UNIVERSITY OF WEST ATTICA

eISSN:3136-7658

Applied Manufacturing Intelligence Journal

Vol. 5, No. 2, April, 2026



Explainable Res2Net-SE Model for Quality Estimation of Robotic Screw Fastening

Recep Urfalı[a,*]

[a] *Faculty of Engineering, Cukurova University, Adana, 01330, Turkey*

Received 09 February 2026; Revised 06 May 2026; Accepted 09 May 2026; Published 13 May 2026

Abstract

Reliable quality estimation for robotic screw fastening is required to prevent loosening, thread damage, seal failure and late-stage product rejection caused by a single fastening error. This paper presents an explainable Res2Net-SE model for screw-fastening quality prediction based on torque-angle curves, axial pressure, spindle speed, depth displacement and visual inspection indicators. The fastening window is aligned at the torque-rise and seating stages, and multi-scale process features are used to represent peak torque, tightening slope, pressure stability, residual vibration and terminal depth. Res2Net blocks learn short-term impulses and gradual fastening trends within the same residual structure, while squeeze-and-excitation weighting enhances the signal channels that provide the most reliable evidence. The model outputs a quality score, confidence value, dominant evidence channel and review priority. A simulated robotic assembly dataset of 4,080 fastening windows was generated for normal fastening, under-tightening, over-tightening, thread slip, tilted screw entry and low-confidence mixed cases. The proposed model achieved 93.4% accuracy and 90.7% macro-F1, outperforming CNN, LSTM, Transformer and ResNet baselines. Torque noise and pressure drift still reduce recall for thread-slip and under-tightening cases. The results show that torque-angle data are more indicative of over-tightening, whereas pressure-depth data are closely associated with tilted entry and incomplete seating. Interpretable multi-scale feature learning can therefore support fastening quality assessment, rework prioritization and traceable robotic assembly inspection.

Keywords: Robotic Screw Fastening; Quality Estimation; Res2Net-SE; Explainable Learning; Assembly Inspection

1. Introduction

Robotic screw fastening is widely used in electronics, vehicles and precision equipment because it improves assembly repeatability and reduces manual variation [1]. Modern manufacturing lines increasingly require each fastening operation to be monitored as a traceable quality event rather than as a simple completion signal [2]. Torque-angle curves provide important evidence about seating, clamping and thread interaction during the tightening process [3]. Vision-assisted assembly systems further show that process signals must be interpreted together with screw position, tool alignment and surface condition [4]. In this context, the quality of a locked screw depends not only on the final torque value but also on the evolution of force, rotation and contact state within the whole fastening window. A screw may reach the nominal torque while still suffering from tilted entry, thread slip, insufficient seating or temporary tool vibration. Therefore, a data-driven quality estimator should recognise the process pattern before a defective product enters the next station.

Corresponding author: recep.ur@cu.edu.tr

https://doi.org/10.68406/amij.2026.vol5iss2np4:43-55

➢ This is an open access article under the CC BY-NC-SA 4.0 license (https://creativecommons.org/licenses/by/4.0/).

© Authors retain all copyrights.

The rule for the final torque of the traditional fastening inspection is a limit. Although the rules are simple to implement, they are not suitable for cases where the same final value can be achieved by several fastening methods. Recently, in the field of intelligent manufacturing, process-level monitoring has become more reliable with the addition of both transient signals and engineering constraints [5]. Research on robotic assembly has also shown that local disturbances, tool wear and part tolerance can alter the concept of the tightening curve [6]. A small-torque peak may be suitable for one material, but there will be thread interference in other places. Tighten in steps to reach the standard pressure; otherwise, if the screw head is not in contact with the surface, it will be inaccurate. The modified items are now judgments of quality estimation and no longer binary threshold problems.

Although deep learning can address the non-linear characteristics of fastening evidence, a general convolutional model may be unable to extract both long-term and short-term patterns at different scales. Multi-scale residual learning can learn local impulses and slow-fastening trends, and channel attention selects which signal sources support the quality judgment [7]. Explainable industrial artificial intelligence needs to output evidence, confidence and review priority from the model, rather than just a pass-or-fail label [8]. A timely inspection of the screw-fastener will prevent disruptions in the production process, damage to tools, and reduced quality of the final product. According to the above experimental data, a model has been constructed to help differentiate between fastener defects and sensing errors. Therefore, this paper introduces an explainable Res2Net-SE model for robotic screw-fastening quality estimation.

The remainder of this paper follows the quality-estimation workflow described above. Section 2 reviews robotic fastening monitoring, multi-scale feature extraction and explainable assembly inspection. Section 3 introduces fastening signal-window construction, the Res2Net-SE feature encoder, confidence estimation and evidence-output design. Section 4 reports simulated data analysis, comparison with baseline models, error and confidence analysis, explainable evidence and engineering review value. Section 5 summarizes the main findings, limitations and future extensions for production-line validation. This organization separates process evidence modelling from production decision support while preserving the connection among fastening class, quality score, confidence, dominant evidence channel and review priority.

2. Related Work

2.1. Robotic Screw Fastening Quality Monitoring

Torque, Angle, Rotation Speed and Tool Position signals are typically employed to observe robotically screwed-in fasteners. Many enterprises still use the high-end end-torque threshold because it is convenient to inspect and meets the requirements of the products. According to the study of process monitoring, it has been found that end-of-line inspection cannot detect transient problems such as thread slip, material deformation or tool-seat mismatch [9]. This defect is relatively serious for automatic assembly; the robot will still take the programmed tightening path even in the event of abnormal mechanical conditions.

All of the above content can be displayed in a torque-angle graph. The inclined surface is connected to the thread contact and stiffness, and the end section is fixed and clamped stably. Research on intelligent fastening control shows that torque-angle patterns can differentiate between under-tightening and over-tightening better than single-point indicators [10]. However, the curve does not always show the actual cause of the defect. A reduced torque of a slant screw may be due to damage to the tool bit or a dirty surface.

Multi-source inspection has thus been required. Axial pressure is tool engagement and part contact. Spindle Current = Drive Load. Displacement and Depth signals indicate whether the screw has reached the position for seating. Detection of head misalignment, surface occlusion and missing parts by vision. Research on robotic assembly inspection shows that, through signal fusion, defect diagnosis can be improved by considering the reliability of each sensor [11]. According to the above instructions, models for torque, pressure, speed, depth and visual inspection indicators will be constructed as auxiliary evidence in this study.

2.2. Multi-Scale Feature Extraction for Industrial Signals

Fastening signals are either slow changes or sudden increases. A slow decline is a gradual decrease, and the local peaks may be caused by temporary contact, a thread jump, or tool vibration. The original one-dimensional convolution operation is either localised to impulses or spread out over a large area, and neither type of evidence can be captured by a single representation. Recent research on multi-scale residual networks has shown that hierarchical receptive fields can improve the recognition of industrial signal patterns with different durations [12]. This attribute is related to screw fastening, and some types of defects only occur in a small range of the fastening window.

