

Applications of (Big) Data Analysis in A/E/C

Ming-Hung Hsu ¹, Ying-Wu Yang ² and Zheng-Yun Zhuang ^{3,*}

¹ Department of Electrical Engineering, National Penghu University of Science and Technology, Penghu 880011, Taiwan; hsu@gms.npu.edu.tw

² College of Landscape and Architecture, Zhejiang A&F University, Hangzhou 311300, China; yangyw06@163.com

³ Department of Civil Engineering, National Kaohsiung University of Science and Technology (NKUST), Kaohsiung 807415, Taiwan

* Correspondence: wayne@nkust.edu.tw; Tel.: +886-7-3814526

This editorial paper provides an overview of the *Buildings* Special Issue (SI), dedicated to the topic “Applications of (Big) Data Analysis in A/E/C” (where A/E/C stands for architecture, engineering, and construction) and the academic papers it includes. For more details about this SI, please refer to https://www.mdpi.com/journal/buildings/special_issues/GV29ERG645 (accessed on 4 May 2023).

Some of this paper’s contents are presented in tabular format for clear organization. Table 1 lists the title and keywords of each article published in the SI. This paper also offers a series of digests based mainly on their abstracts to maintain each study’s original contributions, as claimed by the author(s) at publication. Stemming from these reviews, the paper summarises each study’s application domains within A/E/C, outlines the phase or role each study plays in the theoretical flow of (big) data analysis, and then concludes with some key observations. This allows for identifying ‘hot zones’ of research, in which follow-up research and extended studies should be performed, and ‘cold zones’, awaiting the utilisation and novel application of data and theories/models. Note that in the subsequent text, if not otherwise specified, the discussion of the papers follows the order in which they were published.

Zhuang and Kuo [1] propose and apply a systematic data analysis methodology to analyse experimental data from high-performance concrete (HPC) samples with different admixtures for use as offshore fan foundation grouting materials. Compared to other relevant research, including experimental studies, physics and chemistry of materials studies, and cementitious material portfolio determination studies, this data-driven analysis deeply explores the experimental variables associated with the test data. The authors employ several methods, including correlation analysis, cosine similarity analysis, simple linear regression (SLR) modelling, heat maps, and heat-based tabularised visualisations, to offer a comprehensive and in-depth perspective. The proposed methodology in [1] is easily implementable. The authors validated the results using a pairwise comparison approach (PCA).

The contributions of this work include insights for coherent groups of variables, techniques for double and triple checking, the establishment of a ‘knowledge base’ consisting of 504 SLR predictive models with their effectiveness (significance) and prediction accuracy (data-model fitness) used in practical applications, an alternative visualisation of the results, three data transforms that can be omitted in future analyses, and three valuable theory-linking perspectives (e.g., for the relationships between destructive and non-destructive tests with respect to the variable categories). The implication that some variables are interchangeable will make future experiments less labour-intensive and time-consuming for pre-project HPC material testing.



Citation: Hsu, M.-H.; Yang, Y.-W.; Zhuang, Z.-Y. Applications of (Big) Data Analysis in A/E/C. *Buildings* **2023**, *13*, 1442. <https://doi.org/10.3390/buildings13061442>

Received: 17 May 2023

Accepted: 26 May 2023

Published: 31 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Table 1. Title and keywords for each paper published in this SI.

