HFACS-based Bayesian Network: Machine learning approach to Human Factors in Hydrogen accidents

Edem Yao Tsei[1], David Barry[2] & Duncan Hewat[3]

[1]Cranfield Aerospace Solutions, UK [2]Safety and Accident Investigation Centre, Cranfield University, UK
[3]Ministry of Defence, UK

SUMMARY

This study combines Bayesian Network (BN) machine learning tool and HFACS to analyse safety risks related to human and organisational factors in hydrogen (H2) accidents in the H2tools database to deduce lessons for aviation. The study statistically identifies significant causal associations between human risk factors and their effect on H2 accidents. Ultimately, the research contributes to the existing human factors knowledge gap in understanding H2 accident risk factors and develops a model for proactive H2 safety management in the aviation domain.

KEYWORDS

Bayesian Network, Human Factors Analysis and Classification System, Aviation

Introduction

The modern approach to risk analysis and safety assessment has shifted from traditional methods like HFACS for predicting risks (Shappell & Wiegmann, 2000). Probabilistic reasoning and machine learning tools like BN has played significant role in decision-making and risk prediction (Fenton N & Neil M, 2019), for example analysing complex data sets like accident investigations in aviation, other means of transportation and understanding new forms of energy like Hydrogen (H2). These days, H2, a Sustainable Aviation Fuel (SAF), has been on the frontline in environmental decarbonisation to achieve NetZero by 2050. Interestingly hydrogen has become vital although unpopular historically and safety wise; for example, the 1937 Hindenburg accident was allegedly caused by H2 though currently disputed (Dessler et al., 2005), This unpopular safety history related to accidents in industry, transportation and space exploration led to many studies to unravel the causes of H2 accidents in storage, system design, maintenance and, more importantly, human, and organisational factors, which contributed to over half of H2 accidents in a previous studies (Wen et al., 2022). However, understanding the associations between levels of human factors in H2 accidents is a knowledge gap. Additionally, only few H2 accidents in aviation are recorded in databases like H2tools. This unavailability of aviation-specific data on H2 incidents in aviation makes drawing statistically valid lessons from other industries challenging. Bridging this gap by learning from non-aviation H2 accidents is this study's primary motivation.

The complexity and evolving H2 technologies in aviation calls for a system approach to understanding the roles played by humans and organisations in H2 systems. BN which is built on the Bayesian theorem therefore provides a good risk assessment probabilistic tool that combines quantitative data and qualitative expert knowledge to deduce lessons from past H2 accidents. HFACS based BN has been used extensively in risk assessment and accident investigation (Fenton & Neil, 2018; Basnet et al., 2023; Marsh & Bearfield, 2004; Zhang & Mahadevan, 2020; Barry, 2021).

Methodology

The study adopts three steps to determine if there is an association between human and organisational risk factors of hydrogen accidents and if accident causal factors of non-aviation H2 accidents are related to a recent hydrogen aircraft accident. Step I: Improved on the traditional HFACS framework to obtain HFACS-H2 framework and Expert classification of 100 H2 accident risk factors; Step II: Construct BN model using HFACS-H2 pathways; Step III testing of the BN for sensitivity and validity and comparing with G-HYZA accident report.

The above-mentioned steps as illustrated in the research process flow chart in Figure 1 were modified to suit H2 studies and were replicated on this current study to achieve research objectives with adaptations from Niu et al., (2023).

