8

When a model is not Enough: A Complementary AI Pipeline for Ultra-Safe PCBA Defect Detection

Alberto Faro, Francesco Cancelliere, Robin Faro, and Raffaele Mineo

Deepsensing s.r.l., Italy

Abstract

In safety-critical domains such as automotive and avionics, automated optical inspection (AOI) systems must meet extremely strict reliability thresholds, typically fewer than 10 false negatives per million inspected units, with high statistical confidence. This paper introduces a multi-stage AOI pipeline designed to meet such requirements under real-world imperfections, including imaging noise and process variability. The architecture uses a cascaded structure, where each stage is optimized to balance high recall with progressively improved precision. Through mathematical analysis and large-scale simulations, we show that single-stage AOI systems, even with recall artificially forced to 100%, fail to statistically guarantee zero defect escapes on large batches. This is due to variability in manufacturing conditions and model behaviour, which makes extremely high thresholds impractical without generating unacceptably high false positive rates. In contrast, our multi-stage system achieves the target of fewer than 10 defect escapes per million units with 90% statistical confidence, while maintaining acceptable production yield. The approach has been validated through simulation of a physical prototype with synchronized image acquisition and edge-based real-time processing. This architecture is scalable and applicable to increasingly

stringent industrial settings, offering a practical pathway to combining safety and efficiency in AI-powered visual inspection.

Keywords: edge AI, image recognition, PCBA AOI, Statistical Stability in Industrial Processes, Statistical Control on Large-Scale Production.

8.1 Introduction

Safety-critical industries such as automotive and avionics impose stringent quality requirements on electronic assemblies. In these domains, the cost of a missed defect is not merely financial: it can translate into operational failure, large-scale recalls and ultimately risks to human life. As a result, production systems are expected to achieve extremely low defect escape rates, commonly phrased as fewer than 10 false negatives per million inspected units ($\leq$10 ppm) with high statistical confidence. This requirement is widely recognized in the literature. For instance, *Wu et al.* discuss the challenges of training classifiers in highly imbalanced settings for rare-event detection [1], while *Guo et al.* explore cascaded architectures for anomaly detection without statistical guarantees [2].

When inspection targets printed circuit board assemblies (PCBA), this means that out of millions of boards shipped as "good", virtually none may contain latent defects. Traditional deep-learning-based Automated Optical Inspection (AOI) systems, even when they report high overall accuracy, struggle to meet these order-of-magnitude guarantees at scale.

This tension is rooted in a well-known precision-recall trade-off. Pushing recall toward unity (so that no defective unit is missed) generally requires lowering decision thresholds, which inflates false positives (good boards erroneously flagged as defective). In small laboratory datasets, the extra review load is tolerable; on a high-throughput line, however, every additional false positive triggers manual re-inspection, production slowdowns and work-in-progress (WIP) accumulation. The consequences are immediately visible in two key operational levers: capacity planning and *takt* time.

The **process capability index (*Cpk*)** measures how well a process produces outputs that fit within specification limits relative to its natural variability:

$$\mathbf{C}pk \; = \mathbf{min}[(\mathbf{USL} - -\boldsymbol{\mu})/\mathbf{3}\boldsymbol{\sigma}, (\boldsymbol{\mu} - \mathbf{LSL})/\mathbf{3}\boldsymbol{\sigma}] \tag{8.1}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are the process mean and standard deviation, [LSL, USL] are the lower and upper specification limits. In the context of AOI-driven

quality, *Cpk* captures more than mechanical tolerances: it reflects how stable and predictable the *inspection + rework* process is. When AOI precision is low (i.e. many false positives), the inspection stage injects variability into the production flow: good boards are diverted to re-inspection queues, buffers grow (work in progress increases), operators are reassigned from value-adding tasks to sorting and confirmation, cycle time elongates and variability rises. All of this widens the effective standard deviation σ of the end-to-end process and reduces ***Cpk***. A lower ***Cpk*** means the process delivers fewer units within "on-time, in-spec" boundaries, which in turn results in missed ***takt*** and capacity loss.

In quantitative terms, automotive manufacturers generally require ***Cpk*** $\geq$ 1.67 ($\sim$99.97% process capability, $\sim$30 ppm defects), whereas avionics and aerospace production typically demands ***Cpk*** $\geq$ 2.0-2.33 ($\geq$ 99.999% accuracy, $\leq$ 3.4 ppm defects). Achieving such capability levels requires AOI systems whose effective consistently exceeds 99.995% under process variability, a target that single deep networks rarely sustain under real production variability.

Takt time defines the cadence the line must maintain to satisfy customer demand as follows:

$$\textbf{Takt time = Available production time / Customer demand} \tag{8.2}$$

In a stable flow, inspection adds a small, predictable overhead that can be planned into the cycle time. But when AOI yields frequent false alarms, the effective cycle time becomes stochastic: units are repeatedly stopped, routed and re-tested. Two effects follow:

- **Direct delay:** each false positive consumes additional inspection time. On a line producing 60 boards per minute, even a 2% false-positive rate creates dozens of interruptions per hour, pushing average cycle time beyond *takt*
- **Propagation delay:** re-inspection queues cause blocking and starvation upstream and downstream. The line controller reacts conservatively, throttling feed rates to prevent overflows, which further increases cycle-time variance

As an illustrative example let us consider a line processing one million boards per year with a true defect rate of 0.1% (1.000 defective boards). A single-model AOI tuned to R = 0.999 and P = 0.990 yields roughly one false negative but about 9.990 false positives, that is nearly ten thousand unnecessary stops. If each re-inspection adds 45 seconds, the plant accrues

more than 125 hours of non-value-added time, not counting the ripple effects on buffers and scheduling. ***Cpk*** drops because the variance of completion times increases and the line systematically misses ***takt*** during peaks or tight delivery windows. Even when **re-inspection is performed offline**, the accumulated rework time translates into substantial non-value-added effort and increased work-in-progress. The resulting increase in completion-time variability degrades process capability (**Cpk**), causing the line to miss ***takt time*** during production peaks or tight delivery windows.

