Supplementary Material

1 BENCHMARKS

The SR and digital circuit benchmarks used in this paper are provided here.

Table S1. Synthetic symbolic regression benchmark candidates. Benchmarks are in order by # features.

Dataset	# Features	# Instances
Keijzer-4	1	402
Keijzer-9	1	1102
Keijzer-10	2	10301
Keijzer-14	2	3741
Nguyen-9	2	1020
Nguyen-10	2	1020
Keijzer-5	3	11000
Vladislavleva-5	3	3000
Korns-11	5	20000
Korns-12	5	20000

Table S2. Real-world symbolic regression benchmark candidates. Benchmarks are in order by # features.

Dataset	Short name	# Features	# Instances
Airfoil Self-Noise	`airfoil`	5	1503
Energy Efficiency - Heating	`heating`	8	768
Energy Efficiency - Cooling	`cooling`	8	768
Concrete Strength	`concrete`	8	1030
Wine Quality - Red Wine	`redwine`	11	1599
Wine Quality - White Wine	`whitewine`	11	4898
Boston Housing	`housing`	13	506
Pollution	`pollution`	15	60
Dow Chemical	`dowchem`	57	1066
Communities and Crime	`crime`	127	1994

Table S3. Digital circuit benchmark candidates. Circuits are in alphabetical order.

Circuit	# Input	# Output	# Test cases
5-bit Comparator	10	3	1,024
5-bit Parity	5	1	32
11-bit Multiplexer	11	1	2,048
ALU	12	5	4,096

2 EVOLUTIONARY PARAMETERS

The evolutionary parameters used in this study are provided in Table S4, along with their corresponding values.

Table S4. Evolutionary parameters used in experimentation.

Parameter	Value
# Runs	30
Total Generations	50
Population Size	250
Selection Type	Tournament
Crossover Type	Effective
Crossover Rate	0.9
Mutation Rate	0.01
Initialisation Type	Sensible

3 TRAINING AND TESTING DATA SIZE

Table S5 presents the sizes of both the training and testing datasets employed across all benchmarks in this study.

Table S5. Training data and Testing data size used in different experiments of 24 benchmarks.

	Benchmarks	Training data size							Testing data size
		Baseline	DBS (70%)	DBS (65%)	DBS (60%)	DBS (55%)	DBS (50%)	DBS (45%)	
Synthetic SR	Keijzer-4	282	199	185	170	157	141	128	120
	Keijzer-9	772	540	503	464	426	386	349	330
	Keijzer-10	7211	5049	4689	4328	3968	3606	3247	3090
	Keijzer-14	2619	1835	1704	1572	1443	1310	1180	1122
	Nguyen-9	714	501	466	430	395	357	322	306
	Nguyen-10	714	501	465	428	394	358	322	306
	Keijzer-5	7700	5391	5007	4621	4237	3851	3466	3300
	Vladislavleva-5	2100	1471	1366	1261	1157	1051	946	900
	Korns-11	14000	9802	9103	8401	7703	7002	6304	6000
	Korns-12	14000	9803	9104	8401	7703	7001	6304	6000
Real-world SR	airfoil	1053	739	687	634	581	527	476	450
	heating	538	379	352	326	298	269	245	230
	cooling	538	379	352	326	298	269	245	230
	concrete	721	506	472	435	399	362	328	309
	redwine	1120	786	730	674	619	561	505	479
	whitewine	3429	2402	2230	2058	1887	1715	1544	1469
	housing	355	250	233	215	198	178	161	151
	pollution	42	32	30	26	25	22	21	18
	dowchem	747	524	487	450	412	375	338	319
	crime	1396	979	912	842	771	700	632	598
Circuit	5-bit Comparator	1024	720	672	616	568	512	461	1024
	5-bit parity	32	24	22	20	18	16	16	32
	11-bit Multiplexer	2048	1440	1344	1232	1136	1024	928	2048
	ALU	4096	2877	2669	2461	2253	2045	1856	4096

4 COMPATIBILITY WITH GENETIC PROGRAMMING

To investigate the usability of DBS beyond its typical pairing with GE, we opted to assess it with GP. We selected two benchmarks (one from synthetic SR and another from real-world SR), each showcasing varying levels of complexity. We conducted a total of 420 runs, comprising 30 independent runs for each training dataset across two problems. A population size of 250, generation of 50, crossover rate of 0.9, mutation rate of 0.01, and a maximum tree depth of 50 were used.

