This is the Pre-Published Version.

The following publication Q. Chen, J. Cao and S. Zhu, "Data-Driven Monitoring and Predictive Maintenance for Engineering Structures: Technologies, Implementation Challenges, and Future Directions," in IEEE Internet of Things Journal, vol. 10, no. 16, pp. 14527-14551, 15 Aug., 2023 is available at https://doi.org/10.1109/JIOT.2023.3272535.

Data-Driven Monitoring and Predictive Maintenance for Engineering Structures: Technologies, Implementation Challenges, and Future Directions

Qianyi Chen, Jiannong Cao, *Fellow, IEEE,* and Songye Zhu

***Abstract*—Estimating engineering structures' health conditions and predicting their future behaviors are fundamental problems for a city's safe and efficient operations. Data-driven solutions estimate the health conditions using statistical models generated from measurement data. They have attracted growing interest recently because advances in information and communication technologies (ICT) have enabled numerous real-world measurement data, and the flourishing big data community has provided enormous state-of-the-art data analytics algorithms. Nevertheless, most existing studies remain in numerical simulation while neglecting their real-world implementation, restricting the extensive development of data-driven methods. In this survey, we provide a structural overview of the past decade's data-driven structural health monitoring (SHM) systems and algorithms from the perspective of real-world implementation. Specifically, we cover various aspects of the design and implementation of monitoring systems, including sensing technologies, communication technologies, and processing software. In addition, we classify the used data sources and statistical models with fined details. Under the proposed taxonomy, their limitations and advantages are thoroughly discussed. Based on our insights into existing studies, we clarify two major implementation challenges: the digressive performance in the real-world environment and inefficient computing systems for real-time data analytics. Possible solutions are then proposed to mitigate those challenges and promote the implementation of data-driven methods. Finally, we raise our outlook on future trends and suggest promising directions for further investigation.**

***Index Terms*—Data-driven methods, cyber-physical systems (CPS), structural health monitoring (SHM), predictive maintenance**

I. Introduction

ENGINEERING structures, as one of the biggest infrastructures in "Smart Cities," play a fundamental role in a society's safe and efficient production. Constructing such structures typically results in significant carbon emissions and requires substantial financial, labor, and energy resources. The failure or collapse of them can cause extensive damage to society and pose a significant threat to life. Maintaining the health and extending the service life of structures is of significant social and economic value. Structural Health Monitoring (SHM) has gained increasing attention as a means of achieving this goal. SHM involves implementing a damage detection strategy for aerospace, civil, or mechanical engineering infrastructure [1]. This process includes techniques for regularly measuring dynamic responses and environmental conditions of structures, as well as analyzing data to identify changes in material and geometric properties. SHM has become widely utilized in various structures, including bridges, high-rise buildings, stadiums, wind turbines, offshore oil platforms, and nuclear power plants [2]. Furthermore, its functionalities have extended from condition monitoring to predictive maintenance, which involves anticipating damage and performing maintenance before structures fail.

SHM is an interdisciplinary field that integrates technologies from computer science, mechanics, and engineering disciplines [3]. It also relies on techniques from communication engineering to electronic engineering [4]. For instance, in the case of monitoring the health of a long-span bridge, researchers often install numerous sensors to measure the bridge's displacement and vibration under various conditions, including traffic load, wind forces, and earthquakes. These sensors generate a large amount of streaming data from the local sensor network, which is transmitted to a remote cloud server for immediate decision-making, such as assessing whether the vibration frequency is abnormal. This process relies on sensing technology, wireless communication technology, and dedicated storage and computation devices like cloud servers for efficient streaming data processing. Dedicated data processing algorithms typically incorporate techniques from system identification [5] and big data analytics [1], such as feature engineering, data fusion, and machine learning.

The popularity of SHM has led to an abundance of proposed methods in the literature, which can be classified into two categories: model-based and data-driven methods [1], [6]. In model-based methods, physical or law-based numerical models [1] are built based on mechanical principles, engineering practices, and experience from inspection. Their parameters are predefined with structures' mechanical properties and then updated using measurement data from structures. Afterward, researchers can identify structural damages by tracking changes in mechanical properties, and the prognosis is performed by simulating the degradation of structures in their numerical models. Thanks to extensively validated physical principles, laws, and theorems, model-based approaches have been extensively used in many industrial systems [7]. However, inevitable discrepancies exist between engineering

Q. Chen is with the Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Hong Kong (E-mail: 19074424r@connect.polyu.hk)
J. Cao is with the Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Hong Kong (E-mail: csjcao@comp.polyu.edu.hk)
S. Zhu is with the Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong (E-mail: ceszhu@polyu.edu.hk)

© 2023 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

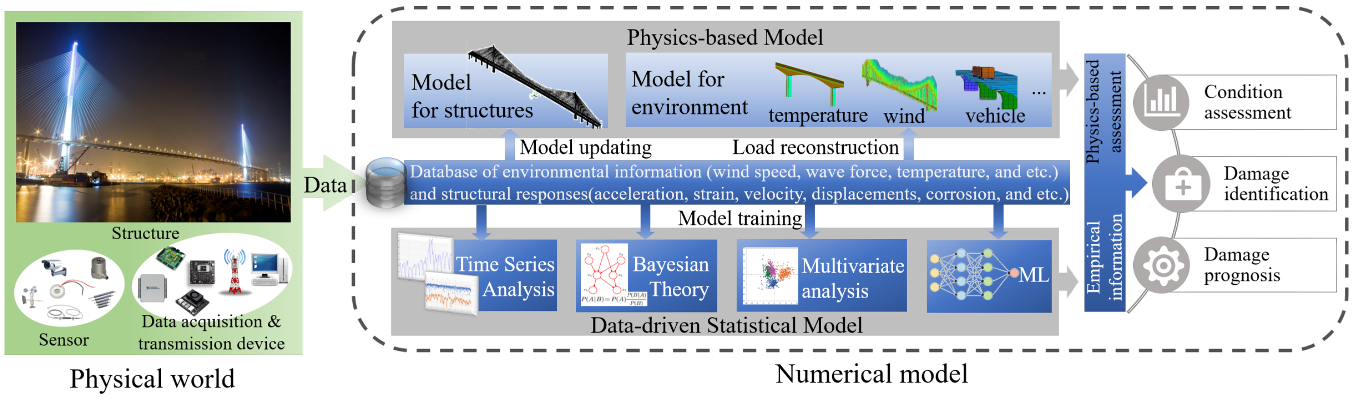

Fig. 1. Comparison between data-driven and model-based methods.

structures and their numerical models. Firstly, those numerical models are built with much-simplified material and geometry properties. Nevertheless, those properties of real engineering structures are always nonlinear and anisotropic. That brings inevitable bias to the predicted responses. Secondly, engineering structures' boundary conditions and ambient environment are ever-changing and unexpected. They are subject to numerous factors which can hardly be simulated with predefined physical models. This may also enlarge the gap between predictions and actual performance.

Unlike model-based methods, data-driven methods process measurement data without physics-based models. Instead, statistical models are built purely with data collected from structures. Then, damage-sensitive features (DSFs) are extracted, and their intrinsic correlation with damages is captured in either a supervised, self-supervised or unsupervised manner. Furthermore, structural health conditions are estimated with the newly sampled data. The comparison between data-driven and model-based methods is presented in Fig. 1. The major differences are the used numerical models and the way to utilize the measurement data. Model-based methods utilize the data to identify the physical properties of structures and reconstruct the load applied to those structures. Then, the corresponding parameters in physics-based models are updated. The environmental influences on structures are simulated with reconstructed load to the updated models. In contrast, data-driven methods abstract the monitoring process into a data mining task while ignoring the exact physical meaning of parameters. They target finding sensitive and robust features to be mapped to specific damages or used as condition evaluating indices for structures. Due to the irregular geometric size, heterogeneous material, and ever-changing environmental conditions, identifying the parameters for a fine-grained physics-based model is either too costly or impossible. As a result, it is fairly challenging to implement model-based methods in real-world engineering structures. Fortunately, recent advances in ICT have made it efficient and economical to collect 'big data' from engineering structures. As a result, there followed a surge in data-driven SHM in the past decade.

In this survey, we summarize studies on data-driven methods from the perspectives of monitoring systems and algorithms. Although several comprehensive reviews [8]–[16] have been published to summarize data-driven SHM methods, they either focus on a specific type of algorithms (such as vibration-based methods [10], [12], neural network [8], machine learning [9], [13], and deep learning [12], [15]), or a specific type of structure (such as bridges [11] and wind turbines [16]). In addition, they mainly clarify the theories of data-driven algorithms while neglect discussions on the overall monitoring systems from a computational perspective.

Different from existing surveys, this paper reviews both the data-driven algorithms and monitoring systems used for a variety of engineering structures. Moreover, we found that while various data-driven methods have been proposed, most of them are designed regardless of their performance in real engineering structures. Although they have been tested with numerical simulation or laboratory experiments, their implementation performance such as latency, accuracy, and robustness in real-world SHM systems are unknown. In fact, those performance might degrade significantly since those data-driven methods usually assume the excitation, damage types, and possible damage locations are known, which can hardly be fulfilled in real-world projects.

To this end, this survey also focuses on the concern of their real-world implementation. Based on our insights into existing studies, we foresee two major implementation challenges: the digressive performance of data-driven methods in the real-world environment and the poor efficiency of current monitoring systems for real-time data analytics. Moreover, we have proposed possible solutions and pointed out directions for further investigation.

This survey is ognized as follows. The design and development of data-driven SHM systems are firstly introduced in section 2. Section 3 clarifies typical procedures in different types of data-driven SHM methods. Key elements in those procedures, including sources of excitation, input data, statistical models, and monitoring objectives, are summarized with abundant examples. Section 4 classifies the input data used in data-driven methods into three types. Then we compare their advantages and disadvantages in real-world implementation with detailed descriptions. Section 5 goes into finer details about five classes of statistical models used in data-

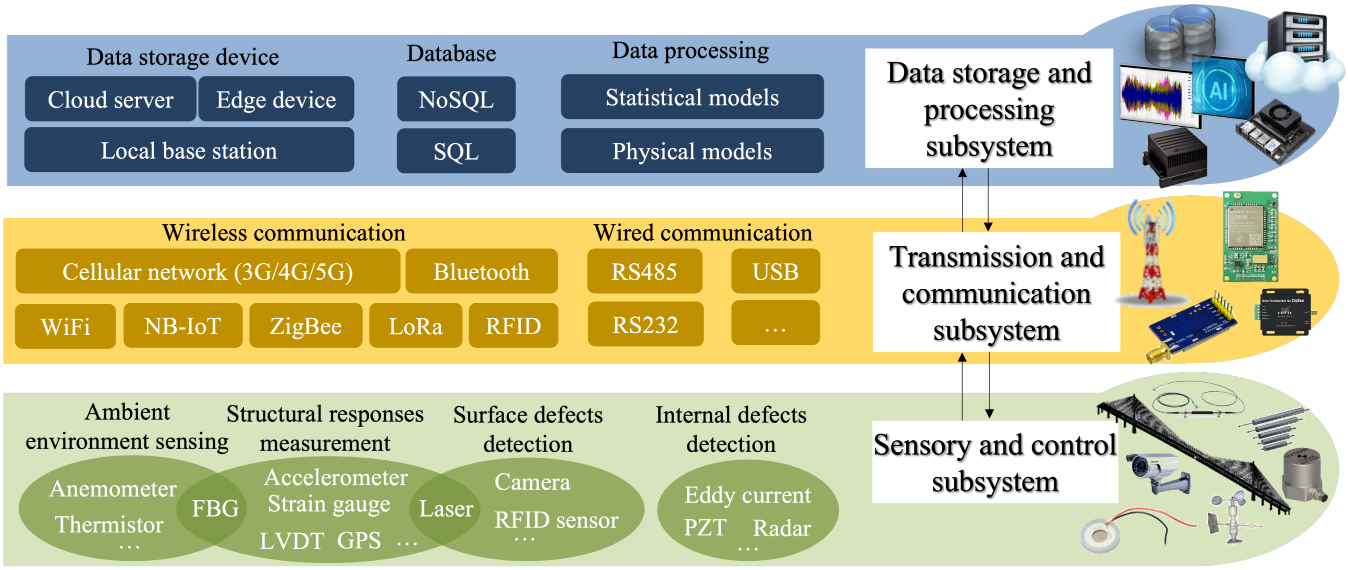

Fig. 2. Deployment architecture of SHM systems.

driven methods. Their performance in numerical simulations, reduced-scale laboratory experiments, and implementation in real structures are clarified. Challenges about existing data-driven algorithms in real-world structures are then discussed in section 6. We then propose fusion methods as a promising solution. Section 7 summarizes the challenges of implementing real data-driven monitoring systems and promising directions in improving those systems. The last section draws conclusions based on the comprehensive review of existing research. Suggestions for the usage of algorithms and design of SHM systems are given, and promising future trends are elaborated.

II. Implementation of Data-driven Methods to Real-world Structures

Although enormous data-driven algorithms have been proposed, most of them are tested in numerical simulation, where environmental factors are available and even controllable. Some algorithms have been validated in laboratory experiments [14], while implementations to real-world engineering structures are much fewer [17]. It is due to following challenges. Firstly, the responses of engineering structures are subject to a variety of excitation, such as temperature, traffic load, and wind load. It is hardly possible to observe and simulate all of them precisely. Secondly, data transmission is always disturbed by unexpected ambient environmental factors. The collected data are usually polluted by noises and sometimes suffer from data missing. Thirdly, the delay caused by data transmission and data processing is inevitable, especially for data processing algorithms with high space and time complexity.

In this section, we start from the implementation of data-driven methods in real-world structures. We first introduce the architecture of real-world SHM systems and clarify related techniques, hardwares and softwares involved in their implementation. Then we review the latest data-driven SHM systems for various real-world engineering structures. The performance of those systems and data-driven algorithms are clarified and summarized. Finally, we compare SHM with several popular areas in the frontier of modern scientific research.

A. Architecture of SHM Systems

As Fig. 2 shows, a SHM system for engineering structure is an Internet of Things (IoT) system [4] with three subsystems:

- Sensory and control subsystem
- Transmission and communication subsystem
- Data storage and processing subsystem

In the sensory and control subsystem of Fig.2, we summarize 13 widely used data acquisition and sensing technology and classify them into 4 categories: ambient environment sensing, structural responses measurement, surface defects detection, and internal defects detection. In ambient environment sensing, anemometers [18] are used to measure wind speed and direction. Thermistors [19] are temperature-sensitive resistors that change resistance with temperature, allowing for the monitoring of temperature variations in structures. A Fiber Bragg Grating (FBG) [20] is the optical sensor that measure strain, temperature, and pressure by monitoring changes in the wavelength of light reflected within the fiber. In structural responses measurement, accelerometers [21] measure acceleration and vibrations in structures. Strain gauges [22] measure the strain (deformation) in structures by converting changes in electrical resistance into strain values. Linear Variable Differential Transformers (LVDT), Global Positioning System (GPS) [23] and Laser-based sensors [24] can measure displacement in structures. Laser-based sensors can also measure surface defects of structures. The more commonly used surface detection device are cameras [25]. They can detect cracks, corrosion, and other signs of damage. RFID sensors [26] are also used to detect both surface crack and corrosion. In internal

TABLE I
SUMMARY TO THE COMMUNICATION TECHNOLOGIES USED IN SHM SYSTEMS

	Communication range	Transmission throughput	Energy consumption	Feature	Ref
RS232	<15m	<20kbps	/	Reliable, secure, and low-latency communication; Limited mobility and flexibility; Require physical cabling infrastructure; High installation and maintenance costs;	Used in most SHM systems
RS485	<1200m	<10Mbps			
USB	<30m	<20 Gbps(USB 3.2) <480 Mbps(USB 2.0)			
RFID	<10m	<100kbps	low	No internal power source; Simple and cost-effective deployment;	[33]
Bluetooth	<100m	<1Mbps	low	Supports point-to-point and mesh network topologies;	[34]
WiFi	<200m	<600Mbps(2.4GHz) <1733Mbps(5GHz)	middle	Supports multiple devices and network topologies; Higher power consumption than most wireless technologies;	[35]
Zigbee	<100m	<250kbps	low	Supports mesh network topologies with a large number of devices;	[36]
LoRa	<14km	<50kbps	low	Ideal for infrequent data transmission over long distances; Supports star network topology;	[37]
NB-IoT	<22km	<200kbps	low	Designed for massive IoT deployments; Operates in licensed spectrum, ensuring reliable communication;	[38]
Cellular network	3G:<5km 4G:<3km 5G:<300m	3G:2.8∼14.4Mbps 4G:300∼450 Mbps 5G:2.5∼4.6 Gbps	high	Requires a subscription and SIM card; Higher power consumption than other wireless technologies;	[39]

Note: / denotes the energy consumption is usually not considered in wired communication

defects detection, eddy current [27] is a non-destructive testing technique used to detect cracks, corrosion, or other defects in metallic structures by inducing electrical currents in the material. Lead(Pb)–zirconate (Zr)–titanate (Ti) (PZT) [28] is a type of sensor that generates electrical charge in response to mechanical stress, enabling the monitoring of vibrations, pressure, or strain in structures. Radar [29], [30] is used to investigate internal defects or damage in concrete, masonry, or other non-metallic structures.

To some extent, the choice in type, number, and location of sensors can decide the monitoring performance of a data-driven SHM system. Generally, the more types and number of sensors, the more accurate observations of structures and surrounding environment are obtained. Nevertheless, the deployment of sensors is also restricted by the cost, requirement of data processing algorithms, and the installation condition of a structure. The layout schema of sensors is a tradeoff between performance and expense. Numerous studies on sensor layout optimization [31] have been proposed to achieve the optimal monitoring performance with a limited number of sensors.

Data collected by sensors are transmitted to the data storage and processing subsystems through a wired or wireless connection. RS485, RS232, and USB cables are widely used for wired communication. RS485 and RS 232 are two serial data standards, while USB is the universal serial bus. For wireless data transmission, the widely used protocols include RFID, Bluetooth, Wifi, Zigbee, LoRa, Cellular network (3G, 4G, 5G), and NB-IoT. The comparison of those protocols used in SHM systems is summarized in the table below. Here in Table I, we only compare the key performance metrics of different protocols. Please refer to a survey of Abdulkarem et. al. [32] for a comprehensive review of the wireless communication technologies used in SHM.

