www.itcon.org - Journal of Information Technology in Construction - ISSN 1874-4753

AUTOMATION FOR HAZOP STUDY: A STATE-OF-THE-ART REVIEW AND FUTURE RESEARCH DIRECTIONS

SUBMITTED: July 2024
REVISED: July 2024
PUBLISHED: September 2024
EDITOR: Robert Amor
DOI: 10.36680/j.itcon.2024.033

Ehab Elhosary, Ph.D Candidate,
Department of Building, Civil and Environmental Engineering, Concordia University, Canada
Department of Structural Engineering, Faculty of Engineering, Tanta University, Egypt
ehab.elhosary@mail.concordia.ca

Osama Moselhi, Professor, Director of Center for Innovation in Construction and Infrastructure Engineering and Management
(CICIEM), Department of Building, Civil and Environmental Engineering, Concordia University, Canada
moselhi @encs.concordia.ca

SUMMARY:*Hazard and Operability Study is a structured and systematic metodology to identify and mitigate potential hazards and operational issues associated with a system, process, or facility. This methodology—dubbed as HAZOP—has been initially applied in the chemical industry and subsequently extended to other process industries. Despite its effectiveness, conventional HAZOP study is time consuming, labor-intensive, expensive,and heavily reliant on human judgement. To address these challenges, intelligent systems and different levels of automation have been developed, including knowledge-based approaches that use domain-specific rules, and expertise and data-driven models that identify potential hazards from historical data patterns. The existing AI HAZOP tools lack both full automation for generating HAZOP reports and a comprehensive knowledge base for detecting hazards and operational malfunctions. This paper provides a detailed literature review on the application of automated HAZOP methodologies across different industries. It summarizes the advancements and contributions made over the past decade, highlighting sophisticated technologies such as powerful knowledge representation formalisms and reasoning techniques. The benefits and shortcomings of existing technologies are discussed and future work directions are proposed.*

KEYWORDS: *Hazard and Operability study, AI HAZOP tools, Knowledge-based approaches, Data-driven models, Comprehensive review.*

REFERENCE: *Ehab Elhosary & Osama Moselhi (2024). Automation for HAZOP study: A state-of-the-art review and future research directions. Journal of Information Technology in Construction (ITcon), Vol. 29, pg. 750-777, DOI: 10.36680/j.itcon.2024.033*

1. INTRODUCTION

HAZOP methodology is considered a promising alternative to other hazard identification methods (such as brainstorming, structured, or semi-structured interviews, the Delphi method, root cause analysis, and business impact analysis) since it provides several advantages, such as a structured method, applicability on a wide range of systems, processes, and procedures for improvements in safety levels (Joubert et al., 2021). The concept of HAZOP study dates back to the mid-1960s at imperial chemical industries (ICI) company in the United Kingdom when a group of engineers was responsible for developing a preliminary version of hazard and operability analysis. However, it was not until 1974 when the Nypro site at Flixborough in North Lincolnshire, England was severely damaged by a large explosion resulting in the death of twenty-eight workers and reported injuries of 36 workers (Heino, 1999). As a result, the very first publication considering the HAZOP study appeared in the same year, defining the principles needed to carry out operability studies and hazard analysis accounting for the increasing complexity of new processes that could not be examined thoroughly using the conventional approaches based on equipment-oriented practices (Lawley, 1974). Two years later, the technical and managerial principles underlying HAZOP studies were introduced, outlining the factors that had to be considered in successful development of HAZOP (Lawley, 1976). In 1977 the Chemical Industries Association published a guide to hazard and operability studies, and the term HAZOP was first used by Kletz in 1983 (Crawley and Tyler, 2015). An international standard—for Hazard and operability studies [BS IEC 61882] was published in 2001 (BSI (British Standards Institute) and updated in 2016 (Joubert et al., 2021).

Subsequently, the HAZOP study has been recognized and received regulatory acceptance, particularly in chemical related industries. However, this was attributed to factors related to the application of HAZOP that include (1) it may be very time-consuming and therefore expensive; (2) it requires a high level of system, process, and procedure documentation; (3) the workshop discussions may be too focused, and so could miss issues with fundamental assumptions and wider or external issues; (4) the process relies heavily on the expertise of the designers, who may find it difficult to be sufficiently objective to spot shortfalls in their designs; (5) inadequate terms of reference or a poor definition of the study's scope may cause difficulties; and (6) it focuses on single events rather than on combinations of possible events. The focus on guide words that allow it to overlook some hazards that are not related to a guide word, and the need to have a trained facilitator, are also seen as limitations (Crawley and Tyler, 2015; Rimkevičius et al., 2016). To address these limitations, intelligent methods and/or systems are deemed required to provide solutions for time and cost reduction and alleviation of human bias and dependencies. Accordingly, expert systems and different levels of automation were widely used in HAZOP studies.

Indeed, Expert Systems offer several benefits when compared with human experts. Expert systems can produce consistent results on the same tasks and produce similar situations consistently, whereas humans can get tired or bored and are influenced by various effects (e.g., their judgments can be easily impacted by new information). Expert systems are inexpensive to operate, easy to reproduce and distribute, and can provide permanent documentation of the decision process. In addition, expert systems may contain knowledge from several human experts, giving them more breadth and robustness than a single expert. Although expert systems enjoy these advantages, they also have many weaknesses that include lack of common-sense knowledge, narrow focus and restricted knowledge, inability to respond creativity to unusual situations and difficulty in adapting to changing environment. (de la O Herrera et al., 2018) reviewed several expert systems such as HAZID, DYNHAZ, Petrohazop, HELPHAZOP, and signed direct graph (SDG) and noticed that in all cases the need for specialists will continue to be necessary due to the expert systems limitations.

In the recent few years, advanced algorithms such as machine learning technology, deep learning and natural language processing (NLP) have been used to determine the causes of the accidents, and describe the accident occurrence processes qualitatively, which helps avoid the recurrence of accidents or minimize the severity of the accident consequences. (Ekramipooya et al., 2023) investigated the effectiveness of NLP and Machine Learning (ML) in HAZOP study and used ML classifiers such as Decision Tree, linear Support Vector Machine, Random Forest, Logistic Regression, Gaussian Naïve Bays, and K-Nearest Neighbors to predict deviations. Natural language processing technology was used to construct a HAZOP report event classification model to ensure consistency in the analysis (Feng et al., 2021). (Wang et al., 2021) conducted active learning to train the deep learning model by mining high-quality samples in dataset through the sampling algorithm, which can reduce the limitation of professional and standard HAZOP text on the model and improve the stability and generalization ability of the model.

Moreover, the utilization of HAZOP method has been extended to other industries including oil and gas, petrochemical plants, environmental engineering, and other domains over the past decade. This noteworthy expansion and progression have captured our interest indicating that the HAZOP method has significant potential for automation in different industries. Such automation can facilitate the analysis of hazards and operability issues, and significantly reduce the overall time consumption associated with traditional HAZOP study. Therefore, there is a need to conduct a review of publications on the automated HAZOP methodology and its application across different industrial sectors to delineate its state of the art and perspectives.

To ensure the usefulness of these studies for industry practitioners and researchers, this study comprehensively reviews the evolution of automation and emerging technologies and algorithms developed to support the application and the practical utilization of HAZOP method across various industries. It is based on a review of publications of the HAZOP methodology from more than 20 diverse scientific journals spanning from 2010 to 2023, aiming to define the current state of the art and perspectives and identify research gaps and propose future research directions accordingly. In this research, keywords including "HAZOP" OR "hazards and operability studies" OR "hazards and operability analysis" OR "operability study and hazards" OR "hazards and operability procedure" OR "automated HAZOP" OR "Intelligent HAZOP" OR "Computer aided HAZOP" were used. The literature was retrieved by searching the contents of titles, abstracts, and keywords in the Scopus database. The Scopus database was selected in this review as it includes the most important literature sources, the highest citations, and abstract numbers compared to Web of Science and Google Scholar (Meho and Rogers, 2008). The objective was to retrieve only original articles on HAZOP technique. Therefore, the publication type was limited to journal articles only and non-English-language publications were excluded. All subject areas on the dashboard were reviewed and publications in Agricultural and Biological Sciences, Social Sciences, Earth and Planetary Sciences, Physics and Astronomy, Mathematics, Medicine, Biochemistry, Genetics and Molecular Biology, Pharmacology, Toxicology, Health Professions, Arts and Humanities, Immunology and Microbiology, Nursing and Neuroscience were excluded. As a result, 347 articles were identified and selected to proceed to the stage of manual literature screening to define eligible articles for further review. The screening process was conducted by checking titles, abstracts, methodology, discussions, and conclusions; only articles that focus on the automation scope of HAZOP method and its application in construction were considered. Studies focused on risk analysis tools such as STAMP, STPA, SIMOPS, fuzzy delphi and fuzzy edas, ...etc. and only indicate HAZOP as a traditional process hazard analysis (PHA) method were also excluded. Moreover, articles that discuss topics such as network tracking of websites, automotive vehicles, security assertion for industrial IoT, ...etc. were excluded. This process yielded 83 publications (75 papers in automation in various industries and 8 papers in areas such as hoisting operations, railway systems, supply chain management, and urban drainage systems) to be used as data in this review. The selected papers include automation of HAZOP in industrial fields such as chemical industry, petrochemical facilities, oil and gas, environmental engineering, and others (including mechanical industry, manufacturing process, production process, safety engineering, …. etc.). Table 1 presents a list of top selected journals, including their CiteScore and impact factor, and the number of publications considered in this review.

Table 1: List of top 10 source journals of HAZOP research.

Journal Title	CiteScore/ Impact Factor	No. of Publications
Journal of loss prevention in the process industries	7.2/3.6	13
Process safety and environmental protection	11.4/6.9	11
Chemical engineering transactions	1.5/1.1	8
Computers and chemical engineering	8.7/3.9	6
Safety science	13.0/4.7	4
Process safety progress	2.3/1.7	3
Processes	5.1/2.8	3
Computers in Industry	18.9/8.2	2
Reliability engineering and system safety	15.2/9.4	2
International Journal of Safety and Security Engineering	1.5/1.089	

In terms of the number of published documents per year, Figure 1 shows the annual research publication trend in terms of HAZOP automation contributions between 2010 and 2023 with a remarkable growth over the period. A total of 75 selected studies were reviewed, among which over 50% were published within the last five years. Relative to the performance of previous years, some years witnessed declined trends in the annual number of research studies. These can be observed in 2011, 2013 and 2014. However, the number of publications has continuously climbed since, peaking in 2023 with 13 studies. The polynomial trend line presented in red color in the figure demonstrates that the automation of HAZOP methodology has increasingly gained attention in recent years, and researchers are actively engaged in developing efficient methods for successful automated HAZOP methodology.

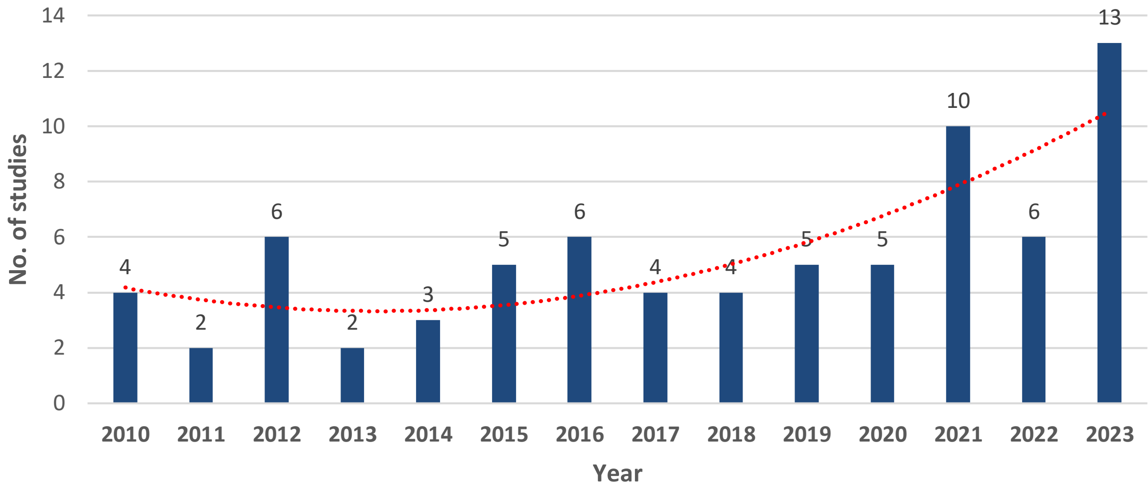

Figure 1: Annual publication trend in automated HAZOP studies from 2010 to 2023.

Organization of this paper: To achieve the paper's objectives, we started with the evolution of HAZOP structure and transitioning from conventional to automation procedures in section 2. Then, a thorough review of extant literature on HAZOP automation endeavors and contributions from 2010 to 2023 is conducted in Section 3. After which, section 4 demonstrates the expansion and application of the HAZOP scope. Opportunities and future work are then presented in Section 5. The outputs of this study will assist researchers in identifying automation HAZOP research scope, trends, and gaps while supporting industrial companies to invest and explore newly developed methods that could be adopted to enhance the overall performance of HAZOP method and solve its pitfalls.

