



Evaluating Testing and Debugging Tools on Real-world Bugs

CS527 - Topics in Software Engineering

Team - Sakshil Verma, Rahul Kumar Veerappa Huleppanour,
Ajitesh Nair, Sanjit Kumar, Urja Khadilkar

Bears

The project creates a detailed database of software bugs and their solutions from open-source projects using the Bears-Benchmark in Java 8. It collects bugs from GitHub via Travis CI builds, storing data such as identifiers, commit hashes, and project URLs. This dataset aids researchers and developers in analyzing bug types, studying fix trends, and developing automated bug detection tools.

Using Evaluation Metrics On The Tests Given In The Dataset

Metrics Used

- CChange (Cyclomatic Complexity Change)
- MChange (Method Change)
- LChange (Line Changes)
- LD (Levenshtein Distance)
- CB (Cyclomatic Complexity of Buggy Code)
- CP (Cyclomatic Complexity of Patched Code)
- CC (Cyclomatic Complexity Change)
- CodeBLEU Score

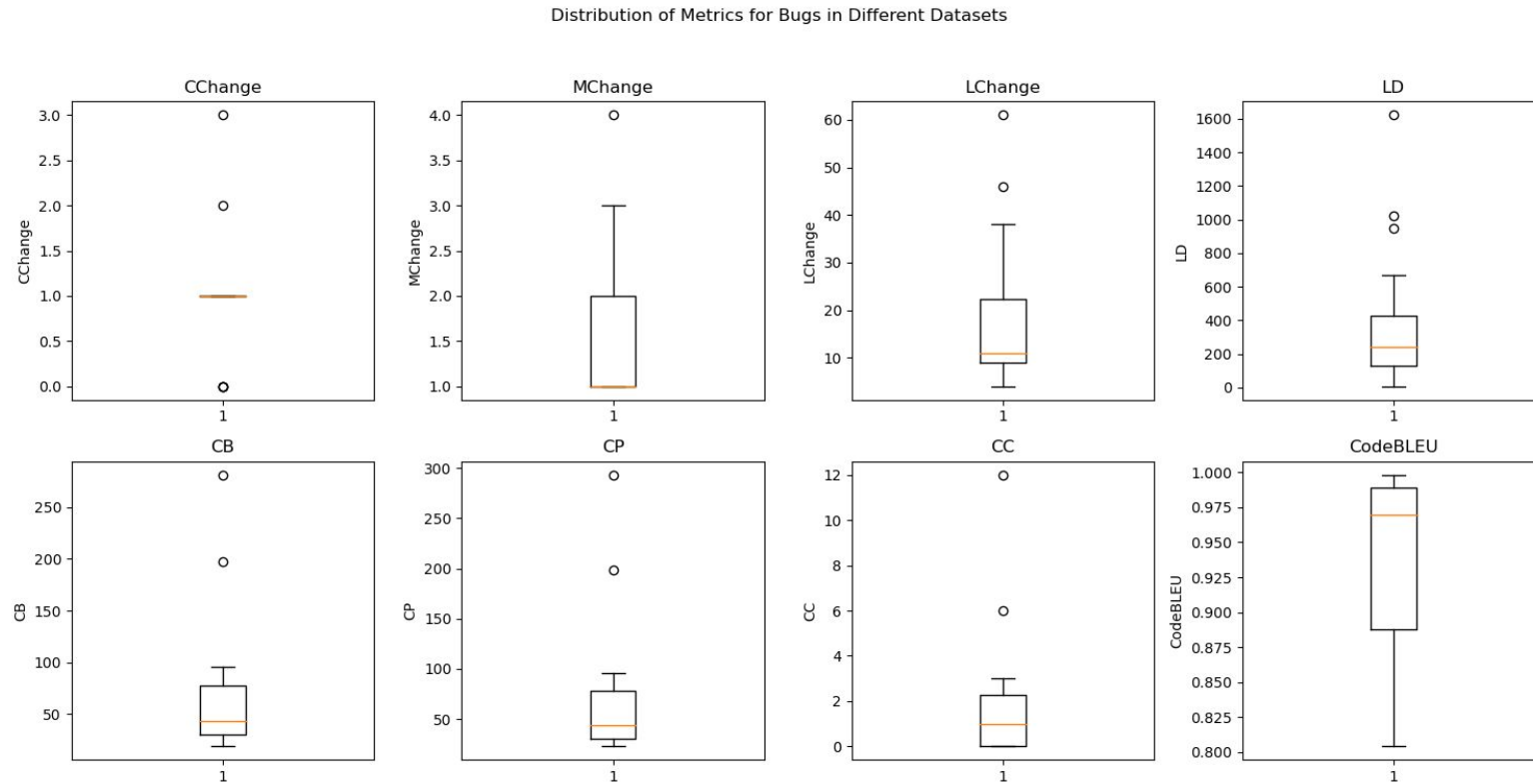
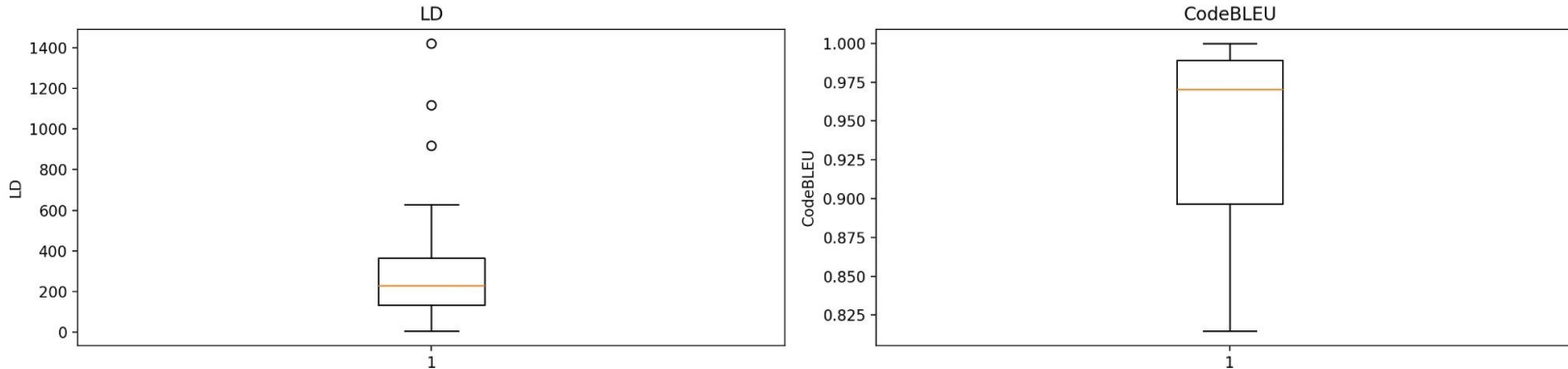


