



Department of Computer Science & Engineering
The National Institute of Engineering, Mysuru

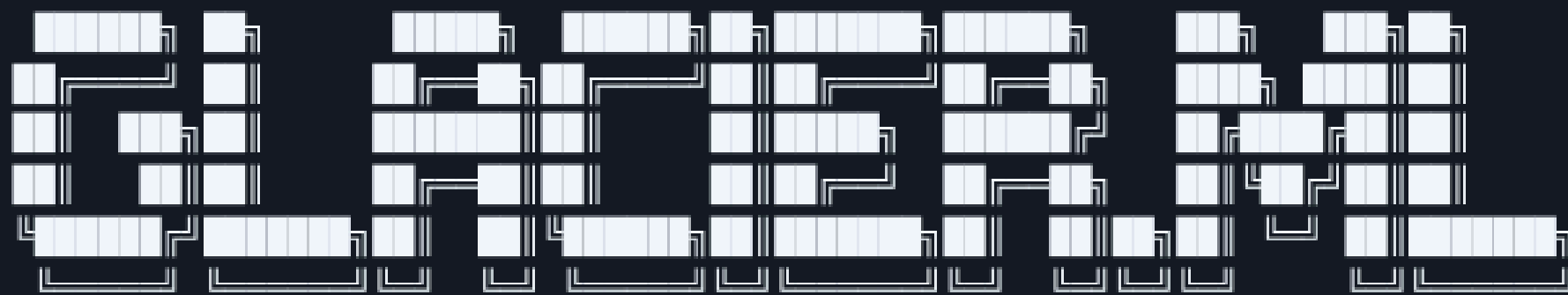
GLACIER.ML

November 28, 2025

Batch: F11

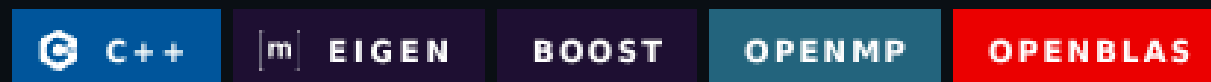
Team: Skandan C.Y (4NI23CI103)

INTRODUCTION



Languages and Frameworks used:

Core stack:



Development and Profiling:



Glacier.ML is a header only Supervised Machine Learning library, built entirely using C++.

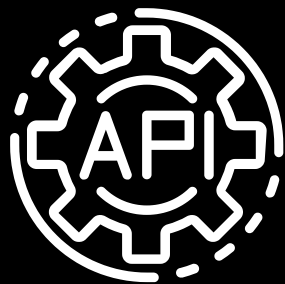
Aims to be a lightweight, fast alternative to **Scikit-learn** in the world of **C++**.

Currently, it includes seven models, with four more in development. Benchmarking shows Glacier.ML consistently outperforms Scikit-learn models in execution speed, while maintaining comparable predictive accuracy.

Link: <https://github.com/skandanyal/Glacier.ML>

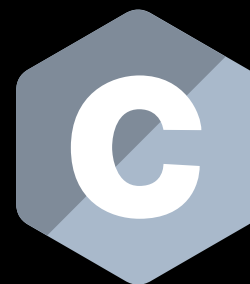
EXISTING SYSTEM

SCIKIT LEARN



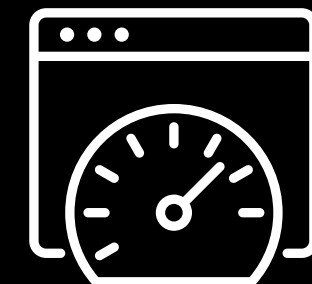
A Production-Grade Pythonic API:

- **High level, industry grade** Python API, used extensively throughout the **world**.
- Contains robust and easy documentation, enabling fast prototyping and useage.



Highly-Optimized C, Cython Core:

- Majorly uses NumPy and SciPy, which themselves are powered by C/C++, followed by Cython and C in performance critical regions and Joblib to implement parallelization.
- Code is highly optimized by numerous researchers and professionals over the years.

