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
 Published by Research Publishing, Singapore.
doi: 10.3850/978-981-94-3281-3_ESREL-SRA-E2025-P6052-cd

Improving classification of imbalanced classes in industrial data: Enhancing defect detection for type IV high-pressure hydrogen vessels

Lina Achour
Roberval, University of Technology of Compiègne, France. E-mail: lina.achour@utc.fr

Salvador Madrigal Castillo
Heudiasyc, University of Technology of Compiègne, France. E-mail: salvador.madrigal-castillo@hds.utc.fr

Wenlong Chen
Heudiasyc, University of Technology of Compiègne, France. E-mail: wenlong.chen@etu.utc.fr

Sebastien Destercke
Heudiasyc, University of Technology of Compiègne, France. E-mail: sebastien.destercke@hds.utc.fr

Zoheir Aboura
Roberval, University of Technology of Compiègne, France. E-mail: zoheir.aboura@utc.fr

Kamel Khellil
Roberval, University of Technology of Compiègne, France. E-mail: kamel.khellil@utc.fr

This study focuses on enhancing defect detection in Type IV high-pressure hydrogen vessels (HPV) using advanced machine learning techniques. Type IV vessels, crucial for hydrogen storage and transportation in industrial and automotive sectors, feature composite materials with polymer liners to ensure mechanical strength and hydrogen tightness at pressures up to 700 bar. Detecting defects in these vessels is critical for maintaining structural integrity and safety. The primary challenge addressed is the significant class imbalance among defect types, where critical defects are infrequent compared to minor defects or defect-free instances. Standard classification models often fail to effectively detect these critical defects due to their bias towards majority classes. To overcome this, we evaluated and compared methods including Weighted Logistic Regression, Weighted Decision Trees, and the Discrete Minimax Classifier (DMC), which adapt their strategies to improve minority class detection. Our findings demonstrate that these adapted algorithms enhance sensitivity to minority classes, particularly in reducing false negatives for critical defect detection. This study emphasizes the importance of tailored machine learning approaches in industrial defect detection, paving the way for safer and more reliable hydrogen storage technologies through optimized predictive maintenance and quality control processes. In particular, we include in our study the recently studied minimax approaches, which have strong theoretical foundations and good empirical performances, as show our experiments.

Keywords: Classification, imbalanced classes, high-pressure type IV hydrogen vessels, DMC, weighted algorithms, defect detection

1. Introduction

1.1. *Context*

Type IV high-pressure hydrogen vessels (HPV) are critical components in hydrogen storage and transportation systems, particularly for industrial applications and hydrogen-powered vehicles. These vessels are made of composites (primarily carbon fibers) with a polymer liner (thermoplastic). The liner in this case does not support the load but ensures hydrogen tightness. The properties of composite materials significantly extend the vessel's fatigue life and ensure its mechanical strength, while their low density reduces the overall mass of the assembly. This combination enables achieving high pressure levels (700 bar). Currently, Type IV hydrogen storage represents the most advanced technology Baramee (2014).

Detecting defects during the manufacturing and testing of Type IV hydrogen vessels is of paramount importance. Even minor defects can compromise the structural integrity of the vessel, leading to potential safety risks. Therefore, implementing effective defect detection systems is essential to ensure that each vessel meets quality and safety standards before use. To improve these detection systems, we propose utilizing machine learning algorithms. By leveraging the power of machine learning, we can enhance the accuracy and efficiency of defect detection, thereby reducing failure risks, increasing vessel durability, and boosting confidence in hydrogen storage technologies.

In the context of defect detection, a major issue is the significant imbalance between defect classes. Serious or critical defects are often much less frequent than minor defects or defect-free instances. This asymmetry in data distribution complicates the task of classification algorithms, as they tend to favor majority classes, thereby neglecting critical defects that, although rare, are of utmost importance for the safety and reliability of vessels Haibo and A (2009).

This class imbalance leads to significant challenges in analyzing and predicting test outcomes. Standard classification models may demonstrate satisfactory overall performance while failing to effectively detect rare but critical defects. This presents a major challenge for engineers and researchers striving to develop precise and reliable defect detection systems. Therefore, adopting specific methods to address this imbalance and enhance model sensitivity to minority defect classes is crucial.