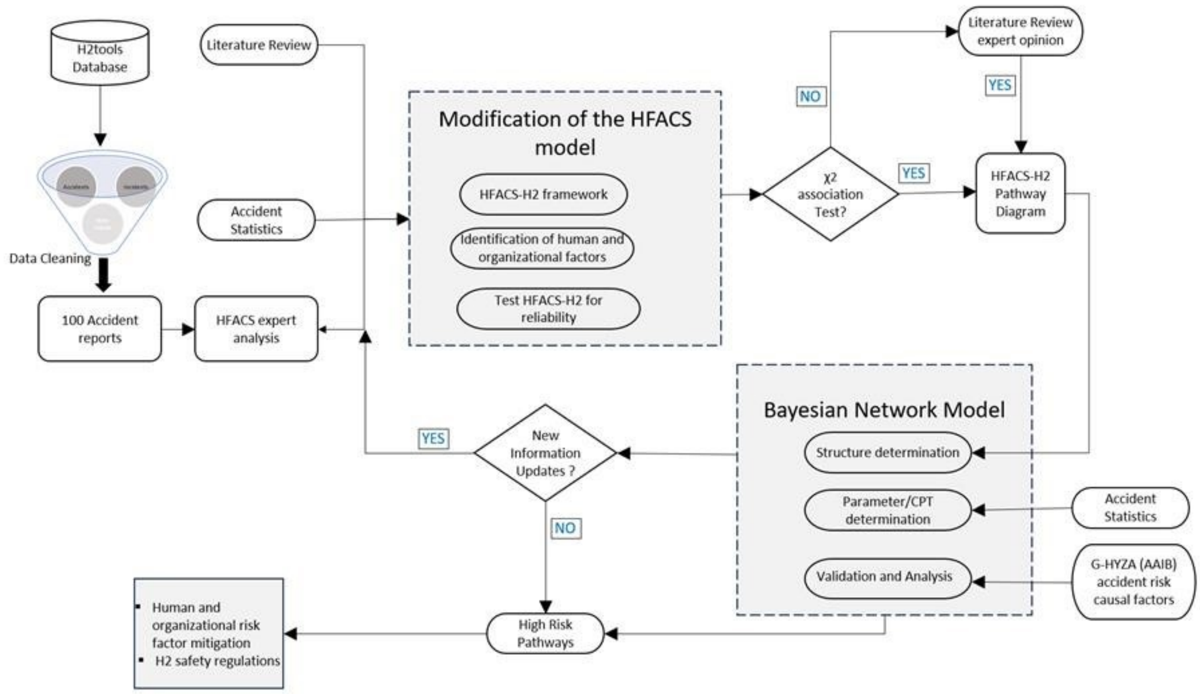

Figure 1: Analysis process of human risk factors of hydrogen fuel accidents (Author's diagram adapted from Niu et al. (2023))

Data Collection

A total of 222 hydrogen accident reports recorded between 1948 and 2019 from H2tools database were studied and filtered to 100 accident reports attributed to human and organisational factors. An exploratory study of the data indicated that the highest Human Factors (HF) related H2 accidents occurred in 1969. Though this could not be explained most 1969 accidents were procedural deficiencies. About 53 % resulted in an ignition, 75 % resulted from H2 leakage, 77 % of the accidents/incidents were discovered during H2 operations with highest source occurring on H2 storage compared to H2 transportation, fuelling, fuel cells and other sources.

Results and Discussions

The results after data classification and interrater reliability verification by two experts trained on the HFACS-H2 model developed for the purpose of the study is shown in Table 1 Table 1: HFACS-H2 risk category frequency table

Level	Risk Category	Times	Frequency
Organisational Influences	Resource Management & Planning **(A1)**	6	6%
	Organisational Process **(A2)**	59	59%
	Organisational Climate **(A3)**	1	1%
Unsafe Supervision	Inadequate Supervision **(B1)**	25	25%
	Failure to Correct a Known Problem **(B2)**	1	1%
	Supervisory violations **(B3)**	20	20%
	Inadequate Emergency Response Plan **(B4)**	6	6%
Pre-Conditions for Unsafe Acts	Poor ventilation Design **(C1)**	6	6%
	Extreme Weather Conditions **(C2)**	2	2%
	Confined Space Installation **(C3)**	2	2%
	Poor Design of Equipment **(C4)**	16	16%
	Poor Installation **(C5)**	4	4%
	Poor Maintenance procedures **(C6)**	23	23%
	Adverse Physiological state **(C7)**	-	-
	Adverse Mental State **(C8)**	-	-
	Physical /Mental limitation **(C9)**	-	-
	Inadequate Training (**C10)**	16	16%
	Inadequate Certification (**C11)**	-	-
	Unfamiliar with H_2 Safety Procedures **(C12)**	8	8%
Unsafe Acts of Operators & Maintenance personnel	Skilled/Technical errors **(D1)**	17	17%
	Decision/Judgement Errors **(D2)**	1	1%
	Perceptual Errors **(D3)**	8	8%
	Attention Errors **(D4)**	10	10%
	Habitual violations **(D5)**	-	-
	Routine violations **(D6)**	14	14%
	Accidental Violations **(D7)**	6	6%