The style of the signal is suitable for Res2Net-style feature extraction, and a residual block that divides several scale groups and connects them hierarchically is used. A few compact modules are used to obtain multi-scale features

instead of several convolution modules. Industrial diagnosis studies have shown that multi-scale residual features improve the robustness of noise and operating condition variations [13]. Therefore, a single model of a robot's fastening can be employed to learn torque peaks, pressure fluctuations, and depth residuals simultaneously.

Attention Mechanisms further enhance feature selection. Squeeze-and-excitation modules generate channel weights based on the global context to increase the weight of some channels and decrease that of others. Recently, studies on manufacturing vision and signal inspection have shown that channel attention is effective when the evidence of defects appears in only a few sources or when different operating states require different features [14]. Generally, an over-tightening of the fastening operation is shown by torque-angle data, and insufficient seating is more often indicated by depth and pressure indicators.

2.3. Explainable Attention in Assembly Inspection

Explainability is now required for practical application in automated quality inspection. A model of the assembly line should be able to predict, and also explain which evidence led to the warning. Research on explainable industrial AI indicates that confidence, attribution of evidence and review priority are required for human-machine cooperation in quality management [15]. It is directly related to screw fastening, so at this time, one must choose among acceptance, re-measurement and rework, or stop the production line.

There are two types of fastener inspections. First, provide the cause at the level of signals; that is to say, why did this occur? Secondly, engineering interpretation is to map the signal pattern to a particular kind of defect. Research on interpretable maintenance decision support indicates that explanations are more helpful when they are connected to repair or inspection actions [16]. Thus, the model will output the quality probability, as well as the main evidence path and review order.

Another problem is also instability. Assembly data may have sensor drift, speed changes, size variations of components, and poor image quality. Research in industrial perception has shown that a model evaluated only on clean samples tends to overestimate the deployment performance [17]. Disturbance may cause the model to mistake low-confidence sensing for actual quality defects in robotic fastening. Therefore, an excellent estimator will show the level of confidence and save a separate re-measurement record.

Based on the above research, process signals, multi-scale features and explainable outputs should be integrated. Most of the current models focus only on the final defect classification or use internal attention without external connections. Recently, quality analytics research has proposed that inspection models should be organised around decision support and auditability [18]. Therefore, this paper constructs an explainable Res2Net-SE model for fastening quality prediction and presents the evidence in a form that is convenient for engineering inspection.

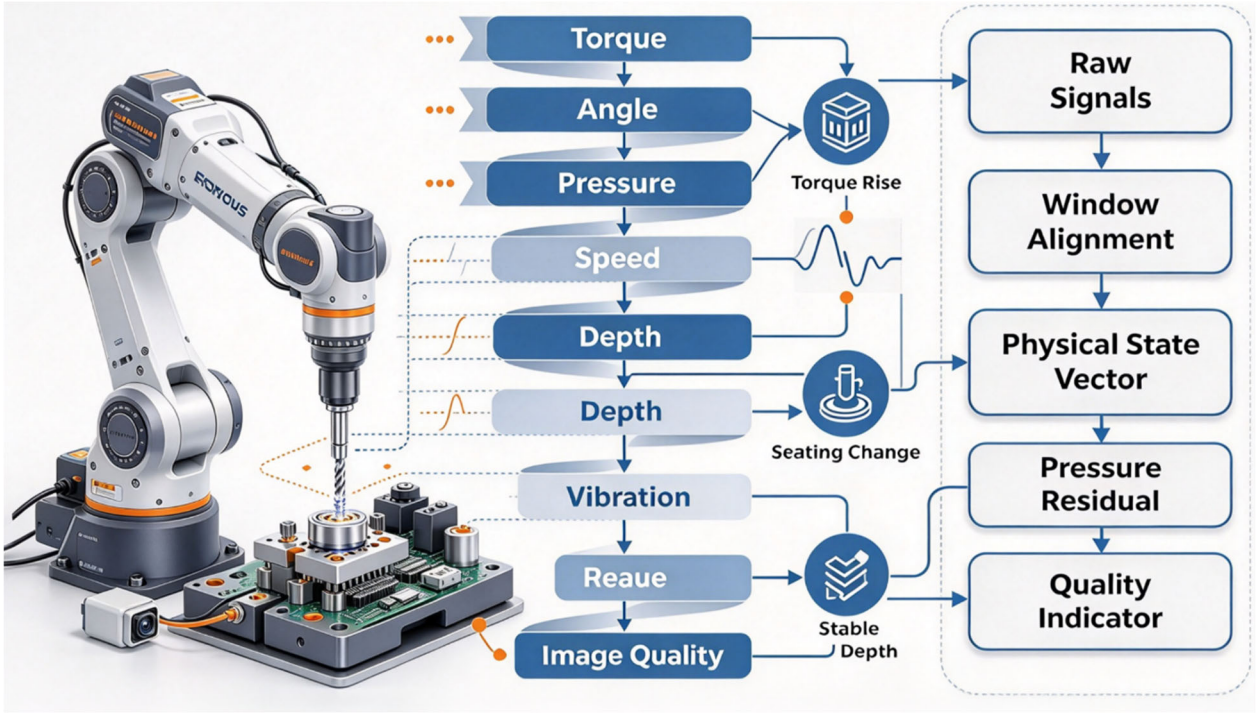


Figure 1. Robotic Screw Fastening Signal Alignment Workflow

3. Material and Method

3.1. Fastening Signal Window Construction

The input of the proposed method is a fastening window obtained from one robotic screw-locking operation. The window begins before torque rise and ends after the seating process. Torque, rotation angle, axial pressure, spindle speed, depth displacement, residual vibration and image-quality indicators are appended to the same aligned record. Figure 1 shows the fastening signal alignment workflow. raw tool signals, pressure evidence and visual inspection indicators are aligned at torque rise and then represented within a shared quality-estimation window.

A certain period is needed for alignment due to different materials, screws and tools, speeds, etc., of fastening. A fixed sample index would consider the slow- and fast-fastening processes to be different events, even if they have the same mechanical meaning. Therefore, the three anchors of this method are: the first torque increase, seating slope change and terminal depth stabilisation. The anchors can be used to compare fastening windows based on the process stage, not on absolute time. The aligned window contains evidence of both the start and end of the engagement.

The first kind is the process-state vector. At each aligned time step, the Torque, Angle, Pressure, Speed, Depth residual and sensing-quality data are stored. A vector is defined as follows:

$$x_t = [T_t, A_t, P_t, S_t, D_t, Q_t] \tag{1}$$

In this expression, T_t denotes torque, A_t denotes rotation angle, P_t denotes axial pressure, S_t denotes spindle speed, D_t denotes depth residual and Q_t denotes signal quality at time t. The vector is not a defect label. It is a compact representation of the physical fastening state that will be processed by the feature encoder.

The pressure residual is used because tool pressure does not always increase linearly with torque. In a normal process, pressure becomes stable after screw seating. In tilted entry or incomplete seating, pressure may fluctuate even if torque reaches the preset threshold. The residual form makes this instability easier to detect.

$$e_t^P = P_t - \text{mean}(P_{t-k:t}) \tag{2}$$

The residual e_t^P compares the current pressure with the local pressure mean over a short window. It helps the model recognise sudden contact changes and separates ordinary tightening slope from unstable tool engagement. This residual is later combined with torque and depth features to support quality estimation.