2. TOWARDS HAZOP AUTOMATION PROCEDURES

Over the years, HAZOP study has been applied in the design phase of the projects to identify and assess potential hazards, causes, consequences, and safeguards. It depends on following a system's Process Flow Diagrams (PFDs) and Piping and Instrumentation Diagrams (P&IDs), dividing the design into subsections with definite boundaries called nodes, so ensuring the analysis of each piece of equipment in the process. A multi-disciplinary team undertakes the analysis, whose members should have sufficient experience and knowledge to answer most questions. The members are selected carefully and are given the authority to recommend any needed changes in design. Executing the method relies on using guidewords (such as NO, MORE, LESS OF) combined with process parameters (e.g., PRESSURE, TEMPERATURE, FLOW) that aim to reveal deviations (such as less flow, more temperature) of the process intention or normal operation. This procedure is applied in a particular node as a part of the system characterized for a nominal intention of the operative parameters. Having determined the deviations, the expert team explores their feasible causes and their consequences. For every pair of cause-consequence, safeguards must be identified that could prevent, detect, control, or mitigate the hazardous situation. Finally, if the safeguards are insufficient to solve the problem, recommendations must be considered (Dunjó et al., 2010).

One of the challenges of the conventional HAZOP methodology is that it is essentially a qualitative method, which is commonly complemented by other Process Hazard Analysis (PHA) tools. Ranking risk assessment tools have been used in the chemical industry to support decision-making to set arrangements and promote mitigation measures to treat risks related to chemical processes and storage of dangerous substances (Milazzo and Aven, 2012). The need to reduce time and effort consuming and human dependencies has been increased over the years for more efforts to improve the performance of the HAZOP application. Therefore, automation of HAZOP studies has been a research topic for more than 30 years. Basically, automated HAZOP is concerned with four main tasks: (1) creating a digital representation of the process plant to automatically identify potential hazardous events, (2) data pre-processing, (3) knowledge representation to transform data and information into machine readable knowledge and (4) reasoning system to infer conclusions regarding cause-and-effect relationships and identify consequences. The framework of HAZOP analysis and transitioning from conventional to automation procedures can be shown in Figure 2.

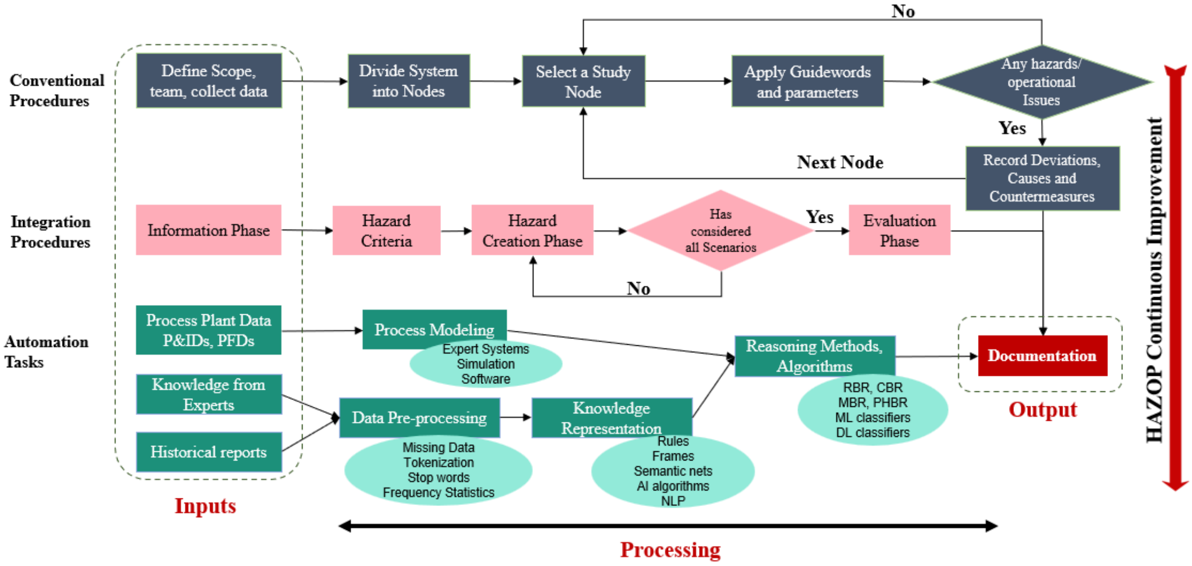

Figure 2: HAZOP framework (Transition from Conventional to Automated Procedures).

2.1 Inputs

The main inputs for both conventional and integrated HAZOP methods consist of the data from the process plants' data being analyzed and the expertise of the HAZOP participants. Typically, the data analyzed in HAZOP workshops includes P&IDs and PFDs, which contain design and engineering details of the plants. The participants contribute their knowledge of processes, equipment, variables, causes, consequences, and safeguards, deriving from their experiences and expertise. These components, along with HAZOP reports from previous projects, have been utilized to automate various aspects of HAZOP process, aiming to streamline operations and reduce time consumption and human dependencies.

2.2 Process Modeling

The process plant must be digitally represented to identify potential hazardous events. Many different design representations were used by process designers to describe a process such as PFD, P&ID, Block flow diagram, Flow sheet, Functional block diagram, Cause and effect diagram (shutdown matrix), Isometric, Engineering data sheet …. etc. (Baybutt, 2016). Often, the process plant is represented using data from P&IDs or PFDs, and qualitatively modeled by listing nodes manually or using a specific modeling language and a graphical editor (J. I. Single et al., 2019). Some approaches are based on graph theory which use Microsoft Visio in a higraph-based modeling representation and require the usage of graphical editors to draw a schematic process plant using a specific modeling language such as LDGHAZOP (Cui et al., 2010), MFM HAZOP Assistant (Rossing et al., 2010) and D-higraphs HAZOP Assistant (Rodríguez and De la Mata, 2012), while methods provide their own graphical editor to model the process. Based on the review in these relevant studies, these methods have challenges. For example, PFD shows only major equipment relationships. P&ID symbols are not to scale and not dimensionally accurate and do not represent physical locations and proximity of each component. Other systems are time-consuming and error-prone, making it more expensive than conventional HAZOP. Other approaches integrate hazard identification into CAD software or process simulators. Simulation and software tools enable engineers to create virtual representations of the plant processes in HAZOP studies, allowing them to simulate various scenarios and analyze the potential hazards associated with each scenario. For instance, the usage of data models that can be extracted from CAD software or process simulators could be done by exporting an XML file which includes all process plant components, connections and instrumentation. Afterwards, this plant representation is processed by the computer aided HAZOP systems (CAHS) and serves as a data basis. This is challenging since objects need to be translated since structure of systems, e.g. CAD software and CAHS, can be very different. Furthermore, this translation process can cause confusion between objects.

2.3 Data Preprocessing

In text analysis and data mining-based methods, prior to classification and conversion of text into numerical form, data pre-processing is essential. This involves several steps. Initially, missing data imputation is conducted to address common occurrences of missing data, which can significantly affect conclusions drawn from the data. The treatment of missing data can be roughly divided into three categories: not processing, direct deletion, and filling (J. Wang et al., 2023). In the cluster process of HAZOP data, blank cells were filled using the word ‘None’, aiding cluster analysis. Subsequently, word segmentation is employed to divide written text into meaningful units, such as words or sentences. Following this, stop words filtering is applied to remove non-essential words like adverbs, adjectives, and conjunctions. The vectorization and word frequency statistics are computed to calculate and record the frequencies of all words in each text element, facilitating the construction of a word index data (Wang and Gu, 2022).

2.4 Knowledge Representation

Knowledge representation involves the transformation of data and information into machine readable knowledge. The quality of the knowledge representation directly influences the inferred results. Therefore, time and effort are required in the thoughtful design of the knowledge representation. Some approaches address this subject superficially. To provide added value to the HAZOP team, solid knowledge regarding safeguards and safety concepts based on sound process safety knowledge is required within the knowledge representation of automated systems (J. I. Single et al., 2019). There are various formalisms for knowledge representation, such as rules (Bassiliades et al., 2011), semantic nets (Studer et al., 2007), frames (Minsky, 1995) and bayesian networks (Hu et al., 2012), ontologies (Aziz et al., 2019) and recently NLP strategies. Rules are formulated to model relationships. The semantic context of strict rules is difficult to map, so there must be a rule for every single relationship, while the context cannot be illustrated. In semantic-nets, knowledge is stored using graphs, where nodes represent objects while the edges (connections) represent the relationships between the objects. Frames can be used to represent stereotypical situations, but frames lack formal semantics, which means that there is no explicit notation to illustrate the meaning of the knowledge. This can lead to ambiguity. Ontologies can be time-consuming and error-prone, making them more expensive than conventional HAZOP studies. Equipment cannot always be accurately modeled, and ontologies require complicated design. To be used for HAZOP automation, ontologies must be carefully established and documented (Daramola et al., 2013; J. I. Single et al., 2019). As a probability based uncertainty inference method, Bayesian networks (BNs) have been adopted extensively to characterize the uncertain interrelationships of various factors, including two steps: establishing the network structure, and calculating network parameters (Gao et al., 2022). For the network structure, the risk factors characterized by nodes and the interrelationships represented by directed arcs need to be identified first. The joint probability distribution is used to characterize the probable relationships among the nodes. Based on the prior probability value of the root node and the conditional probability of the intermediate nodes, the occurrence probability of the target node can be predicted (Gao and Li, 2022). Building BNs requires expert knowledge to specify the network structure and define prior probabilities, which can be subjective and difficult to elicit. Recently, NLP techniques have been applied in HAZOP study, including traditional methods and advanced algorithms and topic modeling, to enable the reasoning systems to infer conclusions

2.5 System Reasoning

To automatically reason potential hazardous events within HAZOP study's conclusions, cause-and-effect relationships or consequences need to be identified based on the knowledge representation. The process of inferring conclusions is called reasoning. (J. I. Single et al., 2019) presented in his review four main reasoning approaches: rule-based, case-based, model-based qualitative and quantitative, and process-history-based. In rule-based reasoning, rules must be accurately aligned with problem information; the strategy fails if rules do not fit (Cameron et al., 2017). The case-based reasoning method is a pattern-based problem-solving method. Knowledge about past situations (cases) is utilized to reuse it in new situations, which means it is a memory-based process. In case-based reasoning (CBR), a significant number of cases are necessary for practical results; thus, this method's quality depends on the size of stored data (Daramola et al., 2013). Model-based quantitative reasoning is based on mathematical models expressed using differential-algebraic equations (DAEs). Model-based qualitative reasoning is concerned with the inference of conclusions in a symbolic manner, without using mathematical relationships (Venkatasubramanian et al., 2003a). Another major group of reasoning techniques is (process) history-based reasoning, where historical process data is analyzed to draw conclusions regarding potential hazards (Venkatasubramanian et al., 2003b). The predictions made by history-based reasoning method depend on the process data's reliability. (Folger and Stein, 2017) classified reasoning techniques as inductive, deductive, or abductive. Figure 3 shows a classification of different reasoning methods. Knowledge-based and data-driven reasoning can be distinguished. While knowledge-based reasoning can be divided into bayesian, deductive and analogical reasoning, data-driven reasoning is usually inductive. Building Bayesian networks involves creating probabilistic models that represent the dependencies between different variables and Bayesian reasoning can manage uncertainty and probabilistic relationships but this depends on having extensive and reliable data to estimate probabilities, which can be time-consuming to gather and validate. Deductive reasoning starts from general facts and rules to derive conclusions, encompassing HermiT reasoner used to create ontology models which determines the relationship between a new concept and other concepts in a network and rule-based reasoning. On the other hand, analogical reasoning is based on heuristics to calculate similarity between a new situation and past cases, but the case based reasoning (CBR) effectiveness depends on the richness of stored cases or knowledge repository of past experiences. Recently, data driven reasoning techniques, including machine learning classifiers and deep learning models have been applied to identify patterns using training data to predict deviations and their corresponding countermeasures in HAZOP studies.

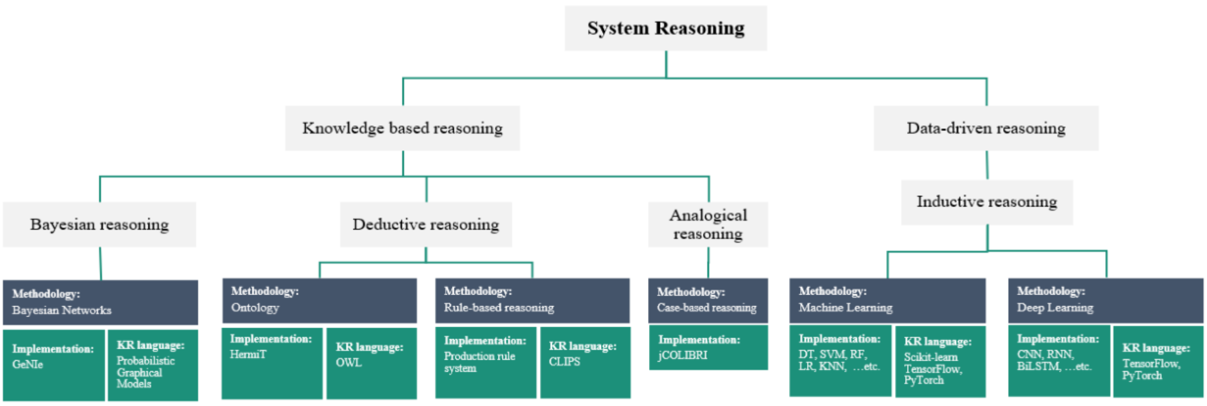

Figure 3: Overview of reasoning techniques used in HAZOP study within the research period.

