Information Fusion 118 (2025) 102983

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Full length article

Next-generation coupled structure-human sensing technology: Enhanced pedestrian-bridge interaction analysis using data fusion and machine learning

Sahar Hassani [a,*], Samir Mustapha [b], Jianchun Li [c], Mohsen Mousavi [a], Ulrike Dackermann [a]

[a] Centre for Infrastructure Engineering and Safety, School of Civil and Environmental Engineering, University of New South Wales, Sydney, NSW, Australia
[b] Laboratory of Smart Structures and Structural Integrity (SSSI), Department of Mechanical Engineering, American University of Beirut, Beirut, Lebanon
[c] School of Civil and Environmental Engineering, University of Technology Sydney, Ultimo, NSW, 2007, Australia

ARTICLE INFO

Keywords:
Crowd monitoring
Structural health monitoring
Sensor fusion
Next-generation sensors
Signal processing
Machine learning

ABSTRACT

The consequences of crowd behavior in high-density pedestrian flows, especially in response to exacerbating incidents, can result in tragic outcomes such as trampling and crushing, making the active monitoring of crowd motion crucial, to provide timely danger warnings and implement preventive measures. This paper proposes a novel approach for crowd behavior monitoring and prediction of bridge loads based on the following innovative solutions: (a) advanced optimized signal processing is leveraged for noise reduction; (b) novel data fusion approaches are proposed to extract the most informative measurement features; (c) fine-tuned machine-learning techniques are implemented for classification and regression tasks. Data from structure-based sensors and wearable devices were utilized to capture movement- and load-sensitive data on a pedestrian bridge, which facilitated the determination of crowd flow, density, and bridge loading information. The proposed monitoring approach explores signal preprocessing methodologies, including variational mode decomposition (VMD), downsampling, principal component analysis, and novel data fusion, to effectively minimize noise and errors in the input data. Data fusion strategies were introduced to significantly enhance the learning models and improve the overall efficiency and resilience of the system. For further analysis, a 2D-convolutional neural network (CNN) approach was initially applied independently to the sensing sources and subsequently extended to fuse multimodal raw, decomposed, and denoised data. The proposed monitoring method was validated using experimental data obtained from crowd simulations conducted on a scaled-down bridge panel, utilizing next-generation coupled structure-human sensing, fiber-optic sensing, and smartphone technology. The results demonstrated a high level of accuracy for crowd monitoring predictions, with the peak testing accuracy reaching 99.62% for single-class crowd flow classification, 98.69% for multiclass crowd flow and density classification, and 98.42% in R^2 score for load estimation when fusing denoised signals using VMD. The proposed 2D-CNN model was compared with an existing adaptive Kalman filter (AKF) fusion technique and various machine learning techniques, including random forest, k-nearest neighbor, support vector machine, XGBoost, and ensemble methods. This comparison unequivocally confirmed the robustness and superiority of the proposed monitoring approach.

1. Introduction

The safety of crowds during events, such as festivals, religious ceremonies, sports competitions, concerts, and political rallies has become an issue of significant importance. These gatherings have the potential to lead to large-scale assemblies where the probability of crowd-related disasters is substantial. Numerous global festivals have encountered substantial challenges related to crowd safety, primarily due to the large turnout of participants. On various occasions, the disregard for crowd safety measures has resulted in the loss of numerous lives [1]. Therefore, understanding crowd safety and implementing related safety measures are vital for planning the diverse array of events that occur each year.

Pedestrian-related disasters often stem from inadequate crowd management, excessive crowd density, and limited entry points. The chaos and panic that often ensue tend to aggravate the magnitude of injuries and casualties, which often surpass the impact of the initial triggering event. Therefore, meticulous crowd management is essential

* Corresponding author.
E-mail address: sahar.hassani@unsw.edu.au (S. Hassani).

https://doi.org/10.1016/j.inffus.2025.102983
Received 11 January 2024; Received in revised form 10 November 2024; Accepted 26 January 2025
Available online 5 February 2025

for ensuring the overall safety and well-being of participants. The spectators' passion during events often leads to spontaneous or coordinated jumping motions, potentially generating harmful vibrations that can affect structural integrity. Synchronized crowd movements instigated by music can exacerbate these vibrations, further challenging structural safety. The variability in the load distribution can also influence the structural response. Consequently, vigilant monitoring during events and assessment of structural behavior are essential to ensure spectator comfort and safety [2].

A rare yet profoundly significant crowd-related catastrophe that can occur during large gatherings is the failure of pedestrian bridges due to excessive loads or deterioration. The factors contributing to such a collapse may include overcrowding, design flaws, material deficiencies, inferior construction, insufficient maintenance, and unexpected environmental conditions. In Phnom Penh, Cambodia, in 2010, approximately 450 people died in a stampede caused by a swaying suspension footbridge [3]. In 2013, 89 people were killed in a stampede on a bridge at the Ratangarh Temple, India [3]. In March 2018, near Florida International University in Miami, U.S.A., six people died, and many others were injured after a 53-meter footbridge collapsed over an eight-lane road [4]. Other examples include a pedestrian bridge linking an island in Prague, Czech Republic, which completely collapsed because of the corrosion of steel ropes [5] and in 2018, the Morandi Bridge in Genoa, Italy collapsed, making it one of the most prominent pedestrian bridges in history. A section of the bridge collapsed onto a busy highway in this catastrophic event, causing numerous deaths and injuries [6]. A framework for combined structure-crowd monitoring was introduced by Mustapha et al. [7,8], whereby a joint methodology was proposed for structural health monitoring (SHM) and crowd monitoring by employing support vector machines (SVM) for classification and regression, using independent inputs from wearable devices and structural sensors.

This paper presents a novel methodology for crowd monitoring on pedestrian bridges by integrating a next-generation coupled structure called human sensing. Pedestrians are equipped with smartphones containing embedded accelerometers, and strain fiber optic sensors (FOS) are installed on the bridge structures. Strain measurements provide insights into the collective behavior of crowds and mobiles, whereas acceleration measurements provide individual-level information. This method uses artificial intelligence to determine pedestrian speeds and predict the load levels on bridges. Data from various sensor sources are fused to ensure redundancy and system robustness. Metrics are further introduced for evaluating machine learning techniques based on multiple crowd attributes and data sources.

The remainder of this paper is organized as follows. Section 2 describes the background, outlining the motivation and objectives of the proposed approach. Section 3 details the conceptual solution approach from a broader perspective, with emphasis on signal processing, data fusion, and the 2D-CNN model. Section 4 describes the complete experimental setup, including the testbed, instrumentation, and crowd replication. Section 5 presents the results of the 2D-CNN classification and regression models. Section 6 compares the proposed approach with existing methods, and Section 7 concludes the paper with closing remarks and suggestions for future work.

2. Background, related work, and motivation

Crowd management [9] is a multifaceted endeavor, encompassing key elements such as crowd modeling for event planning, design of suitable infrastructure, real-time crowd data acquisition to monitor crowds, data analysis for decision-making, and effective implementation of crowd control solutions. The complexity of this process requires the convergence of many fields including physics, computer science, civil engineering, and management. In crowd management, collection, organization, and analysis of crowd data during event execution are of paramount importance in preventing potential crowd disasters. An effective crowd-management system relies on accurate crowd-related information. While existing systems primarily implement vision-based technologies [10] such as closed-circuit television (CCTV) monitoring, a forward-looking perspective takes advantage of vision-based, wireless/radio frequency (RF), and web/social media data mining as three key technologies for crowd information acquisition.

The vision-based approach to crowd monitoring utilizes sources, such as on-site CCTV cameras, aerial and satellite images, and infrared cameras, to retrieve crowd-related information. Researchers have developed crowd counting, monitoring, and congestion analysis systems using advanced artificial intelligence (AI) image processing and computer vision techniques [10,11]. Convolutional neural networks (CNNs) significantly enhance the accuracy and efficiency of image classification systems [12]. Early CNN applications for crowd density estimation included the works of Fu et al. [13] and Wang et al. [14]. A multicolumn CNN was employed by Zhang et al. [15] to produce local density maps, whereas Boominathan et al. [16] captured semantic information to predict crowd density. Sindagi and Patel [17] proposed a model to classify crowd images into various densities, offering a coarse estimation of the number of individuals. Pattern recognition and predictive modeling have also been conducted using SVM and decision trees [18].

RF-based technologies, including mobile phones, radio frequency identification (RFID), wireless sensor networks (WSN), and near-field communication (NFC) offer significant potential for improving crowd management and safety in smart cities [19]. These technologies provide diverse solutions for enhancing urban efficiency such as real-time crowd density, movement, and behavior prediction using mobile phones with RF capabilities. The RF identification (RFID) technology can streamline access control and optimize crowd flow in civil infrastructures and wearable devices. Advancements in noncontact sensing techniques such as drones and robotic sensors, along with wearable devices such as smartphones and smartwatches, can further support crowd monitoring. Equipped with various sensors and wireless interfaces, smartphones can assess the activity status of an individual, facilitate human activity recognition (HAR), and monitor crowd dynamics [7]. These technological advancements offer new opportunities for enhancing crowd management strategies and ensuring urban safety in smart cities.

Data fusion is an effective solution for efficient analysis of the data recorded by diverse sensors, as described by Hassani et al. [20]. Different data-fusion hierarchies can be applied to address specific challenges. The first hierarchy involves data acquisition, raw data fusion, feature extraction, and decision-making using machine-learning outcomes. The second hierarchy includes data acquisition, feature extraction, feature fusion, and machine learning-based decision making. The third hierarchy begins with data acquisition, followed by feature extraction and decision-making using multiple machine learning models, and concludes with decision fusion to produce a final comprehensive decision.

Deep-learning techniques can automatically extract features through intermediate layers; however, they require extensive datasets. For limited datasets, manual design and feature extraction are advised. Feature extraction aims to extract relevant information from raw sensor data for use in machine-learning algorithms. Time-domain features are derived directly from sensor data to capture temporal characteristics [21] based on coefficients or residuals [22]. Frequency-domain features are obtained using Fourier transforms and include metrics such as power spectrum density, dominant and resonance frequencies. Time–frequency features analyze signals using short-time Fourier transform (STFT), wavelet transform (WT), empirical mode decomposition (EMD), and variational mode decomposition (VMD), provide joint time–frequency representations of the signals.

Table 1 compares the proposed approaches and solutions aimed at addressing prevalent gaps in crowd analysis and response prediction. These gaps include insufficient information, inaccurate results, noise, outliers, limited generalization, interpretability issues, dynamic environments, real-time performance constraints, and technical scalability for which various novel advanced data preprocessing, data fusion, and optimized deep learning techniques have been proposed. In this study, these solutions are integrated into a cohesive framework utilizing multimodal data fusion, advanced preprocessing, and optimized deep learning architectures to provide a comprehensive solution to these challenges. This integrated approach aims to develop robust and effective crowd analysis systems capable of handling the complexities of real-world scenarios. This innovative approach provides novel insights into crowd monitoring, signal processing, data fusion, and optimized deep learning models to enhance pedestrian safety and the management of pedestrian bridge infrastructures.

1. **Advanced signal processing techniques:** In crowd management, noisy and error-prone signals are challenging to handle and affect decision-making accuracy. Our methodology addresses this problem using advanced signal processing techniques for feature extraction and denoising, thereby enhancing prediction accuracy. We employed the VMD algorithm for signal decomposition to reveal the underlying characteristics, followed by two denoising methods, VMD and downsampling, to mitigate noise. Relevant features were extracted using VMD and integrated with raw data through fusion methods, thereby improving the robustness and informativeness of the data analysis.
2. **Data fusion techniques:** The proposed method employs data fusion techniques at different levels (signal, feature, and decision) to combine data from different sensors, including wearable and fiber Bragg grating (FBG) sensors, utilizing four distinct methods to enhance accuracy.
3. **Deep learning model:** For signal analysis, a robust 2D-CNN architecture that achieves high accuracy across various classification and regression tasks is proposed. This framework accommodates different combinations of raw data and processed signals, to offer a versatile solution.
4. **Comparison:** The performance of the proposed 2D-CNN architecture on crowd analysis was compared with those of different data fusion methods, such as adaptive Kalman filtering (AKF) fusion and various machine learning models, including random forest, k-nearest neighbors, support vector machines, extreme gradient boosting (XGBoost), and ensemble methods.
5. **Measurement signals:** Effective decision-making in crowd management relies on accurate measurement data. This study focused on collecting essential signals for crowd analysis through next-generation structural sensing by leveraging ubiquitous devices such as smartphones and smartwatches for easy signal measurement. The pedestrian safety during crowded events was enhanced by utilizing wireless networks to handle the data from wearable sensing devices. Structural sensing improves the system accuracy by combining measurements from wearable sensors with data from the fiber-optic sensors installed on bridges. This method processes data in real time to assess crowd loads and movement patterns, thereby significantly enhancing the performance of advanced machine learning algorithms for crowd monitoring, which improves pedestrian safety.
6. **Experiment configuration:** Validation of the proposed technique involved examining various types of pedestrian movements, considering variations in speed, pedestrian numbers, and their weights on a pedestrian bridge equipped with strain sensors. Subsequently, data were collected using sensors such as accelerometers, gyroscopes, and magnetometers from the smartphones carried by the pedestrians. The proposed approach for predicting maximum load effects (MLEs) on bridges was subsequently confirmed using solely smartphones as wearable sensors.

3. Methodological framework: Crafting the proposed approach

Fig. 1 provides an overview of the proposed approach, emphasizing the integration of data collected from smartphones, smartwatches, and structural sensors installed on bridges. This data merging collectively enhances the accuracy and speed of load prediction for improved bridge safety. In this framework, structure-based FBG-FOS strain networks are installed on critical structures, such as bridges, and crowd participants wear devices, such as smartphones and smartwatches. Data collected from different sensors are uploaded to the cloud in real time using existing communication infrastructures for processing and decision-making. Advanced machine learning approaches are employed to determine crowd flow classes and estimates, analyzing the data collected from both structural sensors and mobile accelerometers. Crowd flow classification is based on either a binary decision of motion speed corresponding to fast or slow, or multiclass decision based on fast-heavy, fast-light, slow-heavy, and slow-light, with heavy and light referring to crowd load designations. The total weight of the crowd on the structure is estimated in kilograms based on regression modeling.

Deep learning methods [29,30], particularly CNNs, have gained popularity for their ability to capture complex patterns in data. In this study, we employed a 2D-CNN to address both classification and regression problems. A key feature of our approach is the data fusion of FBG-FOS and accelerometer data implemented at three levels: input (through concatenation and mean), feature (through feature fusion), and decision (through ensembling the results of different machine learning models). For signal decomposition and denoising, the VMD algorithm and downsampling techniques are used. The VMD algorithm also helps extract statistical features from the signals, enhancing the accuracy and efficiency of data analysis by revealing hidden information. The parameters of the 2D-CNN were optimized to improve crowd load predictions and accelerate the calculations with the aim of finding the best set of hyperparameters for optimal model performance. To evaluate the performance and generalization ability of the model, five-fold cross-validation was used within the 2D-CNN framework. This rigorous validation method aids in model selection and hyperparameter optimization, thereby ensuring overall model robustness. The following sections describe different aspects of the proposed methodology.

3.1. Artificial intelligence techniques

In recent decades, significant advances in AI, including machine and deep learning, have substantially transformed the field of data analysis. Machine learning techniques comprise a group of learning methods that automate the process of building analytical models, whereas deep learning methods, one of the most advanced subsets of AI systems, are employed to build intelligent systems and automation models [29,30]. With deep-learning systems, nonlinear relationships can be modeled using a neural network architecture with multiple layers of neurons to retrieve higher-level features from a large amount of data. Deep-learning models have been successfully applied across a wide range of technological areas, including crowd management. In this study, different data analysis tools, including the 2D-CNN deep learning model, and machine learning techniques such as random forest, KNN, SVM, XGBoost, and ensemble methods were employed. The input data acquired by FBG-FOS sensors comprised strain data samples, each spanning 10 s, captured at a rate of 100 Hz. These models were trained using data originating from either a single sensor, resulting in 1000 data points for each sample, or from an amalgamation of data from three sensors located across the bridge, resulting in a sample dimension of 3000 data points. For both the classification and regression tasks, either raw or unprocessed data or data that had undergone decomposition and noise filtering were fed into the model. The smartphone-based accelerometer data were preprocessed and only the magnitude of the triaxial acceleration was analyzed. In both classification and regression scenarios, the sum of the accelerometer data across all individuals within the crowd was used as the input data.

Table 1
Comparison of the proposed approach and solutions in crowd analysis and response prediction in recent literature.

Gaps	Refs	Solution	Description
Insufficient information	[23]	Exploiting information fusion	The proposed data fusion approaches integrate multimodal data sources to augment ground truth data and facilitate robust model training.
Low accuracy	[24]	Improving model accuracy	The proposed DL approach implements robust evaluation and validation to ensure accurate results and mitigate errors in predictions.
Presence of noise and outliers	[25,26]	Advanced data preprocessing	Advanced data preprocessing techniques such as the optimized VMD algorithm are applied to remove noise and outliers from the input data, enhancing the reliability and robustness of the predictive model.
Limited generalization	[25,27]	Robust data analysis	The fine-tuned deep learning model (2D-CNN) employs transfer learning techniques to enhance generalization across diverse crowd scenarios.
Dynamic environments	[28]	Minimizing environmental effects	The proposed signal processing techniques dynamically adjust model parameters based on real-time environmental inputs, improving adaptation to dynamic crowd behavior.
Real-time performance	[27]	Improved computational efficiency	The proposed approach utilizes optimized deep learning architectures and parallel processing techniques to improve real-time performance in crowd analysis applications.
Technical scalability	[25,28]	Novel model architecture	The proposed approach leverages scalable deep learning architectures and distributed computing infrastructure to efficiently handle large-scale crowd analysis tasks.

