MACHINE LEARNING FOR ANALYSIS OF OCCUPATIONAL ACCIDENTS REGISTRATION DATA

May Shayboun[1], Dimosthenis Kifokeris and Christian Koch

Division of Construction Management, Department of Architecture and Civil Engineering, Chalmers University of Technology, Sven Hultinsgatan 6, Gothenburg, Sweden 41296, Sweden

Regardless of the efforts of employers and public organizations to eliminate occupational accidents, the latter is a persistent problem in the construction industry. In the Swedish construction context, there is a desire to identify causes and factors playing a role in work-related accident prevention, as there are large underused databases of collected registrations that represent knowledge on causes and the context of accidents. The aim of the current contribution is to review the application of machine learning (ML) in the improved prevention of accidents and corresponding injuries, and to identify current limitations - and most importantly to answer the question of whether ML actually reveals more than what is currently known about accidents in construction. A systematic literature review on the use of ML for analysing data of accident records was carried out. In the reviewed literature, ML was applied in the prediction of accidents or their outcome, and the extraction or identification of the causes affecting the risks of injuries. ML combined with data mining (DM) techniques such as Natural Language Processing and graph mining, appears to be beneficial in discovering associations between different features and in multiple levels of clusters. However, the literature shows that research on ML in accident prevention is at an early stage. The review of the literature indicates gaps in the justification of methodological choices, such as the choice of ML method and data processing. Moreover, characteristics of the injury rates and severity are shown to be clashing with the mechanisms of the ML classification algorithms. This should probably lead to abandoning severity as a parameter and changing the approach towards the asymmetric data classes (denoted "unbalanced" in ML methodology), leaving space for finding the important causes. An overreliance on internal validity testing and lack of external testing of the algorithms' performance and prediction accuracy persists. Future research needs to focus on methods addressing the problem of data processing, explaining the choice of methods, explaining the results (especially the variance in ML algorithm's performance), merging different data sources, considering more attributes (such as risk management), applying deep learning algorithms, and improving the testing accuracy of ML models.

Keywords: accident registration, Machine Learning, occupational accident prevention

INTRODUCTION

Maintaining a safe workplace and reducing the frequency of serious accidents are continuous important quests in the Swedish and international construction industries. In Sweden, reports show that occupational accidents and near accidents continue to hinder productivity (Berglund *et al.,* 2017). The downward trend has levelled since

[1] shayboun@chalmers.se

Shayboun, M, Kifokeris, D and Koch, C (2020) Machine Learning for Analysis of Occupational Accidents Registration Data *In:* Scott, L and Neilson, C J (Eds) *Proceedings of the 36th Annual ARCOM Conference,* 7-8 September 2020, UK, Association of Researchers in Construction Management, 485-494

2010, but the number of occupational accidents increased in accordance with the increase in the number of employees in the construction industry (frequency of 10.8 accidents/1000 workers) (Samuelsson 2019). In the Swedish context, these low figures are compared to a zero-accident goal, as for example purported by the association "Keep the Zero" (Håll Nollan 2020). Simulation of the workplace to identify various possible risk factors and dangerous situations in advance are two out of many possible tracks of technical solutions (Berglund *et al.,* 2017). There is an opportunity to exploit the large number of gathered registrations of accidents and near accidents in the domain of large contractors. In addition to that, Artificial Intelligence subdivisions such as ML and DM have increasingly been applied in finding underlying patterns and increasing the predictability of the risk of occupational accidents (Bilal *et al.,* 2016, Vallmuur 2015, Hegde and Rokseth 2020). Regardless, Vallmuur (2015) criticized the approach of analysing textual injury records for lacking the description of methodologies in processing the data and training the ML models. While Bilal and Oyedele (2020) suggested that the application of ML surpasses the development of a prototype and reliable models need to be trained based on informed decisions instead of only the expertise and intuition of engineers. Shayboun *et al.,* (2019) concluded that responsibility in action-taking, accountability in decision-making, and the continued crucial need for human reasoning, are important considerations to be taken. The literature review of the current effort is part of a project aiming at applying ML to a data record registered by a large contracting company in Sweden. The aim of this effort is to analyse the current literature in the application of ML to accident reports in the construction industry context. The analysis discretizes the reviewed articles into the following themes: the methodological choices of the ML algorithms, data collection and pre-processing, training of the system and validation performance, analysis and implementation of results, and the managerial implications. Moreover, the question of whether ML can reveal more than what is currently known about accidents in construction is raised.

