

Article

A Dual-Branch Ensemble Learning Method for Industrial Anomaly Detection: Fusion and Optimization of Scattering and PCA Features

Jing Cai ¹, Zhuo Wu ², Runan Hua ³, Shaohua Mao ^{2,4,5}, Yulun Zhang ⁴, Ran Guo ² and Ke Lin ^{2,5,*} ¹ School of Economics and Management, China University of Geosciences, Wuhan 430074, China² Key Laboratory of Geological Survey and Evaluation of Ministry of Education, China University of Geosciences, Wuhan 430074, China³ Wuhan Second Ship Design and Research Institute, Wuhan 430205, China⁴ Faculty of Engineering, China University of Geosciences, Wuhan 430074, China⁵ School of Mechanical Engineering and Electronic Information, China University of Geosciences, Wuhan 430074, China

* Correspondence: linke@cug.edu.cn

Abstract

Industrial visual anomaly detection remains challenging because practical inspection systems must achieve high detection accuracy while operating under highly imbalanced data, diverse defect patterns, limited computational resources, and increasing demands for interpretability. This work aims to develop a lightweight yet effective and explainable anomaly detection framework for industrial images in settings where a limited number of labeled anomalous samples are available. We propose a dual-branch feature-based supervised ensemble method that integrates complementary representations: a PCA branch to capture linear global structure and a scattering branch to model multi-scale textures. A heterogeneous pool of classical learners (SVM, RF, ET, XGBoost, and LightGBM) is trained on each feature branch, and stable probability outputs are obtained via stratified K-fold out-of-fold training, probability calibration, and a quantile-based threshold search. Decision-level fusion is then performed by stacking, where logistic regression, XGBoost, and LightGBM serve as meta-learners over the out-of-fold probabilities of the selected top-K base learners. Experiments on two public benchmarks (MVTec AD and BTAD) show that the proposed method substantially improves the best PCA-based single model, achieving relative F1_score gains of approximately 31% (MVTec AD) and 26% (BTAD), with maximum AUC values of about 0.91 and 0.96, respectively, under comparable inference complexity. Overall, the results demonstrate that combining high-quality handcrafted features with supervised ensemble fusion provides a practical and interpretable alternative/complement to heavier deep models for resource-constrained industrial anomaly detection, and future work will explore more category-adaptive decision strategies to further enhance robustness on challenging classes.

Academic Editor: Antonio
Fernández-Caballero

Received: 25 December 2025

Revised: 24 January 2026

Accepted: 3 February 2026

Published: 5 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.**Keywords:** industrial visual anomaly detection; scattering transform; principal component analysis (PCA); ensemble learning; stacking

1. Introduction

Modern industry is currently undergoing a critical transition toward intelligent and data-driven industrial inspection systems. With the rapid development of the Internet of Things (IoT) and the Industrial Internet of Things (IIoT), sensors, actuators, and control

systems are being deeply integrated to enable smart manufacturing environments with autonomous monitoring and decision-making capabilities [1–3]. In this context, the volume of heterogeneous data generated by industrial production processes is growing rapidly. Timely detection of equipment faults, process anomalies, and product defects from massive data streams has therefore become a core challenge for quality control and operational safety in modern manufacturing.

Anomaly detection (AD) aims to identify instances that deviate from expected patterns in data that are predominantly composed of normal samples. It has been widely applied in surface defect detection, condition monitoring, intrusion detection, and other scenarios [4,5]. In industrial visual inspection, anomalies often manifest as subtle scratches, contamination, wear, or structural deformation. Such defects typically exhibit complex distributions, limited prior knowledge, and severe class imbalance, which impose stringent requirements on the generalization ability and robustness of detection algorithms.

A variety of technical routes have been proposed for anomaly detection, including statistical modeling, traditional machine learning, and deep-learning-based methods [6–8]. Statistical models, such as Gaussian mixture models and Bayesian anomaly detection, rely on strong prior assumptions about the underlying data distribution, and their performance tends to degrade when the true distribution is complex or heavily contaminated by noise [9]. Traditional machine learning methods, such as support vector machines (SVMs) [10], random forests (RFs) [11], isolation forests (IFs) [12], k-nearest neighbors (k-NNs) [13], and clustering-based approaches [14], generally offer good stability and interpretability on small- to medium-scale datasets with relatively stable structures. To alleviate the limitations of hand-crafted features, PCA-based dimensionality reduction and statistic-based feature selection have gradually become key steps for improving the robustness of traditional machine learning pipelines [15].

In recent years, unsupervised and self-supervised deep learning methods have achieved remarkable progress in industrial visual anomaly detection. Representative approaches include reconstruction-based autoencoder/variational autoencoder (AE/VAE) methods [16], synthetic anomaly generation such as CutPaste [17], normalizing-flow-based models like FastFlow [18], and pretrained feature-based methods such as PaDiM [19] and PatchCore [20]. These methods have reached or even surpassed the performance of supervised models on benchmarks like MVTec AD.

However, deep models often require large training datasets and substantial computational resources, which are not always compatible with industrial environments with limited hardware. Moreover, their decision-making processes can be difficult to interpret, and the deployment and maintenance costs further hinder practical adoption [21–24].

Despite the dominance of deep learning on public benchmarks, industrial visual anomaly detection must simultaneously satisfy three tightly coupled requirements: (i) reliable detection under severely imbalanced and scarce labeled anomalies, (ii) low-latency inference on resource-constrained edge/production-line hardware, and (iii) decision transparency for engineering verification. These constraints make it difficult to directly deploy many high-capacity deep models in real inspection pipelines. Therefore, the scientific problem addressed in this work is how to design a resource-efficient yet accurate anomaly detection approach that remains robust under limited and imbalanced labeled defects while providing interpretable decision evidence suitable for industrial deployment.

To address the above problem, this study aims to develop a lightweight, effective, and interpretable ensemble framework for industrial visual anomaly detection in settings where a limited number of labeled anomalous samples are available. Specifically, we aim to exploit complementary handcrafted representations and model diversity to obtain stable

probability estimates and reliable decision boundaries without incurring the computational cost of end-to-end deep models.

Against this background, ensemble learning has gained increasing attention in industrial anomaly detection due to its ability to improve robustness, mitigate overfitting, and enhance generalization by integrating multiple complementary models [25–28]. Although one-class classification methods are widely used in anomaly detection [29–31], in many industrial scenarios a small amount of defect labels can be obtained during routine inspection and quality auditing. Accordingly, this paper focuses on the supervised setting where labeled anomalous samples are available.

To achieve this goal, we propose a dual-branch supervised ensemble learning framework for industrial visual inspection, leveraging complementary handcrafted features and decision-level stacking to improve detection robustness under limited labeled anomalies. Specifically, the framework combines a PCA branch for global structural/statistical features and a scattering transform branch for multi-scale and multi-orientation texture representations. At the model level, we construct a diverse pool of base learners, including SVMs, RFs, extremely randomized trees (ETs), XGBoost, and LightGBM, and employ logistic regression (LR), XGBoost, and LightGBM as meta-learners to perform decision-level stacking fusion.

Combined with a quantile-based adaptive threshold search and probability calibration, the proposed method achieves superior performance over single models and single-branch features on two industrial anomaly detection benchmarks, MVTec AD [32] and BTAD. The subsequent sections present the method design, experimental setup, and result analysis in detail.

The main contributions of this work can be summarized as follows:

1. We propose a dual-branch supervised ensemble framework for industrial visual anomaly detection, which combines a PCA branch for global structural features and a scattering branch for multi-scale texture representations.
2. We design a heterogeneous multi-model pool with decision-level stacking fusion, where SVMs, RFs, ETs, XGBoost, and LightGBM are used as base learners and LR/XGBoost/LightGBM as meta-learners with calibrated probability inputs.
3. We conduct comprehensive experiments and visual analyses on the MVTec AD and BTAD benchmarks, including per-category AUC comparison, PR/ROC curves, and weight heatmaps, to systematically validate the effectiveness and interpretability of the proposed method.

2. Methods and Implementation

2.1. Overall Framework

To address the challenges of highly imbalanced samples, diverse defect patterns, and heterogeneous feature distributions across products/working conditions in industrial visual anomaly detection, while simultaneously balancing detection performance with computational and deployment costs, this paper adopts a dual-branch ensemble learning framework. The overall architecture is illustrated in Figure 1 and consists of three components: a feature engineering layer, a base-learner training layer, and a meta-learner fusion layer. In the feature engineering layer, two parallel feature extraction branches are constructed:

- (1) PCA-based linear feature branch.

