EPH–International Journal of Science & Engineering
Published by: E-Pub Solutions Private Limited
ISSN (Online): **2454-2016**
Volume 12 Issue 03 August 2026
DOI:10.67772/30q43713

PRECISION MANUFACTURING IN AEROSPACE ENGINEERING USING ADVANCED ROBOTIC SYSTEMS

Durvaank Umakant Patel*1

1*10th Grade (High school), Aerospace Engineering, Aviation and Robotics, South Windsor High School, South Windsor, Connecticut, United States, 06074, pateldurvaank@gmail.com

Article History:

Article Type: **Research**
Received Date: **10/06/2026**
Revised Date: **14/07/2026**
Accepted Date: **17/08/2026**
Published Date: **30/08/2026**

Keywords: aerospace manufacturing, robotic assembly, KUKA KR16, defect detection, YOLO11n, precision manufacturing.

ABSTRACT

Automated inspection and robotic process control of the manufacturing process are becoming more integral to the precision aerospace manufacturing process to ensure product quality and assembly consistency. In this study, a complementary framework based on two modules, a robotic fastener-process characterization module and an artificial-intelligence-based aircraft-defect-detection module, were developed. The first module involved analyzing 479 trials of the KUKA KR16 for aeronautical fastener-assemblies based on temporal kinematic and dynamic signals, engineering indicators, deviation measures, and supervised learning for assessment of the assembly outcome. The second module used a leakage-controlled YOLO11n workflow for dent, fastener damage and rupture detection, validation only model selection, locked testing, robustness analysis, bootstrapping uncertainty estimation and object-scale diagnostics and checkpoint-reproducibility controls. There was still the 3-class model with a Macro-F1 of 0.4127, but none of the engineering KPI families were of significance following FDR correction. The chosen YOLO11n-416 detector performed best with test precision of 0.8027, recall of 0.5612 and mAP50–95 of 0.2993, substantially compromised by the performance on the small defects. These results suggest the use of complementary evidence of process stability and quality verification, and not just a single performance measure, to assess the precision of the aerospace manufacturing industry. The framework lays the groundwork for a repeatable approach to the integration of robotic execution analysis with uncertainty-aware visual quality assessment without baseless claims of actual direct robotic-vision fusion or proven dimensional accuracy gains in production.

1. Introduction

The increasing demand for high value structural assemblies, from aerospace manufacturing, becomes more dependent on advanced levels of automation to meet the demands of repeatability, traceability and process stability. Today's aircraft manufacturing is highly complex, has demanding quality requirements, process windows are narrow, and tolerances for assembly errors are tight—making robotic systems an essential part of ensuring aircraft manufacturing consistency. As a result, robotic assembly has become more than just material handling, and is now a technology used for drilling, fastening, riveting, positioning, inspection, and coordinated human–robotic tasks in aerospace applications. The recent reviews demonstrate the development of Robotic Assembly of Aerospace Parts towards higher level of autonomous performance, integration of sensing and adaptive control, but they are still limited by certain structural compliance, part variability, contact uncertainty and the need to assure precision in the presence of varying process conditions (Zhao et al., 2026). Unlike in a typical positioning system, these issues make precision manufacturing a much more complex problem involving the entire systems, since the quality of manufacturing depends on the stability of the robot's motion, the consistency of the force–torque interaction, the repeatability of the assembly trajectory, and the ability to acquire the defects after the completion of the manufacturing process.

This shift is consistent with Industry 4.0, which is changing the way aerospace products are being manufactured through the requirements of automation, digitalization, connected sensing and data-driven decision support. Aerospace composite manufacturing is an example of this, but a survey of the literature also points to the fact that automation levels are still not uniform per process and per factory but many processes are still being performed via hybrid human–machine configurations due to the complexity and flexibility requirements, as well as the need for quality control (Jayasekara et al., 2022). Digital-thread and digital-twin technologies have also increased the scope of manufacturing information to connect process models, production states and lifecycle information to enable system-level monitoring and coordination (Zhang et al., 2022). But automation and digital representation doesn't automatically mean precision is achieved in manufacturing. In cases where robots are physically interacting with the structures of aircraft, and in cases where aircraft simply need to be fastened and aligned with high precision, or where small deviations lead to problems with surface quality, precision should be determined by measurable process behavior and quality outcomes.

In the field of aerospace assembly and inspection, the reliability of process execution and effective quality verification are two issues that are particularly important in the robotized process. The significance of sensing and feedback in robotized structural assembly and inspection with geometric uncertainty and variation in the workpiece is highlighted by research on visual servoing (da Silva Santos et al., 2022). Aside from this, investigations into the riveting of aircraft parts have stressed the importance of real-time modelling of process forces due to their relationship with tool–structure interactions and the ability to offer information on process stability and assembly conditions (Kang et al., 2024). The synergy of humans and robots also tackles trade-offs in the field of aerospace assembly between productivity, ergonomics, resource allocation and operation limitations (Hémono et al., 2025). Even with these improvements, the existing literature is scattered over robotic execution, process monitoring, process optimization and visual inspection, and it is difficult to build an evidence-based framework that links manufacturing behavior with downstream visual inspection of the quality.

One of the limitations related to this is the inspection of the aircraft surface. While automated recognition of dents, cracks, ruptures, fastener damage and other surface anomalies has been greatly advanced by deep learning, the performance of these methods remains highly dependent on the scale, quality, class imbalance, quality of the annotations, and construction of the datasets. For instance, the focus on aircraft surface-defect detection (ASD-YOLO) is highlighting the architectural modification with deformable convolution and attention mechanisms for the current emphasis (Huang et al., 2024). Similarly, the research on instance segmentation has made aircraft-skin inspection closer to providing spatially explicit information on defects instead of just image-

