CARD: Channel Aligned Robust Blend Transformer for Time Series Forecasting

Xue Wang Tian Zhou Qingsong Wen Jinyang Gao Bolin Ding Rong Jin

Code **Paper**

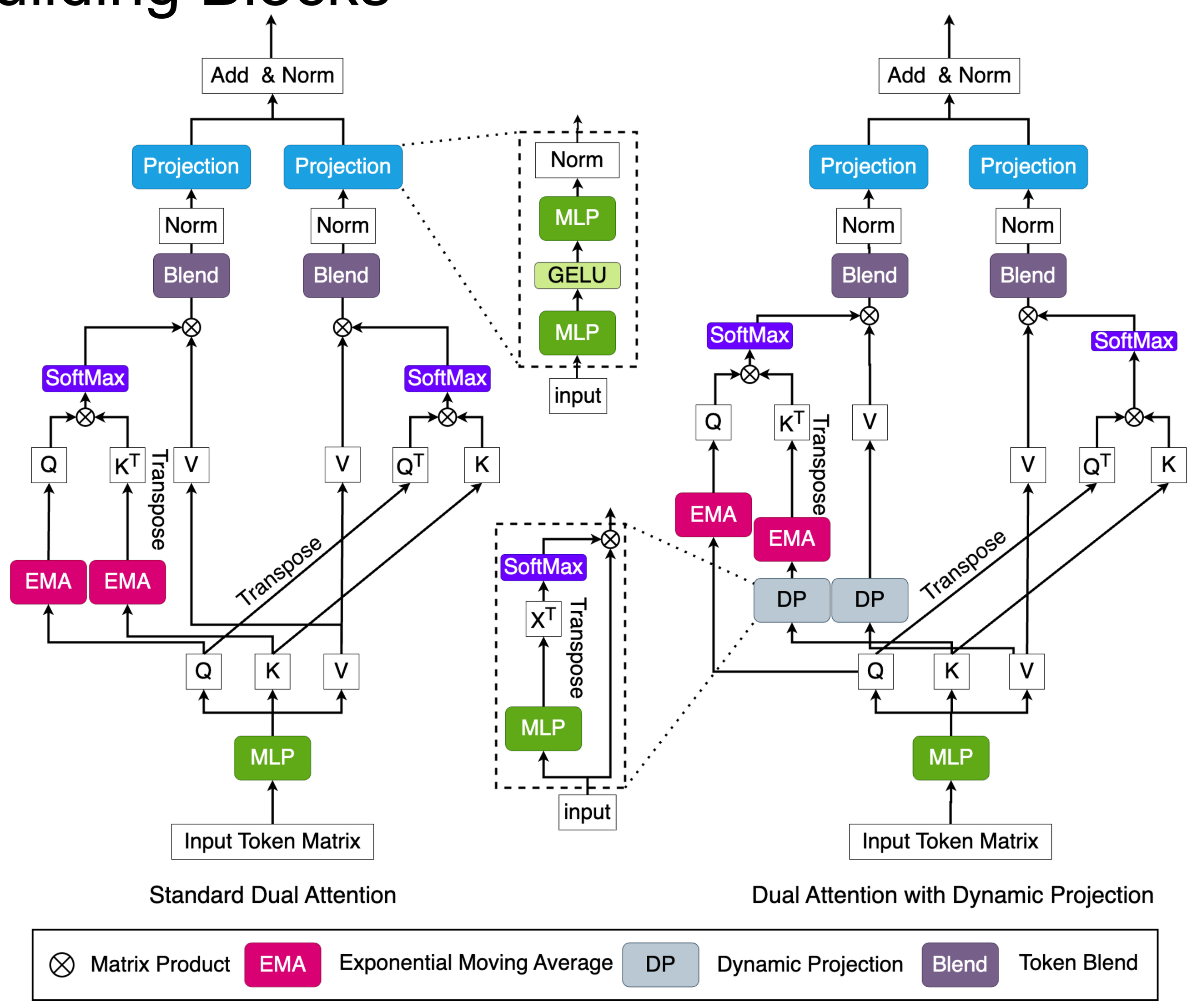
ICLR

Introduction & Contribution

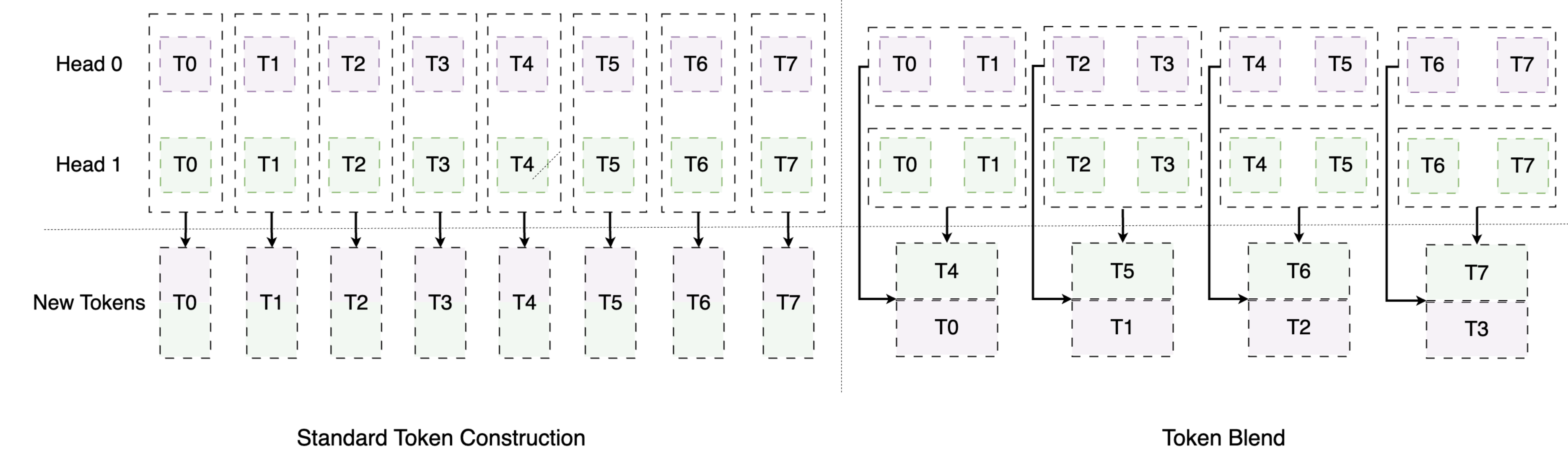
Recent studies have demonstrated the great power of Transformer models for time series forecasting. One of the key elements that lead to the transformer's success is the channel-independent (CI) strategy to improve the training robustness. However, the ignorance of the correlation among different channels in CI would limit the model's forecasting capacity. In this work:

- We propose a **C**hannel **A**ligned **R**obust Blen**d** Transformer (CARD) which efficiently and robustly aligns the information among different channels and utilizes the multi-scale information.
- CARD demonstrates superior performance in several benchmark datasets for forecasting and other prediction-based tasks, outperforming the state-of-the-art models. Our studies have confirmed the effectiveness of the proposed model.
- We develop a robust signal decay-based loss function that utilizes signal decay to bolster the model's ability to concentrate on forecasting for the near future. Our empirical assessment has confirmed that this loss function is effective in improving the performance of other benchmark models as well.