The input RGB image is first converted to grayscale, resized, and standardized. Principal component analysis (PCA) is then applied to retain 95% of the variance (the resulting PCA feature dimensionality is typically around 100–110), followed by F-statistics-based feature selection to remove redundant or weakly discriminative dimensions, since PCA components with high explained variance are not necessarily the most discriminative for anomaly detection. Finally, a robust scaling (RobustScaler)

step is adopted to alleviate the influence of outliers and obtain stable low-dimensional structural features.

(2) Scattering-based nonlinear feature branch.

Based on the Kymatio framework, a 2D Scattering transform is computed to extract multi-scale, multi-orientation wavelet modulus coefficients, which are subsequently aggregated by global pooling, log-transformed, and standardized. The scattering branch provides high-dimensional nonlinear features that are sensitive to local texture, fine-grained structures, and small defects. The dual-branch design leverages PCA to capture global structure, while the scattering transform focuses on detailed textures, enabling complementary modeling of the feature space.

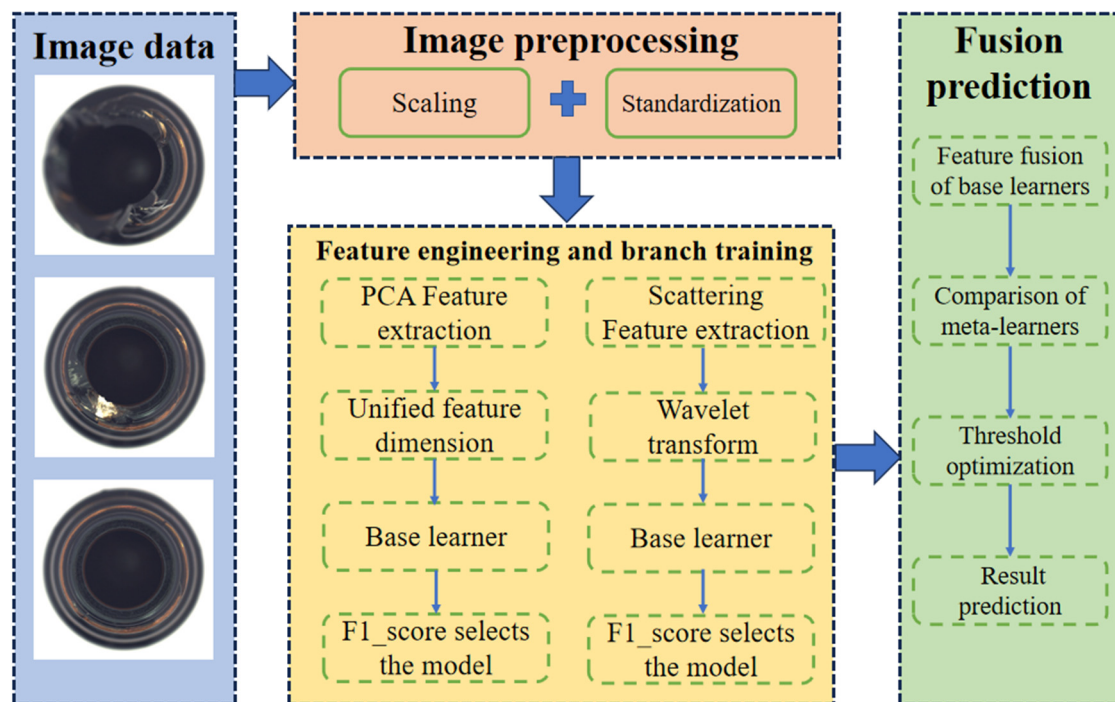


Figure 1. Overall architecture of the proposed dual-branch ensemble system.

In the base-learner training layer, independent pools of base models are constructed on the two feature branches, including support vector machines (SVMs), random forests (RFs), extremely randomized trees (ETs), XGBoost (XGB), and LightGBM (LGBM). For each branch, a stratified K-fold out-of-fold (OOF) training scheme is employed to obtain probability predictions on validation folds. The same data splits and evaluation metrics are used across different models to ensure a fair comparison under a unified validation protocol. Each base model is further tuned by grid search (GridSearchCV) to obtain better performance.

To enhance the effectiveness of the fusion stage, the system automatically selects the top-performing base models on each branch according to their F1_score on the validation set. Concretely, the top-K base learners ($K = 2$ by default) from each branch are retained, and their calibrated probability predictions are used as inputs to the stacking fusion. This selection mechanism adaptively filters out overfitting or weak models and significantly improves the quality of the inputs to the meta-learner.

The base learners in our ensemble are heterogeneous (tree-based models, margin-based classifiers, and boosting methods). Retaining more base learners per branch (larger K) increases the dimensionality of the meta-feature space, which can introduce redundancy and raise the risk of overfitting when anomalous samples are limited. Moreover, empirical

gains typically saturate as K grows, while computational cost and model size increase. Therefore, to balance model diversity, generalization stability, and deployment efficiency in resource-constrained industrial settings, we set $K = 2$ in this study.

Finally, in the meta-learner fusion layer, the validation probabilities of the selected top- K base learners from both branches are concatenated into second-level features and fed into three types of meta-learners—logistic regression (LR), XGBoost, and LightGBM—to perform decision-level stacking. In this way, the complementary relationships among linear models and tree ensembles can be exploited to learn optimal combination weights and improve the overall performance. The fusion stage still adopts a quantile-based adaptive threshold search strategy to determine the optimal decision boundary, thereby ensuring stable and interpretable outputs.

For reproducibility (without repeating the above descriptions), we provide a concise step-by-step pipeline:

1. Split: Stratified train/validation/test.
2. Preprocess: Grayscale, resize, normalize.
3. PCA branch: Fit StandardScaler on the training set, fit PCA to retain 95% variance (≈ 100 – 110 dims), apply to validation/test, and then perform F-test feature selection and RobustScaler.
4. Scattering branch: Extract scattering features; pool/transform; standardize (train-only).
5. Base learners: Train {SVM, RF, ET, XGB, LGBM} on each branch.
6. OOF & calibration: Stratified K -fold OOF probabilities (with calibration) for stacking features.
7. Top- K : Keep top- K base learners per branch by validation F1 ($K = 2$).
8. Stacking: Train meta-learner {LR/XGB/LGBM} on OOF meta-features.
9. Threshold & test: Choose threshold on validation; report test.

All fitted steps (PCA/selection/scaling, calibration, stacking, and thresholding) are learned on training data only (or training folds for OOF) and then applied to validation/test.

2.2. Design of Base Learners and Meta-Learners

At the base-learner level, this work constructs, for both the PCA and scattering branches, a heterogeneous pool of models consisting of SVM, RF, ET, XGBoost (XGB), and LightGBM (LGBM). All base learners are trained with grid-search-based hyperparameter optimization (GridSearchCV), and their outputs are wrapped by CalibratedClassifierCV to obtain well-calibrated probabilities that are suitable for the subsequent probability-based stacking fusion.

- (1) Support vector machine (SVM).

SVM emphasizes margin maximization in high-dimensional spaces and is effective when the decision boundary is nonlinear but data are limited. Class imbalance is handled via class weighting, and probabilities are obtained via calibration.

- (2) Random forest (RF).

RF is a bagging-based ensemble that reduces variance and improves robustness to noise and feature redundancy. It is particularly stable for high-dimensional tabular representations derived from images.

- (3) Extremely randomized trees (ETs).

ETs further increase diversity by introducing stronger randomness in feature/threshold selection. Compared with RFs, ETs often yield lower variance and complementary decision patterns, making them suitable for noisy industrial representations.

(4) XGBoost (XGB).

XGB is a representative gradient-boosting decision tree model that supports regularization and efficient training. It can model complex nonlinear feature interactions and often provides strong ranking performance for anomaly scoring.

(5) LightGBM (LGBM).

LGBM improves efficiency via histogram-based splitting and sampling/bundling strategies, supporting scalable learning on high-dimensional features and achieving high accuracy under limited computing budgets.

