ISSN: 2688-8645

A CNN-LSTM Hybrid Model for Subtle Damage Identification in Structural Health Monitoring

He Bai, Heqi Zhang *, Jincheng Li

School of Civil Engineering, Changchun Institute of Technology, Changchun, Jilin, 130012, China

* Corresponding author: Heqi Zhang

Abstract

Time-series data in structural health monitoring (SHM) contain both local anomaly features, such as crack jumps and strain sudden changes, and long-term evolution dependencies, such as cumulative deformation, which a single model can hardly extract simultaneously. A CNN-LSTM hybrid damage identification model is therefore constructed: one-dimensional convolutional layers extract transient anomaly features through local receptive fields, while LSTM layers model long-term temporal dependencies through gating mechanisms, and the two are cascaded to form an identification pipeline of "local feature extraction - temporal dependency modeling - damage state discrimination". Model training and validation rely on real field data obtained from 27 measurement points over 8.5 months of continuous monitoring, including 24710 crack records, 55832 tilt records and 86181 strain records, and a labeled identification dataset is constructed by injecting artificial subtle damage signals into normal sequences. Comparative experiments show that the CNN-LSTM hybrid model achieves an accuracy of 96.4%, a recall of 94.7%, an F1 score of 0.956 and a false-alarm rate of 1.8%, comprehensively outperforming standalone CNN, standalone LSTM, SVM and the traditional threshold method; compared with the standalone CNN, the accuracy is improved by 4.3 percentage points, the recall by 9.4 percentage points, and the false-alarm rate is reduced by 2.8 percentage points. Ablation analysis indicates that the complementary mechanism of convolutional and recurrent layers is the key to the performance gain. The validity boundaries concerning artificially labeled damage and the lack of cross-point transfer verification are also discussed, providing a basis for applying deep learning methods to the early diagnosis of subtle structural damage.

Keywords

Structural Health Monitoring; CNN-LSTM; Deep Learning; Damage Identification; Time-series Analysis; False-alarm Rate.

1. Introduction

Structural health monitoring (SHM) data are typical high-dimensional time series in which damage manifests in diverse forms: there are both local short-term anomalies, such as instantaneous crack width jumps and strain sudden changes, and long-term evolutionary trends, such as the slow development of cumulative deformation. How to automatically and reliably identify subtle damage from massive monitoring time series is a core issue in intelligent early warning of structural safety.

Traditional damage identification methods are mainly based on fixed threshold discrimination and manual feature extraction: the threshold method is simple and easy to deploy, but it can hardly adapt to the slow drift of structural states and suffers from high false-alarm and miss rates; machine learning methods based on support vector machines (SVM) depend on manually

ISSN: 2688-8645

designed time-domain or frequency-domain features, and the quality of feature engineering directly determines the identification ceiling [12]. In recent years, deep learning methods have been widely introduced into damage identification: convolutional neural networks (CNN) automatically extract spatial or temporal local features through parameter sharing and local receptive fields [12]; long short-term memory networks (LSTM) model long-range temporal dependencies through gating mechanisms and alleviate the gradient vanishing problem [5,7]; generative adversarial networks have also been applied to data augmentation and damage discrimination [7]. Fadel et al. [5] verified the effectiveness of combined CNN and LSTM configurations for bridge damage identification using long-term vibration monitoring data of the Z24 bridge; Li et al. [12] proposed a damage detection method based on one-dimensional dilated convolutions to enlarge the receptive field; Yang et al. [11] proposed a single-sampling identification method using statistical moments as damage indices.

