Compatibility Review of Publicly Available Experimental Data on Common Building Material Temperature Dependent Properties for use with ML Algorithms

Iasonas Bakas[1(✉)], Karolos J. Kontoleon[1], Eleftherios K. Anastasiou[1], Georgios Are-toulis[1], and Müslüm Arıcı[2]

[1] Department of Civil Engineering, Faculty of Engineering, Aristotle University of Thessaloniki, 541 24 Thessaloniki, Greece
iasompak@civil.auth.gr

[2] Mechanical Engineering Department, Engineering Faculty, Kocaeli University, Kocaeli, Turkey

Abstract. The analysis and predictive capabilities of Machine Learning (ML) algorithms are becoming an increasingly invaluable asset in the research domains of heat transfer and building materials. Precision and high performance of such models are heavily reliant on the availability of substantial amounts of data for training and testing purposes. This study concentrates on reviewing available sources of experimental data and assesses compatibility implications with ML algorithm requirements. Focus is given on common building materials with temperature-dependent properties assessed against elevated temperatures or fire scenarios. The structure and consistency of the available information are discussed and compared with the fundamental training data requirements of ML algorithms, focusing on Artificial Neural Networks. The study concludes with a brief discussion on the financial and environmental benefits of data and material circularity and proposes methods of reusing available experimental results in cutting-edge Artificial Intelligence applications.

Keywords: building materials · data analysis · machine learning · information circularity

1 Introduction

1.1 Machine Learning as a Tool for Scientific Research

Machine Learning (ML), under the wider sphere of Artificial Intelligence (AI), is evolving into a valuable tool in scientific research, commercial production and industrial applications. It enables deeper understanding of data, reveals interconnections, supplements or surpasses the capabilities of existing statistical methods of analysis and creates opportunities for circularity and sustainable research and production.

© The Author(s) 2026
L. Bragança et al. (Eds.): ReSBE 2026, LNCE 905, pp. 1424–1431, 2026.
https://doi.org/10.1007/978-3-032-34930-9_166

Fully utilising the potential of such innovative tools relies on understanding and compliance with certain requirements. The algorithm architecture and choice of hyper-parameters are integral to developing a functional and effective model. The reliability and completeness of training and test data are fundamental to ensure underlying information patterns are captured with high levels of precision and confidence. Inherent biases by the model developers need to be filtered and isolated, while accessibility to the end user is essential to ensure uptake of the proposed models and their beneficial features.

This study focuses on the aspect of data availability and quality for training effective ML models. It outlines issues of available datasets, through the prism of building material properties, heat transfer and fire science. It discusses the opportunities created by ML in relation to information circularity and experimental data reuse and invites dialogue on approaches to enhance data depositories and accessibility of information.

1.2 Scientific Knowledge and Data Circularity

Circular Economy (CE) and its ability to "close the loop" on production processes is receiving notable attention due to striking a comfortable balance between economy, societal requirements and environmental consciousness [1]. Machine Learning is already contributing to the development of CE and it is suggested that reverse engineering the principles of circularity to advance ML itself is worth exploring [2]. The current mindset of singularity in information production/consumption is discouraging data sharing, reuse and leads to "fragmentation and lack of interoperability" in the heavily data-reliant field of Machine Learning [3].

Two aspects exacerbate the current difficulty in reusing scientific data in the context of information circularity and subsequently advancing ML. A lack of consistent methodology on recording how and with what purpose in mind datasets have been created hinders future attempts to access and reuse the stored information [4]. Secondly, the infrastructure that facilitates information circularity (at least in the context of reusing scientific data) needs to be enhanced. Data is currently stored in multiple decentralised repositories, behind varying types of access restrictions, making reuse either by researchers or automated information trawlers impractical [5]. Steps to alleviate the situation are being taken, in the form of official guidelines being drawn up (the European AI Act and the AI Right of Bills, for example). Further suggestions involve clearer guidance on data structuring and coherence by scientific journal publishers ahead of publication acceptance and clearer descriptions of dataset features and limitations by researchers who obtain the data in the first place, as well as more descriptive metadata and guidance both for discoverability by machines and organisation by annotator and editorial teams [6]. The core principles of data circularity and reuse are listed on Table 1.

