

Research Article

Data Anomaly Detection through Semisupervised Learning Aided by Customised Data Augmentation Techniques

Xiaoyou Wang,¹ Yao Du,¹ Xiaoqing Zhou ,² and Yong Xia ¹

¹Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China

²College of Civil and Transportation Engineering, Shenzhen University, Shenzhen, China

Correspondence should be addressed to Xiaoqing Zhou; xqzhou@szu.edu.cn

Received 21 January 2023; Revised 2 July 2023; Accepted 14 July 2023; Published 22 July 2023

Academic Editor: Lucia Faravelli

Copyright © 2023 Xiaoyou Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Structural health monitoring (SHM) systems may suffer from multiple patterns of data anomalies. Anomaly detection is an essential preprocessing step prior to the use of monitoring data for structural condition assessment or other decision making. Deep learning techniques have been extensively used for automatic category classification by training the network with labelled data. However, because the SHM data are usually large in quantity, manually labelling these abnormal data is time consuming and labour intensive. This study develops a semisupervised learning-based data anomaly detection method using a small set of labelled data and massive unlabelled data. The MixMatch technique, which could mix labelled and unlabelled data using MixUp, is adopted to enhance the generalisation and robustness of the model. A unified loss function is defined to combine information from labelled and unlabelled data by incorporating consistency regularisation, entropy minimisation, and regular model regularisation items. In addition, customised data augmentation strategies for time series are investigated to further improve the model performance. The proposed method is applied to the SHM data from a real bridge for anomaly detection. Results demonstrate the superior performance of the developed method with very limited labelled data, greatly reducing the time and cost of labelling efforts compared with the traditional supervised learning methods.

1. Introduction

Structural health monitoring (SHM) systems may suffer from multiple patterns of data anomalies caused by sensor faults, communication interference, system malfunctions, and harsh operational or environmental conditions [1–3], making the automatic and reliable structural condition assessment challenging. Therefore, detecting these anomalies is an essential preprocessing step for SHM systems. In recent years, machine learning (ML) and deep learning (DL) techniques have been widely used in the field of SHM for automatic big data processing, such as data compression and recovery [4, 5], response prediction [6, 7], abnormal data detection [8–14], and image-based crack detection [15, 16]. Most studies adopted the supervised learning, that is, the models are trained using the labelled data only. Bao and Li [1] summarized the recent developed supervised learning methods for data anomaly detection. As the representative

studies, Tang et al. [2] developed the convolutional neural network-based anomaly detection method by transforming the time series SHM data into images for model training. Du et al. [8] employed the ResNet18 as the feature extractor to detect anomaly-sensitive features for classification in the context of imbalanced classes [8]. Zhang et al. [3] developed the support vector data description-based method to detect data anomalies caused by sensor faults and extreme events. However, sufficient labelled data are difficult to obtain in many cases because anomalous data are difficult to collect and label manually from oceans of data.

Facing the challenge of limited labelled data, augmenting the quantity of training data is an effective strategy to alleviate model overfitting and enhance model generalisation performance [17, 18]. Data augmentation (DA) is a widely used strategy that leverages existing samples to create new ones through a series of transformations based on the prior knowledge about the invariant characteristics of data against

specific transformations. Enhancing the quantity and quality of training data can facilitate covering unexplored input data space and expanding the decision boundary of the model. DA has been a common and mature technique for application to general image recognition tasks, such as image cropping, mirroring, colour augmentation, scaling, or translation [19–21]. Nevertheless, standard procedures for data anomaly detection, which could be regarded as a type of time series recognition method, have not been well established. The main challenge for DA of time series is to design appropriate operators that can simultaneously generate representative new samples and guarantee that these new samples belong to the same class. Not all transformations are applicable to the time series datasets because of the diversity and complexity of the data samples. Therefore, particular domain knowledge is required for the time series DA.

Another strategy is to take full advantage of limited labelled and sufficient unlabelled data for model training. Semisupervised learning (SSL) is a branch between supervised learning and unsupervised learning. SSL unifies the utilisation of large amounts of available unlabelled data and typically much smaller sets of labelled data; it is particularly suitable when labelled data are extremely scarce to construct a reliable classifier. The underlying precondition of SSL is that the posterior distribution $p(y|x)$ could be inferred with the utilisation of useful information of the marginal distribution $p(x)$ over the input data space [21]. In this case, unlabelled data could be used to integrate information about $p(x)$ and then about $p(y|x)$. This precondition is satisfied in many real-world conditions [21]. Compared with those models trained using limited labelled data only, SSL could help improve the model performance by allowing sufficient unlabelled data to participate in model training.

SSL can be generally grouped into transductive and inductive methods [21, 22]. The former infers the label for unlabelled data directly using graph-based principles, where no classification models are produced. By contrast, the latter aims to construct a reliable classifier to make predictions on any data in the input space, exhibiting high adaptation and generalisation performance, and has arisen a lot of research attention. Inductive methods can be further classified into multiple categories based on how they incorporate unlabelled data. Generic regularisation, entropy minimisation, and consistency regularisation are three widely used techniques to define the unlabelled loss function [21, 22]. The generic regularisation item is usually incorporated in the unified loss function to alleviate model overfitting and improve generalisation performance to unseen data [23]. Entropy minimisation is based on the low-density assumption that the decision boundary should not pass through the high-density areas of input data [24]. Several approaches to achieve low-entropy predictions include directly minimising the entropy items, constructing one-hot labels, or using sharpening functions to force the classifier to output low-entropy predictions on the unlabelled data [24]. Consistency regularisation is correlated with smooth assumption, that is, a small distortion to the input data should not change the distribution of model output [25].

In SHM data anomaly detection, a large amount of unlabelled data may be available. However, current supervised learning-based methods only utilise the information from labelled data for model training. Few existing studies employ SSL methods to exploit the information from unlabelled data [26]. Besides, DA techniques are simply applied by researchers as the data preprocessing tools [27, 28], whereas the combined effects of DA and SSL techniques in improving the model performance for SHM data anomaly detection have not been investigated. Theoretically, DA techniques can assist in the SSL method to some extent. For example, a straightforward approach to add distortion in consistency regularisation-based SSL method is DA, which applies transformations to the input data and assumes that relevant class semantics are unchanged [21]. Nevertheless, due to the diversity and complexity of SHM time series data, DA strategies used in image DA may be inapplicable. Hence, it is worthwhile to develop effective customised DA techniques for SHM time series data.

In this regard, this study develops a novel DL-based data anomaly detection method by fully utilising both labelled and unlabelled data for model training. DA, entropy minimisation, and consistency regularisation strategies are incorporated into the SSL framework to improve model performance. Particularly, customised DA techniques are designed for SHM time series data. The contribution of DA volume and the quantity of labelled data to improve model performance are investigated. The proposed method is applied to a long-span cable-stayed bridge for acceleration data anomaly detection.

2. DA for Time Series

DA for time series can be categorised into four main classes, including random transformation, decomposition, pattern mixing, and generative models [17]. Random transformation is a commonly used method by adding random noise, scaling, random wrapping in time or magnitude dimension, permutation, and window sliding. Decomposition methods extract representative independent features from the original time series, such as trend and season components, and then generate new samples based on extracted features. Pattern mixing is operated by integrating two or more time series to generate new ones. For generative models, the distributions of the extracted features are established and then leveraged to generate new patterns, of which the generative adversarial network is a representative. Amongst all four categories, random transformation-based method is the simplest yet effective; thus, it is adopted in this study. Specifically, the method leverages specific transformation functions $g(\cdot)$ to generate sample x' by $x' \leftarrow g(x)$, where x is a time series $x = x_1, \dots, x_t, \dots, x_T$ with T time steps. Table 1 lists six random transformation-based DA techniques, of which jittering, scaling, and magnitude warping belong to magnitude domain methods, whereas time warping, permutation, and random sampling belong to time domain methods. Magnitude domain transformations convert the raw data along the value/magnitude axis,

whereas time domain transformations mainly adjust the time steps.

