

# Appendix

## 1 Additional Experimental Results

We provide additional experimental results in the rest of this appendix to support our claims in the main paper.

### 1.1 Parameter selection using cross validation

Table 1 demonstrates the best hyperparameter combinations (columns 2–5) for each of the folds. As we can observe, the most prevalent combinations are  $\{S = \text{Ochiai}, \mu = 17, \beta = 20, p = 1e - 5\}$ . Table 1 also shows the mean and median of the percentage improvements in  $\mathcal{EXAM}$  score by ARTEMIS over each of the base SBFL metrics for all faults in the test data using the best hyperparameter combination selected over the training data. As we can see, ARTEMIS performs better than each base SBFL metric in each fold which signifies that ARTEMIS generalizes well on unknown buggy spectrums.

### 1.2 Comparison of ARTEMIS against base SBFL metrics

Table 2 presents the mean and median percentage improvements in developer effort (in terms of  $\mathcal{EXAM}$  score) that ARTEMIS provides over each of the base SBFL metrics on the faults from each project using the most prevalent hyperparameters selected during the cross validation experiments, i.e.,  $S = \text{Ochiai}, \mu = 17, \beta = 20, p = 1e - 5$ .

Tables 3, 4, 5, 6, 7 and 8 demonstrates the percentage of faulty components that are ranked within top- $n$  positions in the ranked lists using any approach on projects *Chart*, *Closure*, *Lang*, *Math*, *Mockito* and *Time* respectively.

Table 1: Comparing the performance of ARTEMIS against each of the base SBFL metrics on test data in the cross validation experiments.

Test data	Best hyperparameter combination				Mean % improvement in $\mathcal{E}\mathcal{X}\mathcal{A}\mathcal{M}$ score on base SBFL metrics					
	$\mu$	$\beta$	$p$	$S$	Ochiai	D <sup>2</sup>	Barinel	Op2	Tarantula	Kulczynski
Chart	17	20	$1e-5$	Ochiai	26%	43%	28%	59%	28%	23%
Closure	6	10	$1e-5$	Op2	13%	13%	41%	8%	41%	18%
Lang	17	20	$1e-5$	Ochiai	15%	12%	30%	24%	30%	19%
Math	17	20	$1e-5$	Ochiai	14%	21%	13%	29%	13%	12%
Mockito	16	20	$1e-5$	Ochiai	27%	25%	24%	30%	24%	31%
Time	17	20	$1e-5$	Ochiai	23%	29%	7%	41%	7%	17%

  

Test data	Best hyperparameter combination				Median % improvement in $\mathcal{E}\mathcal{X}\mathcal{A}\mathcal{M}$ score on base SBFL metrics					
	$\mu$	$\beta$	$p$	$S$	Ochiai	D <sup>2</sup>	Barinel	Op2	Tarantula	Kulczynski
Chart	17	20	$1e-5$	Ochiai	19%	37%	23%	55%	23%	20%
Closure	6	10	$1e-5$	Op2	11%	10%	38%	5%	38%	13%
Lang	17	20	$1e-5$	Ochiai	10%	7%	23%	22%	23%	17%
Math	17	20	$1e-5$	Ochiai	10%	19%	11%	26%	11%	10%
Mockito	16	20	$1e-5$	Ochiai	24%	23%	20%	26%	20%	27%
Time	17	20	$1e-5$	Ochiai	19%	26%	4%	38%	4%	10%

Table 2: Comparing the performance of ARTEMIS against each of the base SBFL metrics on each project using the most prevalent hyperparameters selected during the cross validation experiments, i.e.,  $S = \text{Ochiai}$ ,  $\mu = 17$ ,  $\beta = 20$ ,  $p = 1e-5$ .

Project	Hyperparameter values				Mean % improvement in $\mathcal{E}\mathcal{X}\mathcal{A}\mathcal{M}$ score on base SBFL metrics					
	$\mu$	$\beta$	$p$	$S$	Ochiai	D <sup>2</sup>	Barinel	Op2	Tarantula	Kulczynski
Chart	17	20	$1e-5$	Ochiai	26%	43%	28%	59%	28%	23%
Closure	17	20	$1e-5$	Ochiai	11%	10%	19%	7%	19%	14%
Lang	17	20	$1e-5$	Ochiai	15%	12%	30%	24%	30%	19%
Math	17	20	$1e-5$	Ochiai	14%	21%	13%	29%	13%	12%
Mockito	16	20	$1e-5$	Ochiai	27%	25%	24%	30%	24%	31%
Time	17	20	$1e-5$	Ochiai	23%	29%	7%	41%	7%	17%

  