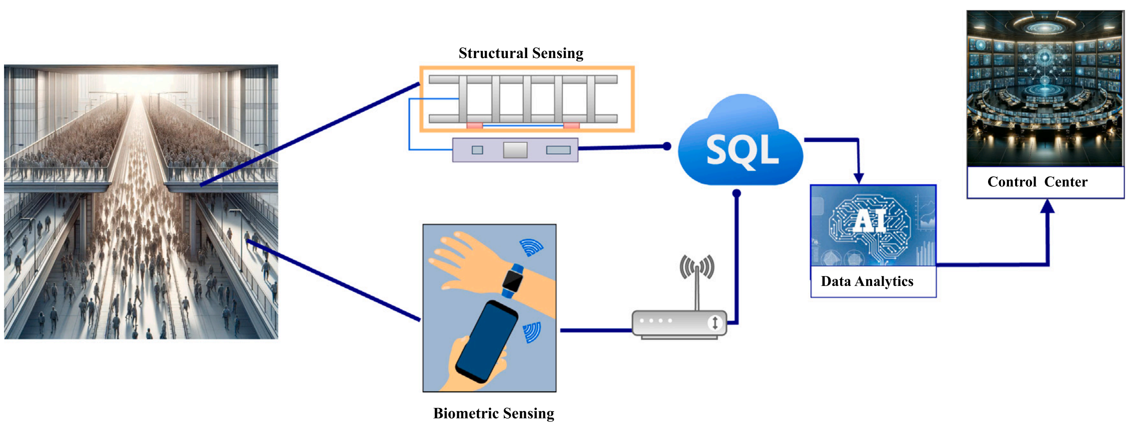

Fig. 1. Conceptualized merged structure and crowd monitoring framework.

3.1.1. Convolutional neural network model

Convolutional neural networks are specialized feed-forward neural networks designed to automatically learn features through filter (or kernel) optimization. They address issues such as vanishing and exploding gradients observed during backpropagation in the earlier neural networks by employing regularized weights across fewer connections. The CNN architecture comprises three distinct layers: convolutional, pooling, and fully connected. The input is transformed into a tensor shape of (number of inputs) × (input height) × (input width) × (input channel). In the convolutional layer, data are abstracted into a feature map, also called an activation map, with the following dimensions: (number of inputs) × (feature map height) × (feature map width) × (feature map channel).

Researchers have demonstrated that CNNs can be effectively used for tasks involving vision-based classification, such as object recognition and digit/character recognition [31,32]. By combining a deep supervised learning architecture with back-propagation, features can be automatically extracted from data without manual, handcrafted feature-based selection [33]. CNNs can autonomously identify and extract globally relevant discriminatory features from the data, thereby eliminating the need for human intervention in the design process. A significant advantage of extracting features from input images is that the process is insensitive to shape distortions and shifts even when textual image inputs are involved [34].

A CNN typically comprises one or more convolutional layers composed of different layers of filters that compute different feature maps. A feature map is generated by convolving each filter with the input and subsequently applying an element-wise nonlinear activation function. Common activation functions include sigmoid, tanh, and rectified linear unit (ReLU). The complexity level of the feature representations increases as the number of convolutions increases. Following the convolutional layer, the pooling layer decreases the resolution of the feature map to introduce shift invariance and efficiency. Average and maximum pooling are the most common pooling layers. After the convolutional and pooling layers, a fully connected layer is integrated. The output layer of a classification CNN typically uses softmax to compute the probability distribution across predicted classes, followed by a classification layer. In regression scenarios, the classification operation is substituted with a regression layer. Notably, CNNs were initially designed for image recognition. However, they can be applied to signals via a one-dimensional variant referred to as 1D CNN [35], which differs from the 2D-CNN [36] used for two-dimensional data.

2D-CNNs are widely used in computer vision tasks because of their ability to capture spatial hierarchies and patterns in the data. At its core, a 2D-CNN employs convolutional layers that apply sliding filters across two-dimensional data, to identify local features, such as edges and textures. Activation functions, often ReLUs, introduce nonlinearity into the network, enhancing its capacity to learn intricate patterns. Pooling layers downsample feature maps and maintain essential information while reducing computational complexity. Fully connected layers perform high-level decision making based on learned features, whereas techniques such as dropout and batch normalization mitigate overfitting and stabilize the training process. 2D-CNNs excel at capturing spatial hierarchies and have become the cornerstone of

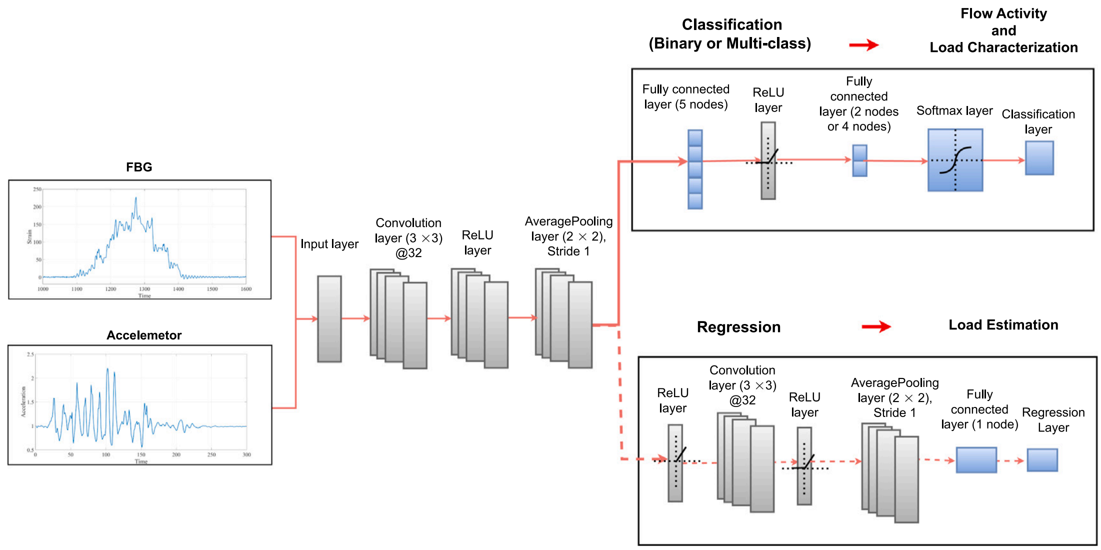

Fig. 2. Proposed 2D-CNN architecture motion speed and load classification through regression of load estimation.

various computer vision tasks, including image classification and object detection.

In this study, we employed a 2D-CNN framework to classify and regress crowd attributes. The model architecture is depicted in Fig. 2. The input to the 2D-CNN model is denoised strain and accelerometer data in raw or decomposed data format. The model comprises an initial input layer, followed by a convolution layer equipped with 32 2D filters, a ReLU layer, and an average-pooling layer for both classification and regression tasks. For classification tasks, the architectural flow continues with a fully connected layer, a ReLU layer, and another fully connected layer connected to a softmax operator, culminating in the classification layer. When performing regression, the network architecture continues with the inclusion of another convolution layer, followed by a ReLU activation function layer, an average-pooling layer, and a fully connected layer with one node, with the final layer substituted with a regression layer. Batch normalization is applied in the classification setup to mitigate the impact of minor load variations, which is a desirable feature for regression but less so for classification.

3.1.2. Cross-validation

Cross-validation is one of the most extensively employed techniques for data resampling and serves two crucial purposes: estimating the true predictive error of models and fine-tuning model parameters, thereby playing a pivotal role in assessing the generalization capabilities of predictive models, while effectively guarding against overfitting. Falling under the umbrella of Monte Carlo methods, cross validation resembles bootstrapping in its resampling approach. The dataset is first decomposed into two subsets for validation and testing, with an 80%–20% split ratio. The validation subset is further divided into five folds, to perform five-fold cross-validation. Each iteration involves splitting the dataset into two complementary subsets. The first subset, known as the training set, comprises four of the five folds (80% of the validation data) and serves as the foundation upon which the model is trained. The model is tested on the second subset, which consists of the remaining fold (20% of the data) against which the performance of the analysis is assessed. This process is repeated five times, with each of the five folds used for testing at each iteration. The validation outcomes yield an optimized model that is further tested on the initial test set. This procedure helps estimate the model performance more robustly than a single train-test split, allowing a more comprehensive evaluation of the optimized model performance on unseen data, providing insights into the overall performance and mitigating the risk of overfitting.

3.2. Signal processing methods

In this study, different signal processing techniques were applied to decompose and denoise the raw data. The data were separated into fundamental components and unwanted noise was removed to enhance data quality. These preprocessing stages are crucial for extracting patterns and meaningful insights from the data. The following section provides a detailed explanation of the specific signal processing methods, their application to raw data, and the benefits they provide for data analysis. Ultimately, these techniques enhance the extraction of meaningful information and aid in the understanding of the underlying information in a dataset.

3.2.1. Decomposition technique: VMD algorithm

In this study, a signal decomposition algorithm called VMD was employed to decompose a signal into multiple narrowband oscillation modes. This enables the extraction of relevant information across various signal bands. VMD was further utilized for denoising and breaking down the raw signal into multiple sub-signals to be input into the 2D-CNN model. This facilitates feature extraction from the constituent intrinsic mode functions (IMFs) derived from a signal, whereby the extracted features are used as inputs to train machine-learning models, enhancing their accuracy and effectiveness in classification or regression tasks.

The VMD algorithm is a dynamic signal decomposition technique that breaks down a signal $f(t)$ into a set of K narrow-band Amplitude-Modulated Frequency-Modulated (AM-FM) or IMFs. An IMF can be expressed in the following manner:

$$u_k(t) = A_k(t)\,\mathbf{cos}(\phi_k(t)), \tag{1}$$

In this context, $u_k(t)$ represents the kth IMF, with $A_k(t)$ and $\phi_k(t)$ referring to instantaneous frequency and phase, respectively. Given the narrow-band nature of each IMF, a Gabor analytical signal can be created from the IMFs, allowing the determination of the instantaneous phase. Each IMF is defined by its center frequency ω_k. To compute u_k and ω_k, VMD optimizes the following augmented Lagrangian:

$$\mathcal{L}(\{u_k\},\{\omega_k\},\lambda) = \alpha \sum_k \left\| \partial_t \left(\delta(t) + \frac{j}{\pi t} * u_k(t) \right) \times e^{-j\omega_k t} \right\|_2^2 + \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle \tag{2}$$

where $\|.\|_2$ is the L^2 norm; $*$ represents convolution; and j is the imaginary unit. The penalty factor α is a denoising factor indicating

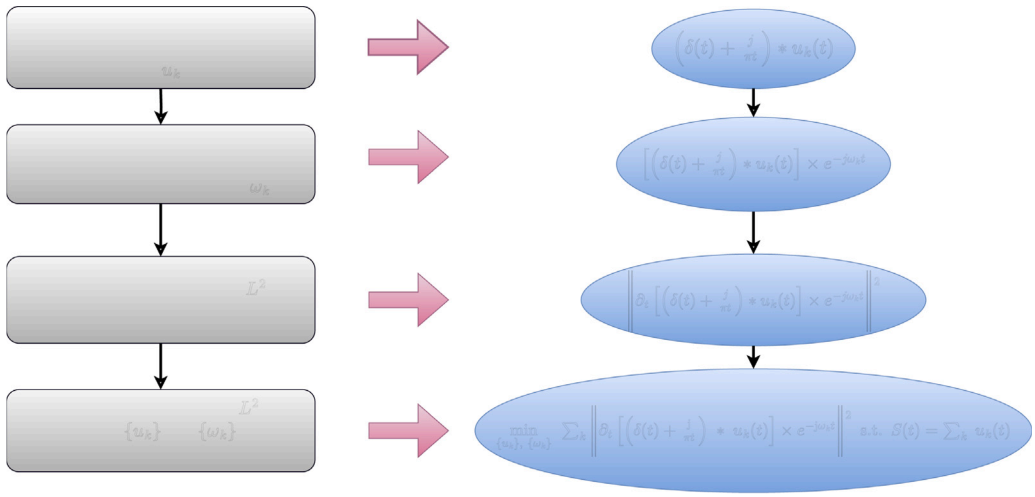

Fig. 3. Procedure for obtaining the objective function of the VMD decomposition algorithm.

Table 2
Selected VMD parameters for the decomposition task.

Parameters	Description	Specified values
p	Number of IMFs	3
α	Denoising factor	10
τ	Time interval	0.1
ϵ	Convergence threshold	10^{-5}
$init$	Center frequency initializer	0
DC	Boolean parameter	0

Table 3
The selected VMD parameters for the denoising task.

Parameters	Description	Specified values
p	Number of IMFs	3
α	Denoising factor	100
τ	Time interval	0
ϵ	Convergence threshold	10^{-5}
$init$	Center frequency initializer	0
DC	Boolean parameter	0

the importance of the first term with respect to the second and third terms in (2).

Fig. 3 presents a hierarchical representation of the steps involved in obtaining the VMD optimization objective outlined in Eq. (2). Consequently, the VMD algorithm has a group of parameters that must be specified *a priori* [37], as follows:

1. K determines the number of IMFs the original signal will be decomposed into.
2. α is a quadratic penalty term and a denoising factor.
3. τ is a time step that determines the speed at which the Lagrangian multiplier accumulates the reconstruction error. Setting τ to a small number, such as 0.1, affects α, rendering denoising ineffective.
4. ϵ is a tolerance parameter and controls the convergence of the algorithm.
5. $init$ initializes the center frequencies. Options are zero ($init = 0$), uniform ($init = 1$), and random ($init = 2$).
6. DC, which is a Boolean parameter, determines whether the first mode is set and maintained at DC (an IMF with zero center frequency).

Fine-tuning of these parameters can be achieved through two main approaches: the first involves the use of parallel optimization algorithms to enhance the utility of VMD-derived features for machine learning tasks, whereas the second relies on domain-specific knowledge and experience. Each approach has its advantages and drawbacks, with the former being ideal for large datasets in mitigating overfitting at the expense of computational costs, whereas the latter is more suitable for situations with limited data, requiring a deep understanding of the problem domain. For this research, the second approach was selected to assign the VMD parameters. The VMD parameters used in the study are listed in Table 2.

3.2.2. Denoising methods

In the context of signal processing and data analysis, denoising methods are employed to remove or reduce unwanted noise or interference from a signal, dataset, or image while preserving the underlying meaningful information. Noise is typically random and leads to unwanted variations in data that not only obscure the true signal but also impede the extraction of useful information. Common denoising techniques include the VMD algorithm, wavelet denoising, low-pass filtering, and downsampling. To reduce the noise in input signals, this study used the VMD algorithm and downsampling.

In addition to data decomposition, VMD can also be used as a denoising technique to reduce noise features from signals or data. There are two methods for performing denoising using VMD. The first approach involves smoothing the input signal before decomposing it, which requires the following:

1. The parameter τ needs to be set to zero. This allows denoising to occur during decomposition.
2. Subsequently, a value larger than zero needs to be specified for the denoising factor α. A larger value of α introduces a more severe denoising effect during decomposition.

The second approach relies on decomposing the signal without smoothing it by setting τ to a small value such as 0.1 and then removing the IMFs with higher-frequency content representing noise in the signal. We used the first strategy to concurrently denoise and decompose signals. Table 3 presents the selected and optimized hyperparameter values for denoising using VMD. A moderate denoising factor of $\alpha = 100$ was considered in this study.

In downsampling, a denoising technique is employed to decrease the volume of data by reducing the signal sampling rate. Downsampling, also known as decimation, reduces the number of samples in a signal by selecting a subset of the original samples at regular intervals. This results in a lower sampling rate and consequently, a signal with reduced data size. Downsampling operations were performed at different rates as follows:

- **Downsampling I:** This involves downsampling by a factor of three, thereby reducing the number of samples in the signal to one-third of the original sampling rate. In addition to reducing high-frequency noise and conserving computational resources, this can affect the fine-grained details.
- **Downsampling II:** This involves downsampling by a factor of five, whereby every fifth sample of the original signal is retained, discarding the rest, which significantly reduces high-frequency noise by further reducing the data size. However, this increases the risk of information loss, particularly if there are high-frequency signal components.
- **Downsampling III:** This involves downsampling by a factor of ten, which is a significant reduction in the sampling rate, retaining only every tenth sample. This can be an effective method for removing high-frequency noise; however, the risk of losing critical signal information is greater, particularly if the original signal contains fine details or rapid changes.

3.3. Data fusion

The primary goal of this study was to integrate and collectively analyze data obtained from various sensing methods using machine learning models to improve the estimation of crowd movement and load characteristics. Instead of relying on data from a single sensor source, combining information from wearable and structure-based sensors provides a more comprehensive understanding of the interplay between the structure and the crowd. Successful fusion of crowd-level strain data with individual-level acceleration is expected to yield classification and regression models that are more resilient to sensor failures. If one sensor fails, the chosen machine learning fusion method can produce reasonable results based on data from the other sources. Ultimately, this fusion can help distinguish between crowd overload, environmental factors, and compromised structures.

Data fusion can occur at input, feature, and decision levels, which have all been investigated in this study. Input-level fusion involves directly combining data from different sensors as a single input for machine learning algorithms. This study explores the concepts of sensor fusion, combining different data from FBG and accelerometer sensors, as well as applying data and feature fusion to integrate data from different as well as within sensors. Decision-level fusion is performed by ensembling different machine-learning models that were trained on the following data:

1. Features extracted from a single input without any fusion.
2. Features extracted from the fusion of diverse raw sensor data.
3. Fused features extracted from diverse raw sensor data.