Interrater Reliability Test of HFACS-H2 Framework

To develop a reliable BN from the HFACS-H2 framework, an interrater reliability was conducted on the classification performed by the two experts McHugh (2012). The coefficient of interrater agreement (Cohen's Kappa) by Cohen (1960) and its significance were computed for all risk categories prior to raters meeting to agree and iron out differences in classification. More than half of the HFACS-H_2 categories recorded a Kappa of 0.60 or more, indicating moderate to almost perfect agreement McHugh (2012). Five out of 26 risk categories had weak agreements. Decision /judgement errors, resource management & planning recorded minimal agreement (0.21- 0.39), probably due to the low frequency of occurrence and sample size. However, Li & Harris (2007) indicated that Cohen's Kappa could be erroneous in small sample sizes and where raters classify many cases into one category, as was the case in this research for Decision /Judgement errors and Resource management & planning. The percentage of interrater reliability for the HFACS-H_2 framework indicates reliability between 81.0% and 100%. This percentage of reliability further indicates an acceptable level of reliability of raters.

Chi-square Test of Association between Risk Categories

The study identified a significant causal relationship between upper and lower HFAC-H2 risk factors, as shown in Table 2. Significant HFAC-H2 pathways deduced in Figure 2 with computed prior and conditional probabilities of causal factors were modelled into a BN in Figure 6 using GeNIe BN software to learn lessons

Table 1: HFACS-H2 Chi ^{2} Association

Risk Factor Association	**Chi^2**	***p***
level-4 association with level 3 and 2		
Organisational Process × Inadequate Emergency plan	6.554	0.016
Organisational Process × Unfamiliar with H_2 safety procedures	5.797	0.004
level-3 association with level 2 and 1		
Supervisory Violations × Poor design of equipment	5.864	0.011
Supervisory Violations × Poor Maintenance Procedures	4.948	0.037
Supervisory Violations × Routine Violations	54.007	0.001
level-2 association with level 1		
Poor Maintenance Procedures × Routine Violations	5.141	0.020
Unfamiliar with H_2 safety procedures × Routine Violations	3.989	0.046
Inadequate training × Skilled Errors	4.858	0.039

.

