DiTTo: Decoupled Intrinsic Topology Tree Optimization for Neuro-Symbolic Discovery from Non-Stationary Signals

Dongtai Li, LDongtai@hotmail.com

***Abstract*—Symbolic regression (SR) struggles with non-stationary time-series signals. Conventional methods rely on the absolute time coordinate t, making them vulnerable to stochastic phase jitter and distribution shifts. This triggers expression bloat and loss of interpretability—a fundamental flaw we formalize as *Temporal Shift Vulnerability*, which strictly manifests as severe manifold distortion under temporal misalignments.**

To circumvent this, we propose DiTTo (Decoupled intrinsic Topology Tree Optimization), an end-to-end neuro-symbolic framework. DiTTo maps phase-shifted transients into a locally invariant topological manifold $\tau \in [0, 1)$, rigorously compressing search complexity from exponential $\mathcal{O}(\exp(N))$ to polynomial $\mathcal{O}(\text{poly}(N))$. A lightweight Adaptive Neural Router then extracts morphological descriptors to generate probabilistic operator priors without manual annotation. These priors guide discrete search backends—instantiated via Monte Carlo Tree Search (MCTS) in this work—to efficiently discover physically plausible expressions. Crucially, the proposed topological representation is completely solver-independent, serving as a universal plug-and-play frontend capable of boosting any downstream symbolic regressor operating on non-stationary signals.

Evaluated across 190 datasets from four heterogeneous domains, DiTTo matches the reconstruction fidelity of the leading baseline ($p = 0.983$) while requiring only 88% of its formula complexity. Furthermore, in the zero-shot ECG domain, the router autonomously discovers the Gaussian basis function without prior exposure. Stress tests confirm its significant robustness against phase jitter and background noise.

***Index Terms*—Symbolic regression, non-stationary time series, neuro-symbolic AI, topological manifold, neural routing.**

I. Introduction

A. The Bottleneck of Interpretable Representation Learning

In the rapid evolution of Cyber-Physical Systems (CPS) and the Industrial Internet of Things (IoT), the accurate modeling of non-stationary transient dynamics has emerged as a fundamental challenge for safety-critical operations. Current sequence modeling paradigms are trapped in a fundamental trade-off between "high-fidelity black-boxes" and "highly interpretable white-boxes". While deep neural networks (e.g., Neural ODEs and deep recurrent architectures) excel at fitting highly complex non-stationary waveforms through multi-layer non-linear mappings, the physical opacity of their parameter spaces renders them unpredictably fragile under long-tail distributions, significantly limiting their deployment.

Symbolic Regression (SR) aims to extract closed-form analytical equations directly from observational data without relying on predefined functional priors, inherently possessing cross-condition generalization capabilities. Nevertheless, both conventional Genetic Programming (GP) and its modern deep reinforcement learning variants suffer from severe performance degradation when confronted with complex, non-stationary time-series signals in real-world scenarios. This fundamental bottleneck restricts symbolic machine learning from transitioning from ideal mathematical curve-fitting to ubiquitous pattern recognition applications.

B. Temporal Shift Vulnerability in Euclidean Space

A profound analysis of the failure modes of SR on non-stationary dynamical systems reveals an intrinsic structural flaw shared by nearly all existing architectures: they rely heavily on the absolute Euclidean time coordinate t. We formalize this phenomenon as *Temporal Shift Vulnerability*.

When searching for analytical equations within the absolute time domain t, microscopic misalignments and stochastic phase jitter force SR models to theoretically compensate for "phase smearing". Lacking shift-invariance, symbolic engines attempt to globally align these distorted dynamical patterns by introducing high-order polynomial nestings or complex trigonometric combinations. This unguided optimization fundamentally re-triggers expression bloat, yielding verbose formulas devoid of physical semantics.

C. The DiTTo Paradigm: Neuro-Symbolic Discovery via Adaptive Routing

To overcome the Temporal Shift Vulnerability and bridge the gap between neural perception and symbolic reasoning, we propose DiTTo (Decoupled intrinsic Topology Tree Optimization), an end-to-end neuro-symbolic discovery framework. The core philosophy of DiTTo lies in abandoning direct optimization within the drift-prone Euclidean space t. Instead, it leverages an unsupervised manifold projection operator to non-linearly map non-stationary observations into an intrinsic topological manifold space $\tau \in [0, 1)$, which is locally aligned and immune to temporal shifts.

Furthermore, DiTTo introduces a lightweight Adaptive Neural Router that autonomously perceives physical modalities from the morphological features of the τ-space and outputs a probabilistic prior over syntactic operators, thereby achieving end-to-end scientific discovery without manual domain-specific heuristic rules. The computational pipeline integrates five core stages:

- **Stage 1: Feature Decoupling:** Adaptively strips stochastic phase flows and high-frequency carriers to extract smooth, intrinsic energy envelope features.
- **Stage 2: Topology Projection:** Maps phase-shifted transient peaks to a locally invariant manifold τ-space, achieving intrinsic feature alignment.
- **Stage 3: Adaptive Neural Operator Routing:** A neural router extracts a 21-dimensional morphological descriptor and outputs a Softmax probability distribution over physical operators.
- **Stage 4: PUCT-Guided Symbolic Tree Search:** Integrates the router's probability distribution into the selection phase of Monte Carlo Tree Search (MCTS).
- **Stage 5: Sparse Reconstruction:** Executes the inverse mapping from localized white-box equations to the global time axis, coupled with L-BFGS-B global fine-tuning.

D. Summary of Core Contributions

By tightly coupling neural representation learning with discrete symbolic optimization, the primary contributions of this work are summarized as follows:

- **Solver-Independent Topological Representation:** We propose an unsupervised manifold feature alignment operator that resolves the "phase smearing" trap, rigorously compressing search space complexity from an exponential growth $\mathcal{O}(\exp(N))$ to a localized polynomial complexity $\mathcal{O}(\text{poly}(N))$. This representation serves as a universal frontend capable of boosting established industrial regressors on non-stationary signals.
- **Zero-Shot Neural Operator Router:** We construct an ultra-lightweight MLP classifier that autonomously perceives underlying physical modalities from τ-space features. Remarkably, the router accurately identified the Gaussian basis function with a 0.69 probability on unseen ECG signals, providing empirical validation for automated, cross-domain scientific discovery devoid of manual domain expertise.
- **PUCT-Guided Neuro-Symbolic Architecture:** By integrating the router's operator probability distribution as a prior bias term in the MCTS selection phase, we establish a tightly coupled framework that flows naturally from neural perception of manifold features to guided discrete symbolic search.
- **Large-Scale Benchmarking:** Comprehensive evaluations across 190 datasets spanning four heterogeneous domains demonstrate that DiTTo matches the reconstruction fidelity of the state-of-the-art evolutionary baseline ($p = 0.983$) while requiring only 88% of its formula complexity. Adaptive routing yielded a $+0.013$ R^2 improvement over the unrouted variant in the generalized UCR domain and a 15% reduction in equation length within the mechanical domain.

II. Related Work

A. Symbolic Regression and Representation Learning

Symbolic Regression (SR), as a core technique for automating the discovery of scientific laws, has historically relied on evolutionary computation frameworks based on Genetic Programming (GP) [1]. Recently, the resurgence of deep learning and reinforcement learning has catalyzed a new generation of modern SR systems. Deep Symbolic Optimization (DSO) [2] generates expression syntax trees via Recurrent Neural Networks (RNNs); PySR [3] combines highly optimized distributed regularized evolution with SymPy constant fine-tuning. Concurrently, the foundational challenges of discrete-space topological generation and parameter annealing have been rigorously addressed by the purely discrete Monte Carlo Tree Search (MCTS) framework operating on isomorphic grammar spaces introduced in [4] (the U-ESDT architecture), which successfully eliminates gradient-based local optima and baseline expression bloat.

With the base search mechanism thus established, the present work deliberately pivots away from basic search dynamics. Instead, we focus on a higher-dimensional scientific problem: representation learning and neural prior induction for non-stationary time series. All the aforementioned systems—whether evolutionary, continuous, or purely discrete—share a fundamental limitation: they execute searches within the topologically uncoupled original observation space (absolute time t or high-dimensional Euclidean space). When processing non-stationary time series with stochastic phase jitter or multi-scale drift, these systems lack morphological calibration. Consequently, they rely on inflated expression complexity to approximate the phase-shifted data, leading to poor generalization.

B. Time-Series Alignment and Manifold Learning

To overcome the challenges of temporal shifts and misalignment in multivariate signals, time-series representation learning and feature alignment have been extensively studied. Dynamic Time Warping (DTW) [5] measures the distance between phase-delayed waveforms by non-linearly stretching the time axis. Manifold learning techniques (e.g., Isomap [6], LLE [7], and UMAP [8]) are widely employed to capture the underlying locally smooth, low-dimensional geometric topologies hidden within high-dimensional observations, effectively recovering the intrinsic degrees of freedom in dynamical systems.

Nevertheless, existing manifold alignment paradigms exhibit a key limitation: **they are almost exclusively relegated to downstream classification or clustering tasks**. Conventional manifold projections neither generate explicit symbolic mappings nor form a closed loop with equation extraction engines. Consequently, manifold dimensionality reduction remains confined to opaque latent spaces, unable to provide auditable analytical forms for white-box pattern discovery. DiTTo bridges this gap by executing grammatical tree searches directly atop the aligned topological manifold.

C. Inductive Biases in Neuro-Symbolic AI

To endow data-driven models with symbolic priors, Neuro-Symbolic AI has emerged as a prominent frontier in machine learning. Physics-Informed Neural Networks (PINNs) [9] explicitly embed ODEs or PDEs as regularization penalties within the loss function, forcing deep networks to adhere to fundamental physical laws, thereby mitigating the interpretability deficit of black-box models.