Before the extensive use of cloud computing, data in SHM systems are stored with a self-designed schema in the local base station deployed on or near the engineering structure. With the emerging demand for storing and processing a large volume of measurement data, cloud computing was ultilized in the monitoring system. Meanwhile, standardized databases, including structured query language (SQL) [40]–[42] and not only SQL(NoSQL) [43]–[45] database management systems were involved in the monitoring system, enabling efficient data storage, access, and management. SQL is applicable for most current SHM systems, especially for small-scale SHM systems with few sensor nodes. Because their data are usually collected in fixed sampling frequency and have unchanged data types. Typical SQL databases used in data-driven SHM systems include PostgreSQL [41] and MySQL [40], [46].With the development of big data analysis, the fusion of multi-modality data has become a growing trend in data-driven SHM. The types and volume of these data can be dynamic as they may come from ever-changing and unexpected data sources. For instance, some SHM systems [44], [47], [48] are designed to store images, videos, and words, such as inspection reports. For such systems, NoSQL databases may be more suitable due to their dynamic schema for unstructured data. Popular NoSQL databases in data-driven SHM include MongoDB [43] and Apache Cassandra [43], [44].

Most of the existing SHM systems store and process data on cloud platforms. Monitoring systems running on local base stations might only support data storage and light-weighted signal conditioning. Computation-intensive data-driven algorithms are performed on other devices in an offline fashion. Domain-specific tools are used to process collected data. Extensively used software includes but not be limited to Matlab for system identification, Python together with various ML libraries for data mining, and universal finite element analysis (FEA) software such as StadPro, ANSYS, Bentley, Adina, and Abaqus. In addition, BIM-related tools are used for building digital models, including Autodesk Revit, Grasshopper, Dynamo, and Rhinoinside. Especially, some systems on local base stations still incorporate enormous computation resources and can support computation-intensive data analysis locally. For instance, for Tsing Ma Bridge [47] in Hong Kong, a dedicated data processing and control system has been built in the central control office of the bridge's administrative department. However, it is usually applicable for crucial large-scale engineering structures where data safety is considered a high priority and the budget is abundant.

Benefited from the elastic storage and computation resources, monitoring systems running on the cloud usually incorporate data storage functions and adequate data analytics tools. As data are continuously generated from sensor clusters and flow into the cloud server, stream data processing frameworks, such as Apache Kafka, Apache Spark, and Apache Flink, are also needed in the cloud for real-time or near real-time data analytics. Moreover, cloud-based monitoring systems can provide three modes of services: infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS). IaaS means the computation infrastructures can be directly utilized for the execution of designed algorithms and scripts. PaaS is mainly for the running of data analytics applications on the fly. SaaS provides user interfaces, web services, and many more services directly to users for efficient remote access, control, and data visualization.

Except for cloud computing, fog and edge computing are also introduced to SHM systems recently. The borderline between fog computing and edge computing is blurred in the monitoring systems of engineering structures. They both aim at migrating computation resources to the distributed devices near the data source, achieving less latency and higher data safety. Dang et al. [42] designed a hybrid-cloud based SHM system that used both the cloud computing layer and fog computing layer. The fog layer's local server can data preprocessing steps and remove insensible data. Then burden in data transmission and computation of cloud servers are much mitigated. Besides, the fog layers can perform encrypted data communication between them and the public cloud, promising the security of data for important engineering structures.

B. Data-Driven SHM Systems for Real-World Structures

Ever since the end of the last century, SHM systems have been extensively deployed in various crucial engineering structures all overall the world. However, their incorporation with data-driven tools starts much later. In Table II, we summarize existing real-world SHM systems that incorporate data-driven algorithms. We mainly summarize the SHM systems for engineering structures like bridges, stadiums, power plants, railways, and other structures. We don't include high-rise buildings because SHM systems' implementation to high-rise buildings have been comprehensively summarized by Shan [49] and Sivasuriyan [50].

Table II summarizes details for each monitoring system and the functionalities of the data-driven method in each system. As the table implies, most of the data-driven algorithms in the monitoring systems are run offline in local base stations instead of online. This is because of the high time and space complexity of most data-driven algorithms. Secondly, most SHM systems are developed for a specific structure or a small population of structures under unified management. Nevertheless, numerous works for the SHM system development are repetitive and trivial. An universal commercial solution is desperately needed. Nevertheless, few commercial integrated solutions for SHM systems are available in the market. Lastly, wired-based monitoring systems still dominate practical engineering structure projects, despite the convenience of wireless sensor nodes in both device installation and maintenance. Since the abundant and stable power supply required by existing wireless sensor nodes are usually inapplicable for most projects. Besides, the data transmission stability of wireless communication is much poorer than wired communication solutions, while stability has a high priority for real-world monitoring projects.

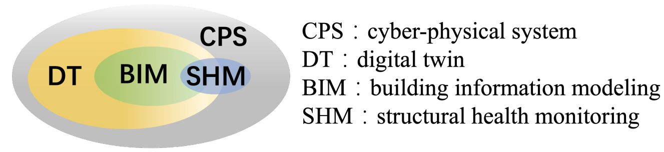

Fig. 3. Relationship between CPS, DT, BIM, and SHM.

C. Comparison of SHM with CPS, DT, and BIM

SHM is closely related to three hot research areas: cyber-physical system (CPS), digital twin (DT), and building information modeling (BIM). To avoid confusion, we distinguish the four concepts before going deeper into data-driven methods in SHM. Their relationships are illustrated in the Venn diagram shown in Fig. 3.

CPS refers to physical and engineered systems whose operations are monitored, coordinated, controlled, and integrated by a computing and communication core [75]. It focuses on the integration of computation and physical process, usually with feedback loops where physical processes affect computations and vice versa [76]. Typical physical systems constitute physical entities and real processes such as civil structures, human bodies, water, transportation, power grid, supply chain, etc. DT refers to a multi-physics, multi-scale, probabilistic, ultra-fidelity simulation that reflects, in a timely manner, the state of a corresponding physical entity based on historical data, real-time sensor data, and physical model. DT [77] is usually regarded as a subset of CPS as it does not include the controlling of physical entities like CPS. Building information modeling (BIM) [78] is developed from computer-aided design (CAD) and used for planning, design, construction, and operation of facilities. BIM can be regarded as DT's application in the building industry.

SHM can be interpreted as CPS's application in monitoring engineering structures, including but not limited to large-scale civil structures, aerospace structures, and machinery. In the meanwhile, SHM has overlaps with both BIM and DT. Nevertheless, it also includes many particular areas that are not in DT and BIM. For instance, SHM involves many non-destructive testing (NDT) methods for defects monitoring. Those technologies don't rely on digital twin. Some pure data-driven SHM methods are not included in BIM because 3D digital models are not needed in those methods.

III. Typical Procedure Of Data-Driven Methods

Similar to the taxonomy proposed by Rytter [79], we divide the objectives of data-driven monitoring and predictive maintenance methods into three levels.

- Damage identification

TABLE II
SHM SYSTEMS WITH DATA-DRIVEN METHODS FOR REAL-WORLD ENGINEERING STRUCTURES

Structures	Sensors, data transmission and storage methods	References and key functionalities
Four bridges in Hong Kong, China	Sensors: Accelerometers, anemometers, strain gauges, displacement sensors, GPS, thermometers, dynamic weight in motion stations, hygrometers, corrosion cells, cameras, elasto-magnetic sensors, and others; SHM System: Wired local base station	[47] incorporates various data-driven algorithms in an off-line SHM system
Yeongjong grand bridge, Seoul	Sensors: Accelerometers, laser displacement sensors, potentiometers, tiltmeters, strain gauges, thermometers, anemometers; SHM system: Data are stored in data loggers on the bridge	[51] use KPCA to learn relationships between responses and environment, and detect anomalies by statistical novelty detection methods.
A suspension bridge, England	Sensors: Anemometers, displacement sensors, thermometers, load cells; SHM system: NA	[52] use five statistical models to predict bridge's natural frequencies with environmental conditions.
Tianjin Yonghe Bridge, China	Sensors: Accelerometers, hygrothermographs, and anemometers; SHM system: Wired local base station	[53] train a Bayesian neural network to predict acceleration and use prediction errors to detect anomalies
Hollandse Brug bridge, Netherlands	Sensors: Temperature sensors, strain sensors, and vibration sensors; SHM system: NA	[54] train a graph neural network to predict future strain with strain data from sensors
Infante D. Henrique bridge, Portugal	Sensors: Accelerometers and thermometers; SHM system: Wired system	[55] use multivariate statistical techniques to extract DSFs from frequencies and identify damages
A PSC box girder bridge, France	Sensors: Accelerometers and thermometers; SHM systems: Local wired system with offline data processing	[56] learn the correlation between temperature and the bridge's model parameters by neural network based regression
A suspension bridge, China	Sensors: Anemometers and accelerometers; SHM system: NA	[18] predict the vortex-induced vibration mode with a trained decision tree
Dashengguan Yangtze Bridge, China	Sensors: Strain gauges and thermometer; SHM system: NA	[57] identify the aging truss with correlation learned from strain and temperature.
25 de Abril Suspension Bridge, Portugal	Sensors: Thermometers, anemometers, strain gauges, accelerometers, displacement sensors, and tiltmeters; SHM system: NA	[58] train a gradient boosting regression tree to predict fatigue damage with environment parameters
Sydney Harbour Bridge	Sensors: Accelerometers; SHM system: Wired Local computing node	[59] use k-means to identify damaged jack arches with the spectral moment
Telegraph and Newburg Road Bridge, US	Sensors: Strain gauges, strain transducers, uniaxial accelerometers, thermistors, and cameras; SHM system: Wireless SHM system with a cloud server	[44], [48] analyze the correlation between traffic load and strain, and propose a data-driven load rating method for damage detection
FuSui bridge, Korea	Sensors: Displacement sensors, temperature sensors, strain sensors, and accelerometers; SHM system: Wired local management station	[60] analyze correlation between strain and temperature for thermal response prediction
A steel girder bridge, Japan	Sensors: Accelerometers; SHM system: A wired temporary system	[61] identify locations of small changes in the structural mass and stiffness
Zhaobaoshan Bridge, China	Sensors: Thermometers, humidity sensors, displacement sensors, accelerometers, and strain gauges; SHM system: NA	[62] detect anomalies using dynamic ICA with strain after removing temperature effects
A high-speed track bridge, France	Sensors: Displacement sensors, accelerometers, temperature gauges, and Q sensors for measuring axle loads; SHM system: NA	[63] use a clustering method to discriminate three different stages of the bridge
G.Meazza stadium	Sensors: Accelerometers; SHM system: NA	[64] find the correspondence between autoregressive parameters and environment conditions
Lillgrund offshore wind farm	Sensors: Accelerometers in the tower and strain gauges on blades and sensors for condition monitoring data; SHM system: Data are stored in a cloud server	[65] use system information, environmental parameters, and vibration data to train a decision tree and detect abnormal operations
Two wind turbines, Germany	Sensors: Accelerometers and wind turbine operation information (power, wind, yaw angle, and etc.); SHM system: Wired system	[66] use polynomial chaos expansion smoothness priors time-varying autoregressive moving average to diagnose the structural condition
An offshore wind turbine, China	Sensors: Accelerometers; SHM system: Wireless system with online data process	[67] train two support vector machines with annotated time-domain and frequency-domain data, respectively, for anomaly detection of the wind turbine.
A suspension railway, Germany	Sensors: Accelerometers, tiltmeters, and velocity sensors; SHM system: Data are stored in local SD card and processed offline	[68] train a support vector machine using pristine track condition and use it to identify health state and detect anomalous signals
Sunrise Movable Bridge, US	Sensors: Accelerometers; SHM system: Wired system	[69] use cross correlation analysis and robust regression analysis to detect damage scenarios caused by leakage and lack of sufficient oil in the gearbox
San Pietro bell-tower, Italy	Sensors: Accelerometers and temperature measurement sensors; SHM system: Wired system with local PC and remote server	[70] use multivariate statistical analysis techniques to remove effects of environmental conditions and detect damages in the form of outliers in frequencies
Consoli Palace, Italy	Sensors: Accelerometers, crack meters, and temperature sensors; SHM system: Wired system with online data processing	[71] train multivariate statistical analysis models with identified modal features and use them to identify anomalies of the structure
A concrete gravity dam, China	Sensors: Displacement sensors, water level, and thermometers; SHM system: Local data storage system with offline data processing	[19] use a support vector machine to learn the correlation between temperature and displacements
La Baells dam, Spain	Sensors: Air temperature sensors, reservoir level sensors, and displacement sensors; SHM system: NA	[72] train boosted regression trees to predict the dam's future behavior, and use displacement prediction errors to detect anomalies
Alto Rabagao dam, Portugal	Sensors: Temperatures sensors, level sensors, displacements sensors, tiltmeters, strain gauges, pressure sensors, and seepage sensors; SHM system: NA	[73] use multiple linear regression (MLR) is to analyze variations in the crest displacements with temperature and reservoir level
Alto Lindoso dam, Portugal	Sensors: Thermometers, reservoir water level, displacements sensors, strains gauges, and pressures sensors; SHM system: NA	[74] use linear regression to simulate variance of time-frequency features of displacements and temperature

- Condition assessment
- Damage prognosis

Damage identification includes identifying the existence, location, type, and extent of local defects in engineering structures. Due to the sparsity of local damages in the whole structure, damage identification is usually performed regularly and after the structure suffers from a sudden degradation or an extreme disaster strike. Data-driven damage identification aims at finding physical defects which cause the change of the condition indices. However, different damage patterns might lead to similar influences on structural conditions. So, the key objective of damage identification is searching for DSFs which are accurately mapped to physical damages such as cracks, a decrease in stiffness, corrosion, etc. Condition assessment refers to estimating a structure's overall serviceability in normal or unfavorable environmental conditions during a long service life [11]. It should be performed regularly to closely track the fluctuation of conditions and detect the abnormality instantly. The condition assessment mainly includes the following tasks:

- Estimating structural condition indices under normal load pattern (temperature load, traffic load, pedestrian load, and wind load)
- Estimating structural condition indices in extreme events (earthquakes, typhoons, collisions, blasts, and overloaded vehicles)
- Estimating structural conditions after retrofitting

Data-driven condition assessment aims at extracting structural condition indices (SCIs) from measurement data to estimate the reliability of structures. Widely used SCIs include parameters of time series analysis models [23], [60], statistical features [80], [81] of raw data, and structural vibration properties [10]. Data-driven condition assessment usually spans a long period of service life. Previously estimated SCIs can be used as the baseline and the long-term variation of SCIs shows the condition of a structure.

The search of SCIs or DSFs in data-driven SHM methods can be interpreted as typical feature engineering tasks. Specifically, a data-driven method aims to select and optimize features that are sensitive to structural health conditions or damages from measurement data. The typical procedures of those data-driven approaches are shown in Fig. 4. Based on the labels used in training samples, those data-driven methods are classified into three types: supervised method, self-supervised method, and unsupervised method. In supervised methods, a statistical model is trained with samples annotated with labels. In SHM, those samples usually include excitation applied to the structure and corresponding structural responses collected by sensors. Labels are known health conditions and damage information. Typical statistical models used in supervised methods include support vector machine (SVM) [82], decision tree [65], ensemble learning model [83], neural network [53], and many deep learning models [12]. Both self-supervised methods and unsupervised methods train statistical models without structural damage information.

Instead of predicting health conditions or damages directly, models in self-supervised methods are trained to predict data in the future steps, while models in unsupervised samples are trained to capture the correlations between variables in samples or to cluster the given samples. Then damages and health conditions are estimated by analyzing the prediction error and outliers in the new coming sample. Widely used statistical models for self-supervised methods include time series analysis models [84], Bayesian-based models [53], and various ML models [54], [85]. Unsupervised methods mainly include multivariate analysis such as principal components analysis (PCA) [70], canonical correlation analysis (CCA) [86], independent component analysis (ICA) [62], and cluster analysis [63]. More details on the statistical models used in data-driven methods are clarified in section 5 of this survey.

Damage prognosis denotes forecasting an engineering system's future damages or health conditions and estimating its remaining service life [87]. It is the primary task in predictive maintenance. Different from the previous two levels, it relies not only on structures' past and present information but also prediction of future environmental conditions. Data-driven damage prognosis relies on the experience from previous data records, current structural conditions, together with the prediction of structural damages and load conditions development in the future. As shown in Fig. 5, an accurate prognostic analysis never divorced from monitoring results in the previous two levels. Furthermore, It requires the integration of both deterministic and probabilistic predictive modeling capabilities [88]. Among the three levels. damage prognosis is the most complicated one due to the lack of comprehensive estimation of current structures and the uncertainty of the future surrounding environment in real applications. Meanwhile, it is regarded as the future direction of SHM [88] owing to its tremendous potential for life-safety and economic benefits during a structure's service life.

Table III summarizes four critical modules in a typical data-driven method. The choices for each module are listed based on the reviewed data-driven methods in this paper. The sources of excitation are usually determined by the used statistical models and the common load conditions of engineering structures. For instance, wind turbines usually work under wind load, while bridges are subject to traffic load and temperature effects. More details about the input data and statistical models are introduced In the following two sections.

IV. Types Of Data Used In Data-Driven Methods

As the name implies, data-driven methods profoundly rely on the data collected from structures. Generally, they require an enormous amount of data to generate their feature spaces. Moreover, the choice of data determines the upper limit of a data-driven method. Specifically, the more sensitive the data is to structures' local defects or changes in overall health condition, the higher accuracy in structural health monitoring will be achieved. We summarize the used data into three levels based on the interoperability of their correlation to structural health conditions and damages.

A. Data with Low Interpretability

Raw acceleration, strain, internal force, and displacement [21] are features that explicitly reflect the conditions of

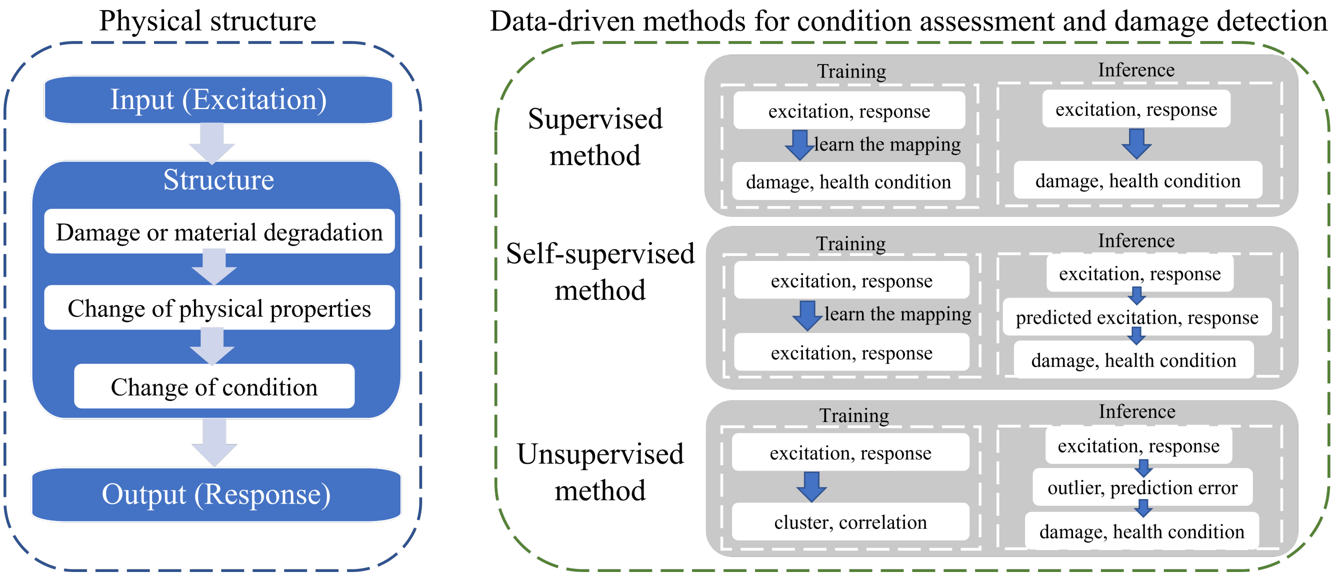

Fig. 4. Procedures of supervised, self-supervised, and unsupervised data-driven methods.