The signal-quality indicator is also retained. Torque noise, visual blur and pressure drift can reduce the reliability of the prediction. Removing such samples would make the model less useful in production, because real stations often contain partial disturbances. The quality indicator allows the model to distinguish weak evidence from a true defect pattern.

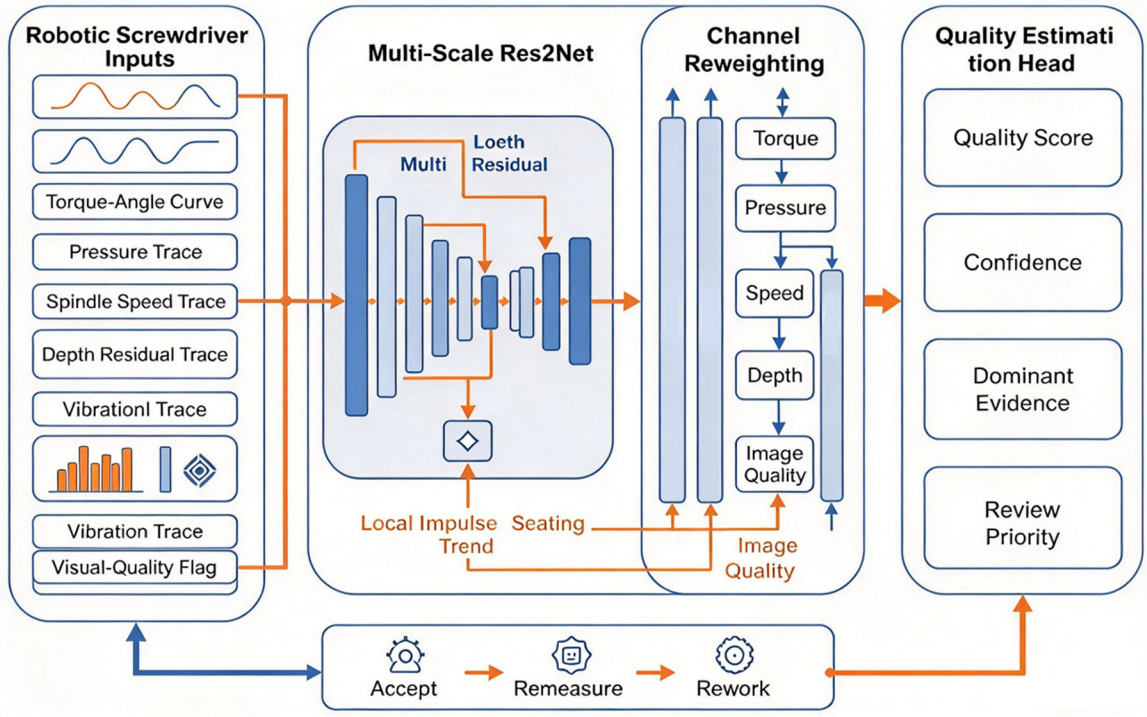


Figure 2. Explainable Res2Net-SE Architecture for Screw-Fastening Quality Estimation

3.2. Res2Net-SE Feature Extraction

Figure 2 shows the explainable Res2Net-SE architecture for screw-fastening quality estimation. The left side receives the aligned fastening signal and visual inspection indicators. The central module uses multi-scale residual groups to extract local torque impulses, pressure oscillations and slower seating trends. The right side applies SE channel recalibration and outputs quality score, confidence and evidence attribution. This architecture makes the prediction traceable because each channel contribution can be related to a specific process variable rather than to an anonymous feature vector.

The Res2Net block divides the input feature map into several scale groups. Each group processes part of the feature channels and exchanges information with the next group. This structure allows the model to capture local torque impulses, pressure oscillation and slow depth trends in the same residual block.

$$F_i = G_i(F_{i-1}) + X_i \tag{3}$$

Here, X_i is the input feature group, G_i is the convolutional transformation at scale i, and F_i is the resulting multi-scale feature. The hierarchical connection allows later groups to include information from earlier groups, so the network can combine short and long fastening patterns without creating a very wide model.

After multi-scale extraction, the squeeze-and-excitation module calculates channel weights. The weights are produced from global context and represent the relative importance of torque, pressure, speed, depth and visual indicators for the current fastening window.

$$w_c = \mathrm{sigmoid}\big(W_2\delta(W_1 g_c)\big) \tag{4}$$

In this formula, g_c is the global descriptor of channel c, W_1 and W_2 are trainable matrices, and w_c increases the contribution of important channels while reducing weakly relevant channels.

The final encoded feature is obtained by multiplying the multi-scale feature with the SE channel weight. This step keeps the learned signal pattern and makes the explanation easier to report, because each channel receives an explicit contribution score.

$$Z_c = w_c F_c \tag{5}$$

The encoded feature is used for both quality estimation and evidence attribution. When torque and angle dominate the output, the model indicates a fastening-force pattern such as over-tightening or under-tightening. When pressure and depth dominate the output, the model indicates seating instability, tilted entry or incomplete seating. When visual quality dominates uncertainty, the model recommends remeasurement rather than immediate rework. This separation of evidence type and review action is central to the explainable design.

3.3. Quality Score and Evidence Output

The model converts encoded features into a quality score. The score estimates whether the fastening process belongs to normal fastening, under-tightening, over-tightening, thread slip, tilted entry or low-confidence mixed cases. The classification layer uses the fused feature and produces class probabilities.

$$p_y = \mathrm{softmax}\big(W_q Z + b_q\big) \tag{6}$$

The probability p_y describes the support for each fastening class. A high value for a defect class indicates strong evidence, but this value alone does not determine the review action. The confidence of the evidence must also be considered because low-quality sensing can produce unstable predictions.

The confidence value is calculated from the prediction distribution and the quality indicator. A sharp probability distribution with reliable signals produces high confidence, while uncertain probabilities or poor signal quality reduce confidence.

$$c = 1 - H\big(p_y\big) + \lambda \mathrm{mean}(Q) \tag{7}$$

In this expression, $H\big(p_y\big)$ denotes prediction entropy, Q is the signal-quality vector and λ is a weighting coefficient. The confidence value is used to separate real defects from cases that require remeasurement.

The final review priority is generated from quality probability and confidence. A severe defect with high confidence is assigned to immediate rework, while a possible defect with low confidence is assigned to remeasurement. A normal result with high confidence is accepted, and a moderate abnormal trend is stored for scheduled audit. The output is therefore designed for production review because it separates defect severity from the reliability of the underlying evidence.

$$R = \text{sigmoid}(w_r^T Z + \alpha(1 - c)) \quad (8)$$

The review score R increases when the encoded feature indicates quality risk or when confidence is low. The method therefore avoids treating an uncertain sample as fully acceptable. At the same time, it does not stop the line unnecessarily when the only problem is poor image quality or temporary signal interruption. The output contains fastening class, quality score, confidence, dominant evidence channel and review priority.