level defect classification, thus proving that automated inspection can also deliver spatially explicit defect information (Meng et al., 2022). Some recent studies tackle such cases by numerically augmenting the amount of data with Fourier-GAN, emphasizing the challenges associated with creating a strong model of aircraft defects from the limited number of labeled images (Li et al., 2025). However, methodological shortcomings, including duplicate leakage, overlap in the source families, choosing thresholds after the data was collected, problematic reporting of the uncertainty, and the degradation of performance between validation and independent testing are not being addressed by stronger model architectures and augmentation strategies.
The main research question is the lack of a well-tested precision-oriented aerospace manufacturing framework, which is dependent on robotics process characterization, independent of the methodology and facilitates AI-driven visual quality inspection. The existing studies make significant contributions to the fields of robotic assembly, sensing, process-force modeling, collaborative optimization, and aircraft-defect detection, but they are typically studied separately. Consequently, little research has been conducted on the myriad ways by which temporal robots' actions, engineering deviation indicators, assembly-outcome discrimination, leakage-controlled defect detection, uncertainty estimation, robustness assessment, and validation-to-test generalization can be analyzed in a single study with a manufacturing focus. This gap is significant because the underlying robotic process cannot be guaranteed by a high performing inspection model and process signals cannot be used to prove that manufactured aircraft surfaces are defect-free.
In this regard the present study fulfills an additional complementary design in two modules. The first module describes the temporal characterization of the KUKA KR16 aeronautical fastener assembly from the perspective of a temporal sensor system consisting of the specific sensor signals position, orientation, force and torque, linear-velocity, and angular-velocity, as well as engineering key performance indicators, nominal successful-assembly deviations, precision-envelope measures, and supervised outcome models. The second module is an aircraft-defect detector YOLO11n with leakage control, based on validation-only model selection, model checkpoint freezing, locked final testing, bootstrap uncertainty, object-scale diagnostics, perturbation robustness and validation-to-test attribution. The study is not physically integrated, but is rather an analytically integrated robotic-vision system. It doesn't make claims of closed-loop vision-guided robot control, direct dimensional metrology or causal improvement of aircraft manufacturing tolerances. It is instead the complementary dimensions of precision-oriented manufacturing process that are evaluated: namely consistency of the process and quality inspection.
Theoretically, this framing has implications as it sees precision manufacturing as a complex construct that includes repeatable robotic behavior, controlled interaction, deviation from nominal manufacturing conditions and reliable defect localization. It offers in practice a repeatable evaluation framework for decoding evidence from the various aspects of robotic processes and inspections while not overstating the integration of different phases of aerospace production systems. The study also makes a methodological contribution through the integration of statistical control of engineering indicators with leakage-aware computer-vision experiment, a fixed test protocol, bootstrap confidence intervals, deterministic auditing for qualitative output, and reproducing the experiment through checkpoints along with a record of the experiment. With this in mind, the first goal was to characterize the aeronautical fastener assembly of the KUKA KR16 using temporal kinematic and dynamic signals and to see if any of the engineering indicators related to precision differ in the assembly results. The second goal is to propose and validate independently a leakage-controlled YOLO11n framework to detect Dent, Fastener Damage and Rupture in validation only model selection and locked testing. The third goal is to incorporate the robotic-process and visual-inspection evidence into a complementary aerospace manufacturing platform, quantify the robustness and uncertainty of the robotic-process, object-scale sensitivity, and generalization without making explicit model fusion.

2. Literature Review

Study on intelligent assembly is gradually transforming the manufacturing performance problem into coordination one related to the autonomous control of the robot, the interactions between the man and the robot, the perception, and the adaptation. Incorporating the notion of state-dependent decision policies, graph-based reinforcement learning (RL) has taken a step in this direction by allowing robots to dynamically adjust collaborative assembly actions, highlighting the importance of reward functions and task graphs while showing that they are susceptible to carefully designed task graphs and reward formulations (Zhang et al., 2022). Temporal reasoning has since been used to extend collaboration to self-organising multi-human–robot systems, with the aim of increasing the flexibility of allocation, while increasing computational and coordination complexity under varying conditions of the shop floor (Li et al., 2023). Recent work in robotic assembly has integrated vision and force sensing into residual reinforcement learning, enhancing contact-aware adaptation, but it continues to depend on the calibration of the sensors, conditions for transfer, and the distribution of training tasks for the specific application (Zhang et al., 2024).

Another research stream involves the precision enhancement from digital representation, multimodal feedback and system calibration. Hybrid drilling robot modelling to accurately represent the robot's geometric shape and correspondence of its state has been established by introducing high precision digital-twin modelling of hybrid drilling robots, however, the accuracy of the modelling is based on the fidelity of the modelling, the ability to identify the robot's parameters and the synchronization of the physical and virtual systems (Kang et al., 2025). A fusion of force and vision is also beneficial for precision assembly, but may result in performance degradation if the uncertainty in the visual information, variation in contact or flexible behavior of the workpiece is outside of the calibrated operating range (Wang et al., 2025). The in situ robotic manufacturing research also reveals that integrated calibration and datum management are helpful to high-precision machining of large aerospace structures, but most of the above-mentioned methods focus on the geometric precision and process control precision of the products. Downstream defect verification is not considered (Zheng et al., 2025).

The development of industrial inspection has been parallel, and the deep visual learning has been the dominant factor. Most of the surveys focus on CNN-based defect detection that highlight good representation-learning capacity along with enduring sensitivity to few labels, class imbalance, background complexity and domain shift (Jha & Babiceanu, 2023). The existing systematic literature provides further empirical evidence of the increasing use of deep architectures to detect surface defects, and of the fact that comparability is limited by the use of different datasets, metrics and validation protocols (Ameri et al., 2024). Transformer-based methods solve the long-range dependency and contextual modelling issues, while hybrid methods help reduce the cost of features interaction, but come with a trade-off of higher computational expenses and data needs (Wang et al., 2023). Defect-aware transformer models further enhance the attention to anomaly relevant regions, but the performance is heavily dependent on the scale of defects, variability in texture and training-domain representativeness (Shang et al., 2023). Specific challenges that are particularly relevant to aircraft-specific reviews include persistent challenges in small-defect localization, illumination variation, small number of annotated samples, and deployment reliability (Li et al., 2025).

This separation makes it difficult to assess the system for an outcome of defect-free trajectories, since defects do not always indicate a system failure, and it is difficult to determine if process indicators are transferable to successful and unsuccessful assembly situations in practice. There is a distinct gap across these themes: Robotic assembly studies are mainly focused on control, collaboration/supervisory control, calibration, or on digital twins, while visual-inspection studies focus on defect recognition separately. Limited work is presented that considers, in a single manufacturing framework, the behavior of robots over time, engineering indicators that are statistically controlled, aircraft defect detection that is leakage aware, uncertainty, robustness, and generalization from validation to test. The present study aims to fill this gap by performing complementary non-fused robotic process characterization and visual quality inspection via AI.

