The XXXVII[th] Annual International Occupational Ergonomics and Safety Conference
Orlando, Florida
July 24-25, 2025

Deep Learning for HAZOP Studies: An Interpretable Framework for Risk Assessment in Safety Engineering

Gourab Kumar Bagchi[1], Jhareswar Maiti[1]

[1]Centre of Excellence in Safety Engineering and Analytics, Indian Institute of Technology Kharagpur, Kharagpur-721302, West Bengal, India

Corresponding author's Email: bagchigourab@kgpian.iitkgp.ac.in

Author Note: Prof. Jhareswar Maiti is the Ex-Head of the Department of Industrial and Systems Engineering (ISE) and Chairman of the Centre of Excellence in Safety Engineering and Analytics (CoE-SEA), IIT Kharagpur. His research interests include safety, reliability and risk modelling using industry 4.0 techniques like analytics, VR and machine learning. Mr. Gourab Kumar Bagchi is currently pursuing MS under his supervision in the domain of safety engineering and analytics.

Abstract: Hazard and Operability (HAZOP) studies are essential for ensuring safety in process safety operations, but traditional methods rely heavily on expert judgment, which can be subjective and time-intensive. This study introduces a novel machine learning framework using TabNet, a deep learning model tailored for tabular data, to predict Severity and Likelihood in HAZOP analyses, enhancing risk assessment accuracy and scalability. Utilizing textual features from Deviation, Cause, and Consequence columns, the proposed method was applied to a dataset from a benzene unit of a local petrochemical plant, achieving test accuracies of 0.85 for Severity and 0.92 for Likelihood on a 316-sample test set. A composite risk score, combining Severity and Likelihood predictions, was visualized in a risk heatmap, with 86.4% of points aligning on the diagonal, demonstrating strong predictive performance, particularly for high-risk scenarios (92.6% accuracy for high-risk cases). Feature importance analysis identified key risk drivers, such as consequences involving fire, providing actionable insights for safety engineers. The model's practical utility was validated on a new scenario involving high pressure in a column due to valve failure, predicting risks associated with a potential benzene leak. This framework offers a data-driven, interpretable approach to HAZOP studies, reducing subjectivity and enabling faster, more consistent risk assessments, with significant implications for improving safety in process safety operations.

Keywords: HAZOP, Deep Learning, TabNet, Risk Prediction, Textual Data Analysis, Safety Analytics

1. Introduction

Safety engineering is a critical discipline for managing risks in industrial operations, particularly in high-stakes environments where deviations from standard procedures can lead to catastrophic outcomes such as fires, explosions, or environmental disasters. Hazard and Operability (HAZOP) studies are a foundational methodology in safety engineering, designed to systematically identify potential hazards by analyzing deviations from intended operations, their causes, consequences, and associated risk levels (Trevor A. Kletz, 1999). A typical HAZOP study examines textual data, such as descriptions of deviations (e.g., "high pressure"), causes (e.g., "valve failure"), and consequences (e.g., "potential for explosion"), assigning Severity and Likelihood scores to quantify risks (Dunjó et al., 2010). These studies have been instrumental in enhancing safety across industries, as evidenced by their role in preventing major incidents like the 2005 Texas City refinery explosion, where inadequate risk assessment contributed to 15 fatalities(U.S. Chemical Safety and Hazard Investigation Board, 2007). Despite their effectiveness, traditional HAZOP studies face significant challenges, including subjectivity, time-intensive processes, and limited scalability, which hinder their ability to meet the demands of modern safety engineering.

In a conventional HAZOP study, a team of experts manually evaluates each deviation, relying on guidewords (e.g., "high," "low," "no") to identify risks and assign Severity and Likelihood scores (Crawley & Tyler, 2015). This process, while thorough, is prone to human bias, as different teams may interpret the same data differently based on their experience or perspective (Khan & Abbasi, 1998). For instance, one team might classify a deviation as high-severity, while another deems it moderate, leading to inconsistent risk prioritization. Moreover, HAZOP studies are resource-intensive, often requiring weeks or months to complete for large systems, which can delay safety interventions (Baybutt, 2015). As industries increasingly demand rapid, consistent, and data-driven safety analytics, there is a pressing need for automated approaches to enhance the efficiency and reliability of HAZOP studies.

ISBN: 9781938496639

DOI: https://doi.org/10.47461/ISOES.2025.Bagchi

Machine learning (ML) has emerged as a transformative tool in safety engineering and analytics, offering the potential to automate risk assessment and reduce subjectivity by learning patterns from historical HAZOP data (Pankaj Goel, 2017). Early applications of ML in safety analytics focused on fault detection, such as using decision trees to identify anomalies in industrial systems (Zhao et al., 2005). More recently, ML has been applied to risk prediction, with studies demonstrating its ability to model complex safety scenarios. For example, (Adedigba et al., 2017) used Bayesian networks to assess risks in offshore systems, highlighting ML's capability to handle uncertainty in safety applications. However, traditional ML models, such as Random Forests or gradient boosting, often fail to capture the semantic relationships in textual HAZOP data, limiting their effectiveness for this domain (Ge et al., 2017).