Machine learning algorithms typically aim to minimize overall classification error. However, in the presence of imbalanced classes, they tend to prioritize the majority class at the expense of the minority class, resulting in significant bias Le et al. (2021) Arvind et al. (2022). This translates into poor prediction performance for the minority class, which is often the most crucial to detect (defects, risks, hazards, etc.).

1.2. *Objective*

The main objective of this study is to improve the accuracy and reliability of defect predictions in Type IV hydrogen vessels, taking into account the asymmetry in data distribution. To achieve this, we have tested and compared various existing methods to enhance the performance of detection systems. By focusing on minimax optimization to improve classification algorithms in imbalanced data contexts, we aim to identify the most effective approaches. By adapting our methodologies to the specific nature of industrial data and integrating advanced machine learning techniques, our goal is to provide comprehensive solutions capable of effectively managing the inherent complexities in industrial data analysis.

2. Related work

2.1. *Defect Detection in Industrial Data - A Recap on Classification*

Classification Murphy (2012) (pages 3-8) is a widely used method in machine learning, involving learning a mapping from inputs X to outputs y, where $y \in \{1, \ldots, C\}$ and C represents the number of classes. If $C = 2$, it is referred to as binary classification (often with $y \in \{0, 1\}$); for $C > 2$, it is called multi-class classification. When an observation can belong to multiple classes simultaneously, it is termed multi-label classification. We will not consider this last problem here, as we consider only one possible defect with different degrees of severity.

Formally, we assume there exists an unknown function f such that $y = f(X)$. The goal of learning is to estimate this function f from a labeled training set (X_i, Y_i), and make predictions on new inputs X_{new} using $\hat{y} = \hat{f}(X_{\text{new}})$, where $\hat{f}$ denotes the estimated f.

Defect detection in industrial data is often framed as a classification problem, where each observation X_i is assigned to a class. Mathematically, this can be formalized by the decision function $f(X_i)$ that assigns a class y_i to each observation based on its extracted features X_i. However, industrial data often exhibits complex variations and noise, making classification challenging. Defects can vary in nature, requiring robust classification models capable of effective generalization on real data while minimizing classification errors.

Classification is perhaps the most prevalent form of machine learning and has been utilized to solve many real-world problems, both interesting and often complex. Data process classification techniques encompass various algorithmic methods such as decision trees, random forests, k-nearest neighbors (KNN) classifiers Murphy (2012) (pages 16-18), and logistic regression Bishop (2006) (pages 205-206).

2.2. *Class imbalance*

Methods for handling class imbalance in classification include several techniques to balance datasets. Over-sampling techniques, such as SMOTE (Synthetic Minority Over-sampling Technique), create synthetic examples by interpolating between instances of the minority class, while ADASYN (Adaptive Synthetic Sampling Approach) generates more examples in areas where classes overlap. Under-sampling methods, like Random Under-sampling, reduce the size of the majority class by randomly removing instances, and Tomek Links remove ambiguous examples. Weighted methods adjust cost functions to give more importance to minority classes. Specific algorithms, such as weighted random forests, adaptive gradient boosting, weighted Logistic Regression and weighted decision trees adapt their learning based on the class imbalance. Hybrid techniques, such as SMOTE combined with Tomek Links or ensemble methods, enhance model performance on imbalanced data by reducing classification errors and improving generalization capability. The Discrete Box-Constrained Minimax Classifier is another approach designed to handle imbalanced datasets and uncertain class proportions. It optimizes classification by balancing class-conditional risks Gilet et al. (2020).

This last recent technique appears particularly interesting, as it is relatively easy to use while benefiting from the strong theoretical apparatus coming with minimax approaches. In particular, sampling techniques such as SMOTE do not come with any theoretical guarantees (we therefore do not explore them in details), and weight-based approaches only provide such guarantees for the binary case. Ths will be observed in our experiments.