The results in Figure S1 show the test score of baseline and different DBS selection budgets.

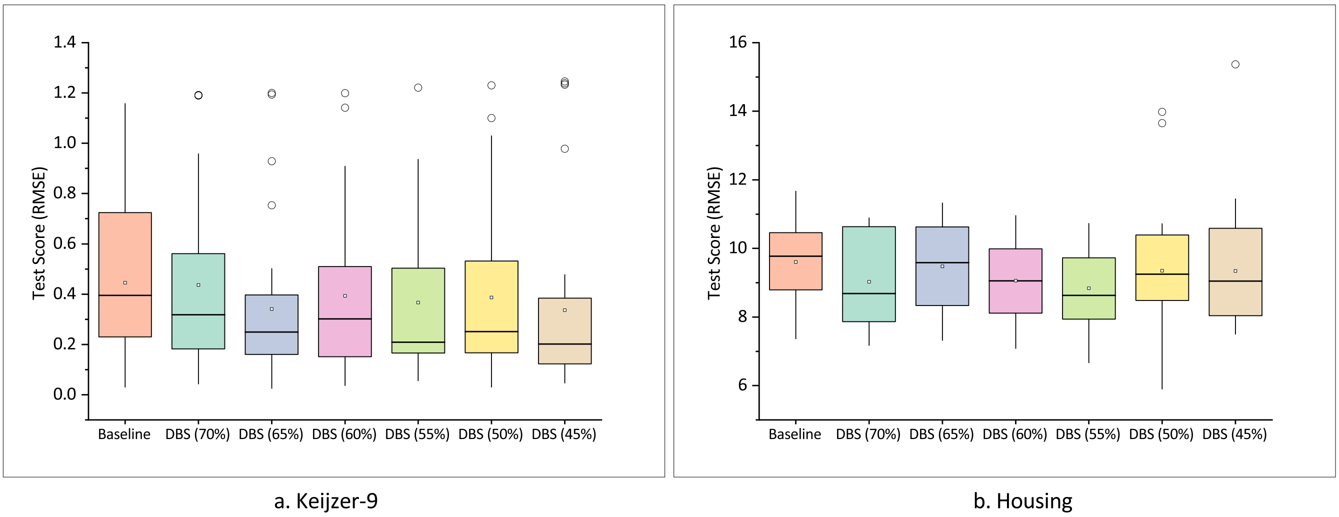

Figure S1: Mean effective individual size of best solutions obtained (across 30 independent runs) on two SR data sets using GP.

We further perform statistical tests on the obtained results with the same approach used earlier. The result in Table S6 shows that DBS performs similar or better to the baseline on different selection budgets. This underscores the potential for DBS to seamlessly integrate with diverse EAs.

Table S6. Statistical test results of DBS evaluated on two SR benchmarks using GP considering a significance level of $\alpha = 0.05$. Values shown are the mean best test scores across 30 independent runs. The symbols +, =, - indicate whether the corresponding results for DBS are significantly better, not significantly different, or worse than baseline, respectively.

Benchmarks	Baseline	DBS (70%)	DBS (65%)	DBS (60%)	DBS (55%)	DBS (50%)	DBS (45%)
Keijzer-9	0.4456	0.43667=	0.3413+	0.393=	0.366=	0.387=	0.3365+
housing	9.6025	9.025=	9.478=	9.058+	8.839+	9.351=	9.3467=

5 COMPARATIVE ANALYSIS

We selected the best test score on a given DBS budget and performed a comparative analysis with RegCNN and RegENN. These algorithms, extensions of CNN and ENN, respectively, are specifically designed for regression tasks.