Table 1. Information and data circularity requirements [5].

Requirement	Practical benefit
Findable	Adequately detailed and accurately labelled meta-data logged on a "searchable" depository to enable human or machine searchers to identify the required dataset.

(*continued*)

Table 1. (*continued*)

Requirement	Practical benefit
Accessible	Once found, the data should be easily accessible for free, through a universally accepted protocol.
Interoperable	Once retrieved, data should be in a universally acceptable "knowledge representation" format, including cross-referencing to other meta-data as required.
Reusable	Data being traceable to its source, with detailed descriptions of content, usage permissions and in line with universally accepted standards for the corresponding domain.

1.3 Scope and Contribution of This Study

The aim of the study is to identify representative sources of available experimental results that can be used for training ML algorithms while discussing opportunities on information circularity. It reviews previously published studies in the field of common building material properties exposed to fire and elevated temperatures, with special focus on concrete applications. This study also assesses the compatibility of such data against the fundamental training requirements of ML algorithms, focusing on Artificial Neural Networks (ANNs). This review is not exhaustive, and it concentrates on the most recent or representative publications available on online indexing platforms.

2 Methodology

The adopted research methodology follows the principles of meta-narrative literature review. The cornerstones of meta-narrative research comprise the initial research question setting; establishing the scoping and search strategy; a method of quality appraisal and data extraction; and finally, the stage of information analysis and narrative synthesis The initial paper selection process involved keyword and title review. Further filtering relied on assessment of abstract relevance, clarity and reproducibility of proposed methodology and finally relevance of conclusions [7].

The research objective for this work is distilled in the following question: "What are the features of the available experimental data of temperature-dependent material properties that can be used for training ANN models predicting thermal behaviour of building elements exposed to fire?". Parallel topics such as ML training data requirements and circularity of existing scientific knowledge are woven into the narrative by reviewing and discussing relevant scientific publications.

The scope of this study has been identified in previous paragraphs. The search strategy utilises online indexing platforms such as Scopus, IEEE, Google Scholar and Web of Science. Combinations of keywords such as "material properties, thermal behaviour, temperature-dependent properties, concrete, elevated temperatures, data circularity, Machine Learning data requirements" were used. The returned results were screened on their relevance and quality before including in the below review.

3 Results

3.1 Experimental Data Availability and Quality

A considerable volume of experimental and analytical research exists in relation to concrete temperature dependent properties when exposed to elevated temperatures [8]. For the purposes of this study 15 published papers have been selected and reviewed in detail. These feature both useful experimental results but also encourage discussion in relation to scientific knowledge and data circularity.

Studies have been focusing on different aspects and parameters of concrete specimen thermal behaviour. The impact of age, in conjunction with the concrete mix characteristics, and environmental conditions was investigated for temperatures up to 60 °C with a Japan developed tester for thermal conductivity [9]. Similarly, an electrical furnace was used to investigate the impact of aggregate type to the thermal conductivity and volumetric change of concrete when exposed to temperatures up to 600 °C [10]. Focusing on the impact of by-product admixtures, the hot wire method was utilised, for heating lightweight concrete specimens to 900 °C [11]. On the other hand, despite adequate details being provided for the fiber content and coarse aggregate mix design of samples, the experimental investigation scope varies considerably from being limited to 25 °C in one instance to extending to 1000 °C in others [12–15]. The experiment time (250 min) is reported only in one investigation of thermal behaviour of fly ash-based concrete for temperatures up to 550 °C with readings at 100 °C intervals [16]. Other studies focus on numerical investigation of cracked concrete or mixes containing asphalt, using experiments for validation but providing nominal information on temperatures developed or time required for the experiment [17, 18]. Although various numerical formulas are provided to describe the development of thermal conductivity, no experimental data is made available on any of the above studies.