3. HAZOP AUTOMATION EVOLUTION

The traditional application of HAZOPs becomes a challenge because of the time consuming, the higher complexity of modern systems and the potential human error of manual processes. To reduce the time required, intelligent methods and/or systems have been developed to automate HAZOP methodology and enhance its performance across various industries. The publications related to HAZOP automation have been reviewed and grouped into sub-categories based on techniques and methods that have been developed within the research period, structure, procedures and the domain of application. These criteria have been selected due to their ability to capture different characteristics of HAZOP methods and various study objectives. The literature distinguishes techniques based on procedures that have been used (e.g., data preprocessing, knowledge representation, process modeling). Accordingly, knowledge-based approaches, encompassing expert systems, computer simulation, integrated-automation tools, and artificial intelligence algorithms and data driven models, including natural language processing techniques, machine learning classifiers and deep learning models were identified as shown in Figure 4. In terms of subject area, most automation techniques have been applied in the chemical industry, oil and gas and petrochemical industry (39, 15 and 9 articles, respectively). Additionaly, 9 articles were found distributed across various areas, including mechanical industry, manufacturing process, production process, design engineering and safety engineering. However, our exploration revealed a conspicuous scarcity of automation studies within the field of environmental engineering, with only three publications. Remarkably, no publications were identified pertaining to automated HAZOP procedures within domains, including railway systems, supply chain management, urban drainage systems, and hoisting operations. Only 8 articles were found regarding the conventional application of HAZOP and its integration with PHA tools in these domains during the research period. This discrepancy underscores an untapped potential for the integration of automation methodologies in these sectors, particularly in environmental engineering and above mentioned domains, where such advancements could yield significant efficiency gains and safety enhancements. Additionally, it's noteworthy that no publications were identified regarding the integration of HAZOP method with the recent digital technologies, such as Building Information Modeling (BIM) and digital twin systems. These technologies offer advanced capabilities for visualization, simulation, and analysis of complex systems, including industrial plants and construction projects. Integrating BIM and digital twin systems with HAZOP study could potentially enhance the effectiveness of hazard identification and risk management processes in industrial plants and construction projects by leveraging the advanced features of these technologies. The overall scope categorization results and the proportions of each category are illustrated in Figure 4. Accordingly, the following sections present in greater detail the literature under each category and the classification results.

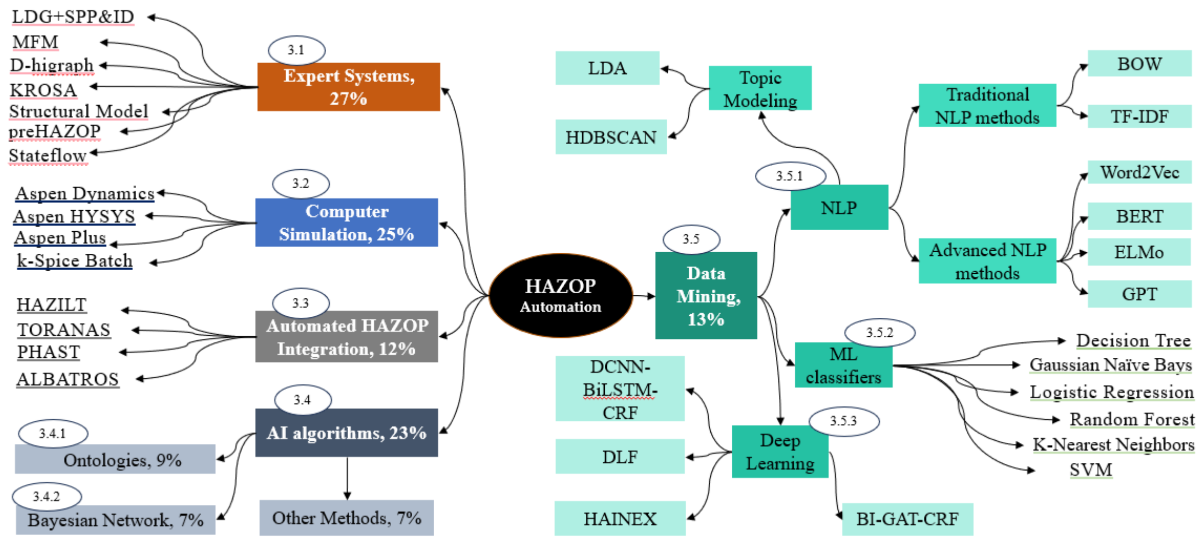

Figure 4: Overview of HAZOP Automation review scope.

Moreover, a time zone map was generated using VOSviewer software (version 1.6.20) as shown in Figure 5, to depict the relationships and trends in the field of automated HAZOP Study within the research period from 2010 to 2023. The map highlights key areas of focus, including the development and application of various tools and algorithms aimed at automating different aspects of HAZOP study. VOSviewer is a visualization tool that offers sufficient features for scientifically mapping literature (van Eck and Waltman, 2010). The visualization capabilities of this software were leveraged to conduct time zone analysis. In this map, each node represents the average year in which the keyword is being used in the literature. The threshold of keyword occurrences is set to three, determined through multiple experiments to enhance the clustering representativeness. Also, it should be noted that a thesaurus file is used to merge similar keywords in the map; for instance, "automation" is replaced with "automated hazop". In this analysis, it is obsrved that around 2018 and 2020, the focus was on describing the process dynamic simulation in case of failures for safety analyses. More recently, automation methods such as data mining including natural language processing and deep learning have received more attention in the HAZOP research field. Furthermore, the heatmap displayed on the lower right-hand side in the figure illustrates the distribution and progression of these algorithms and technologies over the years. The varying intensity of colors in the heatmap highlights the periods of increased research activity and the introduction of new technologies, providing insight into the temporal trends and the growing complexity of tools used in automated HAZOP study. It highlights that the majority of expert systems studies were conducted between 2010 and 2018, while ontological studies reached their peak around 2020. Notably, the data mining techniques including natural language processing algorithms, machine learning classifiers, deep learning models, represent the latest trend in HAZOP automation studies.

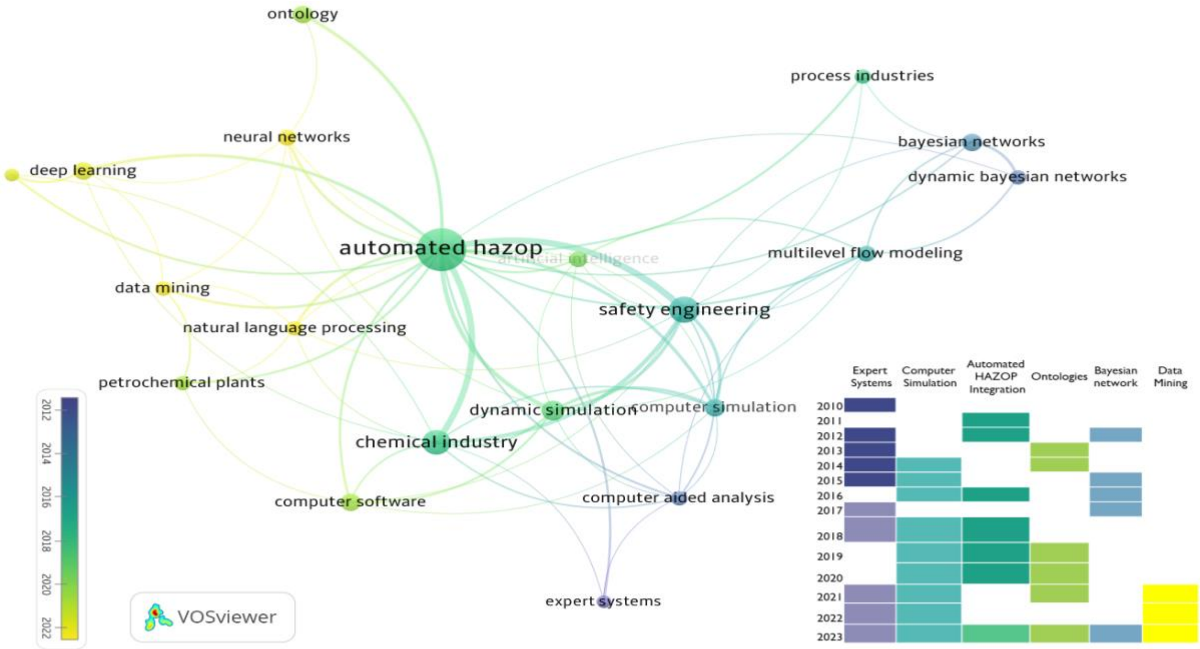

Figure 5: Timeline of HAZOP Automation Categories.

3.1 Expert Systems

Generally, an expert system, a computer program built for the diagnosing, fault-finding, and problem-solving purposes, closely matches the human logic thinking process. The expert system is always composed of knowledge databases, inference engines, knowledge-acquiring regulations, and human interfaces, etc. The information stored in the knowledge database could be employed to deduce new facts. There are some expert systems which have been widely used in chemical process industries, such as the INTEMOR program (real-time intelligent monitoring and accident preventive systems) and the PES program (petroleum production real-time analysis design expert system), etc., because human operations always play a tremendous role in running the plant and the operators need the corresponding guidance of the expert systems, especially in some complex process plants (Liao, 2005). Several researchers have attempted to develop expert systems to resolve the drawbacks of conventional HAZOP such as time-consuming and labour-intensive activity. (Dunjó et al., 2010) reviewed expert systems in chemical process covering a period from 1974 to 2010. Massive efforts made towards this goal, including MFM HAZOP (Rossing et al., 2010), D-higraph HAZOP (Rodríguez and De la Mata, 2012), Structural Model (Boonthum et al., 2014), Knowledge Reuse-Oriented Safety Analysis (KROSA) (Daramola et al., 2013), Stateflow (Chia and Naraharisetti, 2023), and preHAZOP (Oeing et al., 2023). The expert systems category comprises 20 articles, representing 27% of the total (see Figure 4) indicating the most category used over the years during the research period. These systems have been applied across various industries, with 11 articles in the chemical industry, 7 in the oil and gas sector, one study in the petrochemical industry and one article in the process industry.

Examples of research using simulation tools: (Chia and Naraharisetti, 2023) proposed a semi-automated HAZOP study using a computer-aided tool called Stateflow in the oil and gas industry. It used the knowledge base to identify causes of deviations for a wide variety of process in the HAZOP study. This tool started with preparation of raw data and information from two components which are the process plants' data (P&IDs and PFDs) and the knowledge and expertise of the HAZOP participants to achieve computer-aided automation. A protocol and data structure were set up to translate the textual information into a format that the software was able to recognise and manipulate. The second step managed the knowledge base component, where all the information relating the deviations to known industrial causes and consequences were transformed into an organised database library. Once the library of information has been established in the system, the next step would be to set up the digital representation of the other component necessary for the automation, the process plants' data. Using Stateflow's graphical interface, modelling the process plants' data was carried out by designing units or Stateflow blocks which contained unique decision logics, algorithms, and necessary reasoning methods to propagate HAZOP outputs by computation. The final step will be to convert these data outputs back into qualitative expression as the HAZOP worksheet to complete the entire automation process. However, these characteristics entail specific limitations, such as issues with output data quality and the manual effort involved in developing the knowledge database. Table 2 outlines the advantages and disadvantages of the expert systems developed during the study period.

Table 2: Advantages and Disadvantages of expert systems in HAZOP Automation within the research period.

Expert System	Advantages	Disadvantages
Multilevel Flow Model (MFM) (Rossing et al., 2010)	• Describing the purposes and functions of the process and the causal relationships between them and allowing efficient causal reasoning. • Comprising objectives, flow structures, flow functions and a set of means-end relations and causal roles representing purpose related dependencies between functions. They are linked to form a hypergraph.	• Time consuming, depending on the size of the plant, as the user needs to apply the MFM language to model the process plant and its components. (J. I. Single et al., 2019). • The representation of a process plant is complex and error prone for large systems (Taylor, 2017).
LDG HAZOP + SPP&ID (Cui et al., 2010)	• Improving quality of design. • Performing analysis easily at anytime of the whole lifecycle of a plant. • Creating and improving plant configurations efficiently.	• The LDG model library's scale is small. • Needing new data acquisition modules for integration with CAD systems.
D-higraph HAZOP assistant (Rodríguez and De la Mata, 2012)	• Integrating functional and structural information within a single model (MECHHOUD et al., 2017). • Advanced method to model the relationships between process variables.	• Time consuming for large process plants. • Complex and error prone especially for large systems (Taylor, 2017). • More costly than a human made HAZOP study.
KROSA (Daramola et al., 2013)	• Facilitates the knowledge reuse from past studies to support experts by considering the semantic context. • Integration of CBR and ontologies seems to be a promising approach.	• The performance depends on the quality of the domain ontology. • The complexity of process plants and the quality of the process specific knowledge could be a limiting factor.
Structural Model (Boonthum et al., 2014)	• Using a matrix to represent the relationship among all variables in the system. • Using a digraph technique to represent the relationships between causes, consequences, and unit variables to formulate HAZOP digraph model from all relevant parameters.	• Many process variables propagate through plant components, complicating graph models. • The causes and consequences are unclear. • Time-consuming and error-prone, making it more expensive than conventional HAZOP.
PreHAZOP (Oeing et al.,	• Consisting of detecting possible hazardous scenarios in P&Ids using a machine-readable database and carrying out risk assessment for	• This tool is a prototype for feasibility studies. To achieve reliable results, the scenario database needs to be

2023)		these scenarios. • Identifying hazards and reducing design errors and costs, in the early engineering phase while the initial P&ID is being prepared.	extended by expert knowledge. • The safety of the plant is often analyzed at equipment level. Therefore, it is not guaranteed that the weakest component will always be reliably detected.
Stateflow (Chia and Naraharisetti, 2023)		• The time can be reduced with the model library and the knowledge base being set up. • Users could make adjustments during the entire process, such as modifying the plant layout and adding or removal of equipment. • The outputs can be saved and extracted when necessary which overcame the issue of physical HAZOP records that could be easily lost or replaced in a plant over time.	• The effort and time are necessary for knowledge acquisition before the database can be effectively applied to the model due to the large number of existing processes and equipment. • Causes such as operation errors, maintenance of equipment, were not covered in the knowledge base. • Excluding consequences and recommendations from the knowledge base makes the database inadequate for the more complex plants.

