Est. 1841 | YORK ST JOHN UNIVERSITY

Osagie, Efosa and Balasundaram, Rebecca (2026) Investigating Epistemic Uncertainty in PCB Defect Detection: A Comparative Study Using Monte Carlo Dropout. Journal of Experimental and Theoretical Analyses, 4 (12). pp. 93-108.

Downloaded from: https://ray.yorksj.ac.uk/id/eprint/14062/

The version presented here may differ from the published version or version of record. If you intend to cite from the work you are advised to consult the publisher's version: https://www.mdpi.com/2813-4648/4/1/11

Research at York St John (RaY) is an institutional repository. It supports the principles of open access by making the research outputs of the University available in digital form. Copyright of the items stored in RaY reside with the authors and/or other copyright owners. Users may access full text items free of charge, and may download a copy for private study or non-commercial research. For further reuse terms, see licence terms governing individual outputs. Institutional Repositories Policy Statement

RaY

Research at the University of York St John

For more information please contact RaY at ray@yorksj.ac.uk

Research Article

Investigating Epistemic Uncertainty in PCB Defect Detection: A Comparative Study Using Monte Carlo Dropout

Efosa Osagie * and Rebecca Balasundaram

Computer Science and Data Science Department, York St. John University, York YO31 7EX, London, UK; r.balasundaram@yorksj.ac.uk
* Correspondence: e.osagie@yorksj.ac.uk

Abstract

Deep learning models have become central to automated Printed Circuit Board (PCB) defect detection. However, recent work has raised concerns about how reliably these models express confidence in their predictions, particularly when deployed in safety-critical inspection systems. This study conducts an empirical investigation of epistemic uncertainty across representative architectures used in PCB inspection: the two-stage Faster R-CNN detector, the one-stage YOLOv8 detector, and their corresponding classification counterparts, ResNet-50 and YOLOv8-Cls. Monte Carlo Dropout (MCD) was applied during inference to compute predictive entropy, mutual information, softmax variance, and bounding-box variability across multiple stochastic forward passes on both multiclass and binary inspection datasets. On the multiclass SolDef_AI dataset, Faster R-CNN achieved substantially stronger detection performance (mAP = 0.7607, F1 = 0.9304) and lower predictive entropy, with more stable localisation. In contrast, YOLOv8 produced markedly weaker performance (mAP = 0.2369, F1 = 0.3130) alongside higher entropy and greater bounding-box variability. On the binary Jiafuwen datasets, the YOLOv8-Cls model achieved higher overall performance (F1 = 0.6493) compared with the ResNet-50 classifier (F1 = 0.4904), reflecting its strength in simpler binary inspection tasks. Across uncertainty metrics, predictive entropy and mutual information were more sensitive to dataset size, showing higher and more variable values in the smaller multiclass dataset, whereas softmax variance and bounding-box variability appeared more architecture-dependent. These findings demonstrate that architectural choice, dataset structure, and task formulation jointly influence both performance and uncertainty behaviour. By integrating conventional metrics with uncertainty estimates, this study provides a transparent benchmark for assessing model confidence in automated optical inspection of PCBs.

Keywords: PCB defect detection; Monte Carlo Dropout; mutual information; uncertainty quantification; Bayesian approximation; Faster R-CNN; YOLOv8

Academic Editor(s): Name

Received: date
Revised: date
Accepted: date
Published: date

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Printed Circuit Boards (PCBs) are essential components in most electronic systems, including smart devices, industrial systems and components. The design procedures for these PCBs are often resource-intensive, highly expensive, and complex. Furthermore, certain minor faults can indicate impending failures in manufacturing and electronic systems, including solder bridging and open circuits. These issues can lead to serious consequences, including product recalls, safety concerns, and financial losses. There has been

extensive research on this defect-detection challenge, including a review of 270 studies [1,2], yet there remains a persistent quality gap persists in PCB defect detection. Even a single defect can negatively impact product reliability, leading to highly complex regulatory challenges [2–4]. Given these potential consequences, PCB defect detection is seen as a critical aspect of quality assurance and adherence to regulatory compliance. Previous techniques were traditional image-analysis methods [5–7], which were fast and computationally inexpensive. However, these techniques lacked generalisation to environmental variations, such as contrast changes, registration errors, and production inconsistencies, which often led to false positives and false negatives.

In response, Deep Learning (DL), a rapidly growing subset of Artificial Intelligence (AI), has been increasingly applied to PCB defect detection due to its adaptability to such variations identified above. More specifically, Convolutional Neural Networks (CNNs) and their variants have demonstrated strong performance for spatial-based tasks, although early CNN-assisted template-matching and patented approaches [8] remained limited in reliability. The introduction of datasets such as DeepPCB [9] enabled reproducible benchmarking, achieving 98.6% mAP at 62 FPS [9], and rapidly encouraged the development of Faster R-CNN, SSD, and YOLO variants, which were more tailored for PCB inspection and detection. Examples include Faster R-CNN combined with MobileNet for detecting missing holes and shorts [10], LPCB-YOLO achieves a 24% reduction in model size while maintaining 97.0% precision and recall [11], and lightweight architectures achieving 98.1% mAP while outperforming Faster R-CNN and YOLOv5m [12]. These advances varied in scope but were DL-based and supported the practical viability of DL for PCB defect detection. However, ongoing research focuses on optimising inference speed, parameter count, generalisation, and deployment efficiency.

A closer look shows that most PCB detection studies focus primarily on performance metrics, while ignoring predictive uncertainty and the well-documented overconfidence in DL models [13]. Temperature scaling has been shown to improve calibration without affecting accuracy [13], while Monte Carlo Dropout (MCD) provides Bayesian-like uncertainty estimates via multiple stochastic forward passes [14]. For clarity, epistemic uncertainty arises from limited training data and can be reduced by acquiring more samples, whereas aleatoric uncertainty stems from inherent data noise and cannot be eliminated. This study focuses on epistemic uncertainty related to dataset size. Standard uncertainty metrics include predictive entropy, which captures overall uncertainty, and Bayesian Active Learning by Disagreement (BALD), which isolates epistemic uncertainty [15]. Practical guidance suggests that 30–100 stochastic passes can yield reliable estimates [15]. Incorporating these metrics can improve sampling strategies, inspection timing, and the safe deployment of DL-based classifiers in critical PCB workflows. For instance, predictive models in steel manufacturing have reduced inspection volume while maintaining recall on failure samples [16]. Formal guidelines such as ASME VVUQ further emphasise the importance of verification, validation, and uncertainty quantification in computational models [17].

In PCB Automated Optical Inspection (AOI), a wide range of object detection and segmentation architectures has been explored, including SSD, RetinaNet, Mask R-CNN, and transformer-based detectors [8]. This study examines representative architectures used in PCB inspection: the two-stage Faster R-CNN detector, the one-stage YOLOv8 detector, and their corresponding classification counterparts, ResNet-50 and YOLOv8-Cls. Because they represent the two principal detection paradigms used in industrial inspection workflows: two-stage and one-stage detectors [8]. This architectural distinction is central to understanding how uncertainty propagates through different detection pipelines and provides a controlled basis for comparing epistemic uncertainty under MCD. Segmentation-based approaches are excluded because the available datasets used in this

study lack pixel-level masks, and one dataset provides only binary class labels without spatial information. Transformer-based detectors are excluded due to their longer training schedules, large data requirements, and sensitivity to small or imbalanced datasets [18], as well as the practical challenges of applying MCD to attention-based architectures. Unsupervised approaches combining uncertainty indicators with reconstructive and discriminative models [19] do not address uncertainty behaviour within supervised detection pipelines.

The present study, therefore, provides the first controlled comparison of epistemic uncertainty across two major detection paradigms: two-stage and one-stage detectors using MCD under varying dataset conditions. While the multi-class SolDef_AI dataset enables a direct comparison between Faster R-CNN and YOLOv8 as object detectors, the binary Jiafuwen dataset lacks bounding-box annotations and therefore requires a classification-based evaluation. For this reason, this study uses the ResNet-50 and YOLOv8-Cls for the binary task, allowing the examination of uncertainty behaviour consistently across both detection and classification settings. By positioning the analysis around architectural paradigms and task structure, the study offers a focused and interpretable contribution to understanding the behaviour of uncertainty in PCB defect detection.