However, existing studies still show three deficiencies. First, most damage identification models rely on laboratory data or vibration modal parameters, and their generalization on real field monitoring data is insufficient. Second, model evaluation is mostly performed on global vibration data, while systematic research on subtle anomaly identification of quasi-static deformation quantities (cracks, tilt and strain) is scarce. Third, there is a lack of in-depth analysis of the sources and applicability boundaries of model effectiveness, and model selection often remains at the level of performance comparison. To address these issues, this paper constructs a CNN-LSTM hybrid damage identification model and carries out training and validation on field data obtained from 27 measurement points over 8.5 months of continuous monitoring on laboratory columns of a university structural laboratory. The main work includes: (1) designing a hybrid identification pipeline of "local feature extraction - temporal dependency modeling - damage state discrimination" with a complete training scheme; (2) constructing a labeled subtle damage identification dataset and setting up four comparative baselines for performance evaluation; and (3) revealing the sources of the hybrid model effectiveness through ablation and error analysis, and discussing its validity boundaries.

2. Methodology

Structural health monitoring data contain both short-term information reflecting local sudden changes (such as instantaneous crack width jumps and strain sudden changes) and long-term dependence information reflecting the long-term evolution trend of the structure (such as cumulative deformation and cumulative variation). A single model can hardly extract both types of features simultaneously, so a CNN-LSTM hybrid model is introduced to fully exploit the complementary advantages of the two.

CNN performs sliding convolution along the temporal dimension through convolution kernels, and automatically extracts local anomaly features in monitoring sequences, such as crack jumps and strain sudden changes, by virtue of parameter sharing and local receptive fields. Compared with manual feature extraction, convolution can learn multi-scale local patterns in a data-driven manner and exhibits certain robustness to slight phase shifts of the input series. LSTM selectively memorizes and forgets temporal information through three gating mechanisms (input gate, forget gate and output gate), effectively alleviating the gradient vanishing problem of recurrent neural networks on long sequences, and captures long-term evolutionary dependencies in monitoring data, such as the monotonic sustained growth of crack width and the slow drift of tilt angle.

The two are cascaded to form an identification pipeline of "local feature extraction - temporal dependency modeling - damage state discrimination": the convolutional layers first map the raw time series into a short sequence of local anomaly features, the LSTM layers then accumulate the evolutionary information along the time dimension on this feature sequence,

ISSN: 2688-8645

and finally the fully connected layer gives the probability of the normal/damage state. This division of labor enables the model to retain both the sharpness of local anomaly features and the memory of long-term evolutionary trends, fitting the quasi-static deformation data pattern of "significant local sudden changes with slowly varying overall trends".

3. Network Structure and Training

The model consists of one-dimensional convolutional layers, pooling layers, LSTM layers and a fully connected classification layer. Specifically, the convolutional layers adopt multiple one-dimensional kernels to extract local features within the input window; after batch normalization and ReLU activation, the features are fed into a max-pooling layer for dimensionality reduction, then input to the LSTM layers to capture temporal dependencies, and finally passed through the fully connected layer and a Softmax layer to output the probabilities of the two states (normal/damage).

The raw monitoring data are preprocessed by missing-value interpolation, outlier removal, moving-average denoising, time alignment and Z-score standardization, and then divided into fixed-length samples by the sliding window method; each sample corresponds one-to-one with the damage label at the last time step of the window, and the dataset is split into training, validation and test sets at a ratio of 7:2:1. Training adopts the cross-entropy loss function and the Adam optimizer with an initial learning rate of 1e-3 and a batch size of 64; to improve generalization and prevent overfitting, an early-stopping mechanism (training stops when the validation loss does not decrease for 10 consecutive epochs) and a learning-rate decay strategy are introduced.

To make the training process reproducible, the key hyperparameters are set as follows: the sliding window length is 60 sampling points (corresponding to a 1-hour continuous monitoring series) with a window stride of 6; the convolutional part uses 4 one-dimensional convolutional layers with kernel sizes of 8, 8, 4 and 4 and filter numbers of 32, 64, 64 and 128, respectively; the pooling window is 2; the LSTM part consists of 2 unidirectional layers with 64 hidden units; the fully connected part has one hidden layer of 64 units; parameters are initialized with the He initializer, and the maximum number of training epochs is 100. The experiments are implemented in Python 3.10 based on PyTorch 1.13, and training and inference are completed on a workstation equipped with an Intel Core i7 CPU, 16 GB RAM and an NVIDIA RTX 3060 GPU. After early stopping, the actual number of training epochs is 43, and the ratio of training to test samples is about 7:1. The above hyperparameters are fixed after preliminary grid and random search to avoid information leakage from the test set into model selection.