4. Results and Discussion

4.1. Sample Construction and Fastening-State Statistics

A total of 4,080 fastening windows were generated from robotic screw-locking experiments under normal and abnormal conditions [19]. Figure 3 presents the arrangement of sample construction and quality-state statistics. Figure 3(a) shows normal fastening, under-tightening, over-tightening, thread slip, tilted entry and low-confidence mixed cases. Figure 3(b) presents torque-range coverage, showing that several abnormal states overlap with the normal torque range. Figure 3(c) presents pressure fluctuation, where tilted entry and incomplete seating are more unstable. Figure 3(d) presents visual-review categories for identifying low-confidence image conditions. These four subfigures show that fastening quality cannot be represented by a single terminal value.

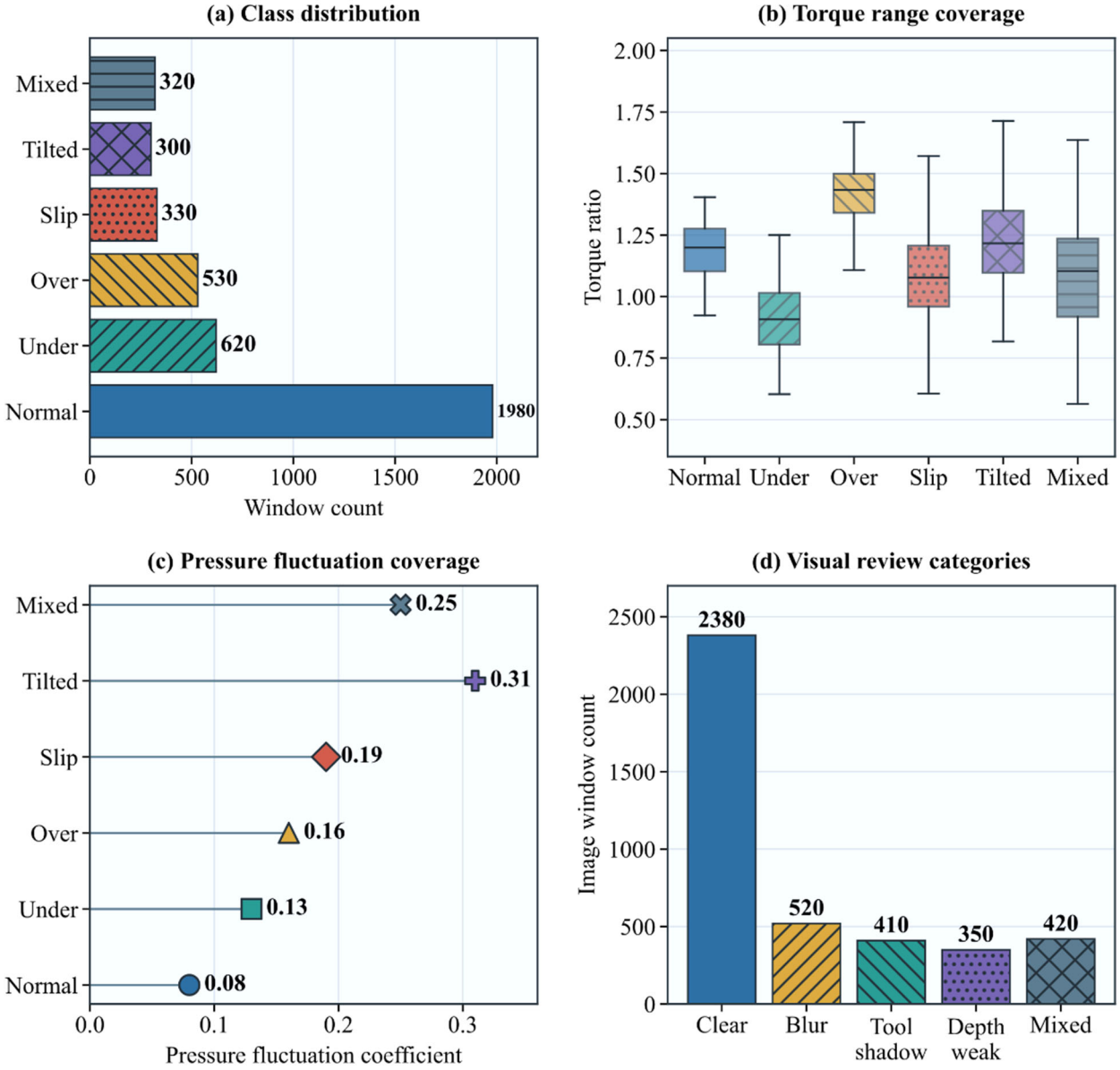


Figure 3. Fastening Sample Construction. (a) Class distribution. (b) Torque range coverage. (c) Pressure fluctuation coverage. (d) Visual review categories.

The sample construction is intentionally uneven because fastening defects occur less often than normal fastening in production. The normal class accounts for about 48.5% of the samples, whereas thread slip and tilted entry are less

frequent. This distribution requires the model to learn the boundary between acceptable process variation and abnormal fastening without inflating performance through an artificially balanced class structure. Distribution [20] is consistent with intelligent inspection research, where rare abnormal states should be treated as difficult engineering cases rather than oversampled without production meaning.

According to the torque-angle curve, over-tightening generally produces a high terminal torque and a steep slope at a high angle, whereas under-tightening shows lower terminal torque and weaker seating evidence. Thread slip may appear as a local torque drop followed by a repeated rise, so the defect can be present before the final torque appears acceptable. Similar observations in robotic fastening monitoring [21] indicate that process-window evidence is more informative than a single final torque value. The model therefore uses the full fastening window to support early inspection and review.

Pressure and depth evidence provide additional discrimination. Tilted entry may create unstable pressure even when the target-torque range is reached, and incomplete seating may appear as a depth residual rather than as a torque anomaly. Process-monitoring studies of automated assembly [22] show that auxiliary signals reduce ambiguity when multiple defect types share similar torque ranges. For this reason, pressure and depth are treated as independent evidence sources, not as secondary annotations added after torque-based classification.

4.2. Model Comparison and Estimation Accuracy

The proposed model is compared with CNN, LSTM, ResNet, Transformer and ResNet-SE baselines. As shown in Figure 4, the proposed structure improves both overall recognition and abnormal-state balance. Figure 4(a) reports accuracy, where the proposed Res2Net-SE model reaches 93.4%. Figure 4(b) reports macro-F1, where the proposed model reaches 0.907 and improves recognition of rare abnormal states. Multi-scale representation [23] is relevant because local thread slip, tilted entry and under-tightening require features at different temporal scales rather than a single convolutional span.