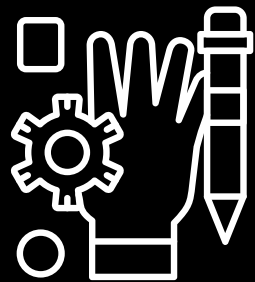


Pythonic Overhead, Dependency hell and Limited Low-Level Efficiency

- Heavy Python overhead: dynamic wrappers and dispatch add latency before C/C++ code runs.
- Limited low-level control: data copies, GIL constraints, and generic parallelism reduce efficiency on large workloads.

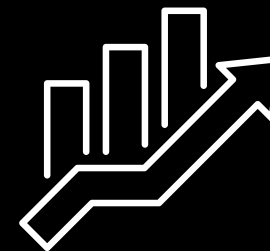
PROPOSED SYSTEM

GLACIER.ML



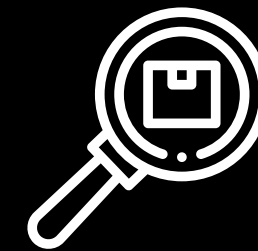
Zero Abstraction, Direct Control of Memory & Threading

- Full compiler visibility via header-only C++20 (inlining, specialization) resulting in near-zero overhead, no interpreter or dynamic layers.
- Explicit authority over data layout, vectorization, cache behavior using **OpenMP** and **OpenBLAS** without Python-level constraints



From Naivete to HPC research

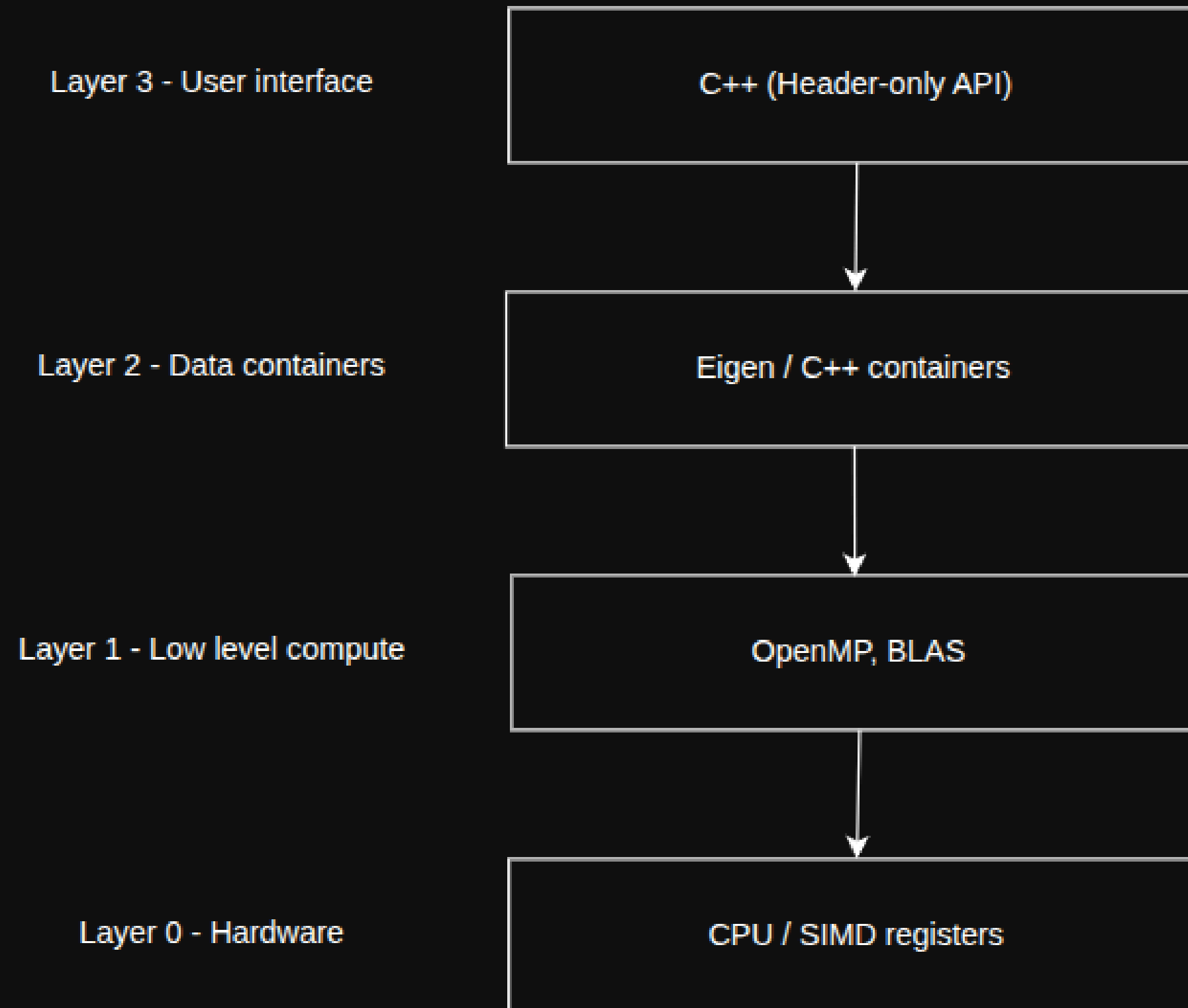
- KNN, once **200x** slower than Scikit-learn, now runs only about **4x** slower, while SVM achieves **4-10x** faster execution.
- Demonstrates a clean path for further research: every gain came from transparent, controllable optimizations, proving the stack can be systematically pushed toward state-of-the-art performance.



