

Multiple runs with different random seeds: as a supplement to the main text

Table. I. Forecasting results under 95% confidence interval evaluation

Model	RMSE [95% CI]	MAE [95% CI]	R ² [95% CI]
LSTM	[0.355, 0.387]	[0.249, 0.273]	[0.918, 0.926]
GRU	[0.356, 0.389]	[0.250, 0.274]	[0.913, 0.927]
BiLSTM	[0.353, 0.385]	[0.250, 0.273]	[0.920, 0.929]
BiGRU	[0.348, 0.380]	[0.250, 0.273]	[0.899, 0.906]
LSTM_QL	[0.247, 0.269]	[0.175, 0.191]	[0.938, 0.949]
GRU_QL	[0.263, 0.287]	[0.167, 0.182]	[0.931, 0.940]
BiLSTM_QL	[0.279, 0.304]	[0.179, 0.195]	[0.932, 0.942]
BiGRU_QL	[0.252, 0.275]	[0.161, 0.176]	[0.938, 0.945]
PPO_Ensemble	[0.247, 0.271]	[0.154, 0.169]	[0.943, 0.953]
HDRL	[0.245, 0.267]	[0.152, 0.165]	[0.951, 0.959]

Table. II. Statistical significance analysis using the Diebold-Mariano test

Model	Difference	DM Statistics	p-value	Significance
LSTM vs. HDRL	+0.1122	29.76	<0.001	
GRU vs. HDRL	+0.1136	27.46	<0.001	
BiLSTM vs. HDRL	+0.1102	27.38	<0.001	
BiGRU vs. HDRL	+0.1055	27.24	<0.001	
LSTM_QL vs. HDRL	-0.0013	-0.28	0.778	✗
GRU_QL vs. HDRL	+0.0185	4.45	<0.001	
BiLSTM_QL vs. HDRL	+0.0326	6.74	<0.001	
BiGRU_QL vs. HDRL	+0.0114	2.51	0.012	

Analysis:

1. Across multiple independent runs with 5 different random seeds on the GEFCom2014 dataset, Table I reports the 95% CIs of RMSE/MAE/R², which provides a complement to the main text. The four backbone baselines show consistently higher error ranges and lower R² than the enhanced methods. After applying the Q-learning adjustment, all base models improve markedly, with RMSE CIs reduced to around 0.247-0.304 and MAE CIs to 0.161-0.195, alongside higher R². Among all methods, HDRL achieves the best and most stable overall performance, indicating robust gains with generalization.

2. Table II further validates these observations using the Diebold-Mariano (DM) test against HDRL. HDRL shows clear statistical evidence of better predictive loss than all four plain backbones (all p < 0.001). Compared with the Q-learning variants, HDRL still outperforms GRU_QL, BiLSTM_QL, and BiGRU_QL with p-values below conventional thresholds (0.001/0.012), while LSTM_QL vs. HDRL is not distinguishable by the DM test, meaning their forecasting accuracy is essentially comparable under this test. Overall, the CI analysis and DM results jointly reinforce that the improvements on GEFCom2014 are reproducible across seeds and supported by statistical testing.

Validate on multiple publicly datasets (the Italian public dataset)

Dataset outline:

Name: Photovoltaic power and weather parameters.

Link: https://ieee-dataport.org/open-access/photovoltaic-power-and-weather-parameters?check_logged_in=1

Contains: PV power production measurements and corresponding weather/irradiance parameters.

Measurement site: SolarTech Lab, Politecnico di Milano, Italy.

Time resolution: 1 minute, and be aggregated one hour-long time resolution.