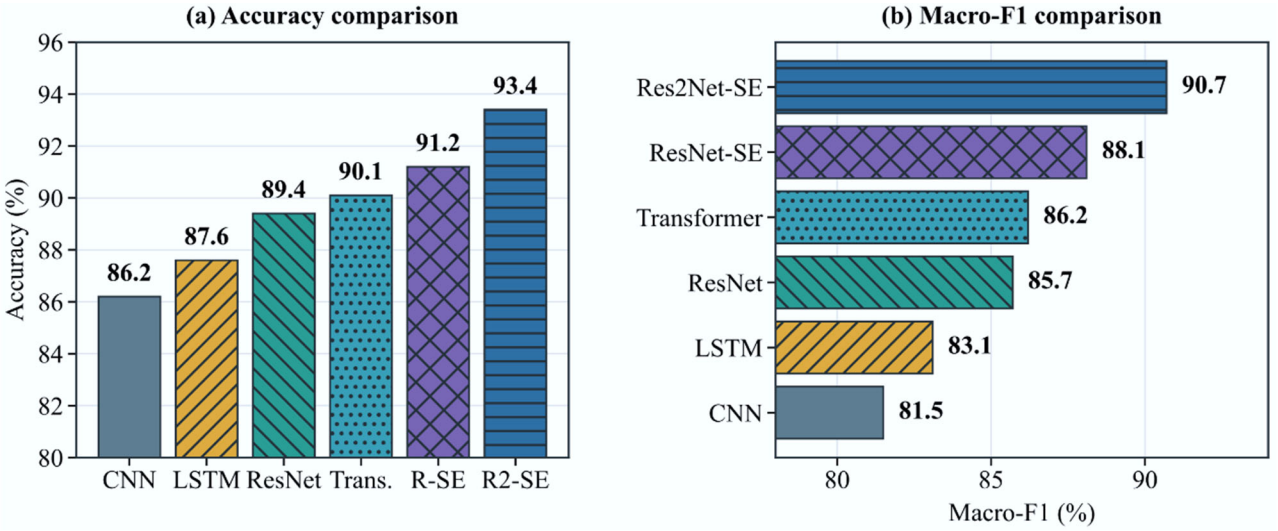


Figure 4. Model Comparison Results. (a) Accuracy comparison. (b) Macro-F1 comparison.

Table 1 presents the comparative results under the same disturbance conditions. The model has a small proportion of review errors and high recall for thread slip and tilted entry. Both are difficult because the final torque values are close to those of the normal samples. Based on the above results, several scales of residual features and channel recalibration can be employed to differentiate between local torque events and widespread seating instability. The mean review error is also lower in the proposed model (0.107) compared with that of the ResNet-SE baseline (0.131); thus, it is more suitable for ranking samples based on inspection priority. At that time, more uncertain and serious cases are generally dealt with first by engineers at the construction site. The numerical difference in production management is also considerable; a small rise in rare-defect recalls will lead to several follow-up inspections. Therefore, in addition to the model's class accuracy and recall, we will also examine the distribution of review errors and confidence. Recently, research on robust industrial diagnosis [24] has also shown that, for rare abnormal situations, overall accuracy is insufficient.

Table 1. Numeric Statistics of Model Performance Under Fastening Disturbances

Model	Accuracy (%)	Macro-F1 (%)	Thread-Slip Recall (%)	Tilted-Entry Recall (%)	Mean Review Error
CNN	86.2	81.5	78.4	76.9	0.184

LSTM	87.6	83.1	80.2	78.5	0.171
ResNet	89.4	85.7	82.6	81.1	0.153
Transformer	90.1	86.2	84.3	82.7	0.146
ResNet-SE	91.2	88.1	86.9	84.8	0.131
Proposed Res2Net-SE	93.4	90.7	89.8	88.3	0.107

Figure 5 shows error and confidence analysis for ambiguous fastening conditions. Figure 5(a) presents the distribution of estimation errors, where most proposed-model errors remain within the acceptable review band. Figure 5(b) presents confidence distribution and shows that low-confidence samples concentrate in mixed cases and visually weak samples. Figure 5(c) presents false alarms caused by transient torque peaks after normal seating impact, and Figure 5(d) presents missed abnormal cases, mainly light thread slips under stable pressure. This error pattern [25] supports the use of confidence as a graded review signal instead of an automatic rework command.

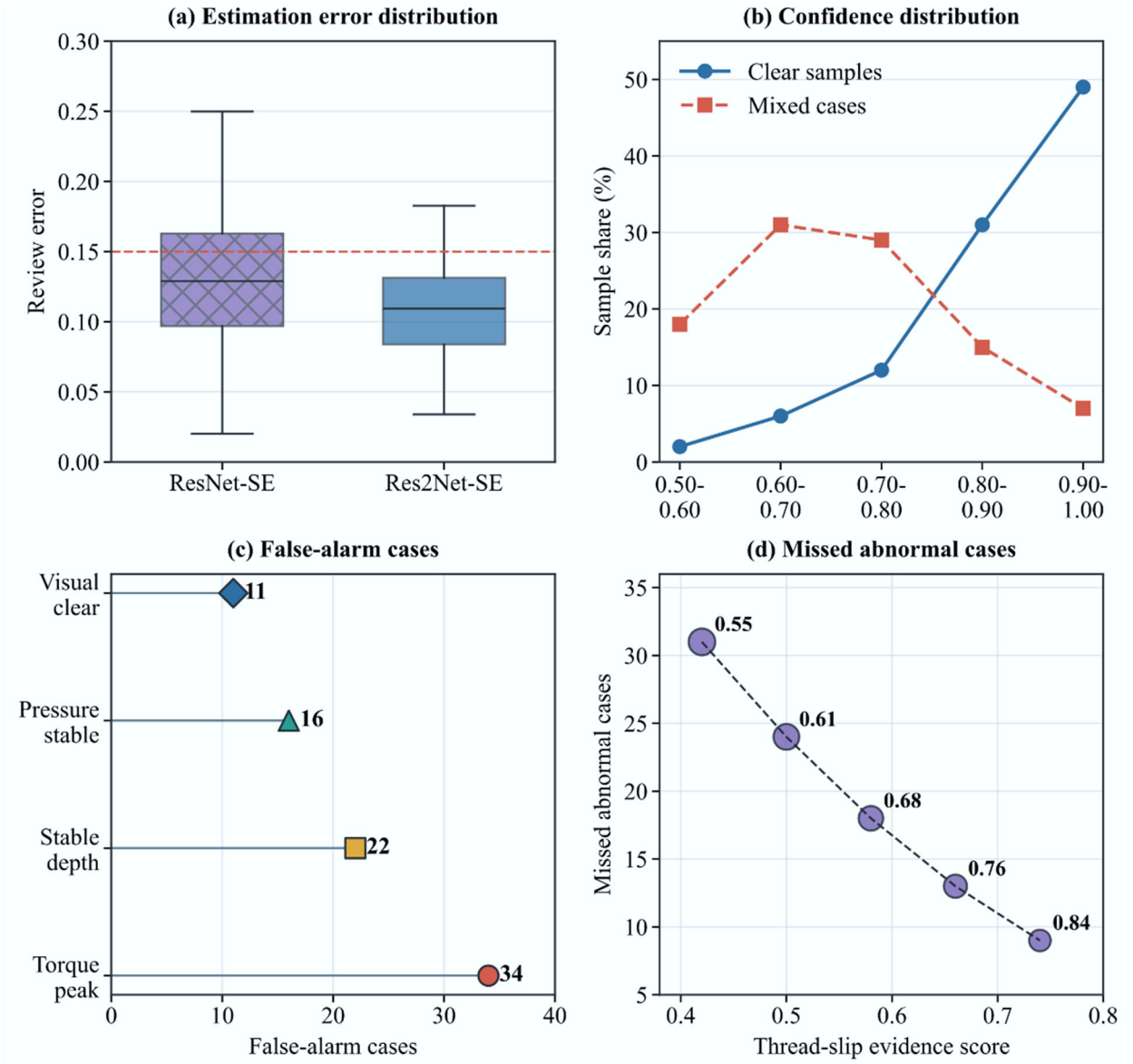


Figure 5. Error and Confidence Analysis. (a) Estimation error distribution. (b) Confidence distribution. (c) False-alarm cases. (d) Missed abnormal cases.

