Science Advances
AAAS

Supplementary Note 7
(Details of Cross Domain Evaluation)
for

Leveraging Data Mining, Active Learning, and Domain Adaptation for Efficient Discovery of Advanced Oxygen Evolution Electrocatalysts

Rui Ding *et al.*

*Corresponding author. Email: junhongchen@uchicago.edu, chenyuxin@uchicago.edu

Supplementary Note Discussion SND 7-1

Corresponding scripts are stored and publicly available in the: "/ML Databases and Scripts/ DFT Domain Adaptation/Cross Domain Evaluation/" directory of the DASH online repository.

Cross Domain Evaluation

Except for the part we have presented in the main text, we also assessed the generalization capabilities of the DFT-ML part across different domains, as well as investigated the potential occurrence of catastrophic forgetting following domain adaptation. This evaluation was particularly focused on the performance of committee models on another domain, namely Committee S's generalization ability on dataset T, Committee T similarly on dataset S, and the adapted Committee S-T on dataset S to check the forgetting issues. Catastrophic forgetting refers to a situation where a neural network, after being trained on a new task, loses the knowledge acquired from previous tasks. To investigate this, the Committee S-T, which underwent domain adaptation (trained on both Dataset S and T), was evaluated on Dataset S. This evaluation aimed to determine whether the adapted models retained their performance on the original dataset (Dataset S) or if they exhibited signs of forgetting the learned patterns and features from their initial training.

The results presented in **Figs. SN7-1~SN7-4** convincingly demonstrate the presence of forgetting issues across different sub-tasks. Notably, while Committee S-T exhibited an exceptionally high R^2 value, exceeding 0.99 on domain B as depicted in **Fig. 6B** of the main text, this performance metric declined to a range of majorly between 0.8 to 0.9 following domain adaptation towards domain B. Despite this reduction, the observed R^2 values are still considerably acceptable, suggesting that Committee S-T has not succumbed to total or 'catastrophic' forgetting and has managed to retain qualitative correctness in its predictions.

This robustness is further supported by the diverse architecture of Committee S-T, which integrates various model types to enhance both adaptability and retention. While domain adaptation introduces challenges such as potential overwriting of previously learned patterns, particularly in highly flexible models, the ensemble approach mitigates these risks by leveraging diversity in learning mechanisms. This ensures that the overall workflow maintains a balance between generalization and domain-specific adaptation.

Furthermore, a stark contrast is observed when comparing these results with the largely negative R^2 values (majorly between -2~-5) and significantly increased MSE/MAE etc. in orders of magnitudes, exhibited by Committee S and Committee T when applied to each other's source dataset. Notably, in statistical terms, a negative R^2 value indicates a model's performance that is significantly worse than a simple, average-based prediction, underscoring the inadequacy of these models in cross-dataset applications. This contrast underscores the significant domain differences between domain S and T. Importantly, given that domain B encompasses all the necessary elements and aligns more closely with the research objectives, the findings reinforce the suitability and cost-effectiveness of Committee S-T as a robust solution for estimating properties from a DFT perspective. This conclusion not only highlights the adaptability of Committee S-T but also its potential as a valuable tool in advancing research in this domain. In conclusion, although the MAE and root mean squared error (RMSE) have indicated our committees' limited ability to directly predict the OER overpotential, the high R^2 value indicates our model's accuracy in capturing the overall trends and relative changes in the data. This level of reliability is crucial for predicting the direction and relative magnitude of changes in crystal structures, particularly given the vast array of candidate doping elements and ratios.

