

# Is Mutation Score a Fair Metric?

Beatriz Souza

Federal University of Campina Grande – Brazil – beatriz.souza@ccc.ufcg.edu.br

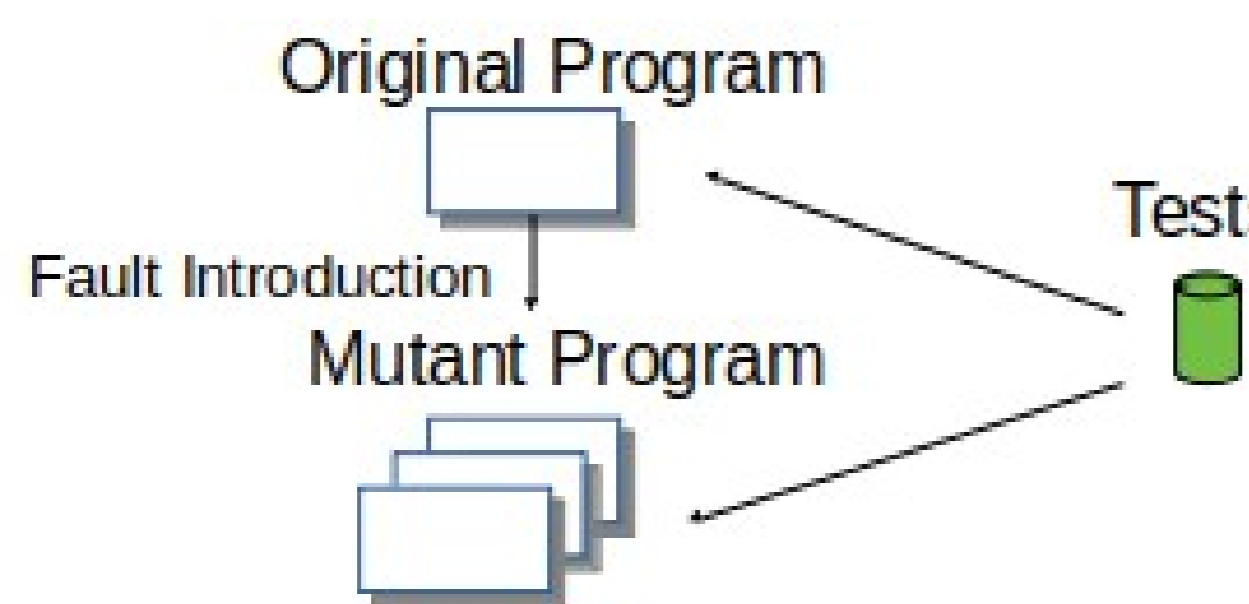
## Problem

- It is not known whether the mutation score is a fair metric to compare test suite effectiveness.
- Consequently, a number of studies using mutation testing may be invalid.

## Study

- Empirical study with manually written and automatically generated test suites, a set of 10 programs and 7 mutation operators to compare their detection ratios with the mutation score.

## How Does Mutation Testing Work?



Mutation Operations Examples	
Name	Transformation Example
Math	$x += n \longrightarrow x -= n$
Increments	$x++ \longrightarrow x--$

