

To comprehensively evaluate the performance of FreFilterTST, we evaluated six real-world datasets covering six domains. The anomaly rates vary from 0.57% to 15.07%, feature dimensions range from 2 to 55, and sequence lengths range from 5,040 to 284,807. This substantial diversity in the datasets enables a comprehensive study of MTSAD (Multivariate Time Series Anomaly Detection) methods. Table 1 lists the statistics for the six multivariate time series.

Our supported metrics fall into two classes: score-based and label-based. The label-based metrics consist of Accuracy (Acc), Precision (P), Recall (R), F1-Score (F1), Range-Precision (R-P), Range-Recall (R-R), Range-F1-Score (R-F), Affiliation-Precision (Aff-P), Affiliation-Recall (Aff-R), and Affiliation-F1-Score (Aff-F). The score-based metrics comprise the Area under the Precision-Recall curve (A-P), Area under the Receiver Operating Characteristic curve (A-R), Range-Area under the PR curve (R-A-P), Range-Area under the ROC curve, Volume under the PR Surface (V-PR), and Volume under the ROC Surface (V-ROC).

For details, the performance of FreFilterTST is reported in Table 2 for label-based metrics and Table 3 for score-based metrics.

To support these metric results, we also conduct a comparison with Modern, a state-of-the-art (SOTA) model from 2024. The details are presented in Table 4.

Dataset	Domain	Dim	AR (%)	Avg TotalLength	Avg TestLength	Description
Calit2	Visitors Flowrate	2	4.09	5,040	2,520	This dataset originates from a monitoring system at the main entrance of a building on the University of California, Irvine (UCI) campus.
MSL	Spacecraft	55	5.88	132,046	73,729	This dataset consists of telemetry data collected from the Curiosity rover of the Mars Science Laboratory (MSL) mission.
NYC	Transport	3	0.57	17,520	4,416	This dataset documents the operational data of taxi services in New York City.
GECCO	Water	9	1.25	5,040	69,261	develop an event detector to accurately predict any kinds of changes in a time series of drinking water composition data.
SEU	Gearbox	29	15.07	284,807	142,404	The dataset covers five conditions: one healthy state and four types of gear faults, which are chipped tooth, missing tooth, root crack, and wear.
PSM	Server Machine	25	11.07	220,322	87,841	This dataset comprises multivariate time series data collected from the server clusters of a large internet company.

Table 1. Introduction of six datasets

Dataset	Accuracy	Recall	Precision	F_Score	adjust_accuracy	adjust_f_score	adjust_precision	adjust_recall
Calit2	0.944	0.135	0.115	0.124	0.957	0.443	0.358	0.581
MSL	0.851	0.095	0.157	0.118	0.939	0.764	0.647	0.932
NYC	0.978	0.0101	1	0.02	1	1	1	1
GECCO	0.985	0.827	0.393	0.533	0.987	0.610	0.439	0.516
SEU	0.981	0.735	0.057	0.106	0.981	0.106	0.057	0.735
PSM	0.731	0.063	0.650	0.115	0.982	0.968	0.966	0.970

Dataset	R_Recall	R_Precision	R_F	affiliation_f	affiliation_precision	affiliation_recall	Fit Time	Infer Time
Calit2	0.221	0.103	0.140	0.807	0.710	0.931	2.52	0.651
MSL	0.239	0.122	0.162	0.723	0.581	0.956	334.094	258.581
NYC	0.208	1	0.344	0.994	1	0.988	8.086	2.831
GECCO	0.851	0.055	0.102	0.912	0.840	0.998	43.308	39.838
SEU	0.711	0.0524	0.0976	0.749	0.618	0.949	198.793	254.629
PSM	0.438	0.558	0.490	0.840	0.774	0.917	190.743	177.558

Table 2. Label Performance of FreFilterTST,The values highlighted in yellow are the final evaluation metrics for this paper.

Dataset	AUC_ROC	AUC_PR	R_AUC_PR	VUS_P_R	R_AUC_ROC	VUS_ROC	Fit Time	Infer Time
Calit2	0.847	0.112	0.126	0.123	0.853	0.847	2.937	0.075
MSL	0.618	0.151	0.244	0.240	0.706	0.695	271.582	0.657
NYC	0.842	0.0849	0.118	0.113	0.838	0.829	4.007	0.015
GECCO	0.966	0.468	0.489	0.487	0.986	0.983	9.006	0.108
SEU	0.957	0.093	0.054	0.052	0.919	0.918	317.959	0.621
PSM	0.647	0.435	0.435	0.437	0.633	0.633	78.139	0.292

Table 3. score Performance of FreFilterTST,The values highlighted in yellow are the final evaluation metrics for this paper.

FreFilterTST												
Dataset	Acc	P	R	F1	R-R	R-P	R-F	A-P	R-A-P	V-P-R	R-A-R	V-ROC
Calit2	0.944	0.114	0.135	0.124	0.221	0.102	0.14	0.112	0.126	0.123	0.853	0.847
MSL	0.851	0.095	0.157	0.118	0.239	0.122	0.162	0.151	0.244	0.240	0.706	0.695
NYC	0.978	0.0101	1	0.02	0.208	1	0.344	0.085	0.118	0.113	0.838	0.829
GECCO	0.985	0.827	0.793	0.533	0.851	0.055	0.102	0.468	0.489	0.487	0.986	0.983
SEU	0.981	0.835	0.735	0.106	0.711	0.052	0.098	0.093	0.054	0.052	0.919	0.918
PSM	0.731	0.630	0.650	0.115	0.438	0.558	0.490	0.435	0.435	0.437	0.633	0.633

Mordern												
Dataset	Acc	P	R	F1	R-R	R-P	R-F	A-P	R-A-P	V-P-R	R-A-R	V-ROC
Calit2	0.889	0.074	0.243	0.114	0.318	0.069	0.094	0.054	0.070	0.070	0.726	0.716
MSL	0.853	0.142	0.091	0.113	0.194	0.121	0.155	0.146	0.224	0.220	0.708	0.701
NYC	0.972	0	0	0	0	0	0	0.020	0.038	0.037	0.598	0.579
GECCO	0.984	0.373	0.779	0.381	0.644	0.084	0.152	0.447	0.459	0.461	0.978	0.975
SEU	0.973	0.811	0.723	0.104	0.693	0.051	0.096	0.088	0.054	0.051	0.917	0.908
PSM	0.722	0.653	0.047	0.089	0.376	0.553	0.448	0.385	0.383	0.384	0.588	0.589

Table 4. Comparison of FreFilterTST and Mordern on Various Metrics,Superior results are highlighted in bold.