In our study, input-level fusion was accomplished in four different ways:

- Concatenating each pair of observations from the FBG sensors and accelerometers into a single signal vector.
- Calculating the mean of the three FBG datasets and concatenating them to accelerometer data in a single signal vector.
- Concatenating each pair of observations from the FBG sensors into a single signal vector.
- Calculating the mean of three FBG datasets into a single signal vector.

The newly formed signal vector was then used in the same manner as the individual FBG and accelerometer data for the 2D-CNN and other machine learning models.

Feature-level fusion is accomplished in two ways:

- In a CNN, feature-level fusion occurs in the convolution layer, which is designed to operate on two separate channels, one per sensor. Similar to how individual FBG and accelerometer signal data arrays are input into the CNN, fusion occurs by concatenating the data arrays from individual sensors along the channel dimensions. The features extracted from each channel are concatenated after passing to the next layer for further processing.
- Feature-level fusion for the random forest, KNN, SVM, XGBoost, and ensemble methods was achieved by concatenating the selected features extracted via VMD decomposition, further subjecting them to the principal component analysis (PCA)-based algorithm discussed in [38].

Fig. 4 displays all the fusion processes executed in this study. As previously stated, our study incorporated a three-level fusion approach at the signal, feature, and decision levels. As shown in the figure, we employed raw data fusion to explore four distinct methods based on concatenation and averaging of different sources. In the realm of feature fusion, we applied two strategies: one within the intermediate layer of a 2D-CNN, and the other by combining selected features based on PCA. Decision fusion is based on an ensemble of models. We implemented this fusion across five scenarios encompassing diverse combinations of decision-, feature-, and raw data-level integration, as illustrated in the figure.

Fig. 5 shows a flowchart of the proposed approach. As shown in the figure, the proposed flowchart begins with data collection from diverse sources including FBGs and smartphones. Subsequently, the raw data are subjected to a series of signal processing techniques, including decomposition and denoising based on VMD, as well as downsampling. After data preprocessing, three distinct levels of fusion are performed on the decomposed and denoised data. Following this procedure, the proposed optimized 2D-CNN model can perform the classification and regression tasks, with the objective of predicting load values and determining pedestrian speed.

4. Experimental setup

An experimental test series was conducted to validate the devised procedure for predicting crowd behavior and associated loads. The experimental setup involved volunteers of different group sizes, who were asked to walk across a pedestrian panel at different speeds (slow and fast). Throughout the experiment, participants carried smartphone devices equipped with accelerometers, which continually recorded acceleration data. To simulate a crowd walking across the model bridge, the measured data were wirelessly transmitted. In addition to wearable devices, the test panel was equipped with three FBG-FOS sensors. To analyze and store the data, both structural sensors and accelerometers were connected to a database server. Fig. 6 shows an overview of the experimental setup.

4.1. Experimental bridge panel

A downsized model of a bridge was designed and constructed as the experimental test structure, as illustrated in Fig. 7. The bridge model measured approximately four meters in length and one meter in width and comprised three steel C-beams. Steel plates were used to connect the C-beams laterally at intervals of 0.95 m. The bridge was supported by a pin and roller at each end. Its upper surface was covered with wooden panels to ensure an even load distribution and provide a comfortable walking surface for volunteer pedestrians.

4.2. Sensors

Small-sized optical fibers, known as FOS were used to instrument the test bridge. FOS modulates the properties of a propagating light wave, such as phase, intensity, wavelength, polarization, and frequency, offering several advantages over traditional strain sensors, owing to the silica core. These advantages include passive nature, minimal losses at optical frequencies, dielectric properties, quasi-and

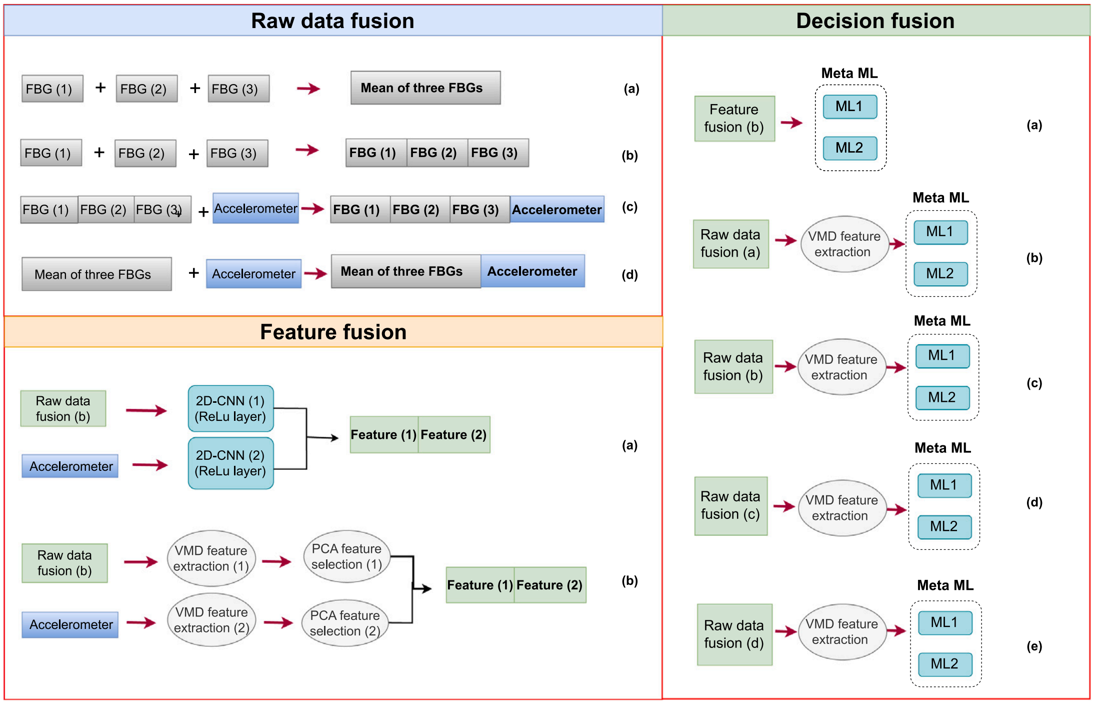

Fig. 4. Proposed fusion process at different levels.

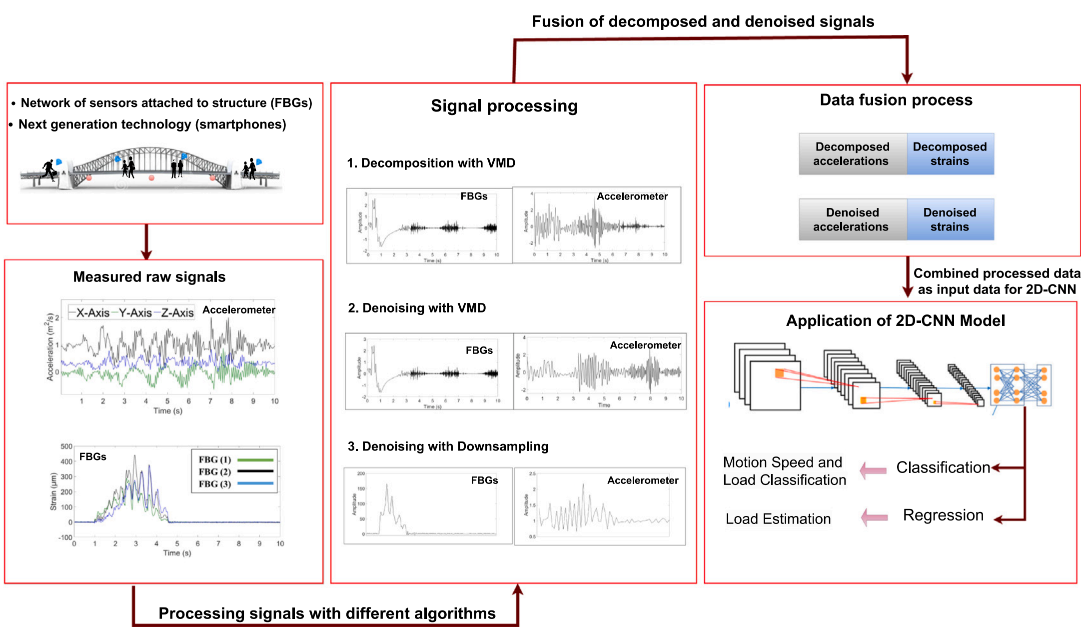

Fig. 5. Overall overview for the proposed framework.

fully distributed sensing points, immunity to electromagnetic interference, chemical inertness, biocompatibility, multiplexing capability, high-temperature resistance, and a compact lightweight structure. Despite their high cost, FOS provides multiplexing and long-distance capabilities of up to 20 km, which justifies the expense compared to conventional sensors.

By incorporating FBG sensors into FOS, FBG-based FOS was obtained, inscribing "wavelength-selective mirrors" into the fiber, to reflect only a specific wavelength from the input light wavelength spectrum. This specific wavelength is referred to as the Bragg wavelength. Changes in parameters, such as strain and temperature, cause alterations in the grating period (Λ), resulting in spectral shifts in the Bragg wavelength, whereby the peaks can be identified and translated into meaningful measurements using an interrogator (data acquisition hardware).

For the test-bridge setup, the sensors were mounted underneath the central C-beam of the bridge using epoxy-bonded brackets. The FBG sensors were procured from Micron Optics and included three Os3610 sensors (25 cm surface-mounted strain sensors). The sensors were arranged in a daisy-chain configuration to collect the structural

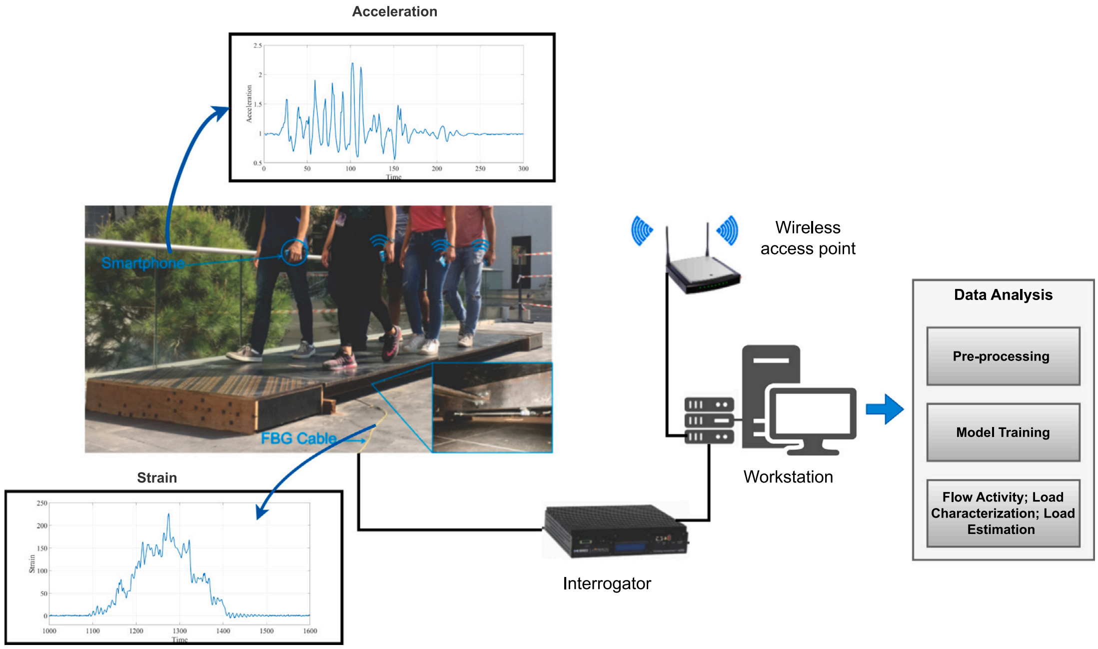

Fig. 6. Overview of the experimental framework.

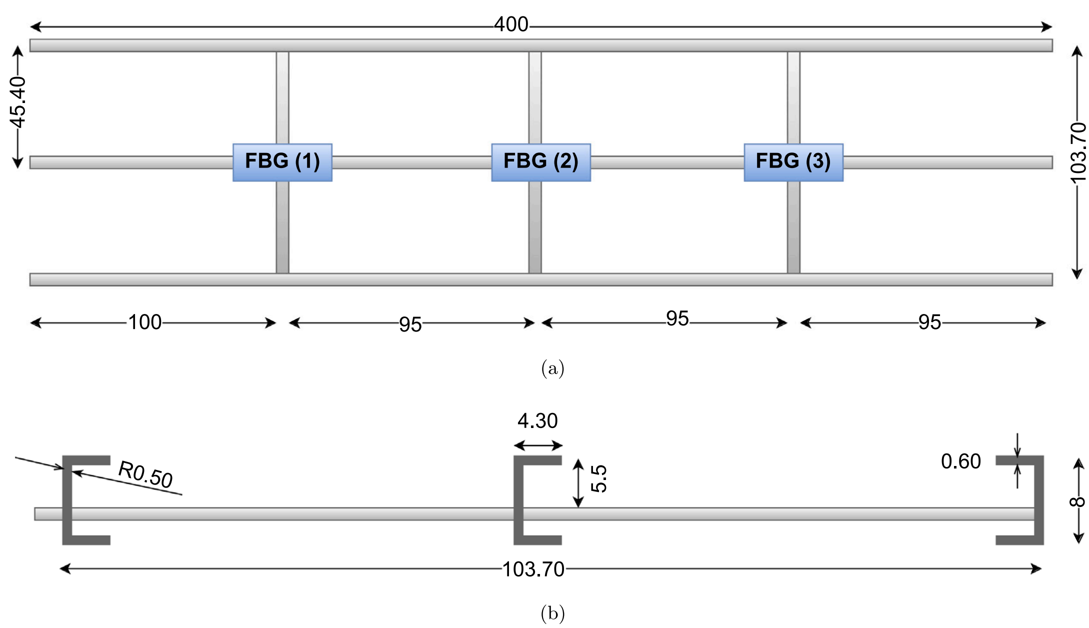

Fig. 7. Experimental test bridge (a) technical drawing (cm) – overhead view, (b) technical drawing (cm) – frontal view.

strain data, as illustrated in Fig. 6. For further information on the FBG-FOS data collection, the reader is referred to [7,8].

Body acceleration measurements were recorded along the x, y, and z axes using a handheld smartphone equipped with built-in accelerometers (Fig. 6). The smartphones were able to record the acceleration forces on all three axes simultaneously, using a customized application. These data, along with individual identification numbers, were transmitted via Wi-Fi to a connected wireless access point at a rate of 20 Hz. An SQL database was used to store and process the data locally.

4.3. Crowd data sampling

Generating a dataset suitable for training machine-learning models requires simulating crowd flows on a pedestrian bridge. Consequently, we assembled diverse groups of volunteers to walk across the model bridge under various conditions, including two distinct walking speeds: leisurely (slow) and brisk (fast), and different group sizes (as indicated in Table 4). Each volunteer was equipped with a smartphone while walking across the bridge. The strain levels of the bridge were recorded using three installed FBG sensors, and individual acceleration data were simultaneously captured using smartphones.

In the dataset, each recorded instance was classified based on factors such as number of pedestrians, number of groups, and speed of walking (slow or fast). Each volunteer group traversed the bridge twice in both forward and backward directions, resulting in a dataset comprising 488 observations. The recorded time stamps from the accelerometers were synchronized with the GPS time, whereas those from the FBG were not. The sensors were aligned using the network time protocol (NTP) [8] to ensure a high degree of temporal accuracy between the two data

Table 4
Composition of volunteers employed for crowd simulation; detailing group sizes, overall number of groups, and average mass associated with each group size.

Group size	Total groups	Mean mass (in kilograms)
1	10	78.3
2	10	154.3
3	10	225.5
4	10	291.8
5	10	392.4
6	10	444.3
8	1	605.9

streams. The minimal time difference between the two streams, which is expected to be within a few milliseconds [39], had no significant impact on the application.

Fig. 8 shows the synchronized strain and acceleration patterns for two scenarios: a group of two individuals with a combined weight of 143.9 kg moving at both fast and slow speeds, and a group of eight individuals with a total mass of 605.9 kg walking at the same speeds. As shown in the figure, the strain signal from the FBG sensors and the acceleration signal from the IMU exhibit distinct patterns influenced by the dynamics of the crowd. The footfalls of the crowd corresponded to discernible peaks in the FBG strain signal as the volunteers crossed the FBG sensors. These peak strain readings are most pronounced over the central bridge sensor "FBG (2)". As expected, a difference in strain is evident between the groups of two and eight individuals, regardless of their walking speed, based on the varying peak amplitudes registered by the FBG sensors. Higher peak amplitude values are indicative of greater total mass and generally larger group sizes.

Upon comparing fast and slow walking, a clear trend emerges: fast walking generates higher-frequency oscillations and shorter activity spans in the FBG strain readings, regardless of group size. Within the same time interval, in the fast motion case, the accelerometer readings along the x and y axes exhibit greater amplitudes than those for the slow-motion case. The outcomes from both sensors underscore the potential of employing sophisticated machine learning algorithms for the accurate classification of motion speed and load, as well as precise estimation of load values.

5. Results and discussions

Simulations were conducted on the crowd and structural movement data recorded during the experiments for the training and evaluation of the proposed crowd attribute estimation methods. The collected dataset included both strain and acceleration data, each paired with the respective labels and was used for the following purposes.

- **Binary classification:** A two-class classification model was employed to determine the speed of crowd motion, distinguishing between fast and slow movements. The dataset was divided evenly into two categories.
- **Multi-class classification:** To provide a more comprehensive characterization of the crowd, multiclass classification models were trained using a labeling system that considered crowd load based on total weight. Categories included slow-light, slow-heavy, fast-light, and fast-heavy, with a 250 kg threshold used to roughly balance the distribution of data labels between light and heavy loads.
- **Regression:** Regression models were utilized to estimate the total crowd mass on the bridge. These models were trained using the total mass data collected during the data acquisition phase.