Ablation experiments further verify the different contributions of the modules. Without the Res2Net branch, recognition of local slip events is reduced because short torque drops and repeated rises are not represented sufficiently. Without SE weighting, tilted-entry recall decreases because pressure and depth evidence receive insufficient emphasis when torque values remain near the normal range. Removing the quality indicator increases false alarms for blurred visual samples because the model cannot separate weak sensing from actual process deviation. Attention-based inspection research [26] similarly shows that channel weighting improves feature selection when evidence is sparse or source-dependent.

4.3. Explainability and Engineering Review Output

As shown in Figure 6, channel contribution represents the amount of model evidence attributed to torque-related and pressure-related sources. Figure 6(a) presents torque-channel contribution for over-tightening and under-tightening, where terminal torque, tightening slope and rotation angle provide the strongest evidence. Figure 6(b) presents pressure-channel contribution for tilted entry and incomplete seating, where pressure fluctuation and depth residual become prominent. The distribution of this evidence [27] is consistent with explainable quality-inspection research, because attribution is meaningful only when it is linked to physical process variables.

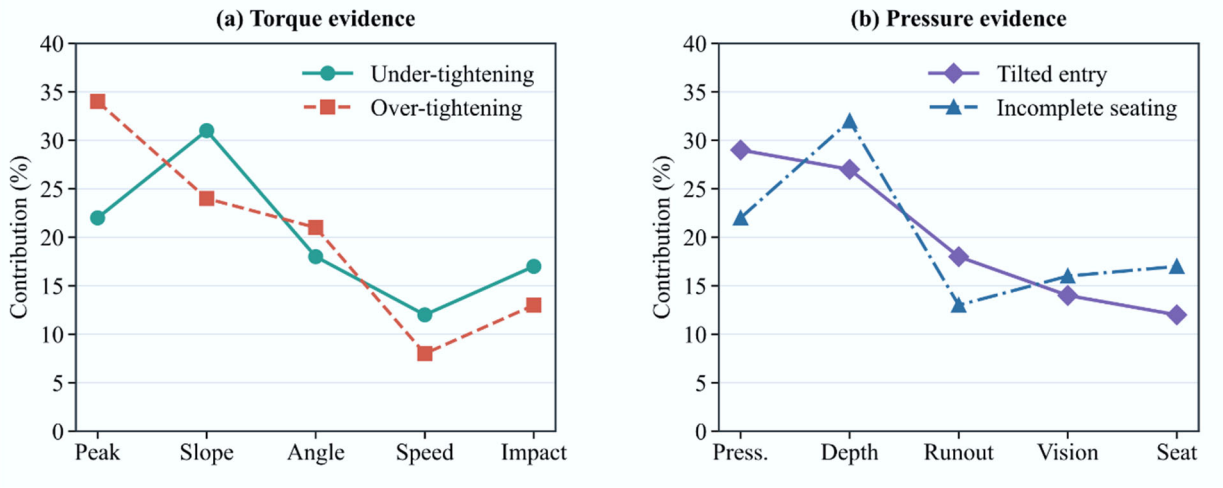


Figure 6. Channel Contribution Analysis. (a) Torque evidence. (b) Pressure evidence.

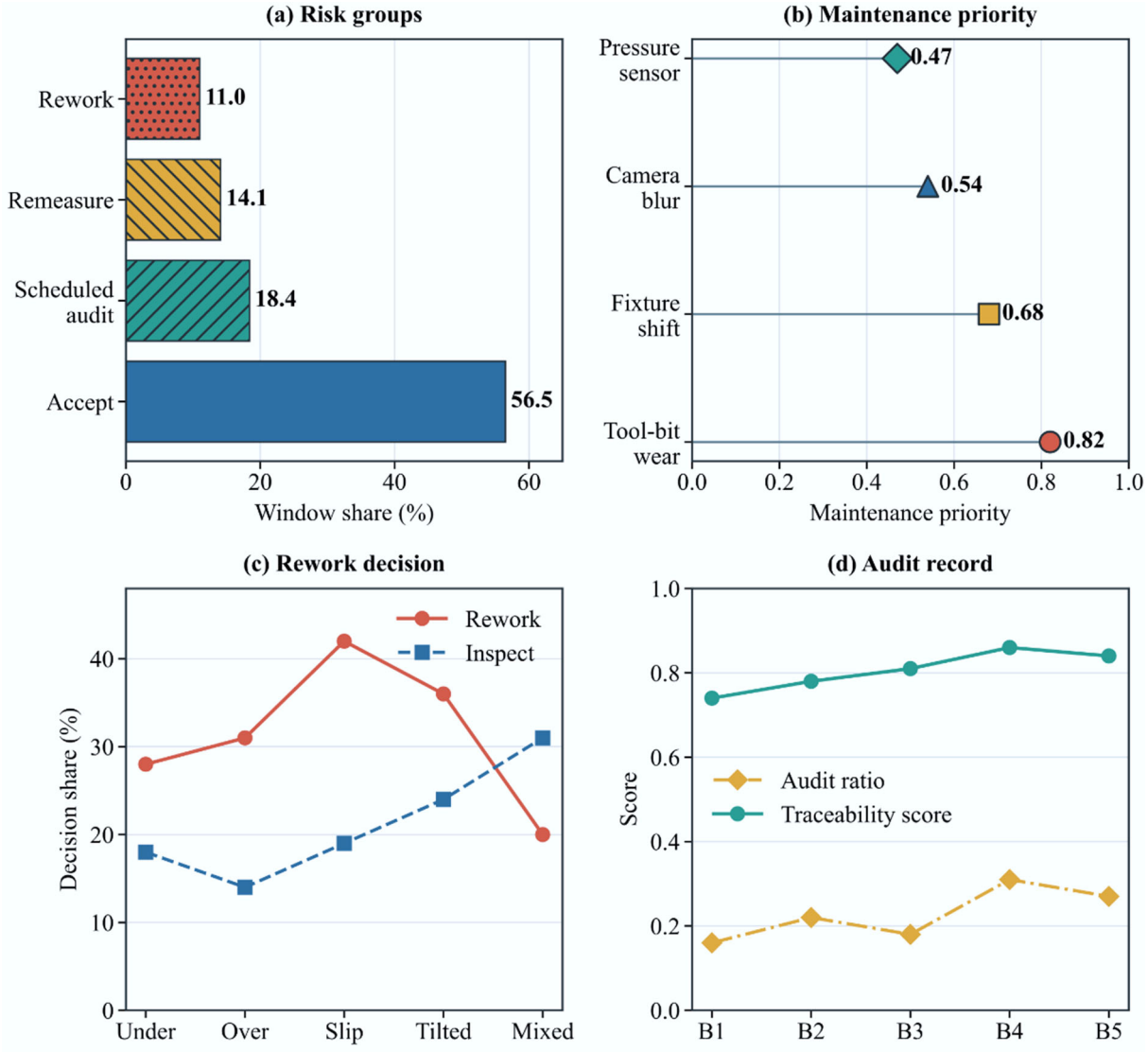


Figure 7. Review Priority Output. (a) Risk groups. (b) Maintenance priority. (c) Rework decision. (d) Audit record.