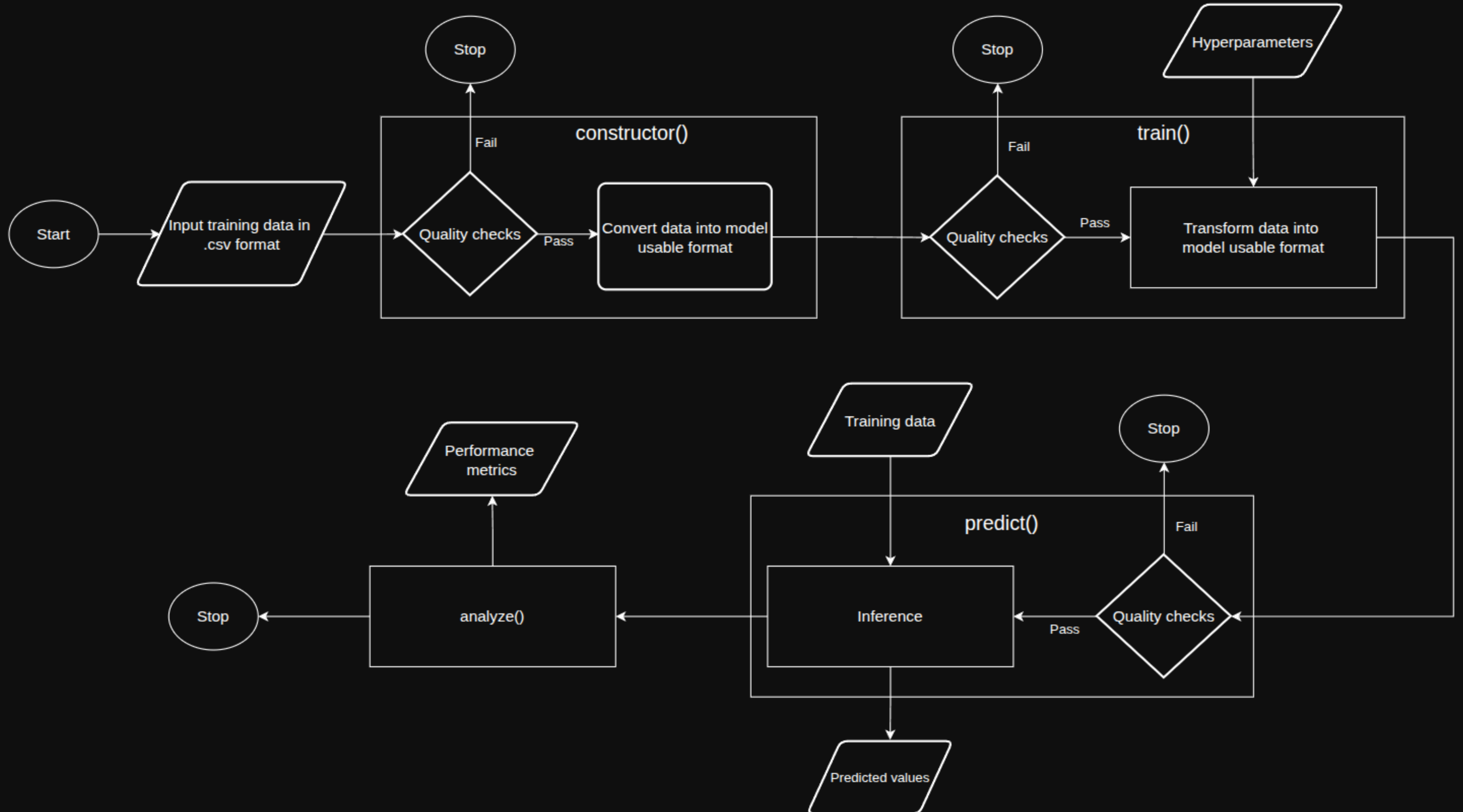
Full Inspectability & Replaceability

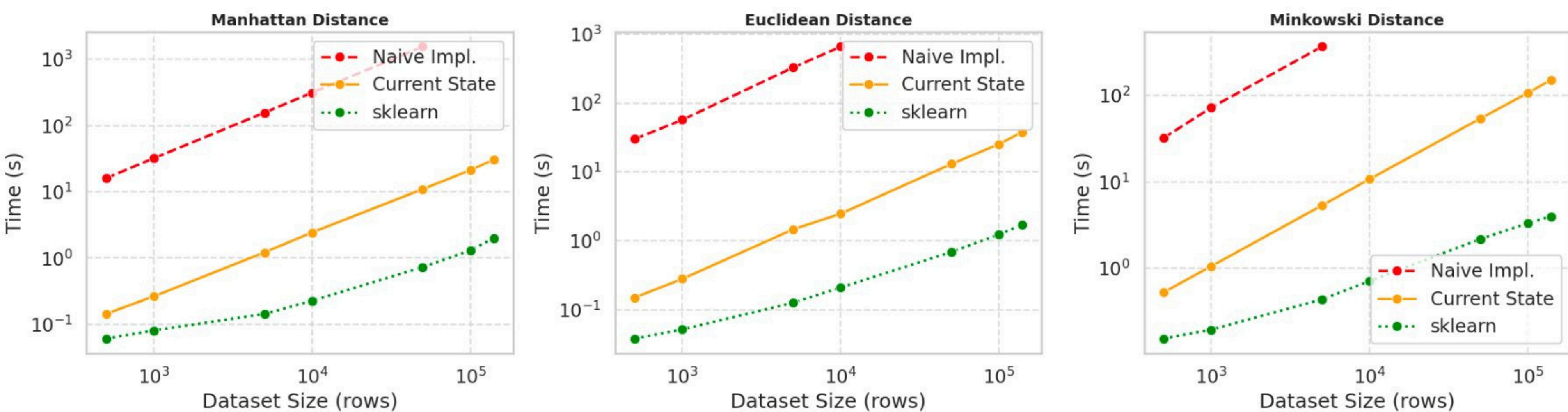
- Each component is directly readable and modifiable with no framework indirection.
- Algorithms, kernels, data layouts, and threading strategies can be rewritten or replaced without legacy constraints.

