DTU Library

Automated performance assessment of prevention through design and planning (PtD/P) strategies in construction

Johansen, K.W.; Schultz, C.; Teizer, J.

Published in:
Automation in Construction

Link to article, DOI:
10.1016/j.autcon.2023.105159

Publication date:
2024

Document Version
Publisher's PDF, also known as Version of record

Automation in Construction 157 (2024) 105159

Contents lists available at ScienceDirect

Automation in Construction

journal homepage: www.elsevier.com/locate/autcon

Automated performance assessment of prevention through design and planning (PtD/P) strategies in construction

K.W. Johansen [a,*], C. Schultz [b], J. Teizer [a]

[a] Department of Civil and Mechanical Engineering, Technical University of Denmark, Denmark
[b] Department of Electrical and Computer Engineering, Aarhus University, Denmark

ARTICLE INFO

Keywords:
Automated construction safety analysis
Performance assessment
Benchmark and ground truth
Construction safety ontology
Construction safety innovation and transformation
Digital safety rule-checking and reasoning
Human-assisted decision-making tools

ABSTRACT

Construction sites are among the most dangerous workplaces due to their complex, dynamic, continuously changing work environment. Many existing workplace safety planning techniques rely on two-dimensional drawings and manual expertise. Such efforts are cumbersome as safety plans quickly become outdated as construction work progresses. There has been significant research into automated safety planning, yet no community-wide standard exists for objectively measuring and comparing automated safety assessment efficacy. To address this, an automated performance assessment framework is proposed. It evaluates input solutions regarding newly formalized quantitative soundness, completeness, and spatial correctness indicators. The ground truth of the deadliest hazard *falls-from-height* is collected through a workshop with domain experts. We validate the proposed framework in a case study, where the performance of our previously developed automated safety planning algorithm is assessed by our new performance assessment framework. The results yield valuable insights into the importance of automated evaluation frameworks that can convince practitioners to invest in human-assisted Prevention through Design and Planning strategies.

1. Introduction

Construction is one of the most dangerous industries due to the continuous changes that occur in the workplace environment [1]. Over time, the previously safest route may have turned into a perilous route, e.g., because of changes in the tower crane's planned tasks, missing fall protection equipment, or debris in the designated pedestrian walk path. Consequently, the workers are responsible and must be aware of, consider, and adapt to new hazardous situations that may not be a part of the safety plan due to the low temporal resolution adopted when undertaking initial safety planning. Safety planning is currently a manual and labor-intensive task. In particular, the standard planning process only covers the overall site layout in a coarse temporal resolution because it would be impossible to generate a new safety plan on every state change of the construction site. The lack of temporal precision and, therefore, the demand for the workers to take over the situation planning, result in thought-provoking statistics.

A report on labor statistics in the US from [2] shows that fatalities in the private construction industry correspond to 21.2% (1008) of fatalities (4764). Furthermore, the report indicates that the predominant reason for fatalities in the construction industry is falls, slips, and trips, which correspond to 36.5% (368). These findings motivate the research and development of safety through design and planning, also referred to as Prevention Through Design and Planning (PtD/P). PtD/P is similar to Prevention through Design (PtD), but the planning aspect lets the approach assess scheduled 4D construction scenarios, where safety-related tasks, such as installation of protective equipment, are identified and added to the schedule when needed. With the emerging research of automating the task, manual work is reduced, and consequently, the temporal resolution increases remarkably. Additionally, the PtD/P strategy provides further insight into the cost and time investment of safety under different circumstances.

Even though PtD/P has been around for at least two decades, the practitioners in the industry are hesitant to adapt to the automated approaches. The reason for this is the lack of support in their current workflow, standardization, and trust in their correctness [3]. Another reason is that each study is carried out in a new non-shared building model, where only a subset of the scenarios is present; thus, comparing different approaches becomes impossible. The lack of standardization is also expressed in both the analysis approach, which should be expected, but more importantly in the outputted safety assessment, where some solutions point out that precautions must be taken. Others inject the

* Corresponding author.
E-mail address: kawj@dtu.dk (K.W. Johansen).

https://doi.org/10.1016/j.autcon.2023.105159
Received 13 March 2023; Received in revised form 24 October 2023; Accepted 26 October 2023
Available online 9 November 2023

prevention equipment into the Building Information Modeling (BIM) model. This is further elaborated in the related work section.

This work builds on, and significantly extends, research in [4], that presents a construction safety ontology for fall- from height hazards, which created the basis for a publicly available benchmark model. In this extension, we add a benchmark model and investigate different approaches to capturing and storing the extracted hazards in the analyzed BIM models. Through a survey, we determine relevant geometric representations of safety hazards and mitigation strategies and propose a common format that supports comparison to a ground truth assessment for correctness and soundness performance assessment. The performance criteria is carefully formed to facilitate a fair comparison. Finally, an application that allows comparison of safety assessment produced by others, e.g., automated approaches, is shared. The comparison is based on the proposed models, collected ground truth, capturing format, and performance criteria.

1.1. Research question

RQ 1: Can we create a common and exchangeable format and approach to capture hazards found in BIM models that is supported by both proprietary and open-source frameworks?

RQ 2: Can we ensure that the format and approach can be used in most approaches described in other Prevention through Design strategies?

RQ 3: Can we create a representative strategy of comparing automated approaches on soundness, completeness, and spatial correctness based on domain experts' ground truth and the proposed hazard-capturing format?

1.2. Contributions

C 1: We propose and validate a formal digital safety framework for automated soundness, completeness, and spatial correctness assessment, which can be used by researchers to measure improvement beyond the state of the art of new safety analysis tools, facilitate their comparison, and convince practitioners to adopt digital safety analysis tools.

C 2: We present a novel approach to automatically quantify an algorithm's performance in identifying hazards compared to a shared ground truth-assessed benchmark model.

C3: We develop and share an online application that utilizes the proposed comparison strategy and quantitative identifiers to automatically assess the soundness, completeness, and spatial correctness of an inputted safety assessment of the presented benchmark model.

We validate our new framework (C1, C2) by developing and sharing two benchmark BIM models, with corresponding sound and complete ground truths concerning the most deadly hazard source *fall from height*.

1.3. Research overview

Fig. 1 shows an overview of this study and its relation to research previously carried out in the research community. The top row captures the activity of formalizing the construction safety regulation into the *construction safety ontology* and *3D benchmark model*. This work was initiated in a previous workshop paper [4], and significantly extended in this work. With the motivation of enabling automated prevention through design researchers and developers to determine the performance of their approach, it is necessary to create an automated and unbiased assessment strategy.

The second row captures the creation of the ground truth, which is initiated with a manual process of ground truth assessment, consisting of domain expert interaction, that is carried out based on a workshop document [5], for which the content and procedure are described in section 5. Section 5 also describes the next step, called *Ground truth digitization*, where the domain expert outputs are digitized, to let them become the basis for the later *Solution assessment*.

The third row describes the steps that are expected to be performed each time one wants to assess the performance of their *Automated safety approach*. The automated approach creates a *safety assessment,* which must be transformed into the format supported by the following *solution assessment* (i.e., the format proposed in section 6.1). To validate the framework we apply our existing safety assessment algorithm, called SafeConAI to produce results that can be compared to the ground truth assessment, described in the previous section.

Having the digitized version of both the ground truth and the inputted solution, allows the *solution assessment* step to produce the *performance assessment.* The assessment consists of the three proposed indicators, i.e., soundness, completeness, and spatial correctness. The solution assessment procedure and performance indicators are further described in section 6.2.

2. Related work

The domain of construction code and regulation checking is an ongoing research topic. The most commonly investigated construction safety rule is regarding *fall from heights* hazards, which are responsible for most fatalities in the construction industry [6–9]. To explore automated prevention through design, one must first define a link between the construction regulation and the Building Information Model (BIM) and afterward define the logic that can check whether the regulation is violated in a given BIM model. The drive behind the efforts is that the current practices are cumbersome and labor-intensive. With the emergence of Digital Twins (DT), the knowledge gap between the current state of the construction site and planning has decreased. As presented in [10], the digital twin and automated safety assessment should even allow the decision-makers at the construction site to analyze different

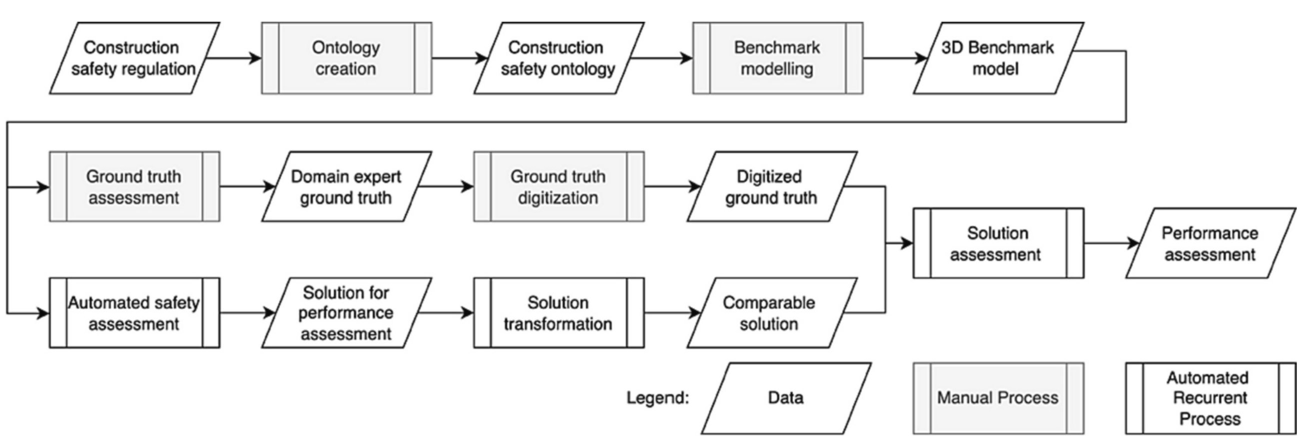

Fig. 1. Overview of the proposed procedure for automated performance assessment of automated safety analysis results.

approaches in terms of cost, time, safety fitness, etc., before selecting the desired construction schedule.

A formal way of capturing construction regulations and building codes is needed such that the computer can (1) interpret the natural language formulation of the content and (2) link this content to concepts and properties in BIM for automated safety analysis in construction. Several construction safety ontologies capture object concepts, properties, and their relationships [11–14]. The aforementioned studies work under a principle that builds directly on the geometries of the BIM model. This means that a hole in a slab is considered a hole even though an element (e.g., stairs) fills the void space. In this study, an approach considering spatial artefacts that are created and changed by the surroundings are considered.

