F1-Score Comparison Across Complexity Levels