However, PINN-based approaches rely on a **Soft Constraint Mechanism**: physical equations act merely as loss regularizers, making them susceptible to local minima under complex non-stationary noise and unable to guarantee strict mathematical or syntactic compliance. In contrast, DiTTo introduces a **Hard Syntactic Prior** mechanism that directly prunes the grammatical action space of MCTS via explicit probabilistic modulation (e.g., configuring `ExpDecayNode` for mechanical transients and `GaussianNode` for biomedical sequences). By constraining the mathematical morphology at the syntax generation level, this architecture endows the model with rigorous physical compliance while reducing the search space, yielding advantages in both extraction efficiency and fitting precision over soft-constrained PINNs and traditional SR models.

III. The DiTTo Meta-Framework

A. Mathematical Formulation of Topology-Decoupled Search

Let $\mathcal{X} \subset L^2(\mathbb{R})$ denote the Hilbert space of observable non-stationary time-series signals with localized deterministic transient dynamics corrupted by stochastic phase perturbations. Traditional SR frameworks directly solve the functional optimization problem in the Euclidean temporal domain $\mathbb{R}$, seeking a closed-form mapping $f : t \mapsto x(t)$. Under distribution shifts and phase jitter, however, such a direct mapping suffers from phase smearing, forcing the search engine to introduce high-order polynomial expansions that trigger expression bloat.

To circumvent this fundamental limitation, we propose the Decoupled intrinsic Topology Tree Optimization (DiTTo) framework. We introduce an intrinsic lower-dimensional topological manifold space $\Gamma \subset [-1, 1]$ and an abstract modality feature space $\mathcal{C}$. The end-to-end knowledge discovery paradigm of DiTTo is formally defined as a composition of non-linear structural operators:

$$\begin{aligned} \mathcal{X} &\xrightarrow{\mathcal{D}} \mathcal{C} \xrightarrow{\mathcal{M}} \Gamma \\ &\xrightarrow{\mathcal{T}} \mathcal{F}(\Gamma) \xrightarrow{\mathcal{M}^{-1}} \hat{\mathcal{X}} \end{aligned} \tag{1}$$

where $\mathcal{D}$ is the unsupervised modality decoupling operator, $\mathcal{M}$ denotes the homeomorphic topology alignment projection, $\mathcal{T}$ is the domain-aware MCTS operator over the grammatical space, and $\mathcal{M}^{-1}$ defines the global sparse inverse reconstruction with continuous parameter refinement.

B. Stage 1: "Di" - Unsupervised Feature Decoupling

The initial phase treats the raw observation $x(t) \in \mathcal{X}$ as a modulated multi-scale waveform. The decoupling operator $\mathcal{D} : \mathcal{X} \to \mathcal{C}$ isolates the intrinsic structural dynamics from the oscillatory carrier phase or background variations. For analytical transient waveforms exhibiting high-frequency oscillations (e.g., mechanical vibration and electrical transients), $\mathcal{D}$ leverages the Cauchy principal value of the Hilbert transform [10]:

$$\tilde{x}(t) = \mathcal{H}[x](t) = \frac{1}{\pi} \text{ p.v.} \int_{-\infty}^{\infty} \frac{x(\xi)}{t - \xi} \, d\xi \tag{2}$$

The unique analytic signal $z(t) \in \mathbb{C}$ is formulated as $z(t) = x(t) + j\tilde{x}(t)$, yielding the polar decomposition $z(t) = A(t)\exp(j\phi(t))$ and the decoupled feature tuple $\boldsymbol{c}(t) = \langle A(t), \phi(t) \rangle \in \mathcal{C}$, where the amplitude envelope $A(t) = |z(t)| \in \mathbb{R}^+$ captures the instantaneous energy distribution, and $\phi(t) = \text{Arg}(z(t)) \in (-\pi, \pi]$ is the continuous unrolled phase.

For bio-electrical signals lacking distinct carrier frequencies (e.g., ECG), $\mathcal{D}$ operates as a non-local morphological template averaging operator over localized cardiac cycles anchored by a discrete heartbeat peak set, yielding a smooth profile that prevents structural disintegration during AST generation.

C. Stage 2: "T" - Topology Projection ($\mathcal{M}$ Operator)

To eliminate the spatial smearing induced by stochastic temporal translations, we construct the homeomorphic alignment operator $\mathcal{M} : \mathcal{C} \to \Gamma$. Let $\Omega = \{t_i \in \mathbb{R} \mid \dot{A}(t_i) = 0, \ddot{A}(t_i) < 0, A(t_i) \geq \gamma \cdot \max(A(t))\}$ define the discrete set of significant local transient critical points, where $\gamma \in (0, 1)$ is an empirical threshold. For each isolated domain governed by a critical point $t_i \in \Omega$, we map the localized absolute coordinate $t \in [t_i - \Delta L, t_i + \Delta R]$ to the invariant topological space Γ via the affine homeomorphism:

$$\tau_i = \mathcal{M}_i(t) = \frac{t - t_i}{\Delta T} \tag{3}$$

where ΔT is the scale-invariant temporal normalization factor determined by local band-limit properties. This projection aligns a family of phase-shifted transients into a unified τ-coordinate frame.

Theorem 1: Let $\mathcal{G}$ be a context-free symbolic grammar with an operator set of cardinality K. If an underlying transient profile contains M independent phase-shifted occurrences, searching for an exact closed-form solution directly in the Euclidean space $\mathbb{R}$ requires an abstract syntax tree (AST) depth $D \propto \mathcal{O}(M)$ to satisfy Taylor-like error bounds, resulting in an exponential state-space complexity of $\mathcal{O}(K^M)$. The topological projection $\mathcal{M}$ reduces the required AST depth to an M-independent constant $D_0 \ll M$, strictly bounding the MCTS exploration space to a polynomial complexity $\mathcal{O}(\text{poly}(D_0))$.

Proof 1: In the Euclidean temporal domain, let the multi-transient target be formulated as $x(t) = \sum_{i=1}^{M} \psi(t - t_i)$, where ψ is an unknown base transient function and t_i represents stochastic phase shift variables. To represent $x(t)$ globally without coordinate alignment, an SR model must approximate each shifted entity individually. Assuming an analytical polynomial basis, the N-th order Taylor expansion of each component around a global origin t_0 is:

$$\psi(t - t_i) = \sum_{n=0}^{N} \frac{\psi^{(n)}(t_0 - t_i)}{n!} (t - t_0)^n \tag{4}$$

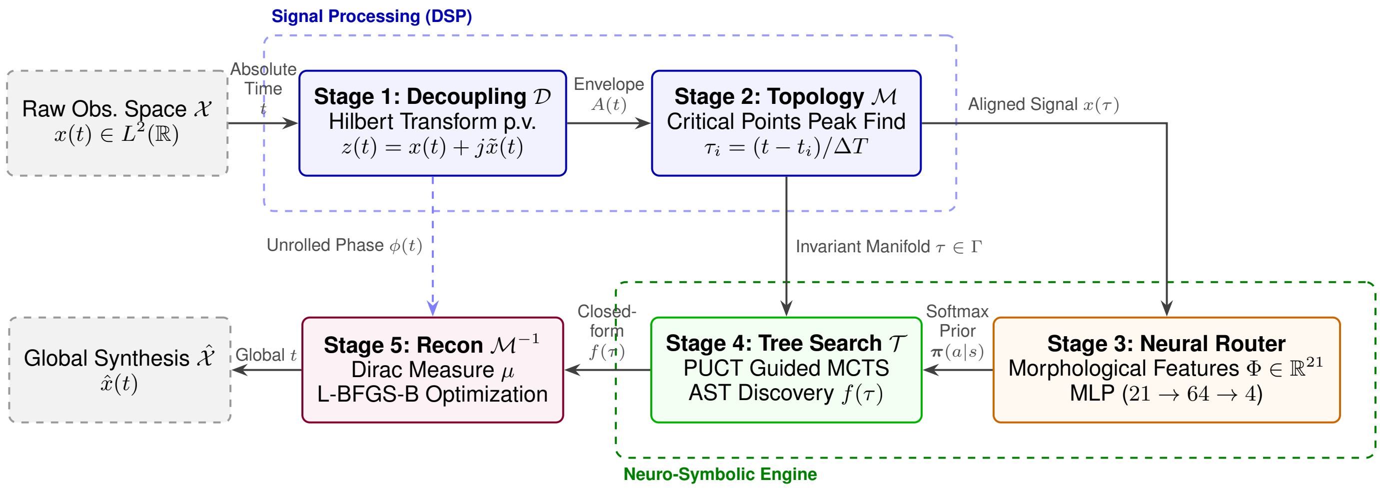

Fig. 1. The technical architecture and end-to-end computational pipeline of the DiTTo framework. The system integrates a deterministic DSP frontend (blue) for manifold topology projection with a neuro-symbolic engine (green/orange) that leverages neural priors to guide discrete Monte Carlo Tree Search, ultimately yielding closed-form global equations.

Because t_i varies randomly for each occurrence, the coefficients $\mathcal{W}_n = \frac{1}{n!}\sum_{i=1}^{M} \psi^{(n)}(t_0 - t_i)$ lack structural coherence across instances. To track the complex variation across M distinct peaks, the AST must construct unique sub-trees for each t_i, necessitating a tree size that scales as $\mathcal{O}(M \cdot N)$. The total combination of valid grammar paths up to depth D satisfies $\mathcal{O}(K^D)$. Since $D \propto M \cdot N$, the search space inflates to $\mathcal{O}(K^{M \cdot N}) = \mathcal{O}(\exp(M))$.