The concept of spatial artefacts has been defined in [15] and adapted to construction safety in [7]. The purpose of spatial artefacts is to incorporate worker experience (e.g. what can be seen, heard, the hazards encountered, etc.) and behavior (e.g. movement, engaging in work tasks, etc.) directly into the BIM model as *objects*, in the form of semantically-rich regions of empty space i.e. on the same ontological level as walls, doors, and other building elements. The existence and geometry of these spatial artefacts can thereby be defined using situation-sensitive rules, so the analysis is both flexible and the analysis results are transparent (i.e. captured as BIM model objects that can be visualized) and portable (i.e. the derived spatial artefacts can be saved in the BIM model) [7,15]. For example, spatial artefacts allow the geometric analysis to consider and incorporate the surroundings of an element, such that the staircase in the previous example is not considered a hole in the slab, but instead, only the edges that are not connected to the stair step are considered dangerous in terms of fall hazard analysis.

The ontologies are, for example, used for safety rule checking, i.e., to determine if some safety hazards are present in the BIM model under investigation. The above examples successfully point out the areas where the safety expert needs to be cautious and apply temporary prevention equipment. There are different approaches to automated safety checking, which have different output formats. The output can be split into two main categories (1) semantic information and (2) geometric model enhancement, which is the most related to this study. Examples of geometric categories are [16,17], where the prevention measures are injected as BIM objects into the BIM model. Injection directly in the BIM model often requires a proprietary program such as Solibri model checker or Revit. As an alternative to proprietary software tools, [18] uses only open-source tools, where the output is captured in two-dimensional (2D) floorplans, which capture the hazards with highlighted lines. Even though those mentioned above are already significant contributions, their completeness, and soundness cannot be assessed without manually going through the identified hazards individually. The reason is that there is currently no straightforward way to compare the results of the different approaches in an environment that contains identified ground truth edge cases and their correct hazard prevention measures.

Benchmarking is commonly used in other domains, such as machine learning and computer vision, where a portion of the data, i.e., test data, is used as ground truth to assess the correctness and soundness of a trained model [19,20]. The adoption of benchmarking provides the stakeholders with a deeper insight into the quality of the hazard identification provided. [3] outlines the importance of the application's capabilities to utilize BIM and exchange information. Furthermore, it points out the gap that exists between practitioners and researchers and the lack of standardizing joint activities. With benchmarking, the comparison of automated safety analysis approaches can be performed qualitatively instead of quantitatively, i.e., approach vs. approach and approach vs. manual assessment. The qualitative comparison should facilitate an improved trust in the automated approaches and, thereby, their application in the real construction industry.

3. Methodology

This section describes the research methodologies of the four activities of this study. The first activity is to define the ontology that utilizes the spatial artifact concept to capture *fall from* heights hazards. The second activity is to create two benchmark models that can be used for the later comparison of automated approaches. The third activity involves interacting with the domain expert to create and digitize a ground truth assessment of the two benchmark models. Finally, the fourth activity is creating an overview of the outputs of automated approaches and investigation of a fair performance criteria.

3.1. Ontology development

3.1.1. Organizing and scoping

The purpose of initiating a formal standardization of a construction safety domain language is to provide an approach that we can use in our future research but also be used by the community to streamline the efforts on automated construction safety assessment. We initiate the domain language with the most straightforward and predominant spatial artifact (i.e., *movement*, *fall*, and *fall hazard space*) and envision the vocabulary extending over time when work progresses in the community. We base our ontology on the Industrial Foundation Classes (IFC) to permit interoperability. Additionally, the IFC structure is similar to graph databases used in the emerging Digital Twins (DTs).

3.1.2. Data collection

We collect the natural language formulation of the construction safety codes from the European Union, Denmark, Germany, and the US regulation. We have chosen the EU regulation to get an overview of Europe, Denmark (where we are located), and Germany to compare similarities within the European countries. Besides the European regulations, we have chosen to consider the US regulations as it should reveal differences and similarities between the two continents.

3.1.3. Data analysis

Based on each of our chosen country and continent regulations, we extract two kinds of information: (1) their definition of when fall protective equipment must be applied, (2) The dimensions of hazard space for different mitigation strategies, and (3) example implementations of fall protection systems. The extracted and analyzed information is assumed to make our ontology applicable for at least the included countries and continents.

3.1.4. Initial ontology development

Our initial ontology is based on the current state of the art, which we refine to ensure further applicability and consensus in the research domain. The ontology focuses on fall hazard scenarios. Based on our data analysis (step 3), we extract the variating factors and define a vocabulary of variables that we extract from the regulation. Subsequently, we define the ontology using spatial artefacts and the vocabulary. Additionally, we propose a strategy to integrate the spatial artefacts into IFC, which only depends on existing IFC-classes, meaning that the ontology is compliant with the IFC4 tools and workflows.

3.1.5. Ontology refinement and validation

To refine and validate our ontology, we develop a benchmark model. Based on the regulations, we carefully create scenarios that will, or will not, require fall hazard mitigation equipment depending on the regulation. We are utilizing the benchmark model and the domain expert assessment to validate our ontology.

3.2. Benchmark model creation

This study proposes two benchmark models with two levels of complexity. The first model (i.e., low complexity model) consists of only six elements and is created with the intent that this should be assessable to all automated approaches. Additionally, the model will not be ambiguous when the domain experts manually assess the ground truth. The second (i.e., edge case model) intends that the automated approaches are assessed regarding their ability to capture the details of the safety regulation. In this model, there are parts of the slabs that are created to represent areas that should or should not be included as protected areas depending on the selected regulation, i.e., the algorithms are tested in centimeter accuracy.

Both models are captured in the non-proprietary IFC schema to ensure useability in all the tools supporting IFC import, for example, Revit and Solibri were mentioned in related work as tools used for automated safety analysis approaches. The non-proprietary format allows domain experts and practitioners to interact with the models in their regular workflows to keep complexity at a minimum.

3.3. Ground truth assessment

To create a ground truth for the two models, it has been decided to involve four domain experts. This ensures that the assessment represents reality, and that the assessment is not biased by the results of the automated approach that the authors have developed. The domain experts are selected from two different countries in Europe.

The interaction has been performed in a three-step procedure described in a workshop document, which is also shared [5]. The steps consist of (1) an introduction to the workshop document and a brief look at the 3D views of the models (also available in the GitHub repository). Additionally, the extracted relevant values from the regulation are discussed to ensure the usage of common values (fall distance, surface dimension, and cover dimension). (2) individual assessment of the two models. (3) common discussion about differences in digitized individual assessment from the domain experts.

A digitization procedure has been created to support different formats of domain expert assessments, e.g., highlight on printed paper, in a pdf reader, or CAD annotations. The digitization process has been included to let the domain expert work in their usual workflow and to avoid putting unnecessary complexity to the task. The process is visualized in Fig. 2, and the result becomes a list of points and their pairwise connections, as shown in the tables in the figure. This information can be used to add lines in the model, which can be used for the discussion. The procedure has been formed to ensure that the authors and the domain experts are not biasing each other. This facilitates the realness of the ground truth assessment.

3.4. Extraction of hazards in construction situations

To create an overview of different approaches to signal hazards, a survey of existing solutions and studies has been performed. The survey mainly concentrates on hazard identification of construction scenarios and covers both studies using semantic and geometric approaches. The overview is presented in a table format that presents the approaches to providing information to the decision-makers at the construction site. During the creation, it is also extracted whether the geometry of the analysis results can be extracted for comparison. In the cases where this is not directly stated in the studies, assumptions are taken based on experience with BIM interfacing and injection. The assumptions are clearly marked with **(assumed)**.

The overview is considered when this work proposes a representation to capture the leading-edge hazards of the analyzed construction situation. The format is used as input to the *Solution Assessment* process shown in Fig. 1. Additionally, it will be considered how the other representations can be transformed into the proposed representation.

3.5. Assessment performance criteria

The *Solution Assessment* process outputs the performance assessment, which consists of a verdict of the performance of an incoming proposed solution compared to the ground truth. In the development of an assessment that enables a fair and equal comparison, a set of specialized performance indicators is proposed. Inspired by the assessment approaches in general software development, it is desired to ensure (1) that all the hazards are found (i.e., the solution is complete) and (2) that all the found hazards exist in the construction situation (i.e., the solution is sound). Besides assessing soundness and completeness, the assessment involves the (3) geometric correctness of hazard identification. The third indicator is assessed in a separate value as it was found in the analysis that not all solutions are geometrically comparable to the ground truth. Even though this is the case, the value is included in the overall assessment as it facilitates the applicability of the safety analysis results in the workflow of the practitioners performing and utilizing safety assessments on the construction sites.

4. Otology development

4.1. Safety regulation collection and analysis

We analyze the European [21], Danish [22], German [23], and US regulations [24]. To ensure that the proposed ontology is representative, we extract the factors that are present in them. We compile the variating

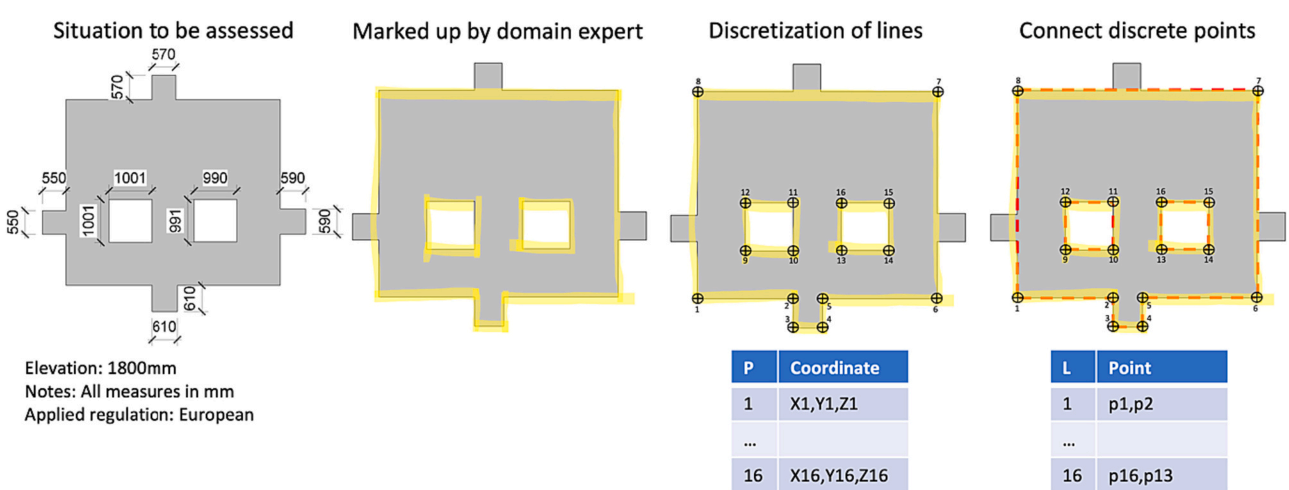

Fig. 2. Digitization process of example domain expert assessment results.

