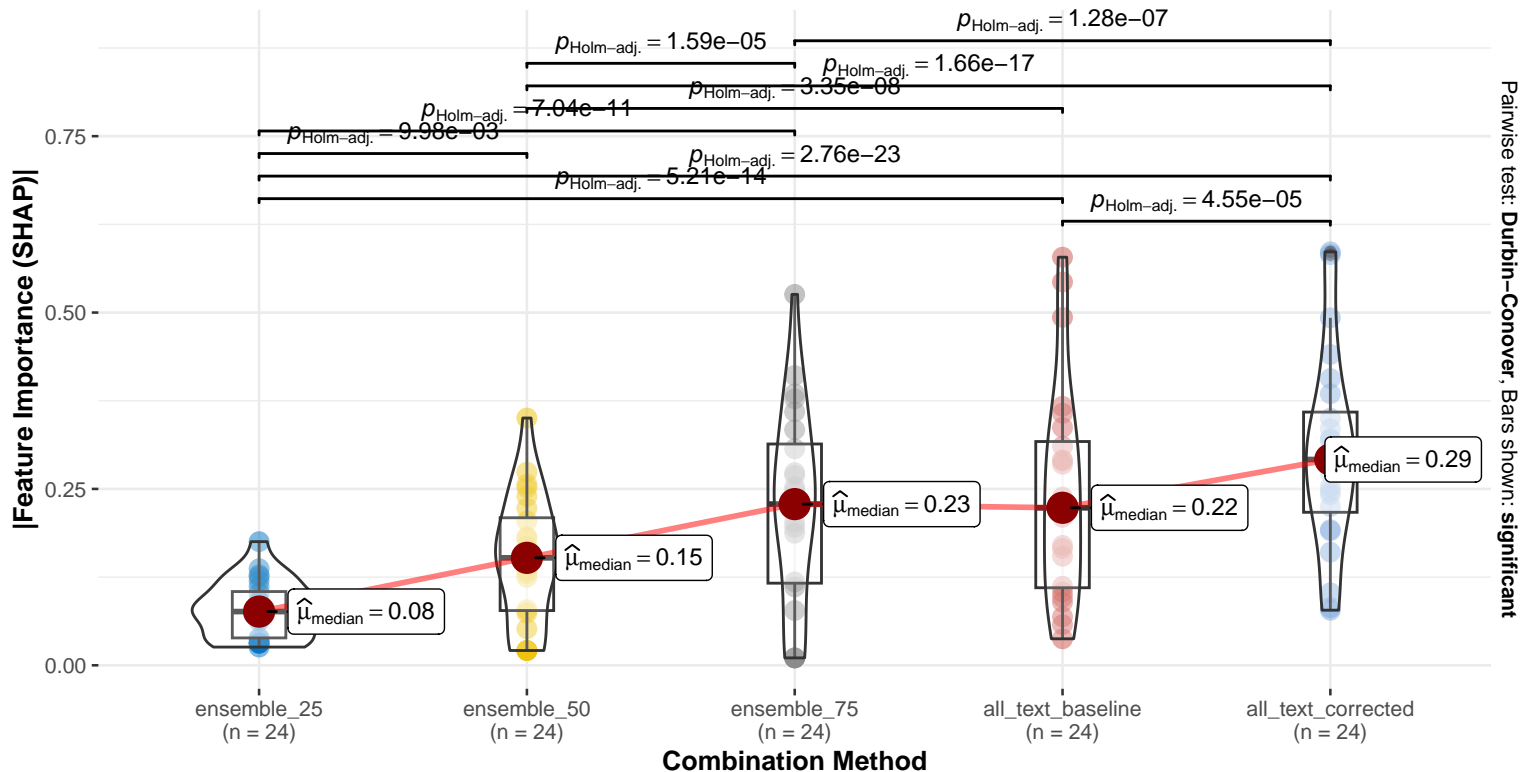**Are the Same Features Always the Most Important?**
Absolute Feature Importance (SHAP) Compared, by Combination Method

$\chi^2_{\text{Friedman}}(4) = 69.23$, $p = 3.29e-14$, $\widehat{W}_{\text{Kendall}} = 0.72$, $\text{CI}_{95\%}$ [0.65, 1.00], $n_{\text{pairs}} = 24$

$p_{\text{Holm-adj.}} = 1.28e-07$

$p_{\text{Holm-adj.}} = 1.59e-05$

$p_{\text{Holm-adj.}} = 1.66e-17$

$p_{\text{Holm-adj.}} = 7.04e-11$

$p_{\text{Holm-adj.}} = 3.35e-08$

$p_{\text{Holm-adj.}} = 9.98e-03$

$p_{\text{Holm-adj.}} = 2.76e-23$

$p_{\text{Holm-adj.}} = 5.21e-14$

$p_{\text{Holm-adj.}} = 4.55e-05$

Pairwise test: **Durbin–Conover**, Bars shown: **significant**

|Feature Importance (SHAP)|

0.75

0.50

0.25

0.00

$\widehat{\mu}_{\text{median}} = 0.08$

$\widehat{\mu}_{\text{median}} = 0.15$

$\widehat{\mu}_{\text{median}} = 0.23$

$\widehat{\mu}_{\text{median}} = 0.22$

$\widehat{\mu}_{\text{median}} = 0.29$

ensemble_25
(n = 24)

ensemble_50
(n = 24)

ensemble_75
(n = 24)

all_text_baseline
(n = 24)

all_text_corrected
(n = 24)

Combination Method

Dataset: prod *Distilbert−stack model excluded