Low-Complexity Approach to Intelligent SHM by Combining Machine Learning Models Using Single-Sensor Data

FIDMA MOHAMED ABDELILLAH, JEAN-FRANCOIS BERCHER
and FRANZISKA SCHMIDT

ABSTRACT

Structural Health Monitoring (SHM) systems, when applied to large civil engineering structures such as bridges, process high-volume data and run computationally intensive algorithms, which typically require important processing power to ensure low inference latency to enable real-time monitoring and rapid decision-making. In this study, we propose a novel resource-efficient approach to optimizing SHM for civil structures. The methodology integrates lightweight machine learning models that rely exclusively on single-sensor data, enabling deployment at the sensor level (smart sensors). By aggregating outputs from multiple sensors, the approach captures spatial information, introduces diversity, and benefits from an averaging effect, significantly improving overall performance compared to individual sensor-based predictions. This single-sensor strategy ensures low computational complexity while maintaining high accuracy, making it particularly suitable for resource-constrained environments. To evaluate the effectiveness of the proposed methodology, we applied it to the Z24 benchmark dataset, a widely recognized SHM resource for civil structures. The objective was to classify various damage scenarios based on data collected from accelerometers deployed on the bridge. The results demonstrate competitive performance with minimal computational complexity, highlighting the scalability and suitability of such an approach for large-scale SHM applications. Ultimately, this study underscores the potential of resource-efficient SHM solutions, contributing to developing sustainable and intelligent monitoring systems.

INTRODUCTION

Traditional structural Health Monitoring (SHM) systems in civil engineering typically rely on fusing high-volume, multi-sensor data at a central node to assess the health of various infrastructure, such as bridges [1]. This approach comes with significant challenges, including heavy communication, storage, and compute costs, which can be

---

Mohamed Abdelillah FIDMA, PhD Student, Email: Mohamed.fidma@univ-eiffel.fr. Experimentation and Modeling for Civil Engineering and Urban (EMGCU), Department of Materials and Structures (MAST), Champs-sur-Marne, France. Jean-Francois BERCHER, Laboratoire d'informatique Gaspard Monge, Univ Gustave Eiffel, CNRS, LIGM, F-77454 Marne-la-Valle´e, France. Franziska SCHMIDT, (EMGCU), Department of Materials and Structures (MAST), Champs-sur-Marne, France

particularly problematic in resource-constrained (edge) deployments where bandwidth and processing power are limited. Additionally, centralized systems are often vulnerable to communication failures, as they depend on a continuous, high-throughput connection between the sensors and the central node [2].

To address these challenges, we propose an alternative strategy: pushing model complexity down onto individual smart sensors. Specifically, we aim to train lightweight classifiers on each sensor, focusing on the local time series data recorded at the given sensor. These per-sensor classifiers will be designed to capture global damage patterns by leveraging a simple aggregation layer that combines outputs from multiple sensors. This approach reduces the computational burden while maintaining effective monitoring capabilities.

In this work, we introduce a damage classification pipeline for the Z24-bridge dataset that operates solely on local single-sensor vibration features, using no cross-sensor preprocessing. This simplifies the architecture, allowing for efficient and scalable deployment on edge hardware. After, we conduct an extensive investigation into the impact of the length of the fed time series segments on model performance. We explore segment lengths ranging from very short windows (64 time samples) to longer windows ($2,048$ time samples). We examine how segment length influences accuracy for both per-sensor classifiers and ensemble models with hyperparameter tuning. We propose an ensemble scheme that fuses per-sensor class probabilities. This aggregation approach helps to capture global damage patterns while maintaining the simplicity and efficiency of local sensor-based classifiers. We present an online inference simulation that profiles the latency of our approach. The simulation demonstrates the feasibility of our method for real-time SHM applications on edge hardware, ensuring that it can be deployed in resource-constrained environments without compromising performance.