Table 1
Variable vocabulary defined through analysis of scoped regulations.

Natural language formulations	Attribute	Symbol	US	EU	German	Danish
The minimum distance, from an elevated surface to a lower surface which an item or a human being could fall onto, which would require a form of fall protection equipment.	Fall distance	f_d	180 cm	200 cm	200 cm	200 cm
The minimum width of a surface, which an agent is allowed to be present on	Surface width	w_s	56 cm	60 cm	60 cm	60 cm
The minimum Height of a space, which is considered walkable	Walk height	h_w	NA	NA	NA	NA
Minimum height of a space considered crawlable	Crawl height	h_c	NA	NA	NA	NA
Maximum width of hole in a surface, where chosen mitigation will be a coverboard, i.e., maximum width of cover boards	Cover width	c_w	1 m	NA	NA	NA
Maximum height of hole in a surface, where chosen mitigation will be a coverboard, i.e., maximum height of cover boards	Cover height	c_h	100 cm	NA	NA	NA
Minimum height of guardrail (aka., Safety railing, safety barrier)	Railing height	r_h	1,1 m	1 m	1 m	1 m
Maximum distance between vertical poles of guardrail installation	Pole distance	p_d	240 cm	NA	200 cm	225 cm
Maximum distance between horizontal boards in guardrail installation	Board distance	b_d	$r_h/2$	47 cm	47 cm	47 cm
Best practice width of applied vertical poles in guardrail installation	Pole width	p_w	5 cm	NA	3 cm	4,5 cm
Best practice height of applied vertical poles in guardrail installation	Pole height	p_h	10 cm	NA	15 cm	7 cm
Best practice width of applied horizontal boards/rails in guardrail installation	Board width	b_w	2,5 cm	NA	3 cm	3,2 cm
Best practice height of applied horizontal boards/rails in guardrail installation	Board height	b_h	15 cm	NA	15 cm	15 cm
Minimum continues force that vertical poles in guardrail installation should withstand	Pole force	p_f	890 N	300 N	300 N	300 N
Minimum continues force that horizontal boards in guardrail installation should withstand	Board force	b_f	890 N	300 N	300 N	300 N

factors into a vocabulary and extract their values for comparison, as shown in Table 1. Fig. 3 shows a graphical representation of the vocabulary variables, which are limited to *falls from height*, where mitigation approaches include safety guardrails and cover panels. Hence, we are not investigating safety nets.

4.2. Definition of ontology for fall from heights

After extracting the variables that change in the European, Danish, German, and US regulations, we define our ontology that captures the construction regulation. Our ontology shown in Fig. 4 is based on spatial artefacts, which capture concepts pertaining to human experience and behavior as semantically rich regions of empty space [25–27].

In a BIM model, spatial artefacts (listed in Table 2) are derived from IfcElements and their spatial relationships. Depending on the point of view, the surface of a slab (for example) may simultaneously introduce a walkable space, fall space, and tumbling space. Thus, extraction of the spatial artefacts is based on the construction regulation, the element relationships according to specific points of view, the location of the IfcElement instance, and the geometry of the IfcElement instance; the location and geometry are extracted from instance's IfcProductRepresentation. Additionally, the relationship between spatial artefacts may introduce *hazard spaces*, e.g., Fall hazard space. Each hazard is mitigated via mitigation equipment, which is a subclass of IfcElement. The individual mitigation strategies have test procedures specified in the safety regulation. The test procedure indirectly captures the attributes of the mitigation system, e.g., dimensions, pole- and bord distances, etc.

4.3. Integration into industry foundation classes

Fig. 10 presents the proposed IFC integration of the proposed ontology. The integration utilizes the *IfcProperty* class and the *IfcRelAssignsToProduct* class to capture information about which products in the BIM model directly generate a given spatial artifact. This startegy is fully compliant with IFC4 and can be processed by all IFC4-compliant tools. Each spatial artifact is implemented as an instance of the *IfcSpatialZone* class. The spatial artifact type is expressed as an instance of *IfcProperty* that selects an enumerated value.

The enumeration of spatial artifact types is implemented as an instance of IfcProperty-Enumeration, with the name "PEnum_SpatialArtefactType". The relationship with existing products in the IFC model that are used to directly generate the spatial artifact is expressed via an instance of *IfcRelAssignsToProduct*; for example, a slab on which a person can walk may be used to derive a movement space. For representing mitigation strategies (e.g., coverings, harnesses, safety nets), we adopt a similar approach by creating instances of the existing class

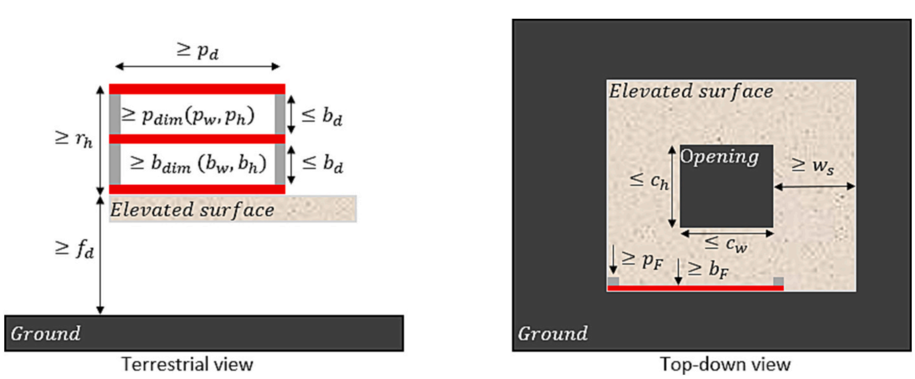

Fig. 3. Illustration of values in Table 1 (horizontal boards colored in red and vertical poles in grey). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

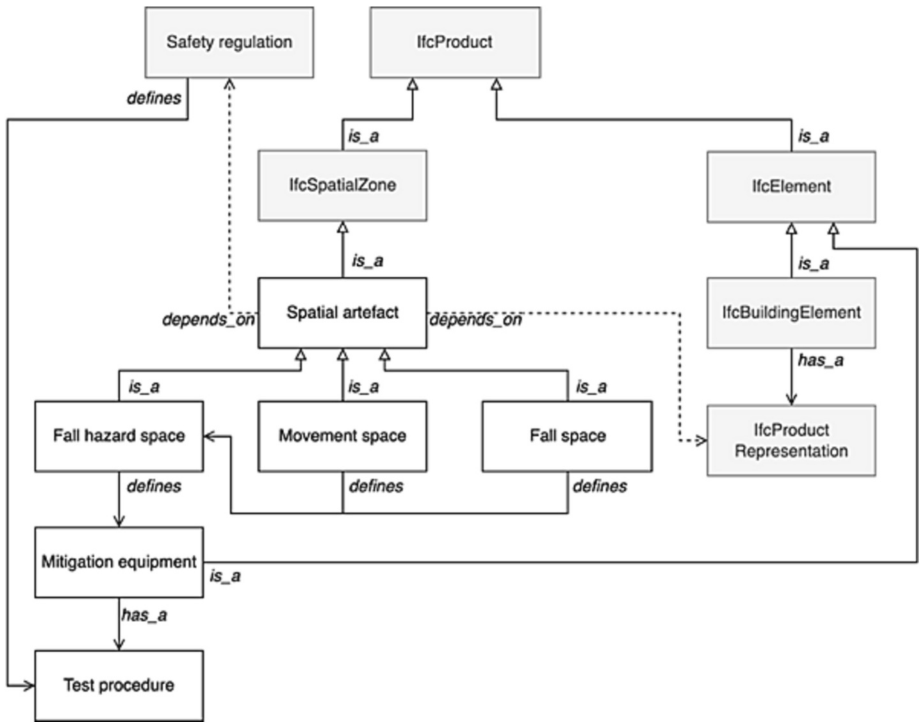

Fig. 4. Diagram of our BIM-based ontology of construction hazards and mitigation interventions.

Table 2
Overview and description of spatial artefacts for fall hazard identification and analysis.

Spatial Artifact	Specialized subclasses	Description	Illustration	Constraints
Movement space		Regions in which an agent (e.g., construction worker, manager, and visitor) can travel		
	Crawlable space	Regions in which an agent can travel crawling.	Fig. 5	$h_c \leq height < h_w$ *and width* $\geq w_s$
	Walkable space	Regions in which an agent can travel upright	Fig. 5	$height = h_w$ *and width* $\geq w_s$
Fall space		Regions in which an object or agent will fall by f_d.	Fig. 5	$F_{z_{lower}} = M_{Z_{lower}} + f_d$
Fall hazard spaces		Regions in which an agent is subject to a fall hazard		
	Leading edge space	Regions where the movement space in its full height intersects with a fall space	Fig. 6	$M_{z_{lower}} \geq F_{z_{lower}} \wedge M_{z_{upper}} \leq F_{Z_{upper}}$
	Offset leading-edge space	Regions where a portion of the movement space intersects with a fall space	Fig. 7	$M_{z_{lower}} + offset_{lower} < M_{z_{lower}} + r_h$
	Offset top leading-edge space	Regions where a portion of the movement space intersects with a fall space	Fig. 8	$M_{z_{upper}} - offset_{upper} < M_{z_{lower}} + h_c$
	Tumbling space	Regions in which an agent can tumble over fall prevention equipment on lower surface	Fig. 9	$z_{upperSurface} - z_{lowerSurface} < f_d \bigwedge width_{lowerSurface} < w_s$

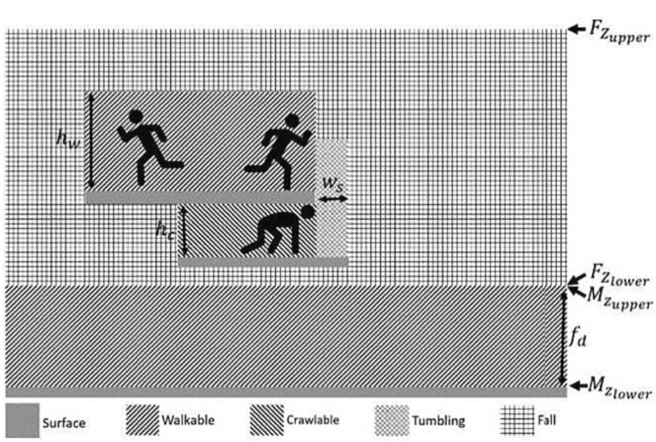

Fig. 5. Illustration of spatial artefacts extracted from IfcElements.