While these expert systems offer valuable capabilities, their limitations in terms of complexity, time consumption, error proneness, integration challenges, and potential costs should be weighed against the benefits they bring to the hazard analysis process. Subsequently, alternative methods and tools have been formulated in automated HAZOP analysis.

3.2 Computer Simulation

HAZOP analysis has become more significant as the complexity of process technology has increased. There is a challenge in describing the process dynamic behavior in case of failures. Consequently, there is a real interest in knowing the system behavior during malfunctions for safety analyses. For this purpose, dynamic simulation is an efficient tool to predict the evolution of variables in chemical processes during deviations from normal operating conditions. In a dynamic HAZOP, the process simulation of the system under examination is integrated with the traditional HAZOP methodology to facilitate hazardous scenario identification and risk assessment. This integration helps mitigate the subjectivity associated with determining the severity and likelihood of hazardous events. Software tools for hazard identification automation based on thorough process simulations require suitable simulation environment and process hazard analysis (PHA) methodology. There are two principal options for simulation environment. The first one employs mathematical models developed specifically for the analyzed unit (Danko et al., 2017). Own mathematical model provides full control over equations solving methods and thus of the process simulation results. The second principal option for simulation environment is the use of commercial process simulators to simulate parts of a chemical plant by simply adding or removing available components and unit operations in the process flow sheet (Janošovský et al., 2019). In recent years, a variety of commercial process simulation software has been utilized in HAZOP studies, including Aspen Plus, Aspen Dynamics, and Aspen HYSYS, which are process simulation software of Aspen Tech company (AspenTech., 2023) and k-Spice, a process simulation software of Kongsberg company (Kongsberg, 2023). This category is found to include 19 articles (25%). Among these, 11 articles focus on the chemical industry, 6 are related to the oil and gas industry, one study pertains to the petrochemical industry and one article is centered on design engineering.

Examples of research using simulation tools: Aspen HYSYS was applied to a case study of an ammonia synthesis plant to detect and report hazardous events and operability problems. Based on simulation data evaluation performed in the semi-automatic manner by the proposed tool, a HAZOP-like report containing deviations and their causes and consequences was generated (Janošovský et al., 2019). (Emami et al., 2022) used Aspen Plus to simulate deviation scenarios and predict process behavior in different situations, quantify unsafe process boundaries and examine the propagation of deviations through the process. The consequences of various deviation scenarios and a detailed analysis of the reasons for the occurrence of the consequences were elaborated. The simulation results demonstrated that new protective layers were designed to prevent various high-risk consequences. Table 3 summarizes the simulation software tools that have been used in HAZOP studies within the research period.

Table 3: Integration of HAZOP with Simulation Software for Hazard Identification.

Model	Software	Purpose of Use	Reference
Dynamic simulation-based quantitative HAZOP	Aspen Plus V11 and Aspen Plus Dynamics V11	• Quantifying deviations and consequences in a hydrocracking unit to reduce risk factors of fire and explosions accidents.	(Yi et al., 2023)
HAZOP automatic hazard analyzer (HAZOP-AHA)	Aspen HYSYS	• Integrating HAZOP with Aspen HYSYS and artificial neural networks to explore the feasibility of deviation quantification and predict the deviation severity. • The results of predicted deviation severity were close to the actual deviation severity, and the accuracy was nearly 100%.	(C. Wang et al., 2022)
Dynamic HAZOP simulation	Aspen Plus	• Investigating the the most frequent disturbances in a natural gas refinery. • Simulating the deviation scenarios to calculate the time required for an accident to occur after the initiation of each deviation. • Analyzing the consequences of each scenario with their risk levels to propose appropriate control structures to reduce the consequences.	(Emami et al., 2022)
Dynamic simulation, HAZOP and Risk assessment Model	Aspen Plus Dynamics	• Combining Aspen dynamic, HAZOP and risk matrices to identify the hazardous scenarios leading to major accidents in a semi-batch reactor. • Quantifying the effects and calculating the corresponding risk level for each scenario.	(Berdouzi et al., 2018)
Simulation based Model	K-Spice	• Generating failure scenarios by considering process equipment deviations with pre-defined failure modes. • Evaluating failure scenarios using dynamic simulations to rank their significance and identify the most critical process parameters and equipment in a system.	(Enemark-Rasmussen et al., 2012)

Although simulation based tools can significantly reduce the risk of overlooking process hazards that are consequences of complicated fault propagation paths, they have some gaps that can be summarized as follows:

- The limited application because of model specificity. In case of hazard identification of another unit, a new set of equations describing the new unit has to be formed and verified.
- The lack of mathematical models of modern hybrid systems and usually insufficient capability of the built-in solvers to investigate complex nonlinear process behaviour.
- HAZOP studies are typically conducted by multidisciplinary teams, including engineers, operators, and safety experts. Simulation tools may not be as user-friendly for non-technical personnel, potentially limiting collaboration and the effectiveness of the study.

3.3 Automated HAZOP Integration

A typical HAZOP provides an identification of accidental events and operability problems by using logical sequences of cause-deviation-consequence of process parameters. However, it doesn't lend itself to rank the hazards and effects of failures and to study the relative effectiveness of the proposed corrective actions. Therefore, several process hazard analysis techniques such as FTA, FMEA, SIL, LOPA, SIF, SIS,… etc. have been integrated with automated HAZOP methodology to assess and rank potential sources of hazards during the various stages of the process. For this integration, several software tools have been developed including HASILT (Cui et al., 2012), TORANAS (Mechhoud et al., 2016), PHAST (Bouafia et al., 2020) and ALBATROS III (Zenier and Antonello, 2023). This category comprises 9 articles accounting for 12% of the total (see Figure 4), distributed across the chemical industry (6 articles), petrochemical industry (2 articles), and one study in environmental engineering.

Examples of research using Integrated Automation Tools: (Mechhoud et al., 2016) developed an automated risk analysis and assessment (TORANAS) in operating petrochemical plants by combining the HAZOP and failure mode and effect analysis (FMEA) methods and assessing the consequences of the accident scenarios. FMEA is a structured method used to identify potential failures of a product or service and determine the failure frequency and impact. The TORANAS software was developed in the form of a graphical interface using Matlab as a coding tool to identify deviations and failure modes, and localization of their causes. They found that this software could heal to decrease human errors, assist the operator to make a good and safe decision and decrease the time utilization in hazard identification. (Zenier and Antonello, 2023) propoesd Albatros III software to minimize wasted time by processing Fault Trees and Minimal Cut Sets (MCS) on the basis of the HAZOP application. Albatros III used MS Excel worksheets for HAZOP writing and Fault Trees with MCS development and reporting. The application of the HAZOP methodology was done according to classical criteria, but using unique codes associated with primary events descriptors. The software allows the choice of predefined guide words for the HAZOP, which can also be integrated by the user, with the opportunity to further qualify the deviation. Table 4 summarizes the integrated automation sofware tools developed in HAZOP automation studies during the research period.

Table 4: Integrated Automation Tools with HAZOP study.

Software	Capabilities	Limitations
HASILT (Cui et al., 2012)	• Integrating HAZOP, LOPA, SRS and SIL to facilitate risk identification, reduce the costs and ensure the data consistency in chemical process industry. • Using case based reasoning (CBR) in developing a new SRS and recommending a SIS that could satisfy the SIL requirement.	• Requiring team to identify the cause and consequence of the deviation on the HAZOP worksheet. • The LOPA analyst has to manually move the IPLs from the Non-IPL Safeguards column to the column of description of IPL because the system cannot distinguish mong them.
TORANAS (Mechhoud et al., 2016)	• Combining HAZOP and FMEA methods to assess the consequences of the accident scenarios in petrochemical plants. • Decreasing the time utilization in hazard identification, human errors and assisting to make a good and safe decision.	• FMEA cannot discover complex failure modes involving multiple failures or subsystems, or to discover expected failure intervals of particular failure modes.
PHAST 7.2 (Bouafia et al., 2020)	• Using quantitative risk assessment based on HAZOP and bow tie to identify and quantify the sources of the unwanted scenarios. • Assesses a wide range of flammable and toxic hazards for process hazard analysis in petrochemical plants.	• Using another tool GRIF-Workshop to construct Bow tie and calculate the frequency of the unwanted events. • Concentrating on assessment of consequences affecting people in surrounding area more than equipment and the environment.
ALBATROS III (Zenier and Antonello, 2023)	• Integrating Fault Trees analysis, MCS and HAZOP application with considerable time saving in the chemical industry. • Verifying the completeness and consistency of the analysis, in cases of complex analysis with many causes / effects sequences involving multiple nodes. • Allowing to check repetitions or duplications or interruptions of the sequences of events.	• Using mandatory parameters (temperature, pressure, flow rate, level, composition) which limits the user discreation • Utilizing MS Excel worksheets for HAZOP writing which could pose challenges in large-scale plants with extensive data due to potential complexity. • Dividing HAZOP worksheet into parameters, causes, and effects and ignoring safeguards and recommendations.

After reviewing the advantages and disadvantages outlined in the table for various software tools that have been developed in HAZOP automation studies, integrating HAZOP method with different process hazard analysis tools, it is evident that while these tools offer numerous benefits, they also share some common challenges and potential issues to be addressed:

- Combining multiple risk assessment tools can increase the complexity of the analysis. This complexity can be challenging to manage, and it may require specialized expertise to ensure that the integrated process is robust and reliable.

- The integration of HAZOP with other risk assessment tools may require cross-training of personnel who are knowledgeable in various methodologies. This can be time-consuming and may result in skills gap. Additionally, this integration may require additional resources in terms of time, money, and personnel. This can be a barrier for organizations with limited resources.
- Different methodologies may yield conflicting conclusions or recommendations, which can be confusing and challenging to resolve.
- Assumptions and criteria of different methodologies may lead to discrepancies or errors in the analysis compatibility between these methodologies.

3.4 Artificial Intelligence Algorithms

Artificial Intelligence algorithms refer to models such as ontologies, bayesian network and other methods which describe the process variables and represent the relationships between them. Ontologies are formal models that use mathematical logic to clarify and define things and can enhance the sharing and exchange of the HAZOP results between computer systems. For example, when a HAZOP produces scenarios that can be reached from several deviations, an ontology-based tool can be used to check for consistency and identify missing parts of a given scenario (Batres et al., 2014). The ontologies category is found to include 7 articles (9%) (see Figure 4).

Examples of research using Ontologies: (Single et al., 2020a) designed an ontology-based model to generate HAZOP worksheets automatically. Ontologies were utilized for representing knowledge from HAZOP domain. The design of a knowledge model of the relevant concepts in the form of ontology was the first step within the conceptualization phase. This means that the concepts and the relationships between them must be identified and modeled carefully. The methodology was applied within a case study to a hexane storage tank as an equipment-based HAZOP analysis. Afterward, the automatically generated results were evaluated and compared to the original HAZOP results. The results showed that the developed ontology and the ontology-based reasoning algorithm are well-suited to generate equipment-specific HAZOP worksheets. (Wu et al., 2013) proposed a domain ontology called Scenario Object Model (SOM) to transfer, share, and computerize information, thereby facilitating the conduct of hazard analysis. They developed computer-aided HAZOP software called CAH, which requires the hazard team to establish the graphical SOM of the target system. This graphical SOM was evolved throughout the hazard analysis process, completing alongside the analysis. The limitations with this model is that it does not consider the process of using ontological concepts to express the dynamic hazard scenarios model or achieve computer-aided dynamic hazard analysis, and it requires experts to establish the graphical SOM of the target system. Table 5 summerizes the state of the art and limitations of ontology works for HAZOP automation.

In addition to ontologies, bayesian networks (BNs) are an ideal infrastructure to model cause-effect chains. The chains consisting of nodes representing random variables, connected by arcs (edges) representing causal relation can be branched and intertwined but the chain must be acyclic, as an effect cannot influence its own cause (Baybutt, 2015). The static Bayesian network can be extended to a dynamic Bayesian network (DBN) model by introducing relevant temporal dependencies that capture the dynamic behaviours of the domain variables between representations of the static network at different times (Hu et al., 2012). The Bayesian Networks category is found to include 5 articles (7%) (see Figure 4), distributed across various industries, with 2 articles in the chemical industry, one article in the process industry, one study in the mechanical industry and one article in environmental engineering. For example, dynamic bayesian network (DBN-HAZOP) model quantified hazard and operability analysis to provide prospective degradation trends of each component and the overall system for maintenance decision making in the mechanical industry. It was developed by integrating the prior knowledge of the interactions and dependencies among components and also the external environment, while the online condition monitoring data which was further to update the parameters of the model. Based on the future degradation trends given by DBN-HAZOP model, a local optimal proactive maintenance practice can be determined for each component by minimizing the expected maintenance cost per time unit. The proposed method was able to improve the accuracy and efficiency of safety management for multicomponent and multi-hazard complex system, providing adequate advised safety-related actions and predictive maintenance plans (Hu et al., 2012).

