RQ1: How does class overlap influence the prediction performance of existing within-project defect prediction models?
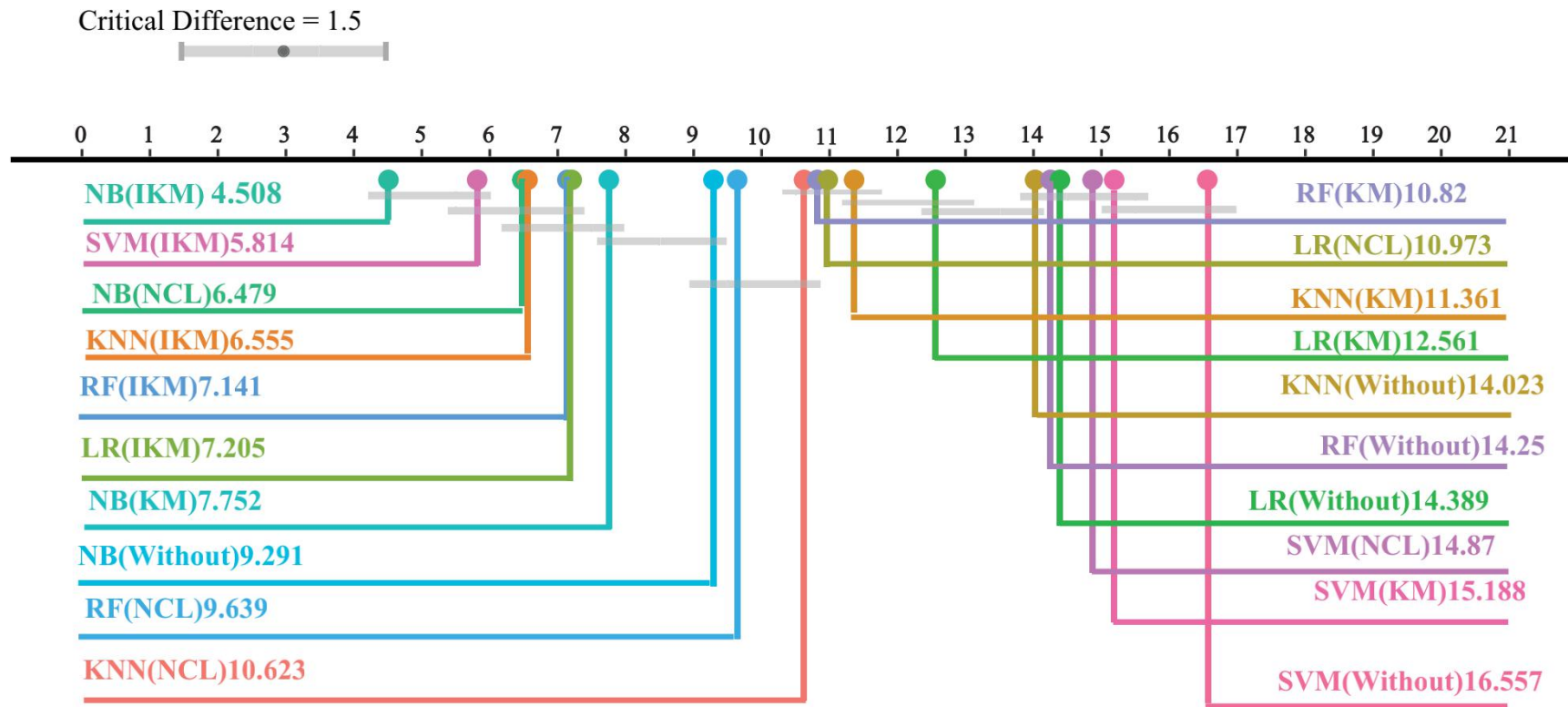