Proceedings of the 35th European Safety and Reliability & the 33rd Society for Risk Analysis Europe Conference
Edited by Eirik Bjorheim Abrahamsen, Terje Aven, Frederic Bouder, Roger Flage, Marja Ylönen

doi: 10.3850/978-981-94-3281-3_ESREL-SRA-E2025-P3665-cd

Simultaneous Prediction of Causes and Consequences in Hydrogen-Related Accidents Using Transformer-Based Multi-Task Learning

July Bias Macedo[1], Plinio Marcio Ramos[1], Luana de Melo Queiroz[1], Dario Valcamonico[2], Márcio das Chagas Moura[1], Isis Didier Lins[1], Piero Baraldi[2], Enrico Zio[2]

[1]*Center for Risk Analysis, Reliability Engineering and Environmental Modeling, Universidade Federal de Pernambuco, Brazil. E-mail: julybias@gmail.com*
[2] *Laboratory of Analysis of Systems for the Assessment of Reliability, Risk and Resilience, Politecnico di Milano, Brazil.*

Hydrogen is emerging as a critical alternative in the global transition to sustainable energy. However, its characteristics - such as high flammability, rapid diffusion, and low ignition energy - pose significant safety challenges in the operation of hydrogen systems. To prevent future accidents, a thorough understanding of hydrogen-related incidents is crucial. This work proposes a novel approach utilizing transformer-based multi-task learning to simultaneously predict accident causes and consequences from unstructured accident narratives within the Hydrogen Incident and Accident Database (HIAD 2.1). By fine-tuning a pre-trained BERT model, the accident narratives are processed to identify root causes while estimating the potential consequences. The proposed multi-task model uses shared representations, enabling efficient learning of both causal and consequence patterns from historical hydrogen accidents. Our results show that multi-task learning enhances the model's ability to generalize across multiple prediction tasks, outperforming single-task models. By suggesting likely causes and consequences, this methodology supports risk identification and assessment, contributing to the safer adoption of hydrogen technologies.

Keywords: Multi-task Learning, Hydrogen Safety, BERT-based Models, Accident Analysis.

1. Introduction

Fossil fuels remain the main source of energy, contributing to global warming. Hydrogen energy is an important solution to addressing global warming challenges and supporting a new energy model. With its versatility and zero-carbon emissions, hydrogen stands out as a key element in the transition to sustainable energy systems (Yang et al. 2024). However, the unique properties of hydrogen, such as low density, high flammability, and explosive characteristics, introduce significant safety risks throughout its production, storage, transportation, and utilization stages (Lu et al. 2024a). These risks necessitate robust strategies for risk management and accident prevention.

Accident narratives, often stored in large industrial databases, are rich with unstructured textual information detailing incidents, causes, and consequences. While these narratives hold valuable insights, manual analysis is time-consuming, error-prone, and infeasible for large datasets (P. M. Ramos et al. 2022; Macêdo, Moura, et al. 2022; Macêdo, Ramos, et al. 2022). Using state-of-the-art transformer-based models like Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al. 2018) allows for the analysis of textual data while preserving contextual relationships between words. With its capability to model complex sentence structures and capture subtle contextual nuances, BERT enables the identification of patterns and interdependencies such as in accident causes and consequences (C. Liu and Yang 2023; Q. Liu, Kusner, and Blunsom 2020).

This paper proposes a novel approach using transformer-based multi-task learning (MTL) to analyze unstructured accident narratives from the Hydrogen Incident and Accident Database (HIAD 2.1). Unlike traditional methods that focus on single aspects of accident analysis, our approach simultaneously predicts causes and consequences, using shared representations to uncover patterns and interdependencies in hydrogen-related incidents. By automating the extraction and interpretation of textual data, this methodology offers insights for risk identification and contributes to the safer adoption of hydrogen technologies.