Figure 1.

Distribution of Metrics for Bugs using Randoop



Distribution of Metrics for Bugs using Evosuite

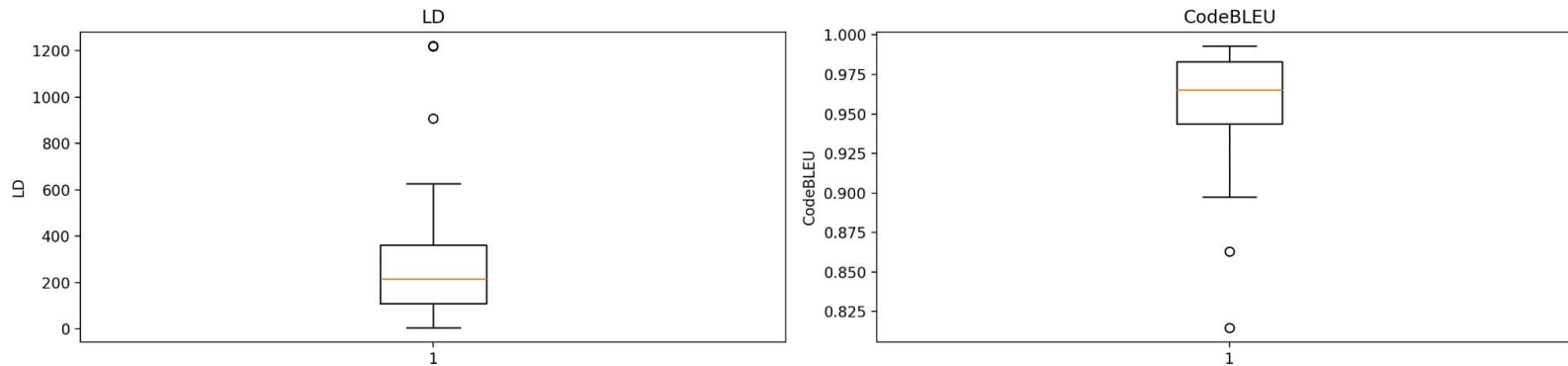


Figure 2.

Using Evaluation Metrics On The Tests Generated Using Automated Test Generation Tools - Randoop And Evosuite