Ref.#	Title	Keywords ¹
[1]	Unravelling the Relations between and Predictive Powers of Different Testing Variables in High Performance Concrete Experiments: The Data-Driven Analytical Methods	high performance concrete (HPC); data-driven analysis; experimental parameters; pre-project material testing; pairwise comparison analysis; correlation analysis; cosine similarity; predictive regression modelling; heat map; variable transform
[2]	An Intelligent Detection Logic for Fan-Blade Damage to Wind Turbines Based on Mounted-Accelerometer Data	energy generation; green energy; wind turbines; fan-blade damage; mechanical condition; accelerometer; monitoring system; fault diagnosis; vibration data; maintenance by prediction; big data analytics
[3]	Reliability Analysis of RC Slab-Column Joints under Punching Shear Load Using a Machine Learning-Based Surrogate Model	reliability analysis; RC slab-column structure; machine learning; Monte Carlo simulation; SHapley Additive exPlanation
[4]	Real-Time Monitoring for Monolithic Movement of a Heritage Curtilage Using Wireless Sensor Networks	wireless sensor network; historic building; real-time monitoring
[5]	Polyethylene (PE)Waste Minimization Study of Cement Mortar with Adding PE Content under Different W/B Ratios	recycling of waste materials; cement mortar; water-to-binder ratio (W/B ratio); waste PE; durability
[6]	Real-Time Monitoring for Effects of Vibration and Temperature of Construction Site on Steel Assembly Bracing of Foundation Pit	steel assembly bracing; foundation pit; environmental temperature; vibration analysis; real-time monitoring
[7]	Automated Semantic Segmentation of Indoor Point Clouds from Close-Range Images with Three-Dimensional Deep Learning	building information model; 3D point cloud; semantic segmentation; deep learning
[8]	Determinants of Data Quality Dimensions for Assessing Highway Infrastructure Data Using Semiotic Framework	highway data quality assessment; data quality dimensions; semiotic framework; decision-making
[9]	Applying Deep Learning and Single Shot Detection in Construction Site Image Recognition	construction image; artificial intelligence; deep learning; object detection; single shot multi-box detector (SSD)

¹ Keywords may be reordered for simplicity of presentation.

Hsu and Zhuang [2] assist industry by establishing a real-time condition-monitoring and fault-detection system with rules for recognising a wind turbine's abnormal operation, mainly caused by different types of fan-blade damage. This system can ensure ideal wind turbine operation by monitoring the health status of the blades, detecting sudden anomalies, and performing maintenance almost in real time. This enables 'maintenance by prediction' actions for unplanned maintenance as a supplement to the 'predictive maintenance' tasks for regular planned maintenance, which is especially significant for wind farms operating in harsh marine or shore environments that are subject to frequent natural disasters (e.g., earthquakes and typhoons). Turbines might fail to endure these because the manufacturers have built them according to the standards developed for areas less prone to natural disasters.

The system's rules are established utilising concepts and methods from data analytics, digital signal processing (DSP), and statistics to analyse data from the accelerometer mounted on the platform of the wind turbine's structure, measuring the vibration signals in three dimensions. The patterns for cases involving fan blade damage are found to establish the rules. By detecting and reporting anomalies effectively, repairs and maintenance can be carried out on faulty wind turbines.

Shen, Shen, and Liang [3] found that reinforced concrete (RC) slab-column structures are prone to punching shear failure, despite their architectural flexibility and easy construction. Punching shear failure is a typical brittle failure, making it challenging to assess

slab-column structure functionality and failure probability. Therefore, predicting punching shear resistance and the corresponding reliability analysis are critical issues in designing RC slab-column structures. The authors used a database containing 610 experimental data for machine learning (ML) modelling to enhance the computational efficiency of the reliability analysis of RC slab-column joints. According to the nonlinear mapping between seven selected input variables and the punching shear resistance of slab-column joints, the study established four ML models, namely the artificial neural network (ANN), decision tree (DT), random forest (RF), and extreme gradient boosting (XGBoost) models.

Based on three performance measures, the authors selected XGBoost as the best prediction model; its root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2) were 32.43, 19.51, and 0.99, respectively. Such advantages are reflected in a comparison with five empirical models introduced. The study visualised the prediction process of XGBoost using SHapley Additive exPlanation (SHAP); the importance sorting and feature dependency plots of the input variables explain the prediction process globally. Furthermore, the paper adopts Monte Carlo simulation with an ML-based surrogate model (ML-MCS) to calibrate the reliability of slab-column joints in a real engineering example. The authors obtained 1,000,000 samples through random sampling and calculated the reliability index (β) of this practical building using Monte Carlo simulation. As a result, they achieved the targeted reliability index under the design provisions. They also conducted a sensitivity analysis of stochastic variables and deeply examined the impact on structural reliability.

