous expansion of the technical thesaurus, supplemented by sub-word embedding, will keep the out-of-vocabulary rate low, further stabilizing performance when new, domain-specific

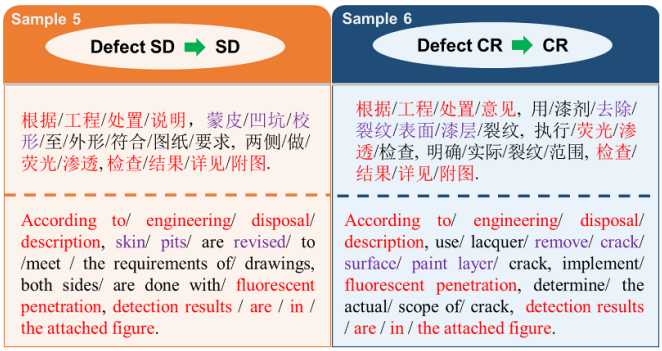


Figure 14. Comparison of the word segmentation result of defects SD and CR.

terminology emerges. These measures require no re-design of the SSAE backbone yet provide a clear potential pathway to higher accuracy and broader applicability.

6. Conclusions and future work

An intelligent diagnosis framework for quality defects in aircraft assembly was proposed. The quality defect records in the form of long Chinese text were segmented into individual words, which were then transformed into vectors by word2vec. Expert knowledge-based semi-supervised attribute annotation was used to give weights to the keywords. Then, the text vectors were obtained by the weighted sum of word vectors. Furthermore, an SSAE network was developed for text classification. The effectiveness and generalization ability of the proposed framework was illustrated by practical aircraft assembly-related data from the assembly line of an aircraft manufacturer.

Although the framework accurately diagnoses quality defects, the numerical information embedded in some text records is currently ignored, leaving part of the data underutilized. Future work will therefore (i) mine these mixed-type data to improve diagnostic accuracy, (ii) continually enlarge the technical-term dictionary through structured expert elicitation to refine keyword weighting, (iii) distill a domain-adapted Deepseek-style model to a sub-1-billion-parameter backbone that can run on-premise, ground it in an aircraft-assembly defect knowledge graph, and then the hybrid approach can be evaluated by (a) benchmarking F1-score, precision, and recall against the current SSAE on a held-out dataset, and (b) conducting an expert-panel assessment of explanation quality before and after language-model integration, and (iv) replace the deterministic SSAE with a variational AE so the system can output calibrated uncertainty intervals alongside defect classes, which are valuable for quality-assurance decisions.

Acknowledgment

This work was supported in part by Natural Science Foundation of Shanghai Municipality 23ZR1428100, and in part by Postdoctoral Fellowship Program of CPSF under Grant

GZC20240989, and China Postdoctoral Science Foundation under Grant 2024M761994.

ORCID iDs

Zirong Wang 0000-0003-0268-3709
Zhen Chen 0000-0003-2590-0307
Tangbin Xia 0000-0001-9121-1716

References

Compare M and Zio E 2025 Opportunities and risks of artificial intelligence for industry 5.0 in the context of reliability and maintenance engineering *J. Reliab. Sci. Eng.* **1** 015001

Deng S, Li Q, Dai R, Wei S, Wu D, He Y and Wu X 2024 A Chinese text classification based on active *Appl. Soft Comput.* **150** 111067

Gao F, Zhang L, Wang W, Zhang B, Liu W, Zhang J and Xie L 2024 Named entity recognition for equipment fault diagnosis based on RoBERTa-wwm-ext and deep learning integration *Electronics* **13** 3935

Hu Z, Zhang X and Xiong H 2024 Two-stage attention network for fault diagnosis and retrieval of fault logs *Expert Syst. Appl.* **249** 123365

Hui Y, Mei X, Jiang G, Zhao F, Ma Z and Tao T 2022 Assembly quality evaluation for linear axis of machine tool using data-driven modeling approach *J. Intell. Manuf.* **33** 753–69

Jiang K, Gao X, Gao H, Han H, Qi Y and Han H 2025 A zero-shot fault diagnosis framework for chillers based on sentence-level text attributes *IEEE Trans. Ind. Inform.* **21** 2878–87