2.1. Magnitude Domain Transformations. Jittering is one of the simplest yet efficient DA methods by adding noise to time series, which can be defined as follows:

$$\mathbf{x}' = x_1 + \varepsilon_1, \dots, x_t + \varepsilon_t, \dots, x_T + \varepsilon_T, \quad (1)$$

where ε is typically assumed to be Gaussian noise added to each time point. The standard deviation of the noise is a hyperparameter, which needs to be predetermined.

Scaling multiplies the entire time series with a scaling parameter α , which can be defined as follows:

$$\mathbf{x}' = \alpha x_1, \dots, \alpha x_t, \dots, \alpha x_T, \quad (2)$$

where α is a hyperparameter of the zoom-in/out factor and can be determined by a Gaussian distribution.

Magnitude warping augments time series by multiplying a smoothed curve to warp the original signal's magnitude, which can be defined as follows:

$$\mathbf{x}' = \alpha_1 x_1, \dots, \alpha_t x_t, \dots, \alpha_T x_T, \quad (3)$$

where $\alpha_1, \dots, \alpha_t, \dots, \alpha_T$ is a random curve sequence generated by cubic spline interpolating with hyperparameters I (the number of knots) and σ (the standard deviation of each knot).

Generally, jittering can be considered as adding Gaussian noise to each data point, whereas magnitude warping smoothly applies varying noise to each data point. Different from these methods, scaling applies the same magnitude transformation to each data point.

2.2. Time Domain Transformations. Time warping is similar to magnitude warping. Warping is operated in the temporal dimension, which can be defined as follows:

$$\mathbf{x}' = x_{\tau(1)}, \dots, x_{\tau(t)}, \dots, x_{\tau(T)}, \quad (4)$$

where $\tau(\cdot)$ is a warping function based on the generated smooth random curve with similar hyperparameters I (the number of knots) and σ (the standard deviation of each knot).

Permutation produces a new sample by reassigning the segment order of the original time series, which can be mainly classified into two: with the same-sized segments or with the random-sized segments. The original time sequence dependency will not be preserved after permutation.

Random sampling is similar to time warping. The difference between these methods is that the former only uses partial data points for interpolating, whereas the latter uses all samples.

The methods adopted in the time domain transformations are similar to those in the magnitude domain transformations, except that the former transformations operate on the time axis, whereas the latter transformations occur on the value axis. In practice, a significant issue is to balance the transformation effect to guarantee the

TABLE 1: DA techniques for time series in this study.

Data transformation	Description
Jittering	Adding random noise
Scaling	Multiplying a random scalar
Magnitude warping	Convolving a smooth curve varying around one
Time warping	Perturbing temporal locations smoothly
Permutation	Permuting data segments randomly
Random sampling	Subsampling and interpolating

generalisation performance, and new samples would be accurately identified.

3. Proposed SSL-Based Anomaly Detection Method

The proposed SSL-based data anomaly detection method is based on MixMatch [23], a superior SSL study conducted by Google. Figure 1 depicts the overall framework of the method. For labelled examples, the classifier will directly predict their labels. For unlabelled examples, a common DA technique is initially introduced. The augmented examples are sent to the classifier, and the predictions are averaged to obtain consistent labels. A label guessing procedure based on a sharpening function is introduced to adjust the averaged predictions as low-entropy predictions, through which the unlabelled examples guess one-hot labels. Another unique technique, named MixUp [23], is introduced to combine the labelled and unlabelled examples with guessing labels to enhance model generalisation and robustness. Subsequently, the processed labelled and unlabelled data are reused to calculate the labelled and unlabelled loss items for model updating. Through these steps, the classifier is expected to learn the underlying useful information from the massive unlabelled data, and an optimal decision boundary is obtained for anomaly classification.

3.1. DA Prior to Input. As shown in Figure 1, DA is implemented on the labelled and unlabelled data, which is consistent with the smoothness assumption. In each batch, only one DA transformation is executed for labelled data, whereas K DA transformations are executed for unlabelled data to force the model to generate the same outputs for unlabelled samples with different augmentations.

3.2. Label Guessing. Each unlabelled sample produces K augmented samples after DA, as described in Figure 2. The augmented samples are sent to the same classifier to produce model predictions. The average of those predicted class distributions over K augmentations is calculated as follows:

$$\bar{q}_b = \frac{1}{K} \sum_{k=1}^K p_{\text{model}}(y | \hat{u}_{b,k}; \theta), \quad (5)$$

where \bar{q}_b denotes the averaged prediction of unlabelled data, $\hat{u}_{b,k}$ denotes the K th augmented unlabelled data, and θ denotes the model parameter.

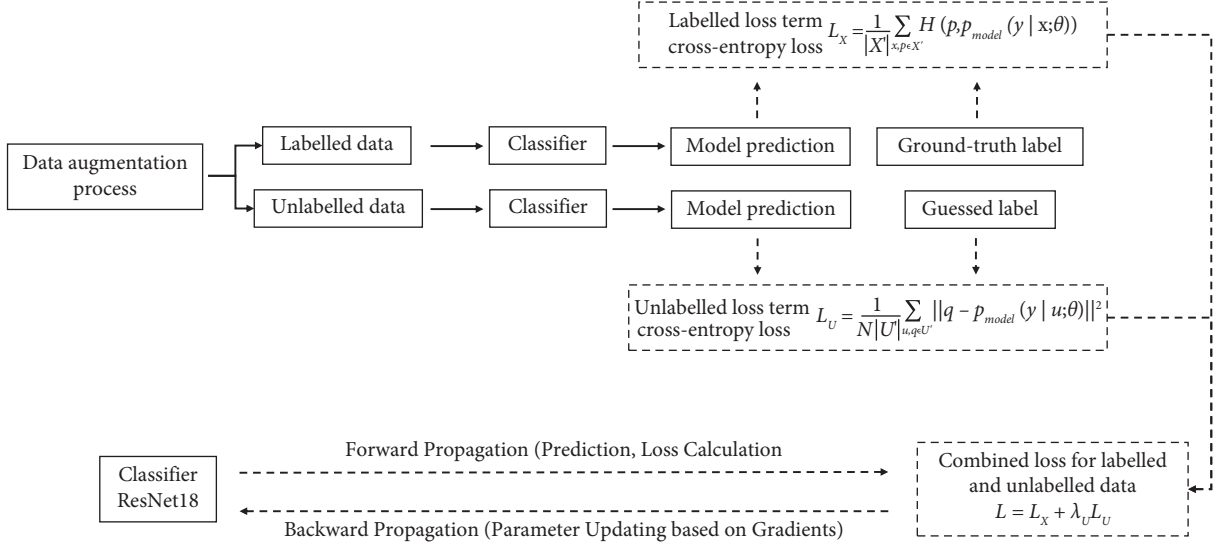


FIGURE 1: Framework of the proposed SSL-based method.

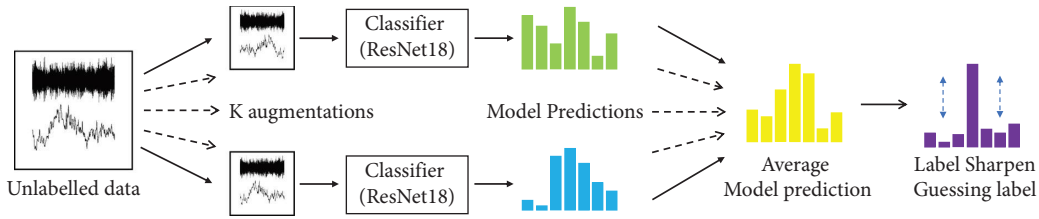


FIGURE 2: Label guessing process for unlabelled data.