2. Theoretical Background

The growing adoption of hydrogen technologies has spurred research into improving safety and mitigating risks. Lu et al. (2024) combined Management Oversight and Risk Tree (MORT) analysis with machine learning to identify causative factors and predict risks of hydrogen leakage. Their approach structured data from 23 real-world cases into a dataset of 30 predefined factors, reducing subjectivity and enabling effective risk modeling. Similarly, Luan et al. (2024) employed a knowledge graph and Bayesian network to structure accident data, improving the accuracy of hydrogen risk assessments by integrating historical information with experts' opinions.

While these studies rely on structured datasets (e.g., created through systematic methodologies like MORT or knowledge graphs) to predict and analyze risks, a significant gap remains in utilizing unstructured accident narratives. Accident narratives, such as those found in HIAD 2.1, often contain valuable contextual information about causes and consequences, which are critical for understanding and preventing future incidents. However, their variability in language and format poses challenges for traditional analysis methods.

Recent studies have explored multi-task learning (MTL) techniques to enhance risk analysis. Zhao et al. (2025) proposed a leakage detection method for hydrogen-blended natural gas pipelines, integrating Computational Fluid Dynamics simulations with an MTL-LSTM model. This approach simultaneously predicted leak location and rate, improving detection accuracy over traditional models. Similarly, Desterro et al. (2023) developed an MTL-based framework for nuclear power plant safety, using a Multi-Task Deep Neural Network combined with a Random Forest to predict the size and location of Loss-of-Coolant Accidents with high precision.

In the chemical industry, Guo, Ai, and Luo (2024) introduced an MTL-based risk assessment method, using the RoPLE model to simultaneously predict risk category, probability, and severity, enhancing dynamic risk evaluation. Another relevant study, by Kim et al. (2023), focused on rail safety, proposing a deep anomaly detection system for rail fasteners using MTL. Their model integrated segmentation and denoising with a deep learning classifier, achieving a 97.57% accuracy in detecting anomalies, demonstrating the potential of automated defect detection in industrial applications.

Furthermore, Jia et al. (2022) applied MTL to bearing fault diagnostics, developing a one-dimensional convolutional neural network (MAM-1DCNN) with an attention-guidance mechanism and multi-scale feature extraction. Their model effectively identified fault types and working conditions, addressing the challenge of reduced network performance when multiple tasks are integrated.

These studies highlight the potential of MTL techniques in accident analysis and risk assessment, demonstrating improved predictive accuracy and adaptability. Integrating such methodologies into hydrogen safety analysis could provide more comprehensive insights into accident causation and prevention strategies.

3. Methodology

The proposed approach employs a fine-tuned BERT model optimized for MTL to predict causes and consequences from incident narratives.

3.1. *Dataset*

HIAD 2.1 is a web-based platform designed to provide accurate information about accidents and incidents involving hydrogen. Its primary goal is to assist stakeholders in understanding these events, identifying preventive measures, and facilitating the safe adoption of hydrogen technology (Daniele, Jennifer Xiaoling, and Moretto 2019). This provides a repository of historical accident narratives that are used to train and evaluate the proposed multi-task learning model (Campari et al. 2023).

The dataset contains detailed narratives of various hydrogen-related accidents. These narratives are organized and categorized using unique identifiers and include information such as text descriptions of the accidents, the presence of injuries or fatalities, the cause of accidents, and other critical attributes. The consequences of each accident are categorized in this paper, based on the information extracted from database, into two types: minor damage and severe damage. Incidents classified as having minor damage involve injured individuals but no fatalities, whereas those with fatalities are categorized as severe damage. These categories are represented as a two-dimensional binary vector, where the first digit indicates the presence of severe damage and the second indicates minor damage. This encoding allows for incidents to exhibit both minor and severe damage simultaneously, for example, an accident that led only to injuries is labeled as [1,0], while an accident that led to injuries and fatalities is labeled as [1,1].