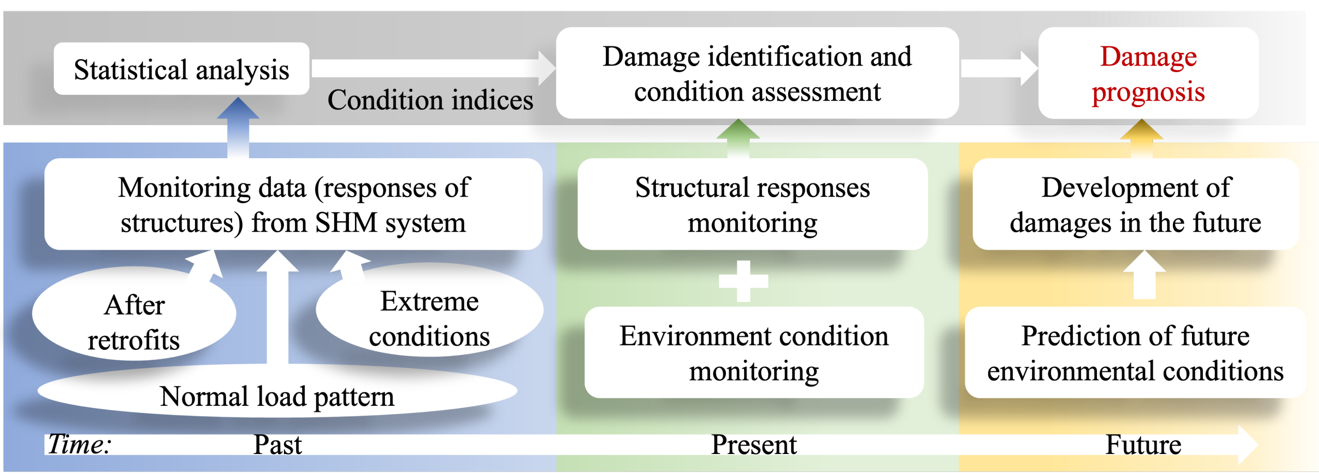

Fig. 5. Typical procedures of data-driven damage identification, condition assessment, and damage prognosis.

TABLE III
SUMMARY OF CRITICAL MODULES IN A DATA-DRIVEN METHOD

Sources of excitation	Input data	Statistical model	Monitoring cbjectives
1.White noises, ambient vibration, random excitation	1.Raw acceleration data	1.Time series analysis: AR, ARMA, NARX, and NARMAX	1.Existence of damage
2.Traffic load, vehicle bridge interaction or human-structure interaction	2.Raw strain, internal force, and displacement	2.Bayesian theory	2.Location of the damage
3.Temperature effects	3.Modal properties	3.Multivariate analysis (PCA, ICA, CCA, regression, cluster analysis)	3.Type of the damage
4.Impact load, hammer excitation	4.Frequency domain features	4.ML based method (SVM, DT, KNN, and Ensemble learning, NN, DNN)	4.Extent of the damage
5.Wind load	5.Time-frequency domain features	5.Hybrid statistical models	5.Condition assessment
6.Earthquake, ground motion, dynamic base excitation	6.Data collected with NDT methods		6.Damage prognosis
7.Uniform pressure, self-weight, and other dead load			
8.Excitations in NDT methods			

engineering structures. Intuitively, the large deformation and abnormal vibration can be taken as signals of danger for structures that suffer from damage. However, they cannot be interpreted as damage-sensitive features with high confidence. Because acceleration, strain, internal force, and displacement are determined by not only the mechanical properties of a structure but also the load applied to the structure. The abnormal deformation and vibration might result from the unexpected external loads applied to structures, but the structures are still healthy and intact. Consequently, they are classified into data with low interpretability in this paper.

Raw acceleration: In traditional structure dynamics, acceleration data are most frequently used to describe the dynamic responses of a structure. Various structural dynamic features representing the dynamic properties of structures can be extracted from raw acceleration data from structures. Nevertheless, most of the data-driven methods cannot extract DSFs directly from them without any preprocessing. Yet, some types of neural networks [53], especially deep neural networks (DNNs) like CNN [89]–[95], have been proven to achieve high prediction accuracy with raw acceleration data. Besides, deep Bayesian belief network (DBBN) [96] has also been validated to detect damage of structure from acceleration responses.

Raw strain, internal force, and displacement: The strain of structures can be measured with FBG and strain gauges. Structual displacements can be monitored with LVDTvacuum laser collimation system, GPS. It is widely accepted that strain is more sensitive to local defects while displacement is a common indicator of the overall structural condition and load patterns. Julian et al. [22], [97] identified structural damages with the strain field measured by FBG. Strain data collected from Canton Tower were used by Wan et al. [98] to predict future strain with Bayesian framework combined with Gaussian process. Internal forces can reflect the structural variation caused by environmental and operational effects, and they are especially important for cable-stayed bridges [85]. Based on the displacements and velocities of different nodes, Li and Yu [99] used online SVM to classify the damaged and healthy conditions of a structure. Kang [19] used SVM to learn the correlation between air temperature and displacements of a concrete gravity dam. Salazar [72] trained boosted regression trees to predict a dam's future behavior using its radial displacement and various environmental factors.

B. Data with Medium Interpretability

Structural modal properties, frequency domain features, and time-frequency domain features are extracted from structural dynamic responses (e.g., raw acceleration and displacement). They have interpretable physical meanings and are able to reveal the intrinsic dynamic properties of structures. Specifically, the degradation of material and defects of structures lead to the variation of their dynamic properties. Compared with raw dynamic responses, dynamic properties are less related to excitations and more subject to structural states themselves. Consequently, they are extensively used as indicators in the monitoring of real-world engineering structures. However, the variation of dynamic properties might also result from the temperature change. Besides, those data are not sensitive to damage precursor and small-scale damages [100]. So, they are classified into data with medium interpretability in this paper.

Structural modal properties: Structural modal properties include frequencies [101], [102], damping ratios [103], mode shapes [101], [104], and mode shape derivatives [102], [105]. Wang and Ni [106] used measured modal frequencies and incomplete mode shape components to identify the occurrence of damages with auto-associative neural network. Frequencies and mode shapes were also processed by Cury and Crémona [107] for detecting structural behaviors based on symbolic data analysis (SDA). Huo et al. [108] chose cross correlation function amplitude (CCFA) as the DSF, wherein CCFA was related to the modal shapes of structures and was computed from dynamic responses of structures. Hu et al. [109] used PCA to detect the damage from the changes of natural frequencies of the structure, wherein PCA can eliminate the variations of frequencies caused by the operational and environmental conditions.

Frequency domain features: Fourier transformation is frequently used to extract frequency domain features and to build the feature space. Lederman et al. [110] and Duan et al. [111] mapped the fast Fourier transform (FFT) magnitude to the bridge's damages with machine learning algorithms. Damage location and damage extent were detected successfully. FRF is widely used to describe structures' dynamic character by representing the relationship between structural responses and excitation in the frequency domain. Fallahian et al. [112] used a couple sparse coding to detect damages in a numerical frame structure, wherein the scale of FRF was reduced via PCA first and then used as the features to be mapped to corresponding damages. As the ratio of two responses in the frequency domain, transmissibility has been proved to be sensitive in detecting structural dynamics behavior changes [113]. Nguyen et al. [114] used a neural network to detect the stiffness loss of the girder section for a bridge in Vietnam. The input of the neural network is the sum of transmissibility for different displacement collection nodes in different frequency ranges. Zhou et al. [115] used PCA to condense the transmissibility obtained from both a clamped-clamped beam and a model of the benchmark structure that were excited by impact force and impact hammer, respectively. The DSFs were constructed with distance-based measures. TF a mathematical representation of the output-to-output relationship for a dynamical system [116]. In the research of Diao et al. [117], TF amplitudes were preprocessed with PCA. After the reduction of dimensionality, the TF was inputted to SVM. Then damage location and extent were identified. Liu et al. [118] used one-dimension CNN to detect damage from structural TF. The acceleration responses are transformed into TF and fed to CNN during training and testing.

Time-frequency domain features: Wavelet transformation [119] is used to transform structural responses from time domain to time-frequency domain, through which the dynamic parameters of structures can be obtained. Shi and Yu [120] used discrete wavelet transform (DWT) to decompose the acceleration responses. Tibaduzia [121] computed DWT coefficients from the dynamic responses collected by PZT. Kim et al. [122] used DWT to eliminate the noises of structural velocity responses and reduce the computation cost of the following analyses. HosseinAbadi et al. [123] conducted the damage identification on both simulated and real steel beams with the guided ultrasonic wave (GUW) generated by PZT, in which wavelet packet transform was used to de-noise the GUW. Besides, wavelet packet distribution [124]–[126] is also used as the damage index for SHM.

C. Data with High Interpretability

NDT methods use capacitive, inductive, piezoelectric, optical effects, and many more non-destructive techniques to detect defects of structures made up of various materials like concrete and steel. Data collected with NDT methods [127], [128] and pictures taken by cameras are of highly interpretability because they can visualize or detect the physical defects (such as cracks, corrosions, and bolt loose) directly. Common techniques used in NDT include PZT, Laser, Eddy Current, acoustic emission [129] , near-field radio frequency reflectivity [130], X-ray [131], and Radar. Unlike NDT, which aims at detecting local defects inside structures, vision-based SHM methods identify defects on the surface of structures based on images. Owing to the tremendous progress of AI-based computer vision, vision-based SHM algorithms [25] have emerged and achieved extensive applications in real-world engineering structures during the last decade. Among all types of data used in SHM, image is one of the most direct indicator in identifying damages. Based on the identified severity, type and location of defects, the conditions of the monitored structures can be estimated. It should be noted here that there is a slight difference between NDT and NDE. NDT is restricted to testing, while NDE includes both testing and the evaluation of the results. In this survey, NDE refers to SHM methods that work with data from NDT methods.

V. Statistical Models Used In Data-Driven Methods

As a ‘bridge’ that links data and monitoring objectives, statistical models play a key role in estimating the conditions of structures. In this section, we introduced commonly used statistical models and summarized recent publications based on those models. Their advantages, disadvantages, and applicable scenarios are also discussed comprehensively.

A. Time Series Analysis

Time series analysis is a critical tool in many fields of research, including control, engineering, and economics. It employs time series models to uncover relationships and patterns within data that are recorded over time, providing insights into trends and behavior. The time series analysis for SHM usually follows three steps [84]: (i) random excitation and structural responses collection, (ii) statistical model fitting, (iii) damage identification and structural health inference. Raw acceleration data and responses collected with PZT sensors are common data sources in time series analysis. In SHM, the parameters of time series models and the residual errors of prediction are two common types of DSF. Based on different manners that time series models can simulate, those methods are basically classified into two categories: linear and nonlinear time series models, as summarized in Table IV. Autoregressive (AR) and autoregressive moving average (ARMA) models are the most frequently used linear models that can simulate and predict the linear responses of structures. Nonlinear Autoregressive with exogenous (NARX) and nonlinear Autoregressive Moving Average with Exogenous Inputs (NARMAX) models can predict the nonlinear structural responses. Obviously, the responses from real structures are always nonlinear, and thus those nonlinear models can simulate the systems with higher accuracy.

Recent research proposed some new DSFs for damage detection, including Ljung-Box statistic of AR model residual sequence [132], Cosh spectral distance of AR model spectrum [132], evolutionary spectra from TFARMA [133], and Kolmogorov–Smirnov (KS) test statistical distance [135]. In addition, innovations in the algorithm design have been made for damage identification, including NARMAX with Echo State Networks (ESNs) [144], AR-ARX two-staged model [134], and NARX neural network [136]. Except for damage detection, ARMA model has also been applied in long-term condition assessment. Van Le and Nishio [23] used an ARIMA model to simulate the global deformation pattern of tower and girder points in Can Tho bridge in Vietnam. They concluded that the AR-MA parameters could be used as features for estimating the bridge’s condition. NARX model can also be used to predict the thermal responses under hard environmental effects [60].

Time series analysis methods have advantages of low computational complexity, simpleness in form, and high interpretability. However, they operate on the premise that alterations in a structure’s integrity will cause corresponding changes in its dynamic responses. Nevertheless, these responses may only exhibit a dominant effect when damages are significant enough to produce discernible changes detectable by sensors. If response variations predominantly stem from excitation sources or environmental factors, time series analysis may struggle to identify defects. Additionally, time series analysis is most effective when a substantial amount of time-history data is available under consistent ambient conditions.

B. Bayesian Theory

Bayesian theory has found widespread applications in data-driven predictive maintenance, including Bayesian (belief) network, Bayesian neural network, Bayesian decision tree (BDT), and Bayesian Inference. Bayesian (belief) network is one type of graphical model embedded with the probability and direct dependency relationship of different variables and events. It is widely used for decision-making based on big data from various domains. Rafiq et al. [138] used dynamic Bayesian belief networks to conduct overall condition-based deterioration profiles based on the inspection results of different elements. Wan et al. [98] used Bayesian framework combined with Gaussian process to predict the future structural stress responses. The method was proved effective with strain data collected from Canton Tower. Flynn et al. [139] applied a Bayesian network to enhance the guided wave based SHM. With the help of Bayesian approach, the missed detections were considerably reduced in the guided wave based damages identification. A bayesian neural network is a neural network with a prior distribution on its weights [143]. Bayesian prior is used to regularize the neural network and control overfitting problems. Pan et al. [96] used DBBN to detect damage of a structure from acceleration responses. Arangio et al. [53], [90] also used Bayesian neural networks to detect damages for a numerical suspension bridge and real cable-stayed bridge based on acceleration time histories. BDT is a type of decision tree with Bayesian theory. Cury and Crémona [107] compared

TABLE IV
SUMMARY OF DATA-DRIVEN METHODS BASED ON TIME SERIES ANALYSIS

Time series analysis	Input Data	Exci-tation	Monitoring Objectives	Implementation	Ref
ARIMA/ AR model	1	1	1	Laboratory experiment 1: truss model and damages are simulated by adding disks; Laboratory experiment 2: concrete bridge model with cracks generated through earthquake excitation	[132]
	2	3	5	Girder and tower displacement of a cable-stayed bridge in Vietnam	[23]
	6	8	1	Steel truss bridge model in the laboratory with loosen bolts	[133]
NARMAX/	2	3	5	Strain and temperature data collected form Fu Sui bridge	[60]
NARX model	2	5	1	Modal properties identified from acceleration and wind turbine operation information (power, wind, yaw angle, rotor revolutions per minute, and temperature) are used for tracking and diagnosing structural condition of two wind turbines in Germany	[66]
	1	2	4	Reduced scale experiment of the bridge deck and field test on real bridge with cracks	[134]
	1	1	2, 4	Numerical models of a shear building and a steel moment-resisting frame, and a 3D six-story steel moment frame with stiffness degradation of different elements	[135]
	1	1	1	8-DOF mass-spring system and 3-story frame structure with linear and nonlinear damages	[136]
	2	3	1	Laboratory footbridge model with loading and unloading of two half-full tanks on the bridge	[137]

Note: Values in columns of Input Data, Excitation, and Monitoring refer to the number of items defined in Table 3. For instance, 1 in Input data column refers to random excitation, which is the first type of excitation in Table III. Ref is the reference

TABLE V
SUMMARY OF DATA-DRIVEN METHODS BASED ON BAYESIAN THEORY

Bayesian-based methods	Input Data	Excit-ation	Monitoring Objectives	Implementation	Ref
Bayesian (belief) network	2	NA	6	Strain data prediction for Canton Tower	[98]
	*	NA	5	Condition assessment for UK masonry arch railway bridges	[138]
	6	8	1	Panel with stiffeners in laboratory and damages were simulated with holes in the panel	[139]
Bayesian neural network	1	1	1	A three-story frame structure in a laboratory with added mass and stiffness reduction	[96]
	1	1	1	Tianjin Yonghe Bridge with damages and deterioration	[53]
	1	2, 5	2, 4	Numerical suspension bridge with stiffness reduction of hangers, cables, and beams	[90]
BDT	3	1	2, 4	A numerical model with reduction of the Young's Modulus and real bridge application with a strengthening of a steel bridge	[107]
Bayesian inference	2	3	5, 6	Real-world monitoring data acquired from a cable-stayed bridge	[140]
	2	2	5	Stress responses generated from a three-dimensional finite element model of an I-10 Twin Span Bridge	[141]
	6	8	6	Experimental datasets from several lab tests are used to verify the proposed framework	[142]

Note: NA denotes that the information is not available; * denotes inspection results are used as input data

the performance of BDT, neural network, and SVM in the structural damage identification. Both raw dynamic responses and processed information (frequencies and mode shapes) were used as input to test the statistical models. Bayesian inference is widely used to quantify uncertainty in data-driven methods in SHM. Ni et al. [140] used Bayesian regression model to interpret the relationship between expansion joint and temperature. Yu et al. [141] used Bayesian inference to predict the future extreme load effect, which was subsequently used in the proposed bridge condition assessment framework. Gobbato et al. [142] proposed a damage prognosis framework for adhesively-bonded joints based on Bayesian inference of nondestructive evaluation (NDE) inspection results.

Bayesian techniques offer advantages in probabilistic quantification of model uncertainty, enabling a better understanding of the confidence levels associated with structural health predictions. They also allow for the integration of prior knowledge or expertise about the system, which can improve the predictive accuracy and reliability of SHM. Those advantages are especially beneficial when data are multi-source but the volume is limited or contain to much noises. Meanwhile, Bayesian techniques rely on the assumed models for data and prior distributions. If these models are not representative of the underlying process, the resulting predictions and inferences can be inaccurate. Besides, they can be computationally intensive since they are designed to generate distribution of variables instead of constant values. This may limit their applicability in real-time monitoring or with limited computational resources.

