Retrospective check (demo): retinal organoids, Afting et al. 2026

Scope: medaka (fish) retinal organoids, imaged 72 h every 30 min; outcome (RPE at the end of culture) assessed ~72 h after the first image; RPE visible by eye from ~44 h

Data trust check: can the headline numbers be trusted? With caution.

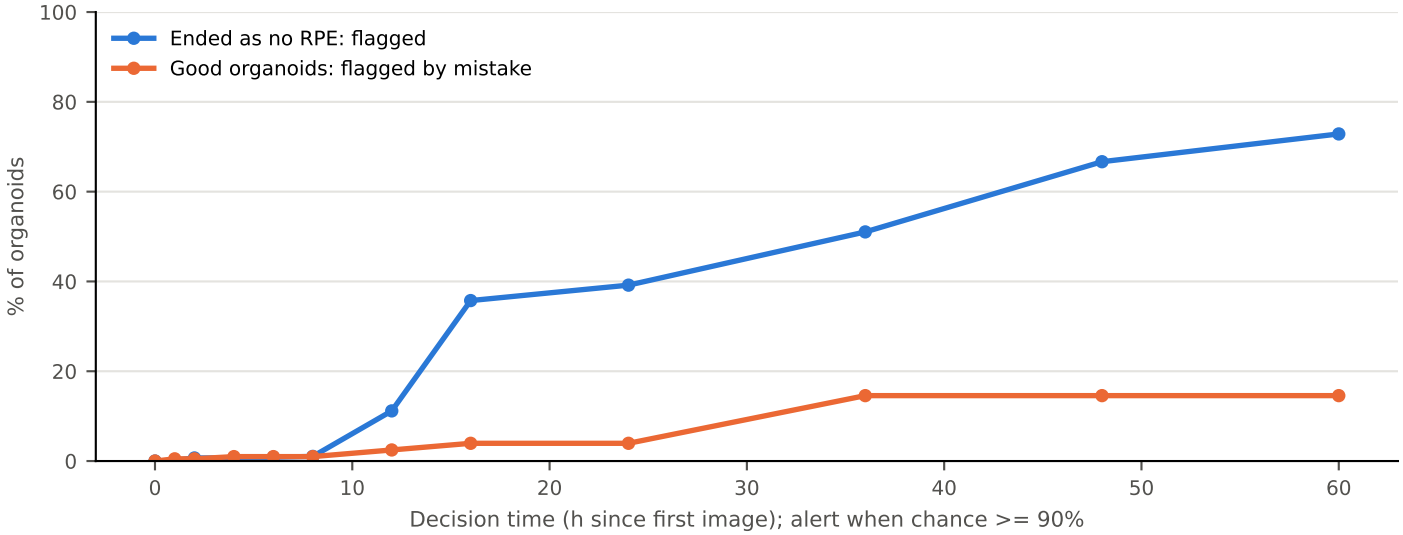
No clear shortcut, but some checks raised warnings. Read the headline numbers with the points below in mind.

Not checkable (not recorded): plate, imaging session, batch, treatment, cell line, well

Bottom line

Of 582 organoids that ended as no RPE, 424 (73%) would have been flagged, typically 48 h before the outcome was assessed. 59 of 405 good organoids (15%) would have been flagged by mistake. About 9 in 10 alerts would have been right (for comparison, 59% of all organoids ended as no RPE, so flagging at random would be right about 6 in 10).

Before 12 h the predictions were no better than chance; alerts raised earlier than that are noise and should be ignored.



Likely fails / likely succeeds / uncertain (measured on held-out experiments)

Decision time	Likely fails	Likely succeeds	Uncertain	Right when called	Coverage
0 h	5%	5%	90%	30%	93%
2 h	11%	9%	79%	58%	91%
6 h	7%	8%	85%	45%	92%
12 h	38%	26%	37%	86%	91%
24 h	35%	27%	38%	85%	90%
48 h	50%	28%	22%	87%	90%
60 h	52%	45%	4%	92%	92%

Target coverage 90% (split conformal). 'Uncertain' means the images support neither call. Coverage holds only for new experiments that resemble the ones used here. Where 'Right when called' is below 70%, treat the calls as uncertain.

Limits

- A look back at finished experiments, not a prospective trial.
- The highlighted alert setting was chosen on this same data, so its numbers are slightly optimistic.
- Fish organoids from one lab: this demonstrates the report, not performance on human organoids.
- The authors' images were already segmented and masked by their own model before release, which makes shape measurement easier than on raw brightfield images.
- RPE becomes visible around 44 h into imaging; alerts after that are closer to detection than forecasting.

987 organoids, 11 experiments/batches; each experiment predicted by a model trained only on the others, using only images up to each decision time. Full detail: retrospective_report.html.