4. Experimental Design and Results

4.1. Data Source and Identification Dataset Construction

The training and validation data come from 27 measurement points (14 strain, 4 crack and 9 tilt) deployed on laboratory columns of a university structural laboratory, monitored continuously for about 8.5 months from December 2025 to August 2026 at an acquisition interval of about 1 minute. After removing abnormal data caused by device disconnection, a total of 24710 crack records, 55832 tilt records and 86181 strain records are obtained. To verify the ability of the model to identify subtle anomalies, artificially constructed subtle damage signals (such as low-amplitude monotonic drift, short-term step and progressive growth patterns) are injected into the normal monitoring series to build a labeled damage identification dataset. The injected amplitudes are set with reference to the statistical characteristics of the measured data, ensuring that they remain at the "subtle damage" level and preventing the task from degenerating into a simple discrimination of significant anomalies.

ISSN: 2688-8645

4.2. Evaluation Metrics and Comparative Baselines

Accuracy, precision, recall and the F1 score are adopted to evaluate model performance, with particular attention to the false-alarm rate and miss rate, so as to measure the reliability of damage identification and early warning in practical engineering. These metrics are computed from the TP, TN, FP and FN of the confusion matrix: accuracy reflects the overall proportion of correct discrimination, precision reflects the proportion of true damage among samples judged as damaged, recall reflects the proportion of true damage samples correctly identified, the F1 score is the harmonic mean of precision and recall, and the false-alarm rate reflects the proportion of normal samples misjudged as damaged.

Four comparative baselines are set: a standalone CNN (convolution-pooling-fully connected only, without temporal memory modules), a standalone LSTM (LSTM-fully connected only, without convolutional feature extraction), a support vector machine (SVM with manually designed time-domain/frequency-domain features as input), and the traditional threshold method (fixed amplitude threshold discrimination). All methods are compared on the same dataset under the same evaluation protocol in terms of accuracy, recall, F1 score and false-alarm rate, and the results are shown in Table 1.

Table 1. Performance comparison of different damage identification methods

Method	Accuracy (%)	Recall (%)	F1 score	False-alarm rate (%)
CNN-LSTM (proposed)	96.4	94.7	0.956	1.8
Standalone CNN	92.1	85.3	0.886	4.6
Standalone LSTM	89.7	83.9	0.868	6.2
SVM	87.5	76.4	0.816	7.9
Traditional threshold method	74.3	58.9	0.658	18.5

4.3. Results Analysis

As shown in Table 1, the CNN-LSTM hybrid model achieves an accuracy of 96.4%, a recall of 94.7%, an F1 score of 0.956 and a false-alarm rate of 1.8%, outperforming the four baselines on all four metrics. Compared with the standalone CNN, the second-best method, the hybrid model improves the accuracy by 4.3 percentage points and the recall by 9.4 percentage points, raises the F1 score by 0.070, and reduces the false-alarm rate by 2.8 percentage points; compared with the traditional threshold method, the recall is improved by 35.8 percentage points and the false-alarm rate is reduced by 16.7 percentage points.

These differences indicate that although CNN can capture local anomalies such as crack jumps and strain sudden changes, it lacks the memory of the evolutionary trend reflected by the accumulation of anomalies over consecutive time steps at the same measurement point, and is therefore prone to misses at the initial stage of an anomaly; LSTM compensates for this deficiency in temporal memory, but responds sluggishly to abrupt amplitude jumps in the absence of a local feature extraction mechanism. When cascaded, the model retains both the sharpness of local anomaly features and the memory of long-term evolutionary trends, thus achieving the best comprehensive performance on the subtle damage identification task, which is sensitive to both series length and instantaneous response. From an engineering perspective, reducing the false-alarm rate to 1.8% means that the proportion of false alarms among the alerts received by maintenance personnel is significantly reduced, which is of direct value for maintaining the credibility of the warning system and avoiding the "cry-wolf" effect; meanwhile, a recall of 94.7% ensures that the vast majority of damage events are captured, and the two together support the reliability of early diagnosis.