Prediction averaging coordinates with consistency regularisation, which expects the model to produce the same outputs over augmented samples. Entropy minimisation is another essential component for obtaining an ideal decision boundary. One effective approach to achieve entropy minimisation is adopting sharpening function to adjust the temperature of the class distribution, which is defined as follows:

$$\text{Sharpen}(p, T)_i = \frac{p_i^{1/T}}{\sum_{j=1}^L p_j^{1/T}}, \quad (6)$$

where p is the averaged categorical distribution, i indicates class information, and T is the temperature hyperparameter. The output of the sharpening function approaches a one-hot distribution if T is close to zero. Therefore, lowering the temperature hyperparameter is beneficial for the model to make low-entropy predictions. The one-hot class distribution after prediction averaging and sharpening is assigned as the guessing label for unlabelled data, as shown in Figure 2.

3.3. MixUp. Apart from traditional DA techniques, MixMatch integrates a new data synthesis method, namely, MixUp to mix labelled and unlabelled data, which can be described as follows:

$$\lambda' = \max(\lambda, 1 - \lambda), \quad (7)$$

$$x' = \lambda' x_1 + (1 - \lambda') x_2, \quad (8)$$

$$p' = \lambda' p_1 + (1 - \lambda') p_2, \quad (9)$$

where λ is a random variable following Beta distribution, (x', p') , and (x_1, p_1) and (x_2, p_2) are the input and label probability of integrated samples and original two input samples, respectively. Equation (9) guarantees that the integrated new sample is closer to x_1 than x_2 to preserve the original order of the sequence because the labelled and unlabelled data are combined and integrated using MixUp.

MixUp plays different roles for labelled and unlabelled data. For labelled data, MixUp serves as a regularisation item given that the data have been combined to generate unseen new samples for training, even potentially mixed with unlabelled data. For unlabelled data, MixUp serves as an additional strong DA technique to enrich the input space.

3.4. Unified Loss Function. MixMatch provides a unified loss function, which gracefully minimises prediction entropy whilst maintaining consistency, thereby achieving

satisfactory or even exceeding the performance of traditional regularisation techniques.

The combined loss for SSL is defined as follows:

$$X', U' = \text{MixUp}(X, U, T, K, \alpha), \quad (10)$$

$$L_X = \frac{1}{|X'|} \sum_{x, p \in X'} H(p, p_{\text{model}}(y | x; \theta)), \quad (11)$$

$$L_U = \frac{1}{C|U'|} \sum_{u, q \in U'} \|q - p_{\text{model}}(y | u; \theta)\|_2^2, \quad (12)$$

$$L = L_X + \lambda_U L_U. \quad (13)$$

As shown in equation (10), a labelled batch X and an equally sized unlabelled batch U could produce augmented labelled batch X' and unlabelled batch U' after MixUp. Subsequently, X' and U' are separately used to calculate labelled and unlabelled loss items. Specifically, equation (11) calculates the typical cross-entropy loss between ground-truth labels and model predictions from X' , and equation (12) calculates the squared L_2 norm between guessing labels and model predictions from U' . The adaptation of L_2 norm causes the model to become less sensitive to inaccurate predictions. Equation (13) combines the labelled and unlabelled loss items, where λ_U is a time-dependent parameter balancing the trade-off between labelled and unlabelled loss items.

The MixMatch method was originally developed for image classification, and it primarily investigated common DA strategies by elastically deforming or adding noise to the image data [23]. However, SHM time series datasets are diverse and complicated. They contain important information related to structural vibration properties and the patterns that may be sensitive to noise, peak magnitude, and time series sequence [7, 29, 30]. Therefore, not all transformation strategies used in image DA are applicable. The core of DA for SHM time series data is to develop suitable operators that can generate representative new samples and ensure that these samples belong to the same class. In this study, customised DA techniques listed in Table 1 will be applied and investigated with specific domain knowledge for SHM time series data. The method developed in this study contributes to enhance the performance of MixMatch in time series data classification tasks, especially for SHM data anomaly detection.

4. Case Study

4.1. Description of the SHM Data. The acceleration data measured by the SHM system of a long-span cable-stayed bridge in China are used to validate the proposed method [31]. A total of 38 accelerometers are available, with the specific layout shown in Figure 3. The acceleration data collected with a sampling frequency of 20 Hz in two months (2012-01-01 to 2012-02-29) are used. The acceleration responses are divided into hourly segments without

overlapping, and then transformed into the image form as the dataset. The entire dataset contains 54,720 ($38 \times 24 \times 60$) segments. Each image is composed of two parts, the hourly time-domain acceleration responses as the upper part and the frequency-domain spectrum as the lower part. The image samples of seven data patterns, including the normal data and six representative anomalies, are shown in Figure 4. The time-domain and frequency-domain information are used within one image because a few patterns may be similar in the time domain and may be misclassified, such as “normal,” “minor,” and “outlier,” whereas “trend” and “drift” are similar in the frequency domain. Therefore, the use of dual-domain image can help the DL models achieve better performance.

The acceleration data in January 2012 (28,272 samples) are used in the semisupervised training process. For each data pattern in January 2012, 200 samples are randomly selected as the balanced labelled data, and the remaining samples are assigned as unlabelled data for training, as listed in Table 2. The proposed model is expected to learn data pattern-sensitive features from 1400 labelled data and a large amount of unlabelled data. The acceleration data in February 2012 (26,448 samples) are randomly separated into two parts with 5,000 and 21,448 samples. The former is used as the validation dataset to determine the training quality, and the latter is used as a blind dataset to validate the effectiveness of the proposed method, as shown in Table 3.

4.2. Results and Analysis. The proposed MixMatch-based method involves multiple hyperparameters. Specifically, the number of augmentation times K for unlabelled data, sharpening temperature T for entropy minimisation, Beta distribution parameter, and trade-off loss parameter λ_U are set to 2, 0.5, 0.75, and 100, respectively. Random horizontal flips and adding random noise are selected as the DA algorithms prior to input. More DA techniques are explored in the later section. The training epoch is set to 100, and all calculations are carried out on a computer with a CPU of InterCore i7-8700 @3.20 Ghz and a GPU of GeForce RTX 2080Ti.

The well-trained model is tested on the blind acceleration data collected in February 2012, containing 21,448 samples. Figure 5 shows the confusion matrix of the results. The precision and recall rates of most data patterns are above 95%, and the overall accuracy is 93.6%, which indicates that the SSL-based model could learn useful information of data patterns under the circumstances of extremely limited labelled data by utilising additional information from unlabelled data. The overall performance is satisfactory considering that only 200 labelled samples are available for each data pattern. Nevertheless, the precision of “minor,” “outlier,” and “drift” are relatively poor. As shown in Figure 5, a relatively large number of “normal” samples are misclassified as “minor” and “outlier,” and “trend” samples are misclassified as “drift.” One possible reason is that the sample number of these falsely identified patterns is the smallest amongst all data patterns. Therefore, the available

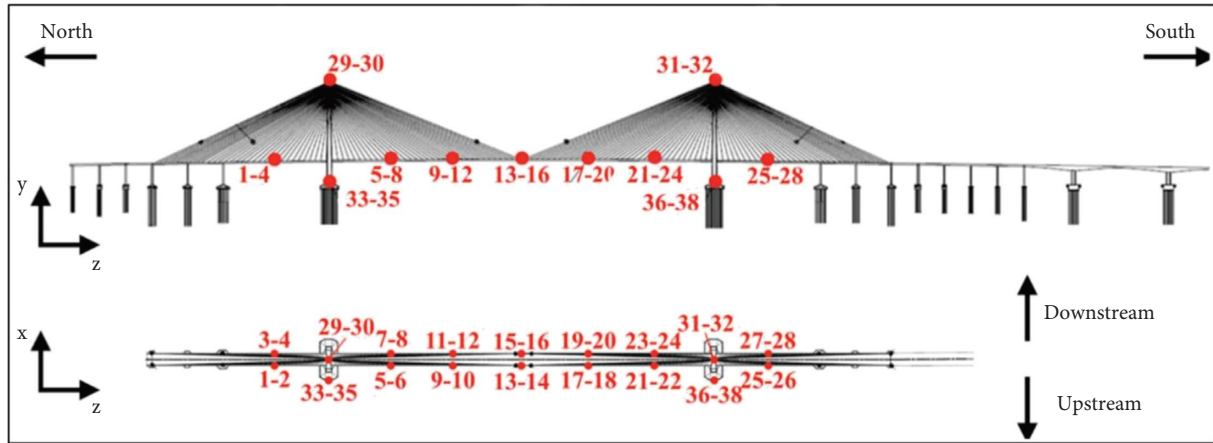


FIGURE 3: Accelerometer layout of a long-span bridge.