Metrics Used

- LD (Levenshtein Distance)
- CodeBLEU Score

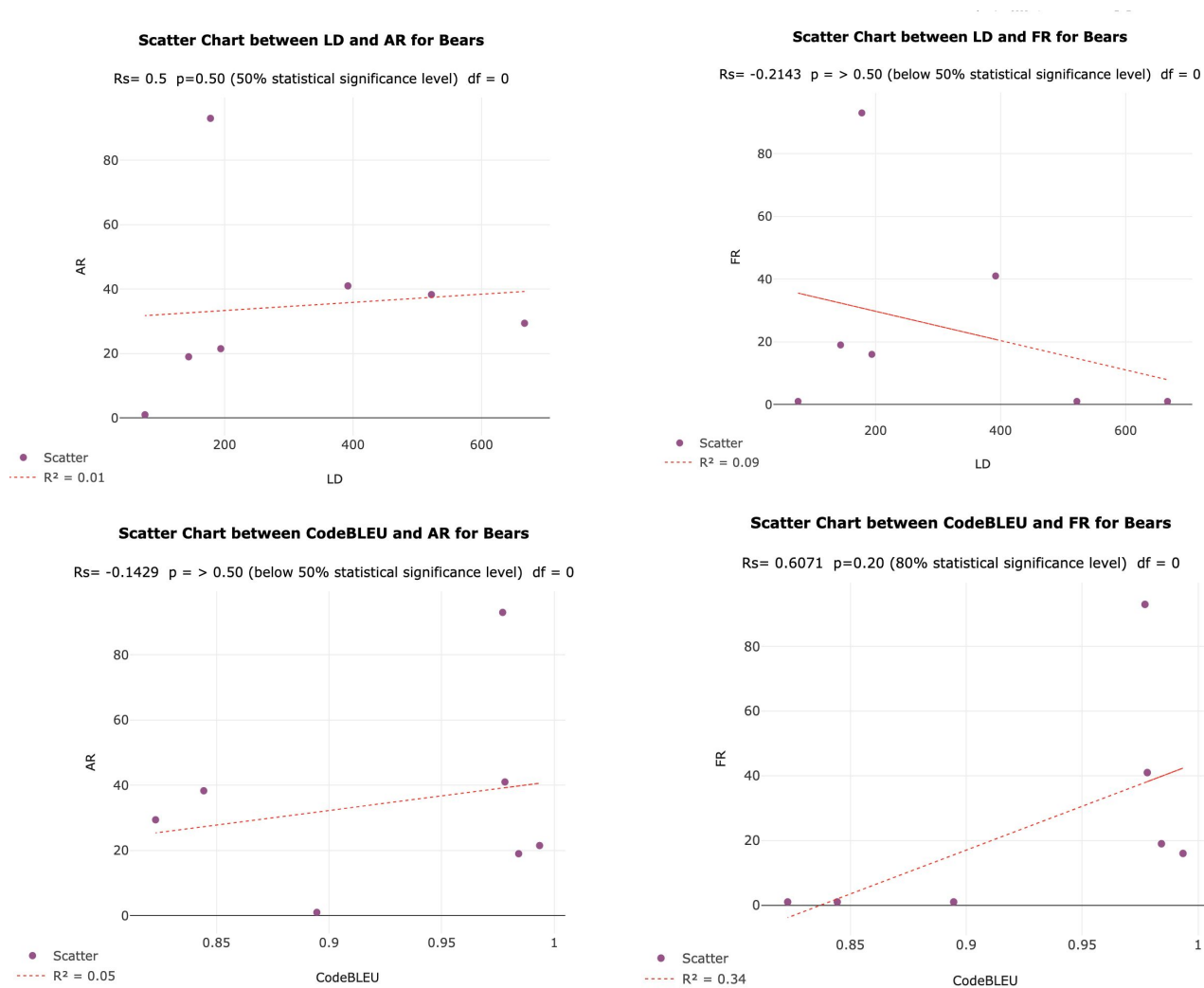


Figure 3.

Bug Localization Analysis:

- Distribution of Levenshtein Distance with Rank
- Distribution of CodeBLEU with Rank
- Bugs Considered: Bears-11, Bears-17, Bears-25, Bears-101, Bears-107, Bears-114.

QuixBugs

A dataset containing 40 Java and Python programs, each containing a single line bug. These buggy programs were built as part of the Quixey Challenge, organized by a startup named Quixey in 2011. These defective programs of classic algorithms were provided to programmers to fix in a minute for a monetary prize. These programs were initially written only in Python. Researchers at MIT and Google translated them into Java to generate a dataset for Multilingual Program Repair. Each bug is accompanied by pass/fail test cases.

Using Evaluation Metrics On The Tests Given In The Dataset

Metrics Used

- CChange (Cyclomatic Complexity Change)
- MChange (Method Change)
- LChange (Line Changes)
- LD (Levenshtein Distance)
- CB (Cyclomatic Complexity of Buggy Code)
- CP (Cyclomatic Complexity of Patched Code)
- CC (Cyclomatic Complexity Change)
- CodeBLEU Score