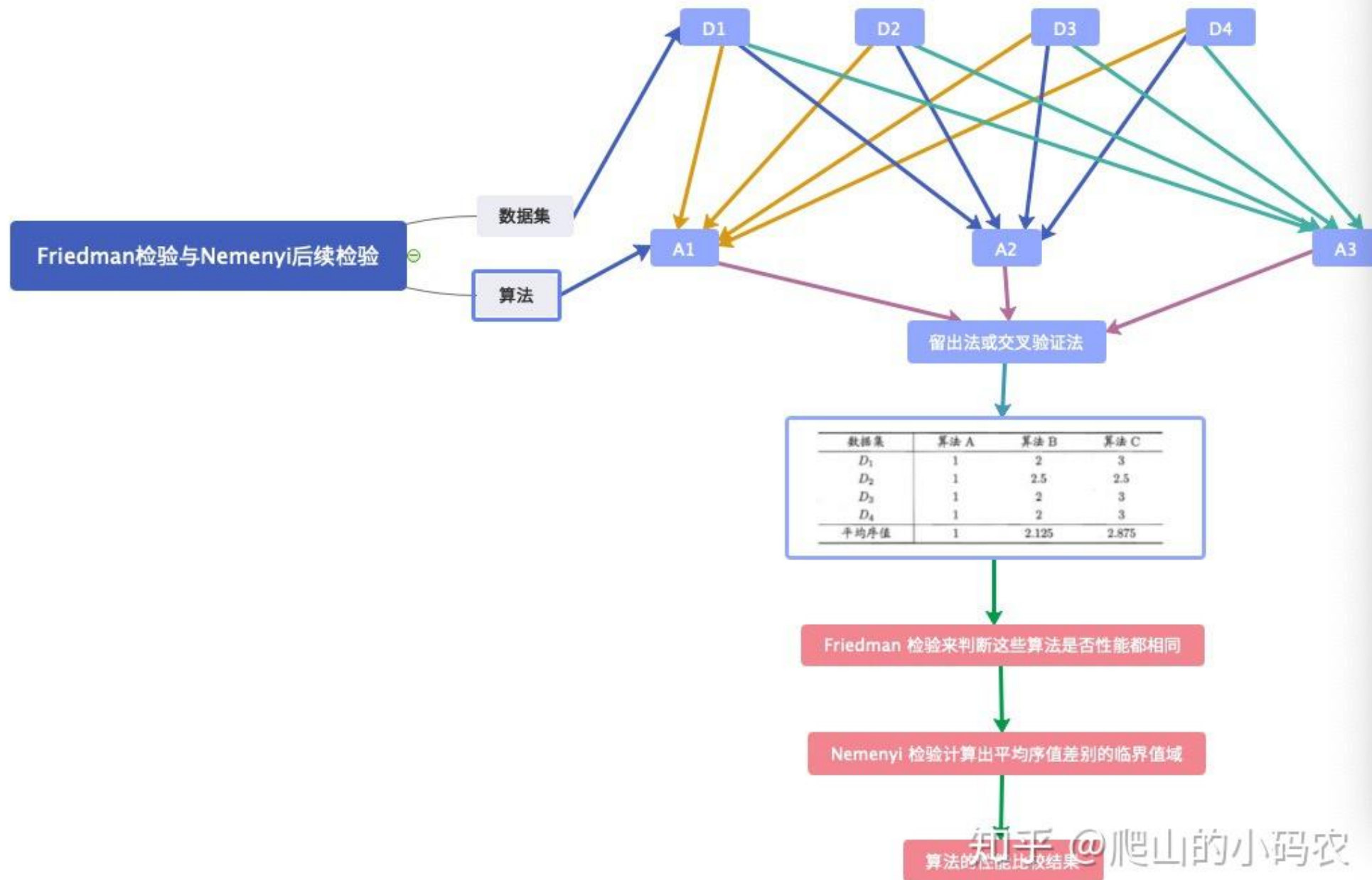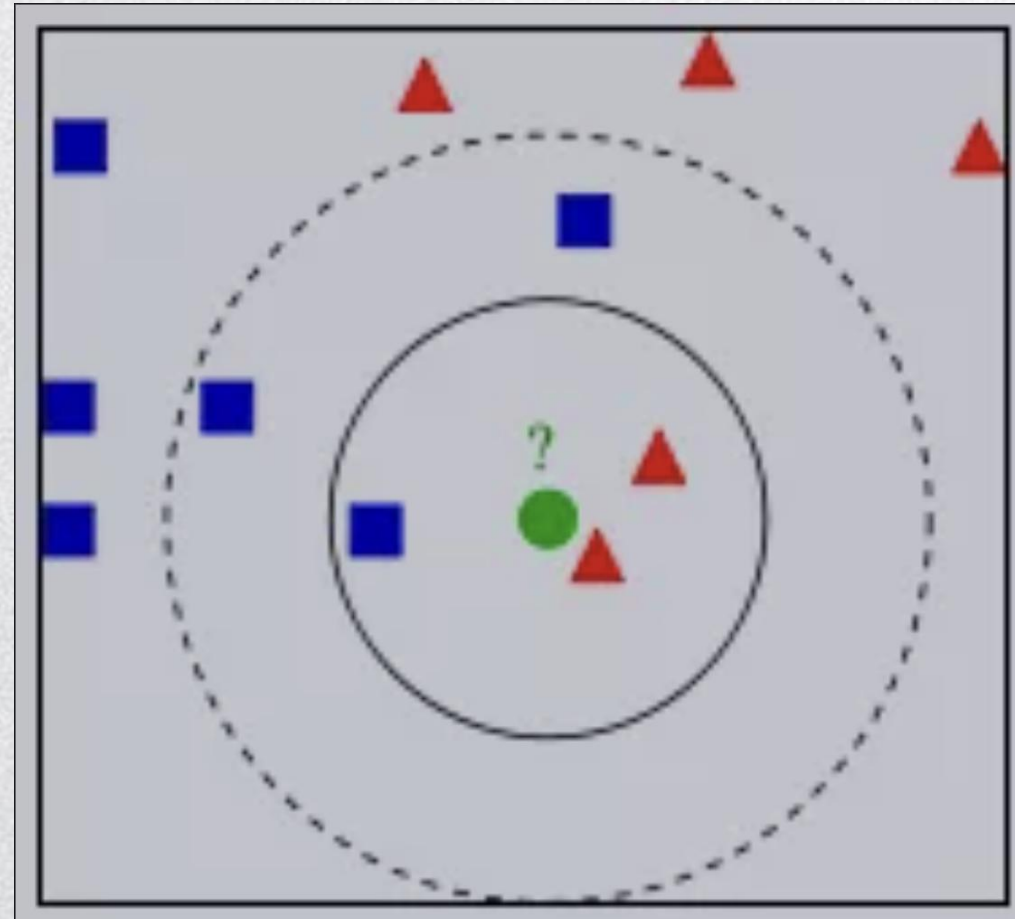

