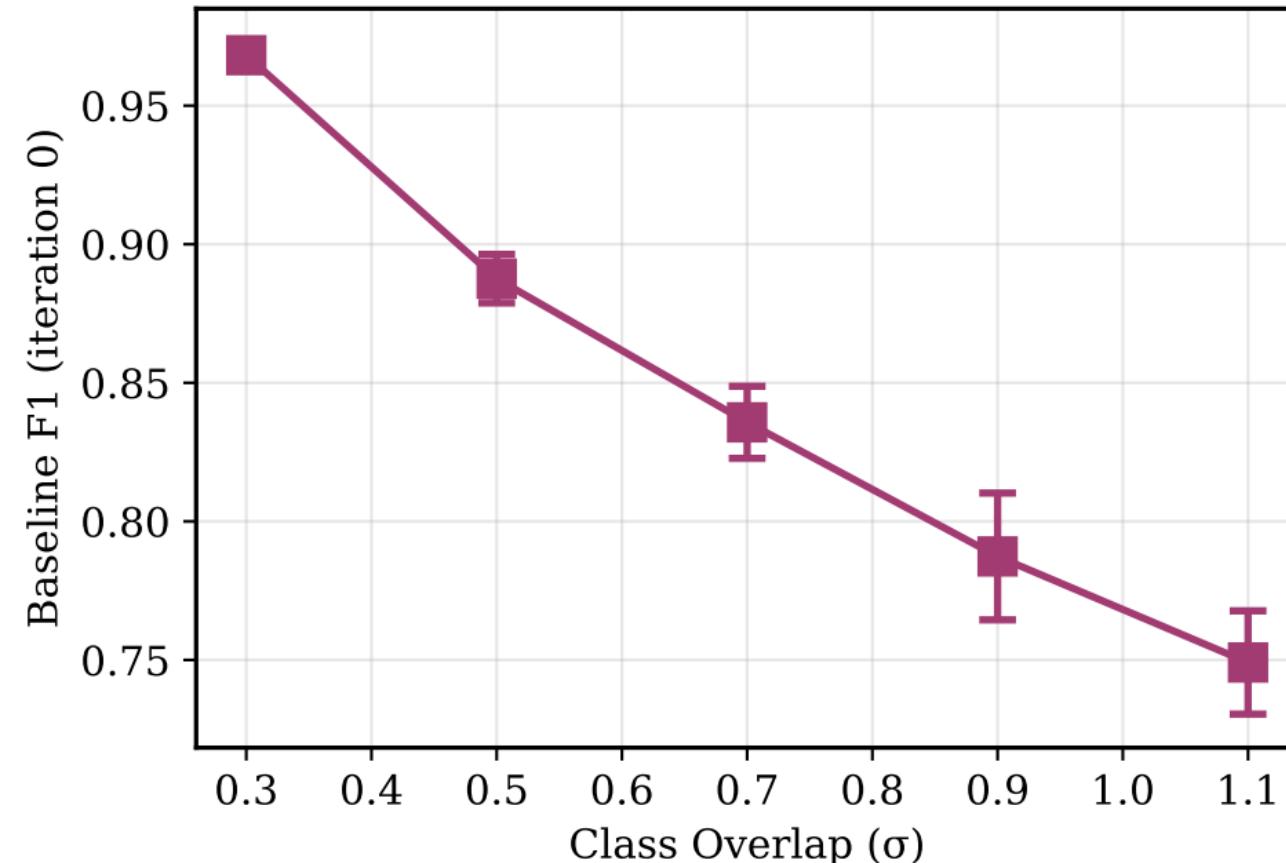


(A) Task Difficulty Effect on F1



**(B) β Remains Low Despite Difficulty
(No False Positives)**