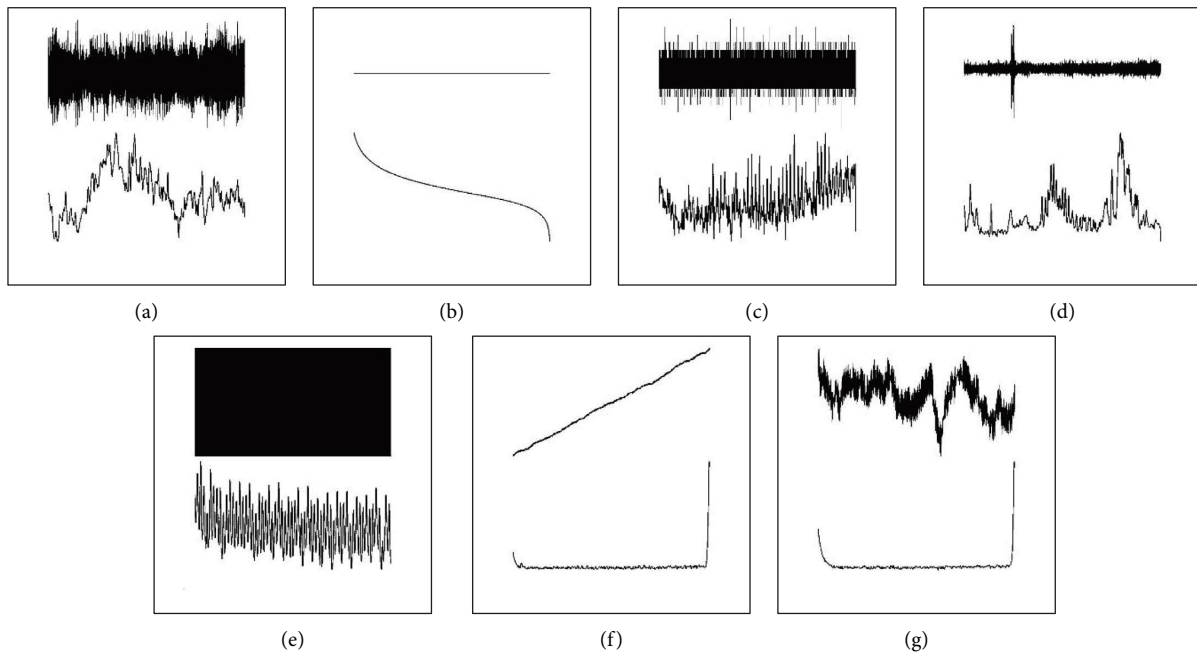


FIGURE 4: Acceleration data patterns: (a) normal, (b) missing, (c) minor, (d) outlier, (e) square, (f) trend, and (g) drift.

TABLE 2: Dataset composition for the SSL training process.

Data pattern	Labelled/unlabelled data composition of each pattern						
	Normal	Missing	Minor	Outlier	Square	Trend	Drift
Labelled data	200	200	200	200	200	200	200
Unlabelled data	13,375	2,742	1,575	327	2,796	5,578	479
Total quantity	13,575	2,942	1,775	527	2,996	5,778	679
Total				28,272/100%			

unlabelled data are insufficient for providing efficient contrastive features against confusing patterns, leading to the poor learning performance.

As a comparative study, ResNet18 is trained in a fine-tuning manner [8] using the same number of labelled samples (200 for each data pattern) but without any

unlabelled data. The confusion matrix of the test results is shown in Figure 6. The classification performance is evidently poor with an overall accuracy of 76.7%, as compared with the SSL-based method. Specifically, the performance of “normal” and “square” patterns is acceptable, revealing that these patterns might be easier for the model to learn than

TABLE 3: Dataset composition for the SSL training process.

Data pattern	Normal	Missing	Minor	Outlier	Square	Trend	Drift
<i>(a) Validation data composition of each pattern</i>							
Quantity	2,208	632	185	87	934	846	108
Percentage (%)	44.16	12.64	3.70	1.74	18.68	16.92	2.16
Total	5,000/100%						
<i>(b) Test data composition of each pattern</i>							
Quantity	10,689	2,335	1,465	244	2,280	3,712	723
Percentage (%)	49.84	10.89	6.83	1.14	10.63	17.31	3.37
Total	21,448/100%						

Predicted \ Labelled	Normal	Missing	Minor	Outlier	Square	Trend	Drift	Recall
Normal	9919	0	559	199	5	0	7	93.0%
Missing	0	2331	2	1	1	0	0	99.8%
Minor	27	0	1418	9	5	0	5	96.9%
Outlier	5	1	5	214	16	0	3	87.7%
Square	20	0	1	0	2259	0	0	99.1%
Trend	16	46	0	0	27	3211	412	86.5%
Drift	0	0	0	0	0	0	723	100%
Precision	99.3%	98.0%	71.4%	50.6%	97.7%	100%	62.9%	93.6%

FIGURE 5: Confusion matrix of classification results in February 2012 based on SSL.

other patterns. The precision of “missing,” “outlier,” and “drift” patterns is relatively low. A considerable number of “trend” samples are misclassified as “missing” or “drift” patterns, probably because these patterns exhibit similar features when only extremely limited labelled data are provided. The precision and recall of “minor” patterns are poor, showing that the model requires more samples to learn distinguishable features of “minor.” In addition, the recall rate of “trend” pattern is only 49.8% because of the misclassification with “missing,” “outlier,” and “drift” patterns. The comparative study indicates that the SSL-based method improves the model performance in comparison with supervised learning when applied to cases with limited labelled data. Nevertheless, SSL cannot further improve the model performance of data patterns with limited labelled and unlabelled data, such as “outlier” and “drift.”

4.3. Effects of DA Volume. A customised DA technique for time series is further developed to expand the available training data to tackle the insufficient unlabelled data problem encountered during SSL training, as introduced in Section 2. Six DA algorithms described in Table 1 are further utilised to investigate the effects of DA. The insufficient and difficult-to-train data patterns are augmented prior to implementing the SSL method, including “minor,” “outlier,” “trend,” and “drift” patterns. Figure 7 presents some DA examples for the “outlier” and “drift” patterns, where the characteristics of the related patterns are effectively enriched.

Nine schemes are designed to investigate their effects on model performance, as shown in Table 4. DA is mainly implemented for “minor,” “outlier,” “trend,” and “drift” patterns, whose performance indices are less satisfactory in

Predicted \ Labelled	Predicted							
	Normal	Missing	Minor	Outlier	Square	Trend	Drift	Recall
Normal	9676	0	0	1013	0	0	0	90.5%
Missing	1	2331	0	2	0	1	0	99.8%
Minor	108	0	39	1116	202	0	0	2.66%
Outlier	0	9	0	235	0	0	0	96.3%
Square	16	0	562	4	1698	0	0	74.5%
Trend	2	1423	0	114	3	1848	322	49.8%
Drift	3	0	0	81	2	4	633	87.6%
Precision	98.7%	62.0%	6.50%	9.16%	89.1%	99.7%	66.3%	76.7%

FIGURE 6: Confusion matrix of classification results in February 2012 based on ResNet18 using extremely limited labelled data.