IfcCivilElement and assigning a property enumerated value (with a custom property enumeration listing the mitigation strategies) to indicate the mitigation strategy class.

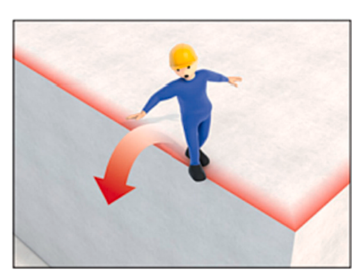

Fig. 6. Leading edge.

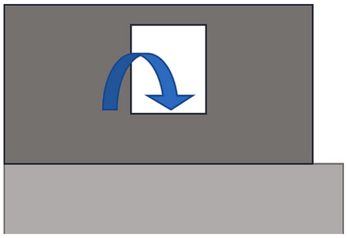

Fig. 7. Offset leading edge.

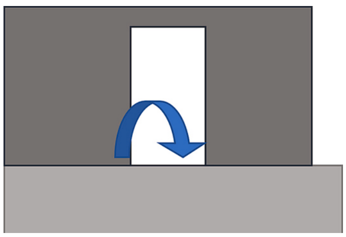

Fig. 8. Offset top leading edge.

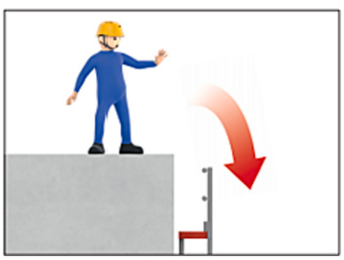

Fig. 9. Tumbling space.

5. Definition of the benchmark model and ground truth capturing

As stated previously, this study consists of two benchmark models. The two models have different complexities. The models, their contents, and their rationale are described below. Both models are captured in the IFC 4 schema and are available in a GitHub repository [5]. Having the benchmark model modeled allow for the subsequent collection and digitization of the ground truth assessment that was gathered through the domain expert workshops.

5.1. BIM benchmarkmodels

5.1.1. Model A - Low complexity model

Fig. 11 (left) shows the 3D view of the low complexity benchmark model, which is a minimum viable solution assessment. The model consists of only a ground plane placed at elevation 0, 4 walls, each 4 m high, and a slab on top. This situation clearly contains one hazardous platform that needs to be protected. This model allows us to clearly understand how the domain experts are assessing the model, capturing the ground truth, and should not lead to any ambiguity.

5.1.2. Model B - Edge-case model

Fig. 11 (right) shows the benchmark model that has been carefully designed to include edge-case scenarios of the regulations that have been investigated for this work. Specifically, it consists of two parts. The first part in front of the stamped line in Fig. 11 is designed such that the first platform's elevation (f_d) is below the threshold for the European and US regulations, the second platform's elevation (f_d) is high enough to be subject to the EU and US regulation, and the third platform's elevation is subject to only the US regulation. Additionally, the platforms have been designed with smaller outgoing platforms, whose widths (w_s) are chosen to be subject to individual regulations. Lastly, the platforms include two openings, one bigger than the allowable coverable dimensions (c_w and c_h) stated for the US regulation, i.e., the larger opening requires guardrails, and the other smaller opening requires a covering. This dimension is not stated for the EU regulation and will be subject to best practices in the ground truth assessment by the domain experts. The other part of the model (behind the line) is designed to capture special cases such as openings in walls and slabs, leading edges, coverable gabs, tumbling spaces, leading edges that are non-orthogonal to the model space, and obstacles in the movement space where the domain experts also must use their best practices.

5.2. Capturing the ground truth

The ground truth that is used in the assessment application was captured through domain expert interaction in the format of the

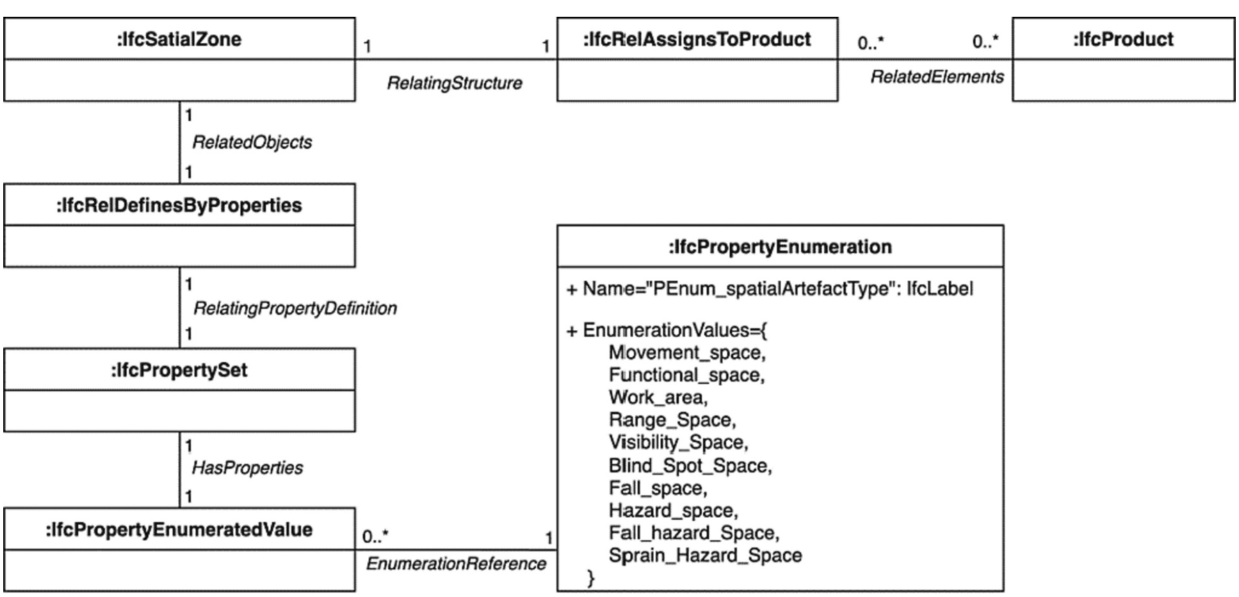

Fig. 10. UML class diagram depicting how instances of spatial artefacts for safety are expressed in standard IFC4.

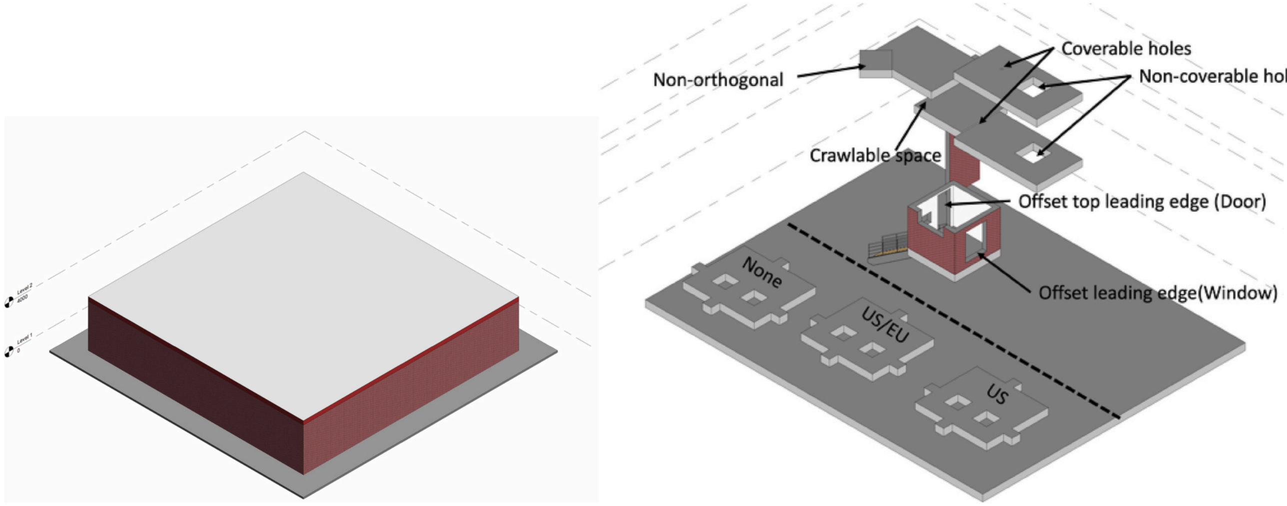

Fig. 11. 3D view of the low complexity benchmark model (left) and edge-case model (right).

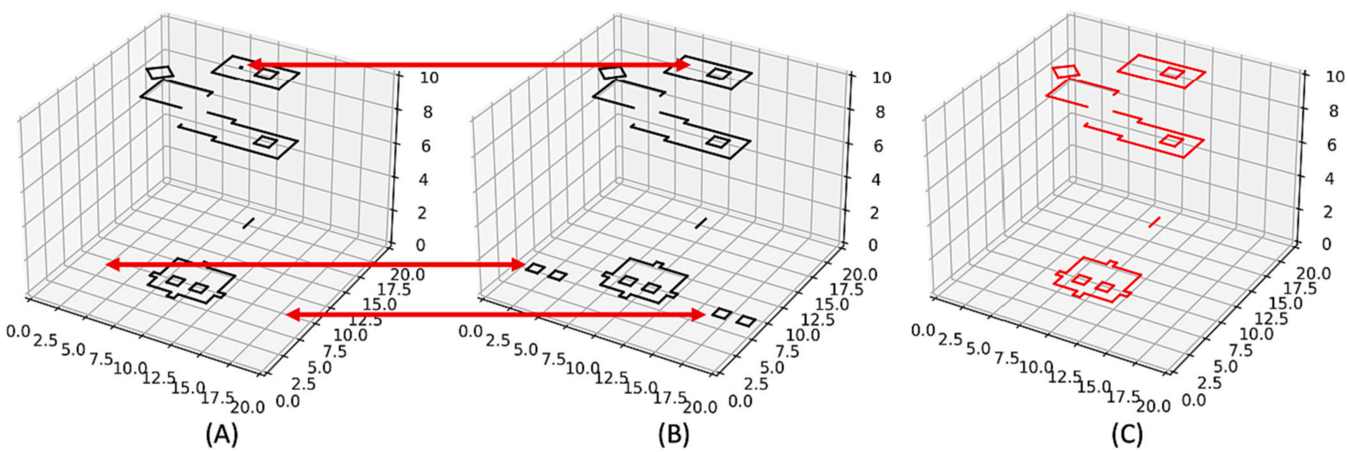

Fig. 12. Examples of two domain expert assessments (A and B) and their differences illustrated with the red-colored arrows. (C) show the resulting ground truth assessment. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

previously described workshop. Due to the availability of the participants, three individual workshops were performed, capturing feedback from four domain experts. Throughout the interaction, we were available to answer questions, that would avoid misunderstandings with regards to the regulation and spatial structure of the models.

