Emerging Risks and Energy Transition in Industrial and Port Environment

Tomaso Vairo[a*], Margherita Pettinato[a], Roberto Setola[b], Andrea P. Reverberi[c], Bruno Fabiano[a]

[a]DICCA – Civil, Chemical and Environmental Engineering Dept. – Genova University, Via Opera Pia, 15 – 16145 Genova, Italy.
[b]Unit of Automatic Control, University Campus Bio-Medico di Roma, Via Alvaro del Portillo 21, 00128, Roma, Italy.
[c]DCCI - Chemistry and Industrial Chemistry Dept., Genova University, via Dodecaneso 31 - 16146, Genova, Italy.
tomvairo@gmail.com

The ongoing transition towards more sustainable and renewable energy sources represents a multi-factorial process posing to socio-technological systems novel hazards, characterized by both epistemic and aleatory uncertainty. In the realm data data-driven approaches, historical data can provide knowledge by deriving from them low-dimensional features. Building on this concept, this paper proposes a comparative analysis of two methodologies for identifying influential factors in the assessment of emerging risk. The former approach is based on the Analytic Hierarchy Process (AHP) integrated with a fishbone diagram, while the latter relies on gradient-boosted decision trees and feature importance, integrated into the fishbone diagram framework. The study focuses on accelerating and mitigating factors associated with accidents and near misses within the context of the energy transition, elucidating insights gained from post-accidental investigations, and covering both the technical and organizational aspects. A set of accident data was collected and utilized, covering both industrial settings and urban industrial port data. The last investigated context is of peculiar interest, considering the significant responsibility to assess emerging operational risks and mitigate potential catastrophic events, mainly connected to hazardous material handling. The paper aims to offer a nuanced understanding of the strengths and limitations of the explored approaches, contributing to attaining enhanced decision-making processes in sustainable energy transitions.

1. Introduction

The global energy landscape is currently in a substantial state of transformation as nations endeavour to shift towards more sustainable and low-carbon energy sources, driven by the imperative to address climate change and reduce greenhouse gas emissions (Schischke et al., 2023). This transition introduces both opportunities and challenges for various industries. While embracing renewable energy sources yields benefits such as decreased environmental impact and enhanced energy efficiency, it simultaneously presents novel and evolving risks that necessitate effective management (Chen et al., 2023). The new technologies and processes entail new processes and personal hazards, and much effort is going into renewal, but safety analyses are still scarce and improvements in the concept of hazard identification and scenario definition are becoming possible due to the digitization of the industry and the availability of data (Pasman et al., 2023). This study focuses on comprehending the factors influencing safety, particularly in terms of accidents and near misses, amidst the ongoing energy transition. As industries adopt new energy technologies and substances, they encounter new operational risks, technological complexities, and shifts in workforce dynamics (Vairo et al., 2023a). Human factor evaluation on safety performance can reveal novel unsafe attitudes and failures in training, supervision and management, needing the enforcement of innovative of safety programs (Fabiano et al.,2022). Identifying and comprehending these factors are crucial to ensuring safety, protecting the environment, and sustaining the reliability and resilience of energy systems. The methodology employed in this study adopts a comprehensive approach to analyse and assess emerging risks associated with the energy transition. Likewise in different

contexts, the main challenge is related to data uncertainty and context interaction complexity (Fabiano et al., 2024). To address this issue, the analysis integrates quantitative data and qualitative insights obtained through a questionnaire. The central goals of this study involve scrutinizing factors that contribute to emerging risks through the utilization of two distinct techniques. The former relies on a customized regression model employed to capture system conditions that may give rise to incidents and near misses, starting from the approach outlined by Vairo et al. (2023b). The model enables correct and early identification of the precursors associated with these conditions. The latter considers the Analytic Hierarchy Process (AHP), utilized within the explored domain to pinpoint the most influential factors. Subsequently, a comparative analysis is conducted to assess and contrast the results obtained from these two methodologies. By achieving these objectives, organizations can proactively address emerging risks associated with the energy transition, ensuring the safety of their operations.