A smaller proportion of studies (26% of the reviewed work) mention that experimental data can be made available upon specific request. Different experimental approaches were used to measure the thermal conductivity and specific heat of three samples of lightweight concrete at temperatures ranging from 20 to 1000 °C [19]. Others focused on the impact of fiber content in ultra-high-performance concrete, reviewing the thermal conductivity of samples at temperatures up to 850 °C [20]. The thermal conductivity, specific heat and thermal expansion of sleeving grout was subject of experimental investigation for temperatures up to 900 °C [21]. The influence of water content and levels of saturation were also reviewed analytically and experimentally for concrete samples heated up to 800 °C [22].

The compatibility and structure of the experimental data obtained through these works would need to be assessed against the fundamental requirements of Machine Learning models.

3.2 Machine Learning and Training Data Requirements

As seen in the previous paragraphs, experimental data on the temperature dependent material properties does exist. However, it is essential that the process of obtaining,

managing and preprocessing this data is assessed, before analysing and hopefully extracting useful conclusions [23]. The reason for such upfront scrutiny is that the quality of obtained information is detrimental to the quality of the ANN model predictions. Ensuring the provided training data is complete, balanced and accurate (both in terms of recorded features and target figures accuracy) is critical for the efficient operation of an ML model [24]. Gaps in the obtained data (either due to randomly missed/mistakenly noted observations, due to systematic recording errors or corruption in the process of storing, retrieving or preprocessing) can skew or significantly mislead the predictions made by an ANN [25, 26]. The presence of noisy data in the training dataset can be similarly impactful. Apart from potential degradation of predictive capabilities, noisy data elevates the complexity of the model leading to increased training times and computational costs [27]. Depending on the application, data of appropriate quality would conform with the criteria listed on Table 2.

Table 2. Machine Learning training and test data requirements.

Requirement	Qualitative explanation	Impact on Machine Learning model
Completeness	Data should not contain: • observations/entries with missing values. • features with missing titles or labels.	Missing observation values can create imbalances in the training data, affecting the quality of the ML predictions. Missing feature labels can lead to unexpected predictions due to erroneous consideration of training data sections.
Balance	A balanced amount of data should be available to represent each class or feature.	Unbalanced volumes of data lead to misrepresentation of data classes affecting the accuracy of the model.
Accuracy	Data should be free from systematic erroneous entries, noisy recordings and systematic outliers.	Although ML can quantify and filter out the impact of outliers, systematically erroneously recorded data, or noisy readings can skew the final predictions of the model. Such data also increases complexity of the model and reduces its predictive efficiency.
Consistency	Data entries should be presented consistently, avoiding variations in spelling, abbreviations, different languages, synonyms etc.	Data presentation inconsistencies can lead to unexpected editing results in the pre-processing phase and could create multiple data categories referring to the same entries or entry categories.
Volume	Adequate quantity of data should be available.	Scarcity of data could lead to unstable ML models, limited range of application or unreliable predictions [28].

Considering the influence of data quality on ML performance, this initial quality assessment should be tailored to the needs of the specific application. Different models have different requirements and tolerance towards quantity and quality of training data [24]. Regression is an example where data scarcity can lead to overfitting or model imbalances. When the available training data is riddled with the above weaknesses, strong domain prior knowledge is essential to understand, assess, interpret and accept or reconsider the generated ANN predictions [29]. A structured regime of recording data, ensuring quality criteria are met at the outset, could mitigate the risks of over recording (increased time, upfront costs) or under recording (data scarcity, noise, missing or skewed data or delays and associated data re-collection costs) [30].

4 Conclusions

This study concentrated on the implications of data availability and circularity when attempting to develop ML models predicting the thermal behaviour of concrete exposed to elevated temperatures. Developing these models necessitates a substantial amount of tabulated data. To ensure functionality, efficiency and reliability of the ML model, the data needs to be consistent, complete and balanced.

A significant number of experimental studies exist, reporting on the thermal behaviour of concrete exposed to fire or elevated temperatures. They report notable results and observations through schematics, diagrams and mathematical models.

However, this wealth of information and data is generally inaccessible for further use in Machine Learning research. The datasets are unavailable or difficult to access. There is also high variation in how and what data is recorded, necessitating convoluted pre-processing, performance or scope compromise or even rendering the intended research not possible.