Our contributions are as follows:

I. We present a controlled, architecture-aware comparison between Faster R-CNN and YOLOv8, demonstrating how the fundamental differences between a two-stage and a one-stage detector influence predictive confidence, calibration behaviour, localisation stability, and sensitivity to dataset size. This shows that architectural design affects not only speed and accuracy but also the reliability of uncertainty estimates in PCB Automated Optical Inspection.

II. We evaluate these models across two complementary inspection settings: the multi-class, Multiview SolDef_AI dataset for object detection, and the unified binary Jiafuwen datasets for image-level classification. The latter employs ResNet-50 and YOLOv8-Cls due to the absence of bounding-box annotations. This dual-task perspective reveals how dataset structure, class granularity, and dataset size shape both performance and uncertainty behaviour, providing a clearer understanding of when and why uncertainty metrics fluctuate across inspection scenarios.

III. We apply Monte Carlo Dropout during inference to compute predictive entropy, mutual information, softmax variance, and bounding-box variability, extending the analysis beyond conventional accuracy-based evaluation. By treating uncertainty as a primary diagnostic signal rather than a post hoc add-on, the study establishes a transparent benchmark for assessing model reliability under varying dataset and architectural conditions.

Section 2 reviewed relevant work in PCB inspection, highlighting that DL-based classifiers and detectors are widely used in the literature while noting the persistent gap in uncertainty estimation. Section 3 outlines the proposed methodologies and experimental setups for both the detection models (Faster R-CNN and YOLOv8) and the classification models (ResNet-50 and YOLOv8-Cls). Section 4 presents the results, including the means and standard deviations of the uncertainty and performance metrics. Section 5 provides a critical discussion of the findings, and Section 6 concludes the paper with recommendations for future research.

2. Related Works

Earlier PCB defect-detection techniques relied mainly on image analysis and template-matching methods, which relied on pixel relationships to detect anomalies [6]. This technique typically involves a pixel-by-pixel comparison of images with reference boards [7]. However, this conventional pixel-matching technique was not adaptable to

environmental variations, leading to faulty detections, though it was effective in controlled conditions. This led to frequent false positives and false negatives [6,7]. Despite the rapid adoption of AI- and DL-based classifiers, the enhancement of DL-based algorithms, scale-adaptive template matching [8], and patented calibration-based techniques, robustness and model reliability remained a significant challenge. This reliance on stable conditions of the input PCB images emphasises the critical role of dataset quantity and quality in shaping the patterns and representations learned by DL-based classifiers. To address this challenge, Wei et al. [20] introduced a CNN-based reference comparison method for bare PCB inspection, demonstrating that deep feature extraction can significantly improve defect classification accuracy when compared to conventional pixel-matching techniques. While effective for feature discrimination, this method does not address predictive uncertainty and model reliability across varying dataset conditions. This further improved the design of CNN-based variant architectures, providing the foundation for modern PCB defect detection. This steady progress can be traced back to LeNet-5, which laid the foundation for hierarchical feature extraction in handwritten digit recognition [21], followed by AlexNet, which introduced ReLU activations (non-linearity) and dropout regularisation techniques in deeper DL-based classifiers for recognition [22]. Then, VGGNet, designed by the Visual Geometry Group (VGG), further increased the practical depth of DL networks by using uniform filters, though at a higher computational cost due to higher parameter size [23]. Then, the GoogLeNet's Inception modules improved computational efficiency by combining multiple receptive fields [24]. A recurrent problem of vanishing gradients through residual connections, enabling deeper and more generalised models, was addressed by the design of ResNet [25]. These architectures became widely adopted backbones in industrial vision tasks, including PCB defect inspection. However, most of these approaches focused solely on improving accuracy in image detection and classification tasks. Very little attention was given to the uncertainty of these large DL-based backbone models. Also, there is a gap in their performance when compared to variations in dataset size or in the reliability of their predictions, evaluated under the same constraints. With their inclusion in critical systems for PCB defect detection, where decisions must be transparent and explainable, and the consequences of incorrect predictions can be severe.

Building on these backbones, single-stage detectors such as SSD and YOLO emerged as dominant frameworks for PCB defect detection [26, 27]. For instance, the SSD mode, which performs localisation and classification in a single forward pass, improved the balance between accuracy and efficiency [26]. It has been widely tested on PCB datasets, including DeepPCB [9]. It has also been enhanced with extensions, including semi-supervised modules for unlabelled datasets [28] and lightweight base models for deployment in resource-constrained environments [29]. However, the SSD model still faces challenges with minor or clustered defects, which results to missing critical anomalies and rarely achieving stable predictions under noisy conditions [30]. To further address these issues, the YOLO (You Only Look Once) family has become particularly popular for PCB detection due to its speed and real-time applicability [31]. Recent adaptations, such as YOLOv5 and YOLOv8, have been fine-tuned for PCB tasks, improving detection of small defects through architectural modifications [32]. There have also been improvements in new variants, such as YOLOv8-DEE [33] and attention-enhanced YOLOv8 [34], that aim to improve detection while reducing inference time. This shows YOLO's applicability and flexibility for AOI and its more common applications. However, these YOLO variants often exhibit overconfidence, a phenomenon commonly observed when there are constraints on noise or distribution-shifted inputs [35]. This consequently leads to false alarms or missed detections [35]. While impressive mAP scores are frequently reported, few studies critically examine the calibration of confidence or the uncertainty behaviour of these detectors. Parallel to detector development, uncertainty quantification in DL has attracted attention

in computer vision, as evidenced by studies [36] that showed that softmax probabilities fail to reflect the true likelihood of correctness. Also, studies [37] distinguished epistemic uncertainty from aleatoric uncertainty, highlighting the importance of empirically evaluating both in critical applications. Industry-supported practical guidance recommends predictive entropy, mutual information, and softmax variance as key metrics, with 30–100 iterations for reliable estimates [15]. Beyond detector-centric approaches, Ustabas Kaya [38] proposed a hybrid optical sensor combined with a DL-classifier to detect micro-scale PCB defects, achieving strong performance under challenging imaging conditions. While this work highlights the benefits of integrating sensing and DL, it focuses solely on accuracy and does not examine uncertainty behaviour or confidence calibration.

Previous studies have proposed various solutions to advance PCB defect detection, with DL-based classifiers being the main approach in recent years [39, 40]. However, these solutions remain limited in scope when their reliability is considered. Wang et al. [39] introduced a lightweight fusion model that improved recognition of tiny defects, while Tang et al. [40] developed PCB-YOLO with attention mechanisms and transformer integration, achieving high accuracy and speed. However, their [39, 40] evaluation focused only on mAP and FPS, without considering uncertainty behaviour. To improve mode efficiency, Tang et al. [41] propose an improved YOLOv8n model (YOLO-SUMAS), which enhances detection performance through multi-module collaborative optimisation via an attention mechanism. Although YOLO-SUMAS achieved 98.8% precision, 99.2% recall, and a mAP@50 of 99.1%, significantly outperforming the original YOLOv8n and other mainstream models, its uncertainty behaviour was not thoroughly examined [41]. Multi-scale techniques coupled with Bayesian feature fusion for bare PCB defect detection [42]. However, their analysis focused primarily on improvements in classification accuracy from feature extraction [42]. Similarly, Kim et al. [43] developed a skip-connected convolutional autoencoder for PCB defect detection, leveraging reconstruction errors to highlight anomalous regions. Although effective for unsupervised defect localisation, this approach does not quantify epistemic uncertainty or assess how confidence varies across different architectures. These limitations are directly addressed by this present study. A comprehensive literature review of 105 PCB defect-detection studies identified trends and limitations but did not provide an empirical evaluation of uncertainty across architectures [44]. To address the limitations of related work, our study integrates MCD into both detection and classification tasks. The aim is to quantify predictive uncertainty metrics systematically and establish a transparent benchmark that links accuracy with uncertainty analysis for risk-sensitive deployment.

3. Methodology

This section presents our investigative approach to estimate epistemic uncertainty during inference using an MCD-based technique.

