(*Supplementary File*) WITNESS: A lightweight and practical approach to fine-grained predictive mutation testing

ZEYU LU, State Key Laboratory for Novel Software Technology, Nanjing University, China
PENG ZHANG, University of Luxembourg, Luxembourg
CHUN YONG CHONG, Huawei Technologies Co., Ltd., China
SHAN GAO, Independent Researcher, China
YIBIAO YANG, YANHUI LI, LIN CHEN, YUMING ZHOU, State Key Laboratory for Novel Software Technology, Nanjing University, China

This document serves as the supplementary file for the paper titled "*WITNESS: A lightweight and practical approach to fine-grained predictive mutation testing*." Section 1 presents the project version along with the sizes of the training, validation, and test sets for experiments conducted under the same-version, cross-version, and cross-project scenarios. Section 2 presents the results of comparisons with PMT.

CCS Concepts: • **Software and its engineering → Software testing and debugging**.

Additional Key Words and Phrases: Mutation Testing, Machine Learning, Kill Matrix, Test Case Prioritization

ACM Reference Format:
Zeyu Lu, Peng Zhang, Chun Yong Chong, Shan Gao, and Yibiao Yang, Yanhui Li, Lin Chen, Yuming Zhou. 2025. (*Supplementary File*) WITNESS: A lightweight and practical approach to fine-grained predictive mutation testing. 1, 1 (August 2025), 8 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 DATASET

We present the project version along with the sizes of the training set, validation set, and test set for experiments conducted under the same-version, cross-version, and cross-project scenarios. The values in Tables 1 through 8 represent mutant-test pairs. Tables 1, 2, 3, and 4 show the details of experiments conducted on mutant-test pairs where mutations occur inside source methods. Meanwhile, Tables 5, 6, 7, and 8 provide information on experiments conducted on all mutant-test pairs, including mutations occurring both inside and outside source methods.

Authors' addresses: Zeyu Lu, State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China, zeyulu@smail.nju.edu.cn; Peng Zhang, University of Luxembourg, Luxembourg, Luxembourg, peng.zhang@uni.lu; Chun Yong Chong, Huawei Technologies Co., Ltd., Hong Kong, China, chong.chun.yong@huawei.com; Shan Gao, Independent Researcher, Hang Zhou, China, gaoshan_cs@outlook.com; Yibiao Yang, Yanhui Li, Lin Chen, Yuming Zhou, State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China, {yangyibiao,yanhuili,lchen}@nju.edu.cn, zhouyuming@nju.edu.cn.

ACM XXXX-XXXX/2025/8-ART
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Table 1. Information on Experiments Conducted on Mutant-Test Pairs with Mutations Occurring Inside Source Methods in the Same-Version Scenario

Project Version	Training Set	Validation Set	Test Set
Chart_1	770,958	93,672	101,643
Chart_5	673,480	85,340	80,626
Chart_10	591,275	70,438	74,307
Chart_15	586,349	72,039	69,490
Chart_20	529,015	67,339	64,164
Chart_25	524,930	66,130	64,304
JacksonCore_25	420,177	51,362	57,348
JacksonCore_20	312,086	39,275	40,665
JacksonCore_15	250,588	31,516	32,202
JacksonCore_10	245,371	30,252	33,312
JacksonCore_5	165,347	20,842	20,000
JacksonCore_1	121,117	14,438	15,063
Gson_15	306,831	35,350	38,240
Gson_10	292,264	35,166	38,064
Gson_5	285,685	35,063	37,743
Gson_1	171,338	21,272	23,850
Lang_1	147,274	17,689	18,759
Lang_10	141,275	16,962	18,086
Lang_20	115,936	13,369	14,747
Lang_30	114,838	14,865	13,956
Lang_40	113,962	14,490	14,788
Lang_50	120,602	14,727	15,230
Lang_60	100,800	12,231	12,666
Cli_30	47,188	6,934	4,844
Cli_20	21,821	2,492	2,227
Cli_10	17,737	1,840	2,485
Cli_1	14,824	1,365	1,828
Csv_15	35,601	5,020	4,722
Csv_10	19,515	2,756	2,652
Csv_5	17,923	2,382	2,663
Csv_1	6,912	920	743

Table 2. Information on Experiments Conducted on Mutant-Test Pairs with Mutations Occurring Inside Source Methods in the Cross-Version Scenario