Table 5: Ontology-based works for HAZOP Automation study.

Method	Purpose of use	Limitations
SOM (Scenario Object Model) (Wu et al., 2013)	• Representing the contents and structures of hazard evaluation information. • Recording all the valid details of a brainstorming session that can be tracked, checked, and revised interactively. • Improving the effect of evaluation through the combination of various scenario-based hazard evaluation methods.	• Requiring experts to establish the graphical SOM of the target system. • Ignoring the dynamic hazard scenarios model and computer-aided dynamic analysis.
ARC (Accident Represe-ntation Chart) (Batres et al., 2014)	• Representing graphical elements of activities, events, physical objects as well as causality, temporal boundings, participation, containment, location, and connectivity. • Capturing descriptions of accidents and facilitating the reuse of past accident data	• Developing domain ontologies require the collaboration of experts that could create and maintain specific ontologies.
Ontology for computer-aided HAZOP systems (J. Single et al., 2019)	• Identifying hazards based on the knowledge structure on different layers of abstraction: (1) Substance, (2) Specific Unit, (3) Abstract Object, and (4) Hazard/Malfunction Propagation. • Assisting HAZOP conductors in performing hazard analysis while increasing the speed of safety assessments and serving as a decision support system.	• The ontologies must be extended and refined with further knowledge. • More complex systems must be conducted to validate this model.
Onto-HAZOP (Single et al., 2020b)	• Supporting participants based on answering competency questions using ontology queries. • Providing knowledge to fill HAZOP worksheets including risk assessment.	• Requiring integration with intelligent query system within a user-friendly application to enable non-experts to train in HAZOP studies.
Knowledge-based system (Single et al., 2020a)	• Using ontology to generate HAZOP worksheets automatically. • Automatically generated but meaningless scenarios can be avoided by modeling causal relationships between HAZOP concepts. • Within the knowledge modeling process, it is particularly important to define relationships between knowledge concepts unambiguously.	• If process units are modeled very roughly, wrong conclusions may be drawn. • In case of a too detailed process unit, the process unit modeling becomes too complicated and time consuming.
ACO-GRNN (Bai et al., 2021)	• Storing the risk propagation path and realizing the standardization of knowledge expression. • Improving accuracy compared with the traditional semantic similarity algorithm.	• Highly dependent on the richness of HAZOP ontology individuals.

Other AI methods such as natural-language information exploration environment for scanning thousands of documents in seconds (Garvin and Kimbleton, 2021), alarm causality corresponding to the deviation causality and associated alarm generation-skipping tracing method (Meng et al., 2021),…etc. have been applied to reduce hazards and optimize the deviation propagation relationship in the HAZOP reports. These methods are found to include 5 articles (7%). Although artificial intelligence algorithms (Ontologies and Dynamic Bayesian Network) are essential to represent complex correlations models and help reduce hazards, there are still gaps which can be summarized as follows:

- Ontologies are created based on human's personal experience and perspective at some point in time and thus can be biased or become outdated.
- Ontologies can be time-consuming and error-prone, making them more expensive than conventional HAZOP studies (J. I. Single et al., 2019).
- Equipment cannot always be accurately modeled, and ontologies require complicated design. To be used for HAZOP study automation, ontologies must be carefully established and documented (Daramola et al., 2013; J. I. Single et al., 2019).

- It is not feasible computationally to design a graphical network (Bayesian network) that is too large, dynamic, inhomogeneous, noisy and incomplete. Therefore, the modelling assumptions are often relaxed and specialised approximation and heuristic methods are widely used.

Table 6: Capabilities and limitations of intelligent HAZOP systems.

Method	Capabilities	Limitations
Expert Systems	• Produce similar situations consistently. • Inexpensive to operate, easy to reproduce and distribute. • Providing permanent documentation of the decision process. • Ability to obtain knowledge from several human experts, giving them more breadth and robustness than a single expert	• MFM focused on identifying the root causes of faulty scenarios. No consequences, and safeguards (Rossing et al., 2010). • D-higraph identified causal trees of deviations. No consequences, and safeguards are considered (Rodríguez and De la Mata, 2012). • Structural models focused on cause/ consequence derived from general heat and mass balance relations. No safeguards (Boonthum et al., 2014). • Stateflow tool did not include consequences derived from the causes as part of the HAZOP study output (Chia and Naraharisetti, 2023). • Lacking automatic knowledge recognition capabilities. • Inability to respond creatively to unusual situations. • Difficulty in adapting to changing environment. • The need for specialists is necessary (de la O Herrera et al., 2018).
Simulation tools	• Describing process dynamic behavior in case of failures. • The integration with HAZOP can facilitate scenario identification and risk assessment. • This integration mitigates the subjectivity associated with determining the severity and likelihood of hazardous events.	• HAZOP studies are typically conducted by multidisciplinary teams. • Simulation tools may not be as user-friendly for non-technical personnel, potentially limiting collaboration and the effectiveness of the study. • In case of hazard identification of another unit, a new set of equations describing the new unit has to be formed and verified (Janošovský et al., 2017).
Integration based tools	• Assessing and ranking potential sources of hazards during the various stages of the process. • Decreasing the time utilization in hazard identification, human errors and assisting to make a good and safe decision.	• Combining multiple risk assessment tools can increase the complexity of the analysis. • The integration may require additional resources in terms of time, money, and personnel. • Different tools may yield conflicting conclusions or recommendations. • Assumptions and criteria of different tools may lead to discrepancies or errors in the analysis.
Ontologies	• A precise specification of domain model based on formal semantics (Grimm et al., 2007). • Ability to represent knowledge unambiguously.	• Time-consuming and error-prone, making them more expensive than conventional HAZOP studies (J. I. Single et al., 2019). • Developing ontologies requires significant effort and expertise to formalize domain-specific knowledge. • Ontologies depend on the availability of a large dataset or HAZOP study reports (Feng et al., 2021).
Bayesian Networks	• Modeling complex process systems based on conditional dependencies and subjective probabilities (Ale et al., 2014). • Managing information from different sources, such as expert judgment, observable information or experience, common causes, and human influence (Villa et al., 2016).	• Building BNs requires expert knowledge to specify the network structure and define prior probabilities, which can be subjective and difficult to elicit. • Treating the failure data as precise and clearly defined numbers which is usually not the case (Zhang et al., 2016).

Building on the aforementioned literature, knowledge-based approaches including expert systems, computer simulation software, integrated automation tools, ontologies, and Bayesian networks, are essential for supporting hazard identification and analysis. These approaches help represent complex correlations models and help reduce hazards. However, acquiring, eliciting, and updating domain-specific knowledge can be challenging, subjective, and time-consuming, requiring expert knowledge, experience, and iterative refinement. Furthermore, existing HAZOP tools lack a comprehensive knowledge base that includes guidewords, parameters, deviations, causes, consequences, safeguards, and recommendations, rendering them inadequate for the more complex plants. Table 6 presents the capabilities and limitations of different knowledge-based approaches used in HAZOP study within the research period. In recent years, data driven models (illustrated in detail in the next section), including machine learning technology, deep learning, and natural language processing (NLP), have been employed to determine the causes of the accidents, and describe the accident occurrence processes qualitatively, which helps avoid the recurrence of accidents or minimize the severity of the accident consequences. Therefore, it is crucial to develop an AI HAZOP tool that uses historical data to systematically build a generic knowledge base applicable to most, if not all, process design plants. This would improve efficiency, reduce the time required to implement HAZOP study and alleviate human bias and dependencies .

3.5 Data Mining Models

Data-driven methods use raw or pre-processed data to train models to draw conclusions and predict system behavior. Historical data, or documents can be used as an information basis. Data Mining methods have been applied to digitalize HAZOP methodology to explore accident cause mechanisms and help promote intelligent analysis. Data Mining is concerned with the automatic or semi-automatic exploration and analysis of large volumes of data, to discover meaningful patterns or rules. During the research period, data mining approaches, including natural language processing, machine learning classifiers and deep learning have been used to improve the accuracy of HAZOP analysis results. The data mining category encompasses 10 articles (13%), with 3 focusing on the chemical industry, 5 on the petrochemical industry, one on environmental engineering and one centered on safety engineering.

3.5.1 Natural Language Processing

Natural Language Processing (NLP), similar to brainstorming, prepares the situation using expert engineers' experience. NLP makes HAZOP study reports readable for machine learning (ML) algorithms. Since ML algorithms cannot directly process text, conversion into numerical form is necessary (Aggarwal, 2018). NLP can process HAZOP study reports, which are text containing experiences from previous HAZOP studies. NLP encompasses conventional strategies, advanced strategies and topic modeling. Advanced text modeling strategies in NLP such as Word2Vec which employs neural networks to assign vectors to words (Mikolov et al., 2013), Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2018), embeddings from language models (ELMo), and Generative pre-trained transformers (GPT) are ideal for analyzing large data corpus. In contrast, conventional strategies such as Bag of Words (BOW) and Term Frequency-inverse Document Frequency (TF-IDF) are well suited to small datasets. Different topic modeling methods exist, including Latent Dirichlet Allocation (LDA), the best-known and most widely used algorithm in large-scale data clustering because of its high efficiency, speed, and influential clustering algorithm technology (Suh, 2021) and recently subsequently, Hierarchical Density-based Spatial Clustering of Applications with Noise (HDBSCAN), have been used for clustering sentences and topic modeling (Grootendorst, 2022).

Examples of research using NLP: (Feng et al., 2021) trained BERT as a NLP model on Chinese texts and used Chinese HAZOP high quality reports to help smaller chemical plants perform security analysis. This study used consequence descriptions as features for classifying consequence severity in HAZOP analysis, utilizing a dataset consisting of 2075 data points. They achieved better results with 88.6%, 88.2% and 88.2% precision, recall and F1-Score respectively compared to Word2Vec + BiLSTM +Attention model which achieved 85.8%, 84.8% and 85.1% (Mikolov et al., 2013). The results would only be useful for Chinese. However, the proposed approach is independent of language and can be applied to HAZOP reports in English or other languages by changing the language model BERT and the language of the training corpus. (Ekramipooya et al., 2023) used BOW to convert causes into feature vectors with a small dataset (42 datapoints) for training the system. This method does not fit to large dataset and complex processes. Also, they did not consider consequences and safeguards and used only causes as features (input) to predict deviations (target) because there is more diversity in the consequences, safeguards, and recommendations than in deviations, so they should first be categorized in terms of their scope. (Wang and Gu, 2022) proposed an intelligent HAZOP analysis method based on data mining to explore the law of accident cause mechanisms, help promote intelligent analysis, and improve the accuracy of analysis results. They used LDA method to identify hidden cause and consequence topics from the HAZOP analysis data and concluded that LDA clustering models can explore potential causes and consequences.

Observing the previous NLP methods used in HAZOP automation analysis, the BERT model (Feng et al., 2021) requires validation and it is difficult to be applied in complex processes due to the unavailability of high-quality reports. Furthermore, They classified the consequences on severity aspect and did not consider the possibility and risk aspects (Z. Wang et al., 2023a). On the other hand, the bag of words (BOW) model (Ekramipooya et al., 2023) did not consider consequences, safeguards and recommendations and used only causes as features (input) to predict deviations (target). Additionally, BOW overlooks the contextual meaning of words, resulting in a loss of semantic information. It treats each word in isolation, assigning equal importance to all terms. Consequently, the resulting feature vectors can be sparse and high-dimensional, posing challenges in terms of computational efficiency and model interpretability. Moreover, BOW fails to account for the sequential order of words in sentences or text. The limitations and advantages of NLP strategies used in automated HAZOP studies are outlined in Table 7.

Table 7: advantages and disadvantages of NLP strategies used in HAZOP Automation study.

Method	Advantages	Drawbacks
BOW (Ekramipooya et al., 2023)	• Simple sentence embedding technique that applies simple formalisms. • Extracting features from the text based on the count of words in a document or sentence (Qader et al., 2019).	• Neglecting the contextual meaning of words, resulting in a loss of semantic information. • Ignoring the grammar and order of the words (Kim et al., 2017).
TF-IDF (Wang and Gu, 2022)	• Processing the words with high frequencies to obtain the word vectors, contributing to the clustering analysis and risk prediction. • Measures the rarity of the words in the text that may hold significant information (Kim et al., 2019).	• Creating sparse vector representations in large document collections with a wide vocabulary leading to computational inefficiencies. • Ignoring semantic similarities between words.
Word2Vec (Mikolov et al., 2013)	• Capturing semantic similarities and relationships between words. • Algorithms such as CBOW and Skipgram are efficient and scalable to large data corpus. • Having the characteristics of low computing cost and high accuracy (Zhao et al., 2022).	• This method could not solve the problem of polysemy or contextual variation, nor could it change according to the context of the word.
BERT (Feng et al., 2021)	• Using a transformer to extract text features which improves the ability to extract semantic features (Tang et al., 2018). • Unlike GPT, BERT is bidirectional which considers both preceding and succeeding words when analyzing a given word or phrase, allowing it to capture contextual nuances more effectively.	• A complicated formalism and dfficult to be applied in complex processes due to the unavailability of high-quality reports. • Requiring large amounts of annotated data to achieve optimal performance.
ELMo (Peng et al., 2021)	• Solving the problem of polysemous words and also suitable for entity recognition of small sample data. • Inferring the word vector of each word through the multi-layer bidirectional LSTM structure. • learning a lot of linguistic knowledge as it is trained on a large dataset.	• Time-consuming in large- datasets compared to BERT as it is based on a character-level convolutional neural network and bidirectional long short term memory (LSTM) architecture. • LSTM struggles with capturing information in very long sentences.