3.1. Model Architectures & Justification

3.1.1. Faster R-CNN

Faster R-CNN is a popular two-stage object detection model that integrates a Region Proposal Network (RPN) with a classification and regression head. This setup offers faster, more efficient end-to-end training and significantly improves detection speed compared to earlier R-CNN variants [45]. In detection, the RPN outputs candidate regions of interest (RoIs), which are then re-processed using bounding-box regression. Then, they are classified into pre-defined object categories. This multi-staged pipeline makes this detection model effective for computer vision tasks which require high localisation precision. An example of this is separating RoI proposal generation from classification, enabling the model to focus on fine-grained spatial features. In its original design, Faster R-

CNN was developed for general object detection benchmarks such as PASCAL VOC and MS COCO [46]. To adapt this architecture for PCB defect detection, several modifications are necessary. In this study, we replaced the backbone network with ResNet-50 and combined it with a Feature Pyramid Network (FPN) [25]. This hybrid set-up has the advantage of enhancing multi-scale feature extraction. This multi-scale enhancement is important because PCB defects are often small, subtle, and distributed across varying scales, requiring rigorous feature representation. After that, we fine-tuned the classification head to match the predefined classes in the datasets. For uncertainty estimation, only the dropout layers were activated during stochastic inference, while all Batch Normalisation layers were kept in evaluation mode to prevent updates to running statistics. This ensures that MCD captures epistemic uncertainty without introducing unintended stochasticity during training.

The relevance of Faster R-CNN to this study lies in its precision-oriented design. Unlike single-stage detectors that prioritise speed, Faster R-CNN's staged processing enables a more detailed examination of uncertainty at multiple points in the detection pipeline, both at the proposal and classification stages. This makes it an ideal candidate for studying how uncertainty metrics behave in an inherently cautious, region-focused model. By adapting its original design to incorporate uncertainty estimation, Faster R-CNN provides a structured framework for understanding how confidence calibration and localisation stability can be improved in PCB Automated Optical Inspection.

3.1.2. YOLOv8

YOLOv8 is the recent variant of the You Only Look Once family of one-stage object detectors [47]. Compared with two-stage detectors such as Faster R-CNN, it conducts both localisation and classification in a single forward pass. This gives it more preference when inference speed and throughput are critical. In this study, its architecture combines an anchor-free detection head, a Cross Stage Partial Darknet (CSPDarknet) backbone, and a Path Aggregation Network FPN (PAN FPN) neck for multi-scale feature fusion, enabling robust detection of small, densely packed objects [48]. These design choices reflect a deliberate shift towards speed and scalability, positioning YOLOv8 as a practical solution for environments where inspection must keep pace with high-volume production. To adapt it for PCB defect detection, the classification head is fine-tuned to reflect binary or multi-class defect categories in the datasets. This adaptation ensures that YOLOv8 can distinguish genuine defects from pseudo-defects and acceptable assemblies from solder-related anomalies. To estimate uncertainty, MCD was integrated into the model. Dropout layers were inserted into the backbone and neck at locations where feature activations are most sensitive to noise, with a dropout probability of 0.2. During stochastic inference, these dropout layers were kept active to generate multiple forward passes. Batch Normalisation layers were kept in evaluation mode to prevent updates to the running statistics, ensuring stable behaviour across passes. Thirty stochastic passes were performed for each image. These implementation details are included to ensure that the uncertainty estimation procedure can be reproduced reliably.

YOLO models are known to exhibit overconfidence under noisy or distribution-shifted inputs [35], and uncertainty quantification provides a mechanism to expose and mitigate these reliability issues. The relevance of YOLOv8 to this study lies in its efficiency-oriented design. While Faster R-CNN offers precision through staged processing, YOLOv8 embodies the trade-off between speed and confidence calibration. Its inclusion, therefore, provides a methodologically meaningful contrast to Faster R-CNN. YOLOv8 represents the high-throughput, industry-oriented detection paradigm, allowing this study to examine how uncertainty behaves across fundamentally different architectural philosophies rather than conducting a detailed architectural performance comparison.

This aligns with the primary aim of the work, which is to analyse uncertainty behaviour rather than to optimise detector design.

Figure 1 outlines the overall workflow used to apply, evaluate, and compare Faster R-CNN and YOLOv8. The aim is not only to measure accuracy but to examine how each model behaves when uncertainty is introduced through stochastic inference. This helps reveal whether a model's speed or architectural simplicity comes at the cost of reliability, which is a practical concern in automated PCB inspection. The framework also provides an early indication of whether calibration or uncertainty-aware strategies could make these detectors more dependable in real inspection settings.

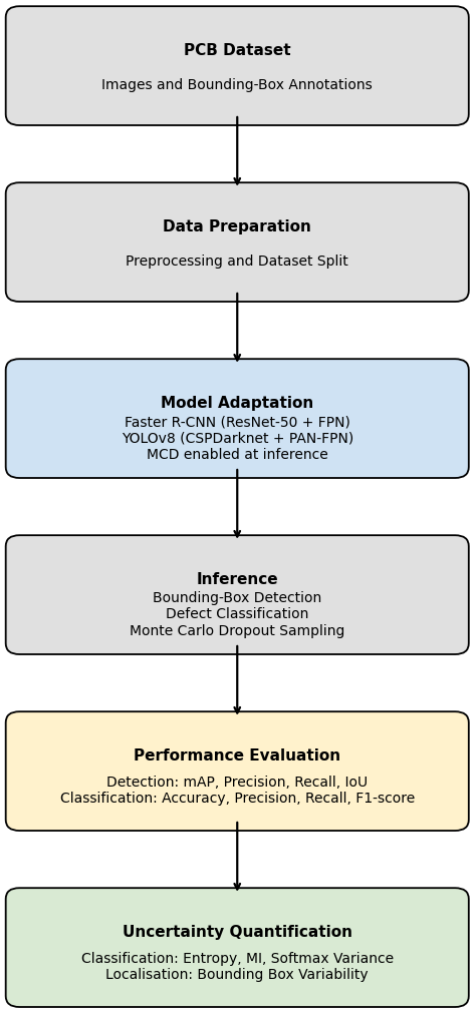


Figure 1. Overview of the proposed methodology for PCB defect detection and uncertainty evaluation.

Although the study focuses on comparing these two detection approaches, the classification experiments on the Jiafuwen dataset follow the same design logic. Because the dataset does not include bounding-box annotations, the detection models cannot be used directly. Instead, their classification counterparts were selected: ResNet-50, which shares the residual-learning foundations of the Faster R-CNN backbone, and YOLOv8-Cls, which adapts the YOLOv8 feature extractor for image-level prediction. These models do not introduce new architectural categories. They simply extend the same two design philosophies into a classification setting, allowing uncertainty to be examined consistently across both detection and classification tasks.

3.2. Uncertainty Metrics

3.2.1. Predictive Entropy

In this study, we use predictive entropy, a measure of the total uncertainty in the output distribution that quantifies the average information content of predictions across multiple stochastic forward passes. We define the number of these passes as explained in the experimental setup subsection in Section 3.4. In line with current understanding, high entropy indicates that the model is not certain about its predictions, whereas low entropy suggests the opposite. This metric is particularly relevant in PCB defect detection, where noisy imaging conditions and rare defect types can lead to unstable detector outputs. Previous studies in DL have shown that this metric can be used as a reliable indicator of overall uncertainty in classification tasks [14]. In industrial inspection, this metric has been routinely used to flag critical cases that need human reinspection to reduce the risk of false negatives [16].

3.2.2. Mutual Information

To capture epistemic uncertainty by comparing predictive entropy with the expected entropy of individual stochastic passes, we used the mutual information metric. Compared with predictive entropy, this metric, MI, isolates the uncertainty attributable to the model's parameters. This distinction is important in PCB defect detection, where epistemic uncertainty often arises from limited and/or imbalanced defect samples. Studies [35] have shown that this metric is effective at identifying regions where the model lacks adequate learned representation. This makes it a valuable tool for highlighting defect categories underrepresented in the training data. In industrial inspection contexts, this metric has been applied to distinguish between systematic model errors and random data variability, guiding retraining strategies [37].

3.2.3. Softmax Variance

We also included the softmax variance, which quantifies dispersion in class probabilities across pre-defined multiple stochastic passes [13]. In line with current understanding, a high-variance metric value indicates instability in the predicted class distribution. This may suggest that the model is sensitive to changes in the dataset or inputs, as well as parameter uncertainty. This metric is particularly relevant to the current investigation, as it can identify where subtle visual differences between genuine defects and pseudo-defects can lead to fluctuating predictions. A study highlighted that modern DL-based classifiers often produce poorly calibrated probabilities, and variance analysis provides a complementary perspective to calibration metrics [13]. In industrial inspection, this metric has been used to flag predictions that appear confident but are inconsistent across multiple passes [49].

