A PROOF OF THEOREM

A.1 Proof of Theorem 4.1

Proof of Theorem 4.1. Based on Eq. (10) and by using the triangle inequality, we have

$$||\nabla_{\Theta}\mathcal{L}_i - \nabla_{\Theta}\mathcal{L}_j|| = ||\nabla_F\mathcal{L}_i\cdot\nabla_{\Theta}F_i - \nabla_F\mathcal{L}_j\cdot\nabla_{\Theta}F_j|| \le \Delta_A + \Delta_B + \Delta_C, \tag{18}$$

where

$$\Delta_A = ||\nabla_{\Theta}F_i\big(\nabla_F\mathcal{L}_i - \nabla_F\mathcal{L}(F_i,Y_j)\big)||, \tag{19}$$

$$\Delta_B = ||\nabla_{\Theta}F_i\big(\nabla_F\mathcal{L}(F_i,Y_j) - \nabla_F\mathcal{L}_j\big)||, \tag{20}$$

$$\Delta_C = ||(\nabla_{\Theta}F_i - \nabla_{\Theta}F_j)\nabla_F\mathcal{L}_j||. \tag{21}$$

For the term Δ_A in Eq. (19), since $\nabla_{\Theta}F_i$ is bounded by B_F and $\nabla_F\mathcal{L}$ is Lipschitz *w.r.t.* Y, we have

$$\Delta_A \le ||\nabla_{\Theta}F_i||\cdot||\nabla_F\mathcal{L}_i - \nabla_F\mathcal{L}(F_i,Y_j)|| \le B_F M_Y||\boldsymbol{Y_i - Y_j}||, \tag{22}$$

where $||\boldsymbol{Y_i - Y_j}|| \le 2(||\boldsymbol{p}_i||^2 + ||\boldsymbol{p}_j||^2)$. Note that $\boldsymbol{p}_i$ and $\boldsymbol{p}_j$ are in a probability simplex, which means that the maximum for $||\boldsymbol{p}_i||^2$ and $||\boldsymbol{p}_j||^2$ is 1. Thus, we have $||\boldsymbol{Y_1 - Y_2}|| \le 2$ and $\Delta_A \le 2B_F M_Y$. Meanwhile, since $\mathcal{L}$ and F are Lipschitz, by using the triangle inequality, we get

$$\begin{aligned}|\mathcal{L}_i - \mathcal{L}_j| &\le |\mathcal{L}_i - \mathcal{L}(F_i,Y_j)| + |\mathcal{L}(F_i,Y_j) - \mathcal{L}_j| \\ &\le L_Y||Y_i - Y_j|| + L_F C_F||\boldsymbol{X}_i - \boldsymbol{X}_j|| \\ &< L_Y||Y_i - Y_j|| + L_F C_F\delta_X \\ &= \Delta_{XY} \le 2L_Y + L_F C_F\delta_X.\end{aligned} \tag{23}$$

Recall that $|\mathcal{L}_i - \mathcal{L}_j| < \delta_L$. Thus, when $\Delta_{XY} \le \delta_L$, we may get a tighter upper bound for $||Y_i - Y_j||$ (i.e., $||Y_i - Y_j|| \le \frac{\delta_L - L_F C_F\delta_X}{L_Y}$) and also Δ_A formulated in Eq. (22). To sum up, we get

$$\Delta_A \le B_F M_Y\Big(\mathbb{I}_{\Delta_{XY}>\delta_L}\cdot 2 + \mathbb{I}_{\Delta_{XY}\le\delta_L}\cdot\min(2, \frac{\delta_L - L_F C_F\delta_X}{L_Y})\Big). \tag{24}$$

Similarly, for the terms Δ_B and Δ_C in Eqs. (20)-(21), we have

$$\begin{aligned}\Delta_B &\le ||\nabla_{\Theta}F_i||\cdot||\nabla_F\mathcal{L}(F_i,Y_j) - \nabla_F\mathcal{L}_j|| \\ &\le B_F M_F||F_i - F_j|| \le B_F M_F C_F||\boldsymbol{X}_i - \boldsymbol{X}_j|| < B_F M_F C_F\delta_X,\end{aligned} \tag{25}$$

and

$$\Delta_C \le ||\nabla_F\mathcal{L}_j||\cdot||\nabla_{\Theta}F_i - \nabla_{\Theta}F_j|| \le B_L M_X||\boldsymbol{X}_i - \boldsymbol{X}_j|| < B_L M_X\delta_X. \tag{26}$$

Substituting Eqs. (24)-(26) back to Eq. (18), we get the inequality in Eq. (11). □

A.2 Proof of Theorem 4.2

Proof of Theorem 4.2. By substituting Eq. (16) to the left-hand side of Eq. (17), the objective of using SDGP becomes

$$\begin{aligned}O_{SDGP} &\triangleq \arg\min_{\Theta}\mathop{\mathbb{E}}_{\boldsymbol{X}\in\mathcal{S}}[\,\mathcal{L}'(\boldsymbol{X};\Theta)\,] \\ &= \int_{\boldsymbol{X}\in\mathcal{S}_1}\frac{1}{1-r}\mathcal{L}(\boldsymbol{X};\Theta)\rho(\boldsymbol{X})d\boldsymbol{X} + \int_{\boldsymbol{X}\in\mathcal{X}_2}\mathcal{L}(\boldsymbol{X};\Theta)\rho(\boldsymbol{X})d\boldsymbol{X}.\end{aligned} \tag{27}$$

By substituting Eq. (13) to the first term on the right-hand side of Eq. (27), we get

$$\begin{aligned}\int_{\boldsymbol{X}\in\mathcal{S}_1}\frac{1}{1-r}\mathcal{L}(\boldsymbol{X};\Theta)\rho(\boldsymbol{X})d\boldsymbol{X} &= \int_{\boldsymbol{X}\in\mathcal{X}_1}\frac{1-Pr(\boldsymbol{X})}{1-r}\mathcal{L}(\boldsymbol{X};\Theta)\rho(\boldsymbol{X})d\boldsymbol{X} \\ &= \int_{\boldsymbol{X}\in\mathcal{X}_1}\mathcal{L}(\boldsymbol{X};\Theta)\rho(\boldsymbol{X})d\boldsymbol{X}.\end{aligned} \tag{28}$$

Finally, by substituting Eqs. (28) and (14) to Eq. (27), we have

$$\begin{aligned}O_{SDGP} &= \int_{\boldsymbol{X}\in\mathcal{X}_1}\mathcal{L}(\boldsymbol{X};\Theta)\rho(\boldsymbol{X})d\boldsymbol{X} + \int_{\boldsymbol{X}\in\mathcal{X}_2}\mathcal{L}(\boldsymbol{X};\Theta)\rho(\boldsymbol{X})d\boldsymbol{X} \\ &= \int_{\boldsymbol{X}\in\mathcal{X}}\mathcal{L}(\boldsymbol{X};\Theta)\rho(\boldsymbol{X})d\boldsymbol{X} = \arg\min_{\Theta}\mathop{\mathbb{E}}_{\boldsymbol{X}\in\mathcal{X}}[\,\mathcal{L}(\boldsymbol{X};\Theta)\,].\end{aligned} \tag{29}$$

□

B DATASET DESCRIPTION

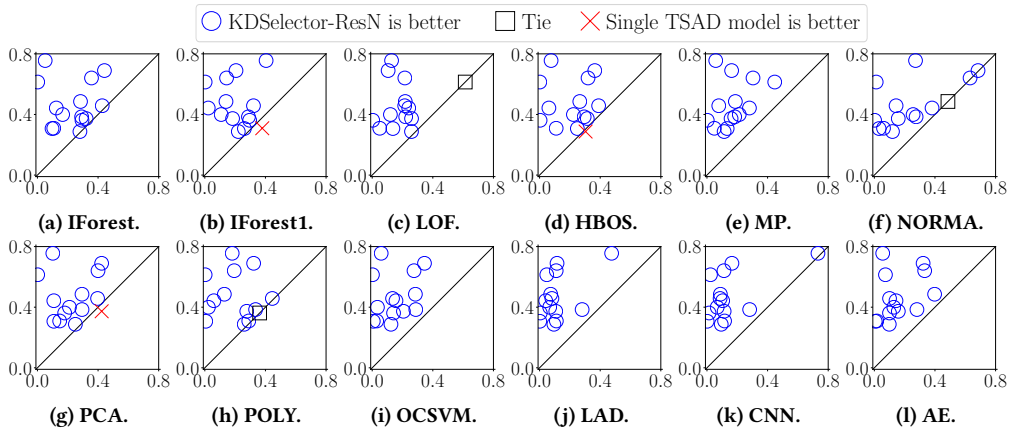
We use the dataset description provided by the TSB-UAD benchmark [53] in the text K_i, which is shown in Table 5.

C COMPARISON WITH SINGLE TSAD MODELS.

Fig. 15 shows the one-versus-one AUC-PR comparison between our KDSelector-ResN with each single TSAD model that is run for all test series. The results further verify the effectiveness of our model selection in improving the detection performance by adaptively selecting suitable TSAD models for different time series.

Table 5: Dataset name, statistics, and description.

Name	# of series	Avg. length	Avg. # of anomalies	Dataset description
Daphnet	45	21760	7.6	contains the annotated readings of 3 acceleration sensors at the hip and leg of Parkinson's disease patients that experience freezing of gait (FoG) during walking tasks.
Dodgers	1	50400	133.0	is a loop sensor data for the Glendale on-ramp for the 101 North freeway in Los Angeles and the anomalies represent unusual traffic after a Dodgers game.
ECG	53	230352	195.6	is a standard electrocardiogram dataset and the anomalies represent ventricular premature contractions. We split one long series (MBA_ECG14046) with length about 1e7 to 47 series by first identifying the periodicity of the signal.
Genesis	6	16220	3.0	is a portable pick-and-place demonstrator which uses an air tank to supply all the gripping and storage units.
GHL	126	200001	1.2	is a Gasoil Heating Loop Dataset and contains the status of 3 reservoirs such as the temperature and level. Anomalies indicate changes in max temperature or pump frequency.
IOPS	58	102119	46.5	is a dataset with performance indicators that reflect the scale, quality of web services, and health status of a machine.
KDD21	250	77415	1.0	is a composite dataset released in a recent SIGKDD 2021 competition with 250 time series.
MGAB	10	100000	10.0	is composed of Mackey-Glass time series with non-trivial anomalies. Mackey-Glass time series exhibit chaotic behavior that is difficult for the human eye to distinguish.
MITDB	32	650000	210.1	contains 48 half-hour excerpts of two-channel ambulatory ECG recordings, obtained from 47 subjects studied by the BIH Arrhythmia Laboratory between 1975 and 1979.
NAB	58	6302	2.0	is composed of labeled real-world and artificial time series including AWS server metrics, online advertisement clicking rates, real time traffic data, and a collection of Twitter mentions of large publicly-traded companies.
OPP	465	31617	2.0	is a dataset devised to benchmark human activity recognition algorithms (e.g., classification, automatic data segmentation, sensor fusion, and feature extraction). The dataset comprises the readings of motion sensors recorded while users executed typical daily activities.
Occupancy	10	5726	18.3	contains experimental data used for binary classification (room occupancy) from temperature, humidity, light, and CO2. Ground-truth occupancy was obtained from time stamped pictures that were taken every minute.
SensorScope (SenSco)	23	27038	11.2	is a collection of environmental data, such as temperature, humidity, and solar radiation, collected from a typical tiered sensor measurement system.
SMD	281	25562	10.4	is a 5-week-long dataset collected from a large Internet company. This dataset contains 3 groups of entities from 28 different machines.
SVDB	115	230400	208.0	includes 78 half-hour ECG recordings chosen to supplement the examples of supraventricular arrhythmias in the MIT-BIH Arrhythmia Database.
YAHOO	367	1561	5.9	is a dataset published by Yahoo labs consisting of real and synthetic time series based on the real production traffic to some of the Yahoo production systems.



(a) IForest. **(b) IForest1.** **(c) LOF.** **(d) HBOS.** **(e) MP.** **(f) NORMA.**

(g) PCA. **(h) POLY.** **(i) OCSVM.** **(j) LAD.** **(k) CNN.** **(l) AE.**

Figure 15: AUC-PR comparison with each single TSAD model.