



ARTICLE

Three-Stage Learning Framework for Compound Fault Diagnosis in Delta 3D Printers via Multi-Output Fusion Ensembles

Lin Fang^{1,2}, Razi Abdul-Rahman^{1,*} and Cheng-Fu Yang^{3,4,*}

¹School of Mechanical Engineering, University Sains Malaysia, Nibong Tebal, Penang, Malaysia

²School of Urban Construction and Intelligent Manufacturing, Dongguan City University, Dongguan, China

³Department of Chemical and Materials Engineering, National University of Kaohsiung, Kaohsiung, Taiwan

⁴Department of Aeronautical Engineering, Chaoyang University of Technology, Taichung, Taiwan

*Corresponding Authors: Razi Abdul-Rahman. Email: mearahman@usm.my; Cheng-Fu Yang. Email: cfyang@nuk.edu.tw

Received: 08 February 2026; Accepted: 06 May 2026; Published: 30 June 2026

ABSTRACT: Parallel mechanisms are extensively employed in industrial logistics, food processing, and medical applications. Due to the strong nonlinearity and cross-axis coupling inherent in closed-chain kinematics, fault diagnostic performance is highly sensitive to signal perturbations and class imbalance under noisy measurement conditions. Furthermore, diagnostic models trained under single-fault scenarios often exhibit notable performance degradation when transferred to compound fault conditions as a result of distribution shift. In this study, a Delta 3D printer, as a representative parallel mechanism, is adopted as the experimental platform. An interpretable three-stage diagnostic framework is proposed, in which compound fault diagnosis is reformulated as a multi-output classification problem that simultaneously predicts the health states of the A-, B-, and C-belts. This formulation avoids explicit enumeration of compound fault classes while preserving maintenance-relevant, belt-level diagnostic information. Under a strict leakage-avoidance protocol, a fusion ensemble integrating LightGBM and XGBoost classifiers is employed to enhance robustness and generalization to previously unseen compound fault combinations. On the compound-fault subset of the Delta 3D printer dataset, the proposed method achieves a multi-output Macro-F1 score of 0.9290, with a 95% bootstrap confidence interval of 0.9198–0.9379. The corresponding belt-wise Macro-F1 scores reach 0.9508, 0.9173, and 0.9189 for the A-, B-, and C-belts, respectively. Moreover, the average inference latency on the compound-fault subset is 0.9305 ms per sample, demonstrating a favorable balance between diagnostic accuracy and computational efficiency for edge-deployment scenarios.

KEYWORDS: Delta 3D printer; fault diagnosis; compound faults; multi output classification; fusion ensemble

1 Introduction

Parallel mechanisms are extensively utilized in high-speed and precision production because of their superior stiffness-to-mass ratio and rapid dynamic response [1–5]. In addition, their structural characteristics provide advantages in positioning accuracy, motion stability, and load-carrying capability, making them suitable for advanced manufacturing applications [6–9]. The Delta 3D printer, as a sample of parallel architecture, utilizes three identical kinematic chains to collaboratively operate the printhead, providing low movement inertia and high motion efficiency [10,11]. Nonetheless, its closed-loop kinematic structure, performance degradation in any single leg may propagate through strong coupling effects and compromise end-effector motion, thereby affecting print quality and operational stability.

Fault diagnosis methods can generally be categorized into model-based and data-driven approaches. Model-based approaches are difficult to deploy to parallel robots, because accurate physical modeling and parameter identification burden is substantial for strongly coupled parallel robots and may not remain valid under varying trajectories and operating conditions [12,13]. Data-driven approaches reduce reliance on explicit physical modeling, but traditional handcrafted features often struggle to capture coupled fault signatures, while recent deep learning and vision-based approaches usually require large labeled datasets and may face limitations in interpretability, efficiency, and robustness in industrial settings [12,14–17]. Image-based techniques utilizing cameras, thermal imaging, or layer-wise visual inspection have demonstrated efficacy in identifying surface defects, deposition irregularities, and geometric discrepancies during the printing process. Concurrently, deep neural models applied to vibration, acoustic, or multisensor signals can enhance nonlinear representation learning and minimize the necessity for manual feature engineering. Nonetheless, these techniques frequently necessitate substantial labeled datasets and may encounter difficulties regarding interpretability, deployment efficiency, and resilience in real-world industrial settings. Notwithstanding these advancements, there has been a paucity of focus on structured compound fault diagnosis in Delta 3D printers, particularly within a cohesive framework that concurrently addresses interpretability, leakage-aware assessment, and scalable fault representation.

In Delta 3D printers, an additional practical challenge arises from the coexistence of single-leg and multi-leg faults. Conventional single-label multi-class formulations implicitly assume that each sample corresponds to a single fault category. This assumption becomes restrictive in the presence of compound faults, as explicit enumeration of all fault combinations rapidly expands the label space and exacerbates data imbalance. More critically, such formulations tend to generalize poorly to unseen fault combinations, limiting their effectiveness in real-world operating environments.

To address these challenges, the present study investigates a Delta 3D printer and develops an interpretable and unified diagnostic framework that integrates systematic preprocessing and feature engineering, imbalance-aware resampling, and an ensemble strategy based on weighted soft voting (WSV). The proposed workflow is structured into three stages. First, standardized preprocessing pipelines and candidate learners are screened to identify robust pipeline–model pairs under leakage-avoidance constraints. Second, the selected models are refined through hyperparameter optimization to enhance stability and generalization. Third, a fusion ensemble is constructed and evaluated under both single-fault and compound-fault conditions using a multi-output formulation, enabling simultaneous prediction of belt-level health states.

This study makes two primary contributions. First, it establishes a unified and interpretable diagnostic pipeline tailored to Delta 3D printers, in which hybrid anomaly detection, feature enhancement and selection, imbalance treatment, and fusion ensembles are integrated into a reproducible workflow that supports fair comparison and deployment-oriented evaluation. Second, it introduces a multi-output fault representation that unifies single-fault and compound-fault modeling within a scalable label space. This formulation avoids combinatorial class explosion and enables systematic evaluation on previously unseen compound fault combinations, thereby improving robustness and practical applicability.

2 Related Work

Parallel mechanisms are characterized by multi-chain coupling, strong nonlinearities, and multi-source error propagation. Their degradation processes often manifest simultaneously as kinematic inaccuracies, dynamic disturbances, and anomalies in end-effector pose signals, which complicates fault isolation and interpretation.

Model-based fault diagnosis relies on accurate kinematic and dynamic modeling together with reliable parameter identification [18]. For example, in six-degree-of-freedom parallel mechanisms, the kinematic

equations may yield multiple real solutions, among which only one corresponds to the physically meaningful configuration [19,20]. In practice, modeling errors, measurement noise, and parameter uncertainty prevent residuals from converging to zero, even under healthy conditions. To address this limitation, Mardt et al. proposed a physics-based statistical framework that evaluates residuals under uncertainty, enabling more robust fault detection [21]. While model-based diagnosis is attractive when high-fidelity models and identifiable parameters are available, nonlinear coupling and parameter uncertainty in parallel mechanisms make residual modeling and threshold calibration difficult to sustain across varying operating conditions, thereby restricting real-time applicability and limiting deployment robustness [19–21].

Driven by the increasing availability of sensor data, data-driven fault diagnosis has become the dominant paradigm in many robotic systems. For Delta-type 3D printers, several studies have explored data-driven diagnosis using end-effector pose measurements and state-monitoring signals. For instance, a pose-monitoring framework combined with a support vector machine was reported to identify typical fault states [22]. To mitigate performance degradation caused by cross-condition or cross-task discrepancies, transfer SVM frameworks were introduced to enhance generalization capability [23]. Other studies employed lightweight models, such as extreme learning machines combined with optimization strategies, to improve classification performance [24], or adopted transfer-learning approaches to alleviate distribution shifts induced by operating-condition variations [25]. To handle more complex nonlinearities and multi-condition factors, deep learning models have also been applied to construct end-to-end or representation-learning-based diagnostic solutions [26]. Collectively, these studies demonstrate the feasibility of combining multi-channel measurements with data-driven classification for Delta platforms. Nevertheless, substantial challenges remain, including inter-class overlap, class imbalance, combinatorial growth of compound fault categories, and limited generalization to unseen fault combinations and operating conditions.

Time-series signals collected from parallel mechanisms are often contaminated by impulsive noise, intermittent jitter, sensor drift, and outliers, which can bias statistical features and distort classification boundaries. Studies on time-series anomaly detection indicate that noise and anomalies are frequently entangled in the data space, motivating the use of smoothing techniques, signal transformations, or robust learning strategies to improve diagnostic reliability [27]. In the context of 3D printing process monitoring, vibration-based anomaly detection has been investigated as an external sensing approach to identify abnormal printing states or irregular machine behavior [28]. In engineering diagnosis, unsupervised methods such as Isolation Forest (IF) and Local Outlier Factor (LOF), which impose weak assumptions on data distributions, are commonly employed for pre-training outlier removal or training-set sanitization to reduce the influence of anomalous samples on model learning [29,30].

Fault-related information is typically distributed across signal statistics in a multi-channel and multi-scale manner, which makes lightweight and interpretable time-domain statistical features attractive for engineering applications. However, feature expansion may introduce redundancy and instability, necessitating robust feature-selection strategies. The ReliefF algorithm family evaluates feature relevance by analyzing nearest-neighbor discrepancies and can be adapted to multi-class problems and noisy environments [31]. Improved variants of ReliefF have also been applied to multi-label or multi-output tasks to enhance stability and capture label correlations [32]. Recursive feature elimination with cross-validation (RFECV) and its variants are widely used to remove redundant features in a classifier-driven manner [33]. A two-stage strategy that combines ReliefF for relevance screening with RFECV for redundancy removal has been shown to improve generalization robustness while preserving interpretability [34]. Variations in operating conditions often lead to imbalanced sample distributions in fault datasets, and direct training under such conditions may cause models to favor majority classes. The synthetic minority over-sampling technique (SMOTE) is a classical data-level balancing method that augments minority-class samples via k-nearest-neighbor

interpolation [35]. The combination of SMOTE and Tomek links further enhances class separability by removing overlapping samples near decision boundaries, and it is commonly used to mitigate inter-class confusion introduced by resampling [36]. In multi-class fault diagnosis, oversampling followed by boundary cleaning can therefore provide clearer class separation for subsequent classifiers.

Conventional classifiers such as support vector machines, k-nearest neighbors, random forests, and gradient boosting trees remain widely used in fault diagnosis due to their robustness with limited data, stable training behavior, and relatively strong interpretability [37–40]. Ensemble learning, particularly stacking, improves generalization by exploiting the complementary strengths of multiple base learners and has consistently demonstrated superior performance over single models in engineering fault diagnosis tasks [41,42]. For hyperparameter tuning, grid search remains a commonly adopted approach for fair model comparison [43]. However, as the hyperparameter space expands, Bayesian optimization and its variants provide more efficient exploration toward near-optimal solutions under limited evaluation budgets [44]. Beyond algorithm selection, diagnostic performance depends on the design of the entire pipeline, including signal segmentation and labeling, preprocessing strategies, feature engineering and selection, class imbalance handling, and evaluation under operating variability. Studies employing attitude multi-sensor signals and pose-related features for Delta-type 3D printers indicate that representation design and preprocessing pipeline choices may exert a greater influence on diagnostic performance than switching among comparable classifiers [22,23,25,45].

A central challenge in fault diagnosis for parallel mechanisms lies in the formulation of single-fault and compound-fault conditions. Single-label multi-class classification assumes one fault category per sample and becomes restrictive when multiple kinematic legs degrade simultaneously. Enumerating compound fault classes leads to rapid expansion of the label space, aggravates class imbalance, and generalizes poorly to previously unseen fault combinations. These limitations motivate multi-output classification approaches that predict leg-level health states for each kinematic chain, enabling compound faults to be represented as combinations of leg-wise states rather than enumerated compound classes [25,45]. In summary, although significant progress has been achieved in data-driven robotic fault diagnosis, existing evidence for parallel mechanisms remains fragmented in several respects. In particular, scalable compound-fault modeling without class enumeration, traceable decision mechanisms in multi-step diagnostic pipelines, and robustness to preprocessing choices, class imbalance, and run-to-run distribution mismatch have not yet been addressed in a unified and reproducible manner [17,25,46]. These gaps motivate the development of an integrated framework that standardizes preprocessing and feature engineering, embeds imbalance handling within model selection under a fixed evaluation protocol, and represents compound conditions through multi-output classification.

3 Methodology

The methodology adopted in this study is structured around a three-stage diagnostic framework and encompasses the experimental platform, dataset construction, and evaluation protocol. The overall design emphasizes reproducibility, leakage avoidance, and robustness assessment under realistic sensing and operating conditions.

3.1 Experimental Platform and Dataset

The experimental testbed developed for Delta 3D printer fault diagnosis consists of a Delta 3D printer, an attitude sensor mounted on the moving platform, and a laptop computer for data logging and experiment control, as illustrated in Fig. 1. Due to the transmission degradation in the belt-driven Delta mechanism affecting the motion attitude and dynamic responsiveness of the end effector, the sensor was positioned on

the upper surface of the moving platform, close to the central area of the end effector, instead of on the base frame. The sensor was securely affixed via a mounting base to reduce relative motion during operation, and its detecting surface was maintained nearly parallel to the platform surface. The mounting location and orientation were same across all trials to guarantee measurement consistency, as shown in Fig. 1c.