These approaches utilize an adaptive threshold to compare an instance's output attribute value with its neighborhood, classifying it based on similarity or dissimilarity with nearby instances. The adaptive threshold adjusts to the standard deviation of neighboring instances. The parameter α determines the algorithm's performance: a larger α results in stricter noise removal, while a smaller α allows instances to be retained if their output values closely match the predicted values.

The authors in (Kordos and Blachnik, 2012) used 50 different values of α. However, in this experiment, we use the optimal values of parameters recommended by the authors: setting k to 9 for both RegCNN and RegENN, while α was set to 0.5 for RegCNN and 5 for RegENN.

We performed 30 independent runs on each training data and tested against the same set of testing data used in previous experiments (see Table S5). The results are shown in Figures S2 and S3.

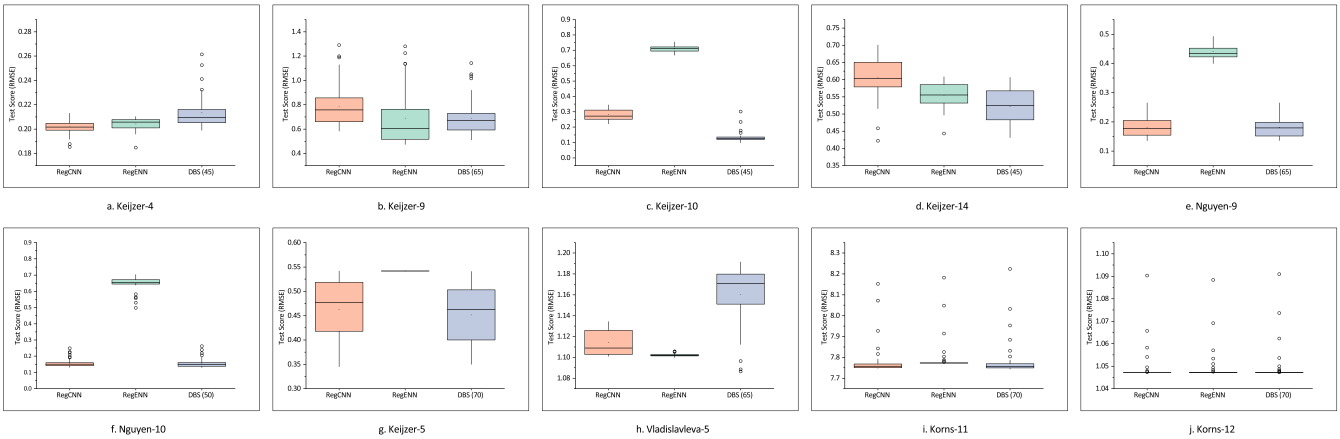

Figure S2: Mean test score of best solutions across 30 independent runs obtained on synthetic SR problems. The best test score of DBS is considered to be compared against RegCNN and RegENN.