All the participants agreed on the assessment of Model A. The purpose of this Model A is to provide a model that sets the scene for the interaction, and this model does not lead to any ambiguity. Therefore, it is possible to determine the individual domain expert's assessment strategy, to ensure a streamlined extraction of leading edges while digitalizing them.

Model B, on the other hand, led to further discussion and clarification. This model introduced matters about best practices, where the individual domain experts reasoned some of identified leading edges that are not strictly needed to comply with the regulation.

In Germany, the regulation states that walkways should be protected, when the elevation exceeds 1 m, but that working area only needs protection if their elevation exceeds 2 m. The workshop document states, "Assume that all platforms and surfaces are working areas and should be reachable to personnel", which removes the ambiguity. Although it should be considered when the algorithms are developed. Additionally, it was raised that an elevation of 997 mm would most likely be considered 1 m (shown in Fig. 12, two lower red-colored arrows). Model B is, as the name suggests, created to contain the edge cases. Thus, it has been chosen to remove the protection equipment that is not strictly needed to comply with the regulation from the ground truth. Additionally, the domain expert did not agree that the small (10 by 10 cm) gap in the topmost slab would necessarily need protection equipment (shown in Fig. 12, top red-colored arrow). Nonetheless, the comments have been noted and will be further elaborated in the discussion and future work sections.

Following the procedure proposed in the methodology section, a digitized ground truth has been extracted based on the domain experts' feedback. The ground truths for each model are captured in a CSV format similar to the format that the inputted solutions should be uploaded in. This format is further described in the following sections.

Table 3 shows a summary of the resulting ground truth assessment, which was mutually decided in the fourth step (i.e., discussion of possible differences in assessments) of the workshop procedure.

Table 3
Ground truth model content.

Model	Number of leading edges	Total length leading edges
Model A	4	80 m
Model B	61	102 m

6. Identification, extraction, and evaluation of hazards

This section described the processes of defining an input format, that support the majority of current PtD/P studies identified in the survey and defining suitable performance indicators that can be used to compare and benchmark solution.

6.1. Extraction of hazards in construction situations

Table 4 shows a survey of the automated safety research considered in this work. The table provides information about how the individual studies output the extracted hazards. The output has also been used to do an overall grouping of the studies. In the first category is the *BIM-geometric*, where the leading edges for some of the cases are stated to be extractable, but for others, they are assumed to be. The reason for assuming that the lines are extractable is that the exact geometry and location must be known to inject those in the BIM. For the second category, *Other formats -geometric,* it is also assumed that the actual locations of the leading edges can be extracted because the work in this category shows results either in 2D plan views or in a 3D visualization. To perform this sort of assessment, the geometry and, therefore, the location must be known; hence the extractability is assumed. The last category *Other formats – Sematic,* has one thing in common, which is that the studies are extracting the sources of the hazards (i.e., the element that creates the hazard) instead of the actual hazard. This means that the location and geometry of the hazards are not extractable in a format that is sufficient for the assessment performed in this study. Additionally, the studies are not particularly extracting fall from height hazards, which may be the reason for the semantic assessment. Nonetheless, they have been included to represent the semantic assessment, which is also something that should be considered when developing a performance assessment strategy.

During the extraction of Table 4, the location of the assessment was analyzed and noted. The *approx. on slab edge* assessment format means that the visuals in the publication show results, where the injected objects are on the slab edge corresponding to the leading edge. As the verdict in this column only stems from visual interpretation, it has been chosen to limit the verdict to approximately on edge. Another reason is that some publications use a strategy to simplify the assessment, which results in less change of direction in the railing objects. Fig. 13, capture the expected deviation that may exist in the assessment category *Approx. at slab edge*. The performance assessment approach that this work

Table 4
Overview of automated prevention through design strategies survey in construction.

Output format	Description	Reference	Extractable	Dimension	Assessment
BIM - geometric					
Objects injected in BIM	Extraction of fall from height hazards and injection of prevention equipment in 3D	[4]	Yes	3D	Approx. on slab edge
Objects injected in BIM	Extraction of fall from height and falling object hazards and injection of prevention equipment in 4D	[28]	Yes	4D	Approx. on slab edge
Objects injected in BIM	Extraction of fall from height hazards and injection of prevention equipment in 4D	[17]	Yes (assumed)	3D	On slab not always at edge
Objects injected in BIM	BIM-based fall hazard identification and prevention in construction safety planning	[29]	Yes (assumed)	4D	On slab not always at edge
Objects injected in BIM	BIM-based fall hazard identification and prevention in construction safety planning	[16]	Yes (assumed)	3D	Approx. on slab edge
Other formats – geometric					
Leading edges superimposed on 2D plan view	BIM-based fall hazard identification and prevention in construction safety planning	[30]	Yes	4D	Approx. on slab edge
Visualization of hazards in Desite MD	Construction site planning of crane lifts and excavation pit stabilization	[9]	Yes (assumed)	3D	Graphically
leading edges in 3D unity	Fall hazard identification and injection of prevention equipment in Unity	[31]	Yes (assumed)	3D	Approx. on slab edge
Other formats – semantic					
Job hazard analysis based on database	Text based report of hazards of task based on database	[32]	No - actual hazard area not extracted	4D	Highlight of BIM-element
Falling objects hazard zones notifications	Tool extracts and reports about falling object possible falling object hazards in 4D construction scenario	[33]	No - actual hazard area not extracted	4D	Highlight of BIM-element
Selection of elements that applies to rule	Checking safety rules and mark elements that applies, which can then be prevented manually	[34]	No - actual hazard area not extracted	3D	Highlight of BIM-element
Selection of elements that applies to rule	Checking safety rules and mark elements that applies, which can then be prevented manually	[35]	No - actual hazard area not extracted	4D	Highlight of BIM-element
Extracts roof elements with too steep slope in Revit	Rule regarding sloped surfaces is used in Revit/dynamo to select the roof-elements that are subject to sloped surfaces exceeding the limits	[36]	No - actual hazard area not extracted	3D	Highlight of BIM-element

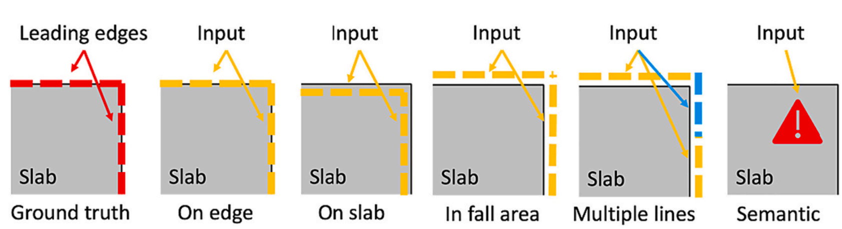

Fig. 13. Visualization of possible deviations in outputted assessment results. The leading edge shows where the actual leading edge is in the scenario, and the lines annotated with input represent the possible inputs for the assessment.

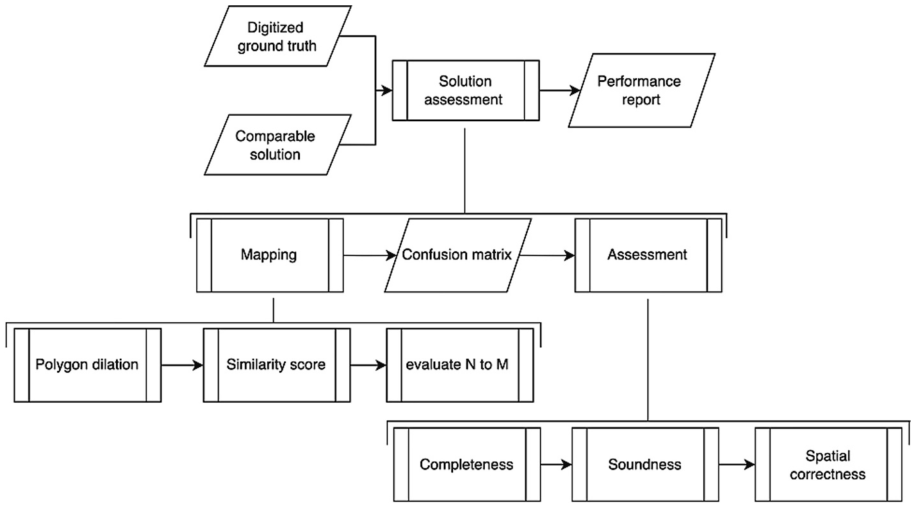

Fig. 14. Detailed flow diagram showing the sequence of steps performed in the proposed solution assessment.

proposes should be capable of handling these situations, and still, provide the user with useful results. To provide useful information even when the assessment contains deviations, it is necessary to dilate the inputted leading edges in specific directions. To accomplish this, the proposed format of capturing the resulting leading edges must incorporate a direction that is used in the subsequent mapping strategy.

The input format for the performance assessment (output of the automated safety assessment algorithm) should therefore be consistent of two points per line. The points' order should be defined counterclockwise (CCW) for the outer perimeter, and clockwise (CW) for lines that are part of holes in the slab. This approach is inspired by the common practice of defining polygons, where the points of the outer perimeter are also defined in a CCW convention. With this approach, the normal, direction towards the slab and direction towards the fall area is extractable with simple operations, which allows for a later parameterized dilatation. If the lines are not defined with the CCW convention, the dilation can only happen based on two parameters, i.e., the along and perpendicular to the identified hazard line.

The lines should be captured in a CSV-file that contains a header and the individual lines each in a new row, as exemplified below. In this example, each line has an index in the *line* column, and two sets of *x, y, and z*, one for *p1* and one for *p2*. The *x, y, and z coordinates* should be described in meters relative to BIM model origin. The CSV- file should be formed using semi-colons (**;**) instead of commas (**,**).

line;	**p1_x;**	**p1_y;**	**p1_z;**	**p2_x;**	**p2_y;**	**p2_z**
0;	−15.55;	21.47;	3.8;	−16.26;	21.47;	3.8

6.2. *Performance assessment criterion and computation*

The performance assessment consists of three different indicators of performance, i.e., completeness, soundness, and spatial correctness indicators. In this section, the definitions and their rationale are described. The sequence of the descriptions follows the sequence shown in Fig. 14. Thus the sequence becomes *Mapping* and *Assessment,* and the latter consists of the three indicators.