3. Methodology

3.1 Research Design and Two-Module Manufacturing Framework

A quantitative, secondary-data engineering design examined precision-oriented aerospace manufacturing through two complementary modules: robotic fastener-process characterization and AI-assisted visual inspection. Module I evaluated KUKA KR16 aeronautical threaded-fastener assembly from kinematic and dynamic signals, whereas Module II developed an object-detection pipeline for aircraft defects. The modules were analyzed independently because the datasets were not generated from a synchronized production line; integration therefore occurred at the manufacturing-system interpretation level rather than through model fusion. Precision was operationalized as process consistency, deviation from nominal successful assembly, stability of robot–workpiece interaction, and reliable defect localization. As illustrated in Figure 1, this design links process execution with quality assurance while preserving evidence bases.

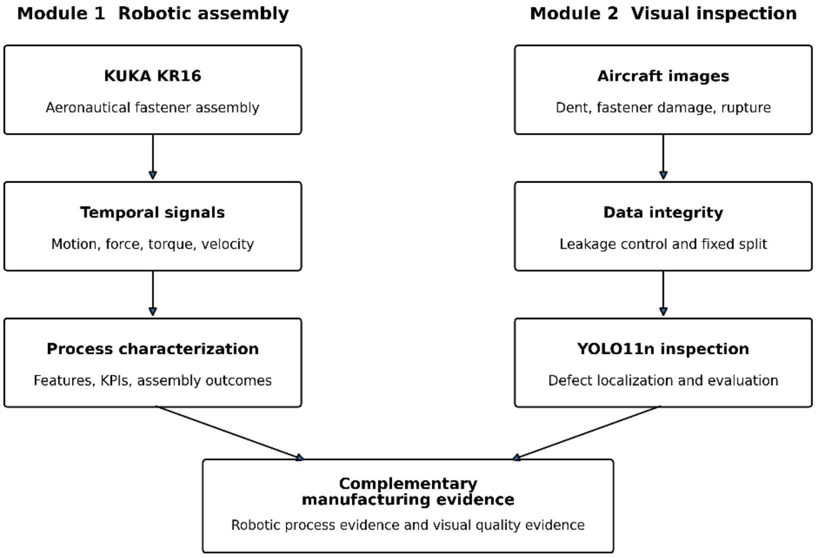

Figure 1. Two-Module Framework for Precision-Oriented Aerospace Manufacturing

Figure 1 shows that robotic process characterization and AI-assisted visual inspection provide complementary evidence for manufacturing execution and quality control. The framework does not imply closed-loop robot–vision integration or causal improvement in dimensional accuracy.

3.2 Aerospace Manufacturing Data Sources and Experimental Boundaries

The robotic dataset comprised 479 labeled KUKA KR16 assembly trials involving aeronautical collars: 306 Mounted, 112 Not-Mounted, and 61 Jammed (Giardini Lahr et al., 2023). All eligible trials available in the secondary dataset were included in the analysis. The vision dataset was the Roboflow UTS Aircraft Defect Detection v3 collection containing Dent, Fastener Damage, and Rupture annotations (University of Technology Sydney, 2024). After duplicate and source-family controls, 6,320 images remained: 4,531 training, 897 validation, and 892 test images. Sample size was determined by complete-case availability in the datasets rather than an a priori power calculation; independent holdouts and bootstrap uncertainty quantified generalization. As summarized in Table 1, the datasets represent complementary manufacturing stages.

Table 1. Data Sources and Experimental Configuration of the Two Manufacturing Modules

Experimental component	Module I: Robotic fastener process characterization	Module II: Aircraft surface defect inspection
Manufacturing function	Robotic execution and assessment of aeronautical fastener assembly	AI-assisted visual inspection of aircraft surface defects
System / model	KUKA KR16 industrial robot	YOLO11n object detector
Primary data type	Multivariate robotic time-series signals	Aircraft surface images with bounding-box annotations

Experimental observations	479 assembly trials	6,803 initial images; 6,320 pixel-unique images retained after leakage control
Outcome / defect categories	Mounted, Jammed, Not-Mounted	Dent, Fastener Damage, Rupture
Class distribution	Mounted: 306; Jammed: 61; Not-Mounted: 112	11,050 annotated defect objects across the final leakage-safe dataset
Input variables	End-effector position and orientation; force and torque; linear and angular velocity	RGB aircraft surface images and corresponding object annotations
Temporal / spatial representation	Approximately 31-s cleaned assembly cycle; engineered temporal and process-level descriptors	Leakage-group-aware image partitions with object-level bounding boxes
Feature representation	2,256 temporal predictors plus engineering KPI families	Image features learned directly by YOLO11n
Primary analytical procedures	Temporal characterization, engineering KPI assessment, three-class outcome prediction, and Jammed vs Not-Mounted binary analysis	Object detection, validation-only model comparison, locked-test evaluation, error analysis, robustness assessment, and bootstrap uncertainty estimation
Training / development partition	Model-specific stratified development procedure applied to robotic trials	4,531 images; 7,654 objects
Validation partition	Used within the robotic predictive workflow where applicable	897 images; 1,696 objects
Final test partition	Locked three-class test: 72 trials; locked binary test: 26 trials	892 images; 1,700 objects
Leakage / integrity control	Trial-level separation and preprocessing of individual temporal records	Exact/near-duplicate auditing, source-family grouping, and zero leakage-group overlap across final partitions
Primary evaluation measures	Accuracy, balanced accuracy, Macro F1, MCC, ROC AUC, engineering KPIs, and statistical significance	Precision, recall, Macro F1, mAP50, mAP75, mAP50-95, object-scale error rates, robustness, and bootstrap confidence intervals
Role in the proposed framework	Characterizes consistency, motion/interaction behavior, and assembly-outcome discrimination	Evaluates defect localization and visual quality-inspection capability
Experimental boundary	Does not establish micron-level dimensional precision or direct process optimization	Does not provide closed-loop robotic control or direct robot-vision fusion

Table 1 shows that the robotic module captures process-level kinematic and interaction data, whereas the vision module captures image-level defect evidence. Their separate sampling frames prevent claims of direct experimental coupling.