The paper is structured according to its status as a literature review commencing with a method and continuing with a themed review. It then proceeds by synthesizing the findings in the discussion, followed by a conclusion.

METHOD

The literature review was conducted using the concept-centric framework augmented by units of analysis (Webster and Watson 2002), and it was based on a search regarding the application of ML to the analysis of accident registries in the construction sector. First, a list of relevant journals was prepared (Webster and Watson 2002), namely Safety Science, International Journal of Occupational Safety and Ergonomics, and Automation in Construction. Then, the concepts of the literature review emerged from using the search terms "occupational accidents", "accident prevention", "machine learning" and "construction projects". This framework was strengthened by using the references-of-references and "snowballing" techniques (Greenhalgh and Peacock 2005) and aiming at a targeted but still comprehensive search (MacLure 2005).

The review was conducted in iterations; 169 publications were abstract-scanned, 54 fully read, and 7 were finally selected for the current effort. The main reason for selecting these few papers was that an in-depth elaboration on central studies in the cross-section of the aforementioned concepts was sought, rather than the accumulation of references that might be peripheral. This work is also is a preliminary part of a project aiming at applying ML to registered accidents by a large contracting company in Sweden. Due to the latter reasons, the literature search started with a broad scope of review and search words. However, the selection of papers for this effort was at a narrow-targeted list that is only related to the use of ML on reported accidents' data in the construction industry.

The themes emerging from the selected ML articles were organized in the following: choice of algorithms, data collection and pre-processing, training of the system and validation performance, analysis and implementation of results, and the managerial implications of the ML modelling. The iterations of the literature review and the emergence of the aforementioned themes followed the abductive reasoning of qualitative research, where observations and explanations of phenomena are developed by working iteratively between theory and data (which, in the present case, is the content of the references accumulated with each iteration), thus facilitating the revision and refinement of earlier conceptions (Bell *et al.,* 2019).

LITERATURE REVIEW

Choice of Algorithm

The literature shows that ML has been methodologically applied for different purposes, with this playing a significant role in the choice of the utilized algorithm(s). For example, Choi *et al.,* (2020) used a variety of algorithms they deemed suitable for forecasting purposes (such as logistic regression (LR), and AdaBoost), in order to predict the likelihood of fatality; Poh *et al.,* (2018) used classifiers (e.g. random forest (RF)) for the level of severity of accidents; Tixier *et al.,* (2016b) used algorithms such as the stochastic gradient tree boosting (STGB) to predict the type of energy involved in the accident, the injury type, the affected body part, and the severity of the injury; and Ayhan and Tokdemir (2019) used artificial neural networks (ANNs) for accident outcome prediction. Moreover, a natural language processing (NLP) algorithm was applied as a method for extracting features from textual data in accident injury reports (Zhang *et al.,* 2019), while Tixier *et al.,* (2016a) developed their own NLP algorithm. DM was also applied in conjunction with ML to graphically view groups of attributes that together lead to risky situations (Tixier *et al.,* 2017).

The predictions related to accidents were mostly treated as a classification problem. Popular classification algorithms were used both in single and ensemble learning models, and included, indicatively, k-nearest neighbour (KNN) and support vector machines (SVM) (Zhang *et al.,* 2019). The ensemble model outperformed all single classifiers with an accuracy of 68% in classifying 11 causes of accidents using the data extracted with NLP (Zhang *et al.,* 2019). On the other hand, the choice of Tixier *et al.,* (2016b) on using both RF and SGTB was explained by the authors to be based on the logic that the purpose of the research is to test if fundamental construction attributes can be used in predicting safety-related outcomes, while simultaneously there is a lack of general rule on which algorithm is better than the other.