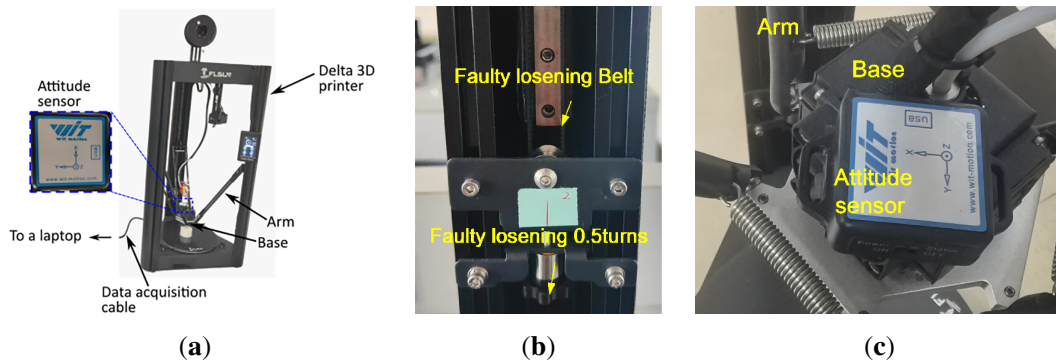


Figure 1: Delta 3D printer system, (a) overview of the experimental platform; (b) loosened belt condition with 0.5 turns; (c) attitude sensor mounted on the moving platform.

The attitude sensor (WT901SDCL BT50) is a compact inertial measurement unit that provides twelve synchronous signal channels, including tri-axial angular velocity, tri-axial vibration acceleration, tri-axial attitude angles, and tri-axial magnetic field intensity. Signal data were acquired at 200 Hz during the standardized R25 cylindrical printing task. A low-cost sensor is intentionally selected to introduce practical measurement noise, thereby enabling a more realistic evaluation of diagnostic robustness under cost-constrained sensing scenarios.

The Delta 3D printer is actuated by three synchronous belts that form three coupled kinematic chains, denoted as Belt A, Belt B, and Belt C. Fault conditions are simulated by loosening the tensioning screw of each belt from 0.25 turns up to 2.0 turns, representing progressive belt tension reduction. An example condition with 0.5 turns of loosening is shown in Fig. 1. Based on combinations of belt states, two datasets are constructed, as summarized in Table 1. A compound fault physically corresponds to the simultaneous loosening of multiple belts. To maintain consistency with the single-fault data structure and to avoid explicit enumeration of compound fault classes, each sample is labeled using a three-output vector that describes the health states of Belt A, Belt B, and Belt C. Each belt state is quantized into nine discrete levels, including the normal condition and increasing degrees of loosening severity. The nine looseness levels are defined by uniformly adjusting the belt tensioning screw from 0.25 to 2.0 turns in increments of 0.25 turn, together with the normal condition. This step size is selected to balance controllability and resolution, as quarter-turn adjustments are practically achievable and repeatable in mechanical tuning. Preliminary experiments confirm that each increment produces consistent and monotonic changes in sensor responses, including vibration intensity, angular velocity, and end-effector attitude stability. These observations indicate that the discretized levels correspond to physically meaningful degradation states. The same adjustment protocol is applied across all three belts and repeated under identical motion conditions to ensure consistency and reproducibility.

Table 1: Fault configuration of delta 3D printer.

ID	Fault Type	Fault Setting	Belt A Turns	Belt B Turns	Belt C Turns
N0	Baseline	Normal	0	0	0
S-A	Single fault	Belt A	1/4 to 2.0	0	0
S-B	Single fault	Belt B	0	1/4 to 2.0	0
S-C	Single fault	Belt C	0	0	1/4 to 2.0
C-AB1	Compound fault	Belt A plus Belt B	1/2	1/2	0
C-AB2	Compound fault	Belt A plus Belt B	3/4	3/4	0
C-AB3	Compound fault	Belt A plus Belt B	1.0	1.0	0
C-AC1	Compound fault	Belt A plus Belt C	1/2	0	1/2
C-AC2	Compound fault	Belt A plus Belt C	3/4	0	3/4
C-AC3	Compound fault	Belt A plus Belt C	1.0	0	1.0
C-BC1	Compound fault	Belt B plus Belt C	0	1/2	1/2
C-BC2	Compound fault	Belt B plus Belt C	0	3/4	3/4
C-BC3	Compound fault	Belt B plus Belt C	0	1.0	1.0

Signal data are sampled at a frequency of 200 Hz. Each sample corresponds to one acquired signal segment comprising twelve channels. During data acquisition, the printer executes a standardized R25 cylindrical printing task for 30 s, and each condition is repeated three times. The resulting dataset contains 7328 single-fault samples and 9177 compound-fault samples. In this study, a “sample” is defined as a feature vector extracted from a short segment of continuous signals rather than an entire trial. Each 30 s recording (sampled at 200 Hz) is uniformly divided into multiple non-overlapping segments with a fixed duration of T seconds, and each segment is treated as an independent sample. For each segment, multi-channel signals are transformed into a low-dimensional feature vector (12D–15D) by extracting representative statistical descriptors (e.g., mean, standard deviation, extrema, and related temporal characteristics). This segmentation-based aggregation explains the larger number of samples relative to the number of runs while maintaining a compact feature representation.

In this study, sliding-window segmentation is not applied to the original time-series signals. Instead, samples are defined as fixed observation units during data acquisition. To prevent adjacent segments originating from the same run from appearing in both training and evaluation sets, data partitioning is performed at the run level. The dataset is divided into training, validation, and test subsets following a 60%, 20%, and 20% protocol. The entire workflow strictly adheres to a no-leakage principle: all fitting, screening, and resampling procedures are conducted exclusively on the training subset, while the validation and test subsets are transformed using artifacts fitted on the training data only. This protocol ensures statistical validity and enables fair and comparable performance evaluation. All experiments are implemented in Python 3.10 using Scikit-learn, XGBoost, LightGBM, and PyTorch. Model training and inference are performed on a workstation equipped with an AMD Ryzen 7 4800H CPU operating at 2.90 GHz and 16 GB of memory.

In a belt-driven Delta 3D printer, belt slackness signifies not merely a mechanical defect but also has immediate engineering implications for the printing process. Decreased belt tension can diminish transmission rigidity and disrupt motion coordination among the three parallel kinematic chains, consequently impairing the trajectory-tracking precision of the end effector. In practical printing, such degradation may manifest as dimensional departure, contour distortion, layer misalignment, and decrease of surface quality

in printed components. Consequently, detecting belt slackness is crucial for preserving print quality and geometric accuracy.

3.2 Model Framework and Staged Objectives

A typical machine-learning-based fault diagnosis workflow includes signal acquisition, feature extraction, model training, and evaluation. However, many data-driven studies focus primarily on classifier design and parameter optimization while keeping upstream preprocessing fixed, which limits interpretability. In addition, although compound faults frequently occur in practice, generalization to unseen compound fault conditions is rarely examined. To address these issues, this study adopts a unified diagnostic pipeline organized into three stages. In the first stage (S1), standardized preprocessing pipelines are systematically screened together with candidate learning algorithms to identify a suitable preprocessing scheme and the two most competitive model families under a leakage-avoidance protocol. In the second stage (S2), the selected models are refined through hyperparameter optimization, and a WSV ensemble is constructed to improve robustness and generalization stability. In the third stage (S3), the framework is evaluated under compound fault conditions using a multi-output formulation, with particular emphasis on generalization to unseen fault combinations. The overall procedure is illustrated in Fig. 2.

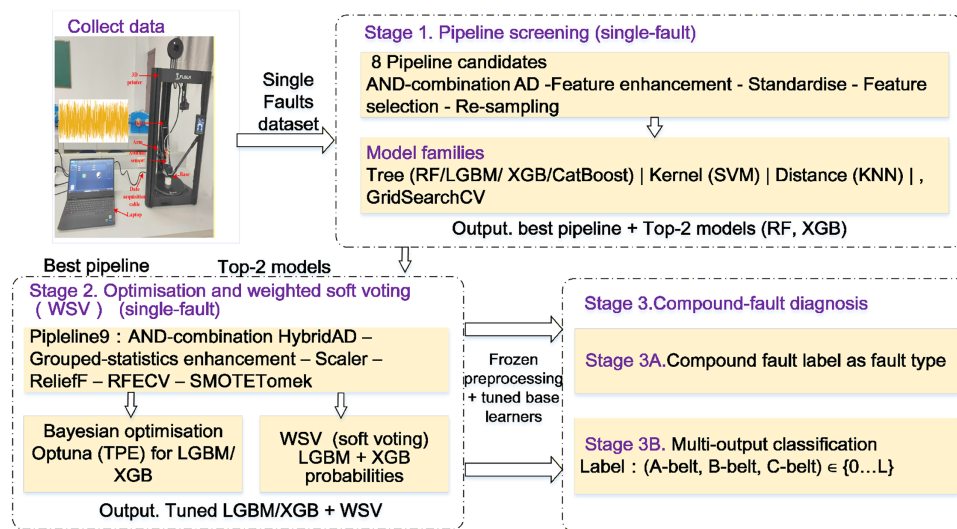


Figure 2: The overall workflow.

3.2.1 Stage S1: Pipeline Screening and Baseline Establishment

The first stage of the proposed framework focuses on establishing a reproducible and robust diagnostic baseline, while identifying an appropriate preprocessing scheme and two candidate model families under a unified evaluation protocol. In practical fault diagnosis, performance is often determined not only by the classifier itself but also by its interaction with upstream preprocessing. Consequently, preprocessing schemes and learning models are evaluated jointly rather than in isolation. Model candidates in this stage prioritize interpretable learning families that are commonly adopted in engineering diagnosis, including tree-based, kernel-based, and distance-based approaches. Accordingly, four representative classifiers are considered: Random Forest (RF), XGBoost (XGB), Support Vector Machine (SVM), and k-Nearest Neighbors (KNN). To provide a more comprehensive evaluation, additional strong tree-based baselines, including LightGBM and CatBoost, are incorporated. These methods are widely recognized for their effectiveness on tabular

data and serve as competitive benchmarks for comparison. All baseline models are implemented under the same preprocessing pipeline and evaluation protocol to ensure fairness. This extended comparison enables a more rigorous assessment of the proposed method against both classical and state-of-the-art tree-based approaches. This selection balances diagnostic performance, interpretability, and suitability for limited-data scenarios.

To mitigate distortion of feature relevance estimation and decision boundaries caused by abnormal signal segments, this stage incorporates a AND-combination anomaly detection strategy (samples removed only if identified by both Isolation Forest and Local Outlier Factor). The proposed AND-combination strategy combines IF and LOF to suppress both globally extreme segments and locally inconsistent patterns in the training data. Let the two methods produce binary decisions on the training split, denoted as p^{IF} and p^{LOF} , where +1 indicates an inlier and -1 indicates an outlier. A keep indicator is then defined as follows to determine whether a sample is retained for subsequent processing.

$$I = 1[p^{\text{IF}} = +1] \wedge [p^{\text{LOF}} = +1] \quad (1)$$

Only training samples satisfying Eq. (1) are kept for subsequent processing. To stabilize representation screening, a lightweight time-domain statistical feature enhancement is introduced on top of the raw 12-channel inputs. For $x^{(1)} \in \mathbb{R}^{12}$, the global mean $\mu(x)$ standard deviation $\sigma(x)$, and $r(x)$ are computed and concatenated, as the below Equations show, yielding $\phi(x) = [x, \mu, \sigma, r] \in \mathbb{R}^{15}$.

$$\mu(x) = \frac{1}{12} \sum_{j=1}^{12} x_j, \quad (2)$$

$$\sigma(x) = \sqrt{\frac{1}{11} \sum_{j=1}^{12} (x_j - \mu(x))^2}, \quad (3)$$

$$r(x) = \max_j x_j - \min_j x_j \quad (4)$$

The mapping is deterministic, adds only three scalars (12 \rightarrow 15), and introduces no trainable parameters. The lightweight enhancement is applied to all four tri-axial sensor groups, and the resulting statistics are appended to the original feature set. This design captures cross-axis magnitude structures and reduces sensitivity to variations in sensor mounting orientation, while remaining computationally lightweight. Following hybrid anomaly detection and the lightweight grouped feature enhancement, Stage 1 employs the ReliefF algorithm to rank features according to their discriminative capability, and subsequently applies SMOTE to alleviate class imbalance in the training data. All procedures described in this subsection are fitted exclusively on the training split and are then applied to the held-out splits to prevent information leakage. Given the training set $D_{\text{train}} = \{(x_t, y_t)\}_{t=1}^m$, where $x_t \in \mathbb{R}^d$ denotes the d -dimensional feature vector and $y_t \in \{1, \dots, C\}$ is the class label, ReliefF is employed to estimate a relevance weight W_j for each feature $j \in \{1, \dots, d\}$ by comparing each sample to its local neighbors. Following the standard multi-class ReliefF formulation, the weight update is defined over the k nearest hits and misses as

$$w_j^{\text{new}} = w_j^{\text{old}} - \frac{1}{mk} \sum_{t=1}^m \sum_{r=1}^k \text{diff}_j(x_t, H_{t,r}) + \frac{1}{mk} \sum_{i=1}^m \sum_{c \neq y_t} \frac{P(c)}{1 - P(y_t)} \sum_{r=1}^k \text{diff}_j(x_t, M_{t,r}(C)), \quad (5)$$

where k is the number of nearest neighbours, $H_{t,r}$ denotes the r -th nearest neighbor of x_t from the same class (nearest hit), $M_{t,r}(c)$ denotes the r -th nearest neighbor of x_t from class $c \neq y_t$ (nearest miss), and $P(c)$ is the empirical class prior estimated from D_{train} . The per-feature difference for continuous features is computed as below:

$$diff_j(x, z) = \frac{|x_j - z_j|}{\max_{t \in D_{\text{train}}}(x_{t,j}) - \min_{t \in D_{\text{train}}}(x_{t,j}) + \varepsilon}, \quad (6)$$

where $\max_{t \in D_{\text{train}}}(x_{t,j})$ and $\min_{t \in D_{\text{train}}}(x_{t,j})$ are the maximum and minimum values of the j -th feature computed from the training split only, and ε is a small constant to avoid division by zero (e.g., $\varepsilon = 10^{-12}$). After obtaining $\{W_j\}_{j=1}^d$, features are ranked in descending order of W_j , and the top $p = \min(10, d)$ features are selected for downstream modelling. In stage 1, the neighborhood size is set to $k = 10$ [31]. SMOTE interpolation. To alleviate class imbalance after feature selection, stage 1 applies Synthetic Minority Over-sampling Technique (SMOTE) on the training split. For a minority-class sample x_i , a synthetic sample is generated by linear interpolation with one of its minority-class nearest neighbors x_z ,

$$x_{\text{new}} = x_i + \delta(x_z - x_i), \delta \sim U(0,1), \quad (7)$$

where δ is a random scalar drawn from a uniform distribution. In stage 1, SMOTE uses $k = 5$ nearest neighbors for neighbor selection [35,39]. Standardisation is then performed by fitting a scaler on the resampled training set and applying the same transformation to the corresponding validation or test set [47]. In summary, stage 1 constructs eight preprocessing schemes by combining the above modules in different configurations. For each preprocessing–classifier combination, hyperparameters are tuned using grid search with five-fold cross-validation, with the weighted F1-score adopted as the evaluation metric. The configuration achieving the highest cross-validated score is retained. Based on the resulting stable pipeline, the two best-performing classifier families are identified. These outcomes—namely, the selected preprocessing pipeline and the top two classifiers under this pipeline—serve as the inputs and baselines for refinement and ensemble construction in stage 2.