2. Materials and methods

Random Forest Regressor
The Random Forest Regressor (RFR) is a machine learning algorithm falling under the ensemble learning category (Breiman, 2001). It is particularly effective for regression tasks, which involve predicting continuous outcomes The algorithm builds multiple decision trees during the training phase and combines their predictions to enhance accuracy and reduce overfitting. The key characteristics are:

1. Ensemble of Decision Trees: The algorithm constructs a multitude of decision trees, each trained on a random subset of the dataset.
2. Random Feature Selection: Random Forest introduces randomness by selecting a subset of features for each tree, preventing individual trees from dominating the overall prediction.
3. Voting Mechanism: During prediction, the algorithm aggregates the outputs of all individual trees through a voting mechanism or averaging, providing a more robust and accurate result.
4. Robust to Overfitting: The ensemble nature of Random Forest makes it robust to overfitting, ensuring reliable predictions on new, unseen data.
5. Feature Importance: Random Forest can provide insights into feature importance, indicating which variables contribute most significantly to the model's predictions.

Within the context of each decision tree, the model computes the importance of nodes using Gini Importance (Gerstorfer et al., 2023). Gini importance helps us understand how influential features are in decision tree models. Decision trees grow by splitting the data into more specific groups based on features, as previously detailed. When the tree considers splitting a node, it calculates the so-called *Gini gain*, which is Mean Decrease Impurity, i.e. the reduction in the classification error within that node achieved by splitting on a particular feature. Figure 1 depicts the designed fishbone addressing emerging risks/challenges within the energy transition context.
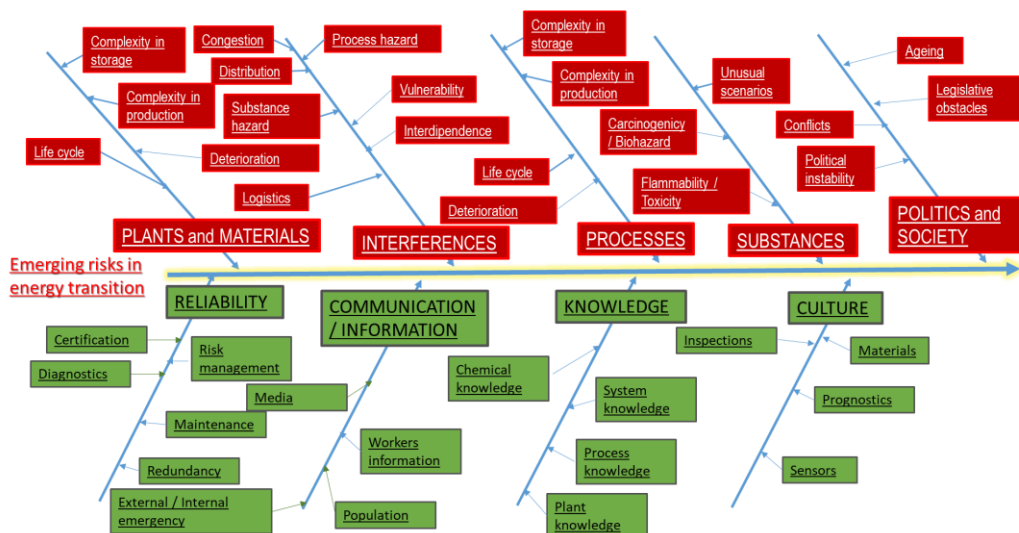

Figure 1: Ah-hoc designed fishbone covering emerging risks possibly connected to energy transition.

Analytic Hierarchy Process (AHP)
AHP is a structured decision-making technique originally developed by Saaty (1980). It helps individuals or groups to evaluate and prioritize multiple criteria, or alternatives in a systematic and quantitative manner,

evidencing its effectiveness especially when there is a need to resort to information provided by human stakeholders and take decisions based on relative preference information. The key characteristics are:

1. Pairwise Comparisons: AHP involves pairwise comparisons, where decision-makers assess the relative importance of criteria or alternatives in relation to each other.
2. Consistency Checks: AHP includes consistency checks to ensure the reliability of the decision-makers' judgments during the pairwise comparisons.
3. Eigenvalue Method: The method calculates an eigenvalue to determine the consistency ratio, aiding in assessing the reliability of the provided judgments.
4. Weighted Sum Method: Once the pairwise comparisons are completed, AHP synthesizes the judgments to calculate weights for each criterion or alternative. Subsequently obtained weights are adopted to make informed decisions and technical/strategic choices.