6.2.1. *Mapping*

The first step that must be performed in the proposed assessment procedure is mapping the individual incoming lines to the lines of the ground truth. The proposed mapping has been selected to be flexible in terms of input, i.e., capable of mapping inputted lines to ground truth lines even when these deviate from the ground truth. The examples of deviations are shown previously in Fig. 13.

Mapping the input to the ground truth is initiated with a parameterized dilation of the ground truth and input lines, consequently turning those into polygons. The parameters consist of the amount of dilation in each direction (i.e., along the line, towards the slab, away from the slab, and in the opposite direction of the line). Subsequently, each line n in the set of input lines $\boldsymbol{N}$ a similarity score is calculated to each line m in the set of ground truth lines $\boldsymbol{M}$.

As shown in Definition 1, the similarity score consists of two components, the intersection of the polygons (i.e., input polygon and ground truth polygon) relative to the union of the polygons. This measure is inspired by the intersection over union (IoU) in computer vision. During the process of computing the intersection and union, the input polygon is projected on the ground truth polygon along the z-axis, which is incorporated as some input may contain the coordinates of the top of the hazard and others the bottom. The other component of the similarity score is the distance between the centroid of the n^{th} input polygon to the m^{th} ground truth polygon relative to the dimension of the complete construction situation. The distance is computed as a 3D distance, which allows the mapping procedure to distinguish lines that are identical for different elevations of the model, and still maintain the previously mentioned relaxation.

The similarity function consequently creates a similarity n by m matrix (i.e., input lines as rows, and ground truth lines as columns), which can be used to map the input lines to the corresponding ground truth lines using the highest similarity score of each row.

$$Similarity\ score(n,m) \stackrel{\text{def}}{=} \frac{area(n \cap m)}{area(n \cup m)} * \left(1 - \frac{dist(centroid(n)\,, centroid(m)\,)}{dimension(benchmarkmodel)}\right) \tag{1}$$

6.2.2. *Assessment indicators*

After the individual input lines have been mapped to the individual ground truth lines, the assessment can be initiated. As mentioned in the methodology, the assessment consists of three parameters, i.e., completeness, soundness, and spatial correctness. Each of the performance indicators and their extraction method and rationale is described in the following.

6.2.2.1. *Completeness identicator.* In the completeness assessment shown in Definition 2, the *input solution* (**N**) is evaluated if it contains all the hazardous leading edges present in the *ground truth* (**M**). The process is based on the similarity matrix, where the approach is to find columns, which does not have a similarity score higher than 1/3. When this is the case, it means that there are no lines in the input that are mapped to the ground truth line, which consequently means that the input solution has not extracted that hazard. Intuitively this indicater captures whether all the hazards in the ground truth are also represented in the assessed solution, and can determine how well the safety algorithm reflects the safety regulation and best practices provided in the manual assessment.

$$Completeness \overset{\text{def}}{=} \frac{\left|\left\{m | m \in M.Similarity\ score(n,m) > \frac{1}{3}\right\}\right|}{|M|} \tag{2}$$

6.2.2.2. Soundness identicator. The soundness measure is shown in Definition 3 and captures the case that all the extracted hazards of the input solution exist in the ground truth assessment. The approach to extract this is similar to the extraction of the completeness indicator but is now based on the rows (*input-lines* (**N**)) instead of the columns (*ground truth lines* (**M**)). The approach is to check if all the lines of the input solution are assigned to lines of the ground truth. Intuitively this indicator represents how many of the hazards in the solution, provided by an algorithm, are also represented in the ground truth assessment, that is, if all the *identified* hazards are indeed *real* hazards.

$$Soundness \overset{\text{def}}{=} \frac{\left|\left\{n | n \in N.Similarity\ score(n,m) > \frac{1}{3}\right\}\right|}{|N|} \tag{3}$$

6.2.2.3. Spatial correctness identicator. Algorithm 1 shows a pseudo code representation of the algorithm used to extract the spatial correctness of the inputted solution captured in the list **N,** compared to the ground truth captured in the list **M**. Variable assignment is denoted with ":=", and the empty list is denoted with "[]". The algorithm loops through **M**, and dilates each line into a polygon (i.e. the standard Minkowski sum operator). The dilation function is used in **lines 3, 6,** and **7** and also happens in a parameterized fashion, such that lines can be dilated by different amounts. The purpose of a parameterized dilation is that it allows the assessment to ignore small deviations that have been found in the PtD/P survey (shown in Fig. 13). Intuitively this indicator captures how precisely the hazards' locations are captured in the input solution compared to the ground truth.

The for loop in **line 4** loops over all the input-line, *I*, in **N**. In **line 5,** it is checked if the label of *I* is equal to the label of *GT*. If this is the case, the *GT_Polygon* is updated to be the Boolean difference to the dilated *I, this is similar to GT_Polygon* not *I*. After the input lines in **N** have been handled, the area of the remaining *GT_Polygon* is computed relative to the area of the initial *GT_Polygon*. The result is captured as a percentage and is added to the list *S* with the list concatenation operator (^).

Algorithm 1. Spatial_Correctness_Score

Input: M, I
Output: S (a list of spatial correctness in percent)
1: *S* = []
2: for each *GT in* **M**:
3: *GT_Polygon* := Dilate(*GT*)
4: for each *I* in **N**:
5: if Label(*I*) == Label(*GT*):
6: *GT_Polygon* := Difference*(GT_Polygon,* Dilate(*I*))
7: *S* := *S*^[(Area(*GT_Polygon*)/Area(Dilate(*GT*)))*100]
8: return *S*

7. Case study

In this case study, we are using the approach proposed in this work to assess the performance of our Prevention through Design and Planning (PtD/P) algorithm called SafeConAI, which is further described in [28,37]. SafeConAI is currently under development, which means that its results must be assessed such that we, as the developer, know where the software does not comply with the regulation. This case study will explain how the output of the work described in this work can be utilized to assess the performance of an automated approach. This section also describes the output of the assessment with examples that showcase how the proposed solution can be used and interpreted.

SafeConAI uses the ontology that is described previously and operates directly on the IFC files, which are freely available on the GitHub repository [5]. The input format is IFC, and the results are captured in an outputted IFC file called the safe BIM. Screenshots of the input benchmark models are shown in Fig. 11 and the resulting enhanced output models are shown in Fig. 15. Besides the outputted BIM model (IFC file), the SafeConAI application creates a CSV file of all the identified leading edges of the input BIM model. The format of the CSV file complies with the format described previously.

The CSV file is uploaded to the automated assessment web application, which generates the report shown in Fig. 16. The annotated box (A) captures the performance indicators in percent, which is also shown in the percentage bar. Additionally, for completeness and soundness, the indexes of the lines that were not sufficiently mapped are shown in the lists. This facilitates debugging. The annotated box (B) captures the graphical representation and consists of four plots. First (upper left), plots the lines of both the input solution and the ground truth or one of these based on the selector at the top. The second (upper right), captures a visualization of the mapping matrix, that allows the user to get a visual

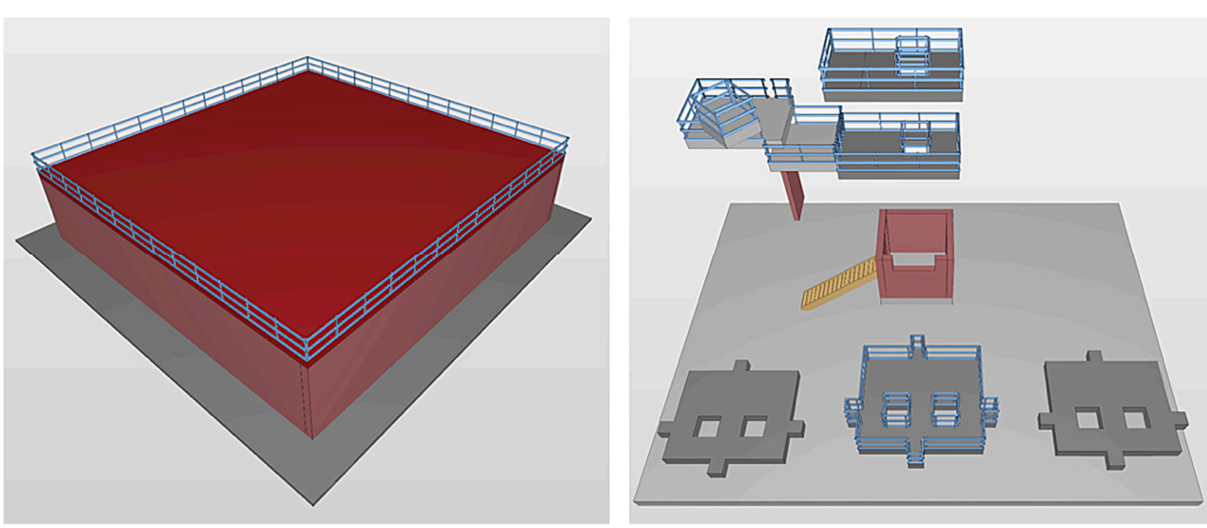

Fig. 15. 3D view of the resulting safe version of the low complexity benchmark Model A (left) and Model B edge-case model (right).