The explanation output does not substitute for engineering review; it organizes the review path. A high torque contribution with high confidence supports rework or tool-parameter modification. A high pressure contribution with moderate confidence indicates possible screw alignment, tool-nose or part-seating problems. Visual-quality contribution supports remeasurement or camera adjustment rather than immediate defect confirmation. Research on human-centred industrial AI [28] shows that evidence-based review improves acceptance of automated inspection systems because engineers can trace the source of each warning.

Figure 7 shows the engineering review-priority output and connects prediction results with rework, maintenance and audit decisions. Figure 7(a) presents risk groups generated from the review score, separating accepted samples from remeasurement samples with weak evidence. Figure 7(b) presents maintenance priority, where repeated thread-slip warnings are linked to tool-bit inspection. Figure 7(c) presents the rework decision group for severe defects, and Figure 7(d) presents audit records for batch traceability. Similar decision-support designs [29] have been proposed for traceable manufacturing quality management.

According to the robustness test, the model can tolerate torque noise, pressure disturbance and partial visual blurring. When torque noise increases, pressure and depth evidence become more important; when pressure drift occurs, torque-angle and visual-quality indicators help identify uncertainty; when image quality declines, the confidence term reduces the authority of the prediction. Robust multimodal fusion research [30] supports this design because disturbance testing should occur before line-level deployment. The quality mask is therefore retained as part of the model rather than used only as a data-cleaning rule.

The engineering value of the proposed model is to separate defect severity from evidence reliability. A high-risk and high-confidence sample can be sent to rework immediately, whereas a high-risk and low-confidence sample should be re-examined before rework. Maintenance decision research [31] indicates that confidence-aware review can reduce redundant manual inspection and improve resource allocation. Digital quality-control research [32] also treats inspection output as a source for process improvement rather than as an isolated final result. In production, the model output can be saved with station number, tool identity, screw batch, fixture version and rework result to support audit trails and later threshold calibration.

5. Conclusion

This study proposed an explainable Res2Net-SE model for robotic screw-fastening quality estimation, and the overall conclusion is that multi-scale process evidence, channel recalibration and confidence-aware review should be integrated for traceable assembly inspection. First, experiments on simulated robotic fastening data show that the proposed model achieves 93.4% accuracy and 90.7% macro-F1, with stronger recognition of thread slip, tilted entry and under-tightening than CNN, LSTM, ResNet, Transformer and ResNet-SE baselines. Second, the comparative results show that torque-angle evidence is more indicative of over-tightening and under-tightening, whereas pressure and depth evidence are more useful for tilted entry and incomplete seating. Third, the main limitation is that the evaluation uses simulated fastening windows and simplified defect categories; therefore, future validation should include long-term production data, various screw materials, tool wear conditions, station-level calibration records and actual rework results.

The proposed output separates fastening class, quality score, confidence, dominant evidence channel and review priority. This structure supports immediate rework for high-confidence severe defects, remeasurement for low-confidence risky windows, acceptance for stable normal samples and scheduled audit for moderate abnormal trends. The separation is important because low-quality sensing can otherwise be confused with actual fastening defects. Evidence attribution also supports tool-parameter adjustment, fixture alignment, camera review and maintenance prioritization.

Future work should connect the quality-estimation output with production-line logs, screw-material variation, tool wear, fixture tolerance, camera illumination and operator-confirmed rework outcomes. Additional experiments should test tool speed, fixture calibration and visual-inspection conditions across multiple stations. These extensions will improve the reliability of the estimator for traceable robotic assembly, closed-loop process control and long-term quality-management audit.

6. Declarations

6.1. Author Contributions

Recep Urfalı contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. All authors have read and agreed with the manuscript before its submission and publication.

6.2. Funding and Acknowledgments

This research received no specific financial support from any funding agency.

6.3. Institutional Review Board Statement

Not applicable.

6.4. Informed Consent Statement

Not applicable.

6.5. Data Availability

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

6.7. Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors utilized Grammarly for linguistic polishing, logical organization and draft revision. The generative AI tool was not involved in research design, data collection, data analysis, or the equation of core conclusions. All outputs generated by AI were reviewed, revised and verified manually by the authors. The authors take full responsibility for all content of this manuscript.

7. References

[1] El Kalach, F., Farahani, M., Wuest, T., & Harik, R. (2025). Real-time defect detection and classification in robotic assembly lines: A machine learning framework. Robotics and Computer-Integrated Manufacturing, 95, 103011. https://doi.org/10.1016/j.rcim.2025.103011

[2] Moreno, M., Rocha, H., Pilastri, A., Moreira, G., Matos, L. M., & Cortez, P. (2025). Screw Process Anomaly Visualization (SPAV): A Python module for local and global machine learning visualizations for screw tightening anomaly detection. Software Impacts, 26, 100786. https://doi.org/10.1016/j.simpa.2025.100786

[3] Pérez-Olguín, I. J. C., Fernández-Gaxiola, C. C., Rodríguez-Picón, L. A., & Méndez-González, L. C. (2024). Characterization and dataset compilation of torque-angle curve behavior for M2/M3 screws. Data, 9(10), 115. https://doi.org/10.3390/data9100115

[4] Liu, Z., Chen, J., Chen, C., Qi, B., Guo, J., & Xu, J. (2024). Target reconstruction and process parameter decision-making for bolt intelligent assembly based on robot and multi-camera. Expert Systems with Applications, 256, 124871. https://doi.org/10.1016/j.eswa.2024.124871

[5] Zhang, X., Wang, H., Wang, C., Liu, M., & Xu, G. (2024). Time-segment-wise feature fusion transformer for multi-modal fault diagnosis. Engineering Applications of Artificial Intelligence, 138, 109358. https://doi.org/10.1016/j.engappai.2024.109358

[6] Shi, Y., Zhu, Z., Lyu, G., & Chen, K. (2025). Application of lightweight CNN-LSTM-attention model in defect detection during thread tightening process. 2025 IEEE 4th International Conference on Computing, Communication, Perception and Quantum Technology, 1-9. https://doi.org/10.1109/CCPQT66408.2025.11383325