AHP was effectively applied in various fields such as project management, resource allocation, and strategic planning. In the context of the study, AHP is employed to identify and prioritize the most influential factors contributing to emerging risks in the energy transition.

2.3 Dataset

The dataset encompassed two main components. Firstly, it included factors and subfactors associated with the conditions of corporate organizations, as evaluated by various hierarchical levels within companies, in response to the challenges presented by the energy transition. Secondly, it incorporated data on the incidence of events such as incidents, near misses, anomalies, and non-conformities, systematically gathered by organizations by firm safety management systems. Selected factors are categorized as accelerating (i.e., the main potentially hazardous factors, referred to both already known and completely new hazards) and mitigating factors (i.e., the main factors useful to reduce the emerging hazards, referred to both already known and completely new factors), and Operative experience factors. Those factors represent the target variables and are included in the model through their probability distributions from the collected data, according to Vairo et al. (2023b).

3. Results and discussion

Regression results
A pivotal advantage offered by the Random Forest Regressor lies in its capability to assess the significance of various features in predicting the target variable.
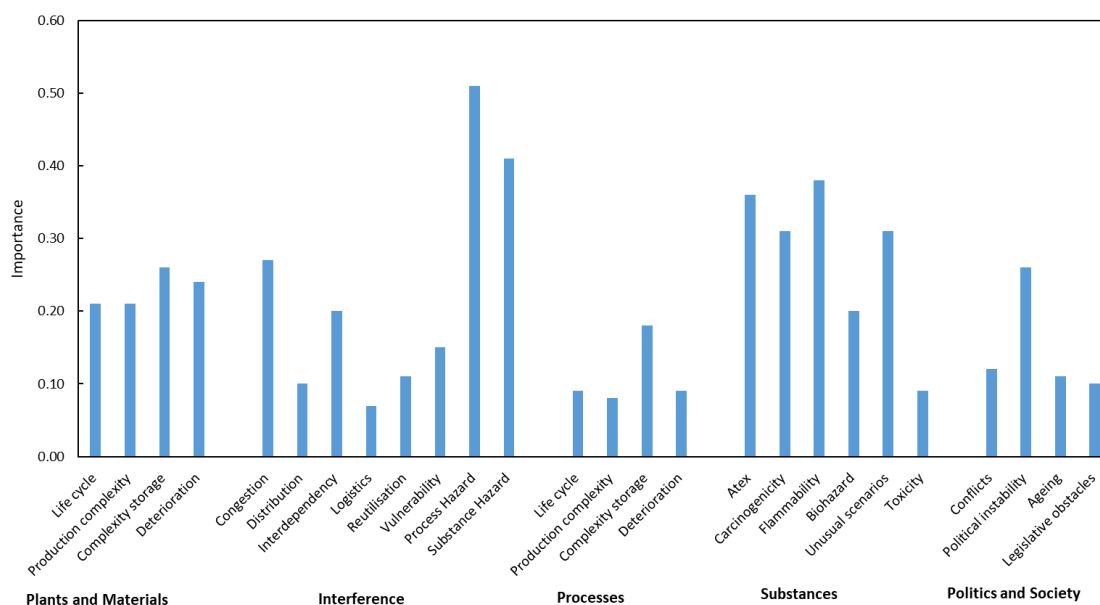