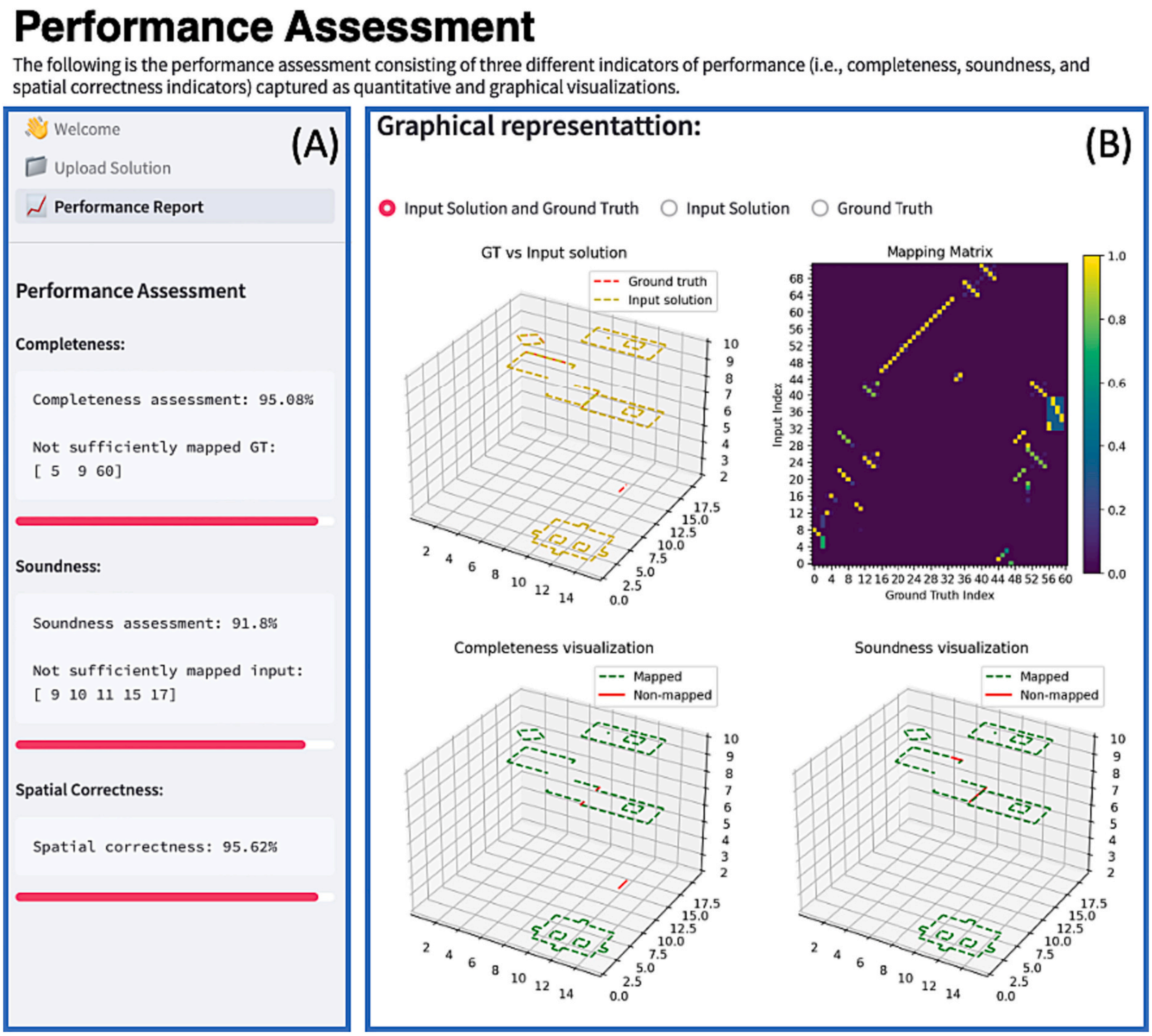

Fig. 16. The resulting performance assessment of the SafeConAI output for Model B. The analysis is done in the web application [38]. The annotated box (A) captures the quantitative values of the performance indicators, and (B) captures the graphical representation.

understanding of the similarity score and, therefore, the mapping of the individual lines. The third (lower left) captures a visualization of the completeness, where the correctly *Mapped* lines are plotted with a dashed green line, and the *Non-mapped* lines (i.e., the lines from the list in the annotated box (A)) are plotted with a solid red line. The same goes for the fourth and last plot (lower right), but for the soundness assessment. Besides the described functionality, it is also possible to get the textual output of the mapping verdict (i.e., a list of input line indexes and the corresponding ground truth index). Likewise, it is possible to get the quantitative representation of the mapping matrix from the web application.

After both Model A and B have been assessed in SafeConAI, we uploaded the solutions as input to the performance assessment application, which yielded the quantitative results presented in Table 5. The results show that SafeConAI achieves 100% across all three indicators for bechmark model A, and scores greater than 91% across all indicators for benchmark model B.

Table 5
Performance assessment of SafeConAI results.

Model	Completeness	Soundness	Spatial Correctness
Benchmark model A	100,0%	100,0%	100,0%
Benchmark model B	95,08%	91,80%	95,62%

8. Discussion

Before going into the discussion of the we emphasise that this work is limited to only *falls from height* hazards, as mentioned in the introduction. However, performing a study where all hazards are considered would be impossible, considering their ambiguity and complexity. Additionally, as mentioned in the related work section, automated safety algorithms are not yet capable of analyzing all hazard types, and thus we focused on the most widespread type to target the most people possible. While the academics researching automated safety assessment algorithms are the target audience of this work and the produced outcome, the practitioners have been kept in mind. Knowing that most safety engineers have limited use and access to BIM models and BIM tools, it has been decided to offer both options in the manual assessment study but still account for the future, where the tendencies are showing that yet more planning work is carried out in BIM, and safety planning is no exception. The proposed work and its indicators is meant for academics desiring to capture their own progress and compare it to the community state of the art. This also means that the indicators can appear a bit abstract to some practitioners. The indicators may, however, be considered when the state of the art advances to a stage where practitioners have to decide which algorithm to use for their specific case.

The safety ontology that we have presented builds on existing research (as detailed in the related work section), by incorporating the knowledge gathered in [11–14], but is now based on spatial artefacts, which enables the algorithms to include more details in the assessed environment. Based on the ontology, it has been possible to create two benchmark models. The two models, especially Model B, include edge cases from the regulation. The edge cases were identified through the analysis and creation of the ontology and are envisioned to challenge the automated safety assessment algorithms.

Through the developed workshop, it was possible to interact with domain experts from the industry, who do safety assessments as part of their daily work tasks. The workshop participant was introduced to the two models and was asked to identify the leading edges. Throughout the workshop, it became clear that the edges in the model that strictly needed protection equipment were all identified, but also that best practice plays a role in safety assessment. Examples are pointed out in the Fig. 12, where some participants identified edges that should be protected, without a clear demand from the regulation. The reasoning was clear "those edges would also lead to at least accidents if they existed in a construction situation". When it was decided not to include the non-demanded leading edges, the advocation was that it should be fair to the automated assessing strategies, which are usually based strictly on the regulation. The leading edges introduced because of best practices should be considered when automated PtD/Palgorithms are developed. Therefore, it is envisioned that a future version of the assessment application would have the option to include optional best practice leading edges, but this could also be something that should be considered on a regulation authorities' level.

The performance assessment application has been made available as an online web application, which the community can use. To use the application, the user must upload an input file following the format described in Section 7. The file format has been decided to be CSV, as this format is a possible export file format from most proprietary BIM modeling tools, which showed to be used for leading-edge assessment in the survey that was performed to get an overview of the current solutions. Besides that, CSV is also widely supported and used in other programming frameworks. After a user has uploaded an input file with a solution, it is possible to go to the "performance report" tab, which computes and presents the performance in the three indicators presented in Section 8. The indicators have been carefully defined to ensure a fair and comparable assessment. The indicators have been inspired by the performance assessment that is used in computer vision. Thus, the inputted lines are turned into polygons representing bounding boxes. This enables us to assess the IoU, which in combination with a normalized distance offset, allows the mapping of incoming lines to the lines in the ground truth even when the ordering is not identical, which cannot be expected.

The case study presents the usage of the assessment application on a result from our PtD/P algorithm (SafeConAI). SafeConAI has not been changed to perform better for this test, and the presented results are the current stage of the algorithm. Nonetheless, it is now visible, that the assessment differs from the ground truth, which can be further investigated. The performance of SafeConAI is shown in Table 5, which for Model A is 100% for completeness, soundness, and spatial correctness. For Model B, the results are 95%, 92%, and 96%, respectively. The completeness of 95% corresponds to a lack of three hazard lines that SafeConAI did not identify, and the soundness of 92% corresponds to five hazards line that were redundant compared to the ground truth. The spatial correctness of 96% indicates that the placement of the identified hazards was accurate. Fig. 16 also shows graphically which of the lines that are were redundant, and which that are missing in SafeConAI's solution. These results enable us to investigate specific scenarios to improve the assessment performance. Based on these numbers and the graphical material provided, it is now up to the developer to look into the reason for missing or redundantly identified hazards. For the practitioner, the numbers would illustrate if all hazards are identified correctly by a piece of software. From the perspective of a safety engineer, the most important indicator is completeness, representing how well an algorithm identifies all hazards. That said, the practitioners would also have an interest in the soundness parameter as it can help to avoid uneccesarry protective equipment and installation time, and consequently, budget.

9. Conclusions and outlook

This paper presents a formal framework for measuring the efficacy of automated safety analysis tools, which comprises three indicators of soundness, completeness and spatial correctness. Our framework is based on our construction safety ontology that, in contrast to other ontologies for construction safety, incorporates spatial artefacts. Spatial artefacts allow for an additional layer of abstraction, where the surroundings in a work area are also considered before the spatial artefacts are created, e.g., that an existing wall should be subtracted from the walkable space spanned by a slab if the wall is standing on the slab, and also that the door hole should not be subtracted from the walkable space.

Based on the safety analysis framework and ontology, we have created two freely available benchmark models to support the research and development community. Model A serves as a simple situation, which should be possible to assess by most algorithms. Therefore, Model A also serves as a minimum viable solution, that can be used to ensure file formatting is as described. The second model, Model B, contains complex edge case scenarios extracted from the regulation and should test the algorithms' capabilities to capture the regulation. The overall purpose of the benchmark models is that the community can download and assess these in their own PtD/P approach. This enables new developers to get started without having to 3D model themselves, and working on a commonly shared model allows them to compare to others' results.

To be able to compare the different solutions based on quantitative identifiers, it is necessary to create a platform that is based on third-party information, i.e., experts in the industry. Therefore, we collected feedback from the practitioners in the industry through our workshop procedure. The procedure is shared [5] to enable others to continue the work, but also to ensure full disclosure.

This work describes a process of digitizing the domain expert feedback, as well as a process for capturing hazards in a common way. The analysis of the capturing format is based on the survey shown in Table 4, where an overview of the existing studies, and their approaches are analyzed and considered in the final proposed format. The input solution (i.e., solution provided by users) and the ground truth (i.e., combined solution from domain experts) use the same format, which can also be interpreted manually.

To compare the input solution and ground truth, it was necessary to develop an approach that is robust in terms of the ordering and segmentation of the lines, i.e., it cannot be assumed that the analysis approaches identify the hazards in the same order and that a line segment corresponds to an identical line in the ground truth. Therefore, we propose a mapping strategy, which makes the basis for the following performance assessment that evaluates the input in terms of proposed completeness, soundness, and spatial correctness indicators. The performance assessment application is shared as an online web application, which ensures availability and easy access that does not require any installation for the individual user.

In this study, it was identified that the so-called best practices play an important role (e.g., applying protection equipment, in areas where there is no strict demand from the regulation). It is envisioned that such best practices should be mapped out and included in the ground truth assessment upon request from the user. Additionally, it would be beneficial to include another indicator, which captures the ability to capture these scenarios. Another improvement to this 3D benchmark model assessment study, would be to use a similar approach to capture a 4D scenario, which would mean that the steps of this study would have to be investigated in a scheduled construction situation.