3.3 Robotic Signal Preprocessing and Precision-Oriented Feature Construction

The trials included time, end-effector position, Euler orientation, force, torque, linear velocity and angular velocity for each robotic trial. As a result of repeated cycle inspection, duplicated first/second segments were found and the unique 31-s interval was retained to avoid pseudo-replication. Derivative-based interpretation was performed with unwrapped Euler angles to reduce spikey angular-velocities caused by discontinuities. Cleaned signals were converted to 2,256 temporal predictors which summarized trajectory, interaction and motion characteristics. Along with the 18 scalar KPIs, 14 nominal Mounted-deviation KPIs and 14 cross-fitted precision-envelope KPIs were used in analysis in a precision-oriented way; the latter are more related to the stability of the process and deviation from it, rather than to the metrology-grade positional accuracy.

3.4 Robotic Assembly Modeling and Statistical Assessment

To break the 479 robotic assembly trials, they separated them into mutually exclusive training, validation, and test sets, which contained 335, 72 and 72 trials, respectively. XGBoost was used to make a classification of the Mounted, Not-Mounted, and Jammed outcome variables, whereas logistic regression was used to give a binary label for Not-Mounted versus Jammed outcomes. The accuracy, balanced accuracy, Macro-F1, Matthews correlation coefficient and ROC-AUC were used to assess the performance. The 18 scalar engineering KPIs, 14 nominal Mounted-deviation KPIs and 14 cross-fitted precision-envelope KPIs were used for statistical assessment. Multiple comparisons were corrected for false-discovery-rate (FDR), and cross-fitting was performed to minimize optimistic bias and maximize the statistical validity of the analysis by separating the estimation of references and deviation assessment.

3.5 Aircraft Defect Dataset Integrity and Leakage Control

Image and label files were checked for consistency, validity of classes, bounding-box geometry, missing records, near duplicate images and labels and source families. Before splitting, related images were pre-grouped to prevent the crossing of borders between training, validation and test sets in terms of derivatives. These were then removed to make an exact duplicate, resulting in the final leakage controlled partition. As shown in Figure 2, only training data could be used to develop the model, candidate comparison to the validation data, and final estimation to the locked test partition.

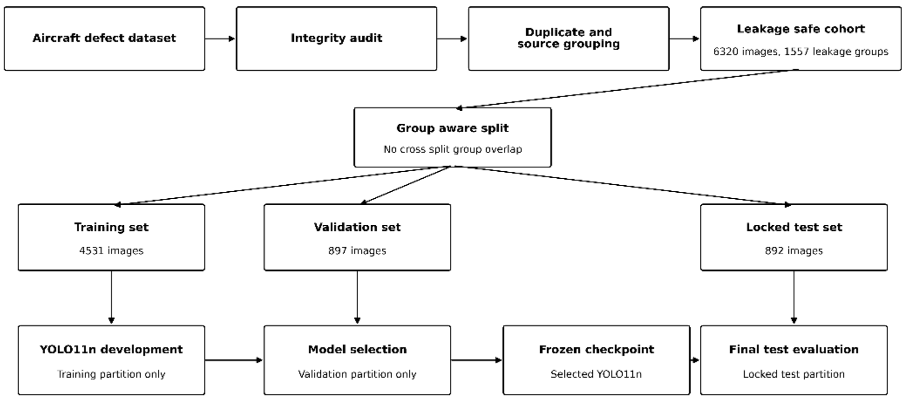

Figure 2. Leakage-Controlled Aircraft Defect Detection Pipeline

Figure 2 shows that duplicate control precedes model development and validation-only selection separates optimization from final testing. This sequencing reduces information leakage and optimistic performance estimation.

3.6 Defect-Detection Model Development and Validation-Only Selection

YOLO11n was ported to Ultralytics using PyTorch. The baseline was a CPU-feasible one with 416-pixel inputs, and a pre-defined 512-pixel candidate was used to test the ability to improve defect localization with higher resolution. The validation precision, recall, Macro-F1, mAP50, mAP75, mAP50–95, object-size diagnostics, and perturbation tests were used for candidate comparison. Only validation mAP50–95 and only after SHA-256 identity had the final checkpoint selected and then it was frozen for test evaluation. Final-test outcomes were not used to select a confidence threshold or configuration.

3.7 Final Evaluation, Robustness, Uncertainty, and Reproducibility

The test result of the frozen detector was evaluated with Precision, Recall, Macro-F1, AP, class-wise AP, mAP50, mAP75, mAP50–95, confusion behavior and CPU inference time. For correct localization and useful overlap, the error analysis was set at 0.25 and 0.10 respectively, when considering IoU. Robustness was tested with blur, brightness, contrast and JPEG perturbations. Image-level bootstrap resampling (300 replications) estimated uncertainty in custom mAP50–95; and validation-to-test change was split into composition/class effects and within-stratum effects. Deterministic sampling, checkpoint hashes, saved predictions, manifest-linked outputs, and artifact SHA-256 records supported reproducibility. Since both sets of data were secondary research material, without any human subjects or personal identifiers, informed consent, confidentiality procedures and institutional human subjects approval was not applicable.

4. Results

4.1 Dataset Integrity and Experimental Cohort Characteristics

The robotic dataset comprised 479 KUKA KR16 aeronautical fastener-assembly trials: 306 Mounted (63.88%), 112 Not-Mounted (23.38%), and 61 Jammed (12.73%). Removal of the duplicated cycle segment retained an approximately 31-s unique assembly interval and produced 2,256 temporal predictors from position, orientation, force, torque, linear-velocity, and angular-velocity channels. For the vision module, exact-duplicate removal and source-family grouping yielded 6,320 leakage-controlled images: 4,531 training, 897 validation, and 892 test images, containing 7,654, 1,696, and 1,700 objects, respectively. No cross-partition overlap remained after finalization procedures.

4.2 Robotic Process Characterization and Engineering Indicators

Raw orientation derivatives contained discontinuity-driven angular-velocity spikes before Euler-angle correction. Unwrapping reduced the maximum apparent angular velocity from approximately 29,970.88 deg/s to approximately 32.04 deg/s. As illustrated in Figure 3, the robotic module was evaluated using time-resolved process behavior rather than terminal states.

