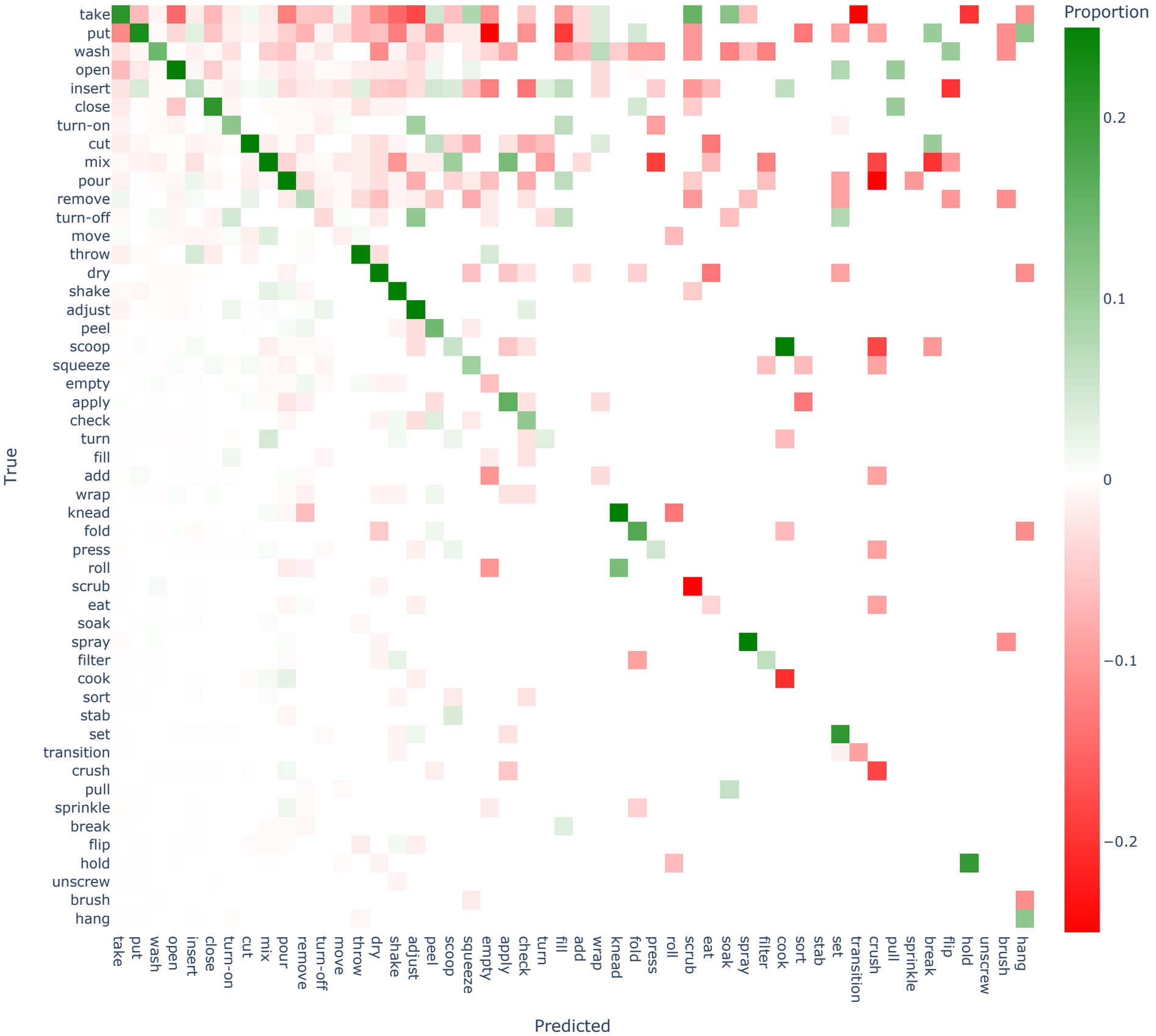Difference in confusion matrix for the top 50 verb classes between **multi-modal ensemble** and AudioSlowFastGRU