According to the diversity theory of ensemble learning, an effective ensemble requires a proper balance between the accuracy and diversity of individual learners. Wood et al. [33] and subsequent studies have further shown that, under the premise of maintaining sufficient accuracy for each model, introducing learners with significantly different structures and low error correlation is crucial for reducing the overall generalization error. Following this principle, SVM, RF, ET, XGB, and LGBM are selected as base learners mainly due to the complementarity of their hypothesis spaces and training mechanisms. SVM emphasizes margin maximization in high-dimensional feature spaces; RF and ET focus on bagging-based variance reduction and robust tree ensembles [34]; XGB and LGBM exploit gradient boosting to model complex nonlinear interactions [35,36]. The combination of these models enriches the diversity of decision boundaries and error patterns and thus leads to more stable overall performance and better generalization when integrated via stacking [37,38].

2.3. Evaluation Metrics

Since this work adopts a supervised learning paradigm for image-level anomaly detection, the evaluation metrics need to reflect both the discriminative capability under class imbalance and the practical detection effectiveness. We treat “anomalous” samples as the positive class. Let TP, FP, TN, and FN denote the numbers of true positives, false positives, true negatives, and false negatives, respectively. Then, the common metrics are defined as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (1)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (3)$$

$$\text{F1_score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

In highly imbalanced anomaly detection tasks, the F1_score, as the harmonic mean of precision and recall, serves as the primary metric to measure the trade-off between detection accuracy and coverage. In our comparisons of different models and configurations, the F1_score is used as the main evaluation criterion, while accuracy is reported as a supplementary reference for overall classification correctness.

To comprehensively evaluate the discriminative ability of the classifiers across different thresholds, we also adopt the area under the receiver operating characteristic curve (AUC) as a threshold-independent metric. Let S^+ and S^- denote the sets of positive (anomalous) and negative (normal) samples, respectively, and let $f(\cdot)$ be the anomaly score output by a model. Then, AUC can be written in the probabilistic form of ranking capability:

$$\text{AUC} = \frac{1}{|P| \cdot |N|} \sum_{x_p \in P} \sum_{x_n \in N} \mathbb{I}(s(x_p) > s(x_n)) \quad (5)$$

where $\mathbb{I}(\cdot)$ is the indicator function, which equals 1 if the condition in parentheses holds and 0 otherwise. This definition shows that AUC is equivalent to the probability that a randomly drawn positive sample receives a higher score than a randomly drawn negative sample, thereby characterizing the global ranking ability of the classifier over the entire threshold range. For scenarios with even more severe class imbalance, we additionally employ the precision–recall (PR) curve and its corresponding average precision (AP) as complementary evaluation measures.

Moreover, to quantify the “alarm strength” of the model and the reasonableness of threshold selection under a fixed decision threshold, we introduce the predicted anomaly rate as an auxiliary statistic. Let N_{pred}^+ be the number of samples predicted as anomalous under a given threshold and let N be the total number of test samples. The predicted anomaly rate is defined as follows:

$$\hat{r}_{\text{anom}} = \frac{N_{\text{pred}, \text{pos}}}{N_{\text{all}}} \quad (6)$$

Correspondingly, the true anomaly rate on the test set is the following:

$$r_{\text{anom}} = \frac{N_{\text{pos}}}{N_{\text{all}}} \quad (7)$$

where N_{true}^+ denotes the number of truly anomalous samples. If, while maintaining a high F1_score, r_{pred} is close to r_{true} , this indicates that the model’s probability outputs and the chosen threshold are reasonably calibrated at the global level; thus, the model is unlikely to produce excessively high false-alarm or miss rates in practice.

In summary, the F1_score is mainly used to assess the actual detection performance at the selected threshold, accuracy provides a reference for overall correctness, AUC and AP characterize the global discriminative ability over all thresholds, and the predicted anomaly rate reflects the deployability of the model in industrial scenarios from the perspective of overall alarm proportion. The subsequent experiments will systematically compare different feature branches, base learners, and stacking configurations based on the above metrics.

3. Experiments

3.1. Dataset Description

In this study, two public industrial anomaly detection datasets, MVTec AD and BTAD, are used to validate the proposed method.

The MVTec Anomaly Detection (MVTec AD) dataset is a standard benchmark in the field of industrial visual anomaly detection. It contains 5354 high-resolution images from 15 industrial product categories (such as bottle, cable, etc.), covering more than 70 types of defects and providing pixel-level ground-truth masks. As mentioned earlier, MVTec AD consists of 4096 normal training images and 1258 anomalous test images, with an anomaly rate of 23.5% in the test set. The products span multiple industrial domains, including textiles, metal parts, pharmaceuticals, and electronic components, while the anomalies involve various defect patterns such as scratches, breaks, contamination, and deformations. Thanks to its rich product types, diverse defect modes, and precise annotations, MVTec AD has become one of the most widely used benchmarks for evaluating both image-level anomaly detection and pixel-level defect localization.

The Bottleneck Anomaly Detection Dataset (BTAD) proposed by Mishra et al. is a real-world industrial dataset that contains three product categories (usually referred to as product 1–3), with a total of 2830 RGB images. It is designed for evaluating both image-level anomaly detection and pixel-level defect segmentation. Similar to MVTec AD, BTAD includes both structural defects and surface defects, such as local breakage, indentations,

scratches, contamination, and geometric deformations. However, the three products in BTAD differ significantly in image resolution and imaging conditions, which further increases the diversity of the dataset. Each anomalous image in BTAD is equipped with a pixel-level defect mask, enabling researchers to assess detection accuracy and defect localization performance under a unified framework. Compared with the multi-category MVTec AD dataset, BTAD focuses more on high-resolution industrial products collected from real production lines with fewer categories, thus posing higher robustness requirements for algorithms under the conditions of small category numbers but diverse defect patterns.

By conducting experiments on both MVTec AD and BTAD, this work is able to comprehensively evaluate the effectiveness of the proposed method from multiple perspectives, including dataset scale, product types, and defect pattern diversity.

3.2. Experimental Setup

Unless otherwise specified, all experiments in this paper adopt a 6:2:2 stratified random split of the data, where 60%, 20%, and 20% of all samples are assigned to the training, validation, and test sets, respectively. Stratified sampling is employed to ensure that the proportion of anomalous samples is approximately consistent across the three subsets, thereby guaranteeing the fairness and comparability of model evaluation. All models under comparison share the same preprocessing, feature extraction, and training procedures.

For threshold selection, a quantile-based search strategy is used. Concretely, in the interval $[0.60, 0.90]$, candidate decision thresholds are generated with a step size of 0.01 according to the empirical quantiles of the predicted probabilities, and the threshold that maximizes the F1_score on the validation set is chosen as the optimal decision threshold for that model. In special cases where the probability distribution is overly concentrated, the search degenerates to a uniform grid over $[0, 1]$ to improve the robustness of threshold optimization. During testing, all models use the corresponding optimal thresholds determined on the validation set, ensuring a fair comparison under a unified decision criterion.

All experiments are conducted in a Python 3.12 environment. The scikit-learn 1.3+ library is used to implement the machine learning algorithms, while numpy 1.24+ and pandas 2.0+ are used for data processing and analysis. The hardware platform consists of an Intel Core i7 processor and 16 GB of RAM running the Windows operating system. All implementations are based on open-source libraries, which facilitates the reproducibility of the experimental results.

4. Results and Discussion

4.1. Overall Results on MVTec AD

On the MVTec AD dataset, we first compare the performance of single models built on the linear PCA feature branch and the nonlinear scattering feature branch and then examine the stacking ensembles with different meta-learners. For clarity, Table 1 reports, for each of the PCA and scattering branches, the three single models with the highest F1_score, as well as the stacking results obtained with logistic regression, LightGBM, and XGBoost as meta-learners.

From Table 1, it can be observed that models based on scattering features are overall clearly superior to those based on PCA features. On the PCA branch, the best-performing model is PCA-ET, with an F1_score of 0.540 and an AUC of 0.780. On the scattering branch, SCAT-LGBM increases the F1_score to 0.703 and the AUC to 0.907, while SCAT-XGB and SCAT-RF also reach F1_scores of 0.682 and 0.673, respectively. This indicates that, on MVTec AD, which is dominated by complex textures and diverse appearance defects, nonlinear scattering features are more suitable than linear PCA features as supervised anomaly detection representations.

Table 1. Performance comparison on the MVTec AD dataset.