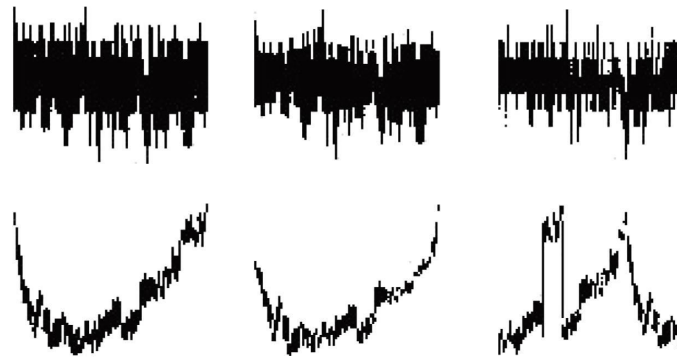


FIGURE 7: DA samples for “outlier” and “drift.”

the abovementioned studies. The blank block refers to no DA operation. After the DA operation, 200 samples are randomly selected as the labelled training data of each pattern, and the remaining samples are used as the supplementary unlabelled training data. The compositions of validation and test data remain unchanged for a straight-forward performance comparison.

Figure 8 depicts the precision/recall curve of each data pattern with different DA compositions, where the horizontal axis represents the situation of SSL without DA and nine situations of SSL with different DA compositions. Figures 8(b) and 8(e) show that the performance of “missing” and “square” patterns slightly fluctuates because they have already achieved considerable performance, as shown in Section 4.2. For the “normal” and “trend” patterns in Figures 8(a) and 8(f), a clear ascending trend of recall

score is found with the increase in unlabelled data, indicating the incremental improvement of model performance on distinguishing “normal” against other patterns. “Minor,” “outlier,” and “drift” patterns are rare classes compared with others, which also performed poorly in the abovementioned studies. As shown in Figures 8(c), 8(d), and 8(g), the supplement of unlabelled data could improve their precision scores, especially indicating from the beginning of each precision curve. Another phenomenon is that the precision score seems to be inversely related with the recall score. In practice, this phenomenon is reasonable. Specifically, considering an extreme situation where only one sample is classified as “minor,” the precision becomes 100%, whereas the recall becomes almost zero. Therefore, the precision and recall score should be balanced according to the actual scenario.

TABLE 4: DA composition strategies in this study.

Data pattern	Normal	Missing	Minor	Outlier	Square	Trend	Drift	Total
No DA	13,575	2,942	1,775	527	2,996	5,778	679	28,272
Aug_1	—	—	10,650	3,162	—	—	4,074	46,158
Aug_2	—	—	10,650	6,324	—	—	4,074	49,320
Aug_3	—	—	10,650	12,648	—	—	4,074	55,644
Aug_4	—	—	10,650	9,486	—	—	8,148	56,556
Aug_5	—	—	10,650	12,648	—	—	8,148	59,718
Aug_6	—	—	10,650	12,648	—	—	12,222	63,792
Aug_7	—	—	10,650	12,648	—	11,562	12,222	75,354
Aug_8	—	—	10,650	15,810	—	—	16,296	82,590
Aug_9	—	—	10,650	18,972	—	—	20,370	89,826

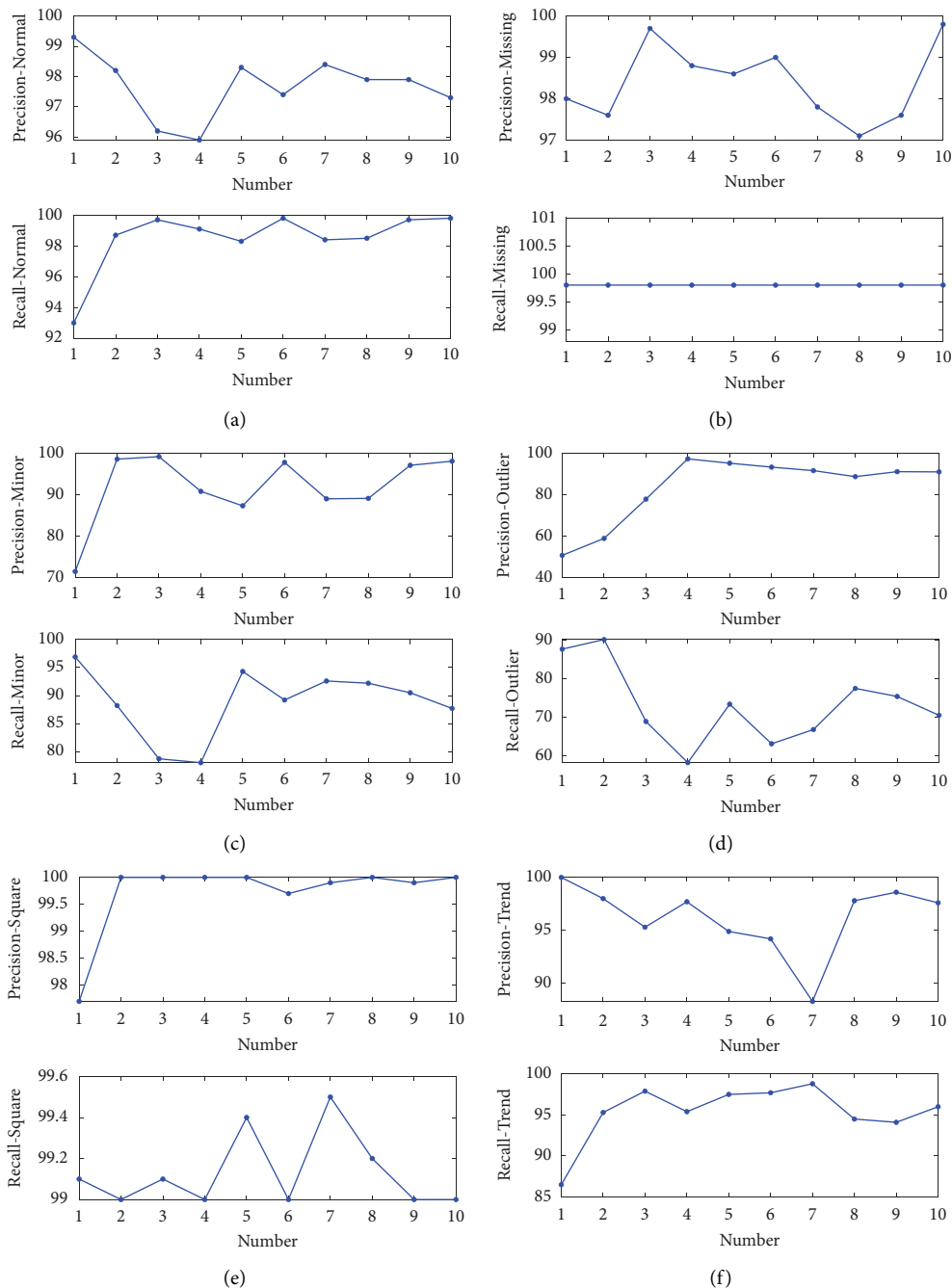


FIGURE 8: Continued.

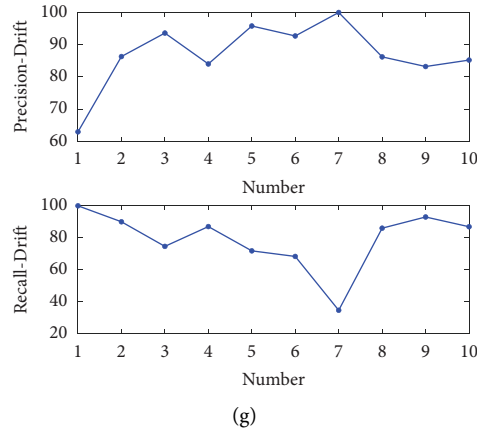


FIGURE 8: Precision/recall curve of each pattern with different DA compositions: (a) normal, (b) missing, (c) minor, (d) outlier, (e) square, (f) trend, and (g) drift.