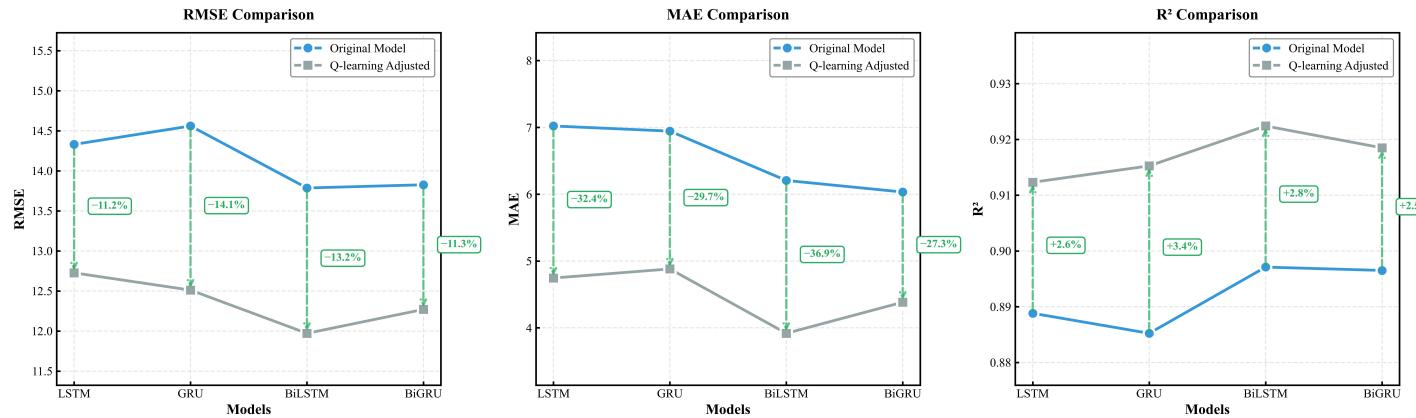


Fig. 1. Comparison of baseline and Q-learning adjusted models

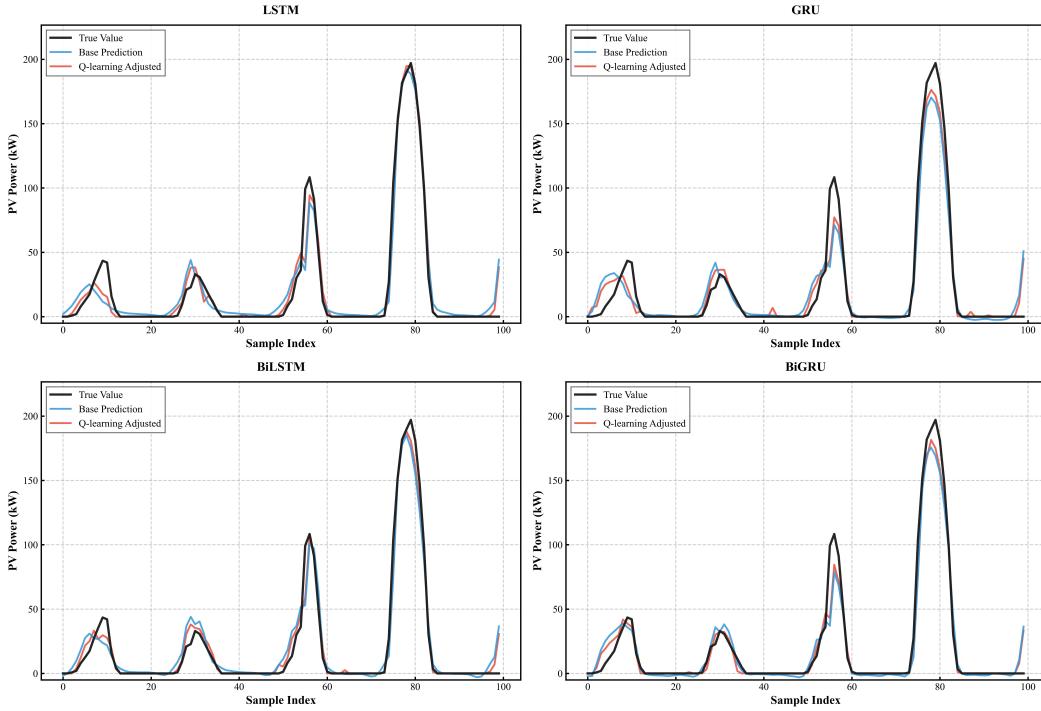


Fig. 2. Qualitative PV forecasting results before and after Q-learning adjustment

Analysis:

Figs. 1-2 show that the proposed Q-learning adjustment consistently improves PV forecasting performance on the new dataset across all four base models, demonstrating the generalization of the adjustment mechanism. More specifically, the Q-learning adjustment reduces RMSE by 11.2-14.1% and MAE by 27.3-36.9%, while increasing R^2 by 2.5-3.4%, indicating a robust error correction effect rather than a dataset specific gain. In Fig. 2, the adjusted outputs track the ground truth more closely than the base predictions, particularly around sharp ramps and high peaks, with reduced over/under-shooting and better peak alignment, confirming that the adjustment can refine bias and transient errors in challenging PV fluctuation.