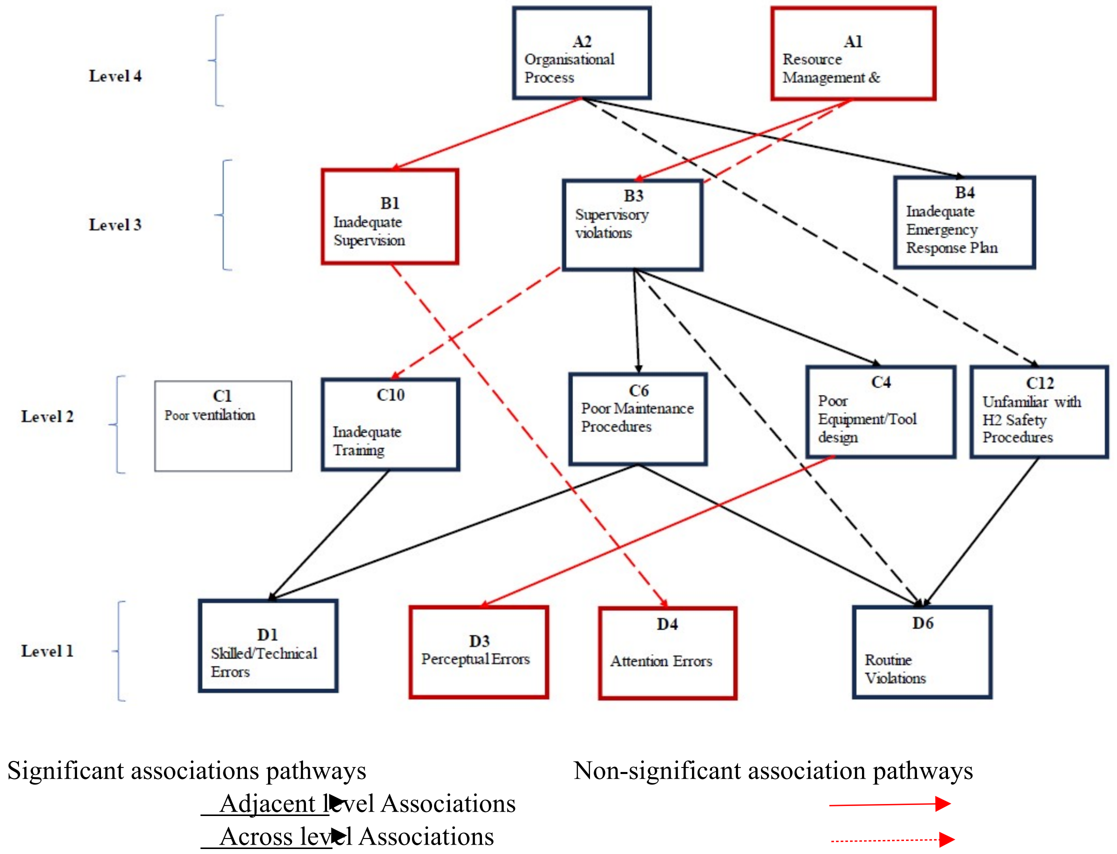

Figure 2: HFACS-H2 pathways of Association

Determination of Bayesian Network Parameters

In most BN applications in risk analysis, the quantitative association between nodes (risk factors) are determined by expert judgements. In the absence of expert judgement or when expert probabilities are unfeasibly large, the probability of occurrence of parameters is determined through a calculation using the frequency of analysed data (Fenton N & Neil M, 2019). The Prior Probabilities (PP) of parent (root) nodes and Conditional Probabilities (CP) of child nodes were therefore determined in this study as was done in previous studies (Niu et al., 2023). The prior probabilities of the root nodes in this study are taken directly from the observed frequency tables. Where 'True' is the probability of a risk factor occurring (6% = 0.06) in the case of organisational process (A1), and the probability of A1 not occurring is 0.94 (1- 0.06). The CP depict the influence level of parent nodes (organisational process) on child nodes B1, B3, B4, C10, C12. Though several methods exist in determining conditional probabilities, this study adopts the simplified causal structures using Noisy OR gates Rohmer (2020) and calculated conditional probabilities by cooper (1984, p.19). According to Rohmer (2020) CP derivation may be based on strong assumptions. For example, this study assumed that a child node is false only if the parent node is false for complex CPs. Table 3 is an extract of CP derivation for Organisational process and Its child node

Table 3: Conditional probability of inadequate emergency response and frequency table

Frequency Statistics	Organisational Process				Conditional probabilities	Organisational Process	
Inadequate		False	True	**Total** 55		False	True
Emergency	False	39	94		False	0.95	0.93
Response	True	2	4	6	True	0.05	0.07
	Total	41	59	100	**Total**	1	1

Bayesian Network Results

When prior and conditional probability outputs were fed into a modelled BN using GeNIe Academic tool, simulated results in Figure 4 indicate behaviour of parameters (Risk factors) related to human and organisational factors in achieving hydrogen safety. From the model in level 1 (unsafe acts of operators), routine violations (27%) have the highest occurrence probability, followed by killed/technical errors (12%) and attention errors (10%). Perceptual errors had the lowest chance of occurrence. These results support previous research (Wen et al., 2022b).

The model also indicates the need to monitor and provide adequate supervision to front-line operators to reduce H_2 procedure violations to improve safety. At the organisational influence level, organisational process recorded a likelihood of more than 50 %, which is comparatively high to other risk factors. Management should, therefore, regularly review and ensure hydrogen safety procedures, operational checks, and processes are adhered to.

