

PART 2 – Cross Dataset Engineering Tasks

Task 11: Outlier Materials Identification

May 1, 2025

Abstract

This report details the analysis of engineering material outliers based on Task 11 (Outlier Materials Identification). The study identifies materials with extreme property combinations—particularly those with high strength but low ductility—and examines inconsistencies in hardness measurements. These findings provide crucial insights for engineering design decisions, highlighting materials that may offer exceptional performance for specific applications while posing potential risks in others. The analysis also reveals important data quality considerations that engineers must account for when making material selection decisions based on this dataset.

1 Introduction

The Significance of Material Outliers

Identifying material outliers is essential for both practical engineering applications and data quality assessment. Materials with extreme property combinations may represent either specialized high-performance options or potential data anomalies requiring verification. This analysis focuses on detecting and characterizing materials that exhibit unusual property combinations—particularly those with exceptional strength coupled with concerning brittleness—as well as materials with inconsistent hardness measurements across different testing methodologies. Understanding these outliers enables engineers to make more informed material selection decisions while recognizing dataset limitations.

2 Material Selection Insights

2.1 High Strength but Low Ductility Materials

The analysis identified 29 materials that combine exceptionally high strength (top 10% of ultimate tensile strength) with very low ductility (bottom 10% of elongation), representing critical outliers in the strength-ductility trade-off relationship.

Table 1: Notable High-Strength, Low-Ductility Material Outliers

Material	Su (MPa)	A5 (%)	Su/A5 Ratio	Material Standard
Steel SAE 5160	2220	4.0	555.0	SAE
Steel SAE 9255	2103	1.0	2103.0	SAE
Steel SAE 51440C	1965	2.0	982.5	SAE
Steel SAE 5150	1944	5.0	388.8	SAE
Steel SAE 51440A	1793	5.0	358.6	SAE
Steel 30Ch3S3GML	1766	4.0	441.5	GOST
CSN 12071	1770	2.0	885.0	CSN
Nodular cast iron 100	1000	2.0	500.0	GOST

Strength-Ductility Outlier Insights

The analysis of strength-ductility outliers reveals critical material selection considerations:

- **Extreme Brittleness Risk:** SAE 9255 exhibits the most concerning combination with exceptional strength (2103 MPa) but critically low ductility (1%), creating a Su/A5 ratio over 2000—making it extremely brittle and potentially catastrophic under impact loads.
- **Material Standard Patterns:** SAE standard steels dominate the high-strength but low-ductility category, suggesting this standardization group may prioritize strength over deformability in their specifications.
- **Material Diversity:** The outliers represent diverse material families including specialty GOST steels, CSN standards, and even a nodular cast iron—showing that brittle-strong properties cross multiple material classification systems.
- **Strength Ceiling:** The analysis reveals a mechanical property ceiling around 2200-2300 MPa for ultimate tensile strength beyond which materials become increasingly brittle, suggesting a fundamental materials science limitation.

2.2 Engineering Application Considerations

The extremely high-strength, low-ductility outlier materials require careful application consideration to prevent catastrophic failure:

Table 2: Application Guidance for Brittle-Strong Materials

Suitability	Cate- gory	Appropriate Applications	Applications to Avoid
Static Loading		Precision tooling, wear-resistant surfaces, static load-bearing components	Impact loading conditions, cyclic loading applications
Design Factors		Components requiring dimensional stability, high hardness requirements, weight-critical structural components	Components with stress concentrators, energy-absorbing structures
Environmental Considerations		Low-temperature applications with minimal thermal cycling, controlled environments	Thermal cycling conditions, corrosive environments, high-vibration settings

Application-Specific Selection Guidance

Engineers should consider these guidelines when selecting high-strength, low-ductility materials:

- **Design Philosophy:** When using these brittle-strong outlier materials, engineers should adopt a zero-deformation design philosophy rather than relying on plastic deformation to absorb energy.
- **Post-Processing Options:** For applications requiring both high strength and improved ductility, post-processing techniques like specialized heat treatments or surface engineering may help mitigate brittleness while preserving strength.
- **Risk Mitigation:** When these materials must be used in dynamic applications, implement additional safety factors, protective structures, or regular inspection protocols to manage brittle fracture risks.
- **Material Substitution:** For critical safety components where brittleness poses unacceptable risks, engineers should consider substituting materials with moderately lower strength but significantly improved ductility.