Models	Accuracy	Precision	Recall	F1_Score	AUC	Predicted Anomaly Rate
PCA-ET	0.750	0.474	0.625	0.540	0.780	0.309
PCA-XGB	0.702	0.419	0.697	0.523	0.783	0.390
PCA-LGBM	0.694	0.409	0.693	0.515	0.784	0.397
SCAT-LGBM	0.868	0.746	0.665	0.703	0.907	0.209
SCAT-XGB	0.857	0.713	0.653	0.682	0.902	0.215
SCAT-RF	0.845	0.665	0.681	0.673	0.887	0.240
Stacking-LR	0.865	0.714	0.705	0.709	0.911	0.232
Stacking-LGBM	0.867	0.774	0.614	0.684	0.886	0.186
Stacking-XGB	0.862	0.709	0.697	0.703	0.894	0.231

On this basis, the stacking models with three meta-learners further improve the overall performance. Using logistic regression as the meta-learner, Stacking-LR achieves an F1_score of 0.709 and an AUC of 0.911 on MVTec AD. While maintaining a high AUC, it still brings a small but consistent improvement compared with the best single model SCAT-LGBM (F1 = 0.703, AUC = 0.907). Stacking-XGB reaches an F1_score of 0.703, close to SCAT-LGBM, with a slightly higher recall, whereas Stacking-LGBM attains an F1_score of 0.684 under the lowest predicted anomaly rate (0.186), making it more suitable for scenarios that are highly sensitive to false alarms. Overall, the combination of dual-branch features, a multi-model pool, and stacking brings about a relative F1_score improvement of approximately 31% on MVTec AD compared with the best PCA single-branch model (0.540 $\rightarrow$ 0.709).

4.2. Overall Results on BTAD

As shown in Table 2, the overall trend on BTAD is consistent with that on MVTec AD: the scattering branch significantly outperforms the PCA branch, and stacking on top of the scattering branch further improves performance. On the PCA branch, PCA-ET achieves an F1_score of 0.606, which is the best result within that branch. On the scattering branch, SCAT-LGBM raises the F1_score to 0.736 and the AUC to 0.962, while SCAT-XGB and SCAT-RF achieve F1_scores of 0.729 and 0.716, respectively. This shows that even on the BTAD dataset, which has fewer categories but is closer to real production-line data, scattering features still demonstrate clear advantages in representing textures and local defects.

Table 2. Performance comparison on the BTAD dataset.

Models	Accuracy	Precision	Recall	F1_Score	AUC	Predicted Anomaly Rate
PCA-ET	0.915	0.647	0.569	0.606	0.926	0.100
PCA-SVM	0.919	0.698	0.517	0.594	0.918	0.085
PCA-LGBM	0.923	0.771	0.466	0.581	0.922	0.069
SCAT-LGBM	0.945	0.813	0.672	0.736	0.962	0.094
SCAT-XGB	0.943	0.796	0.672	0.729	0.953	0.096
SCAT-RF	0.939	0.765	0.672	0.716	0.956	0.100
Stacking-LR	0.949	0.808	0.724	0.764	0.959	0.102
Stacking-LGBM	0.941	0.938	0.517	0.667	0.961	0.063
Stacking-XGB	0.937	0.783	0.621	0.692	0.960	0.091

At the ensemble level, using logistic regression as the meta-learner, Stacking-LR obtains an F1_score of 0.764 and an AUC of 0.959 on BTAD. Compared with the best single model SCAT-LGBM (F1 = 0.736, AUC = 0.962), this corresponds to an absolute F1 improvement of about 2.8 percentage points (a relative improvement of about 3.8%), while maintaining a comparable AUC.

Since the meta-features are condensed representations derived from the base learners, the resulting meta-feature space tends to be close to linearly separable, especially after probability calibration. From a bias–variance perspective, a linear meta-learner such as logistic regression provides a favorable trade-off under limited anomalous samples: it can learn a stable weighted combination of heterogeneous predictors with low variance and reduced risk of meta-level overfitting. This also leads to more reliable probability outputs and, consequently, a more robust validation-based threshold selection. In contrast, boosting-based meta-learners introduce higher functional complexity at the stacking stage and may fit spurious patterns in the meta-features, which is more likely when anomaly labels are scarce. These considerations are consistent with our observations that logistic regression yields more stable performance than XGB and LGBM as the meta-learner on both MVTec AD and BTAD.

4.3. Comparison with Existing Methods on MVTec AD

As shown in Table 3, on the MVTec AD dataset, the proposed method achieves competitive or even clearly superior AUC performance across all 15 subcategories. Taking the arithmetic mean of the per-category AUC as an example, the proposed ensemble model reaches 0.889, which is higher than VT-ADL [39]. Meanwhile, compared with classical methods such as 1-NN, AE-MSE, VAE, and AE-SSIM, the average AUC improvements of the proposed method are 0.249, 0.098, 0.250, and 0.194, respectively. Table 3 reports per-category rank and the average rank across the 15 categories; the proposed method achieves the best Avg. rank = 1.733 with 9/15 category-wise wins. These results verify that, under the same data and evaluation protocol, supervised ensemble learning can significantly improve the discriminative ability of image-level anomaly detection.

Table 3. AUC of different methods on each category of MVTec AD (taken from [40]).

Category	1-NN	AE-MSE	VAE	AE-SSIM	VT-ADL	Ours
Carpet	0.512	0.456	0.501	0.647	0.773	0.908
Grid	0.228	0.582	0.224	0.849	0.871	0.512
Leather	0.446	0.819	0.635	0.561	0.728	0.999
Tile	0.822	0.897	0.870	0.175	0.796	0.876
Wood	0.502	0.727	0.628	0.605	0.781	0.992
Bottle	0.898	0.910	0.897	0.834	0.949	0.995
Cable	0.806	0.825	0.654	0.478	0.776	0.747
Capsule	0.631	0.862	0.526	0.860	0.672	0.943
Hazelnut	0.861	0.917	0.878	0.916	0.698	0.951
Metal Nut	0.705	0.830	0.576	0.603	0.320	0.820
Pill	0.725	0.893	0.769	0.830	0.705	0.845
Screw	0.604	0.754	0.559	0.887	0.928	0.819
Toothbrush	0.675	0.822	0.693	0.784	0.749	1.000
Transistor	0.680	0.728	0.626	0.725	0.549	0.952
Zipper	0.512	0.839	0.549	0.665	0.808	0.970
Means	0.640	0.790	0.639	0.694	0.807	0.889
Avg. rank	4.667	2.333	4.800	3.867	3.600	1.733
#Wins (rank = 1)	0	4	0	0	2	9

From the per-category results, the proposed method obtains the highest AUC on 9 out of the 15 categories, including carpet, leather, wood, bottle, capsule, hazelnut, toothbrush, transistor, and zipper. In particular, on categories such as leather, wood, bottle, and toothbrush, where the texture structure is relatively regular, and defect patterns are relatively clear, the AUC is close to or even reaches 1.0. It should be noted that in a few categories, such as grid, cable, and screw, the AUC of the proposed method is slightly lower than

that of VT-ADL. Visual analysis on these categories shows that their normal samples often exhibit strongly periodic geometric textures or complex background structures, and some minor defects are easily confused with normal texture patterns at the base-learner level, leading to a large overlap between the anomaly score distributions in the meta-features. In addition, the current work adopts a globally unified calibrated threshold for decision making, which may not be optimal for scenarios with large inter-class distribution differences. This is a potential reason why the performance on a few “difficult categories” is slightly worse. In future work, category-adaptive threshold learning or cost-sensitive meta-learning strategies could be introduced to further improve robustness in these categories.

4.4. PR/ROC Curves and Weight Analysis on MVTec AD and BTAD

In addition to ROC-AUC, we report the area under the precision–recall (PR) curve using average precision (AP), which is more informative under severe class imbalance. Figure 2 summarizes PR and ROC comparisons on the MVTec AD validation set, with zoomed views highlighting the high-precision and low-FPR operating regions. Overall, Stacking-LR achieves the best ranking and discrimination performance, yielding the highest AP (0.817) and AUC (0.911). The strongest single-learner SCAT-LGBM remains highly competitive (AP = 0.812, AUC = 0.907), followed by SCAT-XGB (AP = 0.804, AUC = 0.902), while SCAT-RF and the stacking variants with boosting meta-learners show lower overall performance (e.g., AUC = 0.894/0.886 for meta-XGB/meta-LGBM). Importantly, the zoomed regions indicate that several top methods exhibit very close behavior under strict false-alarm constraints, whereas Stacking-LR consistently stays among the leading curves and tends to provide a slightly better recall/TPR at comparable precision/FPR, suggesting a more stable fusion rule than high-capacity boosting meta-learners on this dataset.