In our approach, we utilized three different types of data: raw, decomposed, and denoised, with each serving a specific purpose in enhancing our analysis (see Fig. 9).

- **Decomposed data:** To acquire valuable insights from the original signals, we employed the VMD technique to decompose the signals into three distinct components. This decomposition created a new input dataset with richer information than the original signals. Fig. 10 provides a visual representation of this process, displaying the raw strain and acceleration signals alongside the three decomposed components. The figure shows how the original signals were broken down into distinct components.
- **Denoised data:** To improve the quality of the signals, we applied denoising techniques using VMD in combination with downsampling. Fig. 10 illustrates the outcome of this denoising process, in which the initially decomposed signals were subjected to denoising, resulting in three denoised signal components. These denoised components were then combined to create the final denoised signals. Fig. 11 illustrates the denoising process coupled with various types of downsampling, namely, downsampling I, II, and III. This comparison allowed us to study the impact of downsampling on our data. We found that using a high downsampling factor can lead to loss of information, resulting in reduced accuracy. However, when downsampling was applied judiciously, it effectively removed noise from the input signals, thereby enhancing the accuracy of the results.

As elaborated above and depicted in Fig. 4, six different data fusion approaches were investigated using the following strategies:

- Calculating the mean of the three FBGs.
- Concatenating the data from three FBGs.
- Averaging the data from three FBGs while incorporating acceleration.
- Concatenating the data from three FBGs while incorporating acceleration.
- Implementing feature fusion within the middle layer of a 2D-CNN model.
- Employing feature fusion based on the concatenation of selected features using PCA.
- Employing decision fusion based on the concept of an ensemble model.

5.1. Overview of data analysis

In this study, the analyzed dataset consisted of FBG and accelerometer data paired with the respective labels. For strain data, each observation was evenly distributed among the three FBG sensors. To ensure temporal alignment, the accelerometer data were time-synchronized based on the timestamps of the FBG strain data. The measurements from the three FBG sensors were concatenated or averaged into a single vector to create a sequential strain curve representing crowd flow. To account for the variations in the way volunteers held their smartphones, the accelerometer data needed to be orientation insensitive. This was achieved by summing the norms of the tri-axis smartphone accelerometer for each group size.

For data mining, 2D-CNN models were selected because of their ability to automatically extract features. In this context, the Python libraries Scikit-learn and TensorFlow were employed to formulate the classification and regression problems. The raw FBG and accelerometer signal data served as one-dimensional image inputs for training the 2D-CNN models. The VMD algorithm and downsampling models were applied to the raw sensor data to generate additional inputs for the 2D-CNN models.

To ensure unbiased evaluation, the dataset was subjected to a fivefold cross-validation procedure. During this process, the dataset was randomly divided into training, validation, and test subsets. The training subset was employed for model training, the validation subset for optimizing and fine-tuning of parameters, and the testing subset for evaluating model performance and accuracy. The classification model

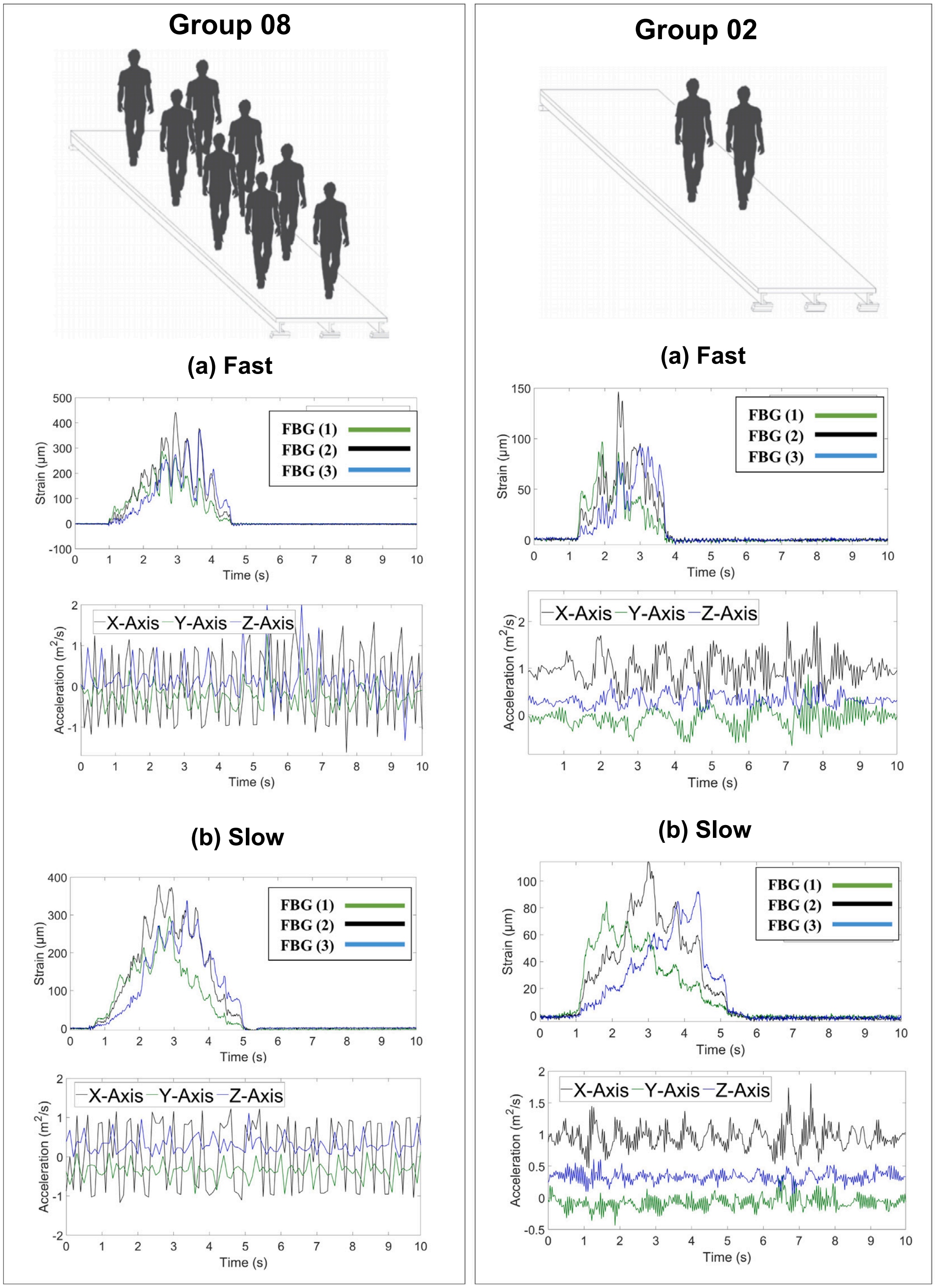

Fig. 8. Examples of strain and acceleration recordings observed in a group of eight and two individuals during (a) high-speed and (b) low-speed motion.

evaluation was based on the percentages of accuracy, recall, precision, and F1 score, whereas the regression model evaluation was determined by the average percentage error between the estimated and actual loads and R-squared (R^2) score. The final assessment of model performance was based on the mean and standard deviation of the resulting scores for each fold. Higher values of recall, F1 score, precision, accuracy, and R^2 score indicate better performance.

To investigate the impact of data fusion, we first employed the dataset from each sensor in isolation to develop individual models. This approach enabled us to assess the independent capabilities of the FBG and accelerometer sensors in classifying and regressing crowd attributes, encompassing parameters such as motion speed, load categorization, and load estimation. Subsequently, the fused signals were analyzed at the signal and feature levels in various states (raw, decomposed, and denoised data).

The following nomenclature represents the different forms of data, including the fusion techniques used to generate the results presented in this section:

- **Acc:** acceleration data.
- **CoS:** concatenation of strains.

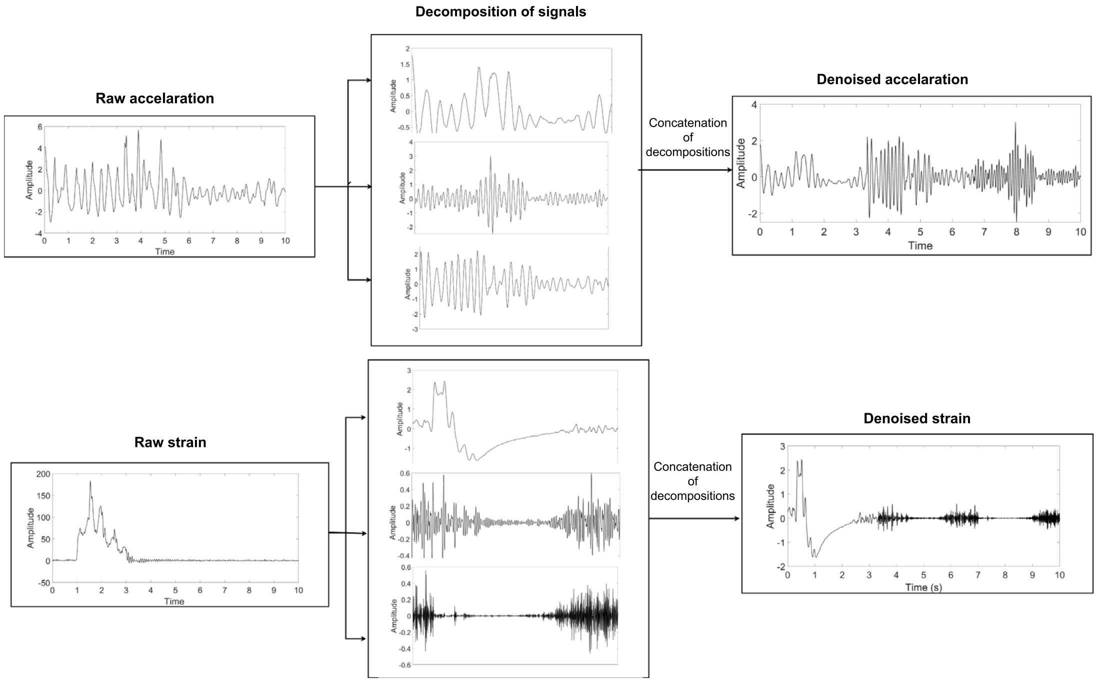

Fig. 9. Decomposition of strain and acceleration signals using the VMD algorithm.

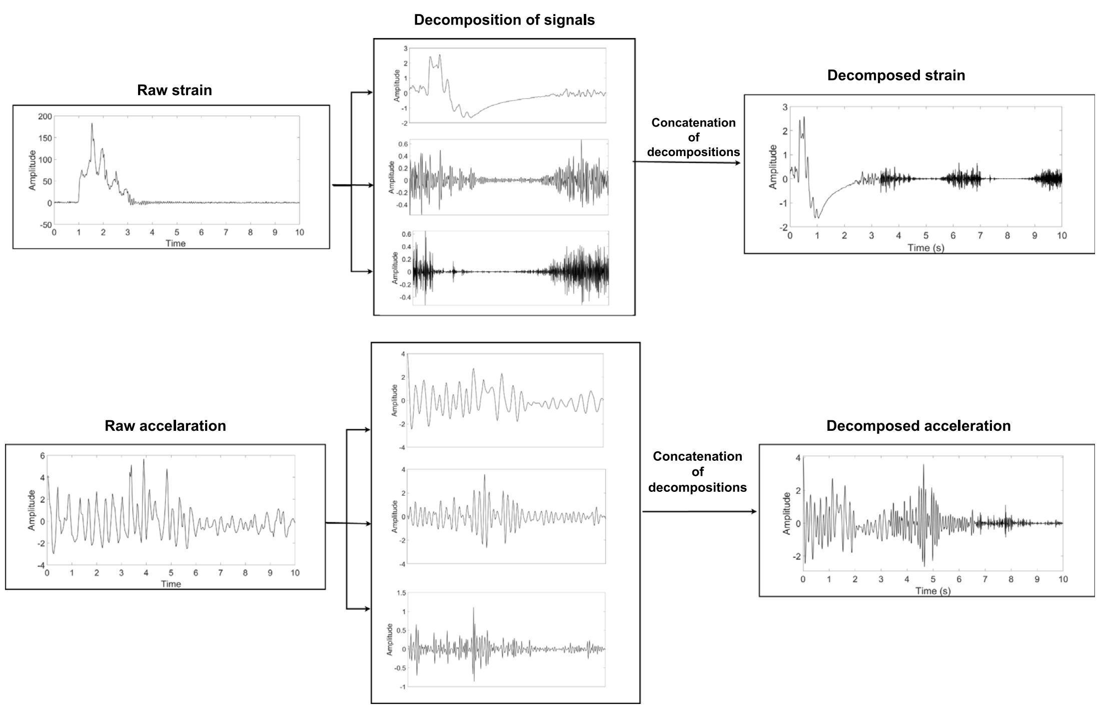

Fig. 10. Denoising of strain and acceleration signals using the VMD algorithm.

- **MoS:** mean of strains.
- **DF-I:** concatenation of strain with acceleration data.
- **DF-II:** mean of strain with acceleration data.
- **FF:** feature fusion.

5.2. Binary classification results

This section discusses the results of binary classification conducted using different types of data combinations and denoising strategies. Fig. 12 shows the validation scores of the binary-class motion speed classification task using both raw and decomposed data. The hyper-parameters for the 2D-CNN models were optimized based on the validation scores. For each data combination, the figure shows the mean results for all the folds. For the raw data, the validation accuracy across all inputs is approximately 96%, except for the accelerometer data which have an accuracy of 94%. The F1, precision, and recall scores for the raw data are also impressive, averaging approximately 93%, 92.5%, and 94.5%, respectively.

Notably, data fusion at each level yielded higher accuracy, recall, precision, and F1 scores than using acceleration or strain data alone. For instance, accuracy scores for DF-II and DF-I outperform those

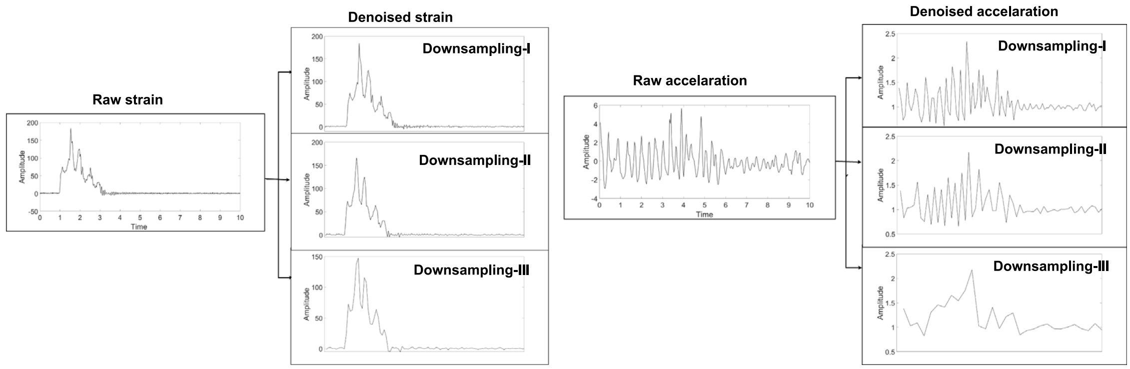

Fig. 11. Denoising of strain and acceleration signals through downsampling.

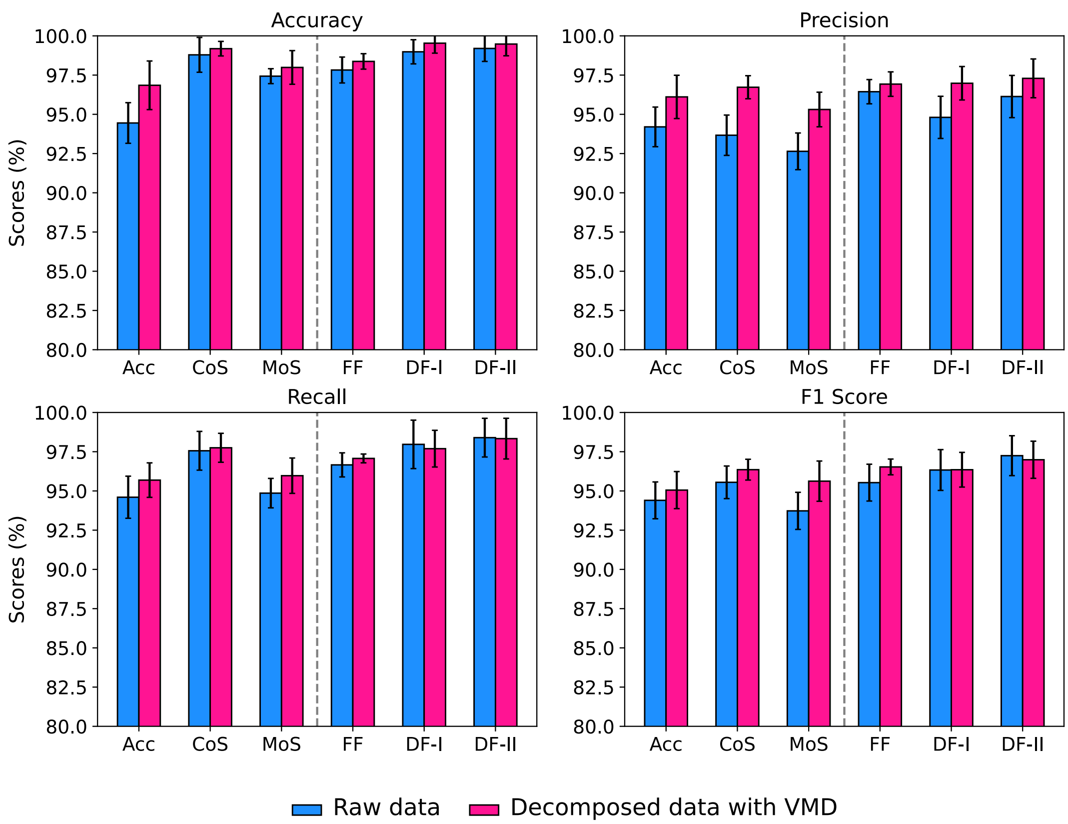

Fig. 12. Validation scores (%) for binary classification using both raw and decomposed data.

of other outcomes. The results for CoS, FF, and MoS followed, with the lowest score attributed to Acc. Data fusion at the signal level outperformed feature fusion. The figure also presents the results of the decomposed data. Here, we observe that the scores in all states of the data are higher than those of the raw data. The accuracy score for decomposed data fusion (DF-II) reached 98.9%, whereas that for FF reached 98.1%. Notably, the performance of DF-II surpasses that of DF-I for both raw and decomposed data. Decomposition consistently led to higher scores in all aspects of validation, including accuracy, precision, recall, and F1 scores.