The upshot is clear: **precision drives *Cpk*** and ***takt time***, while **recall drives safety**. A single classifier rarely sustains both simultaneously at the parts-per-million level. In practice, even classifiers reaching 97-98% accuracy are insufficient when scaled to millions of units. The residual 2-3% error rate, acceptable in consumer electronics, becomes catastrophic in safety-critical manufacturing.

Recent work has explored multi-stage or multi-view AOI pipelines using CNNs [3] or hybrid methods combining detection and filtering [4]. However, these approaches typically optimise either precision or recall, but not both in a statistically robust way. They also lack formal reasoning frameworks that guarantee ppm-level safety under production variability.

The central industrial question addressed in this paper is therefore: Can we raise recall to near unity (no defective PCBA escaping) without triggering unsustainable growth in false positives, re-inspections and ***takt*** time?

Section 2 deals with the problem formalization (single-model limits). In particular, it formalizes the problem by quantifying the theoretical limits of single-model AOI in ultra-safe regimes. In particular, it shows that under ppm-level constraints a single stage cannot simultaneously achieve near-zero False Negatives (FN) and very low False Positives (FP) at production scale. This analysis, grounded in the **Bernoulli/Poisson envelope** and one-sided **Clopper-Pearson bounds** (often referred to as **CP bounds**), lays the foundation for the multi-stage design that follows.

Section 3 introduces the main features of the three-stage pipeline and operational control**.** It presents a three-stage complementary pipeline that decouples safety from throughput: Stage 1 (T) is tuned for high recall and broad coverage of suspicious regions (tolerating some FP); Stage 2 (K1) emphasizes high precision, filtering false alarms and recovering good boards otherwise sent to rework; Stage 3 (K2) performs consistency checks on both "good" and "bad" flows, removing residual rare errors and providing traceable justifications. This decomposition preserves recall early while restoring precision downstream, stabilizing ***Cpk*** and keeping ***takt time*** within

contractual limits. The section also addresses how to control variability in a more complex system, via image registration, illumination checks, inter-board similarity (IBS) and predictive maintenance on the vision/conveyor chain, so that process drift does not degrade the expected ***FN*** and ***FP*** agreed with the customer.

Section 4 points out the validation protocol and certification under worst-case assumptions. A detailed description of the validation and certification protocol is given by combining **Monte Carlo** simulation of realistic defect distributions with a **rare-event** reasoning framework inspired by, yet more elaborate than, the sequential filter approach used at **CERN for the Higgs boson discovery** [5]. Whereas CERN's triggers maximize precision at the expense of recall, our method balances both extremes: it minimizes defect escape (recall ≈ 1) while maintaining production-flow stability (precision $\approx$ 0.98), keeping ***Cpk*** high even under statistical variability.

Let **DPPM** denotes the number of defective units per one million produced units, this section also specifies what can be certified. We report monthly and cumulative point-estimate DPPM for operational monitoring together with one-sided 95% Clopper-Pearson DPPM for certification. We further note that the monitoring and fallback logic, including FP trend analysis, Inter-Board Similarity (IBS), illumination and focus drift detection, safe profiles; golden set reaches and predictive maintenance, helps ensure that the control system operates against worst-case conditions that may still arise in a controlled environment, thereby keeping the process within the agreed false negatives and false positives limits.

Finally, Section 5 previews industrial integration in ***VisioEdgeAI***, a robotic inspection platform where the three-stage logic operates over virtual flows on a single conveyor. Defective candidates are physically removed by a robot arm for laboratory confirmation, while "good" boards continue unimpeded. This architecture preserves high recall, restores high precision, stabilizes ***Cpk*** and protects ***takt time***, aligning AI-powered AOI with the uncompromising reliability thresholds of automotive and avionics production.

Related Work and Novelty of Our Approach

Several studies have proposed improvements to AOI systems using convolutional neural networks (CNNs), transformers and hybrid visual models [6], with a strong focus on enhancing classification accuracy. Other works have explored cascaded CNN pipelines [7] or Bayesian post-classification filters [8] to reduce false positives. However, these approaches rarely address

process-level KPIs such as *takt* time or capability indices (***Cpk***) and generally offer no statistical guarantees regarding ppm-level defect escapes under real-world production variability.

In contrast, the approach proposed in this paper differs fundamentally. We introduce a complementary, multi-stage inspection pipeline, not merely for boosting accuracy, but for formally guaranteeing less than 10 escaped defects per million ($\leq$ 10 ppm) with $\geq$ 90% confidence. Our design explicitly balances sensitivity and selectivity while preserving ***Cpk*** $\geq$ 1.67-2.33, integrating statistical reasoning, explainability and real-time deploy ability on edge AI platforms. To our knowledge, this is the first work to rigorously analyse and simulate these trade-offs under realistic AOI constraints.

8.2 Theoretical Limits of Single-Model AOI

In high-volume digital manufacturing (millions of PCBAs/year), even tiny error rates produce large absolute counts of misclassified units. We adopt the standard metrics:

- **Precision:** fraction of flagged boards that are truly defective
- **Recall:** fraction of truly defective boards that are found
- **True Positives (TP):** defective boards correctly flagged
- **False Positives (FP):** good boards incorrectly flagged (waste)
- **False Negatives (FN):** defective boards missed (dangerous) where:

where:

$$FN = (\mathbf{1} - -\mathbf{Recall})\mathbf{Ndefective}, FP = (\mathbf{1} - -\mathbf{Precison})\mathbf{N_{good}} \tag{8.3}$$

Because typically Ngood $>>$ **Ndefective** even minute deviations from perfect recall or precision amplify into thousands of errors.