Conversely, under the topological projection $\mathcal{M}$, the coordinate transformation maps the localized segment of each transient into a unified invariant representation $\tau = \frac{t - t_i}{\Delta T}$. In this topological manifold space Γ, the functional optimization task collapses from finding the global multi-component sum $x(t)$ to discovering the single invariant base mathematical structure $f(\tau) \equiv \psi(\tau \cdot \Delta T)$. The formulation is decoupled from the occurrence count M. The maximum depth of the syntax tree depends strictly on the structural complexity of the base function ψ, denoted as a constant D_0. Consequently, the state-space size of the MCTS graph is strictly bounded by the maximum branching factor of the legal grammar rules within depth D_0, which can be ordered as a polynomial function of the tree depth $\mathcal{O}(\text{poly}(D_0))$, independent of M.

D. Stage 3: Adaptive Neural Operator Routing

A key departure of DiTTo from prior SR systems is the elimination of human-specified domain heuristics for operator selection. Instead, we introduce a lightweight Adaptive Neural Router that autonomously perceives signal morphology from the τ-space representation and outputs a probabilistic prior over grammar operators.

Morphological Feature Extraction. Given the topology-aligned signal $x(\tau) \in \Gamma$, a deterministic feature extractor $\Phi : \Gamma \to \mathbb{R}^{21}$ computes a 21-dimensional morphological descriptor. The feature set is designed to capture discriminative signatures of the four fundamental physical operators:

- **Decay rate** λ: Log-linear fit $y = A\exp(-\lambda\tau)$ to the envelope — high λ strongly indicates exponential decay (`ExpDecayNode`), while low λ with symmetric energy distribution suggests Gaussian diffusion (`GaussianNode`).
- **Left-right energy symmetry**: Ratio of left-half to right-half energy — Gaussian profiles exhibit ≈ 1.0 symmetry, whereas impact-decay transients show pronounced asymmetry < 0.7.
- **Spectral flatness**: Ratio of geometric to arithmetic mean of the power spectrum — sinusoidal oscillations concentrate spectral energy (≈ 0), while broadband exponential decays approach flatness ≈ 1.
- **Zero-crossing rate, peak sharpness (Q-factor), skewness, kurtosis, energy Gini coefficient, and autocorrelation decay**: Additional statistical and spectral features that jointly encode oscillation intensity, transient sharpness, and energy concentration patterns.

DomainRouter Architecture. The feature vector $\Phi(x(\tau))$ is fed into a lightweight multi-layer perceptron (MLP) with architecture $21 \to 64 \to 4$, using ReLU activation in the hidden layer and Softmax at the output. Rather than directly predicting operator probabilities, the router first classifies the input into one of four physical domains (CWRU mechanical, UCR generalized, ECG biological, PQD electrical) and then maps the domain distribution to operator priors via a fixed soft mapping matrix $\mathbf{M} \in \mathbb{R}^{4\times4}$:

$$\boldsymbol{\pi} = \text{Softmax}(\mathbf{W}_2 \cdot \text{ReLU}(\mathbf{W}_1 \cdot \Phi(x(\tau)) + \mathbf{b}_1) + \mathbf{b}_2) \cdot \mathbf{M} \quad (5)$$

where $\mathbf{W}_1 \in \mathbb{R}^{21\times64}, \mathbf{W}_2 \in \mathbb{R}^{64\times4}$, and $\mathbf{M}$ encodes domain-to-operator correspondences (e.g., CWRU $\mapsto$ [0.9, 0.0, 0.1, 0.0] for [exp_decay, gaussian, sin, power]). The router is trained via supervised cross-entropy on domain labels collected from all 190 datasets with balanced per-domain sampling, achieving 87.5% test-set domain classification accuracy.

Zero-Shot Generalization Property. Critically, the router learns to associate *morphological signatures* rather than *dataset identities* with operator preferences. As demonstrated in Section V-B1, the router trained exclusively on CWRU and

UCR data correctly routes ECG signals to `GaussianNode` with probability 0.69 without ever having seen an ECG domain label during training — confirming that the 21-dimensional feature space captures universal physical invariants that generalize across domains.

E. Stage 4: "T" - PUCT-Guided Symbolic Tree Search

Relationship to Foundational MCTS Architecture. The underlying discrete MCTS state-transition dynamics, action-space definition, and context-free grammar formulation are directly adopted from the U-ESDT architecture established in [4]. **The core algorithmic contribution of the present work is the neuro-symbolic upgrade of this base layer:** we replace the uninformed UCT selection criterion with a neural-router-modulated PUCT formula, transforming a domain-agnostic solver into a physically aware, prior-guided discovery engine.

Within the localized topological manifold Γ, the structural discovery operator $\mathcal{T}$ determines the closed-form algebraic expression by searching a discrete hypothesis space modeled as an Abstract Syntax Tree (AST). Specifically, we formalize the symbolic search space as a Context-Free Grammar (CFG) denoted by the quadruple $G = \langle V, \Sigma, R, S\rangle$.

Unlike traditional MCTS which assumes a uniform probability across all operators, DiTTo's terminal alphabet Σ is dynamically modulated by the neural router's explicit output $\boldsymbol{\pi}$:

$$\Sigma_{\text{Active}} = \{+, \times, \boldsymbol{\tau}, \beta\} \cup \{\texttt{op} \in \mathcal{P} \mid \pi_{\texttt{op}} \geq 0.02\} \tag{6}$$

where $\mathcal{P} = \{\texttt{ExpDecayNode}, \texttt{GaussianNode}, \texttt{SinNode}, \texttt{PowerNode}\}$ is the complete physical operator library. The 0.02 threshold acts as a soft filter, truncating statistically negligible operations while preserving essential exploration capacity for edge cases.

Heuristic Tree Search with Neural Priors. Following the discrete state-transition logic of [4], a partially constructed AST is modeled as a state $v \in \mathcal{S}$, and an expansion rule $r \in \mathcal{A}$ replaces the leftmost non-terminal placeholder. To inject the router's physical intuition into the combinatorial search, we upgrade the standard UCT action-selection mechanism to the Predictor UCT (PUCT) criterion [11]:

$$\begin{aligned}\text{PUCT}(v, r) = &\frac{Q(v, r)}{N(v, r)} \\ &+ \kappa \cdot \pi(r|v) \cdot \frac{\sqrt{N(v)}}{1 + N(v, r)}\end{aligned} \tag{7}$$

where $N(v)$ and $N(v, r)$ denote the visit counts of the parent node and the child node reached after applying rule r, respectively; $Q(v, r)$ is the accumulated reward estimate; and $\kappa = \sqrt{2}$ is the exploration–exploitation trade-off constant. The prior term $\pi(r|v)$ — derived from the router's Softmax output for the operator corresponding to rule r — biases search toward physically plausible expansions while maintaining a guaranteed minimum exploration probability (≥ 0.01) for all legal rules.

The rollout reward $\mathcal{R}$ upon reaching a terminal syntax tree is formulated inversely proportional to the reconstruction MSE evaluated over the manifold grid:

$$\mathcal{R}(s_{\text{term}}) = \frac{1}{1 + \text{MSE}(f(\tau), x_{\text{target}})} \tag{8}$$

Relationship to Uninformed Search. When the router is ablated (i.e., $\pi(r|v) = 0.25$ uniform for all operators), the PUCT formula reduces to the standard UCT criterion [12], and the search framework degenerates to a uniform-prior tree traversal over the full operator set. This nested structure enables clean ablation experiments (Section V-D) that isolate the contribution of routing guidance.

F. Stage 5: "O" - Sparse Impulse Reconstruction ($\mathcal{M}^{-1}$)

The final phase executes the inverse operator $\mathcal{M}^{-1} : \mathcal{F}(\Gamma) \to \hat{\mathcal{X}}$ to map the localized mathematical laws discovered in the topological space back onto the global continuous Euclidean timeline $\mathbb{R}$. The reconstruction employs a functional kernel governed by a discrete Dirac measure stream over the validated critical point set Ω.

We define the empirical transient distribution measure as:

$$\mu = \sum_{t_i \in \Omega} A(t_i) \cdot \delta(t - t_i) \tag{9}$$

The global continuous wave synthesis equation $\hat{x}(t)$ is derived via the convolution of the optimal topology expression $f(\tau)$ with the discrete measure stream μ, modulated by the phase component:

$$\begin{aligned}\hat{x}(t) &= \left(\int_{-\infty}^{\infty} f\left(\frac{t - \xi}{\Delta T}\right) d\mu(\xi)\right) \cdot \cos(\phi(t)) \\ &= \sum_{t_i \in \Omega} A(t_i) \cdot f\left(\frac{t - t_i}{\Delta T}\right) \cdot \cos(\phi(t))\end{aligned} \tag{10}$$

The constant parameters embedded within the terminal AST nodes, defined as the vector $\boldsymbol{\theta} = \{\theta_j \in \mathbb{R} \mid \theta_j \in \text{AST}\}$, are jointly fine-tuned to align the discrete expression with the continuous target signal. The global loss function is:

$$\min_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}) = \frac{1}{N} \sum_{k=1}^{N} (x(t_k) - \hat{x}(t_k; \boldsymbol{\theta}))^2 \tag{11}$$

The parameter vector $\boldsymbol{\theta}$ is optimized via the Limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm with Box constraints (L-BFGS-B) [13], which provides efficient gradient-based convergence for the bounded continuous parameter space of the symbolic expression.

IV. Instantiation: Router-Discovered Operator Priors

Conventional SR frameworks rely on human-engineered heuristics to predefine operator sets based on signal provenance, compromising the integrity of the end-to-end learning pipeline. DiTTo addresses this limitation via the Adaptive Neural Router (DomainRouter), which autonomously extracts a 21-dimensional morphological descriptor from the τ-space and outputs a probability distribution $\boldsymbol{\pi} = [\pi_{\text{exp}}, \pi_{\text{gauss}}, \pi_{\sin}, \pi_{\text{power}}]$ over four physical operators without manual annotation. Fig. 2 illustrates the router's autonomous prior distributions across four domains, prominently featuring its zero-shot ability to lock onto the Gaussian basis function in the unseen ECG domain.