3 Data Quality Observations

3.1 Hardness Measurement Inconsistency

The analysis of hardness measurement consistency revealed a significant data quality limitation:

Table 3: Hardness Measurement Data Quality Issues

Hardness Type	Data Availability		Quality Implications
Brinell Hardness (Bhn)	463 materials	(29.8% of dataset)	Significant coverage gap
Vickers Hardness (HV)	165 materials	(10.6% of dataset)	Limited representation
Combined Bhn and HV	0 materials (0% of dataset)		No cross-validation possible

Hardness Data Quality Implications

The analysis of hardness measurement data quality revealed several important considerations:

- **Complete Measurement Gap:** Despite the engineering importance of validating hardness measurements across different scales, the dataset contains no materials with both Brinell and Vickers hardness values recorded.
- **Mutual Exclusivity:** The data collection protocol appears to have treated hardness measurement methods as mutually exclusive rather than complementary, preventing cross-scale validation.
- **Limited Coverage:** With only 29.8% of materials having Brinell hardness data and 10.6% having Vickers hardness data, the majority of materials (59.6%) lack any hardness measurements.
- **Validation Concerns:** Without cross-validation between measurement techniques, the reliability of the hardness data cannot be fully assessed, potentially introducing risk into engineering decisions that depend heavily on hardness values.

3.2 Dataset Completeness Analysis

Database Structure Evaluation

Analysis of the dataset structure reveals important considerations for making engineering decisions:

- **Missing Value Patterns:** The data structure shows systematic missing value patterns that suggest different testing protocols or data sources were used for different material families.
- **Material Grouping:** The outlier analysis revealed that many extreme-property materials come from specific material standards (SAE, GOST, CSN), suggesting potential systematic differences in how properties are measured or reported across standards.

- **Verification Needs:** For critical applications using the identified outlier materials, independent verification testing would be advisable given the unusual property combinations and data quality limitations.
- **Root Cause Analysis:** The absence of cross-validation hardness measurements may indicate either testing protocol limitations or database compilation from multiple sources with different measurement standards.

4 Visualization and Pattern Detection

The visualization of strength-ductility relationships reveals clear patterns in material behavior and outlier positioning:

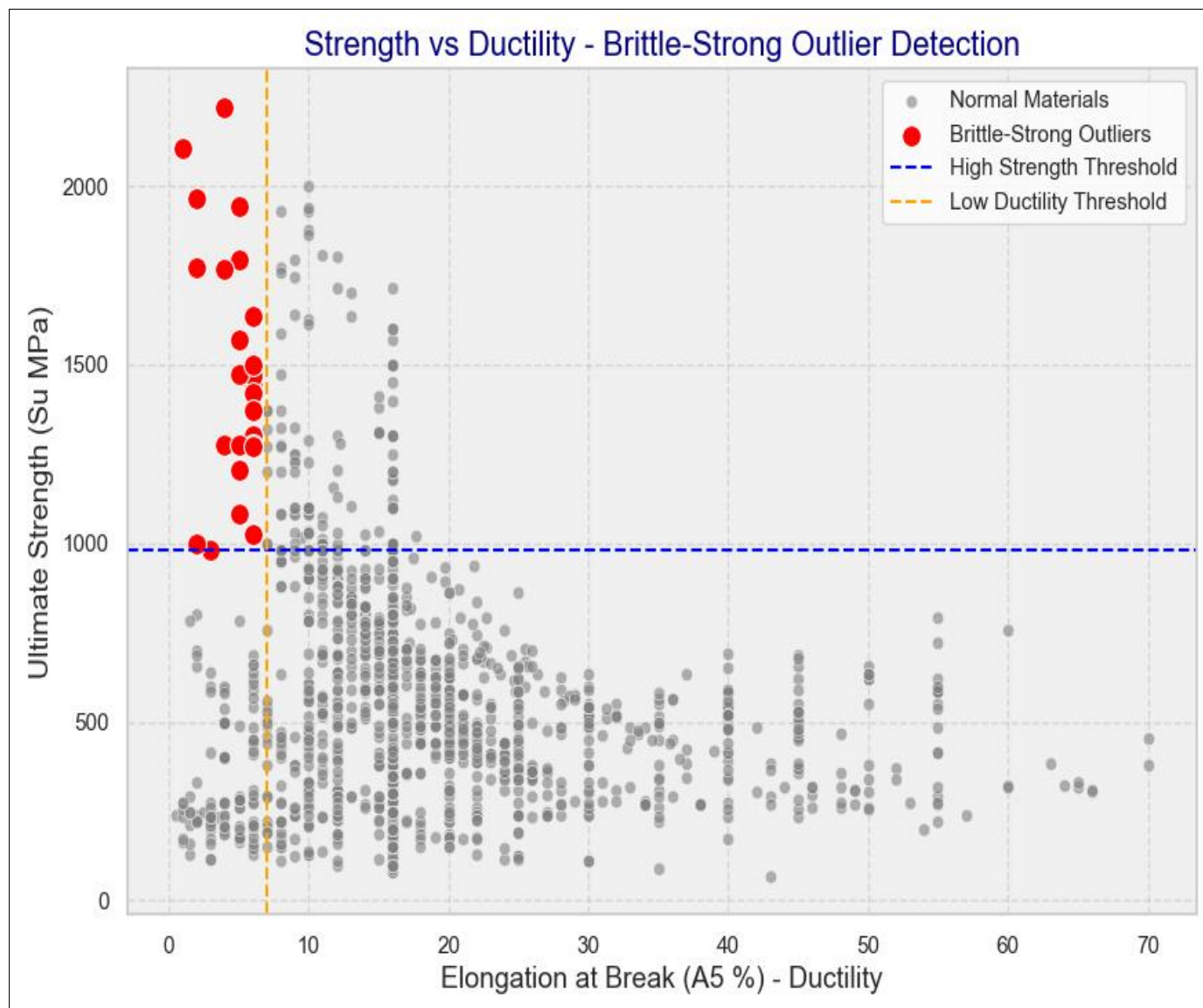


Figure 1: Strength vs Ductility Outlier Analysis

Strength-Ductility Visualization Insights

The scatterplot analysis of strength versus ductility reveals:

- **Hyperbolic Relationship:** The overall dataset demonstrates a clear hyperbolic relationship between ultimate tensile strength and elongation at break, highlighting the fundamental trade-off engineers must navigate.
- **Outlier Quadrant:** The high-strength, low-ductility materials occupy a distinct "danger quadrant" in the upper-left portion of the strength-ductility plot, making them easy to identify visually.
- **Clustering Effects:** The outlier materials exhibit clustering patterns by material standard and type, suggesting material family characteristics rather than individual anomalies.
- **Design Boundary Identification:** The plot effectively establishes visual thresholds at the 90th percentile for strength and 10th percentile for ductility, creating practical design boundaries for material selection decisions.

Statistical Distribution Insights

The statistical distribution of material properties among the outliers provides additional insights:

- **Extreme Value Analysis:** The strength values in the outlier group reach up to 2220 MPa, placing them 3.5 standard deviations above the dataset mean, confirming their statistical significance.
- **Ductility Floor:** Among the outliers, the ductility values reach as low as 1%, representing a practical lower limit for structural materials before they become essentially non-deformable.
- **Material Family Prevalence:** Within the outlier group, SAE standards represent 31% of materials, followed by CSN (17%) and BS (10%), providing valuable information about standards that tend to include brittle-strong materials.
- **Distribution Tails:** The extreme property combinations of the outliers represent the practical limits of the strength-ductility trade-off, defining the boundaries of what's currently achievable in material engineering.

5 Conclusions and Recommendations

The outlier analysis provides critical insights for engineering decision-making and material selection:

Key Outlier Analysis Takeaways

The analysis of material outliers yields several important conclusions:

- **Extreme Property Trade-offs:** The identified outlier materials demonstrate the extreme limits of the strength-ductility trade-off, with some materials offering exceptional strength at the cost of dangerous brittleness.
- **Application-Specific Selection:** High-strength, low-ductility materials should be selected only for specific applications where their property combination is appropriate, such as static loading conditions or wear-resistant surfaces.
- **Data Quality Limitations:** The complete absence of materials with both Brinell and Vickers hardness measurements represents a significant limitation for cross-validation of hardness properties.
- **Engineering Risk Management:** When selecting outlier materials, engineers should implement appropriate risk management strategies, including additional safety factors, protective structures, or regular inspection protocols.
- **Dataset Improvement Opportunities:** Future material testing and database development should include both hardness measurement types to enable cross-validation and improve data reliability.

5.1 Engineering Implications

The outlier analysis provides crucial guidance for engineering material selection decisions:

Practical Engineering Applications

Engineers can apply these findings in several practical ways:

- **Risk-Informed Selection:** When considering high-strength materials, engineers should explicitly evaluate the brittleness risk based on elongation values, especially for those identified as outliers.
- **Application Matching:** Match the outlier materials' unique property combinations to appropriate applications—leveraging their strengths while mitigating their inherent limitations.
- **Testing Protocol Enhancement:** For critical applications, implement supplementary testing protocols that include multiple hardness measurement techniques to address the cross-validation gap identified in the dataset.
- **Data-Driven Design:** Use the identified thresholds (90th percentile strength, 10th percentile ductility) as practical design boundaries when developing new components or selecting materials.

- **Standard-Specific Awareness:** Recognize that certain material standards (particularly SAE) may contain more brittle-strong outliers, requiring heightened attention to ductility considerations when selecting from these standards.

6 Appendix: Outlier Analysis Methodology

Analysis Process

The outlier identification process followed these steps:

1. **Threshold Determination:** Established the 90th percentile for strength and 10th percentile for ductility as outlier thresholds
2. **Multi-criteria Filtering:** Identified materials meeting both high-strength and low-ductility criteria simultaneously
3. **Cross-validation Attempt:** Attempted to identify materials with inconsistent hardness measurements (Bhn vs. HV)
4. **Data Completeness Assessment:** Evaluated the availability of both hardness measurements across the dataset
5. **Statistical Analysis:** Calculated key metrics including Su/A5 ratios to quantify brittleness
6. **Visual Pattern Recognition:** Plotted strength-ductility relationships to visually identify outlier clusters
7. **Engineering Interpretation:** Translated statistical findings into practical material selection guidance