DATASET & PREPROCESSING

The Z24 dataset consists of acceleration time series data collected from multiple sensors strategically positioned along the bridge [3]. The recordings capture the bridge's dynamic response across several progressive damage scenarios, such as pier settlement. This benchmark dataset has been studied recently in [4] for SHM damage classification using neural networks. For this study, we focused on a specific multiclass classification task extracted from the broader set of damage scenarios. In particular, we selected the "lowering of pier" experiment, which involves classifying time series into one of five distinct damage categories. After preprocessing, the dataset for this task contained a total of 1,453 labeled time series. Each sensor's raw time series was first segmented into non-overlapping windows of fixed length. Each window inherits the damage label associated with the original full-length recording. This segmentation step increases the number of available training samples while enabling our lightweight models to operate on smaller, localized data segments. Following segmentation, per-sensor min–max normalization, rescaling each segmented window to the [0,1] range, and train/test split were performed for each sensor. Thus, each sensor has its own fully independent train-test dataset used to train a sensor-specific lightweight classifier.

METHODOLOGY

Single-Sensor Classifier Training

For each sensor, we trained an independent lightweight classifier, using only the segmented and normalized time series from that sensor. No cross-sensor preprocessing or feature fusion was performed before training. We explored a set of lightweight models:

- SupervisedTimeSeriesForest: An interval-based random forest classifier designed for time series data [5].
- TimeSeriesForestClassifier: A variant of TSF with modifications improving efficiency and accuracy [6].
- RocketClassifier: Applies a random convolutional kernel transform followed by ridge regression [7].
- Tiny 1D-CNN: A minimal convolutional neural network with a single Conv1D followed by ReLU, then, GlobalMaxPool and Dense(Softmax) architecture, containing approximately 1,000 parameters, which ensure ultra-low-latency inference on microcontrollers.

We assessed both individual sensor models and the final ensemble using the following metrics: Accuracy, Inference Latency (ms), which is, in this study, the prediction time of all the test segments, and Ensemble Diversity, which was quantified using the Average Pairwise Disagreement among sensor classifiers.

Window-Length Study

In this study, we conducted an extensive analysis of the effect of window length M on classifier performance. Specifically, we varied M from a very short (64 time points) to very large ones (4096 time points), observing how it impacted: single-sensor classification accuracy, ensemble accuracy and inference efficiency. Hyperparameters for each model were tuned individually to ensure fair comparison.

Ensemble Averaging

To capture global damage patterns across the bridge, we combined the outputs of individual sensor classifiers. Sensors are installed along the bridge and grouped into different setups as shown in Figure 1, which do not necessarily share the same structural behavior. Each sensor produced a class-probability vector for a given local segment. These vectors were aggregated using simple averaging (soft aggregation). The ensembling is done across either the sensors within the same setup (31 sensors), scenario called “One setup”” or across specific sensors spaced along the bridge with a fixed step, scenario called “Along bridge”; in this case we fix the step at eight, which results in 37 sensors). The final predicted class was the one with the highest average probability. This aggregation layer is lightweight and enables decentralized processing while maintaining strong global performance.

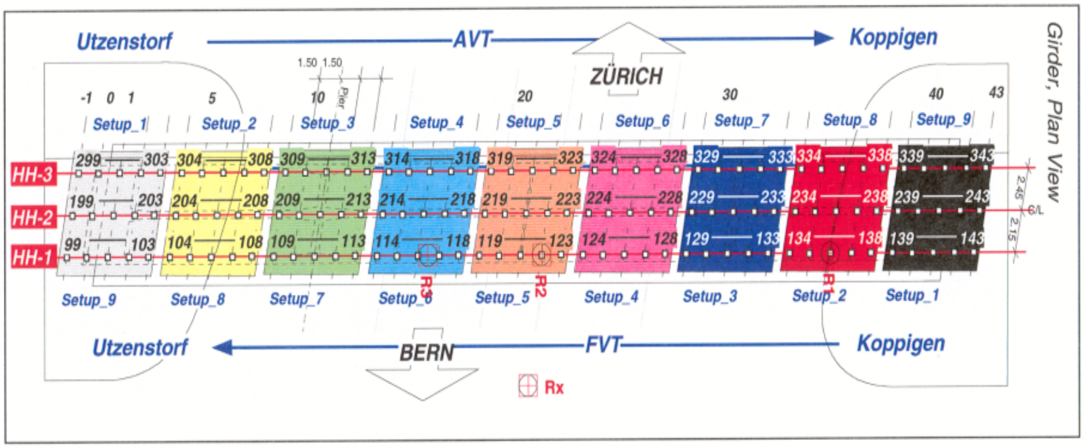

Figure 1. sensor position in the Z24 bridge [3] .