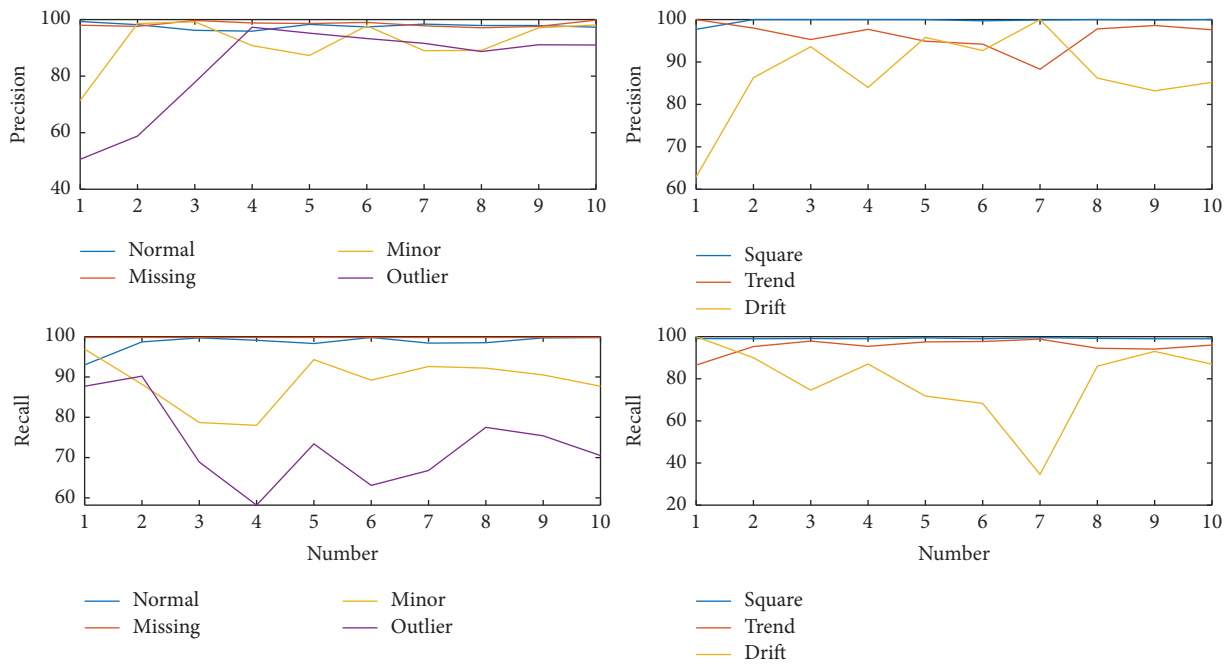


FIGURE 9: Precision/recall curve of all patterns with different DA compositions.

For ease of comparison, all results are summarised in Figure 9. In general, with the increase in unlabelled data of rare classes, the precision score of originally poor patterns exhibits a considerable improvement. In the meantime, a corresponding slight decrease in recall scores is found and is acceptable in practice. The global impact of different DA compositions cannot be evaluated directly, considering that the performance improvements of different data patterns are not closely consistent with each other. Here, a pair of simple yet efficient indices is defined, namely, “ScorePart” and “ScoreGlobal,” as shown in Figure 10. “ScorePart” is the increment of the sum of precision and recall scores of “minor,” “outlier,” “trend,” and “drift” patterns compared with those of pure SSL results without DA, whereas “ScoreGlobal” is the increment of the sum of precision and

recall scores of all patterns compared with those of pure SSL results without DA.

Figure 10 shows that the overall model performance is steadily improved with the supplement of unlabelled data for rare but indiscernible patterns. Nevertheless, a sharp decrease corresponds to the implementation of Aug_6. The results show that the application of Aug_6 changes the original data ratio greatly, resulting in the imbalance against the “trend” pattern. With the additional DA for “trend,” the data composition of each pattern becomes roughly compatible, and the overall performance continues to become better. The highest model performance is obtained at the implementation of Aug_8; then, the overall performance starts to decay, indicating that the DA strategy reaches its boundary in improving model accuracy and no more

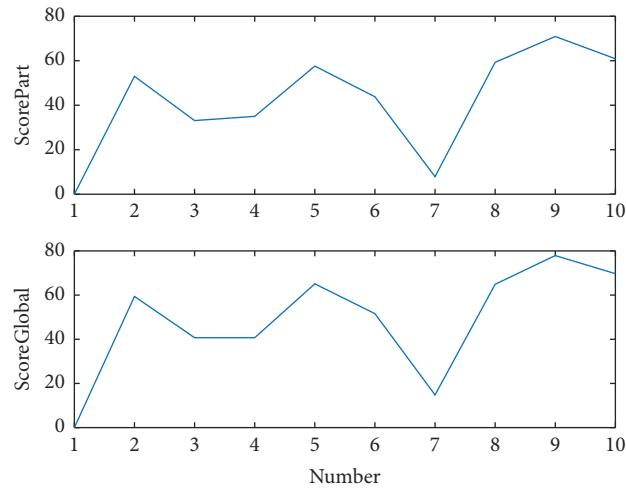


FIGURE 10: Overall model performance improvements with different DA compositions.

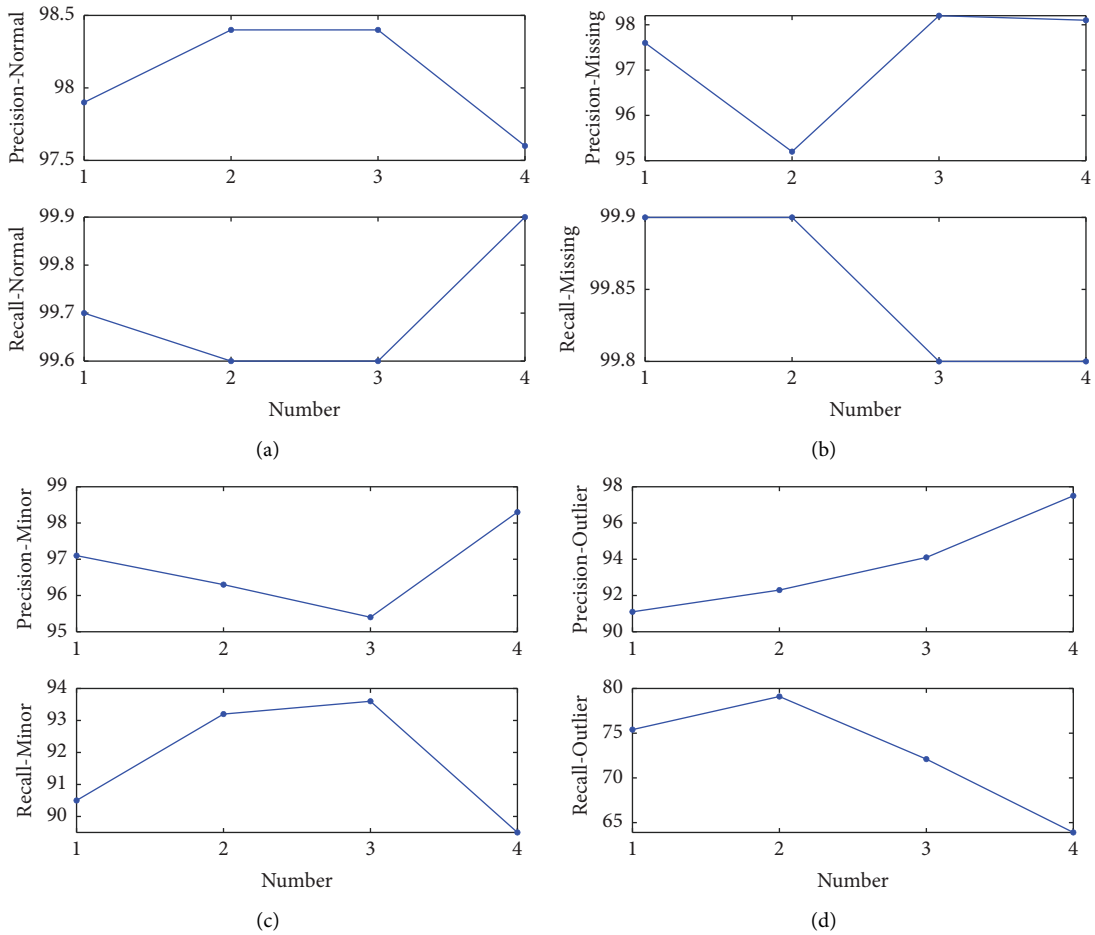


FIGURE 11: Continued.