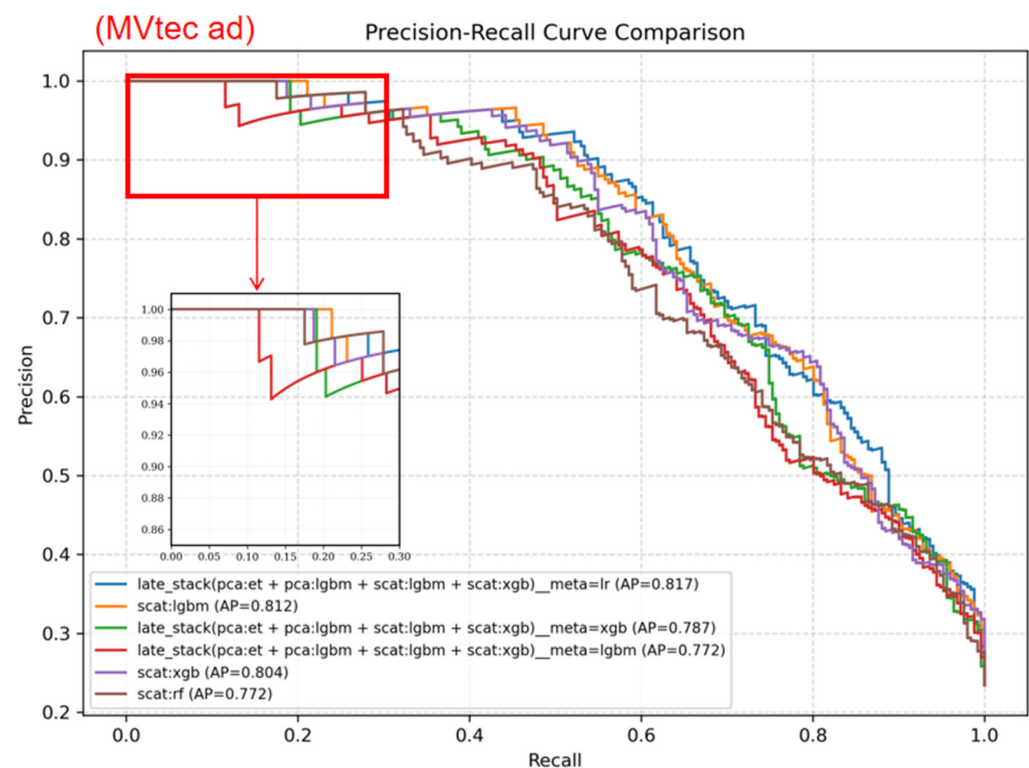


Figure 2. Cont.

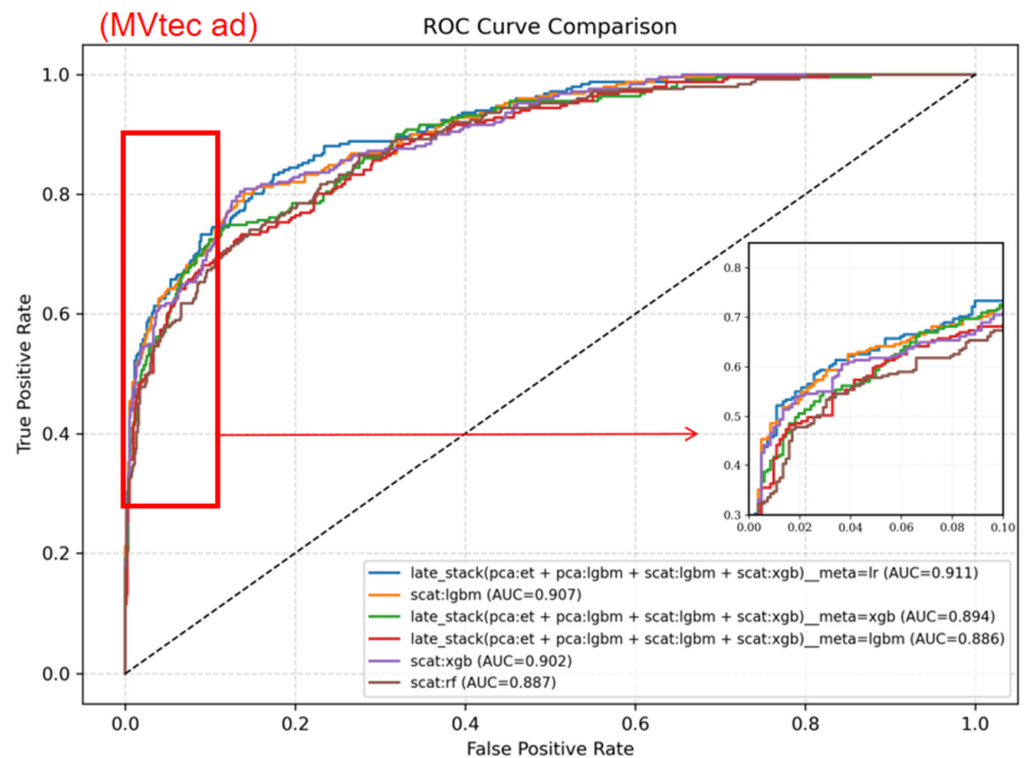


Figure 2. PR/ROC curve comparison on the MVTec AD validation set.

In Figure 3 (BTAD validation set), the PR and ROC curves of the top methods are closely clustered, indicating that the performance gap among strong learners is relatively small on this dataset. On the PR curves, SCAT-LGBM achieves the best ranking performance with the highest AP (0.845), while Stacking-LR remains competitive (AP = 0.838) and is comparable to other strong scattering-based models (e.g., SCAT-ET, AP = 0.835). The ROC curves show a similar trend: the best AUC is obtained by SCAT-LGBM (0.962), followed by SCAT-ET (0.961) and stacking variants (AUC = 0.959–0.960). Overall, these results suggest that BTAD can be well handled by a single strong scattering-based classifier, whereas stacking provides a stable and consistently high-performing alternative, with improvements that are marginal in terms of AP/AUC due to the already saturated discrimination on this benchmark. This is likely because the strongest base learner already offers near-saturated separability on BTAD, leaving limited room for further gains from late fusion.

Figure 4 reports the normalized weights assigned to base learners by different meta-learners. On MVTec AD, the fusion consistently places larger weights on scattering-based predictors than on PCA-based ones. In particular, the LR meta-learner allocates most of the mass to the two scattering models (scat:lgbm = 0.38, scat:xgb = 0.38), while the PCA branch receives smaller contributions (pca:et = 0.11, pca:lgbm = 0.13). This “two-strong-model” weighting pattern suggests that LR is exploiting complementary information from multiple scattering predictors rather than relying on a single expert. In contrast, the XGB meta-learner is more selective and strongly emphasizes scat:lgbm (0.52), with a secondary weight on scat:xgb (0.27), indicating a sharper preference among base learners.

On BTAD, the weight distributions are generally less extreme under LR and LGBM, where the four base models receive comparable shares (e.g., LR: 0.23/0.19/0.29/0.29; LGBM: 0.26/0.26/0.29/0.19). Under XGB, the weights become more selective, but the concentration is not stronger than the MVTec case: scat:et (0.47) is emphasized, while scat:xgb (0.29) remains a stable contributor and the PCA models are downweighted (0.14 and 0.10). Combined with the curve-based results, these observations are consistent with two points: (i) scattering features provide the dominant predictive power on both

datasets; (ii) the best-performing stacking configuration on MVTec (Stacking-LR in Figure 2) corresponds to a weight pattern that aggregates multiple strong scattering predictors, whereas on BTAD (Figure 3) the top methods are already tightly clustered in AP/AUC, and stacking mainly acts as a robust combiner rather than producing a clear margin over the best single model.