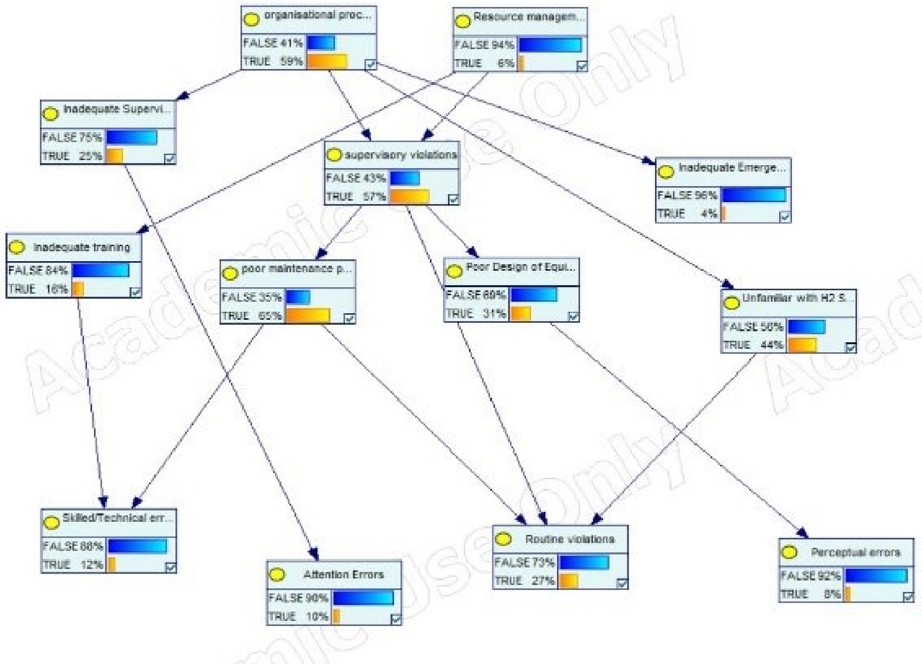

Figure 4: BN modelled structure output from Genie Academic software

In level 3 (unsafe supervision), supervisory violations and inadequate supervision recorded probabilities of 57 % and 25%, respectively. The impact of violations of supervisors in ensuring hydrogen safety procedures may have caused routine violations by front-line operators and Maintenance personnel. This causal effect re-affirms previous studies adapting Reason's theory of active failures of operators caused by latent failures at higher supervisory levels adapted by (Kulsomboon, Tsei, et al.,2023) The model also shows a 6 % chance for resource management and planning. Though this low value could be attributed to low study sample size, it, however, may have impacted inadequate resources for training and planning at the supervision level. Finally, in level 2, which is the preconditions for unsafe acts, poor maintenance procedures (65%), unfamiliar with H_2 safety procedures (44%), poor design of equipment and tools (31%), and inadequate training (16%) were the recorded probabilities in the BN model. Deviations from laid down procedures may have been due to a lack of proper supervision.

Bayesian Network Model Validation

The BN model was validated by splitting the data set into two groups; a training set and a test set (Jamilloux et al., 2021). The 100 H2 accident reports from HFAC-H2 analysis was used for training the BN model and the test data is an official accident report of a hydrogen fuel cell aircraft G-HYZA (AAIB, 2022). Table 4 is the comparison between BN model probabilities and HFAC-H2 risk factor frequencies.