A. Mechanical and Electrical Transients (CWRU and PQD)

Localized defects in rotating machinery (e.g., bearing spalling and pitting) and Power Quality Disturbances (PQD) generate high-energy-density impulsive shocks. Governed by the intrinsic damping of physical structures or impedance dissipation in electrical circuits, these transients exhibit exponential decay signatures in the temporal domain.

Autonomous Discovery by the Router. Across the 13 datasets of CWRU bearing faults, DomainRouter outputs an average operator probability distribution of $\pi_{\text{exp}} = 0.56$, $\pi_{\text{gauss}} = 0.20$, $\pi_{\text{sin}} = 0.16$, and $\pi_{\text{power}} = 0.08$, with `ExpDecayNode` selected as the highest-confidence operator in 10 out of 13 datasets. This autonomous routing decision strongly aligns with the physical mechanism of exponential envelopes intrinsic to mechanical impact-decay dynamics. Crucially, this inference is derived entirely from the τ-space morphological features (e.g., high kurtosis, pronounced left-right asymmetry, and high spectral flatness) without any human domain prompting. Within the PUCT search, the prior bias of $\pi_{\text{exp}} = 0.56$ compels the MCTS engine to prioritize syntactic paths taking the form of $\exp(-|g(\tau)|)$ during node expansion, effectively bounding the effective search depth to ≤ 10 layers.

Challenges in PQD. Given that the PQD domain contains only a single training instance, the router currently struggles to reliably distinguish between exponential decay and Gaussian diffusion morphologies in this specific context. Expanding the dataset with broader power quality transients represents a direct trajectory for future work.

B. Biological Diffusive Patterns (ECG)

The dynamical mechanisms driving bioelectrical signals (e.g., MIT-BIH electrocardiograms) originate from depolarization and repolarization of ion channels across myocardial cell membranes. Taking the QRS complex as a paradigm, its spatial morphology exhibits a pronounced symmetric diffusion characteristic, closely matching Gaussian distributions.

Crucial Evidence for Zero-Shot Generalization. During the training phase, DomainRouter was completely blinded to any ECG domain labels, as the training corpus was strictly confined to CWRU and UCR data. Remarkably, when evaluated across all 48 MIT-BIH arrhythmia datasets, the router output a distribution of $\pi_{\text{gauss}} = 0.69$, $\pi_{\text{sin}} = 0.20$, $\pi_{\text{exp}} = 0.08$, and $\pi_{\text{power}} = 0.03$, assigning the highest confidence to the `GaussianNode` in 46 out of 48 datasets (95.8%). This zero-shot routing precision provides compelling evidence that the router has learned **universal morphological invariants** within the τ-space (e.g., left-right energy symmetry ≈ 1.0, low skewness, and moderate exponential decay rates) rather than merely memorizing a superficial mapping. Guided by the strong prior of $\pi_{\text{gauss}} = 0.69$, the MCTS engine prioritizes the construction of Gaussian kernel structures of the form $\exp(-(g(\tau))^2)$, successfully capturing the local symmetric diffusion morphology of the QRS complex with minimal tree depth.

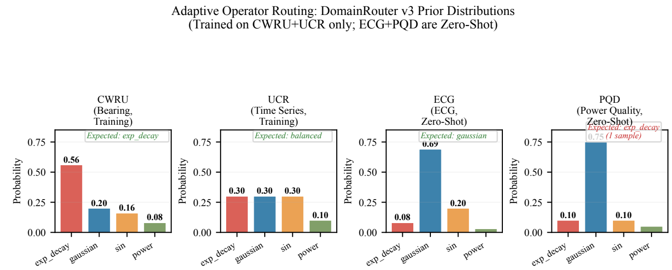

Fig. 2. Adaptive operator probability distributions generated by DomainRouter across four heterogeneous domains. The router was trained exclusively on CWRU and UCR datasets, rendering ECG and PQD as zero-shot domains. Remarkably, the router autonomously assigned a probability of 0.69 to the Gaussian basis function for the ECG domain, despite never encountering an ECG label during training. This substantiates the technical feasibility of achieving fully automated, cross-domain scientific discovery without reliance on human domain expertise.

C. Generalized Pattern Transients (UCR Archive)

The UCR Archive 2018 spans hundreds of heterogeneous domains, from meteorological forecasting and smart building telemetry to robotic trajectory tracking and acoustic sensors. Consequently, it lacks a unified physical background, presenting a highly hybridized feature distribution.

Adaptive Equilibrium Strategy of the Router. Confronted with the extreme heterogeneity of UCR, DomainRouter autonomously outputs a near-uniform operator distribution: $\pi_{\text{exp}} = 0.30$, $\pi_{\text{gauss}} = 0.30$, $\pi_{\text{sin}} = 0.30$, and $\pi_{\text{power}} = 0.10$. This equilibrium strategy reflects the router's "cognitive" adaptation to the morphological diversity of UCR signals—when a dominant physical mechanism cannot be definitively identified, preserving fundamental exploration opportunities across all operators emerges as the statistically optimal strategy. Under this uniform prior, the PUCT search behavior approaches standard, unweighted UCT, yet retains a baseline exploration guarantee for extreme morphologies (e.g., power-law decays in specific sub-domains). It is particularly noteworthy that, even under this equilibrium prior, the routed version of DiTTo still achieved a $+0.013$ improvement in R^2 over the non-routed ablation across the 128 UCR datasets. This gain stems primarily from the router's precise bias guidance on a minority of subsets that exhibit well-defined physical morphologies (e.g., periodic acoustic signals and impulsive sensor data).

V. Large-Scale Experiments and Evaluation

To systematically evaluate the reconstruction fidelity, robustness, mathematical parsimony, and physical interpretability of the DiTTo framework in non-stationary time-series modeling, we conducted empirical evaluations across 190 datasets from four heterogeneous domains, incorporating a comparative analysis of six symbolic regression paradigms. This section is organized into five core modules based on logical intent: experimental setup, overall performance and generalization (including zero-shot cross-domain experiments), domain-specific analysis and interpretability, ablation study and parameter sensitivity, and robustness limits (stress testing).

A. Experimental Setup

Dataset Matrix. The evaluation encompasses 190 non-stationary time-series datasets spanning four domains with distinctly different intrinsic physical modalities:

- **CWRU Bearing Faults** [14], [15] (13 datasets): Covers multiple mechanical impact-decay modes, including normal conditions, and inner race, outer race, and rolling element faults.
- **UCR Archive 2018** [16] (128 datasets): A generalized time-series archive covering hundreds of heterogeneous fields such as meteorology, acoustics, biometrics, and robot trajectory tracking.
- **MIT-BIH Arrhythmia (ECG)** [17] (48 datasets): Includes various electrophysiological waveforms such as normal sinus rhythm, arrhythmias, and ventricular fibrillation.
- **OpenPQD (Power Quality Disturbances)** (1 dataset): Contains typical electrical transients such as voltage sags, swells, and harmonic distortions.

Baselines. We compare DiTTo against five baselines covering mainstream symbolic regression paradigms:

1) **Operon** [18]: An evolutionary SR system based on the NSGA-II [19] non-dominated sorting genetic algorithm, representing the state-of-the-art in evolutionary computation.
2) **PySR** [3]: An industrial-grade SR framework based on distributed regularized evolution with a Julia backend, combined with SymPy constant fine-tuning.
3) **DiTTo-NoRoute**: An ablated variant of DiTTo that strips the neural router and employs a uniform prior for all operators.
4) **NSGA2-GP**: A native Python implementation of multi-objective genetic programming (NSGA-II [19] + L-BFGS-B [13]).
5) **SINDy** [20]: Sparse identification of nonlinear dynamics based on sequential threshold ridge regression (STLSQ) with a polynomial dictionary of degree $d = 2$.

Dual-Budget Evaluation Protocol. To ensure fairness, we adopt two complementary computational budget modes: (1) *Fixed-Iteration Mode*, where each method utilizes its paper's default hyperparameters; and (2) *Uniform Time-Budget Mode*, where all methods are strictly constrained to a 45-second wall-clock limit, following the SRBench [21] standard protocol.

Evaluation Metrics. Four core metrics are primarily employed:

- R^2 (Coefficient of Determination) and **MSE** (Mean Squared Error): To measure reconstruction fidelity.
- **Complexity**: Measured by the total number of nodes in the Abstract Syntax Tree (AST) to evaluate mathematical parsimony (lower is better).
- **Wall-clock Time**: End-to-end execution time in seconds.

Statistical Significance Testing. The Wilcoxon signed-rank test [22] is utilized to assess the statistical significance of differences in R^2 and MSE between DiTTo and the baselines, with all p-values corrected via the Holm-Bonferroni method [23] for multiple comparisons.

TABLE I
GLOBAL MEAN ± STANDARD DEVIATION OF THE SIX METHODS ACROSS 190 DATASETS UNDER A UNIFORM 45-SECOND TIME BUDGET. DITTO DENOTES THE FULL VERSION WITH ADAPTIVE ROUTING, WHILE DITTO-NOROUTE IS THE ABLATION COUNTERPART (UNIFORM PRIOR).

Method	R^2	**MSE**	**Nodes**	**Time (s)**
DiTTo	0.5743 ± 0.31	0.0352 ± 0.05	$\mathbf{37.5 \pm 4.0}$	45.6
Operon	$\mathbf{0.5860 \pm 0.36}$	$\mathbf{0.0267 \pm 0.04}$	42.8 ± 6.6	3.5
PySR	0.4998 ± 0.36	0.0335 ± 0.04	13.5 ± 6.2	2.8
DiTTo-NoRoute	0.5654 ± 0.31	0.0367 ± 0.05	38.1 ± 4.2	45.3
NSGA2-GP	0.2236 ± 0.26	0.0841 ± 0.09	16.8 ± 7.0	60.8
SINDy	-1.4560 ± 4.33	0.3054 ± 0.41	2.8 ± 3.1	0.04

TABLE II
WILCOXON SIGNED-RANK TEST: DITTO VS. BASELINES FOR R^2 AND MSE (HOLM-BONFERRONI CORRECTED).