Fig. 6. The ranks on *bal* for within-project defect prediction (WPDP) methods with post-hoc Nemenyi test. Methods connected by gray lines are not significantly different.

数据集	算法 A	算法 B	算法 C
D_1	1	2	3
D_2	1	2.5	2.5
D_3	1	2	3
D_4	1	2	3
平均序值	1	2.125	2.875

KNN (K-NearestNeighbor)

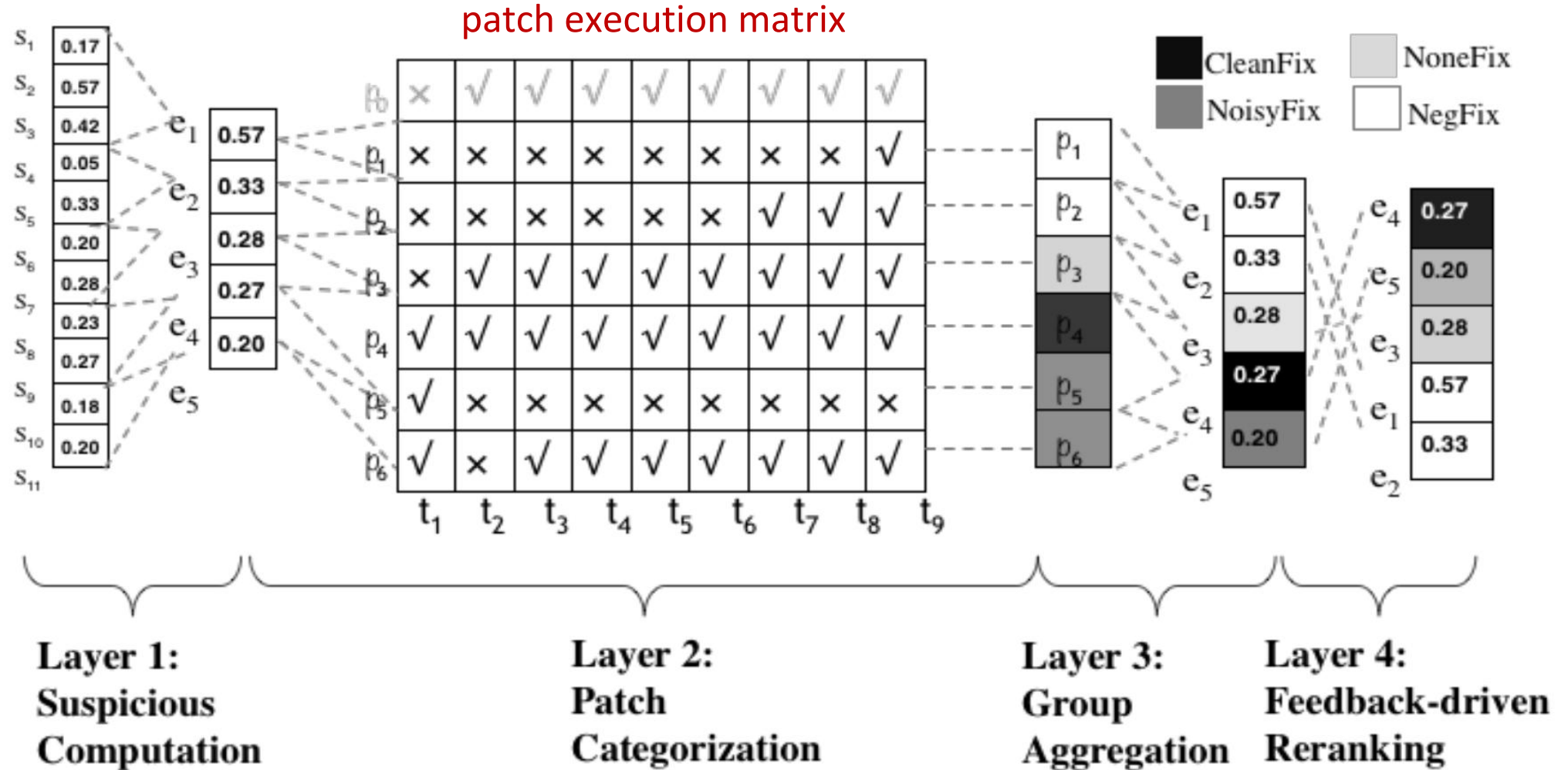|		
《Can Automated Program Repair Refine Fault Localization? A Unified Debugging Approach》	2020	ISSTA
《Using bug report similarity to enhance bug localisation》	2012	WCRE
《A method of nonbug report identification from bug report repository》	2021	ISAROB （International Society of Artificial Life and Robotics ）

2020 ISSTA

patch execution matrix

Figure 3: Overview of ProFL

Can Automated Program Repair Refine Fault Localization? A Unified Debugging Approach

2020 ISSTA

$$\mathbb{G}[e] = \begin{cases} \texttt{CleanFix} \\ \texttt{NoisyFix} \\ \texttt{NoneFix} \\ \texttt{NegFix} \end{cases}$$

if it passes some originally fail_x0002_ing tests while does not fail any originally passing tests

if it passes some originally fail_x0002_ing tests but also fails on some originally passing tests

if it does not impact any originally fail_x0002_ing or passing tests.

if it does not pass any originally failing test while fails some originally passing tests