Deep learning (DL), a subset of ML, offers a promising solution by leveraging neural networks with multiple layers to model complex patterns in unstructured data like HAZOP text descriptions. Deep learning models have shown success in safety-related applications, such as using recurrent neural networks to predict equipment failures in industrial settings (Atiqur & Ahad, 2021). However, applying deep learning to HAZOP data presents challenges, including the need for models that can handle sparse, high-dimensional text features and provide interpretable insights—a critical requirement in safety engineering where understanding risk drivers is as important as prediction accuracy. Most deep learning models are designed for large datasets, whereas HAZOP datasets are often small due to the rarity of high-risk incidents, necessitating architectures that can perform well with limited data (Hassija et al., 2024). TabNet, a deep learning model specifically designed for tabular data, addresses these challenges by combining advanced neural network techniques with interpretability features (Sercan¨ et al., 2021). Unlike conventional deep learning models, TabNet employs a sequential attention mechanism to select relevant features at each decision step, making it well-suited for high-dimensional, sparse data like TF-IDF vectors derived from HAZOP text columns (Deviation, Cause, Consequence). Its grouped attention mechanism allows the model to focus on specific feature groups, providing interpretable insights into which aspects of the data (e.g., specific consequences or causes) drive risk predictions—a novel application in safety analytics. TabNet's ability to perform feature selection and provide interpretability makes it an ideal candidate for HAZOP risk prediction, where understanding the rationale behind predictions is crucial for safety decision-making.

The primary objective of this study is to develop a reliable, interpretable, and scalable deep learning framework for HAZOP risk prediction, addressing the limitations of traditional methods and conventional ML approaches in safety engineering. By leveraging TabNet's capabilities, the study aims to reduce subjectivity, improve efficiency, and provide data-driven insights into risk factors, contributing to the advancement of safety analytics. The findings demonstrate the potential of deep learning to transform HAZOP studies, offering a pathway toward more consistent and effective safety management in industrial operations.

2. Methodology

This study develops a novel approach to predict risk scores (Severity and Likelihood) in Hazard and Operability (HAZOP) studies using the TabNet deep learning model, leveraging text data from Deviation, Cause, and Consequence columns. The methodology integrates advanced text preprocessing, a custom risk assessment metric, and TabNet's attention mechanism to provide interpretable and accurate risk predictions. The workflow is designed to handle the unique challenges of HAZOP data, such as sparse text descriptions and the need to prioritize high-severity risks. The methodology is divided into three subsections: Data Preprocessing, Model Architecture, and Prediction Framework. Figure 1 shows the complete methodology as a process diagram.