Jing X, Wu Z, Zhang L, Li Z and Mu D 2024 Electrical fault diagnosis from text data: a supervised sentence embedding combined with imbalanced classification *IEEE Trans. Ind. Electron.* **71** 3064–73

Kim S, Park H and Lee J 2020 Word2vec-based latent semantic analysis (W2V-LSA) for topic modeling: a study on blockchain technology trend analysis *Expert Syst. Appl.* **152** 113401

Liao Z, Zhan K, Zhao H, Deng Y, Geng J, Chen X and Song Z 2024 Addressing class-imbalanced learning in real-time aero-engine gas-path fault diagnosis via feature filtering and mapping *Reliab. Eng. Syst. Saf.* **249** 110189

Liu G and Guo J 2019 Bidirectional LSTM with attention mechanism and convolutional layer for text classification *Neurocomputing* **337** 325–38

Ma Y, Zheng S, Yang Z, Pan H and Hong J 2025 A knowledge-graph enhanced large language model-based fault diagnostic reasoning and maintenance decision support pipeline towards industry 5.0 *Int. J. Prod. Res.* 1–22

Mikolov T, Chen K, Corrado G and Dean J 2013 Efficient estimation of word representations in vector space (arXiv:1301.3781)

Nan J, Deng B, Cao W, Hu J, Chang Y, Cai Y and Zhong Z 2022 Big data-based early fault warning of batteries combining short-text mining and grey correlation *Energies* **15** 5333

Park M and Park D H 2025 Optimization of a two-dimensional warranty policy for a multi-component system with consideration of periodic preventive maintenance action *J. Reliab. Sci. Eng.* **1** 025004

Peng H and Yang W 2024 Knowledge graph construction method for commercial aircraft fault diagnosis based on logic diagram model *Aerospace* **11** 773

Shafi I, Mazhar M F, Fatima A, Alvarez R M, Miro Y, Espinosa J C M and Ashraf I 2023 Deep learning-based real time defect detection for optimization of aircraft manufacturing and control performance *Drones* **7** 31

Shang H, Wu J, Sun C, Liu J, Chen X and Yan R 2023 Global prior transformer network in intelligent borescope inspection for surface damage detection of aeroengine blade *IEEE Trans. Ind. Inform.* **19** 8865–77

Sheikh A A, Chong E K P and Simske S J 2024 Enhancing defect detection in circuit board assembly using AI and text analytics for component failure classification *IEEE Trans. Compon. Packag. Manuf. Technol.* **14** 1881–90

Tao C, Pan H, Li Y and Zou Z 2015 Unsupervised spectral–spatial feature learning with stacked sparse autoencoder for hyperspectral imagery classification *IEEE Geosci. Remote Sens. Lett.* **12** 2438–42

Wang J, Gao S, Yu L, Zhang D and Kou L 2022 Defect severity identification for a catenary system based on deep semantic learning *Sensors* **22** 9922

Wang J, Li G, Bai H, Yuan G, Li X, Lin B, Zhong L and Zhang X 2024 A hybrid deep learning-based framework for chip packaging fault diagnostics in x-ray images *IEEE Trans. Ind. Inform.* **20** 11181–91

Wen L, Gao L and Li X 2017 A new deep transfer learning based on sparse auto-encoder for fault diagnosis *IEEE Trans. Syst. Man Cybern. Syst.* **49** 136–44

Xu J, Xiang L, Liu Q, Gilmore H, Wu J, Tang J and Madabhushi A 2015 Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images *IEEE Trans. Med. Imaging.* **35** 119–30

Xu W and Tan Y 2019 Semi-supervised text classification by variational autoencoder *IEEE Trans. Neural Netw. Learn. Syst.* **31** 295–308

Xu Z, Chen B, Zhou S, Chang W, Ji X, Wei C and Hou W 2021 A text-driven aircraft fault diagnosis model based on a Word2vec and priori-knowledge convolutional neural network *Aerospace* **8** 112

Yan Q and Wang H 2022 Double-layer Q-learning-based joint decision-making of dual resource-constrained aircraft assembly scheduling and flexible preventive maintenance *IEEE Trans. Aerosp. Electron. Syst.* **58** 4938–52