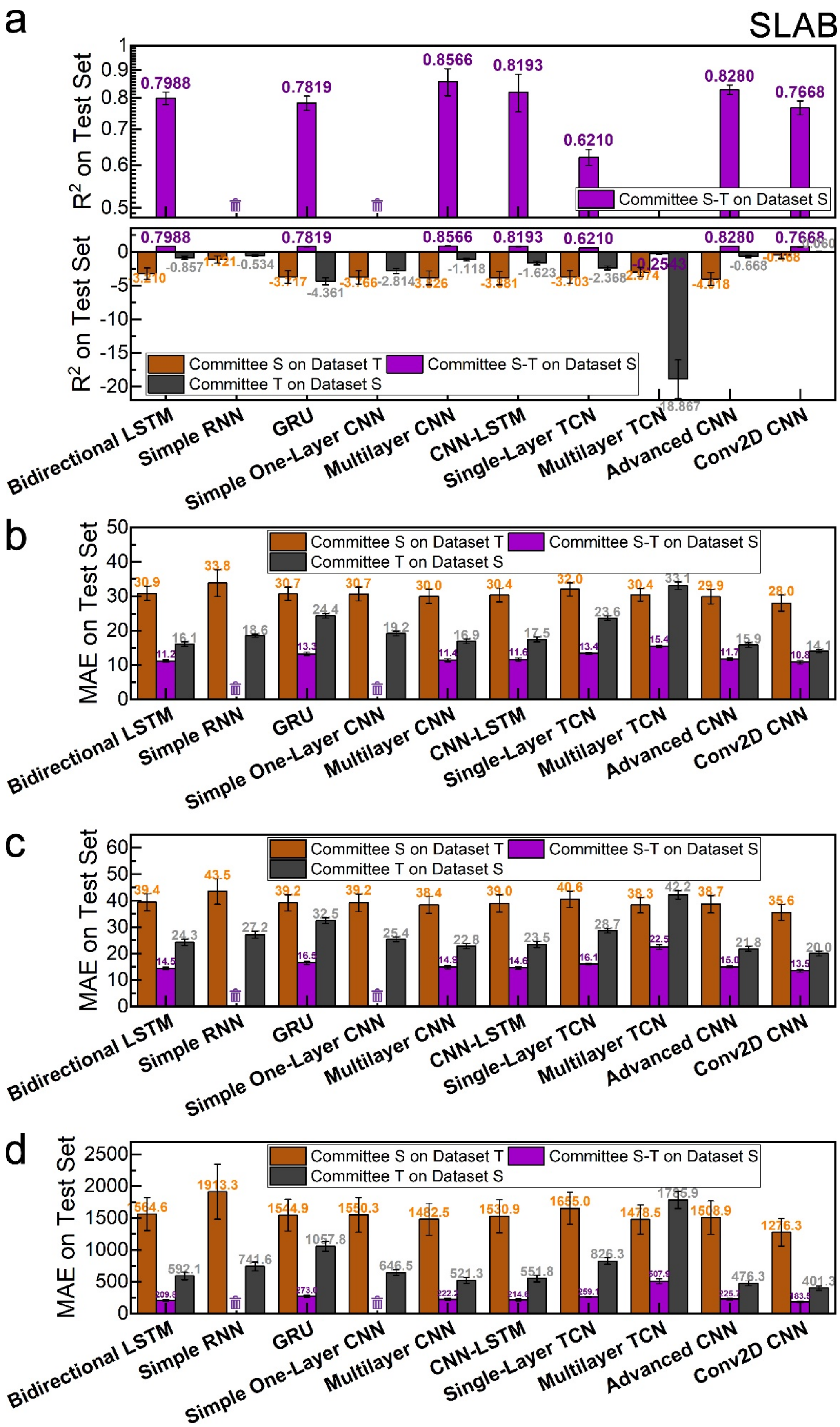

Fig. SN 7-1 Summary of performance metrics in cross-domain evaluation and forgetting test for the committee predicting the energy of the slab.