3. Methodology

In this section, we describe the various methods used to address the issue of class imbalance in defect detection. We explored several approaches, including techniques without class balancing and methods more suited to imbalanced data.

We began by applying standard classification techniques without balancing the classes to evaluate their baseline performance. Among these methods, we used Logistic Regression Peter (2019), a linear classification model that estimates the probability of an observation belonging to a given class. Although simple, this model tends to be biased toward the majority classes due to data asymmetry.

Next, to manage data asymmetry, we considered the anomaly detector using a Discrete Minimax Classifier (DMC) Gilet and Fillatre (2021) (pages 179-187), Gilet et al. (2020) for imbalanced data. This supervised classification method is particularly useful for monitoring the condition of manufactured parts, such as predicting failures from measurements like process parameters. The minimax classifier aims to minimize the maximum conditional risk per class, making the decision rule robust against variations in class proportions.

We also tested Weighted Logistic Regression, which adjusts sample weights based on class distributions, thereby enhancing sensitivity to minority classes. Additionally, we evaluated Weighted Decision Trees, which adapt their splitting criteria to achieve more balanced predictions across all classes, including minorities.

We applied these algorithms to real manufacturing process monitoring data from 12 high-pressure hydrogen vessels. The objective of this study is to predict potential defects in the composite structure of the vessels. The database was segmented into three classes: class 0 represents a null or near-zero defect level in the composite, class 1 corresponds to a moderate defect level, and class 2 to a very high defect level. The segmentation thresholds were defined by experts.

After preprocessing, the database contains 135 measurements in class 0, 36 measurements in class 1, and 9 measurements in class 2. Thus, the class proportions $\pi = [0.75, 0.2, 0.05]$ of the database are highly imbalanced, complicating defect prediction. Each sample is described by 8 anonymized numerical and categorical features. For this database, we employed cross-validation, with 10% of the data used as a test set, to ensure robust evaluation of the model's performance.

Initially, we performed classification with the three initial classes. We compared the performance of four classification algorithms: Logistic Regression, weighted Logistic Regression, weighted decision trees, and DMC. The overall accuracies for these algorithms are 0.74, 0.5, 0.67, and 0.63, respectively, but these values do not fully reflect the efficiency for detecting minority classes. Confusion matrices obtained for each method in this configuration are given in Figure 1

Predicted class

$$\text{True class}\quad \begin{array}{cc} & \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{array}{c} 0\\1\\2 \end{array} & \begin{pmatrix} 0.94 & 0.06 & 0 \\ 0.79 & 0.20 & 0 \\ 1 & 0 & 0 \end{pmatrix} \end{array}$$

Logistic regression (LR)

Predicted class

$$\text{True class}\quad \begin{array}{cc} & \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{array}{c} 0\\1\\2 \end{array} & \begin{pmatrix} 0.50 & 0.27 & 0.24 \\ 0.26 & 0.62 & 0.12 \\ 0.70 & 0.27 & 0.03 \end{pmatrix} \end{array}$$

Weighted LR

Predicted class

$$\text{True class}\quad \begin{array}{cc} & \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{array}{c} 0\\1\\2 \end{array} & \begin{pmatrix} 0.77 & 0.10 & 0.13 \\ 0.38 & 0.46 & 0.16 \\ 0.63 & 0.23 & 0.14 \end{pmatrix} \end{array}$$

Weighted tree

Predicted class

$$\text{True class}\quad \begin{array}{cc} & \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{array}{c} 0\\1\\2 \end{array} & \begin{pmatrix} 0.70 & 0.23 & 0.07 \\ 0.48 & 0.48 & 0.04 \\ 0.58 & 0.26 & 0.17 \end{pmatrix} \end{array}$$