Building Blocks



Dual Attentions over both hidden dimension and token dimension



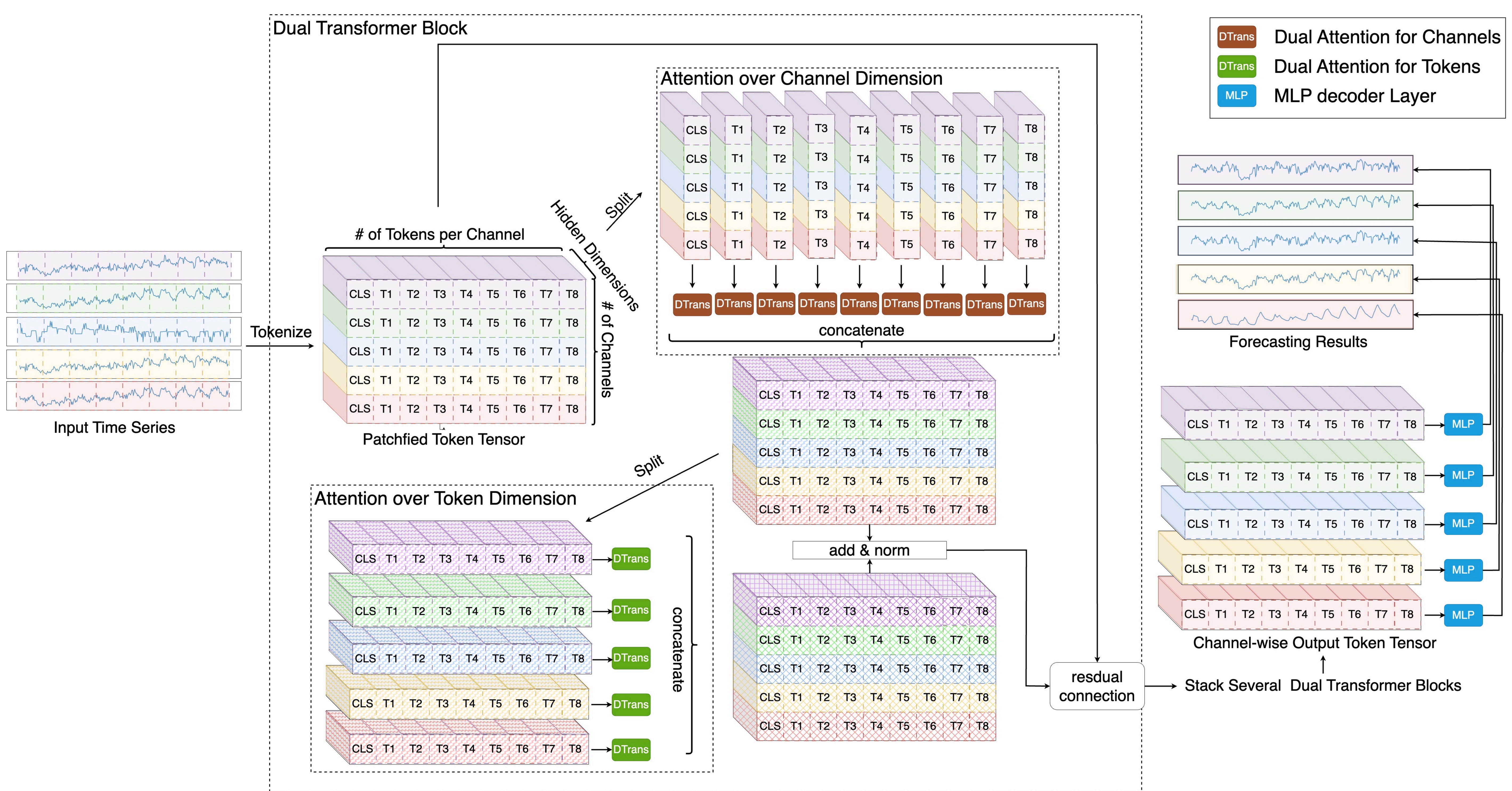
Multi-Resolution with Token Blend

$$\min \mathbb{E}_{\boldsymbol{A}}\left[\frac{1}{L}\sum_{l=1}^{L}\|\hat{\boldsymbol{a}}_{t+l}(\boldsymbol{A})-\boldsymbol{a}_{t+l}(\boldsymbol{A})\|_2^2\right] \Rightarrow \min \mathbb{E}_{\boldsymbol{A}}\left[\frac{1}{L}\sum_{l=1}^{L} l^{-1}\|\hat{\boldsymbol{a}}_{t+l}(\boldsymbol{A})-\boldsymbol{a}_{t+l}(\boldsymbol{A})\|_2^2\right]$$

$$\min \mathbb{E}_{\boldsymbol{A}}\left[\frac{1}{L}\sum_{l=1}^{L}\|\hat{\boldsymbol{a}}_{t+l}(\boldsymbol{A})-\boldsymbol{a}_{t+l}(\boldsymbol{A})\|_1\right] \Rightarrow \min \mathbb{E}_{\boldsymbol{A}}\left[\frac{1}{L}\sum_{l=1}^{L} l^{-1}\|\hat{\boldsymbol{a}}_{t+l}(\boldsymbol{A})-\boldsymbol{a}_{t+l}(\boldsymbol{A})\|_1\right]$$