C. *Multivariate Analysis*

Multivariate analysis (MA) helps in processing high-dimensional data and analyzing correlations and mutual inference between different variables. In existing SHM systems, it is extensively used in analyzing fluctuation of measured structural responses with respect to environmental conditions, damages, and sensor faults. Then, the identified correlations are used for predictive maintenance through clustering and outlier-based anomaly detection.

The most widely used MA models include PCA, CCA, ICA, multivariate regression, and cluster analysis. Their application in data-driven monitoring for engineering structures is summarized in Table VI. PCA searches an orthogonal space from the multivariate observation matrix to transform variates to a new coordinate system with no cross-correlations [144]. It has been utilized to extract DSF [55], [83], [145], compress data volume, and eliminate environmental interference [146] from data. ICA is initially proposed for blind source separation by separating independent multivariate signal vectors that don't follow Gaussian distribution from mixed signals. In SHM, ICA has been used for system identification [147] and damage identification [148]. Huang et al. [62] used a modified dynamic ICA (DICA) to extract components sensitive to anomalies and detect anomalies for both numerical structures and a real-world cable-stayed bridge after training with acceleration and strain data, respectively. CCA works for finding collections of linear combinations of pairs of multivariate data vectors that are maximally correlated with each other and uncorrelated with other pairs [149]. Bhowmik et al. [86] used CCA to obtain vibration modes of structures in real-time.

Similar to other MA models, multivariate regression-based approaches also consists in measuring the variations in operational and environmental actions along with the structural responses and then establishing regression models between them [150]. Henriques et al. [73] used multiple linear regression (MLR) to analyze variations in the crest displacements with temperature and reservoir level of Alto Rabagão dam in Portugal. The deviations of those variations generated from MLR can be used to alert possible damage in the dam. Mata et al. [74] also used linear regression to analyze data collected from a dam, Alto Lindoso dam in Portugal. Instead of raw measurement data, the time-frequency features extracted from displacement and air temperature are used in the regression analysis. Cluster analysis targets to arrange data into automatically generated groups instead of predefined groups. Cury et al. [63] used several symbolic clustering methods to cluster both raw acceleration and modal data of a railway bridge in France. Results show that those symbolic clustering methods achieve more robust results with modal data than raw accelerations. Alamdari [59] used k-means, a clustering algorithm, to identify damaged jack arches for Sydney Harbour Bridge in Australia. Spectral moment, a type of frequency domain feature, is used to form the feature spaces.

Different moving average (MA) models make different assumptions about the underlying data. However, measurement data may not always conform to these assumptions, resulting in uncertain performance of these methods in real-world SHM systems. Consequently, SHM systems typically incorporate multiple multivariate analysis models, which are compared to select the most appropriate candidate model. Gomanducci et al. [55] have used six types of multivariate statistical techniques to learn the correlation between natural frequencies and environmental conditions. The anomaly of structures is identified based on the residual error matrix computed with multivariate statistical techniques. García-Macías [71] has proposed two software solutions that integrate various multivariate statistical techniques for anomaly detection. The two solutions have been used to process acceleration, crack width, and temperature collected from a monumental masonry palace in Italy. Huang et al. [62] also carried out comparative investigations among PCA, ICA, and DICA to detect anomalies. Besides, several different MA models can be utilized jointly for more accurate analysis. Ubertini et al. [70] integrate several multivariate analysis methods to detect damage of a historic structure, San Pietro bell-tower in Italy. A multivariate linear regression (MLR) model is adopted to remove the effects of temperature in identified natural frequencies, and PCA is used to remove the effects of other unmeasured environmental factors.

In conclusion, MA is advantageous in extracting meaningful features from high-dimensional data, which is typical of SHM systems. However, the specific assumptions to measurement data made by MA models, such as linearity, normality, or independence, can restrict their widespread implementation, especially when those assumptions do not hold in real-world SHM applications.

D. Machine Learning Based Methods

1) Classical ML methods: As Table VII implies, the widely employed classical ML methods in SHM include SVM, decision tree, K-nearest neighbor (KNN), and ensemble learning. SVM is one type of supervised learning method. Based on the statistical learning theory, it tries to find the optimal hyperplane for the target data set, aiming at enlarging the margin between different classes as much as possible. In many cases, SVM can successfully distinguish between healthy and damaged conditions by identifying the boundary in high-dimensional feature spaces. SVM has been employed to process various types of data in SHM systems, including traditional modal parameters [82], [151], strains and displacement [99], wavelet transformation of responses [123], guided-wave [123], and TF [117]. Compared with the applications in civil engineering, SVM may obtain more applications in the damage identification for machines, including bearings [152], [153] and gearboxes [154]. However, the normal SVM algorithm can only tackle linear separable conditions. Although, with the aid of kernel methods, SVM can work on linear inseparable conditions. The classification ability of SVM is still limited because of its inherent characteristic.

A decision tree is a sort of diagram that can learn the relationship between features and target outputs through computing their conditional probabilities. Abdallah et al. [65] used a decision tree to detect abnormal operations based on data collected from wind turbines. System information, environmental parameters, and vibration data from wind turbines are used to train the decision tree for the binary classification. Bayane et al. [155] predicted the probability of fatigue failure of a 60-year-old viaduct using a decision tree. Probabilistic models were developed based on codes and strain monitoring data. In addition to predictive maintenance, a decision tree can be helpful design parameters for damage detection systems [156] and predict the vortex-induced vibration mode [18].

Isomap is an Ml model widely used for dimensionality reduction. It incorporates geodesic distance on a weighted graph with the classical multidimensional scaling. Jeong et al. [157] used Isomap to detect the stiffness change of a cantilevered beam in comparison with PCA. They found that Isomap-based damage identification was superior to the PCA-based one. Liu [145] also compared Isomap with other dimensionality reduction methods, including PCA, Laplacian eigenmaps, and autoencoders.

KNN is a simple nonparametric and distance-based method to cluster samples in the feature space. The anomaly of clustering results generated from new coming data gives rise to the alarm of damage. KNN has been proved effective in

TABLE VI
SUMMARY OF DATA-DRIVEN METHODS BASED ON MULTIVARIATE ANALYSIS

Multivariate analysis		Input Data	Excit-ation	Monitoring Objectives	Implementation	Ref
PCA	MLR	3	NA	1	The Infante D. Henrique bridge in Portugal	[55]
		3	1	1	San Pietro bell-tower in Perugia	[70]
	Distance measure	4	4	1	Test 1: a numerical clamped-clamped beam with stiffness reduction; Test 2: the ASCE benchmark model with additional mass, changing braces or loosening bolts	[115]
	Kernel regression	1	2	2, 4	Reduced scale experiment for bridge and vehicle with additional mass	[110]
CCA		1	1	1, 2	2 storey shear building model with different water dispenser and IASC–ASCE benchmark structure	[86]
ICA		1, 2	1	1, 2	A numerical benchmark structure and an actual cable-stayed bridge	[62]
Multivariate regression		2	3	1	Alto Rabagão dam, a concrete dam in Portugal with the crest displacements vary with temperature and reservoir level	[73]
		5	3	1	Alto Lindoso dam, a concrete dam in Portugal with horizontal displacements (radial direction) vary with air temperature	[74]
Cluster analysis		2	4	2	Sydney Harbour Bridge with damaged jack arches	[59]
		1, 3	1	1	A railway bridge with three different stages: before, during, and after strengthening.	[63]

detecting damages using modal parameters identified from a laboratory wooden truss [158] and raw strain data generated in a numerical railway bridge [159]. Ensemble learning refers to a model that combines the prediction of several base learners to improve the performance of classification or prediction. Li et al. [160] compared the performance of ensemble learning (RealAdaBoost) with logistic regression, decision tree, neural network, and SVM in damage detection for a bridge. They found that RealAdaBoost outperformed other classifiers in this task. Ensemble learning algorithms such as random forest and XGBoost were used to detect damages by Huang et al. [83] in the American Society of Civil Engineers (ASCE) benchmark structure, which is a laboratory-scale four-story steel frame. The cross correlation function (CCF) and wavelet packet decomposition (WPD) was used to form the feature space. Ensemble learning achieves higher prediction accuracy in damage type detection compared with SVM.

Classical ML models often have simple architecture supported by clear theoretical basis. Consequently, they can work well with small datasets and are often more interpretable than deep neural networks. These are crucial in real-world SHM projects, as real-world monitoring data are always limited. And engineers need to understand the underlying patterns and relationships to make informed decisions. However, they have limited capacities to capture complex, non-linear relationships in the data, which may lead to suboptimal performance in certain scenarios. Besides, it is recommended to use ensemble learning when the performance of a single classifier is inadequate.Because ensemble learning has been shown to outperform many standalone classifiers in various damage identification scenarios.

2) Neural networks: Neural networks, also known as artificial neural networks (ANN) consists of layers and connections, which simulate the structure of neurons and synapses in the brain of mammals. The development of them is regarded as the beginning of the machine learning discipline [1]. In the last century, researchers began to apply neural networks in the damage identification of SHM [161]. They were continuously studied and used in damage identification and condition evaluation in the past decade. It should be noted that neural networks summarized in this subsection refer to shallow neural networks. Studies on deep neural networks are introduced in the next subsection.

Neural networks are extensively employed to identify damages from measured modal properties [106], DWT coefficients of acceleration responses [120], cross-correlation coefficient of impedance [162], the sum of transmissibility from displacement [114], and sparse coding extracted features from acceleration [163]. They are also used to predict dynamic responses in [53], [89], [136] and their prediction errors are used to evaluate the health condition of engineering structures, such as Tianjin Yonghe Bridge [53].

Compared to classical machine learning models, neural networks can capture more complex, non-linear relationships, which improves their accuracy and generalization in complex tasks. Table VIII demonstrates that neural networks can be applied to both classification and regression problems, making them useful in detecting damage occurrence, extent, and types. However, neural networks, even shallow ones, are often considered ”black boxes” due to their complex structure, which makes it difficult to understand the underlying patterns and relationships. Additionally, neural networks can be more computationally expensive and typically require larger datasets than classical ML models. Furthermore, the feature extraction ability of shallow neural networks is limited by their simple architecture and limited parameters, which means they may not be suitable for tasks involving high-dimensional sequential and spatial data. Those disadvantages can restrict their real-world implementation.

3) Deep learning models: DNN models have more hidden layers and parameters compared to shallow neural networks, which enhances their feature extraction ability and enables them to simulate more complex nonlinear systems. Additionally, the sophisticated structures of different types of DNNs contribute to the feature extraction in the data processing of SHM. Studies have shown that artificial neural networks (ANNs) can simulate structural responses [164], and recurrent neural networks (RNNs) with encoder-decoder architecture can even predict them with attention mechanisms [165]. To some extent, this helps explain why DNNs perform well in data-driven structural health monitoring.

CNNs are distinguished from other DNNs with convolution

TABLE VII
SUMMARY OF DATA-DRIVEN METHODS BASED ON CLASSICAL ML MODELS

ML models	Input Data	Excit-ation	Monitoring Objectives	Implementation	Ref
SVM	3	3	5	Data from Ting Kau Bridge (Hong Kong) were used for analyzing the correlation between temperature and modal properties	[82]
		1	2	Ball-bearing with loose elements and framework numerical model with stiffness reduction	[151]
	2	6	1	A laboratory SDOF mechanical structure with stiffness change	[99]
	5	8	4	A numerical and a real steel beam with cracks of different width	[123]
Decision tree	1	3, 5	1	Data form Lillgrund offshore wind farm comprising 48 wind turbines were used for the classification of normal operation and abnormal vibration	[65]
	2	2	6	Strain from a Crêt de l'Anneau Viaduct were used to predict the probability of fatigue failure	[155]
Isomap	3	1	2	Numerical models for a cantilevered beam with stiffness damage	[157]
KNN	3	1	1	A laboratory wooden truss structure with added point masses as damages	[158]
Ensemble learning	1	NA	1, 2	Acceleration responses of a healthy and a damaged bridge	[160]
	5	1	2, 3	ASEC benchmark structure with loose of stiffness and failed connections	[83]

TABLE VIII
SUMMARY OF DATA-DRIVEN METHODS BASED ON NEURAL NETWORKS

Input Data	Excit-ation	Monitoring Objectives	Implementation	Ref
1	2	1	A numerical single-track railway bridge with damages simulated by removing the bottom flange in a girder section or removing one bracing	[89]
	1	1	Data from real bridge damage with damages	[53]
	1	1	8-DOF mass-spring system and 3-story frame structure with linear and nonlinear damages	[136]
3	NA	1, 2	The 3D FEM model of Tsing Ma Bridge with the damaged bridge deck	[106]
	1	2,4	Numerical bridge model with reduction of Young's modulus and a real steel bridge with strengthening	[107]
4	2	2, 4	Numerical model for Ca-Non Bridge with stiffness decreases in girder	[114]
5	6	1	ASCE benchmark model with broken braces	[120]
1, 3	7	3	A numerical model for a bridge with Young's and shear modulus reduction	[163]
4	8	3, 4	Reduced scale experiment and real bridge with notches and loose bolts	[162]

layers and pooling layers. The joint operation of the two structures works like a filter by eliminating the noises [166] and reserving specific frequency ranges of original data [92]. The latest work on deep learning algorithms in data-driven SHM has been summarized in Table IX. CNNs have been used to identify damage from structural TF [167], acceleration responses [91], [92], [94], [95], [159], spatial-spectral features transformed from raw acceleration [111], strain field [168], electromechanical impedance (EMI) signatures [169], and bridge condition rating data from National Bridge Inventory (NBI) data repository [170]. The aforementioned CNN-based methods were all validated with numerical or laboratory tests. Later, Zhang et al. [61] finished the CNN-based damage identification test in a real steel girder bridge. They found CNN can detect the location of the added mass and stiffness in the bridge with high accuracy. Besides, some researchers also used CNN in anomaly detection during the data collection [171].

RNN is a type of neural network that process sequential data recurrently with one repeated network. RNNs can capture complex temporal patterns and long-range dependencies from variable-length sequences. Therefore, they have been extensively used in SHM for sensor data prediction, structural damage identification, and remaining useful life prediction. Wootton et al. [137], [172] used one type of RNN, echo state networks (ESN), to predict the tilt of a footbridge from temperature force. Robles Urquijo et al. [173] also used long short-term memory (LSTM) to detect the 'no vehicle', 'light vehicle', and 'heavy vehicle' conditions for a highway bridge based on the strain data collected with FBG. Guo et al. [174] chose LSTM to detect bridge conditions. Previous deflection data collected from the girder were used to predict the present deflection value, and the prediction error could depict the health condition. Zhou et al. [175] demonstrated the feasibility of RNN in solving nonlinear inverse problems. Acceleration time-history data and impact load were the output of RNN. Given the high prediction accuracy of RNN in this inverse problem, it is reasonable to expect further application in damage identification, which is another inverse problem.

In the past five years, graph neural networks (GNNs) have been introduced to SHM applications because data collected from multiple sensors in an engineering structure naturally forms a graph. And GNNs are designed to extract features from those graph-shaped data. In other ML models such as CNNs and RNNs that only work with grid-shaped data (matrix), graph-shaped data must be transformed into a matrix before feeding into those models. This transformation may distort or neglect graph-shaped features, leading to a decrease in the accuracy of data-driven methods. As a versatile non-Euclidean machine learning algorithm designed specifically for graphs, GNNs are more suitable for learning from graph-shaped data in SHM systems and have been actively studied in recent years. Tsialiamanis et al. [176] used GNN to predict the natural frequencies of a population of numerical truss structures and use the predicted frequencies to detect the existence of damages. Zhou et al. [177] trained graph convolutional networks (GCNs) to detect bolt losses. The input graph of GCN is electro-mechanical impedance (EMI) readings from PZT sensor networks, and the target node outputs are torque losses of bolts. The proposed GNN model achieved smaller prediction error than all baseline models. Son et al. [178] used a message passing neural network (MPNN) to learn the mapping from cables' tension to the decrease of cables' section areas. Similarly, Li et al. [85] also used cable forces to train a spatiotemporal graph convolutional network (STGCN). However, STGCN is trained with sequences of cable forces in a period of time to predict future cable forces instead of predicting damages directly. The adjacent matrix in STGCN consists of learnable parameters to automatically capture correlations between nodes that are locally connected. STGCN is also utilized by Bloemheuvel et al. [54] for strain prediction of a real-world highway bridge. But this work focus on comparing different graph generation methods for strain prediction instead of damage identification.

Autoencoder is another type of DNN that has been applied in SHM as well. The damage extent and location identification accuracy of deep autoencoder [179] and deep sparse autoencoder [180] methods were also proved with both numerical models and laboratory experiments. Rafiei and Adeli [181] proposed a structural condition assessment method with a deep Boltzmann machine (DBM). After being exposed to 4 levels of earthquake excitation, the structure's overall conditions were classified into four categories accordingly. The reduced scale experiments for a 42-story building proved the performance of the proposed method in evaluating the health indices for both substructures and the entire structure. Compared with neural networks, DNNs have better feature extraction ability. DNN consistently achieves higher prediction accuracy than neural networks with the same datasets. However, as a compute-intensive algorithm, DNNs have higher requirements on computation resources. Real-time monitoring with DNNs has been one of the biggest challenges due to the limited computation resources in existing SHM systems.

To sum up, CNNs are particularly effective at extracting local spatial features from grid-like data structures. However, they may not perform well on non-grid data structures or when the input data lacks spatial locality. RNNs are designed to capture long-range dependencies from sequential data, making them well-suited for time-series analysis. Nevertheless, RNNs can suffer from the vanishing or exploding gradient problem, making it challenging to train deep RNNs. GNNs can handle irregular data structures, making them suitable for SHM tasks where the data is represented as a graph or network. But they are not well-suited for handling grid-like data structures or purely sequential data. Although DNNs have demonstrated impressive performance, they typically require a large amount of labeled training data, which is often unavailable in most SHM projects. Moreover, their interpretability are even lower than shallow neural networks. Those disadvantages limit their implementation in real-world SHM systems.

E. Hybrid Statistical Models

Some data-driven methods integrate various statistical models into a pipeline for sequential feature extraction and mapping. Due to the significance of each model in the pipeline, it would be unfair to categorize such integrated methods into any of the statistical models previously mentioned. In this paper, we refer to these data-driven solutions as hybrid statistical models. Table X summarizes typical combinations of statistical models used in these hybrid approaches.

Dimensionality reduction+Classifier/Regression analysis: PCA is a widely used dimensionality reduction method in SHM. It significantly reduces computation costs and improves the generalization ability of damage identification methods. Vitola et al. [145] used PCA to preprocess dynamic responses from different sensors on aluminum and composite plates with PZT sensors. They then classified the responses into different clusters corresponding to different damages using KNN. Additionally, PCA has been combined with various other feature extraction or mapping methods, such as self-organizing maps [22] and sparse coding [112].