3.2.2 Stage S2: Optimization and WSV Ensemble Fusion

Stage 2 further enhances physical consistency by introducing a group-wise statistical feature enrichment. Specifically, the twelve sensor channels are organized into four tri-axial groups corresponding to acceleration, angular velocity, attitude angles, and magnetic field measurements. For each group, seven intra-group statistical descriptors are computed, including the mean, standard deviation, minimum, maximum, median, range, and skewness. These descriptors are concatenated with the original twelve-dimensional feature vector to form an enhanced representation $x_i^{(2)} \in \mathbb{R}^{40}$. Feature selection is subsequently performed using the ReliefF algorithm, which retains up to 20 features for model construction, i.e., $\min(20, D)$, where D denotes the dimensionality of the enhanced feature space.

For the i -th sample, the original sensor vector is denoted as $x_i = [x_{i,1}, x_{i,2}, \dots, x_{i,12}] \in \mathbb{R}^{12}$, where the 12 channels are partitioned into four tri-axial groups corresponding to acceleration, angular velocity, attitude angles, and magnetic field measurements. Let $g_{i,q} = [g_{i,q,1}, g_{i,q,2}, g_{i,q,3}]$ denote the q -th tri-axial group of sample i , where $q \in \{1, 2, 3, 4\}$, k is axis within each tri-axial group.

The group-wise mean is defined as

$$\mu_{i,q} = \frac{1}{3} \sum_{k=1}^3 g_{i,q,k} \quad (8)$$

The group-wise standard deviation is defined as

$$\sigma_{i,q} = \sqrt{\frac{1}{3} \sum_{k=1}^3 (g_{i,q,k} - \mu_{i,q})^2} \quad (9)$$

The group-wise minimum and maximum are defined as

$$g_{i,q}^{\min} = \min_k g_{i,q,k}, g_{i,q}^{\max} = \max_j v_{i,j} \quad (10)$$

The group-wise median, range, and skewness are defined as

$$m_{i,q} = \text{median}(g_{i,q}), r_{i,q} = g_{i,q}^{\max} - g_{i,q}^{\min}, \text{ and } skew_{i,q} = \frac{1}{3} \sum_{k=1}^3 \left(\frac{g_{i,q,k} - \mu_{i,q}}{\sigma_{i,q} + \varepsilon} \right)^3 \quad (11)$$

Final feature is

$$x_i^{(2)} = [x_i, \{\mu_{i,q}, \sigma_{i,q}, g_{i,q}^{\min}, g_{i,q}^{\max}, m_{i,q}, r_{i,q}, skew_{i,q}\}] \in R40 \quad (12)$$

To further eliminate redundancy and improve model stability, recursive feature elimination with cross-validation (RFECV) is then applied using a Random Forest estimator, with five-fold cross-validation and the weighted F1-score adopted as the evaluation criterion. To mitigate class imbalance, the SMOTE-Tomek resampling algorithm is performed on the training set only while ensuring that all synthetic instances are within the kinematic constraints of the Delta printer [48]. Hyperparameter optimization is carried out over 30 trials using the Tree-structured Parzen Estimator (TPE) sampler implemented in Optuna, where the objective function is defined as the mean weighted F1-score obtained from five-fold cross-validation.

In stage 2, the Top-2 model families, LighBGM(LBGM) and XGBoost (XGB), are further optimized and combined via a WSV ensemble to improve stability and predictive performance. RF is adopted as a robust bagging-based learner with feature randomness, where class probabilities are obtained by averaging the outputs of T trees,

$$\hat{p}(y = c|x) = \frac{1}{T} \sum_{t=1}^T p_t(y = c|x), \quad (13)$$

$$\hat{y} = \arg \max_c \hat{p}(y = c|x), \quad (14)$$

XGB is used as a regularized gradient-boosted tree model, optimizing a penalized objective

$$L = \sum_{i=1}^N l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k), \quad (15)$$

where $l(\cdot)$ denotes the classification loss, such as softmax log-loss and $\Omega(\cdot)$ controls model complexity.

In stage 2, ensemble learning is introduced to further enhance decision robustness by combining the two optimized base learners, LighBGM and XGBoost. A weighted soft voting (WSV) strategy is adopted to fuse class-posterior probabilities without training an additional meta-learner, as formally defined in Eqs. (16) and (17). This design choice maintains model simplicity and computational efficiency while exploiting the complementary predictive characteristics of the two ensemble members. For an input sample x_i , each base learner produces a class-probability vector over the fault-type label space, denoted as $p_{LBGM}(x_i) \in R^K$ and $p_{XGB}(x_i) \in R^K$, where K represents the number of fault types. The final fused probability distribution is then obtained through a weighted aggregation of the individual posterior probabilities.

$$\hat{p}(x_i) = w p_{LBGM}(x_i) + (1 - w) p_{XGB}(x_i), \quad (16)$$

where K is the number of classes and m belongs to the set one and two. The final fault-type prediction is obtained by

$$\hat{y}_i = \arg \max_{k \in \{1, \dots, K\}} \hat{p}_k(x_i) \tag{17}$$

The weighting coefficient w is selected exclusively on the validation set and then fixed during test evaluation to prevent information leakage. Inference latency is reported as the median per-sample runtime, measured on the test features using repeated forward passes after a brief warm-up phase to ensure stable timing estimates.

3.2.3 Stage S3: Compound Fault Diagnosis

Compound faults frequently arise in practical systems, where superposed error sources and intensified signal coupling substantially increase the complexity of data distributions and decision boundaries. Treating each compound fault configuration as an independent target class leads to rapid expansion of the label space and provides limited coverage for previously unseen combinations, thereby restricting scalability and generalization. To address these limitations, stage 3 adopts a multi-output formulation that recasts compound-fault diagnosis as a three-output classification problem. Three parallel outputs are used to independently predict the health states of Belts A, B, and C, while compound fault conditions are represented implicitly through the mapping between belt-wise states and their corresponding combinations. Fig. 3 illustrates the overall workflow of the proposed multi-output formulation.

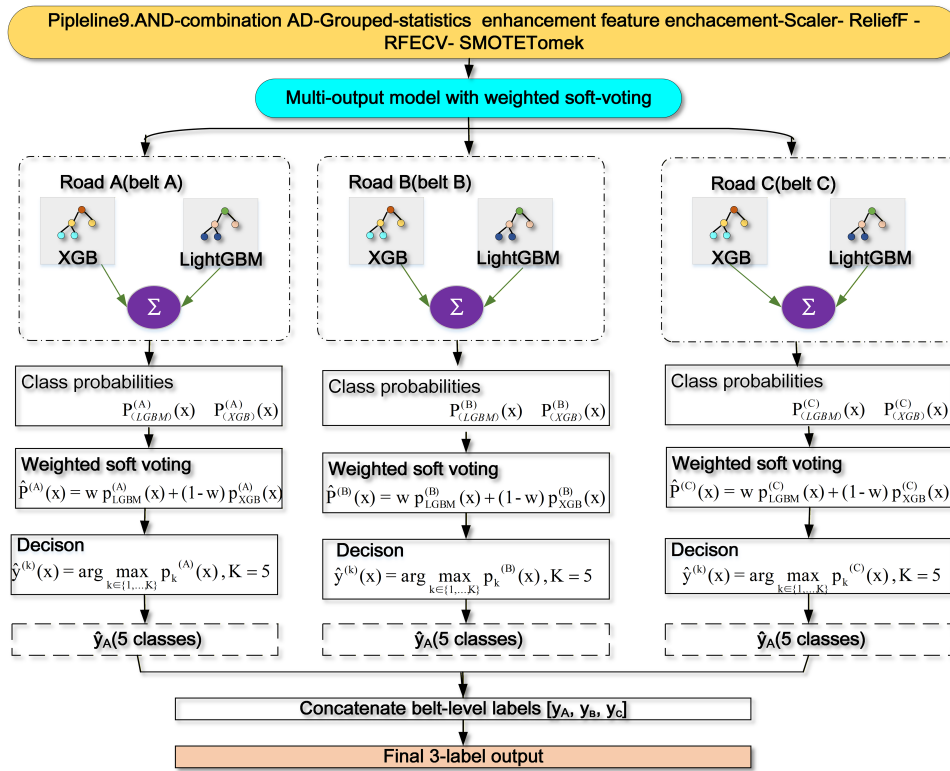


Figure 3: Workflow of the multi-output WSV fusion model.

To rigorously evaluate generalization to unseen compound-fault combinations, a combination-level splitting strategy is adopted instead of a random sample-level split. Specifically, samples are grouped according to their compound-fault combination IDs, and a group-wise partitioning scheme is applied such that all samples from selected combinations are entirely excluded from the training set and reserved for testing. This ensures that the test set contains only compound-fault combinations that are not observed during training. Under this protocol, model performance reflects the ability to generalize to previously unseen fault combinations, providing a more realistic and stringent evaluation of practical deployment scenarios.

Compound-fault diagnosis was formulated as a multi-output classification problem that predicted belt-level looseness states for belts $b \in \{A, B, C\}$. Each sample x_i was associated with label vector $y_i = [y_i^{(A)}, y_i^{(B)}, y_i^{(C)}]^T$, where $y_i^{(b)} \in \{1, \dots, K\}$ and $K = 5$ denotes the number of discrete looseness levels per belt. The implementation followed the standard scikit-learn multi-output setting, where each belt output head was fitted independently for a given base learner. At inference time, the WSV fusion defined in Eqs. (16) and (17) was applied independently to each belt output head to combine Random Forest and XGBoost class-posterior probabilities, yielding belt-wise predictions $\hat{y}_i^{(A)}, \hat{y}_i^{(B)}, \hat{y}_i^{(C)}$. The final compound diagnosis was formed by concatenating the three belt-level predictions, which preserved belt-level interpretability and avoided enumerating compound combinations as a single expanded label space. To characterize post-adaptation performance and robustness under compound-fault conditions, stage 3 adopts the same optimization strategy as stage 2 and conducts adaptation training and evaluation using an identical cross-validation scoring function and the same computational budget. On the test set, overall performance is reported alongside multi-output metrics computed separately for the single-fault and compound-fault subsets. Let the test dataset D_{test} be partitioned into the single-fault subset D_{single} and the compound-fault subset D_{comp} , as defined in Eq. (18). All evaluation metrics are computed on *each* subset using the same metric definitions.

$$\text{Metric}(D) = \text{Metric}(\{x, y\} \in D), D \in \{D_{single}, D_{comp}\}, \quad (18)$$

3.3 Hyperparameter Search Space and Optimization Settings

To ensure consistency, the search spaces and optimization parameters employed for model calibration. Stage S1 employed grid search across specified parameter ranges to facilitate transparent baseline screening, whereas Stages S2 and S3 utilized Bayesian optimization for the most promising candidate models identified in Stage S1. The identical parameter limits were employed in Stages S2 and S3 to maintain consistency. Hyperparameter selection was conducted only on the training data and was informed by the weighted F1-score according to the relevant validation process. The comprehensive search spaces are enumerated in Table 2.

Table 2: Hyperparameter search spaces and optimization settings for the compared models.

Sstage	Model	Hyperparameter and Search Space
Stagel	Random Forest	n_estimators: [50, 100, 150, 200]; max_depth: [3, 5, 8, 10, 15]; min_samples_split: [2, 5, 10, 15, 20]; min_samples_leaf: [1, 2, 5, 8, 10]
	SVM	C: [0.1, 1, 10]; kernel: [linear, rbf]
	KNN	n_neighbors: [3, 5, 7, 9]
	XGBoost	reg_alpha: [0.1, 0.5, 1]; reg_lambda: [0.1, 0.5, 1]; learning_rate: [0.01, 0.05, 0.1]; n_estimators: [50, 100, 150]

(Continued)

Table 2 (continued)

Sstage	Model	Hyperparameter and Search Space
	LightGBM	reg_alpha: [0.1, 0.5, 1]; reg_lambda: [0.1, 0.5, 1]; learning_rate: [0.01, 0.05, 0.1]; n_estimators: [50, 100, 150]
	CatBoost	learning_rate: [0.01, 0.05, 0.1]; iterations: [50, 100, 150]; l2_leaf_reg: [1, 3, 5]; depth: [4, 6, 8]
Stage2-3	LightGBM	learning_rate: [0.02, 0.15]; n_estimators: [150, 500]; num_leaves: [20, 120]; max_depth: [4, 12]; min_child_samples: [5, 30]; subsample: [0.6, 1.0]; colsample_bytree: [0.6, 1.0]; reg_alpha: [0.01, 0.5]; reg_lambda: [0.01, 0.5]; min_split_gain: [0.0, 0.5]; min_data_in_leaf: [5, 30]
Stage2-3	XGBoost	n_estimators: [50, 1000]; max_depth: [3, 12]; learning_rate: [0.005, 0.3]; subsample: [0.4, 1.0]; colsample_bytree: [0.4, 1.0]; gamma: [0.0, 10.0]; reg_alpha: [1e-8, 10.0]; reg_lambda: [1e-8, 10.0]; min_child_weight: [1, 15]; scale_pos_weight: [0.5, 5.0]

Note: Stage S1 used grid search over predefined candidate values, whereas Stages S2 and S3 used Bayesian optimization within the same parameter bounds. Hyperparameter tuning was performed on the training data only.