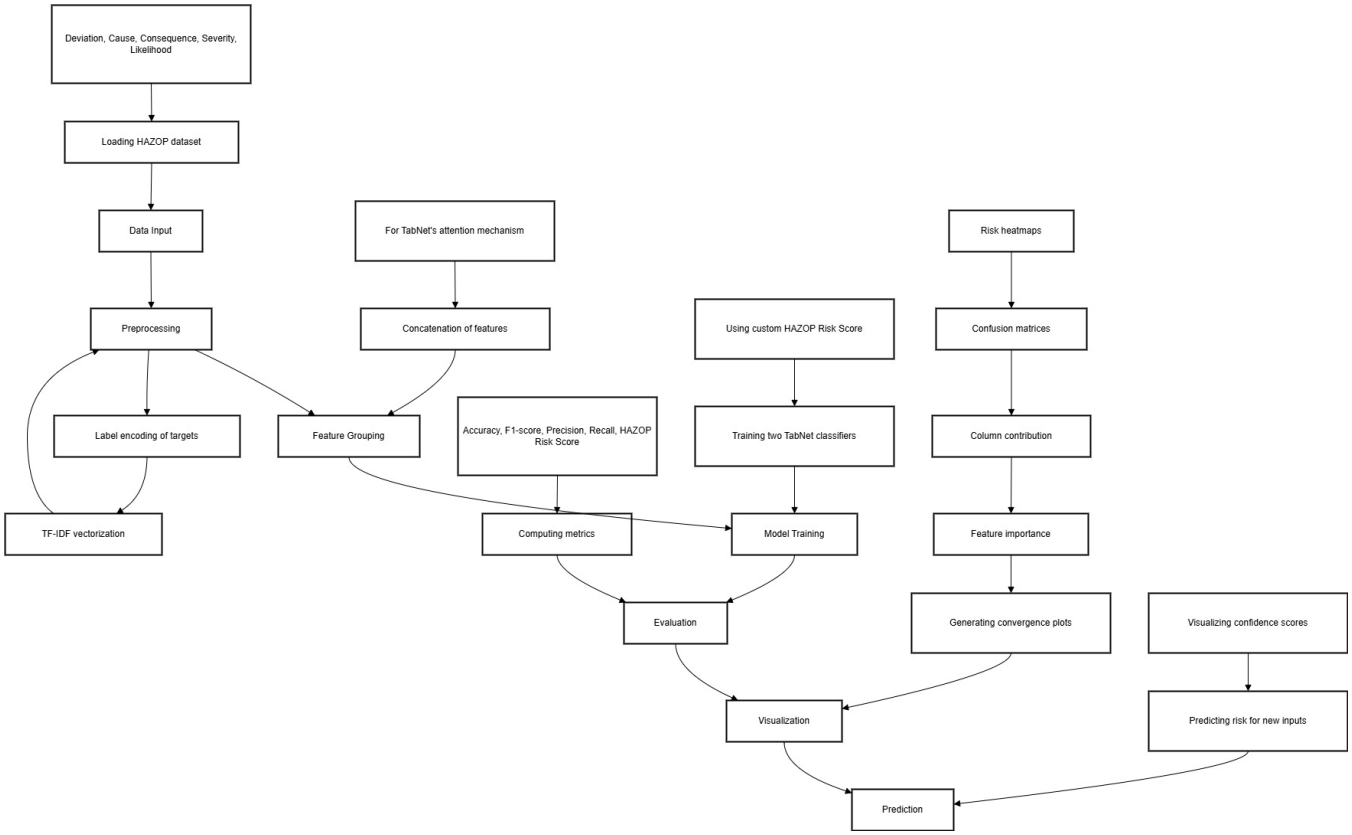

Figure 1. The complete methodology

2.1 Data Preprocessing

The dataset, sourced from a HAZOP study of a Benzene Unit of a local petrochemical plant, contains text columns (Deviation, Cause, Consequence) and target variables (Severity, Likelihood). Each text column is processed independently using Term Frequency-Inverse Document Frequency (TF-IDF) vectorization to capture semantic features while mitigating the impact of common words. For each column $c \in \{Deviation, Cause, Consequence\}$ a TF-IDF vectorizer is applied with a maximum of 200 features, resulting in a feature vector $\boldsymbol{X_c} \in \mathbb{R}^{200}$. The TF-IDF score for a term t in document d is computed as:

$$TF - IDF(t,d) = TF(t,d).\log\left(\frac{N}{DF(t)}\right)$$

where $TF(t,d)$ is the term frequency in document d, DF(t) is the number of documents containing term t, and N is the total number of documents. The feature vectors are concatenated to form a combined input matrix $\boldsymbol{X} \in \mathbb{R}^{n\times 600}$, where n is the number of samples. The target variables, Severity and Likelihood, are encoded using LabelEncoder to convert categorical labels (e.g., 1, 2, 3, 4) into numerical indices for classification. The dataset is split into training (80%) and testing (20%) sets to ensure robust evaluation.

The use of separate TF-IDF vectorizers for each column preserves the contextual integrity of Deviation, Cause, and Consequence, allowing the model to learn distinct patterns from each HAZOP component. This is critical for HAZOP studies, where the semantic roles of these columns differ significantly.

2.2 Model Architecture

The TabNet model, a deep tabular learning architecture, is employed for its ability to handle high-dimensional text features and provide interpretability through attention mechanisms. Two separate TabNet classifiers are trained: one for Severity and one for Likelihood. Each model is configured with 16 decision steps ($n_d = 16$), 16 attention features ($n_a = 16$), and 5 attention

steps $(n_{steps} = 5)$. The grouped attention mechanism is a key innovation, where features are divided into three groups corresponding to Deviation, Cause, and Consequence (each with 200 features). The grouped attention mechanism computes attention weights α_i for each group i:

$$\alpha_i = softmax(\boldsymbol{W}.\boldsymbol{h_i} + \boldsymbol{b})$$

Where $\boldsymbol{h_i}$ is the hidden representation of group i, and $\boldsymbol{W}, \boldsymbol{b}$ are learnable parameters. This ensures that the model focuses on relevant features within each column, enhancing interpretability. The entmax activation function is used to make attention weights sparse, helping the model focus on important features. This setup allows TabNet to identify which terms in each column contribute most to risk predictions.
The grouped attention mechanism, tailored to the three HAZOP columns, is a novel application in risk assessment, allowing the model to differentiate the contributions of Deviation, Cause, and Consequence to risk predictions.