8.2.1 Hypothetical Extreme Case (Proof by Contradiction)

This proof considers a hypothetical extreme case to effectively convey the point in an engineering context. Assume a single-model AOI achieves near-perfect Recall $=$ 0.999 and Precision = 0.999. On a production line with 10^6 boards per year and a 0.1% defect rate:

- ***FN*** = (1 - 0.999) $\times$ 1'000 = 1 missed defect
- ***FP*** = (1 - 0.999) $\times$ 999.000 = 999 false alarms

Even under this near-ideal scenario, ≈ 1.000 total misclassifications occur per million boards. This contradiction shows that statistical amplification renders single-model perfection unattainable at scale.

8.2.2 Threshold Dependence and Trade-off

The precision-recall curve highlights an inherent trade-off. Lowering the classification threshold increases recall (fewer ***FN***) but reduces precision (more FP); raising it has the opposite effect. Figure 8.1 illustrates that no fixed threshold can simultaneously optimize both. This trade-off is not merely theoretical: in AOI systems, lowering thresholds results in excessive re-inspections, while raising them risks letting critical defects pass undetected. Such tension makes static, single-model AOI unsuitable for high-assurance manufacturing contexts.

8.2.3 ROC and Asymptotic Limit

The Receiver Operating Characteristic (ROC) curve further emphasizes this limitation. As the decision threshold is relaxed, true positive rate or recall (***TPR***) increases at the cost of the false positive rate (***FPR***). In the ROC curve shown in Figure 8.2, the log-scaled x-axis highlights that pushing recall toward 1 yields orders-of-magnitude increases in ***FPR***, illustrating

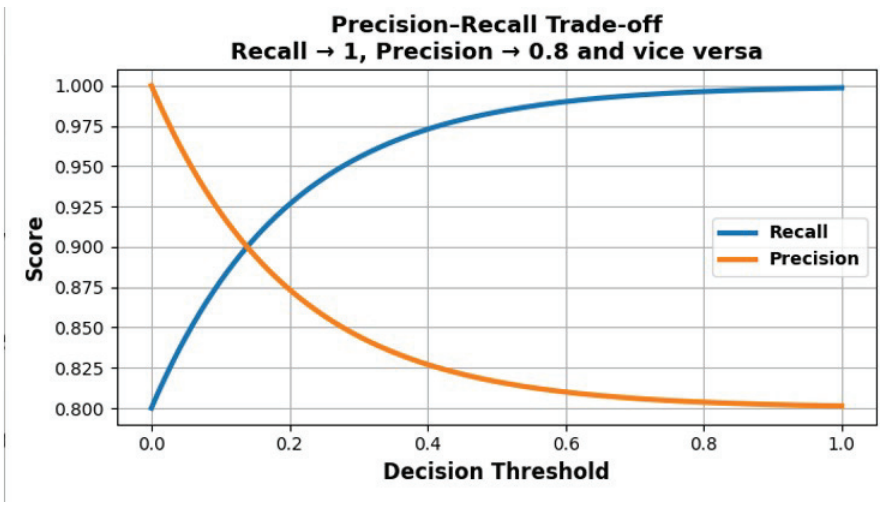

Figure 8.1 Precision and Recall Vs Decision Threshold

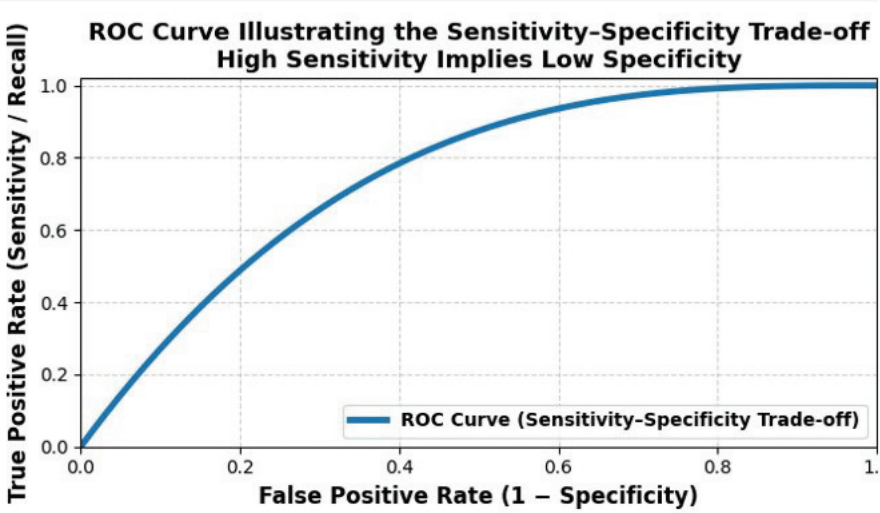

Figure 8.2 ROC curve (log-scaled ***FPR***).

the asymptotic barrier: attempting to saturate specificity (also referred to as selectivity) and sensitivity leads to a collapse in throughput rather than improved operational safety.

8.2.4 Statistical Impact and the Binomial Limit

8.2.4.1 Bernoulli/Poisson envelope

Even if recall is pushed arbitrarily close to 1, any non-zero per-defective escape probability implies a non-zero chance of at least one escape. Under a Bernoulli model (independent trials) [10], the probability of observing zero escapes decreases rapidly as the number of defectives tested grows; on large runs you will typically observe x $\geq$ 1 escapes. Hence, "zero-escape" claims cannot be certified with finite evidence.