The search spaces were selected to equilibrate practical significance, computational expense, and equity. Stage S1 employed compact and interpretable ranges for clear baseline screening, while Stages S2 and S3 utilized wider bounds in Bayesian optimization for the most robust Stage S1 candidates. Aligned ranges were utilized for comparable boosting models to facilitate equitable comparison. Hyperparameter tuning was performed exclusively on the training data, utilizing the weighted F1-score as the principal selection criterion.

3.4 Evaluation Metrics

This study involves both a single-output multi-class classification task and a multi-output classification task under compound-fault conditions. For the single-output multi-class setting, performance is evaluated using Accuracy, Precision, Recall, F1-score, Macro-F1, and the Area Under the Receiver Operating Characteristic Curve (AUC), as shown in Eqs. (19) and (20). The Macro-F1 metric is computed by first evaluating the F1-score for each class independently and then taking their unweighted average, ensuring that all classes contribute equally to the overall assessment and reducing the influence of class imbalance.

$$Accuracy = \frac{1}{N} \sum_{i=1}^N \delta(\hat{y}_i, y_i), precision_c = TP_c / (TP_c + FP_c) \text{ and } recall_c = \frac{TP_c}{TP_c + FN_c} \quad (19)$$

$$F1_c = 2 \times (precision_c \times recall_c) / (precision_c + recall_c) \text{ and } Macro - F1 = (1/c) \times \sum_{c=1}^c F1_c \quad (20)$$

here, δ denotes the kronecker delta, N is the number of samples, TP_c , FP_c , and FN_c denote the numbers of true positives, false positives, and false negatives for class c , respectively, and C denotes the total number of classes. For AUC evaluation in the multi-class setting, a one-vs-rest strategy is adopted to compute class-wise AUC values, which are subsequently macro-averaged to obtain an overall AUC. The true positive rate and false positive rate at a given threshold t are defined as follows.

$$TPR_c(t) = \frac{TP_c(t)}{TP_c(t) + FN_c(t)} \text{ and } FPR_c(t) = \frac{FP_c(t)}{FP_c(t) + TN_c(t)}, \quad (21)$$

The class-wise AUC is given by the area under the corresponding operating characteristic (ROC) curve as below.

$$AUC_c = \int_0^1 TPR_c(FPR_c^{-1}(u)) du, \quad (22)$$

For the multi output task, Hamming loss (HL) is used to quantify the label wise prediction error across the three output channels. HL is defined as in Eq. (23) for the average fraction of mismatched labels over all samples and all outputs.

$$HL = \frac{1}{NQ} \sum_{i=1}^N \sum_{q=1}^Q I(\hat{y}_i^{(q)} \neq y_i^{(q)}), \quad (23)$$

where $I(\cdot)$ denotes the indicator function that equals 1 if the condition holds and 0 otherwise, N is the number of test samples, Q = 3 corresponds to the A belt B belt and C belt outputs, and $y_i^{(q)} \in \{0, 1, 2, 3, 4\}$ denotes the five health states from normal to four looseness levels. By definition, $HL \in [0, 1]$, and a smaller value indicates fewer label mismatches across outputs.

Performance uncertainty on the compound-fault subset is quantified using a nonparametric bootstrap procedure with 1000 resamples. The multi-output Macro-F1 score is computed as the mean of the belt-wise Macro-F1 values, and the corresponding 95% confidence interval is obtained from the 2.5th and 97.5th percentiles of the bootstrap distribution. This section has presented a unified fault diagnosis methodology for a Delta 3D printer, encompassing dataset construction, label definition, preprocessing, feature engineering, model training, and performance evaluation. Building on this methodology, the subsequent section reports stage-wise experimental results and analyses to quantify the contributions of individual modules to diagnostic accuracy and generalization stability.

4 Experiments and Results

This section presents the experimental design and result analysis of the proposed three-stage framework. To ensure a fair evaluation of the contribution of each module, a fixed train-test split is adopted under a consistent random seed and a unified evaluation protocol. All preprocessing operations within the diagnostic pipeline are performed exclusively on the training set, while the test set is used only for the corresponding transformations and performance evaluation. This design effectively prevents information leakage and ensures reproducible and comparable experimental results.

4.1 Experimental Design and Result Analysis of the Stage 1

In stage 1, a total of eight preprocessing schemes are constructed and systematically combined with four representative classifier families, as summarized in Table 3. The feature enhancement module expands the original feature dimensionality from 12 to 15, as detailed in Table 4. Model hyperparameters are tuned using GridSearchCV with five-fold cross-validation. For the ReliefF algorithm, the number of nearest neighbors is set to $k = 10$, while for SMOTE resampling, the number of nearest neighbors is fixed at $k = 5$. These modules are incorporated to address noise suppression, representation stabilization, and class imbalance mitigation in a unified manner. The final preprocessing pipeline and the top two performing classifier families are determined exclusively through the standardized screening protocol applied across Schemes 1–8, ensuring a fair and reproducible comparison under consistent experimental conditions.

Table 3: Eight preprocessing schemes for stage1.

Scheme	Pipeline	Model Column
1	IF-ReliefF-SMOTE-Standardization	Random
2	IF-SMOTE-ReliefF-Standardization	Forest
3	IF-Standardization	XGBoost
4	SMOTE-ReliefF-Standardization	KNN
5	ReliefF-SMOTE-Standardization	SVM
6	Only SMOTE-Standardization	LightGBM
7	Hybrid Anomaly-ReliefF-SMOTE	CatBoost
8	AND-combination Anomaly-Light feature enhancement-ReliefF-SMOTE-Standardization	

Table 4: Light feature enhancement schemes.

Scheme	Input Representation	Input Dimension	Components
1-7	Raw only	12	12 raw channels-Accel 3 + angular velocity 3 + Angle 3 + Mag 3
8	Raw + light_enhance	15	Raw (12) + mean + std + range

To assess the impact of filtering on class distribution, we report the per-class removal rate for each filtering strategy, including different looseness levels. This analysis allows us to examine whether minority classes are disproportionately affected. In addition, an ablation study is conducted to compare multiple configurations: (i) no filtering, (ii) Isolation Forest (IF) only, (iii) Local Outlier Factor (LOF) only, (iv) AND-combination (samples removed only if identified by both IF and LOF), and (v) OR-combination (samples removed if identified by either method). All settings are evaluated under the same protocol to ensure a fair comparison. The results demonstrate that the adopted filtering strategy achieves a balance between noise reduction and class distribution preservation.

[Table 5](#) delineates the impact of various filtering procedures on sample elimination and subsequent classification efficacy. The findings indicate that the advantage of anomaly filtering is limited and highly contingent on the approach employed. Of the strategies evaluated, the parallel AND rule attained the highest test F1-score, the maximum class-wise elimination rate attained was 42.5%. The results suggest that anomaly filtering can significantly modify the effective class distribution. This trend indicates that unsupervised anomaly detectors may eliminate not just corrupted samples but also typical observations characterized by sparse, irregular, or boundary-like distributions. Consequently, anomaly filtering should be considered a pragmatic albeit flawed denoising procedure rather than an unbiased preprocessing operation.

A comparative analysis of preprocessing Schemes 1–8 using XGBoost as a representative classifier is presented in [Table 6](#). The results indicate that class imbalance handling alone is insufficient to achieve satisfactory diagnostic performance. Specifically, Scheme 6, which applies only SMOTE together with standardization, yields the lowest accuracy of 62.41%. This outcome suggests that resampling without addressing noise contamination and representation instability may amplify ambiguous samples near class boundaries rather than improving class separability. Incorporating feature selection leads to a noticeable improvement, as evidenced by the increase in accuracy to 65.14% in Scheme 5 (ReliefF + SMOTE). This gain reflects the benefit of removing weakly informative or redundant features before classification. Further enhancement is observed in Scheme 1, where the addition of anomaly filtering using Isolation Forest (IF)

increases the accuracy to 65.48%. This result highlights the importance of suppressing abnormal segments that can distort feature relevance estimation and decision boundaries.

Table 5: Comparison of anomaly filtering strategies in terms of sample removal and test performance.

Strategy	Train_Before	Train_After	Removed_Count	Removed_Rate (%)	Accuracy	Precision	Recall	F1
AND-combination	5129	4958	171	3.334	0.6603	0.6692	0.6603	0.6524
IF_only	5129	4975	154	3.0025	0.6528	0.6576	0.6528	0.6451
OR-combination	5129	5101	28	0.5459	0.6521	0.6579	0.6521	0.6446
LOF_only	5129	5084	45	0.8774	0.648	0.6548	0.648	0.6393
No_filtering	5129	5129	0	0	0.6487	0.6574	0.6487	0.6389

Table 6: XGBoost result of scheme 1–8.

Scheme	Accuracy	Precision	Recall	F1	AUC	Inference Delay (m/s)
1	0.6548	0.6561	0.6548	0.6524	0.972	0.02
2	0.6562	0.6556	0.6562	0.6536	0.9723	0.02
3	0.6849	0.6876	0.6849	0.6815	0.9759	0.02
4	0.6562	0.6576	0.6562	0.6537	0.9728	0.03
5	0.6514	0.6539	0.6514	0.6488	0.9727	0.02
6	0.6241	0.6252	0.6241	0.6212	0.9686	0.02
7	0.6719	0.6707	0.6719	0.6687	0.9721	0.02
8	0.7101	0.7104	0.7101	0.7076	0.9788	0.02

Replacing IF with a hybrid anomaly detection strategy results in a more pronounced improvement, raising the accuracy to 67.19% in Scheme 7. The hybrid approach more effectively captures both globally extreme and locally inconsistent patterns, thereby providing a cleaner and more representative training set. Finally, the introduction of lightweight feature enhancement on top of Scheme 7 achieves the highest accuracy of 71.01% in Scheme 8. This result demonstrates that resampling alone cannot compensate for noise and representation instability. Instead, anomaly handling and feature enhancement play a critical role in stabilizing feature distributions and providing more reliable discriminative cues prior to feature selection and model training. These findings underscore the necessity of a coordinated preprocessing strategy in fault diagnosis for parallel mechanisms, where noise suppression, representation enrichment, and imbalance mitigation must be jointly considered to achieve robust and generalizable performance.

Fig. 4 illustrates the feature importance derived from the Random Forest model. The tri-axial magnetic-field and angular-velocity channels receive notably high importance scores, indicating that the classifier relies heavily on these signals for fault discrimination. This observation is consistent with the physical characteristics of belt loosening in coupled parallel mechanisms, where changes in belt tension alter both the dynamic response of the moving platform and the associated electromagnetic and inertial signatures. In addition, several lightweight statistical enhancement features, including the mean, range, and standard deviation, appear among the top ten most important features. This result suggests that the proposed enhancement effectively captures cross-axis magnitude structure and variability, thereby providing discriminative cues that improve class separability and stabilize model decision boundaries. It should be noted, however, that Random Forest feature importance reflects model reliance rather than causal influence; further validation through ablation studies or complementary explainability techniques would be required to establish causal relationships.

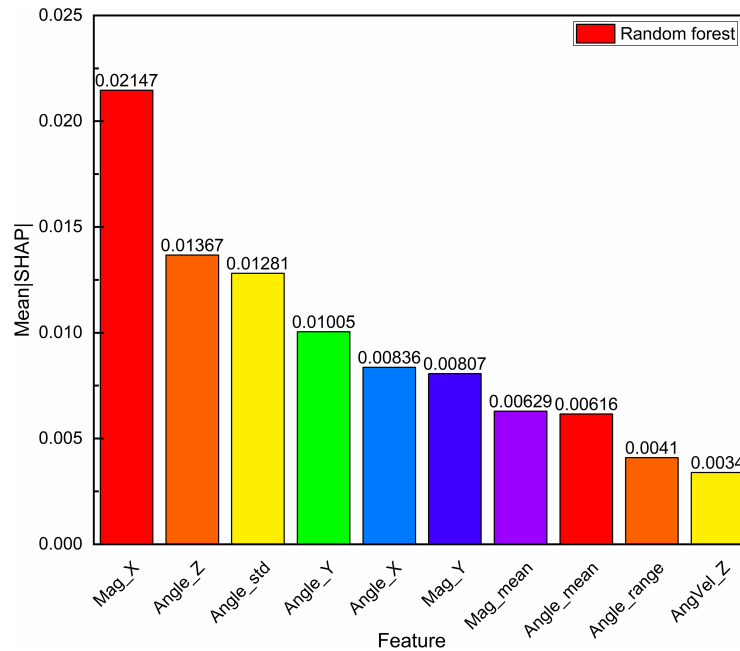


Figure 4: Important feature analysis based on RF.

Under the best-performing preprocessing configuration (Scheme 8), a comparative evaluation of the candidate classifiers is presented in Table 7. Among the evaluated models, LightGBM achieves the highest overall performance, attaining an accuracy of 75.17% and an F1-score of 0.7499, together with a high AUC of 0.984 and a low inference latency of 0.99 ms per sample. This performance can be attributed to LightGBM’s ability to model complex nonlinear decision boundaries while effectively handling feature interactions and residual errors through gradient boosting, making it particularly well suited for the coupled and nonstationary characteristics of parallel mechanism signals. XGBoost also delivers a highly competitive result, achieving an accuracy of 0.71, an F1-score of 0.7499, and an AUC of 0.9788. However, its inference latency of is approximately 0.02 ms per sample, which is lower than that of lightGBM. This difference reflects the higher computational overhead associated with aggregating predictions from a larger ensemble of independently trained trees, leading to a slightly less favorable accuracy–latency trade-off in time-critical deployment scenarios.

Table 7: Stage 1 experimental results of scheme 8.