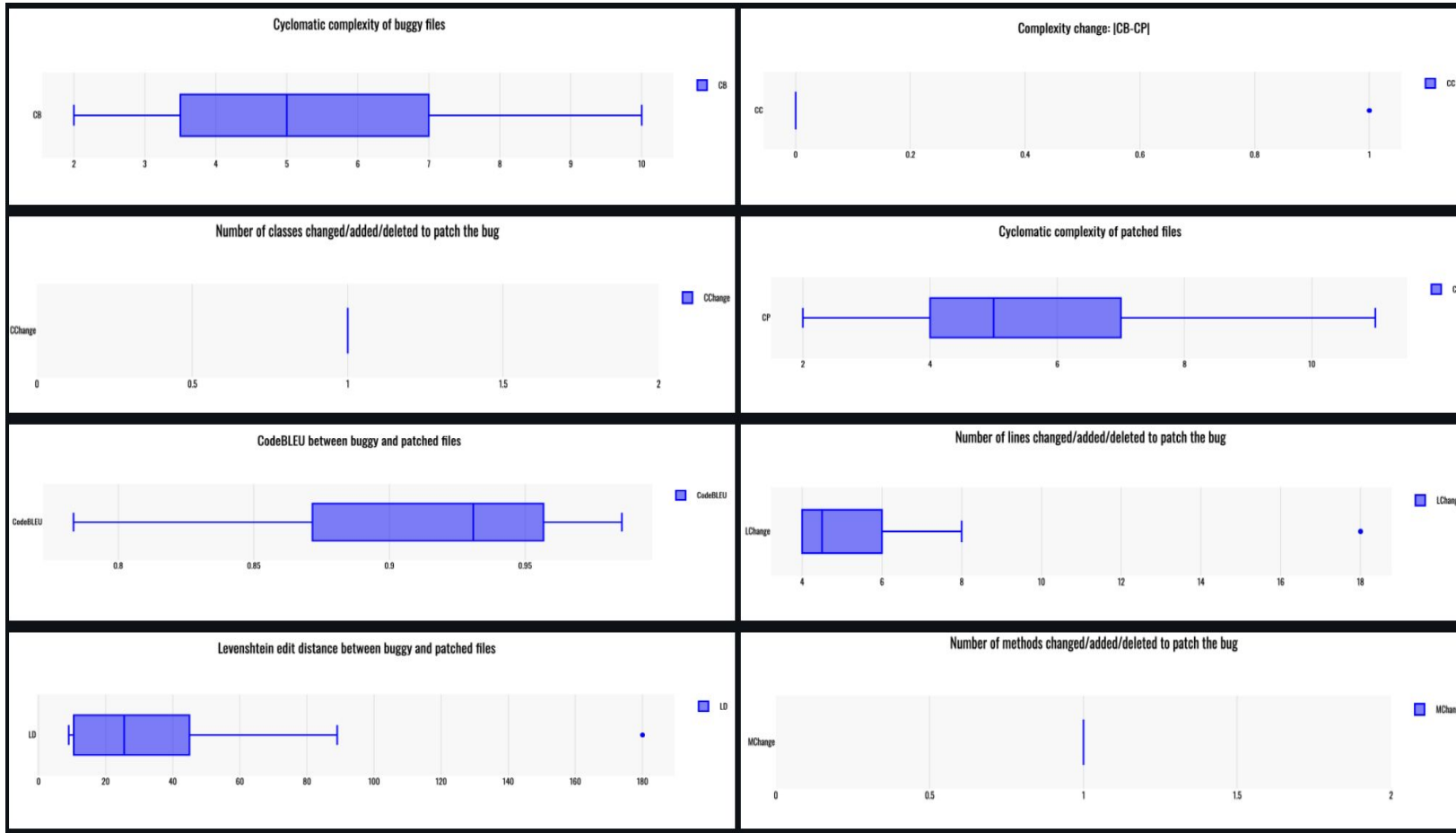


Figure 4.

QuixBugs - Benchmarking Results Using Automated Test Generation Tools

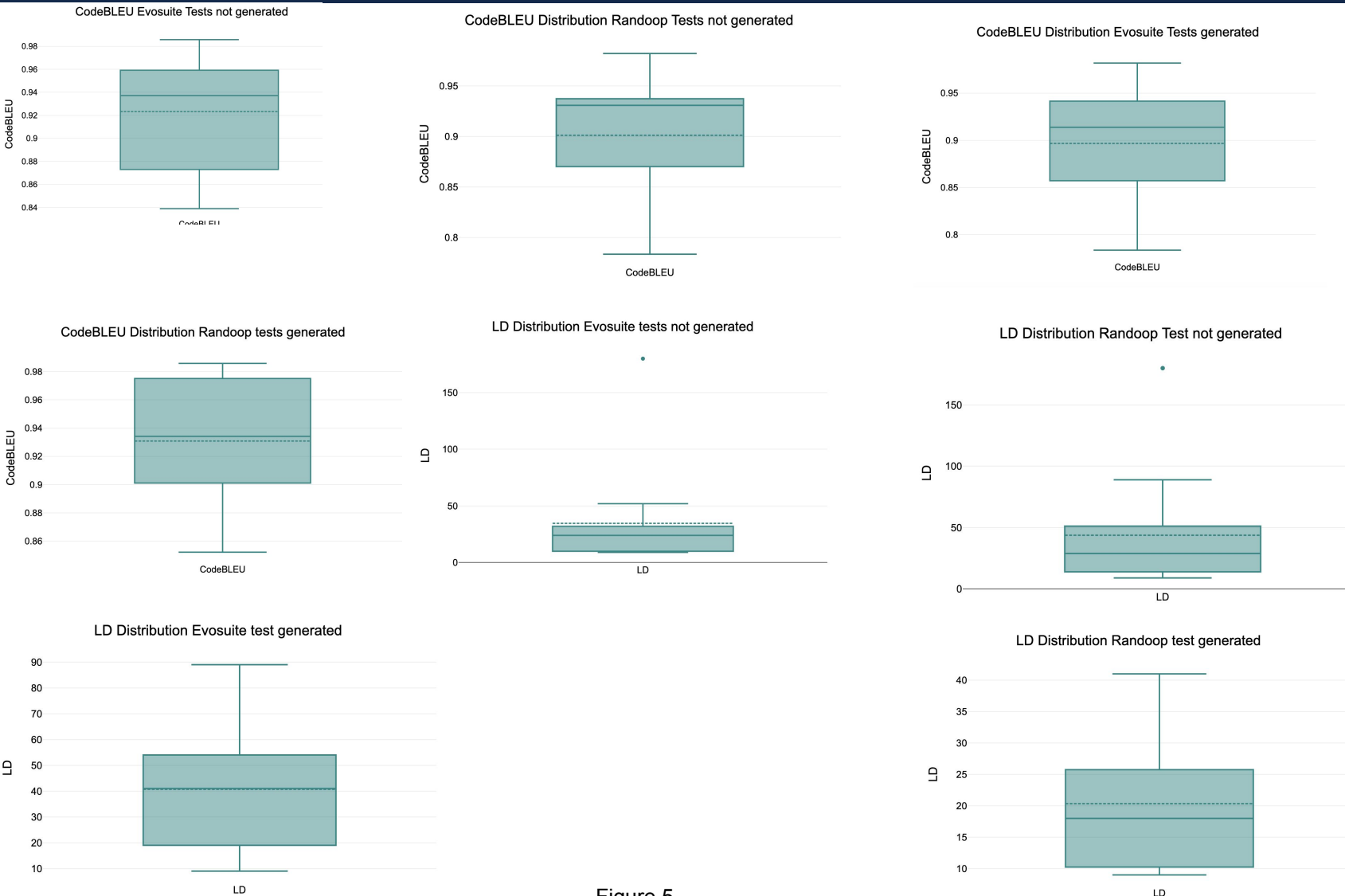


Figure 5.