3.2.4. Bounding Box Variability

We used bounding-box variability to assess inconsistency in spatial orientation or object localisation across stochastic forward passes [50]. In PCB defect detection, this variability in bounding boxes can directly reflect uncertainty in defect positioning. In line with current understanding, high variability indicates that the detector is unsure of the defect's exact location, leading to costly false negatives and unnecessary reinspection. Previous work in industrial vision has shown that this metric can be used as a key indicator of localisation reliability [50]. By including this metric, our work extends uncertainty analysis beyond classification confidence to spatial precision, offering a more holistic perspective on the reliability of DL-based models in PCB AOI. Bounding-box variability was computed directly from the raw detector outputs across stochastic passes, without normalising for box size or image resolution, to preserve each architecture's native localisation behaviour.

For each image, detections from all stochastic passes were aligned using an IoU-based matching strategy. A reference set of boxes was created from the first deterministic forward pass, and boxes from each stochastic pass were matched to these references using an IoU threshold of 0.5. When multiple boxes overlapped the same reference, the box with the highest IoU was selected, and passes without a matching box were treated as missing detections. Only boxes that appeared in at least half of the stochastic passes were retained for uncertainty computation. For each retained detection, class probability vectors across passes were used to compute predictive entropy and mutual information, while bounding-box variability was computed from the variance of the matched coordinates. The aggregated mean box, mode label, and mean confidence were then used for mAP and F1 evaluation.

3.3. Dataset and Defect Taxonomy

The two open-source datasets employed in this study provide complementary perspectives on PCB defect detection and are central to evaluating uncertainty-aware models. The SolDef_AI dataset [51] contains 1150 multi-view images of soldered Surface Mount Technology (SMT) components, with each sample captured from three different perspectives to reveal both placement accuracy and solder joint characteristics. It includes six classes: *background, exc_solder, good, no_good, poor_solder* and *spike*. For YOLOv8 training, the background label was excluded because YOLO does not treat the background as an explicit detection category. The remaining five foreground classes (exc_solder, good, no_good, poor_solder, spike) were used as the detection targets. This detailed labelling scheme differentiates acceptable assemblies from several types of solder-related defects. Although these labels reflect realistic manufacturing conditions, the dataset's limited size and class imbalance may pose challenges for generalisation, especially with deeper models. As a result, SolDef_AI serves as a valuable yet constrained benchmark for PCB defect research, emphasising the importance of careful evaluation protocols and the potential benefits of data augmentation or transfer learning. The second dataset, the Jiafuwen PCB defect dataset [52], comprises four unified inspection data specifications under a common label space. Each dataset is structured as a binary classification task that distinguishes between real defects and pseudo-defects encountered during PCB reinspection. This streamlined labelling scheme reflects practical industrial scenarios where separating genuine manufacturing faults from false positives is critical to maintaining inspection efficiency. While these datasets provide valuable resources for benchmarking defect detection algorithms, the binary nature of the labels may limit the granularity of diagnostic insights compared to multi-class schemes.

The choice of datasets plays a critical role in evaluating uncertainty-aware PCB defect detection, as they complement one another and offer distinct perspectives for this study. The SolDef_AI dataset provides a multi-class, multi-view benchmark that captures realistic solder joint variations, making it well-suited for testing how architectures handle fine-grained defect categorisation under limited and imbalanced data conditions. The Jiafuwen PCB dataset, by contrast, reflects practical industrial re-inspection scenarios in which distinguishing real defects from pseudo-defects is essential. Their combined use strengthens the validity of this work, ensuring that the uncertainty analyses are grounded in both detailed academic benchmarks and practical industrial contexts.

3.4. Experimental Setup

We conducted all experiments in PyTorch 2.0 with GPU acceleration, and we fixed the random seeds to ensure reproducibility. This subsection presents the experimental setup and hyperparameter configurations, organised by dataset to provide a clear and coherent description of the training and inference procedures.

3.4.1. SolDef_AI Dataset: Multi-Class Object Detection

For the SolDef_AI dataset, a custom dataset class was implemented to load images, convert polygon annotations into bounding boxes, and map labels to numerical IDs. The dataset was split into 80% for training and 20% for validation for both models.

For the Faster R-CNN model, we adopted a ResNet-50 backbone with a Feature Pyramid Network (FPN), a configuration widely used for object detection due to its strong multi-scale feature representation [46]. The model was initialised with COCO-pretrained weights, and dropout layers (0.3) were added to the classification and regression heads to enable MCD during inference. We trained the model for 100 epochs using stochastic gradient descent with a learning rate of 0.005, momentum of 0.9, weight decay of 0.0005, and a StepLR scheduler. During inference, dropout was activated for stochastic forward passes, and all Batch Normalisation layers were kept strictly in evaluation mode with fixed running statistics, ensuring that uncertainty estimates reflected dropout-induced epistemic variability rather than batch-normalisation noise. Each validation image was processed with 30 stochastic forward passes. Predictions were aggregated using an IoU threshold of 0.5 and a score threshold of 0.05, and uncertainty metrics were computed as described in Section 3.2. Performance was evaluated using mAP@0.5 and F1-score, with a confidence threshold of 0.5.

For YOLOv8, a pretrained YOLOv8n model was used. Annotations were converted into YOLO-formatted .txt files with 0-indexed class IDs. Training was performed for 100 epochs using the native YOLOv8 training API with a batch size of 8. During inference, internal dropout layers were explicitly activated to enable stochastic passes, while Batch Normalisation layers remained in evaluation mode. Each validation image was processed with 30 stochastic forward passes, and detections were aggregated using an IoU threshold of 0.5 and a score threshold of 0.05. The IoU threshold aligned with the standard mAP@0.5 evaluation metric, while the lower score threshold allowed low-confidence detections to contribute to the uncertainty analysis.

The configuration described above for Faster R-CNN was selected to balance methodological clarity with established best practices in object detection. The use of a ResNet-50-FPN backbone reflects its widely reported effectiveness in capturing multi-scale features, where the combination of deep residual representations and a FPN provides strong localisation and classification performance across varied object sizes [46]. Initialising the model with COCO-pretrained weights leverages broad visual priors, which are particularly beneficial for relatively small PCB datasets commonly encountered in industrial inspection. The integration of MCD builds on prior work demonstrating that stochastic forward passes provide a practical means of estimating epistemic uncertainty without requiring full Bayesian neural networks [14]. This approach has been shown to produce reliable uncertainty estimates in computer-vision tasks [37]. Finally, the use of IoU@0.5 as the detection threshold aligns with standard evaluation protocols in object-detection benchmarks, ensuring comparability with established studies such as PASCAL VOC [53]

3.4.2. Jiafuwen Dataset: Binary Image-Level Classification

The Jiafuwen PCB dataset lacks bounding-box annotations, so the task was formulated as a binary classification problem. The dataset was split into 70% for training, 15% for validation, and 15% for testing, providing a consistent, reproducible evaluation protocol for both classification models.

For the ResNet-50 classifier, ImageNet-pretrained weights were used. The final fully connected layer was replaced with a binary classification head, and a dropout layer (0.5) was integrated to enable MCD during inference. Training was performed for 100 epochs using the Adam optimiser and cross-entropy loss. During inference, dropout was activated for 50 stochastic forward passes per test image, and all Batch Normalisation layers

were explicitly kept in evaluation mode, ensuring that stochasticity originated solely from dropout rather than training-time behaviour. Uncertainty metrics were computed from the resulting softmax distributions. Classification performance was reported with emphasis on the real_defects class, given its operational importance.

For YOLOv8, the classification variant (YOLOv8n-Cls) was used. The pretrained classification model was configured with a dropout probability of 0.5, and dropout was activated during inference to enable stochastic passes. Training was conducted for 100 epochs with checkpoints, and 50 stochastic forward passes were executed per test image. Uncertainty metrics were computed from the resulting softmax probability distributions, and performance was evaluated using the same metrics as for the ResNet-50 classifier.

4. Results

The results presented in this section critically examine the performance of Faster R-CNN, YOLOv8, ResNet-50, and YOLOv8-Cls under uncertainty-aware PCB defect detection across both the SolDef_AI and Jiafuwen datasets. Following the experimental protocols described in Section 3.4, each model was evaluated using both conventional accuracy metrics and the uncertainty measures defined in Section 3.2. The corresponding results are summarised in Tables 1–4.

Table 1. Uncertainty Metrics on the SolDef_AI dataset.