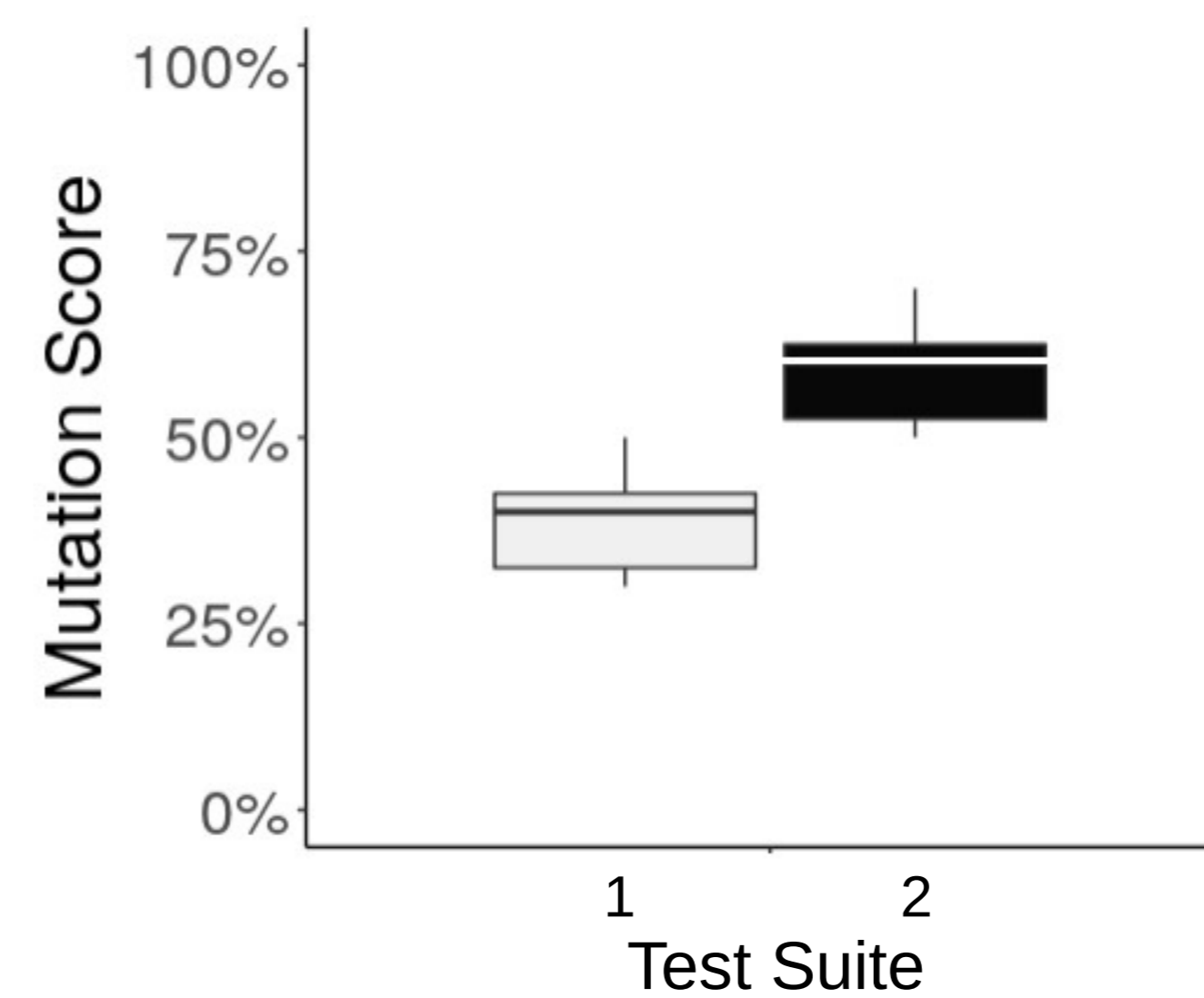