Using Evaluation Metrics On The Tests Generated Using Automated Test Generation Tools - Randoop And EvoSuite

Metrics Used

- LD (Levenshtein Distance)
- CodeBLEU Score

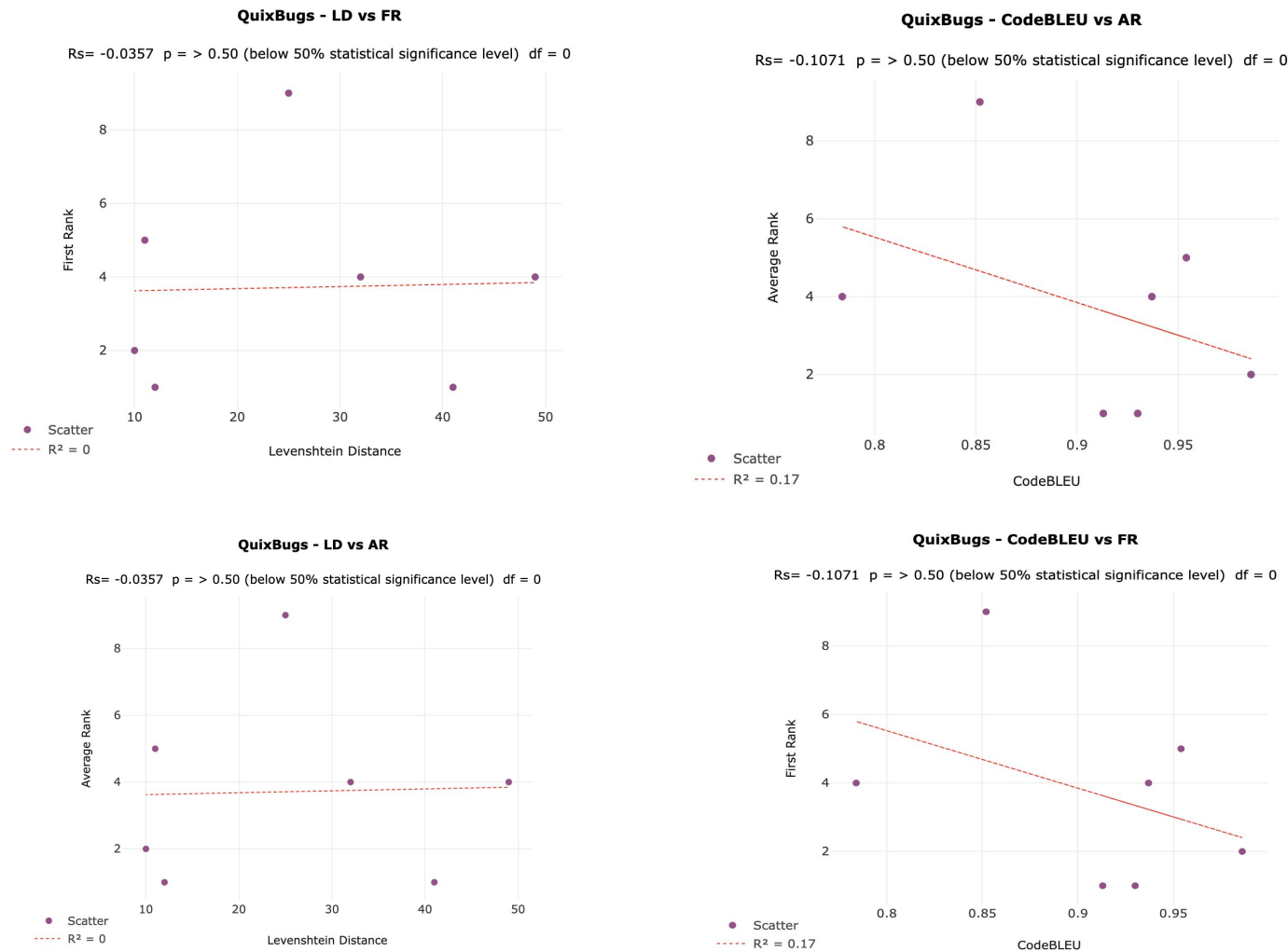


Figure 6.

Bug Localization Analysis:

- Distribution of Levenshtein Distance with Rank
- Distribution of CodeBLEU with Rank
- Bugs Considered: BUCKETSORT, LEVENSHTein, DETECT_CYCLE, FIND_IN_SORTED, FLATTEN, KHEAPSORT, KNAPSACK

Defects4J

Defects4J is a curated database of real-world software bugs specifically designed for software testing research that contains 835 bugs from 17 different open-source java projects. It comes with a framework that helps manage, evaluate and test the bugs based on different evaluation and localization approaches. Each bug in the database is associated with a specific program version where the bug exists.