Shen, Yang, Yang, Yang, Zhu, and Wang [4] argue that since monolithic movement is a promising technology for relocating historical buildings, corresponding real-time monitoring is of great interest due to the buildings' age and poor structural integrity. However, as related research and practical applications remain limited, the paper proposes a wireless sensor network (WSN)-based strategy as a non-invasive approach to monitoring heritage curtilage during monolithic movement. The collected data show that the inclination of the curtilage is almost negligible. With the aid of finite element simulation, the study found that the crack displacement curves changed from -0.02 to 0.07 mm depending on the direction of movement; however, this value is not enough to cause structural cracks. The deformation of the steel underpinning beam, used to reinforce masonry walls and wooden pillars, is related to the stiffness in different directions. In addition, the strain variations of the steel chassis, which bears the vertical loads from the wooden pillars and masonry walls, are less than 0.04%. This indicates that they are kept within the durable range during monolithic movement. The authors thus prove that the WSN-based approach can potentially be used for the real-time monitoring of the monolithic movement of historic buildings.

Lin, Hung, Wang, and Wen [5] researched the durability of cement mortar prepared using different W/B ratios and percentages of waste PE content. The study was based on the logic that waste can be effectively used in concrete, and the characteristics of concrete can be maintained or enhanced, so the economy of waste management can be significantly increased, thus reducing pollution. They mixed the cement mortar with 0%, 1%, 2%, 3%, and 4% of waste PE and 20% of ground-granulated blast-furnace slag (GGBFS) in W/B ratios of 0.4, 0.5, and 0.6. The results show that slump and flow decrease with increasing waste PE content and increase with increasing W/B ratio; therefore, the setting time becomes shorter as the waste PE content rises. Regarding hardened (mechanical) properties, the specimen strength slightly decreased with increasing waste PE content. Still, the specimens performed better at a later age due to the pozzolanic reaction of slag, which can be verified using a scanning electron microscope.

Yang, Zeng, Liu, Yang, and Li [6] propose a real-time monitoring system—including vibration acceleration sensors, temperature sensors, and static and dynamic strain sensors—to monitor the safety status of a steel (assembly) bracing system in a practical project. It uses 5G wireless networking technology to transmit monitoring data to a cloud server for early warning of abnormal changes and development trends. The authors used real-time

monitoring data obtained from a construction site as the inputs for the finite element model. They compared the corresponding results of a numerical simulation with the results from real-time monitoring. The paper concluded that (1) environmental temperature causes significant stress, which can be higher than the initial prestress of the steel bracing system; (2) the stress caused by vertical vibration, mainly from construction vehicles, is not remarkable, but the vertical frequency-weighted acceleration of support vibration is relatively large, which can affect on-site engineering technicians' sense of safety; and (3) the combination of environmental temperature and vertical vibration does not affect the safety of the steel bracing system.

Hsieh and Ruan [7] point out that generating a 3D model from 3D point clouds involves classification, outline extraction, and boundary regularisation for semantic segmentation. In addition, the number of 3D point clouds generated using close-range images is smaller, and they tend to be unevenly distributed. This is not conducive to automated modelling processing. However, the creation of building information models requires acquiring real building conditions. Thus, the authors propose an efficient solution for the semantic segmentation of indoor point clouds from close-range images. They further propose a 3D deep learning framework that achieves better results. The study used a dynamic graph convolutional neural network (DGCNN) 3D deep learning method to learn point cloud semantic features. Moreover, the authors developed more efficient operations to build a module for extracting point cloud features to resolve the problem of inadequate beam and column classifications. They first applied DGCNN to learn and classify the indoor point cloud into five categories: columns, beams, walls, floors, and ceilings. Then, they utilised the proposed semantic segmentation and modelling method to obtain the geometric parameters of each object to be integrated into building information modelling (BIM) software.