Online Inference Simulation

To verify real-time applicability, we simulated the full online operation pipeline by streaming each incoming segment through its corresponding local classifier, then measured inference latency. This analysis provided concrete validation that the proposed methodology could run on real-world edge devices with stringent computational constraints.

EXPERIMENTAL RESULTS

Selection of Single-Sensor Classifier

Figure 2 presents results comparing the four models in terms of classification accuracy (%) and inference latency (ms) across three window lengths (64, 512, and 2048 time points). STSF achieves the highest overall accuracy, but at the cost of significantly increased inference latency. In contrast, the shallow TinyCNN offers the best latency performance while maintaining reasonable accuracy with short windows. On the other hand, the Rocket classifier is sensitive to window length, showing significant drops in performance for shorter windows. Finally, TSFC offers a balanced trade-off between accuracy and latency, particularly at intermediate window sizes. STSC and Rocket benefit from longer windows, which allow them to better select optimal intervals and kernel parameters, respectively. This generally leads to improved accuracy, although the impact on inference latency varies.

Between these 4 models, TimeSeriesForestClassifier (TSFC) offers a compelling balance of accuracy and latency, while providing extensive hyperparameter flexibility, making it the ideal choice for our single-sensor SHM pipeline. Furthermore, [6], its `parallel_backend="loky"` support streamline distributed hyperparameter sweeps and edge-device deployment [6].

