

Refactoring Type	Zero-shot ( $\rightarrow$ Python)	CL Few-shot (Java $\rightarrow$ Python)	CL Few-shot with desc (Java $\rightarrow$ Python)	Native Few-shot (Python $\rightarrow$ Python)
Used in RQ1	○	○		
Used in RQ2		○		○
Used in RQ3		○	○	○
EXTRACT METHOD	0.640 (0.522, 0.828)	0.697 (0.622, 0.793)	0.754 (0.719, 0.793)	<b>0.758</b> (0.676, 0.862)
INLINE METHOD	0.459 (0.359, 0.636)	0.583 (0.539, 0.636)	<b>0.619</b> (0.650, 0.591)	0.520 (0.464, 0.591)
RENAME VARIABLE	0.490 (0.632, 0.400)	0.546 (0.600, 0.500)	0.542 (0.722, 0.433)	<b>0.552</b> (0.571, 0.533)
EXTRACT VARIABLE	0.312 (0.255, 0.400)	0.429 (0.462, 0.400)	0.393 (0.387, 0.400)	<b>0.452</b> (0.438, 0.467)
INLINE VARIABLE	<b>0.250</b> (0.189, 0.368)	0.125 (0.154, 0.105)	0.125 (0.154, 0.105)	0.125 (0.154, 0.105)
Average	0.430	0.501	<b>0.512</b>	0.506

Table 1: F1-Scores for refactoring detection by prompt strategy using GPT-4.1-mini. **Bold** values indicate the highest F1 score for each refactoring type. Values in brackets are Precision and Recall, respectively.