Ship Engine Anomaly Detection: A Multi-Method Approach

Approach

This analysis used three complementary methods to detect anomalies in ship engine sensor data: the Interquartile Range (IQR) statistical method, One-Class Support Vector Machine (SVM), and Isolation Forest. Each method catches different kinds of problems, and together they provide both coverage and confidence for maintenance decisions.

A key decision early in the analysis was to split outliers into "high" and "low" categories for each feature rather than treating all outliers equally. Based on domain knowledge from the dataset documentation (Mohakul, 2022), a high lubrication oil temperature is more severe than a low one, and low oil pressure is more critical than high pressure. This led to introducing severity weights from 1 to 4 for each outlier type, informed by the physical meanings of each sensor reading.

Two extreme outliers in coolant temperature were removed before training SVM because the radial basis function kernel is sensitive to such extremes. This was not necessary for Isolation Forest since it isolates points through random splits rather than learning a boundary shape.

Here is the breakdown of anomalies by method:

- **IQR:** Samples with a weighted severity score of 5 or higher were flagged as anomalies, yielding 420 samples (2.15%).
- **SVM:** Used nu=0.03 with an RBF kernel and gamma='scale', producing 589 anomalies (3.02%).
- **Isolation Forest:** Used contamination=0.03, n_estimators=100, and max_samples='auto' on unscaled data, flagging 586 samples (3.00%).
- **Union of all three methods:** Gave 974 unique anomalies (4.99%), landing precisely in the target range.

Results

The three methods detected anomalies with meaningful overlap: 168 samples were flagged by all three methods, 285 by exactly two methods, and 521 by only one method.

Ship Engine Anomaly Detection Dashboard
Failure Modes by Severity and Detection Confidence

← Higher Confidence Lower Confidence →

Severity	Action	Failure mode	All 3 Methods	ML Overlap	SVM Only	IF Only	IQR Only
9	STOP	Coolant press HIGH + Fuel press HIGH + Lub oil temp HIGH	12	47	15	12	-
9	STOP	Coolant press HIGH + Engine RPM HIGH + Lub oil temp HIGH	18	44	28	24	-
8	STOP	Lub oil pressure LOW + Lub oil temp HIGH	8	31	14	12	-
7	STOP	Coolant pressure HIGH + Lub oil temp HIGH	15	55	22	19	-
7	STOP	Fuel pressure LOW + Lub oil temp HIGH	6	23	20	18	-
7	STOP	Coolant pressure HIGH + Lub oil temp LOW	5	24	12	11	-
6	SLOW	Fuel pressure HIGH + Lub oil temp HIGH	9	29	5	5	-
6	SLOW	Engine RPM HIGH + Lub oil temp HIGH	-	-	12	9	-
6	SLOW	Engine RPM HIGH + Lub oil temp LOW	3	10	18	16	-
5	SLOW	Coolant pressure HIGH + Fuel pressure HIGH	2	7	4	4	-
5	SLOW	Coolant pressure HIGH + Engine RPM HIGH	14	43	14	12	-
4	MONITOR	Lub oil temp HIGH	45	86	45	110	46
4	MONITOR	Engine RPM HIGH + Fuel pressure HIGH	3	12	5	3	-
4	MONITOR	Lub oil pressure LOW	1	5	8	8	-
3	MONITOR	Coolant pressure HIGH	22	80	8	27	-
3	MONITOR	Fuel pressure LOW	-	1	10	8	-
2	MONITOR	Fuel pressure HIGH	12	42	20	16	-
2	MONITOR	Engine RPM HIGH	10	33	15	12	-

Count: 0, 20, 40, 60, 80, 100

Table 1. The dashboard. The highest-confidence anomalies (flagged by all three methods) are listed by severity. The most severe combinations, both scoring severity 9, involve either coolant pressure high, fuel pressure high, and lubrication oil temperature high, or coolant pressure high, engine RPM high, and lubrication oil temperature high. These six STOP-level combinations accounted for 64 total high-confidence critical events.

The most frequently flagged single feature was high lubrication oil temperature, appearing in 155 IF-only detections and 131 overlap detections. However, its standalone severity is only 4 (MONITOR level) because high oil temperature alone, while common, does not require immediate action. This accounts for roughly half of all flagged anomalies but should not cause the crew to stop or slow down. The more severe combinations involving multiple features warrant stronger responses.

Method Comparison

Each method serves a different purpose. IQR provides simple one-dimensional flagging of boundary violations. The crew can quickly see which individual readings are out of range, and with experience they will develop intuition for what is normal during docking, travelling, or rough weather.

SVM learns what "normal" looks like across all features simultaneously and flags points that fall outside that learned shape, even if individual values seem acceptable. The PCA visualisation shows SVM-only anomalies scattered throughout the distribution rather than concentrated at the edges. These represent subtle multivariate patterns where the combination of readings is unusual. Two extreme outliers from coolant temp were removed from SVM training but included in the outlier set.

Isolation Forest isolates extreme points through random partitioning. It tends to overlap more with IQR because both are sensitive to extreme values, but IF can catch extremes that fall just inside IQR boundaries. The PCA shows IF anomalies clustering toward the outer regions of the distribution.

