As DeepDiagnosis identifies issues after integration, for a fair comparison, we computed the total time, i.e., summation of time taken by participants to resolve bugs in data and model stages separately using KUnit, and compared it with DeepDignosis. In our analysis shown in Fig. A, we found that participants took an average of 15 minutes and 12 minutes to debug the DL programs using DeepDiagnosis, and KUnit, respectively. For debugging using DeepDignosis, we observed that the quality of the preprocessed data significantly impacted the debugging time. Participants had to resolve data-related issues before addressing model-specific problems, thereby increasing the debugging time (Task 2 and Task 5). By isolating data and model, KUnit facilitates a more focused and efficient debugging process that can save time and resources during training.

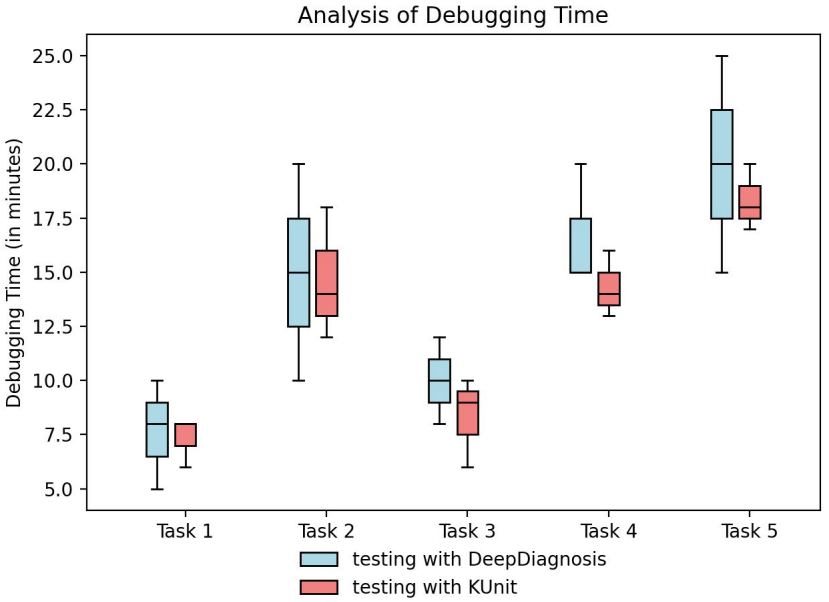

Fig. A

As KUnit enables parallel testing of both data and model, Fig. B presents the time participants took to resolve issues in the data and model stages separately. On average, we found that participants took 6 minutes to debug the data and 7 minutes to debug the model. This approach allows developers to address issues in both components simultaneously, rather than sequentially, streamlining the debugging process.

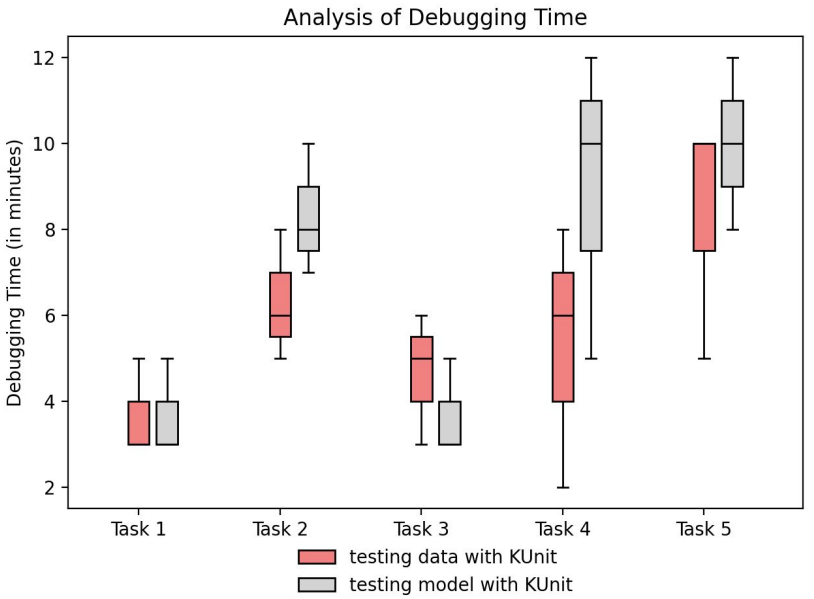

Fig. B