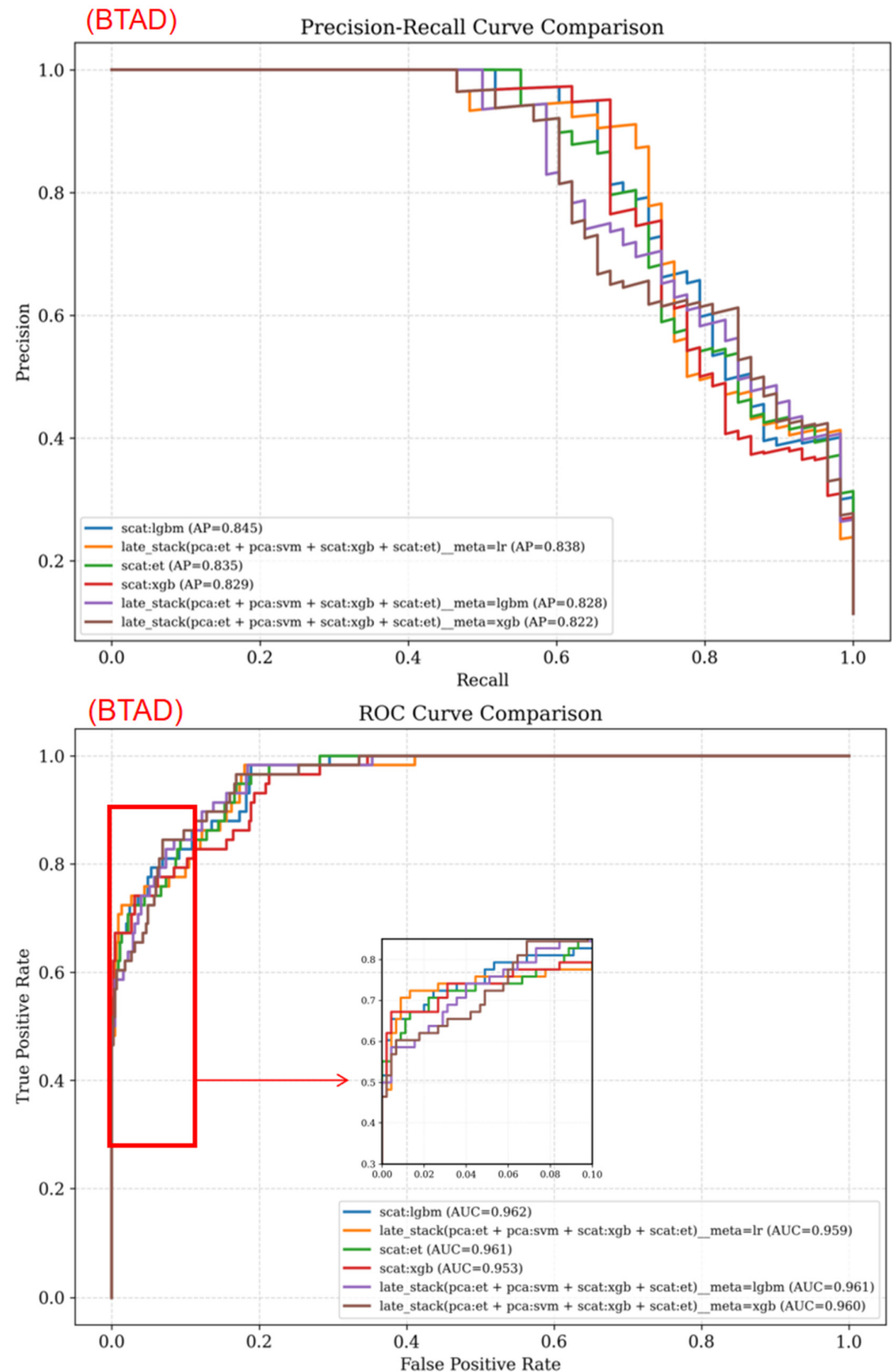


Figure 3. PR/ROC curve comparison on the BTAD validation set.

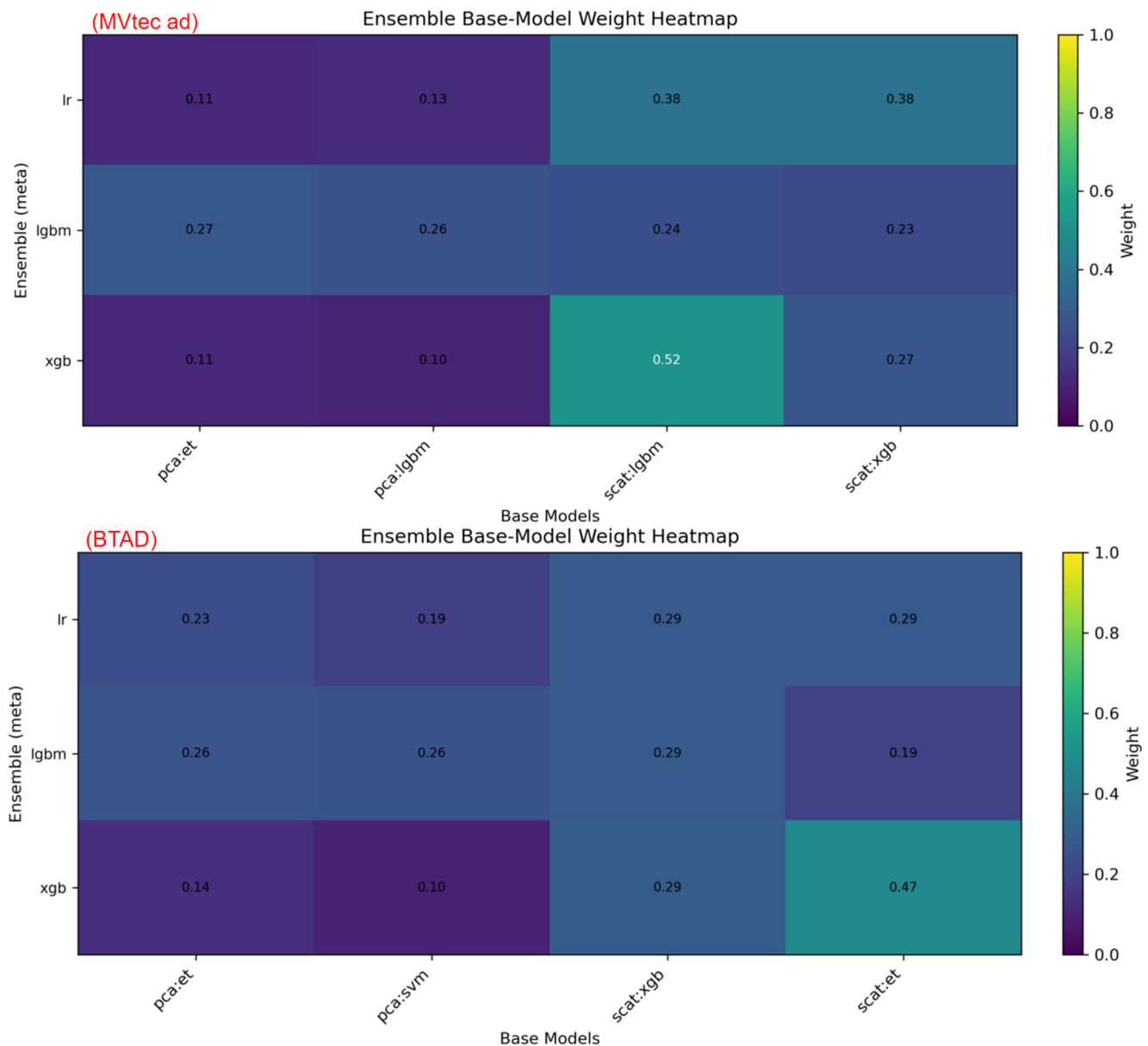


Figure 4. Heatmap of base-learner weights on the MVtec AD and BTAD validation set.

4.5. Theoretical Interpretation and Practical Implications

This subsection provides a unified interpretation of our empirical findings from the perspectives of representation complementarity, small-sample learning, and deployable decision fusion.

These results suggest that, in texture-dominated industrial categories (e.g., scratches, wear, dents), nonlinear and multi-scale scattering features contribute more critically to decision making than PCA features. This is theoretically plausible because scattering representations encode multi-scale and multi-orientation wavelet responses and are effective at capturing local texture and fine-grained structural variations, which often constitute the primary evidence of defects in surface inspection. This interpretation is consistent with Mallat's scattering theory, which shows that scattering representations are locally translation-invariant and provably stable to small deformations. Such stability is particularly relevant for industrial surface inspection, where defects often appear as localized texture perturbations under nuisance variations (e.g., slight misalignment or illumination changes) [39,41]. In contrast, PCA is a linear variance-preserving projection that mainly

emphasizes global directions of maximum variation; such directions are not necessarily the most discriminative for subtle, localized texture anomalies. Nevertheless, PCA can still provide complementary global cues that improve the stability of the ensemble when combined with texture-sensitive scattering features.