Fig. 13 shows the results of the raw and denoised signals across all combinations of input data using the VMD algorithm. The denoised results outperformed both the decomposed and raw data. For example, the accuracy score for DF-II reached an impressive 99.80%, significantly surpassing that of the raw data. The precision scores are also notably higher, exceeding 98.2%. Denoised signals for all data combinations consistently achieved validation scores above 95%. The highest scores in accuracy, precision, recall, and F1 score were achieved by DF-II followed by DF-I.

Fig. 14 further explores the outcomes of denoising, considering three types of downsampling and various data combinations. Downsampling-I consistently outperformed raw data signals, with an accuracy score of 98% for DF-II. Downsampling-II exhibited mixed results, with some scores surpassing those of the raw data and others falling short. However, downsampling-III, yielded lower scores than raw data as some useful information was lost. Overall, the results suggest that for our dataset, downsampling-I and downsampling-II demonstrate better performance in noise removal, aligning with our goal of denoising input signals. The highest scores in accuracy, precision, recall, and F1 score were achieved by DF-I followed by DF-II. Regarding the standard deviation of the results, downsampling-II introduced consistent fluctuations across different folds, leading to unstable accuracy results. However, slight or moderate downsampling can be beneficial, and other metrics show less severe fluctuations in the downsampling-II results.

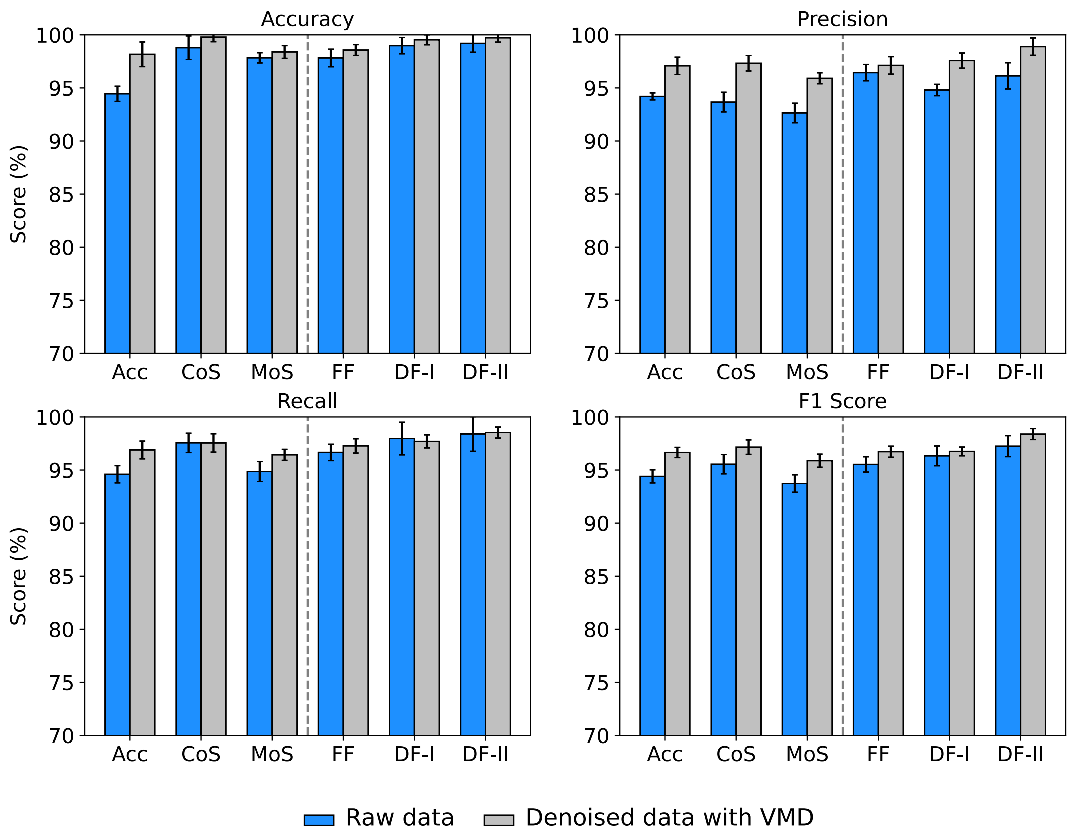

Fig. 13. Validation scores (%) for binary classification using both raw and denoised data with the VMD algorithm.

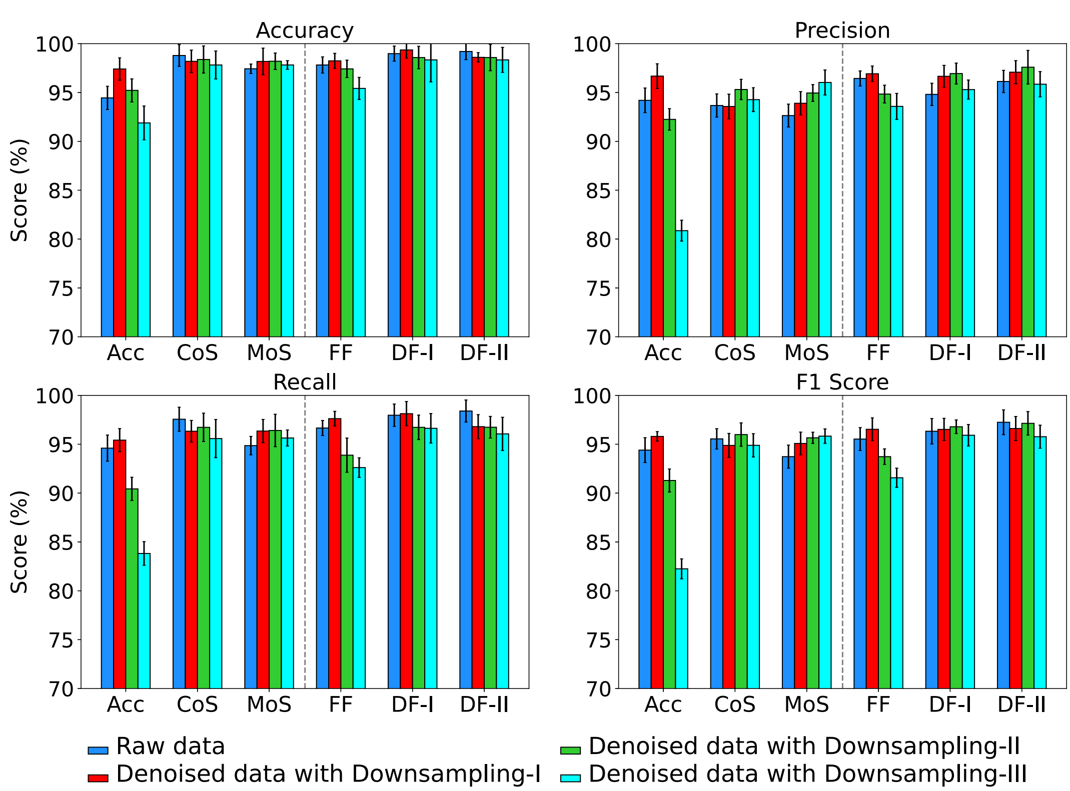

Fig. 14. Validation scores (%) for binary classification using both raw and denoised data from different types of downsampling.

Table 5
Test scores for different combinations of raw, decomposed, and denoised data for the binary-class classification task.

Type	Raw	Decomposed with VMD	Denoising with VMD	Downsampling-I	Downsampling-II	Downsampling-III
Test accuracy (%)						
Acc	94.45	96.85	98.17	97.41	95.22	91.89
CoS	98.19	99.28	99.39	98.50	98.20	97.82
MoS	97.43	97.99	98.39	98.18	98.20	97.12
FF	97.82	98.37	98.87	98.25	97.42	95.42
DF-I	98.98	99.33	99.53	99.36	98.88	98.34
DF-II	99.20	99.48	99.62	98.59	98.57	97.02
Test precision (%)						
Acc	94.20	96.11	97.08	96.68	92.25	80.86
CoS	93.67	96.73	97.32	95.57	94.31	93.07
MoS	92.64	95.31	95.91	94.90	93.94	92.03
FF	96.44	96.93	97.12	96.92	94.84	93.58
DF-I	94.81	96.98	97.58	96.66	96.14	95.30
DF-II	96.13	97.29	98.89	97.08	96.59	95.85
Test recall (%)						
Acc	94.60	95.69	96.89	95.41	90.43	83.82
CoS	97.56	97.75	97.95	96.33	96.10	95.58
MoS	94.86	96.97	96.83	96.45	96.41	95.64
FF	96.66	97.07	97.97	97.12	93.88	92.61
DF-I	97.17	97.69	97.83	97.13	96.73	96.63
DF-II	98.40	98.43	98.63	96.80	96.74	96.06
Test F1 (%)						
Acc	94.40	95.05	96.65	95.80	91.29	82.25
CoS	95.55	96.36	97.16	95.99	94.99	94.89
MoS	93.73	95.62	95.88	94.93	94.66	94
FF	95.53	96.11	96.73	96.02	93.67	91.43
DF-I	96.33	96.68	96.85	96.42	96.19	95.22
DF-II	97.25	98.28	98.99	96.61	96.13	95.77

As detailed in Section 3.1.2, we employed five-fold cross-validation to identify the optimal hyperparameters for the 2D-CNN model. The next step was to evaluate model performance on a separate test set comprising 20% of the data which were excluded from the training process. Table 5 provides an overview of the test scores for various combinations of raw, decomposed, and denoised data using VMD and downsampling. As shown in the table, the highest score (indicated by the blue number) was achieved with denoised data using VMD for all data combinations. The next highest scores were associated with decomposed data and denoising using downsampling-I. For instance, the highest accuracy scores are obtained by denoising the data using VMD for the states DF-II, DF-I, and FF, respectively. As previously mentioned, this study examined various data fusion states as inputs, and the results in the table demonstrate that data fusion also enhances the test scores. Notably, DF-II achieved the highest scores on all tests (accuracy = 99.62%, precision = 98.89%, F1 = 98.63%, and recall = 98.99%).

5.3. Multiclass classification results

This section presents the results of the multi-classification conducted on different types of data combinations and denoising strategies. Fig. 15 shows the validation scores of multiclass motion speed and load classification (fast-heavy, fast-light, slow-heavy, and slow-light) using both raw and decomposed data. For raw data, the validation accuracy across all inputs was consistently approximately 97%, except for accelerometer data, which achieved 94% accuracy. The F1, precision, and recall scores for the raw data were also approximately 92%, 93%, and 94%, respectively. Notably, the fusion of raw data at each level yielded higher accuracy, recall, precision, and F1 scores than using acceleration or strain data alone. The validation scores for the decomposed data surpassed those for the raw data. Furthermore, the performance of the fused data exceeded that of the individual data sources. For instance, when comparing the accuracy scores, FF (98.25%), DF-I (98.02%), and DF-II (97.90%) outperformed the other data combinations. Regarding the standard deviation of the results across different folds, a smaller deviation is apparent in the decomposition results, confirming the effectiveness of the decomposition strategy. However, in a handful of instances decomposition was not as effective, as in the case of F1 scores associated with MoS and recall obtained for CoS.

Fig. 16 illustrates the results of the input data denoised using the VMD algorithm. Notably, the denoising process of VMD was conducted concurrently along with decomposition. The results indicate that the VMD algorithm is significantly more effective when used for denoising. For example, when performing VMD denoising, the accuracy scores for DF-I and FF reached 98.5% and 99%, respectively, significantly surpassing those of the raw data. The denoised signals for all combinations of data states consistently achieved accuracy, F1, precision, and recall scores above 99%, 93.34%, 94.28%, and 94.67%, respectively.

Fig. 17 further explores the outcomes of various data combinations denoised via three types of downsampling. Downsampling-I consistently outperformed raw data, with an accuracy score of 98.3% for DF-II. Downsampling-II exhibited mixed results, with some scores surpassing raw data and others falling short. Downsampling-III, however, yielded lower scores than raw data, as some useful information was lost. Overall, the results suggest that on our dataset, Downsampling-I and Downsampling-II demonstrated better performance in removing noise, aligning with our goal of denoising input signals. In the multiclass context, the outcomes of Downsampling-III were less favorable than those of Downsampling-I in the context of binary classification. The underlying explanation for this discrepancy lies in the nature of multiclass classification, which required handling four distinct classes. For a comprehensive analysis, each class requires a substantial amount of data, such that a high downsampling factor leads to data loss, thereby reducing the available data for each class. In contrast, in binary classification, which involves only two classes, the omission of some data does not significantly impact the analysis since decision-making is a simpler process with fewer classes to consider.

Table 6 presents an overview of the testing scores for various combinations of raw and decomposed data for multiclass classification. As depicted in the table, the top score (indicated by the blue number) corresponds to denoised data obtained using VMD for all data

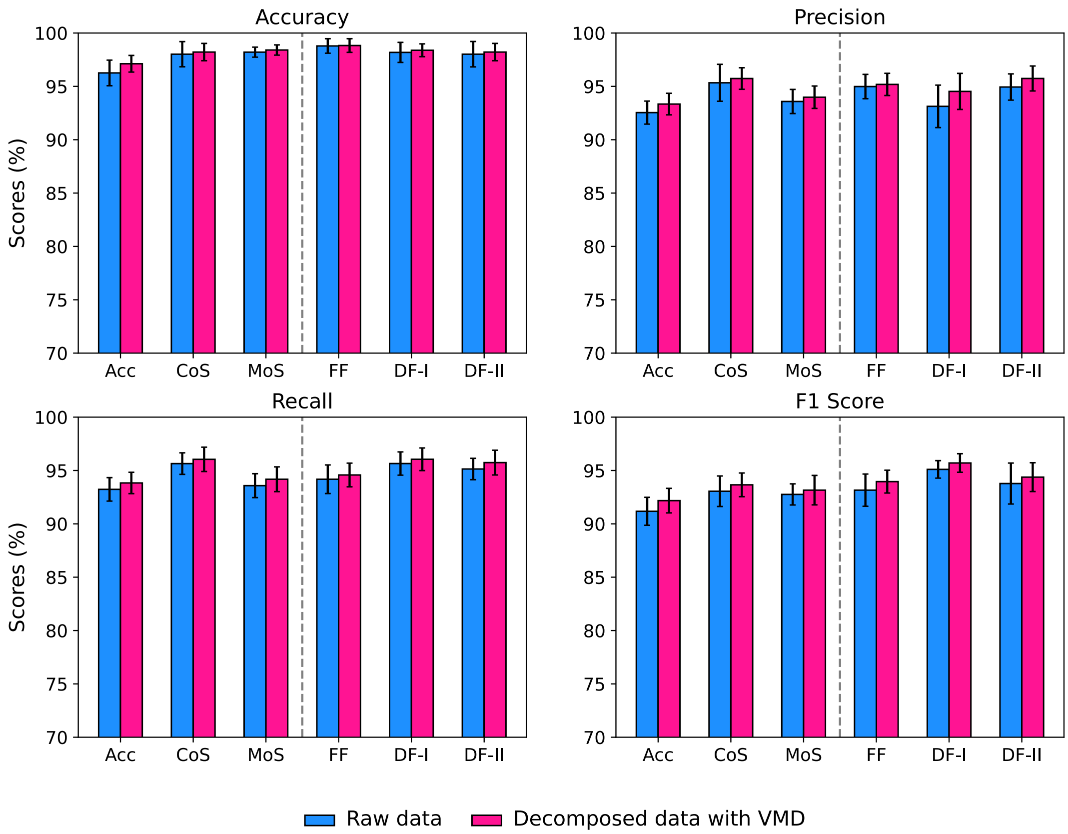

Fig. 15. Validation scores (%) for multiclass classification using both raw and decomposed data.