When all three methods agree, confidence is highest. When only one method flags a reading, it may warrant logging and monitoring rather than immediate action.

PCA Visualisation Evaluation

The 2D PCA projection, despite the presence of 12 outlier types, was based on only six features, capturing 55.3% of the variance (PC1: 35.9%, PC2: 19.4%). This visualisation was intentionally displayed alongside the Interquartile Range (IQR) analysis, which provided a distinct, almost quadrant-like breakdown of the data space. This side-by-side comparison facilitated a clearer interpretation of the detection methods: IQR anomalies appeared in clear outer quadrants, representing boundary violations. SVM anomalies were more dispersed, highlighting a nuanced detection across the space, while Isolation Forest (IF) anomalies clustered at distribution edges, often with substantial IQR overlap, indicating their focus on more extreme value-based outliers in the outer 'quadrants'.

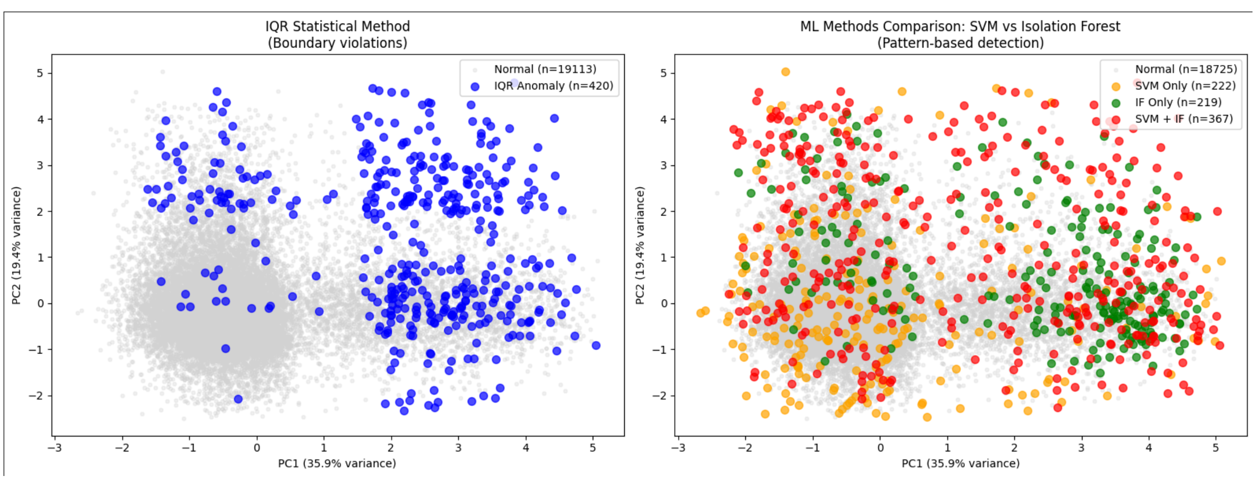

Table 2 shows the results of the 2D Principal Component Analysis (PCA) projection used to evaluate the detection patterns of the three methods. The projection captured 55.3% of variance (PC1: 35.9%, PC2: 19.4%) and effectively distinguished the anomalies: IQR clustered in clear outer quadrants (boundary violations), SVM was more dispersed, and Isolation Forest clustered at the distribution edges with

substantial IQR overlap. The side-by-side plots are useful for method comparison, though two dimensions cannot fully represent the six-dimensional relationships.

The side-by-side PCA plots proved useful for understanding what each method catches. However, two dimensions cannot fully represent six-dimensional relationships, so some anomalies that appear within the main cluster in 2D may actually be distant in the original feature space.

Conclusions and Recommendations

The multi-method approach identified 974 unique anomalies (4.99%) with confidence tiers based on method agreement. For ship engine maintenance, the following actions are recommended:

STOP (Severity 7-9): Combinations involving multiple critical features require immediate load reduction. These include any pairing of high coolant pressure, high fuel pressure, low oil pressure, and extreme oil temperatures. Approximately 64 highest-confidence events fall into this category.

SLOW DOWN (Severity 5-6): Combinations like high fuel pressure with high oil temperature, or high engine RPM with temperature anomalies, warrant load reduction and investigation. Roughly 28 high-confidence events require this response.

MONITOR (Severity 2-4): Single-feature anomalies such as high oil temperature alone should be logged and watched for patterns. Though high oil temperature appears frequently, it does not require stopping or slowing down unless combined with other issues.

The dashboard (Table 1) can serve as a practical reference for maintenance crews. When a reading seems off, they can check whether it is an IQR flag (simple boundary violation), an SVM flag (subtle pattern), or an IF flag (extreme value). The overlap columns show detection confidence, helping crews prioritise their response.

Based on frequency data, crews should expect roughly 50 readings per 1000 requiring some attention and about 5 per 1000 requiring immediate action. Routine inspections should prioritise the lubrication system (most common in high-severity modes), coolant system (appears in 7 of the top 10 failure combinations), and fuel system components.

References

Mohakul, D. (2022). Ship engine data for condition-based predictive maintenance. IEEE Dataport. https://dx.doi.org/10.21227/1xth-3t82