Project Version	Training Set	Validation Set	Test Set
Csv_1 → Csv_5	7,585	990	22,968
Csv_1 → Csv_10	7,585	990	24,923
Csv_1 → Csv_15	7,585	990	45,343
Csv_5 → Csv_10	20,838	2,130	24,923
Csv_5 → Csv_15	20,838	2,130	45,343
Csv_10 → Csv_15	22,754	2,169	45,343
Cli_1 → Cli_10	16,211	1,806	22,062
Cli_1 → Cli_20	16,211	1,806	26,540
Cli_1 → Cli_30	16,211	1,806	58,966
Cli_10 → Cli_20	20,125	1,937	26,540
Cli_10 → Cli_30	20,125	1,937	58,966
Cli_20 → Cli_30	24,298	2,242	58,966
Lang_60 → Lang_50	113,299	12,398	150,559
Lang_60 → Lang_40	113,299	12,398	143,240
Lang_60 → Lang_30	113,299	12,398	143,659
Lang_60 → Lang_20	113,299	12,398	144,052
Lang_60 → Lang_10	113,299	12,398	176,323
Lang_60 → Lang_1	113,299	12,398	183,722
Lang_50 → Lang_40	135,777	14,782	143,240
Lang_50 → Lang_30	135,777	14,782	143,659
Lang_50 → Lang_20	135,777	14,782	144,052
Lang_50 → Lang_10	135,777	14,782	176,323
Lang_50 → Lang_1	135,777	14,782	183,722
Lang_40 → Lang_30	130,910	12,330	143,659
Lang_40 → Lang_20	130,910	12,330	144,052
Lang_40 → Lang_10	130,910	12,330	176,323
Lang_40 → Lang_1	130,910	12,330	183,722
Lang_30 → Lang_20	129,633	14,026	144,052
Lang_30 → Lang_10	129,633	14,026	176,323
Lang_30 → Lang_1	129,633	14,026	183,722
Lang_20 → Lang_10	130,467	13,585	176,323
Lang_20 → Lang_1	130,467	13,585	183,722
Lang_10 → Lang_1	157,758	18,565	183,722
Gson_1 → Gson_5	194,606	21,854	358,491
Gson_1 → Gson_10	194,606	21,854	365,494
Gson_1 → Gson_15	194,606	21,854	380,421
Gson_5 → Gson_10	318,878	39,613	365,494
Gson_5 → Gson_15	318,878	39,613	380,421
Gson_10 → Gson_15	322,300	43,194	380,421
JacksonCore_1 → JacksonCore_5	135,346	15,272	206,189
JacksonCore_1 → JacksonCore_10	135,346	15,272	308,935
JacksonCore_1 → JacksonCore_15	135,346	15,272	314,306
JacksonCore_1 → JacksonCore_20	135,346	15,272	392,026
JacksonCore_1 → JacksonCore_25	135,346	15,272	528,887
JacksonCore_5 → JacksonCore_10	184,229	21,960	308,935
JacksonCore_5 → JacksonCore_15	184,229	21,960	314,306
JacksonCore_5 → JacksonCore_20	184,229	21,960	392,026
JacksonCore_5 → JacksonCore_25	184,229	21,960	528,887
JacksonCore_10 → JacksonCore_15	278,255	30,680	314,306
JacksonCore_10 → JacksonCore_20	278,255	30,680	392,026
JacksonCore_10 → JacksonCore_25	278,255	30,680	528,887
JacksonCore_15 → JacksonCore_20	282,252	32,054	392,026
JacksonCore_15 → JacksonCore_25	282,252	32,054	528,887
JacksonCore_20 → JacksonCore_25	351,063	40,963	528,887
Chart_25 → Chart_20	586,304	69,060	660,518
Chart_25 → Chart_15	586,304	69,060	727,878
Chart_25 → Chart_10	586,304	69,060	736,020
Chart_25 → Chart_5	586,304	69,060	839,446
Chart_25 → Chart_1	586,304	69,060	966,273
Chart_20 → Chart_15	594,722	65,796	727,878
Chart_20 → Chart_10	594,722	65,796	736,020
Chart_20 → Chart_5	594,722	65,796	839,446
Chart_20 → Chart_1	594,722	65,796	966,273
Chart_15 → Chart_10	655,009	72,869	736,020
Chart_15 → Chart_5	655,009	72,869	839,446
Chart_15 → Chart_1	655,009	72,869	966,273
Chart_10 → Chart_5	657,846	78,174	839,446
Chart_10 → Chart_1	657,846	78,174	966,273
Chart_5 → Chart_1	755,019	84,427	966,273

Table 3. Information on Experiments Conducted on Mutant-Test Pairs with Mutations Occurring Inside Source Methods in the Cross-Project Scenario (One-to-One Prediction)