Initially, the events were classified into eight categories based on their causes: Unknown, System design error, Material/manufacturing error, Installation error, Job factors, Human factors, Management factors, and Environmental factors. In this paper, events labeled as "Unknown" or solely attributed to "Environmental factors" were excluded from the analysis. The remaining categories were regrouped into "Human factors" and "Non-human factors." Specifically, events marked with "Job factors," "Human factors," and "Management factors" were classified under "Human factors," while "System design error," "Material/manufacturing error," "Installation error," and "Environmental factors" were categorized as "Non-human factors".

3.2 *Multi-Task Learning Framework*

Our model is based on an MTL framework utilizing BERT as the shared feature extractor. BERT is a transformer-based deep learning model that processes textual input bidirectionally, capturing both left and right contextual dependencies. The architecture consists of multiple self-attention layers that enable efficient representation learning from unstructured text. Specifically, our implementation employs BERT-base, which consists of:

- 12 transformer layers (encoders)
- 768 hidden units per layer
- 12 self-attention heads
- 110 million parameters.

The training process was designed to optimize the multi-task BERT model for accident narrative analysis. A custom dataset class was implemented to preprocess the textual data, ensuring proper tokenization and structured labeling. Specifically, BERT tokenizer was applied to convert raw text into tokenized sequences, including special tokens (such as [CLS] and [SEP]) and attention masks and binary labels were assigned separately for cause and consequence classification tasks.

The model architecture consists of shared BERT encoder that transforms input sequences into dense vector representations, and two task-specific classification heads, each composed of a fully connected layer (768 × 1), responsible for predicting causes and consequences, respectively. In addition, dropout layers were added to mitigate overfitting.

To train the model, the dataset was divided into training (80%) and test (20%) subsets. Training was conducted over multiple epochs, using the AdamW optimizer and a binary cross-entropy loss function. During each training step, the following procedure was followed:

- Tokenized input sequences, along with corresponding attention masks, were fed into the shared BERT encoder.
- The encoder output was passed through the cause and consequence classification layers.
- Separate loss values were computed for each task and subsequently aggregated to update model weights via backpropagation.

After each epoch, predictions were generated for the test dataset, and the model's performance was assessed using accuracy and F1-score for both classification tasks. Accuracy offers a general measure of correctness, reflecting the proportion of correctly classified samples across all categories. However, it can be misleading in imbalanced datasets, where the model may perform well overall while failing on minority classes. Thus, F1-score is particularly useful in cases with class imbalance, as it balances precision and recall, ensuring that both false positives and false negatives are considered. A high F1-score indicates that the model is making correct predictions across different classes.

Moreover, to assess the effectiveness of the MTL approach, we will compare its performance with a single-task learning (STL) model. The STL model consists of separate BERT-based classifiers trained independently for cause and consequence prediction, without shared representations. By evaluating both models using the same dataset and performance metrics, we aim to analyze the benefits of parameter sharing in MTL and its impact on predictive accuracy. This comparison will provide insights into whether MTL enhances generalization and improves performance in extracting causal relationships from accident narratives.

4. Results and Discussion

The dataset comprised 221 samples for training and 56 for testing. We conducted experiments with different training durations (3, 5, 6, 10, and 15 epochs) to evaluate the impact of training length on model performance. The best results for cause (task a) and consequence (task b) prediction using MTL and STL are summarized in Table 1.

Table 1. Performance of multi-task model on test data.

Model	Task	Accuracy	F1-score
MTL	*a*	93.67	95.02
	b	87.56	76.19
STL	*a*	58.93	67.00
	b	66.00	54.25

There was a significant performance in the MTL experiment improvement for cause prediction. The initial accuracy of approximately 54% with 3 epochs increased to over 93% after 10 epochs, with an F1-score reaching 0.95. For consequence prediction, the results exhibited greater variability, with accuracy fluctuating between 56% and 66% in the early epochs and ultimately reaching 87% by the 10th epoch. Precision and recall for consequences also showed higher instability, particularly in the initial training cycles.