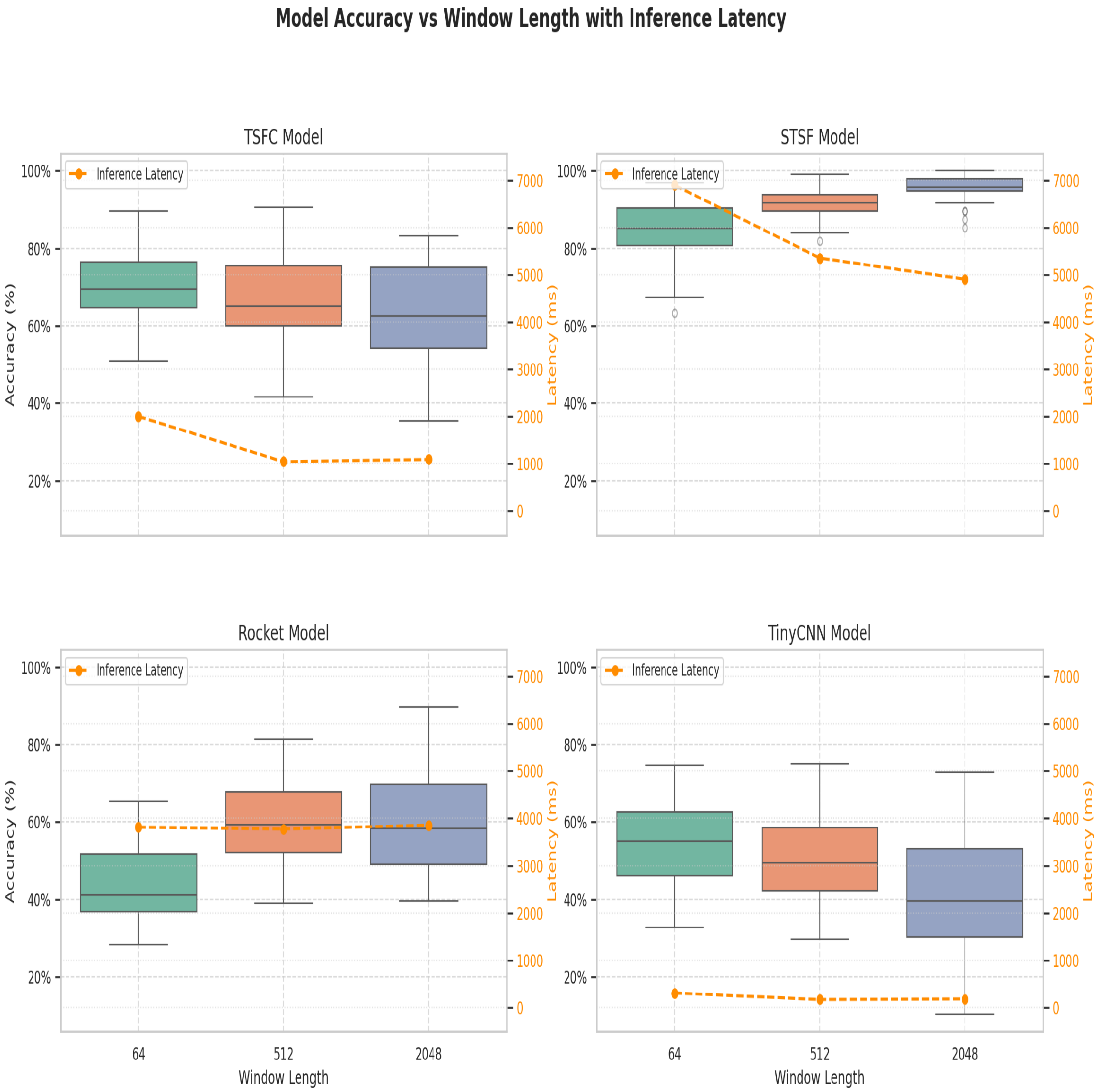

Figure 2. Model Accuracy vs Window length with inference latency.

Impact of Window Length and TSFC Hyperparameter Tuning

To ensure a fair comparison, the prediction time reported refers to the average prediction time per segment in the test set. As shown in Figure 3, accuracy generally decreases as window length increases, although utilizing more estimators consistently boosts accuracy at every window size. Prediction time tends to grow with longer windows, often quite sharply when windows become large, and adding more trees further lengthens inference, even on short windows. This computational cost escalates when both the window length and the number of estimators are high.

Ensemble learning – Aggregation of the sensors prediction models

We considered TSFC single-sensor models computed for $M = 1024$ and fixed TSFC hyperparameters `Min_Interval` = 8, `Max_Interval` = 32. In the TSFC model, we

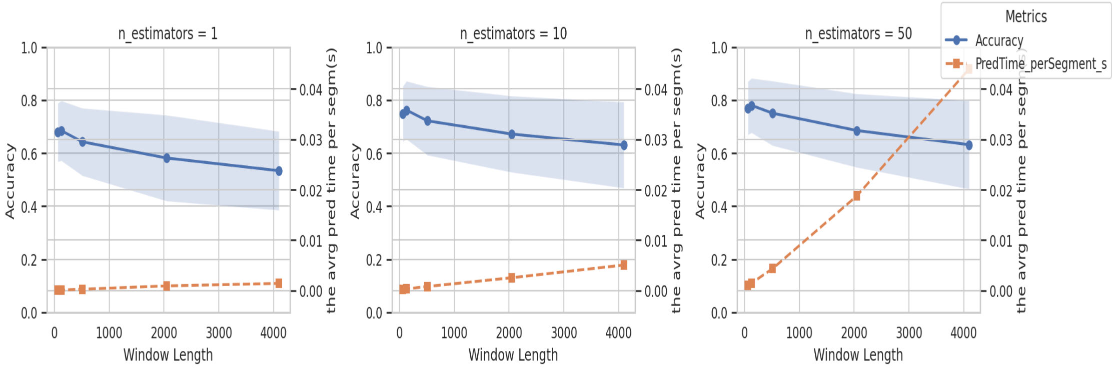

Figure 3. TSFC Accuracies & prediction times across varying number of estimators and window length.