Data Collection and Pre-Processing

The data used in Zhang *et al.,* (2019) consisted of 16323 accident reports related to construction sites that are registered in the Occupational Safety and Health Administration (OSHA), collected between 1983 and 2016. But this data was not labelled (i.e. the instances were not initially attributed into the sets of specific classes), and the selection of the dataset shifted to 1000 labelled records published by previous research. Choi *et al.,* (2020) collected the data through the Ministry of Employment and Labour in the Republic of Korea. The dataset contained 137323 injuries and 2846 deaths, and included information about the age, sex, length of service for each injured worker, the type of construction, employer scale, and the data of the accident. However, Choi *et al.,* (2020) encountered limitations in their access to certain parts of the dataset, due to the related data protection law in the Republic of Korea. Poh *et al.,* (2018) had the advantage of expanding the data type. The data covered 27 construction projects from a single contracting company in Singapore (consisting of 19 building projects and 8 infrastructure projects) over a period of seven years (from 2010 to 2016); it included 785 safety monthly inspection records, 418 accident cases, and their corresponding monthly project-related attributes. Another approach was to collect data through a structured template (Ayhan and Tokdemir 2019). The templates consisted of six categories for accident causes, such as human factors, workplace factors, the course of an accident, and time of occurrence.

As described above, there can be challenges and variations in the data collection, size, variety, and structure (e.g. labelled vs unlabelled instances). Critically, these challenges also tie with challenges in handling and processing the data. A common practice found in the accident registry analysis is the handling and processing of textual data, as can be found in e.g., Tixier *et al.,* (2016a). Zhang *et al.,* (2019) applied NLP to extract the causes of accidents and the objects which contributed to the accidents, from labelled accident reports - but found that the performance of the NLP was not achieving the full potential of extraction. The result of the classifier was not very satisfactory with an accuracy of only 68%; Zhang *et al.,* (2019) explained this result by arguing that natural language is not precise and that developing comprehensive rules to cover all meanings of different expressions is not feasible. To avoid the previous limitations of NLP, Tixier *et al.,* (2016a) developed a new domain-related NLP algorithm to automatically scan and extract features from unstructured accident records. The motivation of programming a new NLP algorithm over the existing ones was that it was based on hand-coded rules and dictionaries of keywords related to the domain of accidents, which resulted in higher levels of accuracy. Tixier *et al.,* (2016a) reached a precision of 95% in scanning 80 attributes, 7 injury types, 9 energy sources, and 5 body parts, after having a team of relevant experts review the algorithmic results in order to find true positives, false positives, and false negatives.

The challenge in handling and processing data does not end at the extraction of features from textual data but can also be found in the data featuring classes with a large variation in the number of instances they include (so-called "unbalanced" classes). This problem places a challenge in ML because the training of the model can fall short to recognize the more sparsely populated classes. Class variation in terms of volume of data has been found in the injury severity, energy type involved, and body parts injured. In Choi *et al.,* (2020), the injury data was represented 48 times more than the fatality data. The authors approached this as a challenge that needs to be tackled; in doing so, they suggested three methods for resampling: random oversampling (ROS), random under sampling (RUS), and the synthetic minority oversampling technique (SMOTE). They ended up choosing ROS because it was a better fit with the categorical values in the dataset. Poh *et al.,* (2018) also encountered the same phenomena, as the instances in the dataset included a total of 35 “Major Accident” cases, 336 “Minor Accident” cases, and 256 “No Accident” cases. The authors applied SMOTE to overcome this problem. In the dataset of Tixier *et al.,* (2016b), the class “pressure” in energy type and “neck” in body parts were disproportionately represented. Stratified oversampling was used to reduce this effect.

The disadvantage of oversampling methods is that they reduce the accuracy of the majority class. Tixier *et al.,* (2016b) looked for a balance in the overall error with resampling proportion tuning integrated into the parameter optimization of the algorithm. In general, the three cases of unbalanced classes were approached using some sort of oversampling technique without clear justification of this choice or deeper explanation of the implications. The critique of the notion of “unbalanced” when oversampling is used, lies in the implicit assumption that a phenomenon should generate balanced datasets, which is not the case in the causes and consequences of accidents. Moreover, the critique of oversampling is that the ML designer moves into an unknown ground by assuming similarities in different parts of the studied phenomenon. Future conceptual development of ML for accident analysis needs to correct these faulty assumptions and find ways where ML can support the understanding of accidents.