Figure 2: Accelerating factors importance calculated according to the Regression Model.
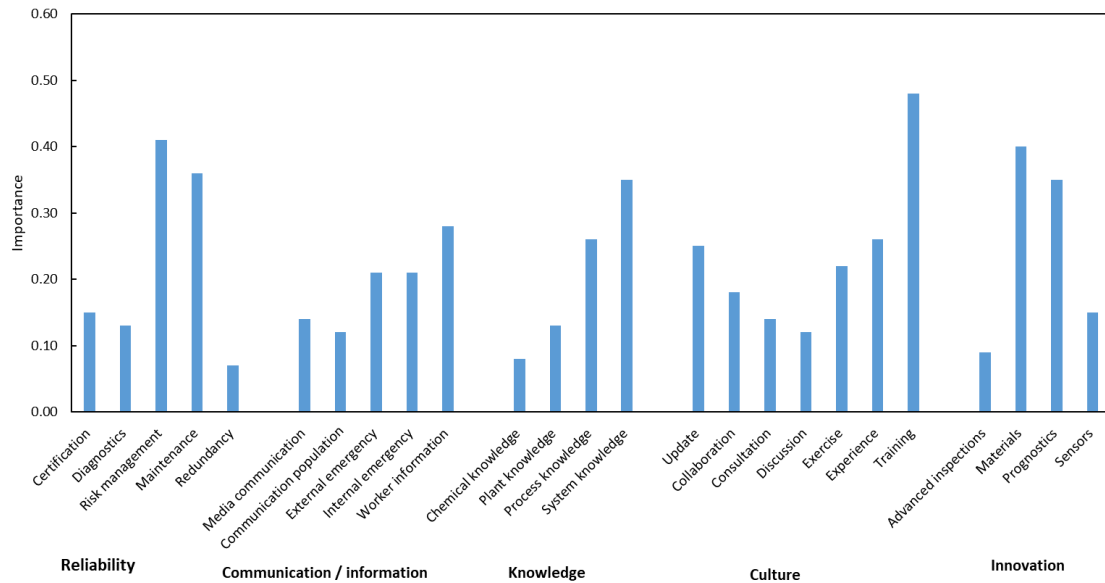
Figure 3: Mitigating factors importance according to the Regression Model.

The analysis allows pinpointing the factors and subfactors exerting the most profound influence on the occurrence of accidents and near misses. The outcomes provide an objective indication of the elements of the identified categories demanding heightened attention to effectively manage the risks associated with the energy transition. Feature importance is computed based on the total reduction of mean squared error (MSE) attributed to each feature across all decision trees within the Random Forest. These importance values are normalized to ensure their collective sum equals 1 and the feature leading to the highest Gini gain is selected for the split. Gini importance doesn't just pick the splitting features but allows measuring their overall contribution. It calculates the average Gini gain across all the splits in the tree where a particular feature was used. According to this approach, it is possible to quantify the actual feature contribution to the final classifications. In other words, by analysing Gini importance, one can assess which features have the biggest impact on selected model choices and consequently can focus attention on the most relevant ones increasing the quality of a possible solution to a problem. The feature importance from the regression model is graphically depicted in Figures 2 and 3, respectively accounting for the identified accelerating factors and mitigating factors.

AHP analysis

Decision weights and priorities are obtained from the decision maker's assessments of how each item of a decision problem compares with respect to any other item at the same hierarchy level.

Results of the analysis obtained by AHP analysis are shown in Figures 4 and 5 respectively referred to accelerating and mitigating factors. Being clear that expert opinion for hazard analysis is affected by uncertainty, as commented by Baybutt (2017) different relevant cognitive biases may affect the analysis team and require proper attenuation measures.

From the obtained results this facet is evident, as there are some accordance and some differences between the two explored approaches. The most relevant accelerating factor, determined by the Regression model, is Process Hazard compared with Deterioration identified by AHP. The most relevant mitigating factor, determined by the Regression model, is Training, compared to System knowledge identified as the most relevant one by applying AHP.

The accordance and differences can be explained by the different fallacies inherent to the two methodologies.