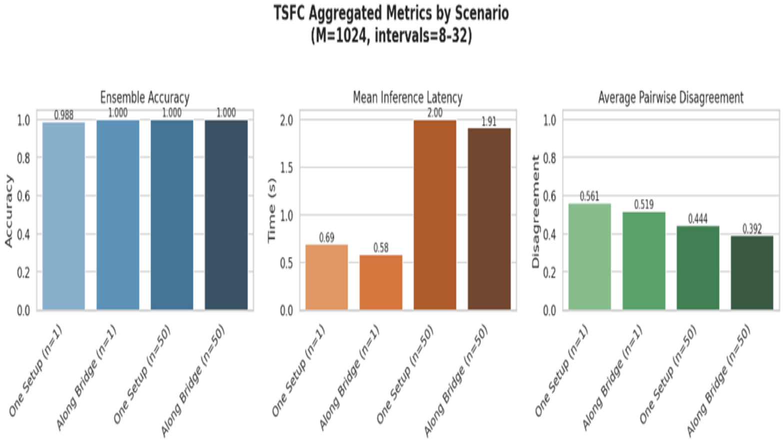

Figure 4. Ensemble accuracy, mean inference latency, and average pairwise disagreement across different configurations.

used different numbers n of estimators with $n = 1$ and $n = 50$. Then we aggregated the class-probability outputs of these single-sensor models for both the “Along Bridge” and “One Setup” deployments.

Figure 4 shows that all configurations achieved very high accuracies with ensemble accuracies approaching 1.000 (100%) for $n = 1$ and $n = 50$. Single-estimator systems are extremely lightweight and fast, making them suitable for near real-time applications. However, larger ensembles ($n = 50$) approximately double or triple the prediction time. Additionally, ensemble disagreement decreases as the number of estimators increases: with only one estimator, there is more significant variability among model predictions, whereas larger ensembles ($n = 50$) produce more consistent and stable outputs, as the models tend to agree more.

Inference simulation

To assess the real-time applicability of the proposed TSFC framework, we simulated the full online inference pipeline. In this setup, each window was sequentially streamed through its corresponding local classifier, emulating the operation of a real-world edge

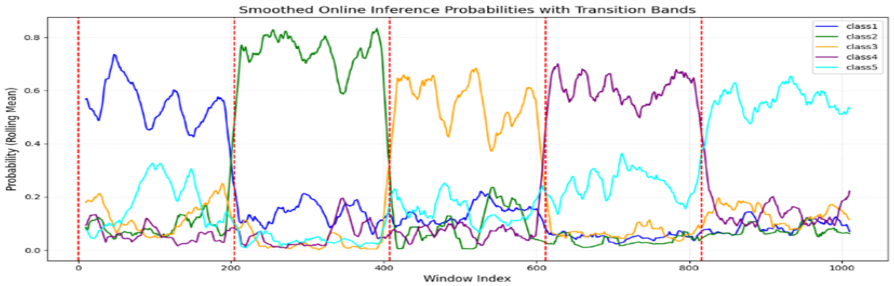

Figure 5. Smoothed Online Inference Probability with Transition Bands.

deployment. We simulated a live data stream by concatenating the previously unseen final 20% of each data block, sliding 75%-overlapping windows through the model, and making one `predict_proba` call per window to fully replicate online inference conditions. The predicted class probabilities from all sensors were then aggregated, averaged, and plotted as a time series to visualize how quickly and confidently the ensemble model "locks on" to each damage class.

The results, presented in Figure 5, demonstrate the framework's performance over 1,000 window indices. Within each structural block, the corresponding class rapidly dominates, and transitions between classes are smoothed by applying a 30%-overlapping rolling mean to the ensemble probabilities, effectively reducing noise while preserving detection accuracy.

DISCUSSION

We propose a low-complexity, resource-efficient approach to structural health monitoring (SHM) that leverages lightweight, single-sensor classifiers to support edge-based deployment while preserving global performance through ensemble aggregation. The results of extensive experimentation using the Z24 bridge benchmark dataset substantiate the feasibility of deploying such models on low-power devices without significant trade-offs in classification accuracy.