Training of the System and Validation Performance

In Tixier *et al.,* (2016b), SGTB outperformed RF, achieving high performance of the Rank Probability Skill Score (RPSS) in predicting the energy type. The superiority of the SGTB might be explained by the fact that it reduces error by reducing variance and bias, while RF only reduces variance. However, predicting injury severity was not successful. Either additional layers of attributes were required (such as the amount of energy released), or injury severity could be a result of random components of similar events.

In the work of Poh *et al.,* (2018) and Choi *et al.,* (2020), RF outperformed other classification algorithms, such as, indicatively, SVM, KNN, and AdaBoost. The classification by Poh *et al.,* (2018) into “No accident”, “Minor accident” and “Major accident”, achieved an accuracy of 78%, while in Choi *et al.,* (2020) the value of the Area Under the Receiver Operating Characteristic Curve (AUROCC) metric was 0.9198; this was considered as satisfactory, as the ideal value of AUROCC is 1.

Ayhan and Tokdemir (2019) chose to apply only ANNs and conventional multiple regression to predict the outcome of accidents. The conventional multiple regression failed compared to the ANNs, but 13 different iterations were tried. The ANNs’ performance was evaluated with R-square values and mean percentage errors. Considerable difference was found between training and testing accuracy, as the testing accuracy dropped by 50% for the fatality class.

Analysis and Implementation of Results

More methodological advancement can be found in the work by Tixier *et al.,* (2017), as they proposed the use of graph mining and hierarchal clustering on principle components (HCPC) to analyse 4387 injury reports. HCPC is an unsupervised data mining technique that groups observations into levels of clusters (Tixier *et al.,* 2017). The clusters were manually inspected to find relevant safety clashes and organized by the authors in main themes; e.g., the congested workplace and confined workplace combined increased the risk of many other different attributes. The injury reports were automatically scanned by the developed NLP algorithm of Tixier *et al.,* (2016a) for 80 binary attributes, which were also identified by previous research. The authors based their work on five different algorithms and analysed them in terms of centrality, closeness, and betweenness. To use the algorithms, the data was split into subsets of injury type, namely struck by or against, caught in or compressed, fall on the same or to lower level, overexertion, and exposure to harmful substance. The results were extensive; to mention a few, it was found that the improper procedure/inattention plays a central role in the misuse of tools such as hammer and rebar. This was explained by the authors by noting that the human-related errors were shared across all groups of attributes. Moreover, confined workspace, working at height, and scaffolding, were closely related, which implied that constrained working space was a potential environment for falling. Also, a strong centrality appeared between the piping and the unpowered tool, bolt, and steel sections, which implied that these were major contributors to the caught in between or compressed injuries. The close proximity of welding to improper body positioning and working overhead and scaffold, showed that the possibility for workers to be injured through exposure to a harmful substance, was amplified when adopting non-natural body positioning.

The fuzzy inference decision-making system of Ayhan and Tokdemir (2019) was tied up with three courses of action; if a lost workday or a fatality was predicted, the action was to stop construction, then check the method of statement of work to find the direct cause of the prospective incident, and finally set up a research team to seek out the root cause.

However, discussing principles of the implementation of results is not enough to capture the experience of employing the algorithms in the prediction of accidents or decision-making support. There is a need to investigate the users' experiences and the external validation of the ML models. Most importantly, it is crucial to understand the practical implication of applying the ML models for testing the accuracy of training, testing, and observing the change in workplace safety processes.

Managerial implications

The practical implication of the ML analysis and utilization of results is very important in understanding how construction can benefit from the research. Few authors discussed that; Tixier *et al.,* (2017) suggested that safety knowledge in the form of binary attributes can be used in combination with Building Information Modelling (BIM), as attributes can be assigned to physical elements and spaces to automatically identify and report potential hazards in the design phase. Poh *et al.,* (2018) suggested using the RF model in the cases where the input regarding each project in the company is used to predict the projects in which generate high risks. Choi *et al.,* (2020) conceptually presented a system based on access control systems on construction sites, where the data of age, length of service, construction type and the season were used as input. The model of fatality prediction could identify safety managers, workers, contractors, and work teams who were at a high risk of major accidents.

