# Evaluating the Implementation of Machine Learning Techniques in the South African Built Environment

Peter Adekunle[1], Cliton Aigbavboa[1], Kenneth Otasowie[1], Matthew Ikuabe[1]

[1] acidb Centre for Excellence & Sustainable Human Settlement and Construction Research Centre, Faculty of Engineering and the Built Environment, University of Johannesburg, South Africa.
Adekunlepeter90@gmail.com

## Abstract

The future of machine learning (ML) in building may seem like a general idea that will take decades to materialize, but it is far closer than previously believed. In reality, the built environment has progressively increased interest in machine learning. Although it could appear to be a very technical, impersonal approach, it can make things more personable. Instead of eliminating humans from the equation, machine learning allows people to do their real work more efficiently. It is, therefore, vital to evaluate the factors influencing the implementation of machine learning techniques in the South African built environment. The study's design was one of a survey. In South Africa, construction workers and professionals were given one hundred fifty (150) questionnaires, of which one hundred and twenty-four (124) were returned and deemed eligible for the study. The collected data were analyzed using percentages, mean item scores, standard deviation, and Kruskal-Wallis. The results demonstrate that the top factors influencing the adoption of machine learning are knowledge level and a lack of understanding of its potential benefits. In comparison, lack of collaboration among stakeholders and lack of tools and services are the key hurdles to deploying machine learning within the South African built environment. The study concluded that ML adoption should be promoted to increase safety, productivity, and service quality within the built environment.

## Keywords

Machine learning, Implementation, Built environment, Construction Stakeholders

## 1. Introduction

Machine learning (ML) and artificial intelligence (AI) are two examples of digital tools. The whole point of ML is to use computers or other technologies to perform activities that often need human thought (Brynjolfsson & Mitchell, 2017). Moreover, machine learning, a subset of AI, recognizes data trends using mathematical algorithms based on statistical models to make predictions or provide descriptions (Sarker, 2021). It's important to recognize that learning is the process of turning experience into knowledge that can be applied generally. The concept of machine learning was highlighted by Kudelina et al., (2021) in that a computer is programmed using a particular mathematical model (fit on example data or historical data), developed using statistical techniques, to automatically suggest solutions to the user when problems arise, without the user having to provide solutions or anticipate them for all situations. Additionally, these mathematical algorithms are computer-understood instructions that tell computers how to convert inputs into outputs. As a result, these algorithms make sure that statistical models are fitted to data in order to explain the data and produce useful information that can be applied in the real world.

Data-driven machine learning techniques are not a recent development. The fundamental ideas behind ML rely on example data or previous data (Ihme et al., 2022). This data can be in a variety of formats, including video, sensor, text (both historical and current), and numerical. Several techniques that operate on a set of guiding principles can be used interchangeably on data and are not constrained to a particular format. There are primarily two uses for these machine learning techniques (Wan et al., 2020). This is specifically for learning from historical data and making predictions (or learning) about current occurrences based on that knowledge. Moreover, there are two categories that can be used to differentiate between various machine learning algorithms: supervised learning (SL) and unsupervised learning (UL) (Siam et al., 2019). Algorithms that fall under the SL category must learn how to map inputs to their appropriate outputs using a labelled dataset that is provided by a supervisor. Similar to this, algorithms in UL are simply supplied input without a supervisor. Regressions and classification are SL techniques, but clustering is an unsupervised learning technique. Regression and clustering are the machine learning techniques most frequently employed in the construction sector, according to Bilal et al., (2016). These techniques are frequently used in conjunction with other machine learning (ML) techniques including natural language processing (NLP) and information retrieval, however depending on the data that is available, ML techniques can also be employed alternately.

Since the adoption of ML approaches in developing nations like South Africa has been obviously lagging, numerous studies on the application of ML algorithms in the construction industry have been done remarkably in rich countries (Carbonneau et al., 2008; Musumeci et al., 2019) . Also, substantial study was done on the use of ML approaches in construction (Mirzaei et al., 2022; Baduge et al., 2022). This can be divided into two distinct parts. The first part focuses more on knowledge extraction by creating a program that can be used to extract and acquire key details about different accidents and their attributes. The second part involves using the ML method to predict the nature of accidents and the potential types of injuries they may cause. Natural language processing tool (NLP), a machine learning algorithm, can be used to extract knowledge from accident data. Use machine learning techniques, such as random forest (RF) and stochastic gradient boosting (SGTB), which employ an algorithm to build regression and classification models to predict the type of injury and the type of energy. The system can also be used to forecast which body part, under specific conditions, is most likely to sustain damage. With this understanding, safety precautions can be adopted and applied correctly.