Discrete minimax

Fig. 1. 3 class confusion matrices

Logistic Regression demonstrates high overall accuracy (0.74), but its confusion matrix reveals limited performance in detecting minority classes, with high false negative rates for classes 1 and 2 frequently misclassified as class 0. It is particularly striking for the most imbalanced default class (class 2), for which all predictions are of class 0. In contrast, Weighted Logistic Regression adopts a balanced approach. Despite lower overall accuracy (0.50), it significantly improves the detection of minority classes, reducing false negatives and enhancing recall for these classes. It however remains quite poor for the most important default (only 0.03 good recognitions). Weighted Decision Trees strike a middle ground with an overall accuracy of 0.67, showing improved performance in detecting classes 1 and 2 compared to standard Logistic Regression, albeit with some misclassification remaining. The Discrete Minimax Classifier achieves an overall accuracy of 0.63, with its confusion matrix indicating notable improvement in detecting minority classes despite residual errors. It is also the best on the most important default (class 2). These results highlight the importance of using adapted classification algorithms and weighting techniques to enhance the detection of minority classes in imbalanced datasets. They demonstrate the potential for these methods to provide a more equitable distribution of correct predictions across all classes. Yet, their performance on very poorly represented classes remains limited, which is notably due to the combination of imbalance and a data set of limited size, which severely limits the possibility of learning from the less represented class. The accuracy per class for each tested algorithm is shown in Figure 2.

To improve the robustness of our model and increase the number of samples in the minority classes, we merged classes 1 and 2 into a single class, 'Class 1.' This new classification consists of: Class 0 (no defect or near-zero defect level) and Class 1 (presence of defects at a moderate or high level). After this merger, the classification algorithms were reapplied, and the confusion matrices as well as overall accuracies were recalculated: Logistic Regression at 0.74, weighted Logistic Regression at 0.64, weighted decision trees at 0.74, and DMC at 0.64. The confusion matrices

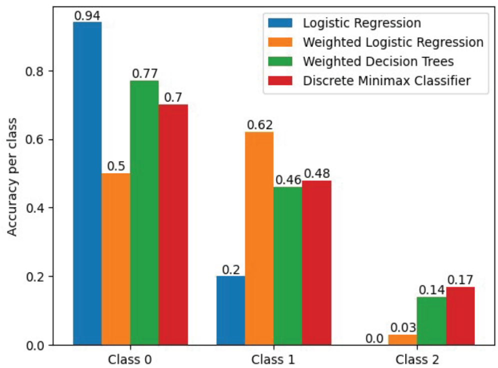

Fig. 2. Accuracy per class for each method with three classes

obtained for each method in this configuration are given in Figure 3.

$$\begin{array}{cc} & \text{Predicted class} \\ & \begin{array}{cc} 0 & 1 \end{array} \\ \text{True class} \begin{array}{c} 0 \\ 1 \end{array} & \begin{pmatrix} 0.93 & 0.07 \\ 0.83 & 0.17 \end{pmatrix} \end{array} \qquad \begin{array}{cc} & \text{Predicted class} \\ & \begin{array}{cc} 0 & 1 \end{array} \\ \text{True class} \begin{array}{c} 0 \\ 1 \end{array} & \begin{pmatrix} 0.64 & 0.36 \\ 0.37 & 0.63 \end{pmatrix} \end{array}$$

Logistic regression (LR) **Weighted LR**

$$\begin{array}{cc} & \text{Predicted class} \\ & \begin{array}{cc} 0 & 1 \end{array} \\ \text{True class} \begin{array}{c} 0 \\ 1 \end{array} & \begin{pmatrix} 0.83 & 0.17 \\ 0.53 & 0.47 \end{pmatrix} \end{array} \qquad \begin{array}{cc} & \text{Predicted class} \\ & \begin{array}{cc} 0 & 1 \end{array} \\ \text{True class} \begin{array}{c} 0 \\ 1 \end{array} & \begin{pmatrix} 0.64 & 0.36 \\ 0.38 & 0.62 \end{pmatrix} \end{array}$$