Ayhan and Tokdemir (2019) also acknowledged the vagueness of ANNs in understanding the results. Therefore, the fuzzy set theory was suggested to achieve a more trustworthy prediction. Specifically, the authors suggested a fuzzy inference method by comparing expert module and the prediction, then taking the worst outcome. As explained in the previous subsection, the fuzzy inference decision-making system was tied up with specific courses of action.

DISCUSSION

The evaluation metrics and performance of the models appear to be different between the literature efforts reviewed. This contributes to hindering the possibility for current and future benchmarking of results in comparison to previous efforts. The same applies to the choice of algorithms. The RF outperformed other classification algorithms (Poh *et al.,* 2018). However, the SGTB outperformed the RF classifier in

Tixier *et al.,* (2016b). Overall, the approach of choosing the classification algorithm is experimental and rarely coupled with an analysis for the logic behind choosing algorithms, and the reasoning for higher prediction capability in one algorithm compared to others.

The literature review shows a variation in the data sources and limitations related to the availability and the structure of the data. Zhang *et al.,* (2019) preferred to use a dataset that is 16 times smaller than the available one, due to the limitation of labelling. Although this option had an advantage in the applicable ML classification, it did not exploit the available large data sample. Increasing the sample size can potentially increase precision in the sample (Bell *et al.,* 2019). Although the size of the dataset is not the only indicator of quality, it is worth to investigate whether the types of accidents in smaller datasets are representative of the cases that are found in larger volumes of data, and how unlabelled data can be exploited in the first place, especially with high volume availability. Currently, it looks like lost potential.

Moreover, extracting features from free written textual data is found to be complicated and immature (Zhang *et al.,* 2019, Tixier *et al.,* 2016a). Although Tixier *et al.,* (2016a) showed that a domain-specific NLP is yielding promising results, the success of the algorithm is depending on the quality of reports and the quality of the textual data. It is not expected from the algorithm to detect misspelled or missing words and it is not known how the algorithm would perform if applied to extract features from a different set of data other than the one used for developing the algorithm. It was also noted by the authors that the quality of the reports used was high, as they were short and very well written. Ayhan and Tokdemir (2019) argued that structured templates for collecting data about occupational accidents have an advantage compared to the free text data collection. Carefully defined templates provide a ready categorization of attributes of work events compared to the free text that might be categorised by occupation health professionals offsite. Predefined templates might be advantageous from a pragmatic point of view, but it can also be argued that an unstructured form of accident reports can allow the possibility to attain further and deeper information and reduce bias.

Methodological issues related to unbalanced datasets were found in the literature (Choi *et al.,* 2020, Poh *et al.,* 2018, Tixier *et al.,* 2016b). This is a ML learning classification problem that clashes with the occupational accidents problem. The method choices for tackling the unbalanced classes are problematic both in the resampling techniques and the definition of the ML task. Multiple methods of resampling were applied by the authors, with little explanation of the implications or the disadvantages of using these methods. The resampling techniques increase the frequency of the underrepresented class, but assuming regularity of causes in areas that are less well known seems to be problematic. The elements for causes patterns are not automatically the same; for example, the risk of one machine is not the same as the risk of another, and the same accident outcome does not necessarily emanate from the same course of events. It is crucial to search for other techniques to manage the unbalance with minimum change to the data. There is also a need for metrics and evaluation for the common resampling techniques that are to be applied.