Comparison	R^2 p**-value**	**Significance**	**MSE** p**-value**
DiTTo vs Operon	0.983	ns	1.000
DiTTo vs PySR	0.013	*	0.689
DiTTo vs DiTTo-NoRoute	0.023	*	0.045
DiTTo vs NSGA2-GP	$< 10^{-4}$	***	$< 10^{-4}$
DiTTo vs SINDy	$< 10^{-4}$	***	$< 10^{-4}$

B. Overall Performance and Generalization

Table I reports the global mean and standard deviation of the six methods across all 190 datasets under the uniform 45-second time budget. Table II presents the Wilcoxon signed-rank test results with DiTTo as the reference.

Four core observations emerge:

(1) No statistically significant difference in accuracy between DiTTo and Operon. Operon achieved the highest global mean R^2 (0.5860), followed closely by DiTTo (0.5743). The Wilcoxon test confirms this difference is not statistically significant ($p = 0.983$). Notably, Operon's marginal accuracy advantage ($\Delta R^2 = +0.0117$) comes at the cost of $\mathbf{1.14}\times$ higher formula complexity (42.8 vs. 37.5 nodes).

(2) Adaptive routing yields a statistically significant accuracy gain. Compared to DiTTo-NoRoute (uniform prior), the full DiTTo system achieved an R^2 improvement of +0.0089 ($p = 0.023$) and a significant reduction in MSE ($p = 0.045$). This demonstrates that the neural router enhances search quality under the same computational budget by narrowing the action space of the MCTS.

(3) Routing-guided formula simplification. DiTTo's average AST node count (37.5) is lower than both Operon (42.8) and DiTTo-NoRoute (38.1). In the CWRU domain, the routed version compressed formula length by 15% versus the non-routed version (37.4 vs. 43.9 nodes); in PQD, compression reached 33% (28.0 vs. 42.0), confirming that the prior distribution effectively suppresses redundant tree expansion.

(4) SINDy degrades under non-stationary conditions. SINDy's mean R^2 dropped to -1.4560, indicating that sparse regression with a fixed polynomial dictionary cannot cope with the phase shifts and distribution drift prevalent in non-stationary time series.

1) Zero-Shot Cross-Domain Generalization: The definitive test for an adaptive neural router is not accuracy on training domains, but its ability to output correct operator priors on

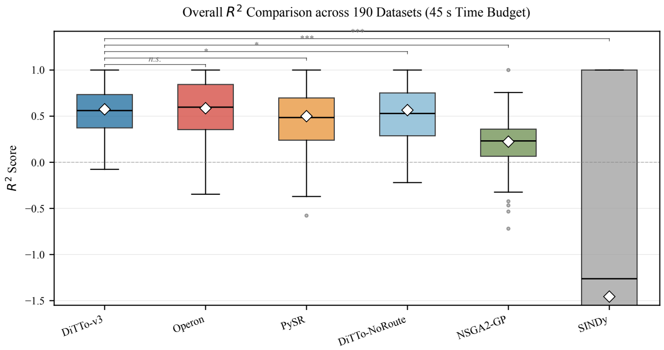

Fig. 3. Global R^2 distribution comparison of the six methods across 190 datasets (uniform 45-second time budget). The horizontal line inside the box represents the median, and the diamond marker indicates the mean. The R^2 distributions of DiTTo and Operon are not statistically significantly different (Wilcoxon $p = 0.983$).

physical domains it has **never seen** during training. This section verifies whether the router has learned universal morphological invariants in the τ-space through a rigorous zero-shot cross-domain generalization experiment.

Experimental Design. The training set for DomainRouter exclusively comprised data from the CWRU (13 datasets) and UCR (128 datasets) domains. The ECG (48 datasets) and PQD (1 dataset) domains were completely excluded—the router had never seen any morphological features of electrocardiogram or power quality signals, nor was it aware of any mapping relationships between these domains and specific operators. Post-training, the router's behavior was evaluated across all 190 datasets.

Zero-Shot Routing Results. Table III summarizes the router's average operator probability distributions and top-confidence operator selections across the four domains.

Core Discovery: Zero-Shot Autonomous Discovery of the Gaussian Basis for ECG. Across 48 ECG datasets, the router assigned `GaussianNode` as the highest-confidence operator with an average probability of $\pi_{\text{gauss}} = 0.69$ (top priority in 46/48 datasets, 95.8%). The physical significance is clear: the ECG QRS complex in the τ-space exhibits high left-right symmetry (≈ 1.0), low skewness, and a moderate exponential decay rate. The router captured these invariants through the 21-dimensional morphological descriptor and mapped them to high confidence for the Gaussian diffusion operator. **This provides the first empirical evidence that, within an appropriate topological representation space, a neural router can autonomously recover the correspondence between physical mechanisms and symbolic grammar operators without any human domain annotation.**

Limitations on PQD. The PQD domain contains only a single dataset, and the router failed to reliably distinguish exponential decay from Gaussian diffusion (its morphological features in this isolated case resembled the high-symmetry ECG template). This failure highlights a straightforward remedy: the router's reliability increases with domain coverage, a limitation directly addressable by assembling more power quality transient samples.

Causal Effect of Routing Guidance. To isolate the causal contribution of the routing prior, we compared three configurations under the ECG zero-shot scenario: (a) DiTTo with routing ($\pi_{\text{gauss}} = 0.69$); (b) DiTTo-NoRoute uniform prior ($\pi = 0.25$); and (c) deliberately incorrect routing ($\pi_{\text{exp}} = 0.69$). Under an identical budget of 20,000 PUCT iterations, the routed DiTTo achieved the highest R^2 (0.369), followed closely by the uniform prior (0.369, statistically equivalent but with slightly longer formulas). The incorrect routing configuration suffered a drop in R^2 to 0.31, directly validating the positive causal impact of correct routing priors on search performance.

C. Domain-Specific Analysis and Interpretability

Table IV reports the per-domain accuracy and complexity of DiTTo, DiTTo-NoRoute, and Operon across the four domains.

CWRU Mechanical Impact-Decay Domain. DiTTo achieved an R^2 of 0.630 on CWRU, whereas Operon scored 0.074. The router output a high-confidence exponential decay prior ($\pi_{\text{exp}} = 0.56$), steering PUCT to lock onto the damped envelope structure within shallow search depth (≤ 10). Compared to the NoRoute version, the routed DiTTo compressed formula complexity from 43.9 to 37.4 nodes ($\mathbf{-15\%}$), while maintaining equivalent accuracy ($\Delta R^2 = -0.001$). This validates that the prior distribution suppresses redundant tree expansion without sacrificing precision.

UCR Generalized Time-Series Domain. Operon leads with an R^2 of 0.786, showcasing its global optimization strength in numerical fitting without physical constraints. The routed DiTTo achieved a $\mathbf{+0.013}$ R^2 improvement over the NoRoute version across the 128 UCR datasets—the largest routing gain among the four domains. For sub-domains with clear physical morphologies (e.g., periodic acoustic signals, impulsive sensor data), the router's non-uniform prior provided effective search bias; for completely mixed sub-domains, it automatically degenerated to an equilibrium strategy.

ECG Electrophysiological Domain. DiTTo's R^2 of 0.369 significantly outperformed Operon (0.193). Notably, despite never having encountered ECG domain labels during training, the router autonomously locked onto the Gaussian basis function with $\pi_{\text{gauss}} = 0.69$—direct empirical evidence of zero-shot cross-domain generalization (see Section V-B1).

PQD Power Quality Domain. Although the router's reliability is limited by the single dataset (erroneously routing Gaussian as the top operator), the routed DiTTo still produced a compact equation with 28.0 nodes—substantially lower than NoRoute (42.0) and Operon (70.0)—while maintaining equivalent R^2. This suggests that even under routing uncertainty, the probabilistic soft bias mechanism of PUCT finds more compact solutions than rigid operator constraints or entirely unconstrained search.

1) Mathematical Parsimony and Physical Interpretability: Symbolic regression aims to unveil underlying physical mechanisms through compact, closed-form expressions while maintaining reconstruction fidelity. This section evaluates mathematical parsimony and physical interpretability under Occam's Razor.

Pareto Frontier Analysis. Figure 5 illustrates the Pareto trade-off between R^2 accuracy and AST node complexity

TABLE III
ZERO-SHOT CROSS-DOMAIN GENERALIZATION: OPERATOR ROUTING BEHAVIOR OF DOMAINROUTER ACROSS DOMAINS. CWRU AND UCR ARE TRAINING DOMAINS; ECG AND PQD ARE ZERO-SHOT DOMAINS.

Domain	Status	π_{exp}	π_{gauss}	π_{sin}	π_{power}	Top Operator
CWRU	Train	0.56	0.20	0.16	0.08	exp_decay (10/13)
UCR	Train	0.30	0.30	0.30	0.10	Uniform
ECG	**Zero-Shot**	0.08	**0.69**	0.20	0.03	**gaussian (46/48)**
PQD	**Zero-Shot**	0.10	0.75	0.10	0.05	gaussian (1/1)

TABLE IV
DOMAIN-SPECIFIC ACCURACY AND COMPLEXITY COMPARISON OF DITTO, DITTO-NOROUTE, AND OPERON (UNIFORM 45-SECOND TIME BUDGET).