Project Version	Training Set	Validation Set	Test Set
Chart_1 → Csv_15	871,235	95,038	45,343
JacksonCore_25 → Csv_15	473,585	55,302	45,343
Gson_15 → Csv_15	337,957	42,464	45,343
Lang_1 → Csv_15	164,217	19,505	45,343
Cli_30 → Csv_15	53,226	5,740	45,343
Chart_1 → Cli_30	871,235	95,038	58,966
JacksonCore_25 → Cli_30	473,585	55,302	58,966
Gson_15 → Cli_30	337,957	42,464	58,966
Lang_1 → Cli_30	164,217	19,505	58,966
Csv_15 → Cli_30	41,267	4,076	58,966
Chart_1 → Lang_1	871,235	95,038	183,722
JacksonCore_25 → Lang_1	473,585	55302	183,722
Gson_15 → Lang_1	337,957	42,464	183,722
Cli_30 → Lang_1	53,226	5,740	183,722
Csv_15 → Lang_1	41,267	4,076	183,722
Chart_1 → Gson_15	871,235	95,038	380,421
JacksonCore_25 → Gson_15	473,585	55302	380421
Lang_1 → Gson_15	164,217	195,05	380,421
Cli_30 → Gson_15	53,226	5,740	380,421
Csv_15 → Gson_15	41,267	4,076	380,421
Chart_1 → JacksonCore_25	871,235	95,038	528,887
Gson_15 → JacksonCore_25	337,957	42,464	528,887
Lang_1 → JacksonCore_25	164,217	19,505	528,887
Cli_30 → JacksonCore_25	53,226	5,740	528,887
Csv_15 → JacksonCore_25	41,267	4,076	528,887
JacksonCore_25 → Chart_1	473,585	55,302	966,273
Gson_15 → Chart_1	337,957	42464	966,273
Lang_1 → Chart_1	164,217	19,505	966,273
Cli_30 → Chart_1	53,226	5,740	966,273
Csv_15 → Chart_1	41,267	4,076	966,273

Table 4. Information on Experiments Conducted on Mutant-Test Pairs with Mutations Occurring Inside Source Methods in the Cross-Project Scenario (Many-to-One Prediction)

Project Version	Training Set	Validation Set	Test Set
JacksonCore Gson Lang Cli → Csv	1,039,033	112,963	45,343
JacksonCore Gson Lang Csv → Cli	1,025,464	112,909	58,966
JacksonCore Gson Cli Csv → Lang	913,641	99,976	183,722
JacksonCore Lang Cli Csv → Gson	738,220	78,698	380,421
Gson Lang Cli Csv → JacksonCore	602,586	65,866	528,887
Chart Cli Csv → Lang	961,346	109,236	183,722
Chart Cli Csv → Gson	961,346	109,236	380,421
Chart Cli Csv → JacksonCore	961,346	109,236	528,887

Table 5. Information on Experiments Conducted on All Mutant-Test Pairs in the Same-Version Scenario

Project Version	**Training Set**	**Validation Set**	**Test Set**
Chart_1	882,566	109,514	112,330
Chart_5	777,009	100,082	90,974
Chart_10	684,779	84,809	85,929
Chart_15	679,424	84,837	81,401
Chart_20	615,698	79,942	75,086
Chart_25	613,008	78,173	74,306
JacksonCore_25	542,571	66,164	72,574
JacksonCore_20	394,409	52147	49,159
JacksonCore_15	325,328	40,480	41,486
JacksonCore_10	319,662	39,104	41,448
JacksonCore_5	218,283	28,840	25,440
JacksonCore_1	158,847	19,736	19,783
Gson_15	333,338	43,705	36,257
Gson_10	319,887	36,806	40,384
Gson_5	312,543	37,676	39,541
Gson_1	179,800	21,685	24,691
Lang_1	179,144	22,574	22,230
Lang_10	171,453	22,064	22,413
Lang_20	142,451	16,999	18,723
Lang_30	138,411	17,588	15,673
Lang_40	136,516	17,291	17,100
Lang_50	141,676	17,012	17,113
Lang_60	118,554	14,489	14,201
Cli_30	47,188	6,934	4,844
Cli_20	21,821	2,492	2,227
Cli_10	17,737	1,840	2,485
Cli_1	14,824	1,365	1,828
Csv_15	38,281	5,282	5,261
Csv_10	20,740	2,763	2,890
Csv_5	19,087	2,440	2,815
Csv_1	7205	954	792

Table 6. Information on Experiments Conducted on All Mutant-Test Pairs in the Cross-Version Scenario