Many other dimensionality reduction methods have also been applied in SHM algorithms, including Isomap [167], Random Projection [182], autoencoder [167], and Laplacian Eigenmaps [167]. Liu et al. [167] analyzed the performance of four different dimensionality reduction methods in a data-driven SHM method based on acceleration responses from a bridge and a vehicle. It was observed that non-linear and non-convex dimensionality reduction methods (e.g., stacked autoencoder) achieved the best identification accuracy. Different applications of dimensionality reduction methods in data-driven SHM are summarized in Table X. Dimensionality reduction methods can eliminate noises or non-primary information. Besides, it can be used as preprocessing method in hybrid statistical models for improving the robustness of data-driven methods and reducing the computation.

Time series analysis + SVM/PCA: Time series analysis models are used to predict time-history data. More details on these models will be illustrated in Section 4.1. Parameters and residual errors (RE) of these analysis models can be used as DSFs. Gui et al. [185] selected AR parameters and RE in autoregressive model as the DSFs. Kim et al. [122] also choose AR parameters of the responses to form the DSFs. Datteo et al. [64] analyzed the data (from August 13th, 2015 to April 19th, 2016) of a stadium and used AR parameters of acceleration data DSF. The results indicated that the principal component analysis (PCA) method could classify different responses of the stadium under different events like empty, hosting concerts, or hosting football games.

Based on existing research, we recommend selecting appropriate statistical models based on the data, monitoring objectives, and characteristics of the monitored structures. Table XI summarizes the recommended scenarios for each statistical model.

VI. Implementation Challenges of Data-Driven Algorithms

After reviewing existing literature, we found that the assumptions and premise of many proposed data-driven algorithms might not aligned with engineering practices. Besides, those algorithms have not been tested with real measurement data or in real engineering structures. Due to their requirements in complete observation of the structure, a stable surrounding environment, and sufficient computation

TABLE IX
SUMMARY OF DATA-DRIVEN METHODS BASED ON DEEP LEARNING

DL models	Input Data	Excit-ation	Monitoring Objectives	Implementation	Ref
CNN	4	1	2	Numerical four-story frame structure with stiffness reduction	[167]
	1	6	3	Numerical steel frame buildings with reduced beam sections	[91]
	1	1	2, 4	A numerical beam structure with a decrease in cross-section area	[92]
	1	1	4	ASCE benchmark experimental data wherein damages were simulated with the removal of the diagonal braces, loosen bolts	[95]
	1	1	2, 4	Reduced scale steel frame with loosen bolts	[94]
	4	5	2, 4	Numerical arch bridge and damages are simulated with a decreased cross-section area	[111]
	1	4	2	A laboratory model of a short steel girder bridge with additional mass and a real long steel girder bridge (in-service) with additional mass and steel plate.	[61]
	6	8	2	Laboratory aluminum plate with an added metallic nut	[169]
	2	7	2	FEM of two channels welded with steel plate and damages are simulated with cracks	[168]
	*	NA	5	CNN is trained and tested on National Bridge Inventory (NBI) database from the Year 1992 to 2017 by the Federal Highway Administration (FHWA) in the United States	[170]
	2	2	2,4	A numerical model of a steel railway truss bridge and damages are simulated with reduced cross-sectional areas of elements	[159]
RNN	2	3	1	Laboratory footbridge model with loading and unloading of two half-full tanks on the bridge	[137], [172]
	2	2	5	Strain data from FBG sensors in a highway bridge since 2000	[173]
	2	NA	1	A real bridge in Hubei Province, China	[174]
	1	4	5	Numerical and lab experiments to identify impact load history	[175]
GNN	3	NA	1	Numerical two-dimensional trusses to predict the natural frequencies under different temperature	[176]
	6	8	2,4	A twin-bolt plate with multiple bolt losses	[177]
	2	7	2,3	Numerical bridges and damages are simulated by reducing the cross-sectional areas of cables	[178]
	2	NA	2	A cable-stayed bridge in China. The cable forces are monitored for damage identification	[85]
	2	NA	NA	A large highway bridge in the Netherlands after retrofits.	[54]
Deep sparse autoencoder	3	4	2, 4	Numerical steel frame with element stiffness reduction and laboratory concrete bridge with cracks	[180]
Deep autoencoder	3	4	2, 4	Numerical and laboratory steel frame structures with element stiffness reduction	[179]
DBM	4	1	5	Reduced scale laboratory experiment for a 42-story concrete high-rise building	[181]

Note: * denotes bridge condition rating data

TABLE X
SUMMARY FOR HYBRID STATISTICAL MODELS

Hybrid Statistical Models		Input Data	Excit-ation	Monitoring Objectives	Implementation	Ref
PCA+	Selforganizing maps	2	7	1	Laboratory experiment for wing section with skin cutting & cracks	[22], [97]
	Sparse coding	4	4	2, 4	Numerical two-span two-story frame with column stiffness reduction	[112]
	KNN	6	8	2	Damage is simulated with added mass. Experiments include: Aluminum Rectangular Profile with four damage locations, Aluminum Plate with three damage locations, and Composite Plate with three damage locations	[145]
Random Projection+	Bayesian	3, 4	NA	1	Benchmark dataset from a three-story building structure with gaps and added mass, Z24 bridge with damage introduced deliberately	[182]
Stacked autoencoder+	Linear regression	1	2	4	Numerical model & lab bridge experiment with concentrated mass at mid-span	[183]
Time series+	SVM	5	6	1	Three-story laboratory building with degradation of floor stiffness	[122]
		1	1	2	Laboratory eight-DOF mass-spring system with nonlinear damages simulated with installed bumper	[184]
		1	1	1	Frame structure with the adjustment of the gap between the bumper and column	[185]
	PCA	1	2	5	Real stadium with different events	[64]

resources, their accuracy and robustness can degrade in real-world monitoring projects. In this section, we first summarize the implementation challenges of data-driven algorithms and then propose our solutions and insights for future directions.

A. Challenges in Monitoring Algorithms

The primary implementation challenge of data-driven monitoring algorithms comes from the field collected data. Unlike model-based monitoring methods, data-driven methods are more vulnerable to the quality and quantity of data. Nevertheless, data collected from real structures are always limited, unlabelled, and unbalanced. Consequently, it is incredibly challenging to implement data-driven algorithms in real-world projects.

Firstly, data-driven methods require data from both intact and damaged conditions to generate the feature space, from which they can automatically capture the relationship between input data and DSFs. However, most engineering structures have never encountered failures. Data from structures with defects are much less than data from intact structures. Therefore, data-driven methods trained with imbalanced data might not be able to make accurate predictions. Secondly, most samples from structures are unlabeled because the health conditions of an actual engineering structure are always unknown. In

TABLE XI
RECOMMEND APPLICABLE SCENARIOS OF STATISTICAL MODELS IN DATA-DRIVEN SHM

Statistical models		Computational complexity	Interpritability	Recommended applicable scenarios
Time series analysis		Low	High	Analyse patterns and trends for time sequences collected from structures situated in relatively stable environmental conditions
Bayesian theory		/	High	Predict the likelihood of damages and make optimal decision under dynamic environment with high uncertainty
Multivariate analysis		Medium	High	Extract correlations and mutual inference between different variables
Classical ML methods		Medium	Medium	Accomplish simple classification, clustering, and regression tasks based on datasets with small quantities of data
Neural networks		Medium	Low	Identify non-linear relationships between variables when computation resources is limited
Deep learning models	RNN	High	Low	Learn sequential patterns from high-dimensional time sequences
	CNN	High	Low	Learn spatial patterns from high-dimensional graph shaped data
	GNN	High	Low	Learn spatial patterns from high-dimensional grid-like data
Hybrid statistical models		/	/	Applicable for SHM projects with large and diverse datasets, complex data relationships, and varied data types.

Note: / denotes the computational complexity or interpritability is unevaluable;

contrast, samples generated from numerical simulation can be labeled. However, unexpected errors might be introduced because of the gap between real structures and numerical simulation. Those unlabeled data make it quite challenging to train supervised data-driven methods. Although some data-driven methods can detect the degradation of engineering structures in an unsupervised manner, they can only detect the existence of degradation instead of the extent of degradation. Thirdly, the field monitoring data is limited with respect to the variable environmental conditions. A statistical model usually fails to make an accurate prediction when the environmental conditions are not seen in the training samples. So, data-driven approaches might not be robust to the unseen environment where the engineering structures are operating [186].

Moreover, most data-driven monitoring approaches operate like a "black box," which lacks physical interpretation, and their performance in real projects is unreliable. When damages are misidentified, it is difficult to find out where the mistake comes from and explain why it happens. It is even harder to update the data-driven algorithms if we want to prevent the recurrence of similar mistakes.

B. Challenges in Predictive Maintenance Algorithms

As clarified in section III, damage prognosis represents the future trend of health monitoring for engineering structures. It enables the planning of regular maintenance, guiding of retrofits, and ultimately prolonging the service life of structures. However, it is the most challenging objective among the six monitoring objectives since it requires reliance on both current monitoring results and predictions of future behaviors. The summary of existing studies in section V also supports this argument. Only a small fraction of existing data-driven methods are proposed to achieve objective 6, which refers to damage prognosis in Table III. Instead, most methods are designed to detect the existence and extent of one specific type of damage, remaining in the first several levels for predictive maintenance.

In addition to the challenges similar to data-driven monitoring, the implementation challenges of data-driven predictive maintenance also involve the following aspects. Firstly, it places higher demands on the input data and statistical models since predicting future failure and RUL requires the exact degree of damages concerning the engineering structures' functional failure. Secondly, most real-world engineering structures have large spatial scales and multiple complicated substructures, leading to complex failure modes. As a result, it is nearly impossible to fully understand the precise physics of failure [187]. Thirdly, damage prognosis is performed on top of predictions of structures' behavior and environmental conditions. However, environmental factors are volatile, and the degradation of structures is usually nonlinear and unexpected, increasing the prediction errors in future damages. Fortunately, data-driven methods are appropriate for making predictions under great uncertainty, especially when the mechanism of engineering structures' (like composite structures) damage evolution is not fully understood [188]. Through incorporating statistical models (such as Non-Homogenous Hidden Semi Markov Model (NHHSMM) [188], Gaussian process-based predictive model [189], Bayesian Neural Network [129], etc.), data-driven methods can predict the RUL for engineering structures.

C. Possible Solutions: Fusion Methods

Fusion methods have been widely studied in the broad sense of SHM [190]. They represent a promising way of bringing together the strengths of different data sources and models. Consequently, they lead to a complete image of the present damage, reduce ambiguity, and increase confidence in the results [191]. As Fig. 6 shows, the fusion operations can occur during raw data collection [191], feature extraction [107], [182], model updating [192], [193], and decision making [138], [194]. Unlike existing frameworks that only include data, feature, and decision-level fusion, we include model-level fusion in our fusion methods and will clarify it with fined details later.

Data-level fusion refers to the integration of raw data from multiple sensors [191], [195], [196]. Those sensors measure the same physical quantities, and the physical meaning of measured raw data is the same. Data-level fusion helps enlarge the source of information and reduce the ambiguity in the further data processing.

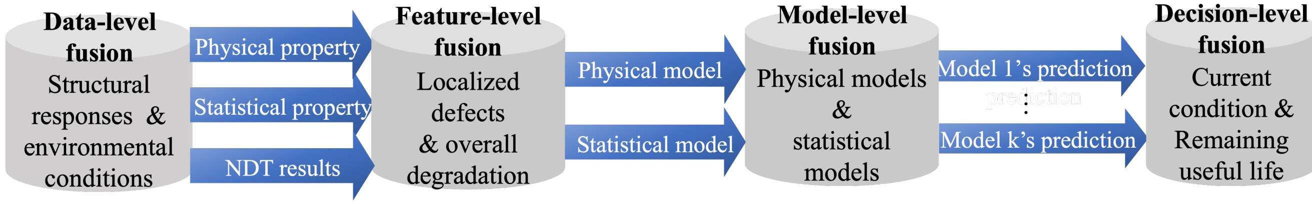

Fig. 6. Fusion methods for the monitoring and predictive maintenance of engineering structures.

Feature-level fusion aims at combining features extracted from heterogeneous data or from the same data but with different data featuring methods. Different features describe the properties of structures from different aspects. For instance, physical and statistical properties from raw data can indicate the overall health conditions, while the NDT results can reflect the localized defects. Frequency domain features like frequencies can describe the vibration patterns, while time domain features reveal the amplitude variation of data. The integration of those macro and micro features helps draw an exhaustive and precise condition assessment of the whole structure. Dang et al. [197] integrate features in time and frequency domains from acceleration data, demonstrating that the feature-level fusion method can achieve highly accurate damage detection with lower time and space complexity. So, the proposed method is more practical for real-world engineering projects.

Model-level fusion combines statistical models in data-driven methods and physical models in model-based methods. It makes full use of available resources and achieves higher accuracy when neither precise physical models nor sufficient measurement data are available in the monitoring of large/complex engineering structures. A model-level fusion method use known physical properties of engineering structures to compensate for the insufficient measurement data. Meanwhile, it also improves the usability of data-driven methods since their reliance on data is much reduced. Physics-guided ML models [198], sometimes also called physics-informed/enhanced ML models [199], [200], belong to model-level fusion methods. They embed ML models with given prior knowledge of the physical properties of the corresponding physical systems. Knowledge in several formats is utilized, including physics-guided loss function, architecture design inspired by physical systems, and parameter values in physical systems. Guided by prior physical knowledge, the searching spaces for parameter updating are dramatically reduced, and fewer data are required for model training. Moreover, ML models are less prone to make physically implausible predictions owing to the inductive bias [201] introduced by physical systems. In decision-level fusion methods, predictions generated by different methods are integrated to form the final decision through particular fusion rules. How to aggregate those predictions and make robust decisions remains an open research issue [190].

In summary, regardless of where the fusion occurs, a fusion method can always achieve higher accuracy and greater generalization ability. At times, it can also reduce time and space complexity by integrating multiple data or features simultaneously in a more efficient manner. Considering these advantages, we believe that the four levels of fusion represent a promising direction for the implementation of data-driven methods.

VII. Implementation Challenges Of Current SHM Systems

Current SHM systems are alarmingly inefficient in supporting the execution of complex data analytic algorithms under acceptable performance. SHM systems in real-world engineering structures are required to be responsive to anomalies and robust to node failures. Nevertheless, unlike the prosperity of research on data-driven algorithms, much less attention has been paid to optimizing the operational performance of SHM monitoring systems, such as latency, robustness, and data security issues. As Table II implies, most real-world data-driven monitoring systems don't support real-time data processing. Instead, data are first stored in databases and then processed offline once users send the request. That can be risky for engineering structures because damage cannot be identified instantly. To this end, we put forward promising solutions and future directions to promote the implementation of online monitoring systems.

A. *Deficiencies of Current Monitoring Systems*

Fig. 7 shows a currently adopted cloud-based monitoring system, in which a server/station is deployed either near the structure (private cloud) or on a remote server on the internet (public cloud). Data are transmitted to and stored in a centralized server, where SHM applications are provided through web services. The robustness of this system is limited and data are not well protected, hindering the wide implementation of such a system. Firstly,a cloud-based monitoring system consist of numerous sensors. Those sensors inevitably suffer from aging or external impacts in outdoor environments, such as strong wind, heavy rainfall, and even earthquake. However, the centralized computing paradigm adopted by cloud-based SHM systems is vulnerable to node failures. They might not work normally in a fierce environment but the engineering structure should be monitored more closely during that period. Secondly, data collected by SHM systems are valuable because they reveal the structure's operational condition. They belong to the structure's governing body and their leakage can be dangerous to the normal operation of the structure. Nevertheless, similar to most IoT applications, data in cloud-based SHM systems are natively vulnerable to attacks because data

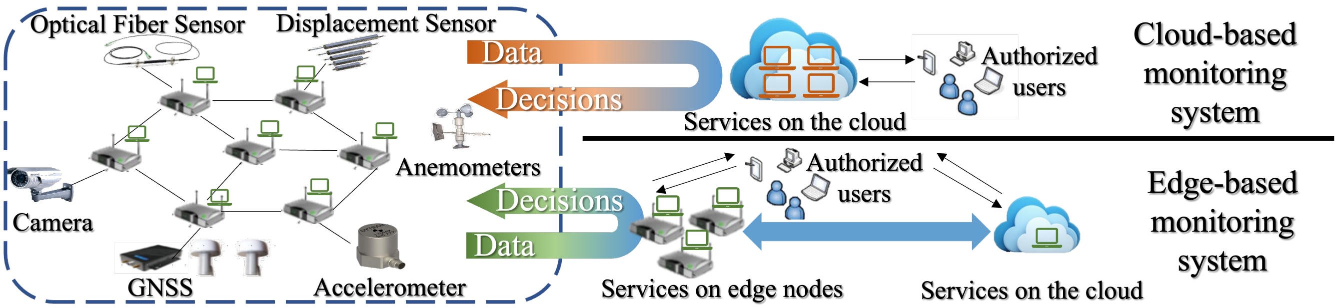

Fig. 7. SHM system based on cloud computing and edge computing.

are usually transferred over global network before processing. There are many potential security holes to be fixed but little attention has been paid to data security issues in SHM.

Reducing the latency of a monitoring system is also challenging. The latency is usually related to data transmission efficiency, the server's computation resources, and the data processing algorithms' complexity. Systems based on remote cloud servers have abundant and elastic computation capacities. But they suffer from high data transmission latency because the data volume is enormous large and the transmission distance is enormous long. Systems based on local base stations or private cloud servers have much-reduced data transmission latency and increased privacy. But their computation resources are usually limited and fixed. They might be inefficient or too expensive to execute complex data-driven algorithms.

Moreover, current data-driven algorithms are usually computationally demanding for SHM systems, especially for those based on structural modal properties, frequency domain features, and time-frequency domain features. For instance, the extraction of modal properties usually relies on system identification methods, including the eigensystem realization algorithm [202] and subspace system identification [203]. They all involve complex operations on large-scale matrixes and have high time and space complexity. The recently proposed DL-based solutions, such as CNN, RNN, and GNN, are computationally demanding as well. They even require extensive or specific hardware for computation acceleration, e.g., high-performance GPU and TPU. Otherwise, the data processing time can be extremely long, increasing the risk when damages strike engineering structures.

Developing and implementing real-time monitoring systems is much more expensive than numerical simulation. They are more demanding in the efficiency, stability, and data security of the system. However, few researchers study it due to its miscellaneous programming tasks concerning limited research value. Few developers are willing to develop specialized tools for it considering its relatively small market and the lack of a mature monetization mode. Till now, SHM systems are designed only for a few crucial engineering structures and managed by governments for public safety instead of commercial use.