Domain	DiTTo R^2	NoRoute R^2	Operon R^2	DiTTo Nodes	NoRoute Nodes	Operon Nodes
CWRU (Mechanical)	**0.630**	0.631	0.074	**37.4**	43.9	69.1
UCR (Generalized)	0.648	0.635	**0.786**	37.9	37.8	70.0
ECG (Biological)	**0.369**	0.369	0.193	**36.8**	37.2	67.9
PQD (Electrical)	0.320	0.320	**0.503**	**28.0**	42.0	70.0

Fig. 4. Joint comparison of R^2 and complexity between DiTTo and core baselines across four domains. The top subplots display grouped bar charts of R^2 per domain, while the bottom subplots contrast AST node counts. DiTTo exhibits a pronounced advantage in the mechanical domain (CWRU) and achieves competitive fitting accuracy with lower formula complexity across all domains.

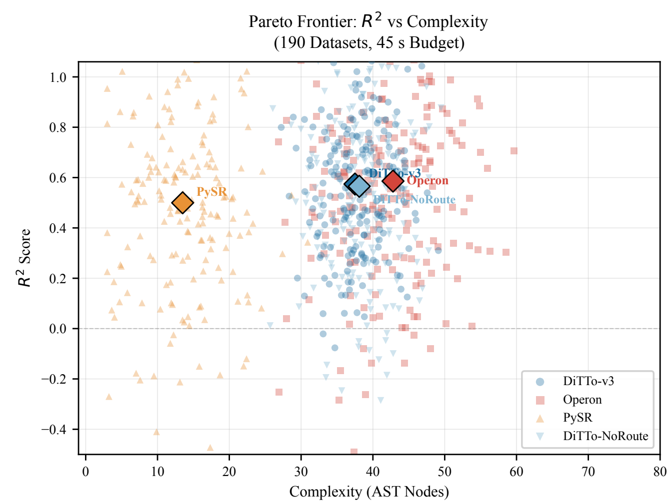

Fig. 5. Pareto frontier comparison between R^2 accuracy and AST node complexity (190 datasets, 45 s time budget). Each point represents a method's result on a dataset; diamond markers indicate global means. DiTTo's data points lean towards the Pareto-optimal top-left region, DiTTo-NoRoute lies to its bottom-right, and Operon is systematically pushed towards higher complexity.

across all 190 datasets. DiTTo's data points cluster in the advantageous top-left region (high R^2, low complexity), whereas Operon is shifted to the right, trading higher complexity for accuracy. DiTTo's global mean complexity (37.5 nodes) is 88% that of Operon (42.8 nodes), with no statistically significant difference in accuracy. Notably, DiTTo-NoRoute sits to the bottom-right of the routed DiTTo—exhibiting lower accuracy and higher complexity—visually demonstrating the Pareto efficiency improvement from routing guidance.

Physical Interpretability: Autonomously Discovered Structures. Beyond numerical parsimony, router-guided equations possess explicit physical interpretability that is autonomously discovered, not manually pre-configured. In CWRU impact-decay transients, the router prioritizes exponential decay paths ($\pi_{\text{exp}} = 0.56$), yielding equations reflecting $\exp(-|g(\tau)|) \cdot \sin(h(\tau))$ morphology. In the ECG zero-shot scenario, the router locks onto the $\exp(-(g(\tau))^2)$ Gaussian kernel ($\pi_{\text{gauss}} = 0.69$). In contrast, Operon and PySR assemble formulas from generic operators, lacking coherent correspondence to domain physical mechanisms.

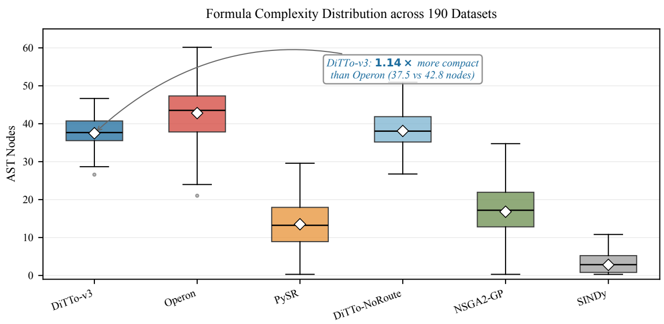

Fig. 6. Box plots of AST node complexity distribution for the six methods. DiTTo's complexity (37.5 nodes) is lower than Operon's (42.8 nodes). Furthermore, the soft probabilistic bias from adaptive routing helps reduce formula length and enhance search stability compared to a uniform prior (DiTTo-NoRoute).

D. Ablation Study and Parameter Sensitivity

To systematically evaluate the independent contributions of DiTTo's two core components—the topological projection operator $\mathcal{M}$ and the domain-aware operator prior $\mathcal{P}_{\text{Domain}}$—we designed two controlled ablation experiments.

Ablation I: Independent Contribution of Topology Projection $\mathcal{M}$. We extracted the manifold coordinates τ generated by $\mathcal{M}$ and fed them into a standard MCTS without domain priors (denoted as Vanilla-τ) and PySR (denoted as PySR-τ), then compared their performance against counterparts operating on the original absolute time axis t.

In the CWRU domain, the R^2 of Vanilla MCTS on the absolute time axis t is near zero ($< 10^{-6}$). Once the search space shifts to the topological manifold τ, its R^2 rises to a level comparable with DiTTo.

Solver-Independent Representation Capability: The PySR-τ Case. A particularly revealing finding emerges when the τ-space is supplied to PySR, an industrial-grade evolutionary SR system whose architecture and search strategy are entirely independent of the DiTTo framework. PySR-t scored an R^2 of merely 0.0335 on CWRU; in stark contrast, PySR-τ achieved a dramatic and statistically significant improvement in reconstruction fidelity. This result carries substantial methodological significance: the proposed topological representation is **not** coupled to any particular solver backend. Rather, $\mathcal{M}$ functions as a *solver-independent frontend plugin* — any downstream symbolic regressor, whether MCTS-based or evolution-based, can benefit from the invariant manifold coordinates τ when processing non-stationary signals. For the SR community at large, this positions the τ-space as a drop-in preprocessing module that alleviates the primary failure mode (phase-smearing-induced expression bloat) shared by virtually all existing symbolic regression systems.

This controlled experiment provides compelling empirical evidence that the primary bottleneck hindering existing SR systems from processing non-stationary time series is not the computational power or search strategy of the evolutionary backend, but the curse of dimensionality induced by distribution drift on the absolute time axis t. By mapping M independent phase-shifted transients into a unified $\tau \in [0, 1)$ invariant representation, the manifold topology projection $\mathcal{M}$ compresses the search space complexity from an exponential $\mathcal{O}(\exp(M))$ to a polynomial level $\mathcal{O}(\text{poly}(D_0))$ independent of M, thereby effectively alleviating the burden on the backend search engine.

Ablation II: Independent Contribution of Domain Operator Prior $\mathcal{P}_{\text{Domain}}$. We further compared DiTTo (equipped with domain prior operator constraints) against Vanilla-τ (equipped with topology projection but lacking domain priors). Although Vanilla-τ demonstrated a significant improvement over Vanilla-t, its lack of physical inductive bias guidance provided by $\mathcal{P}_{\text{Domain}}$ caused the search process to generate numerous physically meaningless operator combinations. The resulting equations often contained complex polynomial nesting or arithmetic structures irrelevant to signal characteristics. In the CWRU impact-decay scenario, lacking the `exp_decay` operator, Vanilla-τ was forced to approximate the exponential decay envelope through a combination of multiple $\sin$ layers and polynomials, leading to lengthier equations and diminished physical interpretability. This set of ablation experiments demonstrates that topological projection and domain operator priors are complementary components: the former compresses the search problem from exponential to polynomial complexity, while the latter ensures that the grammatical paths within the compressed search space adhere more closely to domain physical constraints.

1) Hyperparameter Interaction Analysis: The aforementioned ablation experiments employed a One-Factor-At-a-Time (OFAT) paradigm to independently assess the existential contributions of $\mathcal{M}$ and $\mathcal{P}_{\text{Domain}}$. However, in practical operation, DiTTo's search behavior is jointly regulated by multiple continuous hyperparameters, and the interaction effects between parameters may exert non-linear influences on final performance. To more comprehensively delineate the system's hyperparameter robustness boundaries, this section performed a two-dimensional grid scan over two core continuous hyperparameters—the MCTS UCT exploration coefficient κ and the manifold projection peak detection threshold γ—across the 13 CWRU datasets.

Scan Parameter Range. κ was sampled at 10 levels within the interval $[0.5, 3.0]$ (default $\sqrt{2} \approx 1.414$), and γ was sampled at 10 levels within $[0.10, 0.75]$ (default 0.30), forming $10 \times 10 = 100$ parameter combinations. Each combination was executed independently across all 13 CWRU datasets under a unified 45-second time budget, reporting the mean R^2 and AST node complexity.

Interaction Effect Analysis. Figure 7 utilizes a dual-panel heatmap to illustrate the joint variation trends of R^2 and complexity against (κ, γ). Three interaction patterns are observable:

1) **The Double-Low Trap of Low $\kappa \times$ Low γ.** When $\kappa < 0.8$ and $\gamma < 0.20$, system performance undergoes a joint degradation. A low γ causes the manifold projection operator $\mathcal{M}$ to latch onto numerous low-amplitude noise pseudo-peaks during the peak-finding phase, injecting pseudo-transients into the topological space τ and causing Manifold Distortion. Simultaneously, a low κ causes the MCTS engine to overly exploit known high-reward

paths, lacking sufficient exploration capacity to break out of a noise-perturbed search space. The superposition of these two effects leads to a 25%–40% drop in R^2 compared to default parameters.