With the STL, improvements throughout training were more limited, especially for consequence prediction. The initial accuracy for causes was around 25% in the 3-epoch test, remaining relatively constant. With 5 and 10 epochs, there was slight variation, but the results did not reach the levels achieved by the multi-task model. In terms of accuracy and F1-score, the multi-task model consistently outperformed the single-task approach, suggesting that the multitask learning strategy facilitated more efficient knowledge transfer, particularly in cause prediction.

The superior performance of the multi-task model can be attributed to its ability to leverage shared representations across tasks, leading to improved generalization. The strong improvement in cause prediction suggests that shared learning benefits this task more significantly, potentially due to the availability of more consistent patterns in the data. However, the greater variability in consequence prediction indicates that this task might require additional refinements, such as tailored loss weighting, data augmentation, or architectural modifications to enhance stability. These findings highlight the advantages of multi-task learning in complex predictive modeling and suggest directions for further optimization to ensure robust performance across all prediction targets.

4. Conclusion

This study demonstrates how unstructured data from HIAD 2.1 can be transformed into valuable insights, highlighting its potential as a valuable resource for advancing hydrogen risk management and safety research.

The multi-task BERT model provides a robust framework for analyzing hydrogen-related accidents by simultaneously predicting causes and consequences. The encoding of binary consequence categories and the regrouping of causal factors into human and non-human categories enhance the interpretability of the model.

The model's ability to predict causes accurately suggests that the shared encoder effectively learns patterns relevant to causal relationships. The difficulty in predicting consequences aligns with previous findings in the literature, such as P. Ramos et al. (2024) and P. M. Ramos et al. (2022) which underscores the challenges in capturing complex causal relationships and their downstream effects. The variability observed in consequence prediction suggests that additional contextual information or external constraints might be necessary to improve model robustness. Future research could explore integrating domain-specific knowledge or hybrid modeling approaches to address these persistent challenges

References

Campari, Alessandro, Antonio Javier Nakhal Akel, Federico Ustolin, Antonio Alvaro, Alessandro Ledda, Patrizia Agnello, Pietro Moretto, Riccardo Patriarca, and Nicola Paltrinieri. 2023. "Lessons Learned from HIAD 2.0: Inspection and Maintenance to Avoid Hydrogen-Induced Material Failures." *Computers and Chemical Engineering* 173 (February): 108199. https://doi.org/10.1016/j.compchemeng.2023.108199.

Daniele, Melideo, Wen Jennifer Xiaoling, and Pietro Moretto. 2019. "HIAD 2.0 - Hydrogen Incidents and Accidents Database." In *Proceedings of the International Conference on Hydrogen Safety ICHS*.

Desterro, Filipe S.M., Victor H.C. Pinheiro, Cláudio M.N.A. Pereira, and Roberto Schirru. 2023. "Prediction of LOCA's Break Size and Location Based on Random Forest and Multi Tasking Deep Neural Network." *Nuclear Engineering and Design* 415 (December): 112711. https://doi.org/10.1016/j.nucengdes.2023.112711.

Devlin, Jacob, Ming-wei Chang, Lee Kenton, and Toutanova Kristina. 2018. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding." *ArXiv Preprint ArXiv:1810.04805*.

Guo, Yanjun, Xinbo Ai, and Wenbin Luo. 2024. "A Multi-Task Learning Risk Assessment Method for the Chemical Process Industry." *Process Safety and Environmental Protection* 186 (June): 980–94. https://doi.org/10.1016/j.psep.2024.04.030.

Jia, Linshan, Tommy W. S. Chow, Yu Wang, and Yixuan Yuan. 2022. "Multiscale Residual Attention Convolutional Neural Network for Bearing Fault Diagnosis." *IEEE Transactions on Instrumentation and Measurement* 71: 1–13. https://doi.org/10.1109/TIM.2022.3196742.

Kim, Beomjun, Younghoon Jeon, Jeong-Won Kang, and Jeonghwan Gwak. 2023. "Multi-Task Transfer Learning Facilitated by Segmentation and Denoising for Anomaly Detection of Rail Fasteners." *Journal of Electrical Engineering & Technology* 18 (3): 2383–94. https://doi.org/10.1007/s42835-022-01347-1.