As widely acknowledged, pairwise comparison by AHP allows facing problems like e.g. decision support, multi-criteria decision making and estimating subjective probabilities estimate of future events, but may encounter problems related to data sensitivity issues, as discussed by Huang (2002). As previously reported, Gini feature importance helps to overcome and mitigate the biases in identifying feature importance summarized as follows.
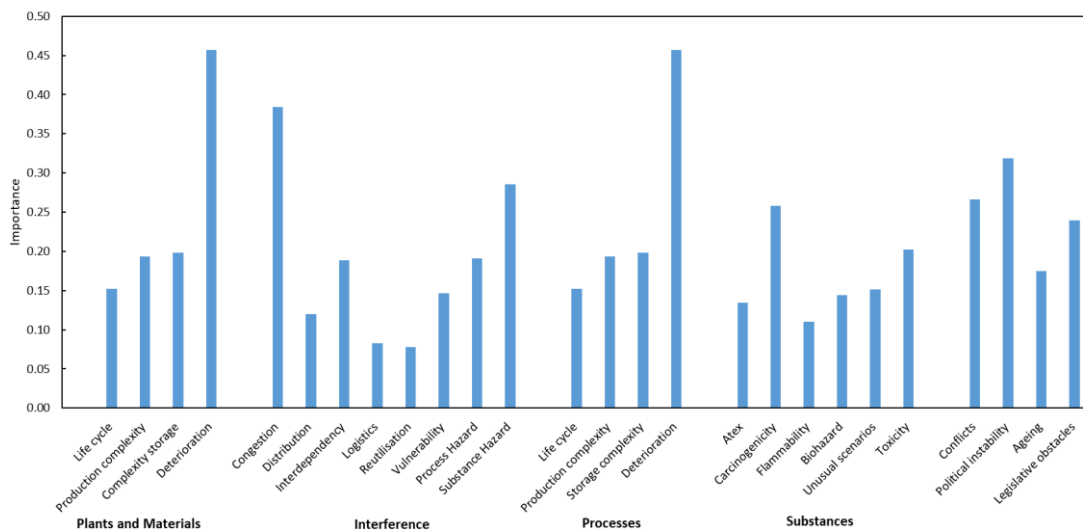
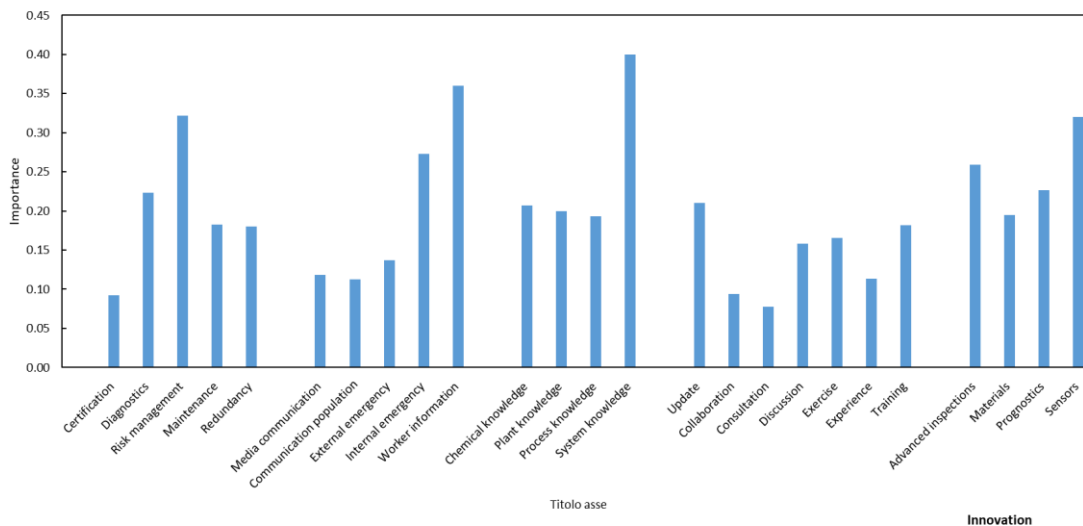Figure 4: Accelerating factors importance obtained by AHP.

Figure 5: Mitigating factors importance obtained by AHP.

- *Reduced Subjectivity.*
  AHP relies on human experts to judge feature importance. The resulting figure can be subjective, influenced by the experts' background, experience, and even personal biases. Gini feature importance uses objective metrics based on the data itself leading to more consistent results.
- *Automatic Assessment.*
  AHP requires dedicated effort and expertise from humans to assess each feature, making it time-consuming and potentially expensive. Gini feature importance is automatically calculated during the model-building process, saving time and resources.
- *Data-Driven Insights.*
  AHP may overlook hidden patterns or relationships not readily apparent to the experts, leading to missed insights. Gini feature importance captures the impact of features based on the actual data, potentially revealing hidden relationships that might not be obvious to human experts.

Nevertheless, assessment of Gini feature importance exhibits a number of inherent limitations, i.e.:

- *Bias towards high cardinality features.*
  Features with many categories can exhibit higher Gini feature importance, simply due to having more opportunities for splitting, not necessarily because they are inherently more important.
- *No consideration is given to external knowledge.*

Gini feature importance might not capture the importance of features based on real-world knowledge, which is not an inherent facet of the data.