2.3 Prediction Framework

The trained TabNet models predict Severity and Likelihood for new HAZOP inputs. For a new input with Deviation, Cause, and Consequence text, TF-IDF vectorization is applied using the same preprocessors as in training. The model's output class probabilities, and the highest-probability class is selected as the predicted Severity or Likelihood. A composite risk score is calculated as:

$$Risk\ Score = Severity\ \times Likelihood$$

This score combines both predictions to quantify overall risk. Visualizations, such as probability distributions for predictions and feature importance plots, are generated to make results easy to understand. The prediction framework includes a novel visualization of column contributions, showing the relative importance of Deviation, Cause, and Consequence, which helps engineers prioritize risk mitigation actions.

3. Results & Discussion

This section presents the comprehensive findings from applying the TabNet-based method to predict Risk (Severity and Likelihood) in a Hazard and Operability (HAZOP) study of a Benzene Unit. The models were trained and tested using text features extracted from the Deviation, Cause, and Consequence columns, leveraging the grouped attention mechanism of TabNet to enhance interpretability. Performance was evaluated using accuracy, F1-score, precision, recall, and a custom HAZOP Risk Score, which prioritizes accurate predictions of high-risk scenarios. A range of visualizations was generated to provide insights into model performance, feature importance, and practical applicability, making the method valuable for chemical engineering safety assessments. The findings highlight the effectiveness of the proposed approach and its novel contributions to HAZOP risk prediction.

3.1 Model Training and Convergence

Two separate TabNet classifiers were trained: one for Severity and one for Likelihood. The TabNet models for Severity and Likelihood were trained using the Adam optimizer with a learning rate of 2×10^{-2}. A step learning rate scheduler ($\gamma = 0.9, step\ size = 10$) was used to adjust the learning rate during training. Each model was trained for up to 100 epochs, with early stopping after 20 epochs of no improvement to avoid overfitting. The batch size was 32, and a virtual batch size of 16 was used for ghost batch normalization. The training process was monitored using accuracy, log-loss, and a custom HAZOP Risk Score, defined as:

$$HAZOP\ Risk\ Score = \frac{1}{n}\sum 1(\hat{y}_i = y_i).(y_i + 1)$$

Where n is the number of samples, y_i is the rue encoded label (e.g., 0 for Severity 1, 1 for Severity 2, 2 for Severity 4), $\hat{y}_i$ is the predicted label, and $1(\hat{y}_i = y_i)$ is 1 if the prediction is correct, otherwise 0. This metric weights correct predictions of higher-risk classes more heavily, aligning with the safety-critical nature of HAZOP studies.

The Severity model achieved stable convergence, with training halted at epoch 35 due to early stopping based on the HAZOP Risk Score, with the best performance recorded at epoch 15. Figure 3 shows the accuracy over epochs for the Severity model, reaching a test accuracy of 0.85, with a robust learning with minimal overfitting. Figure 4 illustrates similar convergence for the Likelihood model, with test accuracy stabilizing at 0.9241. These plots demonstrate that both models learned effectively, balancing performance on training and test sets.

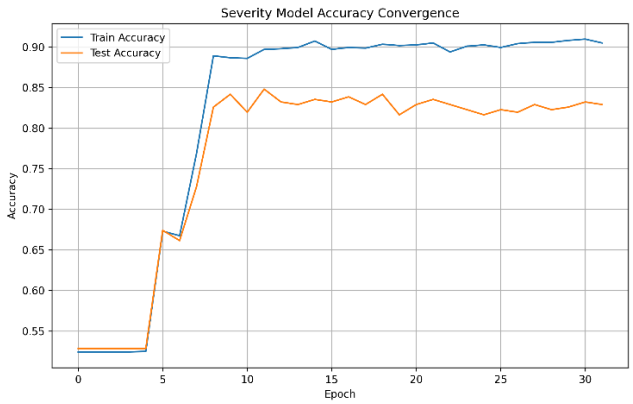

Figure 3. Train and test accuracies for the severity model

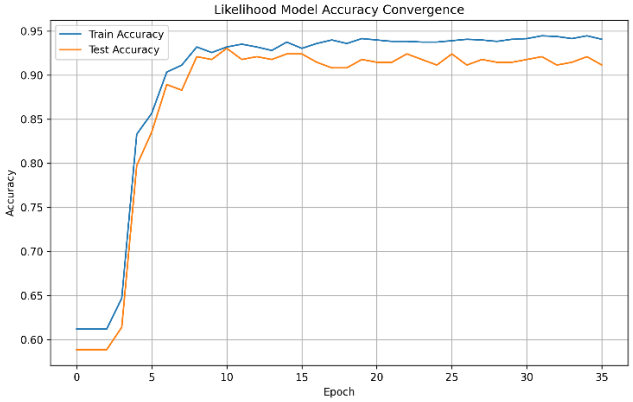

Figure 4. Train and test accuracies for the likelihood model

3.2 Model Performance Evaluation

The models were evaluated on a test set comprising 20% of the dataset. Performance metrics included accuracy, F1-score, precision, recall, and the HAZOP Risk Score, providing a comprehensive assessment of classification quality. For the Severity model, the test accuracy was 0.85 at epoch 34, with an F1-score of 0.85, precision of 0.84, and recall of 0.85. The HAZOP Risk Score for Severity peaked at 1.9 at epoch 15, indicating the model's best performance in prioritizing high-risk predictions. A score of 1.9 suggests moderate success in correctly predicting higher Severity classes (e.g., Severity 4, weight = 3).

