F1 Scores per Label (Three-way comparison, Test Splits)