B. *Possible Solutions*

Improvements to SHM systems are desperately needed to mitigate the above deficiencies. Wireless sensor networks (WSN) [204], [205] are used first and edge computing [206], [207] has been introduced to SHM recently. In SHM, researchers use WSN to reduce cost and improve efficiency when managing a large number of sensors. Edge computing can reduce the latency dramatically by migrating data computation and storage from a cloud server to distributed edge devices. The overall system's robustness is much improved due to distributed computation resources. Data security is improved as well because raw data can be encrypted and authenticated key establishment [208] can be performed in edge devices. Although resource congestion might also happen in edge clusters since the distributed computation and storage resources are relatively weaker than the centralized resources in cloud servers. The possible solutions include developing high-performance edge clusters and optimizing the distributed data processing algorithms. However, high-performance edge clusters also occupy large spaces and require an enormous power supply, making them inconvenient in implementation. Consequently, reducing the delay of data processing algorithms under constraint computation resources has been a promising research direction.

However, research on optimizing the complexity or resource demand of SHM algorithms has just started. Liu et al. [205] adopted distributed modal analysis approaches in WSN-based SHM to reduce requirements on computational resources. Wu et al. [209] used transfer learning and network pruning technique to deploy DCNN-based defects detection on edge devices. With limited computation resources on edge nodes, the proposed model compression method achieved high accuracy and acceptable time cost for crack detection and corrosion detection. Chen et al. [207] have designed a lossy data compression algorithm for edge devices in SHM systems. It helps reduce the latency for data transmission analytics. Moreover, the proposed physics-enhanced PCA is proven to preserve modal properties for acceleration responses with much reduced time and space complexity. Testoni et al. [210] designed a sensor network system where acceleration sensor nodes could acquire and reprocess data such as frequency spectrum pick extraction. Besides, the data and power are exchanged in the same bus based on data-over-power (DoP) communication. The real-time requirements of

SHM applications are subsequently fulfilled. Still, the system architecture and distributed algorithms need further optimization. Fortunately, the ever-increasing number of heterogeneous IoT devices (such as smartphones, edge gateways, and multi-access edge computing servers) in Smart City infrastructures can provide ubiquitous computing resources, enabling future real-time health monitoring and predictive maintenance.

Furthermore, a standalone SHM system might only contribute partially to the comprehensive assessment of the overall health conditions due to the absence of various environmental information. A "Smart City" in the future must constitute multiple intelligent systems except for SHM systems. Intelligent transportation systems (ITS) [211], and Automated driving systems might help locate and identify vehicles, which are subsequently used to identify the exact excitations on civil engineering structures like bridges. A GPS [212] can be implemented in SHM to monitor structure displacement. Integrated with a meteorological monitoring system [213], an SHM system might be able to prepare for extreme weather in the future.

VIII. Conclusion

This survey provides an overview of the latest studies on data-driven monitoring and predictive maintenance for engineering structures from an implementation perspective. Firstly, we introduce the system architecture and typical procedures of data-driven solutions with detailed explanations. Then, we classify the data and statistical models used in data-driven methods, clarify the advantages and limitations of different models, and discuss their implementation in real-world projects. We propose solutions to address two major challenges in implementing existing data-driven methods, including the digressive performance in the real-world environment and inefficient computing systems for real-time data analytics. Finally, to provide guidance for future research on data-driven solutions, we summarize our insights into the future trends as follows:

- From single statistical model to hybrid statistical model: Combining statistical models can meet multiple requirements for SHM tasks, such as feature extraction, noise elimination, dimensionality reduction, feature mapping, and damage prognosis, making them more robust for real-world data analytics and warranting further investigation.
- From private datasets to public datasets with standard objects: Public datasets with standard objects and evaluation metrics can advance SHM research as they enable researchers to identify state-of-the-art algorithms with higher efficiency and accuracy. Existing data-driven SHM methods are predominantly evaluated using private datasets collected from specific structures, making it difficult to compare their performance.
- From pure data-driven to physics-enhanced data-driven methods: Physical models of engineering structures have been well studied and extensively validated for decades. Integrating those physical principles, laws, and theorems into statistical models provides a promising way to find robust and generalized DSFs.
- From centralized remote server to distributed nearby edge devices: Edge-based and WSN-based solutions are promising for achieving real-time SHM since they rely on distributed computing, and algorithms distributed among local devices generate less delay in data processing and transmission.

Limitations in existing ICT also restrict the usability of data-driven SHM. We summarize following noteworthy research directions in ICT to push the horizon of the extensive implementation of data-driven SHM.

- Improve the accuracy and durability of monitoring systems. Data-driven methods heavily rely on field-collected data, and long-term and precise monitoring data can help these methods capture long-term patterns and achieve higher accuracy. Therefore, there is an urgent need to improve the accuracy of existing sensors, develop new sensors, and enhance systems' durability to environmental conditions in data-driven methods.
- Optimize current wireless communication solutions. While wireless monitoring systems are more convenient to install and maintain, they are rarely chosen due to the instability and inefficiency of existing wireless communication solutions or their high energy consumption. Therefore, improvements in existing ICT (e.g., 5G) are necessary for the wider use of wireless monitoring systems in the future.
- Improve the numerical simulation technology of engineering structures. Real-world measurement data are always imbalanced and unlabeled, since most engineering structures operates normally during their service period. Researchers always need to generate damage scenarios use physical models or laboratory experiments. Consequently, developing fine-grained numerical simulation technology (i.e. DT) for engineering structures can significantly contribute to data-driven methods.
- Enhance the interpretability of DL algorithms. While DL algorithms excel in feature extraction ability compared to other statistical methods, their lack of interpretability severely hinders their development. Enhancing their interpretability is a guaranteed path to their extensive implementation.

Acknowledgments

The authors are grateful for the financial supports from the Key-Area Research and Development Program of Guangdong Province (2019B111106001), the Research Grants Council of Hong Kong through the Theme-based Research Scheme (T22-502/18-R), the Innovation and Technology Commission of Hong Kong through Smart Railway Technology and Applications (No. K-BBY1), the Hong Kong Polytechnic University through the Strategic Importance Project (No. ZE2L), and the Research Institute for Artificial Intelligence of Things, the Hong Kong Polytechnic University. The findings and opinions expressed in this paper are from the authors alone and are not necessarily the views of the sponsors.

REFERENCES

[1] C. R. Farrar and K. Worden, *Structural health monitoring: a machine learning perspective*. John Wiley & Sons, 2012.

[2] H.-N. Li, T.-H. Yi, L. Ren, D.-S. Li, and L.-S. Huo, "Reviews on innovations and applications in structural health monitoring for infrastructures," *Structural Monitoring and Maintenance*, vol. 1, no. 1, pp. 1–45, 2014.

[3] W. Ostachowicz and A. Güemes, *New trends in structural health monitoring*. Springer Science & Business Media, 2013, vol. 542.

[4] C. A. Tokognon, B. Gao, G. Y. Tian, and Y. Yan, "Structural health monitoring framework based on internet of things: A survey," *IEEE Internet of Things Journal*, vol. 4, no. 3, pp. 619–635, 2017.

[5] I. Takewaki, M. Nakamura, M. Nakamura, and S. Yoshitomi, *System identification for structural health monitoring*. WIT Press, 2012.

[6] R. J. Barthorpe, "On model-and data-based approaches to structural health monitoring," Ph.D. dissertation, University of Sheffield, 2010.

[7] J. Luo, M. Namburu, K. Pattipati, L. Qiao, M. Kawamoto, and S. Chigusa, "Model-based prognostic techniques [maintenance applications]," in *Proceedings AUTOTESTCON 2003. IEEE Systems Readiness Technology Conference.* Ieee, 2003, pp. 330–340.

[8] D. An, N. H. Kim, and J.-H. Choi, "Practical options for selecting data-driven or physics-based prognostics algorithms with reviews," *Reliability Engineering &amp System Safety*, vol. 133, pp. 223–236, jan 2015.

[9] A. Malekloo, E. Ozer, M. AlHamaydeh, and M. Girolami, "Machine learning and structural health monitoring overview with emerging technology and high-dimensional data source highlights," *Structural Health Monitoring*, vol. 21, no. 4, pp. 1906–1955, aug 2021.

[10] R. Hou and Y. Xia, "Review on the new development of vibration-based damage identification for civil engineering structures: 2010–2019," *Journal of Sound and Vibration*, vol. 491, p. 115741, jan 2021.

[11] L. Sun, Z. Shang, Y. Xia, S. Bhowmick, and S. Nagarajaiah, "Review of bridge structural health monitoring aided by big data and artificial intelligence: From condition assessment to damage detection," *Journal of Structural Engineering*, vol. 146, no. 5, may 2020.

[12] G. Toh and J. Park, "Review of vibration-based structural health monitoring using deep learning," *Applied Sciences*, vol. 10, no. 5, p. 1680, mar 2020.

[13] M. Flah, I. Nunez, W. B. Chaabene, and M. L. Nehdi, "Machine learning algorithms in civil structural health monitoring: A systematic review," *Archives of Computational Methods in Engineering*, vol. 28, no. 4, pp. 2621–2643, jul 2020.

[14] D. A. T. Burgos, R. C. G. Vargas, C. Pedraza, D. Agis, and F. Pozo, "Damage identification in structural health monitoring: A brief review from its implementation to the use of data-driven applications," *Sensors*, vol. 20, no. 3, p. 733, jan 2020.

[15] M. Azimi, A. Eslamlou, and G. Pekcan, "Data-driven structural health monitoring and damage detection through deep learning: State-of-the-art review," *Sensors*, vol. 20, no. 10, p. 2778, may 2020.

[16] M. Martinez-Luengo, A. Kolios, and L. Wang, "Structural health monitoring of offshore wind turbines: A review through the statistical pattern recognition paradigm," *Renewable and Sustainable Energy Reviews*, vol. 64, pp. 91–105, oct 2016.

[17] P. Cawley, "Structural health monitoring: Closing the gap between research and industrial deployment," *Structural Health Monitoring*, vol. 17, no. 5, pp. 1225–1244, 2018.

[18] S. Li, S. Laima, and H. Li, "Data-driven modeling of vortex-induced vibration of a long-span suspension bridge using decision tree learning and support vector regression," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 172, pp. 196–211, jan 2018.

[19] F. Kang, J. Li, and J. Dai, "Prediction of long-term temperature effect in structural health monitoring of concrete dams using support vector machines with jaya optimizer and salp swarm algorithms," *Advances in Engineering Software*, vol. 131, pp. 60–76, may 2019.

[20] X. W. Ye, Y. H. Su, and J. P. Han, "Structural health monitoring of civil infrastructure using optical fiber sensing technology: A comprehensive review," *The Scientific World Journal*, vol. 2014, pp. 1–11, 2014.

[21] S. Das, P. Saha, and S. K. Patro, "Vibration-based damage detection techniques used for health monitoring of structures: a review," *Journal of Civil Structural Health Monitoring*, vol. 6, no. 3, pp. 477–507, apr 2016.

[22] J. Sierra-Perez, M. A. Torres-Arredondo, G. Cabanes, A. Güeme, and L. E. Mujica, "Structural health monitoring by means of strain field pattern recognition on the basis of pca and automatic clustering techniques based on som," *IFAC-PapersOnLine*, vol. 48, no. 28, pp. 987–992, 2015.

[23] H. V. Le and M. Nishio, "Time-series analysis of GPS monitoring data from a long-span bridge considering the global deformation due to air temperature changes," *Journal of Civil Structural Health Monitoring*, vol. 5, no. 4, pp. 415–425, may 2015.

[24] A. B. Mehrabi and S. Farhangdoust, "A laser-based noncontact vibration technique for health monitoring of structural cables: Background, success, and new developments," *Advances in Acoustics and Vibration*, vol. 2018, pp. 1–13, jun 2018.

[25] C.-Z. Dong and F. N. Catbas, "A review of computer vision–based structural health monitoring at local and global levels," *Structural Health Monitoring*, vol. 20, no. 2, pp. 692–743, 2021.

[26] J. Zhang, G. Y. Tian, A. M. Marindra, A. I. Sunny, and A. B. Zhao, "A review of passive rfid tag antenna-based sensors and systems for structural health monitoring applications," *Sensors*, vol. 17, no. 2, p. 265, 2017.

[27] G. Chen, W. Zhang, Z. Zhang, X. Jin, and W. Pang, "A new rosette-like eddy current array sensor with high sensitivity for fatigue defect around bolt hole in SHM," *NDT &amp E International*, vol. 94, pp. 70–78, mar 2018.

[28] W. Na and J. Baek, "A review of the piezoelectric electromechanical impedance based structural health monitoring technique for engineering structures," *Sensors*, vol. 18, no. 5, p. 1307, apr 2018.

[29] J.-M. Munoz-Ferreras, Z. Peng, Y. Tang, R. Gomez-Garcia, D. Liang, and C. Li, "A step forward towards radar sensor networks for structural health monitoring of wind turbines," in *2016 IEEE Radio and Wireless Symposium (RWS).* IEEE, jan 2016.

[30] I. Morris, H. Abdel-Jaber, and B. Glisic, "Quantitative attribute analyses with ground penetrating radar for infrastructure assessments and structural health monitoring," *Sensors*, vol. 19, no. 7, p. 1637, apr 2019.

[31] W. Ostachowicz, R. Soman, and P. Malinowski, "Optimization of sensor placement for structural health monitoring: A review," *Structural Health Monitoring*, vol. 18, no. 3, pp. 963–988, 2019.

[32] M. Abdulkarem, K. Samsudin, F. Z. Rokhani, and M. F. A. Rasid, "Wireless sensor network for structural health monitoring: A contemporary review of technologies, challenges, and future direction," *Structural Health Monitoring*, vol. 19, no. 3, pp. 693–735, jul 2019.

[33] J. Zhang, G. Tian, A. Marindra, A. Sunny, and A. Zhao, "A review of passive RFID tag antenna-based sensors and systems for structural health monitoring applications," *Sensors*, vol. 17, no. 2, p. 265, jan 2017.

[34] J. Yin, Z. Yang, H. Cao, T. Liu, Z. Zhou, and C. Wu, "A survey on bluetooth 5.0 and mesh," *ACM Transactions on Sensor Networks*, vol. 15, no. 3, pp. 1–29, aug 2019.

[35] G. Heo, B. Son, C. Kim, S. Jeon, and J. Jeon, "Development of a wireless unified-maintenance system for the structural health monitoring of civil structures," *Sensors*, vol. 18, no. 5, p. 1485, 2018.

[36] J.-L. Lee, Y.-Y. Tyan, M.-H. Wen, and Y.-W. Wu, "Applying ZigBee wireless sensor and control network for bridge safety monitoring," *Advances in Mechanical Engineering*, vol. 10, no. 7, p. 168781401878739, jul 2018.

[37] C. Li and Z. Cao, "Lora networking techniques for large-scale and long-term iot: A down-to-top survey," *ACM Computing Surveys (CSUR)*, vol. 55, no. 3, pp. 1–36, 2022.

[38] I. Bisio, C. Garibotto, F. Lavagetto, and A. Sciarrone, "A novel IoT-based edge sensing platform for structure health monitoring," in *IEEE INFOCOM 2022 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS).* IEEE, may 2022.

[39] L. Chettri and R. Bera, "A comprehensive survey on internet of things (iot) toward 5g wireless systems," *IEEE Internet of Things Journal*, vol. 7, no. 1, pp. 16–32, 2019.

[40] K. Y. Koo, N. D. Battista, and J. M. Brownjohn, "Shm data management system using mysql database with matlab and web interfaces," in *5th international conference on structural health monitoring of intelligent infrastructure (shmii-5), cancún, méxico*, 2011, pp. 589–596.

[41] Y. Zhang, S. M. O'Connor, G. W. van der Linden, A. Prakash, and J. P. Lynch, "SenStore: A scalable cyberinfrastructure platform for implementation of data-to-decision frameworks for infrastructure health management," *Journal of Computing in Civil Engineering*, vol. 30, no. 5, sep 2016.

[42] H. Dang, M. Tatipamula, and H. X. Nguyen, "Cloud-based digital twinning for structural health monitoring using deep learning," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 6, pp. 3820–3830, jun 2022.

[43] S. Jeong, Y. Zhang, S. O'Connor, J. P. Lynch, H. Sohn, and K. H. Law, "A nosql data management infrastructure for bridge monitoring," *Smart Structures and Systems*, vol. 17, no. 4, pp. 669–690, 2016.

[44] S. Jeong, R. Hou, J. P. Lynch, H. Sohn, and K. H. Law, "A scalable cloud-based cyberinfrastructure platform for bridge monitoring," *Structure and Infrastructure Engineering*, vol. 15, no. 1, pp. 82–102, 2019.

[45] Y. Qin, R. Xiao, Y. Wang, and K. H. Law, "A bridge information modeling framework for model interoperability," in *Computing in Civil Engineering 2019*. American Society of Civil Engineers, jun 2019.

[46] K. Smarsly, D. Hartmann, and K. H. Law, "A computational framework for life-cycle management of wind turbines incorporating structural health monitoring," *Structural Health Monitoring*, vol. 12, no. 4, pp. 359–376, jun 2013.

[47] Y. L. Xu and Y. Xia, *Structural health monitoring of long-span suspension bridges*. CRC Press, 2011.

[48] R. Hou, "Bridge structrural health monitoring using a cyber-physical system framework," Ph.D. dissertation, 2020.

[49] J. Shan, H. Zhang, W. Shi, and X. Lu, "Health monitoring and field-testing of high-rise buildings: A review," *Structural Concrete*, vol. 21, no. 4, pp. 1272–1285, apr 2020.

[50] A. Sivasuriyan, D. S. Vijayan, W. Górski, Ł. Wodzyński, M. D. Vaverková, and E. Koda, "Practical implementation of structural health monitoring in multi-story buildings," *Buildings*, vol. 11, no. 6, p. 263, jun 2021.

[51] C. K. Oh, H. Sohn, and I.-H. Bae, "Statistical novelty detection within the yeongjong suspension bridge under environmental and operational variations," *Smart materials and structures*, vol. 18, no. 12, p. 125022, 2009.

[52] I. Laory, T. N. Trinh, I. F. Smith, and J. M. Brownjohn, "Methodologies for predicting natural frequency variation of a suspension bridge," *Engineering Structures*, vol. 80, pp. 211–221, dec 2014.

[53] S. Arangio and F. Bontempi, "Structural health monitoring of a cable-stayed bridge with bayesian neural networks," *Structure and Infrastructure Engineering*, vol. 11, no. 4, pp. 575–587, 2015.

[54] S. Bloemheuvel, J. van den Hoogen, and M. Atzmueller, "A computational framework for modeling complex sensor network data using graph signal processing and graph neural networks in structural health monitoring," *Applied Network Science*, vol. 6, no. 1, pp. 1–24, 2021.

[55] G. Comanducci, F. Magalhães, F. Ubertini, and Á. Cunha, "On vibration-based damage detection by multivariate statistical techniques: Application to a long-span arch bridge," *Structural Health Monitoring*, vol. 15, no. 5, pp. 505–524, aug 2016.