The focus of the reviewed literature was on the outcome of an accident such as predicting the likelihood of fatality (Choi *et al.,* 2020), classification of the severity level of an accident (Poh *et al.,* 2018), predicting the type of injury, body part affected, and the injury severity (Tixier *et al.,* 2016b), and the prediction of accident outcome (Ayhan and Tokdemir 2019). It can be learned from Tixier *et al.,* (2016b) that predicting severity level fails compared to the type of energy that is involved in an accident, a result consistent with accident research (Rollenhagen 2011). To distinguish between minor and major accidents is maybe influenced by the obligation of the industry to mostly record serious injuries (Oswald *et al.,* 2018). But severity might be less important information to predict compared to predicting the occurrence of an accident regardless of the outcome. It was suggested that the target of the ML modelling task is a crucial step in the design of applied ML (Bilal and Oyedele 2020). The unbalanced data can be mitigated by alternatively defining the goal of the ML model to focus on the risks associated with an accident instead of the accidents' outcome. This would give a better understanding of the distant events leading to the accident, instead of focusing on the outcome which might not be present in high frequency.

Few authors suggested the application of the ML models in decision making. However, theoretically discussing the implementation of results is not enough. To capture the experience of employing the algorithms in predicting accidents or decision-making support, implementation trials are required. Bilal and Oyedele (2020) suggest that applied ML methods include further steps other than modelling successful predictors. Interpretation and production deployment are the two main steps necessary for the evaluation of ML modelling. There is a need to investigate the users' experience and the external validation of the ML models. Most importantly, to understand the practical implication of applying the ML models for testing the accuracy of training and observing the change in the safety processes of the workplace. This would provide an indication of credibility and trust when the decision-making is supported by the ML recommendation.

In our problem statement, we questioned whether ML actually reveals more than what is currently known about accidents in construction. Realizing the broadness of the question, it can be posited that ML might add knowledge when comparing to local knowledge represented by health and safety professionals, because of the large volumes covered. At a time, local knowledge might be much richer in appreciating the complexity of causes and factors involved in actual accidents, provided that the local health and safety personnel have been involved in reporting. Comparing to research on causes behind accidents (Berglund 2017, Ringdahl 2013, Jørgensen 2002, Rollenhagen 2011) it appears that the deeper layers of causes, such as management strategy, industrial norms, contracts, and wage systems are poorly covered by the ML applications. A likely reason for this is the quality and character of the registered data used.

CONCLUSION

This contribution set out to review the application of ML for the improved prevention of accidents and related injuries and to identify current limitations. It was questioned whether ML reveals more than what is currently known about accidents in the construction domain. A systematic literature review on the use of ML for analysing accident records was carried out. The literature contains ML applications using data from registered accidents and their deployment in the prediction of accidents or their outcome. As the ML system intends to extract or identify causes affecting the risks of injuries, a series of ML and data mining techniques have been used. However, the research on ML in accident prevention is at an early stage. And there were identified gaps in the justification of methodological choices, such as the choice of ML method and data pre-processing; which appear to be of an experimental character (trial and error). Moreover, the characteristics of the accident's rates and severity showed to be clashing with approaches employed in the use of ML classification algorithms. The articulated need for "balancing" data according to severity of accidents should in the future be abandoned to the benefit of a focus on risks, as severity is a difficult, if not impossible analytical category. The use of oversampling appeared to be misguided as the patterns of accidents in fewer data-covered areas cannot be easily identified. Rather, other sources for causation such as systematic accident analysis of singular accidents should be employed. Furthermore, an overreliance on internal validity testing and a lack of external testing of algorithms' performance and prediction accuracy benchmarks persists. By mitigating these issues, future research might be able to focus on systematizing causes related to, for example, risk and energy, and thereby finding other and more important causes. Future research needs to focus on methods addressing the problem of data pre-processing, explaining the choice of methods, employing a mixed method approach merging several quantitative and also qualitative data sources, and explaining the results (especially the variance in ML algorithm's performance). Research should be commenced into investigating more attributes (such as risk analysis), applying deep learning algorithms, and improving the testing accuracy of ML models.

REFERENCES

Ayhan, B U and Tokdemir, O B (2019) Predicting the outcome of construction incidents, *Safety Science*, **113**, 91-104.

Bell, E, Bryman, A and Harley, B (2019) *Business Research Methods 5th Edition*, Oxford: Oxford University Press.