Model	Accuracy	Precision	Recall	F1	AUC	Inference Delay
RF	0.7053	0.7056	0.7053	0.7012	0.976	0.04
SVM	0.5553	0.5644	0.5553	0.5554	–	0.95
KNN	0.5082	0.526	0.5082	0.5113	0.8329	0.04
XGBoost	0.7101	0.7104	0.7101	0.7076	0.9788	0.02
LightGBM	0.7517	0.7537	0.7517	0.7499	0.984	0.09
CatBoost	0.5648	0.5626	0.5648	0.5609	0.9603	0.04

In contrast, Support Vector Machine and k-Nearest Neighbors exhibit substantially lower predictive performance, with F1-scores of 0.5554 and 0.5113, respectively. The reduced performance of these methods

can be explained by their limited capacity to capture high-order feature interactions and multi-scale dependencies inherent in parallel mechanism fault signatures. Moreover, SVM incurs a markedly higher inference latency of 0.95 ms per sample, primarily due to the reliance on support vector evaluations during prediction, which renders it unsuitable for latency-sensitive or edge-deployed diagnostic applications. These results underscore the advantage of tree-based ensemble methods for fault diagnosis in Delta 3D printers, particularly when both diagnostic accuracy and computational efficiency are critical. The findings further motivate the selection of XGBoost and Random Forest as base learners for subsequent ensemble construction in the proposed three-stage framework.

Fig. 5 summarizes the stage 1 comparison of Random Forest, Support Vector Machine, k-Nearest Neighbors, XGBoost, LightGBM and CatBoost under a unified evaluation protocol, reporting both classification performance and inference latency measured in milliseconds per sample. Among the evaluated methods, LightGBM and XGBoost demonstrate the strongest overall performance, with all major metrics exceeding 0.7, while maintaining inference latencies of 0.09 and 0.02 ms per sample, respectively. This favorable balance can be attributed to the ability of tree-based ensemble methods to model nonlinear feature interactions efficiently without incurring substantial computational overhead at inference time. In contrast, Support Vector Machine exhibits a markedly higher inference latency of 0.95 ms per sample without delivering corresponding performance gains, which significantly limits its suitability for online or real-time deployment scenarios. Based on these results, LightGBM and XGBoost are selected as the candidate base learners for subsequent hyperparameter optimization and ensemble integration within the proposed three-stage framework.

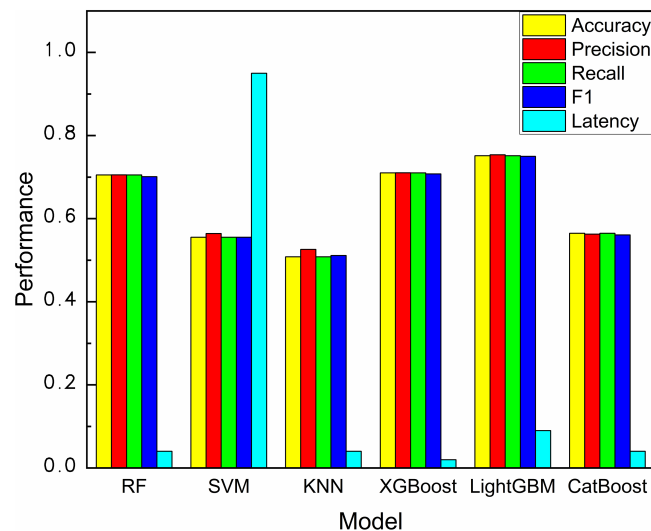


Figure 5: Comparison of performance metrics for different models.

4.2 Experimental Design and Result Analysis of the Stage 2

In Stage 2, the top-performing LightGBM and XGBoost models identified in Stage 1 are selected as the two base learners for further refinement. This stage follows the pipeline defined in the Methodology and is evaluated under a strictly leakage-free protocol (Table 8, Scheme 9). All trainable components are fitted exclusively on the training split, while the validation and test sets are transformed using parameters learned from the training data only; SMOTETomek resampling is applied solely to the training set. To enhance representation capacity and robustness to operating variations, seven additional time–frequency statistical features are incorporated on top of the existing feature set. Key hyperparameters are subsequently optimized

using Bayesian optimization within a fixed computational budget. Based on the optimized base learners, a WSV ensemble is constructed using w probability predictions, enabling principled model fusion while mitigating information leakage. Under this unified pipeline, LightGBM, XGBoost, and their soft-voting fusion are compared. Hyperparameters are tuned via cross-validation using the mean weighted F1-score as the objective, and final performance is reported on the held-out test set.

Table 8: Optimization scheme of stage2.

	Pipeline	Model Column
Scheme 9	Train (fit only). AND-combination (optional)-Grouped-statistics enhancement-Scaler.fit_transform- ReliefF.fit_transform-RFECV.fit_transform- SMOTETomek (train only)	LightGBM/ XGBoost /WSV
	Val/test (transform only). Grouped-statistics enhancement-Scaler.transform- ReliefF.transform-RFECV.transform	

To prevent data leakage during cross-validation, all preprocessing steps—including scaling, feature selection, and resampling (SMOTE)—are integrated into a unified Pipeline. Within each cross-validation fold, these transformations are fitted exclusively on the training subset and subsequently applied to the validation subset, ensuring that no information from validation data is used during training. No global fitting or preprocessing is performed prior to cross-validation. This design guarantees a leakage-free evaluation and ensures the validity of the reported results.

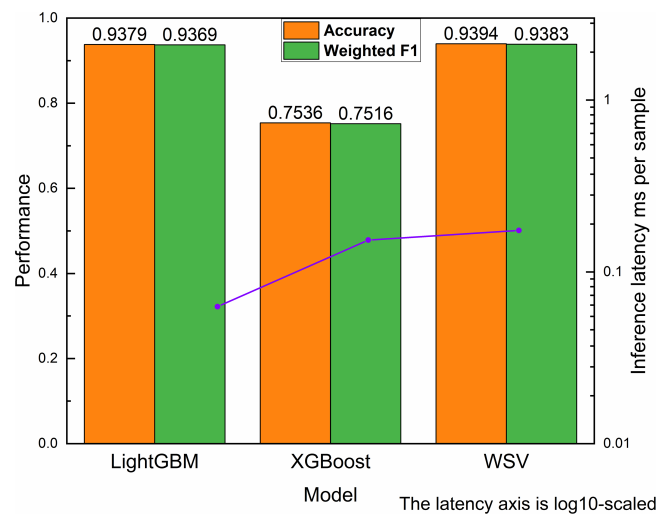
To ensure consistent and fair latency reporting, the measurement scope is explicitly defined to include the entire inference pipeline, namely feature extraction, scaling, feature selection, and model prediction. For the proposed WSV method, latency is measured based on the final deployed model only, without redundant computation of multiple base models during inference. All methods are evaluated under the same hardware environment and measurement protocol, ensuring a consistent and comparable latency assessment.

The impact of Bayesian optimization is summarized in [Table 9](#) and [Fig. 6](#), which indicate that its effectiveness is strongly dependent on both the learner and the adopted search space. With 30 Optuna trials and a time budget of 600 s, LightGBM benefits substantially from optimization: accuracy improves from 0.7517 to 0.93, and the weighted F1-score increases from 0.7499 to 0.9369. This gain is achieved without an increase in inference cost, as latency decreases from 0.09 to 0.0627 ms per sample. In contrast, XGBoost shows only modest improvement after optimization, with accuracy improves from 0.7101 to 0.7221 and weighted F1-score from 0.7076 to 0.7196, while inference latency increases from 0.02 to 0.0973 ms per sample. These results indicate that Bayesian optimization is highly model-dependent and that a broader or more flexible search does not necessarily translate into the same level of benefit across different learners.

[Table 8](#) illustrates that Bayesian optimization influences the two candidate learners in distinct manners. LightGBM significantly benefits from optimization, while XGBoost has only modest improvements. Notwithstanding this disparity, the weighted soft-voting ensemble attains the highest overall accuracy and Macro-F1, suggesting that the two optimized boosting models preserve a level of complementing predictive performance. The ensemble’s latency exceeds that of individual models, indicating that the advantage of fusion primarily results in a marginal performance enhancement rather than improved computing efficiency.

Table 9: Comparison of outcomes between Stage 2 and Stage 1.

Model	Config	Accuracy	Macro-F1	AUC	Inference Delay ms Sample
LightGBM	Grid search	0.7517	0.7499	0.984	0.09
LightGBM	Bayesian	0.9379	0.9369	0.9936	0.0627
XGBoost	Grid search	0.7101	0.7076	0.9788	0.02
XGBoost	Bayesian	0.7221	0.7196	0.9734	0.0973
WSV	Bayesian	0.9394	0.9383	0.9900	0.1742

**Figure 6:** Comparison of performance metrics for different models for stage 2.

The results of Recursive Feature Elimination with Cross-Validation (RFECV) applied in stage 2 are presented in Fig. 7. The RFECV curve exhibits a rapid increase in cross-validation performance as the number of selected features grows from a minimal subset to approximately 4 features, after which performance reaches a plateau. Beyond this range, further feature expansion yields diminishing returns and may introduce redundancy or mild overfitting. This behavior confirms the effectiveness of the two-stage feature selection strategy adopted in stage 2, in which ReliefF is first used to identify discriminative features and RFECV is subsequently employed to remove redundancy. Together, these steps contribute to stable generalization performance by balancing representational richness and model complexity.

To prevent overinterpretation of a singular important metric, feature relevance in Stage 2 was analyzed from four complementary viewpoints, as seen in Fig. 8: LightGBM important ranking, permutation importance, XGBoost mean absolute SHAP values, and XGBoost five-fold stability study. Throughout these analyses, various magnetic-field-related characteristics, such as Mag_Min, Mag_Skew, Mag_Y, and Mag_Std, consistently emerge as the highest-ranked predictors. This consistency indicates that the observed ranking pattern is not simply a byproduct of a singular model or relevance metric. These findings are construed as markers of model dependence and predictive correlation rather than as proof of causal physical determinants. The recognized significance pattern aligns with the hypothesis that belt looseness may influence the consistency of end-effector motion and the responses of connected sensors; nevertheless, the interpretability analysis does not demonstrate causality.

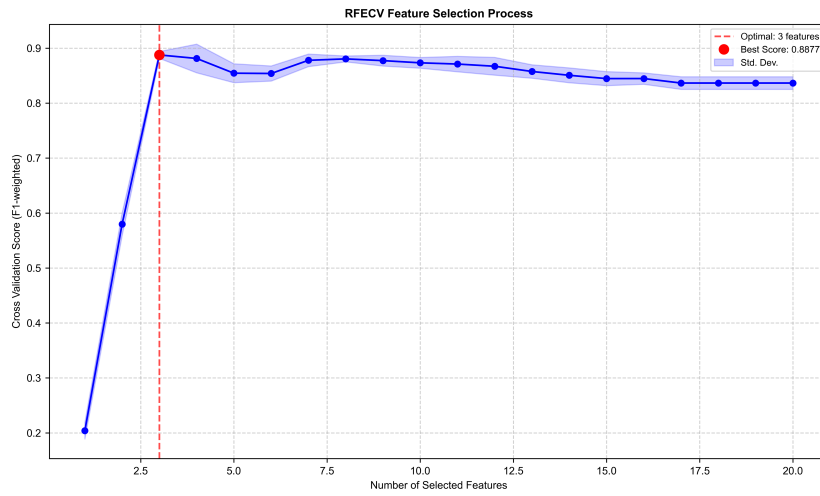


Figure 7: Rfcv_feature_selection for stage 2.

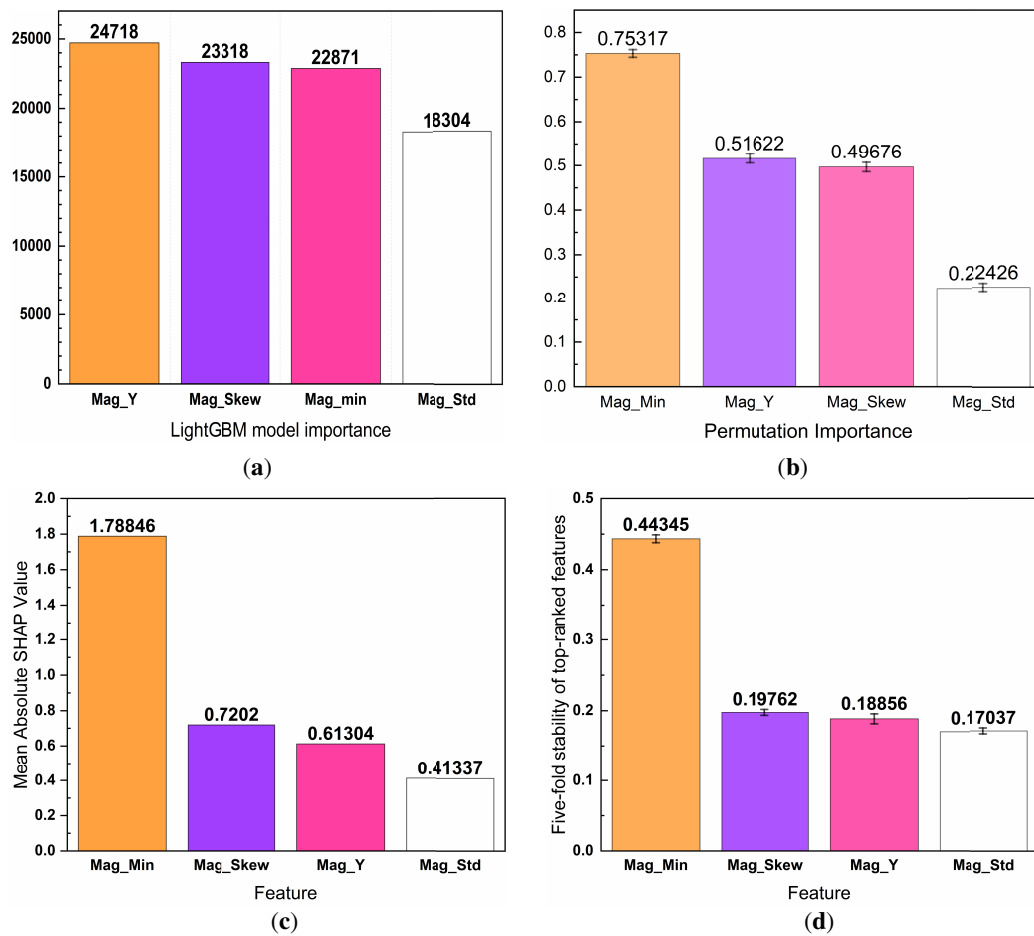


Figure 8: Complementary feature-relevance analyses in Stage 2: (a) LightGBM feature importance, (b) permutation importance, (c) XGBoost mean absolute SHAP values, (d) XGBoost five-fold stability analysis.