Validate on multiple publicly datasets (the Italian public dataset)

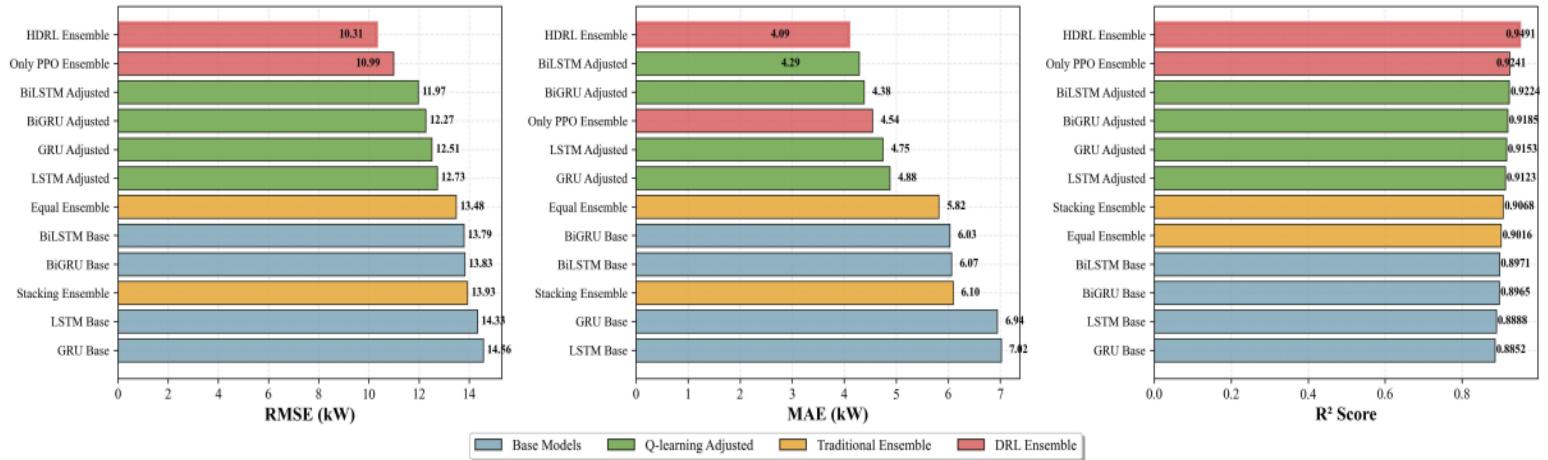


Fig. 3. Overall Performance comparison across different methods

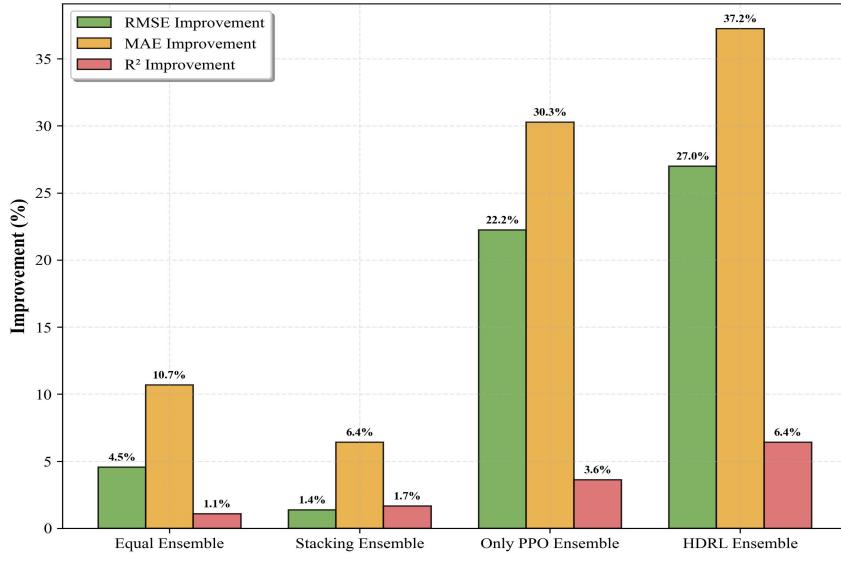


Fig. 4. Relative improvements of ensemble strategies in RMSE, MAE, and R²

Analysis:

In the Fig. 3 and Fig. 4, overall metric comparisons and relative improvement summaries are provided. The RMSE/MAE/R² bar charts indicate that the proposed HDRL ensemble scheme consistently achieves lower RMSE and MAE and higher R² than the base models and the conventional ensembles, suggesting the improvement is not limited to a single dataset. The improvement percentage figure further confirms stable gains across different ensemble settings with larger benefits for more advanced ensemble schemes, demonstrating that proposed method effectively exploits model generalization ability and generates robust performance.