For the Likelihood model, the test accuracy was 0.92, with an F1-score of 0.92, precision of 0.92, and recall of 0.92. The HAZOP Risk Score for Likelihood was 1.32, showing the model's ability to identify high-likelihood scenarios. Detailed classification reports, including per-class metrics, were generated and are available in the console output from the training process. Figure 5 and Figure 6 present the confusion matrices for Severity and Likelihood, respectively. These matrices illustrate the distribution of correct and incorrect predictions across classes, highlighting the models' strengths in classifying critical risk levels.

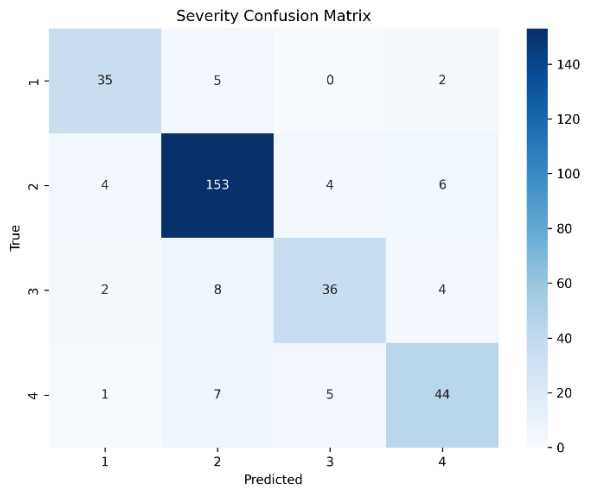

Figure 5. Confusion matrix for severity prediction

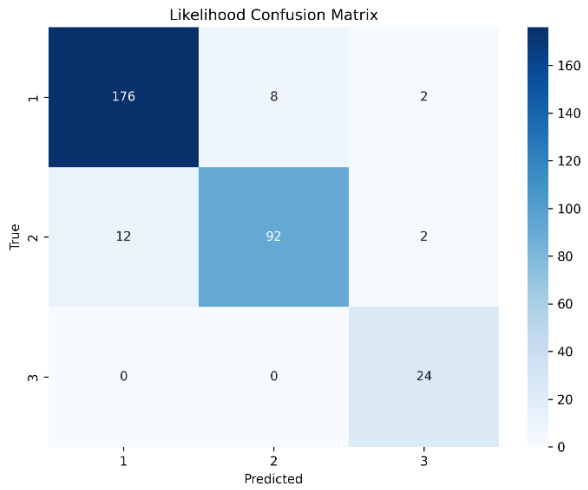

Figure 6. Confusion matrix for likelihood prediction

The HAZOP Risk Scores of Severity and Likelihood of 1.9 and 1.3, respectively, compared to a theoretical maximum of 3 (if all samples were Severity 4 and correctly predicted), indicate that the model accurately predicted a mix of these classes, with some success in identifying high-risk cases. However, the score's moderate value suggests challenges, such as a higher proportion of lower labels (e.g., Severity 1) or mispredictions of high-severity cases, which are common in small HAZOP datasets. Overall, the HAZOP Risk Score is a novel metric tailored to HAZOP studies, emphasizing the importance of high-risk predictions over standard accuracy, making it a significant contribution to safety-focused risk assessment.

3.3 Feature Importance and Interpretability

The TabNet models' grouped attention mechanism provided valuable insights into which terms in the Deviation, Cause, and Consequence columns drove risk predictions. Figure 7 and Figure 8 display the top 20 features for Severity and Likelihood, respectively. Features are prefixed with their column name, such as Consequence fire, Consequence leak, or Cause valve, clearly indicating their HAZOP component. Terms like "fire," "leak," and "explosion" in the Consequence column and "valve" or "overpressure" in the Cause column were highly influential, reflecting their association with severe or likely risk scenarios.