Table 10 demonstrates that the highest-ranked XGBoost characteristics exhibit considerable stability throughout the five folds. All four features demonstrate minimal standard deviations and low coefficients of variation, with stability scores consistently exceeding 0.96. Mag_Min exhibits the highest average significance and stability score, indicating that it is the most reliably influential predictor inside the XGBoost model. The remaining features, such as Mag_Skew, Mag_Y, and Mag_Std, exhibit consistent importance rankings throughout resampling folds. This outcome suggests that the recognized feature pattern is resilient to data partitioning and is hence improbable to be a split-specific artifact.

Table 10: XGBoost five-fold stability analysis.

Feature	Mean_Importance	Std_Importance	CV	Stability_Score
Mag_Min	0.4434	0.0056	0.0125	0.9876
Mag_Skew	0.1976	0.0041	0.0210	0.9794
Mag_Y	0.1886	0.0070	0.0371	0.9642
Mag_Std	0.1704	0.0043	0.0252	0.9755

The confusion matrix of the WSV ensemble in Stage 2, shown in Fig. 9, exhibits a strong diagonal dominance on the test set, indicating that the majority of fault classes are reliably and consistently distinguished. The remaining misclassifications are largely confined to adjacent classes, which can be attributed to partial overlap in feature distributions caused by the continuous and gradual nature of belt-loosening severity. From a modeling perspective, such behavior is expected when fault progression follows a smooth physical continuum rather than discrete state transitions. Importantly, this error pattern suggests that misclassifications arise primarily from inherently ambiguous boundary cases, rather than from systematic confusion between physically unrelated fault categories, reflecting a well-structured decision space learned by the ensemble.

The multi-class receiver ROC curves of the WSV ensemble for stage 2 are presented in Fig. 10. The micro-averaged AUC reaches approximately 0.99, while most individual classes achieve AUC values close to unity, demonstrating strong class separability at the probabilistic output level. A small subset of classes exhibits slightly reduced AUC values, which is consistent with the neighboring-class confusions observed in the confusion matrix. This correspondence indicates that the residual classification difficulty is localized to a limited number of boundary classes where physical differences between fault states are subtle. Overall, these results confirm that the WSV ensemble effectively captures the dominant discriminative structure of the fault space, and that remaining performance limitations are governed primarily by intrinsic signal ambiguity rather than deficiencies in model capacity or training strategy. Overall, Stage 2 demonstrates the effectiveness of Plan 9 when combined with WSV ensemble learning, yielding stable improvements in overall performance together with a well-founded feature selection outcome. The residual errors are predominantly concentrated in boundary samples corresponding to neighboring severity levels, indicating that the dominant challenge arises from local decision uncertainty rather than global model inadequacy. Motivated by this observation, Stage 3 adopts a more stringent compound-fault evaluation protocol to assess generalization performance under previously unseen fault combinations.

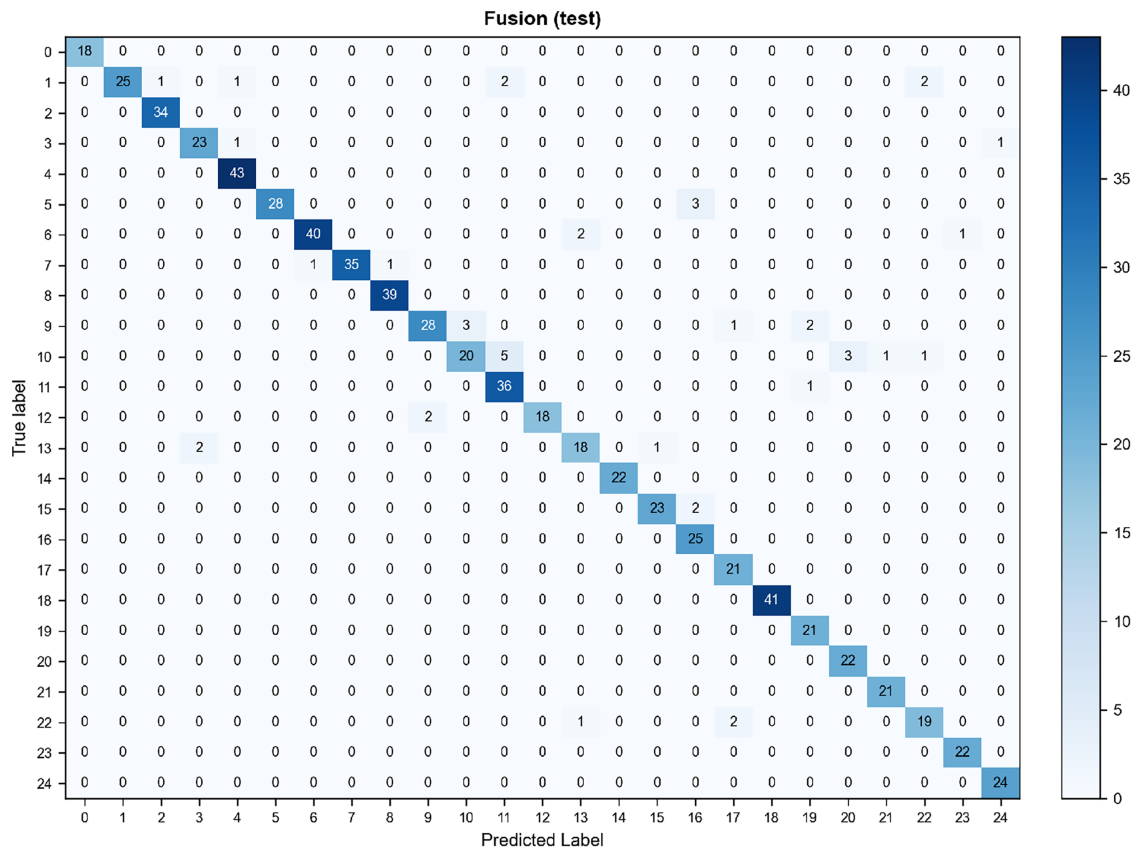


Figure 9: Confusion matrix of WSV for stage 2.

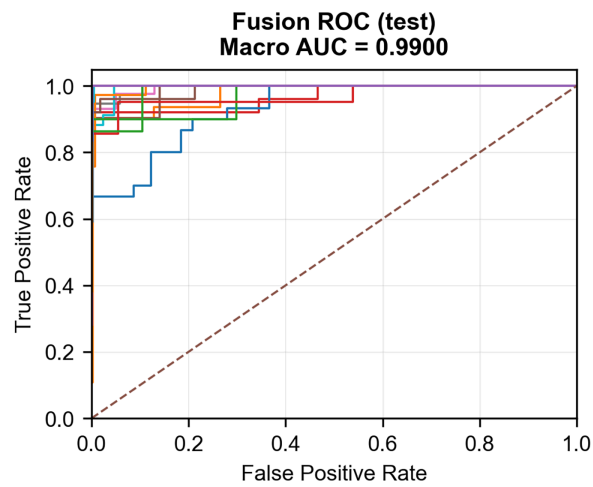


Figure 10: Roc_curves of WSV for stage 2.

4.3 Experimental Design and Result Analysis of the Stage 3

Stage 3 introduces real compound-fault samples to evaluate the adaptability of the proposed framework under distribution shift and changes in label composition. To this end, compound-fault diagnosis is reformulated as a three-output classification problem that jointly predicts the health states of Belts A, B, and C for each

sample. This formulation preserves belt-level diagnostic information while avoiding explicit enumeration of compound fault classes, thereby providing a scalable representation for mixed fault conditions. Compound-fault samples are partitioned into training, validation, and test sets using a 70–10–20 split and are combined with single-fault samples to form a mixed dataset. The resulting training set contains 11,112 samples, including 5189 single-fault samples and 6423 compound-fault samples. The validation set comprises 1641 samples, with 733 single-fault and 908 compound-fault instances, while the test set includes 3312 samples, consisting of 1466 single-fault and 1846 compound-fault samples. This composition reflects a realistic operating scenario in which single and compound faults coexist and jointly define the data distribution.

Table 11 reports the direct-transfer baseline results. In this setting, the preprocessing pipeline and feature selectors optimized in stage 2 are reused, and the single-fault models trained in stage 2 are directly evaluated on the stage 3 mixed test set without any adaptation. The findings indicate that direct transfer is viable, although its efficacy is significantly contingent upon the model employed. LightGBM demonstrates superior performance, whereas XGBoost is significantly less effective, and the WSV ensemble fails to surpass the most robust individual learner. These data indicate that, while direct transfer may be effective under the existing technique, reliable compound-fault detection remains significantly reliant on the chosen model. This necessitates the introduction of the multi-output formulation in the ensuing study, aimed at delivering a more systematic representation of compound fault states and enhancing diagnostic consistency under interrelated degradation scenarios.

Table 11: Direct transfer baseline for stage 2 to stage 3.

Model	Accuracy	Macro-F1	Macro-AUC	Inference Latency (ms/Sample)
LightGBM	0.9505	0.9507	0.9910	0.1249
XGB	0.8166	0.8158	0.9771	0.0826
WSV	0.9505	0.9507	0.9910	0.1249

Tables 12–14 encapsulate the Stage 3 multi-output findings from three complimentary viewpoints: overall performance, subset-specific performance, and belt-level performance. Table 11 illustrates that XGBoost attains commendable overall performance with a Macro-F1 score of 0.9285 and a Hamming loss of 0.0573, but the weighted soft-voting (WSV) model enhances these metrics marginally to 0.9295 and 0.0571, respectively. Despite the slight improvement over the most robust individual model, the multi-output approach offers a systematic depiction of compound faults by concurrently forecasting the health statuses of Belt A, Belt B, and Belt C. The WSV model exhibits remarkable consistency across both single-fault and compound-fault subgroups, achieving Macro-F1 scores of 0.9305 and 0.9290, respectively, which signifies robust diagnostic performance under diverse fault scenarios.

Table 12: Stage 3 multi output results on overall and subset test sets.

Model	Overall Macro-F1	Overall Hamming Loss	Single-Fault Macro-F1	Compound-Fault Macro-F1	Inference Latency (ms/Sample)
LightGBM	0.7662	0.2223	0.8186	0.7633	0.0345
XGB	0.9285	0.0573	0.9298	0.9280	0.0985
WSV	0.9295	0.0571	0.9305	0.9290	0.1451

Table 13: Belt level performance of the multi output models on the overall test set.

Model	A-Belt Macro-F1	B-Belt Macro-F1	C-Belt Macro-F1
LightGBM	0.8317	0.7141	0.7526
XGB	0.9525	0.9177	0.9154
WSV	0.9514	0.9179	0.9191

Table 14: Belt level performance of the multi output models on the compound fault test subset.

Model	A-Belt Macro-F1	B-Belt Macro-F1	C-Belt Macro-F1
LightGBM	0.8293	0.7115	0.7492
XGB	0.9518	0.9171	0.9151
WSV	0.9508	0.9173	0.9189

Tables 13 and 14 report the belt-wise Macro-F1 scores of the multi-output models on the overall test set and the compound-fault subset, respectively. Across all evaluated learners, Belt A consistently achieves the highest Macro-F1, whereas Belt B exhibits the lowest scores. This pattern indicates non-uniform diagnostic separability across belts under an identical sensing configuration and preprocessing pipeline, highlighting inherent differences in belt-wise fault observability. In a Delta parallel mechanism, the end-effector attitude sensor captures the superposed dynamic response of three strongly coupled transmission chains rather than isolated belt-specific signatures.

Consequently, belt-wise observability is governed by a combination of belt-to-sensor transfer characteristics, axis-dependent sensitivity of the inertial measurements, and trajectory-dependent excitation of the mechanism. Belts that induce stronger or more directionally aligned perturbations in the measured attitude and magnetic-field signals tend to yield higher diagnostic separability, whereas belts whose effects are partially masked by coupling or symmetry exhibit lower classification performance. Among the evaluated models, XGBoost consistently remains the best-performing learner across belts. The weighted soft-voting fusion produces results identical to those of XGBoost, indicating that the validation-selected global fusion weight is effectively dominated by the stronger base learner under the current data split and feature representation. This outcome further suggests that, in the present setting, performance gains are primarily driven by improved representation and output formulation rather than by averaging complementary decision boundaries.

Fig. 11 presents the confusion matrix of the WSV model evaluated on the stage 3 compound-fault subset. The residual misclassifications are predominantly concentrated between neighboring severity levels rather than across distant fault categories, indicating that the principal limitation in the compound-fault setting arises from local decision boundary uncertainty rather than structural collapse of class separability. This behavior is consistent with the continuous progression of belt loosening, where adjacent severity levels exhibit highly similar signal characteristics under coupled operating conditions. Among the three outputs, Belt B exhibits a slightly broader off-diagonal dispersion, suggesting comparatively weaker separability under compound disturbances. This observation implies that the fault signatures associated with Belt B are more strongly masked by inter-belt coupling effects or exhibit reduced sensitivity in the measured sensor channels, making fine-grained severity discrimination more challenging. Overall, the confusion pattern confirms that the proposed multi-output framework preserves meaningful fault structure under compound conditions,

with remaining errors largely attributable to intrinsic physical ambiguity rather than deficiencies in the modeling approach.