It becomes apparent that action is needed to ensure that existing scientific knowledge becomes reusable in the context of building fire design and Machine Learning development. Possible actions would include, but not be limited to:

- Develop data sharing and storing protocols for fire testing results.
- Establish publicly available datasets drawing information from real time data collection sensors in existing buildings or lab experiments.
- Standardise the way data is recorded, labelled and enriched with meta-data.
- Obtain feedback through reviewing current fire research community needs and align data recording, reporting and sharing accordingly.

Considering that ML and AI in general are becoming indispensable tools in research, a general mindset shift might be necessary, where data curation and circulation become deliverables on their own. Future research could focus on preprocessing specific datasets, assessing existing knowledge gaps and establishing methodologies for effective use of existing knowledge. Standardinsing methods for tackling said gaps with bootstrapping techniques, FE modelling, and crossover review of existing experimental sets would be another essential future research avenue.

References

1. Ghisellini, P., Cialani, C., Ulgiati, S.: A review on circular economy: the expected transition to a balanced interplay of environmental and economic systems. J. Clean. Prod. **114**, 11–32 (2016). https://doi.org/10.1016/j.jclepro.2015.09.007
2. Despotovic, M., Glatschke, M.: Challenges and opportunities of artificial intelligence and machine learning in circular economy (n.d.)
3. Nahmadova, F.: Data circularity: how circular data ecosystems can be leveraged for better decision-support (n.d.)
4. Gebru, T., et al: Datasheets for datasets (2018)
5. Wilkinson, M.D., et al.: Comment: the FAIR Guiding Principles for scientific data management and stewardship. Sci. Data. **3**, 160018 (2016). https://doi.org/10.1038/sdata.2016.18
6. Giner-Miguelez, J., Gómez, A., Cabot, J.: On the readiness of scientific data papers for a fair and transparent use in machine learning. Sci. Data. **12**(1), 440 (2025). https://doi.org/10.1038/s41597-025-04402-4
7. Bakas, I., Kontoleon, K.J.: Meta-narrative review of artificial intelligence applications in fire engineering with special focus on heat transfer through building elements. Fire. **6**(7), 261 (2023). https://doi.org/10.3390/fire6070261
8. Asadi, I., Shafigh, P., Abu Hassan, Z.F., Mahyuddin, N.B.: Thermal conductivity of concrete – a review. J. Build. Eng. **20**, 81–93 (2018). https://doi.org/10.1016/j.jobe.2018.07.002
9. Kim, K.H., Jeon, S.E., Kim, J.K., Yang, S.: An experimental study on thermal conductivity of concrete. Cem. Concr. Res. **33**(3), 363–371 (2003). https://doi.org/10.1016/S0008-8846(02)00965-1
10. Lucio-Martin, T., Roig-Flores, M., Izquierdo, M., Alonso, M.C.: Thermal conductivity of concrete at high temperatures for thermal energy storage applications: experimental analysis. Sol. Energy. **214**, 430–442 (2021). https://doi.org/10.1016/j.solener.2020.12.005
11. Al-Sibahy, A., Edwards, R.: Thermal behaviour of novel lightweight concrete at ambient and elevated temperatures: experimental, modelling and parametric studies. Constr. Build. Mater. **31**, 174–187 (2012). https://doi.org/10.1016/j.conbuildmat.2011.12.096
12. Liu, K., Lu, L., Wang, F., Liang, W.: Theoretical and experimental study on multi-phase model of thermal conductivity for fiber reinforced concrete. Constr. Build. Mater. **148**, 465–475 (2017). https://doi.org/10.1016/j.conbuildmat.2017.05.043
13. Xue, C., Yu, M., Xu, H., Xu, L., Saafi, M., Ye, J.: Experimental study on thermal performance of ultra-high performance concrete with coarse aggregates at high temperature. Constr. Build. Mater. **314** (2022). https://doi.org/10.1016/j.conbuildmat.2021.125585
14. Kodur, V.K.R., Sultan, M.A.: Effect of temperature on thermal properties of high-strength concrete (n.d.). http://irc.nrc-cnrc.gc.ca/ircpubs
15. Yu, M., et al.: Experimental and numerical investigation on thermal properties of alkali-activated concrete at elevated temperatures. J. Build. Eng. **74** (2023). https://doi.org/10.1016/j.jobe.2023.106924
16. Wang, W., Lu, C., Li, Y., Li, Q.: An investigation on thermal conductivity of fly ash concrete after elevated temperature exposure. Constr. Build. Mater. **148**, 148–154 (2017). https://doi.org/10.1016/j.conbuildmat.2017.05.068
17. Shen, L., Ren, Q., Zhang, L., Han, Y., Cusatis, G.: Experimental and numerical study of effective thermal conductivity of cracked concrete. Constr. Build. Mater. **153**, 55–68 (2017). https://doi.org/10.1016/j.conbuildmat.2017.07.038
18. Mirzanamadi, R., Johansson, P., Grammatikos, S.A.: Thermal properties of asphalt concrete: a numerical and experimental study. Constr. Build. Mater. **158**, 774–785 (2018). https://doi.org/10.1016/j.conbuildmat.2017.10.068