8.2.4.2 Worst-case confidence bound (Clopper-Pearson, exact; per-unit DPPM)

For certification we use distribution-free one-sided Clopper-Pearson bounds. If escapes are observed over N_total shipped/tested units, let ***rU*** be the 95% upper bound on the delivered escape rate, report:

$$\mathrm{DPPM}_{95\%} = \boldsymbol{rU}\ 10^6 \tag{8.4}$$

Table 8.1 Clopper-Pearson bounds

Observed escapes	95% worst-case escapes (count)	DPPM95%
0	∼3 .00	∼ 3.00
1	∼ 4.74	∼ 4.74
2	∼ 6.30	∼ 6.30
3	∼ 7.75	∼ 7.75

Two practical cases (per-unit view):

- Zero events (x = 0) "rule of three": DPPM95% = $\mathbf{3} * 10^6$ / Ntotal , i.e. 3 **DPPM** when Ntotal = 10^6
- At least one event (x $>$ 0) bound rises above the "3 DPPM" floor

For large Ntotal, the 95% worst-case counts (then scaled to DPPM at Ntotal =10^6) are reported in Table 8.1. Such values closely match the exact Clopper-Pearson upper bound for large Ntotal. For 95% confidence, multiply the counts (and DPPM) by ∼1.5 (i.e. 3 -> 4.6).

Implications for certification *Bernoulli/Poisson* view explains why zero escapes are implausible at scale; the *Clopper-Pearson* bound provides the formal worst-case DPPM to report. As soon as x $\geq$1, the 95% bound exceeds the ∼ 3 DPPM floor of the zero-event case**.**

8.2.5 Avionic Safety Constraint (3 ppm Limit)

In avionics and safety-critical electronics, the allowable number of escaped defects is typically below three per million units (3 ppm). Let p denote the per-defective residual escape probability, i.e., the probability that a truly defective unit is misclassified as acceptable. For N_defective = 1.000, the expected residual number *E* of residual escapes must satisfy:

$$E[FN\text{residual}] = N_{\text{defective}} \cdot p \leq 3 \Rightarrow p \leq 0.003 \tag{8.5}$$

We may relate this bound to the classifier performance, by expressing:

$$\mathbf{p} = (\mathbf{1} - \mathbf{Precision})\boldsymbol{\epsilon} \leq 0.003 \tag{8.6}$$

where ϵ denotes the defect prevalence introduced above. Assuming ϵ = 0.01, the above constraint yelds a required **Precision $\geq$ 0.97.** Therefore, if Precision drops below 0.97, **even Recall = 1 is not enough** to meet safety standards. Moreover, boosting recall to 1 typically inflates ***FP*** to unsustainable levels. Figure 8.3 (Avionic Acceptance Zone) plots this safety threshold: only high-precision systems ($>$ 0.97) can meet avionic requirements even

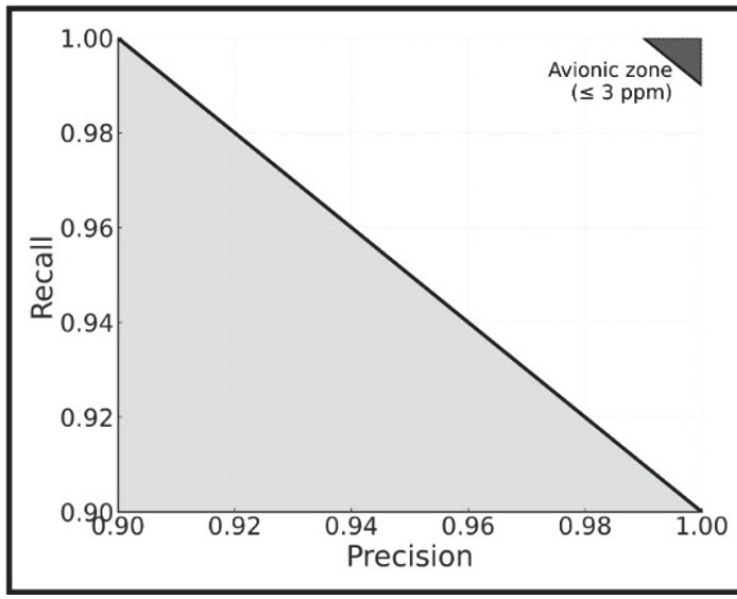

Figure 8.3 Avionic Acceptance Zone.

under perfect recall. This confirms that single-model AOI fails to deliver both safety and efficiency, especially in stringent sectors.

8.2.6 Empirical AOI Performance and Practical Limits

Real-world AOI systems, as reported in literature [3, 4], cluster around:

- Precision = 0.93-0.99
- Recall = 0.88-0.97

Recent surveys on these subjects, e.g., [9] show that even well-tuned single-stage AOIs cannot meet both perfect sensitivity and precision. Therefore, current AOI systems operate far from safety-optimal zones, validating the need for reflective multistage architectures like the one proposed in §3. This mainly depends on illumination, optics, surface complexity, and defect type.

8.2.7 Key Points

This section has shown that:

- Statistical leakage prevents hard guarantees of zero-defect escape

- Recall and precision are intrinsically in conflict
- For high-volume, safety-critical applications, single classifiers fail to meet both targets
- Achieving Recall ≈ 1 inflates false positives, blocking the line
- Conversely, tightening precision causes defect escapes, violating safety

Hence, a **modular, synergistic architecture with staged inspection logic** becomes necessary. Section 3 introduces such a design: a **three-stage AI pipeline** that balances sensitivity and selectivity across inspection layers.

8.3 Multi-stage Architecture and Statistical Stability in Industrial Processes

A robust way to detect rare defects at scale is to **stage** decisions: push **recall** high at the front gate (to protect safety), then **recover precision** downstream (to protect throughput). Unlike a one-way filter, the chain is **reflective**: later stages can correct early over-flags and recheck borderline "good" items.