Model	Predictive Entropy	Mutual Information	Softmax Score Variance	Bounding Box Variance
Faster R-CNN	0.0959 ± 0.1145	0.0079 ± 0.0182	0.0012 ± 0.0085	304.6492 ± 1381.5215
YOLOv8	0.2718 ± 0.1145	0.0007 ± 0.0082	0.0001 ± 0.0038	23.1839 ± 450.3332

Table 2. Performance Metrics on the SolDef_AI dataset.

Model	Mean Average Precision (mAP)	mAP_50	mAP_75	F1-Score	Precision	Recall
Faster R-CNN	0.7607	0.9612	0.8507	0.9304	0.9071	0.9549
YOLOv8	0.2369	0.2772	0.2708	0.3130	0.3711	0.2707

Table 3. Uncertainty Metrics on the Jiafuwen PCB dataset.

Model	Predictive Entropy	Mutual Information	Softmax Score Variance
ResNet-50	0.4611 ± 0.1132	0.1335 ± 0.0713	0.0712 ± 0.0341
YOLOv8-Cls	0.4205 ± 0.2072	0.0121± 0.0072	0.0087 ± 0.0077

Table 4. Performance Metrics on the Jiafuwen PCB dataset.

Model	F1-Score	Precision	Recall
ResNet-50	0.4904	0.4685	0.5144
YOLOv8-Cls	0.6493	0.5156	0.8766

5. Discussion

As shown in Tables 1 and 2, Faster R-CNN and YOLOv8 clearly differ on the SolDef_AI dataset. Faster R-CNN demonstrated strong detection performance, achieving an mAP of 0.7607, an F1-score of 0.9304, a precision of 0.9071, and a recall of 0.9549. These results indicate robust localisation and classification, suggesting that the two-stage detector effectively balances sensitivity and specificity and produces reliable detections across defect categories. In contrast, YOLOv8 showed substantially reduced performance, with mAP = 0.2369, F1-score = 0.3130, precision = 0.3711, and recall = 0.2707. This performance

gap highlights the limitations of single-stage detectors in multi-class PCB defect detection when the dataset is small and class distributions are imbalanced. The relative weakness of YOLOv8 aligns with previous findings that region-proposal architectures often outperform single-stage models on complex detection tasks requiring fine-grained localisation [23,39]. The uncertainty metrics in Table 1 further reinforce this interpretation. Faster R-CNN maintained low predictive entropy (0.0959 ± 0.1145) and modest mutual information (0.0079 ± 0.0182), reflecting stable confidence calibration across stochastic passes. Although Faster R-CNN exhibited a higher mean bounding-box variance (304.6492 ± 1381.5215), most predictions clustered near the lower end of the distribution, indicating generally stable localisation with occasional outliers. YOLOv8, by comparison, produced higher predictive entropy (0.2718 ± 0.1145) and lower mean bounding-box variance (23.1839 ± 450.3332), but with a disproportionately large spread relative to its mean. This pattern suggests that while YOLOv8's average localisation shifts were smaller, its localisation behaviour was less consistent across stochastic passes, reflecting instability in both confidence and spatial predictions. These findings imply that YOLOv8 not only struggled to achieve high detection accuracy but also produced less reliable confidence estimates.

Literature on Bayesian deep learning emphasises that predictive entropy and mutual information are critical for distinguishing epistemic uncertainty from aleatoric noise [14]. The higher predictive entropy observed in YOLOv8 indicates that its predictions are less trustworthy, particularly in environments where false positives or false negatives incur operational costs. The broader implication is that models with poorly calibrated confidence not only reduce detection reliability but also undermine the safe integration of automated inspection systems into industrial workflows, where uncertainty must function as a safeguard against costly or hazardous decision-making.

Figure 2a,b provide visual reinforcement of the findings reported in Tables 1 and 2. Panel (a) illustrates the performance metrics side by side, making the superiority of Faster R-CNN immediately apparent across mAP, F1-score, precision, and recall. Panel (b) complements this by showing uncertainty values with error bars, where Faster R-CNN's tighter distributions contrast sharply with YOLOv8's wider spread and instability.

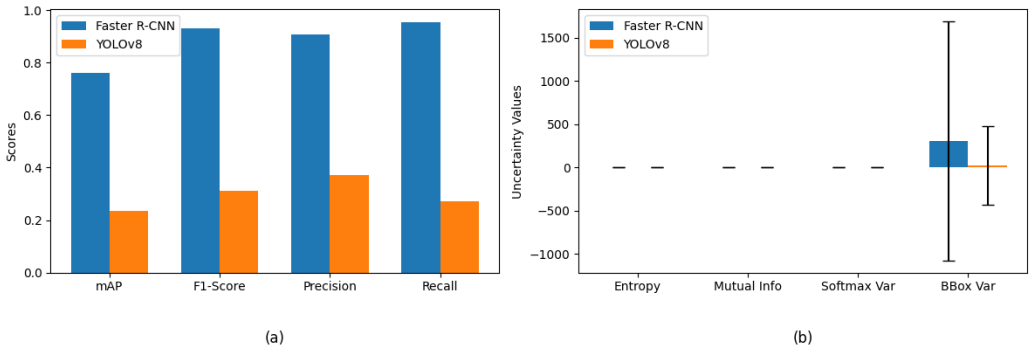


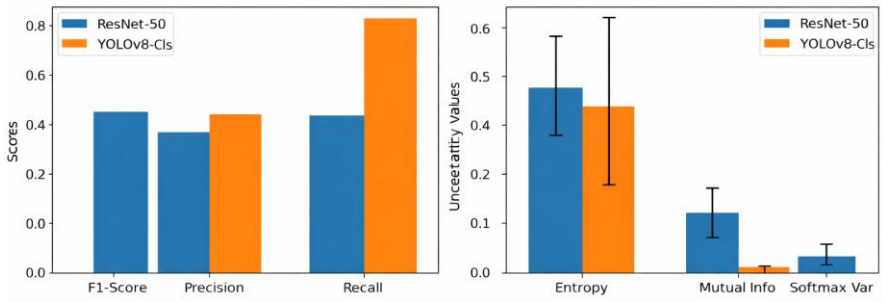
Figure 2. Performance and Uncertainty Metrics on SolDef_AI. (a) Performance comparison of Faster R-CNN and YOLOv8 across standard detection metrics. (b) Uncertainty comparison showing entropy-based measures and bounding-box variance

Jointly, these visuals confirm the numerical results and make the implications tangible. Faster R-CNN not only achieves stronger detection accuracy but also maintains more stable confidence calibration. YOLOv8, by contrast, shows both weaker performance and greater variability in its uncertainty behaviour. The size and diversity of the SolDef_AI dataset also play a decisive role in shaping these outcomes. With relatively few samples distributed across multiple defect classes, the dataset amplifies the challenges of multi-

class detection and makes confidence calibration particularly difficult for single-stage detectors such as YOLOv8. Two-stage architectures like Faster R-CNN are better equipped to compensate for data scarcity through region-proposal mechanisms, which explains their more stable performance. In practice, these findings highlight the importance of uncertainty-aware evaluation: DL-based models must be judged not only by their accuracy metrics but also by how reliably they express confidence, especially when datasets are small. Poorly calibrated uncertainty in small, multi-class datasets can negatively impact defect detection in industrial manufacturing.

Similarly, the results in Tables 3 and 4 highlight the performance differences between the ResNet-50 classifier and YOLOv8-Cls on the binary Jiafuwen PCB dataset. ResNet-50 achieved an F1-score of 0.4904, precision of 0.4685, and recall of 0.5144, indicating balanced but modest performance. YOLOv8-Cls, however, demonstrated stronger classification capability, with an F1-score of 0.6493, precision of 0.5156, and recall of 0.8766. The high recall achieved by YOLOv8-Cls suggests that it is more effective at identifying true defects, which is an essential requirement in industrial inspection contexts where missing a defect carries significant operational risk. These findings are consistent with previous studies, which have noted that lightweight single-stage architectures often perform well on binary classification tasks due to their efficiency and direct optimisation of confidence scores [29]. The uncertainty metrics in Table 3 provide further insight into these outcomes. ResNet-50 exhibited higher predictive entropy (0.4611 ± 0.1132) and mutual information (0.1335 ± 0.0713) compared with YOLOv8-Cls, reflecting greater sensitivity to epistemic uncertainty. This suggests that ResNet-50 flags cases where predictions depend strongly on stochastic sampling more readily, which is valuable for risk-sensitive deployment but also indicates difficulty in confidently separating defect from non-defect classes. YOLOv8-Cls, by contrast, showed lower mutual information (0.0121 ± 0.0072) and softmax variance (0.0087 ± 0.0077), indicating more consistent predictions across stochastic passes. Combined with its higher recall, this makes YOLOv8-Cls more suitable for binary PCB inspection tasks where sensitivity to defect detection is paramount, even if its uncertainty calibration is less expressive.

