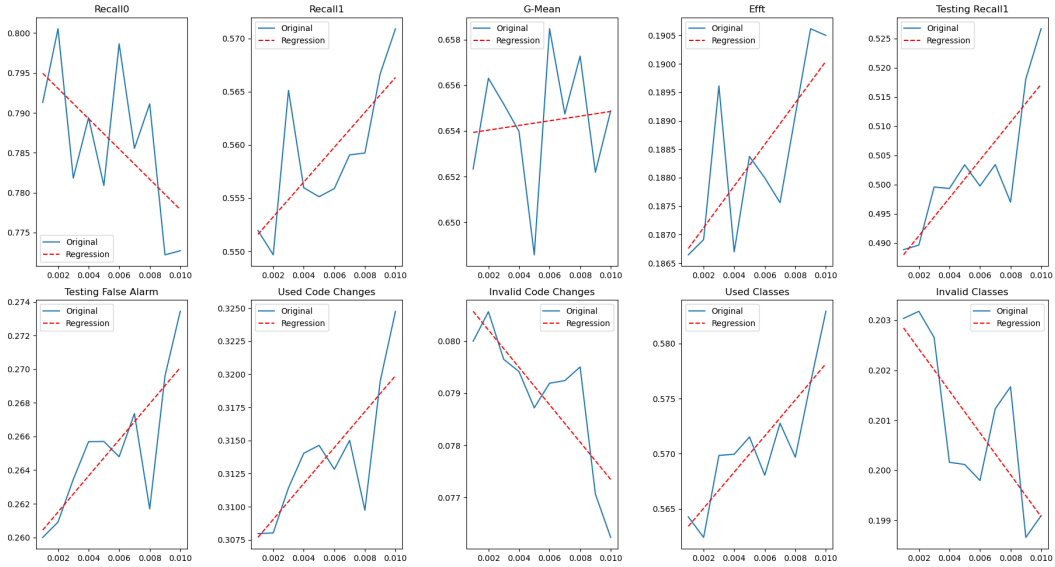
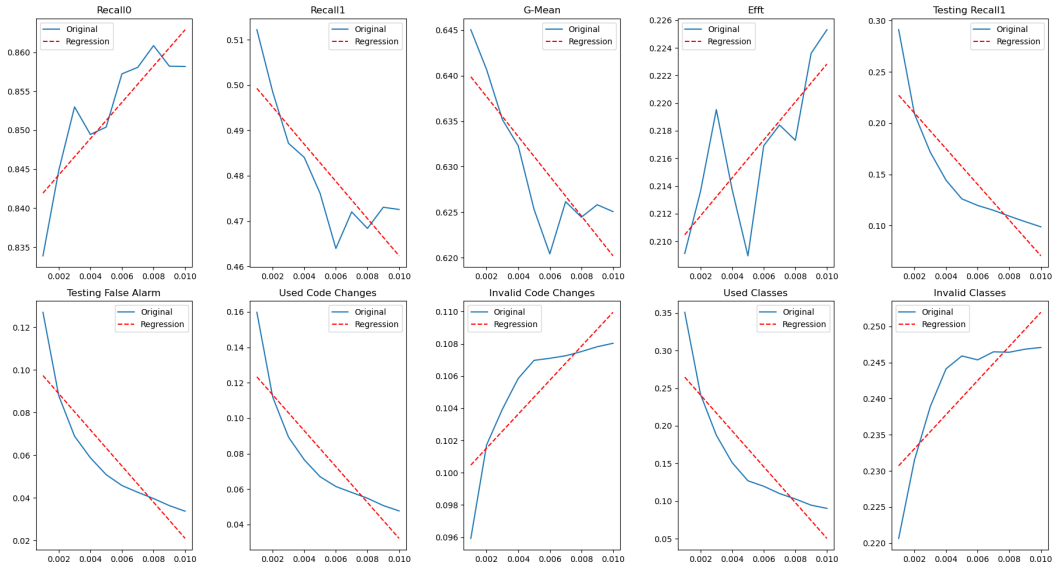


(a) $\mathcal{S}_4$



(b) $\mathcal{S}_5$

Fig. 1. The horizontal axis is the parameter values and the vertical axis is the corresponding metric values. The blue solid line (Original) is a line graph of the raw data, and the red dashed line (Regression) is a regression curve fitted to the raw data, showing the trend of the raw data with the hyperparameter.

Figure 1(a) and 1(b) demonstrates the effect of hyperparameters on the $\mathcal{S}_4$ and $\mathcal{S}_5$ separately in terms of various metrics.

For $\mathcal{S}_4$, when the value of hyperparameters is larger Recall 1 and Testing Recall 1 are also larger, indicating that AuDITee's ability to find defects increases, but Used Code Changes and Used Classes also become larger, indicating that AuDITee's resource conservation rate decreases. For $\mathcal{S}_5$, when the value of hyperparameters is larger Recall 1 and Testing Recall 1 are lower, indicating that AuDITee's ability to find defects decreases, but Used Code Changes and Used Classes also become lower, indicating that AuDITee's resource conservation rate increases. This supports the viewpoints we mentioned in the paper when analyzing α in $\mathcal{S}_4$ and $\mathcal{S}_5$.

Compared to $\mathcal{S}_5$, $\mathcal{S}_4$ focuses more on improving performance, while $\mathcal{S}_4$ is more focused on resource savings.

We also analyze each metric introduced in the paper and aim to find the overall optimal parameters (balancing performance and resource savings). Specifically, we rank each parameter under each metric (the lower the rank, the better) and ultimately select the parameter with the smallest average rank to present. It is important to note that parameters significantly impact various metrics of the model, and we only present the overall optimal results in the subsequent experiments. If a developer is only concerned with a specific metric, they can select better results accordingly.

We display the parameters we ultimately chose for different projects in table 1.

Table 1. The result of analyzing selector $\mathcal{S}_4$ and $\mathcal{S}_5$.

	project	hyperparameter	Recall 1	Recall 0	G-Mean	EFF_t	TR_1	$FalseAlarm_T$	UCC	IRCC	UC	IRC
S4	JGroups	0.003	0.6263	0.6678	0.6448	0.2003	0.4463	0.2050	0.2400	0.0757	0.4937	0.2901
	Broadleaf	0.007	0.3934	0.8635	0.5823	0.1633	0.3790	0.2063	0.2252	0.0775	0.4678	0.1570
	Tomcat	0.004	0.6808	0.8057	0.7401	0.2087	0.7064	0.3884	0.4839	0.0813	0.7671	0.1541
S5	JGroups	0.001	0.5608	0.7665	0.6555	0.2102	0.2983	0.1270	0.1495	0.0865	0.3747	0.2888
	Broadleaf	0.001	0.3289	0.8889	0.5402	0.1678	0.1952	0.1003	0.1122	0.0882	0.2867	0.1815
	Tomcat	0.009	0.6297	0.8617	0.7365	0.3172	0.1963	0.0532	0.0884	0.1260	0.1422	0.2139

In the paper, we will use the selected parameters to compare with other methods and conduct a more in-depth analysis.