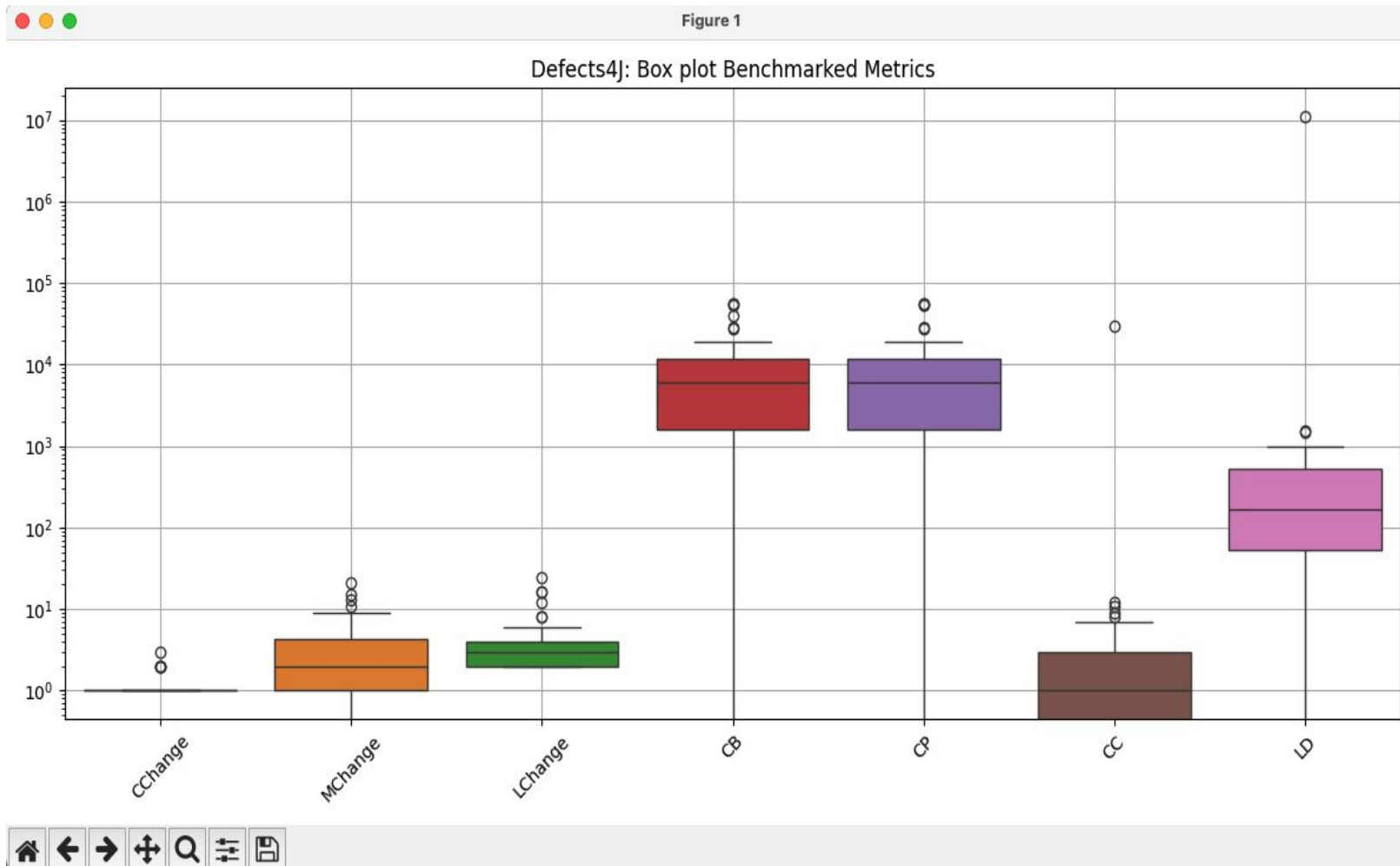


Figure 7.

Using Evaluation Metrics On The Tests Given In The Dataset

Metrics Used

- CChange (Cyclomatic Complexity Change)
- MChange (Method Change)
- LChange (Line Changes)
- LD (Levenshtein Distance)
- CB (Cyclomatic Complexity of Buggy Code)
- CP (Cyclomatic Complexity of Patched Code)
- CC (Cyclomatic Complexity Change)
- CodeBLEU Score

Compilation issues with the generated bugs.

Firstly, the generated bugs via Randoop, Evosuite did not build and run with the projects natively (had compilation errors).

Identification of ‘failing bugs’

Secondly, because of these compilation issues *identification* of the failing bugs were not possible and therefore we did not extract the information for benchmarked metrics associated with such generated *failing tests*.