3.5.2 Machine Learning Classifiers

Machine learning (ML) classifiers offer a powerful approach for reasoning and achieving high accuracy predictions of hazards, causes, and consequences. One of the major advantages of ML algorithms compared to other reasoning techniques lies in their diversity, making them suitable for a wide range of applications. Several ML classifiers such as Decision Tree (DT), linear Support Vector (SVM) Machine, Random Forest (RF), Logistic Regression (LR), Gaussian Naïve Bays (GNB), and K-Nearest Neighbors (K-NN) have been used in HAZOP study for reasoning and high accuracy predictions.

Examples of research using ML classifiers: (Ekramipooya et al., 2023) used ML classifiers such as Decision Tree (DT), linear Support Vector (SVM) Machine, Random Forest (RF), Logistic Regression (LR), Gaussian Naïve Bays (GNB), and K-Nearest Neighbors (K-NN) to predict deviations using causes as features. Decision Tree outperformed other classifiers with 92% accuracy. The drawback in this study is that the HAZOP study automation's performance depends on the input dataset's quality and reliability. Additionaly, increasing the number of labels will reduce this study's quality (if the dataset's size remains constant). The most accurate results will be acquired when a data set with the least number of labels is available. Also, the performance of this method depends on the length of the sentence since this approach uses sentence embedding. Increasing the length of the sentence may change the performance of the classification. (Wang and Gu, 2022) used the naive Bayes algorithm, a method for establishing a risk prediction model trained to calculate the likelihoods, severities, and risk levels. The analysis incorporated parameters, guidewords, causes and consequences as features to predict likelihood, severity, and risk levels, utilizing a dataset comprising 5503 records. The prediction accuracy was above 80% which can verify the causation of the accident and realize that the intelligent prediction of the risk can help to solve the mutual independence between HAZOP reports, and comprehensively identify accident problems. The accuracy of this approach is low compared to other methods such as decision tree.

3.5.3 Deep Learning

Despite efforts to enhance HAZOP through the development of automatic hazard and operability expert systems, including computer-aided or computer-automated analysis methods, there are still many shortcomings in the communication and storage of analysis results. The computer-aided method lacks a unified form of communication, struggles to share data between different software types, and lacks automatic knowledge recognition capabilities. Consequently, it becomes challenging to reuse and share information across various software types or even within the same software across different projects, teams, or stages. Faced with a large number of unstructured data and information in the form of paper documents or electronic documents generated by HAZOP analysis, there is an urgent need for automation technologies such as deep learning are to help people quickly find the information they need in mass information sources.

Examples of research using deep learning: (Peng et al., 2021) introduced a deep learning framework for Chinese HAZOP documents to perform a named entity recognition (NER) task to identify the entities of material and equipment s in the HAZOP document of a coal indirect liquefaction project in the petrochemical industry. The preprocessed data are input into ELMo and a double convolution neural network (DCNN) model to capture rich character features. Additionally, a bidirectional long short-term memory (BiLSTM) network was employed to extract long-distance semantic information. Ultimately, the results were decoded by a conditional random field (CRF), and then output. The findings from this model showcased that the accuracy reached 90.83%, the recall rate reached 92.46%, and the F-value reached 91.76%. (Z. Wang et al., 2023b) presented a deep learning and multifractal analysis (DLF) model to classify hazard events (HaE). They used BERT to vectorize HaE, generating HaE time series. Subsequently, a new multifractal analysis method termed variant of multifractal detrended fluctuation analysis (HmF- DFA) was developed to win HaE fractal series by analyzing HaE time series. Finally, they designed a new hierarchical gating neural network (HGNN) to process HaE fractal series to classify HaE based on severity, possibility, and risk. The precision, recall, and F1-Score for the severity aspect exceeded 80%, whereas those for the possibility and risk aspects were below 80%. The results indicate that DLF's performance is relatively inferior in classifying hazard events in HAZOP reports. (He et al., 2020) utilized generative adversarial network (GAN)-based semi-supervised learning method for real-time risk warning of process industries. They applied fuzzy HAZOP to estimate the risk of systems quantitatively based on the process data. Based on deep convolutional network structure and adversarial learning algorithm, a deep neural network of GAN for the accident risk warning is constructed to improve the risk prediction performance and overcome the prob- lem of insufficient data with labels for risk prediction. The results indicated that this model is favorable in predicting risk for process industries. Table 8 illustrates the deep learning models that have been developed in automated HAZOP analysis within the research period .

Table 8: Summary of works on deep learning in HAZOP Automation studies.

Method	Capabilities	Limitations
Fuzzy HAZOP-DCGAN-CNN (He et al., 2020)	• Constructing real-time warning system by using GAN to address the problem of the scarce labaled data for real time risk assessment. • Applying numerous unlabeled process samples from distributed control systems (DCSs) to improve the model performance.	• Focusing on predicting frequency, consequence impact, and risk level and ignoring automating HAZOP elements, including deviations, causes, consequenses, safeguards, and recommendations. • GANs typically require large datasets to produce high-quality outputs.
ELMo-DCNN-BiLSTM-CRF (Peng et al., 2021)	• Identifying the entities in textual data of a chinese HAZOP document and classifying them to categories, such as people, places, equipment and materials. • Improving entity recognition difficulty caused by plenty of polysemous words in petrochemical HAZOP text. • Making the reuse and sharing of Chinese HAZOP document information more convenient and automatic and contributes to information standardization.	• The recognition accuracy of Chinese NER is not as good as English as the character vectors were used instead of word vectors as input for training and the character sequence cannot express enough semantic nformation. • Combining multiple neural networks can increase the computational resources required for training and inference. • The interpretability of this complex model may be reduced, making it harder to diagnose errors or biases.
BERT-BiLSTM-Attention (Feng et al., 2021)	• Classifying the severity of consequence events of superior HAZOP reports to ensure consistency. • Providing security knowledge references for small chemical companies that lack security knowledge and security experts.	• Failing to consider the possibility aspect and risk aspect, only the severity aspect. • The research object is not the hazard event, but the consequence in hazard event, which weaken the power of HAZOP in redevelopment.
DLF (Z. Wang et al., 2023a)	• Exploring HaE classification from three aspect: severity, possibility, and risk. • Promoting the experts to conduct safety analysis and explore the fractal properties hidden in various time series and quasi-time series.	• Some hazard events may be incomplete due to the quality of technical reports, and DLF model may be less robust. • Suitable for hazards with certain time series attributes, but it may be powerless for ordinary hazard. • The evaluation results for possibility and risk aspects are very low.
HAINEX (IBERT, BiLSTM, CRF&IL) (Z. Wang et al., 2022)	• Extracting the industrial safety knowledge (ISK) in HAZOP reports based on data layer. • Improving HAZOP execution efficiency and the security of the system.	• Some entities and relationships were redundan due to the complexity of the process. • Ignoring the problem of ambiguity which the literal meaning of words may be very close.
BI-GAT-CRF (Zhao et al., 2022)	• Determining entity boundaries and recognize named entity categories in Chinese text. • Solving the vocabulary ambiguity and improves the accuracy of named entity recognition tasks. • Reducing labor and time costs and improving the completeness of the analysis results.	• Based on character vectors of the HAZOP chinese text which cannot express enough semantic nformation.

While deep learning methods have the potential to enhance automated HAZOP studies, they have significant disadvantages and challenges. Deep learning models typically require large datasets for training. Moreover, deep learning models are sensitive to the quality and representativeness of the data they are trained on. Inaccurate or biased data can lead to incorrect hazard identifications and operability assessments.

The literature on NLP methods, ML classifiers and deep learning models reveals a variety of input data and results influenced by factors such as feature definition, feature text length, target variables and the number of datapoints utilized in the analysis. For example, (Feng et al., 2021) used consequence descriptions and 2075 samples in the BERT+BiLSTM+Attention model to Classify the severity of consequence events. (Wang and Gu, 2022) incorporated parameters, guidewords, causes and consequences as features to predict likelihood, severity, and risk levels, utilizing a dataset comprising 5503 records in the TF-IDF + LDA + GNB model. Process node, deviation, consequences, causes, measures, and suggestions were incorporated in the BERT+HmF-DFA+HGNN model to predict possibility, severity, and risk level, utilizing 5869 hazard events (Z. Wang et al., 2023a). It is noteworthy that these studies often overlooked the full automation of HAZOP report elements, including node descriptions, node components, guidewords, parameters, deviations, causes, consequences, safeguards, and recommendations, focusing only on automating specific aspects of HAZOP reports. Therefore, further research is imperative to advance the automation of HAZOP study, aiming to identify the most suitable NLP strategies and classifiers tailored to specific industrial sectors, as well as to automate all elements of HAZOP reports.

On the other hand, data mining models, which are based on extracted data from HAZOP reports, lack crucial technical details embedded in design drawings or P&IDs. These details include the identification of major and minor components within each node, the intricate connections between components, and the spatial arrangement of entities, including their position, orientation, and proximity to understand physical layout. Integrating these elements holds the potential to significantly advance the comprehensiveness and effectiveness of automated HAZOP study for new projects. This initiative shows potential to enhance the efficiency, accuracy, and applicability of automated HAZOP methodology, facilitating informed decision-making for industry professionals involved in hazard identification and process safety across varios industries.

4. HAZOP SCOPE EXPANSION AND APPLICATIONS

The research on HAZOP has demonstrated significant growth, showing an exponential increase over the years and extending its scope to various sectors. A limited number of scholars (8 papers) have applied HAZOP analysis within various domains, comprising 6 articles utilizing the conventional HAZOP method and 2 articles integrating HAZOP with other process hazard analysis tools for further analysis. In our review, we observed that the conventional HAZOP method has been employed for hazard identification in diverse areas including railway systems, supply chain management, asset management of urban drainage systems, demolition of large bulk materials, and hoisting operations of prefabricated buildings, while researchers have employed an integrated approach, combining HAZOP with other Process Hazard Analysis (PHA) tools, notably in the domain of railway systems, to enable more comprehensive analyses. In our exploration of HAZOP automation, we have not encountered any publications utilizing automated HAZOP within these areas.

Examples of various domains applying HAZOP methodology: (Zhu et al., 2022) used HAZOP to analyze construction workers' unsafe behaviours in hoisting operations of prefabricated buildings to mitigate unsafe behaviours. A questionnaire survey was first performed to gather information on safety risks and influencing factors during each stage of hoisting operations, and the SPSS software was used to analyze the reliability of the survey results. Next, HAZOP was applied to identify the deviation and change of the unsafe behaviors as well as their causes, consequences, and countermeasures. Finally, a case study was presented (using wristband sensor, surveillance cameras and drone footage) to verify the effectiveness of the countermeasures through a comparison and evaluation method from experimental economics. This study therefore provided an innovative method and a theoretical foundation for reducing unsafe behaviours in hoisting operations of prefabricated buildings. HAZOP method was performed on the Chinese Train Control System3 (CTCS3) to guarantee the operational safety of the high-speed railway system. The CTCS-3 on board system is responsible for receiving the data and command information from the track side then calculates the speed profile and safe operation of the moving train. Four models (reference model, functional hierarchical model, state diagram, and sequence diagram) were developed to describe the system design intent and operational situation and used to provide a basis for the hazard identification process. During the examination session, any deviations between the design intent and the operational situations were identified by experts based on the systems model. The causes of deviations and consequences were also determined to help mitigate the risk and propose safety actions and measures (Li et al., 2015). This HAZOP study in the CTCS-3 onboard system had successfully identified high-accuracy hazards with respect to the increasing operational speed and expanding railway of the Chinese high-speed railway system. Moreover, HAZOP method was presented to the area of supply chain management (SCM), more specifically to supply chain risk management

(SCRM), with objective on identification of risks in organizations modelled with the Collaborative Planning, Forecasting and Replenishment (CPFR) model. The research implication showed that the method like HAZOP, designed to identify hazards in other fields (i.e., chemical engineering) can be successfully used in SCRM (Mitkowski and Zenka-Podlaszewska, 2014). Table 9 summarizes the application of HAZOP methodology in the construction domain subareas throughout the research timeframe.

Table 9: Application of HAZOP Methodology in the Construction Domain Subareas.