8.3.1 AOI Architecture (T -> K1 -> K2)

In brief, the multi-stage AOI comprises three stages:

1. **T (front gate):** a high-recall filter that flags anything suspicious
2. **K1 (reject refinement):** re-analyses flagged items to recover good units and reduce false positives
3. **K2 (final gate):** re-checks the "good" stream to catch residual escapes (false negatives)

The flows among T, K1 and K2 are as follows:

- **From T** (front gate)
  - KO_T (predicted defective = ***TPT*** + ***FPT*** -> K1
  - OK_T (predicted good = ***TNT*** + ***FNT*** -> K2
- **At K1** (refines rejects from T)
  - OK_K1 (now predicted good = ***TNK1*** + ***FNK1*** -> K2
  - KO_K1 (still predicted defective = ***TPK1*** + ***FPK1*** -> scrap/repair (exits the system)
- **At K2** (final gate on the accepted stream)
  - KO_K2 (predicted defective = ***TPK2*** + ***FPK2*** -> scrap/repair
  - OK_K2 (predicted good = ***TNK2*** + ***FNK2*** -> ship to customer

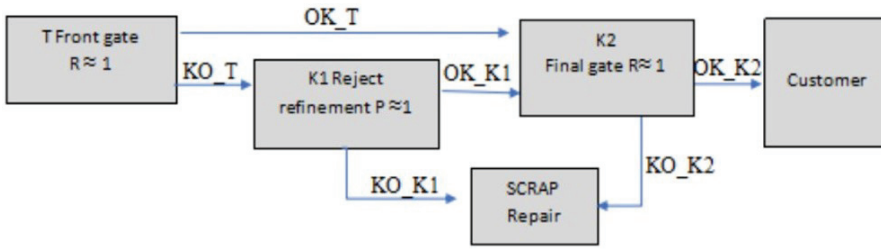

Figure 8.4 Block diagram of the three-stage AOI pipeline.

Table 8.2 Results (point estimates

Scenario	FP (good scrapped)	FN (escapes)	DPPM ($=$ FN at $N = 10^{\circ}$)
T only	**556**	**0**	**0**
T $+$**K1**	**50**	**100**	**100**
T $+$**K1** $+$ **K2**	**52**$(= 50 + 2)$	**1**	**1**

Figure 8.4 shows the AOI pipeline pointing out that escapes are the ***FNK2*** contained in OK_K2, K1 recovers good units from KO_T stream (cuts ***FP***), while K2 catches residual defects that passed through T (reduces ***FN***).

8.3.2 Point-estimate simulation (10^6 boards; prevalence 0.5%)

Because point estimates reflect real on-line behaviour, they may be used to reveal the superiority of the three-stage AOI over T-only and alternative configurations. This is illustrated by the example below.

Assumptions per stage

- T: Recall = **1.00**, Precision = **0.90**
- K1: Recall = **0.98**, Precision = **0.99**
- K2: Recall = **0.99**, Precision = **0.98**
- Population: N = 10^6, prevalence = 0.5%,

where prevalence is defined as the number D of defective boards divided by the total production volume, => D = 5.000 and good boards G = 995.000

Computation sketch:

- T: ***TPT*** = 5.000; ***FNT*** = 0; ***FPT*** $\approx$ 556
- K1 (on KO_T): ***TPK1*** = 4.900; ***FNK1*** = 100; ***FPK1*** $\approx$ 50
- K2 (on OK_T + OK_K1): defectives-in = 100 $\Rightarrow$ ***FNK2*** = 1; ***FPK2***$\approx$ 2

Result Interpretation

Table 8.2 is a deterministic operating estimate given nominal (***Recall, Precision***). For reporting/certification, it is suitable to use one-sided Clopper-Pearson bounds (see §2.4.2 and [11]):

- **T only** with x = 0 escapes cannot claim "0 DPPM"; the 95% floor is $\sim$ **3** DPPM at N = 10^6
- **T+K1** with x = 100 escapes implies a 95% upper bound $>$ 100 DPPM ($\sim$116 DPPM via Poisson/CP)
- **T+K1+K2** with x = 1 implies $\approx$ 4.74 DPPM95%

Table 8.2 shows that the multi-stage chain significantly reduces the false-reject burden. in our 1M/0.5% simulation ***FP*** drops by $\sim$10$\times$ versus T-only (556 -> 52 per million), while escapes fall to $\sim$1 DPPM by point estimate (T+K1+K2), instead of 100 DPPM with T+K1. Therefore, multi-stage AOI makes the operating risk tiny ($\sim$1 DPPM by point estimate) and the certifiable risk acceptably low ($\sim$4.7 DPPM). A single stage cannot substantiate a zero-escape claim (its 95% floor is $\sim$3 DPPM even if x = 0).

8.3.3 Keeping FN and FP under control (few, frequent actions)

Controlling ***FP*** and a few stable **process variables** is essential to keep the line statistically stable and to prevent hidden increases in ***FN***. Routine monitoring, paired with simple, pre-agreed actions, keeps day-to-day performance aligned with the one-sided 95% **Clopper-Pearson** bounds and prevents drift. Suggested actions (derived from the above):

- **Report two numbers (monthly + cumulative): Point DPPM** (what happened) and **CP-bounded DPPM95%** (what you can certify) is important for evaluating the status of the production line as will be discussed in the next section
- **Watch the FP trend:** simple per-shift FP chart with a stop-and-check rule (run a small golden set before changing thresholds)
- **Clear threshold split:** give **recall** to **T** (lower its threshold if FN risk rises) and **precision** to **K1/K2** (they filter the extra FP)
- **Weekly recall probe:** inject a few known-defect boards; if recall drops, lower threshold of T slightly and keep K2 strict until stable
- **Deterministic repeatability checks:** track **inter-board similarity (IBS)** and re-scan one unchanged board; if IBS falls or repeatability breaks, switch to a safe profile, re-run the golden board and correct illumination/focus/registration before resuming

These few controls keep the section actionable, tie directly to the formal bounds in §2 and avoid over-engineering.