04. SYSTEM DESIGN

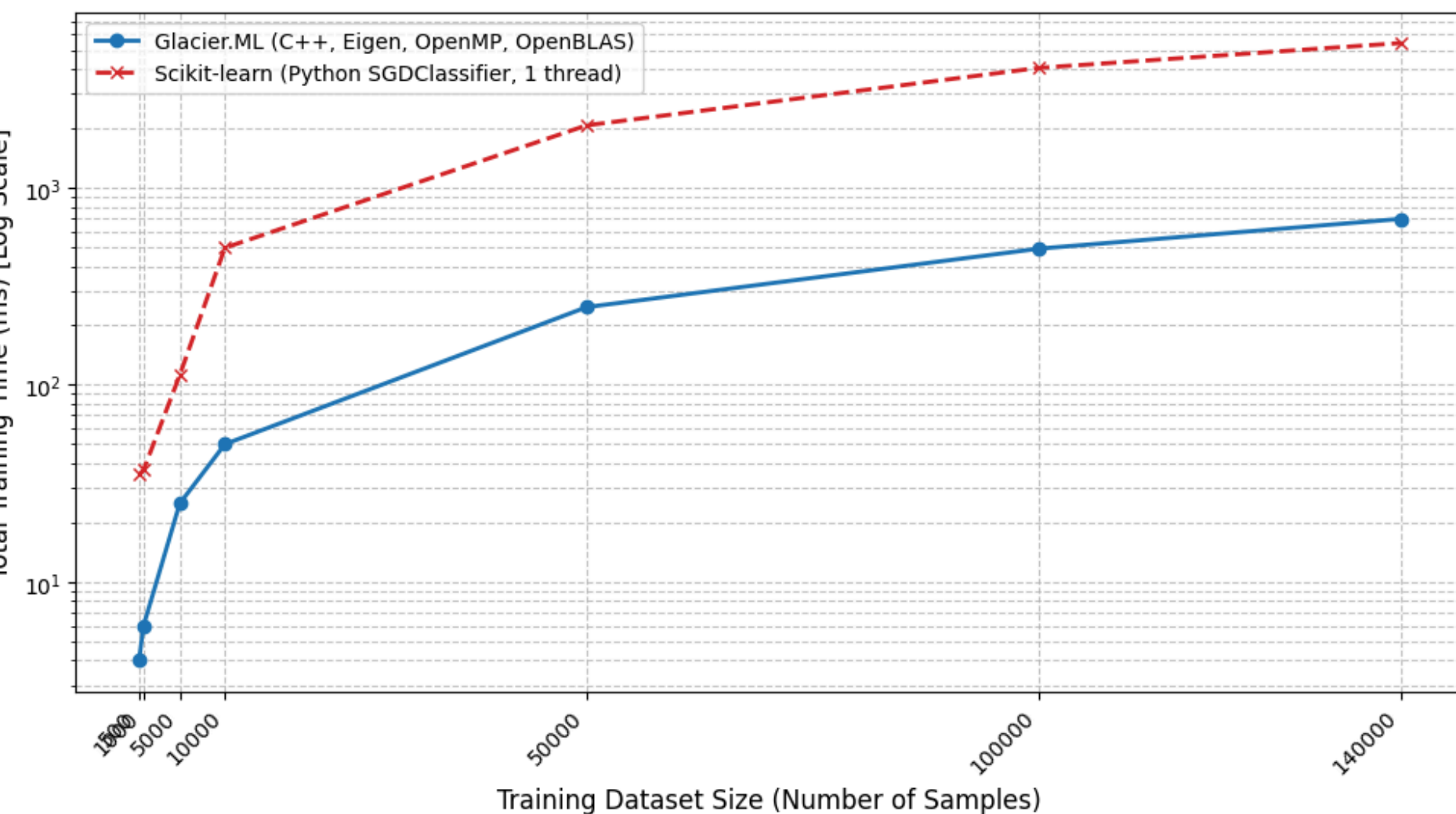


CONTROL FLOW DIAGRAM

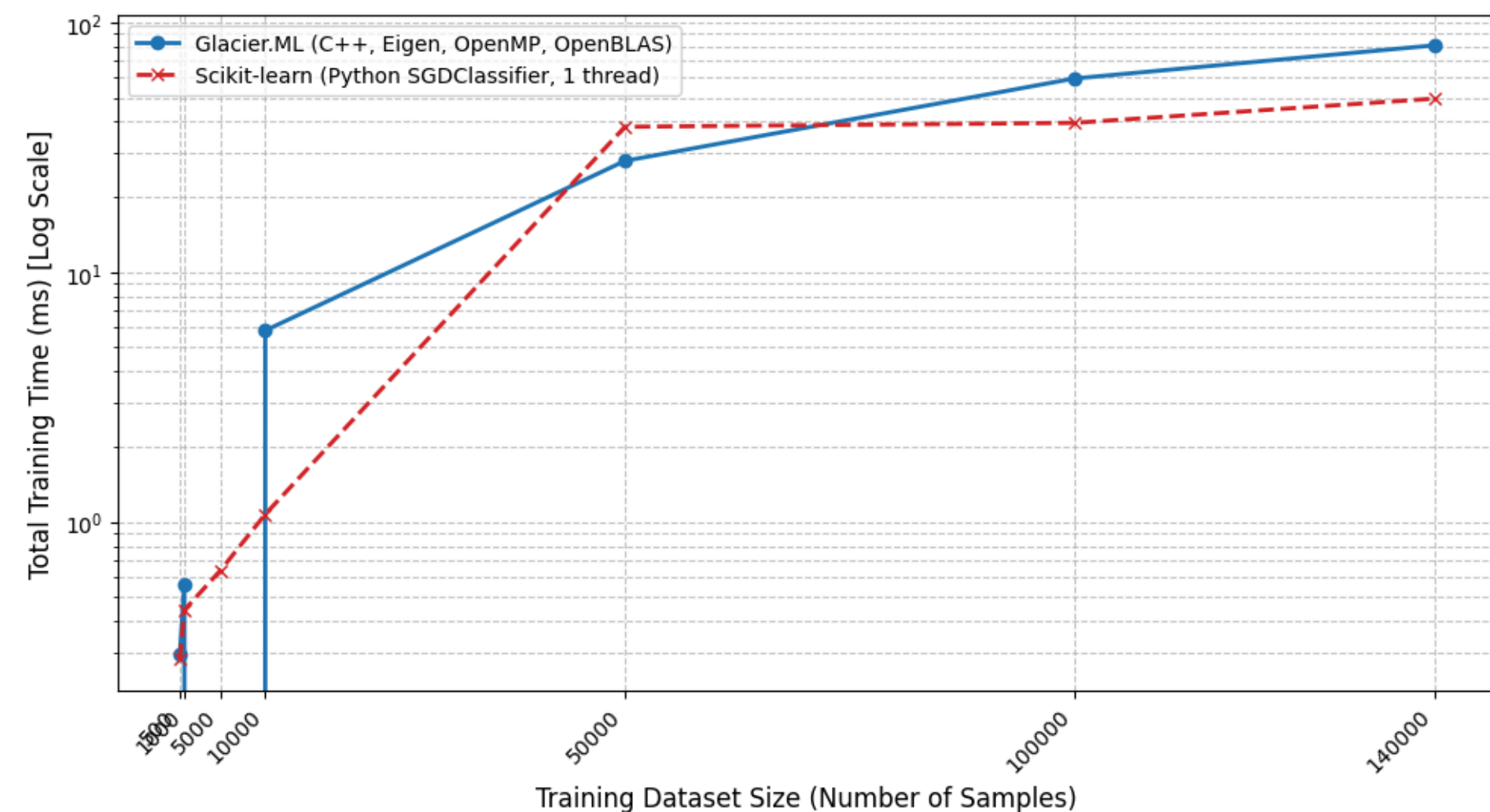




Performance Benchmark: Glacier (C++) vs. Scikit-learn (Python)



Performance Benchmark: Glacier (C++) vs. Scikit-learn (Python)



THANK YOU!