Field	Purpose of use	References
Urban drainage systems	• Assessing sewer failure mechanisms and defects responsible for the structural/operational failures of sewer elements to identify information needed for sewer asset management.	(Stanić et al., 2014)
CPFR supply chain management	• Identification of risks in organizations modelled with the CPFR model. • Promoting the experts to conduct safety analysis and treat unforeseen hazards.	(Mitkowski and Zenka-Podlaszewska, 2014)
Demolition of large bulk materials	• Conducting a safety risk assessment prior to the dismantling of large bulk materials handling machines depending on the project's work method statements, construction regulations, as well as existing standards such as AS2601: The demolition of structures, to develop guide words.	(Joubert et al., 2021)
Hoisting Operations of Prefabricated Buildings	• Analyzing construction workers' unsafe behaviors to mitigate unsafe behaviors in hoisting operations through survey questionnaires of unsafe behaviors of hoisting workers and results verification by using a comparison and evaluation method from experimental economics through a case study experiment.	(Zhu et al., 2022)
Railway systems	• Using human factor HF-HAZOP to analyze the technical aspects, the procedural issues and human factor elements of rail systems at the worksite utilising an expert team.	(Noorudheen et al., 2013)
	• Identifying the hazards of a CTCS-3 onboard system during examination sessions considering the basis of the functions of the onboard system and a scenario of temporary speed restriction.	(Li et al., 2015)
	• Integration of HAZOP, risk graph and safety integrity level/tolerable hazard rate (SIL/THR) for hazard identification, impact analysis, and selection of the risk acceptance principle of real signalling equipment in railway control systems.	(Szmel et al., 2019)
	• HAZOP, dynamic fault tree (DFT) and Monte Carlo Simulation (MCS) were used for quantitative and qualitative analysis of hazard sources identification and reliability sensitivity analysis of CTCS3 on-board subsystem.	(Shi and Chen, 2020)

While the application of HAZOP in various sub-areas within the industry has proven effective in hazard identification process, these applications used conventional method which is time consuming and susceptible to human errors. Additionaly, the integration of HAZOP with other PHA techniques introduces challenges encompassing the complexity of the analysis which may require specialized expertise to ensure that the integrated process is robust and reliable, using different methods which may yield conflicting conclusions or recommendations and using assumptions and criteria of different methodologies which may lead to discrepancies or errors in the compatibility analysis between these methodologies. Therefore, there is an urgent requirement to develop intelligent and automated HAZOP models designed for these domains to alleviate these challenges and to reduce conventional time consumption and human dependency.

5. OPPORTUNITIES AND FUTURE DIRECTIONS

Over the years, HAZOP has been applied to various industrial sectors, utilizing diffrent levels of automation to facilitate HAZOP applications. However, the method still needs research to address the gaps highlighted earlier in this paper. Future research in the following fields is expected to improve and expand HAZOP applications.

Expansion to other industries: Whilst automated HAZOP has traditionally been used in the chemical engineering, oil and gas and petrochemical industries, its application is expanding to other sectors such as environmental engineering and other domains, including railway systems, supply chain management, urban drainage systems, and hoisting operations. The review showed that these domains applied HAZOP methodology in terms of conventional and integrated methods (8 papers) and just three papers in environmental engineering used automated approach. Therefore, these areas have the potential to develop automated approaches to improve efficiency, accuracy and safety. Furtheremore, HAZOP method can be used to analyze the hazards associated with construction tasks such as but not limited to excavation and trenching, dewatering systems, demolition and dismantling, MEP works, utilities installation,.etc. Automating HAZOP of these tasks can bring several benefits, including increased efficiency, improved safety, and more consistent hazard identification and mitigation.

Development of AI HAZOP tool: The existing AI HAZOP tools lack a comprehensive knowledge base for detecting hazards and operational malfunctions. Therefore, it is crucial to figure out an AI tool to systematically build up a generic knowledge base that could be applied to most, if not all process design plants in general.

Integration with Digital Technologies: The integration of HAZOP with digital technologies can enhance the accuracy and efficiency of hazard identification. Digital twins and advanced analytics can help in real-time monitoring and prediction of potential hazards. HAZOP can also be integrated with BIM technologies to analyze and visualize potential hazards and issues in 3D models of construction projects, improving the clarity and understanding of risks.

Comprehensive Automation of HAZOP Studies: To enhance efficiency, it is crucial to consider automating all components of HAZOP reports, including node name and description, main and minor equipment, guide words, parameters, causes, deviations, safeguards, and recommendations, ensuring a thorough automated HAZOP process.

Integration of P&IDs and HAZOP reports: Exploring the integration of data extracted from both P&IDs and HAZOP reports using data mining techniques can enhance hazard identification and mitigation for new projects. This integrated approach has the potential to significantly improve the comprehensiveness and effectiveness of automated HAZOP study.

Comparative Analysis of Data Mining models: Conducting a comparative analysis of data mining models to assess their efficacy in automating HAZOP reports and determining the most suitable model for specific applications. This analysis aims to expedite hazard identification and enhance safety levels in different industerial sectors.

Utilization of Natural Language Processing: NLP techniques can be used to analyze textual data, including safety reports, incident logs, and documentation. NLP makes HAZOP study reports readable for machine learning (ML) algorithms. NLP can process HAZOP study reports, which are text containing experiences from previous projects. The existing literature on NLP showcases divergent input data and results. Consequently, further research is warranted to ascertain the most suitable NLP model tailored to specific applications in industries. Topic modeling methods such as LDA and HDBSCAN can be used for clustering sentences and categorizing consequences, safeguards, and recommendations.

Expermintation with Machine learning classifiers for risk assessment: Machine learning algorithms can assist in analyzing data, identifying patterns, and predicting potential hazards. Machine learning can help in proactive risk assessment and decision-making during the HAZOP process. A significant advantage of ML techniques is their diversity that makes them suitable for various applications. Decision Tree, Support Vector Machine, Random Forest, Logistic Regression, Gaussian Naïve Bays, and K-Nearest Neighbours can be applied to predict likelihood, severity and risk level aspects.

6. CONCLUSIONS

The traditional HAZOP methodology cannot meet the requirements of process risk assessment due to increasingly complex processes using the recent technology and stringent safety analysis requirements. To address these limitations, this paper intially developed a framework for the HAZOP structure, illustraing the transition from conventional to automation procedures. The framework was then used to categorize automated HAZOP method into knowledge-based approaches (expert systems, computer simulation tools, integrated automation tools, artificial intelligence algorithms) and data-driven models. A total of 83 journal papers were extracted from the Scopus database and used as the basis of the analysis, exploring the main research contributions and trends. The research gaps were then identified accordingly, and potential future research directions were suggested. The results show that the automated HAZOP has been expanded to other industrial sectors, such as environmental engineering, but no applications of HAZOP automation have been observed in domains such as railway systems, supply chain management, urban drainage systems, and hoisting operations. Additionally, relatively few studies have applied conventional and integrated HAZOP to these domains. Therefore, these areas have a significant potential to utilize automated HAZOP to its tasks for identifying potential hazards and increase projects safety levels. Furtheremore, existing HAZOP tools lack a comprehensive knowledge base for detecting hazards and operational malfunctions, proposing that mining historical data could develop this knowledge base. On the other hand, data mining algorithms, encompassing NLP, ML classifiers and deep learning models have not received sufficient attention from researchers. The literature studies exhibit various techniques and models with diverse input data and results. Consequently, further research is required to identify the most suitable method for specific applications within the HAZOP automation process. Additionally, current data mining models, which are based on extracted data from HAZOP reports, lack crucial technical details embedded in design drawings or P&IDs. Integrating these details holds the potential to significantly advance the comprehensiveness and effectiveness of automated HAZOP study for new projects. This research aims to assist in identifying patterns, predicting potential hazards, and facilitating proactive risk assessment and decision-making.

REFERENCES

Aggarwal, C.C., 2018. Machine Learning for Text. Springer, IBM T. J. Watson Research Center, Yorktown Heights, NY, USA, https://doi.org/10.1007/978-3-319-73531-3.

Ale, B., van Gulijk, C., Hanea, A., Hanea, D., Hudson, P., Lin, P.H., Sillem, S., 2014. Towards BBN based risk modelling of process plants. Saf Sci 69, 48–56. https://doi.org/10.1016/j.ssci.2013.12.007

AspenTech., 2023. Aspen Technology Engineering Products - Aspen Plus, Aspen Dynamics, and Aspen HYSYS. https://esupport.aspentech.com/ [WWW Document].

Aziz, A., Ahmed, S., Khan, F.I., 2019. An ontology-based methodology for hazard identification and causation analysis. Process Safety and Environmental Protection 123, 87–98. https://doi.org/10.1016/j.psep.2018.12.008

Bai, Y., Gao, D., Peng, L., 2021. Hazop ontology semantic similarity algorithm based on aco-grnn. Processes 9. https://doi.org/10.3390/pr9122115

Bassiliades, N., Governatori, G., Paschke, A., 2011. Rule-Based Reasoning. Programming, and Applications, vol. 6826, Springer Berlin Heidelberg, Berlin, Heidelberg. https:// doi.org/10.1007/978-3-642-22546-8. .

Batres, R., Fujihara, S., Shimada, Y., Fuchino, T., 2014. The use of ontologies for enhancing the use of accident information. Process Safety and Environmental Protection 92, 119–130. https://doi.org/10.1016/j.psep.2012.11.002

Baybutt, P., 2016. Design intent for hazard and operability studies. Process Safety Progress 35, 36–40. https://doi.org/10.1002/prs.11718

Baybutt, P., 2015. Competency requirements for process hazard analysis (PHA) teams. J Loss Prev Process Ind 33, 151–158. https://doi.org/10.1016/j.jlp.2014.11.023

Berdouzi, F., Villemur, C., Olivier-Maget, N., Gabas, N., 2018. Dynamic simulation for risk analysis: Application to an exothermic reaction. Process Safety and Environmental Protection 113, 149–163. https://doi.org/10.1016/j.psep.2017.09.019

Boonthum, N., Mulalee, U., Srinophakun, T., 2014. A systematic formulation for HAZOP analysis based on structural model. Reliab Eng Syst Saf 121, 152–163. https://doi.org/10.1016/j.ress.2013.08.008

Bouafia, A., Bougofa, M., Rouainia, M., Medjram, M.S., 2020. Safety Risk Analysis and Accidents Modeling of a Major Gasoline Release in Petrochemical Plant. Journal of Failure Analysis and Prevention 20, 358–369. https://doi.org/10.1007/s11668-020-00826-9

BSI (British Standards Institute)., 2001. Hazard and operability studies (HAZOP studies)—Application guide. BS IEC 61882:2001. , London.

Cameron, I., Mannan, S., Németh, E., Park, S., Pasman, H., Rogers, W., Seligmann, B., 2017. Process hazard analysis, hazard identification and scenario definition: Are the conventional tools sufficient, or should and can we do much better? Process Safety and Environmental Protection 110, 53–70. https://doi.org/10.1016/j.psep.2017.01.025

Chia, M.F., Naraharisetti, P.K., 2023. HAZOP using Stateflow software: Methodology and case study. Process Safety and Environmental Protection 179, 137–156. https://doi.org/10.1016/j.psep.2023.09.005

Crawley, F., Tyler, B., 2015. HAZOP: Guide to best practice guidelines to best practice for the process and chemical industries. Elsevier, Amsterdam, Netherlands.

Cui, L., Shu, Y., Wang, Z., Zhao, J., Qiu, T., Sun, W., Wei, Z., 2012. HASILT: An intelligent software platform for HAZOP, LOPA, SRS and SIL verification. Reliab Eng Syst Saf 108, 56–64. https://doi.org/10.1016/j.ress.2012.06.014

Cui, L., Zhao, J., Zhang, R., 2010. The integration of HAZOP expert system and piping and instrumentation diagrams. Process Safety and Environmental Protection 88, 327–334. https://doi.org/10.1016/j.psep.2010.04.002

Danko, M., Frutiger, J., Jelemenský, Ľ., Sin, G., 2017. Monte Carlo Based Framework to Support HAZOP Study. Computer Aided Chemical Engineering. 40, 2233–2238.

Daramola, O., Stålhane, T., Omoronyia, I., Sindre, G., 2013. Using ontologies and machine learning for hazard identification and safety analysis. . Managing requirements knowledge. Springer 117–141.

de la O Herrera, M.A., Luna, A.S., da Costa, A.C.A., Lemes, E.M.B., 2018. Risk Analysis: A generalized Hazop methodology state-of-the-art, applications, and perspective in the process industry. Vigilância Sanitária em Debate 6, 106. https://doi.org/10.22239/2317-269x.00990

Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference 1, 4171–4186.

Dunjó, J., Fthenakis, V., Vílchez, J.A., Arnaldos, J., 2010. Hazard and operability (HAZOP) analysis. A literature review. J Hazard Mater. https://doi.org/10.1016/j.jhazmat.2009.08.076

Ekramipooya, A., Boroushaki, M., Rashtchian, D., 2023. Application of natural language processing and machine learning in prediction of deviations in the HAZOP study worksheet: A comparison of classifiers. Process Safety and Environmental Protection 176, 65–73. https://doi.org/10.1016/j.psep.2023.06.004

Emami, M., Hejazi, B., Karimi, M., Mousavi, S.A., 2022. Quantitative risk assessment and risk reduction of integrated acid gas enrichment and amine regeneration process using Aspen Plus dynamic simulation. Results in Engineering 15. https://doi.org/10.1016/j.rineng.2022.100566

Enemark-Rasmussen, R., Cameron, D., Angelo, P.B., Sin, G., 2012. A simulation based engineering method to support HAZOP studies. pp. 1271–1275. https://doi.org/10.1016/B978-0-444-59506-5.50085-7

Feng, X., Dai, Y., Ji, X., Zhou, L., Dang, Y., 2021. Application of natural language processing in HAZOP reports. Process Safety and Environmental Protection 155, 41–48. https://doi.org/10.1016/j.psep.2021.09.001

Folger, R., Stein, C., 2017. Abduction 101: Reasoning processes to aid discovery. Human Resource Management Review 27, 306–315. https://doi.org/10.1016/j.hrmr.2016.08.007

Gao, P., Li, W., 2022. Integration of HAZOP and Bayesian network in city gas explosion emergency response processes. Emergency Management Science and Technology 2, 1–9. https://doi.org/10.48130/emst-2022-0019

Gao, P., Li, W., Sun, Y., Liu, S., 2022. Risk assessment for gas transmission station based on cloud model based multilevel Bayesian network from the perspective of multi-flow intersecting theory. Process Safety and Environmental Protection 159, 887–898. https://doi.org/10.1016/j.psep.2022.01.036

Garvin, T., Kimbleton, S., 2021. Artificial intelligence as ally in hazard analysis. Process Safety Progress 40, 43–49. https://doi.org/10.1002/prs.12243

Grimm, S., Hitzler, P., Abecker, A., 2007. Knowledge Representation and Ontologies. in Studer, semantic web services: concepts, technologies and applications, Springer Berlin, Heidelberg, Berlin, Heidelberg.