From a broader perspective, feature-level complementarity is closely related to the small-sample and class-imbalance regime commonly encountered in industrial anomaly detection. When anomalous samples are scarce, learning highly complex representations may increase estimation uncertainty and sensitivity to noise. Decomposing representation learning into complementary linear (PCA-based global structure) and nonlinear (scattering-based textures) components provides a structured way to control effective model complexity while preserving discriminative information, which helps stabilize learning under limited anomalies without relying on end-to-end deep feature learning.

Alternative analytic late fusion strategies, such as copula-based fusion or nonlinear α -integration, have been shown to achieve theoretical optimality under criteria such as least-mean-square error (LMSE) or minimum probability of error (MPE) when reliable joint probability estimation is feasible [42]. However, these approaches typically rely on sufficiently large sample sizes and relatively stable distributional assumptions. In industrial anomaly detection, where anomalous samples are scarce and defect patterns are highly heterogeneous, such assumptions are often difficult to satisfy in practice. From a sample-size perspective, it is also important to ask how many labeled anomalies are required to reach a target probability of error under severe class imbalance. This question is fundamentally linked to the (unknown) Bayes risk and therefore remains difficult to answer reliably when defect modes are heterogeneous. Recent work has proposed proxy learning curves for the Bayes classifier to extrapolate performance from finite-sample observations and thereby approximate the sample size needed for a desired error level. While a rigorous sample-size estimation is beyond the scope of this study, this line of analysis offers a principled direction for future deployment planning and data-collection budgeting [43,44]. In contrast, stacking formulates decision-level fusion as a supervised learning problem on calibrated probability outputs, allowing combination weights and error correlations among base learners to be learned directly from data; this data-driven property makes stacking more robust and deployable under the small-sample and imbalanced conditions considered in this work. In addition, the proposed dual-branch decomposition yields favorable computational characteristics: PCA requires a one-time eigen-decomposition during training and only a linear projection during inference, while the scattering transform relies on predefined wavelet filters without trainable parameters, resulting in a fixed and deterministic feature extraction cost. The stacking stage further operates on low-dimensional calibrated probabilities, introducing negligible additional overhead. Overall, the proposed framework provides a practical trade-off among detection performance, interpretability, and computational cost, making it well suited for resource-constrained industrial inspection scenarios.

5. Conclusions

Industrial visual anomaly detection remains challenging in practice because inspection systems must achieve reliable detection under severely imbalanced and scarce defect samples, heterogeneous defect patterns, limited computing resources, and increasing requirements for interpretability and deployability. To address these constraints, this work proposes a dual-branch supervised ensemble framework that combines complementary handcrafted representations (PCA-based global structural cues and scattering-based multi-scale texture descriptors) and integrates a heterogeneous pool of classical learners through probability-calibrated stacking.

The main scientific advancement of this study is the demonstration that supervised ensemble learning equipped with complementary handcrafted features can provide a reproducible, interpretable, and resource-efficient alternative (or complement) to heavy deep anomaly detection models in industrial settings. Instead of relying on end-to-end deep feature learning, our framework decomposes representation learning into a linear global branch and a nonlinear texture branch and performs late fusion on calibrated probability outputs. This design yields a robust decision mechanism under limited anomalies and class imbalance, while keeping inference lightweight and operationally transparent. In this sense, our approach also differs from and complements representative baselines frequently used in anomaly detection, such as nearest-neighbor methods (e.g., 1-NN/k-NN) and reconstruction-based models (e.g., AE/VAE), as well as recent deep baselines such as VT-ADL: rather than modeling normality or relying on heavy end-to-end training, we explicitly leverage supervised defect cues when a small number of anomaly labels is available.

Experiments on MVTec AD and BTAD show clear improvements over PCA-only supervised baselines under comparable inference complexity (e.g., the best PCA-based single model improves from $F1 = 0.540$ to 0.709 on MVTec AD and from $F1 = 0.606$ to 0.764 on BTAD). Moreover, the per-category AUC comparisons against literature-reported methods further confirm that the proposed fusion is competitive across categories, while remaining lightweight for deployment. Beyond aggregate metrics, per-category AUC comparisons, PR/ROC analyses, and ensemble-weight visualizations consistently indicate that the scattering branch provides dominant discriminative evidence for texture-heavy defects, while the PCA branch supplies complementary global cues that stabilize fusion. The results also suggest that, when meta-features are low-dimensional calibrated probabilities, logistic regression can serve as a particularly stable meta-learner, likely due to a favorable bias–variance trade-off at the stacking level.

This manuscript is intended to be directly reusable: researchers and practitioners can replicate the pipeline with standard machine-learning toolkits, replace individual modules (feature branches, base learners, calibration strategies, meta-learners), and deploy the method in production-line environments where GPUs or large-scale defect annotations are unavailable. From a practical perspective, the proposed framework can be integrated into resource-constrained industrial inspection systems, providing a favorable trade-off among detection performance, computational cost, and interpretability, which can further reduce rework and waste and improve operational safety.

This study focuses on supervised image-level anomaly detection and requires at least a limited number of labeled anomalies. Future work will explore self/semi-supervised extensions, more category-adaptive decision strategies (e.g., adaptive thresholding), and anomaly localization to further enhance robustness on challenging classes and broaden applicability across varying products and working conditions.

Author Contributions: Conceptualization, K.L.; methodology, K.L., J.C. and Z.W.; software, J.C. and Z.W.; validation, Y.Z., R.H. and R.G.; formal analysis, K.L.; investigation, J.C.; resources, K.L.; data curation, Y.Z.; writing—original draft preparation, J.C.; writing—review and editing, Z.W. and K.L.; visualization, R.G.; supervision, R.H.; project administration, S.M.; funding acquisition, S.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China under Grant Nos. 52576145 and 52406162, the Natural Science Foundation project of Hubei Province under Grant No. 2025AFB913, the Postdoctoral Fellowship Program of CPSF under Grant Number GZC20241593, the China Postdoctoral Science Foundation under Grant Number 2024M753029, the Postdoctoral Project of Hubei Province under Grant Number 2024HBBHCXA090, “CUG Scholar” Scientific Research Funds at China University of Geosciences (Wuhan) (Project No. 2023097), and the

Fundamental Research Funds for the Central University, China University of Geosciences (Wuhan) under Grant No. 162301242612. The authors deeply appreciate all of the support.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The MVTec AD dataset and the BTAD dataset are publicly available through their official project pages. The code, scripts, and configuration files used in this study will be made available from the corresponding author upon reasonable request. (MVTec AD: <https://www.mvtec.com/company/research/datasets/mvtec-ad/downloads>, accessed on 26 October 2025. BTAD: <https://github.com/pankajmishra000/VT-ADL>, accessed on 26 October 2025).