Data availability

All data has been shared directly in manuscript

Acknowledgment

The research presented in this paper was funded by the European Union Horizon 2020 research and innovation program under grant agreement nos. 958398 and 958310.

References

[1] A. Pinto, I.L. Nunes, R.A. Ribeiro, Occupational risk assessment in construction industry – overview and reflection, Saf. Sci. 49 (5) (Jun. 2011) 616–624, https://doi.org/10.1016/j.ssci.2011.01.003.

[2] B. of labour statistics BLS, Fatal Occupational Injuries by Industry and Event or Exposure, all United States, 2020, 2020 https://www.bls.gov/iif/oshwc/cfoi/cftb0340.htm#cfoi_at_a1.f.2 (accessed Feb. 16, 2022).

[3] M. Fargnoli, M. Lombardi, Building information modelling (BIM) to enhance occupational safety in construction activities: research trends emerging from one decade of studies, Buildings 10 (6) (Jun. 2020), https://doi.org/10.3390/buildings10060098. Art. no. 6.

[4] K.W. Johansen, C. Schultz, J. Teizer, BIM-based fall Hazard ontology and benchmark model for comparison of automated prevention through design approaches in construction safety, in: 29th International Workshop on Intelligent Computing in Engineering (EG-ICE), Jan. 2022, https://doi.org/10.7146/aul.455.c231.

[5] K.W. Johansen, C. Schultz, J. Teizer, BIM-Based3DBenchmark, Jan. 09, 2023. Accessed: Jan. 09, 2023. [Online]. Available: https://github.com/kakke14/BIM-Based3DBenchmark.

[6] R. Collins, S. Zhang, K. Kim, J. Teizer, Integration of Safety Risk Factors in BIM for Scaffolding Construction, International Society for Computing in Civil and Building Engineering (ICCCBE),(Jun (2014) 307–314, https://doi.org/10.1061/9780784413616.039.

[7] B. Li, C. Schultz, J. Teizer, O. Golovina, J. Melzner, Towards a unifying domain model of construction safety, health and well-being: SafeConDM, Adv. Eng. Inform. 51 (Jan. 2022), 101487, https://doi.org/10.1016/j.aei.2021.101487.

[8] J. Melzner, S. Zhang, J. Teizer, H.-J. Bargstädt, A case study on automated safety compliance checking to assist fall protection design and planning in building information models, Constr. Manag. Econ. 31 (6) (Jun. 2013) 661–674, https://doi.org/10.1080/01446193.2013.780662.

[9] K. Schwabe, J. Teizer, M. König, Applying rule-based model-checking to construction site layout planning tasks, Autom. Constr. 97 (Jan. 2019) 205–219, https://doi.org/10.1016/j.autcon.2018.10.012.

[10] J. Teizer, K.W. Johansen, C. Schultz, The Concept of Digital Twin for Construction Safety, Construction Research Congress (CRC, Mar. 2022, pp. 1156–1165, https://doi.org/10.1061/9780784483961.121.

[11] Y. Lu, Q. Li, Z. Zhou, Y. Deng, Ontology-based knowledge modeling for automated construction safety checking, Saf. Sci. 79 (Nov. 2015) 11–18, https://doi.org/10.1016/j.ssci.2015.05.008.

[12] H.-H. Wang, F. Boukamp, Ontology-based representation and reasoning framework for supporting job Hazard analysis, J. Comput. Civ. Eng. 25 (6) (Nov. 2011) 442–456, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000125.

[13] S. Zhang, F. Boukamp, J. Teizer, Ontology-based semantic modeling of construction safety knowledge: towards automated safety planning for job hazard analysis (JHA), Autom. Constr. 52 (Apr. 2015) 29–41, https://doi.org/10.1016/j.autcon.2015.02.005.

[14] Z. Zhou, Y.M. Goh, L. Shen, Overview and analysis of ontology studies supporting development of the construction industry, J. Comput. Civ. Eng. 30 (6) (Nov. 2016) 04016026, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000594.

[15] M. Bhatt, C. Schultz, M. Huang, The Shape of Empty Space: Human-centred cognitive foundations in computing for spatial design, in: 2012 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC), Sep. 2012, pp. 33–40, https://doi.org/10.1109/VLHCC.2012.6344477.

[16] G. Tekbas, G. Guven, "BIM-based automated safety review for fall prevention," in Advances in Building Information Modeling, in: S. Ofluoglu, O.O. Ozener, U. Isikdag (Eds.), Communications in Computer and Information Science, Springer International Publishing, Cham, 2020, pp. 80–90, https://doi.org/10.1007/978-3-030-42852-5_7.

[17] S. Zhang, J. Teizer, J.-K. Lee, C.M. Eastman, M. Venugopal, Building information modeling (BIM) and safety: automatic safety checking of construction models and schedules, Autom. Constr. 29 (Jan. 2013) 183–195, https://doi.org/10.1016/j.autcon.2012.05.006.

[18] B. Li, R.O. Nielsen, K.W. Johansen, J. Teizer, P.G. Larsen, C. Schultz, "Towards digital twins for knowledge-driven construction progress and predictive safety analysis on a construction site," in Leveraging Applications of Formal Methods, Verification and Validation: Tools and Trends, in: T. Margaria, B. Steffen (Eds.), Lecture Notes in Computer Science, Springer International Publishing, Cham, 2021, pp. 153–174, https://doi.org/10.1007/978-3-030-83723-5_11.

[19] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, ImageNet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, Jun. 2009, pp. 248–255, https://doi.org/10.1109/CVPR.2009.5206848.

[20] A. Xuehui, Z. Li, L. Zuguang, W. Chengzhi, L. Pengfei, L. Zhiwei, Dataset and benchmark for detecting moving objects in construction sites, Autom. Constr. 122 (Feb. 2021) 103482, https://doi.org/10.1016/j.autcon.2020.103482.

[21] E. standards ES, Temporary Edge Protection Systems - DS_EN 13374_2013+A1_2018, ES, European standards, 2018.

[22] B. for B. & A. BFA, Håndbogen - Arbejdsmiljø i bygge og anlæg. Branchearbejdsmiljørådet for Bygge & Anlæg. https://www.haandbogen.info/wp-content/uploads/Håndbogen-2020-print.pdf, 2020.

[23] BG-Bau, Absturzsicherungen auf Baustellen Seitenschutz/Absperrungen | BG BAU - Berufsgenossenschaft der Bauwirtschaft, 2021 (accessed Feb. 2022), https://www.bgbau.de.

[24] OSHA, 1910.29 - Fall Protection Systems and Falling Object Protection-Criteria and Practices. | Occupational Safety and Health Administration, 2019 (accessed Feb. 11, 2022), https://www.osha.gov/laws-regs/regulations/standardnumber/1910/1910.29.

[25] M. Bhatt, F. Dylla, J. Hois, Spatio-terminological Inference for the Design of Ambient Environments, 2009, p. 391, https://doi.org/10.1007/978-3-642-03832-7_23.

[26] C. Schultz, M. Bhatt, Multimodal spatial data access for architecture design assistance | AI EDAM | Cambridge Core, Camb. Univ. Press 26 (2) (2012) 177–203, https://doi.org/10.1017/S0890060412000066.

[27] C. Schultz, M. Bhatt, InSpace3D: a middleware for built environment data access and analytics, Proc. Comput. Sci. 18 (Jan. 2013) 80–89, https://doi.org/10.1016/j.procs.2013.05.171.

[28] K.W. Johansen, C. Schultz, J. Teizer, Hazard ontology and 4D benchmark model for facilitation of automated construction safety requirement analysis, Comput. Aided Civ. Inf. Eng. (2023), https://doi.org/10.1111/mice.12988.

[29] S. Zhang, K. Sulankivi, M. Kiviniemi, I. Romo, C.M. Eastman, J. Teizer, BIM-based fall hazard identification and prevention in construction safety planning, Saf. Sci. 72 (Feb. 2015) 31–45, https://doi.org/10.1016/j.ssci.2014.08.001.

[30] B. Li, C. Schultz, J. Teizer, O. Golovina, J. Melzner, Towards a unifying domain model of construction safety, health and well-being: SafeConDM, Adv. Eng. Inform. 51 (Jan. 2022), 101487, https://doi.org/10.1016/j.aei.2021.101487.

[31] G. Hongling, Y. Yantao, Z. Weisheng, L. Yan, BIM and safety rules based automated identification of unsafe design factors in construction, Proc. Eng. 164 (Jan. 2016) 467–472, https://doi.org/10.1016/j.proeng.2016.11.646.

[32] R. Takim, M.H. Zulkifli, A.H. Nawawi, Integration of automated safety rule checking (ASRC) system for safety planning BIM-based projects in Malaysia, Procedia Soc. Behav. Sci. 222 (Jun. 2016) 103–110, https://doi.org/10.1016/j.sbspro.2016.05.195.

[33] K. Kim, Y. Cho, S. Zhang, Integrating work sequences and temporary structures into safety planning: automated scaffolding-related safety hazard identification and prevention in BIM, Autom. Constr. 70 (Jul. 2016), https://doi.org/10.1016/j.autcon.2016.06.012.

[34] C.K.I.C. Ibrahim, P. Manu, S. Belayutham, A.-M. Mahamadu, M.F. Antwi-Afari, Design for safety (DfS) practice in construction engineering and management research: a review of current trends and future directions, J. Build. Eng. 52 (2022), https://doi.org/10.1016/j.jobe.2022.104352.

[35] Z. Jin, J. Gambatese, D. Liu, V. Dharmapalan, Using 4D BIM to assess construction risks during the design phase, Eng. Constr. Archit. Manag. 26 (11) (Jan. 2019), https://doi.org/10.1108/ECAM-09-2018-0379. Art. no. 11.

[36] J. Yuan, X. Li, X. Xiahou, N. Tymvios, Z. Zhou, Q. Li, Accident prevention through design (PtD): integration of building information modeling and PtD knowledge base, Autom. Constr. 102 (Jun. 2019) 86–104, https://doi.org/10.1016/j.autcon.2019.02.015.

[37] K.W. Johansen, C. Schultz, J. Teizer, Automated spatiotemporal identification and dissemination of work crews' exposure to struck-by hazards, in: Digital Transformation of Health and Safety in Construction, Jun. 2023, pp. 1–10, https://doi.org/10.24840/978-972-752-309-2.

[38] K.W. Johansen, C. Schultz, J. Teizer, Performance Assessment Application, 2023 (accessed Feb. 23, 2023), https://kakke14-assessmentgui-0–welcome-w3jlfn.streamlit.app/.