On these grounds, it is suggested to adopt a combined approach, to combine Gini feature importance with other methods, like permutation importance, or domain knowledge to validate and refine the insights from the feature importance and increase considerably the quality of a solution to a given problem or analysis.

4. Conclusions

As a concluding remark, even though both the AHP and RFR exhibit certain limitations, their distinct strengths make them valuable tools in different aspects of the decision-making process. Recognizing the advantages and shortcomings of each methodology allows for a more comprehensive and robust decision-making strategy. The structured and systematic AHP approach to pairwise comparisons provides a clear framework for eliciting preferences and prioritizing criteria, but possible biases must be considered. Despite potential sensitivity to input values and concerns about rank reversal, AHP remains a widely used method for decision analysis, particularly in situations where human judgment and qualitative factors play a crucial role. RFR, with its ability to handle complex relationships in large datasets and provide insights into feature importance, offers a data-driven approach to decision support. The algorithm ability to manage non-linearity and capture intricate patterns makes it particularly effective in identifying the precursors of undesired situations. Although challenges such as sensitivity to hyperparameters exist, it is worth noting that many of the flaws in the regression model are more easily addressed through careful tuning, regularization techniques, and feature engineering. Given the complementary strengths of AHP and RFR, a pragmatic and effective approach to strengthen the decision-making process in real-world applications involves leveraging both techniques synergistically to generate rankings that are not in contrast with the available information. AHP can be employed in the early stages of decision formulation to structure preferences and establish a clear hierarchy of criteria and the consistency level in decision-making, while RFR finds suitable and practical application for predictive modelling and understanding the importance of features within complex systems.

Acknowledgments

References

Baybutt P., 2017, The validity of engineering judgment and expert opinion in hazard and risk analysis: the influence of cognitive biases, Process Safety Progress, 37(2), 205–210.

Breiman L., 2001, Random Forests, Machine Learning, 45, 5–32.

Chen W., Zou W., Zhong K., Aliyeva A., 2023, Machine learning assessment under the development of green technology innovation: A perspective of energy transition, Renewable Energy, 214, 65-73.

Dyer J.S., 1999, Remarks on the Analytic Hierarchy Process, Management Science, 36, 249–258.

Fabiano B., Guastaferro M., Pettinato M., Pasman H.J., 2024, Towards strengthening resilience of organizations by risk management tools: A scientometric perspective on COVID-19 experience in a healthcare and industrial setting. Canadian Journal of Chemical Engineering 102, 1705-1725.

Fabiano B., Pettinato M., Currò F., Reverberi A.P., 2022, A field study on human factor and safety performances in a downstream oil industry, Safety Science, 153, 105795.

Gerstorfer Y., Hahn-Klimroth M., Krieg L., 2023, A Notion of Feature Importance by Decorrelation and Detection of Trends by Random Forest Regression, Data Science Journal, 22,1-14.

Huang Y.-F., 2002, Enhancement on sensitivity analysis of priority in analytic hierarchy process. International Journal of General Systems 31, 531–542.

Pasman H., Sripaul E., Khan F., Fabiano B., 2023, Energy transition technology comes with new process safety challenges and risks, Process Safety and Environmental Protection, 177, 765-794.

Saaty T.L., 1980, The Analytic Hierarchy Process, McGraw-Hill, New York, US.

Schischke A., Papenfuß P., Brem M., Kurz P., Rathgeber A.W., 2023, Sustainable energy transition and its demand for scarce resources: Insights into the German Energiewende through a new risk assessment framework, Renewable and Sustainable Energy Reviews, 176, 113190.

Vairo T., Cademartori D., Clematis D., Carpanese M.P., Fabiano B., 2023a, Solid oxide fuel cells for shipping: A machine learning model for early detection of hazardous system deviations, Process Safety and Environmental Protection, 172, 184-194.

Vairo T., Agnello P., Currò F., Fabiano B., 2023b, Unveiling the Achilles' heel: detecting organizational weaknesses in the energetic transition challenge, Chemical Engineering Transactions, 105, 499-504.