How littlEye AI learns — and why it sometimes answers wrongly

This short guide explains how to read the model's results, why the model occasionally states nonsense with great confidence, and how to teach it with your own images, step by step, so that it improves exactly where you need it to.

1 · Why the model called normal lung “Heart 100 %”

A real-world experience: a web image of normal lung → the model said Heart 100 %; an image of grey hepatization in lobar pneumonia → Liver 81.8 %. This is not a random malfunction — there are three understandable reasons behind it:

- Scale. If a model is trained on patches at roughly the 20× objective (field of view ~0.5 mm). An overview image of a whole lobe at low magnification is a completely unknown world to the model — alveoli appear as fine texture rather than the structures it knows.
- The model MUST pick one of its classes. “100 %” does not mean the certainty of a pathologist — it only means “of everything I know, this most resembles heart”. The model cannot say “I don't know” or “I have never seen this”.
- The model knows only what was in its training. If the dataset held normal histology only, Grey hepatization — an airless lobe packed with cells — logically resembles liver to the model: dense, cellular, without alveolar spaces. The model has never seen pneumonia, so it picked the nearest thing it knows.

> **The golden rule**
>
> The model can only assess what it was trained on: the same scale (~20× objective; in practice a microphotograph at a total magnification of 100×–200×), the same stain (H&E;) and microphotograph quality — not a web JPEG. The more your image differs from the training images, the less the result means.

2 · How learning works — a dataset, not magic

A Core ML model cannot be fine-tuned — a finished model is baked. Learning in littlEye AI therefore works through the dataset: everything you want the model to learn is saved as a labelled image, and the model is retrained from scratch on the whole dataset. It takes minutes and has a great advantage: every new model comes with honest, measurable metrics on a held-out test.

3 · The learning loop — step by step

1. Evaluate an image (Classify image → From photos / From files).
2. Disagree with the result? Below the result there is a section “Learn from this image”: choose the correct organ and diagnosis and tap Save to dataset. The image is stored together with the correct answer.
3. In the organ's dataset, open the image detail and fill in the magnification and stain; if the image contains several findings, mark the location of the finding with a crop — the crop is then used for training.
4. Once you have collected at least 5 images per class this way (we recommend 30+), open Training and retrain the model. You may delete the older model of the same name.

5. Check the test % and the confusion matrix — you will see exactly what the model still confuses, and therefore which images to add next time.

This IS the learning

The loop evaluate → correct → save → retrain → verify is precisely how the app learns and remembers your corrections. With every iteration the model becomes a little more yours — and by sharing the dataset (.le3set) your corrections reach your colleagues as well.

4 · Editing diagnoses after training

- Add a new diagnosis: in the organ section tap Add next to “Diagnoses (training classes)” — the new class is immediately available for labelling and for the learning loop.
- Reassign an image: in the image detail simply choose a different diagnosis — the change takes effect at the next training run.
- Diagnoses bind to a model only through training. A finished model has its classes baked in — renaming or adding a diagnosis in the dataset changes only the next model you train.

5 · What to realistically expect from how many images

Images per class	What to expect
5–10	the model gets going, but treat the results as indicative only
30+	a usable working model on your own data
100+ from different cases	more stable behaviour also on slides from elsewhere
+ always	a “Normal” class, one magnification, one stain, validation on independent images

Where to start without a baseline model

Real diagnostic power comes only from a model trained on your own slides through the learning loop above. Recommended approach: pick one organ you actually assess, fill it with your own images (Normal + pathologies from your autopsies) and train. The first usable model arrives sooner than you expect — three different cases per class is enough to begin.

The Michigan-derived models — deleted

We have deleted the models derived from the Michigan collection. We asked the University of Michigan for a position and the answer was that our use is commercial and the CC BY-NC-SA licence does not cover it, even where the images are not part of the app. We accepted that position. The same applies to models derived from TCGA-UT. The workflow above works just as well without them.

Ján Šikuta, MD, PhD. · forensika.eu@icloud.com · littlEye AI 2.2.3 · August 2026