Project Version	Training Set	Validation Set	Test Set
Csv_1 → Csv_5	8,143	808	24,342
Csv_1 → Csv_10	8,143	808	26,393
Csv_1 → Csv_15	8,143	808	48,824
Csv_5 → Csv_10	21,973	2,369	26,393
Csv_5 → Csv_15	21,973	2,369	48,824
Csv_10 → Csv_15	23,904	2,489	48,824
Cli_1 → Cli_10	16,211	1,806	22,062
Cli_1 → Cli_20	16,211	1,806	26,540
Cli_1 → Cli_30	16,211	1,806	58,966
Cli_10 → Cli_20	20,125	1,937	26,540
Cli_10 → Cli_30	20,125	1,937	58,966
Cli_20 → Cli_30	24,298	2,242	58,966
Lang_60 → Lang_50	132,599	14,645	175,801
Lang_60 → Lang_40	132,599	14,645	170,907
Lang_60 → Lang_30	132,599	14,645	171,672
Lang_60 → Lang_20	132,599	14,645	178,173
Lang_60 → Lang_10	132,599	14,645	215,930
Lang_60 → Lang_1	132,599	14,645	223,948
Lang_50 → Lang_40	158,804	16,997	170,907
Lang_50 → Lang_30	158,804	16,997	171,672
Lang_50 → Lang_20	158,804	16,997	178,173
Lang_50 → Lang_10	158,804	16,997	215,930
Lang_50 → Lang_1	158,804	16,997	223,948
Lang_40 → Lang_30	152,448	18,459	171,672
Lang_40 → Lang_20	152,448	18,459	178,173
Lang_40 → Lang_10	152,448	18,459	215,930
Lang_40 → Lang_1	152,448	18,459	223,948
Lang_30 → Lang_20	152,901	18,771	178,173
Lang_30 → Lang_10	152,901	18,771	215,930
Lang_30 → Lang_1	152,901	18,771	223,948
Lang_20 → Lang_10	160,635	17,538	215,930
Lang_20 → Lang_1	160,635	17,538	223,948
Lang_10 → Lang_1	193,306	22,624	223,948
Gson_1 → Gson_10	202,320	23,856	397,077
Gson_1 → Gson_15	202,320	23,856	413,300
Gson_5 → Gson_10	339,824	49,936	397,077
Gson_5 → Gson_15	339,824	49,936	413,300
Gson_10 → Gson_15	350,596	46,481	413,300
JacksonCore_1 → JacksonCore_5	179,063	19,303	272,563
JacksonCore_1 → JacksonCore_10	179,063	19,303	400,214
JacksonCore_1 → JacksonCore_15	179,063	19,303	407,294
JacksonCore_1 → JacksonCore_20	179,063	19,303	495,715
JacksonCore_1 → JacksonCore_25	179,063	19,303	681,309
JacksonCore_5 → JacksonCore_10	244,052	28,511	400,214
JacksonCore_5 → JacksonCore_15	244,052	28,511	407,294
JacksonCore_5 → JacksonCore_20	244,052	28,511	495,715
JacksonCore_5 → JacksonCore_25	244,052	28,511	681,309
JacksonCore_10 → JacksonCore_15	358,662	41,552	407,294
JacksonCore_10 → JacksonCore_20	358,662	41,552	495,715
JacksonCore_10 → JacksonCore_25	358,662	41,552	681,309
JacksonCore_15 → JacksonCore_20	367,049	40,245	495,715
JacksonCore_15 → JacksonCore_25	367,049	40,245	681,309
JacksonCore_20 → JacksonCore_25	449,177	46,538	681,309
Chart_25 → Chart_20	688,058	77,429	770,726
Chart_25 → Chart_15	688,058	77,429	845,662
Chart_25 → Chart_10	688,058	77,429	855,517
Chart_25 → Chart_5	688,058	77,429	968,065
Chart_25 → Chart_1	688,058	77,429	1,104,410
Chart_20 → Chart_15	691,828	78,898	845,662
Chart_20 → Chart_10	691,828	78,898	855,517
Chart_20 → Chart_5	691,828	78,898	968,065
Chart_20 → Chart_1	691,828	78,898	1,104,410
Chart_15 → Chart_10	760,818	84,844	855,517
Chart_15 → Chart_5	760,818	84,844	968,065
Chart_15 → Chart_1	760,818	84,844	1,104,410
Chart_10 → Chart_5	768,412	87,105	968,065
Chart_10 → Chart_1	768,412	87,105	1,104,410
Chart_5 → Chart_1	869,105	98,960	1,104,410

Table 7. Information on Experiments Conducted on All Mutant-Test Pairs in the Cross-Project Scenario (One-to-One Prediction)