Weighted tree **Discrete minimax**

Fig. 3. 2 class confusion matrices

The Logistic Regression demonstrates robust overall accuracy but reveals substantial limitations in minority class detection, notably Class 1, where an 83% false negative rate suggests a strong bias towards the majority class. This observation aligns with logistic regression's tendency to optimize predictions based on dominant class frequencies, thereby compromising sensitivity to minority class instances. In contrast, Weighted Logistic Regression and Discrete Minimax Classifier present adaptive strategies that prioritize mitigating the impact of class imbalance. Weighted Logistic Regression achieves a lower overall accuracy yet significantly reduces false negatives for Class 1 to 37%, highlighting its effectiveness in enhancing minority class detection through strategic weighting adjustments. Similarly, Discrete Minimax Classifier shows notable improvements with a false negative rate of 38% for Class 1, indicating its capability to balance prediction accuracy across diverse class distributions. Weighted Decision Trees exhibit a balanced approach with a high overall accuracy, yet the confusion matrix reveals challenges in achieving parity in minority class detection compared to the weighted approaches. This suggests potential avenues for optimizing decision tree methodologies to better address class imbalances and enhance sensitivity to minority classes. The accuracy per class for each tested algorithm is shown in Figure 4.

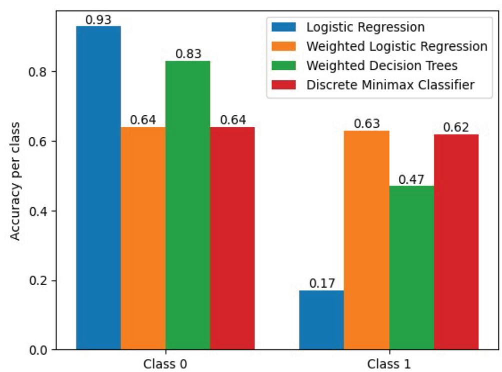

Fig. 4. Accuracy per class for each method with two classes

4. Conclusion

In conclusion, our study highlights the critical importance of accurately detecting defects in Type IV hydrogen vessels, requiring customized approaches to overcome the significant imbalance between defect classes. The application of the Discrete Minimax Classifier (DMC) significantly improved the ability to identify critical defects, both in the cases of three and two classes, thereby ensuring better safety and reliability of hydrogen storage systems. These findings encourage further exploration of algorithmic applications in other industrial contexts for the detection of various types of defects and emphasize the need for extensive data collection to enrich and generalize the models. Looking ahead, the continuous integration of advanced machine learning methods in quality control processes promises to optimize monitoring and preventive maintenance, thereby enhancing confidence in cutting-edge technologies such as high-pressure hydrogen vessels. Our study has also emphasized the critical need in having a sufficient number of representatives of default classes, as even advanced techniques such as minimax struggle to recognize classes having only a handful of representatives. A possibility would be to include more severe cost for the misclassification of such defect, yet this would also degrade our overall accuracy.

Acknowledgement

We thank Cyprien GILET for his help and advises on this project, David RUGGI for having shared the data about default detections of hydrogen vessels, and Mathieu JONGBLOEDT for his management of the project at OPmobility.

References

Arvind, K., G. Shivani, S. Nishant, and B. Arpit (2022). A review on unbalanced data classification. In *Proceedings of International Joint Conference on Advances in Computational Intelligence: IJCACI 2021*, pp. 197–208. Springer.

Baramee, P. (2014). *Conception et durabilité de réservoirs en composites destinés au stockage de l'hydrogène*. Ph. D. thesis, Ecole Nationale Supérieure des Mines de Paris.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Gilet, C., S. Barbosa, and L. Fillatre (2020). Discrete box-constrained minimax classifier for uncertain and imbalanced class proportions. *IEEE Transactions on Pattern Analysis and Machine Intelligence 44*(6), 2923–2937.

Gilet, C. and L. Fillatre (2021). Anomaly detection with discrete minimax classifier for imbalanced dataset or uncertain class proportions. In *Advances in Condition Monitoring and Structural Health Monitoring: WCCM 2019*, pp. 179–187. Springer.

Haibo, H. and G. E. A (2009). Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering 21*(9), 1263–1284.

Le, W., H. Meng, L. Xiaojuan, Z. Ni, and C. Haodong (2021). Review of classification methods on unbalanced data sets. *Ieee Access 9*, 64606–64628.

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. The MIT Press.

Peter, M. (2019). *Generalized linear models*. Routledge.