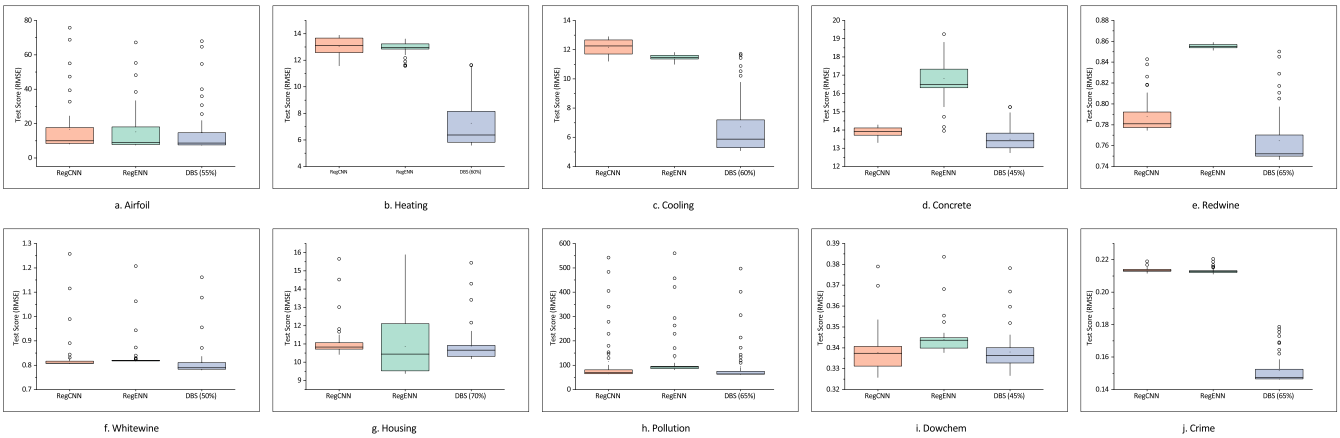

Figure S3: Mean test score of best solutions across 30 independent runs obtained on real-world SR problems. The best test score of DBS is considered to be compared against RegCNN and RegENN.

We further perform statistical tests using the methodology and hypotheses used in the previous results analysis. The findings are shown in Tables S7 and S8. As the best results of DBS are used, we denote it as DBS*. The DBS* selection budgets are indicated in Figures S2 and S3.

Table S7. Results of the Wilcoxon test for SR **synthetic benchmarks** considering a significance level of $\alpha = 0.05$. Values shown are the mean best test scores across 30 independent runs. The symbols +, =, - indicate whether the corresponding results for DBS are significantly better, not significantly different, or worse than the RegCNN and RegENN.

Benchmarks	RegCNN	RegENN	DBS*		
	score	score	score	Vs. RegCNN	Vs. RegENN
Keijzer-4	0.201769	0.204216	0.2135	$-$	$-$
Keijzer-9	0.782613	0.690005	0.6893	$+$	$=$
Keijzer-10	0.281525	0.709491	0.133	$+$	$+$
Keijzer-14	0.607142	0.555018	0.522	$+$	$+$
Nguyen-9	0.180262	0.439797	0.1812	$=$	$+$
Nguyen-10	0.157474	0.648281	0.1562	$=$	$+$
Keijzer-5	0.463181	0.541545	0.4515	$=$	$+$
Vladislavleva-5	1.11412	1.102172	1.1599	$-$	$-$
Korns-11	7.778168	7.791103	7.7806	$=$	$=$
Korns-12	1.048843	1.048922	1.0491	$+$	$+$

Table S8. Results of the Wilcoxon test for SR **real-world benchmarks** considering a significance level of $\alpha = 0.05$. Values shown are the mean best test scores across 30 independent runs. The symbols +, =, - indicate whether the corresponding results for DBS are significantly better, not significantly different, or worse than the RegCNN and RegENN.

Benchmarks	RegCNN	RegENN	DBS*		
	score	score	score	Vs. RegCNN	Vs. RegENN
airfoil	16.7552	15.0828	14.8581	$+$	$=$
heating	13.0144	12.9269	7.2609	$+$	$+$
cooling	12.1646	11.4581	6.7121	$+$	$+$
concrete	13.8908	16.8237	13.497	$+$	$+$
redwine	0.7877	0.8553	0.7647	$+$	$+$
whitewine	0.8315	0.8357	0.8126	$+$	$+$
housing	11.0758	10.8613	10.8883	$+$	$=$
pollution	113.8127	126.0512	95.3476	$+$	$+$
dowchem	0.3377	0.3441	0.338	$=$	$+$
crime	0.2136	0.2131	0.1519	$+$	$+$