8.4 Multi-stage AOI System and Statistical Control on Large-Scale Production

To meet ultra-low escape rates (automotive, avionics), we deploy a three-stage AOI that separates concerns: a high-recall front gate, a high-precision refinement on rejects and a final check on the "good" stream to catch residual escapes. All claims are reported with one-sided Clopper-Pearson bounds (see §2.4).

8.4.1 Annotation and classification-only strategy (no object detection)

In our implementation each component is pre-indexed from CAD/BOM and cropped into a fixed patch (one component per patch). This removes object detection and turns the task into pure classification ("conforming/non-conforming"). Benefits: higher accuracy (the model focuses on conformity), full traceability (ID/position known) and stability (fixed ROIs).

8.4.2 Three-stage pipeline and threshold policy

In details, the AOI used in our prototype *VisioEdgeAI* are as follows:

AOI-1 - Tiny-autoencoder Filter (High Recall)

A Tiny-autoencoder trained on golden boards detects any deviation from normality. Its threshold is adjusted to ensure **recall $\approx$ 1.0**, meaning that no defective component is missed, even at the cost of many false positives.

AOI-2 - KMeans Refiner (High Precision)

Only the flagged components are passed to an unsupervised method such as KMeans, which clusters them and isolates rare, subtle defect types. This stage acts as a **second filter**, drastically reducing the false positive rate by operating on a semantically aligned, low-dimensional space.

AOI-3 - Final Classifier (False Negative Recovery)

Despite the high recall of AOI-1, **residual false negatives** may still occur due to statistical effects or rare defect patterns. The third stage acts as a **fall back detector**, specifically optimized for **FN recovery** via a lightweight

classifier trained on known defect prototypes. This final step ensures that even statistically rare errors are caught with high probability. Threshold policy is to allocate recall headroom to T and precision headroom to K1 (they filter the extra FP). Stress P increases at K1 whereas K2 reinforces recall on critical cases, improving robustness to statistic variation.

8.4.3 Analytical comparison with certification hooks (single vs. multi-stage)

Given the practical scope of our study, this section explains the engineering meaning of the three KPIs used in our analysis, i.e. Point, Bernoulli and Clopper-Pearson, so that operational performance and certification claims are interpreted consistently.

- Point (operational estimate): FP/FN expected on $\boldsymbol{N}$ units from the nominal ($\boldsymbol{R}$, $\boldsymbol{P}$) at each stage (i.e. "what happened" in production)
- Bernoulli (likelihood at scale): probability of at least one escape given per-defective escape $\boldsymbol{PFN}$ = 1 - R and the number of defectives i.e. "how likely an escape is" under current setting. At production scale, the probability of exceeding a tolerable number of escapes increases as the probability that a single defect escapes increases.
- Clopper-Pearson (certification bound): one-sided 95% upper bound on the delivered escape rate from the observed escapes x on N units; reported as DPPM95% = rU 10^6 i.e. "what you can safely claim"

The following example (same setting unless noted) may clarify how they are computed in practice. Assume N $= 10^6$ boards, prevalence $= 0.5\%$ defectives D $= 5.000$, good G $= 995.000$.

1) Point (operational) - T only

Take RT = 0.98 and PT = 0.90

- $\boldsymbol{FN}$ = (1 - $\boldsymbol{RT}$) $*$ D = (1 – 0.98) $*$ 5000 = 100 DPPM
- $\boldsymbol{TP}$ = $\boldsymbol{RT}$ D = 0.98 $*$ 5000 = 4900
- $\boldsymbol{FP}$ = $\boldsymbol{TP}$ (1 - $\boldsymbol{PT}$) / D $\approx$ 544

2) Bernoulli (engineering likelihood of $\geq$1 escape)

Per defective escape PFN = 1 - R, with n defectives reaching the gate Probability of zero escapes: $(1 - \boldsymbol{PFN})^n$; hence $\boldsymbol{P}(\geq 1) = 1 - (1 - \boldsymbol{PFN})^n$

if $\boldsymbol{RT}$ = 0,999 => $\boldsymbol{PFN}$ = 0,001, n = 5000 then $\boldsymbol{P}(\geq 1) \approx 99{,}33\%$

if $\boldsymbol{RT}$ = 0,98 => $\boldsymbol{PFN}$ = 0,02, n = 5000 then $\boldsymbol{P}(\geq 1) \approx 100\ \%$ i.e., virtually certain at this scale

3) Clopper-Pearson (certification, 95% upper bound)

Report DPPM 95%= ***rU*** 10^6, where ***rU*** is the CP 95% one-sided upper bound on the escape rate given observed escapes x over N. For N = 10^6

- x = 0 $\Rightarrow \sim$ 3.00 DPPM (zero-event "rule of three")
- x = 1 $\Rightarrow \sim$ 4.74 DPPM
- x = 2 $\Rightarrow \sim$ 116 DPPM (CP one-sided 95%)

These examples clearly demonstrate that an AOI using T-only could be enough for efficient certified production only if R is close to 1.

However, another example is suitable for illustrating that this is not completely true neither under this extreme condition. Indeed, the following Table 8.3, obtained using the same previour formulas, certifies that three stage AOI achieves similar ***FN*** but at a lower ***FP***. In this experiment we assume. N= 10^6 boards; prevalence 0.5% => 5.000 defectives, good G = 995.000 ,

Nominal stage metrics:

- T: ***R*** = 1, ***P*** = 0,90
- K1: ***R*** = 0,98, ***P*** = 0,99
- K2: ***R*** = 0,99, ***P*** = 0,98

Before discussing the results, we first clarify the table's columns and the computation sketches (Point and Bernoulli) used in this experiment; the Clopper–Pearson computation sketch follows the description provided above.