Signal Decay based MSE/MAE Loss Functions

Architecture Overview of CARD



Experimental Results

Models	CARD		PatchTST		MICN		TimesNet		Crossformer		Dlinear		LightTS		FilM		ETSformer		FEDformer	
Metric	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTm1	**0.383**	**0.383**	0.395	0.408	0.387	0.411	0.400	0.406	0.435	0.417	0.403	0.407	0.435	0.437	0.408	0.399	0.429	0.425	0.448	0.452
ETTm2	**0.271**	**0.316**	0.283	0.327	0.284	0.340	0.291	0.333	0.609	0.521	0.350	0.401	0.409	0.436	0.287	0.328	0.292	0.342	0.305	0.349
ETTh1	0.443	**0.429**	0.455	0.444	**0.440**	0.462	0.458	0.450	0.486	0.481	0.456	0.452	0.491	0.479	0.461	0.456	0.452	0.510	**0.440**	0.460
ETTh2	**0.367**	**0.390**	0.384	0.406	0.402	0.437	0.414	0.427	0.966	0.690	0.559	0.515	0.602	0.543	0.384	0.406	0.439	0.452	0.437	0.449
Weather	**0.240**	**0.262**	0.257	0.280	0.243	0.299	0.259	0.287	0.250	0.310	0.265	0.317	0.261	0.312	0.269	0.339	0.271	0.334	0.309	0.360
Electricity	**0.169**	**0.258**	0.216	0.318	0.187	0.295	0.192	0.295	0.273	0.363	0.212	0.300	0.229	0.329	0.223	0.303	0.208	0.323	0.214	0.327
Traffic	**0.450**	**0.278**	0.488	0.327	0.542	0.316	0.620	0.336	0.593	0.332	0.625	0.383	0.622	0.392	0.639	0.389	0.621	0.396	0.610	0.376

Long term Forecasting with 96 input length

Models		CARD	PatchTST	MICN	TimesNet	N-HiTS	N-BEATS	ETSformer	LightTS	Dlinear	FEDformer	Autoformer	Informer
Yearly	SMAPE	**13.215**	13.258	14.935	13.387	13.418	13.436	18.009	14.247	16.965	13.728	13.974	14.727
	MASE	**2.972**	2.985	3.523	2.996	3.045	3.043	4.487	3.109	4.283	3.048	3.134	3.418
	OWA	**0.778**	0.781	0.900	0.786	0.793	0.794	1.115	0.827	1.058	0.803	0.822	0.881
Quarterly	SMAPE	**9.958**	10.179	11.452	10.100	10.202	10.124	13.376	11.364	12.145	10.792	11.338	11.360
	MASE	**1.163**	1.212	1.389	1.182	1.194	1.169	1.906	1.328	1.520	1.283	1.365	1.401
	OWA	**0.876**	0.904	1.026	0.890	0.899	0.886	1.302	1.000	1.106	0.958	1.012	1.027
Monthly	SMAPE	**12.414**	12.641	13.773	12.670	12.791	12.667	14.588	14.014	13.514	14.260	13.958	14.062
	MASE	**0.907**	0.930	1.076	0.933	0.969	0.937	1.368	1.053	1.037	1.102	1.103	1.141
	OWA	**0.856**	0.867	0.983	0.878	0.899	0.880	1.149	0.981	0.956	1.012	1.002	1.024
Others	SMAPE	**4.522**	4.851	6.716	4.891	5.061	4.925	7.267	15.880	6.709	4.954	5.458	24.460
	MASE	**3.021**	3.238	4.717	3.302	3.216	3.391	5.240	11.434	4.953	3.264	3.865	20.960
	OWA	**0.962**	1.021	1.451	1.035	1.040	1.053	1.591	3.474	1.487	1.036	1.187	5.879
Avg	SMAPE	**11.614**	11.807	13.130	11.829	11.927	11.851	14.718	13.252	13.639	12.840	12.909	14.086
	MASE	**1.553**	1.590	1.896	1.585	1.613	1.599	2.408	2.111	2.095	1.701	1.771	2.718
	OWA	**0.832**	0.834	0.980	0.851	0.861	0.855	1.172	1.051	1.051	0.918	0.939	1.230

Short term M4 forecasting

Models	CARD	PatchTST	MICN	TimesNet	Crossformer	ETSformer	LightTS	Dlinear	FEDformer	Stationary	Autoformer	Informer
SMD	**0.872**	0.866	0.800	0.858	0.778	0.831	0.825	0.771	0.851	0.847	0.851	0.855
MSL	0.817	0.823	0.816	**0.852**	0.820	0.850	0.790	0.849	0.786	0.775	0.791	0.841
SMAP	**0.857**	0.695	0.656	0.715	0.674	0.695	0.692	0.693	0.708	0.711	0.711	0.699
SWaT	**0.945**	0.909	0.875	0.921	0.886	0.849	0.933	0.875	0.932	0.799	0.927	0.814
PSM	0.957	0.951	0.933	**0.975**	0.921	0.918	0.972	0.936	0.972	0.973	0.933	0.771
Avg	**0.890**	0.849	0.816	0.864	0.816	0.829	0.842	0.825	0.849	0.821	0.843	0.789

Forecasting based anomaly detection