In a review, Hegde & Rokseth, (2020) provided examples of standard machine learning (ML) approaches that can be used to support risk assessment in the construction sector. These approaches include classification algorithms like ANN, SVM, DTs, and others. Input data sources for ML models used in risk assessment in the construction industry were described in the paper. The Occupational Health and Safety Act (OHSA) provided the following text-based injury and incident reports, videos, photographs, and accident data (Fang et al., 2020). The use of other algorithms including stochastic regression, multiple regressions, evolutionary algorithms, and nave Bayes in the construction industry was comprehensively defined by Baduge et al., (2022). Unsupervised learning has been used in various construction-related applications, despite the fact that the majority of research in the field has centered on supervised learning techniques. By examining how machine learning methods are used in the South African built environment, this work advances the study of machine learning approaches.

## 2. Methods

This study was started to assess how machine learning techniques were being used in the built environment. The use of a quantitative methodology inspired the survey design, which led to the use of a structured questionnaire survey. Closed-ended questions designed to elicit feedback on the application of machine learning techniques in the construction industry were utilized in the structured questionnaire created by Adekunle et al., (2022). The study's participants are academics and professionals from the South African construction sector who work for various organizations. For data gathering within the study population to obtain wider coverage, the study used random sampling and a snowball sampling approach. The questionnaire In order to obtain enough data, one hundred and thirty-five (135) questionnaires were retrieved after being distributed from April 2022 to October 2022, and each one was evaluated for analytic readiness. The questionnaire was split into two sections: one half evaluated the respondents' demographics, and the other gave respondents access to a myriad of factors impacting the adoption of ML as gleaned from a thorough assessment of the literature. The dependability of the research tool was assessed using Cronbach's alpha. The study's research instrument's reliability was demonstrated by the value of 0.875 that was obtained.

## 3. Results

According to the respondents' backgrounds, 26% of them were academicians, 10% were architects, 7% were electrical and electronics engineers, 10% were mechanical engineers, 17% were civil engineers, 14% were quantity surveyors, and 16% were project managers for construction. 30% of respondents have one to five years' experience, 54% have six to ten years' experience, and 16% have between eleven and fifteen years' experience.

The parameters impacting the adoption of machine learning in the built environment are displayed in table 1 below, going from greatest mean to lowest mean. As can be seen, measures with the same mean were ranked according to how much they deviated from it (standard deviation, or SD). As compared to the mean value of the set, the standard deviation (SD) measures the degree of scattering in a set of data. A big SD is the consequence of data that deviates significantly from the mean, whereas a small SD indicates that all of the values are relatively close to the mean. In order to compare respondents' perspectives based on their years of experience, a Kruskal-test Walli's was also done. The complexity of ML results and trend visualisation, two factors impacting the adoption of ML in the South African built environment, were shown to not significantly deviate from the mean values. This is due to the fact that the p-values are higher than 0.05, which is consistent with a study by Greenland et al., (2016) that found no significant difference between variables with p-values more than 0.05 and vice versa. The mean values of each group of respondent are statistically different for the other parameters.

**Table 1.** Factors influencing adoption of Machine Learning

| Identified factors | $\bar{x}$ | $\sigma X$ | R | P-Values |
|---|---|---|---|---|
| The Level of knowledge of ML methods | 4.04 | 0.947 | 1 | 0.048 |
| Potential benefits uncertainty | 3.88 | 0.895 | 2 | 0.000 |
| Characteristics of a dataset | 3.86 | 0.857 | 3 | 0.000 |
| Complex nature of ML results | 3.86 | 0.808 | 4 | 0.629 |
| The level of formal education | 3.86 | 0.923 | 5 | 0.032 |
| The choice of ML technique to use | 3.86 | 0.948 | 6 | 0.028 |
| Accuracy | 3.86 | 1.050 | 7 | 0.045 |
| Level of trustworthiness in the ML algorithm | 3.80 | 0.833 | 8 | 0.000 |
| The Heterogeneous nature of data | 3.80 | 0.969 | 9 | 0.000 |
| Understanding digital technology | 3.80 | 1.030 | 10 | 0.013 |
| Data quality | 3.76 | 1.098 | 11 | 0.005 |
| Trend visualisation | 3.72 | 0.882 | 12 | 0.088 |
| User interface of ML tools | 3.70 | 0.863 | 13 | 0.049 |
| Time spent on developing suitable model | 3.64 | 1.005 | 14 | 0.000 |
| Unjustifiable benefits | 3.62 | 1.028 | 15 | 0.000 |

