Train a small model
- consider size vs. accuracy trade-offs.
- use only supported operations
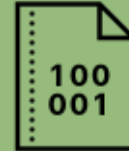
Convert and optimise
- with TensorFlow Lite Converter
- optimise size and speed
- evaluate accuracy

Convert to C byte array
- use xxd to create hexdump
- generate a .c and .h file

Run inference
- Use C++ api to run the model with TensorFlow Lite Interpreter on a microcontroller.