According to the experimental results, the overall accuracy rates of the three experimental sections of Area_1 in the Stanford 3D semantic dataset test results were 86.9%, 97.4%, and 92.5%. The segmentation accuracy of corridor 2F in a building was 94.2%. In comparing the length with the actual on-site measurement, they found the root mean square error to be ± 0.03 m. Thus, the method is capable of automatic semantic segmentation from 3D point clouds with indoor close-range images.

Krishna, Ruikar, and Jha [8] identified the critical data quality dimensions affecting highway projects' decision-making process to address the data quality issues posed by the rapid accumulation of highway infrastructure data and its widespread reuse in decision-making. The authors propose addressing these issues by examining data quality, using various approaches to enhance data quality, and making decisions based on data quality information. Firstly, they conducted a state-of-the-art review of data quality frameworks applied in multiple fields to identify suitable frameworks for highway infrastructure data. Next, they identified data quality dimensions of the semiotic framework from the literature and conducted interviews with highway infrastructure stakeholders to finalize the data quality dimension. Then, they used a questionnaire survey to identify the critical data quality dimensions for decision-making. They also identified the importance of each critical dimension at each decision-making level in the highway infrastructure project. This 'semiotic data quality framework' provides a theoretical foundation for developing data quality dimensions to assess subjective data quality. However, further research is required to find effective ways to evaluate current satisfaction with data quality at various decision-making levels.

Lung and Wang [9] observed that although most construction site workers take photos of construction activities, the site manager relies mainly on manual labour to assess construction progress, quality control, and field management to facilitate job site coordination and productivity management. Moreover, it often takes a great deal of time to process the many photos taken, so in most cases, the image data are processed passively and used only for reference. However, using computerised tools, these photos could serve as aids for project management, including construction history records, quality, and schedule

management. Thus, the authors propose an image recognition system for construction activities by incorporating image recognition through deep learning, using the powerful image extraction ability of a convolutional neural network (CNN) to automatically extract contours, edge lines, and local features via filters and feed feature data to the network for training in a fully connected way. The system is effective in image recognition, which helps identify subtle differences. The authors adjusted the parameters and structure of the neural network for use with a CNN. They selected objects like construction workers, machines, and materials for a case study. A CNN can be used to extract individual features for training, which improves recognizability and helps project managers make decisions regarding construction safety, job site configuration, progress control, and quality management, thus enhancing the efficiency of construction management.

Table 2 below summarises the application domains of the papers reviewed above. Table 3 then summarises the phases of (big) data analytics touched on by the SI papers.

Table 2. Application domains of the SI papers.

Application Domain	SI Papers	Count (Multiple Answers)
Structures and Structural Eng.	[2–4,6]	4
Materials or Geotechnical Eng.	[1,5]	2
Construction Design, Measure, Information Systems and BIM	[3,6–8]	4
Transport/Energy Planning and Related Building Management	[1,2,4,8,9]	5

Table 3. Phases of big data analytics touched on by the SI papers.

Phase Involved or Resolved	SI Papers	Count (Multiple Answers)
Data Collection and Curation	[4–6,8,9]	5
Data Pre-processing	[1,2,4,6,7]	5
Data Analysis	[1,3–6,8]	6
Prediction	[1–3,7,9]	5
Forecasting	N/A	0
Decision-making and Support	[2,3,5,8,9]	5

Tables 2 and 3 show that the SI papers involve highly diversified application domains and phases. However, several observations can be made:

1. Big data theories and techniques have been applied less to the materials or geotechnical (earth) engineering domains than to the other three domains, thus pointing to opportunities for future research;
2. More and more studies (including 6/9 reviewed here, as shown in Table 2) involve two interdisciplinary application domains;
3. As seen in Table 3, an increasing number of studies applying the theories or methods of big data analysis involve two or even three phases within the methodological flow of big data. This trend is noteworthy;
4. Rarely do studies address the forecasting function (0/9 in this review). This also indicates an opportunity for its first-time application in A/E/C studies;
5. By contrast, many papers contribute to the data collection and curation phase (5/9). Still, for the application of (big) data analysis in A/E/C, it is possible that many datasets still need to be collected and/or curated (e.g., using proper IoT or sensor devices). However, much room remains for subsequent analytical or knowledge exploration studies following this initial phase.