## 4. Discussion

The results of this study indicate that significant aspects mainly related to the types of data formats used in the construction business, the degree of expertise, and the opinion of professionals regarding machine learning techniques. According to Bilal et al., (2016) data in the construction industry come in a variety of formats and frequently need to be cleaned. This may have a significant impact on how machine learning is used and implemented, which in turn affects adoption. This is because experts who are unfamiliar with ML approaches won't know which data to utilize them on. Which will then have an impact on how others perceive them, and which can affect how extensively they are adopted in the building sector. It's crucial to keep in mind that for any organization at any given time, the trade-off analysis is not between implementing or not implementing a new technology or process (as in the case of ML techniques); rather, it is between implementing it right away or delaying that decision until a later time (Rana et al., 2014). This distinction is crucial since adoption is influenced by organizational and environmental factors as well as direct advantages and limitations of a particular technology at a given period. The survey also showed that enterprises now have access to a sizable and continually expanding amount of data that can be used to improve processes, goods, and services by providing helpful insights (Ranjan & Foropon, 2021). Machine learning techniques have a great potential to help businesses with this. The use of these strategies in industry is now far from ideal, despite evidence of their value in academia and the availability of instruments. This study shows that in order to hasten the adoption of ML-based approaches in industry, it is necessary to improve our comprehension of the information requirements of industry in this area. According to Zhang et al., (2021), a technology acceptance model offers a practical method to achieve this goal. Labor-intensive industries like construction run counter to modern trends like an aging population. The greatest obstacles to the construction industry's profitability are labor cost losses and a shortage of on-site personnel as a result of the demographic dividend's progressive absence and rising labor expenses. Using machine learning and other artificial intelligence technologies to increase the level of automation in the construction industry is an unavoidable development given these circumstances (Kuang et al., 2021). While evaluating the use of machine learning in construction, there are some key factors to take into account. First off, since the technology in this area is already relatively advanced and the vision-related component is the most commonly used component in the construction sector, the field of computer vision will advance quickly (Ding et al., 2022). Due to the consistent nature of the tasks that object detection and action recognition models are trained for, they typically perform well in subsequent projects. Deep learning will become more significant as safety issues in construction or automatic detection become more prevalent; shallow learning will still be used for processing before deep learning is used. Second, data is essential for machine learning applications in the construction industry (Vadyala et al., 2022). Establishing a public data source for the building sector is crucial. A construction-related dataset could have a similar impact on research in the field of construction automation as the general-purpose dataset ImageNet, which is similar in design. Construction stakeholders can concentrate more on deep learning algorithms with these kinds of open data sets.

## 5. Conclusions

The first conference on the topic of machine learning took place in 1956, so the field of study is hardly new. Nonetheless, it has gained greater attention in the last ten years and has a significant impact on the advancement of automated technologies in the built environment. However, due to factors including knowledge level, ambiguity about possible advantages, and dataset features outlined in the study, implementation has only had a limited amount of success. To power the insights that may be gleaned from ML, a large amount of data must be present, and over the last few years, this amount of data has increased dramatically. Throughout the past few years, there has been significant investment in construction technology. A significant percentage of that expenditure was used to digitize various processes involved in the construction cycle. Building information modeling (BIM) models have altered how structures are planned, project management, issue management, and operations management processes have all moved to the cloud and are becoming more sensorised and automated. The utility of ML-based applications will increase with the availability of data in the construction industry. It is possible to automatically rank topics using ML, specifically construction language analysis. The algorithms are able to comprehend and forecast complex situations, such as whether a problem could lead to an intrusion if it is not resolved. It is advised that ML approaches be adopted because they will be crucial to solving important environmental and physical modeling issues in the built environment in the future of scientific modeling.

## References


Adekunle, P., Aigabvboa, C., Thwala, W., Akinradewo, O., & Oke, A. (2022). Challenges confronting construction information management. *Frontiers in Built Environment*, *8*(December). https://doi.org/10.3389/fbuil.2022.1075674

Baduge, S. K., Thilakarathna, S., Perera, J. S., Arashpour, M., Sharafi, P., Teodosio, B., Shringi, A., & Mendis, P. (2022). Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications. *Automation in Construction*, *141*(June), 104440. https://doi.org/10.1016/j.autcon.2022.104440

Bilal, M., Oyedele, L. O., Qadir, J., Munir, K., Ajayi, S. O., Akinade, O. O., Owolabi, H. A., Alaka, H. A., & Pasha, M. (2016). Big Data in the construction industry: A review of present status, opportunities, and future trends. *Advanced Engineering Informatics*, *30*(3), 500–521. https://doi.org/10.1016/j.aei.2016.07.001