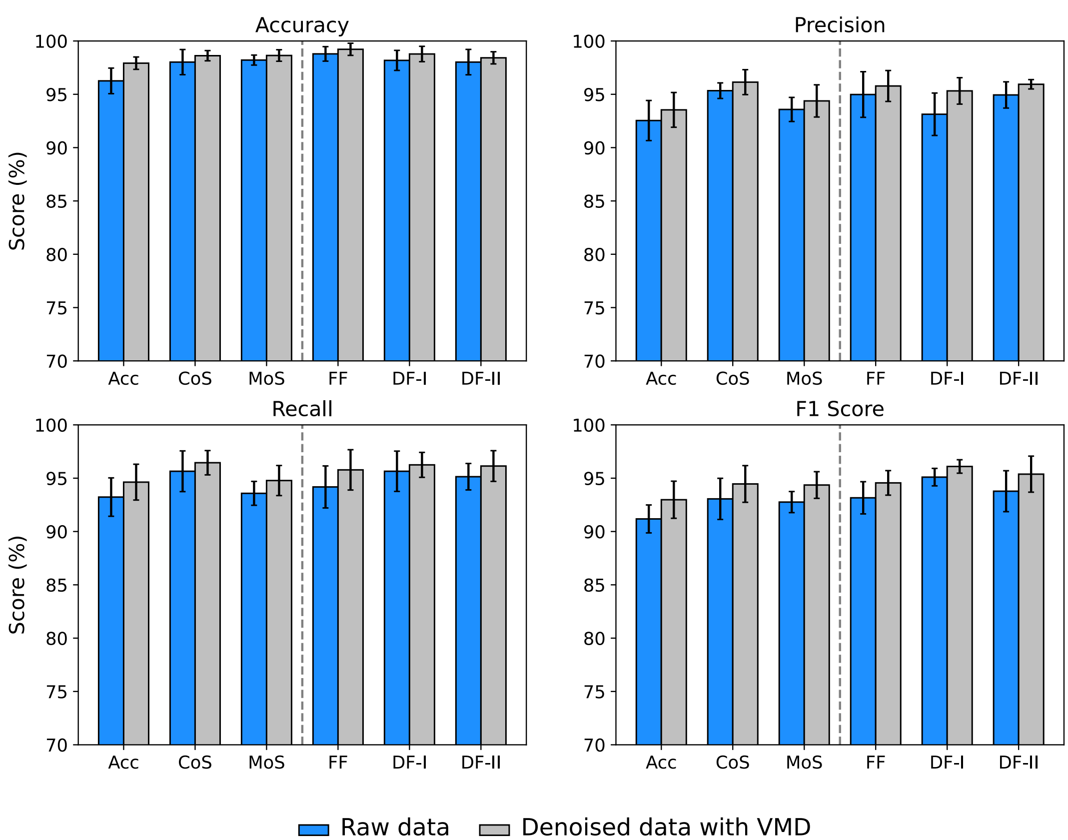

Fig. 16. Validation scores (%) for multiclass classification using both raw and denoised data with the VMD algorithm.

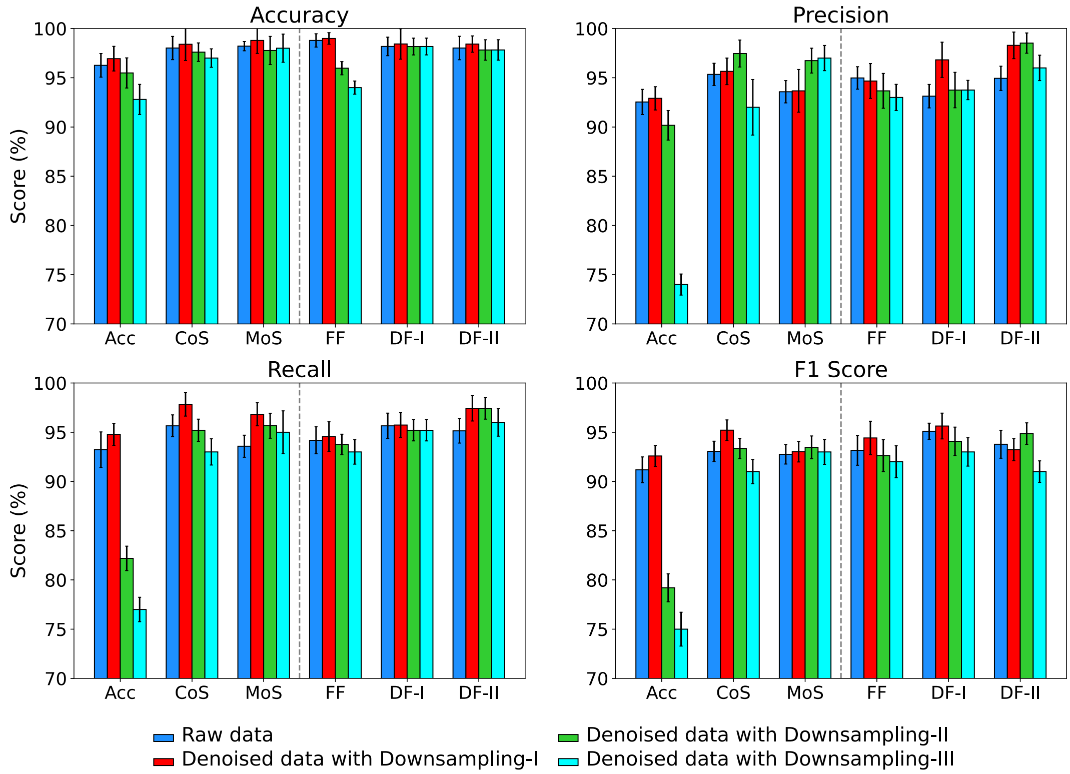

Fig. 17. Validation scores (%) for multiclass classification using both raw and denoised data with downsampling.

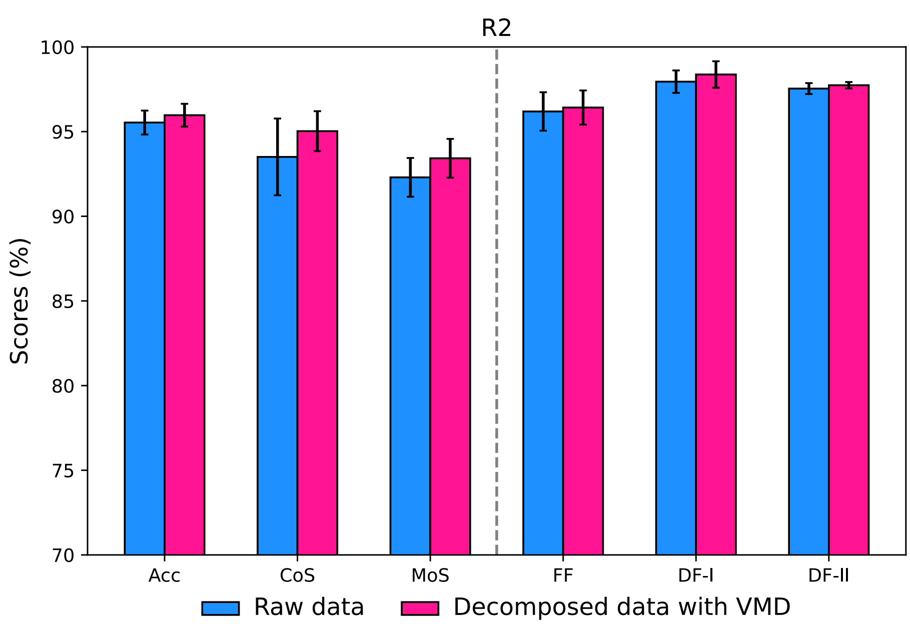

Fig. 18. Validation R^2 (%) for the regression task using both raw and decomposed data.

combinations. The highest scores are associated with the decomposed data and denoised data using Downsampling-I. For instance, the highest accuracy scores are achieved by denoising the data using VMD for DF-I, DF-II, and FF, respectively. The results in the table demonstrate that data fusion enhances the test scores for any type of data (raw, decomposed, and denoised). Notably, DF-I achieves the highest scores on all tests (accuracy = 98.69%, precision = 98.46%, F1 = 97.66%, and recall = 96.10%).

5.4. Regression results

This section discusses the regression analysis results. Fig. 18 shows the validation R^2 of the regression task for load prediction on the bridge using both raw and decomposed data. Regarding raw data, the validation R^2 across all inputs is consistently approximately 93%. As depicted in the figure, the R^2 scores for the decomposed data outperform those for the raw data, and data fusion also exhibits superior performance

Table 6
Test scores for different combinations of raw, decomposed, and denoised data for the multiclass classification task.

Type	Raw	Decomposed with VMD	Denoising with VMD	Downsampling-I	Downsampling-II	Downsampling-III
Test accuracy (%)						
Acc	96.26	97.12	97.92	96.45	95.49	92.79
CoS	98.03	98.24	98.61	98.19	97.61	96.02
MoS	98.21	98.42	98.65	97.80	97.79	96.02
FF	97.79	97.83	98.33	96.13	96.00	94.05
DF-I	98.19	98.39	98.69	97.02	96.18	95.18
DF-II	98.03	98.22	98.42	98.21	95.80	96.20
Test precision (%)						
Acc	92.55	93.38	93.79	92.21	80.17	74.01
CoS	95.39	95.79	96.14	94.65	93.46	92.02
MoS	93.59	93.99	94.39	93.67	91.76	90.05
FF	94.96	95.18	95.79	93.75	93.67	93.08
DF-I	93.14	94.53	98.46	96.82	93.72	93.10
DF-II	96.94	96.74	97.94	97.58	95.29	95.07
Test recall (%)						
Acc	93.24	93.83	94.64	90.79	82.19	77.00
CoS	95.66	96.05	97.45	96.14	95.21	93.01
MoS	93.58	96.18	96.78	96.33	95.66	95.02
FF	94.18	94.58	95.78	93.56	93.17	93.03
DF-I	95.67	96.06	97.66	95.74	95.20	95.23
DF-II	95.14	96.74	96.44	97.13	97.11	96.00
Test F1 (%)						
Acc	91.18	92.18	92.97	89.60	79.20	75.00
CoS	93.06	93.66	94.46	93.21	92.36	91.00
MoS	92.76	93.16	94.36	93.22	93.01	93.00
FF	93.16	93.96	94.56	92.66	92.42	92.00
DF-I	95.10	95.70	96.10	94.63	93.09	92.10
DF-II	93.78	94.38	95.38	92.22	91.86	91.00

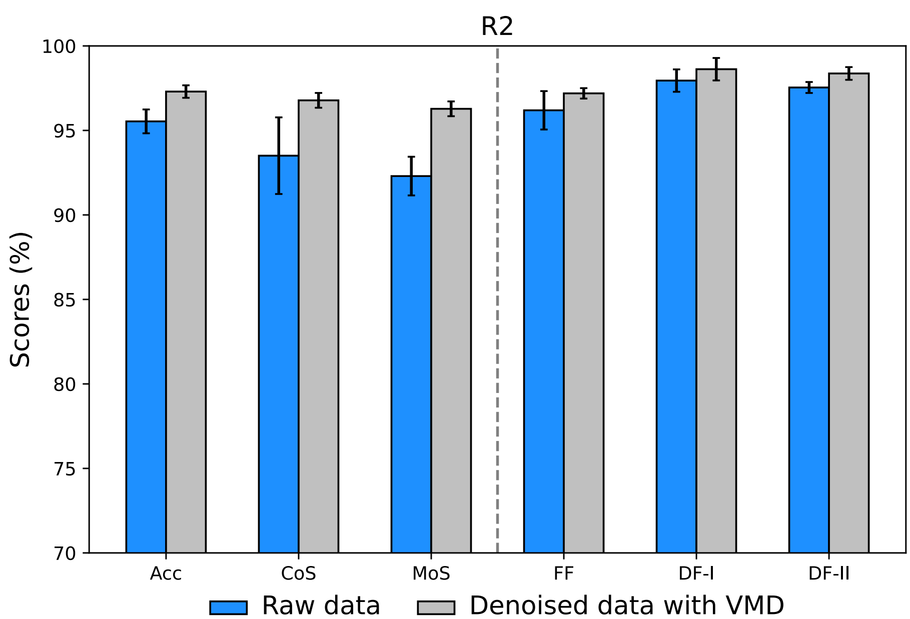

Fig. 19. Validation R^2 (%) for the regression task using both raw and denoised data with the VMD algorithm.

compared to individual data sources. For instance, the highest R^2 score of 98% is achieved for DF-I with the decomposed data as input.

Fig. 19 shows the denoising results across all combinations of input data using the VMD algorithm. The denoised results outperform both the decomposed and raw data. For example, the R^2 score for DF-I reached 98% with the denoised data, significantly surpassing the raw data state. Denoised signals for all combinations of data consistently achieved validation R^2 scores greater than 96.5%.

Fig. 20 shows the results of denoising for three types of downsampling and various data combinations. Downsampling-I consistently outperformed raw data signals, with a R^2 score of 97.9% for DF-I.

Table 7 provides an overview of the test R^2 for various combinations of raw and decomposed data for the regression problem of load prediction. As depicted in the table, the top score (indicated by the blue number) corresponds to the denoised data obtained using VMD for all data combinations. The highest R^2 scores are associated with the decomposed data and denoising using Downsampling I. For instance, for the data denoised using VMD, the R^2 scores are: DF-I (98.42%), DF-II (98.17%), and FF (97.00%). The results in the table demonstrate that data fusion enhances the test scores at any level (signal and feature) for any type of data (raw, decomposed, or denoised data).

Notably, the results in the table clearly indicate the superior performance of DF-II compared to other forms of data combinations. This

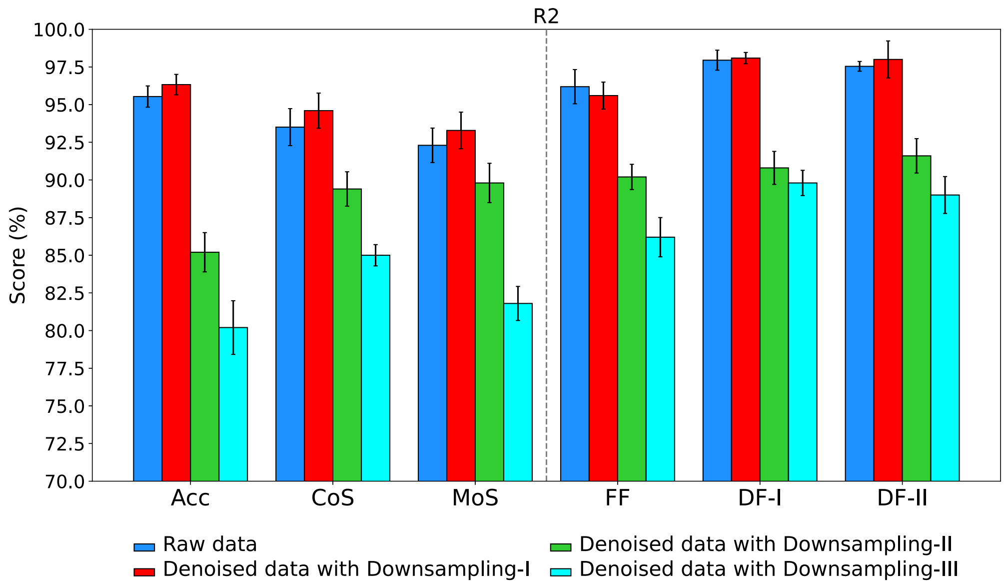

Fig. 20. Validation R^2 (%) for the regression task using both raw and denoised data using downsampling.

Table 7
Test R^2 score (%) for different combinations of raw, decomposed, and denoised data for the regression task.

Type	Raw	Decomposed with VMD	Denoising with VMD	Downsampling-I	Downsampling-II	Downsampling-III
Acc	95.57	95.97	97.30	95.78	85.20	80.20
CoS	93.51	95.03	96.78	94.60	89.40	85.00
MoS	92.30	93.43	96.28	93.29	89.80	81.80
FF	96.20	96.42	97.00	95.60	90.20	86.20
DF-I	97.54	97.74	98.17	97.00	91.60	89.00
DF-II	97.95	98.37	98.42	98.09	90.80	89.80

comparison is illustrated visually in a Taylor diagram in Fig. 21. The Taylor diagram is a specialized tool designed to evaluate datasets or models considering parameters such as the correlation coefficient, root mean square error (RMSE), and standard deviation. In this diagram, we analyze four models, denoted as M1, M2, M3, and M4, representing DF-II with raw data, DF-II with Downsampling-I, DF-II with decomposed data, and DF-II denoised using VMD, respectively.

Overall, the results indicate that M4 is the closest to the ground truth. The figure displays the correlation coefficient as a vital metric for comparison. Notably, M4 exhibits the highest correlation coefficient, indicating a stronger alignment with the ground truth data. The distance of each model point from the origin on the Taylor diagram is the root mean square difference (RMSD), with shorter distances indicating better agreement with the reference. As observed in the figure, M4 has the shortest distance from the reference, reinforcing its accuracy. In addition to RMSD, we considered the standard deviations of the models. Models with lower standard deviations are generally preferable when they are in the proximity of the reference. In this context, M4 exhibits lower standard deviations than the other models, further emphasizing its superiority. In summary, the results depicted in the Taylor diagram validate that M4, which represents DF-II for data denoised using VMD, outperforms the other models, showing higher correlation coefficients, lower RMSD, and reduced standard deviation, all collectively affirming its superior accuracy and alignment with ground truth data.

Fig. 22 shows the load predictions achieved using the denoised data with the DF-II model compared with the perfect regression line. The aim was to highlight the outstanding accuracy achieved with the proposed model using specially preprocessed input signals. The "perfect regression line" represents the ideal results, with every prediction aligning perfectly with the actual load values. This hypothetical line served as a benchmark for accuracy in our analysis. Upon close examination of the figure, the regression results of the model based on DF-II with denoised data consistently produced load predictions that closely approximated this ideal "perfect regression line". That is, the predictions of our model display an outstanding R^2, mirroring the actual load values with minor deviations.

6. Comparison with existing approaches

In this section, we compare the proposed approach with existing methodologies for condition monitoring. The goal is to evaluate and emphasize the advantages of our method across various metrics. First, our innovative data fusion techniques are compared with the established AKF method, as detailed by Liu et al. [40] and Pellegrini et al. [41]. This comparison showcases the effectiveness of our data fusion approach relative to the traditional AKF methodology. Next, the performance of our finely tuned 2D-CNN model is evaluated against conventional machine learning techniques employed by Mustapha et al. [7,8]. These techniques include random forest, KNN, SVM, XGBoost, and ensemble methods. The objective was to demonstrate the superior performance of the optimized 2D-CNN model compared to traditional methods within the framework of the proposed approach.