ISSN: 2688-8645

4.4. Error Analysis and Computational Efficiency

To further examine the sources of false alarms and misses of the hybrid model, the misclassified samples in the test set are categorized by anomaly pattern. The results show that false alarms are mainly concentrated in segments of the normal series with small but directionally consistent random fluctuations, which resemble low-amplitude drift-type damage signals within a local window and constitute the main source of false alarms; misses mainly occur at the initial stage of the injected damage signals, when the accumulated anomaly is not yet distinguishable from normal fluctuations. This distribution corroborates the mechanism analysis in Section 4.3: the convolutional layers are sensitive to instantaneous amplitudes, while the recurrent layers require a certain accumulation window to form a stable discrimination, so misses are likely at the first few sampling points of an anomaly, but the recall recovers rapidly as the window advances. In engineering practice, fusing the warning criterion with the judgment results of multiple consecutive windows can further reduce early-stage misses without significantly raising the false-alarm rate.

In terms of computational efficiency, the inference time of the hybrid model for a single sample on the RTX 3060 workstation is on the millisecond scale, which satisfies the real-time identification requirement for the online monitoring data stream at an interval of about 1 minute; the total training time (43 epochs) is on the order of several minutes, and the whole pipeline can be retrained within hours after hyperparameter updating, demonstrating engineering feasibility of rolling updates as monitoring data keep accumulating. By contrast, the training time of the SVM on high-dimensional window features grows markedly with the sample size; although the traditional threshold method is the simplest computationally, its performance gap is significant as noted above. Considering both performance and deployment cost, the CNN-LSTM hybrid model is a reasonable choice that balances accuracy and efficiency.

5. Discussion: Validity Boundaries of the Hybrid Model

The CNN-LSTM outperforms the standalone CNN, the standalone LSTM, the SVM and the traditional threshold method on all four metrics, with the false-alarm rate reduced to 1.8%. This advantage is directionally consistent with the conclusion drawn by Fadel et al. [5] on long-term vibration monitoring data of the Z24 bridge, but its origin differs: that study relies on global vibration modal parameters of the bridge, whereas this paper relies on local quasi-static deformation time series. This difference suggests that the value of the hybrid model lies not in the network structure itself, but in its adaptation to a specific data pattern of "significant local sudden changes with slowly varying overall trends". When the separability of damage in the feature space at a single time step is low but a cumulative trend emerges over the whole observation window, the local receptive fields of the convolutional layers capture instantaneous anomalies while the memory mechanism of the recurrent layers accumulates low-amplitude but persistent directional deviation; the absence of either component leads to misses. In other words, for identification tasks based on global modal parameters, the source of model effectiveness may differ, and the conclusions of this paper should not be extrapolated directly.

It should be pointed out that the comparative experiments in this paper have three boundaries. First, the damage labels are obtained by injecting artificially constructed subtle damage signals into normally measured series, rather than from an actual loading-to-failure process; therefore, the reported accuracy and recall reflect the ability of the model to discriminate known anomaly patterns within existing normal data and do not yet cover the diversity of real damage deformation. Actual damage patterns in engineering, such as slow cracking coupled with ambient temperature, may be more complex than the injected signals. Second, the test set and the training set come from different time windows of the same measurement-point series, and

no cross-point or cross-structure transfer verification has been conducted; the generalization of the model after measurement-point replacement or structure-type change remains to be examined. Third, the SVM and the threshold method among the baselines are configured with their respective applicable input forms, and their hyperparameters are tuned independently but with a smaller search budget than that of the hybrid model, which may lead to a potentially unfair comparison with the baselines.

Future research should conduct cross-domain validation on reserved independent measurement points and apply an equivalent tuning budget to all methods; meanwhile, genuine anomaly cases identified from the field data (such as the anomalous evolution of crack point-3 and the strain sudden changes at strain point-4) can be used to further enrich the real damage sample library, so as to examine and improve the identification ability of the model under real damage patterns and promote its evolution from "identifying injected anomalies" to "identifying real damage".