As a concluding remark, we would like to highlight the two areas for future studies in the above list. Researchers may consider focusing on the current research ‘hot spot’ and conducting research based on the existing solid ground; alternatively, they may wish to

propose new ways to apply the theories or methods to new cases (to fill in vacant areas, i.e., the ‘cold spot’). Either way, researchers may find the papers in this SI helpful.

Author Contributions: Conceptualization, Z.-Y.Z., Y.-W.Y. and M.-H.H.; Writing—original draft preparation, M.-H.H.; Writing—review and editing, Z.-Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This editorial essay was funded by the National Science and Technology Council (Taiwan, ROC), grant number 111-2410-H-992-011, the Natural Science Foundation of Zhejiang Province (China), grant number LGG20E080002, LQ21A020009, the Science Foundation of Zhejiang Sci-Tech University (China), Grant number 19052460-Y, and the Department’s Quality Improvement and Leap Projects of National Kaohsiung University of Science and Technology (NKUST) (Taiwan, ROC), grant number 112TSD00-1.

Data Availability Statement: This editorial essay involves no data outside the quoted SI, nor any data that are not publicly citable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhuang, Z.Y.; Kuo, W.T. Unravelling the Relations between and Predictive Powers of Different Testing Variables in High Performance Concrete Experiments: The Data-Driven Analytical Methods. *Buildings* **2022**, *12*, 1545. [\[CrossRef\]](#)
2. Hsu, M.H.; Zhuang, Z.Y. An intelligent detection logic for fan-blade damage to wind turbines based on mounted-accelerometer data. *Buildings* **2022**, *12*, 1588. [\[CrossRef\]](#)
3. Shen, L.; Shen, Y.; Liang, S. Reliability Analysis of RC Slab-Column Joints under Punching Shear Load Using a Machine Learning-Based Surrogate Model. *Buildings* **2022**, *12*, 1750. [\[CrossRef\]](#)
4. Shen, L.; Yang, B.; Yang, Y.; Yang, X.; Zhu, W.; Wang, Q. Real-Time Monitoring for Monolithic Movement of a Heritage Curtilage Using Wireless Sensor Networks. *Buildings* **2022**, *12*, 1785. [\[CrossRef\]](#)
5. Lin, K.T.; Hung, C.C.; Wang, H.Y.; Wen, F.L. Polyethylene (PE) Waste Minimization Study of Cement Mortar with Adding PE Content under Different W/B Ratios. *Buildings* **2022**, *12*, 2117. [\[CrossRef\]](#)
6. Yang, Y.; Zeng, H.; Liu, X.; Yang, B.; Li, Y. Real-Time Monitoring for Effects of Vibration and Temperature of Construction Site on Steel Assembly Bracing of Foundation Pit. *Buildings* **2023**, *13*, 450. [\[CrossRef\]](#)
7. Hsieh, C.S.; Ruan, X.J. Automated Semantic Segmentation of Indoor Point Clouds from Close-Range Images with Three-Dimensional Deep Learning. *Buildings* **2023**, *13*, 468. [\[CrossRef\]](#)
8. Krishna, C.M.; Ruikar, K.; Jha, K.N. Determinants of Data Quality Dimensions for Assessing Highway Infrastructure Data Using Semiotic Framework. *Buildings* **2023**, *13*, 944. [\[CrossRef\]](#)
9. Lung, L.W.; Wang, Y.R. Applying Deep Learning and Single Shot Detection in Construction Site Image Recognition. *Buildings* **2023**, *13*, 1074. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.