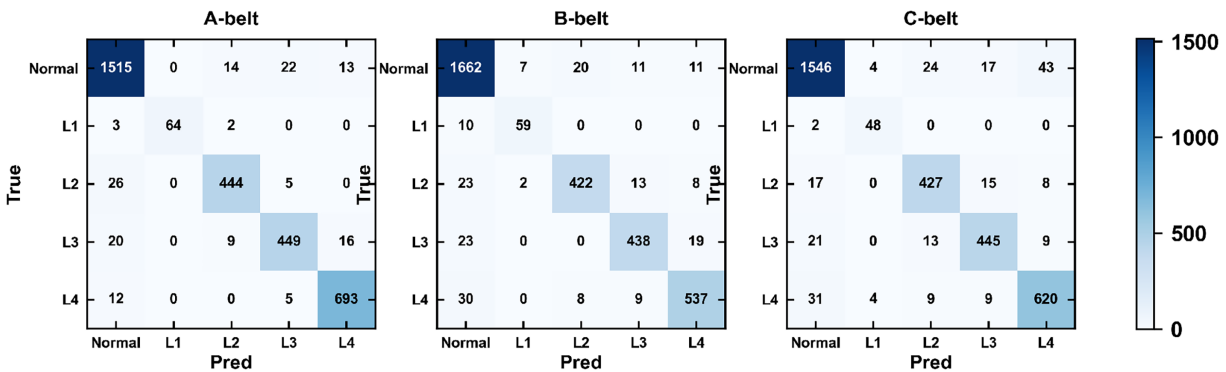


Figure 11: WSV compound fault confusion matrix for stage 3.

Fig. 12 presents the belt-wise multiclass ROC curves of the WSV model evaluated on the Stage 3 compound-fault subset. The macro-averaged AUC reaches 0.991 for Belt A, 0.981 for Belt B, and 0.982 for Belt C. For all three belts, most class-specific ROC curves cluster near the upper-left corner, indicating strong separability at the probabilistic output level even under compound-fault conditions. The comparatively lower AUC observed for Belt B suggests that the remaining classification ambiguity is concentrated in a limited number of more challenging classes when distribution shift is present. This behavior is consistent with earlier belt-wise analyses, in which Belt B exhibited weaker separability due to stronger coupling effects and reduced sensor observability. Importantly, the high AUC values across all belts demonstrate that, despite increased complexity under compound faults, the proposed multi-output framework maintains robust probabilistic discrimination, with residual performance degradation primarily driven by intrinsic physical ambiguity rather than a loss of global model discriminative capacity.

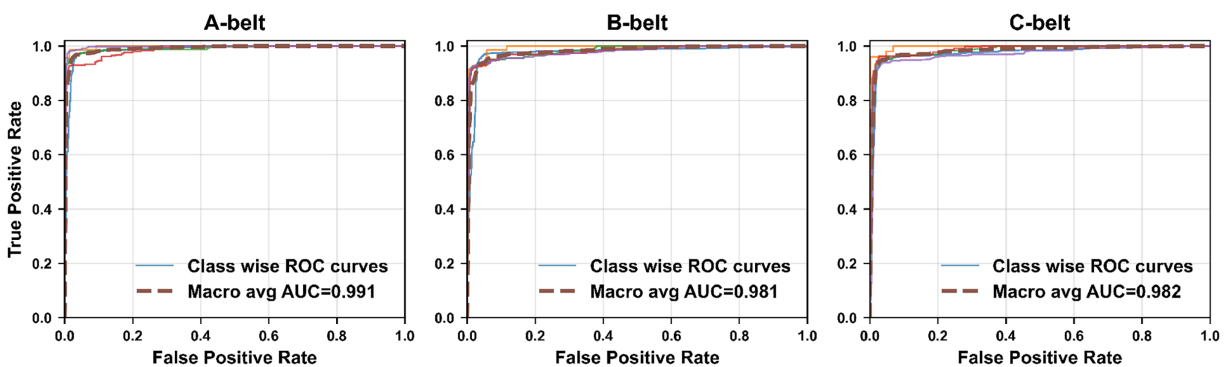


Figure 12: WSV compound fault AUC curves for stage 3.

Fig. 13 presents the belt-wise 95% bootstrap confidence intervals (CIs) of the WSV model on the compound-fault test subset for F1-score, precision, recall, and accuracy, where each belt corresponds to an individual output head in the multi-output formulation. Across all metrics, the confidence intervals are consistently narrow, indicating that the belt-level performance estimates are statistically stable under resampling and are not dominated by a small number of test instances. Among the three outputs, Belt A exhibits the highest central estimates across all evaluated metrics, whereas Belt B yields the lowest values. This

pattern suggests persistent belt-dependent differences in diagnostic separability under an identical sensing configuration and preprocessing pipeline. From an interpretive perspective, these differences are consistent with asymmetric belt-to-sensor transfer characteristics and coupling effects in the Delta mechanism, which influence the observability of fault signatures at the end-effector. Overall, the narrow confidence intervals reinforce the robustness of the proposed multi-output framework, while the belt-wise trends highlight intrinsic physical factors that govern fault discriminability under compound operating conditions.

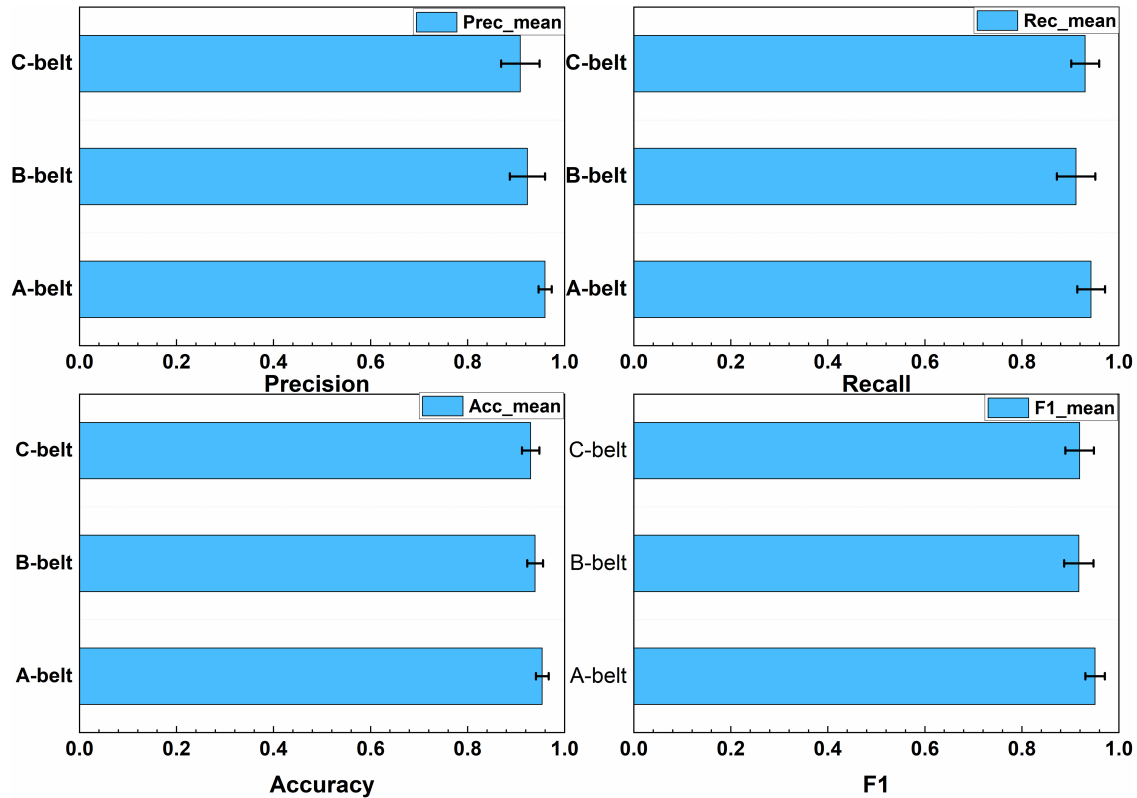


Figure 13: Belt-wise 95% confidence intervals on the compound-fault subset in stage 3 (WSV).

Table 15 further summarizes the overall three-output performance using the Macro-F1 (3-out) metric. The WSV model attains a Macro-F1 (3-out) score of 0.9290, with a corresponding 95% confidence interval of 0.9198–0.9379, demonstrating stable and consistent compound-fault recognition capability across bootstrap resamples. The observed belt-wise performance disparity aligns with belt-dependent observability in a coupled Delta mechanism, where diagnostic sensitivity is primarily governed by transmission characteristics and axis-directional responsiveness of the sensing modalities, rather than by geometric proximity alone. This result reinforces the suitability of the multi-output formulation for compound-fault diagnosis, as it preserves belt-level distinctions while maintaining robust aggregate performance under uncertainty.

Table 15: Macro-F1 (3-out) with 95% confidence interval on the compound-fault subset in stage 3 (WSV).

Metric	Estimate	Lower 95% CI	Upper 95% CI
Macro-F1 (3-out)	0.9290	0.9198	0.9379

Note: 95% CIs are obtained via bootstrap resampling on the test compound-fault subset.

4.4 Summary of Findings

Fig. 14 illustrates an exemplary printed specimen generated in the designated printing task. The figure is presented to demonstrate the physical printing result and the practical experimental framework of the investigation, rather than to establish a direct visual correlation between defect severity and final print quality. Under the current experimental conditions, mild and moderate belt-loosening scenarios represent progressive mechanical degradation and do not consistently produce clearly distinguishable macroscopic defects in the printed product. Representative samples under selected fault conditions were visually inspected; however, no consistent macroscopic defect pattern was observed for these mild and moderate degradation levels. This observation further supports the necessity of sensor-based diagnosis during the printing process, as reliance on post hoc visual inspection alone is insufficient to reliably characterize early-stage faults. Accordingly, the primary aim of this study is to identify mechanically relevant degradation states from multisensor dynamic responses throughout the printing process, rather than to classify faults solely based on the visual appearance of the finished printed components.



Figure 14: Representative printed sample from the adopted Delta printing task.

Table 16 displays a progressive component analysis instead of a rigid same-task ablation. The initial three configurations assess the cumulative benefits of preprocessing and fusion in a single-fault context, while the final two configurations investigate the transferability of the optimized pipeline to the compound-fault task and the impact of the subsequent multi-output reformulation on diagnostic behavior within a structured compound-fault framework. The chosen Stage 1 preprocessing pipeline yields a little enhancement from C1 to C2 compared to the baseline. A far greater improvement is noted in C3, suggesting that the enhanced LightGBM–XGBoost weighted fusion is a primary factor in the ultimate single-fault performance. This streamlined pipeline in C4 is directly applied to the compound-fault problem, yielding robust results and demonstrating that the acquired representation and fusion technique maintain significant predictive capability in more intricate fault scenarios. C5 redefines the compound-fault task as a multi-output problem. This reformulation, while not enhancing the flat aggregate measure compared to C4, offers a more authentic depiction of compound defects, facilitates belt-wise understanding, and accommodates Hamming-loss-based structured evaluation. Consequently, the significance of the last stage resides in systematic fault modeling rather than merely augmenting a singular aggregate score.

The staged experiments establish a reproducible and transparent decision pathway from preprocessing design to model selection and generalization assessment. Stage 1 identifies the most effective preprocessing configuration together with the two strongest base learners under a unified evaluation protocol. Stage 2 demonstrates that Bayesian optimization substantially improves single-fault diagnostic performance and that ensemble learning, through weighted soft voting, yields the most robust overall results. Stage 3 further

delineates the generalization boundary under compound-fault conditions and provides empirical support for a multi-output formulation with a WSV backbone as the final configuration. Despite these gains, the remaining performance limitation is primarily associated with belt-wise ambiguity induced by strong inter-belt coupling. This observation suggests that future improvements are more likely to arise from enhancing belt-level separability-through refined sensing strategies, trajectory design, or representation learning—rather than from further increasing classifier complexity.

Table 16: Progressive evaluation of the main methodological components across task settings.

ID	Configuration	Accuracy	Macro-F1	Hamming Loss	Inference Latency (ms/Sample)
C1	Baseline single model on single-fault task	0.6849	0.6815	–	0.02
C2	C1 + selected Stage 1 preprocessing	0.7101	0.7076	–	0.02
C3	C2 + optimized LightGBM–XGBoost weighted fusion	0.9394	0.9383	–	0.1742
C4	Direct transfer of C3 to compound-fault task	0.9505	0.9507	–	0.1249
C5	C4 + multi-output reformulation	0.929	0.929	0.071	0.1451

5 Conclusion

This paper proposes an interpretable three-stage fault diagnosis framework for delta-type parallel mechanisms, using a Delta 3D printer as a representative experimental platform. The framework targets two practical bottlenecks in diagnosing parallel mechanisms: pipeline sensitivity under noisy and imbalanced measurements, and limited robustness when transferring from single-fault identification to compound-fault scenarios under distribution shift. To address these issues, a leak-free and fully aligned preprocessing pipeline is adopted, in which feature enhancement and two-stage feature selection are performed using training data only, followed by model training with representative learners and ensemble fusion. To avoid explicit compound-class enumeration while preserving maintenance-relevant belt-level information, stage 3 reformulates compound-fault diagnosis as a multi-output classification task. The direct transition from Stage 2 to Stage 3 is viable for certain learners, although its efficacy is evidently contingent upon the paradigm employed. Conversely, the Stage 3 multi-output formulation offers a more organized depiction of compound failures by explicitly forecasting the belt-specific states of the three transmission chains. In this context, the weighted soft-voting model attains a Macro-F1 score of 0.9295 on the comprehensive test set, along with very consistent outcomes on the single-fault and compound-fault subsets, which yield Macro-F1 scores of 0.9305 and 0.9290, respectively. The little disparity between these two subset results suggests that the suggested structural formulation preserves consistent diagnostic efficacy across diverse fault circumstances, while facilitating belt-wise interpretation via multi-output prediction. Confusion-matrix analysis shows that the remaining errors are mainly concentrated around adjacent severity levels rather than across distinct fault categories, suggesting that the dominant limitation arises from local decision

uncertainty. These observations indicate that future work should focus on uncertainty-aware decision rules, probability calibration, and evaluation across additional platforms and operating conditions, in order to further strengthen cross-scenario generalization.

