# PatchCamelyon (PCam) Dataset Overview

## 1. Introduction

PatchCamelyon (PCam) is a benchmark dataset for binary classification in digital histopathology. It derived from the Camelyon16 challenge dataset and focuses on detecting metastatic cancer in l node tissue sections.

The dataset is specifically designed for patch-level classification rather than whole-slide analysis, mak computationally efficient and suitable for deep learning experimentation.

---

## 2. Dataset Structure

Each sample in the PCam dataset consists of:

1. A 96 × 96 RGB image patch.
2. A binary label:
3. 0 → Benign (no tumor tissue present in the central region)
4. 1 → Malignant (tumor tissue present in the central region)

The label is determined based on whether metastatic tissue is present in the central 32 × region of the patch.

---

## 3. Dataset Statistics

1. Total images: Approximately 327,680 patches.
2. Training set: Approximately 262,144 images.
3. Validation set: Approximately 32,768 images.
4. Test set: Approximately 32,768 images.
5. Image resolution: 96 × 96 pixels.
6. Color format: RGB (3 channels).
7. Balanced binary classification task.

---

## 4. Medical Context

The images are extracted from whole-slide images (WSIs) of lymph node sections stained using h and eosin (H&E). These WSIs are extremely high resolution (gigapixel scale), and PCam provides cropped patches from these slides.

The objective is to detect small metastatic regions, which makes the task a fine-grained visual problem. The dataset is commonly used to evaluate image classification models in medical AI research.

---

## 5. Research Significance

PCam is widely used because:

1. It provides a realistic medical classification problem.
2. It is small enough for rapid experimentation.
3. It supports benchmarking of deep learning models.
4. It is suitable for transfer learning studies.
5. It enables research in hybrid classical–quantum machine learning systems.

---

## 6. Relevance to Quantum Machine Learning Pipelines

In hybrid classical–quantum workflows:

1. A classical convolutional neural network (CNN) extracts feature embeddings.
2. The embedding vector can be normalized.
3. The normalized vector may be interpreted as a quantum state representation.
4. Quantum classifiers can then perform inference using engineered observables or interference-based scoring.

For example, a `32`-dimensional embedding corresponds to a `5`-qubit quantum state (since `2^5 = 32`).

This makes PatchCamelyon suitable for experimental evaluation of quantum-enhanced classification models.

---

## 7. Conclusion

PatchCamelyon is a compact, well-structured medical imaging dataset designed for binary metastas detection at the patch level. Its manageable size, clear labeling strategy, and medical relevance standard benchmark in both classical and emerging quantum machine learning research.