When considered jointly, the results in Tables 3 and 4 show that dataset structure strongly influences the trade-off between model performance and uncertainty behaviour. In this binary classification task, YOLOv8-Cls achieved high recall and consistent confidence, whereas ResNet-50's greater epistemic sensitivity reflects a more cautious but less operationally efficient approach. These findings reinforce the importance of evaluating both accuracy and uncertainty. While YOLOv8-Cls may be preferable in contexts where defect-detection sensitivity is critical, ResNet-50's richer uncertainty signals may be advantageous in scenarios where risk management requires explicit identification of uncertain predictions. Figure 3a,b present visual representations of these results on the Jiafuwen dataset.



(a) (b)

Figure 3. Performance and Uncertainty Metrics on Jiafuwen Dataset. (a) Performance comparison of ResNet-50 and YOLOv8-CIS (b) Uncertainty comparison across entropy-based measures.

Figure 3a,b provide visual reinforcement of the results reported in Tables 3 and 4 for the Jiafuwen dataset. Panel (a) illustrates the performance metrics side by side, clearly showing YOLOv8-Cls's superiority in recall and overall F1-score compared to the ResNet-50 classifier. Panel (b) complements this by showing the uncertainty values with error bars, where ResNet-50's higher predictive entropy and mutual information contrast with YOLOv8-Cls's lower values, highlighting the different ways each architecture expresses confidence and uncertainty in binary classification tasks.

The implications of these visuals are significant. In binary classification tasks, YOLOv8's stability and high recall make it operationally attractive, ensuring that fewer defects are missed. However, its lower epistemic sensitivity means that it provides less information about when predictions are uncertain, which could limit its usefulness in risk-sensitive contexts where uncertainty awareness is critical for decision-making. ResNet-50, by contrast, offers richer uncertainty signals that could be leveraged to flag ambiguous cases for human review, even if its overall classification performance is weaker. These visuals also reinforce the observation that dataset structure strongly influences model behaviour: binary tasks simplify the decision boundary, allowing lightweight architectures like YOLOv8-Cls to perform better, while deeper classifiers such as ResNet-50 tend to be more uncertainty-aware but less operationally efficient.

We provide additional visual representations of the uncertainty metrics for models on SolDef_AI in Figure 4 below:

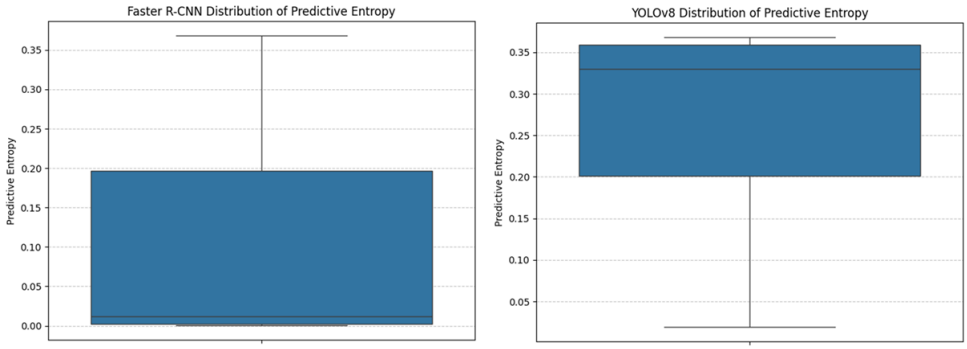


Figure 4. Boxplots of predictive entropy for Faster R-CNN and YOLOv8 on the SolDef_AI dataset.

To support Figure 4, the bounding box variance is shown below Figure 5. Figure 5 shows that Faster R-CNN and YOLOv8 differ in spatial consistency, with Faster R-CNN showing broader but steadier localisation behaviour, while YOLOv8 exhibited more unstable variance patterns.

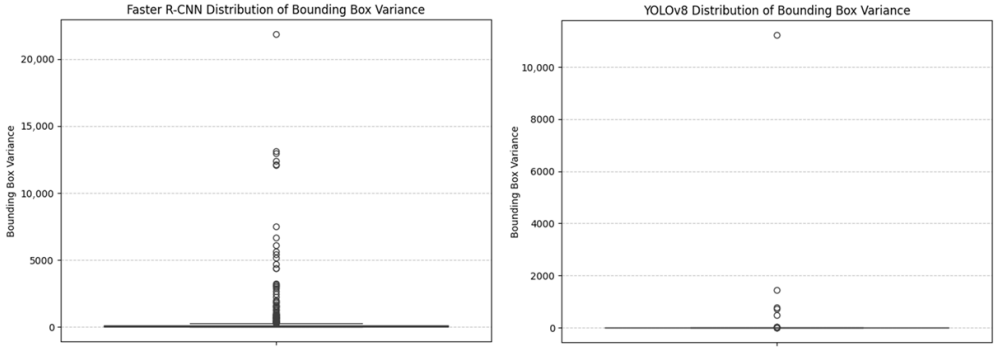


Figure 5. Boxplots of bounding box variance for Faster R-CNN and YOLOv8 on the SolDef_AI dataset.

Overall, Overall, the results across both datasets show that architectural choices and dataset structure jointly shape the balance between detection performance and uncertainty estimates. These findings provide a transparent benchmark for understanding how performance and uncertainty interact in PCB defect detection and offer a practical foundation for risk-sensitive deployment strategies in DL-based inspection systems. In this study, uncertainty thresholds were not fixed in advance. Instead, they were determined empirically by analysing the distribution of entropy and mutual information values on the validation set and identifying where higher uncertainty aligned with misclassifications or unstable localisation. This approach reflects how thresholds would be selected in practice, since uncertainty values do not generalise across datasets or architectures. Across the uncertainty metrics, predictive entropy and mutual information were more sensitive to dataset size, with higher and more variable values observed in the smaller multiclass dataset. In contrast, softmax variance and bounding-box variability were less affected by sample quantity and appeared to reflect architectural behaviour more than dataset size. This distinction suggests that some uncertainty measures respond strongly to data scarcity, while others primarily capture model-specific characteristics.

Study Limitations & Constraints

Despite the insights gained, several limitations must be acknowledged. The evaluation was restricted to two detection architectures, Faster R-CNN and YOLOv8, and two classification models, ResNet-50 and YOLOv8-Cls. This limits the scope of the investigation, since other architectures (such as SSD or transformer-based detectors) may produce different uncertainty behaviours. The datasets used, although complementary, remain relatively small and imbalanced, restricting the generalisability of the findings to broader industrial settings. Uncertainty quantification was performed using MCD with a fixed number of stochastic passes, which may not capture the full range of epistemic uncertainty compared with alternative approximation methods. A further limitation is the absence of additional uncertainty-aware baselines such as deep ensembles, Test-Time Augmentation uncertainty, or calibration-based methods. Implementing these approaches was not feasible within the scope of this unfunded study due to limited computational resources and time constraints, as these methods require training multiple independent models, thereby significantly increasing computational costs. Future work can extend this analysis by incorporating a broader set of uncertainty estimation. Similarly, only the YOLOv8n variant was evaluated in this study. Larger variants were not included because of their substantially higher computational and memory requirements, which, combined with the repeated stochastic inference required for MCD, would exceed the available resources. As a result, YOLOv8n was selected as a lightweight baseline that allowed all uncertainty experiments to be conducted consistently within the practical constraints of the study. Future work can explore larger YOLOv8 variants to assess how model capacity interacts with uncertainty behaviour. The study also did not include training-strategy ablations, such as varying fine-tuning depth, alternative learning-rate schedules, or extended training runs, because these experiments would have required substantially more computational time and resources than were available. These choices may influence both performance and uncertainty estimates, and a more exhaustive exploration of training dynamics remains an important direction for future work. In addition, uncertainty was not evaluated in a decision-making context, such as selective prediction or risk–coverage analysis. These techniques require additional modelling stages and larger validation sets to reliably estimate coverage–risk trade-offs, which were beyond the scope of this study.