Table 4: Probabilities training testing data frequencies

HFACS-H_2 Risk Factors	BN model Probabilities	G-HYZA Frequencies
Organisational process	59	57
Resource management &planning	6	29
Organisational Climate	1*	29
Inadequate supervision	25	42
Supervisory violations	57	29
Failure to correct a known problem	1*	29
Inadequate emergency response plan	4	29
Poor design of equipment/tools	31	29
Inadequate training	16	14
Unfamiliar with H_2 safety procedures	44	42
Routine violations	27	29

**Was not included in the BN model because of very low frequencies.*

*** Individual frequencies do not add up to 100%*

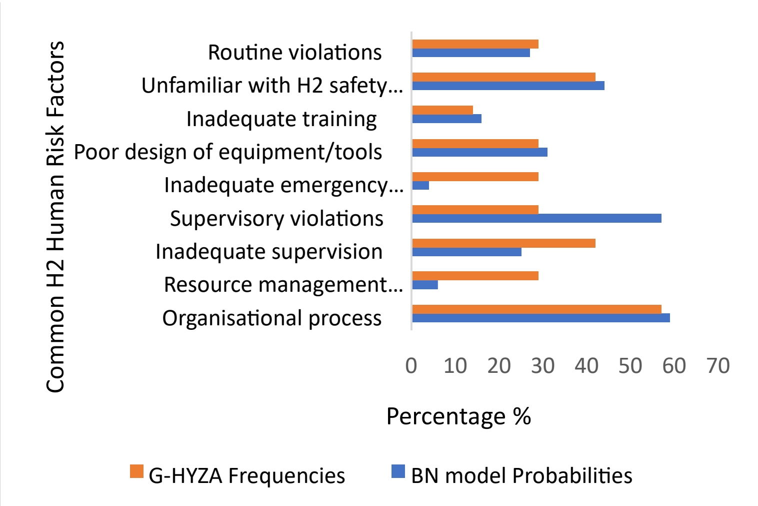

Figure 5: A Histogram comparing BN model probabilities with accident causal frequencies of G-HYZA

There is minor difference between the BN model and the G-HYZA accident analysis in Figure 5. From the comparison, organisational process (59% vs 57%); poor design of equipment (31% vs 29%); inadequate training (16% vs 14%); unfamiliar with H_2 safety procedures (44% vs 42%); and routine violations (27% vs 29%) recorded close frequencies both in the developed BN model probabilities and the test case. In level 3, inadequate supervision and supervisory violations occurred in both instances but at different frequencies. Finally, the test case recorded higher frequencies in resource management and inadequate emergency response plan compared to the BN model probabilities. The similarities between the model and the test case suggest that the model is reliable and could be used to improve hydrogen safety related to human and organisational factors in aviation and non-aviation domains. The results also support hydrogen safety recommendations for the aerospace industry by Wen et al. (2022b). For safe hydrogen application in aviation, hydrogen safety procedures and documentation (organisational process) should be updated regularly by management and communicated to operators to mitigate risk factors at the organisational level.

Furthermore, adequate training, compatible design and installation of hydrogen equipment are required to minimise human errors associated with inadequate training and poor equipment /tool design.

It is important to note that the AAIB accident report on G-HYZA identified seven broad accident causal factors. Classification and percentage conversion may have introduced data collection and measurement biases. Future studies should take advantage of BN flexibility to validate with additional hydrogen-fuelled aircraft incident reports in the future when data becomes available.

Reverse Inference

Table 5: Most likely Induced Paths

Unsafe Acts of Operators	Most likely Induced Path
Routine violations	Organisational process →Supervisory violation→Poor maintenance procedures→Routine violations
Attention errors	Organisational process→Inadequate supervisison→Attention errors
Perceptual errors	Organisational process→supervisory violations→Poor design of equipment→Perceptual errors
Skilled/Technical errors	Organisational process→supervisory violations→Poor maintenance procedures→Skilled/Technical errors

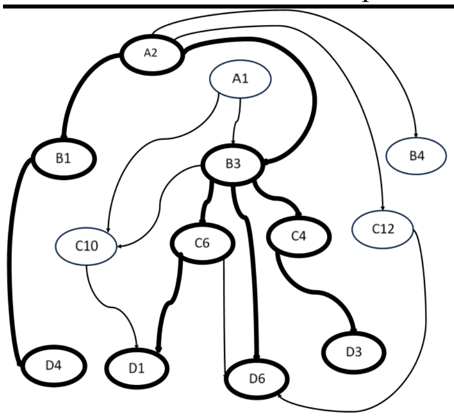

Figure 6: Most likely induced Acyclic graph

One advantage of BN in risk assessment is the flexibility to update influencing factors to simulate probabilities of occurrence of other node factors like unsafe acts of frontline operators leading to an accident. Therefore, by updating unsafe act nodes through reverse inference helps to predictively determine probable HF accident-causing pathways in an H2 system.