- **Point** = operational estimate from (***R, P***)
- **Bernoulli** (engineering view; [10]): probability of observing at least one escape at the stage that determines delivery (K1 if no K2; K2 if present), computed with n defectives reaching that stage and per-defective escape probability ***PFN***. It shows how likely an escape is in practice under the assumed operating point
- **Clopper-Pearson** (certification view; [11]): one-sided 95% upper bound on delivered DPPM from the observed count x of escapes (per §2.4.2): with x = 0 the bound is the familiar $\sim$3 DPPM floor at 10^6; with x = 1 it becomes $\approx$ 4.74 DPPM; with x = 100 it is $\approx$ 116 DPPM

Table 8.3 Results (point, Bernoulli, Clopper-Pearson)

Scenario	FP (point)	FN (point)	Point DPPM	Bernoulli Pr [$\geq$ 1 escape]	CP-bounded DPPM (95%)
T only	556	0	0	0%	$\approx$ 3.00
T + K1	50	100	100	$\approx$ 100%	$\approx$ 116
T + K1+ K2	2	1	1	$\approx$ 63%	$\approx$ 4.74

Computation sketch (Point)

- T: ***TPT*** = 5.000 ***FNT*** = 0 ***FPT*** $\approx$ 556
- K1: ***KOT TPT*** = 4.900 ***FNK1*** = 100 ***FPK1*** $\approx$ 50
- K2: ***OKT*** + ***OKK1*** defectives-in = 100 ***FNK2*** = 1 ***FPK2*** $\approx$ 2

Computation sketch (Bernoulli)

- T: 0% (since ***RT*** = 1)
- K1: n = 5.000, ***RT*** = 0,02 ***PFN*** = 1 - 0,98 ^{5000} => ***PFN*** $\approx$ 100 %
- K2: n = 100, ***RT*** = 0,01 ***PFN*** = 1 - 0,99 ^{100} => ***PFN*** $\approx$ 63 %

With that in mind, the table confirms that a T-only system cannot occupy the upper-right (high R, high P) corner at scale: raising ***PT*** lowers ***FP***, but any ***RT*** < 1 generates ***FN***; no single threshold achieves both tiny ***FN*** and tiny ***FP***. Adding K1 (emphasizing precision) recovers most ***FP***; adding K2 (emphasizing recall) collapses residual ***FN*** with negligible ***FP*** increase. Therefore, a multi-stage design is necessary, both operationally (point estimates and formally Clopper-Pearson bounds). As outlined in §3, it is also clear that reporting these two numbers monthly and cumulatively may give clear indication about maintenance. Point DPPM clarifies "what happened" and CP-bounded DPPM (95%) points out "what you can certify".

The monthly estimate captures drift; the cumulative bound tightens with volume (approximatively $3 * 10^6$ / Ncum) in the zero-event case), thus aligning operational monitoring with certification requirements and avoiding overclaims after quiet months.

Fixed per-component crops make the decision about conformity only; this tightens feature distributions and improves both ***P*** and ***R*** versus detection-based pipelines. We then stress stages by design: K1 toward precision on the reject stream P increases, K2 toward recall on the accepted stream ***R*** increases. The result is a stable operating point with FP $\ll$ T-only (i.e. 556 -> $\sim$52 per million) and escapes $\sim$ 1 DPPM by point estimate, paired with defensible CP bounds per §2.4.

8.5 Concluding remarks

The approach detailed in this paper has been adopted in the final design of the robotized conveyor AOI called *VisioEdgeAI*, both in the choice of a three-stage processing pipeline while boards pass under the camera and in the systematic reduction of variability sources that could jeopardize the targets.

Three-stage AOI is on the line. As each PCBA traverses the imaging station, the system executes: T (high-recall front gate), K1 (precision-oriented refinement of rejects) and K2 (recall-oriented final check on the accepted stream). This triplex design replaces earlier two-step error-recognition schemes and proved necessary to achieve low escapes with low false rejects at production scale. Two-step AOI remains a valid option for consumer-grade electronics lots where the risk and volume profile are less stringent [12]. Variability control by design has been taken into account. To minimize context variability:

- Detection-free, classification-only processing: components are pre-indexed from CAD/BOM; fixed crops (one component per patch) remove localization noise and stabilize the feature space
- Pose/placement tolerance: exact mechanical centering of the PCBA is not required; pre-processing alignment registers each image to a stable reference before inference
- Illumination stability: per-shift and daily checks track inter-board brightness drift; thresholds switch to a safe profile if limits are exceeded, followed by recalibration

Compute platforms are another aspect under study to achieve good time performance. Real-time execution is feasible on embedded GPU modules (i.e. NVIDIA Jetson Orin/TX2-class); FPGA-based pipelines are under evaluation to further reduce latency by exploiting data-parallel primitives in the image pipeline.

Algorithmic targets and tuning is under study. K1 and K2 are being characterized to deliver baseline $R \geq 0.98$ and $P \geq 0.98$ at nominal settings, enabling controlled "stress" toward P increasing at K1 or R increasing at K2 without materially penalizing the other metric. Early runs on small lots showed zero observed escapes, consistent with the design. Further classification algorithms are also under study to improve performance and assure a certain degree of statistical independence among them [13].

The dataset used in these preliminary runs was obtained from a production board. From images of this board, we extracted per-component image patches and generated PCBAs with representative defect conditions (e.g., missing components, rotations, burn marks, and overstressed or aged components). Approximately one hundred PCBAs of the same type were provided by HTS, with whom we are collaborating on the development of the VisioEdgeAI prototype. Each board contains around 40 components (resistors, capacitors, inductors, and integrated circuits).

The training dataset comprises 500 PCBA images derived from the best-quality image of the original HTS board. Each defective image includes four defective components, covering a range of defect types such as rotated components, missing components, burned components, imperfect solder joints, and incorrect components (i.e., components mounted in the wrong locations). Testing was conducted using defective images not included in the training set, as well as HTS boards on which defects were intentionally introduced.