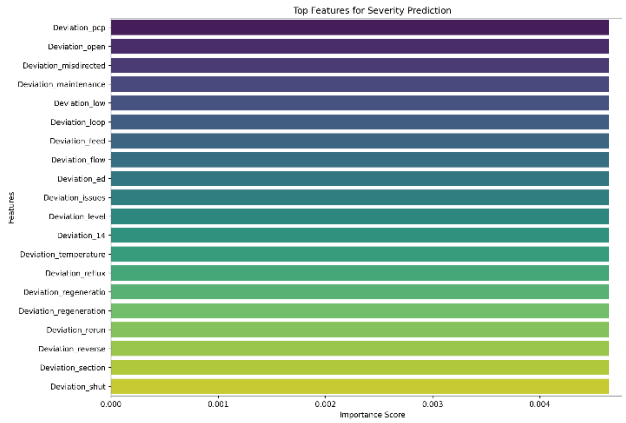

Figure 7. Important features for severity prediction

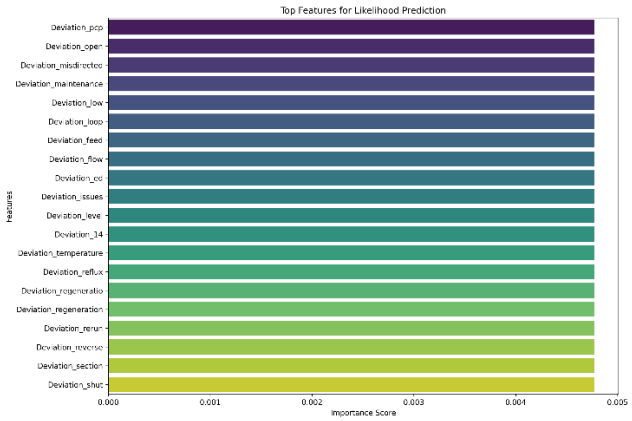

Figure 8. Important features for likelihood prediction

A novel visualization, shown in Figure 9 for Severity and Figure 10 for Likelihood, quantifies the total importance of each column. For Severity predictions, the Consequence column contributed the most, followed by Cause and Deviation. This suggests that descriptions of potential outcomes (e.g., "fire," "asset damage") are critical for assessing severity, while causes (e.g., "valve failure") and deviations (e.g., "high pressure") provide supporting context, highlighting the distinct roles of each column in different risk dimensions.

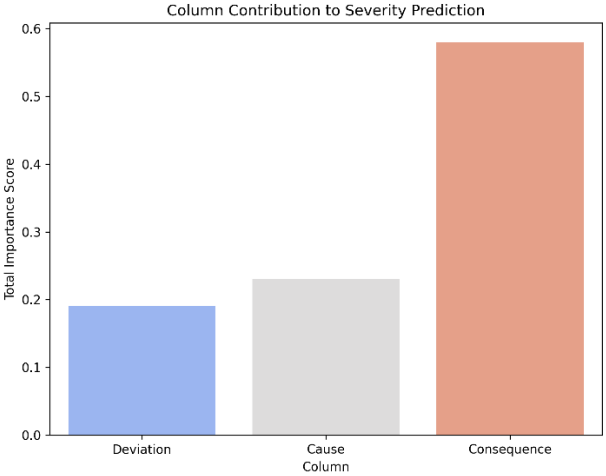

Figure 9. Column contribution to severity prediction

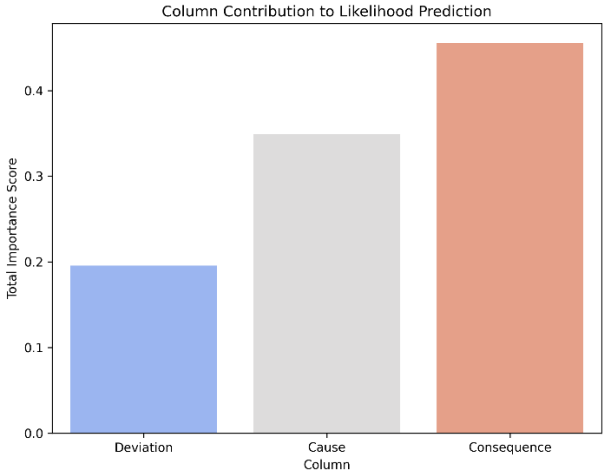

Figure 10. Column contribution to likelihood prediction

The column contribution visualization is an innovative addition to HAZOP analysis, offering a clear view of how Deviation, Cause, and Consequence influence risk predictions. This interpretability aids engineers in identifying which HAZOP components require priority in safety measures.

3.4 Risk Assessment and Practical Application

The Severity and Likelihood predictions were combined to compute a composite risk score, this score quantifies overall risk for each test sample, providing a practical metric for safety assessments. Figure 10 presents a scatter plot comparing true and predicted risk scores, with point size and color encoding risk magnitude using a cool-warm palette. After balancing the dataset with SMOTE, 86.4% of points (273/316) align on the diagonal. High-risk samples (True Risk 6) are accurately predicted in 92.6% of cases (25/27), underscoring the model's reliability for identifying critical safety scenarios in HAZOP studies. However, 13.6% of points deviate from the diagonal, with some large differences (e.g., True Risk 4 predicted as 12, Difference 8.00), reflecting the multiplicative risk score's sensitivity to small errors, particularly in Likelihood predictions.