6. Conclusion

(1) A CNN-LSTM hybrid damage identification model is constructed, in which the convolutional layers capture local anomalies and the recurrent layers model long-term temporal dependencies, forming a complete identification pipeline of "local feature extraction - temporal dependency modeling - damage state discrimination", and a reproducible network structure and training scheme are provided.

(2) Based on an identification dataset built from 27 measurement points over 8.5 months of continuous field monitoring, the CNN-LSTM hybrid model achieves an accuracy of 96.4%, a recall of 94.7%, an F1 score of 0.956 and a false-alarm rate of 1.8%, comprehensively outperforming the standalone CNN, the standalone LSTM, the SVM and the traditional threshold method, providing methodological support for the early diagnosis of subtle structural damage.

(3) Ablation analysis shows that the performance advantage of the hybrid model stems from the complementary adaptation of the convolutional and recurrent layers to the data pattern of "significant local sudden changes with slowly varying overall trends"; its effectiveness is subject to three boundaries, namely artificially constructed labels, lack of cross-domain verification and asymmetric baseline tuning budgets, and cross-point transfer verification and real damage samples should be introduced in future work.

Acknowledgments

Changchun Institute of Technology's National College Student Innovation Training Program project: "Research on an AI-Based Health Monitoring System for Civil Engineering Structures" (202511437043).

References

[1] Tang, D. Y., Wei, X. B., Sun, Z. H., et al. (2023). Review of structural health monitoring systems. Intelligent Building and Smart City, (8), 6–11.

[2] Jiao, K., Lai, H. L., Peng, Z. X., et al. (2021). Research on key technologies of an IoT-based life-cycle safety operation service system for existing buildings. Industrial Construction, 51(10), 201–210.

[3] Bai, B., Chen, Y., Du, C. Q., et al. (2020). Intelligent safety monitoring building information model based on IoT technology. Industrial Construction, 50(4), 175–179.

[4] Luo, Y. Z., & Zhao, J. Y. (2022). Research status and prospects of structural health monitoring of spatial structures. Journal of Building Structures, 43(10), 16–28.

[5] Fadel, Y., & Zhu, J. (2023). Classification and regression-based convolutional neural network and long short-term memory configuration for bridge damage identification using long-term

monitoring vibration data. Structural Health Monitoring, 22(6), 4027–4054. https://doi.org/10.1177/1475921723116047.

[6] Tokognon, C. A., Gao, B., Tian, G. Y., et al. (2017). Structural health monitoring framework based on internet of things: a survey. IEEE Internet of Things Journal, 4(3), 619–635. https://doi.org/10.1109/JIOT.2017.2674466.

[7] Azimi, M., Eslamlou, A. D., & Pekcan, G. (2020). Data-driven structural health monitoring and damage detection through deep learning: state-of-the-art review. Sensors, 20(10), 2778. https://doi.org/10.3390/s20102778.

[8] Liu, Y. F., Qi, Y., Li, B. L., et al. (2022). Intelligent identification of sensitive service indices for multi-scenario structural operation and maintenance. Journal of Building Structures, 43(10), 1–15.

[9] Yue, Q. R., Liu, X. G., & Chen, H. B. (2022). Research and application progress of intelligent diagnosis for structural service safety. Journal of Building Structures, 43(10), 41–49.

[10] Xia, Z. Q., Ma, L. Y., Shan, J. Z., et al. (2023). Two-stage detection method for building facade spalling and cracks based on computer vision. Journal of Building Structures, 44(2), 207–216.

[11] Yang, Y., Wang, Z. W., Ling, Y., et al. (2023). Damage identification of frame structures based on statistical moment theory and Bayesian method. Journal of Building Structures, 44(2), 217–226.

[12] Li, R., & Zhang, C. (2021). Structural damage detection method based on one-dimensional dilated convolutional neural network. Industrial Construction, 51(10), 170–176.