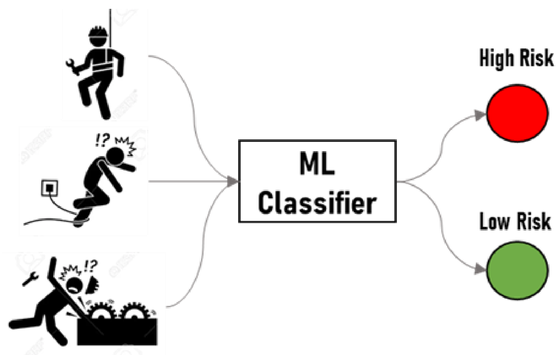
A data analytics framework for identifying and characterizing incident clusters in the diagnosis of risk factors in construction

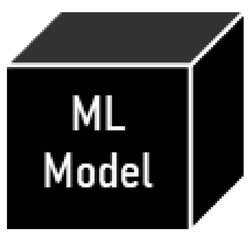
1st Aswin Ramaswamy Govindan[1], 2nd Xinming Li[2]
[1]PhD Student, Department of Mechanical Engineering, University of Alberta,
[2]Assistant Professor, Department of Mechanical Engineering, University of Alberta

Introduction

In construction, workplace hazards can cause injuries, costly workers' compensation claims, and harm productivity and product quality. To tackle these problems, it is crucial to create a machine learning based risk diagnosis framework that can identify and comprehend the often hidden and unclear occupational risk factors.



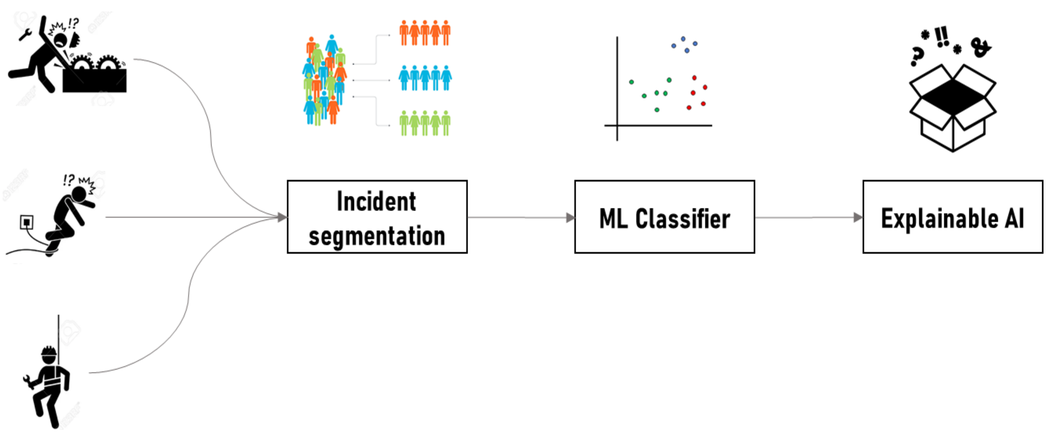
Label dependent Risk Classifiers



Black-box ML models

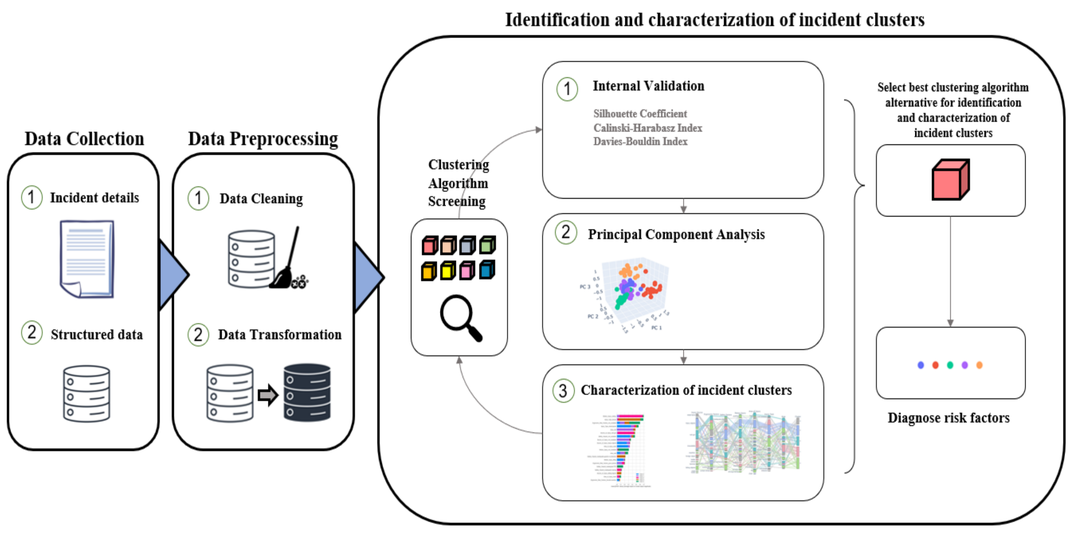
Objective & Framework

To address the issues, this project proposes an interpretable ML-based occupational risk diagnosis framework that does not require pre-labeled datasets.



Simplified Occupational Risk Diagnosis Framework

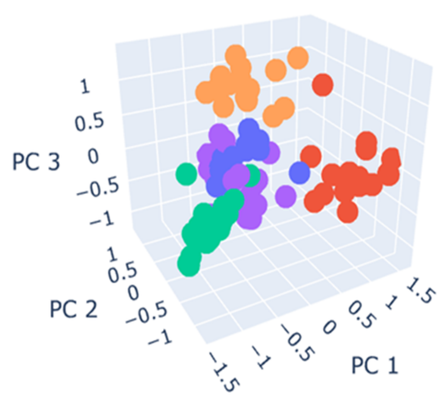
The framework consists of three modules: (1) data collection, where incident details are gathered and organized; (2) data preprocessing, which prepares the data for analysis; and (3) identification and characterization of incident clusters. This involves labeling unlabeled incident data using clustering algorithms and interpreting the clusters using Shapley Additive Explanations (SHAP) to diagnose risk factors.



Detailed Occupational Risk Diagnosis Framework

Validation & Results

The framework was tested in a modular building manufacturing facility and proved to be highly effective in diagnosing risk factors by identifying and characterizing incident clusters. By using an XGBoost classifier on the clustered data, we achieved an impressive F1 score of 0.92, which confirms the reliability and accuracy of the SHAP values used for risk diagnosis.



Cluster 0
Inadequate Training related Nailing Injuries

Cluster 1
Roof-related Ergo-Safety Injuries

Cluster 2
WMSD injuries from heavy lifting, back strain/sprain

Cluster 3
Machine Guard Injuries and Bruising-Related Injuries

Cluster 4
Roof-related injuries due to insufficient PPE

Occupational Risk Diagnosis

Conclusion

In conclusion, our interpretable machine learning-based risk diagnosis framework is a valuable tool for understanding and interpreting risk diagnosis, particularly when there is a lack of pre-labeled incident data.

Interpretable Risk Diagnosis

Label-free Risk Diagnosis

Robust Risk Profiling

Furthermore, the risks workers encounter in the construction industry can lead to injuries, higher expenses, and lower productivity. However, our framework enables businesses to recognize various groups of worker incident risks and comprehend the traits of each group. This understanding empowers businesses to take proactive steps to effectively lessen incident risks and enhance worker safety. By doing this, businesses can minimize costs, and improve productivity. In summary, our framework has the potential to stimulate growth in the construction industry.

Acknowledgement

NSERC CRSNG
ALBERTA INNOVATES