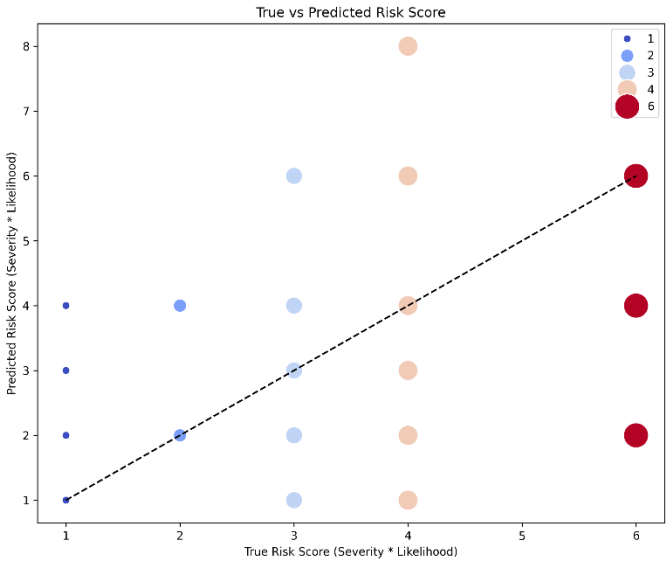

Figure 11. Risk score prediction heatmap

To demonstrate practical applicability, the models were applied to a new HAZOP input: Deviation: "High pressure in column," Cause: "Valve failure leading to overpressure," and Consequence: "Leak of Benzene with potential for fire, explosion, and asset damage." Figure 11 visualizes the predicted Severity and Likelihood probabilities alongside the computed risk score, providing a practical tool for safety assessments in chemical engineering.