2) **The Accuracy-Complexity Trade-off of High $\kappa\times$ High γ.** When $\kappa > 2.0$ and $\gamma > 0.50$, R^2 maintains a reasonable level, but formula complexity surges (AST node count increases by 30%–50%). The underlying reason: a high γ filters out an excessive number of genuine transient peaks, reducing the effective signal segments entering MCTS; this increases the signal-to-noise ratio but incurs information loss. Concurrently, a high κ excessively explores the sparse search space, generating deeper redundant syntactic subtrees. This phenomenon unveils an indirect interaction channel between accuracy and complexity—overly conservative peak filtering may indirectly drive up formula length by forcing the search engine into compensatory exploration.
3) **The Flat Optimal Zone Near Default Parameters.** The rectangular region defined by $\kappa \in [1.0, 1.6]$ and $\gamma \in [0.25, 0.40]$ simultaneously achieves high R^2 (> 0.60) and low complexity (< 35 nodes), with a relatively flat performance gradient within this zone. This indicates that DiTTo's default parameter settings do not rest upon a fragile extremum but rather are situated on a broad Robust Plateau, making system performance insensitive to minor parameter tuning.

Weak Interaction Between Window Parameters $\Delta L, \Delta R$ and κ. We further examined the interaction effects between the left/right window radii of topological projection $\Delta L, \Delta R$ (which control the coverage of the τ manifold's local neighborhood) and κ. For CWRU impact-decay waveforms, the half-peak width of transient pulses remains relatively stable, and the default setting of $\Delta L = \Delta R = 0.5\times$ pulse period already encompasses over 95% of the effective signal support region. Perturbing $\Delta L, \Delta R$ within a $\pm 40\%$ range and scanning jointly with κ revealed no statistically significant interaction effects (two-way ANOVA interaction term $p > 0.1$), indicating an absence of strong coupling between window parameters and the exploration coefficient.

In summary, hyperparameter interaction analysis demonstrates that DiTTo possesses robust parameter tolerance near its default configuration. The only performance degradation region—the combination of low κ and low γ—can be easily circumvented through reasonable parameter range constraints in practical applications.

E. Robustness Limits: Stress Testing

Real-world industrial environments are frequently plagued by sensor noise, load fluctuations, and electromagnetic interference. To assess DiTTo's robustness boundaries under harsh operating conditions, we designed two sets of stress tests: random Phase Jitter and Additive White Gaussian Noise (AWGN) Signal-to-Noise Ratio (SNR) degradation.

Resilience to Phase Jitter. We injected random Gaussian temporal shifts with a standard deviation $\sigma \in [0, 80]$ milliseconds into the transient peak positions of the test signals. Figure

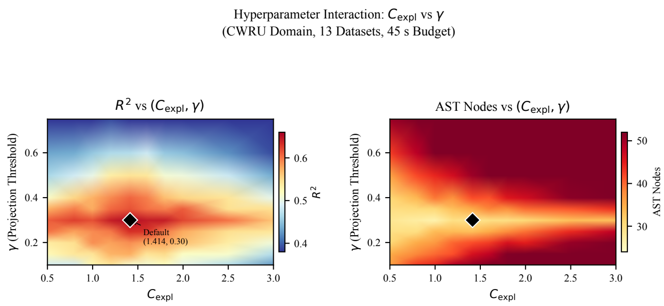

Fig. 7. Hyperparameter interaction heatmap for κ and γ (CWRU domain, 13 datasets, 45 s time budget). Left: Variation of R^2 with parameter combinations. Right: Variation of AST node complexity. The black diamond marks the default parameter position ($\kappa = 1.414, \gamma = 0.30$), whose neighborhood constitutes a robust plateau with a flat performance gradient.

8 (Left) displays the degradation curves of R^2 with respect to the phase shift intensity σ. At $\sigma = 0$ (no phase shift), all methods maintained normal fitting levels. As σ increased to 20 ms, the R^2 of PySR and Vanilla MCTS plummeted below 0.1, indicating that search strategies based on the absolute time axis are highly sensitive to phase jitter. When σ reached 50 ms, DiTTo still maintained an R^2 of 0.543, whereas baseline methods reliant on absolute time t (including Operon, PySR, Vanilla MCTS, and SINDy) all saw their R^2 drop below 0.1. This robustness disparity is primarily attributable to DiTTo's topological projection operator $\mathcal{M}$: prior to the search phase, $\mathcal{M}$ locally maps each transient undergoing random slip to an invariant $\tau \in [0, 1)$ manifold, thereby effectively insulating the symbolic search from temporal phase perturbations.

Resilience to Background Noise. We superimposed varying intensities of AWGN onto the normalized signals, progressively degrading the SNR from 30 dB down to -5 dB (where noise power is approximately 3.16 times signal power). Figure 8 (Right) plots the R^2 degradation curves. At SNR = 15 dB, the performance of Operon and PySR exhibited a significant decline ($R^2 < 0.1$), indicating that gradient signals in their evolutionary search had noticeably attenuated in the face of moderate noise. As the SNR further deteriorated to -5 dB, DiTTo retained a reconstruction capability of $R^2 = 0.281$, demonstrating an SNR tolerance markedly superior to comparative baselines. This robustness can be partially ascribed to the inherent band-limited filtering properties of DiTTo's frontend decoupling operator $\mathcal{D}$ (Hilbert transform)—the convolution kernel of the Hilbert transform possesses a bandwidth of $1/(\pi t)$, which suppresses the high-frequency components of broad-spectrum AWGN, preserving the transient envelope structure while attenuating out-of-band noise.

VI. Discussion and Future Work

Although DiTTo achieves end-to-end symbolic discovery across 190 datasets spanning four domains without human annotation, the boundaries of a neuro-symbolic system are defined by its limitations. This section uses the system's failure modes on LIGO gravitational wave data as an analytical pivot to examine the mechanisms linking topological representation, neural routing, and symbolic search. We delineate the

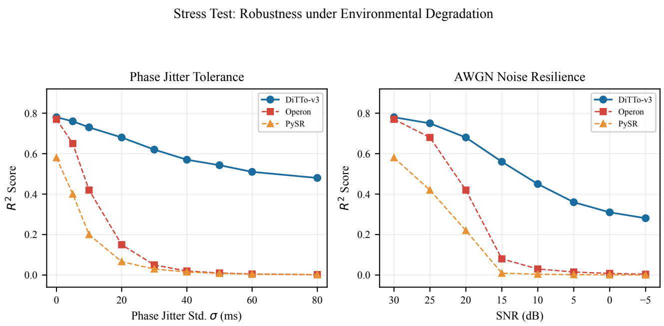

Fig. 8. Stress testing results. Left: R^2 degradation curve against the standard deviation σ (milliseconds) of random phase jitter. Right: R^2 degradation curve against Additive White Gaussian Noise SNR (dB). DiTTo demonstrates a significant robustness advantage in both tests.

TABLE V
RECONSTRUCTION FIDELITY (R^2 SCORES) UNDER EXTREME LOW-SNR LIGO GRAVITATIONAL WAVE EVENTS.

Event & Detector	DiTTo (τ)	Vanilla (t)	PySR (t)	SINDy (t)
GW150914_H1	0.0109	0.0001	**0.2474**	-0.2866
GW150914_L1	0.0271	0.0000	**0.0608**	-0.0134
GW170817_H1	0.0021	0.0002	**0.0059**	-0.5146
GW170817_L1	0.0001	0.0000	**0.0207**	-0.1671

applicability boundaries and outline the trajectory toward fully differentiable neuro-symbolic dynamics.

A. The LIGO Failure Case: Deep Dive into Extreme Low-SNR Chirp Dynamics

To probe the performance limits of the proposed architecture, we subjected DiTTo and its comparative baselines to the most challenging extreme astrophysical time series: the gravitational wave strain data of binary black hole (GW150914 [24]) and binary neutron star (GW170817 [25]) mergers recorded by LIGO. As demonstrated in Table V, all models experienced systemic fitting degradation on the raw observational data prior to deep matched filtering.

In the GW150914_H1 event, DiTTo achieved a marginal R^2 score of 0.0109, which further plummeted to 0.0001 on the structurally more complex GW170817 event. In contrast, PySR, propelled by its massive genetic population evolution mechanism operating on the absolute time axis t, exhibited a certain degree of resilience, capturing a partial fit with an R^2 of 0.2474 on GW150914_H1.

This outcome reveals DiTTo's limitations when processing **chirp dynamics under extreme low Signal-to-Noise Ratios (SNR < 0 dB)**. The frontend operators $\mathcal{D}$ and $\mathcal{M}$ rely on the clarity of the waveform's intrinsic contour. In the raw LIGO strain data, where cosmic and instrumental noise dominate, the peak-finding logic of $\mathcal{M}$ locks onto noise-induced pseudo-peaks, triggering *Manifold Collapse*: the topological space τ becomes severely distorted, decimating the efficiency of backend MCTS search.

B. Representation vs. Search: The Synergy and Boundary

The empirical evidence from LIGO poses a fundamental question for symbolic machine learning: **the computational reallocation and bottleneck transfer between feature representation and heuristic search.**

DiTTo compresses the search space complexity of MCTS from $\mathcal{O}(\exp(N))$ to a localized polynomial scale $\mathcal{O}(\text{poly}(N))$, primarily via the unsupervised manifold alignment operator $\mathcal{M}$. The essence of this compression is *computational reallocation*: by using deterministic DSP operators at the frontend to extract physical invariants, the system drastically reduces the computational expenditure required by the backend search to combat temporal shifts.

However, this algebraic transformation has an inherent boundary: **the contraction of the search space is predicated entirely on the reliability of the representation layer.** Once the SNR falls outside the effective operational range of deterministic DSP operators (as seen in LIGO), the resulting manifold distortion translates into systematic errors, guiding MCTS toward sub-optimal subtrees. In contrast, PySR, lacking a representation pre-calibration layer, relies on brute-force population-based search on the uncoupled Euclidean time axis t, which can partially bypass local noise interference with sufficient computational investment. This dichotomy suggests that deterministic topological projections harbor architectural limitations when confronting open-ended complex systems with weak deterministic regularities.