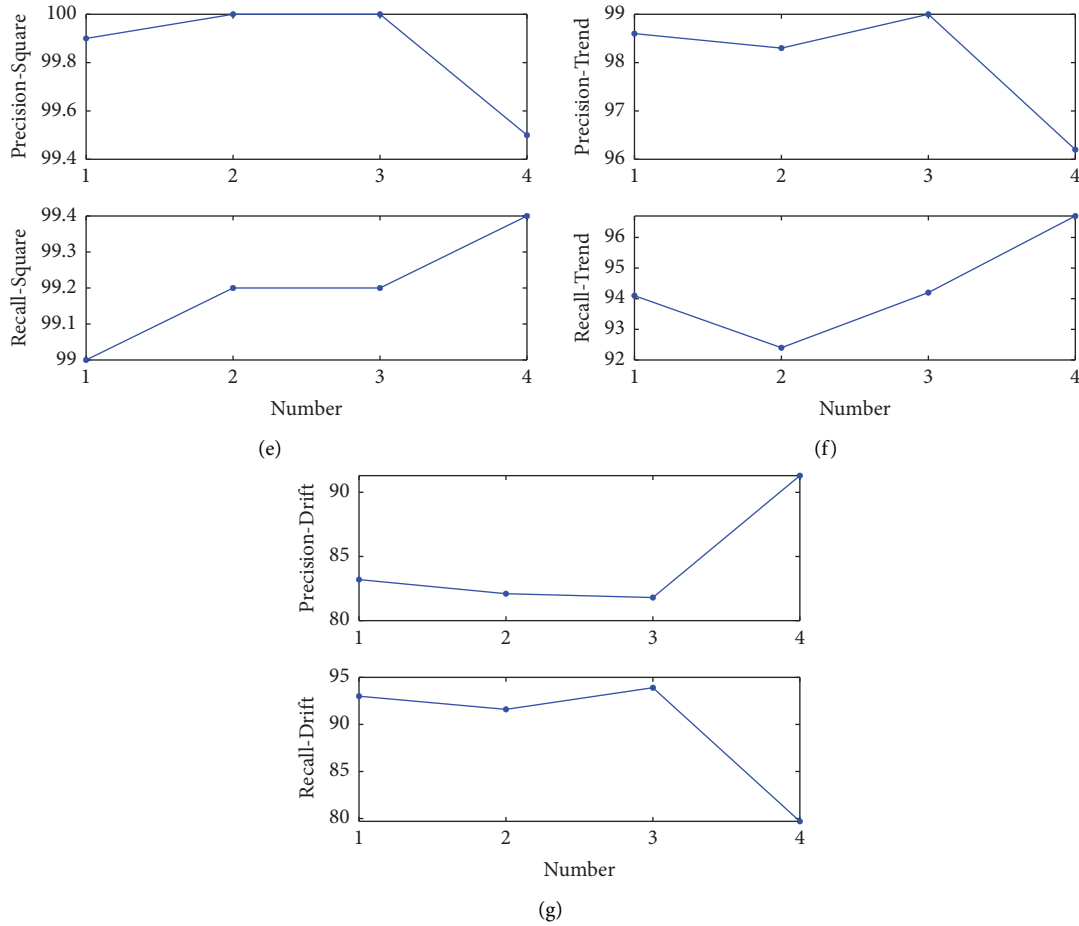


FIGURE 11: Precision/recall curve of each pattern with different numbers of labelled data: (a) normal, (b) missing, (c) minor, (d) outlier, (e) square, (f) trend, and (g) drift.

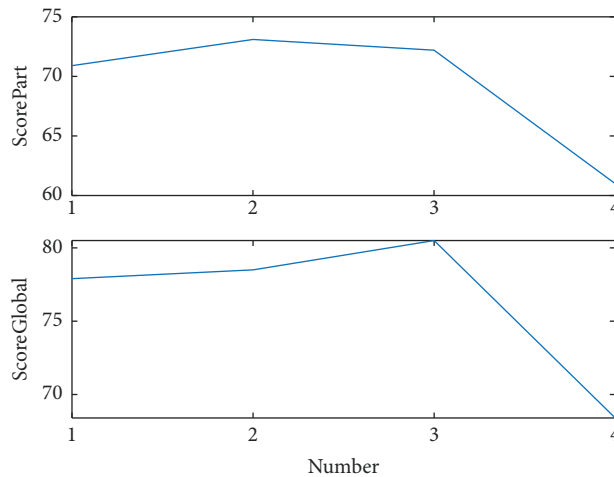


FIGURE 12: Overall model performance improvements with different numbers of labelled data.

benefits can be obtained by augmenting extra unlabelled data.

4.4. Effects of Labelled Data Quantity. The influence of model performance with different numbers of labelled training data is further investigated. Four different situations are

considered; with 200, 300, 400, and 500 labelled training data for each pattern, the number of the entire training data is unchanged. The baseline is determined as 200 labelled training data for each pattern with the implementation of Aug_8 in the former analysis. The results are shown in Figures 11 and 12.

Similarly, a clear inverse correlation is found between precision and recall of all data patterns, with the increase in the labelled training data. The precisions of the “minor,” “outlier,” and “drift” patterns that are originally rare and poorly classified are ascending as the labelled data are added, accompanied with descending recalls. For “normal” and “trend” patterns that satisfactorily classified themselves, whilst containing a relatively large number of misclassifications with other patterns, their precisions are slightly descending as the labelled data are added, accompanied with ascending recalls. “Missing” and “square” patterns are naturally easier to be distinguished than other data patterns owing to their simple yet conspicuous features. A possible explanation for this inverse correlation phenomenon is that the model training of data patterns is a progressive process. For data patterns with fewer labelled training data, the model seems to be cautious of class prediction. Therefore, only limited “confirmed” samples could be predicted, resulting in high precision score but low recall score. This status becomes constant until the training data are increased sufficiently for the model to “confidently” identify these data patterns. Subsequently, with the extra increase in labelled training data, model predictions become “audacious,” resulting in a relatively large number of falsely misclassified samples, presented by a slight decay of precision and growth of recall. Nevertheless, this slight decay of precision is insignificant to some extent because the model has learned reliable and robust features of data patterns. More detailed and rigorous investigations of this interesting phenomenon might be investigated in the future.

Overall, a preliminary increasing and then decreasing trend is observed with the increase in labelled training data for each pattern, as shown in Figure 12. This finding indicates that the assistance of labelled data for SSL model training is restricted. The redundant labelled data may not further improve the model performance. This finding is consistent with that of investigating different-scale unlabelled data for model performance in Section 4.3. In summary, the analysis in Sections 4.3 and 4.4 reveals that the labelled and unlabelled data play essential roles in the SSL training process, and the increase in the volume of training data and the DA technique have limited effects on improving the model performance.

5. Conclusions

This study develops a novel method for detecting SHM data anomalies with limited labelled data and massive unlabelled data. To fully exploit the information of unlabelled data, the MixMatch technique is employed to define a unified loss function that comprises consistency regularisation, entropy minimisation, and regular model regularisation items, which combine information from both labelled and unlabelled data to enhance model generalisation and robustness. The proposed method is applied to the SHM data from an actual bridge. The results indicate that the SSL-based method improves the model performance compared to supervised learning when applied to the case with limited labelled data.

In addition, customised DA strategies for SHM time series data are investigated to improve the model performance, including jittering, scaling, magnitude warping, time warping, permutation, and random sampling. The effects of DA volume and labelled data quantity in improving model performance are studied. Results reveal that the increase in the volume of training data and the DA technique contribute to improve the classification accuracy. Nevertheless, the DA strategy has its boundary in improving model accuracy and no more benefits can be obtained by augmenting extra unlabelled data upon the boundary. This study will provide valuable insights into time series DA and data anomaly classification in the context of limited labelled data.

Data Availability

The data are provided by the first International Project Competition for SHM (IPC-SHM, 2020).