Sensitivity analysis

A sensitivity analysis conducted on the BN model indicates that a 10 % increase in most sensitive parameters (Poor maintenance procedures, Inadequate supervision, Poor equipment design) led to about (37.9 %,71.3%, 76.8%) increase in Skilled, Attention and Perceptual errors, respectively. The BN model could be used to proactively develop H2 Safety Management Systems for aviation. The most sensitive node factors contributing to accidents include poor H2 maintenance procedures, inadequate supervision of H2 operations and poor design of H2 equipment.

Lessons Learnt

Following the analysis and discussions of 100 H2 accident causal factors and BN risk modelling, lessons learnt from risk factors are summarized in Table 6.

Table 6: Risk factor Lessons Table

Risk factor analysis	Risk factors
High probability risk factors	Organisational process, supervisory violations, inadequate supervision, poor maintenance procedures, unfamiliar with H2 safety procedures, routine violations
Model comparison and validation factors	Organisational process, poor design of equipment, inadequate training, unfamiliar with H_2 safety procedures, routine violations
Reverse inference factors	Organisational process, supervisory violations, and inadequate maintenance procedures
Most Sensitive node factors	Poor maintenance procedures, inadequate supervision, poor design of equipment

Conclusion

The research identified a significant statistical association between human and organisational factors that contributed to 100 H2 accidents and developed a BN model to deduce lessons for the aviation industry. The findings were similar to G-HYZA (H2-fuel cell aircraft) accident. The research was however limited by H2 data unavailability and inadequate time to elicit CPT parameters from H2 experts. Future studies should combine hydrogen expert knowledge and data to improve validity of the research.

References

Barry, D. J. (2021). Estimating runway veer-off risk using a Bayesian network with flight data. *Transportation Research Part C: Emerging Technologies*, *128*. https://doi.org/10.1016/j.trc.2021.103180

Cooper, G. (1984). *NESTOR: A Computer-Based Medical Diagnostic Aid that Integrates Causal and Probabilistic Knowledge*.

Dessler, A. J., Overs, D. E., & Appleby, W. H. (2005). *The Hindenburg Fire: Hydrogen or Incendiary Paint?*

Fenton N, & Neil M. (2019). *Risk Assessment and Decision Analysis with Bayesian Networks*.

Kulsomboon, P. Tsei, E.Y., Rebbapragada, G., & Li, W.-C. (2023). *Lessons Learned – the similarities and differences of human factors in Aircraft Maintenance between JL123 and* CI611

Li, W.-C., & Harris, D. (2007). *Pilot Error and Its Relationship with Higher Organizational Levels: HFACS Analysis of 523 Accidents*.

Marsh, W., & Bearfield, G. (n.d.). *Using Bayesian Networks to Model Accident Causation in the UK Railway Industry*.

McHugh, L. M. (2012). *Interrater_reliability_The_kappa_statistic*.

Shappell, S. A., & Wiegmann, D. A. (2000). *The Human Factors Analysis and Classification System-HFACS*

Rohmer, J. (2020). *Uncertainties in conditional probability tables of discrete Bayesian Belief Networks: A comprehensive review*. https://doi.org/10.1016/j.engappai.2019.103384ï

Wen, J. X., Marono, M., Moretto, P., Reinecke, E. A., Sathiah, P., Studer, E., Vyazmina, E., & Melideo, D. (2022b). Statistics, lessons learned and recommendations from analysis of HIAD 2.0 database. *International Journal of Hydrogen Energy*, *47*(38), 17082–17096.

Zhang, X., & Mahadevan, S. (2020). Bayesian neural networks for flight trajectory prediction and safety assessment. *Decision Support Systems*, *131*. https://doi.org/10.1016/j.dss.2020.113246