6.1. Comparative assessment of existing data fusion models

In scenarios in which noise exhibits non-Gaussian or non-stationary characteristics, AKF typically outperforms simple signal averaging. The adaptive nature of the Kalman filter dynamic parameter adjustment effectively filters out noise while preserving crucial signal components. Consequently, we conducted a performance comparison between our proposed data-fusion technique and the AKF-based signal-fusion approach. Recent studies, such as those by Liu et al. [40] and Pellegrini et al. [41], highlight the use of AKF for sensor signal fusion, underscoring their relevance in contemporary research. The AKF is an advanced variant of the Kalman filter, with the variance parameter based on the discrepancy between the predicted and measured values. This adaptive

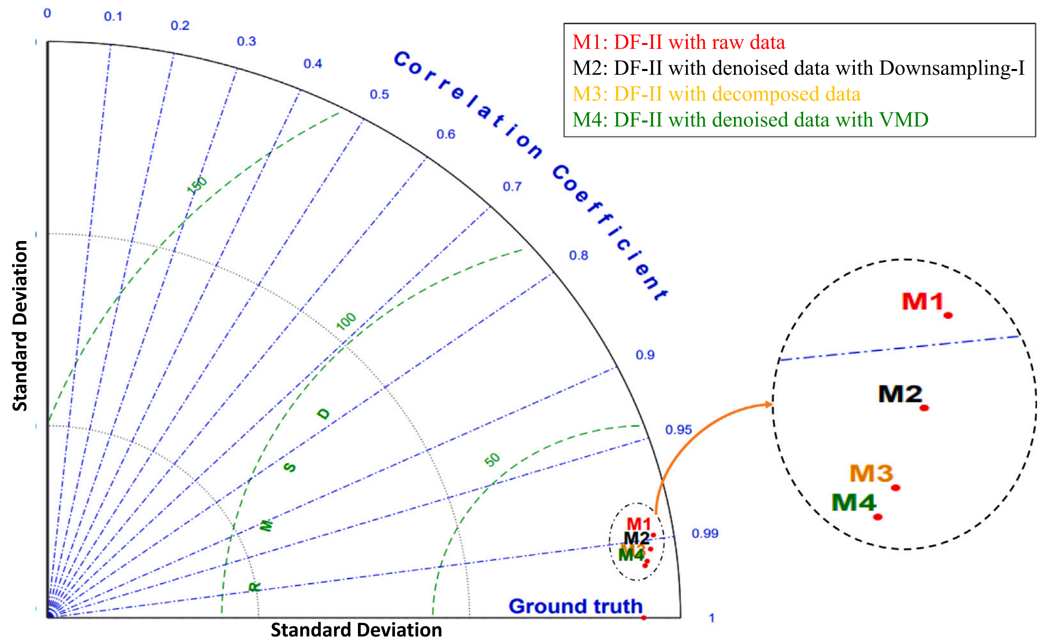

Fig. 21. Taylor diagram for DF-II with different types of data.

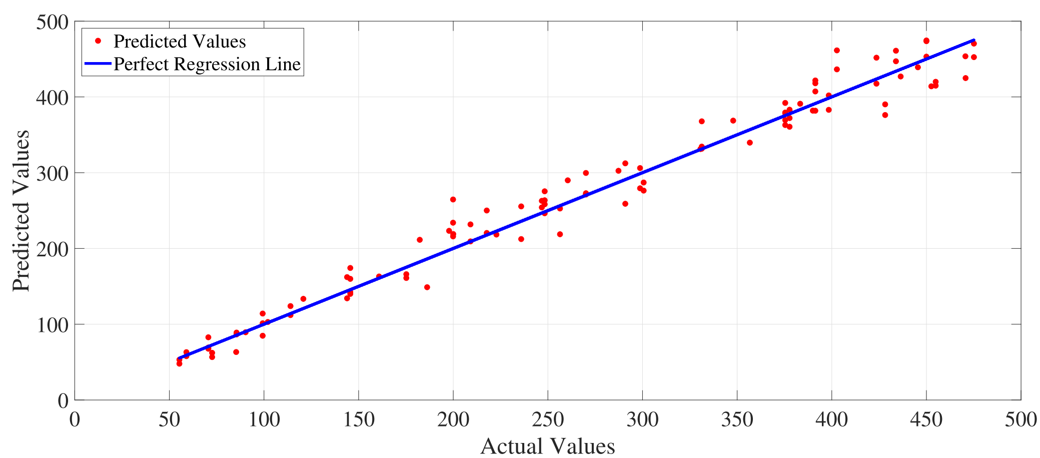

Fig. 22. Comparison of the predicted load value for DF-II with denoised data vs. the perfect regression line.

mechanism helps manage the uncertainties associated with dynamic system changes.

The AKF consists of two main stages:

1. **Prediction Update:** In this phase, the a priori estimate and error covariance are updated according to the process variance.
2. **Measurement Update:** The blending factor is calculated using the prior error covariance and measurement variance to determine the contribution of the measurement to the final estimate. The updated estimate and error covariance are adjusted based on the measurement and blending factor.

This adaptive filtering approach fine-tunes the process variance by accounting for the deviation between the measurement and updated estimate, to adjust to evolving system dynamics. We developed a Python program to implement AKF to fuse multiple signals into a unified estimate. The program iterates over subsets of input data while optimizing the hyperparameters of the AKF algorithm to enhance the accuracy of the fusion process.

The conventional Kalman filter can be expressed through the following equations [42]:

$$\hat{x}^k_{k|k-1} = U_k^{k-1} \hat{x}^{k-1} \tag{3}$$

$$P^k_{k|k-1} = U_k^{k-1} P^{k-1} (U_k^{k-1})^T + C_k Q_k C_k^T \tag{4}$$

$$K_k = P^k_{k|k-1} H_k^T (H_k P^k_{k|k-1} H_k^T + R_k)^{-1} \tag{5}$$

$$\hat{x}^k = \hat{x}^k_{k|k-1} + K_k m_k \tag{6}$$

$$P^k = (I - K_k H_k) P^k_{k|k-1} \tag{7}$$

Here, m_k represents the innovation (or measurement residual), given by:

$$m_k = z_k - H_k U_k^{k-1} \hat{x}^{k-1}$$

In these equations, various variables are involved:

- $\hat{x}^{k-1}$: State estimate at time step $k-1$.
- P^{k-1}: Covariance matrix of estimation errors at time step $k-1$.
- U_k^{k-1}: State transition matrix from time step $k-1$ to k.
- C_k: Process noise matrix at time step k.
- Q_k: Process noise covariance matrix at time step k.

- H_k: Measurement matrix at time step k.
- R_k: Measurement noise covariance matrix at time step k.
- K_k: Kalman gain matrix at time step k.
- $\hat{x}^k_{k|k-1}$: Predicted state estimate at time step k, considering measurements up to time step $k-1$.
- $P^k_{k|k-1}$: Predicted covariance matrix of estimation errors at time step k, considering measurements up to time step $k-1$.
- $\hat{x}^k$: Updated state estimate at time step k, after incorporating the measurement at time step k.
- P^k: Updated covariance matrix of estimation errors at time step k, after incorporating the measurement at time step k.

In addition, a common AKF [42] based on current measurements is often employed in the Russian aviation industry. This filter adapts the estimate of the state noise covariance matrix Q^k using the Maximum Likelihood (ML) criterion:

$$C_k Q^k C_k^T = K_k m_k m_k^T K_k^T + P_k - U_k^{k-1} P_{k-1} U_k^{k-1T} \tag{8}$$

In engineering calculations, to ensure the positive semi-definiteness of Q and simplify computations, the terms $P_k - U_k^{k-1} P_{k-1} U_k^{k-1T}$ in Eq. (8) might be neglected, yielding the approximation:

$$C_k Q^k C_k^T = K_k m_k m_k^T K_k^T \tag{9}$$

Consequently, in implementing the algorithm, the term $C_k Q_k C_k^T$ in the second equation of the previous step can be replaced by an approximate estimate obtained in the preceding step. This adaptation makes the algorithm suitable for implementation on onboard computers, owing to its swift response and minimal computational requirements.

Table 8 presents a comparison of the test scores between the proposed data fusion approaches and the AKF fusion method across different tasks: binary and multiclass classification, as well as regression. The results are shown for the denoised signals using VMD (see Section 3.2.2), and the fine-tuned 2D-CNN model (see Section 3.1.1) is used for the analysis tasks, outlined as follows:

- **Binary-class classification:** The proposed data fusion approaches consistently outperformed the AKF method in terms of accuracy and F1 score. For instance, the proposed approach achieved an accuracy of 98.17% compared to AKF's 88.94%, and "DF-II" achieved the highest accuracy of 99.62%.
- **Multi-class classification:** Similar to binary-class classification, the proposed methods demonstrated superior performance over AKF, with "DF-II" achieving an accuracy of 98.42% compared to AKF's 83.12%.
- **Regression:** In regression tasks, the proposed data fusion approaches also outperformed AKF, with "DF- II" achieving an R^2 score of 98.42%, in contrast to AKF's score of 79.85%.

Overall, the results indicate that the proposed data fusion approaches consistently provided higher accuracy, F1 score, and R^2 than the AKF fusion method across all tasks. This suggests that the proposed fusion methods are more effective in handling noise and extracting useful information from the data, leading to better performance in various types of tasks.

6.2. Comparative assessment of machine learning models

In this section, various machine learning techniques—random forest, KNN, SVM, XGBoost, and ensemble methods—are investigated to evaluate the performance of the 2D-CNN. As in previous analyses, these ML techniques were employed for classification and regression tasks using different combinations of input and signal types. The inputs included data from individual sources (FBGs and accelerometers) as well as fused data from the input, feature, and decision levels. Feature extraction was conducted using the VMD algorithm. The VMD parameters and their designated values are listed in Table 2. Seven statistical features were extracted from the IMF decomposition results of a given signal, as outlined below:

1. The first quartile of the instantaneous frequency (IF) signal is defined as follows:
$$\mathrm{IF}_k(t) = \frac{\mathrm{d}\phi_k(t)}{\mathrm{d}t} \tag{10}$$
where $\phi_k(t)$ indicates the instantaneous phase of the k**th** IMF, as presented in (1).
2. second quartile of the IF signal,
3. third quartile of the IF signal,
4. center frequency (ω) of the IMF.
5. Kurtosis of the IF signal,
6. variance of the IF signal,
7. skewness of the IF signal,

The extracted features are labeled as presented in Table 9.

The raw signals were decomposed into three IMFs using the VMD algorithm, resulting in a total of 21 features. To select the most effective and uncorrelated features, a PCA-based feature selection algorithm was applied, as discussed in [38], thereby selecting 15 features as the optimal subset.

In our evaluation, we used Monte Carlo cross-validation [43–45] with 100 iterations. Fig. 23 provides a comprehensive summary of the key metrics, including accuracy, precision, recall, and F1 scores, for binary-class classification across various data combinations and models. Our approach incorporates an ensemble model that combines the decisions of the top-performing models using XGBoost as the metamodel. Our findings indicate that both XGBoost and the ensemble model consistently outperformed the other models. Notably, the highest accuracy of 91% was achieved through decision fusion using the ensemble framework, particularly with the data fusion model DF-II. However, as illustrated in the figures, the testing scores for binary classification fall short of the results obtained with the 2D-CNN models (Table 5).

In multiclass classification, as illustrated in Fig. 24, the highest test accuracy reached 77% when using the ensemble decision model (decision fusion) in conjunction with the raw data fusion model (DF-II). Despite this notable achievement, the results fall significantly short of the multiclass testing scores obtained using the 2D-CNN models (Table 6).

In Fig. 25, the test R^2 scores are displayed across all models. Notably, the highest R^2 score of 63% is achieved by the decision model within the ensemble combined with the data fusion model (DF-II). However, these results, although commendable, fall short compared to the R^2 scores achieved by the 2D-CNN models (Table 7). Fig. 26 presents the perfect regression line plot for all models. Our observations based on the figure show that none of the models considered can perfectly predict the load values when compared with the proposed 2D-CNN model (Fig. 22).

These results highlight the superiority of the 2D-CNN model over other ML models [7,8] in binary and multiclass classification, as well as regression tasks on crowd movement and bridge load prediction.

7. Conclusions

This paper proposes an innovative monitoring methodology for coupled structure-human sensing. The overarching goal was to enhance pedestrian safety through the comprehensive monitoring of crowd movement and structural performance. In our approach, the synergy of the insights acquired from both structural responses and human behavior was integrated with state-of-the-art signal processing, machine learning, and multimodal sensor data fusion techniques at various levels, opening new avenues for advanced data analysis. This novel approach facilitates improved crowd and structural management by

Table 8
Comparison of proposed data fusion approaches with AKF fusion. The test scores are for the binary class, multilabel, and regression tasks.

Type of data fusion	Binary-class classification		Multi-class classification		Regression
	Accuracy (%)	F1 (%)	Accuracy (%)	F1 (%)	R^2 (%)
AKF	88.94	84.85	83.12	81.54	79.85
Acc	98.17	96.65	97.62	92.97	97.30
CoS	99.39	97.16	98.61	94.46	96.78
MoS	98.39	95.88	98.65	94.36	96.28
FF	98.87	96.73	98.33	94.56	97.00
DF-I	99.53	96.85	98.69	96.10	98.17
DF-II	99.62	98.99	98.42	95.38	98.42

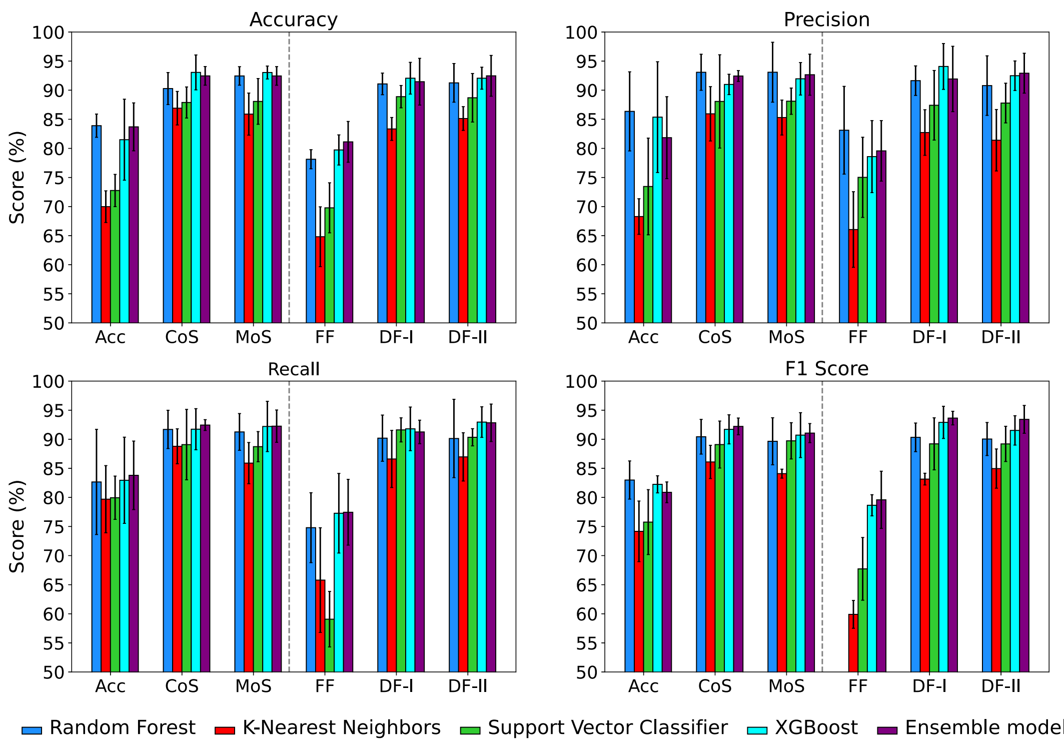

Fig. 23. Testing scores (%) for binary classification using VMD features for Random Forest, KNN, SVM, XGBoost, and ensemble methods.

Table 9
Description of features labels.

Index	Features	Description
1	$p1\_IF_i$	First quartile of IF for IMF_i
2	$p2\_IF_i$	Second quartile of IF for IMF_i
3	$p3\_IF_i$	Third quartile of IF for IMF_i
4	k_IF_i	Kurtosis of IF for IMF_i
5	v_IF_i	Variance of IF for IMF_i
6	cf_IF_i	Center frequency of IMF_i
7	sk_IF_i	Skewness of IF for IMF_i

establishing effective early warning systems for predicting safety risks. By leveraging an integrated understanding of crowd behavior and structural responses, the proposed methodology paves the way for proactive measures, to mitigate potential hazards and ensure the safety and resilience of infrastructures in crowded environments.

This proof-of-concept study was validated through laboratory experiments involving diverse groups on a bridge model and simulations conducted on crowd movements. The bridge behavior was monitored using embedded FBG sensors, whereas crowd movements were tracked using smartphone-integrated accelerometers. By leveraging fusion techniques and employing 2D-CNNs, novel algorithms were developed to classify crowd flow and density, as well as to estimate bridge loading with enhanced robustness. Various data pre-processing strategies, including decomposition using VMD, denoising through VMD, and different levels of downsampling, were explored to cleanse the data. A comparative analysis was conducted with existing data fusion approaches, such as AKF fusion and various machine learning models, including random forest, KNN, SVM, XGBoost, and ensemble methods, highlighting the advantages of the proposed approach. The developed methodology not only facilitates the estimation of crowd load and mobility parameters but also lays the foundation for a comprehensive crowd management system driven by artificial intelligence-based decision-making. The key findings of this study underscore the efficacy of the proposed approach in enhancing crowd monitoring and management capabilities, providing improved safety and efficiency in crowded environments. As a result,

1. Data fusion emerged as a highly effective approach for achieving superior performance in crowd analysis. The most effective fusion strategies were raw data fusion with the 2D-CNN model (DF-I and DF-II), feature fusion using 2D-CNN (FF), decision fusion employing an ensemble of machine learning models, and feature fusion incorporating selected features through PCA.
2. Denoising, particularly using VMD decomposition demonstrated remarkable effectiveness. Although low-level downsampling was observed to enhance the results compared to VMD, aggressive downsampling may compromise the model performance by sacrificing information integrity.