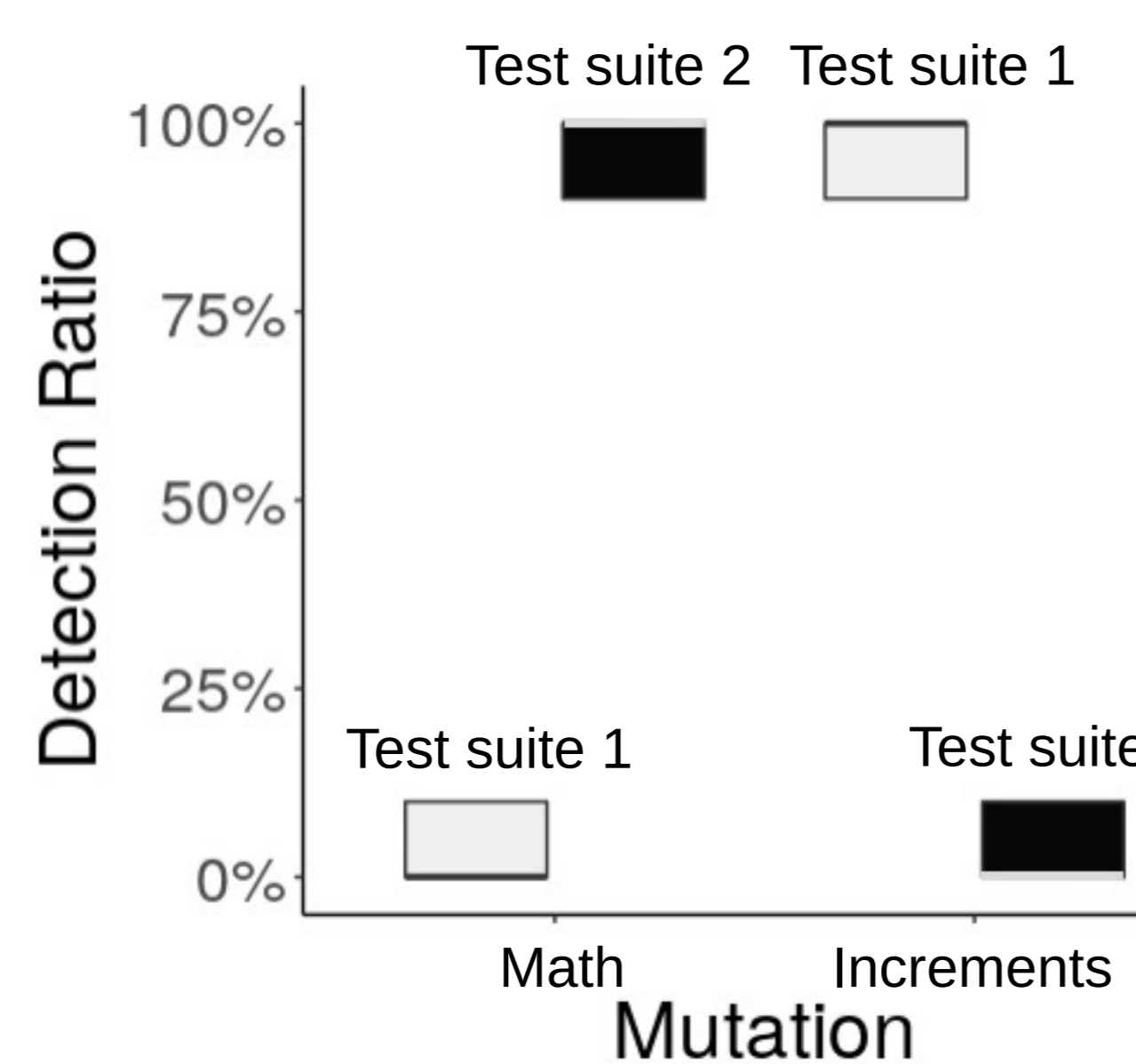
### Mutation Score Definition