Project	Hyperparameter values				Median % improvement in $\mathcal{E}\mathcal{X}\mathcal{A}\mathcal{M}$ score on base SBFL metrics					
	$\mu$	$\beta$	$p$	$S$	Ochiai	D <sup>2</sup>	Barinel	Op2	Tarantula	Kulczynski
Chart	17	20	$1e-5$	Ochiai	19%	37%	23%	55%	23%	20%
Closure	17	20	$1e-5$	Ochiai	7%	8%	12%	4%	12%	11%
Lang	17	20	$1e-5$	Ochiai	10%	7%	23%	22%	23%	17%
Math	17	20	$1e-5$	Ochiai	10%	19%	11%	26%	11%	10%
Mockito	16	20	$1e-5$	Ochiai	24%	23%	20%	26%	20%	27%
Time	17	20	$1e-5$	Ochiai	19%	26%	4%	38%	4%	10%

Table 3: Percentage of faulty components that are ranked withinin top-n for project Chart.

top-n	ARTEMIS	Ochiai	D <sup>2</sup>	Barinel	Op2	Tarantula	Kulczynski
$n = 1$	1%	0%	0%	1%	0%	1%	0%
$n = 2$	6%	5%	5%	5%	0%	5%	5%
$n = 5$	19%	19%	17%	13%	2%	13%	19%
$n = 10$	34%	30%	26%	21%	7%	21%	31%
$n = 20$	43%	42%	33%	41%	13%	41%	42%
$n = 50$	72%	64%	54%	70%	40%	70%	66%

Table 4: Percentage of faulty components that are ranked withinin top-n for project Closure.

top-n	ARTEMIS	Ochiai	D <sup>2</sup>	Barinel	Op2	Tarantula	Kulczynski
$n = 1$	20%	16%	20%	6%	20%	6%	14%
$n = 2$	35%	27%	31%	10%	35%	10%	24%
$n = 5$	59%	41%	45%	22%	57%	22%	39%
$n = 10$	76%	55%	61%	41%	71%	41%	55%
$n = 20$	84%	82%	82%	69%	82%	69%	82%
$n = 50$	100%	98%	98%	100%	98%	100%	100%

Table 5: Percentage of faulty components that are ranked withinin top-n for project Lang.

top-n	ARTEMIS	Ochiai	D <sup>2</sup>	Barinel	Op2	Tarantula	Kulczynski
$n = 1$	5%	5%	5%	1%	1%	1%	2%
$n = 2$	15%	8%	11%	3%	5%	3%	2%
$n = 5$	27%	21%	26%	9%	17%	9%	19%
$n = 10$	49%	45%	44%	28%	34%	28%	36%
$n = 20$	72%	71%	72%	58%	67%	58%	67%
$n = 50$	92%	89%	90%	86%	88%	86%	89%

Table 6: Percentage of faulty components that are ranked withinin top-n for project Math.

top-n	ARTEMIS	Ochiai	D <sup>2</sup>	Barinel	Op2	Tarantula	Kulczynski
$n = 1$	8%	7%	5%	7%	4%	7%	7%
$n = 2$	14%	11%	8%	13%	7%	13%	10%
$n = 5$	27%	22%	20%	25%	14%	25%	22%
$n = 10$	52%	46%	40%	44%	34%	44%	47%
$n = 20$	73%	65%	60%	66%	52%	66%	66%
$n = 50$	88%	82%	80%	87%	75%	87%	83%

Table 7: Percentage of faulty components that are ranked withinin top-n for project Mockito.

top-n	ARTEMIS	Ochiai	D <sup>2</sup>	Barinel	Op2	Tarantula	Kulczynski
$n = 1$	26%	24%	26%	22%	14%	22%	24%
$n = 2$	40%	29%	29%	33%	17%	33%	29%
$n = 5$	69%	57%	53%	55%	52%	53%	55%
$n = 10$	81%	76%	72%	76%	71%	76%	71%
$n = 20$	95%	88%	88%	95%	88%	95%	88%
$n = 50$	100%	100%	100%	100%	100%	100%	100%

Table 8: Percentage of faulty components that are ranked withinin top-n for project Time.

top-n	ARTEMIS	Ochiai	D <sup>2</sup>	Barinel	Op2	Tarantula	Kulczynski
$n = 1$	18%	18%	18%	12%	7%	12%	18%
$n = 2$	23%	23%	23%	23%	11%	23%	23%
$n = 5$	43%	36%	34%	43%	21%	43%	26%
$n = 10$	54%	45%	43%	52%	27%	52%	52%
$n = 20$	64%	54%	48%	64%	48%	64%	59%
$n = 50$	84%	79%	79%	82%	73%	82%	82%