[56] A. Cury, C. Cremona, and J. Dumoulin, "Long-term monitoring of a psc box girder bridge: Operational modal analysis, data normalization and structural modification assessment," *Mechanical Systems and Signal Processing*, vol. 33, pp. 13–37, 2012.

[57] Y.-L. Ding, G.-X. Wang, Y. Hong, Y.-S. Song, L.-Y. Wu, and Q. Yue, "Detection and localization of degraded truss members in a steel arch bridge based on correlation between strain and temperature," *Journal of Performance of Constructed Facilities*, vol. 31, no. 5, oct 2017.

[58] Z. Sun, J. Santos, and E. Caetano, "Data-driven prediction and interpretation of fatigue damage in a road-rail suspension bridge considering multiple loads," *Structural Control and Health Monitoring*, may 2022.

[59] M. M. Alamdari, T. Rakotoarivelo, and N. L. D. Khoa, "A spectral-based clustering for structural health monitoring of the sydney harbour bridge," *Mechanical Systems and Signal Processing*, vol. 87, pp. 384–400, mar 2017.

[60] J. W. Hu and M. R. Kaloop, "Single input-single output identification thermal response model of bridge using nonlinear ARX with wavelet networks," *Journal of Mechanical Science and Technology*, vol. 29, no. 7, pp. 2817–2826, jul 2015.

[61] Y. Zhang, Y. Miyamori, S. Mikami, and T. Saito, "Vibration-based structural state identification by a 1-dimensional convolutional neural network," *Computer-Aided Civil and Infrastructure Engineering*, vol. 34, no. 9, pp. 822–839, apr 2019.

[62] H.-B. Huang, T.-H. Yi, and H.-N. Li, "Anomaly identification of structural health monitoring data using dynamic independent component analysis," *Journal of Computing in Civil Engineering*, vol. 34, no. 5, sep 2020.

[63] A. Cury, C. Crémona, and E. Diday, "Application of symbolic data analysis for structural modification assessment," *Engineering Structures*, vol. 32, no. 3, pp. 762–775, mar 2010.

[64] A. Datteo, F. Lucà, and G. Busca, "Statistical pattern recognition approach for long-time monitoring of the g.meazza stadium by means of AR models and PCA," *Engineering Structures*, vol. 153, pp. 317–333, dec 2017.

[65] I. Abdallah, V. Dertimanis, H. Mylonas, K. Tatsis, E. Chatzi, N. Dervili, K. Worden, and E. Maguire, "Fault diagnosis of wind turbine structures using decision tree learning algorithms with big data," in *Safety and Reliability – Safe Societies in a Changing World*. CRC Press, jun 2018, pp. 3053–3061.

[66] S. Bogoevska, M. Spiridonakos, E. Chatzi, E. Dumova-Jovanoska, and R. Höffer, "A data-driven diagnostic framework for wind turbine structures: A holistic approach," *Sensors*, vol. 17, no. 4, p. 720, mar 2017.

[67] A. Zhang, M. Li, and L. Zhou, "Structural health monitoring of offshore wind turbine based on online data-driven support vector machine," in *2018 IEEE 7th Data Driven Control and Learning Systems Conference (DDCLS)*. IEEE, may 2018.

[68] D. F. Hesser, K. Altun, and B. Markert, "Monitoring and tracking of a suspension railway based on data-driven methods applied to inertial measurements," *Mechanical Systems and Signal Processing*, vol. 164, p. 108298, feb 2022.

[69] F. N. Catbas and M. Malekzadeh, "A machine learning-based algorithm for processing massive data collected from the mechanical components of movable bridges," *Automation in Construction*, vol. 72, pp. 269–278, dec 2016.

[70] F. Ubertini, G. Comanducci, and N. Cavalagli, "Vibration-based structural health monitoring of a historic bell-tower using output-only measurements and multivariate statistical analysis," *Structural Health Monitoring*, vol. 15, no. 4, pp. 438–457, jul 2016.

[71] E. García-Macías and F. Ubertini, "MOVA/MOSS: Two integrated software solutions for comprehensive structural health monitoring of structures," *Mechanical Systems and Signal Processing*, vol. 143, p. 106830, sep 2020.

[72] F. Salazar, M. Á. Toledo, J. M. González, and E. Oñate, "Early detection of anomalies in dam performance: A methodology based on boosted regression trees," *Structural Control and Health Monitoring*, vol. 24, no. 11, p. e2012, apr 2017.

[73] J. Mata, "Interpretation of concrete dam behaviour with artificial neural network and multiple linear regression models," *Engineering Structures*, vol. 33, no. 3, pp. 903–910, mar 2011.

[74] J. Mata, A. T. de Castro, and J. S. da Costa, "Time–frequency analysis for concrete dam safety control: Correlation between the daily variation of structural response and air temperature," *Engineering Structures*, vol. 48, pp. 658–665, mar 2013.

[75] R. R. Rajkumar, I. Lee, L. Sha, and J. Stankovic, "Cyber-physical systems," in *Proceedings of the 47th Design Automation Conference on - DAC '10*. ACM Press, 2010.

[76] E. A. Lee, "Cyber physical systems: Design challenges," in *2008 11th IEEE international symposium on object and component-oriented real-time distributed computing (ISORC)*. IEEE, 2008, pp. 363–369.

[77] E. Negri, L. Fumagalli, and M. Macchi, "A review of the roles of digital twin in CPS-based production systems," *Procedia Manufacturing*, vol. 11, pp. 939–948, 2017.

[78] S. Azhar, "Building information modeling (BIM): Trends, benefits, risks, and challenges for the AEC industry," *Leadership and Management in Engineering*, vol. 11, no. 3, pp. 241–252, jul 2011.

[79] A. Rytter, "Vibrational based inspection of civil engineering structures," 1993.

[80] Y.-L. Ding, G.-X. Wang, P. Sun, L.-Y. Wu, and Q. Yue, "Long-term structural health monitoring system for a high-speed railway bridge structure," *The Scientific World Journal*, vol. 2015, pp. 1–17, 2015.

[81] N. Dervilis, K. Worden, and E. Cross, "On robust regression analysis as a means of exploring environmental and operational conditions for SHM data," *Journal of Sound and Vibration*, vol. 347, pp. 279–296, jul 2015.

[82] Y. Ni, X. Hua, K. Fan, and J. Ko, "Correlating modal properties with temperature using long-term monitoring data and support vector machine technique," *Engineering Structures*, vol. 27, no. 12, pp. 1762–1773, oct 2005.

[83] D. Huang, D. Hu, J. He, and Y. Xiong, "Structure damage detection based on ensemble learning," in *2018 9th International Conference on Mechanical and Aerospace Engineering (ICMAE)*. IEEE, jul 2018.

[84] S. D. Fassois and F. P. Kopsaftopoulos, "Statistical time series methods for vibration based structural health monitoring," in *New Trends in Structural Health Monitoring*. Springer Vienna, 2013, pp. 209–264.

[85] S. Li, J. Niu, and Z. Li, "Novelty detection of cable-stayed bridges based on cable force correlation exploration using spatiotemporal graph convolutional networks," *Structural Health Monitoring*, vol. 20, no. 4, pp. 2216–2228, 2021.

[86] B. Bhowmik, T. Tripura, B. Hazra, and V. Pakrashi, "Real time structural modal identification using recursive canonical correlation analysis and application towards online structural damage detection," *Journal of Sound and Vibration*, vol. 468, p. 115101, mar 2020.

[87] C. R. Farrar, H. Sohn, F. M. Hemez, M. C. Anderson, M. T. Bement, P. J. Cornwell, S. W. Doebling, J. Schultze, N. Lieven, and A. Robertson, "Damage prognosis: current status and future needs," *Los Alamos National Laboratory, LA*, vol. 176, pp. 177–178, 2003.

[88] C. R. Farrar and N. A. Lieven, "Damage prognosis: the future of structural health monitoring," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 365, no. 1851, pp. 623–632, 2007.

[89] A. C. Neves, I. González, J. Leander, and R. Karoumi, "Structural health monitoring of bridges: a model-free ANN-based approach to damage detection," *Journal of Civil Structural Health Monitoring*, vol. 7, no. 5, pp. 689–702, nov 2017.

[90] S. Arangio and J. Beck, "Bayesian neural networks for bridge integrity assessment," *Structural Control and Health Monitoring*, vol. 19, no. 1, pp. 3–21, 2012.

[91] A. Ibrahim, A. Eltawil, Y. Na, and S. El-Tawil, "A machine learning approach for structural health monitoring using noisy data sets," *IEEE Transactions on Automation Science and Engineering*, vol. 17, no. 2, pp. 900–908, 2019.

[92] Y. zhou Lin, Z. hua Nie, and H. wei Ma, "Structural damage detection with automatic feature-extraction through deep learning," *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 12, pp. 1025–1046, nov 2017.

[93] H. Khodabandehlou, G. Pekcan, and M. S. Fadali, "Vibration-based structural condition assessment using convolution neural networks," *Structural Control and Health Monitoring*, p. e2308, dec 2018.

[94] O. Abdeljaber, O. Avci, S. Kiranyaz, M. Gabbouj, and D. J. Inman, "Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks," *Journal of Sound and Vibration*, vol. 388, pp. 154–170, feb 2017.

[95] O. Abdeljaber, O. Avci, M. S. Kiranyaz, B. Boashash, H. Sodano, and D. J. Inman, "1-d CNNs for structural damage detection: Verification on a structural health monitoring benchmark data," *Neurocomputing*, vol. 275, pp. 1308–1317, jan 2018.

[96] H. Pan, G. Gui, Z. Lin, and C. Yan, "Deep BBN learning for health assessment toward decision-making on structures under uncertainties," *KSCE Journal of Civil Engineering*, vol. 22, no. 3, pp. 928–940, mar 2018.

[97] J. Sierra-Perez, A. Güemes, and L. E. Mujica, "Damage detection by using fbgs and strain field pattern recognition techniques," *Smart materials and structures*, vol. 22, no. 2, p. 025011, 2012.

[98] H.-P. Wan and Y.-Q. Ni, "Bayesian modeling approach for forecast of structural stress response using structural health monitoring data," *Journal of Structural Engineering*, vol. 144, no. 9, sep 2018.

[99] X. Li and W. Yu, "Data stream classification for structural health monitoring via on-line support vector machines," in *2015 IEEE First International Conference on Big Data Computing Service and Applications*. IEEE, mar 2015.

[100] Y. Liu and S. Nayak, "Structural health monitoring: State of the art and perspectives," *JOM*, vol. 64, no. 7, pp. 789–792, jul 2012.

[101] A. Teughels and G. D. Roeck, "Structural damage identification of the highway bridge z24 by FE model updating," *Journal of Sound and Vibration*, vol. 278, no. 3, pp. 589–610, dec 2004.

[102] J. Maeck and G. De Roeck, "Dynamic bending and torsion stiffness derivation from modal curvatures and torsion rates," *Journal of Sound and Vibration*, vol. 225, no. 1, pp. 153–170, 1999.

[103] G. Kawiecki, "Modal damping measurement for damage detection," *Smart Materials and Structures*, vol. 10, no. 3, pp. 466–471, jun 2001.

[104] B. Koh and S. Dyke, "Structural health monitoring for flexible bridge structures using correlation and sensitivity of modal data," *Computers &amp Structures*, vol. 85, no. 3-4, pp. 117–130, feb 2007.

[105] H. Zhu, L. Li, and X.-Q. He, "Damage detection method for shear buildings using the changes in the first mode shape slopes," *Computers &amp Structures*, vol. 89, no. 9-10, pp. 733–743, may 2011.

[106] J. Wang and Y. Ni, "Refinement of damage identification capability of neural network techniques in application to a suspension bridge," *Structural Monitoring and Maintenance*, vol. 2, no. 1, pp. 77–93, mar 2015.

[107] A. Cury and C. Crémona, "Pattern recognition of structural behaviors based on learning algorithms and symbolic data concepts," *Structural Control and Health Monitoring*, vol. 19, no. 2, pp. 161–186, 2012.

[108] L. sheng Huo, X. Li, Y.-B. Yang, and H.-N. Li, "Damage detection of structures for ambient loading based on cross correlation function amplitude and SVM," *Shock and Vibration*, vol. 2016, pp. 1–12, 2016.

[109] W.-H. Hu, E. Caetano, and Á. Cunha, "Structural health monitoring of a stress-ribbon footbridge," *Engineering Structures*, vol. 57, pp. 578–593, dec 2013.

[110] G. Lederman, Z. Wang, J. Bielak, H. Noh, J. Garrett, S. Chen, J. Kovačević, F. Cerda, and P. Rizzo, "Damage quantification and localization algorithms for indirect SHM of bridges," in *Bridge Maintenance, Safety, Management and Life Extension*. CRC Press, may 2014, pp. 640–647.

[111] Y. Duan, Q. Chen, H. Zhang, C. B. Yun, S. Wu, and Q. Zhu, "Cnn-based damage identification method of tied-arch bridge using spatial-spectral information," *Smart Struct. Syst*, vol. 23, no. 5, pp. 507–520, 2019.

[112] M. Fallahian, F. Khoshnoudian, and S. Talaei, "Application of couple sparse coding ensemble on structural damage detection," *Smart Structures and Systems*, vol. 21, no. 1, pp. 001–14, 2018.

[113] A. P. Urgueira, R. A. Almeida, and N. M. Maia, "On the use of the transmissibility concept for the evaluation of frequency response functions," *Mechanical Systems and Signal Processing*, vol. 25, no. 3, pp. 940–951, apr 2011.

[114] D. H. Nguyen, T. T. Bui, G. De Roeck, and M. A. Wahab, "Damage detection in ca-non bridge using transmissibility and artificial neural networks," *Structural Engineering and Mechanics*, vol. 71, no. 2, pp. 175–183, 2019.

[115] Y.-L. Zhou, N. M. Maia, and M. Abdel Wahab, "Damage detection using transmissibility compressed by principal component analysis enhanced with distance measure," *Journal of Vibration and Control*, vol. 24, no. 10, pp. 2001–2019, 2018.

[116] W.-J. Yan, M.-Y. Zhao, Q. Sun, and W.-X. Ren, "Transmissibility-based system identification for structural health monitoring: Fundamentals, approaches, and applications," *Mechanical Systems and Signal Processing*, vol. 117, pp. 453–482, feb 2019.

[117] Y. Diao, X. Men, Z. Sun, K. Guo, and Y. Wang, "Structural damage identification based on the transmissibility function and support vector machine," *Shock and Vibration*, vol. 2018, 2018.

[118] T. Liu, H. Xu, M. Ragulskis, M. Cao, and W. Ostachowicz, "A data-driven damage identification framework based on transmissibility function datasets and one-dimensional convolutional neural networks: Verification on a structural health monitoring benchmark structure," *Sensors*, vol. 20, no. 4, p. 1059, 2020.

[119] G. Facchini, L. Bernardini, S. Atek, and P. Gaudenzi, "Use of the wavelet packet transform for pattern recognition in a structural health monitoring application," *Journal of Intelligent Material Systems and Structures*, vol. 26, no. 12, pp. 1513–1529, 2015.

[120] A. Shi and X.-H. Yu, "Structural damage detection using artificial neural networks and wavelet transform," in *2012 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA) Proceedings*. IEEE, jul 2012.

[121] D.-A. Tibaduiza, M.-A. Torres-Arredondo, L. Mujica, J. Rodellar, and C.-P. Fritzen, "A study of two unsupervised data driven statistical methodologies for detecting and classifying damages in structural health monitoring," *Mechanical Systems and Signal Processing*, vol. 41, no. 1-2, pp. 467–484, dec 2013.

[122] Y. Kim, J. W. Chong, K. H. Chon, and J. Kim, "Wavelet-based AR–SVM for health monitoring of smart structures," *Smart Materials and Structures*, vol. 22, no. 1, p. 015003, nov 2012.

[123] H. Z. HosseinAbadi, R. Amirfattahi, B. Nazari, H. R. Mirdamadi, and S. A. Atashipour, "GUW-based structural damage detection using WPT statistical features and multiclass SVM," *Applied Acoustics*, vol. 86, pp. 59–70, dec 2014.

[124] X.-L. Peng, H. Hao, and Z.-X. Li, "Application of wavelet packet transform in subsea pipeline bedding condition assessment," *Engineering Structures*, vol. 39, pp. 50–65, jun 2012.

[125] C. D. Duan, Y. Y. Liu, and Q. Gao, "Structure health monitoring using support vector data description and wavelet packet energy distributions," in *Applied Mechanics and Materials*, vol. 135. Trans Tech Publ, 2012, pp. 930–937.

[126] R. Yan and R. X. Gao, "Wavelet domain principal feature analysis for spindle health diagnosis," *Structural Health Monitoring*, vol. 10, no. 6, pp. 631–642, jan 2011.

[127] R. Yang, Y. He, and H. Zhang, "Progress and trends in nondestructive testing and evaluation for wind turbine composite blade," *Renewable and Sustainable Energy Reviews*, vol. 60, pp. 1225–1250, jul 2016.

[128] P. Rizzo, "Nde/shm of underwater structures: a review," *Advances in Science and Technology*, vol. 83, pp. 208–216, 2013.

[129] T. Loutas, N. Eleftheroglou, and D. Zarouchas, "A data-driven probabilistic framework towards the in-situ prognostics of fatigue life of composites based on acoustic emission data," *Composite Structures*, vol. 161, pp. 522–529, feb 2017.

[130] C. Tang, G. Y. Tian, K. Li, R. Sutthaweekul, and J. Wu, "Smart compressed sensing for online evaluation of cfrp structure integrity,"

IEEE Transactions on Industrial Electronics, vol. 64, no. 12, pp. 9608–9617, 2017.

[131] S. Senck, M. Happl, M. Reiter, M. Scheerer, M. Kendel, J. Glinz, and J. Kastner, "Additive manufacturing and non-destructive testing of topology-optimised aluminium components," *Nondestructive Testing and Evaluation*, vol. 35, no. 3, pp. 315–327, 2020.

[132] R. Yao and S. N. Pakzad, "Autoregressive statistical pattern recognition algorithms for damage detection in civil structures," *Mechanical Systems and Signal Processing*, vol. 31, pp. 355–368, aug 2012.

[133] X. Fan, J. Li, and H. Hao, "Piezoelectric impedance based damage detection in truss bridges based on time frequency ARMA model," *Smart Structures and Systems*, vol. 18, no. 3, pp. 501–523, sep 2016.

[134] S. Mustapha, Y. Hu, K. Nguyen, M. M. Alamdari, P. Runcie, U. Dackermann, V. Nguyen, J. Li, and L. Ye, "Pattern recognition based on time series analysis using vibration data for structural health monitoring in civil structures," *Electronic Journal of Structural Engineering*, 2015.