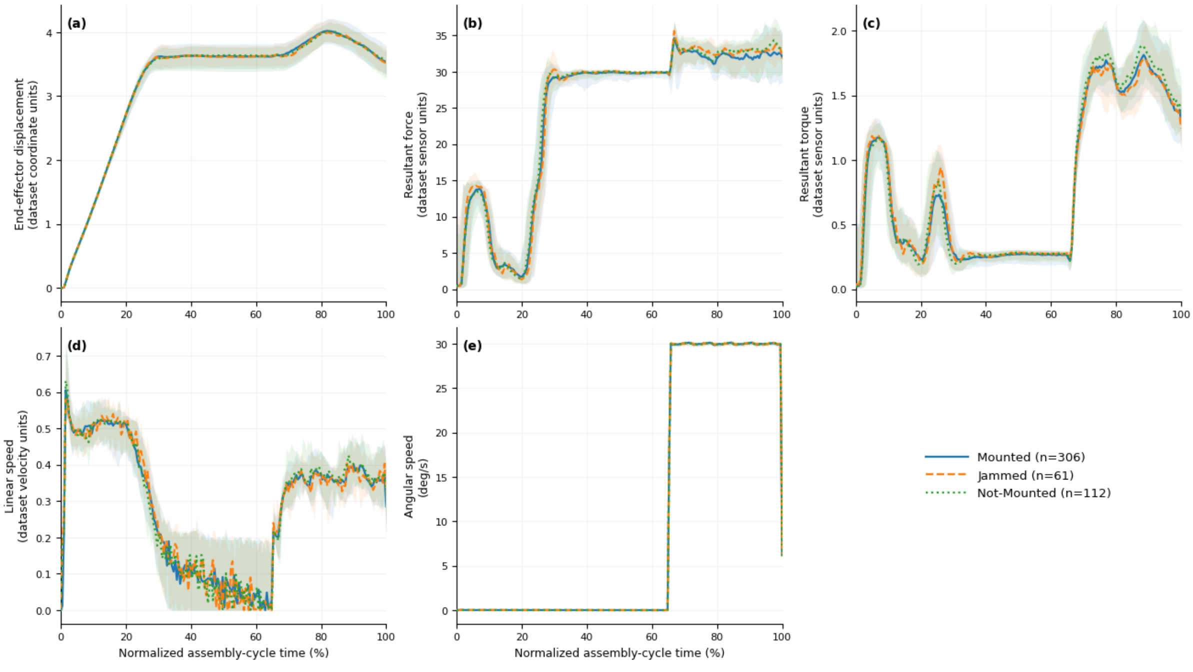

Figure 3. Temporal Characterization of the KUKA KR16 Aeronautical Fastener Assembly

Process

Figure 3 shows synchronized kinematic and dynamic signals across the cleaned cycle and the removal of orientation-wrap artifacts. These trajectories formed the basis for temporal predictors and precision-oriented engineering indicators.

Three KPI families were evaluated: 18 scalar engineering KPIs, 14 nominal Mounted-deviation KPIs, and 14 cross-fitted precision-envelope KPIs.

4.3 Robotic Assembly Prediction and Statistical Significance

The three-class XGBoost model, evaluated on 72 held-out trials, achieved accuracy 0.4444, balanced accuracy 0.4623, Macro-F1 0.4127, MCC 0.1294, and ROC-AUC 0.5981. Binary Not-Mounted-versus-Jammed logistic regression, evaluated on 26 held-out trials, achieved accuracy 0.5000, balanced accuracy 0.5131, Macro-F1 0.4933, MCC 0.0249, and ROC-AUC 0.4771. As summarized in Table 2, no KPI family retained statistical significance after FDR correction.

Table 2. Predictive and Statistical Results of KUKA KR16 Robotic Fastener Assembly

Analysis	**Result**
Three-class XGBoost	Macro-F1 0.4127; ROC-AUC 0.5981
Binary logistic regression	Macro-F1 0.4933; ROC-AUC 0.4771
Scalar engineering KPIs	0/18 FDR-significant
Mounted-deviation KPIs	0/14 FDR-significant
Precision-envelope KPIs	0/14 FDR-significant

Table 2 shows limited held-out discrimination for both predictive tasks. It also confirms that none of the 46 engineering indicators remained significant after multiple-testing correction.

4.4 Detector Validation Performance and Final Model Selection

Validation-only comparison favored the YOLO11n-416 baseline. It achieved precision 0.787841, recall 0.625000, Macro-F1 0.696123, mAP50 0.703433, mAP75 0.316805, and mAP50–95 0.355858. The 512 candidate achieved precision 0.824151, recall 0.619632, Macro-F1 0.707406, mAP50 0.694385, mAP75 0.302805, and mAP50–95 0.348042. The baseline was therefore selected on the predefined primary criterion, validation mAP50–95, with a margin of 0.007816, and its checkpoint was frozen before final testing. As shown in Table 3, the higher-resolution candidate achieved a slightly higher Macro-F1 but did not improve the primary selection metric.

Table 3. Validation-Based Detector Selection and Final Test Performance

Stage	**Input**	**Precision**	**Recall**	**Macro-F1**	**mAP50**	**mAP75**	**mAP50–95**
Baseline validation	416	0.7878	0.6250	0.6961	0.7034	0.3168	0.3559
512 candidate validation	512	0.8242	0.6196	0.7074	0.6944	0.3028	0.3480
Final baseline	416	0.8027	0.5612	0.6606	0.6437	0.2188	0.2993

test							

Table 3 shows that the baseline retained the highest validation mAP50–95 and became the final checkpoint.

4.5 Final Aircraft Defect Detection Performance

On 892 test images containing 1,700 objects, the frozen baseline achieved official detector-evaluation precision of 0.802650, recall of 0.561210, Macro-F1 of 0.660559, mAP50 of 0.643735, mAP75 of 0.218843, and mAP50–95 of 0.299262. Class-wise AP50–95 was 0.292440 for Dent, 0.279877 for Fastener Damage, and 0.325470 for Rupture; recalls were 0.590902, 0.501819, and 0.590909. CPU inference time was 32.29 ms/image. Custom mAP50–95 was 0.300116; 300 bootstrap replicates yielded mean 0.303086, SD 0.012329, and 95% CI [0.280361, 0.328255].