In this study, faults are defined as belt looseness induced by controlled screw rotation, which provides a clear, repeatable, and quantitatively adjustable experimental setting. While this definition enables precise labeling and systematic variation of fault severity, it represents a specific type of degradation and may not fully capture the complexity of real-world fault mechanisms, such as wear, misalignment, or material aging. Nevertheless, the proposed framework is sensor-driven and model-agnostic, and can be extended to other fault types given appropriate data. Future work will focus on validating the approach under more diverse and realistic fault conditions to further enhance its generalizability.

Acknowledgement: None.

Funding Statement: This research was funded by Guangdong University Research Project with Grant No. 2022KTSCX219 and Yuejiao Gao Han with Grant No. [2021]4.

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Lin Fang, Razi Abdul-Rahman, Cheng-Fu Yang; Methodology, Lin Fang, Razi Abdul-Rahman, Cheng-Fu Yang; Software, Lin Fang; Validation, Lin Fang; Formal analysis, Lin Fang; Investigation, Lin Fang, Razi Abdul-Rahman, Cheng-Fu Yang; Data curation, Lin Fang; Writing—original draft preparation, Lin Fang, Razi Abdul-Rahman, Cheng-Fu Yang; Writing—review and editing, Razi Abdul-Rahman, Cheng-Fu Yang; Visualization, Lin Fang; Supervision, Razi Abdul-Rahman, Cheng-Fu Yang; Project administration, Lin Fang; Funding acquisition, Lin Fang. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: Due to the nature of this research, participants of this study did not agree for their data to be shared publicly, so supporting data is not available.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Russo M, Zhang D, Liu XJ, Xie Z. A review of parallel kinematic machine tools: design, modeling, and applications. *Int J Mach Tools Manuf.* 2024;196(1):104118. doi:10.1016/j.ijmachtools.2024.104118.
2. Bruno Siciliano LS. *Robotics: modelling, planning and control.* London, UK: Springer; 2009.
3. Khan ZH, Khalid A, Iqbal J. Towards realizing robotic potential in future intelligent food manufacturing systems. *Innov Food Sci Emerg Technol.* 2018;48(1):11–24. doi:10.1016/j.ifset.2018.05.011.
4. Zhang H, Peeters J, Demeester E, Duflou JR, Kellens K. A CNN-based fast picking method for WEEE recycling. *Procedia CIRP.* 2022;106(21):264–9. doi:10.1016/j.procir.2022.02.189.
5. Fang S, Cao J, Zhang Z, Zhang Q, Cheng W. Study on high-speed and smooth transfer of robot motion trajectory based on modified S-shaped acceleration/deceleration algorithm. *IEEE Access.* 2020;8:199747–58. doi:10.1109/ACCESS.2020.3035430.
6. Zhang Z, Meng Q, Cui Z, Yao M, Shao Z, Tao B. Machine learning applications in parallel robots: a brief review. *Machines.* 2025;13(7):565. doi:10.3390/machines13070565.
7. Yang C, Ye W, Li Q. Review of the performance optimization of parallel manipulators. *Mech Mach Theory.* 2022;170(2):104725. doi:10.1016/j.mechmachtheory.2022.104725.
8. Hassan G, Chemori A, Gouttefarde M, El Rafei M, Francis C, Hervé PE, et al. A new augmented RISE feedback controller for pick-and-throw applications with PKMs. *IFAC-PapersOnLine.* 2022;55(38):19–25. doi:10.1016/j.ifacol.2023.01.128.

9. McClintock H, Temel FZ, Doshi N, Koh JS, Wood RJ. The milliDelta: a high-bandwidth, high-precision, millimeter-scale Delta robot. *Sci Robot*. 2018;3(14):eaar3018. doi:10.1126/scirobotics.aar3018.
10. Zakharov OV, Pugin KG, Ivanova TN. Modeling and analysis of delta kinematics FDM printer. *J Phys Conf Ser*. 2022;2182(1):012069. doi:10.1088/1742-6596/2182/1/012069.
11. Edoimioya N, Chou CH, Okwudire CE. Vibration compensation of delta 3D printer with position-varying dynamics using filtered B-splines. *Int J Adv Manuf Technol*. 2023;125(5):2851–68. doi:10.1007/s00170-022-10789-w.
12. Fathi K, van de Venn HW, Honegger M. Predictive maintenance: an autoencoder anomaly-based approach for a 3 DoF delta robot. *Sensors*. 2021;21(21):6979. doi:10.3390/s21216979.
13. Jaber AA, Bicker R. Development of a condition monitoring algorithm for industrial robots based on artificial intelligence and signal processing techniques. *Int J Electr Comput Eng*. 2018;8(2):996. doi:10.11591/ijece.v8i2.pp996-1009.
14. Zhang S, Sun Z, Li C, Cabrera D, Long J, Bai Y. Deep hybrid state network with feature reinforcement for intelligent fault diagnosis of delta 3-D printers. *IEEE Trans Ind Inform*. 2020;16(2):779–89. doi:10.1109/TII.2019.2920661.
15. Dorian V, Louis G, Elodie C, Louise TM, Sébastien D. Machine learning based fault anticipation for 3D printing. *IFAC-PapersOnLine*. 2023;56(2):2927–32. doi:10.1016/j.ifacol.2023.10.1414.
16. Liu S, Jiang H, Wu Z, Li X. Data synthesis using deep feature enhanced generative adversarial networks for rolling bearing imbalanced fault diagnosis. *Mech Syst Signal Process*. 2022;163:108139. doi:10.1016/j.ymssp.2021.108139.
17. Zhang L, Lin J, Shao H, Yang Z, Liu B, Li C. An unsupervised end-to-end approach to fault detection in delta 3D printers using deep support vector data description. *J Manuf Syst*. 2024;72(1):214–28. doi:10.1016/j.jmsy.2023.11.020.
18. Ahmad M, Mohd-Mokhta R. A survey on model-based fault detection techniques for linear time-invariant systems with numerical analysis. *Pertanika J Sci Technol*. 2022;30(1):53–78. doi:10.47836/pjst.30.1.04.
19. Bortoff SA. Object-oriented modeling and control of delta robots. In: *Proceedings of the 2018 IEEE Conference on Control Technology and Applications (CCTA); 2018 Aug 21–24; Copenhagen, Denmark*. New York, NY, USA: IEEE; 2018. p. 251–8. doi:10.1109/CCTA.2018.8511395.
20. Kumar S, Wöhrle H, de Gea Fernández J, Müller A, Kirchner F. A survey on modularity and distributivity in series-parallel hybrid robots. *Mechatronics*. 2020;68(1):102367. doi:10.1016/j.mechatronics.2020.102367.
21. Mardt F, Bischof P, Thielecke F. Design methodology for robust model-based fault diagnosis schemes and its application to an aircraft hydraulic power package. *PHM Soc Eur Conf*. 2022;7(1):315–28. doi:10.36001/phme.2022.v7i1.3339.
22. He K, Yang Z, Bai Y, Long J, Li C. Intelligent fault diagnosis of delta 3D printers using attitude sensors based on support vector machines. *Sensors*. 2018;18(4):1298. doi:10.3390/s18041298.
23. Guo J, Wu J, Sun Z, Long J, Zhang S. Fault diagnosis of delta 3D printers using transfer support vector machine with attitude signals. *IEEE Access*. 2019;7:40359–68. doi:10.1109/ACCESS.2019.2905264.
24. Li X, Guo J, Jia X, Zhang S, Liu Z. Intelligent fault diagnosis of delta 3D printers using attitude sensors based on extreme learning machines. *Int J Performability Eng*. 2019;15(12):3196. doi:10.23940/ijpe.19.12.p11.31963208.
25. Qin YX, Hong Y, Long JY, Yang Z, Huang YW, Li C. Attitude data-based deep transfer capsule network for intelligent fault diagnosis of delta 3D printers. *J Phys Conf Ser*. 2022;2184(1):012017. doi:10.1088/1742-6596/2184/1/012017.
26. Verana M, Nwakanma CI, Lee JM, Kim DS. Deep learning-based 3D printer fault detection. In: *Proceedings of the 2021 Twelfth International Conference on Ubiquitous and Future Networks (ICUFN); 2021 Aug 17–20; Jeju Island, Republic of Korea*. New York, NY, USA: IEEE; 2021. p. 99–102. doi:10.1109/icufn49451.2021.9528692.
27. Choi K, Yi J, Park C, Yoon S. Deep learning for anomaly detection in time-series data: review, analysis, and guidelines. *IEEE Access*. 2021;9:120043–65. doi:10.1109/ACCESS.2021.3107975.
28. Isiani A, Weiss L, Bardaweel H, Nguyen H, Crittenden K. Fault detection in 3D printing: a study on sensor positioning and vibrational patterns. *Sensors*. 2023;23(17):7524. doi:10.3390/s23177524.
29. Liu FT, Ting KM, Zhou ZH. Isolation forest. In: *Proceedings of the 2008 Eighth IEEE International Conference on Data Mining; 2008 Dec 15–19; Pisa, Italy*. New York, NY, USA: IEEE; 2009. p. 413–22. doi:10.1109/ICDM.2008.17.

30. Luan S, Gu Z, Freidovich LB, Jiang L, Zhao Q. Out-of-distribution detection for deep neural networks with isolation forest and local outlier factor. *IEEE Access*. 2021;9:132980–9. doi:10.1109/ACCESS.2021.3108451.
31. Aggarwal N, Shukla U, Saxena GJ, Rawat M, Bafila AS, Singh S, et al. Mean based relief: an improved feature selection method based on ReliefF. *Appl Intell*. 2023;53(19):23004–28. doi:10.1007/s10489-023-04662-w.
32. Sun L, Yin T, Ding W, Qian Y, Xu J. Multilabel feature selection using ML-ReliefF and neighborhood mutual information for multilabel neighborhood decision systems. *Inf Sci*. 2020;537(4):401–24. doi:10.1016/j.ins.2020.05.102.
33. Nemat Saberi A, Belahcen A, Sobra J, Vaimann T. LightGBM-based fault diagnosis of rotating machinery under changing working conditions using modified recursive feature elimination. *IEEE Access*. 2022;10:81910–25. doi:10.1109/ACCESS.2022.3195939.
34. Nie L, Wu R, Ren Y, Tan M. Research on fault diagnosis of HVAC systems based on the ReliefF-RFECV-SVM combined model. *Actuators*. 2023;12(6):242. doi:10.3390/act12060242.
35. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *Jair*. 2002;16:321–57. doi:10.1613/jair.953.
36. Yang X, Xu X, Wang Y, Liu S, Bai X, Jing L, et al. The fault diagnosis of a plunger pump based on the SMOTE + tomeK link and dual-channel feature fusion. *Appl Sci*. 2024;14(11):4785. doi:10.3390/app14114785.
37. Cerrada M, Zurita G, Cabrera D, Sánchez RV, Artés M, Li C. Fault diagnosis in spur gears based on genetic algorithm and random forest. *Mech Syst Signal Process*. 2016;70:87–103. doi:10.1016/j.ymssp.2015.08.030.
38. Maincer D, Benmahamed Y, Mansour M, Alharthi M, Ghoneim SSM. Fault diagnosis in robot manipulators using SVM and KNN. *Intell Autom Soft Comput*. 2023;35(2):1957–69. doi:10.32604/iasc.2023.029210.
39. Qiu C, Zhang L, Li M, Zhang P, Zheng X. Elevator fault diagnosis method based on IAO-XGBoost under unbalanced samples. *Appl Sci*. 2023;13(19):10968. doi:10.3390/app131910968.
40. Yan X, Jia M. A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing. *Neurocomputing*. 2018;313:47–64. doi:10.1016/j.neucom.2018.05.002.
41. Du M, Wang Z, Zhang Z, Zhang X. A grid fault diagnosis method based on stacking algorithm. *J Phys Conf Ser*. 2023;2477(1):012067. doi:10.1088/1742-6596/2477/1/012067.
42. Al-Haddad LA, Jaber AA, Al-Haddad SA, Al-Muslim YM. Fault diagnosis of actuator damage in UAVs using embedded recorded data and stacked machine learning models. *J Supercomput*. 2024;80(3):3005–24. doi:10.1007/s11227-023-05584-7.
43. Yao L, Fang Z, Xiao Y, Hou J, Fu Z. An intelligent fault diagnosis method for lithium battery systems based on grid search support vector machine. *Energy*. 2021;214(3):118866. doi:10.1016/j.energy.2020.118866.
44. Zulfqar M, Gamage KAA, Kamran M, Rasheed MB. Hyperparameter optimization of Bayesian neural network using Bayesian optimization and intelligent feature engineering for load forecasting. *Sensors*. 2022;22(12):4446. doi:10.3390/s22124446.
45. Wang M, Sun Z. Intelligent fault diagnosis of delta 3D printers using local support vector machine by a cheap attitude multi-sensor. In: *Proceedings of the 2020 Prognostics and Health Management Conference (PHM-Besançon)*; 2020 May 4–7; Besançon, France. New York, NY, USA: IEEE; 2020. p. 21–7. doi:10.1109/phm-besancon49106.2020.00011.
46. Yang Z, Gjorgjevskij D, Long J, Zi Y, Zhang S, Li C. Sparse autoencoder-based multi-head deep neural networks for machinery fault diagnostics with detection of novelties. *Chin J Mech Eng*. 2021;34(1):54. doi:10.1186/s10033-021-00569-0.
47. Carvalho M, Pinho AJ, Brás S. Resampling approaches to handle class imbalance: a review from a data perspective. *J Big DATA*. 2025;12(1):71–128. doi:10.1186/s40537-025-01119-4.
48. Shabrina Assyifa D, Luthfiarta A. SMOTE-tomek re-sampling based on random forest method to overcome unbalanced data for multi-class classification. *Inf J Ilm Bid Teknol Inf Dan Komun*. 2024;9(2):151–60. doi:10.25139/inform.v9i2.8410.