

## 

## 

## 

## 

## 

## 

## 

## 

## 

## **1. Introduction**

In the maritime industry, accurate classification of vessel deficiencies is paramount for ensuring safety and compliance with international standards. Multiple Subject Matter Experts (SMEs) often annotate deficiencies with varying severity levels. These discrepancies necessitate a systematic methodology to derive a consensus severity that prioritizes safety while accounting for expert opinions and contextual factors.

This paper presents a structured methodology for determining consensus severity, incorporating a weighted point system based on three primary factors: the content of the deficiency text (def\_text), the vessel type (VesselGroup), and the vessel’s age. Additionally, the methodology integrates plurality voting and industry-standard assumptions to ensure the system is aligned with real-world maritime safety practices.

## **2. Methodology**

### **2.2. Weighted Point System**

To resolve ties and ambiguities, a weighted point system was developed, assigning scores based on three key factors:

#### **2.2.1. Deficiency Text (def\_text)**

Using a TF-IDF (Term Frequency-Inverse Document Frequency) vectorizer, 50 keywords were extracted from the def\_text column to identify significant terms associated with each severity level:

* **High**: Keywords like "fracture," "leakage," and "corrosion" were strongly correlated with high-risk deficiencies.
* **Medium**: Keywords like "deviation" and "repair" indicated moderate risk.
* **Low**: Keywords like "routine" or "monitoring" were associated with minor issues.
* **Not a Deficiency**: Terms like "not applicable" or "non-critical" correlated with no actionable deficiencies.

For each keyword found in the def\_text, severity-specific points were added:

* High: +5 points
* Medium: +3 points
* Low: +1 point
* Not a Deficiency: +0 points
* Null (omitted)

#### **2.2.2. Vessel Type (VesselGroup)**

Based on domain knowledge, vessel types were scored to reflect their associated risk levels:

* Tanker: +10 (high-risk due to hazardous cargo)
* Gas Carrier: +9
* Container: +8
* Dry Bulk: +7
* Passenger: +6
* General Cargo: +5
* Fishing: +4
* Others: +3

#### **2.2.3. Vessel Age**

A vessel’s age significantly impacts its likelihood of deficiency. Scoring was distributed as follows:

* 0-5 years: +2 points (newer vessels are less prone to issues)
* 6-10 years: +4 points
* 11-15 years: +6 points
* 16-20 years: +8 points
* 21-30 years: +10 points (peak risk due to wear and tear)
* 31-40 years: +8 points
* 41-50 years: +5 points
* 50 years: +2 points (reduced activity diminishes risk).

### **2.3 Tie-Breaking Process**

#### **2.3.1. Plurality Voting**

* SME annotations were grouped by PscInspectionId and deficiency\_code.
* Plurality (most votes) was prioritized over majority (more than 50%) as per maritime industry standards.
* If one severity level had the most votes, it was directly assigned as the consensus severity.

#### **2.3.2. Point-Based Scoring**

In cases where votes were tied or evenly distributed, the weighted point system was applied:

1. Scores from def\_text keywords, VesselGroup, and age were combined for each severity level.
2. The severity with the highest total score was selected.
3. Example:
   * def\_text: "Leakage in hull structure" (High = +5, Medium = +3, Low = +0).
   * VesselGroup: Tanker (High = +10).
   * age: 12 years (High = +6).
   * Total Scores: High = 21, Medium = 6, Low = 0 → **Consensus: High**.

### **2.4 Addressing Special Cases**

#### **2.4.1. Fewer or More Than 3 SMEs**

* For fewer than 3 SMEs, all votes were considered, and plurality determined the outcome.
* For more than 3 SMEs, the plurality rule applied unless there was a tie, where the point-based system was used.

#### **2.4.2. Inclusion of "Not a Deficiency"**

* "Not a Deficiency" was treated as a valid severity level but assigned 0 points in the point-based system.
* This ensured that true deficiencies were prioritized over non-critical cases.

## **3. Assumptions**

1. **Risk Minimization Principle**: Safety-critical outcomes (e.g., High severity) were prioritized.
2. **Uniform Weight Distribution**: Keywords, vessel types, and age were weighted proportionally to reflect their respective contributions to risk.
3. **Data Integrity**: Missing or null values were excluded to avoid bias.
4. **Domain Knowledge**: Scoring for vessel types and age ranges was based on maritime industry standards and best practices.

## **4. Evaluation**

### **4.1. Strengths**

* **Fairness**: Combines plurality voting with a point-based system to balance SME opinions and contextual evidence.
* **Robustness**: Handles diverse scenarios, including ties and varying SME counts.
* **Scalability**: Applicable to large datasets and adaptable to other safety-critical industries.

### **4.2. Limitations**

* **Subjectivity in Scoring**: The weighting of factors (e.g., age points) may require further validation.
* **Data Dependence**: Results are influenced by dataset quality and completeness.
* **Keyword Generalization**: TF-IDF keywords may miss nuanced phrases not captured in the top 50 terms.

## **5. Conclusion**

The proposed methodology effectively integrates SME opinions, statistical insights, and domain knowledge to determine consensus severity for vessel deficiencies. By prioritizing plurality voting and applying a weighted point system, the approach ensures fairness, accuracy, and alignment with maritime safety standards. Future iterations could enhance keyword extraction and incorporate machine learning techniques for greater precision.