[135] K. Roy, B. Bhattacharya, and S. Ray-Chaudhuri, "ARX model-based damage sensitive features for structural damage localization using output-only measurements," *Journal of Sound and Vibration*, vol. 349, pp. 99–122, aug 2015.

[136] L. Yan, A. Elgamal, and G. W. Cottrell, "Substructure vibration NARX neural network approach for statistical damage inference," *Journal of Engineering Mechanics*, vol. 139, no. 6, pp. 737–747, jun 2013.

[137] A. J. Wootton, J. B. Butcher, T. Kyriacou, C. R. Day, and P. W. Haycock, "Structural health monitoring of a footbridge using echo state networks and NARMAX," *Engineering Applications of Artificial Intelligence*, vol. 64, pp. 152–163, sep 2017.

[138] M. I. Rafiq, M. K. Chryssanthopoulos, and S. Sathananthan, "Bridge condition modelling and prediction using dynamic bayesian belief networks," *Structure and Infrastructure Engineering*, vol. 11, no. 1, pp. 38–50, 2015.

[139] E. B. Flynn, M. D. Todd, A. J. Croxford, B. W. Drinkwater, and P. D. Wilcox, "Enhanced detection through low-order stochastic modeling for guided-wave structural health monitoring," *Structural Health Monitoring*, vol. 11, no. 2, pp. 149–160, 2012.

[140] Y. Ni, Y. Wang, and C. Zhang, "A bayesian approach for condition assessment and damage alarm of bridge expansion joints using long-term structural health monitoring data," *Engineering Structures*, vol. 212, p. 110520, jun 2020.

[141] Y. Yu and C. S. Cai, "Prediction of extreme traffic load effects of bridges using bayesian method and application to bridge condition assessment," *Journal of Bridge Engineering*, vol. 24, no. 3, mar 2019.

[142] M. Gobbato, J. Kosmatka, and J. Conte, "Developing an integrated structural health monitoring and damage prognosis (SHM-DP) framework for predicting the fatigue life of adhesively-bonded composite joints," in *Fatigue and Fracture of Adhesively-Bonded Composite Joints*. Elsevier, 2015, pp. 493–526.

[143] R. M. Neal, *Bayesian learning for neural networks*. Springer Science & Business Media, 2012, vol. 118.

[144] T.-H. Yi, H.-B. Huang, and H.-N. Li, "Development of sensor validation methodologies for structural health monitoring: A comprehensive review," *Measurement*, vol. 109, pp. 200–214, oct 2017.

[145] J. Vitola, F. Pozo, D. Tibaduiza, and M. Anaya, "A sensor data fusion system based on k-nearest neighbor pattern classification for structural health monitoring applications," *Sensors*, vol. 17, no. 2, p. 417, feb 2017.

[146] H. Sohn, K. Worden, and C. R. Farrar, "&lttitle&gtnovelty detection under changing environmental conditions&lt/title&gt," in *SPIE Proceedings*, S.-C. Liu, Ed. SPIE, jul 2001.

[147] Y. Yang and S. Nagarajaiah, "Time-frequency blind source separation using independent component analysis for output-only modal identification of highly damped structures," *Journal of Structural Engineering*, vol. 139, no. 10, pp. 1780–1793, oct 2013.

[148] C. Zang, M. I. Friswell, and M. Imregun, "Structural damage detection using independent component analysis," *Structural Health Monitoring*, vol. 3, no. 1, pp. 69–83, mar 2004.

[149] D. S. Wilks, *Statistical methods in the atmospheric sciences*. Academic press, 2011, vol. 100.

[150] C. Cremona and J. Santos, "Structural health monitoring as a big-data problem," *Structural Engineering International*, vol. 28, no. 3, pp. 243–254, jul 2018.

[151] K. Worden and A. J. Lane, "Damage identification using support vector machines," *Smart Materials and Structures*, vol. 10, no. 3, pp. 540–547, jun 2001.

[152] A. Soualhi, K. Medjaher, and N. Zerhouni, "Bearing health monitoring based on hilbert–huang transform, support vector machine, and regression," *IEEE Transactions on instrumentation and measurement*, vol. 64, no. 1, pp. 52–62, 2014.

[153] X. Zhang, Y. Liang, J. Zhou, and Y. zang, "A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM," *Measurement*, vol. 69, pp. 164–179, jun 2015.

[154] T. Praveenkumar, M. Saimurugan, P. Krishnakumar, and K. Ramachandran, "Fault diagnosis of automobile gearbox based on machine learning techniques," *Procedia Engineering*, vol. 97, pp. 2092–2098, 2014.

[155] I. Bayane and E. Brühwiler, "Investigation of fatigue damage in a RC deck slab due to moving loads through long-term monitoring of a road viaduct," in *Bridge Maintenance, Safety, Management, Life-Cycle Sustainability and Innovations*. CRC Press, apr 2021, pp. 3721–3727.

[156] L. Long, S. Thons, and M. Dohler, "The effects of shm system parameters on the value of damage detection information," in *EWSHM 2018, 9th European Workshop on Structural Health Monitoring*, 2018, pp. 10–p.

[157] M. Jeong, J.-H. Choi, and B.-H. Koh, "Isomap-based damage classification of cantilevered beam using modal frequency changes," *Structural Control and Health Monitoring*, vol. 21, no. 4, pp. 590–602, 2014.

[158] H. Sarmadi and A. Karamodin, "A novel anomaly detection method based on adaptive mahalanobis-squared distance and one-class kNN rule for structural health monitoring under environmental effects," *Mechanical Systems and Signal Processing*, vol. 140, p. 106495, jun 2020.

[159] F. Parisi, A. Mangini, M. Fanti, and J. M. Adam, "Automated location of steel truss bridge damage using machine learning and raw strain sensor data," *Automation in Construction*, vol. 138, p. 104249, jun 2022.

[160] Z. Li, J. Guo, W. Liang, X. Xie, G. Zhang, and S. Wang, "Structural health monitoring based on realadaboost algorithm in wireless sensor networks," in *International Conference on Wireless Algorithms, Systems, and Applications*. Springer, 2014, pp. 236–245.

[161] F.-K. Chang, *Structural health monitoring: current status and perspectives*. CRC Press, 1998.

[162] J. Min, S. Park, C.-B. Yun, C.-G. Lee, and C. Lee, "Impedance-based structural health monitoring incorporating neural network technique for identification of damage type and severity," *Engineering Structures*, vol. 39, pp. 210–220, jun 2012.

[163] J. Guo, X. Xie, R. Bie, and L. Sun, "Structural health monitoring by using a sparse coding-based deep learning algorithm with wireless sensor networks," *Personal and Ubiquitous Computing*, vol. 18, no. 8, pp. 1977–1987, aug 2014.

[164] D. M. Sahoo and S. Chakraverty, "Functional link neural network learning for response prediction of tall shear buildings with respect to earthquake data," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 1, pp. 1–10, 2017.

[165] T. Li, Y. Pan, K. Tong, C. E. Ventura, and C. W. de Silva, "Attention-based sequence-to-sequence learning for online structural response forecasting under seismic excitation," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 4, pp. 2184–2200, 2021.

[166] R.-T. Wu and M. R. Jahanshahi, "Deep convolutional neural network for structural dynamic response estimation and system identification," *Journal of Engineering Mechanics*, vol. 145, no. 1, p. 04018125, 2019.

[167] T. Liu, H. Xu, M. Ragulskis, M. Cao, and W. Ostachowicz, "A data-driven damage identification framework based on transmissibility function datasets and one-dimensional convolutional neural networks: Verification on a structural health monitoring benchmark structure," *Sensors*, vol. 20, no. 4, p. 1059, feb 2020.

[168] N. S. Gulgec, M. Takáč, and S. N. Pakzad, "Convolutional neural network approach for robust structural damage detection and localization," *Journal of Computing in Civil Engineering*, vol. 33, no. 3, may 2019.

[169] M. de Oliveira, A. Monteiro, and J. V. Filho, "A new structural health monitoring strategy based on PZT sensors and convolutional neural network," *Sensors*, vol. 18, no. 9, p. 2955, sep 2018.

[170] H. Liu, "Smart structural condition assessment methods for civil infrastructures using deep learning algorithm," Ph.D. dissertation, University of Maryland, College Park, 2018.

[171] Z. Tang, Z. Chen, Y. Bao, and H. Li, "Convolutional neural network-based data anomaly detection method using multiple information for structural health monitoring," *Structural Control and Health Monitoring*, vol. 26, no. 1, p. e2296, 2019.

[172] A. J. Wootton, C. R. Day, and P. W. Haycock, "An echo state network approach to structural health monitoring," in *2015 International Joint Conference on Neural Networks (IJCNN)*. IEEE, jul 2015.

[173] I. R. Urquijo, A. Q. Incera, S. V. Vaerenbergh, D. Inaud, and J. L. Higuera, "Risks and opportunities of using fibre optic sensors for

long term infrastructure health monitoring systems in an 18 year old installation," in *International Conference on Smart Infrastructure and Construction 2019 (ICSIC)*. ICE Publishing, jan 2019.

[174] A. Guo, A. Jiang, J. Lin, and X. Li, "Data mining algorithms for bridge health monitoring: Kohonen clustering and LSTM prediction approaches," *The Journal of Supercomputing*, vol. 76, no. 2, pp. 932–947, oct 2019.

[175] J. Zhou, L. Dong, W. Guan, and J. Yan, "Impact load identification of nonlinear structures using deep recurrent neural network," *Mechanical Systems and Signal Processing*, vol. 133, p. 106292, nov 2019.

[176] G. Tsialiamanis, C. Mylonas, E. Chatzi, N. Dervilis, D. J. Wagg, and K. Worden, "Foundations of population-based shm, part iv: The geometry of spaces of structures and their feature spaces," *Mechanical Systems and Signal Processing*, vol. 157, p. 107692, 2021.

[177] L. Zhou, S.-X. Chen, Y.-Q. Ni, and A. W.-H. Choy, "Emi-gcn: a hybrid model for real-time monitoring of multiple bolt looseness using electromechanical impedance and graph convolutional networks," *Smart Materials and Structures*, vol. 30, no. 3, p. 035032, 2021.

[178] H. Son, V.-T. Pham, Y. Jang, and S.-E. Kim, "Damage localization and severity assessment of a cable-stayed bridge using a message passing neural network," *Sensors*, vol. 21, no. 9, p. 3118, 2021.

[179] C. S. N. Pathirage, J. Li, L. Li, H. Hao, W. Liu, and P. Ni, "Structural damage identification based on autoencoder neural networks and deep learning," *Engineering Structures*, vol. 172, pp. 13–28, oct 2018.

[180] C. S. N. Pathirage, J. Li, L. Li, H. Hao, W. Liu, and R. Wang, "Development and application of a deep learning–based sparse autoencoder framework for structural damage identification," *Structural Health Monitoring*, vol. 18, no. 1, pp. 103–122, 2019.

[181] M. H. Rafiei and H. Adeli, "A novel unsupervised deep learning model for global and local health condition assessment of structures," *Engineering Structures*, vol. 156, pp. 598–607, feb 2018.

[182] T. Rogers, K. Worden, R. Fuentes, N. Dervilis, U. Tygesen, and E. Cross, "A bayesian non-parametric clustering approach for semi-supervised structural health monitoring," *Mechanical Systems and Signal Processing*, vol. 119, pp. 100–119, mar 2019.

[183] J. Liu, S. Chen, M. Bergés, J. Bielak, J. H. Garrett, J. Kovačević, and H. Y. Noh, "Diagnosis algorithms for indirect structural health monitoring of a bridge model via dimensionality reduction," *Mechanical Systems and Signal Processing*, vol. 136, p. 106454, 2020.

[184] C. K. Oh and H. Sohn, "Damage diagnosis under environmental and operational variations using unsupervised support vector machine," *Journal of sound and vibration*, vol. 325, no. 1-2, pp. 224–239, 2009.

[185] G. Gui, H. Pan, Z. Lin, Y. Li, and Z. Yuan, "Data-driven support vector machine with optimization techniques for structural health monitoring and damage detection," *KSCE Journal of Civil Engineering*, vol. 21, no. 2, pp. 523–534, jan 2017.

[186] T. Tinga and R. Loendersloot, "Physical model-based prognostics and health monitoring to enable predictive maintenance," in *Predictive Maintenance in Dynamic Systems*. Springer International Publishing, 2019, pp. 313–353.

[187] C. Hu, B. D. Youn, and P. Wang, "Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life," in *2011 IEEE Conference on Prognostics and Health Management*. IEEE, jun 2011.

[188] N. Eleftheroglou, D. Zarouchas, T. Loutas, R. Alderliesten, and R. Benedictus, "Structural health monitoring data fusion for in-situ life prognosis of composite structures," *Reliability Engineering &amp System Safety*, vol. 178, pp. 40–54, oct 2018.

[189] Y. Liu, S. Mohanty, and A. Chattopadhyay, "Condition based structural health monitoring and prognosis of composite structures under uniaxial and biaxial loading," *Journal of Nondestructive Evaluation*, vol. 29, no. 3, pp. 181–188, jun 2010.

[190] R.-T. Wu and M. R. Jahanshahi, "Data fusion approaches for structural health monitoring and system identification: past, present, and future," *Structural Health Monitoring*, vol. 19, no. 2, pp. 552–586, 2020.

[191] A. A. Broer, R. Benedictus, and D. Zarouchas, "The need for multi-sensor data fusion in structural health monitoring of composite aircraft structures," *Aerospace*, vol. 9, no. 4, p. 183, 2022.

[192] R. Rocchetta, M. Broggi, Q. Huchet, and E. Patelli, "On-line bayesian model updating for structural health monitoring," *Mechanical Systems and Signal Processing*, vol. 103, pp. 174–195, 2018.

[193] M. Rabiei and M. Modarres, "A recursive bayesian framework for structural health management using online monitoring and periodic inspections," *Reliability Engineering &amp System Safety*, vol. 112, pp. 154–164, apr 2013.

[194] A. Decò and D. M. Frangopol, "Real-time risk of ship structures integrating structural health monitoring data: Application to multi-objective optimal ship routing," *Ocean Engineering*, vol. 96, pp. 312–329, 2015.

[195] N. Li, Y. Lei, N. Gebraeel, Z. Wang, X. Cai, P. Xu, and B. Wang, "Multi-sensor data-driven remaining useful life prediction of semi-observable systems," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 11, pp. 11 482–11 491, 2020.

[196] J. Hou, H. Jiang, C. Wan, L. Yi, S. Gao, Y. Ding, and S. Xue, "Deep learning and data augmentation based data imputation for structural health monitoring system in multi-sensor damaged state," *Measurement*, vol. 196, p. 111206, 2022.

[197] H. V. Dang, H. Tran-Ngoc, T. V. Nguyen, T. Bui-Tien, G. De Roeck, and H. X. Nguyen, "Data-driven structural health monitoring using feature fusion and hybrid deep learning," *IEEE Transactions on Automation Science and Engineering*, vol. 18, no. 4, pp. 2087–2103, 2020.

[198] J. Willard, X. Jia, S. Xu, M. Steinbach, and V. Kumar, "Integrating physics-based modeling with machine learning: A survey," *arXiv preprint arXiv:2003.04919*, vol. 1, no. 1, pp. 1–34, 2020.

[199] F.-G. Yuan, S. A. Zargar, Q. Chen, and S. Wang, "Machine learning for structural health monitoring: challenges and opportunities," *Sensors and smart structures technologies for civil, mechanical, and aerospace systems 2020*, vol. 11379, p. 1137903, 2020.

[200] M. Lesjak, "Prediction of chaotic systems with physics-enhanced machine learning models," 2020.

[201] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi, M. Malinowski, A. Tacchetti, D. Raposo, A. Santoro, R. Faulkner *et al.*, "Relational inductive biases, deep learning, and graph networks," *arXiv preprint arXiv:1806.01261*, 2018.

[202] A. Almunif, L. Fan, and Z. Miao, "A tutorial on data-driven eigenvalue identification: Prony analysis, matrix pencil, and eigensystem realization algorithm," *International Transactions on Electrical Energy Systems*, vol. 30, no. 4, p. e12283, 2020.

[203] H. Shokravi, H. Shokravi, N. Bakhary, S. S. Rahimian Koloor, and M. Petr, "Health monitoring of civil infrastructures by subspace system identification method: an overview," *Applied Sciences*, vol. 10, no. 8, p. 2786, 2020.

[204] M. Z. A. Bhuiyan, J. Wu, G. Wang, J. Cao, W. Jiang, and M. Atiquzzaman, "Towards cyber-physical systems design for structural health monitoring," *ACM Transactions on Cyber-Physical Systems*, vol. 1, no. 4, pp. 1–26, oct 2017.

[205] X. Liu, J. Cao, W.-Z. Song, P. Guo, and Z. He, "Distributed sensing for high-quality structural health monitoring using wsns," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 3, pp. 738–747, 2014.

[206] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE internet of things journal*, vol. 3, no. 5, pp. 637–646, 2016.

[207] Q. Chen, J. Cao, and Y. Xia, "Physics-enhanced pca for data compression in edge devices," *IEEE Transactions on Green Communications and Networking*, 2022.

[208] M. E. S. Saeed, Q.-Y. Liu, G. Tian, B. Gao, and F. Li, "Akaiots: authenticated key agreement for internet of things," *Wireless Networks*, vol. 25, pp. 3081–3101, 2019.

[209] R.-T. Wu, A. Singla, M. R. Jahanshahi, E. Bertino, B. J. Ko, and D. Verma, "Pruning deep convolutional neural networks for efficient edge computing in condition assessment of infrastructures," *Computer-Aided Civil and Infrastructure Engineering*, vol. 34, no. 9, pp. 774–789, 2019.

[210] N. Testoni, C. Aguzzi, V. Arditi, F. Zonzini, L. D. Marchi, A. Marzani, and T. S. Cinotti, "A sensor network with embedded data processing and data-to-cloud capabilities for vibration-based real-time SHM," *Journal of Sensors*, vol. 2018, pp. 1–12, jul 2018.

[211] S. M. Khan, S. Atamturktur, M. Chowdhury, and M. Rahman, "Integration of structural health monitoring and intelligent transportation systems for bridge condition assessment: Current status and future direction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 8, pp. 2107–2122, aug 2016.

[212] G. M. Guzman-Acevedo, G. E. Vazquez-Becerra, J. R. Millan-Almaraz, H. E. Rodriguez-Lozoya, A. Reyes-Salazar, J. R. Gaxiola-Camacho, and C. A. Martinez-Felix, "GPS, accelerometer, and smartphone fused smart sensor for SHM on real-scale bridges," *Advances in Civil Engineering*, vol. 2019, pp. 1–15, jun 2019.

[213] M. Benghanem, "Measurement of meteorological data based on wireless data acquisition system monitoring," *Applied Energy*, vol. 86, no. 12, pp. 2651–2660, dec 2009.