Disclosure

Xiaoyou Wang and Yao Du are co-first authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by the Key-Area Research and Development Program of Guangdong Province (Project no. 2019B111106001) and RGC-GRF (Project no. 15217522). The authors would like to thank the organisers of the first International Project Competition for SHM (IPC-SHM, 2020) for generously providing invaluable data from an actual structure. Special thanks to Professor Hui Li and Professor Billie F. Spencer Jr., Co-Chairs of IPC-SHM, 2020.

References

- [1] Y. Bao and H. Li, “Machine learning paradigm for structural health monitoring,” *Structural Health Monitoring*, vol. 20, no. 4, pp. 1353–1372, 2021.
- [2] Z. Tang, Z. Chen, Y. Bao, and H. Li, “Convolutional neural network-based data anomaly detection method using multiple information for structural health monitoring,” *Structural Control and Health Monitoring*, vol. 26, no. 1, Article ID e2296, 2019.
- [3] Y. Zhang, X. Wang, Z. Ding, Y. Du, and Y. Xia, “Anomaly detection of sensor faults and extreme events based on support vector data description,” *Structural Control and Health Monitoring*, vol. 29, no. 10, Article ID e3047, 2022.
- [4] M. Azimi and G. Pekcan, “Structural health monitoring using extremely compressed data through deep learning,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 35, no. 6, pp. 597–614, 2020.
- [5] G. Fan, J. Li, and H. Hao, “Lost data recovery for structural health monitoring based on convolutional neural networks,” *Structural Control and Health Monitoring*, vol. 26, no. 10, Article ID e2433, 2019.
- [6] L. M. Sun, Z. Q. Shang, Y. Xia, S. Bhowmick, and S. Nagarajaiah, “Review of bridge structural health

- monitoring aided by big data and artificial intelligence: from condition assessment to damage detection,” *Journal of Structural Engineering*, vol. 146, no. 5, Article ID 04020073, 2020.
- [7] X. W. Ye, T. Jin, and C. B. Yun, “A review on deep learning-based structural health monitoring of civil infrastructures,” *Smart Structures and Systems*, vol. 24, no. 5, pp. 567–585, 2019.
 - [8] Y. Du, F. Sun, R. Hou, X. Y. Wang, W. Tian, and Y. Xia, “HiCBin: binning metagenomic contigs and recovering metagenome-assembled genomes using Hi-C contact maps,” *Genome Biology*, vol. 23, no. 1, pp. 63–75, 2022.
 - [9] X. Wang and Y. Xia, “Knowledge transfer for structural damage detection through re-weighted adversarial domain adaptation,” *Mechanical Systems and Signal Processing*, vol. 172, Article ID 108991, 2022.
 - [10] X. Xu, Z. D. Qian, Q. Huang, Y. Ren, and B. Liu, “Probabilistic anomaly trend detection for cable-supported bridges using confidence interval estimation,” *Advances in Structural Engineering*, vol. 25, no. 5, pp. 966–978, 2022.
 - [11] X. Wang, L. Li, J. L. Beck, and Y. Xia, “Sparse Bayesian factor analysis for structural damage detection under unknown environmental conditions,” *Mechanical Systems and Signal Processing*, vol. 154, Article ID 107563, 2021.
 - [12] Z. Y. Fan, Q. Huang, Y. Ren, Z. Y. Zhu, and X. Xu, “A cointegration approach for cable anomaly warning based on structural health monitoring data: an application to cable-stayed bridges,” *Advances in Structural Engineering*, vol. 23, no. 13, pp. 2789–2802, 2020.
 - [13] Y. Xu, W. Qian, N. Li, and H. Li, “Typical advances of artificial intelligence in civil engineering,” *Advances in Structural Engineering*, vol. 25, no. 16, pp. 3405–3424, 2022.
 - [14] X. Y. Wang, L. F. Li, W. Tian, Y. Du, R. R. Hou, and Y. Xia, “Unsupervised one-class classification for condition assessment of bridge cables using Bayesian factor analysis,” *Smart Structures and Systems*, vol. 29, no. 1, pp. 41–51, 2022.
 - [15] Y. Xu, Y. Bao, J. Chen, W. Zuo, and H. Li, “Surface fatigue crack identification in steel box girder of bridges by a deep fusion convolutional neural network based on consumer-grade camera images,” *Structural Health Monitoring*, vol. 18, no. 3, pp. 653–674, 2019.
 - [16] B. Kim and S. Cho, “Image-based concrete crack assessment using mask and region-based convolutional neural network,” *Structural Control and Health Monitoring*, vol. 26, no. 8, Article ID e2381, 2019.
 - [17] B. K. Iwana and S. Uchida, “An empirical survey of data augmentation for time series classification with neural networks,” *PLoS One*, vol. 16, no. 7, Article ID e0254841, 2021.
 - [18] Z. Hussain, F. Gimenez, D. Yi, and D. Rubin, “Differential data augmentation techniques for medical imaging classification tasks,” *AMIA annual symposium proceedings*, vol. 2017, pp. 979–984, 2017.
 - [19] L. Perez and J. Wang, “The effectiveness of data augmentation in image classification using deep learning,” 2017, <https://arxiv.org/abs/1712.04621>.
 - [20] A. Mikołajczyk and M. Grochowski, “Data augmentation for improving deep learning in image classification problem,” in *Proceedings of the 2018 International Interdisciplinary PhD Workshop (IIPhDW)*, pp. 117–122, IEEE, Swinoujscie, Poland, May 2018.
 - [21] J. E. Van Engelen and H. H. Hoos, “A survey on semi-supervised learning,” *Machine Learning*, vol. 109, no. 2, pp. 373–440, 2020.
 - [22] X. J. Zhu, *Semi-supervised Learning Literature Survey*, University of Wisconsin-Madison, Madison, WI, USA, 2005.
 - [23] D. Berthelot, N. Carlini, I. Goodfellow, N. O. A. Papernot, and C. A. Raffel, “Mixmatch: A holistic approach to semi-supervised learning,” *Advances in Neural Information Processing Systems*, vol. 32, 2019.
 - [24] Y. Grandvalet and Y. Bengio, “Semi-supervised learning by entropy minimization,” *Advances in Neural Information Processing Systems*, vol. 17, 2004.
 - [25] V. Verma, K. Kawaguchi, A. Lamb, J. Kannala, Y. Bengio, and D. Lopez-Paz, “Interpolation consistency training for semi-supervised learning,” 2019, <https://arxiv.org/abs/1903.03825>.
 - [26] P. Martakis, A. Movsessian, Y. Reuland, and E. Chatzi, “A semi-supervised interpretable machine learning framework for sensor fault detection,” *Smart Structures and Systems*, vol. 29, pp. 251–266, 2021.
 - [27] S. Li, L. Jin, Y. Qiu, M. Zhang, and J. Wang, “Signal anomaly detection of bridge SHM system based on two-stage deep convolutional neural networks,” *Structural Engineering International*, vol. 33, pp. 74–83, 2021.
 - [28] Y. Zhang, Z. Tang, and R. Yang, “Data anomaly detection for structural health monitoring by multi-view representation based on local binary patterns,” *Measurement*, vol. 202, Article ID 111804, 2022.
 - [29] Y. Xia and R. Lin, “Improvement on the iterated IRS method for structural eigensolutions,” *Journal of Sound and Vibration*, vol. 270, no. 4–5, pp. 713–727, 2004.
 - [30] K. Liu, S. S. Law, Y. Xia, and X. Q. Zhu, “Singular spectrum analysis for enhancing the sensitivity in structural damage detection,” *Journal of Sound and Vibration*, vol. 333, no. 2, pp. 392–417, 2014.
 - [31] Y. Bao, J. Li, T. Nagayama, Y. Xu, B. F. Spencer, and H. Li, “The 1st international project competition for structural health monitoring (IPC-SHM, 2020): A summary and benchmark problem,” *Structural Health Monitoring*, vol. 20, no. 4, pp. 2229–2239, 2021.