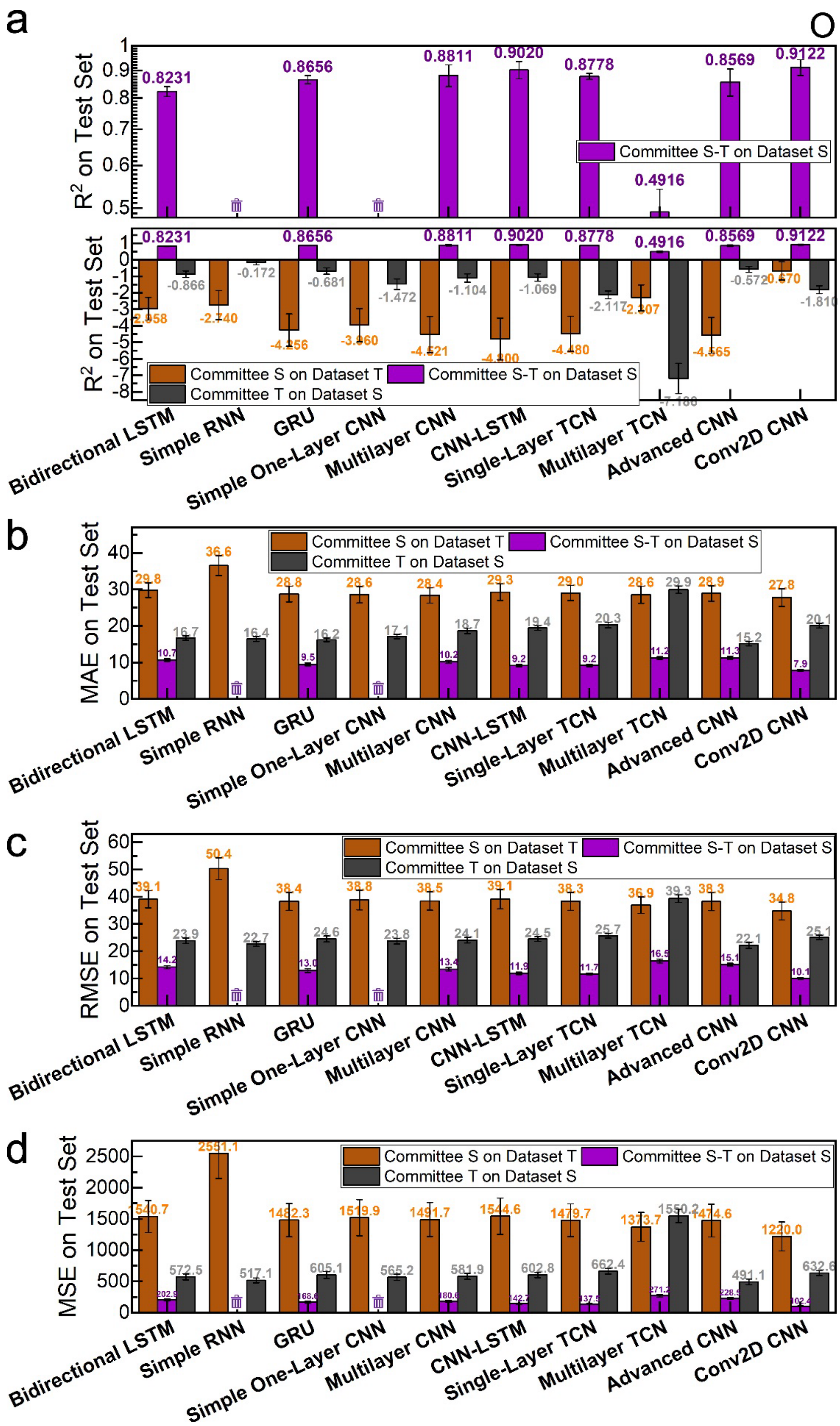

Fig. SN 7-2 Summary of performance metrics in cross-domain evaluation and forgetting test for the committee predicting the energy of the slab with O specie adsorbed.