Yang B, Lei Y, Jia F, Li N and Du Z 2020 A polynomial kernel induced distance metric to improve deep transfer learning for fault diagnosis of machines *IEEE Trans. Ind. Electron.* **67** 9747–57

Yang N and Zhang Y 2022 Railway fault text clustering method using an improved Dirichlet multinomial mixture model *Math. Probl. Eng.* **2022** 7882396

Ye Z, Guo B and Qi D 2023 BERT-TriF: an inductive short text classification model for power equipment defect records *Eng. Rep.* **5** e12671

Yu M, Quan T, Peng Q, Yu X and Liu L 2021 A model-based collaborate filtering algorithm based on stacked autoencoder *Neural Comput. Appl.* **34** 2503–11

Yu Z, Wei Y, Niu B and Zhang X 2024 Automatic condition monitoring and fault diagnosis system for power transformers based on voiceprint recognition *IEEE Trans. Instrum. Meas.* **73** 1–11

Zhang Q, Zhang Y, Luo Q, Yu C, Yu N, Wang Q and Ke Y 2024 Cloud-edge-end-based aircraft assembly production quality monitoring system framework and applications *J. Manuf. Syst.* **75** 116–31

Zhang X, Wang T, Ma L and Mahadevan S 2025 Reliability engineering, risk management, and trustworthiness assurance for AI systems *J. Reliab. Sci. Eng.* **1** 022001

Zhou S, Wei C, Li P, Liu A, Chang W and Xiao Y 2021 A text-driven aircraft fault diagnosis model based on Word2vec and stacking ensemble learning *AEROSPACE* **8** 357

Zhou Y, Xu B, Xu J, Yang L and Li C 2016 Compositional recurrent neural networks for Chinese short text classification *2016 IEEE/WIC/ACM Int. Conf. on Web Intelligence (WI)* pp 137–44



Zirong Wang received the BS degree in Quality and Reliability Engineering from Beihang University in 2022. In the Department of Industrial Engineering & Management, Shanghai Jiao Tong University, he received the MS degree in Industrial Engineering in 2025 and continued to pursue a PhD degree.

His current research interests include reliability modeling, battery health management.



Zhen Chen received the BS degree in Quality and Reliability Engineering from Beihang University in 2014 and a PhD degree in Mechanical Engineering (Industrial Engineering) from Shanghai Jiao Tong University in 2019. He is now an assistant professor in the Department of Industrial Engineering & Management, Shanghai Jiao Tong University, China. He worked at the Department of Energy, Politecnico di Milano, Italy, as a postdoc.

His research interests include reliability modeling, condition monitoring and fault prognostics, battery health management.



Zhaoxiang Chen received the BS degree in industrial engineering from the University of Electronic Science and Technology of China in 2017, the MS degree in control science and engineering from Beihang University in 2020, and a PhD degree in Mechanical Engineering (Industrial Engineering) from Shanghai Jiao Tong University in 2023.

He is currently a Postdoctoral with the Department of Industrial Engineering & Management, Shanghai Jiao Tong University. He is currently a Visiting Scholar with the Department of Industrial Systems Engineering and Management, National University of Singapore. His research interests include reliability assessment and maintenance decisions for manufacturing systems.



Tangbin Xia received the PhD degree in Mechanical Engineering (Industrial Engineering) from Shanghai Jiao Tong University in 2014 and was a co-cultured PhD with the University of Michigan. Currently, he is an Associate Professor and serves as Deputy Director of the Department of Industrial Engineering at Shanghai Jiao Tong University.

He worked at ISyE, Georgia Institute of Technology as a Visiting Scholar. His major research interests are intelligent maintenance systems, prognostics & health management, and advanced manufacturing.



Ershun Pan received BS and MS degrees in Mechanical design and manufacturing from Northeastern University, Shenyang, Liaoning, China, in 1997 and the PhD degree in mechanical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2000.

He is now a professor and the header in the Department of Industrial Engineering & Management, Shanghai Jiao Tong University, China. His current research interests include quality control, reliability engineering and maintenance policy, production system planning and design, and lean manufacturing technology.