The impact of window length on classification performance highlighted important trade-offs. Shorter windows, while reducing model complexity and memory requirements, do not diminish accuracy. Importantly, TSFC proved flexible in adapting to different window lengths through hyperparameter tuning, with ensemble configurations further mitigating the limitations of any single window size choice. Our ensemble averaging strategy demonstrated that even when weak individual classifiers operate with limited local information, aggregating their outputs can yield very robust global predictions. This validates the central hypothesis of the study, that local, independent models can collectively infer global structural states. Interestingly, the accuracy of the ensemble remained consistently high regardless of the number of estimators per sensor, although using larger ensembles (e.g., $n = 50$) significantly reduced the disagreement of the ensemble and improved prediction stability at the cost of longer inference time. Finally, the online inference simulation confirmed that the system is capable of real-time operation under realistic streaming conditions. The ensemble's ability to rapidly lock on to the correct damage class and maintain high confidence over time reinforces its applicability to safety-critical infrastructure monitoring scenarios.

In general, the findings highlight the practicality and scalability of the proposed SHM framework. The system enables decentralized monitoring and efficient bandwidth usage for large-scale deployments by offloading computation to smart sensors and relying on minimal coordination.

CONCLUSION & PERSPECTIVES

In this study, we present a decentralized, resource-efficient SHM pipeline that operates on single-sensor vibration segments, uses lightweight TSFC classifiers tuned for optimal accuracy-latency trade-offs, and aggregates outputs via simple averaging to capture global damage patterns. Our extensive experiments on the Z24 benchmark show that carefully chosen window lengths and forest sizes yield per-sensor accuracies above 95% with inference speedups over 60%, while ensemble averaging drives overall accuracy to nearly 100%. Online inference simulations demonstrate robustness to sensor relocation and real-time practicality on edge hardware.

While single-sensor models minimize communication and computation [1], they rely on predetermined segmentation and per-sensor normalization. Future work should explore adaptive windowing and online normalization to handle non-stationary vibration signals. Finally, evaluating the generalizability of this framework across a wider range of SHM datasets and damage scenarios remains an important direction for future validation.

REFERENCES

1. Zinno, R., S. S. Haghshenas, G. Guido, and A. VItale. 2022. “Artificial intelligence and structural health monitoring of bridges: A review of the state-of-the-art,” *IEEE Access*, 10:88058–88078.
2. Payawal, J. M. G. and D.-K. Kim. 2024. “A Review on the Latest Advancements and Innovation Trends in Vibration-Based Structural Health Monitoring (SHM) Techniques for Improved Maintenance of Steel Slit Damper (SSD),” *IEEE Access*.
3. Department of Civil Engineering, KU Leuven. n.d., “Z24 Bridge: Structural Health Monitoring Benchmark,” https://bwk.kuleuven.be/bwm/z24.
4. Dabbous, A., R. Berta, M. Fresta, H. Ballout, L. Lazzaroni, and F. Bellotti. 2024. “Bringing Intelligence to the Edge for Structural Health Monitoring: The Case Study of the Z24 Bridge,” *IEEE Open Journal of the Industrial Electronics Society*, 5:781–794, doi: 10.1109/OJIES.2024.3434341.
5. Cabello, N., E. Naghizade, J. Qi, and L. Kulik. 2020. “Fast and accurate time series classification through supervised interval search,” in *2020 IEEE international conference on data mining (ICDM)*, IEEE, pp. 948–953.
6. Nurlatifa, H., T. B. Adji, I. Ardiyanto, A. P. Sidhawara, and S. Wibirama. 2024. “Time Series Classification for Eye Tracking-Based Visual-Verbal Learning Style,” in *2024 16th International Conference on Information Technology and Electrical Engineering (ICITEE)*, IEEE, pp. 91–96.
7. Dempster, A., F. Petitjean, and G. I. Webb. 2020. “ROCKET: exceptionally fast and accurate time series classification using random convolutional kernels,” *Data Mining and Knowledge Discovery*, 34(5):1454–1495.