4.6 Error Structure, Object Scale, Robustness, and Generalization

At the fixed confidence threshold of 0.25 and IoU 0.50, 784 of 1,700 objects were correct detections (46.12%), 630 were low-confidence correct localizations (37.06%), 170 were localization errors (10.00%), seven were class confusions (0.41%), and 109 were misses or produced no useful prediction (6.41%). Correct-detection performance increased markedly with object size, from 14.50% for small objects to 47.88% for medium and 54.62% for large objects. The corresponding miss rates were particularly high for very small targets, reaching 84.00% for objects with a minimum dimension below 4 pixels and 61.90% for those measuring 4–7 pixels.

These diagnostic metrics differ from the official detector metrics reported in Section 4.5 because they were calculated using the prespecified object-level operating point of confidence 0.25 and IoU 0.50. Under this diagnostic protocol, macro precision was 0.701028, macro recall was 0.606796, and Macro-F1 was 0.649696. Validation-to-test performance declined by 8.49% for mAP50, 30.92% for mAP75, and 15.90% for mAP50–95. As illustrated in Figure 4, the largest relative generalization decline occurred at the stricter localization threshold, while correct-detection performance increased substantially from small to medium and large objects.

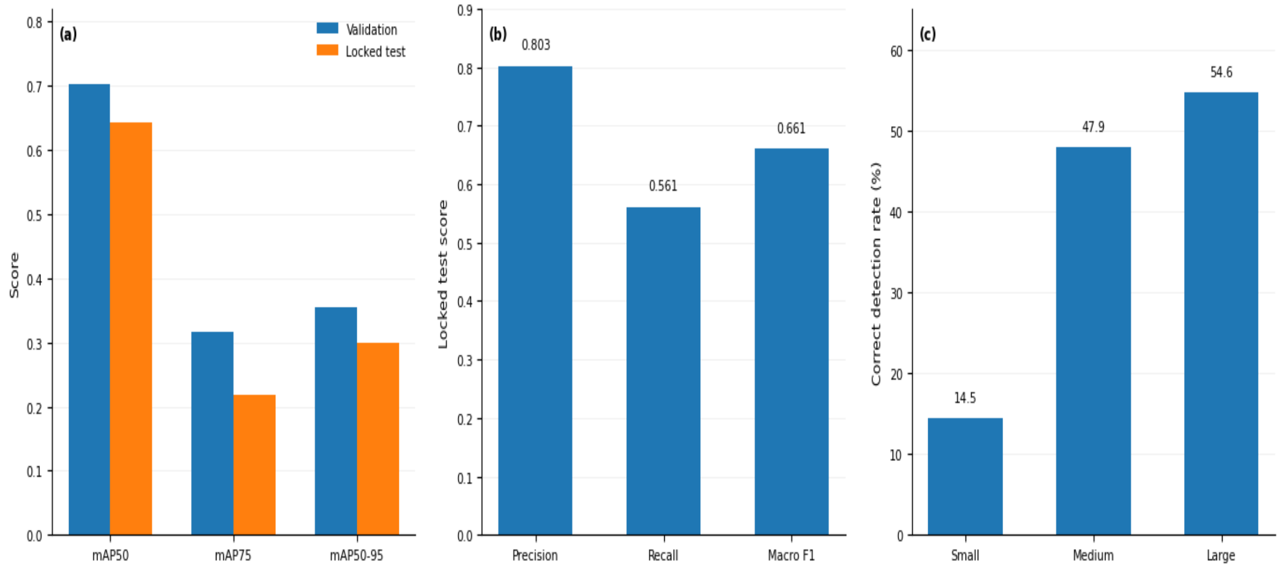

Figure 4. Validation-to-Test Generalization, Fixed-Threshold Detection Performance, and Object-Scale Sensitivity

Figure 4 shows the decline in localization performance from validation to test, the fixed operating-point detection metrics, and the marked dependence of correct-detection rate on object size. The strongest relative deterioration occurred at mAP75, while small objects exhibited the lowest correct-detection rate.

Correct-detection rate declined from 0.646226 on validation to 0.461176 on test, corresponding to an observed change of −0.185050. The composition effect was −0.000005, whereas the within-stratum effect was −0.185045, indicating that the observed decline was almost entirely associated with performance deterioration within class–size strata rather than changes in their relative composition. In the perturbation analysis, the 512 candidate exceeded the baseline under only two of six non-clean conditions. Table 4 consolidates these diagnostic findings.

Table 4. Uncertainty, Error Distribution, Robustness, and Generalization of the Final Detector

Indicator	Result
Bootstrap custom mAP50–95	0.300116 [0.280361, 0.328255]
Correct / low-confidence / localization / confusion / miss	46.12 / 37.06 / 10.00 / 0.41 / 6.41%
mAP50–95 validation-to-test change	−0.056596 (−15.90%)
mAP75 validation-to-test change	−0.097962 (−30.92%)
Composition / within-stratum effect	−0.000005 / −0.185045
512 candidate better than baseline	2/6 perturbations

Table 4 shows the bootstrap interval around custom mAP50–95 and the concentration of generalization loss within class–size strata. It also records the limited perturbation advantage of the 512 candidate.

4.7 Integrated Manufacturing Evidence and Reproducibility

The final package retained two evidence layers: KUKA KR16 robotic process execution and YOLO11n aircraft visual inspection. No direct model fusion was performed. Deterministic qualitative sampling selected 12 test examples: three correct detections, three low-confidence correct localizations, two localization errors, two class confusions, and two misses/no-useful-prediction cases. As illustrated in Figure 5, selection used reproducible hash-based ranking rather than manual choice.