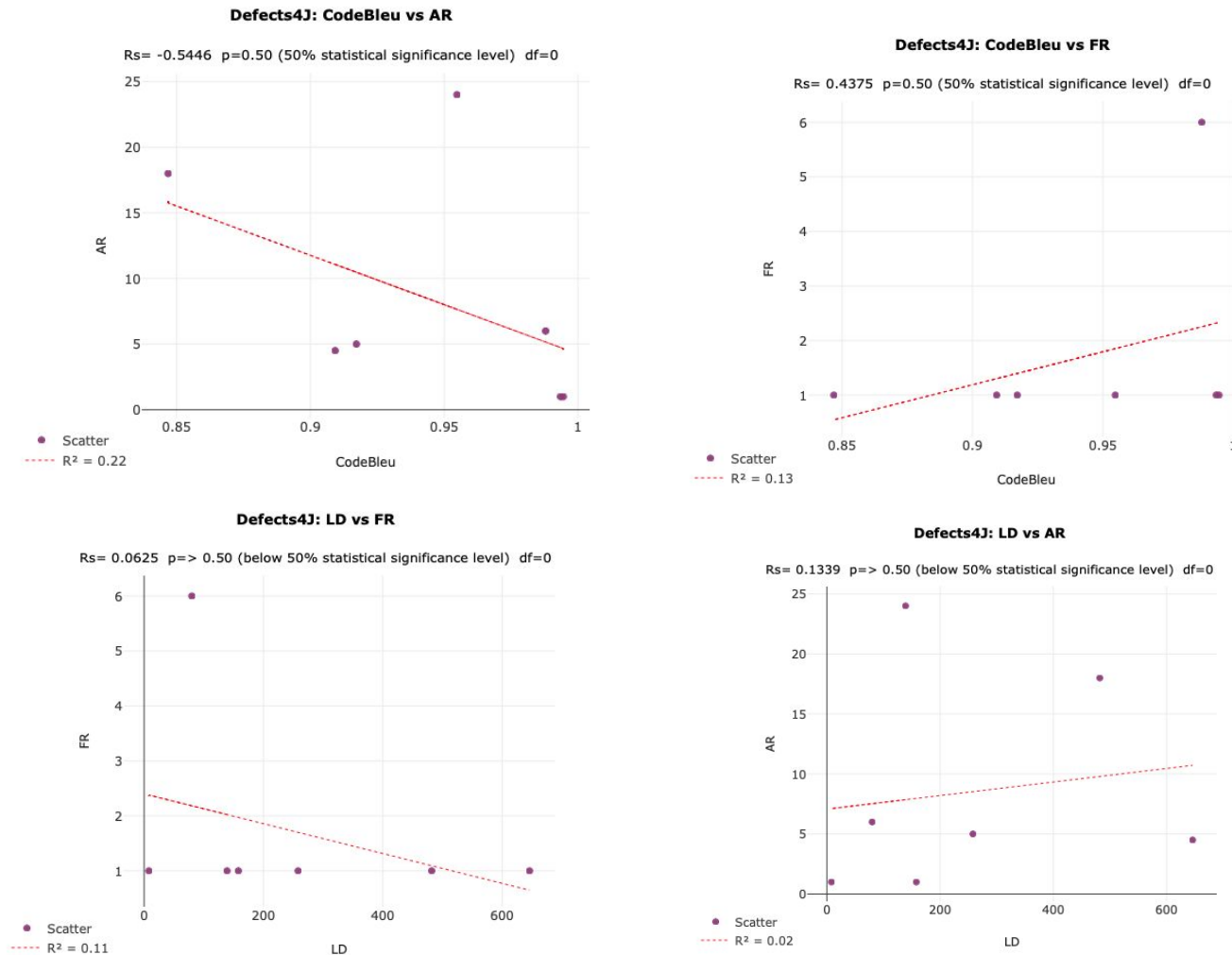


Figure 8.

Bug Localization Analysis:

- Distribution of Levenshtein Distance with Rank
- Distribution of CodeBLEU with Rank
- Bugs Considered: Cli_18, Cli_20, Cli_22, Cli_24, Cli_26, Cli_28, Cli_30

BugSwarm

BugSwarm is a toolset that helps create a continuously growing set of real-world reproducible build failures and fixes. The BugSwarm toolset has mined a dataset of ~3000 bugs and the tests run on them. This dataset mainly consists of bugs arising from build failures, and not test failures like the other datasets involved.

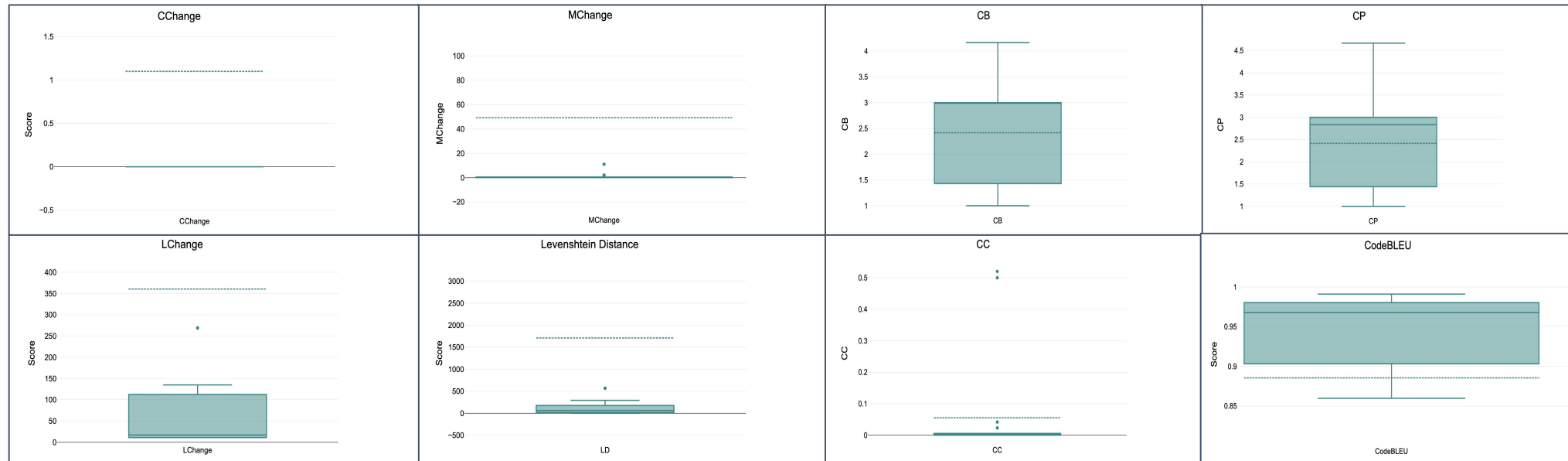


Figure 9.

Using Evaluation Metrics On The Tests Given In The Dataset

- CChange (Cyclomatic Complexity Change)
- MChange (Method Change)
- LChange (Line Changes)
- LD (Levenshtein Distance)
- CB (Cyclomatic Complexity of Buggy Code)
- CP (Cyclomatic Complexity of Patched Code)
- CC (Cyclomatic Complexity Change)
- CodeBLEU Score

Errors with generating tests

Both Randoop and Evosuite were unable to generate tests for the 20 bugs selected, even after changing the bugs involved and adding the correct classpaths due to unresolved dependency issues. In addition, while Evosuite generated tests for 1 bug only, it was after multiple attempts due to system constraints, and hence was unable to generate tests for the remaining bugs.

JaCoCo coverage reports

Each module generated its own coverage report, which prevented us from obtaining a unified project-wide view. Additionally, the lack of information about failing tests in the reports hindered our ability to calculate suspiciousness scores effectively.



