Appendix 1

Figure A1 shows the distribution of R-hat values across PBNN (0, 4) parameters aggregated over all active learning steps for four different case studies: ESOL, FreeSolv, Steel fatigue, and HTEM datasets. All cases demonstrate good convergence characteristics, with the majority of parameters having R-hat values close to 1.0. The distributions exhibit a right-skewed pattern, which is expected in MCMC convergence diagnostics. There are, however, variations between datasets - particularly, the Steel fatigue case shows a wider spread of R-hat values, which correlates with more volatile NLPD values and slower Coverage convergence in early active learning steps. Nevertheless, most of the weights and biases fall within the the range $1.0 < \text{R-hat} < 1.1$, which is traditonally considered to indicate good convergence.

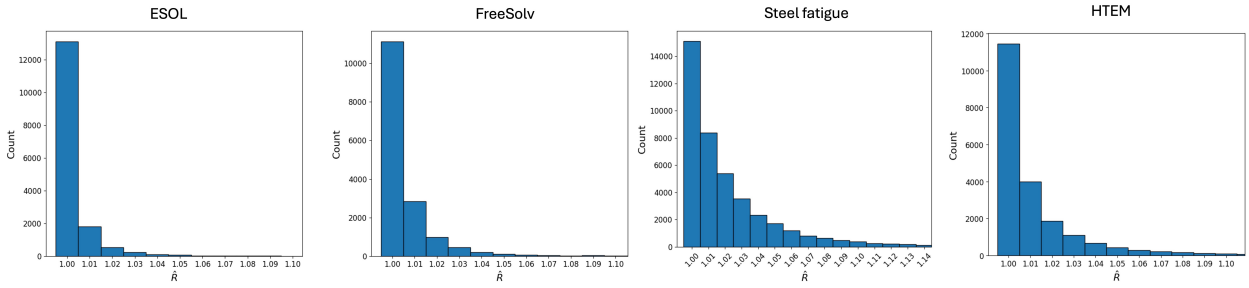

Figure A1: Gelman-Rubin 'R-hat' values over all active learning steps for ESOL, FreeSolv, Steel fatigue, and HTEM datasets.