Liu, Chang, and Shiwu Yang. 2023. "A Text Mining-Based Approach for Understanding Chinese Railway Incidents Caused by Electromagnetic Interference." *Engineering Applications of Artificial Intelligence* 117 (January). https://doi.org/10.1016/j.engappai.2022.105598.

Liu, Qi, Matt J. Kusner, and Phil Blunsom. 2020. "A Survey on Contextual Embeddings." *ArXiv* arXiv prep. http://arxiv.org/abs/2003.07278.

Lu, Zihan, Yingbin Cao, Yu Zou, Xin Li, Fuqiang Yang, Nima Khakzad, and Chao Chen. 2024a. "Dynamic Risk Analysis of Fire and Explosion Domino Accidents at Hydrogen Refueling Stations Using Dynamic Bayesian Network." *International Journal of Hydrogen Energy* 95 (December): 546–57. https://doi.org/10.1016/j.ijhydene.2024.11.234.

———. 2024b. "Dynamic Risk Analysis of Fire and Explosion Domino Accidents at Hydrogen Refueling Stations Using Dynamic Bayesian Network." *International Journal of Hydrogen Energy* 95 (December): 546–57. https://doi.org/10.1016/j.ijhydene.2024.11.234.

Luan, Tingting, Hongru Li, Kai Wang, Xue Zhang, and Xiaoyun Li. 2024. "Knowledge Graph-Based Bayesian Network Risk Assessment Model for Hydrogen Accidents." *International Journal of Hydrogen Energy* 81 (September): 927–41. https://doi.org/10.1016/j.ijhydene.2024.07.339.

Macêdo, July Bias, Márcio das Chagas Moura, Diego Aichele, and Isis Didier Lins. 2022. "Identification of Risk Features Using Text Mining and BERT-Based Models: Application to an Oil Refinery." *Process Safety and Environmental Protection* 158 (February): 382–99. https://doi.org/10.1016/j.psep.2021.12.025.

Macêdo, July Bias, Plinio Ramos, Caio Bezerra Souto Maior, Márcio Das Chagas Moura, Isis Didier Lins, and Romulo Vilela. 2022. "Identifying Low-Quality Patterns in Accidents Reports from Textual Data." *International Journal of Occupational Safety and Ergonomics*, 1–27. https://doi.org/10.1080/10803548.2022.2111847.

Ramos, Plinio, July Bias Macedo, Caio Souto Maior, Márcio José das Chagas Moura, and Isis Didier Lins. 2024. "Enhancing Preventive Measures in Hydrogen Industries: Root Cause Identification Using NLP Techniques." In *17th International Conference on Probabilistic Safety Assessment and Management & Asian Symposium on Risk*

Assessment and Management (PSAM17&ASRAM2024). Sendai, Japan.

Ramos, Plínio MS, July B Macedo, Caio BS Maior, Márcio C Moura, and Isis Didier Lins. 2022. "Combining BERT with Numerical Variables to Classify Injury Leave Based on Accident Description." *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, December, 1748006X2211401. https://doi.org/10.1177/1748006X221140194.

Yang, Zhen, Xiaochuan Wang, Jincheng Hu, Chaoyu Xu, and He Deng. 2024. "Safety Analysis of Hydrogen Explosion Accident in Underground Hydrogen Storage Gas Injection Station." *International Journal of Hydrogen Energy* 92 (November): 1143–59. https://doi.org/10.1016/j.ijhydene.2024.10.215.

Zhao, Jinpeng, Yunlong Bai, Junlei Li, Wenkai Cu, Wenjing Zhou, Yonghai Zhang, and Jinjia Wei. 2025. "A Leakage Detection Method for Hydrogen-Blended Natural Gas Pipelines in Utility Tunnels Based on Multi-Task LSTM and CFD Simulation." *International Journal of Hydrogen Energy* 97 (January): 1335–47. https://doi.org/10.1016/j.ijhydene.2024.12.031.