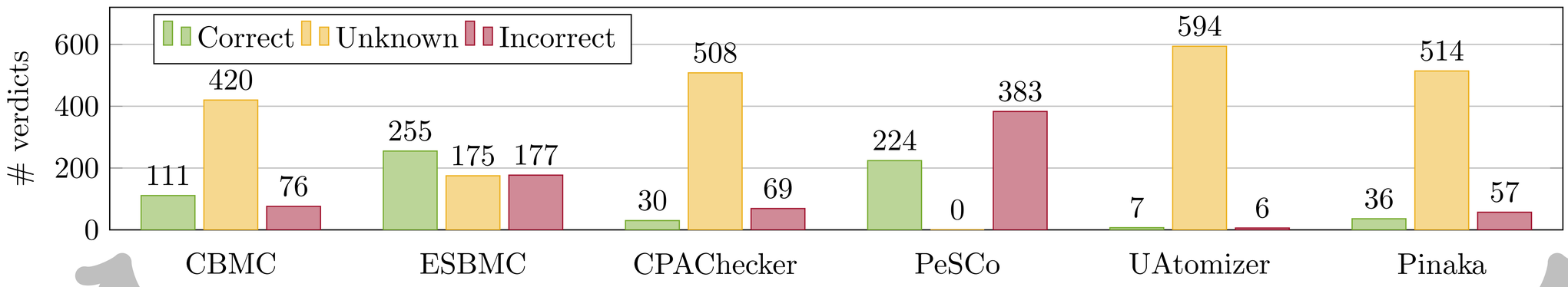
NeuroCodeBench: a plain C neural network benchmark for software verification

Edoardo Manino, Rafael Sá Menezes, Fedor Shmarov, Lucas C. Cordeiro
The University of Manchester, Manchester, United Kingdom and Universidade Federal do Amazonas, Manaus, Brazil



Incorrect verdicts due to incomplete support of `math.h`

Existing software verifiers struggle on neural code!

Unknown verdicts due to timeouts on large instances

Neural code challenges existing software verifiers:
float operations, calls to `math.h`, nested loops, multi-dimensional arrays.

We release *NeuroCodeBench*, a benchmark of neural code verification:
6 categories, 14 functions from `math.h`, 32 neural networks, 607 properties;
safe/unsafe verdicts are either known a priori or independently verified.

Benchmark Category	Safe	Unsafe
`math_functions`	33	11
`activation_functions`	40	16
`hopfield_nets`	47	33
`poly_approx`	48	48
`reach_prob_density`	22	13
`reinforcement_learning`	103	193
Total	293	314

Paper submitted to the 2023 AFRiTS workshop. Find out more at `github.com/emanino/plain_c_nn_benchmark`