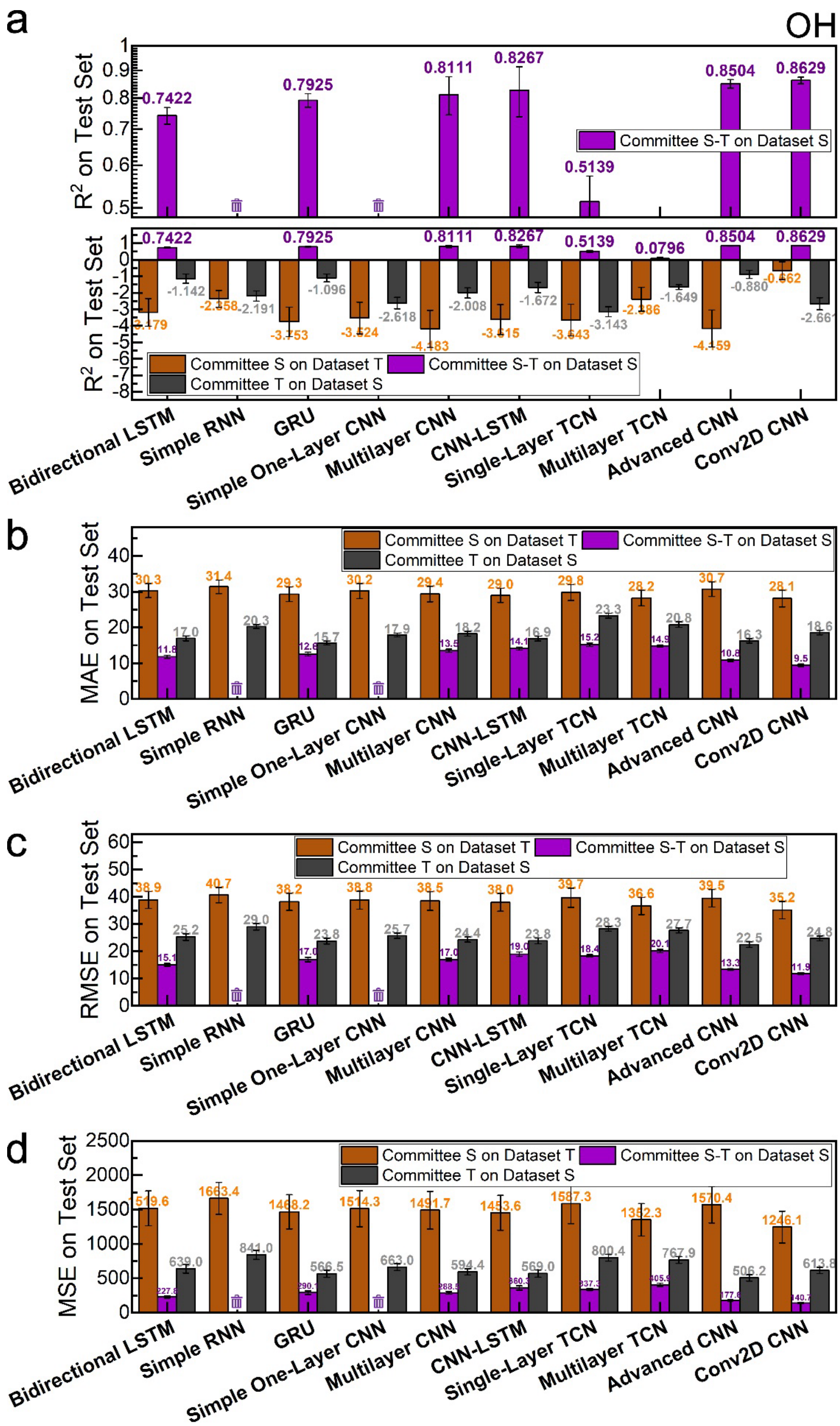

Fig. SN 7-3 Summary of performance metrics in cross-domain evaluation and forgetting test for the committee predicting the energy of the slab with OH specie adsorbed.