Statistical hypothesis testing was not performed, as the aim of the study was to characterise uncertainty behaviour rather than to make inferential claims about statistically significant differences between models. The experiments were conducted under controlled conditions, and distribution shifts or noise injections were not explored. To ensure the results remained valid despite these constraints, random seeds were fixed, hyperparameters were documented, and multiple stochastic passes were applied consistently to reduce variance in the uncertainty estimates. Performance metrics were reported alongside uncertainty measures to provide a balanced view of model behaviour. Another notable limitation concerns the scale dependence of the bounding-box variability metric. Because variability was computed directly from raw detector outputs without normalising for object size or image resolution, the resulting values reflect architecture-specific localisation stability rather than a scale-invariant measure. This choice preserves native detector behaviour but limits comparability across objects of different sizes or across detectors with differing receptive-field characteristics. Future work could incorporate size-normalised or resolution-normalised variants to enable fairer cross-architecture comparison. Bounding-box variability also did not correlate with false negatives. In the detection task, false negatives occurred when the model failed to produce a bounding box for a defect, which prevented variability from being computed. In the binary classification task, no bounding boxes are generated, so the variability metric is not applicable. As a result, bounding-box variability reflects localisation stability only for successful detections and cannot be used to explain false-negative behaviour in either setting. Even with these limitations, the study contributes a transparent benchmark that integrates conventional performance metrics with uncertainty analysis. It establishes a baseline for future work that can be extended with larger datasets, a broader range of architectures, and more realistic deployment conditions where experimental factors can be varied systematically.

6. Conclusions

This study has experimentally demonstrated that Monte Carlo Dropout can be used to investigate epistemic uncertainty in PCB defect detection, and the results reveal a clear relationship between model architecture and uncertainty behaviour. On the multi-class SolDef_AI dataset, Faster R CNN achieved substantially stronger detection performance (mAP = 0.7607, F1 = 0.9304) while maintaining lower predictive entropy and generally stable localisation. These characteristics resulted in more reliable uncertainty estimates. In contrast, YOLOv8 produced markedly lower multi-class performance (mAP = 0.2369, F1 = 0.3130) and exhibited higher predictive entropy and less consistent localisation across stochastic passes. These outcomes reflect the challenges single-stage detectors face when trained on small, imbalanced multi-class datasets. On the binary Jiafuwen dataset, YOLOv8-Cls achieved stronger classification capability (F1 = 0.6493, recall = 0.8766) compared with the ResNet-50 classifier (F1 = 0.4904). The high recall achieved by YOLOv8-Cls indicates greater sensitivity to true defects, which is critical in industrial inspection scenarios where missed defects carry significant operational risk. ResNet-50, by comparison, exhibited higher predictive entropy and mutual information. This pattern signals greater epistemic sensitivity but also greater difficulty in confidently separating defect from non-defect classes in this binary setting. Across both datasets, predictive entropy and mutual information were more sensitive to dataset size and showed larger fluctuations in the smaller multi-class dataset. Softmax variance and bounding box variability appeared more dependent on architectural behaviour than on sample quantity. These findings highlight that accuracy alone is not sufficient for evaluating inspection systems. Uncertainty awareness is essential for risk-sensitive deployment.

Future work will extend beyond epistemic uncertainty to hybrid approaches that also capture aleatoric effects. This will provide a fuller picture of both model-driven and data-

driven risks. Additional directions include formal statistical hypothesis testing to assess the significance of differences in uncertainty across architectures, as well as robustness evaluations under controlled distribution shifts and noise injection scenarios. Such analyses will clarify how uncertainty behaves when imaging conditions degrade or deviate from the training distribution. While recent studies show that region-based CNNs can be adapted for PCB defect detection [54], this study's results emphasise that accuracy alone is not sufficient. In safety-critical settings, including PCB inspection, healthcare, and autonomous systems, uncertainty must be integrated into both model design and result interpretation to ensure that AI systems are not only high-performing but also trustworthy and operationally reliable.

Author Contributions: Conceptualisation, E.O.; methodology, E.O.; experiments, E.O.; result interpretation, E.O.; literature research, R.B.; writing—original draft preparation, E.O. and R.B.; writing—review and editing, E.O.; project administration, E.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original datasets used in this study are openly available in the Jiafuwen PCB defect datasets at https://www.kaggle.com/datasets/jiafuwen77/multiple-datasets-on-pcb-defects](Accessed 02 November 2025). The SolDef_AI dataset at [https://www.kaggle.com/datasets/mauriziocalabrese/soldef-ai-pcb-dataset-for-defect-detection](Accessed 02 November 2025).

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

References

1. Chen, X.; Wu, Y.; He, X.; Ming, W. A comprehensive review of deep learning based PCB defect detection. *IEEE Access* **2023**, *11*, 112345–112367.
2. Bhattacharya, A.; Cloutier, S.G. End to end deep learning framework for printed circuit board manufacturing defect classification. *Sci. Rep.* **2023**, *13*, 12345–12356.
3. Lang, D.; Lv, Z. SEPDNet: Simple and effective PCB surface defect detection method. *Sci. Rep.* **2025**, *15*, 10919. https://doi.org/10.1038/s41598-024-84859-2.
4. Oliveira, G.G. de et al. System for PCB defect detection using visual computing and deep learning. *IET Circuits Devices Syst.* **2023**, *17*, 456–467.
5. Wei, Z.; Yang, F.; Zhong, K.; Yao, L. PCB-YOLO: Enhancing PCB surface defect detection with coordinate attention and multi-scale feature fusion. *PLoS ONE* **2025**, *20*, e0323684. https://doi.org/10.1371/journal.pone.0323684.
6. Ling, L.; Isa, N.A.M. A review of automated optical inspection techniques for PCB defect detection: Image analysis, template matching, and challenges in real world manufacturing. *J. Electron. Manuf. Syst.* **2023**, *45*, 115–130.
7. Sonar, P.; Patil, S.; Kumar, R. Template based PCB defect localisation under varying illumination and registration noise. *Int. J. Imaging Vis. Eng.* **2023**, *18*, 221–234.
8. Liu, Y.; Zhang, H.; Chen, Q. Scale adaptive template matching with CNN generated features for industrial print inspection. *IEEE Trans. Ind. Inform.* **2023**, *19*, 8421–8433.
9. Tang, X.; Li, Y.; Li, X. DeepPCB: A dataset for printed circuit board defect detection. In Proceedings of the CVPR Workshops, Long Beach, CA, USA, 16–20 June 2019; pp. 1172–1180.
10. Yan, H.; Zhang, H.; Gao, F.; Wu, H.; Tang, S. Deep learning model enhancements for PCB surface defect detection. *Electronics* **2024**, *13*, 4626.
11. Yang, H.; Dong, J.; Wang, C.; Lian, Z.; Chang, H. PCES-YOLO: High-Precision PCB Detection via Pre-Convolution Receptive Field Enhancement and Geometry-Perception Feature Fusion. *Appl. Sci.* **2025**, *15*, 7588. https://doi.org/10.3390/app15137588.

12. Bhattacharya, S.; Cloutier, R. A systematic review of PCB quality and inspection challenges in modern manufacturing. *J. Electron. Manuf.* **2023**, *45*, 112–129.
13. Guo, C., Pleiss, G., Sun, Y. & Weinberger, K.Q.. (2017). On Calibration of Modern Neural Networks. Proceedings of the 34th International Conference on Machine Learning, in Proceedings of Machine Learning Research 70:1321-1330 Available from https://proceedings.mlr.press/v70/guo17a.html. (Accessed 10 November 2025). .
14. Gal, Y. & Ghahramani, Z.. (2016). Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. Proceedings of The 33rd International Conference on Machine Learning, in Proceedings of Machine Learning Research 48:1050-1059 Available from https://proceedings.mlr.press/v48/gal16.html (Accessed 12 November 2025).
15. Amazon Web Services. *AWSPG Practical Guidance for Uncertainty Estimation in Deep Learning Systems*; Amazon Web Services Technical Report; Amazon Web Services: Seattle, WA, USA, 2025. Available from https://d1.awsstatic.com/APG/quantifying-uncertainty-in-deep-learning-systems.pdf (Accessed 13 November 2025).
16. Song, Y., Li, F., Wang, Z., Zhang, B., & Zhang, B. (2024). Uncertainty Quantification of Data-driven Quality Prediction Model For Realizing the Active Sampling Inspection of Mechanical Properties in Steel Production. International Journal of Computational Intelligence Systems, 17(1). https://doi.org/10.1007/s44196-024-00451-6
17. ASME. *ASME VVUQ: Verification, Validation, and Uncertainty Quantification Standards for Computational Models*; American Society of Mechanical Engineers: New York, NY, USA, 2024. Available from https://www.asme.org/codes-standards/publications-information/verification-validation-uncertainty (Accessed 10 November 2025).
18. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-End Object Detection with Transformers (Version 3). *arXiv* **2020**, https://doi.org/10.48550/ARXIV.2005.12872.