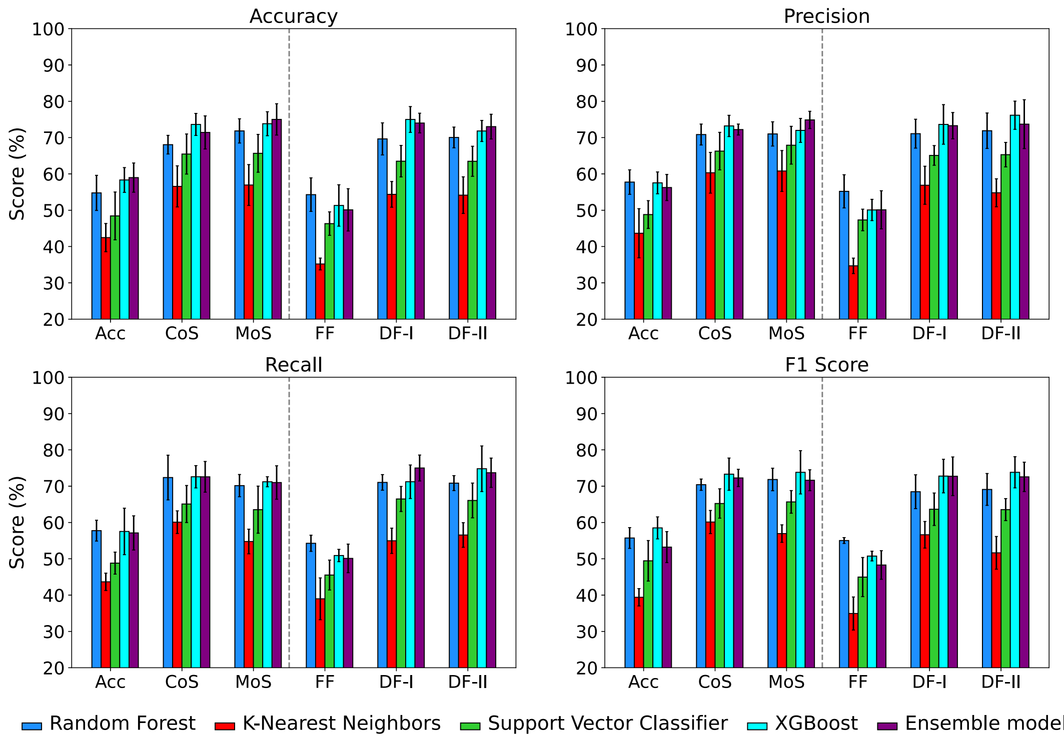

Fig. 24. Testing scores (%) for multiclass classification using VMD features for random forest, KNN, SVM, XGBoost, and ensemble methods.

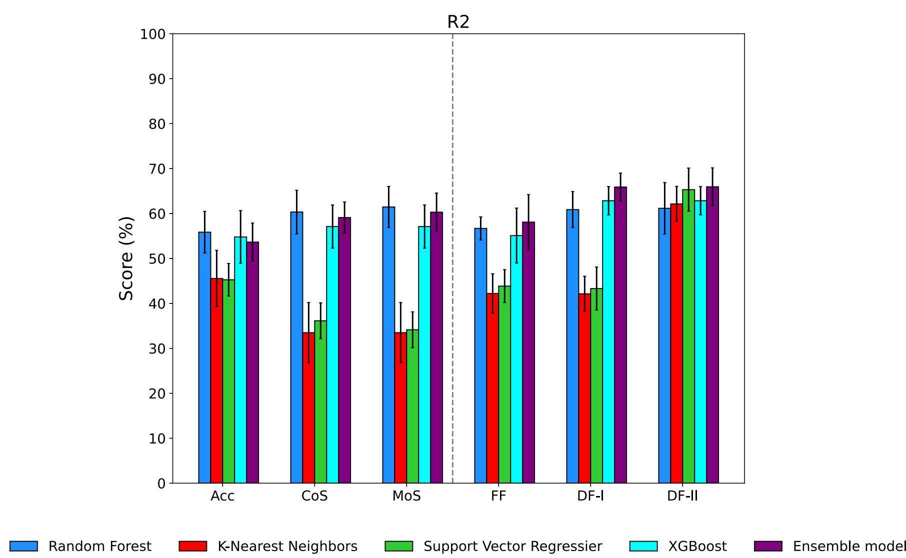

Fig. 25. Testing R^2 scores (%) for regression task using VMD features for random forest, KNN, SVM, XGBoost, and ensemble methods.

3. The fine-tuned and optimized 2D-CNN model exhibited superior performance compared to other state-of-the-art machine learning models because of its ability to extract high-level features from the input data.

Future research endeavors can focus on extending these strategies to encompass advanced structural health monitoring, with particular emphasis on structural degradation prognosis and performance assessment. By leveraging the methodologies established in this study, researchers can develop techniques to predict the deterioration of structures over time, evaluate their behavior under varying conditions, and acquire insights into the expected end-of-service life. With potential to explore data-fusion strategies and observe how structures respond to crowd distribution and motion, this research offers valuable insights into the interaction between crowds and infrastructure, facilitating the development of proactive measures, thereby enhancing structural safety and resilience in crowded environments.

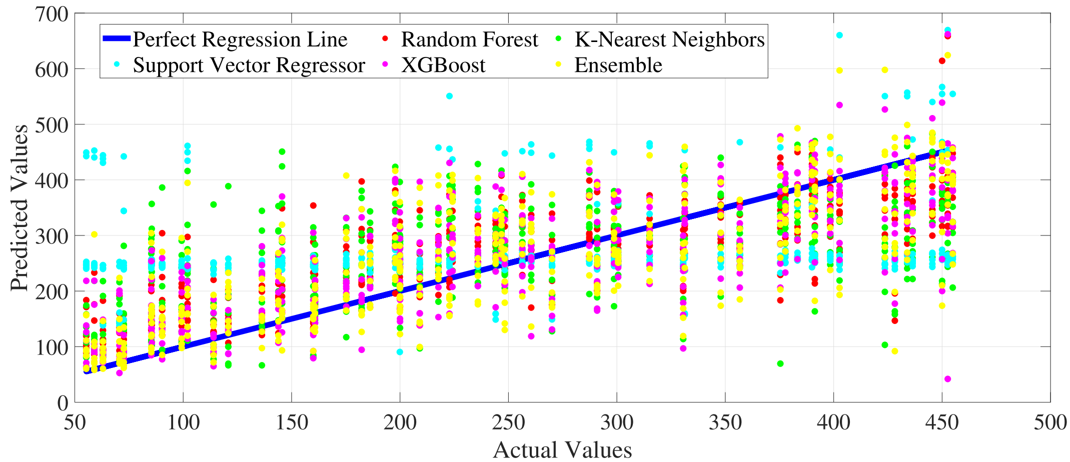

Fig. 26. Comparison of the predicted load value using VMD features vs. the perfect regression line for random forest, KNN, SVM, XGBoost, and ensemble methods.

CRediT authorship contribution statement

Sahar Hassani: Conceptualization, Visualization, Validation, Data curation (pre-processing and analysis), Methodology, Developing AI frameworks, Investigation, Formal analysis, Data analysis, Coding, Writing – review & editing, Writing – original draft. **Samir Mustapha:** Conceptualization, Visualization, Writing – review & editing, Data curation (experimental data), Funding acquisition (experimental data). **Jianchun Li:** Conceptualization, Visualization, Writing – review & editing, Supervision. **Mohsen Mousavi:** Conceptualization, Visualization, Writing – review & editing. **Ulrike Dackermann:** Conceptualization, Visualization, Writing – review & editing, Funding acquisition, Project administration, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Dr Samir Mustapha would like to acknowledge the financial support received from General Electric (GE), award # 103311, where the data was collected.

Data availability

Data will be made available on request.

References

[1] K. Still, M. Papalexi, Y. Fan, D. Bamford, Place crowd safety, crowd science? Case studies and application, J. Place Manag. Dev. 13 (4) (2020) 385–407.
[2] M. Haghani, M. Coughlan, B. Crabb, A. Dierickx, C. Feliciani, R. van Gelder, P. Geoerg, N. Hocaoglu, S. Laws, R. Lovreglio, et al., A roadmap for the future of crowd safety research and practice: Introducing the Swiss Cheese Model of Crowd Safety and the imperative of a Vision Zero target, Saf. Sci. 168 (2023) 106292.
[3] G. Still, Crowd safety and risk analysis - crowd disasters, 2018, Date Accessed (12 December 2019).
[4] T.G.L. To, S.C. Supply, Florida bridge collapse kills six (????).
[5] D. News, Prague bridge collapse leaves 4 injured, 2017, Date Accessed (2 December 2019).
[6] I. Fernandez, C. Callerame, EMDR intervention after a disaster: The Morandi Bridge collapse, Emerg. Care J. 15 (3) (2019).
[7] S. Mustapha, A. Kassir, K. Hassoun, B.A.A. Modad, H. Abi-Rached, Z. Dawy, Joint crowd management and structural health monitoring using fiber optic and wearable sensing, IEEE Commun. Mag. 57 (4) (2019) 62–67.
[8] S. Mustapha, A. Kassir, K. Hassoun, Z. Dawy, H. Abi-Rached, Estimation of crowd flow and load on pedestrian bridges using machine learning with sensor fusion, Autom. Constr. 112 (2020) 103092.
[9] K.S. Kaswan, V. Balu, A. Ojha, A. Sharma, D. Vekariya, A.K. Marandi, Deep learning algorithms and mechanisms in navigation for vehicular crowd management systems in real time for smart transportation, Soft Comput. (2023) 1–12.
[10] L.-Q. Xu, A. Anjulan, Crowd congestion analysis, 2010, US Patent App. 12/735, 819.
[11] N. Ilyas, A. Shahzad, K. Kim, Convolutional-neural network-based image crowd counting: Review, categorization, analysis, and performance evaluation, Sensors 20 (1) (2019) 43.
[12] N. Jmour, S. Zayen, A. Abdelkrim, Convolutional neural networks for image classification, in: 2018 International Conference on Advanced Systems and Electric Technologies (IC_ASET), IEEE, 2018, pp. 397–402.
[13] M. Fu, P. Xu, X. Li, Q. Liu, M. Ye, C. Zhu, Fast crowd density estimation with convolutional neural networks, Eng. Appl. Artif. Intell. 43 (2015) 81–88.
[14] C. Wang, H. Zhang, L. Yang, S. Liu, X. Cao, Deep people counting in extremely dense crowds, in: Proceedings of the 23rd ACM International Conference on Multimedia, 2015, pp. 1299–1302.
[15] Y. Zhang, D. Zhou, S. Chen, S. Gao, Y. Ma, Single-image crowd counting via multi-column convolutional neural network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 589–597.
[16] L. Boominathan, S.S. Kruthiventi, R.V. Babu, Crowdnet: A deep convolutional network for dense crowd counting, in: Proceedings of the 24th ACM International Conference on Multimedia, 2016, pp. 640–644.
[17] V.A. Sindagi, V.M. Patel, Cnn-based cascaded multi-task learning of high-level prior and density estimation for crowd counting, in: 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS, IEEE, 2017, pp. 1–6.
[18] A. Alia, M. Maree, M. Chraibi, A. Toma, A. Seyfried, A cloud-based deep learning system for improving crowd safety at event entrances, 2023, arXiv preprint arXiv:2302.08237.
[19] G.P. Hancke, B. de Carvalho e Silva, G.P. Hancke Jr., The role of advanced sensing in smart cities, Sensors 13 (1) (2012) 393–425.
[20] S. Hassani, U. Dackermann, M. Mousavi, J. Li, A systematic review of data fusion techniques for optimized structural health monitoring, Inf. Fusion (2023) 102136.
[21] A. Entezami, A. Entezami, Feature extraction in time domain for stationary data, Struct. Heal. Monit. Time Ser. Anal. Stat. Distance Meas. (2021) 17–45.
[22] E. Figueiredo, G. Park, J. Figueiras, C. Farrar, K. Worden, Structural Health Monitoring Algorithm Comparisons Using Standard Data Sets, Technical Report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2009.
[23] I. Ahmed, M. Ahmad, A. Ahmad, G. Jeon, IoT-based crowd monitoring system: Using SSD with transfer learning, Comput. Electr. Eng. 93 (2021) 107226.
[24] M. Al-Sa'd, S. Kiranyaz, I. Ahmad, C. Sundell, M. Vakkuri, M. Gabbouj, A social distance estimation and crowd monitoring system for surveillance cameras, Sensors 22 (2) (2022) 418.
[25] F. Venuti, F. Tubino, Human-induced loading and dynamic response of footbridges in the vertical direction due to restricted pedestrian traffic, Struct. Infrastruct. Eng. 17 (10) (2021) 1431–1445.
[26] B. Fu, X. Wei, An intelligent analysis method for human-induced vibration of concrete footbridges, Int. J. Struct. Stab. Dyn. 21 (01) (2021) 2150013.
[27] X. Wei, A simplified method to account for human-human interaction in the prediction of pedestrian-induced vibrations, Struct. Control. Heal. Monit. 28 (7) (2021) e2753.
[28] S. Nakamura, T. Kawasaki, A method for predicting the lateral girder response of footbridges induced by pedestrians, J. Constr. Steel Res. 65 (8–9) (2009) 1705–1711.

[29] A. Entezami, H. Shariatmadar, An unsupervised learning approach by novel damage indices in structural health monitoring for damage localization and quantification, Struct. Heal. Monit. 17 (2) (2018) 325–345.
[30] A. Entezami, H. Shariatmadar, S. Mariani, Fast unsupervised learning methods for structural health monitoring with large vibration data from dense sensor networks, Struct. Heal. Monit. 19 (6) (2020) 1685–1710.
[31] S.A. Singh, A.S. Kumar, K. Desai, Comparative assessment of common pre-trained CNNs for vision-based surface defect detection of machined components, Expert Syst. Appl. 218 (2023) 119623.
[32] P. Chakraborty, S. Roy, S.N. Sumaiya, A. Sarker, Handwritten character recognition from image using CNN, in: Micro-Electronics and Telecommunication Engineering: Proceedings of 6th ICMETE 2022, Springer, 2023, pp. 37–47.
[33] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.
[34] M. Elleuch, R. Maalej, M. Kherallah, A new design based-SVM of the CNN classifier architecture with dropout for offline Arabic handwritten recognition, Procedia Comput. Sci. 80 (2016) 1712–1723.
[35] M.R. Ahmed, S. Islam, A.M. Islam, S. Shatabda, An ensemble 1D-CNN-LSTM-GRU model with data augmentation for speech emotion recognition, Expert Syst. Appl. 218 (2023) 119633.
[36] N. Balamurugan, K. Maithili, T. Babu, M. Adimoolam, Stage-wise categorization and prediction of diabetic retinopathy using ensemble learning and 2D-CNN, Intell. Autom. Soft Comput. 36 (1) (2023).
[37] D. Zosso, Variational mode decomposition, MATLAB central file exchange, 2020, URL https://www.mathworks.com/matlabcentral/fileexchange/44765-variational-mode-decomposition. (Retrieved 27 August 2020).
[38] M. Mousavi, M.S. Taskhiri, A.H. Gandomi, Standing tree health assessment using contact–ultrasonic testing and machine learning, Comput. Electron. Agric. 209 (2023) 107816.
[39] P.A. Crossley, H. Guo, Z. Ma, Time synchronization for transmission substations using GPS and IEEE 1588, CSEE J. Power Energy Syst. 2 (3) (2016) 91–99.
[40] Y. Liu, X. Fan, C. Lv, J. Wu, L. Li, D. Ding, An innovative information fusion method with adaptive Kalman filter for integrated INS/GPS navigation of autonomous vehicles, Mech. Syst. Signal Process. 100 (2018) 605–616.
[41] S.d. Pellegrini, F.C. Trigo, R. Lima, Adaptive Kalman filter-based information fusion in electrical impedance tomography for a two-phase flow, Mech. Syst. Signal Process. 150 (2021) 107326.
[42] L. Zhang, W. Shaoping, M.S. Selezneva, K.A. Neusypin, A new adaptive Kalman filter for navigation systems of carrier-based aircraft, Chin. J. Aeronaut. 35 (1) (2022) 416–425.
[43] G. Shan, Monte Carlo cross-validation for a study with binary outcome and limited sample size, BMC Med. Inform. Decis. Mak. 22 (1) (2022) 1–15.
[44] M. Mousavi, A.H. Gandomi, M. Abdel Wahab, B. Glisic, Monitoring onsite-temperature prediction error for condition monitoring of civil infrastructures, Struct. Control. Heal. Monit. 29 (12) (2022) e3112.
[45] L.V. Ho, D.H. Nguyen, M. Mousavi, G. De Roeck, T. Bui-Tien, A.H. Gandomi, M.A. Wahab, A hybrid computational intelligence approach for structural damage detection using marine predator algorithm and feedforward neural networks, Comput. Struct. 252 (2021) 106568.