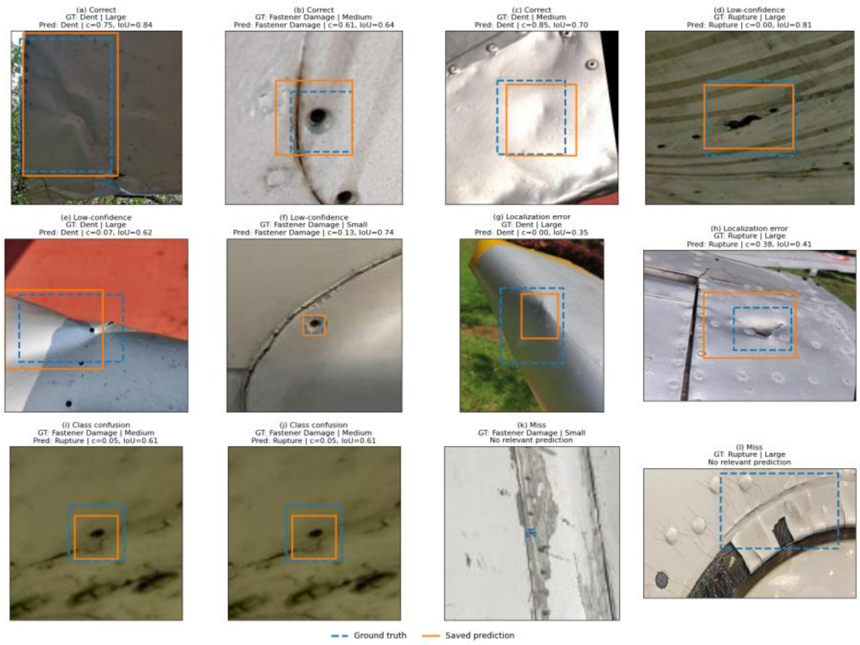

Figure 5. Deterministic Qualitative Examples from the Final Aircraft-Defect Test Set

Figure 5 shows predefined outcome categories using ground-truth and saved prediction boxes. Panels display confidence and IoU without changing the fixed threshold.

Final auditing checked nine lineage records and 26 consistency conditions; all 26 passed. Twenty-nine current authoritative artifacts were hashed, one historical size-specific artifact was isolated as superseded, and the robotic and visual modules remained separate complementary manufacturing components.

5. Discussion

The results show that the quality of the execution of the robot, and visual inspection accuracy are not enough to reduce the precision-oriented manufacturing of aerospace products. The robotic module revealed that temporal and engineering representations were not helpful in achieving good separation between Mounted, Jammed, and Not-Mounted outcomes. This implies that assembly-state variations occur in both across similar kinematic and interaction patterns and in summarzing indicators which do not exist in isolation. This could explain why the process of aerospace robotic assembly is still susceptible to compliance, uncertainty of contact, and process variations, making deterministic failure state characterization challenging (Zhao et al., 2026). The lack of FDR significant KPI families also suggests that process-level indicators should not be interpreted as statistically valid evidence of loss of precision when they show deviations from their descriptions, as can be seen, further reinforcing conservative interpretation of process-level indicators.

Another failure of the classification was the poor performance of the classification on higher dimensional temporal descriptors in the case of heterogeneous or transient failure mechanisms. Real-time force modelling demonstrated that force data in aircraft riveting has useful information of the process, but how process transient phases and structural responses are represented is very important (Kang et al., 2024). Examples of short-duration signatures that could be masked in the present study include those associated with jamming or partial mounting. This inspires the development of sequence models that are trajectory aware, feature extraction specific to the phases of a sequence, and multimodal temporal fusion, over the growth of static feature spaces.

The pattern was reversed for the visual-inspection module; the overall precision was not as good, while the recall and strict localization was not as good. This indicates that detection was generally correct with some confidence but there were a small number of defects that were not detected or scored with confidence, and/or were not localized accurately. This behavior is in line with the confirmed aircraft surface detection research obtained that the architectural improvements enhance the representation of the aircraft surface while small, irregular and low-contrast defects are hard to be located consistently (Huang et al., 2024). The sharp drop at more stringent IoU levels is significant as it differentiates between coarse defect recognition and precise localization of the defect in the image, with the latter being more important if inspection results are used for decisions on maintenance or production.

The scale dependence is another example of the fact that pixel support is a key limitation of the reliability of inspection. Resolution and representation of features by pyramids are still limiting factors as the rate of correct detection of small objects was very low compared with medium and large defects. This relates to high precision digital-twin and robotic manufacturing systems where geometric fidelity and sensing resolution ranges is the dichotomy between operationally useful model outputs and statistically acceptable model outputs (Kang et al., 2025). The near zero change in composition effect in the validation-to-test decline is informative because the composition effect was not due to a change in test-set mixture, but rather from within class-size strata. Thus, the deterioration is attributed to the loss of within-stratum transferability, which is not merely rebalancing of the datasets.

From a methodological standpoint results indicate separation of development and selection (validation only), locked testing, bootstrap uncertainty, and post hoc diagnostics. Heterogeneous validation practice and dataset construction have been identified as challenges for reliable comparison from industrial defect-detection reviews (Ameri et al., 2024). To alleviate this worry, the workflow described here includes the checkpoint feature, enabling the identification of the model and the avoidance of reassignment of historical results to the last detector. Both theoretically and practically, the study can serve as a foundation for a multilevel approach to precision manufacturing that includes two different types of evidence: process consistency and defect verification.

Several limitations remain. The two modules were not a single physical entity, the robot didn't give dimensional metrology and the visual detector worked only with fixed datasets and CPU-experimentation. The literature on aircraft skin inspection also states that the deployment reliability is dependent on the condition in which the aircraft is deployed, which is beyond what can be represented in the benchmark images (Li et al., 2025). The findings from this paper needs to be incorporated in the future work by considering synchronized robot–vision acquisition, sequence-based robotic models, force–vision fusion, external multi-site validation, higher resolution small-object detectors, dimensional tolerance measurements, and prospective factory evaluation. Finally, closed-loop systems need to consider if feedback from the inspection could be used to adapt the parameters of robotic assembly tasks and if the combined effect of the feedback and robotic actions could measurably improve manufacturing quality.