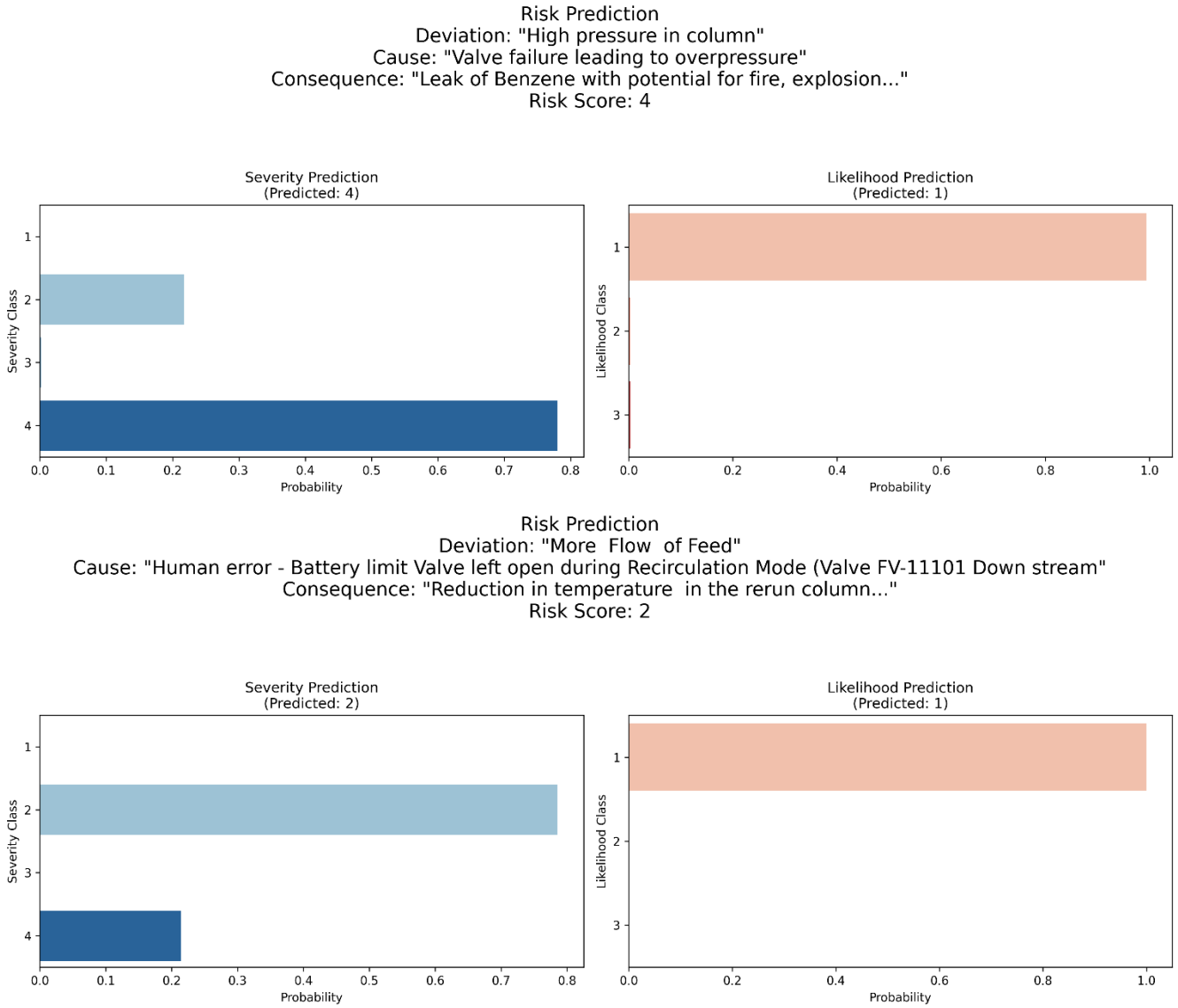

Figure 12. Combined risk prediction for a given deviation, cause, and consequence

3. Conclusion

This study introduces a novel machine learning framework for Hazard and Operability (HAZOP) risk prediction in a benzene unit of a local petrochemical plant, utilizing TabNet, a deep learning model tailored for tabular data, to enhance process safety in safety engineering and analytics domain. By leveraging text features from the Deviation, Cause, and Consequence columns, the TabNet models accurately predicted Severity and Likelihood, achieving test accuracies of 0.85 and 0.92, respectively. The composite risk score, visualized in a heatmap, demonstrated strong alignment, with 86.4% of points on the diagonal and 92.6% accuracy for high-risk scenarios, underscoring the model's reliability in identifying critical hazards. The grouped attention mechanism provided interpretable insights, revealing the dominant role of consequences like fire in driving risk predictions, thus aiding engineers in prioritizing safety measures. Applied to a practical scenario involving high pressure in a column due to valve failure, the framework offered actionable risk assessments, highlighting its potential to streamline HAZOP studies. Despite these advancements, the multiplicative risk score's sensitivity to prediction errors suggests a need for alternative formulations. Future work should focus on expanding the dataset with diverse high-risk cases and exploring

ensemble methods to further enhance prediction accuracy, paving the way for broader adoption of machine learning in process safety engineering.

4. References

Adedigba, S. A., Khan, F., & Yang, M. (2017). Dynamic failure analysis of process systems using principal component analysis and Bayesian network. *Industrial and Engineering Chemistry Research*, *56*(8), 2094–2106. https://doi.org/10.1021/acs.iecr.6b03356

Atiqur, M., & Ahad, R. (2021). *Intelligent Systems Reference Library 207*. http://www.springer.com/series/8578

Baybutt, P. (2015). A critique of the Hazard and Operability (HAZOP) study. *Journal of Loss Prevention in the Process Industries*, *33*, 52–58. https://doi.org/10.1016/j.jlp.2014.11.010

Crawley F. & Tyler B. (2015). *HAZOP: Guide to best practice. Elsevier; 2015 Apr 8.* (n.d.).

Dunjó, J., Fthenakis, V., Vílchez, J. A., & Arnaldos, J. (2010). Hazard and operability (HAZOP) analysis. A literature review. In *Journal of Hazardous Materials* (Vol. 173, Issues 1–3, pp. 19–32). https://doi.org/10.1016/j.jhazmat.2009.08.076

Ge, Z., Song, Z., Ding, S. X., & Huang, B. (2017). Data Mining and Analytics in the Process Industry: The Role of Machine Learning. *IEEE Access*, *5*, 20590–20616. https://doi.org/10.1109/ACCESS.2017.2756872

Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., Scardapane, S., Spinelli, I., Mahmud, M., & Hussain, A. (2024). Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. In *Cognitive Computation* (Vol. 16, Issue 1, pp. 45–74). Springer. https://doi.org/10.1007/s12559-023-10179-8

Khan, F. I., & Abbasi, S. A. (1998). Techniques and methodologies for risk analysis in chemical process industries. In *Journal of Loss Prevention in the Process Industries* (Vol. 11).

Pankaj Goel, A. D. M. S. M. (2017). *Application of Big Data analytics in process safety and risk management*. IEEE. https://doi.org/10.1109/BigData.2017.8258040

Sercan¨, S., Arık, S., & Pfister, T. (2021). *TabNet: Attentive Interpretable Tabular Learning*. www.aaai.org

Trevor A. Kletz. (1999). *Hazop and Hazan Identifying and assessing process industry hazards*.

U.S. Chemical Safety and Hazard Investigation Board. (2007). *Investigation Report: Refinery Explosion and Fire (15 Killed, 180 Injured) Report No. 2005-04-I-TX*.

Zhao, C., Bhushan, M., & Venkatasubramanian, V. (2005). Phasuite: An automated HAZOP analysis tool for chemical processes: Part I: Knowledge engineering framework. *Process Safety and Environmental Protection*, *83*(6 B), 509–532. https://doi.org/10.1205/psep.04055