Berglund, L, Johansson, J, Johansson, M, Nygren, M, Rask, K, Samuelson, B and Stenberg M (2017) *Risker Och Säkerhetsarbete I Byggbranschen En Kunskapssammanställning Baserad På Internationell Forskning (in Swedish)*, Stockholm: Arbetsmiljöverket.

Bilal, M, Oyedele, L O, Qadir, J, Munir, K, Ajayi, S O, Akinade, O O, Owololabi, H A, Alaka, H A and 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..

Bilal, M and Oyedele, L O (2020) Guidelines for applied machine learning in construction industry - A case of profit margins estimation, *Advanced Engineering Informatics*, **43**, 101013.

Choi, J, Gu, B, Chin, S and Lee, J S (2020) Machine learning predictive model based on national data for fatal accidents of construction workers, *Automation in Construction*, **110**, 102974.

Greenhalgh, T and Peacock, R (2005) Effectiveness and efficiency of search methods in systematic reviews of complex evidence: Audit of primary sources, *British Medical Journal*, **331**(7524), 1064-1065.

Hegde, J and Rokseth, B (2020) Applications of machine learning methods for engineering risk assessment - A review, *Safety Science*, **122**, 104492.

Håll Nollan (2020) *Keep the Zero* [in Swedish], Available from https://www.hallnollan.se/ [Accessed 9th April 2020].

Jørgensen, K (2002) *En Taxonomi for Arbejdsulykker: En Systematisk Beskrivelse Af Årsagssammenhænge [in Danish]*, Copenhagen: DTU Lyngby.

MacLure, M (2005) 'Clarity Bordering on Stupidity': Where's the Quality in Systematic Review? *Journal of Education Policy*, **20**(4), 393-416.

Oswald, D, Zhang, R P, Lingard, H, Pirzadeh, P and Le, T (2018) The use and abuse of safety indicators in construction, *Engineering, Construction and Architectural Management*, **25**(9), 1188-1209.

Poh, C Q, Ubeynarayana, C U and Goh, Y M (2018) Safety leading indicators for construction sites: A machine learning approach, *Automation in Construction*, **93**, 375-386.

Ringdahl, H (2013) *Guide to Safety Analysis for Accident Prevention*, Stockholm: IRS Riskhantering AB.

Rollenhagen, C (2011) Event investigations at nuclear power plants in Sweden: Reflections about a method and some associated practices, *Safety Science*, **49**(1), 21-26.

Samuelsson, B (2019) *Arbetsskador I Byggverksamhet 2018: Privat Och Offentlig Verksamhet [in Swedish]*, Stockholm: Byggindustrins Centrala Arbetsmiljöråd.

Shayboun, M, Kifokeris, D and Koch, C (2019) Construction planning with machine learning, *In*: Gorse, C and Neilson, C J (Eds.) *Proceedings of the 35th Annual ARCOM Conference*, 2-4 September 2019, Leeds Beckett University Association of Researchers in Construction Management, 699-708.

Tixier, A J P, Hallowell, M R, Rajagopalan, B and Bowman, D (2016a) Automated content analysis for construction safety: A natural language processing system to extract precursors and outcomes from unstructured injury reports, *Automation in Construction*, **62**, 45-56.

Tixier, A J P, Hallowell, M R, Rajagopalan, B and Bowman, D (2016b) Application of machine learning to construction injury prediction, *Automation in Construction*, **69**, 102-114.

Tixier, A J P, Hallowell, M R, Rajagopalan, B and Bowman, D (2017) Construction safety clash detection: Identifying safety incompatibilities among fundamental attributes using data mining, *Automation in Construction*, **74**, 39-54.

Vallmuur, K (2015) Machine learning approaches to analysing textual injury surveillance data: A systematic review, *Accident Analysis and Prevention*, **79**, 41-49.

Webster, J and Watson, R T (2002) Analysing the past to prepare for the future: Writing a literature review, *Mis Quarterly*, **26**(2), xiii-xxiii.

Zhang, F, Fleyeh, H, Wang, X and Lu, M (2019) Construction site accident analysis using text mining and natural language processing techniques, *Automation in Construction*, **99**, 238-248.