6. Conclusion

In this study, a two-module setup for precision-oriented aerospace manufacturing was introduced: KUKA KR16 robotic fastener-process characterization and AI-assisted aircraft defect inspection. The robotic analysis revealed that temporal kinematic and dynamic descriptors, engineering KPIs, nominal-deviation measures, and precision-envelope indicators were useful for process characterization, but less effective at distinguishing the different types of aircraft assembly, the visual module showed that YOLO11n was able to provide good precision in defect detection but poor recall, stringent localization requirements, and greater sensitivity to small defects. Together these results demonstrate that the manufacturing precision should be considered a multi-layer phenomenon which includes not only positional accuracy, but also process consistency and interaction stability as well as independent quality verification. In practice, the work presented here suggests that the use of RPA and visual AI can be applied as additional layers of decision-support in aerospace manufacturing, provided that uncertainty, scale of objects and generalization and provenance of checkpoints are clearly managed. Leakage-proof data governance, set validation and test procedures and small-defect monitoring, and human review for low-confidence or uncertain data detections, should thus be a priority for industrial deployment. The framework also provides a repeatable ground for comparing manufacturing processes without suggesting or assuming model fusion or, unsupported, improvements in dimensional tolerances or accuracy. Sequenced temporal learning, force–vision fusion, dimensional metrology, higher resolution detection architectures, external multi-site validation, and prospective factory validation should be taken up in future research and work. Closed-loop inspection feedback in such work should eventually be able to prove that it can tune the robotic assembly parameters and lead to measurable improvements in the manufacturing quality, repeatability and defect prevention.

References

1. Ameri, R., Hsu, C. C., & Band, S. S. (2024). A systematic review of deep learning approaches for surface defect detection in industrial applications. Engineering Applications of Artificial Intelligence, 130, 107717.
2. da Silva Santos, K. R., Villani, E., de Oliveira, W. R., & Dttman, A. (2022). Comparison of visual servoing technologies for robotized aerospace structural assembly and inspection. Robotics and Computer-Integrated Manufacturing, 73, 102237.
3. Giardini Lahr, G. J., Segreto Silva, T. H., Moreira, G. R., Ajoudani, A., Boaventura, T., & Caurin, G. A. P. (2023). Robotic assembly data of threaded fasteners: Aeronautical collars (Version 1) [Dataset]. Mendeley Data. https://doi.org/10.17632/26674p3hvg.1
4. Hémono, P., Chabane, A. N., & Sahnoun, M. H. (2025). Multi objective optimization of human–robot collaboration: A case study in aerospace assembly line. Computers & Operations Research, 174, 106874.
5. Huang, B., Ding, Y., Liu, G., Tian, G., & Wang, S. (2024). ASD-YOLO: An aircraft surface defects detection method using deformable convolution and attention mechanism. Measurement, 238, 115300.
6. Jayasekara, D., Lai, N. Y. G., Wong, K. H., Pawar, K., & Zhu, Y. (2022). Level of automation (LOA) in aerospace composite manufacturing: Present status and future directions towards industry 4.0. Journal of Manufacturing Systems, 62, 44-61.
7. Jha, S. B., & Babiceanu, R. F. (2023). Deep CNN-based visual defect detection: Survey of current literature. Computers in Industry, 148, 103911.

8. Kang, R., Hu, J., Shi, Z., Zhang, J., Li, Z., Li, Z., & Tian, W. (2025). A high-precision digital twin modeling approach for the serial-parallel hybrid drilling robot in aircraft assembly. Robotics and Computer-Integrated Manufacturing, 94, 102986.
9. Kang, Y., Wang, T., Song, S., Li, G., & Chen, Y. (2024). Real-time modeling of the riveting process forces for aircraft panel structures. Journal of Manufacturing Processes, 124, 90-109.
10. Li, H., Hou, Z., Wang, C., & Lu, T. (2025). A review of automated aircraft skin defects detection. Measurement, 120019.
11. Li, H., Wang, C., & Liu, Y. (2025). Aircraft skin defect detection based on Fourier GAN data augmentation under limited samples. Measurement, 245, 116657.
12. Li, S., Zheng, P., Pang, S., Wang, X. V., & Wang, L. (2023). Self-organising multiple human–robot collaboration: A temporal subgraph reasoning-based method. Journal of manufacturing systems, 68, 304-312.
13. Meng, D. I. N. G., Boer, W. U., Juan, X. U., & Hongfu, Z. U. O. (2022). Visual inspection of aircraft skin: Automated pixel-level defect detection by instance segmentation. Chinese Journal of Aeronautics, 35(10), 254-264.
14. Shang, H., Sun, C., Liu, J., Chen, X., & Yan, R. (2023). Defect-aware transformer network for intelligent visual surface defect detection. Advanced Engineering Informatics, 55, 101882.
15. University of Technology Sydney. (2024). Aircraft Defect Detection Dataset [Open source dataset]. Roboflow Universe. Roboflow. https://universe.roboflow.com/university-of-technology-sydney-21uto/aircraft-defect-detection
16. Wang, B., Zhang, J., & Wu, D. (2025). Force–vision fusion fuzzy control for robotic batch precision assembly of flexibly absorbed pegs. Robotics and Computer-Integrated Manufacturing, 92, 102861.
17. Wang, J., Xu, G., Yan, F., Wang, J., & Wang, Z. (2023). Defect transformer: An efficient hybrid transformer architecture for surface defect detection. Measurement, 211, 112614.
18. Zhang, Q., Zheng, S., Yu, C., Wang, Q., & Ke, Y. (2022). Digital thread-based modeling of digital twin framework for the aircraft assembly system. Journal of Manufacturing Systems, 65, 406-420.
19. Zhang, R., Lv, J., Li, J., Bao, J., Zheng, P., & Peng, T. (2022). A graph-based reinforcement learning-enabled approach for adaptive human-robot collaborative assembly operations. Journal of Manufacturing Systems, 63, 491-503.
20. Zhang, Z., Wang, Y., Zhang, Z., Wang, L., Huang, H., & Cao, Q. (2024). A residual reinforcement learning method for robotic assembly using visual and force information. Journal of Manufacturing Systems, 72, 245-262.
21. Zhao, L., Li, X., Zhao, H., Zhang, J., & Ding, H. (2026). A comprehensive review of robotic assembly of aerospace components. Journal of Manufacturing Systems, 88, 350-375.
22. Zheng, Y., Liu, W., Zhang, Y., Han, L., Li, J., & Lu, Y. (2025). Integration and calibration of an in situ robotic manufacturing system for high-precision machining of large-span spacecraft brackets with associated datum. Robotics and Computer-Integrated Manufacturing, 94, 102928.