Project Version	Training Set	Validation Set	Test Set
Chart_1 → Csv_15	993,321	111,089	48,824
JacksonCore_25 → Csv_15	617,904	63,405	48,824
Gson_15 → Csv_15	370,722	42,578	48,824
Lang_1 → Csv_15	201,028	22,920	48,824
Cli_30 → Csv_15	53,226	5,740	48,824
Chart_1 → Cli_30	993,321	111,089	58,966
JacksonCore_25 → Cli_30	617,904	63,405	58,966
Gson_15 → Cli_30	370,722	42,578	58,966
Lang_1 → Cli_30	201,028	22,920	58,966
Csv_15 → Cli_30	43,786	5,038	58,966
Chart_1 → Lang_1	993,321	111,089	223,948
JacksonCore_25 → Lang_1	617,904	63,405	223,948
Gson_15 → Lang_1	370,722	42,578	223,948
Cli_30 → Lang_1	53,226	5,740	223,948
Csv_15 → Lang_1	43,786	5,038	223,948
Chart_1 → Gson_15	993,321	111,089	413,300
JacksonCore_25 → Gson_15	617,904	63405	413,300
Lang_1 → Gson_15	201,028	22,920	413,300
Cli_30 → Gson_15	53,226	5,740	413,300
Csv_15 → Gson_15	43,786	5,038	413,300
Chart_1 → JacksonCore_25	993,321	111,089	681,309
Gson_15 → JacksonCore_25	370,722	42,578	681,309
Lang_1 → JacksonCore_25	201,028	22,920	681,309
Cli_30 → JacksonCore_25	53,226	5,740	681,309
Csv_15 → JacksonCore_25	43,786	5,038	681,309
JacksonCore_25 → Chart_1	617,904	63,405	1,104,410
Gson_15 → Chart_1	370,722	42,578	1,104,410
Lang_1 → Chart_1	201,028	22,920	1,104,410
Cli_30 → Chart_1	53,226	5,740	1,104,410
Csv_15 → Chart_1	43,786	5,038	1,104,410

Table 8. Information on Experiments Conducted on All Mutant-Test Pairs in the Cross-Project Scenario (Many-to-One Prediction)

Project Version	Training Set	Validation Set	Test Set
JacksonCore Gson Lang Cli → Csv	1,234,919	142,604	48,824
JacksonCore Gson Lang Csv → Cli	1,225,346	142,035	58,966
JacksonCore Gson Cli Csv → Lang	1,078,665	123,734	223,948
JacksonCore Lang Cli Csv → Gson	907,865	105,182	413,300
Gson Lang Cli Csv → JacksonCore	671,677	73,361	681,309
Chart Cli Csv → Lang	1,093,766	118,434	223,948
Chart Cli Csv → Gson	1,093,766	118,434	413,300
Chart Cli Csv → JacksonCore	1,093,766	118,434	681,309

Table 9. Average Predictive Performance of PMT and WITNESS

	PMT			WITNESS			
	Same-Version	**Cross-Version**	**Cross-Project**	**Same-Version**	**Cross-Version**	**Cross-Project**	
Mutant Killing Prediction	0.62	0.69	0.37	**0.69**	**0.73**	**0.38**	Precision
	0.52	0.47	0.14	**0.70**	**0.73**	**0.43**	Recall
	0.56	0.54	0.17	**0.69**	**0.72**	**0.39**	F1-score
Mutation Score Prediction	4.70	8.51	15.66	**4.07**	**4.15**	**7.84**	APE

2 COMPARISON WITH PMT

As PMT [1] is a coarse-grained predictive mutation testing approach, we compare it in terms of mutant killing prediction (the overall killing results of each mutant by the entire test suite) and mutation score prediction. Since PMT is suitable for predicting mutants occurring outside source methods, we compare PMT and WITNESS on all mutants. The experimental setup and evaluation metrics align with those described in our paper.

Table 9 presents the average predictive performance of PMT and WITNESS in mutant killing prediction and mutation score prediction. Better values in Table 9 are highlighted in bold. Clearly, WITNESS outperforms PMT across all evaluation metrics. The comparison results of Seshat indicate that Seshat outperforms PMT. Since the results in our study show that WITNESS outperforms Seshat, WITNESS is expected to outperform PMT as well.

REFERENCES

[1] Jie Zhang, Lingming Zhang, Mark Harman, Dan Hao, Yue Jia, and Lu Zhang. 2018. Predictive Mutation Testing. *IEEE Transactions on Software Engineering* 45, 9 (2018), 898–918.