Thank You

Lorem Ipsum Dolor Sit Amet Consectetur Adipiscing

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem Ipsum Dolor Sit Amet Consectetur Adipiscing

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

Lorem Ipsum Dolor Sit Amet Consectetur Adipiscing

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Lorem Ipsum Dolor Sit Amet Consectetur Adipiscing

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

IMAGE / GRAPHIC

Lorem Ipsum Dolor Sit Amet Consectetur Adipiscing

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

IMAGE / GRAPHIC

Lorem Ipsum Dolor Sit Amet Consectetur Adipiscing

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

Lorem Ipsum Dolor Sit Amet Consectetur Adipiscing

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

Lorem Ipsum Dolor Sit Amet Consectetur Adipiscing

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

IMAGE / GRAPHIC

IMAGE / GRAPHIC

IMAGE / GRAPHIC

Lorem Ipsum Dolor Sit Amet Consectetur Adipiscing

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

Lorem Ipsum Dolor Sit Amet Consectetur Adipiscing

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

Lorem Ipsum Dolor Sit Amet Consectetur Adipiscing

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

Lorem Ipsum Dolor Sit Amet Consectetur Adipiscing

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

IMAGE / GRAPHIC



CHART / GRAPH

(Short commentary or relevant information about the contents of the chart/graph.)

IMAGE (Send to back so Block I is not covered.)

(Text here relevant to the photo for a larger call out on a given idea or message.)
(Keep this text simple and short on one or two lines.)



IMAGE

Lorem ipsum dolor sit amet consectetur adipiscing elit

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

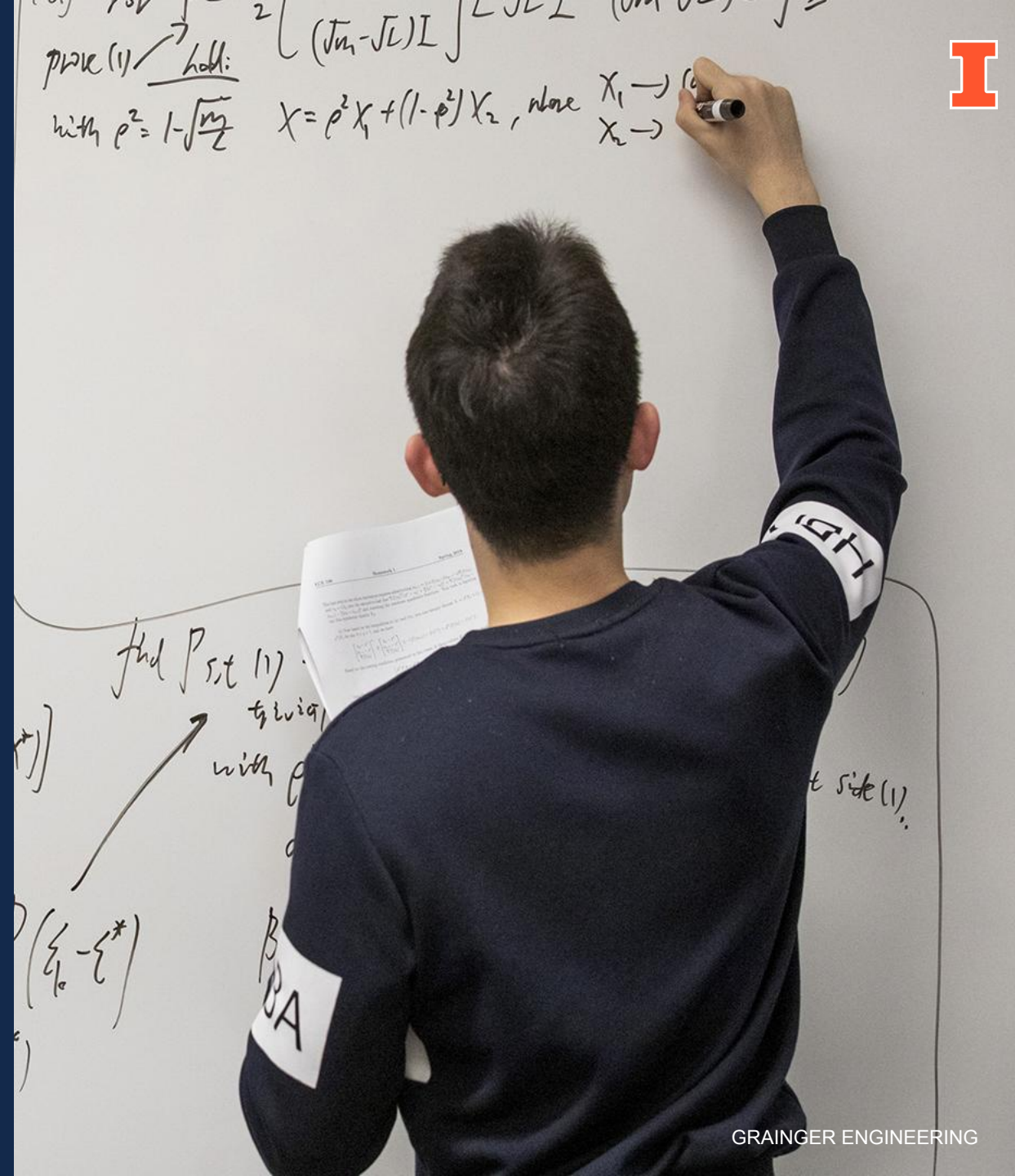
Lorem ipsum dolor sit amet consectetur adipiscing

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

IMAGE

(Use a photo for the entire slide like the example shown here. Send the photo to the back so the Block I and footer text is not covered.)

(Slide the blue box and the text left or right to fit over your background image. Use this text box for a call out or caption to the image.)





(Closing slide option. Suggestions for last slide)

Summary
Thank You
Questions
Contact Information

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.



The Grainger College of Engineering

UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN