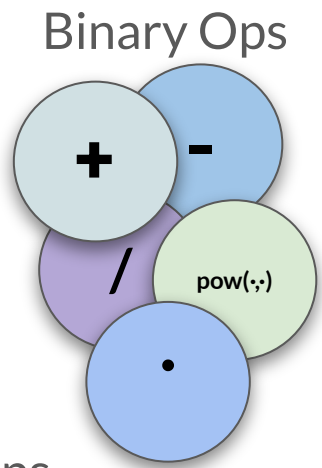
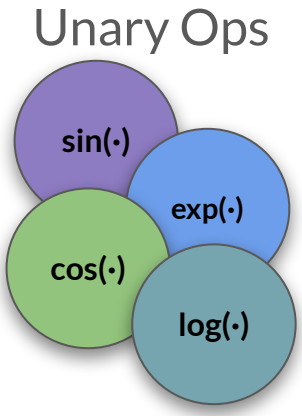
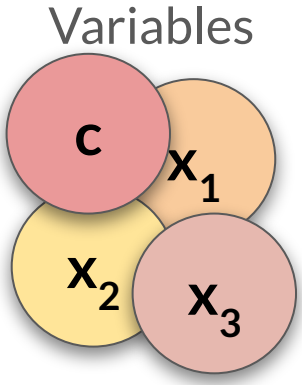
Complexity Aware Deep Symbolic Regression

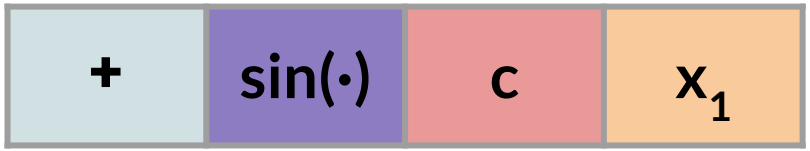
Zachary Bastiani, Robert M. Kirby, Jacob Hochhalter, Shandian Zhe
University of Utah

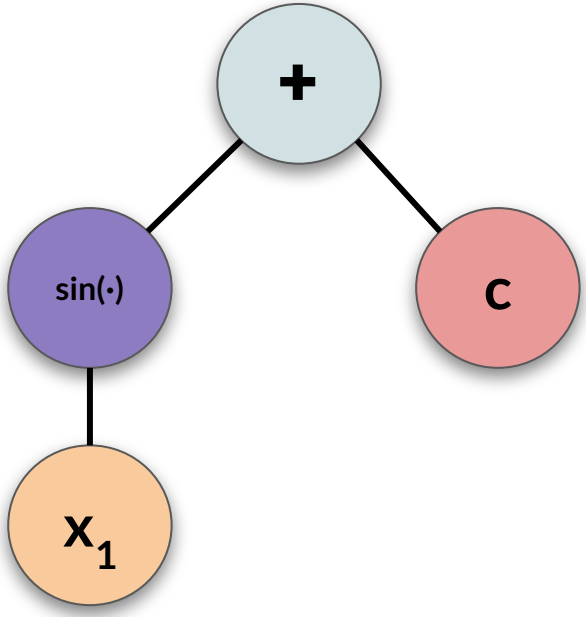
Symbolic Regression: Problem Statement







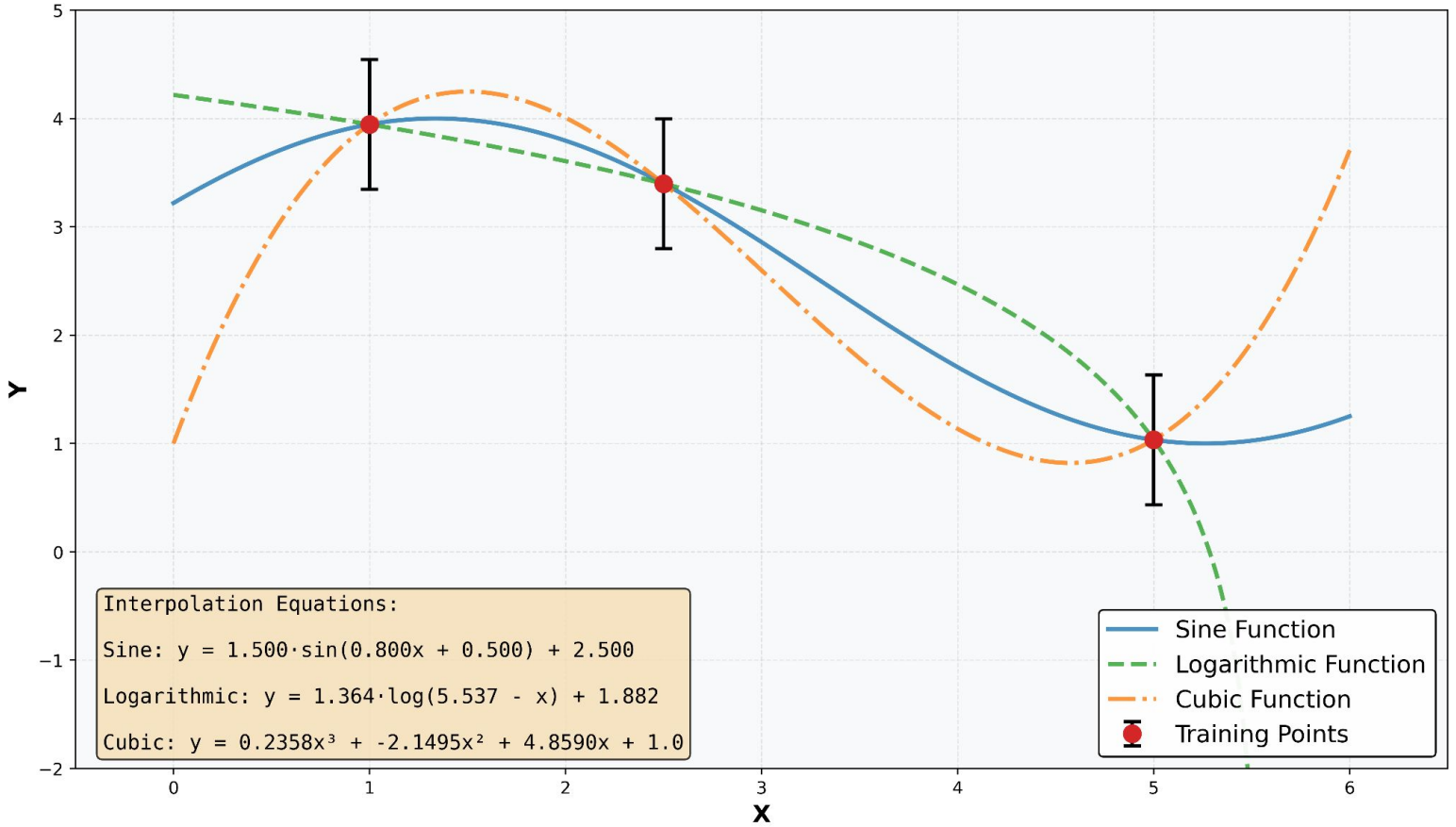




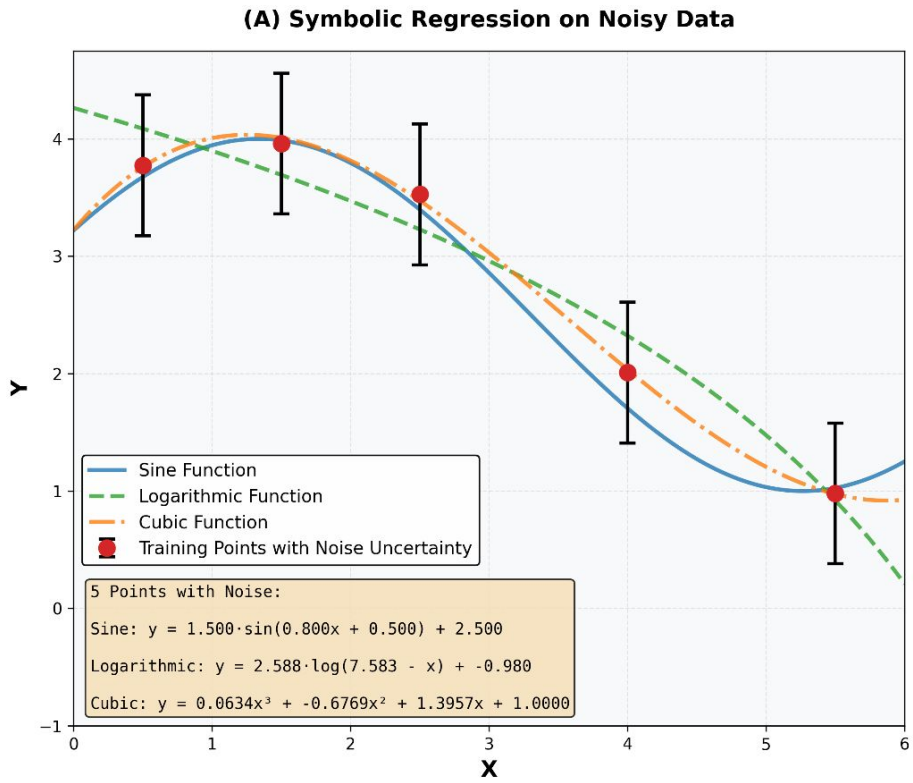
Deep Symbolic Regression: Reinforcement Learning

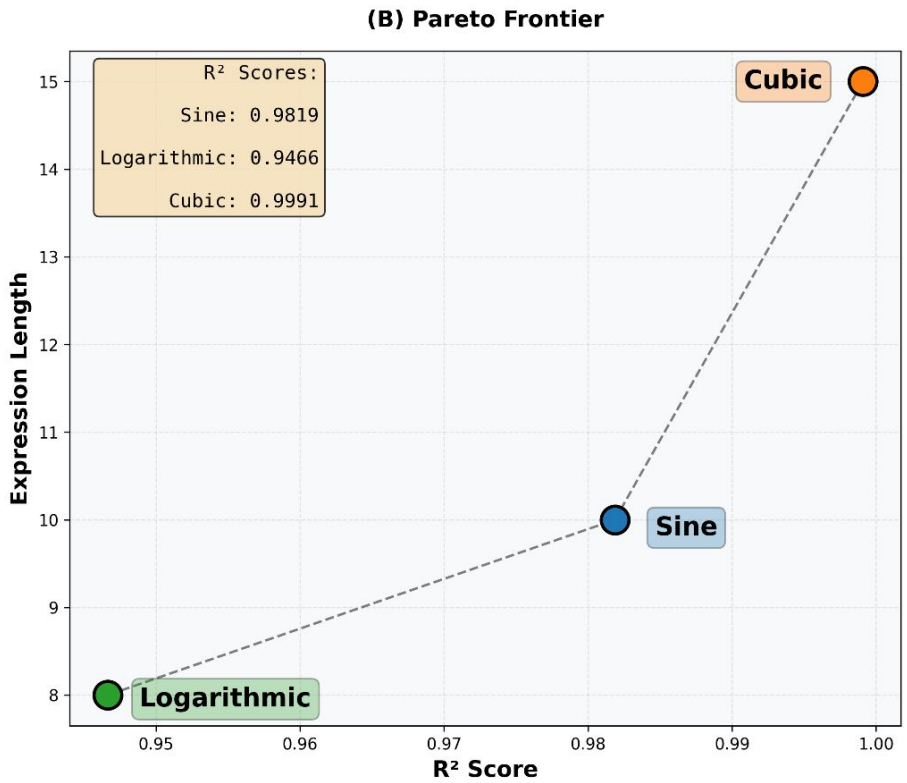
Petersen, B. K., Landajuela, M., Mundhenk, T. N., Santiago, C. P., Kim, S. K., and Kim, J. T. (2021). Deep symbolic regression: Recovering mathematical expressions from data via risk-seeking policy gradients, ICLR

The Interpolation Problem



Pareto Frontier





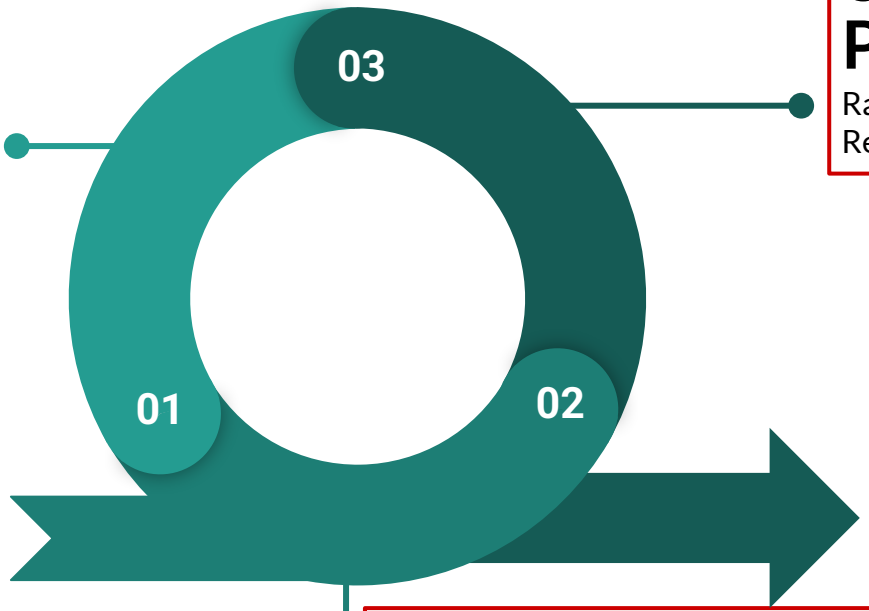
Complexity Aware Deep Symbolic Regression (CADSR)

Contribution 3: Actor

Frequency-based Transformer with Structure-Aware PE

Contribution 2: Policy

Ranked Risk-Seeking Group Relative Policy Optimization



Dataset

Observed data sampled from the underlying expression

Discovered Expression

The “best” expression discovered

Contribution 1: Reward Signal

Bayesian Information Criterion (BIC)

Contribution 1: Bayesian Information Criterion (BIC)

$$BIC = k\log(S) + 2\log(\prod_{i=1}^{N} p((y)_i|\tau,(\boldsymbol{x})_i))$$

Expression Length

Dataset Size

Expression Likelihood

Expression	BIC Reward	Rank
Cubic	-9.3994	1
Sine	-2.4555	2
Logarithmic	-0.2785	3

Contribution 2: Group Relative Policy Optimization

Trust Region

Regularization

$$\nabla \hat{J}(\theta, \alpha) = \frac{1}{\alpha B} \nabla_\theta \sum_{i=1}^{B} \sum_{j=1}^{|\tau^{(i)}|} \boxed{\mathrm{TR}[A_{ij} \frac{p_\theta(\tau^{(i)})}{p_{\theta_{\text{old}}}(\tau^{(i)})}, \epsilon]} - \boxed{\beta D_{KL}[p_\theta(\tau^{(i)}) || p_{\theta_{\text{ref}}}(\tau^{(i)})]}$$

Ranked Reward:

$$A_{ij} = \lambda \cdot \mathrm{ReLU}\left(1 - \frac{|\{\tau^{(k)} : R(\tau^{(k)}) > R(\tau^{(i)})\}|}{\alpha B / 100}\right)$$

Contribution 2: Group Relative Policy Optimization

Trust Region

Regularization

$$\nabla \hat{J}(\theta, \alpha) = \frac{1}{\alpha B} \nabla_\theta \sum_{i=1}^{B} \sum_{j=1}^{|\tau^{(i)}|} \mathrm{TR}[A_{ij} \frac{p_\theta(\tau^{(i)})}{p_{\theta_{\text{old}}}(\tau^{(i)})}, \epsilon] - \beta D_{KL}[p_\theta(\tau^{(i)}) || p_{\theta_{\text{ref}}}(\tau^{(i)})]$$

Ranked Reward:

$$A_{ij} = \lambda \cdot \mathrm{ReLU}\left(1 - \frac{|\{\tau^{(k)} : R(\tau^{(k)}) > R(\tau^{(i)})\}|}{\alpha B / 100}\right)$$

Contribution 2: Group Relative Policy Optimization

Trust Region

Regularization

$$\nabla \hat{J}(\theta, \alpha) = \frac{1}{\alpha B} \nabla_\theta \sum_{i=1}^{B} \sum_{j=1}^{|\tau^{(i)}|} \left[\mathrm{TR}[A_{ij} \frac{p_\theta(\tau^{(i)})}{p_{\theta_{\text{old}}}(\tau^{(i)})}, \epsilon] - \beta D_{KL}[p_\theta(\tau^{(i)}) || p_{\theta_{\text{ref}}}(\tau^{(i)})] \right]$$

Ranked Reward:

$$A_{ij} = \lambda \cdot \mathrm{ReLU}\left(1 - \frac{|\{\tau^{(k)} : R(\tau^{(k)}) > R(\tau^{(i)})\}|}{\alpha B / 100}\right)$$

Contribution 2: Group Relative Policy Optimization

Trust Region

Regularization

$$\nabla \hat{J}(\theta, \alpha) = \frac{1}{\alpha B} \nabla_\theta \sum_{i=1}^{B} \sum_{j=1}^{|\tau^{(i)}|} \left[\mathrm{TR}\left[A_{ij} \frac{p_\theta(\tau^{(i)})}{p_{\theta_{\text{old}}}(\tau^{(i)})}, \epsilon \right] - \beta D_{KL}\left[p_\theta(\tau^{(i)}) || p_{\theta_{\text{ref}}}(\tau^{(i)}) \right] \right]$$

Ranked Reward:

$$A_{ij} = \lambda \cdot \mathrm{ReLU}\left(1 - \frac{|\{\tau^{(k)} : R(\tau^{(k)}) > R(\tau^{(i)})\}|}{\alpha B / 100} \right)$$

Contribution 2: Ranked Risk-Seeking Advantage Matrix

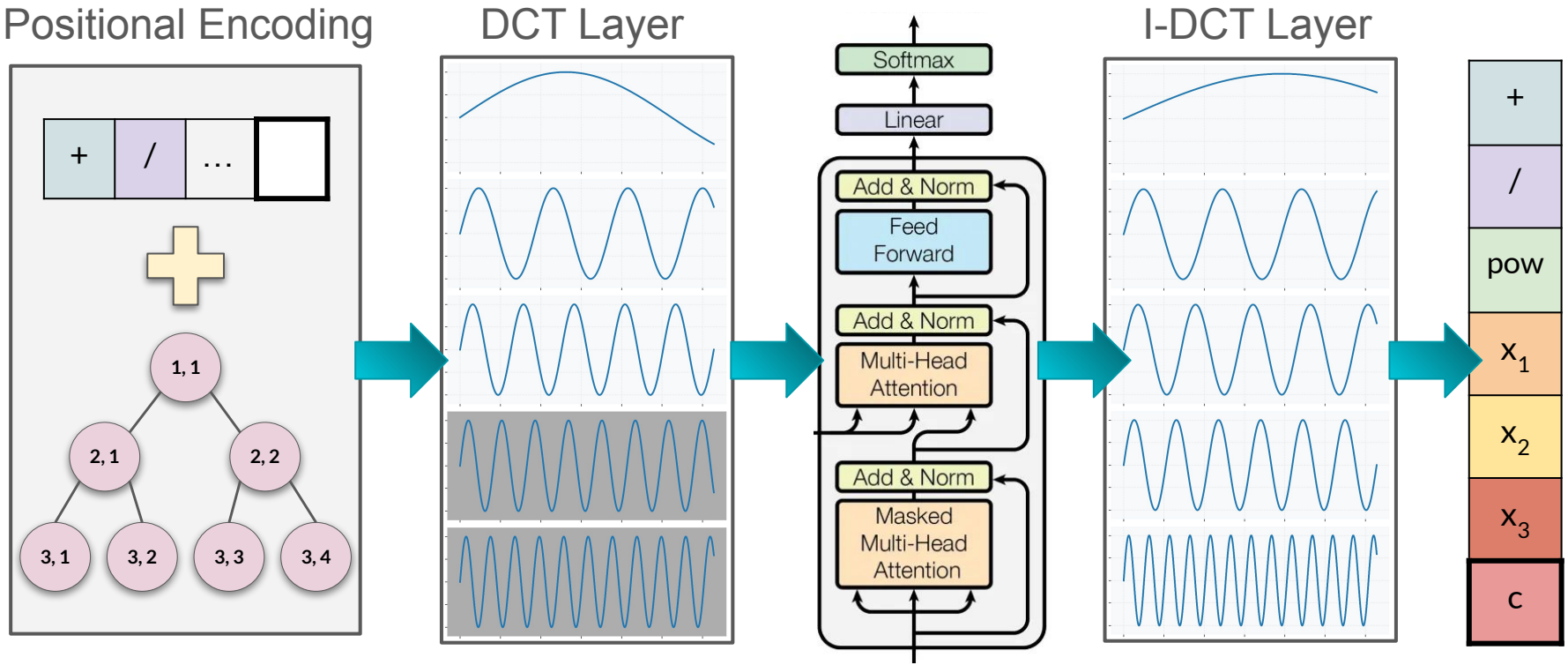
Lemma 1: Given any continuous mapping, *f*, that can map unbounded reward function values to a bounded domain (e.g., (-∞, ∞) -> [0,1]), suppose the reward function is continuous, there always exists a set of distinct rewards values that numerically create a tail barrier in the risk-seeking policy.

Lemma 2: By using the step function for reward mapping, the policy gradients with our BIC reward is unbiased and will not encounter any tail barrier.

Ranked Reward:

$$A_{ij} = \lambda \cdot \mathrm{ReLU}\left(1 - \frac{|\{\tau^{(k)}: R(\tau^{(k)}) > R(\tau^{(i)})\}|}{\alpha B/100}\right)$$

Contribution 3: DCT Layer and Structure Aware PE

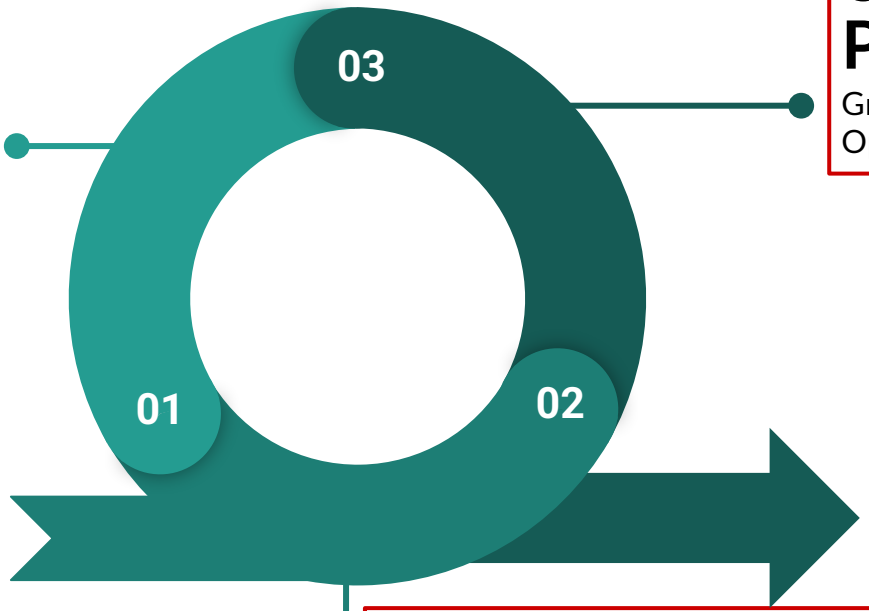


Scribano, C., Franchini, G., Prato, M., and Bertogna, M. (2023). Dct-former: Efficient self-attention with discrete cosine transform. Journal of Scientific Computing, 94(3).

Complexity Aware Deep Symbolic Regression (CADSR)

Contribution 3:
Actor
Frequency-based Transformer with Structure-Aware PE

Contribution 2:
Policy
Group Relative Policy Optimization

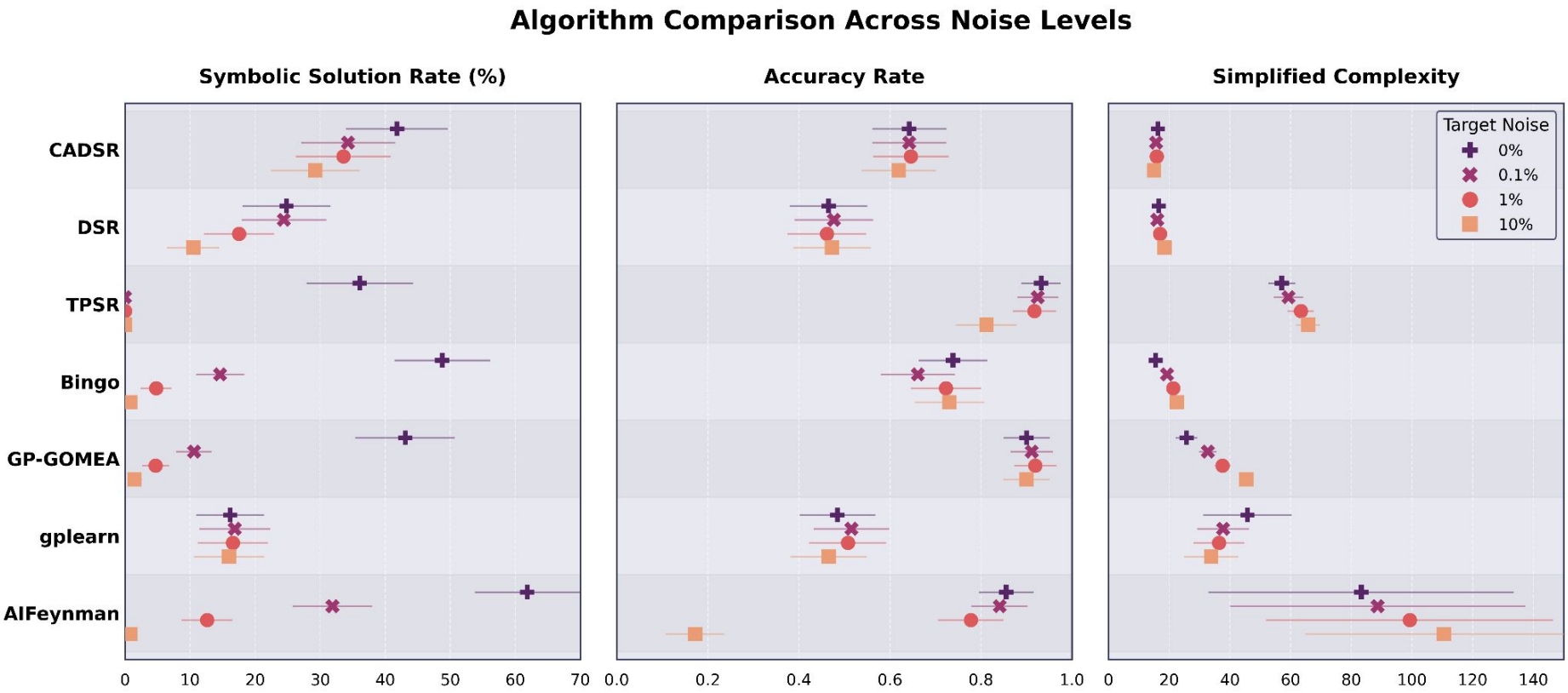


Dataset
Observed data sampled from the underlying expression

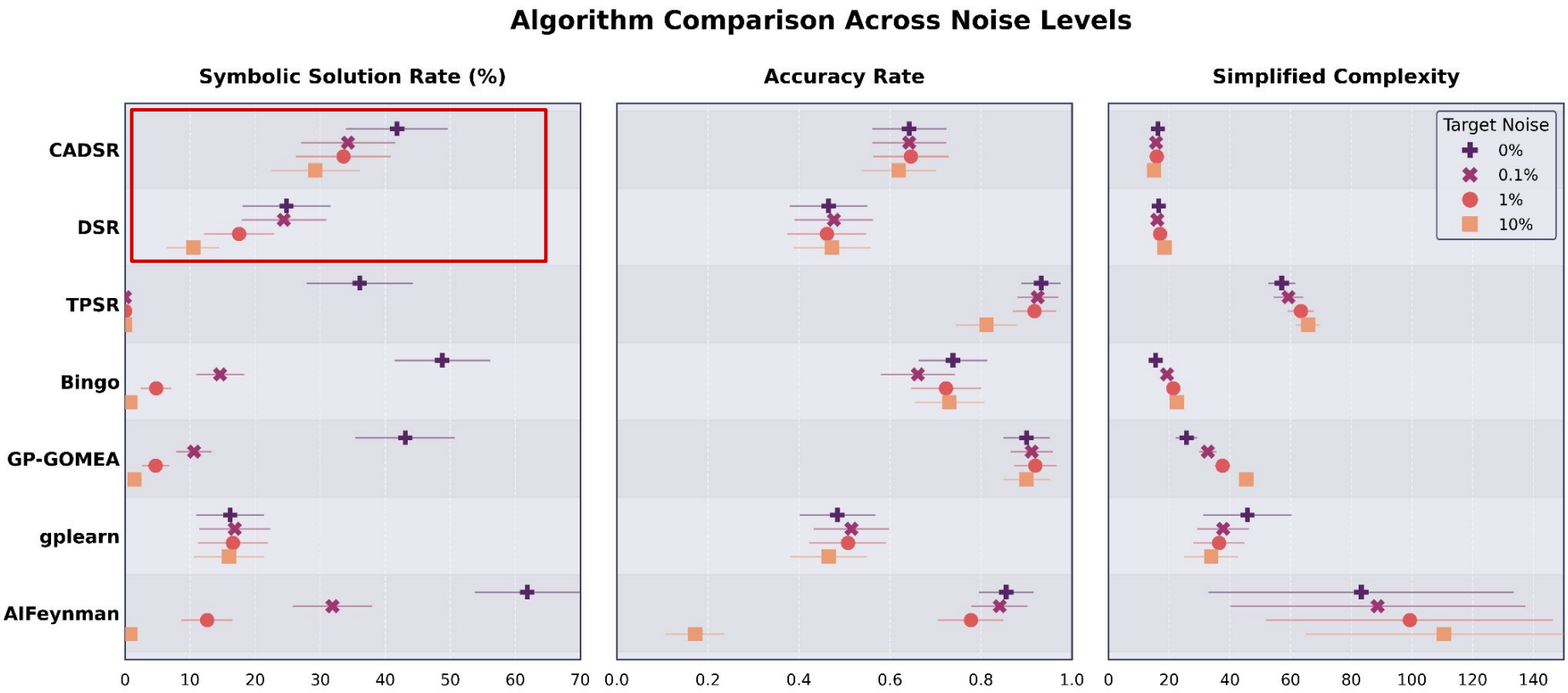
Discovered Expression
The “best” expression discovered

Contribution 1:
Reward Signal
Bayesian Information Criterion (BIC)

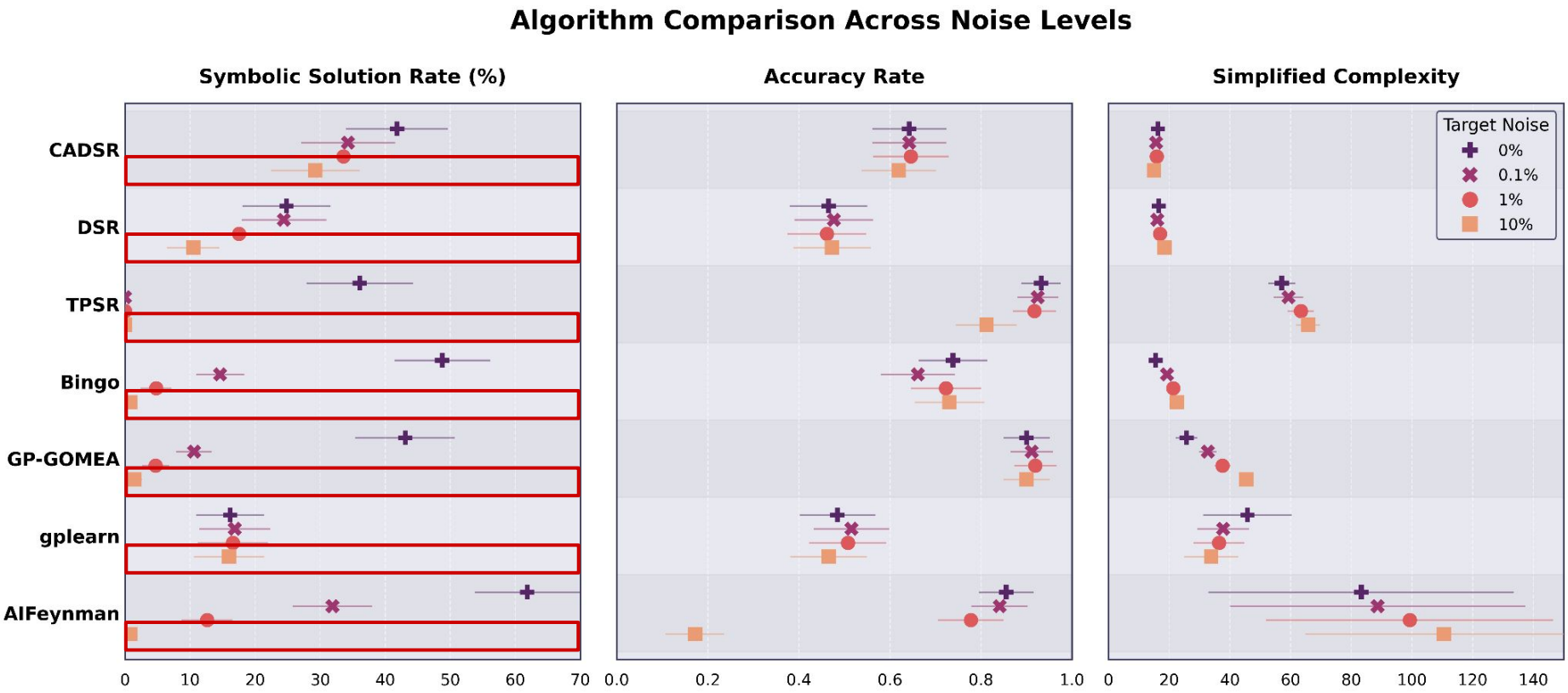
Results


Algorithm Comparison Across Noise Levels
Symbolic Solution Rate (%)
Accuracy Rate
Simplified Complexity
CADSR
DSR
TPSR
Bingo
GP-GOMEA
gplearn
AIFeynman
Target Noise
0%
0.1%
1%
10%
0
10
20
30
40
50
60
70
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
120
140

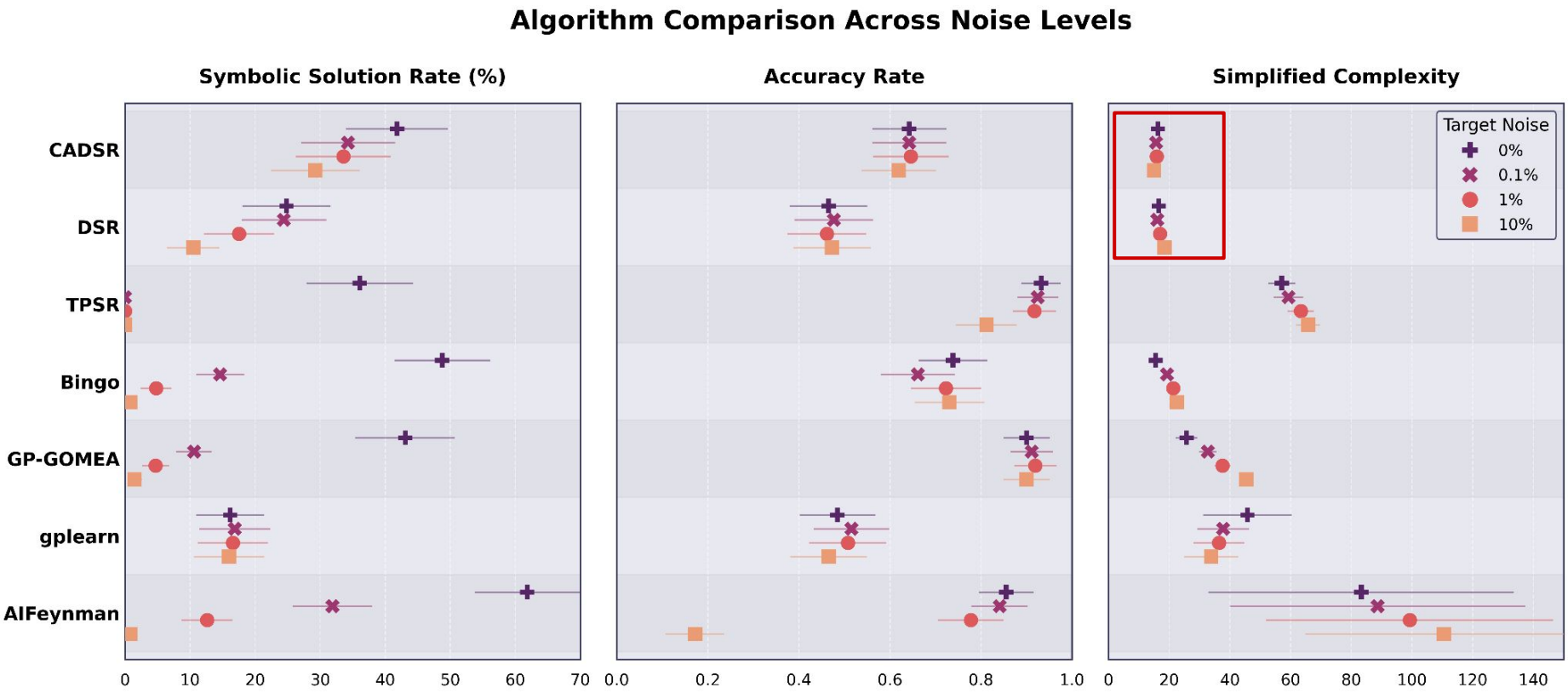
Results


Algorithm Comparison Across Noise Levels
Symbolic Solution Rate (%)
Accuracy Rate
Simplified Complexity
CADSR
DSR
TPSR
Bingo
GP-GOMEA
gplearn
AIFeynman
Target Noise
0%
0.1%
1%
10%
0
10
20
30
40
50
60
70
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
120
140

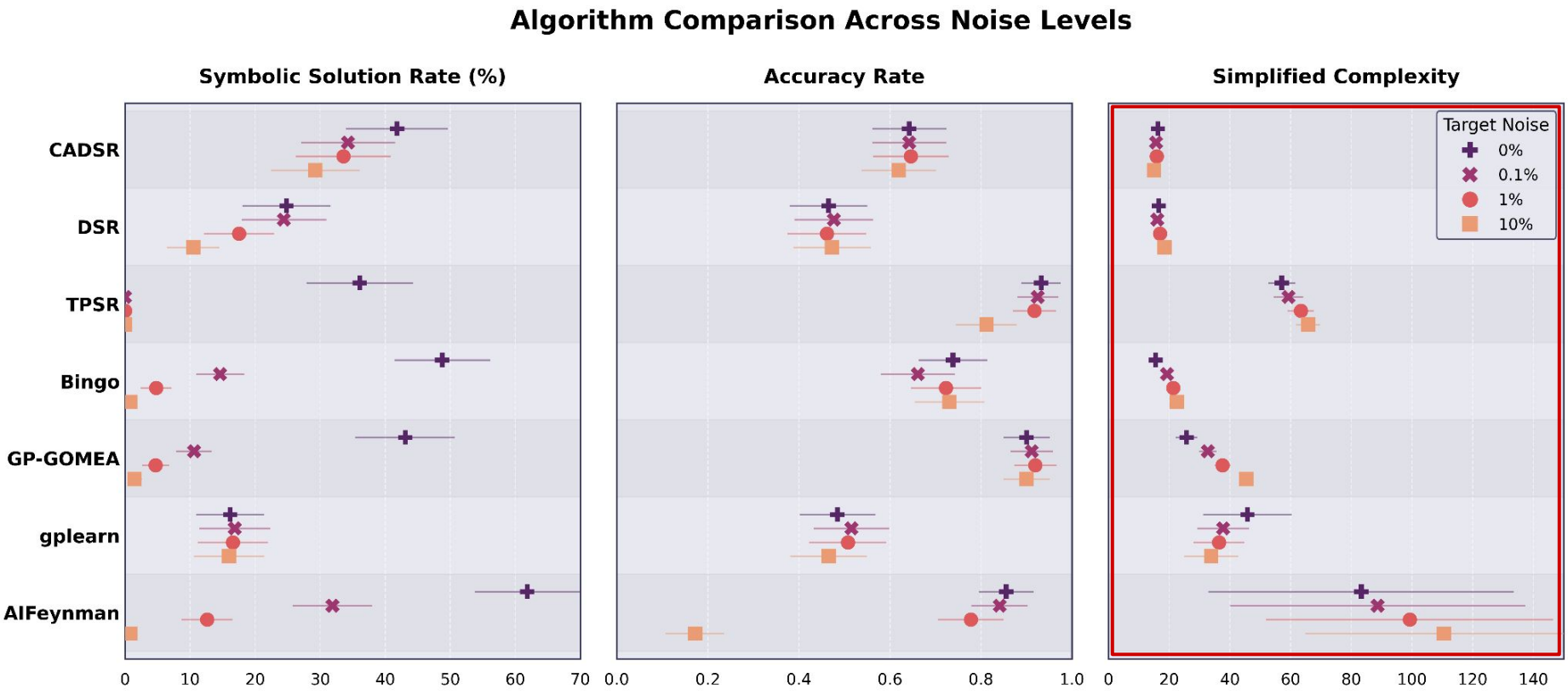
Results


Algorithm Comparison Across Noise Levels
Symbolic Solution Rate (%)
Accuracy Rate
Simplified Complexity
CADSR
DSR
TPSR
Bingo
GP-GOMEA
gplearn
AIFeynman
Target Noise
0%
0.1%
1%
10%
0
10
20
30
40
50
60
70
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
120
140

Results


Algorithm Comparison Across Noise Levels
Symbolic Solution Rate (%)
Accuracy Rate
Simplified Complexity
CADSR
DSR
TPSR
Bingo
GP-GOMEA
gplearn
AIFeynman
Target Noise
0%
0.1%
1%
10%
0
10
20
30
40
50
60
70
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
120
140

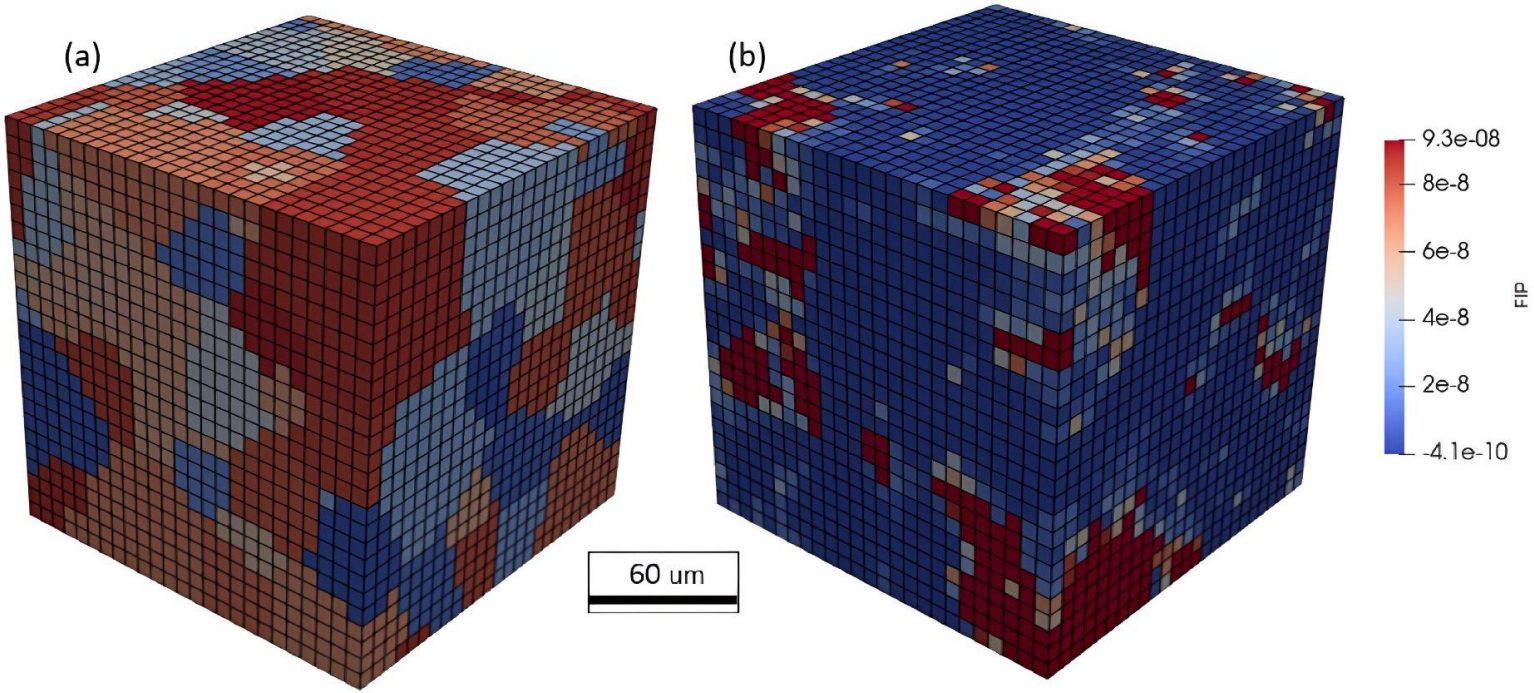
Results


Algorithm Comparison Across Noise Levels
Symbolic Solution Rate (%)
Accuracy Rate
Simplified Complexity
CADSR
DSR
TPSR
Bingo
GP-GOMEA
gplearn
AIFeynman
Target Noise
0%
0.1%
1%
10%
0
10
20
30
40
50
60
70
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
120
140

Fracture Mechanics

Microstructure

FIP Values



Cooper K. Hansen, Gary F. Whelan, Jacob D. Hochhalter, Interpretable machine learning for microstructure-dependent models of fatigue indicator parameters, International Journal of Fatigue, 2024, 108019.

Fracture Mechanics

Method	Train R^2	Test R^2	Runtime (min)	Complexity
CADSR	0.636	0.626	193.06	22.250
DSR	0.622	0.616	124.65	24.625
TPSR	0.043	0.046	11.12	40.750
GP-GOMEA	0.645	0.636	124.66	73.000
DySymNet	0.584	0.456	396.11	1696.875

CADSR Expression:

$$\mathrm{FIP} = -44.7 + 66.1 \cdot S_{\max} - \frac{0.0442}{\mathrm{PCG}} - \frac{36.4 \cdot S_{\mathrm{var}} - 4.58}{S_{\mathrm{avg}}}$$

Fracture Mechanics

Method	Train R^2	Test R^2	Runtime (min)	Complexity
CADSR	0.636	0.626	193.06	22.250
DSR	0.622	0.616	124.65	24.625
TPSR	0.043	0.046	11.12	40.750
GP-GOMEA	0.645	0.636	124.66	73.000
DySymNet	0.584	0.456	396.11	1696.875

CADSR Expression:

$$\mathrm{FIP} = -44.7 + 66.1 \cdot S_{\max} - \frac{0.0442}{\mathrm{PCG}} - \frac{36.4 \cdot S_{\mathrm{var}} - 4.58}{S_{\mathrm{avg}}}$$

Fracture Mechanics

Method	Train R^2	Test R^2	Runtime (min)	Complexity
CADSR	0.636	0.626	193.06	22.250
DSR	0.622	0.616	124.65	24.625
TPSR	0.043	0.046	11.12	40.750
GP-GOMEA	0.645	0.636	124.66	73.000
DySymNet	0.584	0.456	396.11	1696.875

CADSR Expression:

$$\mathrm{FIP} = -44.7 + 66.1 \cdot S_{\max} - \frac{0.0442}{\mathrm{PCG}} - \frac{36.4 \cdot S_{\mathrm{var}} - 4.58}{S_{\mathrm{avg}}}$$

Fracture Mechanics

Method	Train R^2	Test R^2	Runtime (min)	Complexity
CADSR	0.636	0.626	193.06	22.250
DSR	0.622	0.616	124.65	24.625
TPSR	0.043	0.046	11.12	40.750
GP-GOMEA	0.645	0.636	124.66	73.000
DySymNet	0.584	0.456	396.11	1696.875

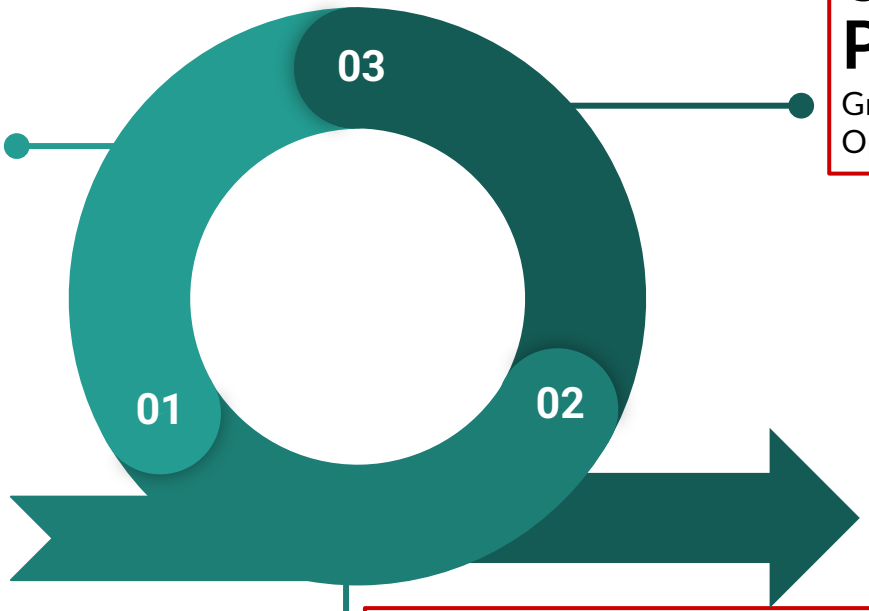
CADSR Expression:

$$\mathrm{FIP} = -44.7 + 66.1 \cdot S_{\max} - \frac{0.0442}{\mathrm{PCG}} - \frac{36.4 \cdot S_{\mathrm{var}} - 4.58}{S_{\mathrm{avg}}}$$

Complexity Aware Deep Symbolic Regression (CADSR)

Contribution 3:
Actor
Frequency-based Transformer with Structure-Aware PE

Contribution 2:
Policy
Group Relative Policy Optimization



Dataset
Observed data sampled from the underlying expression

Discovered Expression
The “best” expression discovered

Contribution 1:
Reward Signal
Bayesian Information Criterion (BIC)

Questions

Code, Results, and Dataset
are available at
https://github.com/ZakBastiani/CADSR