$$MS(P, TS) = K / (T - E)$$

P = program under test  
 TS = test suite  
 K = killed mutants  
 T = total mutants  
 E = equivalent mutants

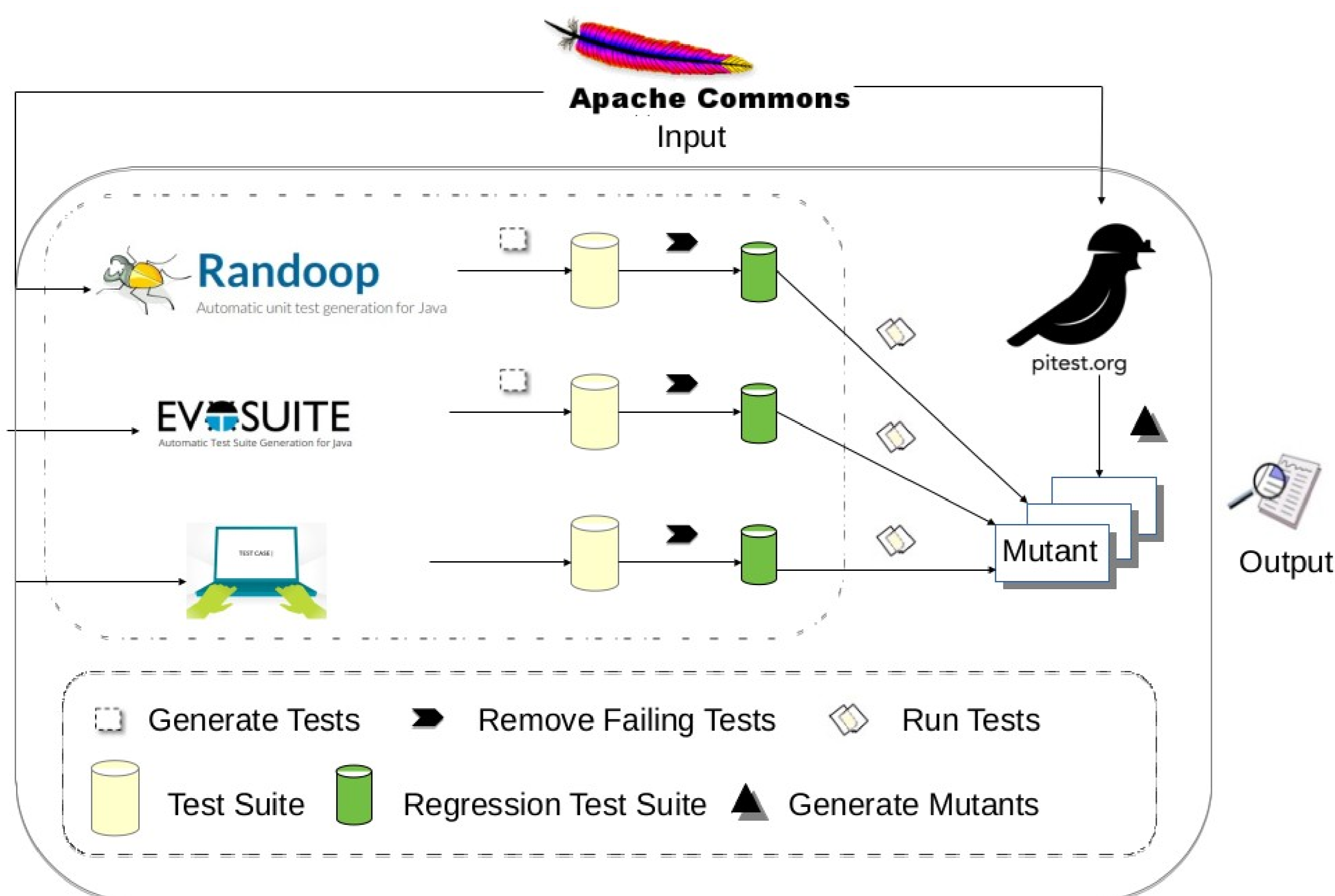
## Motivating Example

1000 mutants created by 2 mutation operators: 600 by Math and 400 by Increments.



It would not be fair to be comparing their effectiveness using mutation score!