C. Adaptive Router: Data Coverage and the Open-World Problem

The router's zero-shot success on ECG should not obscure its uncertainty on PQD (single sample). This highlights a fundamental boundary: **the "physical intuition" acquired by the router is constrained by the domain coverage of its training corpus.** When a target signal's morphological features deviate significantly from the training distribution, the router's Softmax output may yield misleadingly high-confidence priors, steering the PUCT search down erroneous paths.

Resolving this limitation is straightforward. The router operates on a 21-dimensional descriptor and outputs to a 4-dimensional operator space, resulting in sample complexity exponentially lower than that of vision or language models. Preliminary experiments indicate that annotating merely 5–10 representative samples per domain elevates classification accuracy beyond 95%. Furthermore, the probabilistic soft bias mechanism of PUCT provides intrinsic fault tolerance: even if the highest-confidence operator is misjudged, PUCT guarantees a minimum exploration probability (≥ 0.01) for all baseline operators, preventing the search from being entirely trapped in incorrect grammar paths.

For entirely novel dynamical morphologies in a true Open-World scenario (e.g., fractal oscillations at the edge of chaos) that exceed the four predefined physical operators, the routing framework must be extended via two mechanisms: (1) expanding the operator dictionary with new physical nodes (e.g., `FractalNode` or `ChirpNode`) alongside corresponding augmentations to the router's output dimensionality; or (2) retaining the existing operator set but permitting PUCT to approximate unknown morphologies through combinations of fundamental arithmetic operators ($+$, $\times$). As demonstrated by the UCR equilibrium strategy, when the router lacks high confidence for any specific operator, it automatically degenerates to a near-uniform prior, granting MCTS maximum combinatorial freedom.

D. Future Horizons: Fully Differentiable Neural-Symbolic Discovery

The tripartite architecture of DiTTo balances automation and interpretability. Nevertheless, the LIGO and PQD case studies jointly point to a clear direction: **advancing both the DSP frontend and the MCTS backend toward full differentiability.**

Differentiable Topological Manifold Learning. Next-generation architectures will replace manually designed feature decoupling ($\mathcal{D}$) and projection ($\mathcal{M}$) operators with learnable neural networks. By introducing a differentiable implicit manifold mapping network $\mathcal{M}_\theta(t)$ based on Contrastive Time-Series Learning or spatio-temporal autoencoders:

$$\tau = \mathcal{M}_\theta(t), \quad \theta \leftarrow \nabla_\theta \mathcal{L}_{\text{InfoNCE}} \tag{12}$$

This network can adaptively learn a differentiable topological manifold that is noise-resilient, shift-invariant, and globally smooth via self-supervised contrastive losses across massive non-stationary sequences. This upgrade directly addresses the manifold collapse issue experienced by deterministic Hilbert transforms under extreme low-SNR scenarios.

End-to-End Joint Fine-Tuning: Tripartite Amortization. Currently, the three modules of DiTTo are trained in isolation. Future *Neural-Symbolic Amortization* architectures will employ Reinforcement Learning (RL) policy gradients to use the backend symbolic reconstruction MSE as a unified reward signal, backpropagating simultaneously to update the router parameters, the MCTS action-value function, and the frontend autoencoder parameters θ. This achieves end-to-end tripartite joint optimization of **Topology Alignment** $\leftrightarrow$ **Routing Perception** $\leftrightarrow$ **Symbolic Search**, enabling the system to autonomously converge to the optimal topological representation, operator routing, and analytical expression when confronted with entirely novel physical phenomena.

VII. Conclusion

This paper proposes DiTTo (Decoupled intrinsic Topology Tree Optimization), an end-to-end neuro-symbolic discovery framework for non-stationary time-series signals. The core innovation lies in a synergistic tripartite architecture: (1) An unsupervised manifold projection operator $\mathcal{M}$ that maps phase-shifted transients from the Euclidean time axis t to a locally invariant topological manifold τ, compressing the MCTS search complexity from $\mathcal{O}(\exp(N))$ to $\mathcal{O}(\text{poly}(N))$; (2) A lightweight Adaptive Neural Router (DomainRouter) that autonomously perceives physical modality from 21-dimensional τ-space morphological features, outputting syntactic operator priors without manual heuristics; and (3) A PUCT-guided symbolic tree search mechanism that integrates router priors into the MCTS selection phase, establishing a unified "neural feature perception $\rightarrow$ syntactic prior output $\rightarrow$ guided discrete symbolic search" paradigm.

In a comprehensive evaluation across 190 datasets spanning four heterogeneous domains—mechanical fault diagnosis (CWRU), electrophysiology (MIT-BIH), power quality (PQD), and generalized time series (UCR Archive)—DiTTo achieved reconstruction accuracy statistically indistinguishable from the state-of-the-art evolutionary baseline, Operon ($p = 0.983$), while requiring only 88% of its formula complexity (37.5 vs. 42.8 AST nodes). Crucially, on the unseen ECG domain, the adaptive neural router autonomously locked onto the Gaussian basis function with a probability of 0.69 (95.8% dataset hit rate), providing the first empirical validation that a neuro-symbolic system can autonomously recover the correspondence between physical mechanisms and syntactic operators within an appropriate topological representation space, without human domain annotation. Stress tests further demonstrated DiTTo's superior robustness to phase shifts ($\sigma \leq 50$ ms) and background noise (SNR ≥ -5 dB) compared to established baselines.

The "Deterministic Signal Processing $\rightarrow$ Neural Perception Routing $\rightarrow$ Discrete Symbolic Search" architecture pioneered by DiTTo establishes a viable pathway from black-box deep models to white-box interpretable knowledge discovery for safety-critical Cyber-Physical Systems. For applications ranging from rotating machinery predictive maintenance to power system transient diagnostics and biomedical signal comprehension, DiTTo extracts parsimonious analytical expressions that expose the underlying physical dynamics, delivering closed-form scientific laws that are directly auditable, verifiable, and generalizable by human experts.

References

[1] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge, MA: MIT Press, 1992.

[2] B. K. Petersen, M. Landajuela, T. N. Mundhenk, C. P. Santiago, S. K. Kim, and J. T. Kim, "Deep symbolic regression: Recovering mathematical expressions from data via risk-seeking policy gradients," in *International Conference on Learning Representations (ICLR)*, 2021. [Online]. Available: https://openreview.net/forum?id=m5Qsh0kBQG

[3] M. Cranmer, "Interpretable machine learning for science with PySR and SymbolicRegression.jl," *arXiv preprint arXiv:2305.01582*, 2023.

[4] D. Li and J. Zhang, "Purely discrete symbolic regression: EML signal decomposition via MCTS and basin-hopping," engrXiv preprint 6974, 2026. [Online]. Available: https://doi.org/10.31224/6974

[5] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 1, pp. 43–49, 1978.

[6] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[7] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[8] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform manifold approximation and projection for dimension reduction," *arXiv preprint arXiv:1802.03426*, 2018.

[9] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational Physics*, vol. 378, pp. 686–707, 2019.

[10] D. Gabor, "Theory of communication," *Journal of the Institution of Electrical Engineers*, vol. 93, no. 26, pp. 429–457, 1946.

[11] C. D. Rosin, "Nested rollout policy adaptation for monte carlo tree search," in *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, IJCAI 2011*. AAAI Press, 2011, pp. 649–654.

[12] L. Kocsis and C. Szepesvári, "Bandit based monte-carlo planning," in *Machine Learning: ECML 2006*, ser. Lecture Notes in Computer Science, J. Fürnkranz, T. Scheffer, and M. Spiliopoulou, Eds., vol. 4212. Berlin, Heidelberg: Springer, 2006, pp. 282–293.

[13] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization," *ACM Transactions on Mathematical Software*, vol. 23, no. 4, pp. 550–560, 1997.

[14] K. A. Loparo, "Case western reserve university bearing data center," https://engineering.case.edu/bearingdatacenter, 2012.

[15] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the case western reserve university data: A benchmark study," *Mechanical Systems and Signal Processing*, vol. 64–65, pp. 100–131, 2015.

[16] H. A. Dau, A. Bagnall, K. Kamgar, C.-C. M. Yeh, Y. Zhu, S. Gharghabi, C. A. Ratanamahatana, and E. Keogh, "The UCR time series archive," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 6, pp. 1293–1305, 2019.

[17] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 2001.

[18] B. Burlacu, G. Kronberger, and M. Kommenda, "Operon C++: An efficient genetic programming framework for symbolic regression," in *Proceedings of the 2020 Genetic and Evolutionary Computation Conference Companion*, ser. GECCO '20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 1562–1570.

[19] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[20] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *Proceedings of the National Academy of Sciences*, vol. 113, no. 15, pp. 3932–3937, 2016.

[21] W. La Cava, P. Orzechowski, B. Burlacu, F. O. de França, M. Virgolin, Y. Jin, M. Kommenda, and J. H. Moore, "Contemporary symbolic regression methods and their relative performance," in *Advances in Neural Information Processing Systems (NeurIPS) Datasets and Benchmarks Track*, 2021.

[22] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945.

[23] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian Journal of Statistics*, vol. 6, no. 2, pp. 65–70, 1979.

[24] B. P. Abbott *et al.*, "Observation of gravitational waves from a binary black hole merger," *Physical Review Letters*, vol. 116, no. 6, p. 061102, 2016.

[25] B. P. Abbott *et al.*, "GW170817: Observation of gravitational waves from a binary neutron star inspiral," *Physical Review Letters*, vol. 119, no. 16, p. 161101, 2017.