19. Martínez-Martínez, J.E., Álvarez Rabanal, F.P., Lázaro, M., Alonso-Martínez, M., Alvear, D., Del Coz-Díaz, J.J.: Assessment of lightweight concrete thermal properties at elevated temperatures. Appl. Sci. **11**(21), 10023 (2021). https://doi.org/10.3390/app112110023
20. Chen, Y., et al.: Experimental study on the thermal properties of a novel ultra-high performance concrete reinforced with multi-scale fibers at elevated temperatures. Constr. Build. Mater. **366**, 130229 (2023). https://doi.org/10.1016/j.conbuildmat.2022.130229
21. Xue, C., Su, C., Yu, M., Xu, H., Ye, J.: Experimental study on the thermal-mechanical properties and degradation of sleeve grouting material at elevated temperatures. J. Mater. Civ. Eng. **33**(2), 04020443 (2021). https://doi.org/10.1061/(asce)mt.1943-5533.0003493
22. Deng, J., Cui, J., Surahman, R., Tu, M., Wang, Y.: Experimental and analytical study on heat transfer of concrete with different degrees of saturation under elevated temperatures. Fire Mater. **49**(3), 269–279 (2025). https://doi.org/10.1002/fam.3270
23. Xie, J., Sun, L., Zhao, Y.F.: On the data quality and imbalance in machine learning-based design and manufacturing—a systematic review. Engineering. **45**, 105–124 (2025). https://doi.org/10.1016/j.eng.2024.04.024
24. Mohammed, S., et al.: The effects of data quality on machine learning performance on tabular data. Inf. Syst. **128**, 102549 (2025). https://doi.org/10.1016/j.is.2025.102549
25. Zhou, Y., Aryal, S., Bouadjenek, M.R.: Review for handling missing data with special missing mechanism (2024). http://arxiv.org/abs/2404.04905
26. Liu, Q., Ma, W.: Navigating data corruption in machine learning: Balancing quality, quantity, and imputation strategies (2025). http://arxiv.org/abs/2412.18296
27. Gupta, S., Gupta, A.: Dealing with noise problem in machine learning data-sets: a systematic review. In: Procedia Computer Science, pp. 466–474. Elsevier BV (2019). https://doi.org/10.1016/j.procs.2019.11.146
28. Bakas, I., Kontoleon, K.J.: Performance evaluation of artificial neural networks (ANN) predicting heat transfer through masonry walls exposed to fire. Appl. Sci. **11**(23), 11435 (2021). https://doi.org/10.3390/app112311435
29. Naser, M.Z.: A review of machine learning with small and limited data. J. Big Data. **13**, 18 (2026). https://doi.org/10.1186/s40537-025-01346-9
30. Mahmood, R., Lucas, J., Alvarez, J.M., Fidler, S., Law, M.T.: Optimizing data collection for machine learning (n.d.)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits any noncommercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if you modified the licensed material. You do not have permission under this license to share adapted material derived from this chapter or parts of it.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