## Empirical Study Procedure

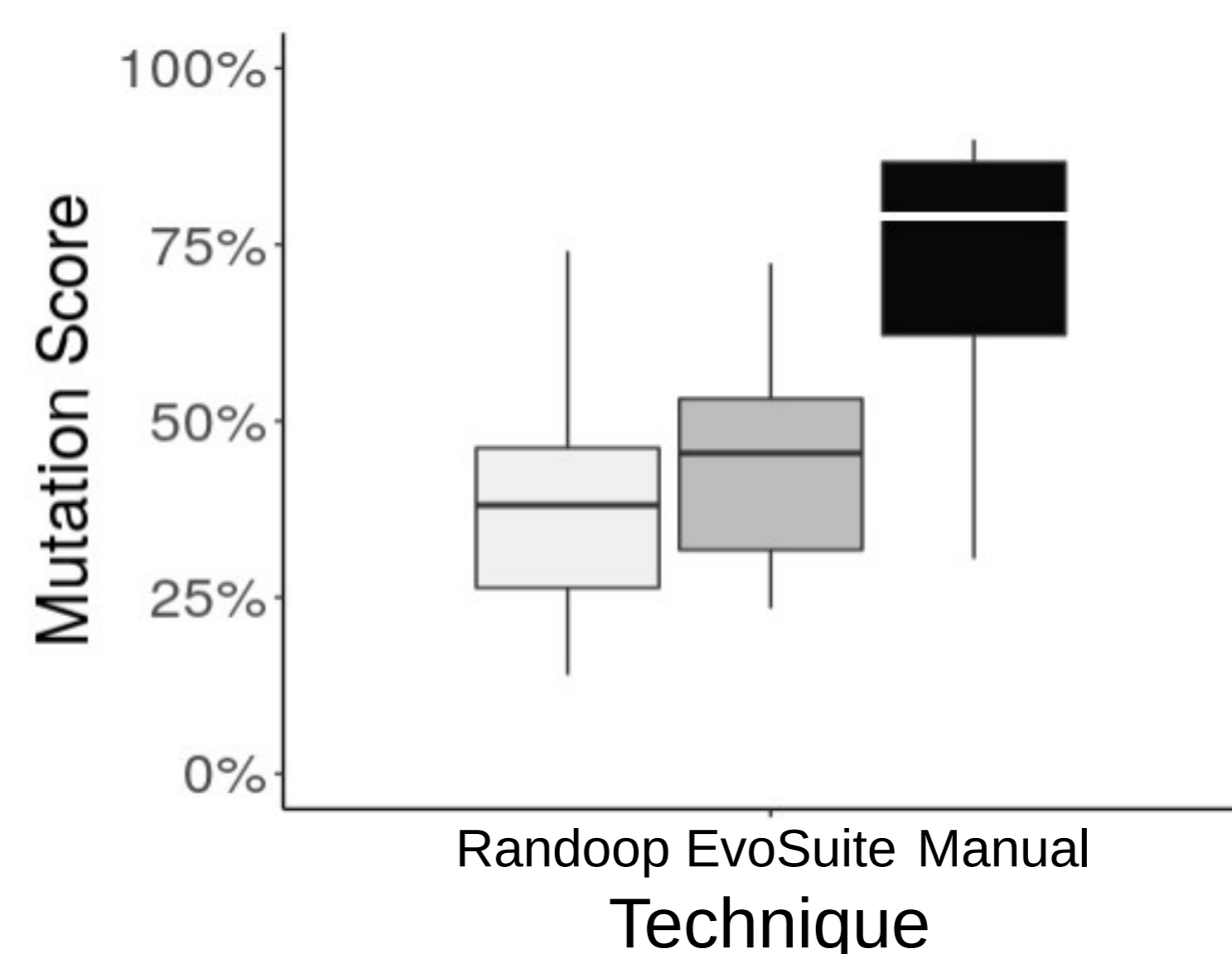
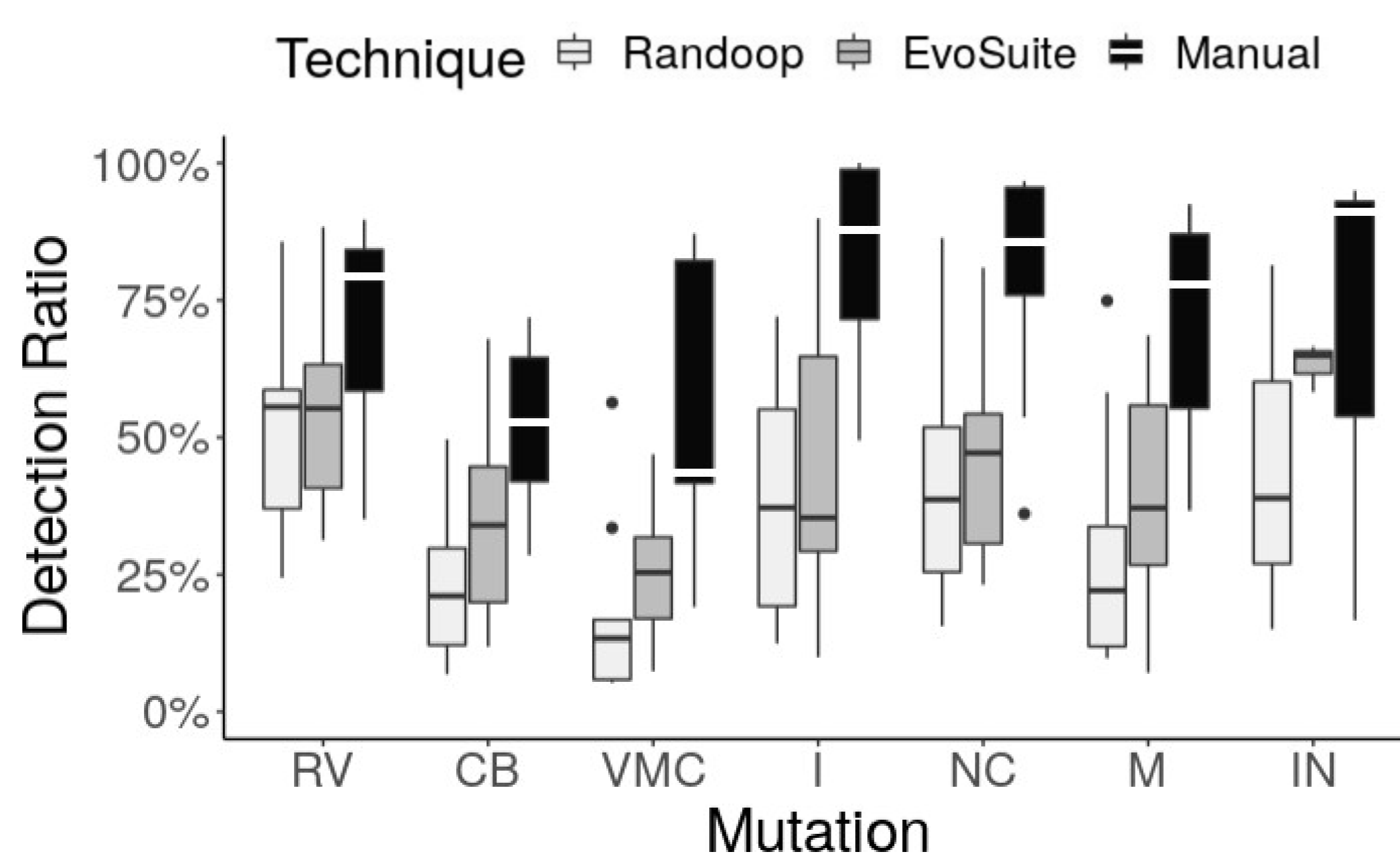


### Projects and Test Suite Description

Project	# of Classes	LOC	Randoop's #tests	EvoSuite's #tests	Manual's #tests
BCEL	426	30812	5023	3435	118
CLI	25	2790	257	455	409
Codec	80	8325	3200	1254	901
CSV	14	1742	2209	102	321
Email	21	2815	3211	134	190
File Upload	47	2425	3676	332	82
Imaging	403	32607	2444	3501	91
Lang	238	27646	2328	3101	4118
Statistics	35	2665	2331	749	386
Validator	73	7409	2716	1092	536
<b>Total</b>	<b>1308</b>	<b>119236</b>	<b>27399</b>	<b>14159</b>	<b>7152</b>

## Study Results

42231 mutants generated by 7 mutation operators.



Mutation Score is a fair metric!

## Acknowledgements

I would like to thank Patricia Machado, Vinicius Matos, Rohit Gheyi, Tiago Massoni and my colleagues.