[7] Deng, L., Zhang, Y., Zhao, C., & Wang, G. (2024). Dual-path multi-scale attention residual network for fault diagnosis of rolling bearings under complex operating conditions. Measurement Science and Technology, 35(8), 086106. https://doi.org/10.1088/1361-6501/ad3f39

[8] Cummins, L., Sommers, A., Ramezani, S. B., Mittal, S., Jabour, J., Seale, M., & Rahimi, S. (2024). Explainable predictive maintenance: A survey of current methods, challenges and opportunities. IEEE Access, 12, 57574-57602. https://doi.org/10.1109/ACCESS.2024.3391130

[9] Zemzemoglu, M., Unel, M., & Tunc, L. T. (2024). Enhancing automated fiber placement process monitoring and quality inspection: A hybrid thermal vision-based framework. Composites Part B: Engineering, 285, 111753. https://doi.org/10.1016/j.compositesb.2024.111753

[10] Yu, J., & Hu, H. (2024). A multi-scale convolutional neural network with self-knowledge distillation for bearing fault diagnosis. Machines, 12(11), 792. https://doi.org/10.3390/machines12110792

[11] Zhao, F., Zhang, C., & Geng, B. (2024). Deep multimodal data fusion. ACM Computing Surveys, 56(9), 1-36.

https://doi.org/10.1145/3649447

[12] Shen, Q., & Zhang, Z. (2024). Fault diagnosis method for bearing based on attention mechanism and multi-scale convolutional neural network. IEEE Access, 12, 12940-12952. https://doi.org/10.1109/ACCESS.2024.3357113

[13] Su, G., Wang, H., Zhang, Y., Zhang, W., & Lin, X. (2024). Simple and deep graph attention networks. Knowledge-Based Systems, 293, 111649. https://doi.org/10.1016/j.knosys.2024.111649

[14] Chen, W., Yang, K., Yu, Z., Shi, Y., & Chen, C. L. P. (2024). A survey on imbalanced learning: Latest research, applications and future directions. Artificial Intelligence Review, 57(6), 137. https://doi.org/10.1007/s10462-024-10759-6

[15] Bitterling, C., Loske, D., & Klumpp, M. (2025). Improving visual inspection accuracy: Explainable AI for human-centric quality control. IFAC-PapersOnLine, 59(10), 1450-1455. https://doi.org/10.1016/j.ifacol.2025.09.244

[16] Gaudreault, J. G., & Branco, P. (2024). Empirical analysis of performance assessment for imbalanced classification. Machine Learning, 113(8), 5533-5575. https://doi.org/10.1007/s10994-023-06497-5

[17] Tekden, A. E., Erdem, A., Erdem, E., Asfour, T., & Ugur, E. (2024). Object and relation centric representations for push effect prediction. Robotics and Autonomous Systems, 174, 104632. https://doi.org/10.1016/j.robot.2024.104632

[18] Nguyen, H. T. T., Nguyen, L. P. T., & Cao, H. (2025). XEdgeAI: A human-centered industrial inspection framework with data-centric explainable edge AI approach. Information Fusion, 116, 102782. https://doi.org/10.1016/j.inffus.2024.102782

[19] Xiao, X., Li, C., He, H., Huang, J., & Yu, T. (2025). Rotating machinery fault diagnosis method based on multi-level fusion framework of multi-sensor information. Information Fusion, 113, 102621. https://doi.org/10.1016/j.inffus.2024.102621

[20] Liang, T., Liu, T., Wang, J., Zhang, J., & Zheng, P. (2025). Causal deep learning for explainable vision-based quality inspection under visual interference. Journal of Intelligent Manufacturing, 36(2), 1363-1384. https://doi.org/10.1007/s10845-023-02297-9

[21] Lee, Y.-S., Eun, S., Shin, D., & Lee, J.-H. (2026). CNN-based thread defect detection system for manufacturing process automation. Journal of the Korea Society of Computer and Information, 31(3), 71-78. https://doi.org/10.9708/jksci.2026.31.03.071

[22] Sun, K., & Yin, A. (2025). Multi-sensor temporal-spatial graph network fusion empirical mode decomposition convolution for machine fault diagnosis. Information Fusion, 114, 102708. https://doi.org/10.1016/j.inffus.2024.102708

[23] Wang, K., & Hao, H. (2025). Process industrial component's intelligent fault diagnosis based on multi-sensor information interaction and multi-level feature adaptive fusion framework. Measurement Science and Technology, 36(10), 106105. https://doi.org/10.1088/1361-6501/ae09c1

[24] Han, J., Chen, Z., Zhou, D., Hu, B., Xia, T., & Pan, E. (2025). Unsupervised motion-based anomaly detection with graph attention networks for industrial robots labeling. Engineering Applications of Artificial Intelligence, 146, 110298. https://doi.org/10.1016/j.engappai.2025.110298

[25] Baek, J., Seo, J., Yu, E., & Joung, J. (2025). Assessing quality risk scores using explainable artificial intelligence for advanced vision inspection. Journal of Intelligent Manufacturing. https://doi.org/10.1007/s10845-025-02751-w

[26] Hütten, N., Gomes, M. A., Hölken, F., Andricevic, K., Meyes, R., & Meisen, T. (2024). Deep learning for automated visual inspection in manufacturing and maintenance: A survey of open-access papers. Applied System Innovation, 7(1), 11. https://doi.org/10.3390/asi7010011

[27] Lymperaiou, M., & Stamou, G. (2024). A survey on knowledge-enhanced multimodal learning. Artificial Intelligence Review, 57(10), 284. https://doi.org/10.1007/s10462-024-10825-z

[28] Khan, M. A., Vasques, C. M. A., & Cavadas, A. M. S. (2026). AI-powered computer vision industrial quality inspection systems: A practice review. Encyclopedia, 6(9), 194. https://doi.org/10.3390/encyclopedia6090194

[29] Rahman, M. A., Shahrior, M. F., Iqbal, K., & Abushaiba, A. A. (2025). Enabling intelligent industrial automation: A review of machine learning applications with digital twin and edge AI integration. Automation, 6(3), 37. https://doi.org/10.3390/automation6030037

[30] Gao, Z., Jiang, X., Xu, X., Shen, F., Li, Y., & Shen, H. T. (2024). Embracing unimodal aleatoric uncertainty for robust multimodal fusion. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 26866-26875. https://doi.org/10.1109/CVPR52733.2024.02538

[31] Cummins, L., Sommers, A., Ramezani, S. B., Mittal, S., Jabour, J., Seale, M., & Rahimi, S. (2024). Explainable predictive maintenance: A survey of current methods, challenges and opportunities. IEEE Access, 12, 57574-57602. https://doi.org/10.1109/ACCESS.2024.3391130

[32] Kondo, R. E., Andrade, W. J., de Mello Henequim, C., Lazzaretti, A. E., de Souza Britto, A., de Freitas Rocha Loures, E.,

Santos, E. A. P., & Reynoso-Meza, G. (2024). An industrial edge computing architecture for local digital twin. Computers & Industrial Engineering, 193, 110257. https://doi.org/10.1016/j.cie.2024.110257