Defect modeling and data augmentation will be used to evaluate long batches suitable for false-negative (FN) certification, which will follow the one-sided Clopper–Pearson framework discussed in §2 (i.e., zero observed events still imply a finite upper bound).

The paper stressed how maintenance and certification integration is also one important aim. To achieve both operational stability and certification credibility, we aim at integrating systematic/predictive maintenance into the imaging-conveyor chain and report both Point DPPM and 95% CP-bounded DPPM within an IATF QMS. Component maintenance, very relevant in this context (i.e. [14], reduces drift in illumination, focus and transport dynamics) primary drivers of FP/FN variance in digital inspection, while certification framing, i.e. [15], aligns internal control with customer-facing, audit-grade claims.

On the algorithmic side, patch-based anomaly detectors (i.e. *PatchCore*; *PaDiM*) and self-supervised backbones (i.e. *DINO-style ViTs*) will be considered to provide high-recall, high-precision building blocks; for imbalanced supervised stages, Focal Loss can further mitigate false decisions. Taken together, these elements, triplex AOI, detection-free classification, engineered variability control and certification-grade reporting, compose a coherent system that meets the operational and formal requirements set out in this work. The resulting architecture is explicitly designed to decouple safety from throughput, allowing recall to be pushed toward unity without inducing unsustainable false-positive rates, while preserving process capability and takt-time stability at production scale.

These conclusions are further supported by recent few-shot SMD inspection benchmarks, such as the PCB-SAID dataset [16], which empirically documents how rare and fine-grained defect distributions expose the practical limits of single-model AOI systems. In this sense, PCB-SAID provides complementary experimental evidence that aligns with the statistical and process-level arguments developed in this paper, reinforcing the need for multi-stage, reliability-oriented inspection pipelines in safety-critical electronics manufacturing.

Acknowledgements

This research was conducted as part of the EdgeAI "Edge AI Technologies for Optimised Performance Embedded Processing" project, which has received funding from Chips JU under grant agreement No 101097300. The Chips JU receives support from the European Union's Horizon Europe research and innovation program and Austria, Belgium, France, Greece, Italy, Latvia, Netherlands and Norway.

References

[1] Berthold Horn, Robot vision. Cambridge, Mass.: Mit Press; New York, 1986.

[2] J. Wu, J. M. Rehg, and M. Mullin, "Learning a Rare Event Detection Cascade by Direct Feature Selection," Advances in Neural Information Processing Systems, vol. 16, 2025, Accessed: Dec. 14, 2025. https://papers.neurips.cc/paper/2353-learning-a-rare-event-detection-cascade-by-direct-feature-selection

[3] X. Guo, J. W. Gichoya, S. Purkayastha, and I. Banerjee, "CVAD: A generic medical anomaly detector based on Cascade VAE," arXiv.org, 2021. https://arxiv.org/abs/2110.15811

[4] Masoud Jalayer, Reza Jalayer, Amin Kaboli, C. Orsenigo, and C. Vercellis, "Automatic Visual Inspection of Rare Defects: A Framework based on GP-WGAN and Enhanced Faster R-CNN," arXiv (Cornell University), Jan. 2021, https://doi.org/10.1109/iaict52856.2021.9532584.

[5] G. Grosso et al., "Triggerless data acquisition pipeline for Machine Learning based statistical anomaly detection," arXiv.org, 2023. https://arxiv.org/abs/2311.02038

[6] K.-J. Wang, H. Fan-Jiang, and Y.-X. Lee, "A multiple-stage defect detection model by convolutional neural network," Computers & Industrial Engineering, vol. 168, p. 108096, Jun. 2022, https://doi.org/10.1016/j.cie.2022.108096.

[7] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," Mechanical Systems and Signal Processing, vol. 115, pp. 213–237, Jan. 2019, https://doi.org/10.1016/j.ymssp.2018.05.050.

[8] D. Luo, Y. Cai, Z. Yang, Z. Zhang, Y. Zhou, and X. Bai, "Survey on industrial defect detection with deep learning," Zhongguo kexue, vol.

52, no. 6, pp. 1002–1002, Jun. 2022, https://doi.org/10.1360/ssi-2021-0336.

[9] C. Cantone, A. Faro, "An Overview of the Automated Optical Inspection (AOI) Edge AI Inference System Solutions," in Advancing Edge Artificial Intelligence - System Contexts, River Publishers, 2024, pp. 153-176

[10] J.K. Blitzstein and J. Hwang, Introduction to Probability, Second Edition. CRC Press, 2019.

[11] M. Thulin, "The cost of using exact confidence intervals for a binomial proportion," Electronic Journal of Statistics, vol. 8, no. 1, pp. 817–840, 2014, https://doi.org/10.1214/14-ejs909.

[12] R. Faro, A. Strano, F. Cancelliere, "Discovering and classifying digital and wooden industries products' defects at the edge by a yolo/resnet-based approach and beyond," European Conference on EDGE AI Technologies and Applications (EEAI 2024), Cagliari, 2024.

[13] K. Roth, Latha Pemula, J. Zepeda, Bernhard Schölkopf, T. Brox, and P. V. Gehler, "Towards Total Recall in Industrial Anomaly Detection," 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2022, https://doi.org/10.1109/cvpr52688.2022.01392.

[14] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," Mechanical Systems and Signal Processing, vol. 20, no. 7, pp. 1483–1510, Oct. 2006, https://doi.org/10.1016/j.ymssp.2005.09.012.

[15] AS9100 (9100:2016) - Quality Management Systems - Requirements for Aviation, Space and Defense Organizations (IAQG/SAE).

[16] R. Mineo, A. Sorrenti, R. Faro, G. Mineo, F. Cancelliere, A. Faro, "PCB-SAID: A Low-Cost Camera-Based Dataset for Few-Shot SMD Assembly Inspection", 2025 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 1351-1357, Oct 2025.