19. Chen, C.; Wu, Q.; Zhang, J.; Xia, H.; Lin, P.; Wang, Y.; Tian, M.; Song, R. U2D2PCB: Uncertainty-Aware Unsupervised Defect Detection on PCB Images Using Reconstructive and Discriminative Models. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 5017710. https://doi.org/10.1109/tim.2024.3386210.
20. Wei, P.; Liu, C.; Liu, M.; Gao, Y.; Liu, H. CNN-based reference comparison method for classifying bare PCB defects. *J. Eng.* **2018**, *2018*, 1528–1533. https://doi.org/10.1049/joe.2018.8271.

21. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324.
22. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105.
23. Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition (Version 6). arXiv. https://doi.org/10.48550/ARXIV.1409.1556
24. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2014). Going Deeper with Convolutions (Version 1). arXiv. https://doi.org/10.48550/ARXIV.1409.4842
25. He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition (Version 1). arXiv. https://doi.org/10.48550/ARXIV.1512.03385
26. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). SSD: Single Shot MultiBox Detector. In Lecture Notes in Computer Science (pp. 21–37). Springer International Publishing. https://doi.org/10.1007/978-3-319-46448-0_2
27. Ali, M.L.; Zhang, Z. The YOLO Framework: A Comprehensive Review of Evolution, Applications, and Benchmarks in Object Detection. *Computers* **2024**, *13*, 336. https://doi.org/10.3390/computers13120336.
28. Wan, X.; Li, J.; Xu, P. DE SSD: A semi supervised single shot detector for PCB defect detection using unlabeled samples. *Sensors* **2022**, *22*, 4521.
29. Lim, J.H.; Park, S.; Kim, H. Lightweight SSD with MobileNetV2 backbone for real time PCB defect detection on edge devices. *IEEE Access* **2023**, *11*, 55678–55690.
30. Kang, L., Ge, Y., Huang, H., & Zhao, M. (2022). Research on PCB defect detection based on SSD. In 2022 IEEE 4th International Conference on Civil Aviation Safety and Information Technology (ICCASIT) (pp. 1315–1319). IEEE. 2022 IEEE 4th International Conference on Civil Aviation Safety and Information Technology (ICCASIT). https://doi.org/10.1109/iccasit55263.2022.9986754
31. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2015). You Only Look Once: Unified, Real-Time Object Detection (Version 5). arXiv. https://doi.org/10.48550/ARXIV.1506.02640
32. Xiao, T.; Liu, H.; Zhang, J. Enhanced PCB defect detection using hybrid CNN–transformer models and YOLO based frameworks. *Pattern Recognit. Lett.* **2024**, *178*, 25–33.

33. Yi, H.; Chen, L.; Wu, Z. YOLOv8 DEE: An enhanced detection engine for small scale PCB surface defects. *IEEE Trans. Ind. Inform.* **2024**, *20*, 6123–6134.
34. Bai, Q.; Zhou, M.; Lin, Y. Attention augmented YOLOv8 for high precision PCB defect detection. *J. Manuf. Syst.* **2025**, *79*, 201–214.
35. Calabrese, G.; Rossi, M.; Bianchi, F.; Romano, A. Uncertainty aware defect detection in industrial visual inspection: Challenges and opportunities. *J. Manuf. Syst.* **2025**, *78*, 112–129.
36. Catak, F.O.; Yue, T.; Ali, S. Uncertainty-aware Prediction Validator in Deep Learning Models for Cyber-physical System Data. *ACM Trans. Softw. Eng. Methodol.* **2022**, *31*, 1–31. https://doi.org/10.1145/3527451.
37. Kendall, A.; Gal, Y. What uncertainties do we need in Bayesian deep learning for computer vision? *Adv. Neural Inf. Process. Syst. (NeurIPS)* **2017**, *30*, 5574–5584. https://doi.org/10.48550/arXiv.1703.04977.
38. Ustabas Kaya, G. Development of hybrid optical sensor based on deep learning to detect and classify the micro-size defects in printed circuit board. *Measurement* **2023**, *206*, 112247. https://doi.org/10.1016/j.measurement.2022.112247.
39. Wang, Y.; Chen, Z.; Wang, J.; Shang, P.; Sowmya, A.; Sun, C. Printed circuit board defect detection based on lightweight deep learning fusion model. *Sensors* **2025**, *25*, 7403. https://doi.org/10.3390/s25247403.
40. Tang, J.; Liu, S.; Zhao, D.; Tang, L.; Zou, W.; Zheng, B. PCB-YOLO: An improved detection algorithm of PCB surface defects based on YOLOv5′. *Sustainability* **2023**, *15*, 5963. https://doi.org/10.3390/su15075963.
41. Tang, Y.; Liu, R.; Wang, S. YOLO-SUMAS: Improved Printed Circuit Board Defect Detection and Identification Research Based on YOLOv8. *Micromachines* **2025**, *16*, 509. https://doi.org/10.3390/mi16050509.
42. Han, X.; Li, R.; Wang, B.; Lin, Z. Defect identification of bare printed circuit boards based on Bayesian fusion of multi-scale features. *PeerJ Comput. Sci.* **2024**, *10*, e1900. https://doi.org/10.7717/peerj-cs.1900.
43. Kim, J.; Ko, J.; Choi, H.; Kim, H. Printed Circuit Board Defect Detection Using Deep Learning via A Skip-Connected Convolutional Autoencoder. *Sensors* **2021**, *21*, 4968. https://doi.org/10.3390/s21154968.
44. Alawandi, S.; Mallibhat, K.; Kudachi, U.; Beedanal, A. PCB Defects: A Unified Survey of Trends, Detection Techniques, and Limitations through Systematic Literature Review. *J. Electron. Test.* **2025**, *41*, 709–787. https://doi.org/10.1007/s10836-025-06198-y.
45. Vu, T.; Jang, H.; Pham, T.X.; Yoo, C.D. Cascade RPN: Delving into High-Quality Region Proposal Network with Adaptive Convolution (Version 2). *arXiv* **2019**, https://doi.org/10.48550/ARXIV.1909.06720.
46. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R CNN: Towards real time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031.
47. Varghese, R., & M., S. (2024). YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness. In 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS) (pp. 1–6). IEEE. 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS). https://doi.org/10.1109/adics58448.2024.10533619
48. Cao, Y.; Luo, H.; Wang, M.; Wang, Y.; Yan, H. Enhanced YOLOv8 for accurate and efficient floating object detection on water surfaces. *Sci. Rep.* **2025**, *16*, 2907. https://doi.org/10.1038/s41598-025-32817-x.
49. Fontana, E.; Gupta, R.; Lin, T. Noise induced uncertainty in industrial CNN inspection systems. *Eng. Appl. Artif. Intell.* **2024**, *128*, 107345.
50. Calabrese, M.; Rossi, F.; Bianchi, L. Evaluating robustness and uncertainty in CNN based industrial inspection systems. *IEEE Trans. Ind. Inform.* **2025**, *21*, 1455–1468.
51. Fontana, G., Calabrese, M., Agnusdei, L., Papadia, G., & Del Prete, A. (2024). SolDef_AI: An Open Source PCB Dataset for Mask R-CNN Defect Detection in Soldering Processes of Electronic Components. Journal of Manufacturing and Materials Processing, 8(3), 117. https://doi.org/10.3390/jmmp8030117
52. Wen, J. Multiple Datasets on PCB Defects. Kaggle. 2024. Available online: https://www.kaggle.com/datasets/jiafuwen77/multiple-datasets-on-pcb-defects (Accessed on 04 November 2025).

53. Everingham, M.; Van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes (VOC) challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. https://doi.org/10.1007/s11263-009-0275-4.
54. Weerakkody, K.D.; Balasundaram, R.; Osagie, E.; Alshehabi Al-Ani, J. Automated Defect Identification System in Printed Circuit Boards Using Region-Based Convolutional Neural Networks. *Electronics* **2025**, *14*, 1542. https://doi.org/10.3390/electronics14081542.

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.