Grootendorst, M., 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure.

He, R., Li, X., Chen, Guoming, Chen, Guoxing, Liu, Y., 2020. Generative adversarial network-based semi-supervised learning for real-time risk warning of process industries. Expert Syst Appl 150. https://doi.org/10.1016/j.eswa.2020.113244

Heino, P., 1999. Fluid property reasoning in knowledge-based hazard identification . Espoo: Technical Research Centre of Finland.

Hu, J., Zhang, L., Liang, W., 2012. Opportunistic predictive maintenance for complex multi-component systems based on DBN-HAZOP model. Process Safety and Environmental Protection 90, 376–388. https://doi.org/10.1016/j.psep.2012.06.004

Janošovský, J., Danko, M., Labovský, J., Jelemenský, Ľ., 2019. Software approach to simulation-based hazard identification of complex industrial processes. Comput Chem Eng 122, 66–79. https://doi.org/10.1016/j.compchemeng.2018.05.021

Janošovský, J., Danko, M., Labovský, J., Jelemenský, Ľ., 2017. The role of a commercial process simulator in computer aided HAZOP approach. Process Safety and Environmental Protection 107, 12–21. https://doi.org/10.1016/j.psep.2017.01.018

Joubert, F., Steyn, E., Pretorius, L., 2021. Using the HAZOP Method to Conduct a Risk Assessment on the Dismantling of Large Industrial Machines and Associated Structures: Case Study. J Constr Eng Manag 147. https://doi.org/10.1061/(asce)co.1943-7862.0001942

Kim, D., Seo, D., Cho, S., Kang, P., 2019. Multi-co-training for document classification using various document representations: TF–IDF, LDA, and Doc2Vec. Inf Sci (N Y) 477, 15–29. https://doi.org/10.1016/j.ins.2018.10.006

Kim, H.K., Kim, H., Cho, S., 2017. Bag-of-concepts: Comprehending document representation through clustering words in distributed representation. Neurocomputing 266, 336–352. https://doi.org/10.1016/j.neucom.2017.05.046

Kongsberg, 2023. Kongsberg high-technology systems - k-Spice. https://www.kongsbergdigital.com/.

Lawley, H.G., 1976. Size up plant hazards this Way. Hydrocarbon Processing 55, 247–261.

Lawley, H.G., 1974. Operability Studies and Hazard Analysis. Chem Eng Prog 70, 105–116.

Li, K., Yao, X., Chen, D., Yuan, L., Zhou, D., 2015. HAZOP study on the CTCS-3 onboard system. IEEE Transactions on Intelligent Transportation Systems 16, 162–171. https://doi.org/10.1109/TITS.2014.2329692

Liao, S.H., 2005. Expert system methodologies and applications-a decade review from 1995 to 2004. Expert Syst Appl 28, 93–103. https://doi.org/10.1016/j.eswa.2004.08.003

MECHHOUD, E.-A., RODRIGUEZ, M., ZENNIR, Y., 2017. Automated dependability analysis of the HDPE Reactor using D-higraphs HAZOP assistant. Algerian Journal of Signals and Systems 2, 255–265. https://doi.org/10.51485/ajss.v2i4.51

Mechhoud, E.-A., Rouainia, M., Rodriguez, M., 2016. A new tool for risk analysis and assessment in petrochemical plants. Alexandria Engineering Journal 55, 2919–2931. https://doi.org/10.1016/j.aej.2016.05.013

Meho, L.I., Rogers, Y., 2008. Citation counting, citation ranking, and h-index of human-computer interaction researchers: A comparison of scopus and web of science. Journal of the American Society for Information Science and Technology 59, 1711–1726. https://doi.org/10.1002/asi.20874

Meng, Y., Song, X., Zhao, D., Liu, Q., 2021. Alarm management optimization in chemical installations based on adapted HAZOP reports. J Loss Prev Process Ind 72. https://doi.org/10.1016/j.jlp.2021.104578

Mikolov, T., Chen, K., Corrado, G., Dean, J., 2013. Efficient Estimation of Word Representations in Vector Space.

Milazzo, M.F., Aven, T., 2012. An extended risk assessment approach for chemical plants applied to a study related to pipe ruptures. Reliab Eng Syst Saf 99, 183–192. https://doi.org/10.1016/j.ress.2011.12.001

Minsky, M., 1995. A framework for representing knowledge. . Computation and Intelligence, ed. G.F. Luger, American Association for Artificial Intelligence: Menlo Park, CA 163–189.

Mitkowski, P.T., Zenka-Podlaszewska, D., 2014. HAZOP method in identification of risks in a CPFR supply chain. Chem Eng Trans 39, 445–450. https://doi.org/10.3303/CET1439075

Noorudheen, N., McClanachan, M., Toft, Y., Dell, G., 2013. Keeping track workers safe: A socio-technical analysis of emerging systems and technology. Proc Inst Mech Eng F J Rail Rapid Transit 227, 517–528. https://doi.org/10.1177/0954409713501654

Oeing, J., Holtermann, T., Welscher, W., Severins, C., Vogel, M., Kockmann, N., 2023. preHAZOP: Graph-Based Safety Analysis for Early Integration into Automated Engineering Workflows. Chem Ing Tech 95, 1083–1095. https://doi.org/10.1002/cite.202200222

Peng, L., Gao, D., Bai, Y., 2021. A study on standardization of security evaluation information for chemical processes based on deep learning. Processes 9. https://doi.org/10.3390/PR9050832

Qader, W.A., Ameen, M.M., Ahmed, B.I., 2019. An Overview of Bag of Words;Importance, Implementation, Applications, and Challenges, in: Proceedings of the 5th International Engineering Conference, IEC 2019. Institute of Electrical and Electronics Engineers Inc., pp. 200–204. https://doi.org/10.1109/IEC47844.2019.8950616

Rimkevičius, S., Vaišnoras, M., Babilas, E., Ušpuras, E., 2016. HAZOP application for the nuclear power plants decommissioning projects. Ann Nucl Energy 94, 461–471. https://doi.org/10.1016/j.anucene.2016.04.027

Rodríguez, M., De la Mata, J.L., 2012. Automating HAZOP studies using D-higraphs. Comput Chem Eng 45, 102–113. https://doi.org/10.1016/j.compchemeng.2012.06.007

Rossing, N.L., Lind, M., Jensen, N., Jørgensen, S.B., 2010. A functional HAZOP methodology. Comput Chem Eng 34, 244–253. https://doi.org/10.1016/j.compchemeng.2009.06.028

Shi, L., Chen, L., 2020. Hazard recognition and reliability analysis of CTCS-3 on-board subsystem. Comput Commun 151, 145–153. https://doi.org/10.1016/j.comcom.2019.12.025

Single, J., Schmidt, J., Denecke, J., 2019. Computer-aided hazop studies: Knowledge representation and algorithmic hazard identification, in: WIT Transactions on the Built Environment. WITPress, pp. 55–66. https://doi.org/10.2495/SAFE190061

Single, J.I., Schmidt, J., Denecke, J., 2020a. Ontology-based computer aid for the automation of HAZOP studies. J Loss Prev Process Ind 68. https://doi.org/10.1016/j.jlp.2020.104321

Single, J.I., Schmidt, J., Denecke, J., 2020b. Ontology-based support for hazard and operability studies. International Journal of Safety and Security Engineering 10, 311–319. https://doi.org/10.18280/ijsse.100302

Single, J.I., Schmidt, J., Denecke, J., 2019. State of research on the automation of HAZOP studies. J Loss Prev Process Ind. https://doi.org/10.1016/j.jlp.2019.103952

Stanić, N., Langeveld, J.G., Clemens, F.H.L.R., 2014. HAZard and OPerability (HAZOP) analysis for identification of information requirements for sewer asset management. Structure and Infrastructure Engineering 10, 1345–1356. https://doi.org/10.1080/15732479.2013.807845

Studer, R., Grimm, S., Abecker, A., 2007. Semantic Web Services: Concepts, Technologies, and Applications. Springer, Berlin and Heidelberg.

Suh, Y., 2021. Sectoral patterns of accident process for occupational safety using narrative texts of OSHA database. Saf Sci 142. https://doi.org/10.1016/j.ssci.2021.105363

Szmel, D., Zablocki, W., Ilczuk, P., Kochan, A., 2019. Method for Selecting the Safety Integrity Level for the Control-Command and Signaling Functions. Sustainability (Switzerland) 11. https://doi.org/10.3390/su11247062

Tang, G., Müller, M., Rios, A., Sennrich, R., 2018. Why Self-Attention? A Targeted Evaluation of Neural Machine Translation Architectures.

Taylor, J.R., 2017. Automated HAZOP revisited. Process Safety and Environmental Protection 111, 635–651. https://doi.org/10.1016/j.psep.2017.07.023

van Eck, N.J., Waltman, L., 2010. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 84, 523–538. https://doi.org/10.1007/s11192-009-0146-3

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S.N., 2003a. A review of process fault detection and diagnosis Part II: Qualitative models and search strategies. Computers and Chemical Engineering 27, 313–326.

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S.N., 2003b. A review of process fault detection and diagnosis part i: quantitative model-based methods . Computers and Chemical Engineering 27, 293–311.

Villa, V., Paltrinieri, N., Khan, F., Cozzani, V., 2016. Towards dynamic risk analysis: A review of the risk assessment approach and its limitations in the chemical process industry. Saf Sci. https://doi.org/10.1016/j.ssci.2016.06.002

Wang, C., Wang, J., Jiapeng Li, Chen, F., Zhi, Y., Wang, Z., 2022. RESEARCH ON QUANTIFICATION OF HAZOP DEVIATION BASED ON A DYNAMIC SIMULATION AND NEURAL NETWORK. International Journal of Industrial Engineering: Theory, Applications and Practice 29, 959–978.

Wang, F., Gu, W., 2022. Intelligent HAZOP analysis method based on data mining. J Loss Prev Process Ind 80. https://doi.org/10.1016/j.jlp.2022.104911

Wang, J., Yang, M., Li, T., Jiang, X., Lu, K., 2023. Research and Application of Improved Multiple Imputation Based on R Language in Fire Prediction. Fire 6, 235. https://doi.org/10.3390/fire6060235

Wang, Z., Wang, B., Ren, M., Gao, D., 2023a. A new hazard event classification model via deep learning and multifractal. Comput Ind 147. https://doi.org/10.1016/j.compind.2023.103875

Wang, Z., Wang, B., Ren, M., Gao, D., 2023b. A new hazard event classification model via deep learning and multifractal. Comput Ind 147, 103875. https://doi.org/10.1016/j.compind.2023.103875

Wang, Z., Zhang, B., Gao, D., 2022. A novel knowledge graph development for industry design: A case study on indirect coal liquefaction process. Comput Ind 139. https://doi.org/10.1016/j.compind.2022.103647

Wang, Z., Zhang, B., Gao, D., 2021. Text mining of hazard and operability analysis reports based on active learning. Processes 9. https://doi.org/10.3390/pr9071178

Wu, C., Xu, X., Zhang, B., Na, Y., 2013. Domain ontology for scenario-based hazard evaluation. Saf Sci 60, 21–34. https://doi.org/10.1016/j.ssci.2013.06.003

Yi, J., Wang, H., Zhang, J., 2023. Dynamic simulation-based quantitative hazard and operability process hazard analysis for a hydrocracking unit. Process Safety Progress. https://doi.org/10.1002/prs.12548

Zenier, F., Antonello, F., 2023. ALBATROS III: an Integrated Software to Obtain the Fault Tree, SIL Level and MCS from the Hazop. Chem Eng Trans 99, 139–144. https://doi.org/10.3303/CET2399024

Zhang, M., Song, W., Chen, Z., Wang, J., 2016. Risk assessment for fire and explosion accidents of steel oil tanks using improved AHP based on FTA. Process Safety Progress 35, 260–269. https://doi.org/10.1002/prs.11780

Zhao, Y., Zhang, B., Gao, D., 2022. Construction of petrochemical knowledge graph based on deep learning. J Loss Prev Process Ind 76. https://doi.org/10.1016/j.jlp.2022.104736

Zhu, L., Ma, H., Huang, Y., Liu, X., Xu, X., Shi, Z., 2022. Analyzing Construction Workers' Unsafe Behaviors in Hoisting Operations of Prefabricated Buildings Using HAZOP. Int J Environ Res Public Health 19. https://doi.org/10.3390/IJERPH192215275