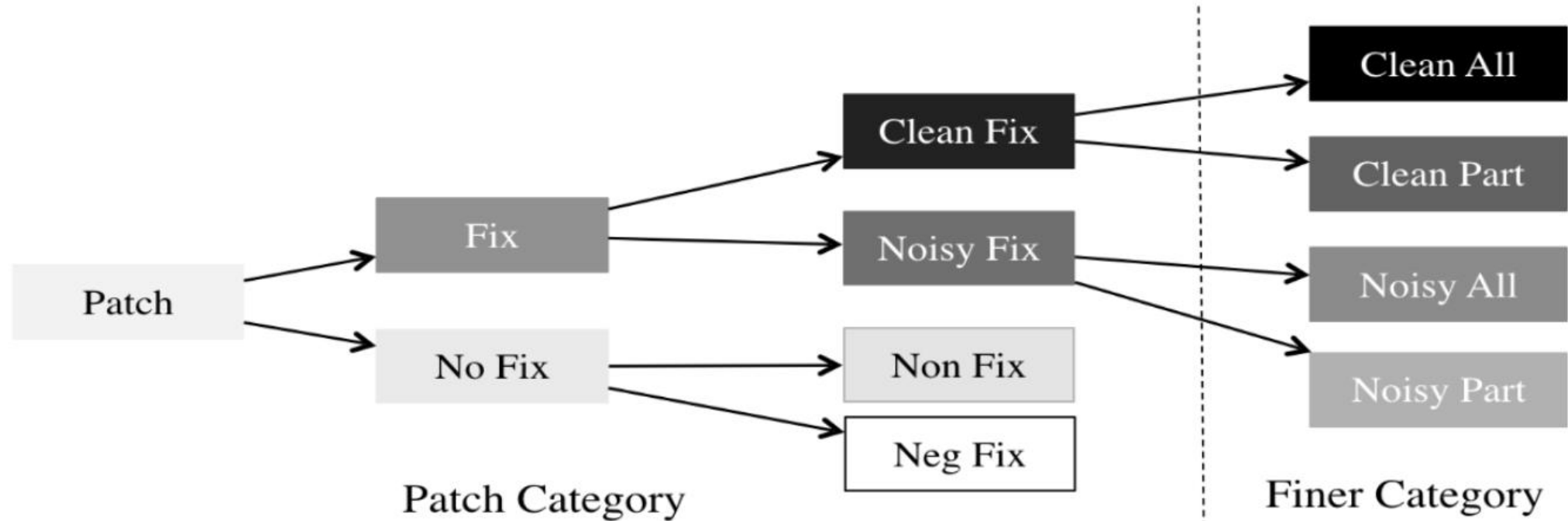2020 ISSTA

CleanFix > NoisyFix > NoneFix > NegFix

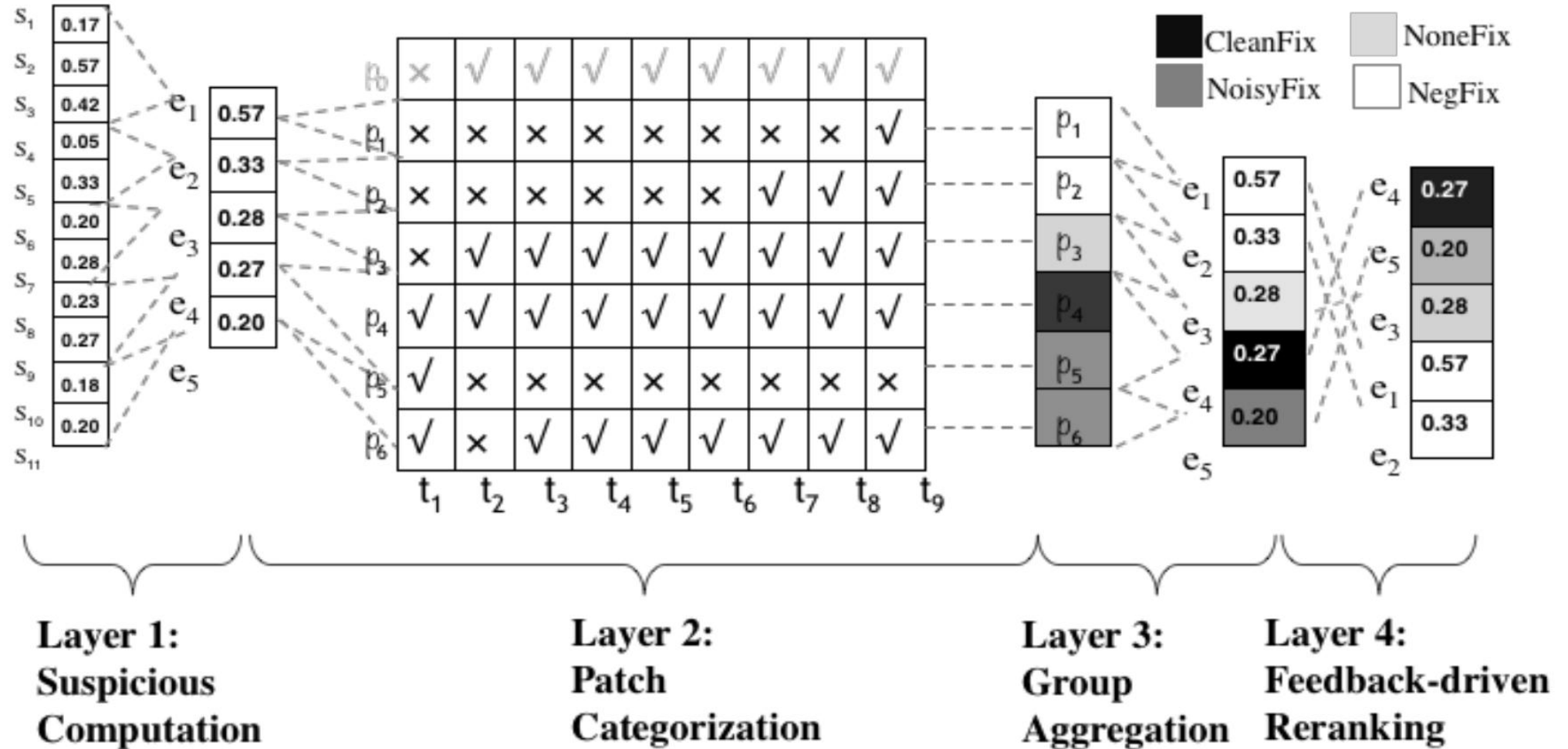Figure 4: Patch categorization tree

2020 ISSTA

Figure 3: Overview of ProFL

Using bug report similarity to enhance bug localisation

If duplicate bugs reports, which by definition are fixed in the same source location,

can be detected through the use of similar language, can bugs which are in the same

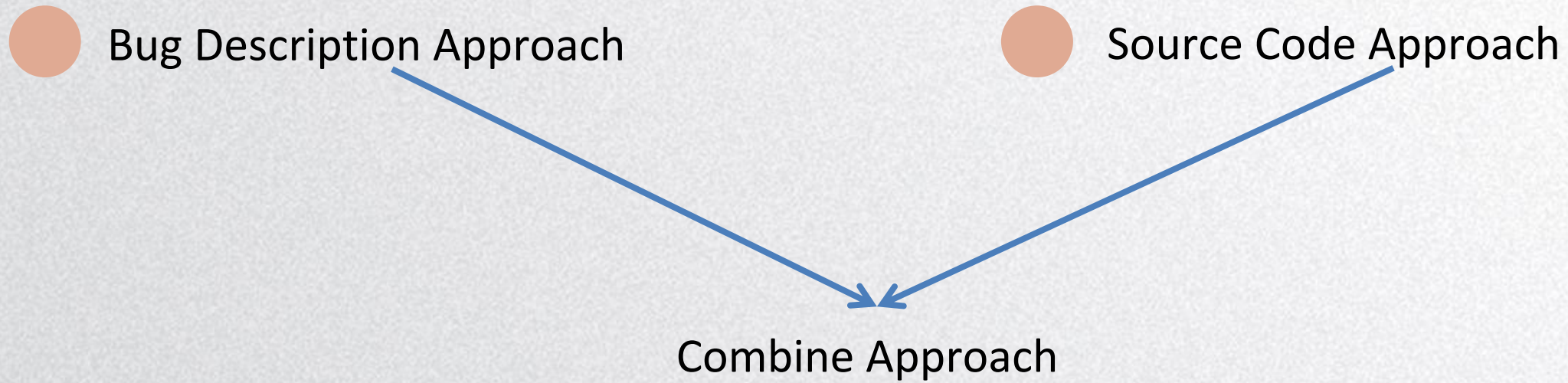*location but not duplicates be detected in the same way?*

jEdit1 project:

- 1659666 — perl.xml patch for quote-like operators
- 1760646 — perl colorization
- 1807549 — Perl escape still not working correctly

org.gjt.sp.jedit.syntax.TokenMarker.markTokens(LineConte xt,TokenHandler, Segment)

Using bug report similarity to enhance bug localisation

● Bug Description Approach   ● Source Code Approach

Combine Approach

A method of nonbug report identification from bug report repository

BR Optimization

BR Triage

Bug Fixing

● Bug report pre-processing

● Bug report representation and term weighting——
tf、tf-idf、tf-igm、tf-icf、tf-icff

● Non-bug report identifiers modeling——
Traditional binary-class SVM、Schölkopf methodology、Support vector data description (SVDD)