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

References

1. Aboah Boateng, E.; Bruce, J.W. Unsupervised machine learning techniques for detecting PLC process control anomalies. *J. Cybersecur. Priv.* **2022**, *2*, 220–244. [\[CrossRef\]](#)
2. Boateng, E.A.; Bruce, J.W.; Talbert, D.A. Anomaly detection for a water treatment system based on one-class neural network. *IEEE Access* **2022**, *10*, 115179–115191. [\[CrossRef\]](#)
3. Kagermann, H.; Wahlster, W. Ten years of Industrie 4.0. *Science* **2022**, *4*, 26. [\[CrossRef\]](#)
4. Jeffrey, N.; Tan, Q.; Villar, J.R. Using ensemble learning for anomaly detection in cyber–physical systems. *Electronics* **2024**, *13*, 1391. [\[CrossRef\]](#)
5. Lin, Y.; Chang, Y.; Tong, X.; Yu, J.; Liotta, A.; Huang, G.; Song, W.; Zeng, D.; Wu, Z.; Wang, Y. A survey on RGB, 3D, and multimodal approaches for unsupervised industrial image anomaly detection. *Inf. Fusion* **2025**, *121*, 103139. [\[CrossRef\]](#)
6. Nassif, A.B.; Talib, M.A.; Nasir, Q.; Dakalbab, F.M. Machine learning for anomaly detection: A systematic review. *IEEE Access* **2021**, *9*, 78658–78700. [\[CrossRef\]](#)
7. Pang, G.; Shen, C.; Cao, L.; Hengel, A.V.D. Deep learning for anomaly detection: A review. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–38. [\[CrossRef\]](#)
8. Weston, D.J.; Hand, D.J.; Adams, N.M.; Whitrow, C.; Juszczak, P. Plastic card fraud detection using peer group analysis. *Adv. Data Anal. Classif.* **2008**, *2*, 45–62. [\[CrossRef\]](#)
9. Samariya, D.; Thakkar, A. A comprehensive survey of anomaly detection algorithms. *Ann. Data Sci.* **2023**, *10*, 829–850. [\[CrossRef\]](#)
10. Hosseinzadeh, M.; Rahmani, A.M.; Vo, B.; Bidaki, M.; Masdari, M.; Zangakani, M. Improving security using SVM-based anomaly detection: Issues and challenges. *Soft Comput.-A Fusion. Found. Methodol. Appl.* **2021**, *25*, 3195–3223. [\[CrossRef\]](#)
11. Primartha, R.; Tama, B.A. Anomaly detection using random forest: A performance revisited. In Proceedings of the 2017 International Conference on Data and Software Engineering (ICoDSE), Palembang, Indonesia, 1–2 November 2017; pp. 1–6.
12. Lesouple, J.; Baudoin, C.; Spigai, M.; Tournet, J.-Y. Generalized isolation forest for anomaly detection. *Pattern Recognit. Lett.* **2021**, *149*, 109–119. [\[CrossRef\]](#)
13. Ying, S.; Wang, B.; Wang, L.; Li, Q.; Zhao, Y.; Shang, J.; Huang, H.; Cheng, G.; Yang, Z.; Geng, J. An improved KNN-based efficient log anomaly detection method with automatically labeled samples. *ACM Trans. Knowl. Discov. Data (TKDD)* **2021**, *15*, 1–22. [\[CrossRef\]](#)
14. Yin, C.; Zhang, S.; Yin, Z.; Wang, J. Anomaly detection model based on data stream clustering. *Clust. Comput.* **2019**, *22*, 1729–1738. [\[CrossRef\]](#)
15. Abdi, H.; Williams, L.J. Principal component analysis. *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2*, 433–459. [\[CrossRef\]](#)
16. Bergmann, P.; Batzner, K.; Fauser, M.; Sattlegger, D.; Steger, C. The MVTec anomaly detection dataset: A comprehensive real-world dataset for unsupervised anomaly detection. *Int. J. Comput. Vis.* **2021**, *129*, 1038–1059. [\[CrossRef\]](#)
17. Li, C.-L.; Sohn, K.; Yoon, J.; Pfister, T. Cutpaste: Self-supervised learning for anomaly detection and localization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 9664–9674.
18. Yu, J.; Zheng, Y.; Wang, X.; Li, W.; Wu, Y.; Zhao, R.; Wu, L. Fastflow: Unsupervised anomaly detection and localization via 2d normalizing flows. *arXiv* **2021**, arXiv:2111.07677. [\[CrossRef\]](#)
19. Defard, T.; Setkov, A.; Loesch, A.; Audigier, R. Padim: A patch distribution modeling framework for anomaly detection and localization. In Proceedings of the International Conference on Pattern Recognition, Virtual, 10–15 January 2021; pp. 475–489.
20. Roth, K.; Pemula, L.; Zepeda, J.; Schölkopf, B.; Brox, T.; Gehler, P. Towards total recall in industrial anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 14318–14328.
21. Khalil, R.A.; Saeed, N.; Masood, M.; Fard, Y.M.; Alouini, M.-S.; Al-Naffouri, T.Y. Deep learning in the industrial internet of things: Potentials, challenges, and emerging applications. *IEEE Internet Things J.* **2021**, *8*, 11016–11040. [\[CrossRef\]](#)

22. Le, Q.; Miralles-Pechuán, L.; Kulkarni, S.; Su, J.; Boydell, O. An overview of deep learning in industry. *Data Anal. AI* **2020**, 65–98. [\[CrossRef\]](#)
23. Mjahad, A.; Rosado-Muñoz, A. Robust Industrial Surface Defect Detection Using Statistical Feature Extraction and Capsule Network Architectures. *Sensors* **2025**, 25, 6063. [\[CrossRef\]](#)
24. Rolih, B.; Ameln, D.; Vaidya, A.; Akcay, S. Divide and conquer: High-resolution industrial anomaly detection via memory efficient tiled ensemble. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 17–18 June 2024; pp. 3866–3875.
25. Aburomman, A.A.; Reaz, M.B.I. A survey of intrusion detection systems based on ensemble and hybrid classifiers. *Comput. Secur.* **2017**, 65, 135–152. [\[CrossRef\]](#)
26. Dasari, A.K.; Biswas, S.K.; Thounaojam, D.M.; Devi, D.; Purkayastha, B. Ensemble learning techniques and their applications: An overview. In Proceedings of the International Conference on Communications and Cyber Physical Engineering 2018, Hyderabad, India, 24–25 January 2023; pp. 897–912.
27. Galar, M.; Fernandez, A.; Barrenechea, E.; Bustince, H.; Herrera, F. A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches. *IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.)* **2011**, 42, 463–484. [\[CrossRef\]](#)
28. Shirley, J.J.; Priya, M. A comprehensive survey on ensemble machine learning approaches for detection of intrusion in iot networks. In Proceedings of the 2023 International Conference on Innovations in Engineering and Technology (ICIET), Muvattupuzha, India, 13–14 July 2023; pp. 1–10.
29. Aly, M.; Behiry, M.H. Enhancing anomaly detection in IoT-driven factories using Logistic Boosting, Random Forest, and SVM: A comparative machine learning approach. *Sci. Rep.* **2025**, 15, 23694. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Baimukhanov, S.; Ali, H.; Yazici, A. Enhancing ML-based anomaly detection in data management for security through integration of IoT, cloud, and edge computing. *Expert. Syst. Appl.* **2025**, 293, 128700. [\[CrossRef\]](#)
31. Liu, Y.; Zhu, L.; Ding, L.; Huang, Z.; Sui, H.; Wang, S.; Song, Y. Selective ensemble method for anomaly detection based on parallel learning. *Sci. Rep.* **2024**, 14, 1420. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Bergmann, P.; Fauser, M.; Sattlegger, D.; Steger, C. MVTec AD—A comprehensive real-world dataset for unsupervised anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 9592–9600.
33. Wood, D.; Mu, T.; Webb, A.M.; Reeve, H.W.; Luján, M.; Brown, G. A unified theory of diversity in ensemble learning. *J. Mach. Learn. Res.* **2023**, 24, 1–49.
34. Ahmad, M.W.; Reynolds, J.; Rezgui, Y. Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees. *J. Clean. Prod.* **2018**, 203, 810–821. [\[CrossRef\]](#)
35. Łoś, H.; Mendes, G.S.; Cordeiro, D.; Grosso, N.; Costa, H.; Benevides, P.; Caetano, M. Evaluation of XGBoost and LGBM performance in tree species classification with sentinel-2 data. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 5803–5806.
36. Ngo, G.; Beard, R.; Chandra, R. Evolutionary bagging for ensemble learning. *Neurocomputing* **2022**, 510, 1–14. [\[CrossRef\]](#)
37. Dietterich, T.G. Ensemble methods in machine learning. In Proceedings of the International Workshop on Multiple Classifier Systems, Cagliari, Italy, 21–23 June 2000; pp. 1–15.
38. Kazienko, P.; Lughofer, E.; Trawiński, B. Hybrid and ensemble methods in machine learning J. UCS special issue. *J. Univers. Comput. Sci.* **2013**, 19, 457–461.
39. Bruna, J.; Mallat, S. Invariant scattering convolution networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, 35, 1872–1886. [\[CrossRef\]](#)
40. Mishra, P.; Verk, R.; Fornasier, D.; Picciarelli, C.; Foresti, G.L. VT-ADL: A vision transformer network for image anomaly detection and localization. In Proceedings of the 2021 IEEE 30th International Symposium on Industrial Electronics (ISIE), Kyoto, Japan, 20–23 June 2021; pp. 1–6.
41. Mallat, S. Group invariant scattering. *Commun. Pure Appl. Math.* **2012**, 65, 1331–1398. [\[CrossRef\]](#)
42. Salazar, A.; Safont, G.; Vergara, L.; Vidal, E. Graph regularization methods in soft detector fusion. *IEEE Access* **2023**, 11, 144747–144759. [\[CrossRef\]](#)
43. Dalton, L.A. Optimal ROC-based classification and performance analysis under Bayesian uncertainty models. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2015**, 13, 719–729. [\[CrossRef\]](#)
44. Klein, A.; Falkner, S.; Springenberg, J.T.; Hutter, F. Learning curve prediction with Bayesian neural networks. In Proceedings of the International Conference on Learning Representations, Toulon, France, 24–26 April 2017.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.