Brynjolfsson, E., & Mitchell, T. (2017). *What can machine learning do? Workforce implications* (Vol. 358, Issue 6370). https://doi.org/10.1126/science.aap8062 View

Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research*, *184*(3), 1140–1154.

https://doi.org/10.1016/j.ejor.2006.12.004

Ding, Y., Ma, J., & Luo, X. (2022). Applications of natural language processing in construction. *Automation in Construction*, *136*(September 2021), 104169. https://doi.org/10.1016/j.autcon.2022.104169

Fang, W., Ding, L., Love, P. E. D., Luo, H., Li, H., Peña-Mora, F., Zhong, B., & Zhou, C. (2020). Computer vision applications in construction safety assurance. *Automation in Construction*, *110*(December 2019), 103013. https://doi.org/10.1016/j.autcon.2019.103013

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology*, *31*(4), 337–350. https://doi.org/10.1007/s10654-016-0149-3

Hegde, J., & Rokseth, B. (2020). Applications of machine learning methods for engineering risk assessment – A review. *Safety Science*, *122*(October 2019), 104492. https://doi.org/10.1016/j.ssci.2019.09.015

Ihme, M., Chung, W. T., & Mishra, A. A. (2022). Combustion machine learning: Principles, progress and prospects: Combustion machine learning. *Progress in Energy and Combustion Science*, *91*(March 2021), 101010. https://doi.org/10.1016/j.pecs.2022.101010

KUANG, L., LIU, H., REN, Y., LUO, K., SHI, M., SU, J., & LI, X. (2021). Application and development trend of artificial intelligence in petroleum exploration and development. *Petroleum Exploration and Development*, *48*(1), 1–14. https://doi.org/10.1016/S1876-3804(21)60001-0

Kudelina, K., Vaimann, T., Asad, B., Rassõlkin, A., Kallaste, A., & Demidova, G. (2021). Trends and challenges in intelligent condition monitoring of electrical machines using machine learning. *Applied Sciences (Switzerland)*, *11*(6). https://doi.org/10.3390/app11062761

Mirzaei, K., Arashpour, M., Asadi, E., Masoumi, H., Bai, Y., & Behnood, A. (2022). 3D point cloud data processing with machine learning for construction and infrastructure applications: A comprehensive review. *Advanced Engineering Informatics*, *51*(March 2021), 101501. https://doi.org/10.1016/j.aei.2021.101501

Musumeci, F., Rottondi, C., Corani, G., Shahkarami, S., Cugini, F., & Tornatore, M. (2019). *A Tutorial on Machine Learning for Failure Management in Optical Networks*. *37*(16), 4125–4139.

Rana, R., Staron, M., Hansson, J., Nilsson, M., & Meding, W. (2014). A framework for adoption of machine learning in industry for software defect prediction. *ICSOFT-EA 2014 - Proceedings of the 9th International Conference on Software Engineering and Applications*, 383–392. https://doi.org/10.5220/0005099303830392

Ranjan, J., & Foropon, C. (2021). Big Data Analytics in Building the Competitive Intelligence of Organizations. *International Journal of Information Management*, *56*(August 2020), 102231. https://doi.org/10.1016/j.ijinfomgt.2020.102231

Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, *2*(3), 1–21. https://doi.org/10.1007/s42979-021-00592-x

Siam, A., Ezzeldin, M., & El-Dakhakhni, W. (2019). Machine learning algorithms for structural performance classifications and predictions: Application to reinforced masonry shear walls. *Structures*, *22*(June), 252–265. https://doi.org/10.1016/j.istruc.2019.06.017

Vadyala, S. R., Betgeri, S. N., Matthews, J. C., & Matthews, E. (2022). A review of physics-based machine learning in civil engineering. *Results in Engineering*, *13*(October 2021), 100316. https://doi.org/10.1016/j.rineng.2021.100316

Wan, S., Qi, L., Xu, X., Tong, C., & Gu, Z. (2020). Deep Learning Models for Real-time Human Activity Recognition with Smartphones. *Mobile Networks and Applications*, *25*(2), 743–755. https://doi.org/10.1007/s11036-019-01445-x

Zhang, X., Wang, Y., & Li, Z. (2021). User acceptance of machine learning models – Integrating several important external variables with technology acceptance model. *International Journal of Electrical Engineering Education*, *3*. https://doi.org/10.1177/00207209211005271

.