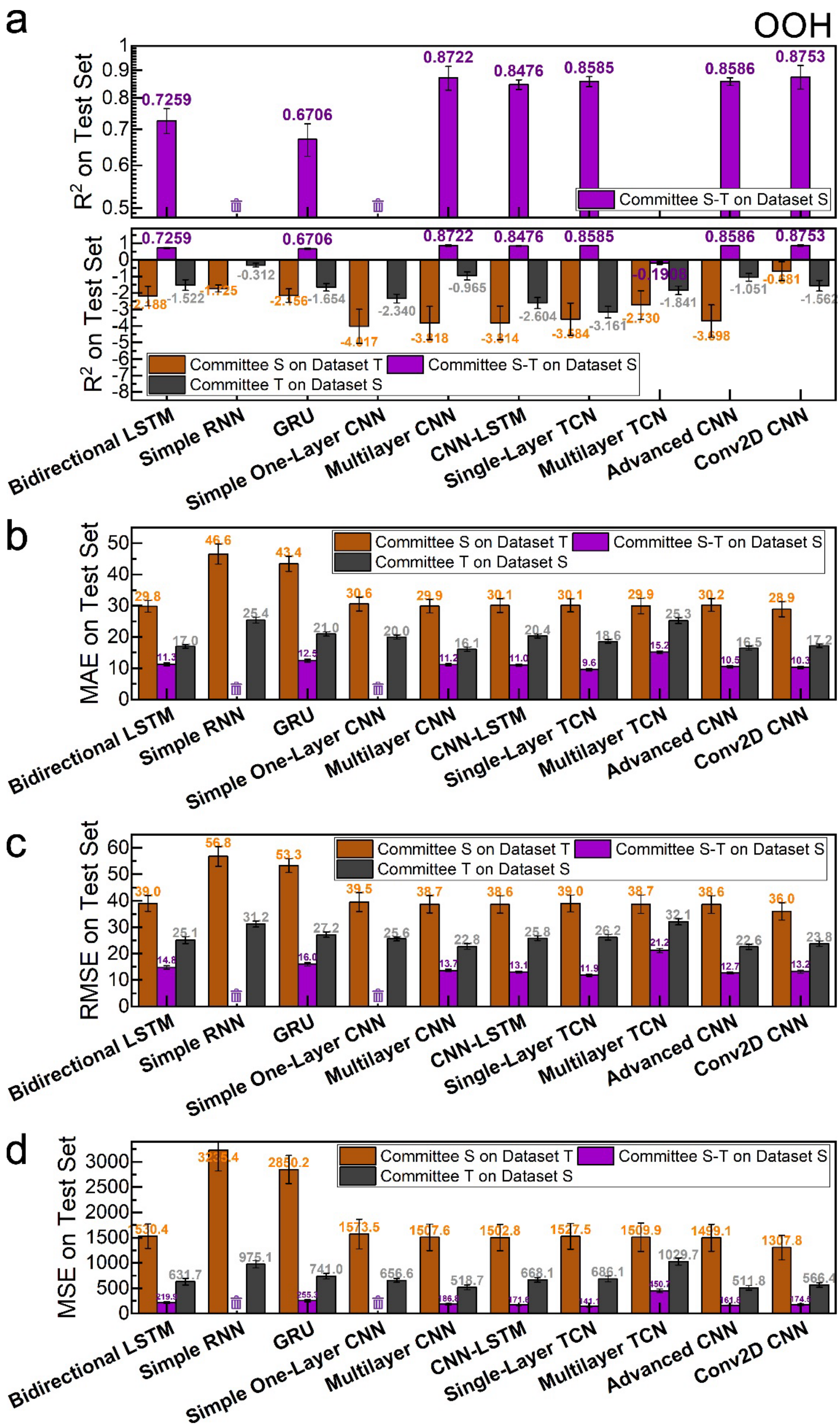

Fig. SN 7-4 Summary of performance metrics in cross-domain evaluation and forgetting test for the committee predicting the energy of the slab with OOH specie adsorbed.